Development and characterization of two new tools for plant genetic engineering: A CRISPR/Cas12a-based mutagenesis system and a PhiC31-based gene switch

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Joan Miquel Bernabé Orts

Advisor
Diego Orzáez Calatayud

Valencia, October 2019

El Dr. Diego Orzáez Calatayud, Científico Titular del Consejo Superior de Investigaciones Científicas, perteneciente al Instituto de Biología Molecular y Celular de Plantas (IBMCP, UPV-CSIC) de Valencia, CERTIFICA que Joan Bernabé Orts, ha realizado bajo su dirección en el Instituto de Biología Molecular y Celular de Plantas, el trabajo titulado "Development and characterization of two new tools for plant genetic engineering: A CRISPR/Cas12a-based mutagenesis system and a PhiC31-based gene switch", y que autoriza su presentación para optar al grado de Doctor en Biotecnología. Y para que así conste, firma el presente certificado en

Valencia a 3 de octubre de 2019.

Dr. Diego Orzáez Calatayud

Agradecimientos

El doctorado no es solo el título académico más alto que una universidad otorga, es también una experiencia de enriquecimiento personal y profesional que cambia la vida del doctorando. Un viaje lleno de emociones y de personas que lo acompañan. Es por ello que quiero dedicar unas líneas de reconocimiento a todas aquellas personas que han formado parte de esta experiencia.

En primer lugar, a mi director Diego Orzaez que me brindó la oportunidad de realizar la tesis bajo su dirección. A tú lado he aprendido a ser un profesional de la ciencia (aunque no te guste este término). Gracias por tener siempre un momento para mí. En segundo lugar, a Toni Granell por transmitirme tu dilatada experiencia y por los consejos e interés que siempre has tenido hacia mi persona.

A mis compañeros del laboratorio 2.10. A Marta Vazquez por instruirme en los momentos iniciales de mi tesis que coincidieron con el final de la suya. Todo doctorando que este acabando sabrá que en esos momentos finales todo supone un esfuerzo extra. Yo en ese momento no era consciente de ello, pero ahora si lo soy, así que muchas gracias de corazón. Por esta misma razón también tengo que agradecer a Estefania Huet los momentos la ayuda prestada. A Julve por llenar de humor el laboratorio. Aunque muchas veces ni siquiera entendiese tus chistes, tu simple risa era contagiosa. A Sara Selma, apareciste más adelante y te ganaste a pulso un lugar en el laboratorio. Gracias Sara por entender mejor que nadie mi humor absurdo y ser conocedora de todas las coletillas de los Simpsom. A Paloma Juárez, por narrarme sus experiencias de postdoc en países más fríos. A Elena Moreno, por la inestimable ayuda prestada en mis experimentos finales con las hairy roots y por las luces LED tan molonas que me hizo tu padre (gracias a él también!). A las que siempre han estado y estarán ahí: a Asun Fernandez y a Silvia

Presa. Asun eres una fuente inestimable de sabiduría, la visnu del 2.10 que tan pronto hace una PCR como te mete caña con los *deliverables*. Silvia, espero que sigas repartiendo caña merecida a todo el mundo ya que sin ti la entropía del laboratorio llegaría a niveles insospechados. Muchas gracias por invitarnos a Flor y a mi a todas tus aventuras en la montaña. Gracias también a Victor Garcia, el bioinformático que tanta falta hacía por el labo y que tanto me ayudó. Lástima no haber coincidido más tiempo.

Dentro del 2.10 tú Flor mereces un punto y a parte para dedicarte unas líneas especiales que ni por asomo reflejaran todo el amor que siento por ti. Fuiste la última persona que conocí en mis primeras andanzas por el IBMCP (estabas en un cursillo de NGS creo recordar) y la que más expectación creó ya que todo el mundo hablaba de ti. Y es que lo bueno se hace esperar. Tus risos danzantes y esponjosos me cautivaron y desde entonces no me he separado de ti. Eres mi mayor descubrimiento, la mujer de mi vida, mi mejora amiga, la estrella polar que me guía en las noches más oscuras. Gracias por todo tu apoyo y amor incondicional, en especial durante esta fase final de escritura.

También Alfredo (A.K. tachino) merece unas líneas a parte, nuestra relación no se entiende sin tu amistad. Gracias por legarme tu trabajo al que debo un capitulo de esta tesis, por ser tan buen anfitrión en EEUU (aunque nos hicieses dormir en un sótano húmedo jaja) y por hacer un hueco en tu apretada agenda para vernos cada vez que vienes a España con los días contados. No importa lo lejos que estemos, nuestra amistad prevalecerá.

A todos los estudiantes que han pasado por el 2.10: Ivan, Alicia, Victor, Javi, Agos, Maria, Borja…aparecisteis casi todos "en trompa" e hicisteis que Sara y yo nos pluriemplearamos. Gracias por todo el buen rollo, humor e ilusión que transmitís. He aprendido tanto o más de vosotros como yo os pueda haber transmitido. Sin duda sois la parte más divertida y enriquecedora del doctorado.

Al “Little Italy” del 2.10 conformado por Francesco, Silvia, Camilla y Gian Franco entre otros. ¡Tomar cafes en los que el italiano es la lengua sin salir del IBMCP no tiene precio!

A los compañeros del 2.07 por incluirme en vuestro grupo durante los momentos iniciales de la tesis. Con especial cariño a Noel y Marco con los que descubrí Doñana y la Salamanca universitaria. ¡Gracias por escucharme y aconsejarme siempre tan bien chicos! Agradecer también a Eugenio su ayuda desinteresada y su tiempo, siempre dispuesto a enseñar. También a Jorge, David y Asier con los que tantos momentos divertidos de cafe hemos compartido.

Cuando llegué cuatro labos ya se estaba desmantelando. Pero su espíritu siempre ha permanecido vivo en personas como Maria Ángeles (nuestra abuela adoptiva), MD y Cris Urbez. Gracias por vuestras palabras de animo y apoyo y por vuestra empatia. Llenáis de calidez los largos pasillos del IBMCP.

Por supuesto a los servicios del IBMCP. A Ana y Eugenio de secuenciación, con el que tanto tiempo pasé al final de la tesis mientras utilizaba el LAS3000. Agradecer también a Marisol toda la ayuda prestada y conocimientos transmitidos acerca del uso del confocal.

Finalmente y no por ello menos importante agradecer a la familia su infinito apoyo. A la familia política Jorge, Andrea, Belen y Agus. Os escogería sin dudar. Gracias por todo vuestro apoyo y por todos los asados habidos y por haber. Gràcies a Miquel per la portada de la tesis y a la tia Anna, sempre disposats a fer de personal shopping. A la meua germana Roser per estar sempre ahí, eres una lleona i no hi ha obstacle que et supere. Papa, Mama, iaios…no seria res del que sóc avui si no fora per tots els valors i educació que m’heu brindat a força del vostre sacrifici. A la mama per eixir abans del treball i batallar amb mi per ensenyarme mates, física, química…Al iaio Pepe per ensenyarme la bellessa que s’oculta darrere d’un solc ben fet.

Al iaio Pepe,

Summary

Plant breeding aims to provide plants with improved traits or novel features that could help to overcome sustainability goals. To this end, plant biotechnology needs to incorporate new genetic engineering tools that combine increased precision with higher breeding power. The recently discovered genome editing tools based on CRISPR/Cas9 technology have opened the way to modify plant´s genomes with unprecedented precision. On the other hand, new synthetic biology approaches based on modularity and standardization of genetic elements have enabled the construction of increasingly complex and refined genetic devices applied to plant breeding. With the ultimate goal of expanding the toolbox of plant breeding techniques, this thesis describes the development and adaptation to plant systems of two new breeding tools: a site-specific nuclease (SSNs), and a modular gene switch for the regulation of transgene expression.

In a first part, this thesis describes the adoption of the SSN CRISPR/Cas12a for plant expression and compares the efficiency of *Acidaminococcus* (As) and *Lachnospiraceae* (Lb) Cas12a variants with the previously described *Streptococcus pyogens* Cas9 (SpCas9) in eight *Nicotiana benthamiana* loci using transient expression experiments. LbCas12a showed highest average mutagenesis activity in the loci assayed. This activity was also confirmed in stable genome editing experiments performed in three different model plants, namely *N. benthamiana*, *Solanum lycopersicum* and *Arabidopsis thaliana*. For the latter, off-target effects in Cas12a-free segregating lines were discarded at genomic level by deep sequencing. Collectively, the results show that LbCas12a is a viable alternative to SpCas9 for plant genome engineering.

In a second part, this work describes the engineering of a new reversible genetic switch aimed at controlling gene expression in plants with higher precision than traditional

inducible systems. This switch, based on the bacteriophage PhiC31 recombination system, was built as a modular device made of standard DNA parts and designed to control the transcriptional state (on or off) of two genes of interest by alternative inversion of a central DNA regulatory element. The state of the switch can be externally and reversibly operated by the action of the recombination actuators and its kinetics, memory, and reversibility were extensively characterized in *N. benthamiana* using both transient expression and stable transgenics.

Altogether, this thesis shows the design and functional characterization of refined tools for genome engineering and synthetic biology in plants that now has been expanded with the CRISPR/Cas12a gene editing system and the phage PhiC31-based toggle switch.

Resumen

La mejora genética vegetal tiene como objetivo la obtención de plantas con rasgos mejorados o características novedosas que podrían ayudar a superar los objetivos de sostenibilidad. Para este fin, la biotecnología vegetal necesita incorporar nuevas herramientas de ingeniería genética que combinen una mayor precisión con una mayor capacidad de mejora. Las herramientas de edición genética recientemente descubiertas basadas en la tecnología CRISPR/Cas9 han abierto el camino para modificar los genomas de las plantas con una precisión sin precedentes. Por otro lado, los nuevos enfoques de biología sintética basados en la modularidad y la estandarización de los elementos genéticos han permitido la construcción de dispositivos genéticos cada vez más complejos y refinados aplicados a la mejora genética vegetal. Con el objetivo final de expandir la caja de herramientas biotecnológicas para la mejora vegetal, esta tesis describe el desarrollo y la adaptación de dos nuevas herramientas: una nueva endonucleasa específica de sitio (SSN) y un interruptor genético modular para la regulación de la expresión transgénica.

En una primera parte, esta tesis describe la adaptación de CRISPR/Cas12a para la expresión en plantas y compara la eficiencia de las variantes de *Acidaminococcus* (As) y *Lachnospiraceae* (Lb) Cas12a con *Streptococcus pyogens* Cas9 (SpCas9) descritos anteriormente en ocho loci de *Nicotiana benthamiana* usando expresión transitoria. LbCas12a mostró la actividad de mutagénesis promedio más alta en los loci analizados. Esta actividad también se confirmó en experimentos de transformación estable realizados en tres plantas modelo diferentes, a saber, *N. benthamiana, Solanum lycopersicum* y *Arabidopsis thaliana*. Para este último, los efectos mutagénicos colaterales fueron analizados en líneas segregantes sin la endonucleasa Cas12a, mediante secuenciación del genoma descartándose efectos indiscriminados. En conjunto, los

resultados muestran que LbCas12a es una alternativa viable a SpCas9 para la edición genética en plantas.

En una segunda parte, este trabajo describe un interruptor genético reversible destinado a controlar la expresión génica en plantas con mayor precisión que los sistemas inducibles tradicionales. Este interruptor, basado en el sistema de recombinación del fago PhiC31, fue construido como un dispositivo modular hecho de partes de ADN estándar y diseñado para controlar el estado transcripcional (encendido o apagado) de dos genes de interés mediante la inversión alternativa de un elemento regulador central de ADN. El estado del interruptor puede ser operado externa y reversiblemente por la acción de los actuadores de recombinación y su cinética, memoria y reversibilidad fueron ampliamente caracterizados en experimentos de transformación transitoria y estable en *N. benthamiana*.

En conjunto, esta tesis muestra el diseño y la caracterización funcional de herramientas para la ingeniería del genómica y biología sintética de plantas que ahora ha sido completada con el sistema de edición genética CRISPR/Cas12a y un interruptor genético reversible y biestable basado en el sistema de recombinación del fago PhiC31.

Resum

La millora genètica vegetal té com a objectiu l'obtenció de plantes amb trets millorats o característiques noves que podrien ajudar a superar els objectius de sostenibilitat. Amb aquesta finalitat, la biotecnologia vegetal necessita incorporar noves eines d'enginyeria genètica que combinen una major precisió amb una major capacitat de millora. Les eines d'edició genètica recentment descobertes basades en la tecnologia CRISPR/Cas9 han obert el camí per modificar els genomes de les plantes amb una precisió sense precedents. D'altra banda, els nous enfocaments de biologia sintètica basats en la modularitat i l'estandardització dels elements genètics han permès la construcció de dispositius genètics cada vegada més complexos i sofisticats aplicats a la millora genètica vegetal. Amb l'objectiu final d'expandir la caixa d'eines biotecnològiques per a la millora vegetal, aquesta tesi descriu el desenvolupament i l'adaptació de dues noves eines: una nova endonucleasa específica de lloc (SSN) i un interruptor genètic modular per a la regulació de l'expressió transgènica .

En una primera part, aquesta tesi descriu l'adaptació de CRISPR/Cas12a per a l'expressió en plantes i compara l'eficiència de les variants de *Acidaminococcus* (As) i *Lachnospiraceae* (Lb) Cas12a amb la ben establida *Streptococcus pyogens* Cas9 (SpCas9), en vuit loci de *Nicotiana benthamiana* usant expressió transitòria. LbCas12a va mostrar l'activitat de mutagènesi mitjana més alta en els loci analitzats. Aquesta activitat també es va confirmar en experiments de transformació estable realitzats en tres plantes model diferents, a saber, *N. benthamiana*, *Solanum lycopersicum* i *Arabidopsis thaliana*. Per a aquest últim, els efectes mutagènics col·laterals van ser analitzats en línies segregants sense l'endonucleasa Cas12a, mitjançant seqüenciació completa del genoma i descartant efectes indiscriminats. En conjunt, els resultats mostren que LbCas12a és una alternativa viable a SpCas9 per a l'edició genètica en plantes.

En una segona part, aquest treball descriu un interruptor genètic reversible destinat a controlar l'expressió gènica en plantes amb major precisió que els sistemes induïbles tradicionals. Aquest interruptor, basat en el sistema de recombinació del bacteriòfag PhiC31, va ser construït com un dispositiu modular fet de parts d'ADN estàndard i dissenyat per controlar l'estat transcripcional (encès o apagat) de dos gens d'interès mitjançant la inversió alternativa d'un element regulador central d'ADN. L'estat de l'interruptor pot ser operat externa i reversiblement per acció dels actuadors de recombinació i la seva cinètica, memòria i reversibilitat van ser àmpliament caracteritzats en experiments de transformació transitòria i estable en *N. benthamiana*.

En conjunt, aquesta tesi mostra el disseny i la caracterització funcional d'eines per a l'enginyeria del genòmica i biologia sintètica de plantes que ara ha sigut completat amb el sistema d'edició genètica CRISPR/Cas12a i un interruptor genètic biestable i reversible basat en el sistema de recombinació del bacteriòfag PhiC31.

Table of contents

CHAPTER I 1

1. A brief overview of plant breeding techniques 1

2. Genome engineering and site-specific nucleases in plants 3

2.1. Genome engineering promoted by double-strand breaks 3

2.2. Democratizing genome engineering with CRISPR/Cas systems 5

2.2.1. Meganucleases 6

2.2.2. Zinc finger nucleases 6

2.2.3. Transcription activator-like effectors 6

2.2.4. CRISPR/Cas systems 7

2.3. Exploiting the naturally occurring diversity of CRISPR systems 7

3. Expanding the horizons of plant breeding through synthetic biology 11

3.1. A brief introduction to synthetic biology and plant synthetic biology 12

3.1.1. Synthetic biology facilitated by modular cloning 13

3.1.2. Synthetic biology hits the plant kingdom 15

3.2. Site-specific recombination 17

3.3. Uses of site-specific recombination in plant biotechnology 20

3.3.1. Transgene excision 20

3.3.2. Gene stacking 21

3.4. Gene switches 22

3.5. Memory switches in plant systems 23

4. References 26

OBJECTIVES 37

CHAPTER II 41

1. Introduction 43

2. Results 45

2.1. A GB-assisted cloning strategy for plant Cas12a constructs 45

2.2. Transient expression of CGEM provides efficient targeted mutagenesis in *N. benthamiana* leaves 48

2.3. Modifications in the crRNA DR loop affect RGEN activity but cannot compensate low efficiency editing 50

2.4. Cas12a as a genome editing tool in *N. benthamiana*, tomato and Arabidopsis 51

2.5. A meta-analysis of Cas12a mutagenesis profile shows a high frequency of small-medium size deletions 57

2.6. Whole genome sequencing of LbCas12a-mutated *A. thaliana* plants revealed no appreciable off-target effects 58

3. Discussion 61

4. Experimental procedures 66

4.1. GBparts construction 66

4.2. Guide RNA assembly on level 1 66

4.3. Cloning in α and Ω-level destination vectors 67

4.4. *Nicotiana benthamiana* transient assays 67

4.5. *Nicotiana benthamiana* stable transformation 68

4.6. *Solanum lycopersicum* stable transformation 68

4.7. *Arabidopsis thaliana* stable transformation 68

4.8. Mutagenesis detection of on-target sites 68

4.9. Mutagenesis detection of off-target sites 69

5. References 71

6. Supplementary information 78

7. Supplementary tables 89

CHAPTER III 97

1. Introduction 99

2. Results 103

2.1. Design of a modular reversible genetic switch for plant systems 103

2.2. Characterization of the register modules in stably transformed *N. benthamiana* plants 107

2.3. Stable and reversible memory storage over a full SET/RESET cycle in whole plants 109

2.4. Chemical induction of PhiC31 controls SET operation in *N. benthamiana* hairy roots 112

3. Discussion 115
4. Experimental procedures 120
4.1. Construction and assembly of the GoldenBraid phytobricks 120
4.2. Time-dependent characterization of reporter expression in transiently expressed and integrated register modules 121
4.3. Confocal laser microscopy 122
4.4. Firefly and renilla luciferase luminescence quantification 122
4.5. Generation of *N. benthamiana* transgenic plants 122
4.6. Generation of *N. benthamiana* transgenic hairy roots 123
4.7. Estradiol induction experiments 124
4.8. YFP quantification 124
4.9. Genomic DNA extraction and PCR amplification 125
5. References 126
6. Supplementary figures 130
7. Supplementary tables 136
CHAPTER IV 141
1. Uses of CRISPR/Cas12a in precision plant breeding and beyond 143
1.1. Targeting non-coding regulatory regions with Cas12a 143
1.2. New multiplexing capacities with Cas12a 145
1.3. Expanding the toolbox of transcriptional regulators 147
2. Applications of the phage PhiC31-based toggle switch 149
2.1. Control of the production of toxic or detrimental compounds. 149
2.2. Regulation of transgene expression in field crops 150
2.3. A DNA-based memory device for recording cellular events 151
3. Combined uses of toggle switches and Cas12/Cas9-based transcriptional regulators 153
4. References 157
CONCLUSIONS 161

CHAPTER I

General Introduction

1. A brief overview of plant breeding techniques

For more than 10,000 years humans have exploited the naturally occurring plant genetic variability for the artificial selection of desirable traits. During the major part of that period, humans carried out unconscious genetic screenings of the genetic diversity generated by spontaneous mutations and DNA recombination (Fig. 1A). This **selective breeding** led to the modern domesticated crops with increased yield but also impoverished genetic variability.

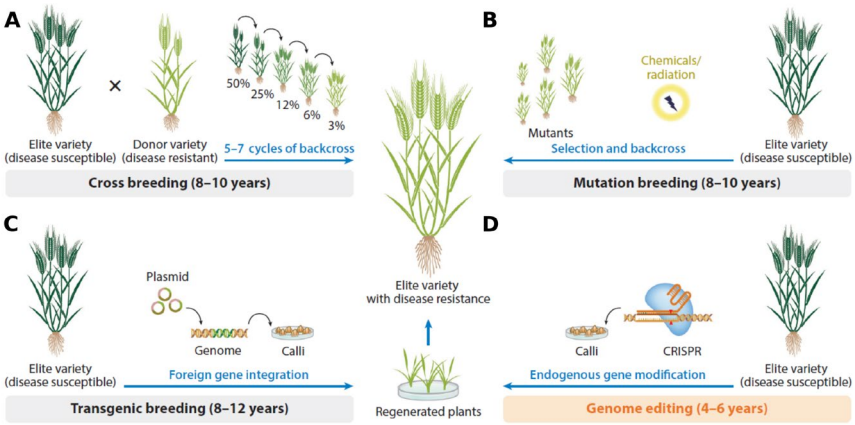


Figure 1: an overview of different breeding methods currently used in modern agriculture. A) Introgression of a certain trait (e.g. disease resistance) into an elite variety involves an initial crossing with a donor variety and several backcrosses of the progeny and the recipient line to incorporate the desired trait while keeping the genetic background of the elite variety. **B)** Mutation breeding uses physic and chemical mutagenic agents to induce genetic variability. Seeds of an elite variety are mutated and the progeny of them screened for a mutated allele of interest. **C)** Transgenic breeding uses various gene delivery methods (*Agrobacterium*-mediated transformation, callus bombardment, protoplast transformation) to transfer distant (*trans*-genesis) or related (*cis*-genesis) genetic constructs to the plant DNA chassis. **D)** Precision mutagenesis facilitated by site-specific nucleases like CRISPR/Cas systems allows the targeted modification of precise traits accelerating the breeding process. Various transformation protocols are used to deliver the components of the CRISPR/Cas systems including plasmid transformation or DNA-free methods by delivery of the Cas endonuclease loaded with the gRNA. Adapted from [9].

In the twentieth century, the notion that genetic information resided in DNA, impelled researchers to use mutagenic agents such as radiation or chemical agents to accelerate the breeding process by inducing genetic variability in the domesticated crops. That was the beginning of **mutation breeding** [1, 2] which resulted in wheat varieties with significantly improved yields that were key to the Green Revolution in the 1970s (Fig. 1B). However, the most important limitation of mutagenic breeding was its stochastic nature. Whereas generating a large collection of mutants was straightforward, now the bottleneck was implementing powerful forward genetic screens to identify the candidate mutants. In part, this was solved by TILLING (targeting induced local lesions in the genome) [3], a refinement of mutation breeding that links mutagenesis to a DNA-analysis method that selects a candidate plant from thousands of mutagenized plants. Although this strategy speeded-up the screening process, it is still costly and labor-intensive and, in many cases, the obtention of specific alleles known to confer certain phenotypes is impossible.

In parallel, **transgenic breeding** (Fig. 1C) came with the development of the *Agrobacterium*-mediated plant transformation protocol in the 1980s [4]. This allowed the introduction of foreign (*trans*-genesis) or related genes (*cis*-genesis) into the plant genome. Now, a plant could be endowed with novel traits or functions impossible to be implemented through traditional breeding such as herbicide or pest resistance in the crop field [5, 6]. This also catapulted functional genetic studies in model organisms such is *A. thaliana*, by fostering forward and reverse genetic analysis of T-DNA insertional mutants [7] or fluorescent tagged proteins [8], for example. Eventually, this valuable information could be extrapolated to crop species for the selection of orthologs for targeted breeding. More recently, with the advance of the modular cloning technologies the single-gene transgenic approaches have advanced to more complex multigenic constructs facilitating the development of plant

synthetic biology as will be introduced later in this General Introduction.

Only recently, with the increasing information provided by functional genomics and the discovery of site-specific mutagenic agents (reviewed in section two of this Chapter I), selective breeding and mutation breeding are gradually moving from random genetic modifications to **precision plant breeding** (Fig. 1D) as we will discuss in detail in section two of this General Introduction. Precision plant breeding takes advantage of all the information gathered in model and crop plants to target specific genes involved in the breeding process through site-specific DNA modification.

2. Genome engineering and site-specific nucleases in plants

2.1. Genome engineering promoted by double-strand breaks

Genomes of eukaryotic organisms are composed of billions of DNA bases. The ability to change these DNA bases at precisely predetermined locations holds tremendous value not only for molecular biology but also for biotechnology and plant breeding. A strategy to achieve precise DNA modifications consists in the introduction of targeted double strand breaks (DSBs) at designated locations in the plant genome. DSBs are highly toxic lesions, which are repaired via non-homologous end joining (NHEJ) or homology-directed repair (HDR) [10, 11].

In somatic plant cells, the vast majority of DSBs are repaired by NHEJ mechanism which involves re-ligation of the free ends. Although NHEJ is often precise, small deletions or more rarely insertions can be introduced at the junction of the repaired chromosome (Fig. 2A left). If the sequence modification causes a

frameshift mutation or alters key amino acid residues of the gene product, a knockout can be created. Broken chromosome ends can also be joined to unrelated DNA molecules producing knock-in modifications (Fig. 2A, right). If two breaks are introduced simultaneously in the chromosome, targeted gene deletions or other rearrangements can result (Fig. 2B). Nevertheless, the intrinsic DSB repair machinery has the potential to remove breaks also in an error-free manner by HDR [12, 13]. In HDR, a repair template is used as a source of sequence homology that is copied to the broken chromosome to restore its integrity (Fig. 2C). This can be exploited by making use of extrachromosomal templates provided, for example, via *Agrobacterium* T-DNA, transfected plasmids, biolistic delivery or viral replicons. Hence, as the user can define the repair template, HDR offers an unprecedented ability to harness the plant's genome.

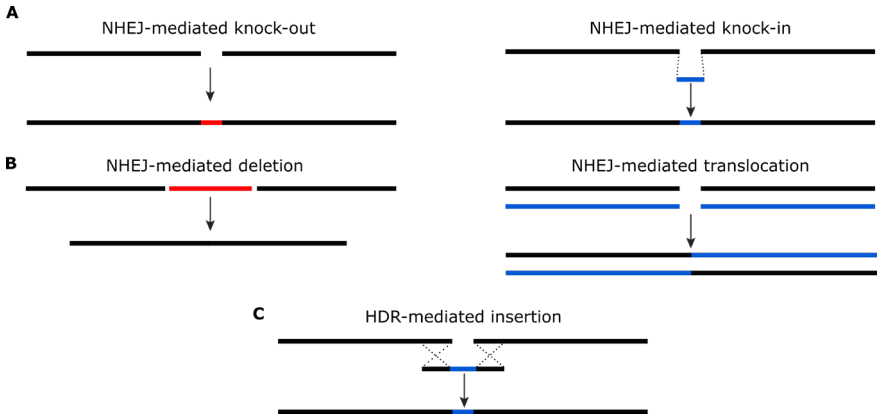


Figure 2: DSB repair mechanisms. A) After a DNA break, the chromosome can be repaired by NHEJ re-joining the ends and producing small indels which eventually knock-out the gene product (left); or, less frequently, by introducing unrelated small DNA fragments (right). **B)** When two breaks are introduced the NHEJ's output is different and can produce chromosomal deletions (left) or large inter-chromosomal DNA re-arrangements (right). **C)** HDR-mediated repair uses homology templates that lead to gene replacement or gene insertion.

The development of site-specific nucleases (SSNs) that are able to introduce DSB at specific locations on the plant genome has fostered the field of precision plant breeding. In the next sections we will present the state-of-the-art of these techniques.

2.2. Democratizing genome engineering with CRISPR/Cas systems

Introduction of DSBs at the desired locus relies on the use of SSNs. Currently, the toolkit of genome engineers contains four classes of SSNs: meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effectors (TALEN) and CRISPR/Cas systems (Fig. 3). These molecular tools display different features such as size or specificity, which determine their programmability and make them more or less suitable for certain purposes (i.e. viral delivery) [14]. The ideal editing tool should be easy-to-engineer and versatile enough to cover any desired genomic locus.

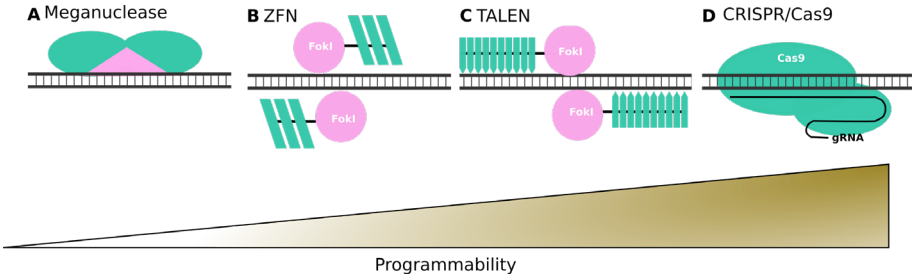


Figure 3: comparison between the different SSNs used in genome engineering. A) Meganucleases are restriction enzymes that recognize long stretches of DNA sequences. **B-C)** ZFNs and TALENs are protein domains that recognize 3 or 1 bp respectively, fused to the FokI endonuclease domain. Specificity is based on a protein-DNA interrogation which requires of engineering a new endonuclease for each target. Dimerization of FokI at the target site is required in order to cut the DNA. **D)** CRISPR/Cas9 system is RNA-guided endonuclease which specificity relies on an RNA-DNA interaction very easy to reprogram.

2.2.1. Meganucleases

Meganucleases (the endonucleases that recognize long stretches of DNA of 14-40 bp) were the first SSNs used in plants. They led to the discovery of fundamental information in the DNA reparation mechanisms in plants [12] and the demonstration that genes can be knocked out in a targeted manner [13]. However, the difficulty for engineering their large recognition sequence by protein redesign limited drastically the applicability of these SSNs [15, 16].

2.2.2. Zinc finger nucleases

The lack of meganucleases re-programmability was solved with zinc fingers, which are protein motifs that bind to DNA in a sequence-specific manner. Each zinc finger module binds a 3-bp DNA sequence [17]. Therefore, unlike meganucleases, multiple zinc finger domains can be arranged to create custom DNA-binding specificities. Nevertheless, zinc finger domains lack endonuclease activity *per se,* so they need to be fused to the Fok I endonuclease domain to create an effective zinc finger nuclease (ZFNs) [18]. In addition, Fok I require homodimerization to cleave the DNA at the target site. Therefore, two separate ZFNs that target two proximal sites need to be engineered, doubling the size of the construct and complicating the design. Commonly, ZFNs are composed of 3-6 zinc finger modules and target specifically 9-18 base long sequences [19].

2.2.3. Transcription activator-like effectors

The discovery of TALEs proteins from *Xanthomonas sp.* allowed a fine selection of the targeted sequence. These proteins contain DNA-binding modular domains which specifically recognize one single base instead of three as in zinc fingers [20, 21]. Typically TALEs are designed to target >30 bp sequences, which make them the most specific of all SSNs [22]. Again, this

DNA recognition modules must be fused to the Fok I endonuclease domain and peer-engineered to cut the DNA.

2.2.4. CRISPR/Cas systems

Despite the modularity of the DNA-binding domains of ZFNs and TALEs permits the design of user-customized target selection, this specificity is determined by a protein-DNA interaction that results in highly repetitive chimeric proteins, which need to be re-engineered for each target. This is an expensive and time-consuming process not affordable for all laboratories. That is why the most recent addition to the SSNs repertoire, the CRISPR/Cas systems, has democratized genome engineering by making the design and construction process of targeted-nucleases easy and accessible for laboratories with low expertise in molecular biology. CRISPR/Cas is a dual system composed of a Cas endonuclease and a guide RNA (gRNA) that directs the ribonucleoprotein complex to the complementary genome target [23, 24]. Therefore, unlike meganucleases, ZFNs or TALEs, the specificity of CRISPR/Cas depends on the RNA-DNA interaction conferred by the gRNA. Furthermore, the Cas protein has an intrinsic endonuclease activity, so no FokI fusion is needed. In consequence, directing CRISPR/Cas to a certain genome locus is extremely simple, requiring only the cloning of a 20 nt gRNA instead of complicated protein re-designing protocols. For all these reasons, CRISPR has revolutionized the field of genome engineering.

2.3. Exploiting the naturally occurring diversity of CRISPR systems

Owing to its low cost, easy execution and efficiency, CRISPR/Cas systems have been rapidly adopted as an editing platform for researchers of all fields, proving to be effective in virtually all the organisms tested. CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated) modules are adaptive immune systems that are present in

prokaryotic organisms and provide sequence-specific protection against foreign nucleic acid sequences [25-27]. The evolutionary arms race between prokaryotes and environmental mobile genetic elements yielded various CRISPR-type immune responses. These CRISPR systems are classified in different classes and types according to the structure of the Cas genes that are typically adjacent to CRISPR arrays [28]. For biotechnological purposes, the class 2 CRISPR systems are the most attractive because are easier to engineer. They involve a single Cas effector protein in comparison to class 1 systems, which often include several Cas proteins. This Cas effector protein is guided by a gRNA to a complementary genomic locus next to a protospacer adjacent motif sequence (PAM) which is essential for target recognition [29].

Since the demonstration that CRISPR systems can be used for genome editing in eukaryotic cells [30, 31], researchers are actively mining different CRISPR systems from the microbial diversity, what has led to the identification of different Cas effector proteins with divergences in their size, PAM requirements and substrate preferences. CRISPR revolution started with the class 2 type II-A CRISPR/Cas9 system from *Streptococcus pyogens,* which has a relatively simple 5-'NGG-3' PAM sequence. Smaller Cas9 orthologs thought for viral delivery have been identified and tested in eukaryotic cells, showing to be at least as robust as SpCas9 is. These include Cas9 orthologs from *Neisseria meningitidis* (NmCas9) [32], *Campylobacter jejuni* (CjCas9) [33], *Staphylococcus aureus* (SaCas9) [34] and *Streptococcus thermophilus* (StCas9) [35]. The latter two (SaCas9 and StCas9) were also evaluated in plants [36]. However, the complexity of the PAM sequences of these orthologs (i.e. 5'-NNGRRT-3' of SaCas9) drastically limited their adoption by the scientific community. Therefore, due to its PAM simplicity, SpCas9 has become the RNA-guided endonuclease (RGEN) of choice in most laboratories. Unfortunately, this choice constrains the target selection to guanine-rich regions. The recent discovery of a novel class 2 type

V-A CRISPR system could contribute to solving this limitation with its Cas effector protein termed Cas12a (formerly Cpf1), which has a preference for thymine-rich PAM [37]. Orthologs from *Acidaminococcus sp.* (AsCas12a) and *Lachnospiraceae bacterium* (LbCas12a) have shown to be active in eukaryotic cells [38].

Both CRISPR systems, type II-A of Cas9 and type V-A of Cas12a, have been proposed to have evolved through independent but remarkably similar pathways [39, 40]. In consequence, Cas9 and Cas12a are structurally unrelated proteins with different features besides the distinct PAM sequence as exemplified in Figure 4 [41].

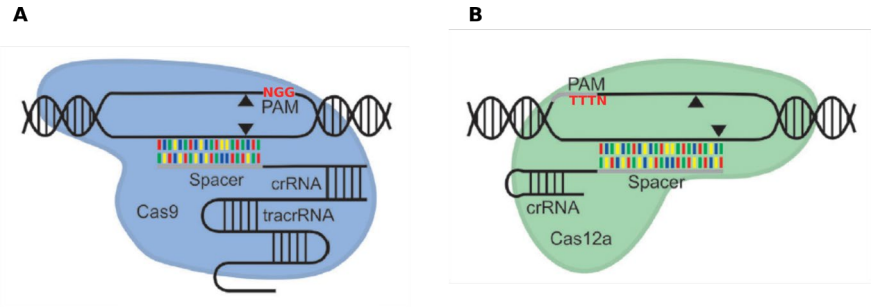


Figure 4: schematic representation of the main CRISPR class 2 systems used in genome engineering. **A)** Cas9 endonuclease recognizes a 5′-NGG-3′ PAM sequence with the aid of a gRNA introducing a blunt DSB in the DNA sequence. The gRNA is formed by a 20 nt PS sequence complementary to the genome, followed by the RNA scaffold. The scaffold is formed by the crRNA and the tracRNA which has been combined for biotechnological purposes in a single gRNA of ~100 nt. **B)** Cas12a endonuclease recognizes a 5′-TTTN-3′ PAM sequence introducing a staggered DSB. Note that in this case, the orientation of the gRNA is the opposite. The gRNA is formed by a short recognition scaffold (crRNA) and a variable spacer sequence (18-23 nt) complementary to the genome. As a result, the size of the gRNA is ~40 nt. Adapted from [44].

Thus, Cas9 recognizes a ~100 nt-long gRNAs scaffold with the targeted sequence (protospacer, PS) placed at its 5′ end. The Cas9 scaffold is formed by two small RNAs (tracrRNA and crRNA), which can be joined in a single molecule to facilitate biotechnological applications [23]. In nature, these RNAs are part

of a pre-crRNA which is processed by the RNase III accessory protein [42]. On the other hand, Cas9 creates a blunt DSB at 3 pb from the PAM sequence (Fig. 4A). In contrast, Cas12a has a completely different gRNA arrangement, target recognition and digestion pattern [37]. In opposite to Cas9, Cas12a endonuclease employs smaller gRNAs (~40 bp) with a shorter scaffold ubicated at the 5′ part of the transcript whereas the PS is present at the 3′ side. Furthermore, Cas12a has an intrinsic RNase activity that utilizes to self-process its own pre-crRNAs [43]. Thus, Cas12a has a dual DNase and RNase activity which can be exploited to target multiple loci at the same time [38]. Finally, this nuclease introduces a DSB at distal positions from the PAM sequence generating ends with 5′ overhangs (Fig. 4B).

All the characteristics described above make of Cas12a an interesting tool for certain goals of genome engineering. First, the T-rich PAM sequence opens new opportunities on gene editing by enabling access to regulatory sequences or A/T rich genomes. Second, the shorter scaffold and the RNase activity of the Cas12a might speed-up the obtention of edited crops at multiple loci by facilitating the construction of polycistronic gRNA expression cassettes targeting multiple loci. Finally, the unique mutation signature of Cas12a could favor the obtention of knock-out mutants and trigger HDR reparation mechanism, which could facilitate precise DNA editing and gene targeting.

CRISPR/Cas9 has been rapidly adopted by the plant scientific community transferring the information acquired from the reversal genetic studies in plant model organisms to actual advances in developing valuable traits for crop breeding. Introduction of mutations by NHEJ with CRISPR/Cas9 system has been used to accelerate the domestication process of wild relatives of modern tomato cultivars [45] in few years by targeting important genes involved in the domestication [46-48]; generating new sources of variability by engineering regulatory sequences of quantitative trait loci [49]; and engineering the

apomixis trait in rice producing clonal seeds from elite cultivars, what could change the hybrid seed business paradigm cheapening the obtention of seeds which could benefit farmers of developing countries [50, 51]. These just to cite some relevant examples of all the accomplishments that can be achieved with this system, which have been reviewed elsewhere [52].

In contrast with the extended use of Cas9 in plants, at the beginning of this thesis, the use of CRISPR/Cas12a was only described in mammalian cells, and no adaptations to plant cells were available. Only during the course of the thesis, the first examples of CRISPR/Cas12a edited plants started to appear in a few plant species, but direct comparisons between Cas9 and Cas12a in the same experimental system remained absent. Therefore, in this thesis we decided to set up a method for CRISPR/Cas12a editing in plants and to compare this with previous CRISPR/Cas9 editing tools established in our lab. The results of this work will be described in Chapter II.

3. Expanding the horizons of plant breeding through synthetic biology

In the first two sections of this Chapter I we have described the traditional plant breeding methods and discussed to what extent new methods, and particularly genome editing methods, can serve to accelerate this process. In this third part, we will try to move forward and explore the limits of plant breeding. In theory, with genetic engineering it would be possible to provide crops with functionalities beyond those available in the natural genetic pool. Some examples are "intelligent" plants capable of responding to weather forecast alerts, with e.g. production of protective anti-freezing proteins, or plants programmed to respond to the remote detection of a pest with the production of defensive compounds. These examples involve equipping plants

not with a single transgene or a small modification, but with genetic circuits and complex metabolic pathways involving many genes. Some authors defend that, to reach the levels of sophistication required to design this type of plants, it is necessary to incorporate engineering principles similar to those that allowed the industrial revolution and later the computer engineering to take off. These aspects are addressed by the discipline known as synthetic biology and represent a new horizon for plant biotechnology.

3.1. A brief introduction to synthetic biology and plant synthetic biology

Synthetic biology is a multidisciplinary field that aims to apply the basic engineering principles of abstraction, modularity, and standardization to molecular biology and genetics [53]. This new discipline investigates our ability to fabricate practical organisms with new or improved features to address a spectrum of urgent, real-world issues; that could for example clean hazardous waste [54], use plants to sense chemicals [55], produce biofuel [56] or recognize and destroy tumors [57]. In another direction, synthetic biologists also seek to reprogram many aspects of the biological systems. From the molecular level, by engineering artificial bases or amino acids with different physicochemical properties, to the *de novo* designed proteins with new structures not found in nature [58] or re-programming the genetic code to reduce the number of codons [59]. However, engineering these bits of the biological systems is complicated and still incipient. Nowadays, the major advances in synthetic biology have been achieved by considering the proteins and their regulatory elements as modular genetic building blocks. As an emerging discipline, synthetic biology is inspired by other well-established engineering disciplines like mechanical engineering or microelectronics [60, 61]. The fast development of both fields has been achieved by the adoption of standardized interchangeable parts and modular construction methods. The

standardization of the different parts has been crucial to making them easy to connect in a predictable way, building up increasingly complex outputs that accelerated the construction of more sophisticated devices. Currently, synthetic biology is in a similar situation.

3.1.1. Synthetic biology facilitated by modular cloning

One of the challenges of the discipline is to collect, characterize, and catalog DNA elements from living systems with potential utility. These standard parts and its associated information are deposited in bioregistries, which may serve also as collection points (biorepositories), although this latter functionality may become irrelevant as DNA synthesis becomes sufficiently cheap to build them from scratch. Besides standardization of genetic building blocks, a universal connecting method is also required to assemble these parts progressively in more complex genetic modules with predictable behavior. The BioBrick standard is the most famous example of bioregistry and assembly method for bacterial synthetic biology [62]. Other assembly methods based on Type IIS restriction enzymes and specialized in eukaryotes have emerged during these years, such MoClo [63] or GoldenBraid (GB) [64]. The latter system has been specialized in plant synthetic biology (PSB) and is based on the Golden Gate assembly system [65]. Golden Gate relies on BsaI-mediated restriction-ligation reaction to combine several genetic parts in a single step to construct single-use genetic modules. GB expands the reusability of these genetic modules by incorporating an additional Type IIS restriction enzyme (BsmbI) that allow the binary assembly of the genetic modules. The iterative use of both enzymes, BsmbI and BsaI, and the use of two destination plasmids establish a loop cloning schema which accelerates the construction of increasingly complex multigenic constructs (Fig. 5). In addition, GB's bioregistry (https://gbcloning.org) provides thousands of interchangeable genetic parts (phytobricks) deposited by the users that can be

assembled *in silico* using the software-assisted web tools to articulate a variety of constructs for plant multigenic engineering.

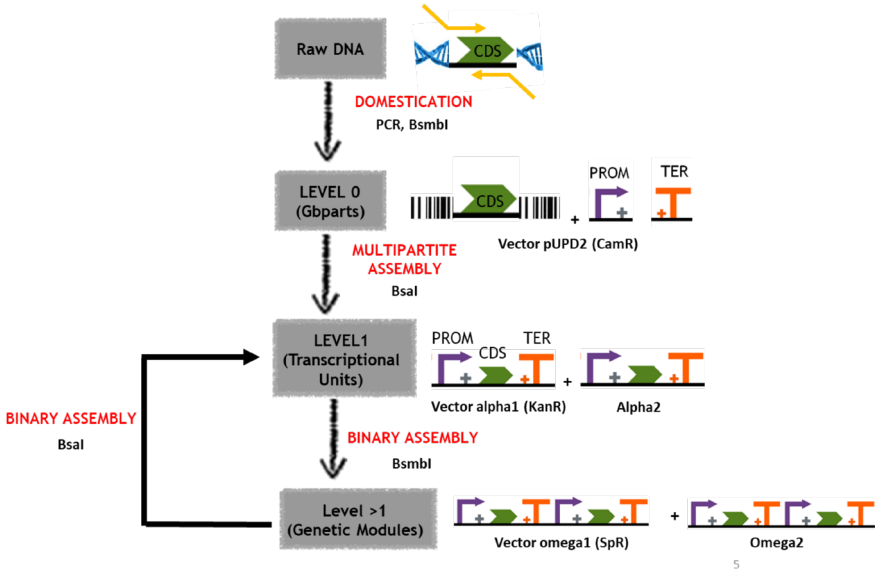


Figure 5: overview of GB cloning schema. Individual genetic parts referred as phytobricks are cloned in the Level 0 universal domestication plasmid 2 (pUPD2) through a BsmbI-mediated restriction-ligation reaction. To this end, phytobricks are PCR-amplified from a raw DNA source (genomic DNA or plasmids, for instance) and labeled with 4-nt overhangs and BsmbI sites. In this example, the overhang defines de phytobrick as a coding sequence (CDS), but extensive grammar defines multiple genetic elements [66]. Level 0 parts are used in a BsaI-mediated multipartite assembly to create a transcriptional unit (TU) in Level 1 alpha vectors. Subsequently, two TUs allocated in complementary alpha1 and alpha2 vectors can be combined in a binary assembly, within a Level≥1 omega vector using BsmbI to create a genetic module. Finally, two genetic modules in complementary omega1 and omega2 vectors can be combined in Level 1 through a BsaI reaction. This BsaI/BsmbI iterative loop can be repeated to create increasingly complex constructs.

3.1.2. Synthetic biology hits the plant kingdom

The major advances in synthetic biology have been accomplished predominantly in microbial DNA chassis. Since the initial milestone publications showing the first synthetic circuits in bacteria, the toggle switch, the repressilator and the autoregulatory negative-feedback circuit [67-69], the field has expanded notoriously to the point that some of the most striking biotechnological advances are linked to this field. Films recorded in the bacterial DNA [70], *de novo* synthesized and assembled genomes [70-73], engineered bacteria capable of invading and destroy tumoral cells [57] or the commercial production of antimalarial drug artemisinin by Amyris using engineered yeast strain [74]. All of these are accounted among the hallmark publications that exemplify the importance of this discipline, which is likely to continue growing and gaining importance in the future. Despite these organisms are relatively easy and fast to engineer, their unicellular nature poses some limitations to the extension of the challenges that synthetic biologists can tackle down with these systems. That is the reason why synthetic biology has advanced to multicellular systems, remarkably to plants [75].

Given its undoubtedly importance in world ecosystems, there are multiple reasons to practice synthetic biology in plants. First, plants are the most important source of primary metabolites (sugars, lipids, and proteins) that sustain the food chain. Second, the rich diversity of plants produce a broad number of secondary metabolites of value for medicine and industry. Third, plants are also a source of raw material for construction or energetic purposes. Finally, plants are ideal biofactories to produce valuable compounds such as biopharmaceuticals or metabolites for several reasons: biosafety (no human pathogens replicate in plants), scalability (cheap and easy to grow in comparison to bacterial or human cell cultures) and similar post-translational modifications as mammalian cells [76]. All these reasons have spurred the incipient field of PSB which aims to accomplish new

engineering goals more complex and precise than single-gene approaches followed in conventional transgenic breeding. Notable examples of the diverse and ambitious goals pursued by PSB account engineering plants as biosensors to detect explosives in the soil [55]; produce cereals that do not require chemical fertilizers by implementing nitrogen-fixation traits [77]; improve crop yield by turning C3 metabolism into C4 or providing alternative pathways for carbon dioxide fixation [78, 79]; or engineering plants for biofuels production towards green energy sources [80].

Most of the previously cited cases are examples of plant metabolic engineering. These approximations involve the construction of genetic constructs which are then delivered to the plant DNA chassis via conventional transformation procedures. However, certain goals cannot be accomplished by simply stacking constitutively expressed genes and transferring them to the plant genome. Sometimes transgene expression needs to be regulated to, for example, avoid deleterious effects on plant growth or to adjust the transcript levels for obtaining a better output. To this end, PSB practitioners need more refined tools to achieve their goals such as regulatory elements, programmable transcriptional factors or gene switches [81]. As a contribution to the development of PSB, in this thesis we aimed to create one of the most basic yet strongly needed elements required for gene circuits design: a memory switch. To do so, we made use of site-specific recombinases from bacterial origin, a resource that has been earlier used in plants for other purposes. Next, we will briefly describe the basis of site-specific recombination and their previous uses in plants.

3.2. Site-specific recombination

Genome engineering technologies are not limited to the introduction of DSB in the DNA sequence. As described in section 2.2. of this Chapter I, targeted nucleases can introduce a DNA rupture at specified sites that trigger the NHEJ or HDR reparation mechanisms, which end in diverse sequence modifications. Recombinases are another class of DNA-modifying enzymes that promote a variety of high-fidelity DNA modifications expanding thus the portfolio of manipulations that can be pursued.

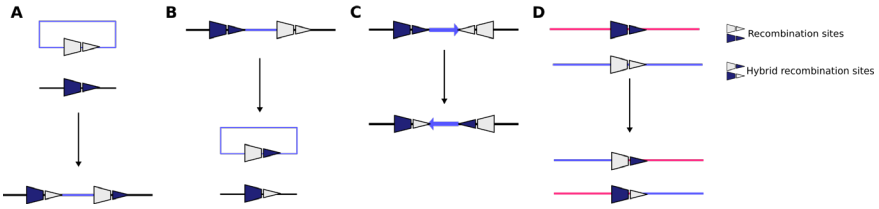


Figure 6: rearrangements mediated by the SSR enzymes within DNA substrates with complementary recombination sites. When the recombination sites share the same orientation (head-to-tail) integration **(A)** or excision **(B)** of a circular substrate occur. If the recombination sites hold an opposite orientation (head-to-head), then the DNA segment in between is inverted **(C)**. Exchange is also possible between recombination sites allocated in different linear DNA strands **(D)**.

Site-specific recombination (SSR) involves a reciprocal exchange between defined DNA sites [82]. Depending on the initial arrangement of the recombination sites and the topology of the DNA molecules (circular or linear), SSR has one of three possible outcomes: integration, excision or inversion (Fig. 6). Integration occurs between linear and circular DNA substrates bearing complementary recombination sites and results in the inclusion of the circular one, generating new hybrid sites (Fig. 6A). Excision is the opposite reaction and can also occur when both recombination sites allocated in the same linear molecule present head-to-tail orientation (Fig. 6B). On the contrary, if both recombination sites present confronted orientation (head-to-

head) the inversion of the flanked segment takes place (Fig. 6C). Finally, recombination can also take part between sites allocated in different linear DNA strands (Fig. 6D).

SSR enzymes are highly specialized proteins that promote DNA rearrangements between specific target sites. Virtually all the identified SSR enzymes can be classified in one of two evolutionarily distinct groups: the tyrosine (e.g. *Cre*, *FRP*) and the serine recombinases (e.g. Bxb1, PhiC31) [83], named by the amino acid residue that forms a covalent protein-DNA bond in the reaction intermediate. Mechanistically, tyrosine recombinases cleave single strands forming a covalent 3'-phosphotyrosine bond with the DNA backbone and rejoin the strands via a Holliday-junction-like intermediate state. In contrast, serine integrases make DSBs in the DNA forming 5'-phosphoserine bonds with the DNA backbone in advance of strand exchange and re-ligation [84]. In addition, recombination reactions mediated by the serine integrases are highly directional and are only reversed in the presence of an accessory protein called recombination directionality factor (RDF) [85, 86]. In opposite, tyrosine recombinase-mediated reactions are freely bidirectional, which limit their applicability [87]. For this reason, serine integrases have become the most popular recombinases for biotechnological purposes.

Serine integrases are encoded by temperate bacteriophages and catalyze recombination between attP (phage) or attB (bacteria) attachment sites generating hybrid attL (left) and attR (right) sites. To this end, integrase must recognize and bind to attP and attB and bring these sites together in a synapse before DNA cleavage, strand exchange and ligation of the products (Fig. 7) [88].

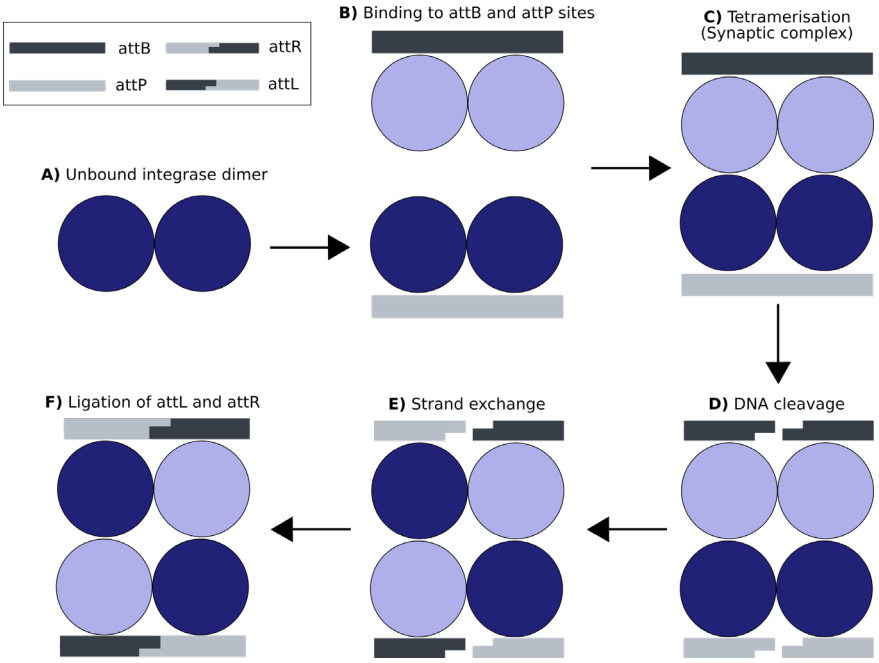


Figure 7: model of the serine recombination mechanism. A) In free solution, serine integrases are dimers with affinity to the attachment (att) attB and attP sites. **B)** Binding to the att sites produces a conformational change, inducing the tetramerization of the binary complexes bringing closer the complementary att sites establishing thus the synaptic complexes **(C)**. **D)** This triggers the scission of both DNA strands leaving complementary 2-bp overhangs. **E)** Next, the integrase subunits rotate 180° in relation to each other swapping the split att sites. **F)** Finally, the half-sites are re-ligated creating hybrid attR and attL sites.

Serine recombinases dimerize in solution and bind to the att sites with one subunit contacting each half-site (Fig. 7A-B). The bound dimers experience a conformational change that makes them interact with their counterparts, bringing attP and attB sites together in a synaptic complex that contains a tetramer of integrase subunits (Fig. 7C). Then, serine residues cleave the att sites generating half-sites with 3' 2-bp overhangs (Fig. 7D). Next, integrase subunits rotate in relation to each other, effectively swapping half-sites (Fig. 7E). Finally, if the 2-bp overhangs on the swapped half-sites are complementary, they will re-ligate to make

attL and attR sites (Fig. 7F). This reaction can be reversed in the presence of RDFs that in nature regulates the choice between lysogeny and lysis by channeling integrase recombination toward phage integration or excision. It is thought that integrase-RDF interaction takes place at the synaptic complex and induce a conformational change that inhibits the attBxattP recombination and stimulate attLxattR recombination [89, 90].

3.3. Uses of site-specific recombination in plant biotechnology

The use of site-specific recombinases to manipulate eukaryotic genomes began nearly three decades ago. In plant biotechnology recombinases have mostly been used for removing selectable markers, inserting and stacking genes at a predefined genomic locus [91-93], taking advantage of the SSR mechanisms of excision and integration (Fig. 6).

3.3.1. Transgene excision

Excision can be used to remove undesirable sequences from the plant chromosomes. Selection markers and reporter genes are essential to enhance the efficiency of the current transformation procedures, but these transgenic traits are unnecessary in commercial crops [94]. For biosafety concerns, the excision of transgenic traits in pollen has been proposed to avoid the dissemination of unwanted sequences in nature [95]. The first recombination systems used for transgenic trait removal were *Cre-lox* from *E. coli* bacteriophage P1 and *FLP-FRT* from yeast, which are tyrosine recombinases easy to engineer [96, 97]. The strategy for marker selection excision consist on flanking the selectable marker with the recombination sites and then introducing the recombinase by secondary transformation or by plant hybridization. As a result, the selectable marker flanked by the recombination sites will be excised from the genome yielding a plant free of undesired antibiotic resistance genes. A notable variation of this scheme included the selection marker together

with an inducible recombinase, both flanked by recombination sites, to control the precise moment in which the excision takes place and extending this technology to vegetatively propagated crops [98, 99].

3.3.2. Gene stacking

Conventional plant transformation methods involve random DNA integrations of single or multiple copies of the transgene whose sequence can be truncated during the process. These situations can be a source of variability and comprise the stability of transgene expression [100]. Moreover, the random integration may cause the disruption of endogenous genes creating undesirable collateral effects. All these reasons prompted the utilization of SSR systems for precise transgene insertion, ensuring single-copy transformation events with no gain or loss of sequences allocated at the specified locus. This level of accuracy has been observed in several studies in plant and mammalian cells, indicating high fidelity of the SSR systems [101, 102]. Therefore, the strategy consist on pre-establish a landing pad, which is a genomic locus endowed by a recombination site adjacent to the recombinase. Then, the complementary recombination site together with the gene of interest is delivered expecting both to be integrated at the landing pad.

The problem of tyrosine recombinases such as *Cre-lox* recombination system is that the reaction is freely reversible so eventually an integration event can be expelled from the genome. To favor the integration in front of the excision, several recombination site mutants were developed, which enhanced the unidirectionality of the *Cre*-mediated recombination [103]. However, the reaction was still residually bidirectional, limiting further gene additions to the pre-defined locus. To meet modern agriculture demands of more complex products for different geographies, the possibility to continue adding more traits at the same locus would be of interest. However, despite being harnessed to enhance insertion, freely reversible systems such as

Cre-lox tyrosine recombinase are not suitable for this purpose [104]. This problem was overcame by serine recombinases (Bxb1 and PhiC31) that catalyze a strictly unidirectional reaction involving nonidentical sites, attP and attB, that can only be reversed in the presence of a helper protein (RDF) yielding hybrid attL/attR sites, as described in the previous section. Hence, by sequentially adding attP or attB sites into the integration locus unlimited rounds of gene stacking could be done endowing an elite line of multiple linked traits [105-108].

3.4. Gene switches

Organisms have evolved to sense and adapt to external cues by re-adjusting their cellular functions. During this process, multiple naturally occurring gene switches capable of detecting a vast range of molecules and trigger a cellular response emerged. Taking advantage of these gene switches researchers have engineered orthogonal inducible expression systems to control transgene activity or study genes that have deleterious effects on development. These tools have become pivotal to advancing disciplines such as PSB whose importance has been previously addressed in this General Introduction.

Conventional inducible systems used routinely in plants are based on the control of transcription factors (activators or repressors) that regulate the production of a downstream gene of interest. Most inducible strategies in plants rely on the use of chemically-triggered systems such ethanol [109], dexamethasone [110], estradiol [111], copper [112, 113] or insecticides [114, 115], although more recently also optogenetic systems are starting to be explored in plant systems [116]. In chemical systems, the supply of the inducer triggers the expression or translocation to the nucleus of the transcription factors, which subsequently activate or repress the expression of the gene of interest. The activation output will be maintained as long as the inducer keep being administered. Hence, when the inducer is removed, the

output decay (Fig. 8A). In contrast, a bistable toggle switch retains a memory of the presence of the trigger. Therefore, once the trigger is applied, the output signal will change to "on" status, and such status will be sustained even if this inducer is retired. Only when an "off trigger" is applied, the system returns to its original status (Fig. 8B) Therefore, in contrast to the non-memory inducible systems, memory switches only need a limited supply of the trigger signal to sustain the output indefinitely. In the following section we will focus on this type of switches.

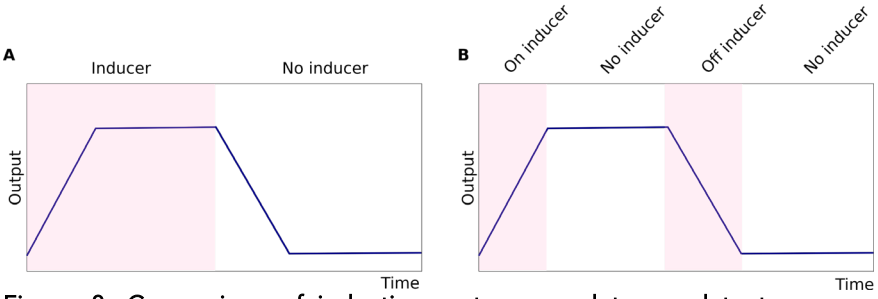


Figure 8: Comparison of induction systems used to regulate transgene expression. A) Classical induction systems output is determined by the presence or absence of inducers. The output will last as long as the inducer is supplied. **B)** In comparison, bistable toggle switches can alternate between two different states recording the memory of the stimulus and sustaining the output in its absence.

3.5. Memory switches in plant systems

A bistable genetic toggle switch allows a living organism to exist in two different stable states (bistable) and switch from one state to another as desired [117]. Bistable switches play an important role in decision-making at different phases of natural biological signaling and transcriptional networks [118], for example in the control of asymmetrical root development [118, 119] and shoot meristem homeostasis [120] in *A. thaliana*. One of the first synthetic gene regulatory networks assembled was a genetic toggle switch in *E. coli* named as the repressilator whose memory mechanism was based on double-negative feedback

loops [67]. Since that pioneering work, a number of different types of synthetic bistable switches have been constructed in bacterial or mammalian cells [121].

In plants, this toggle switch behavior has been achieved in tobacco plant protoplasts by Matias Zurbriggen´s group with a red light-responsive gene expression system [122]. This light-controlled toggle switch is based on the red/far red-dependent interaction of a split transcriptional activator. This is formed by the *A. thaliana* phytocrome B (PhyB) fused to an activator domain (VP16) and the phytocrome-interacting factor 6 (PIF6) attached to a DNA-binding domain, which is allocated upstream of a minimal promoter regulating the expression of a gene of interest (Fig. 9). Under far-red illumination (740 nm), both components of the toggle switch remain dissociated and therefore gene of interest (goi) is not expressed. Red light illumination (660 nm) triggers a conformational change in PhyB that reconstitutes the split transcriptional activator thus activating the expression of the gene of interest, as long as the protein-protein interaction is maintained.

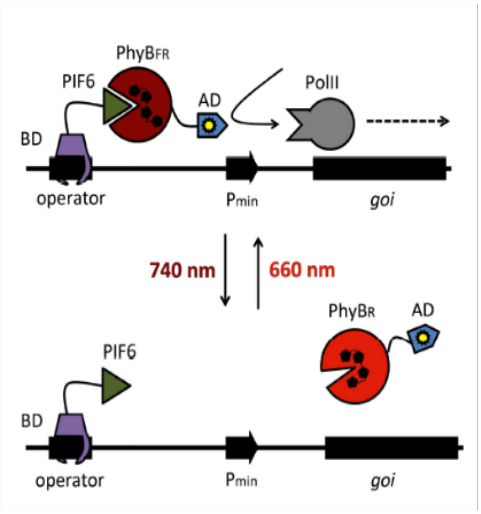


Figure 9: red-light responsive toggle switch for plant systems. The light-inducible system of Zurbriggen lab consists of a split transcriptional activator (BD:PIF6 and PhyB:VP16) which assembles upon red light illumination (660 nm) activating the expression of a downstream goi. As the dissociation rate of both restored transcriptional activator is very low, the system preserves this state until the proteins are degraded or are illuminated with far-red light (740 nm) breaking the interaction of PIF6 and PhyB. Adapted from [122].

Besides stable protein-protein interactions, memory can be installed in genetic systems in several ways, one of the most effective ones consisting in the introduction of covalent changes in the DNA. Therefore, of special interest would be to create a toggle switch based on SSR, whose stability is ensured by a covalent change in the DNA rather than in a passenger protein-protein interaction. This would create a memory state that virtually would last forever and could be transmitted over generations, avoiding any metabolic burden or instability inherent to the constant expression of transcriptional activators or repressors, typical of the classical inducible systems [123, 124].

In plants, SSR has only been applied for transgenic trait excision or insertion as has been previously described in this General Introduction. In Chapter III of this thesis, the applications of SSR in plants have been expanded by engineering a new bistable and reversible toggle switch based on the phage PhiC31 recombination system machinery.

4. References

1. Muller HJ: **Types of visible variations induced by X-rays inDrosophila**. *Journal of Genetics* 1930, **22**(3):299-334.
2. Stadler LJ: **On the Genetic Nature of Induced Mutation in Plants: A Haplo-viable deficiency in maize. II**; 1933.
3. McCallum CM, Comai L, Greene EA, Henikoff S: **Targeted screening for induced mutations**. *Nature Biotechnology* 2000, **18**(4):455-457.
4. Herrera-Estrella L, Depicker A, Van Montagu M, Schell J: **Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector**. *Nature* 1983, **303**(5914):209-213.
5. Jordan MC, McHughen A: **Glyphosate tolerant flax plants from Agrobacterium mediated gene transfer**. *Plant Cell Rep* 1988, **7**(4):281-284.
6. Stewart CN, Jr., Adang MJ, All JN, Boerma HR, Cardineau G, Tucker D, Parrott WA: **Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic Bacillus thuringiensis cryIAc gene**. *Plant Physiol* 1996, **112**(1):121-129.
7. Alonso JM: **Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana**. *Science* 2003, **301**(5633):653-657.
8. Sancho-Andrés G, Soriano-Ortega E, Gao C, Bernabé-Orts JM, Narasimhan M, Müller AO, Tejos R, Jiang L, Friml J, Aniento F *et al*: **Sorting Motifs Involved in the Trafficking and Localization of the PIN1 Auxin Efflux Carrier**. *Plant Physiol* 2016, **171**(3):1965-1982.
9. Chen K, Wang Y, Zhang R, Zhang H, Gao C: **CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture**. *Annu Rev Plant Biol* 2019, **70**:667-697.
10. Puchta H, Dujon B, Hohn B: **Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination**. *Proc Natl Acad Sci U S A* 1996, **93**(10):5055-5060.
11. Voytas DF, Gao C: **Precision genome engineering and agriculture: opportunities and regulatory challenges**. *PLoS biology* 2014, **12**(6):e1001877.

12. Puchta H: **The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution**. *Journal of Experimental Botany* 2004.
13. Salomon S, Puchta H: **Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells**. *EMBO J* 1998, **17**(20):6086-6095.
14. Liu W, Yuan JS, Neal Stewart C: **Erratum: Advanced genetic tools for plant biotechnology**. *Nature Reviews Genetics* 2014, **15**(1):62-62.
15. Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH, Veillet AL: **Mutations altering the cleavage specificity of a homing endonuclease**. *Nucleic Acids Res* 2002, **30**(17):3870-3879.
16. Rosen LE, Morrison HA, Masri S, Brown MJ, Springstubb B, Sussman D, Stoddard BL, Seligman LM: **Homing endonuclease I-CreI derivatives with novel DNA target specificities**. *Nucleic Acids Res* 2006, **34**(17):4791-4800.
17. Klug A, Rhodes D: **Zinc Fingers: A Novel Protein Fold for Nucleic Acid Recognition**. *Cold Spring Harbor Symposia on Quantitative Biology* 1987, **52**(0):473-482.
18. Kim YG, Cha J, Chandrasegaran S: **Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain**. *Proc Natl Acad Sci U S A* 1996, **93**(3):1156-1160.
19. Guan X, Stege J, Kim M, Dahmani Z, Fan N, Heifetz P, Barbas CF, Briggs SP: **Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors**. *Proceedings of the National Academy of Sciences* 2002, **99**(20):13296-13301.
20. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U: **Breaking the code of DNA binding specificity of TAL-type III effectors**. *Science* 2009, **326**(5959):1509-1512.
21. Moscou MJ, Bogdanove AJ: **A simple cipher governs DNA recognition by TAL effectors**. *Science* 2009, **326**(5959):1501.
22. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ *et al*: **A TALE nuclease architecture for efficient genome editing**. *Nat Biotechnol* 2011, **29**(2):143-148.

23. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**(6096):816-821.
24. Gasiunas G, Barrangou R, Horvath P, Siksnys V: **Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria**. *Proc Natl Acad Sci U S A* 2012, **109**(39):E2579-2586.
25. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR provides acquired resistance against viruses in prokaryotes**. *Science* 2007, **315**(5819):1709-1712.
26. Marraffini LA, Sontheimer EJ: **CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA**. *Science* 2008, **322**(5909):1843-1845.
27. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J: **Small CRISPR RNAs guide antiviral defense in prokaryotes**. *Science* 2008, **321**(5891):960-964.
28. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH *et al*: **An updated evolutionary classification of CRISPR-Cas systems**. *Nat Rev Microbiol* 2015, **13**(11):722-736.
29. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C: **Short motif sequences determine the targets of the prokaryotic CRISPR defence system**. *Microbiology* 2009, **155**(Pt 3):733-740.
30. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA *et al*: **Multiplex genome engineering using CRISPR/Cas systems**. *Science* 2013, **339**(6121):819-823.
31. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9**. *Science* 2013, **339**(6121):823-826.
32. Hou Z, Zhang Y, Propson NE, Howden SE, F. Chu L, Sontheimer EJ, Thomson JA: **Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis**. *Proceedings of the National Academy of Sciences* 2013, **110**(39):15644-15649.
33. Kim E, Koo T, Park SW, Kim D, Kim K, Cho H-Y, Song DW, Lee KJ, Jung MH, Kim S *et al*: **In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni**. *Nature Communications* 2017, **8**:14500.

34. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS *et al*: **In vivo genome editing using Staphylococcus aureus Cas9**. *Nature* 2015, **520**(7546):186-191.
35. Xu K, Ren C, Liu Z, Zhang T, Zhang T, Li D, Wang L, Yan Q, Guo L, Shen J *et al*: **Efficient genome engineering in eukaryotes using Cas9 from Streptococcus thermophilus**. *Cell Mol Life Sci* 2015, **72**(2):383-399.
36. Steinert J, Schiml S, Fauser F, Puchta H: **Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus**. *Plant J* 2015, **84**(6):1295-1305.
37. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A *et al*: **Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system**. *Cell* 2015, **163**(3):759-771.
38. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS *et al*: **Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array**. *Nat Biotechnol* 2017, **35**(1):31-34.
39. Koonin EV, Makarova KS, Zhang F: **Diversity, classification and evolution of CRISPR-Cas systems**. *Curr Opin Microbiol* 2017, **37**:67-78.
40. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI *et al*: **Diversity and evolution of class 2 CRISPR-Cas systems**. *Nat Rev Microbiol* 2017, **15**(3):169-182.
41. Swarts DC, Jinek M: **Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing**. *Wiley Interdiscip Rev RNA* 2018:e1481.
42. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E: **CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III**. *Nature* 2011, **471**(7340):602-607.
43. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E: **The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA**. *Nature* 2016, **532**(7600):517-521.
44. Kumlehn J, Pietralla J, Hensel G, Pacher M, Puchta H: **The CRISPR/Cas revolution continues: From efficient gene editing for**

crop breeding to plant synthetic biology. *J Integr Plant Biol* 2018, **60**(12):1127-1153.

45. Evans LT: **Crop Evolution, Adaptation and Yield**: Cambridge University Press; 1996.
46. Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D, Van Eck J, Lippman ZB: **Rapid improvement of domestication traits in an orphan crop by genome editing**. *Nat Plants* 2018, **4**(10):766-770.
47. Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C: **Domestication of wild tomato is accelerated by genome editing**. *Nat Biotechnol* 2018.
48. Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, Freschi L, Voytas DF, Kudla J, Peres LEP: **De novo domestication of wild tomato using genome editing**. *Nat Biotechnol* 2018.
49. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB: **Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing**. *Cell* 2017, **171**(2):470-480.e478.
50. Waltz E: **Asexual crops whet industry appetite**. *Nat Biotechnol* 2019, **37**(2):109-110.
51. Wang C, Liu Q, Shen Y, Hua Y, Wang J, Lin J, Wu M, Sun T, Cheng Z, Mercier R *et al*: **Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes**. *Nat Biotechnol* 2019, **37**(3):283-286.
52. Sedeek KEM, Mahas A, Mahfouz M: **Plant Genome Engineering for Targeted Improvement of Crop Traits**. *Front Plant Sci* 2019, **10**:114.
53. Endy D: **Foundations for engineering biology**. *Nature* 2005, **438**(7067):449-453.
54. Cases I, de Lorenzo V: **Genetically modified organisms for the environment: stories of success and failure and what we have learned from them**. *Int Microbiol* 2005, **8**(3):213-222.
55. Antunes MS, Ha S-B, Tewari-Singh N, Morey KJ, Trofka AM, Kugrens P, Deyholos M, Medford JI: **A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity**. *Plant Biotechnol J* 2006, **4**(6):605-622.
56. Savage DF, Way J, Silver PA: **Defossiling fuel: how synthetic biology can transform biofuel production**. *ACS Chem Biol* 2008, **3**(1):13-16.
57. Anderson JC, Christopher Anderson J, Clarke EJ, Arkin AP, Voigt CA: **Environmentally Controlled Invasion of Cancer Cells by**

Engineered Bacteria. *Journal of Molecular Biology* 2006, **355**(4):619-627.

58. Sismour AM, Michael Sismour A, Benner SA: **Synthetic biology**. *Expert Opinion on Biological Therapy* 2005, **5**(11):1409-1414.
59. Fredens J, Wang K, de la Torre D, Funke LF, Robertson WE, Christova Y, Chia T, Schmied WH, Dunkelmann DL, Beránek V: **Total synthesis of Escherichia coli with a recoded genome**. *Nature* 2019, **569**(7757):514.
60. Haseloff J, Ajioka J: **Synthetic biology: history, challenges and prospects**. *J R Soc Interface* 2009, **6 Suppl 4**:S389-391.
61. Way JC, Collins JJ, Keasling JD, Silver PA: **Integrating Biological Redesign: Where Synthetic Biology Came From and Where It Needs to Go**. *Cell* 2014, **157**(1):151-161.
62. Knight T: **Idempotent Vector Design for Standard Assembly of Biobricks**. 2003.
63. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S: **A modular cloning system for standardized assembly of multigene constructs**. *PLoS One* 2011, **6**(2):e16765.
64. Sarrion-Perdigones A, Falconi EE, Zandalinas SI, Juárez P, Fernández-del-Carmen A, Granell A, Orzaez D: **GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules**. *PLoS One* 2011, **6**(7):e21622.
65. Engler C, Marillonnet S: **Combinatorial DNA Assembly Using Golden Gate Cloning**. *Synthetic Biology* 2013:141-156.
66. Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology**. *Plant Physiol* 2013, **162**(3):1618-1631.
67. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli**. *Nature* 2000, **403**(6767):339-342.
68. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators**. *Nature* 2000, **403**(6767):335-338.
69. Becskei A, Serrano L: **Engineering stability in gene networks by autoregulation**. *Nature* 2000, **405**(6786):590-593.
70. Shipman SL, Nivala J, Macklis JD, Church GM: **CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria**. *Nature* 2017, **547**(7663):345-349.
71. Dymond JS, Richardson SM, Coombes CE, Babatz T, Muller H, Annaluru N, Blake WJ, Schwerzmann JW, Dai J, Lindstrom DL *et al*:

Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* 2011, **477**(7365):471-476.
72. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM *et al*: **Creation of a bacterial cell controlled by a chemically synthesized genome**. *Science* 2010, **329**(5987):52-56.
73. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA *et al*: **Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome**. *Science* 2008, **319**(5867):1215-1220.
74. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D *et al*: **High-level semi-synthetic production of the potent antimalarial artemisinin**. *Nature* 2013, **496**(7446):528-532.
75. Liu W, Neal Stewart C: **Plant synthetic biology**. *Trends in Plant Science* 2015, **20**(5):309-317.
76. Buyel JF: **Plant molecular farming–integration and exploitation of side streams to achieve sustainable biomanufacturing**. *Frontiers in plant science* 2018, **9**.
77. Rogers C, Oldroyd GED: **Synthetic biology approaches to engineering the nitrogen symbiosis in cereals**. *J Exp Bot* 2014, **65**(8):1939-1946.
78. Caemmerer Sv, von Caemmerer S, Quick WP, Furbank RT: **The Development of C4 Rice: Current Progress and Future Challenges**. *Science* 2012, **336**(6089):1671-1672.
79. South PF, Cavanagh AP, Liu HW, Ort DR: **Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field**. *Science* 2019, **363**(6422).
80. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD: **Biomass recalcitrance: engineering plants and enzymes for biofuels production**. *Science* 2007, **315**(5813):804-807.
81. Baltes NJ, Voytas DF: **Enabling plant synthetic biology through genome engineering**. *Trends Biotechnol* 2015, **33**(2):120-131.
82. Merrick CA, Zhao J, Rosser SJ: **Serine Integrases: Advancing Synthetic Biology**. *ACS Synth Biol* 2018, **7**(2):299-310.
83. Grindley NDF, Whiteson KL, Rice PA: **Mechanisms of site-specific recombination**. *Annu Rev Biochem* 2006, **75**:567-605.

84. Gaj T, Sirk SJ, Barbas CF, 3rd: **Expanding the scope of site-specific recombinases for genetic and metabolic engineering**. *Biotechnol Bioeng* 2014, **111**(1):1-15.
85. Ghosh P, Wasil LR, Hatfull GF: **Control of phage Bxb1 excision by a novel recombination directionality factor**. *PLoS Biol* 2006, **4**(6):e186.
86. Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MCM: **A phage protein that binds ϕC31 integrase to switch its directionality**. *Mol Microbiol* 2011, **80**(6):1450-1463.
87. Landy A: **Dynamic, structural, and regulatory aspects of lambda site-specific recombination**. *Annu Rev Biochem* 1989, **58**:913-949.
88. Smith MCM, Brown WRA, McEwan AR, Rowley PA: **Site-specific recombination by phiC31 integrase and other large serine recombinases**. *Biochem Soc Trans* 2010, **38**(2):388-394.
89. Mandali S, Gupta K, Dawson AR, Van Duyne GD, Johnson RC: **Control of Recombination Directionality by the Listeria Phage A118 Protein Gp44 and the Coiled-Coil Motif of Its Serine Integrase**. *J Bacteriol* 2017, **199**(11).
90. Fan H-F, Hsieh T-S, Ma C-H, Jayaram M: **Single-molecule analysis of φC31 integrase-mediated site-specific recombination by tethered particle motion**. *Nucleic Acids Research* 2016, **44**(22):10804-10823.
91. Petolino JF, Kumar S: **Transgenic trait deployment using designed nucleases**. *Plant Biotechnol J* 2016, **14**(2):503-509.
92. Srivastava V, Thomson J: **Gene stacking by recombinases**. *Plant Biotechnol J* 2016, **14**(2):471-482.
93. Ow DW: **The long road to recombinase-mediated plant transformation**. *Plant Biotechnology Journal* 2016, **14**(2):441-447.
94. Hohn B, Levy AA, Puchta H: **Elimination of selection markers from transgenic plants**. *Current Opinion in Biotechnology* 2001, **12**(2):139-143.
95. Hare PD, Chua N-H: **Excision of selectable marker genes from transgenic plants**. *Nature Biotechnology* 2002, **20**(6):575-580.
96. Dale EC, Ow DW: **Intra- and intramolecular site-specific recombination in plant cells mediated by bacteriophage P1 recombinase**. *Gene* 1990, **91**(1):79-85.

97. Odell J, Caimi P, Sauer B, Russell S: **Site-directed recombination in the genome of transgenic tobacco**. *MGG Molecular & General Genetics* 1990, **223**(3).
98. Sugita K, Kasahara T, Matsunaga E, Ebinuma H: **A transformation vector for the production of marker-free transgenic plants containing a single copy transgene at high frequency**. *The Plant Journal* 2000, **22**(5):461-469.
99. Zuo J, Niu Q-W, Møller SG, Chua N-H: **Chemical-regulated, site-specific DNA excision in transgenic plants**. *Nature Biotechnology* 2001, **19**(2):157-161.
100. Wilson C, Bellen HJ, Gehring WJ: **Position Effects on Eukaryotic Gene Expression**. *Annual Review of Cell Biology* 1990, **6**(1):679-714.
101. Ow DW: **Recombinase-directed plant transformation for the post-genomic era**. *Functional Genomics* 2002:183-200.
102. Sauer B: **Cre/lox: One More Step in the Taming of the Genome**. *Endocrine* 2002, **19**(3):221-228.
103. Albert H, Dale EC, Lee E, Ow DW: **Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome**. *The Plant Journal* 1995, **7**(4):649-659.
104. Ow DW: **Site-Specific Gene Stacking Method**. *Plant Biotechnology 2002 and Beyond* 2003:215-218.
105. Ow DW: **Recombinase-mediated Gene Stacking as a Transformation Operating SystemF**. *Journal of Integrative Plant Biology* 2011, **53**(7):512-519.
106. Wang Y, Yau Y-Y, Perkins-Balding D, Thomson JG: **Recombinase technology: applications and possibilities**. *Plant Cell Reports* 2011, **30**(3):267-285.
107. Yau Y-Y, Wang Y, Thomson JG, Ow DW: **Method for Bxb1-Mediated Site-Specific Integration In Planta**. *Methods in Molecular Biology* 2011:147-166.
108. Hou L, Yau Y-Y, Wei J, Han Z, Dong Z, Ow DW: **An Open-Source System for In Planta Gene Stacking by Bxb1 and Cre Recombinases**. *Molecular Plant* 2014, **7**(12):1756-1765.
109. Roslan HA, Salter MG, Wood CD, White MRH, Croft KP, Robson F, Coupland G, Doonan J, Laufs P, Brian Tomsett A *et al*: **Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana**. *The Plant Journal* 2001, **28**(2):225-235.

110. Aoyama T, Chua NH: **A glucocorticoid-mediated transcriptional induction system in transgenic plants**. *Plant J* 1997, **11**(3):605-612.
111. Zuo J, Niu QW, Chua NH: **Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants**. *Plant J* 2000, **24**(2):265-273.
112. Mett VL, Podivinsky E, Tennant AM, Lochhead LP, Jones WT, Reynolds PH: **A system for tissue-specific copper-controllable gene expression in transgenic plants: nodule-specific antisense of aspartate aminotransferase-P2**. *Transgenic Res* 1996, **5**(2):105-113.
113. Saijo T, Nagasawa A: **Development of a tightly regulated and highly responsive copper-inducible gene expression system and its application to control of flowering time**. *Plant Cell Rep* 2014, **33**(1):47-59.
114. Padidam M, Gore M, Lu DL, Smirnova O: **Chemical-inducible, ecdysone receptor-based gene expression system for plants**. *Transgenic Res* 2003, **12**(1):101-109.
115. Unger E, Mark Cigan A, Trimnell M, Xu R-J, Kendall T, Roth B, Albertsen M: **Transgenic Research**. 2002, **11**(5):455-465.
116. Chatelle C, Ochoa-Fernandez R, Engesser R, Schneider N, Beyer HM, Jones AR, Timmer J, Zurbriggen MD, Weber W: **A Green-Light-Responsive System for the Control of Transgene Expression in Mammalian and Plant Cells**. *ACS Synth Biol* 2018, **7**(5):1349-1358.
117. Medford JI, Prasad A: **Towards programmable plant genetic circuits**. *Plant J* 2016, **87**(1):139-148.
118. Kassaw TK, Donayre-Torres AJ, Antunes MS, Morey KJ, Medford JI: **Engineering synthetic regulatory circuits in plants**. *Plant Science* 2018, **273**:13-22.
119. Cruz-Ramírez A, Díaz-Triviño S, Blilou I, Grieneisen VA, Sozzani R, Zamioudis C, Miskolczi P, Nieuwland J, Benjamins R, Dhonukshe P *et al*: **A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division**. *Cell* 2012, **150**(5):1002-1015.
120. Schoof H, Lenhard M, Haecker A, Mayer KFX, Jürgens G, Laux T: **The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes**. *Cell* 2000, **100**(6):635-644.

121. Lebar T, Bezeljak U, Golob A, Jerala M, Kadunc L, Pirš B, Stražar M, Vučko D, Zupančič U, Benčina M *et al*: **A bistable genetic switch based on designable DNA-binding domains**. *Nat Commun* 2014, **5**:5007.
122. Müller K, Siegel D, Rodriguez Jahnke F, Gerrer K, Wend S, Decker EL, Reski R, Weber W, Zurbriggen MD: **A red light-controlled synthetic gene expression switch for plant systems**. *Mol Biosyst* 2014, **10**(7):1679-1688.
123. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK: **Synthetic recombinase-based state machines in living cells**. *Science* 2016, **353**(6297):aad8559.
124. Siuti P, Yazbek J, Lu TK: **Synthetic circuits integrating logic and memory in living cells**. *Nature Biotechnology* 2013, **31**(5):448-452.

OBJECTIVES

All the information reviewed above lead us to propose a research plan directed to expand the plant biotechnological toolkit with new tools for precise genome editing and transcriptional regulation. To develop this general objective, a set of specific objectives were planned:

1. To design and develop new phytobricks for precise plant DNA manipulation.
    i. To expand the genome engineering tools with all the necessary elements to perform site-specific mutagenesis with CRISPR/Cas12a endonuclease (Chapter II).
    ii. To design a reversible and biestable memory switch based on the site-specific recombination machinery of the phage PhiC31 (Chapter III).
2. To extensively characterize these tools through transient and stable expression experiments in different model plants:
    i. To evaluate different Cas12a endonuclease orthologs and crRNA expression strategies (Chapter II).
    ii. To compare the mutagenesis efficiency of CRISPR/Cas12a with the well-stablished CRISPR/Cas9 (Chapter II).
    iii. To assess the genome-wide specificity of CRISPR/Cas12a (Chapter II).
    iv. To test the biestability and memory features of the phage PhiC31-based toggle switch (Chapter III).
    v. To demonstrate the inducibility of the PhiC31 integrase (Chapter III).

CHAPTER II

Assessment of Cas12a-mediated gene editing efficiency in plants

Bernabé-Orts, J. M., Casas-Rodrigo, I. , Minguet, E. G., Landolfi, V. , Garcia-Carpintero, V. , Gianoglio, S. , Vázquez-Vilar, M. , Granell, A. and Orzaez, D. (2019). Plant Biotechnol J, 17: 1971-1984. https://doi.org/10.1111/pbi.13113

1. Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) is an adaptive immune system, which defends prokaryotes against invader nucleic acids [1-3]. Components of Class 2 Type II CRISPR systems have been engineered into easily programmable RNA-guided endonucleases (RGENs), setting the stage for genome editing approaches that will be widely available to the research community [4]. The arrangement of CRISPR-based tools typically involves two components: the Cas9 effector protein and a ~100 nt small guide RNA (sgRNA), which is a combination of the crRNA and tracrRNA originally encoded by the CRISPR loci [5]. To date, the most widely used Cas9 orthologue has been SpCas9 (*Streptococcus pyogens*). The tandem SpCas9-sgRNA targets genomic sequences of 20 nt upstream to a 5′-NGG-3′ protospacer adjacent motif (PAM) and introduces a double-strand break (DSB) in many organisms [6-9]. The DSB takes place near the PAM sequence and originates blunt ends that, in plants, will be repaired preferentially by non-homologous end joining (NHEJ), occasionally introducing indel mutations [10].

The canonical CRISPR/Cas9 RGEN system offers a variety of applications in plant biotechnology, from crop breeding to plant synthetic biology. The efficiency of SpCas9 in genome editing has been demonstrated in dozens of plant species, where it has been used to produce new traits of agronomic importance, including resistance to abiotic stress, herbicides, etc. [11-13]. Despite its success, further expansion of the CRISPR toolbox with new RGEN tools will be necessary to overcome inherent limitations of SpCas9. One of these limitations is the strict PAM-dependence, which constrains the availability of target sites. For instance, non-coding regions, which are important in breeding for the generation of artificial quantitative trait variability [14, 15], are relatively poor in 5′-NGG-3′ sites [16]. Other current limitations of SpCas9, like its large size for viral delivery, or the low efficiency in

gene targeting caused by blunt DSB, could eventually be overcome with the introduction of alternative editing tools.

Recently, a new putative Class 2 Type V CRISPR system has been identified [17]. This locus encodes for a different Cas effector protein, Cas12a (Cpf1), and a single crRNA of 42 nt; The CRISPR/Cas12a system displays different features from that of SpCas9 [18], making it an interesting complement to SpCas9 in plant genome engineering. *Acidaminococcus sp.* Cas12a (AsCas12a) and *Lachnospiraceae bacterium* Cas12a (LbCas12a) are the most commonly used Cas12a orthologues, which have been proven to be effective in mammalian cells [17, 19, 20]. Both RGENs recognize a 5′-TTTN-3′ PAM sequence and introduce a staggered DSB at the end of the protospacer (PS) sequence, which has been reported to favor gene insertions [17]. Finally, Cas12a is smaller than SpCas9, which might facilitate viral delivery [21].

CRISPR/Cas12a engineering activity was recently demonstrated in plant cells. First, targeted mutagenesis of rice and tobacco genomes was achieved using the *Franciscella novicida* Cas12a (FnCas12a) orthologue [22]. Most recently, T0 mutant lines in rice have been obtained using AsCas12a and LbCas12a.[23-26]. Also, a DNA-free approach, delivering AsCas12a and LbCas12a loaded with crRNA, was validated in wild tobacco and soybean protoplasts [27]. However, the decision to use Cas9 or Cas12a is difficult to make since few studies have assessed the reliability and robustness of Cas12a as an alternative to Cas9 in plants. Here we report a comprehensive assessment of Cas12a activity in plants through the analysis of more than 130 mutagenesis events in three different plant species. Our results indicate that Cas12a is a viable alternative to Cas9 for genome editing in plants, although further efforts will be required to make the efficiency of the editing activity more predictable.

2. Results

2.1. A GB-assisted cloning strategy for plant Cas12a constructs

GoldenBraid 3.0 (GB3.0) is a software-assisted multigene cloning system (https://gbcloning.upv.es) based on type IIS restriction enzymes [28] and conforming to the phytobrick standard [29]. GB3.0 was earlier adapted to plant gene editing with SpCas9 [30]. Here, all the necessary elements to carry out targeted mutagenesis with AsCas12a and LbCas12a were adapted to GB3.0. Figure 1a depicts the cloning workflow followed to create the so-called Cas12a gene editing module (CGEM). Cloning a functional crRNA requires the following three components: (i) an invariant 5´element comprising the *At*U6-26 RNApol III promoter fused to either the AsCas12a or LbCas12a direct repeat (DR) (GB1442 or GB1443 standard phytobricks respectively); (ii) two partially complementary oligonucleotides containing the PS sequence and flanked by 4 nt overhangs complementary to the adjacent elements and (iii) a 3´element comprising the signals for correct termination and processing of the crRNA (GB1444 or similar). All Level 0 elements are combined into a Level 1 composite part through a cyclic restriction-ligation Golden Gate-like multipartite BsaI reaction, to build the entire crRNA TU. Finally, the assembled crRNA cassette is joined with pre-assembled As/LbCas12a TUs (GB1440, GB1441) on Level>1 with a BsmBI-mediated restriction-ligation reaction, producing the final CGEM. The CGEM can be directly used for transient expression or attached to a selection marker (e.g. kanamycin resistance gene GB1181) for stable transformation experiments. All the necessary elements to perform plant targeted mutagenesis with Cas12a will be deposited in Addgene and in the GB3.0 repository.

As a first optimization step, we analyzed two design alternatives concerning 3´ RNA processing signals. Since superfluous nucleotides at the 3' end of the crRNA derived from the poly-T U6 terminator could negatively affect the efficiency or the specificity of the Cas12a, we compared the efficiency of a self-processing hepatitis delta virus (HDV) ribozyme that removes the poly-A tail from the transcript leaving no additional nucleotides at the 3' position of the crRNA with a classical construct carrying unprocessed U6 termination signal, which conserves a spurious tail of adenines at the 3´position of the Cas12 crRNA (Fig. 1b). The introduction of a self-cleavable ribozyme in the 5´end of the transcript was unnecessary in our design, as we added a 5'-G to the Cas12a DR, which is compatible with the RNApol III promoter *At*U6-26 transcription start site. We evaluated the efficiency of both constructs on the *N. benthamiana XT1* target locus (*Niben101Scf04205g03008.1*) using a transient assay. Mutation at the *XT1* target site can be easily scored by the loss of an *Eco*RI site. Restriction enzyme (RE) analysis revealed that both AsCas12a and LbCas12a introduced mutations in *XT1* (Fig. 1b), however statistically significant differences were observed between the HDV and U6 terminator strategies for both nucleases, with higher activity observed for the HDV crRNA construct. For this target, LbCas12a was the most effective nuclease, exhibiting an activity 6-fold (HDV setup) and 11-fold (U6 terminator setup) higher than AsCas12a.

Recently, a multiplexing strategy was described for Cas12a based on its ability to process multimeric crRNAs containing two or more PS sequences separated by DRs [31]. To test this function in plants, we designed a double PS construct simultaneously targeting the *XT1* and *XT2* (*Niben101Scf04551g02001.1*) loci and tested its mutagenesis efficiency in *N. benthamiana* (Fig. 1c). Because mutagenesis at the *XT2* target site does not produce a restriction enzyme polymorphism, editing efficiency for each target was estimated using the T7 endonuclease I (T7E1) mismatch cleavage assay. As shown, the self-processing strategy

was successful in guiding Cas12a to both target loci and the efficiencies scored were similar with both strategies (HDV and multiplexing).

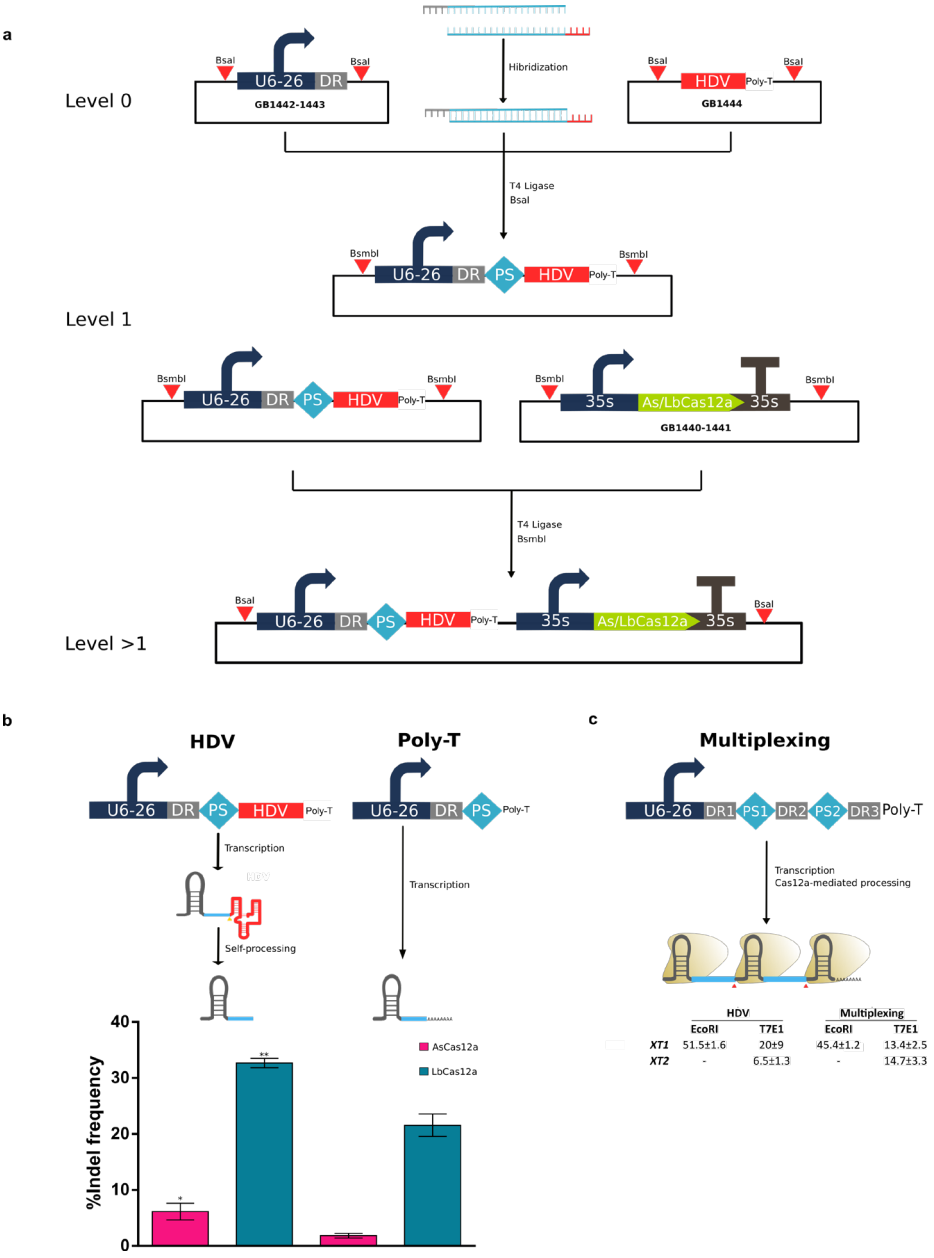


Figure 1: Cas12a gene editing module (CGEM) assembly and strategy of the crRNA expression. (a) Level 1 cloning of the crRNA TU through BsaI restriction-ligation reaction requires three elements: *At*U6-26:DR (GB1442, GB1443), partially complementary oligos of the PS and HDV termination signal (GB1444). BsmbI- mediated binary assembly into Level>1 of the crRNA TU and the As/LbCas12a TU (GB1440, GB1441) to create the Cas12a gene editing module. **(b)** Schematic of the constructs used for the expression of a single crRNA and the final structure of the transcript. Efficiency comparison between HDV and U6 poly-T approach expressed as the percentage of undigested band after *Eco*RI digestion. Statistical *t*-test analysis showed difference for the As and the LbCas12a (pvalue<0.05). Error bars represents SEM; *n*=3. **(c)** Representation of the construct used for the expression of multiple crRNAs in the same transcript, and table showing the comparison of the efficiencies between HDV and multiplexing.

2.2. Transient expression of CGEM provides efficient targeted mutagenesis in *N. benthamiana* leaves

To assess to what extent the differences between AsCas12a and LbCas12a efficiency observed for the *XT1* locus could be generalized, similar comparisons were made at several positions of the *N. benthamiana* genome using the HDV strategy. Furthermore, SpCas9 constructs targeting the same locus were added for comparison. In total, we examined 8 chromosomal sites, four of them designed to have completely overlapping target sequences between As/LbCas12a and SpCas9 and the remaining four with a partially overlapping PS (Fig. 2a). Mutagenesis efficiency for each RGEN-locus pair was scored in transient transformation assays in *N. benthamiana* using the T7EI mismatch cleavage assay. Each experimental point comprised three biological replicates for which three different leaves were infiltrated and pooled (see Fig. S3a). Results plotted in Fig. 2a revealed strong differences in the mutagenesis efficiency of RGENs depending on the target, ranging from undetectable levels to a maximum above 30% efficiency for LbCas12a at the *TFL1 14.1* locus. The Cas12a editing results obtained with T7EI assay for *XT1* locus were compatible with those obtained previously with the RE assay, although the T7E1 assay is less sensitive than restriction polymorphisms. Only one of the assayed

targets (*XT2A*) was mutated by all three RGEN at similar levels (9%). On the contrary, the *XT2B* locus was only efficiently mutagenized by SpCas9. For the targets assayed, LbCas12a was the most reliable RGEN (Fig. 2b), causing detectable indels in 7 out of 8 targets and displaying the higher mutagenesis average, as compared to SpCas9 and AsCas12a (SpCas9 = 5±2%; AsCas12a = 3±2%; LbC12a =17±6%).

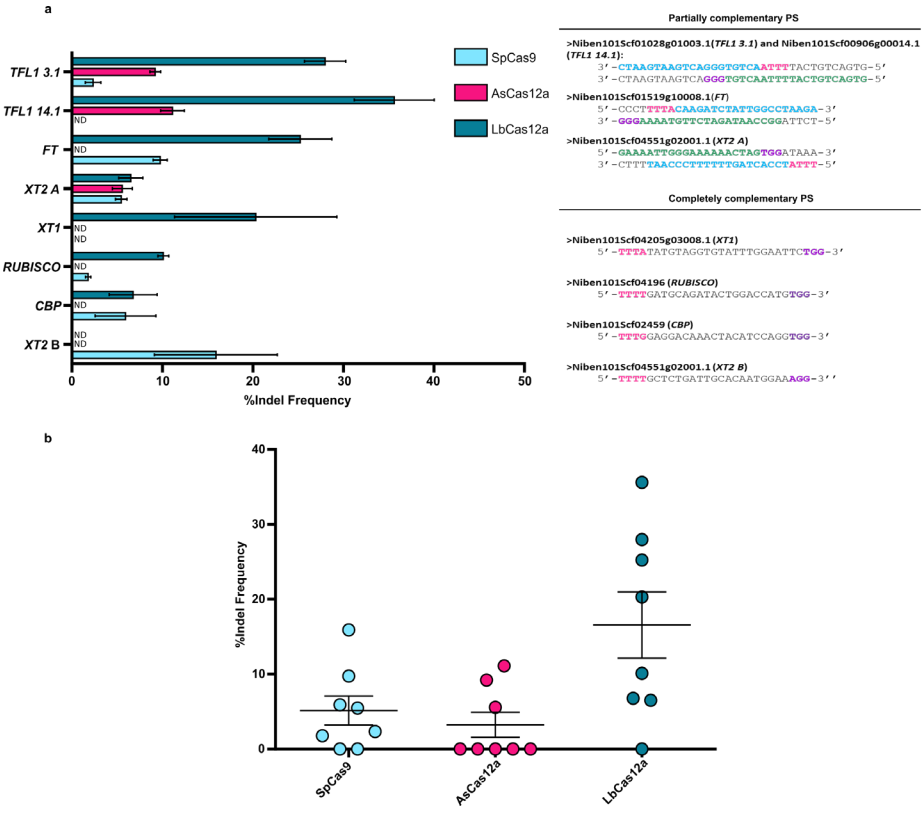


Figure 2: Comparison of the RGEN at several loci of *N. benthamiana*. (a) Mutation frequency for individual loci targeted with the three RGEN and determined by T7E1 mismatch cleavage assay. The sequence for each locus is detailed on the right part of the figure, with the Cas12a PAM highlighted on pink, the PS of Cas12a on light blue, the Cas9 PAM on purple and the PS on dark green. Error bars indicate SEM; *n*=3. ND: not detected. **(b)** Summary of the mutated and no mutated loci for each endonuclease. Mean indel frequencies ± SEM are shown.

2.3. Modifications in the crRNA DR loop affect RGEN activity but cannot compensate low efficiency editing

The crRNA structure includes a DR region whose proper folding is important for nuclease activity [20, 32]. The DR region of AsCas12a and LbCas12a are identical except in the distal loop, which contains a 5′-UCUU-3′ tetra-nucleotide in AsDR and a 5′-UAAGU-3′ penta-nucleotide in the case of LbDR (Fig. 3a).

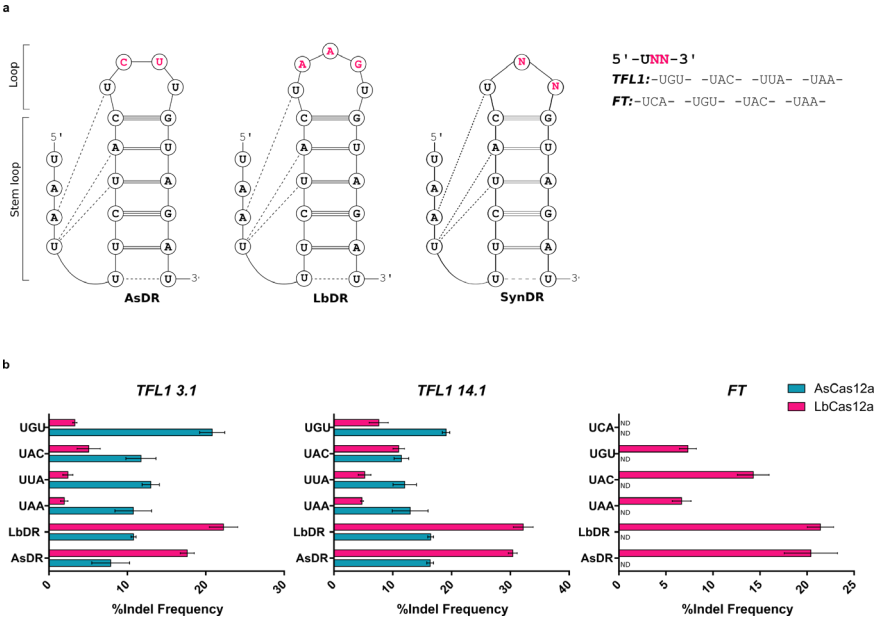


Figure 3: The DR loop sequence influences Cas12a activity. (a) DR structure of the AsCas12a (AsDR), LbCas12a (LbDR) and synthetically engineered (synDR). Highlighted in pink the different nucleotides of the three structures. The varying nucleotides of the synDR loops are also indicated (5′-UNN-3′) for each target. **(b)** Mutagenesis efficiency results of the wtDRs and synDRs at *TFL1* and *FT* targets. All the DRs structures were evaluated for each endonuclease. Error bars represent SEM; n=3. ND mean not detected.

Cas12a variants are known to accept crRNAs from other species [17] demonstrating a certain flexibility in the DR loop. A possible factor influencing Cas12a efficiency is the stability of the DR region. The stability can be affected by the presence of additional complementary regions between the DR and the PS. Therefore, we decided to test the possibility of reducing the size of the DR loop, thus minimizing interactions with PS. Several simplified synthetic DR loops (synDRs) were designed with different abilities to disrupt/conserve the DR integrity, as assessed by the RNA secondary structure prediction software ViennaRNA [33] and MFold [34] that calculate the stability of DR loop-DR/PS interactions based on the predicted –ΔG (Table S3). As shown in Fig. 3b, all crRNA structures (wtDRs and synDRs) exhibited detectable mutagenesis levels at *TFL1 3.1* and *TFL1 14.1* targets, whereas only LbCas12a was active at the *FT* locus. Notably, the AsCas12a-synDR 5′-UGU-3′ combination outperformed native AsCas12a-AsDR at the *TFL1 3.1* locus, reaching levels similar to those obtained with LbCas12a in its native combination. However, none of the synDR loops assayed was capable of increasing AsCas12a activity at the *FT* target to detectable levels. In general, LbCas12a performed better with AsDRs and LbDRs than with synDRs containing shorter loops, whereas AsCas12a activity was unaffected or slightly improved by trinucleotide loops.

2.4. Cas12a as a genome editing tool in *N. benthamiana*, tomato and Arabidopsis

As the next step in the evaluation of the mutagenesis activity in transient assays, we assessed the ability of GB3.0 Cas12a constructs to perform genome editing using standard stable transformation protocols in three different plant species. To this end, we used a generic gene editing module represented in Fig. 4a, containing three TUs: the kanamycin selection marker, the crRNA expression cassette and the TU encoding SpCas9,

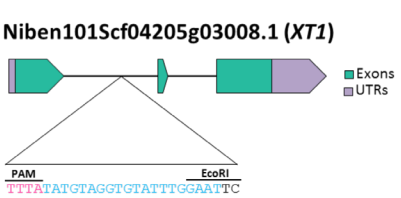
AsCas12a or LbCas12a nucleases. Stable transformation of *N. benthamiana* was first used to confirm the results obtained in leaf transient experiments targeting the *XT1* locus using AsCas12a and LbCas12a (Fig. 4b). Genotyping of the T0 transgenic generation revealed that 54% of the 37 kanamycin-resistant plants carrying the LbCas12a construct had mutations at the target site (Fig. S1a). In contrast, no edited plants were found for the AsCas12a construct (see Fig. S1b). TIDE analysis of the T0 edited lines revealed that the majority of the mutated lines were bialellic (60%); line 22 even showed 4 different mutations, consistent with an early chimera derived from two differentially edited progenitor cells (Fig. 4b).

AsCas12a, LbCas12a and SpCas9, were also assayed in tomato plants in a parallel experiment where all three editing constructs were designed to target the same locus *Solyc01g079620* (*MYB12*). Fig. 4c shows the precise targeting sites for each RGEN and the TIDE analysis for the mutated lines. Only one chimeric plant was recovered from the AsCas12a experiment, showing a 9 bp deletion present in approximately 10% of the copies of the genomic DNA samples analyzed. In contrast, two edited plants were recovered for both LbCas12a and SpCas9 nucleases. Both SpCas9-edited plants were biallelic while only one of theLbCas12-mutated plants had mutations in the two alleles.

Figure 4: Mutagenesis data extracted from stable transformations of different model plants. (a) Generic GEM encompasses three TUs: kanamycine selection marker (*NptII*), the crRNA expression cassette and the RGEN TU. **(b)** Representation of the *XT1* loci of *N. benthamiana* with the targeted sequence, and table collecting TIDE data from the mutated lines. Marked lines (*) were not analyzed with TIDE. **(c)** Genomic loci of the *S. lycopersicum MYB12* targeted gene, with the three RGEN, and results of the TIDE analysis of the edited plants. The "overall efficiency" represents the sum of all the individual mutations traced by TIDE. The "mutation efficiency" specifies the type of mutations associated to its efficiency in brackets. Only mutations which p-value<0.001 have been considered in "mutation efficiency".

a

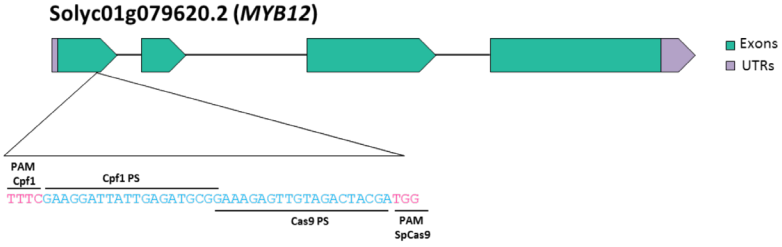
b



N. benthamiana

Line T0	Overall Efficiency	Mutation Efficiency
LbCas12a 2	98.1	-16 (57.9), -10 (40.2)
LbCas12a 4	94.4	-7 (49.5), -1 (42.4)
LbCas12a 6	60.4	-9 (56.7)
LbCas12a 7	94.1	-40 (56.7), -7 (36.6)
LbCas12a 10	54.7	-7 (54.2)
LbCas12a 13	99.4	-10 (98)
LbCas12a 14	97.3	-9 (45.9), -6 (50.5)
LbCas12a 17	94.8	-8 (42.7), -2 (50.1)
LbCas12a 18	51.2	-8 (50.5)
LbCas12a 20*	-	-83, -14, -7
LbCas12a 21	97.6	-10 (50.7), -8 (46.1)
LbCas12a 22	93.8	-12 (9.9), -9 (15.1), -8 (17.7), -6 (50.1)
LbCas12a 31	93.6	-17 (51.5), -7 (41.3)
LbCas12a 33	95.3	-5 (37.9), -30 (56.6)
LbCas12a 34*	-	-83, -14, -9, -7
LbCas12a 35	93.7	-10 (48.1), -8 (42.7)
LbCas12a 37	66.1	-7 (63.2)

c



S. lycopersicum

Line T0	Overall Efficiency	Mutation Efficiency
AsCas12a 2	8.8	-9 (7.1)
LbCas12a 3	82.7	-6 (38.3), -2 (43.5)
LbCas12a 5A	49.1	-2 (48.8)
SpCas9 6	95.5	-2 (49.6), -1 (45.7)
SpCas9 8	96.4	-3 (50.1), +1 (46)

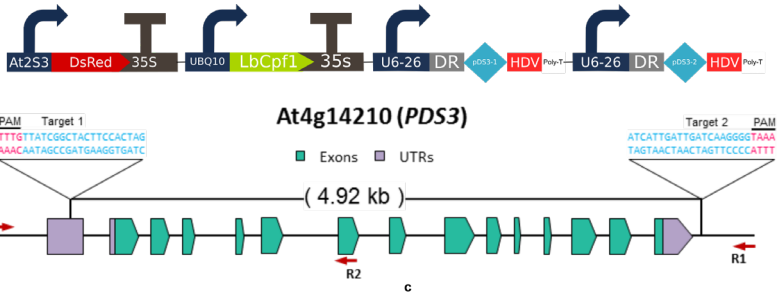
Finally, we assayed LbCas12a activity in Arabidopsis. In this case, we also decided to determine the ability of LbCas12a to produce large deletions in the genome. To this end, we designed a CGEM with the LbCas12a under the transcriptional control of the *A. thaliana* Ubiquitin10 (UBQ10) promoter and two crRNAs directed towards two target sites flanking the Arabidopsis *PDS3* locus (see Fig. 5a). INDELs flanking non-coding regions are unlikely to produce loss-of-function mutations. Therefore, with this design, only editing events producing large deletions in both *PDS3* alleles are expected to generate the albino phenotype. The CGEM construct also contained a DsRED marker gene under the control of a seed-specific promoter to track the presence (DsRED(+)) or absence (DsRED(-)) of the transgene. Despite the strict constraints imposed, we could observe white somatic spots, consistent with bialellic loss-of-function chimeras, in 5 out of 50 T1 stably transformed plants (scored as DsRED(+), Fig. S2b). The analysis of white somatic spots confirmed that they corresponded to different large deletions in the *PDS3* locus (Fig. 5b). Three of the identified chimeric lines (PDS-1, PDS-2 and PDS-3) were grown to the T2 generation and, as expected, some of them (lines PDS-1 and PDS-3) produced offspring showing a completely albino phenotype, consistent with *PDS3* loss-of-function (Fig. 5c). The albino phenotype was present both in DsRED(+) and DsRED(-) offspring (Fig. 5c) indicating that, at this point, *PDS3* mutations segregated independently of the CGEM T-DNA. The presence of loss-of-function deletions in T2 DsRED(-) seedlings, thus regarded as germ-line-associated, was confirmed by PCR (Fig. 5c). Interestingly, an allele with a full deletion of the *PDS3* gene (indicative of concomitant activity in targets 1 and 2) was detected in T2 pools derived from white individuals from line PDS-1, in addition to a partial deletion of the locus. Partial deletions with some additional rearrangements were also detected in pools derived from the T2 generation of lines PDS-2 and PDS-3 (PCR B, Fig. 5c). The structure of deletion-containing PCR bands was analyzed by Sanger sequencing and resolved (Fig. S2c), revealing a number of rearrangements as a result of Cas12a mutagenesis,

predominantly in target 1. In addition to large deletions, small indels ranging from 3 to 34 nt were also identified as expected in green seedling pools using TIDE analysis (Fig. 5d).

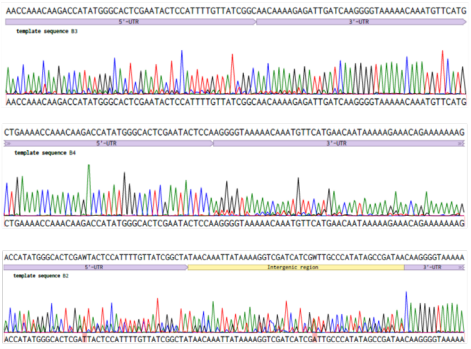
The presence of deletion-containing alleles in DsRED(-) seedlings was clearly indicative of the presence of Cas12a mutations in the germline. An additional confirmation for this germline-association was obtained with a segregation analysis of the T3 seeds from the PDS-1-118 T2 DsRED(-) line. Attending to the number of plants with the recessive albino phenotype, the expected 3:1 segregation hypothesis was initially discarded by X^2 statistic (Fig. S2e). However, the PCR analysis of green T3 seedlings (heterozygous and wt) in the PDS-1-118 line revealed the presence of the loss-of-function deletion in a 1:2 proportion, a segregation which is compatible with germ-line inheritance (Fig. S2e). Combined, these results indicate that the observed *PDS3* deletions are germline-associated, and that the segregation of albino plants is probably distorted by reduced fertility and viability of the homozygous *PDS3* loss-of-function mutation, as suggested previously [35].

Figure 5: LbCas12a is capable to induce large deletions in *A. thaliana*. (a) CGEM used for *A. thaliana* stable transformation and locus chosen to study the ability of LbCas12a to produce large deletions, showing the two target sites on the UTRs flanking the *PDS3* gene. Arrows indicate primers used for the amplifications. **(b)** Alignments of three different complete deletions found in the white spots of the T1 generation visualized with Benchling. The complete gene have been deleted and the UTRs joined. **(c)** Analysis of albino phenotype (*pds* mutants, white seedlings) segregation in T2 seedlings grown in MS plates. Segregation analysis of DsRED(+) and DsRED(-) seeds from the three lines (PDS-1, PDS-2 and PDS-3) and PCR amplification of white or green DsRED(-) seedlings pools from the three different lines. **(d)** TIDE mutation analysis of DsRED(-) and DsRED(+) from pools of green plants from PDS-1, 2, 3 lines. Both, target 1 (T1) and target 2 (T2) were analyzed. "Overall efficiency" and "mutation efficiency" means the same than Fig. 4.

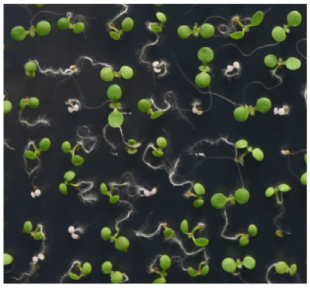
a



b



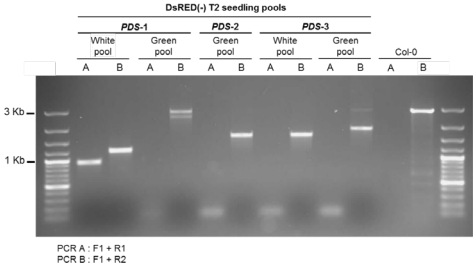
c



	DsRED(-)		DsRED(+)	
	WT	*pds*	WT	*pds*
***PDS*-1**	217	13	430	10
***PDS*-2**	215	0	430	1
***PDS*-3**	200	22	390	50

d

Green Pools	Overall Efficiency	Mutation Efficiency
PDS-1-DsRED(+) T1	92.2	-34 (6.7), -6 (30.2), -4 (46.1)
PDS-1-DsRED(+) T1	56.7	-9 (6.8), -7 (9.8), -3 (35.6)
PDS-1-DsRED(-) T1	72.4	-10 (11.9), -6 (29.6), -4 (13.2)
PDS-1-DsRED(-) T2	22.3	-9 (6.2), -4 (15.4)
PDS-2-DsRED(+) T1	71.9	-15 (6.2), -10 (7.4), -7 (27.3), -4 (15.8), -3 (14.7)
PDS-2-DsRED(+) T2	15.5	-3 (12.3)
PDS-2-DsRED(+) T1	46.6	-17 (8,3), -14 (16.9), -8 (6.6), -4 (14.3)
PDS-2-DsRED(-) T2	23.6	-3 (22.2)
PDS-3-DsRED(+) T1	27.1	-17 (26.2)
PDS-3-DsRED(-) T1	13.9	-17 (13.7)



2.5.A meta-analysis of Cas12a mutagenesis profile shows a high frequency of small-medium size deletions

To get insight into the compared signatures of Cas12a and Cas9, we performed a meta-analysis collecting all TIDE mutagenesis data produced for *N. benthamiana* (stable and transient experiments), tomato and Arabidopsis (stable experiments), (Fig. 6). A total of 272 mutagenesis events were compiled: 137 for Cas12a and 135 for Cas9. To represent the mutation landscape, we assigned an arbitrary value of one to each mutation event detected in a TIDE experiment. Then, we summed all the events with a certain INDEL size and divided it by the number of total events detected in order to calculate the mutation frequency of each INDEL size. The results are plotted in Fig. 6. As observed, the Cas12a deletion profile was different from that of Cas9, showing a tendency to produce larger deletions. In contrast, Cas9 displayed small insertions (+1, +2, and +4 bp) and small deletions more frequently. Summarizing, our meta-analysis illustrates that Cas12a produces a deletion-enriched mutagenesis profile in comparison to that produced by SpCas9.

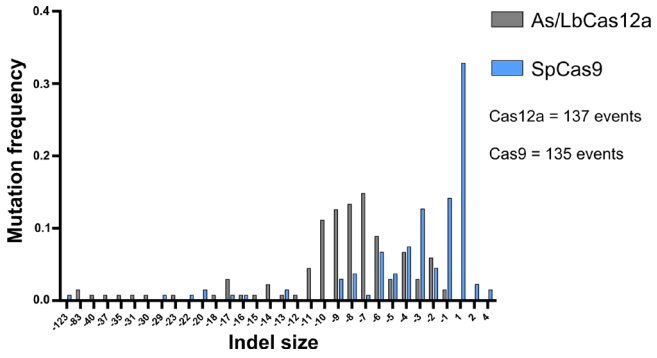


Figure 6: Meta-analysis of *N. benthamiana, S. lycopersicum* and *A. thaliana* mutation events. Compilation of all the TIDE data produced in the laboratory for Cas12a and Cas9 from transient and stable expression experiments of three plant species (*N. benthamiana, S. lycopersicum* and *A. thaliana*) to compare mutagenesis profile of both endonucleases.

2.6. Whole genome sequencing of LbCas12a-mutated *A. thaliana* plants revealed no appreciable off-target effects

An important feature for editing nucleases is the extent of off-target mutagenesis. Usually, this parameter is evaluated by examining candidate off-target loci by PCR. However, recent studies have reported large DNA rearrangements near the target loci, which remain undetected with traditional target-specific PCR analysis [36, 37]. Also, indiscriminate single strand DNA (ssDNA) DNAse activity of Cas12a upon activation by target-specific cleavage has been observed *in vitro* [38, 39].These observations have led to the suggestion that increased mutagenesis rates could occur elsewhere in the genome, particularly in areas with transient ssDNA formation, such as active DNA replication and transcriptional sites. To carefully evaluate the potential off-target effects of LbCas12a, we performed Whole Genome Sequencing (WGS) of eight selected Cas12a-edited Arabidopsis lines.

Four green individuals from the PDS-3 line (PDS-3-111, 113, 114, 115) and four of PDS-1 (PDS-1-117, 118, 119, 123) DsRED(-) T2 plants bearing different mutations at target site 1 including large deletions in heterozygosis, were each re-sequenced at 20X coverage. It is important to note that these plants had segregated Cas12a and therefore observed mutations are necessarily germline-associated. The genotype of the *PDS3* locus for each individual is highlighted in Fig. S2d. As a reference, two DsRED(-) T1 plants (WT-104 106) and a pool of five DsRED(-) T1 plants (including WT-104 and WT-106), in which Cas12a was never present, were also included for resequencing; the pool was sequenced at a deeper coverage (60X) to assess the population variability (named as POOL).

Once all on-target mutations present in the sequenced lines were confirmed by WGS, we proceed with the off-target analysis. The first off-target study consisted in the analysis of homologous sites. Using Cas-OFFinder software [40] set for a maximum of four mismatches, three putative off-target sites for target 1 were found. Software tools failed to detect SNPs, INDELs and structural variations (SVs) at these positions in any of the analyzed plants (Table 1a).

Table 1: off-target effects of LbCas12a in DsRED(-) T2 PDS-1 and PDS-3 *A.thaliana* mutants.

(a) Off-target effects found at three different off-target sites predicted by Cas-OFF finder software. The PAM sequence is highlighted, and the mismatches are shown in lowercase.

Off-target sequence	SNPs	INDELs	SVs
TTTCaTATgaGCTACTTtCACTAG	0	0	0
TTTATTATgaGCTACTTCCACaAG	0	0	0
TTTATTtTtGGCTACTcCCAtTAG	0	0	0

(b) Variations found in the WGS analysis. The numbers outside the parenthesis indicates the variations found when comparing with the reference genome. Inside the parenthesis events obtained comparing each edited plant vs WT samples.

DsRED(-) plant	SNPs	INDELs	SVs
PDS-3-111	1577 (24)	460 (30)	5 (3)
PDS-3-113	1535 (25)	474 (35)	6 (4)
PDS-3-114	1549 (24)	469 (36)	3 (1)
PDS-3-115	1566 (23)	471 (31)	2 (0)
PDS-1-117	1558 (8)	442 (7)	2 (1)
PDS-1-118	1569 (9)	441 (5)	8 (8)
PDS-1-119	1575 (4)	450 (23)	3 (1)
PDS-1-123	1579 (11)	452 (13)	3 (1)
WT-104	1533	437	2
WT-106	1565	423	2

SNPs were analyzed with GATK-HC, VARSCAN and FREEBAYES. INDELs with GATK-HC, VARSCAN, PINDEL. Structural Variations with (SVs) with Delly.

Next, we searched for unspecific enrichment of mutation rates in the Cas12a-"treated" plants as compared to their relatives not subjected to Cas12a activity. Table 1b shows the results obtained with this approach. The columns SNPs, INDELs and SVs show the total number of consensus variations found in each T2 plant compared with the reference genome (as detected by all programs used in the analysis, Fig. S4). We also traced possible background mutations of PDS plants comparing them against WT and POOL samples. The data obtained discards massive increases in "indiscriminate" mutagenesis possibly associated with Cas12a activity. Note for instance that in line PDS-1-117, only 16 of all detected variations could not be traced back to the genomes of its Cas12a-free T1 relatives (8 out of 1558 SNPs, 7 out of 442 INDELs and 1 out of 2 SVs). This low number of putative "new" mutations is fully compatible with development-associated spontaneous events.

Interestingly, the WGS also revealed multiple structural variations surrounding the target site in the PDS-1-118 line, which would have remained undetected by classical PCR analysis. The reads observed in this line can be explained as the result of a duplication of approximately 13Kb and two large deletions of 2 and 4Kb. (Fig. S5). This event was not found in sibling lines, clearly suggesting an association with the Cas12a signature.

3. Discussion

Plant genome engineering facilitated by RGEN is expected to play an important role in the adaptation of agriculture to XXI century challenges. After the first wave of tools based on Cas9, which was shown capable of producing edited mutants in virtually every species tested, a second generation of refined tools is being progressively incorporated to the genome editing toolbox. These new tools are emerging in parallel with a number of modular cloning methods that facilitate the customization of often multiplexed editing constructs [41-44]. Thus, following the same design principles adopted earlier to integrate CRISPR/Cas9 into GB modular cloning [30], we adapted Cas12a tools to the so-called phytobrick standard, adopting both crRNA and Cas12a TUs as level 1 structures to maximize the exchangeability while preserving the combinatorial potential.

Mutagenesis efficiency was evaluated using transient *N. benthamiana* assays, based on our previously described detailed standard procedures for the analysis of Cas9-based mutagenesis (see [30]). We chose the T7E1 assay for the estimation of the mutagenesis efficiency since it is a cost-effective method. It should be noted that this procedure has a lower sensitivity than RE polymorphism analysis, therefore actual efficiencies below a certain threshold (approximately 5%) should not be discarded for some targets. Also, it has been previously described that T7E1 can underestimate mutagenesis rates in highly efficient targets or in DNA populations showing low mutation variability (e.g. homozygous mutations). However, our analysis is unlikely to be affected by this T7E1 bias as they were performed on complex cell populations with maximum estimated efficiency rates of 30%, a range of activities in which the T7E1 assay can be considered a reliable estimation method [45]. Our results, far from showing a clear front-runner nuclease, indicated a strong target-dependent efficiency. Interestingly, all loci analyzed were efficiently mutated with at least one RGEN indicating general accessibility of the

selected loci. On the contrary, all three RGENs produced undetectable mutagenesis levels in at least one selected locus. In general, LbCas12a and SpCas9 showed more reliable results than AsCas12a, producing detectable mutation rates in 7 and 6 of the 8 assayed targets respectively, while AsCas12a produced detectable mutation rates only in three targets. The average and maximum mutagenesis rates were displayed by LbCas12a, indicating that this enzyme can be a fair competitor or even outperform SpCas9 for certain applications. Our findings are consistent with previous publications in mammalian cells reporting high variability among the evaluated targets, with SpCas9 as the most robust of the three RGEN in Kim D. et al., 2016 [19] and LbCas12a in Kleinstiver BP et al., 2015 [46]. In plants, Endo A. et al., 2016 [22], targeted four different positions of the *NtSTF1* gene with FnCas12a, but only one crRNA produced mutations.

In our hands, SpCas9 showed lower dispersion in the scored activities, which grouped between 5% and 20%, whereas LbCas12a showed more variable results. This observation might reflect the larger size of the SpCas9 sgRNA (~100 bp) as compared to AsCas12a and LbCas12a crRNAs (~40 bp). This larger size might "protect" the SpCas9 sgRNA against disruption of the secondary structure towards unproductive/unstable conformations that prevent mutagenesis [47]. In contrast, the smaller size of the Cas12a crRNAs might facilitate the adoption of unproductive/unstable conformations, explaining the broader range of editing efficiencies. We explored the possibility of creating improved structures by engineering DRs with a simplified loop of three nucleotides, trying to minimize the interactions between the DR and the PS that could impair crRNA function. Although certain improvements in mutagenesis activities were obtained with synthetic DR loops when combined with AsCas12a, in general the strategy failed to obtain higher efficiencies for LbCas12a, or to "turn on" inactive combinations as exemplified with *FT* and AsCas12a. Despite this, our data serves to

demonstrate that engineering of DR loop can be used to modulate Cas12a efficiency, producing a range of activities that can be used in transcriptional regulation approaches, for example.

Although efficiency studies using transient assays are highly informative, they need to be confirmed by generating stably edited plants. We first assayed *N. benthamiana* as a way to corroborate the editing efficiencies observed in transient assays. LbCas12a efficiency for the *XT1* locus was surprisingly high compared with a previous report in a solanaceous plant (tobacco) using FnCas12a, where no bialellic mutants were recovered [22]. In contrast, we could not recover mutant lines from AsCas12a transformants. LbCas12a has also been reported to produce high mutation rates with a high ratio of bialellic plants in rice stable transformation [22-25, 48]. Together, there is now enough data to conclude that the robustness of LbCas12a is a general feature in plants. Regarding the comparison between LbCas12a and SpCas9 for stable genome editing, recent experiments in maize showed higher editing rates with SpCas9 as compared to LbCas12a targeted to the *glossy2* locus [32], in line with the data obtained in our tomato experiment. However, these observations reflect only the results obtained with a very limited number of genome targets and therefore it would be premature to draw conclusions from these findings, as our transient expression data show a strong loci-dependent effect on RGENs activities, with LbCas12a producing higher mutagenesis rates in the majority of the loci assayed. It is also possible that *in vitro* culture conditions during stable transformation could account for some of the differences observed. Recently, a strong temperature dependency for AsCas12a has been reported, showing a severe decrease in activity below 37ºC [49]. This effect could explain in part the lower performance of AsCas12a in stable transformation, as the transformation/regeneration processes take place at 22-25ºC. This would discard AsCas12a in practice for plant editing procedures involving *in vitro* culture, unless transformation

conditions are adapted to maximize editing (rather than transformation) efficiency.

The meta-analysis plot shows that SpCas9 tends to produce small indels of few nucleotides in contrast with the Cas12a enzymes tested, which induce a broader range of deletions but no insertions. Deletions are often associated with microhomology-mediated end joining (MMEJ) repair [50], which depends on the specific target sequence that marks the formation of the microhomologies. It should be taken into account that the TIDE data used in this meta-analysis derives from a heterogenic collection of *N. benthamiana*, *S. lycopersicum* and *A. thaliana* targets. For this reason, biases due to the mutation signature of certain targets (determined by the local micro-homologies) or by the organism considered, can be discarded. In addition, the reported profiles are consistent with the literature, as other authors have shown a prevalence of deletions with Cas12a in comparison with SpCas9 in mammalian cells [19] and plants [22]. The distinctive mutation signature of Cas12a has interesting functional implications, considering for example that larger deletions are more prone to generate loss-of-function mutants and remove regulatory operators in promoter regions, two features that can be exploited in breeding practices [51]. Furthermore, whereas DSB in Cas9 takes place at a position proximal to the PS sequence, Cas12a cleaves at a distal position, thus allowing target conservation after cleavage, which allegedly promotes larger deletions by MMEJ or gene insertions via homology direct repeat (HDR) [49, 52]. Whether the Cas12a signature will enhance gene targeting (GT) efficiency in plants is a possibility that remains to be tested, either alone, or in combination with other GT-enhancing methods, such as geminivirus replicons [53].

Finally, we searched for putative off-target effects not only in related homology sites, but also elsewhere in the genome using WGS analysis. We could not detect homologous off-targets in any of the plants under analysis. Moreover, we do not find evidences in any of the analyzed lines of enhanced genome-wide mutagenesis to occur as a result of Cas12a activation as suggested previously [38, 39]. Just the opposite, WGS data showed that all analyzed lines had variations with the reference genome and with their own relatives in the range of those observed in Cas12a-free plants. Therefore, we can conclude that Cas12a can generate transgene-free edited plants (e.g. plant PDS-1-117) whose levels of "new" mutations not present in the background are negligible and indistinguishable from spontaneous mutations caused during development. A similar conclusion was obtained recently in rice Tang X. et al., 2018 [54]. Nevertheless, we did observe a relatively large distortion of the alignment of Illumina reads compatible with a duplication event in the vicinity of the target site in at least one plant (PDS-1-118). This structural change would have remained undetected using only PCR-based off-target detection methods. The same line showed an increased number of SVs. These types of events, which are not unusual in traditional plant breeding practices, do not necessarily comprise breeding value to the resulting variety. However, given the strict scrutiny to which genome editing procedures are subjected, we conclude that re-sequencing of "elite" lines produced with Cas12a is advisable to discard lines in which rearrangements could occur.

Taken together, our data show that LbCas12a is an effective RNA-guided endonuclease in a broad number of plant species, it is amenable for modular cloning and multiplexing, and shows efficiencies comparable with classical SpCas9 with similarly low off-target effects and a characteristic tendency to produce larger deletions.

4. Experimental procedures

4.1. GBparts construction

GBparts employed in this study were created through the domestication strategy described in [28]. Plasmids pY010 (pcDNA3.1-hAsCpf1, Addgene plasmid # 69982) and pY016 (pcDNA3.1-hLbCpf1, Addgene plasmid # 69988) kindly provided by Feng Zhang laboratory, served as a template for the construction of GB1338 and GB1339 by PCR amplification, using the Phusion High Fidelity DNA polymerase (Thermo Scientific). GB1442 and GB1443 were also obtained by PCR amplification of GB1001. An amount of 40 ng of PCR products were subsequently cloned into pUPD2 plasmid to create the above mentioned GBparts through a BsmBI restriction-ligation reaction. Partially complementary ultramers with sticky ends were used to domesticate GB1444. Separate ultramers were resuspended to a final concentration of 1 uM, then 5 uL of each were mixed and incubated for 30 min at room temperature to facilitate the hybridization. Finally, 1 uL of this mix was used to set up the domestication reaction following the same protocol than for the PCR products. Once cloned, GBparts were verified by restriction enzyme (RE) analysis and confirmed by sequencing. All the GBparts are listed in the Table S1.

4.2. Guide RNA assembly on level 1

The design of the guide RNA and the assembly of the expression cassettes on level 1 were performed as described in [30]. On the particular case of Cas12a, the BsaI-mediated restriction-ligation reaction included a complementary pair of oligos of the target (Supplementary Table S2) flanked by 4 nucleotide overhangs complementary to GB1442 and GB1443. Guide RNA constructs were confirmed by RE-analysis and subsequent sequencing. The produced constructs are detailed on the Table S1.

4.3. Cloning in α and Ω-level destination vectors

Level 1 assemblies were performed through Golden Gate-like multipartite BsaI restriction-ligation reactions to obtain TUs from basic domesticated level 0 parts. Similarly, several TUs were combined on level >1 with bipartite BsmBI or BsaI-mediated reactions to create modules. These assembly protocols are detailed in [42]. The generated constructs are included in the Table S1. The sequences are available at http://gbcloning.upv.es/search/feature with the GB database ID.

4.4. *Nicotiana benthamiana* transient assays

Transient expression experiments were performed to test the mutagenesis efficiency at several loci as described in [30, 42]. Each locus was assessed separately for the three RGEN (SpCas9, AsCas12a and LbCas12a). To this end we mixed equal volumes of *Agrobacterium* cultures of the P19 suppressor of silencing, the RGEN and the crRNA/sgRNA. These cultures were first grown from glycerol stock for two days to saturation, then 10 uL were sub-cultivated for 16h. Next, the cultures were pelleted, resuspended in the agroinfiltration (buffer 10 mM MES, pH 5.6, 10 mM MgCl2, and 200 µM acetosyringone) and adjusted to an optical density of 0.2 at 600 nm (estimated 8 active T-DNA copies per cell [42]. Finally, the three cultures (P19, RGEN and crRNA/sgRNA) were mixed at equal parts to prepare the agroinfiltration mixture. Three independent samples (plants of 4-5 weeks-old grown in a stable condition of 24 °C (light)/20 °C (darkness) with a 16-h-light/8-h-dark photoperiod) were infiltrated to assess the effect of each RGEN at each locus. Three consecutive leaves (second to fifth) were infiltrated in each plant. Five days post-infiltration one sample per infiltrated plant was collected. Each sample consisted in 6 pooled leaf-discs, 2 per infiltrated leaf, collected with a 0.5 cm cork-borer (approximately 150 mg of tissue). Immediately the samples were frozen in liquid nitrogen. Control plants were infiltrated using the same mixture than the

targeted locus but with an unrelated crRNA. Samples were grinded with a Retsch Mixer Mill MM400 for 1 min at 30 Hz and stored at -80ºC for subsequent genomic DNA extraction (gDNA).

4.5. *Nicotiana benthamiana* stable transformation

N. benthamiana stable transformation was performed following a modification of transformation and *in vitro* regeneration of leaf-discs method [55, 56]. All *in vitro* steps were carried out in a long day growth chamber (16 h light/ 8 h dark, 24ºC, 60-70% humidity, 250 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Samples for genotyping were collected once the plants were sufficiently developed to harvest 150 mg of tissue, which was frozen in liquid nitrogen and stored at -80ºC until extraction of gDNA.

4.6. *Solanum lycopersicum* stable transformation

Tomato transformation was performed following an adapted protocol of [57]. Growth conditions and sampling was conducted as described for *N. benthamiana* transformats above.

4.7. *Arabidopsis thaliana* stable transformation

Wild-type Col–0 plants were transformed by the floral-dip method [58], with minor modification: 1 minute dipping into a solution (Sucrose 5% + 0.2 ml Silwet77/liter) containing *Agrobacterium tumefaciens*. Fluorescent seeds, containing the transgene, were identified in a Leica microscope DMS1000.

4.8. Mutagenesis detection of on-target sites

Genomic DNA was extracted from transient and stable expression experiment samples following the C-TAB protocol [59]. The obtained gDNA was used for PCR amplification of the selected locus using MyTaq™ DNA Polymerase (Bioline) and a pair of primers flanking the targeted sites (Table S2). PCR product was confirmed by 1% gel electrophoresis and purified employing QIAquick PCR purification kit (QIAGEN), following the

manufacturer's protocol. The purified PCR product was utilized to detect mutagenesis by three well established methods: T7E1 assay, RE-analysis and TIDE. For the T7E1 assay (New England Biolabs), 250 ng of the PCR product was submitted to a denaturation-reannealing process using the thermocycler (95 ºC for 5 min; 95 to 85 ºC at 2 ºC/sec: 85 to 25 ºC at 0.1 ºC/sec; 4 ºC hold). The digestion product was visualized in a 2% electrophoresis gel ran for 45 min (Fig. S3b-d). The mutagenesis efficiency was estimated measuring the intensity of the non-digested and the digested bands with ImageJ, and the data obtained was treated as described in [60]. RE-analysis was set up with 500 ng of PCR product and *Eco*RI (Fermentas) to assess mutagenesis at the *XT1* locus. Mutated samples confirmed by either method were sequenced and analyzed by TIDE to characterize the indel size and the corresponding efficiency. All experimental points represent the average of three independent samples.

4.9. Mutagenesis detection of off-target sites

Genomic DNA was extracted from 100 mg of *A. thaliana* rosetta leaves following the C-TAB protocol [59]. Paired-end reads were obtained by sequencing the extracted DNA with Illumina NextSeq 550 platform. Resulting raw sequences are available at the NCBI Sequence Read Archive (SRA). Read quality was assessed with FASTQC [61] and sequences were cleaned using TRIMMOMATIC [62], keeping reads with an average quality value of 29 in phred 33 scale and a minimum length of 50 nt. Then, these reads were mapped against *A. thaliana* reference genome TAIR10 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.4/) using read mapper BWA-MEM [63]. Mapping statistics can be found in Table S4a. Quality scores were recalibrated using GATK best practices [64]. For this purpose and for further steps in this pipeline a collection of SNPs and INDELs was downloaded from NCBI dbSNP database (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/

archive/arabidopsis_3702/VCF/). SNPs were detected using GATK-HC v4 [64], VARSCAN2 [65] and FREEBAYES [66]. For the PDS and the WT samples a merged set was made using SNPs detected by all three programs using Bcftools [67]. For the POOL sample, GATK and FREEBAYES were used to produce the set of SNPs. INDELs were detected using GATK-HC, VARSCAN2 and PINDEL [68]. Those INDELs detected in two out of three programs were used to construct a set of INDELs using Bcftools and GATK tools. Due to the ploidy of the POOL sample, we omitted the use of VARSCAN2 to the previously described detections. The SVs were detected using Delly [69]. Filtering steps were applied in SNPs, INDELs, and SVs detection using the parameters specified in Table S4b. Using these sets all the SNPs, INDELs and SVs were numbered using Bcftools and VCFtools [70]. For the PDS samples, variations not found in the WT samples nor in the POOL were taken. These variations found were checked using IGV. Potential off-targets were detected using Cas-OFFinder [40] allowing a maximum of 4 nt mismatches searching variations in these regions using BEDtools [71], extending the off-target region 50nt in each side and reviewed with IGV.

5. References

1. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR provides acquired resistance against viruses in prokaryotes**. *Science* 2007, **315**(5819):1709-1712.
2. Marraffini LA: **CRISPR-Cas immunity in prokaryotes**. *Nature* 2015, **526**(7571):55.
3. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH: **An updated evolutionary classification of CRISPR–Cas systems**. *Nature Reviews Microbiology* 2015, **13**(11):722.
4. Hsu PD, Lander ES, Zhang F: **Development and applications of CRISPR-Cas9 for genome engineering**. *Cell* 2014, **157**(6):1262-1278.
5. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**(6096):816-821.
6. Lyznik LA, Gordon-Kamm WJ, Tao Y: **Site-specific recombination for genetic engineering in plants**. *Plant Cell Reports* 2003, **21**(10):925-932.
7. Bortesi L, Fischer R: **The CRISPR/Cas9 system for plant genome editing and beyond**. *Biotechnol Adv* 2015, **33**(1):41-52.
8. Yu Z, Ren M, Wang Z, Zhang B, Rong YS, Jiao R, Gao G: **Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila**. *Genetics* 2013, **195**(1):289-291.
9. Tabebordbar M, Zhu K, Cheng JK, Chew WL, Widrick JJ, Yan WX, Maesner C, Wu EY, Xiao R, Ran FA: **In vivo gene editing in dystrophic mouse muscle and muscle stem cells**. *Science* 2016, **351**(6271):407-411.
10. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F: **Genome engineering using the CRISPR-Cas9 system**. *Nat Protoc* 2013, **8**(11):2281-2308.
11. Soda N, Verma L, Giri J: **CRISPR-Cas9 based plant genome editing: Significance, opportunities and recent advances**. *Plant Physiology and Biochemistry* 2017.
12. Yin K, Gao C, Qiu J-L: **Progress and prospects in plant genome editing**. *Nature plants* 2017, **3**(8):17107.

13. Khan MHU, Khan SU, Muhammad A, Limin H, Yang Y, Chuchuan F: **Induced Mutation and Epigenetics Modification in Plants for crop improvement by targeting CRISPR/Cas9 technology**. *Journal of cellular physiology* 2017.
14. Kumar J, Gupta DS, Gupta S, Dubey S, Gupta P, Kumar S: **Quantitative trait loci from identification to exploitation for crop improvement**. *Plant Cell Reports* 2017:1-27.
15. Zsögön A, Cermak T, Voytas D, Peres LEP: **Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: case study in tomato**. *Plant Science* 2016.
16. Wang Y, Liu X, Ren C, Zhong G-Y, Yang L, Li S, Liang Z: **Identification of genomic sites for CRISPR/Cas9-based genome editing in the Vitis vinifera genome**. *BMC plant biology* 2016, **16**(1):96.
17. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A *et al*: **Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system**. *Cell* 2015, **163**(3):759-771.
18. Zaidi SS, Mahfouz MM, Mansoor S: **CRISPR-Cpf1: A New Tool for Plant Genome Editing**. *Trends Plant Sci* 2017, **22**(7):550-553.
19. Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS: **Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells**. *Nat Biotechnol* 2016, **34**(8):863-868.
20. Kim HK, Song M, Lee J, Menon AV, Jung S, Kang YM, Choi JW, Woo E, Koh HC, Nam JW *et al*: **In vivo high-throughput profiling of CRISPR-Cpf1 activity**. *Nat Methods* 2017, **14**(2):153-159.
21. Wu WY, Scott DA, Severinov K, van der Oost J, Zhang F: **Multiplex gene editing by CRISPR-Cpf1 through autonomous processing of a single crRNA array**. *Nat Biotechnol* 2017, **35**(1):31-34.
22. Endo A, Masafumi M, Kaya H, Toki S: **Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida**. *Sci Rep* 2016, **6**:38169.
23. Yin X, Biswal AK, Dionora J, Perdigon KM, Balahadia CP, Mazumdar S, Chater C, Lin HC, Coe RA, Kretzschmar T *et al*: **CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental gene EPFL9 in rice**. *Plant Cell Rep* 2017, **36**(5):745-757.
24. Tang X, Lowder LG, Zhang T, Malzahn AA, Zheng X, Voytas DF, Zhong Z, Chen Y, Ren Q, Li Q *et al*: **A CRISPR-Cpf1 system for**

efficient genome editing and transcriptional repression in plants. *Nat Plants* 2017, **3**:17018.

25. Xu R, Qin R, Li H, Li D, Li L, Wei P, Yang J: **Generation of targeted mutant rice using a CRISPR-Cpf1 system**. *Plant Biotechnol J* 2017, **15**(6):713-717.
26. Wang M, Mao Y, Lu Y, Tao X, Zhu JK: **Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System**. *Mol Plant* 2017, **10**(7):1011-1013.
27. Kim H, Kim ST, Ryu J, Kang BC, Kim JS, Kim SG: **CRISPR/Cpf1-mediated DNA-free plant genome editing**. *Nat Commun* 2017, **8**:14406.
28. Sarrion-Perdigones A, Falconi EE, Zandalinas SI, Juarez P, Fernandez-del-Carmen A, Granell A, Orzaez D: **GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules**. *PLoS One* 2011, **6**(7):e21622.
29. Patron NJ, Orzaez D, Marillonnet S, Warzecha H, Matthewman C, Youles M, Raitskin O, Leveau A, Farré G, Rogers C: **Standards for plant synthetic biology: a common syntax for exchange of DNA parts**. *New Phytologist* 2015, **208**(1):13-19.
30. Vazquez-Vilar M, Bernabe-Orts JM, Fernandez-Del-Carmen A, Ziarsolo P, Blanca J, Granell A, Orzaez D: **A modular toolbox for gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard**. *Plant Methods* 2016, **12**:10.
31. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS *et al*: **Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array**. *Nat Biotechnol* 2017, **35**(1):31-34.
32. Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK: **Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize**. *Plant biotechnology journal* 2019.
33. Lorenz R, Bernhart SH, Zu Siederdissen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL: **ViennaRNA Package 2.0**. *Algorithms for Molecular Biology* 2011, **6**(1):26.
34. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction**. *Nucleic acids research* 2003, **31**(13):3406-3415.
35. Qin G, Gu H, Ma L, Peng Y, Deng XW, Chen Z, Qu L-J: **Disruption of phytoene desaturase gene results in albino and dwarf**

phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell research* 2007, **17**(5):471.

36. Kosicki M, Tomberg K, Bradley A: **Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements**. *Nature biotechnology* 2018, **36**(8):765.
37. Adikusuma F, Piltz S, Corbett MA, Turvey M, McColl SR, Helbig KJ, Beard MR, Hughes J, Pomerantz RT, Thomas PQ: **Large deletions induced by Cas9 cleavage**. *Nature* 2018, **560**(7717):E8.
38. Li S-Y, Cheng Q-X, Liu J-K, Nie X-Q, Zhao G-P, Wang J: **CRISPR-Cas12a has both cis-and trans-cleavage activities on single-stranded DNA**. *Cell research* 2018, **28**(4):491.
39. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, Doudna JA: **CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity**. *Science* 2018, **360**(6387):436-439.
40. Bae S, Park J, Kim J-S: **Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases**. *Bioinformatics* 2014, **30**(10):1473-1475.
41. Sarrion-Perdigones A, Vazquez-Vilar M, Palaci J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology**. *Plant Physiol* 2013, **162**(3):1618-1631.
42. Vazquez-Vilar M, Quijano-Rubio A, Fernandez-Del-Carmen A, Sarrion-Perdigones A, Ochoa-Fernandez R, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data**. *Nucleic Acids Res* 2017, **45**(4):2196-2209.
43. Peccoud J, Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S: **A Modular Cloning System for Standardized Assembly of Multigene Constructs**. *PLoS ONE* 2011, **6**(2).
44. Engler C, Gruetzner R, Kandzia R, Marillonnet S: **Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes**. *PloS one* 2009, **4**(5):e5553.
45. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM: **A survey of validation strategies for CRISPR-Cas9 editing**. *Scientific reports* 2018, **8**(1):888.
46. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR *et al*: **Engineered CRISPR-**

Cas9 nucleases with altered PAM specificities. *Nature* 2015, **523**(7561):481-485.

47. Xie K, Zhang J, Yang Y: **Genome-wide prediction of highly specific guide RNA spacers for CRISPR–Cas9-mediated genome editing in model plants and major crops**. *Molecular plant* 2014, **7**(5):923-926.
48. Wang M, Mao Y, Lu Y, Tao X, Zhu J-k: **Multiplex gene editing in rice using the CRISPR-Cpf1 system**. *Molecular Plant* 2017.
49. Moreno-Mateos MA, Fernandez JP, Rouet R, Vejnar CE, Lane MA, Mis E, Khokha MK, Doudna JA, Giraldez AJ: **CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing**. *Nature communications* 2017, **8**(1):2024.
50. Bae S, Kweon J, Kim HS, Kim J-S: **Microhomology-based choice of Cas9 nuclease target sites**. *Nature methods* 2014, **11**(7):705-706.
51. Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK: **Cis-regulatory elements used to control gene expression in plants**. *Plant Cell, Tissue and Organ Culture (PCTOC)* 2016, **127**(2):269-287.
52. Tóth E, Weinhardt N, Bencsura P, Huszár K, Kulcsár PI, Tálas A, Fodor E, Welker E: **Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells**. *Biology direct* 2016, **11**(1):46.
53. Cermak T, Baltes NJ, Cegan R, Zhang Y, Voytas DF: **High-frequency, precise modification of the tomato genome**. *Genome Biol* 2015, **16**:232.
54. Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X: **A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice**. *Genome biology* 2018, **19**(1):84.
55. McCormick S, Niedermeyer J, Fry J, Barnason A, Horsch R, Fraley R: **Leaf disc transformation of cultivated tomato (L. esculentum) using Agrobacterium tumefaciens**. *Plant Cell Reports* 1986, **5**(2):81-84.
56. Horsch R, Fry J, Hoffman N, Eichholtz D, Rogers Sa, Fraley R: **A simple and general method for transferring genes into plants**. *Science* 1985, **227**:1229-1232.
57. Ellul P, Garcia-Sogo B, Pineda B, Rios G, Roig L, Moreno V: **The ploidy level of transgenic plants in Agrobacterium-mediated transformation of tomato cotyledons (Lycopersicon esculentum**

L. Mill.) is genotype and procedure dependent. *Theoretical and Applied Genetics* 2003, **106**(2):231-238.

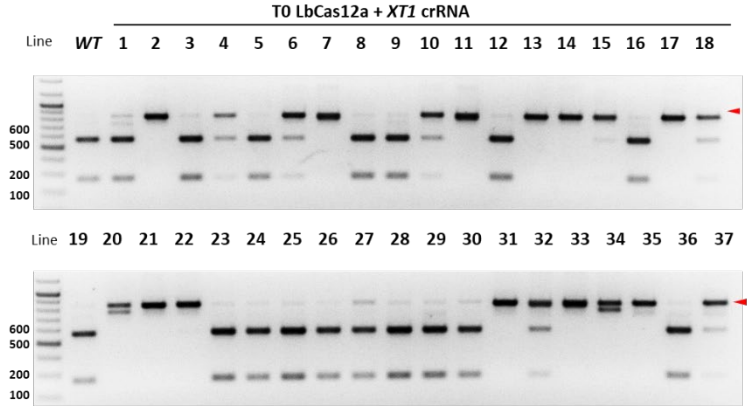
58. Clough SJ, Bent AF: **Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana**. *The plant journal* 1998, **16**(6):735-743.
59. Murray MG, Thompson WF: **Rapid isolation of high molecular weight plant DNA**. *Nucleic acids research* 1980, **8**(19):4321-4326.
60. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ: **A rapid and general assay for monitoring endogenous gene modification**. *Engineered Zinc Finger Proteins: Methods and Protocols* 2010:247-256.
61. Andrews S: **FastQC: a quality control tool for high throughput sequence data**. 2010.
62. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
63. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM**. *arXiv preprint arXiv:13033997* 2013.
64. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome research* 2010.
65. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: **VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing**. *Genome research* 2012.
66. Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing**. *arXiv preprint arXiv:12073907* 2012.
67. Li H: **A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data**. *Bioinformatics* 2011, **27**(21):2987-2993.
68. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z: **Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads**. *Bioinformatics* 2009, **25**(21):2865-2871.
69. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO: **DELLY: structural variant discovery by integrated paired-end and split-read analysis**. *Bioinformatics* 2012, **28**(18):i333-i339.

70. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156-2158.
71. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.

6. Supplementary information

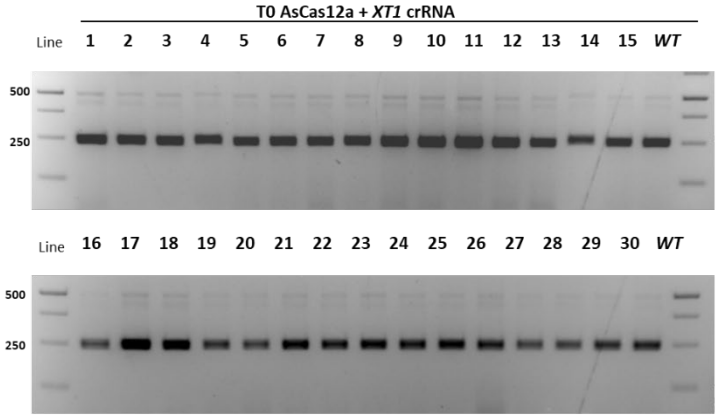
Figure S1: Genotyping results of the *N. benthamiana* transformed lines with the CGEM targeting *XT1* locus.

a) Genotyping of LbCas12a CGEM lines.



20/37 = 54% mutated
12/20 = 60% bialelic

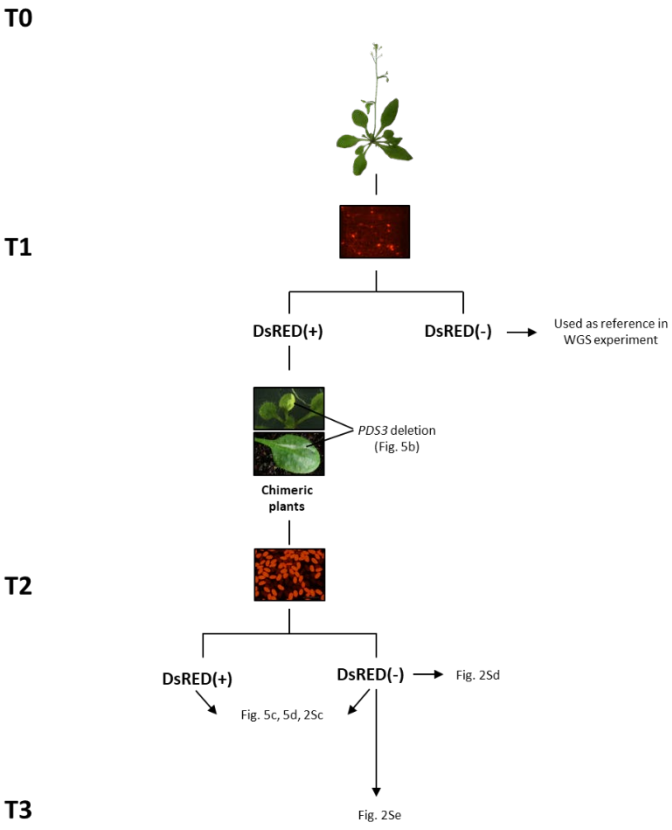
b) Genotyping of AsCas12a CGEM lines.



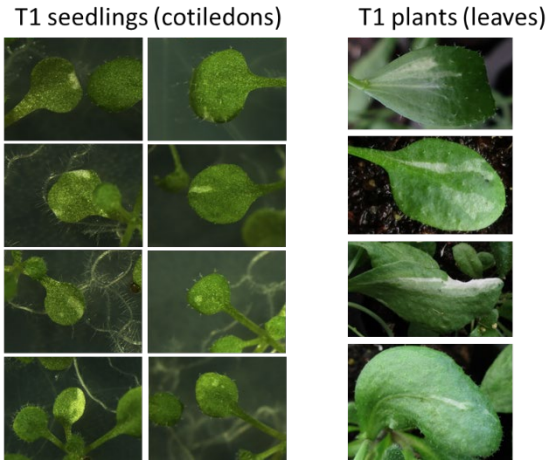
0 mutated lines

Figure S2: Summary of the information regarding *A. thaliana* experiments.

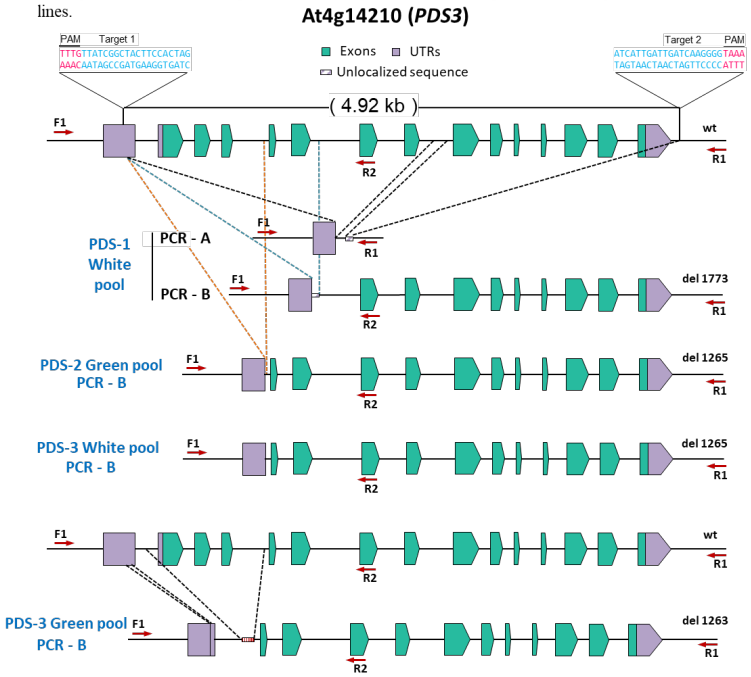
a) Pipeline followed in the experiments with *A. thaliana*.



b) White spots found in the cotiledons (left) and leaves (right) of DsRED(+) T1 lines compatible with *PDS3* deletion.



c) Large deletions and re-arragements found in pools of white and green T2 seedlings of PDS-1,2,3 lines.



d) Mutations found in DsRED(-) T2 PDS-1 and PDS-3 plants. The highlighted plants were selected for Whole Genome Sequencing to perform the off-targets experiments. "Del" stands for deletion, "ins" for insertion. When heterozygous both alleles are separated by slash.

	Mutations	
DsRED(-) Plants	**Target 1**	**Target 2**
PDS-1-32	del6	wt
PDS-1-33	wt/del8	wt
PDS-1-34	del4/del6	wt
PDS-1-35	del6	wt
PDS-1-38	del4	wt/del3
PDS-1-40	wt/del8	wt/ins22
PDS-1-41	del81	wt/del3
PDS-1-55	del4/del6	wt/del3
PDS-1-116	del4/del7	wt
PDS-1-117	del11/del34	wt
PDS-1-118	del5/del1773	wt
PDS-1-119	del6	wt
PDS-1-120	del10/del42	wt
PDS-1-121	wt/del17	wt
PDS-1-122	wt/del9	wt
PDS-1-123	del7/del10	wt
PDS-3-56	wt	wt
PDS-3-59	wt/del1265	wt
PDS-3-60	wt/del17	wt
PDS-3-61	del4/del1265	wt
PDS-3-62	wt	wt
PDS-3-63	wt/del17	wt
PDS-3-64	wt/del1265	wt
PDS-3-65	wt	wt
PDS-3-108	del7/del17	wt/del3
PDS-3-109	wt	wt
PDS-3-110	del7/del17	wt
PDS-3-111	wt/del1265	wt
PDS-3-112	wt	wt
PDS-3-113	wt/del1265	wt
PDS-3-114	wt/del17	wt
PDS-3-115	del17	wt

e) Segregation analysis of the DsRED(-) T3 seeds. Table 1 shows the green:albino seedlings ratio and Table 2 the no-deletion:deletion ratio.

Table 1: green:albino ratio	Phenotype seeds T3			Expected ratio 3:1		χ^2	Hypothesis
Plant T2	Green	White	No germinated	Green	White	*p-value*	3:1
PDS-1-111	85	13	3	74	25	0.0073	Discarded
PDS-1-113	85	8	0	70	23	0.0003	Discarded
PDS11-118	116	20	4	102	34	0.0056	Discarded

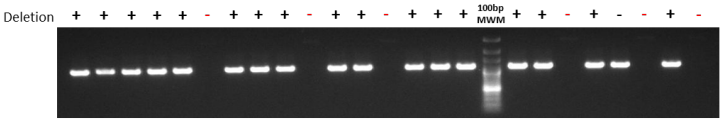
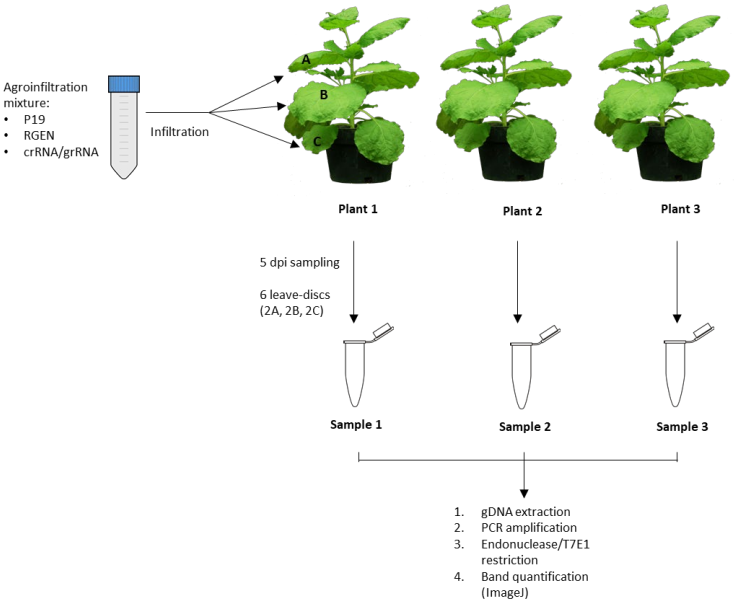


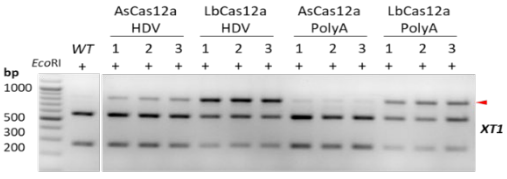
Table 2: no-deletion:deletion ratio	Seeds T3 - Presence Large Deletion		Expected ratio 1:2		χ^2	Hypothesis
Plant T2	No presence	Presence	No presence	Presence	*p-value*	1:2
PDS-1-118	6	18	8	16	0.3865	Approved

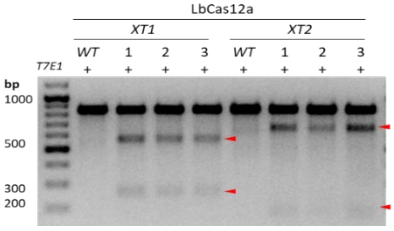
Figure S3: Summary of the *N. benthamiana* transient experiments.

a) Experimental workflow followed to the estimation of the mutagenesis efficiency for a single locus. An agroinfiltration mixture containing the P19 suppressor of silencing, the RGEN and the crRNA/gRNA was infiltrated in three different well stablished leaves (A, B, C) of three biological replicates (Plant 1-3). Five days post infiltration 6 discs per plant were collected (2 discs from A, 2 from B and 2 from C) to obtain three independent replicates (Sample 1-3). Finally, following the gDNA extraction, the locus of interest was amplified by PCR. The product of the PCR was purified and submitted to a T7E1 assay to quantify the undigested and the digested bands using ImageJ. The final efficiency is calculated as the mean of the three replicates.

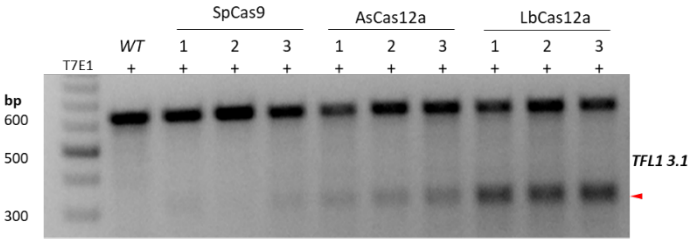


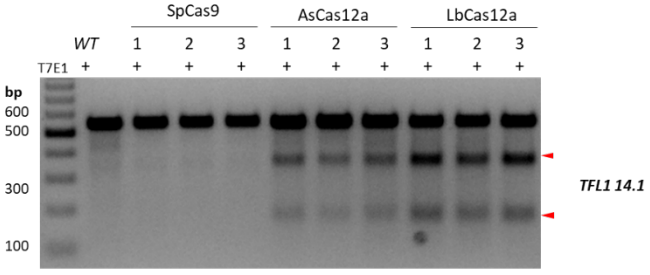
b) *Eco*RI digestion of the HDV and polyA crRNAs comparation (up) and T7E1 assay of multiplexing constructs used on Figure 1b and 1c for indel frequency estimation.

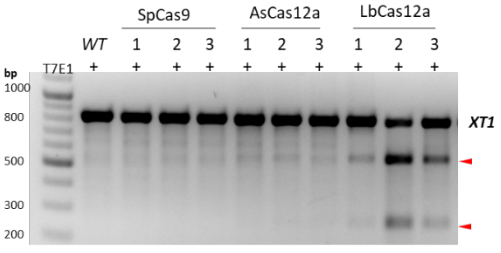


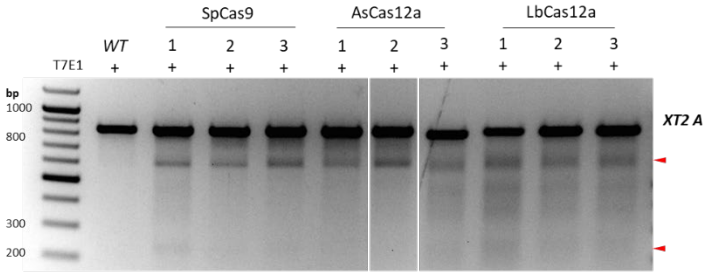


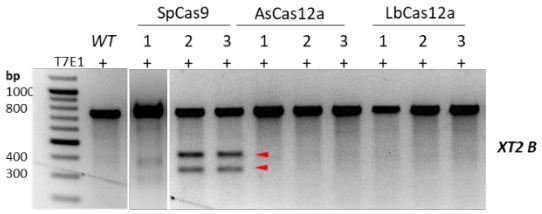
c) T7E1 assays used in the Figure 2 quantifications. The locus analyzed is detailed on the right side of the gel and the digested bands are indicated by the red arrow.

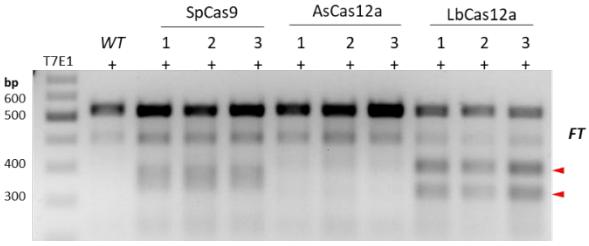


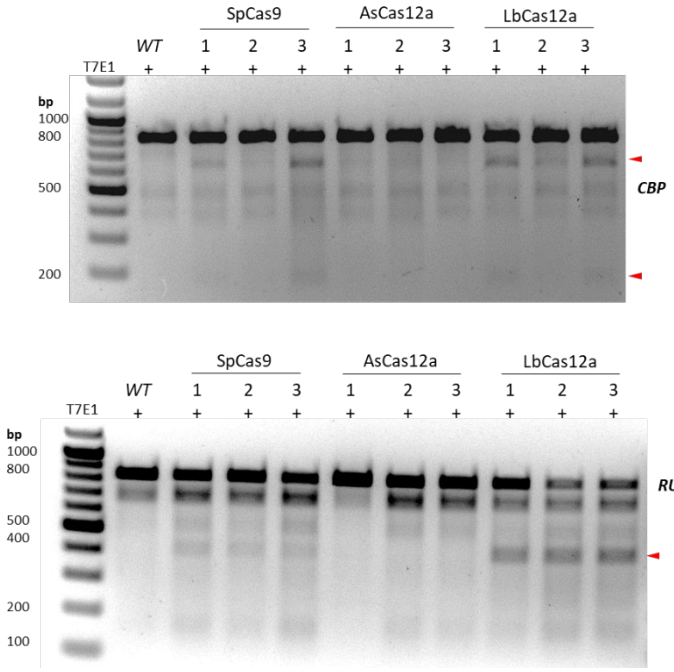





SpCas9
AsCas12a
LbCas12a
WT
1
2
3
T7E1
+
bp
1000
800
300
200
XT2 A


SpCas9
AsCas12a
LbCas12a
WT
1
2
3
T7E1
+
bp
800
400
300
XT2 B


SpCas9
AsCas12a
LbCas12a
WT
1
2
3
T7E1
+
bp
600
500
400
300
FT


SpCas9
AsCas12a
LbCas12a
WT
1
2
3
T7E1
+
bp
1000
800
500
200
CBP
SpCas9
AsCas12a
LbCas12a
WT
1
2
3
T7E1
bp
1000
800
500
400
200
100
RUBISCO

d) T7E1 assays used in the Figure 3 quantifications. The locus analyzed is detailed on the right side of the gel and the digested bands are indicated by the red arrow.

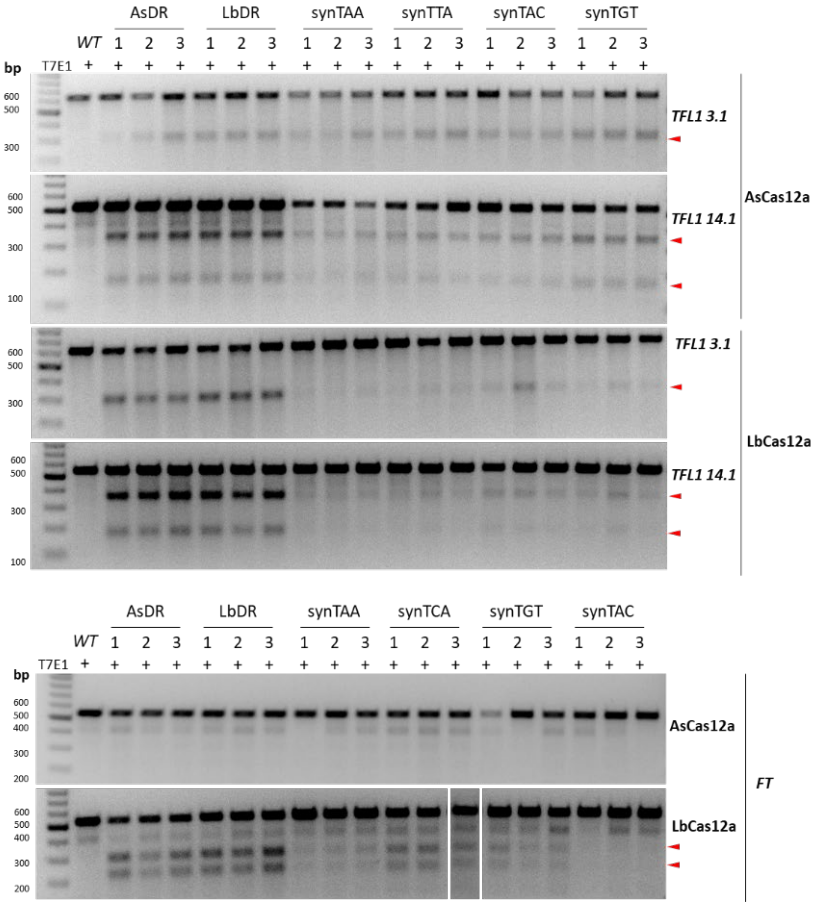
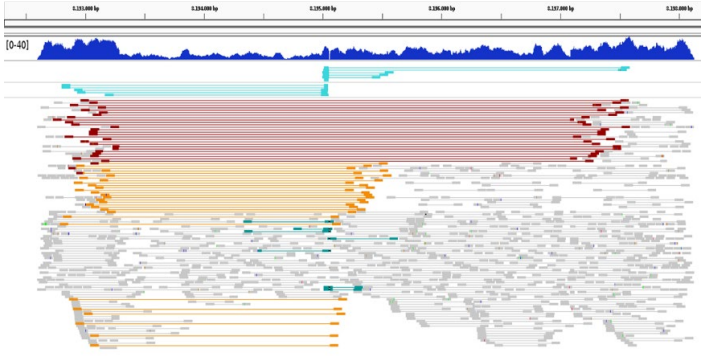


Figure S4: Venn diagrams of SNPs and INDELs detected with different software tools.

Figure S5: Structural variations and large deletions found in a duplication event surrounding the target 1 of PDS-1-118. Mates of blue reads have the same strand orientation and their insert sizes have a length of 2kb, as expected of an inversion event associated to a larger deletion event. Two large deletions of 4.6 Kb (read reads) and another one of 2,4 Kb (orange reads) can also be found. A small deletion of 5bp is also shown (green reads). The sequencing depth as number of reads is shown in blue.



7. Supplementary tables

Table S1: Constructs generated in this study. Sequences are accessible at GB cloning website using the GB database ID.

Level 0 GBparts		
GB database ID	**Name**	**Category**
GB1438	AsCpf1	CDS
GB1439	LbCpf1	CDS
GB1442	AtU6-26::AsDR	Other
GB1443	AtU6-26::LbDR	Other
GB1444	HDV	Other
GB2478	AtUBQ10	PROM+5UTR

Level 1 GBparts		
GB database ID	**Name**	**Category**
GB1440	35s::AsCpf1::35s	TU
GB1441	35s::LbCpf1::35s	TU
GB1753	AtU6-26::AsDR:gXT1::HDV	TU
GB1754	AtU6-26::AsDR:gXT1::PolyA	TU
GB1755	AtU6-26::AsDR::gXT2 A::HDV	TU
GB1756	AtU6-26::AsDR::gXT2 B::HDV	TU
GB1757	AtU6-26::AsDR::gTFL1::HDV	TU
GB1758	AtU6-26::AsDR:gFT::HDV	TU
GB1759	AtU6-26::AsDR:gCBP::HDV	TU
GB1760	AtU6-26::AsDR::gScRub::HDV	TU
GB1761	AtU6-26::LbDR::gXT1::HDV	TU
GB1762	AtU6-26::LbDR::gXT1::PolyA	TU
GB1763	AtU6-26::LbDR::gXT2 A::HDV	TU
GB1764	AtU6-26::LbDR::gXT2 B::HDV	TU
GB1765	AtU6-26::LbDR::TFL1::HDV	TU
GB1766	AtU6-26::LbDR::gFT::HDV	TU
GB1767	AtU6-26::LbDR::CBP::HDV	TU
GB1768	AtU6-26::LbDR::ScRub::HDV	TU
GB1769	AtU6-26::gXT1::gXT2	TU
GB1770	AtU6-26::gXT2 A::psgRNA	TU
GB1771	AtU6-26::gXT2 B::psgRNA	TU
GB1772	AtU6-26::gTFL1::psgRNA	TU
GB1773	AtU6-26::gFT::psgRNA	TU
GB1774	AtU6-26::gCBP::psgRNA	TU

GB1775	AtU6-26::gScRub::psgRNA	TU
GB1776	AtU6-26::synTAADR::TFL1::HDV	TU
GB2050	AtU6-26::synTTADR::TFL1::HDV	TU
GB2051	AtU6-26::synTACDR::TFL1::HDV	TU
GB2052	AtU6-26::synTGTDR::TFL1::HDV	TU
GB2053	AtU6-26::synTAADR::FT::HDV	TU
GB2054	AtU6-26::synTACDR::FT::HDV	TU
GB2055	AtU6-26::synTGTDR::FT::HDV	TU
GB2056	AtU6-26::synTCADR::FT::HDV	TU
GB2479	AtUBQ10::LbCpf1::Tnos	TU
GB2480	AtU6-26::gPDS3-D1::HDV	TU
GB2481	AtU6-26::gPDS3-D2::HDV	TU
GB2482	At2S3::DsRED::T35S	TU
GB2483	DsRED(TU)_LbCpf1(TU)_guides	MODULE

Leve>1 GBparts		
GB database ID	Name	Category
GB1629	NptII-AsCpf1-XT1	Module
GB1630	NptII-LbCpf1-XT1	Module
GB1179	NptII-sgRNA_MYB12-hCas9	Module
GB1780	NptII-crRNA_MYB12-AsCpf1	Module
GB1781	NptII-crRNA_MYB12-LbCpf1	Module

Table S2: oligonucleotides used in this study.

Name	Sequence (5'-3')
crRNA TFL1 Cpf1 +	AGATACTGTGGGACTGAATGAATC
crRNA TFL1 Cpf1 -	GGCCGATTCATTCAGTCCCACAGT
crRNA TFL1 Cas9 +	ATTGTGACTGTCATTTTAACTGT
crRNA TFL1 Cas9 -	AAACACAGTTAAAATGACAGTCA
crRNA FT Cpf1 +	AGATCAAGATCTATTGGCCTAAGA
crRNA FT Cpf1 -	GGCCTCTTAGGCCAATAGATCTTG
crRNA FT Cas9 +	ATTGGCCAATAGATCTTGTAAAA
crRNA FT Cas9 -	AAACTTTTACAAGATCTATTGGC
crRNA Cas9 XT2B +	ATTGCTCTGATTGCACAATGGAA
crRNA Cas9 XT2B -	AAACTTCCATTGTGCAATCAGAG
crRNA Cpf1 XT2B +	AGATGCTCTGATTGCACAATGGAA

crRNA Cpf1 XT2B -	GGCCTTCCATTGTGCAATCAGAGC
crRNA XT2A Cpf1 +	AGATTCCACTAGTTTTTTCCCAAT
crRNA XT2A Cpf1 -	GGCCATTGGGAAAAAACTAGTGGA
crRNA XT2A Cas9 +	ATTGAAAATTGGGAAAAAACTAG
crRNA XT2A Cas9 -	AAACCTAGTTTTTTCCCAATTTTC
crRNA Cpf1 Rubisco +	AGATGATGCAGATACTGGACCATG
crRNA Cpf1 Rubisco -	GGCCCATGGTCCAGTATCTGCATC
crRNA Cas9 Rubisco +	ATTGATGCAGATACTGGACCATG
crRNa Cas9 Rubisco -	AAACCATGGTCCAGTATCTGCAT
crRNA Cpf1 CBP +	AGATGAGGACAAACTACATCCAGG
crRNA Cpf1 CBP -	GGCCCCTGGATGTAGTTTGTCCTC
crRNA Cas9 CBP +	ATTGAGGACAAACTACATCCAGG
crRNA Cas9 CBP -	AAACCCTGGATGTAGTTTGTCCT
crRNA XT1 Cpf1 +	AGATTATGTAGGTGTATTTGGAAT
crRNA XT1 Cpf1 -	GGCCATTCCAAATACACCTACAT
crRNA XT1 Cas9 +	ATTGTGTAGGTGTATTTGGAATTC
crRNA XT1 Cas9 -	AAACGAATTCCAAATACACCTAC
synDR TFL1 F	ATTGTAATTTCTACTAAGTAGATACTGTGGGACTGAA TGAATC
synDR TFL1 R	GGCCGATTCATTCAGTCCCACAGTATCTACTTAGTA GAAATTA
LbCpf1 XT1 multiplexing+	CTCGATTGTAATTTCTACTAAGTGTAGATTATGTAGG TGTATTTGGAATTAATTTCTAC
LbCpf1 XT1 multiplexing -	CTCAGTAGAAATTAATTCCAAATACACCTACATAATC TACACTTAGTAGAAATTACAAT
LbCpf1 XT2 multiplexing +	CTCGCTACTAAGTGTAGATTCCACTAGTTTTTTCCCA ATTAATTTCTACTAAGTGTAGATTTTTTTTTCGCT
LbCpf1 XT2 multiplezing -	CTCAAGCGAAAAAAAAATCTACACTTAGTAGAAATT AATTGGGAAAAAACTAGTGGAATCTACACTTAGTAG
Cpf1 XT1 polyA +	AGATTATGTAGGTGTATTTGGAATTTTTTTTT
Cpf1 XT1 polyA -	AGCGAAAAAAAAAATTCCAAATACACCTACATA
crRNA MYB12 Cas9 +	ATTGAAAGAGTTGTAGACTACGA
crRNA MYB12 Cas9 -	AAACTCGTAGTCTACAACTCTTT

crRNA MYB12 Cpf1 +	AGATGAAGGATTATTGAGATGCGG
crRNA MYB12 Cpf1 -	AAACCCGCATCTCAATAATCCTTC
crRNA synDR_TTA TFL1 +	AAACCATGGTCCAGTATCTGCAT
crRNA synDR_TTA TFL1 -	GGCCGATTCATTCAGTCCCACAGTATCTACTAAGTA GAAATTA
crRNA synDR_TAC TFL1 +	ATTGTAATTTCTACTACGTAGATACTGTGGGACTGAA TGAATC
crRNA synDR_TAC TFL1 -	GGCCGATTCATTCAGTCCCACAGTATCTACGTAGTA GAAATTA
crRNA synDR TGT TFL1 +	ATTGTAATTTCTACTGTGTAGATACTGTGGGACTGAA TGAATC
crRNA synDR TGT TFL1 -	GGCCGATTCATTCAGTCCCACAGTATCTACACAGTA GAAATTA
crRNA synDR TAA FT +	ATTGTAATTTCTACTAAGTAGATCAAGATCTATTGGC CTAAGA
crRNA synDR TAA FT -	GGCCTCTTAGGCCAATAGATCTTGATCTACTTAGTAG AAATTA
crRNA synDR TAC FT +	ATTGTAATTTCTACTACGTAGATCAAGATCTATTGGC CTAAGA
crRNA synDR TAC FT -	GGCCTCTTAGGCCAATAGATCTTGATCTACGTAGTA GAAATTA
crRNA synDR TGT FT +	ATTGTAATTTCTACTGTGTAGATCAAGATCTATTGGC CTAAGA
crRNA synDR TGT FT -	GGCCTCTTAGGCCAATAGATCTTGATCTACACAGTA GAAATTA
crRNA synDR TCA FT +	ATTGTAATTTCTACTCAGTAGATCAAGATCTATTGGC CTAAGA
crRNA synDR TCA FT -	GGCCTCTTAGGCCAATAGATCTTGATCTACTGAGTA GAAATTA

PCR oligonucleotides	
ID	Sequence (5'-3')
NbXT1 F	ACCCTGATCACTCTCGTC
NbXT1 R	TTGTATTCTATATCTGCCAACATTC
TIDE seq XT1	CCACTCTTCAATCACTTCCC

NbXT2 F	GGCGATAACACCGTCTC
NbXT2 R	AGACATCAAAAGTTATAGGAATAATC
TIDE seq XT2	GGTGACGGCGGATGGTTTAG
NbTFL1-3.1 F	GCTTACTGTGTCCTGTATAAAACTG
NbTFL1-3.1 R	GTAATCTCATATGTGGTATCGGAG
TFL1-3.1 seq	CAGCAACTATAAATGAGGCTC
NbTFL1-14.1 F/seq	GAGTAGATATTCCAAGACAGCAAC
NbTFL1-14.1 R	CACTGGAAAACCTCTGTAACAAC
NbFT F	CTAGAAAACCTATGGCTATAAGGG
NbFT R	GTTCTCGAGAGGTATAATATAGGC
FT seq	CACAAGCACGCATAGAAC
NbXT2B F	TGCACGGTTGTCCGAGTTTG
NbXT2B R	TTACTTGTGAATTGCTCTCTGGT
NbRubisco F	AGCGAAATTGAGTACCTCTTGAA
NbRubisco R	ACTCATATAGGCCACACAACAAT
NbCBP AB F	TGTCTAGACTGGTGCATTACTTC
NbCBP AB R	GTTGCCAAAAGGATCACTCAAAT
SlMYB12 F	GATGGACTGCAGAAGAAGATCAA
SlMYB12 R	AACATCGAAATTTGTACCTGAACT
AtPDS F1	CGAACCGACCCGAGAAGAGA
AtPDS R1	CTATTTCAGGTCGCCGCTCA
AtPDS R2	GCCGACCATGGCTGGCAAA

Table S3: stability of the different crRNAs of *TFL1* and *FT* loci. The stability has been measured as the -ΔG predicted by the Vienna RNA and Mfold software tools. For each loci, the sequence among the synthetic crRNAs are identical, only varying two nucleotides of the direct repeat loops (5'-UNN-3') which generates a different structure with different stability.

-ΔG TFL1 crRNAs		
5'-UNN-3'	Vienna RNA	Mfold
UAA	4	4.7; 4.1; 4
UUA	4	4.7; 4.3; 4.1; 4; 3.9
UAC	3.7	4.7; 4.6; 4.1; 4
UGU	5.6	6

-ΔG FT crRNAs		
5'-UNN-3'	Vienna RNA	Mfold
UAA	4.1	4.9; 4.9; 4.7; 4
UAC	3.9	4.9; 4.7; 4.5; 4.3; 4
UGU	3.3	4.9; 4.7; 4.3; 4
UCA	7.3	7; 6.5

	-ΔG TFL1 WT crRNAs	
	Vienna RNA	Mfold
AsDR	5.5	4.1; 4.4; 4.5; 4.9
LbDR	5.8	4.9; 5; 5.4

	-ΔG FT WT crRNAs	
	Vienna RNA	Mfold
AsDR	6.15	4.3; 4.4; 4.9; 5.1
LbDR	5.9	4.9; 5.6

Table S4: mapping stats and filtering parameters used in the WGS off-target analysis.

a) Read and mapping stats against the reference genome TAIR10.

Samples	Raw Reads	Clean Reads	% Remaining Reads	Properly Mapped reads	% Clean reads mapped
POOL	218876308	189905488	86,76 %	142472371	75,02 %
PDS-1-111	53615620	45729312	85,29 %	33841836	74,00 %
PDS-1-113	48045816	41316210	85,99 %	30879446	74,74 %
PDS-1-114	49526888	42845404	86,51 %	32132445	75,00 %
PDS-1-115	49177422	42282212	85,98 %	31058272	73,45 %
PDS-3-117	55059068	46790330	84,98 %	35722381	76,35 %
PDS-3-118	78323478	68581142	87,56 %	52700597	76,84 %
PDS-3-119	41661252	35192082	84,47 %	26186410	74,41 %
PDS-3-123	59969898	51610012	86,06 %	38797014	75,17 %
WT-104	38125478	32252932	84,60 %	23280192	72,18 %
WT-106	55280774	47793828	86,46 %	36127369	75,59 %

b) Parameters used in GATK, FREEBAYES, VARSCAN2, PINDEL and DELLY programs.

	PARAMETERS
GATK	QD < 2.0, FS > 60.0 MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0
FREEBAYES	-n 2, -m 20, -C 2, -F 0.2
VARSCAN2	By default
PINDEL	By default
DELLY	-p -a 0.04 -r 1 -f germline

CHAPTER III

A reversible memory switch for plant synthetic biology based on the phage PhiC31 integration system

Bernabé-Orts Joan Miquel*, Quijano-Rubio Alfredo*, Mancheño-Bonillo Javier, Moles-Casas Victor, Selma Sara, Granell Antonio, Orzaez Diego

bioRxiv 656223; doi: https://doi.org/10.1101/656223

*These authors contributed equally to this work.

1. Introduction

Plant synthetic biology (PSB) is an established but continuously growing field that combines the engineering principles of decoupling, abstraction and standardization with plant biology to provide plant systems with new traits and functions [1]. Plants naturally offer a myriad of metabolic resources that could be harnessed by modern synthetic biology to tackle current challenges in alimentation, energy production, pest management, or production of valuable biologicals for medicine and industry [2]. Recent efforts have been focused on introducing metabolic pathways to improve traits such as vitamin production or glycolate metabolism [3, 4], fixation of atmospheric nitrogen [5], increased yield or improving photosynthesis [6, 7]. Nevertheless, the implementation of these functions requires precise regulation of the new genetic networks to ensure an optimal use of the plant resources and avoid deleterious effects. The introduction of complex programmed genetic networks in bacterial and mammalian systems has been a great success, achieving engineered organisms that behave as cellular computing devices that have substantial biotechnological and therapeutic impact [8]. This success has been possible thanks to the previous development and extensive characterization of genetic devices and molecular tools that, combined, generate increasingly complex functions. However, the translation of these complex genetic circuits to plants has been challenging due to the limited availability of well-documented genetic elements as well as the technical limitations of high-throughput plant transformation and circuit characterization [1, 9]. Current efforts in PSB are focused on developing reliable DNA assembly standards (phytobricks) to allow more efficient biodesign and characterization of plant functions [10, 11].

Among the diverse synthetic biology tools developed in the past years, basic memory devices such as e.g. transcriptional toggle switches have been pivotal in allowing the design of increasingly complex genetic circuits [12]. Traditional inducible control systems lacking memory functions require a constant supply of actuators (activators or repressors) for a sustained control of cell outputs, often imposing a metabolic burden to the cell [13]. In contrast, toggle switches provide the ability to respond to punctual external inputs with a sustained response. Applied to crop plants, gene memory devices could facilitate the deploy of new layers of genetic control over agronomically-relevant outputs such as flowering time, stress response or the biosynthesis of added value metabolites. Furthermore, memory systems circumvent the need for continuous addition and monitoring of external inducers, which can be expensive and difficult to control, especially in large settings as crop fields.

While several non-memory chemically-inducible systems have been adapted to plant biotechnology, such as ethanol [14], glucocorticoids [15, 16], copper [17, 18] or insecticides as methoxyfenozide [19], memory devices adapted and/or engineered for the plant chassis are almost absent. As a notable exception, Müller *et al.*, 2014 [20] developed a red light-inducible toggle switch for plant cell protoplasts, based on the interaction of *Arabidopsis thaliana* PhyB-PIF6 proteins. This approach showed remarkable spatiotemporal resolution of gene expression. However, this system relies on light for activation, thus, its adaptation to whole plants can be challenging in applications in which the illumination conditions cannot be strictly controlled, as in open field applications. Additionally, the memory of this optogenetic system is based on the slow dissociation rate of the PhyB-PIF6 protein-protein interaction, which will result in eventual inactivation due to the decay and dilution of the protein components, making this system unsuitable for applications in which information needs to be maintained for long periods of time or transmitted over generations. Therefore, alternative gene

memory switches for plants that allow long-term and inheritable memory storage are necessary.

Synthetic genetic memory can be built through diverse mechanisms, such as transcription-based double negative feedback loops, positive feedback loops or DNA recombination [21, 22]. Integrases are a group of DNA-actuating enzymes found in temperate bacteriophages that catalyze site-specific recombination (SSR) [23]. Serine integrases catalyze the recombination of attP and attB attachment sites, generating hybrid attR and attL sites in a strictly unidirectional reaction. This process can be reversed in the presence of an excisionase or recombination directionality factor (RDF), which in combination with the integrase can catalyze the recombination of the attR-attL sites into attP-attB again [24]. This specific DNA recombination event can result in a variety of DNA-rearrangements (integration, excision, inversion and translocation) depending on the topology of the initial DNA molecules and the orientation of the recombination sites. SSR mechanisms of diverse integrases have been extensively exploited in bacteria and mammalian cells to engineer reversible synthetic memory devices and complex logic circuits. Examples of applications include counting cellular events [25], store and rewrite biological data in the chromosome [26], control the flow of the RNA polymerase along the DNA and thus amplifying the expression of a reporter gene [27] and to engineer complex regulatory circuits [28-30]. Although SSR systems have been used in plants before, all previous efforts have been focused in transgene engineering for crop breeding rather than in the design of memory devices for the control of gene expression [31, 32]. Some of their uses include the removal of foreign DNA [33], stacking agronomic-valuable traits [34], and chloroplasts engineering [35].

Here, we present the first toggle memory switch for whole plants based of the phage PhiC31 serine integrase and its cognate recombination directionality factor, RDF [36]. The switch is designed to control the transcription of two genes of interest, one of which is initially ON, while the second one remains in OFF status. The state of the switch changes by limited supply of integrase to the cell and the memory of the new status is maintained until intentionally reversed by a combined new supply of integrase and RDF. The components of the genetic switch were designed for plant expression and standardized by using the GoldenBraid (GB) DNA assembly platform and software to allow easy adaptation of this tool to other PSB applications [10]. We extensively tested the kinetics, memory and reversibility of this genetic device in *N. benthamiana* through transient and stable transformation experiments using transgenic plants and hairy roots. Additionally, we coupled the integrase expression to an estradiol-inducible promoter as a proof of principle of externally-induced activation of the switch.

2. Results

2.1. Design of a modular reversible genetic switch for plant systems

To design a reversible genetic switch for plant expression, we adapted a previously reported strategy [26] to the phage PhiC31 integrase and its cognate RDF [36]. This genetic switch comprises a flippable DNA promoter element flanked by the PhiC31 attP and attB recombination sites in opposing orientations, named as register PB (Figure 1A).

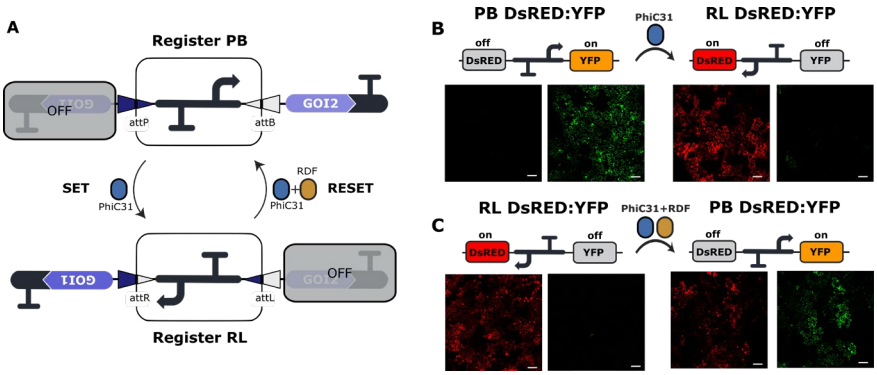


Figure 1: Design and functional validation of the plant toggle switch based on the phage PhiC31 integration system. (A) An invertible plant promoter element (Register) works as a toggle switch for the regulated expression of two genes of interest (GOI1 and GOI2). The promoter orientation can be inverted by action of the PhiC31 integrase (SET) which catalyzes site-specific recombination of the attP and attB sites flanking the promoter. This event results in a change in the expression status of the two GOIs and the creation of the chimeric attR and attL sites. Expression of the integrase and recombination directionality factor (RDF) catalyzes recombination of attR and attL to reset the toggle switch to its original state (RESET). The genetic parts encoding the PhiC31 integrase and RDF can be found in the GoldenBraid collection, GB1531 and GB1508 respectively. **(B, C)** CLM images of WT *N. benthamiana* leaves infiltrated with the toggle switch alone (left) or in combination with the recombination actuators (PhiC31 or PhiC31 + RDF), (right). Reporter genes expression (DsRED and YFP) was analyzed in the initial and the commuted states of the switch during the SET (B) and RESET operations (C). Images were taken 3 days post-infiltration. The scale bar represents 100 µm.

In the initial state, the switchable promoter drives the expression of a gene of interest (GOI2) positioned downstream of the register PB in the direct DNA strand. The orientation of the promoter can be changed by action of the PhiC31 integrase, which catalyzes attPxattB recombination into attR and attL respectively (SET), flipping the DNA segment would produce a new register RL. As a result, the register RL now drives the expression of a second GOI located in the reverse strand (GOI1), while GOI2 expression would be turned off. Our goal was to design a reversible genetic switch, therefore to enable the register RL to be reversed to the PB state we introduced the PhiC31 integrase and RDF factor, which combined catalyze attRxattL recombination into attP and attB (RESET), resulting in the reconstitution of the initial register PB and expression of GOI2 while GOI1 is inactivated. To maximize the exchangeability and reusability of this this genetic device for PSB applications, we took advantage of the modularity of GB to structure our genetic switch in three standard and interchangeable parts: (i) the direct coding sequence (CDS) encoding GOI2, (ii) the registers PB or RL, and (iii) the reverse CDS encoding GOI1 (Fig. S1A). Reverse and direct CDSes can be easily created from any Transcriptional Unit (TU) conforming the phytobrick standard, as e.g. those available at the GB collection (Fig. S1B and S1C). The register comprised the strong CaMV35S promoter (P35S) and a tomato Metallothionein B terminator (MTB) sequence located upstream and in opposite orientation to act as an insulator to avoid any leaky backward expression. Recombination sites were inserted flanking the terminator and promoter elements. Since incorrect insertion of the att site close to the transcription start site (TSS) or ribosome binding site (RBS) could hinder efficient expression, this was inserted in the 5′ untranslated region (5′UTR) of the P35S, downstream of the TATA box and upstream of the RBS, keeping them intact (Fig. S1A). The registers were named according to the 5′-3′ order of the att sites: register PB (attP-attB sequences, GB1494) and the register RL (attR-attL, GB1506). The three elements were assembled in a single step Golden Gate-like

reaction to create the so-called register module (RM). The actuators of the SET and RESET operations were adapted for *in planta* expression and incorporated to the GB collection: Pnos:PhiC31:Tnos (GB1531) and P35S:RDF:T35S (GB1508).

To test the functionality of the toggle switch in plants, yellow fluorescent protein (YFP) and DsRED were selected as direct and reverse reporters respectively. The reporter genes were then combined with the PB and RL registers to create the PB DsRED:YFP and RL DsRED:YFP RMs. Next, we performed *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves by co-infiltrating each RM (PB or RL DsRED:YFP) with its appropriate actuators (PhiC31 or PhiC31+RDF) to evaluate the SET and RESET operations by confocal laser microscopy (CLM), (Fig. 1B-1C). As expected, leaf agroinfiltration of the PB DsRED:YFP register module alone yielded bright YFP expression, with no DsRED expression detected. In contrast, co-infiltration of PB DsRED:YFP with PhiC31 integrase (SET operation) resulted in robust activation of DsRED expression and de-activation of YFP expression in most cells, as shown in Fig. 1B. Similarly, agroinfiltration of RL DsRED:YFP module alone produced only DsRED (+) cells, whereas co-expression of RL DsRED:YFP + Integrase + RDF (RESET operation) resulted in activation of YFP expression (Fig. 1C). As can be observed, the RESET process was not complete, resulting in some cells still expressing DsRED. Co-expression of RL DsRED:YFP with PhiC31 alone did not change the reporter expression, proving the necessity of RDF to enable attRxattL recombination (data not shown). Altogether, with these results we obtained a snapshot of the two states of our plant-customized toggle switch, confirming its functionality.

To monitor quantitatively the expression kinetics of the SET and RESET operations, DsRED was substituted by the firefly luciferase (LUC or Fluc) as a reporter gene, producing four new RMs, namely PB LUC:YFP, PB YFP:LUC, RL LUC:YFP and RL YFP:LUC (Fig. 2 and Fig. S2). Additionally, a constitutively-expressed renilla luciferase (Rluc) was assembled to each RM as internal reference (Fig. S2).

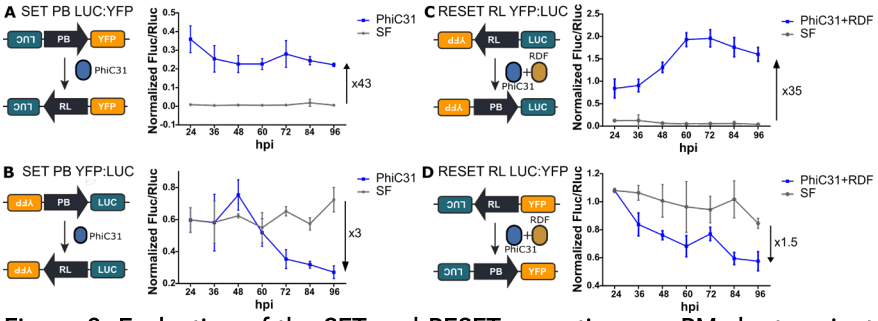


Figure 2: Evaluation of the SET and RESET operations on RMs by transient expression in WT *N. benthamiana* leaves. (A-B) SET of the PB LUC:YFP (A) and the PB YFP:LUC (B) mediated by the PhiC31 integrase. **(C-D)** RESET of the RL YFP:LUC (C) and RL LUC:YFP (D) mediated by the combined action of PhiC31 integrase and RDF. **(A, C)** The recombination processes switch on Fluc reporter expression resulting in an increase of the Fluc/Rluc ratios (Rluc was expressed as a constitutive internal control). **(B, D)** The recombination reactions switch off Fluc expression resulting in a decrease of the Fluc activity over the time. All the graphs represent the evolution of Fluc/Rluc ratios at intervals of 12 hours post infiltration (hpi) for 96h. Leaves were treated with the recombination actuators (blue line) or with an empty vector named SF used as negative control (grey line). Each point represents the mean of Fluc/Rluc ± SD of three leaves in different plants.

RMs were transiently co-expressed in WT *N. benthamiana* leaves with its appropriate effector (PhiC31 for SET or PhiC31+RDF for RESET) or with a non-coding DNA fragment (stuffer fragment, SF) as a negative control, and luciferase levels were recorded over time starting at 24 hours post-infiltration (hpi). As shown in Fig. 2A, SET activation of Fluc was strong and rapid, producing values of 43-fold Fluc/Rluc induction levels from the start of the measurement This indicated that recombination was efficiently taking place almost concomitantly with the transient transformation process. RESET activation of Fluc in the RL

YFP:LUC module was also readily detectable by 24 hpi (Fig. 2C), although the maximum Fluc/Rluc signal of 35-fold induction was not observed until 60 hpi, indicating slower kinetics of this operation involving two actuators. Fig. 2B and 2D show the deactivation of Fluc using SET and RESET operations respectively. In both cases only partial deactivation of Fluc was observed, reaching approximately 75% signal decrease in the case of SET and only 40% Fluc deactivation for RESET as compared to non-operated agroinfiltrated control.

2.2. Characterization of the register modules in stably transformed *N. benthamiana* plants

While transient transformation enables quick characterization of molecular tools developed for plants, we sought to develop a genetic switch that could be stably introduced into genetically modified plants to suit multiple applications. *Agrobacterium*-mediated stable transformation involves the random integration of foreign transfer DNA (T-DNA) in the plant genome, which may result in unpredicted changes in the performance of the introduced genetic element. In order to further characterize the performance of our genetic switch in the genomic context, we generated transgenic *N. benthamiana* lines carrying the four aforementioned RMs (PB LUC:YFP, PB YFP:LUC, RL LUC:YFP and RL YFP:LUC; Fig. S2). T0 plants were screened based on YFP and Fluc expression levels, and then one line of each RM with a single copy of the transgene was selected to perform subsequent experiments on its T1 progeny (Fig. S3). SET and RESET operations were performed by *Agrobacterium*-mediated transient delivery of the recombination actuators. Optimal levels of each actuator were experimentally determined by titrating the optical density (OD) of the respective *Agrobacterium* cultures, a parameter which is known to correlate with the amount of T-DNA delivered to the transformed cells [10] (Fig. S4). Agroinfiltrated leaves were sampled and analyzed for Fluc and Rluc expression every 24h for 7 days (Fig. 3).

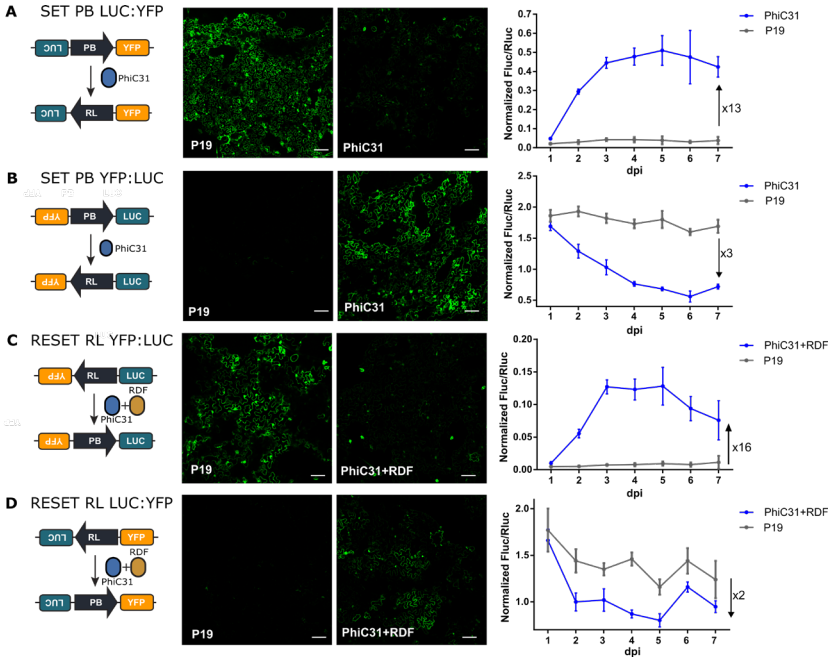


Figure 3: Evaluation of the SET and RESET operations in transgenic *N. benthamiana* with RMs integrated in their genome. (A-B) SET recombination of the PB LUC:YFP (A) and PB YFP:LUC (B) mediated by the PhiC31 integrase. **(C-D)** RESET recombination process mediated by the combined action of PhiC31 and RDF of the RL YFP:LUC (C) and RL LUC:YFP (D) RMs. **(A, C)** In the initial state the YFP reporter is actively expressed as seen by the CLM image infiltrated with P19 (negative control). Infiltration of the recombination actuators flips the register cassette inactivating YFP expression (PhiC31 or PhiC31+RDF images) and activating the Fluc expression (graph, blue line). **(B, D)** In the initial state YFP expression is off as demonstrated by the P19 control. Agroinfiltration of actuators switches on YFP expression (PhiC31 or PhiC31+RDF images) and switches off Fluc expression (graph, blue line). CLM images were taken at 5 days post infiltration (dpi); scale bars represent 100 μm. All the graphs represent the evolution of Fluc/Rluc values for 7 dpi of leaves transiently transformed with actuators (blue lines) or P19 as negative control (gray line). Experimental points show the mean of normalized Fluc/Rluc values of three agroinfiltrated leaves ± SD.

As anticipated, the RMs where Fluc expression turns on upon performing operations (SET PB LUC:YFP and RESET RL YFP:LUC) responded to actuator delivery with a rapid increase in Fluc/Rluc levels, reaching a maximum at 5 days post infiltration (dpi). Concomitantly with Fluc activation, YFP levels decreased markedly when measured in the same infiltrated areas at 5 dpi (Fig. 3A and 3C, Fig. S5B). Reporter Fluc/Rluc induction levels for the transgenic lines were 13-fold and 15-fold for the SET and RESET operation respectively. Additionally, while Fluc levels were sustained over time for the SET operation (Fig. 3A), Fluc/Rluc signal showed a slow decrease after 5 dpi for the RESET operated RMs (Fig. 3C). This could be caused by an excess of PhiC31 integrase not bound to RDF, which would cause multiple RESET/SET recombination cycles decreasing the stability of the measurements. On the other side of the equation, those RM conformations where Fluc turns off upon commutation (SET PB YFP:LUC and RESET RL LUC:YFP), showed a moderate but consistent decrease of Fluc/Rluc levels in parallel to a clear activation of the YFP signal when commuted by agroinfiltration (Fig. 3B and 3D, Fig. S5A). Altogether, these results show that both SET and RESET operations take place correctly in the genomic context of stably transformed plants.

2.3. Stable and reversible memory storage over a full SET/RESET cycle in whole plants

A distinctive feature of a PhiC31-based memory device is the long-term stability of commutations based on DNA recombination [37]. Our primary goal was to create a switch that could alternate between two different configurations and whose memory status be transmitted to the progeny. To this end, we set out to test the memory and reversibility of the switch by performing a full SET/RESET cycle in *N. benthamiana* lines with the RMs integrated in the genome. As successive agroinfiltrations of the actuators would result in excessive damage to the plants for accurate analysis, we decided to generate calluses from

commutated leaves after the first SET operation, and then to perform the second RESET operation on fully regenerated plants. In the first place, the efficiency of the SET and RESET operations was tested in the context of *in vitro Agrobacterium*-mediated plant leaf disc transformation. RL LUC:YFP and PB YFP:LUC lines were used to follow the course of the transformation through YFP expression activation (Fig. S6A). Our approach showed that the SET operation produced twice as much fluorescent callus than the RESET (Fig. S6B-C). However, these transformations ended with highly chimeric calluses which eventually would yield plants with mixed states of the RMs, probably because the recombination took place in an advanced stage of the callus generation process. We reasoned that performing the SET or RESET operation by agroinfiltration prior to the *in vitro* regeneration would increase the initial population of commutated cells, raising the chances to obtain fully commutated calluses and subsequently fully commutated plants. For the same reason, we limited the experiment to the most efficient SET operations. Therefore we conducted the experiment illustrated in the Fig. 4A, using the aforementioned PB YFP:LUC stable line to perform a SET operation by transient expression of PhiC31 in the leaf. Then, at 5 dpi when the peak of recombination was reached, fluorescent discs were excised and regenerated *in vitro* to obtain stably switched calli. As previously shown, this resulted in chimeric calli with variable efficiency of the SET operation and thus, variable YFP expression levels. Then, the calli were induced to differentiate organogenesis and to obtain explants, which resulted in six plant lines (P1-P6). In order to assess the configuration of the RM in each line (either the initial PB YFP:LUC or the commutated RL YFP:LUC state), regenerating lines were genotyped using specific primer pairs for PB or RL state. The presence of integrated PhiC31 was also analyzed in the same lines (Fig. 4B). Plants P1 and P2 did not show evidence of commutation events (only PB configuration detected). P1 showed no presence of PhiC31, while P2 showed only a weak band. P5 and P6 were apparently chimeric plants in which both configurations coexisted, P6 showing only a faint band

for commuted configuration compatible with the absence of PhiC31 integration. Finally, P4 and P7 were positive only for RL and PhiC31, indicating a complete switch to the RL configuration. Thus, P7 explant was transferred to soil and used to test the reversibility of the genetic device.

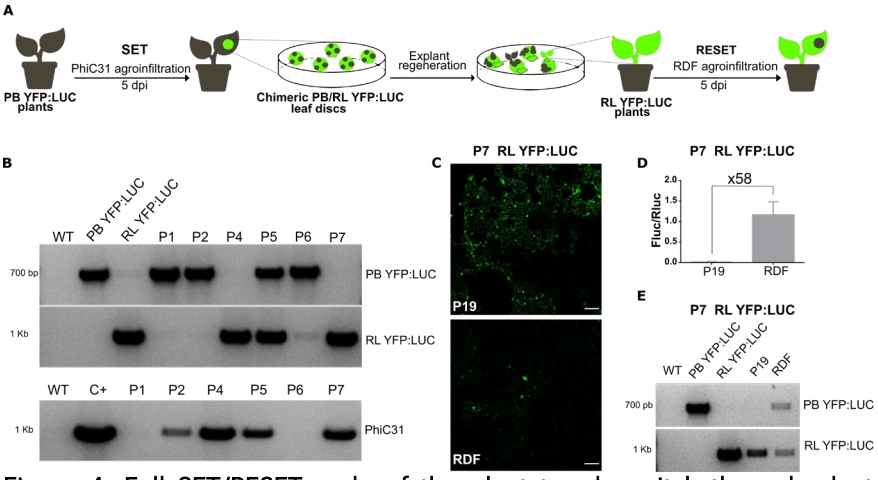


Figure 4: Full SET/RESET cycle of the plant toggle switch through plant transformation and regeneration process. (A) Representation of the experiment conducted to assess the bistability and reversibility of the toggle switch in *N. benthamiana* PB YFP:LUC transgenic lines (YFP off, Fluc on). PB YFP:LUC plants were agroinfiltrated with the PhiC31 integrase to induce a SET to RL YFP:LUC. Five days post infiltration fluorescent leaf discs were collected, sterilized and then cultivated *in vitro* until explant regeneration. Explants genotype was then analyzed by PCR, to select totally commutated RL YFP:LUC plants. RL YFP:LUC plants (such as P7) were then agroinfiltrated with PhiC31+RDF to operate the RESET to the original PB YFP:LUC configuration and demonstrate the reversibility of the system. **(B)** Agarose electrophoresis gels showing the genotyping results of the regenerated plants (P1-P7). The upper panel reflects the results for a specific pair of primers for the amplification of the original PB YFP:LUC configuration. The middle panel shows the results when amplifying with a specific pair of primers for the commutated RL YFP:LUC state. The lower panel shows amplification results of integrated PhiC31 integrase. **(C)** CLM images of agroinfiltrated P7 with P19 (negative control) and RDF (RESET) after 5 dpi. Scale bars represent 100 µm. **(D)** Fluc/Rluc ratio of P7 agroinfiltrated with P19 or RDF at 5 days post induction. Values represent the mean Fluc/Rluc values for three different agroinfiltrated leaves ± SD. **(E)** Genotyping results of agroinfiltrated P7 with P19 and RDF at 5 dpi.

At maturity, P7 leaves were agroinfiltrated with RESET effector (RDF) or with a control (P19) and analyzed for YFP (Fig. 4C) and Fluc expression (Fig. 4D). Remarkably, YFP expression significantly decreased at 5 dpi in RESET samples (Fig. 4C) and Fluc activity was induced 58-fold over the un-induced background levels, reaching approximately 60% of the Fluc/Rluc ratios typically observed in the original PB YFP:LUC plant (Fig. 4D). These results were also genotypically confirmed by PCR-amplification of the commuted tissue, as compared to control samples (Fig. 4E). Altogether, these experiments showed that the toggle switch, once integrated in the genome can effectively perform a full cycle of SET/RESET operations.

2.4. Chemical induction of PhiC31 controls SET operation in *N. benthamiana* hairy roots

Master regulators are essential for engineering synthetic gene networks in which metabolic fluxes can be deviated among different pathways. So far, we have demonstrated that our plant-custom-made PhiC31-based toggle switch can alternate between two defined states and maintain memory, a key feature of a master regulator. However, we relied on agroinfiltration to induce a state change by expression of the SET and RESET effectors, which does not allow precise spatiotemporal control of expression. We sought to achieve tighter control over the recombination process by coupling the expression of PhiC31 to an estradiol-inducible system, which was named estradiol-inducible PhiC31 (EI PhiC31). To this end, a GB module was assembled, encompassing three transcriptional units encoding for (i) a chimeric trans-activator (estradiol receptor (ER) fused in N-terminal to the LexA binding domain (LexABD) and to GAL4 activation domain (GAL4AD) in C-terminal); (ii) an inducible PhiC31 with the LexA operator and the mini35S promoter; and (iii) a DsRED fluorescent marker (Fig. 5A). In this system, the trans-activator is constituvely expressed and remains in the cytoplasm. Upon addition of estradiol, the trans-activator localizes to the nucleus and binds to the DNA operator,

inducing the expression of PhiC31 integrase. Expression of PhiC31 results in a SET operation which is evidenced by Fluc luminescence.

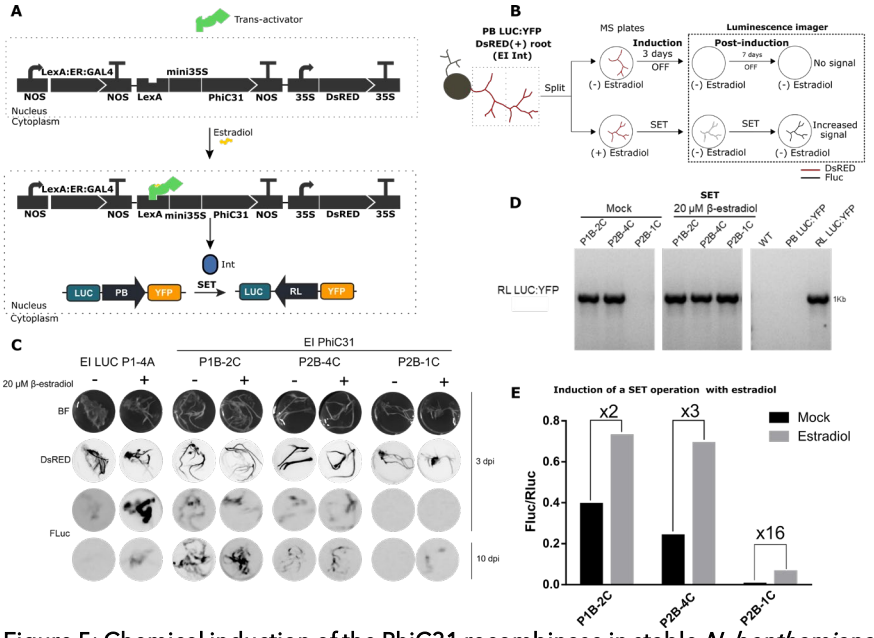


Figure 5: Chemical induction of the PhiC31 recombinase in stable *N. benthamiana* PB LUC:YFP hairy roots. (A) Representation of the estradiol-inducible (EI) system EI PhiC31. In absence of estradiol, the constitutively expressed chimeric trans-activator is confined to the cytoplasm. Upon addition of estradiol it localizes to the nucleus where it induces the expression of the PhiC31 integrase. This enables the SET operation of the PB LUC:YFP register module turning on Fluc expression. **(B)** Diagram of the EI PhiC31 experiment in hairy roots. DsRED(+) *N. benthamiana* PB LUC:YFP hairy roots transformed with EI PhiC31 were divided and incubated in MS plates in presence (SET) or absence of estradiol (Mock) for 3 days. After chemoluminescence imaging, roots were transferred to new estradiol-free plates where they remained for 7 days before imaging them again to measure Fluc activity. **(C)** Images of the EI LUC and different EI PhiC31 hairy roots used in the induction experiment. First row images correspond to bright field images (BF), second row are fluorescence images to detect the DsRED marker, third and fourth are luminescence images (FLuc). A LAS3000 imager was used to take all the images at 3 and 10 dpi. Chemoluminescence images were taken with the "Ultra" mode and 1 sec of exposition. **(D)** Genotyping results of uninduced (left panel) and induced (central panel) EI PhiC31 hairy roots, including negative (WT and PB LUC:YFP) and positive (RL LUC:YFP) controls (right panel). A specific pair of primers for RL LUC:YFP amplification was used. **(E)** Quantification of the Fluc/Rluc ratios of induced and mock EI PhiC31 hairy root lines. Bars represent the mean of three technical replicates for an individual root.

We used *A. rhyzogenes* to stably transform the EI PhiC31 system into the previously described PB LUC:YFP line. This transformation system enables the regeneration of roots which are an ideal system for high-throughput chemical screening. We performed the estradiol-induction (EI) experiments as shown in Fig. 5B. Successfully transformed DsRED(+) roots were incubated with or without estradiol for 3 days. Then the roots were evaluated by image analysis of the luminescence resulting from activity of the Fluc reporter. Subsequently, the roots were transferred to new estradiol-free MS plates and incubated for 7 days before imaging again. In parallel, WT *N. benthamiana* was transformed with a construct bearing an estradiol-inducible Fluc reporter as a positive control with no memory device (EI LUC). As shown in Fig. 5C and Fig. S7, roots transformed with EI LUC had an increase of luminescence at 3 days post-induction (3 dpi) and lost most of the signal at 10 dpi. On the other hand, EI PhiC31 root lines maintained activity over time up to day 10, as expected for a memory switch. Unexpectedly, several EI PhiC31 root lines (depicted here lines 2C and 4C) showed high Fluc background in the absence of estradiol, probably due to leaky activation of the estradiol inducible system. Only weak Fluc lines (line 1C) showed no leaky background, and although Fluc intensity at 3dpi was almost undetectable, it reached moderate levels at 10dpi, indicative of a memory-based sustained activation. In order to confirm this, we performed specific PCR-amplification of the RL LUC:YFP configuration (Fig. 5D). P1B-2C and P2B-4C showed RL LUC:YFP amplification before and after induction with estradiol, confirming that the leakiness was caused by an unwanted early switching of the RM. On the other hand, P2B-1C showed no amplification of RL LUC:YFP before induction but strong amplification after induction with estradiol. This result supported our hypothesis that P2B-1C increase in Fluc activity was due to an estradiol-inducible SET operation of the switch. Subsequently, we quantified the luminescence at 10 dpi comparing induced and non-induced roots for each line (Fig. 5E). P1B-2C and P2B-4C lines showed a slight increase of 2-3 fold after induction, while P2B-1C

showed a 16-fold induction and a substantially lower Fluc/Rluc ratio compared to the leakiness-induced roots. In conclusion, these results show that a chemically-inducible PhiC31 actuator that can control the activation of a genetic memory switch is achievable. However, further optimization of the induction system is required to avoid leaky expression of the PhiC31 integrase. A number of factors including the genomic localization of the transgene or its copy number could result in low basal or sporadic expression of the PhiC31 integrase, resulting in an unwanted permanent change in the toggle switch.

3. Discussion

Here we sought to expand the molecular toolbox for PSB by designing a bistable and reversible toggle switch for whole plants. Reports on tools to engineer genetic memory in plants are currently very limited, with a red-light controlled memory switch for protoplasts the only case reported [20]. While this system provides great spatiotemporal resolution of gene induction, its adaptation to full plants is challenging due to strict light requirements as well as limitations performing long-term and inheritable memory storage. Serine integrases are powerful tools that can induce a stable change in the DNA and have extensively been used for many applications, including the design of toggle switches for bacteria and mammalian cells [38]. We adapted the previously described bacterial switch by Bonnet et al., 2012, which was based on the Bxb1 phage integration system [26]. We decided to use the PhiC31 phage integrase and RDF because this integrase has been previously used for genomic engineering in plants, and also suited better the grammar of our GB cloning system. The architecture of the switch was designed to be fully modular, so it can be easily adapted to control the expression of any GB-domesticated downstream gene by performing a single-step GB assembly reaction.

We built multiple RMs to characterize the behavior of the switch at the SET and RESET operations. First, the genetic parts were tested by transient expression of the RMs and the SET and RESET actuators (PhiC31 and PhiC31+RDF), and the evolution of reporter expression in WT *N. benthamiana* was analyzed, both transiently as well as in stably transformed plants. In all experiments, the initial state of the switch showed strong expression of the active reporter while no noticeable backward expression of the inactive reporter was observed by luminescence or CLM imaging. This suggests that the design of the genetic switch successfully avoided leaky backward expression. When the actuators were applied via agroinfiltration, we observed good performance of reporter activation either by SET or RESET operation. Activation reached levels up to 43-fold over uninduced controls by transient expression and 15-fold (60% of the fully active state) in the case of transgenic lines, and it was sustained over days. However, when analyzing inactivation of a reporter gene after SET or RESET, we observed slow decrease in the signal, in occasions reaching only a 40-75% decrease compared to the non-commutated switch. This effect was probably due to the combined effect of slow decay of the reporters and the incomplete switching on the RMs of all the cells analyzed, especially during the RESET operation involving RDF. Phage RDFs have been studied due to their ability to reverse the directionality of a SSR reaction via direct interaction with the cognate integrase, i.e. to catalyze attRxattL reaction [36, 39]. However, this reaction has shown to be inefficient when the stoichiometry of the integrase and RDF is not optimal because integrase unbound to the RDF can catalyze the attPxattB reaction, entering a bidirectional cycle of PB-RL-PB recombination. Bonnet *et al.*, 2012 also observed this effect when developing their bacterial toggle switch based on Bxb1 and Xis (RDF) as separate TUs. In fact, these authors postulated that the system enters in a bidirectional regime if the concentration of the excisionase is too low relative to the integrase. Therefore, these authors tested several combinations of genetic constructs to increase the excisionase expression levels

and reduce the half-life of the integrase achieving an unidirectional reaction with 90% efficiency. We followed a similar approach by increasing RDF expression levels through the CaMV35 promoter, which has about 10-fold higher expression than the Nopaline Synthase promoter used to express the PhiC31 integrase [40]. Additionally, we optimized the OD600 of the *A. tumefaciens* cultures during transient expression to optimize the efficiency of the RESET operation. However, we still saw incomplete attRxattL recombination of all induced cells, as seen by fluorescent imaging of leaf cells and transformed calluses. Recently, Olorunniji et al., 2017 reported higher attRxattL recombination efficiency by creating a fusion of PhiC31 integrase to RDF [41]. Using this new Integrase-RDF fusion as the RESET operator could help to solve bidirectionally issues in future iterations of this plant switch, by achieving higher efficiency off attRxattL recombination without the need of high throughput testing of multiple genetic constructs to balance expression levels of both components.

After individual characterization by agroinfiltration of the SET and RESET operations, we decided to test the performance of a full SET/RESET cycle to evaluate the memory of the switch over plant transformation and regeneration processes. However, a drawback of the agroinfiltration method is the cellular damage produced by the bacteria during the process. This could impose an observational effect which might underestimate the efficiency of the observed recombination and precludes the evaluation of a full SET/RESET cycle by subsequent agroinfiltrations of the actuators in the same leaf. Therefore, we decided to induce SET in a stable PB YFP:LUC plant, to then successfully regenerate *in vitro* new fully switched RL YFP:LUC plants. We monitored the regeneration process with YFP and DNA genotyping finding a number of un-commutated and partially commuted (chimeric) plants. Nevertheless, 30% of the regenerated plants showed complete commutation into RL YFP:LUC RM and constitutive expression YFP, demonstrating that the switch is able to record

the SET event keeping its memory. This also indicates that the SET operation can be fully completed in plant systems as it is in bacteria [41]. Presumably, such change at the DNA level of somatic T0 cells should be transmitted to the progeny as other authors reported in *A. thaliana* for Bxb1 [42] and PhiC31 [43]. Subsequently, we induced RESET on the new generated RL YFP:LUC P7 plant, showing increase of Fluc expression and decrease of YFP. This demonstrated that P7 cells returned to the original PB YFP:LUC, showing a successful SET/RESET cycle over the same RM integrated in the genome. Nevertheless, further optimizations need to be done to overcome the partial efficiency of the RESET, as suggested by the observation of un-commutated RL registers.

Finally, we coupled the expression of PhiC31 integrase to an estradiol-inducible system, as a proof of concept of externally inducing the SET operation to control the toggle switch in *benthamiana* hairy roots. Several transgenic root lines showed unwanted activation of the switch due to leaky expression of the integrase. Chemically-inducible systems with low basal expression have been reported in plants [14-17, 19]; however most of these systems were tested in *Arabidopsis*, and it is possible that chassis-specific conditions (e.g. phytosterol contents) could also determine background expression levels of actuators, which can produce the commutation of the RM. This should be taken into consideration in the design of complex circuits based on PhiC31. Regulation of site-specific DNA excision has been previously reported with the tyrosine recombinase Cre/loxP system in *A. thaliana* using the XVE induction system [44]. In this study, the authors counter-selected the leaky transformation events by flanking the inducible recombinase and the selection marker with the loxP sites. If leaky, the recombinase would excise this cassette and therefore the transformats would be lost in a selection media. In this work we did not apply negative selection pressure, so all the transformation events were kept. Luckily, we were able to recover one root line with low basal

expression but still commutable upon estradiol induction. After activation, low but sustained expression of the Fluc reporter was observed. Similar results were obtained by Hoff et al., 2001 when using heat-shock regulated Cre/loxP [45]. They found the system was leaky but still inducible, and that induction worked better in lines with low reporter gene expression like was our case. In order to circumvent this issue, different strategies could be implemented, such as the use of degradation tags to regulate the half-life of the integrase or an induction system relying on logic gates to regulate expression of the operators. In this regards, a split PhiC31 integrase that can be reconstituted after trans splicing of both components has been developed recently [46].

Summarizing, in this work we created the first modular memory switch for plants making use of the phage PhiC31 integration system. This switch showed to be bistable, reversible and operable, although future efforts are needed in order to optimize the RESET reaction and avoid unwanted activation. Even though the complexity of big multicellular organisms can be a challenge for extensive characterization of new molecular tools, new advances in the field will allow harnessing their potential. We expect that this system will provide a new tool for PSB that will enable engineering plants with more complex gene networks and circuits.

4. Experimental procedures

4.1. Construction and assembly of the GoldenBraid phytobricks

A complete list of all phytobricks used in this work can be found in Table S1 and their sequence information and assembly history is available at the GoldenBraid online repository (https://gbcloning.upv.es/). The list of oligonucleotides used to build them can be found on Table S2. The level 0 parts of the PhiC31 integrase and RDF were assembled as previously described to obtain the SET and RESET actuators (GB1531 and GB1531+GB1508 respectively) [10]. The registers were designed and assembled as described in Figure S1. In detail, register PB Level 0 part (GB1495) was assembled in a 2 step PCR, (i) addition of attP to the 5′ inverse sequence of TMtb using primers ALF15EN04 and ALF15NOV03 (ii) addition of the attB to the 3′ of P35s (excluding its RBS) using primers ALF15EN05 and ALF15NOV04. Both parts (i) and (ii) were assembled at Level 0 in pUPD2 through a BsmbI-mediated restriction-ligation reaction. Register RL (GB 1514) was amplified from genomic DNA (gDNA) of transiently transformed *N. benthamiana* with register PB and PhiC31 integrase using primers ALF15DIC06 and ALF15DIC07. The reporter genes to be assembled into the register module were amplified from a pre-assembled TU of the GoldenBraid collection as described in Fig. S1B-C. In brief, TUs containing a P35S, followed by a CDS (DsRED, YFP or Fluc) and a Terminator were amplified using primers to obtain a 5′UTR(RBS):CDS:Terminator amplicon to be assembled as a Level 0 part; either to be used as the reverse CDS or direct CDS for the register. Reverse CDSs were amplified using primers ALF15DIC08 and ALF15DIC10 depending on the terminator used in the TU, Tnos and T35s respectively. Direct CDSs were amplified using the primer pairs ALF15DIC09 and ALF15DIC10 depending on the terminator used in the TU, Tnos and T35s respectively. The register modules (RMs) were assembled in a Level 1 reaction

using three Level 0 parts: (i) reverse CDS, (ii) register PB or RL and (iii) direct CDS. All assemblies were performed as regular GB reactions [47].

4.2. Time-dependent characterization of reporter expression in transiently expressed and integrated register modules

Transient expression experiments were performed to test the SET and RESET recombination processes as described in the main text. For the initial characterization by transient expression, equal volumes of Agrobacterium cultures transformed with the register modules (GB1523, GB1524, GB1527, GB1528) and their appropriate operators were mixed (GB1531 (SET), GB1508+GB1531 (RESET) or GB0106 (SF negative control) and agroinfiltrated in wild type (WT) N. benthamiana plants. For the transgenic lines, only the aforementioned operators or GB0108 (P19) as negative control were used. These cultures were first grown from glycerol stocks for two days until saturation, then 10 uL were sub-cultivated for 16h. Next, the cultures were pelleted, resuspended in the agroinfiltration buffer (10 mM MES, pH 5.6, 10 mM MgCl2, and 200 µM acetosyringone) and adjusted to the appropriated optical density (Table S3). Four to five weeks-old N. benthamiana plants cultivated under a 24°C (light)/20°C (darkness) 16-h-light/8-h-dark photoperiod were used. One plant per experimental point was used (1 to 7 days post infiltration (dpi)). Three consecutive leaves (third to fifth from the base) were infiltrated, half with the recombination mixture and half with the negative control. At each time-point, one sample per infiltrated leaf was collected using a 0.8 cm cork-borer (approximately 20 mg of tissue for the luminescence quantification assays and 150 mg for the gDNA extraction) and immediately frozen in liquid nitrogen. Samples were grinded with a Retsch Mixer Mill MM400 for 1 min at 30 Hz and stored at -80°C for subsequent analysis.

4.3. Confocal laser microscopy

At the stage of 5 dpi agroinfiltrated leaves of N. benthamiana were examined under a ZEISS 780 AxioObserver Z1 confocal laser microscope to visualize the yellow fluorescent protein YFP (λex = 514 nm; λem = 518-562 nm) and DsRED (λex = 561 nm; λem = 563-642 nm) fluorescence. Images of 9-16 tiles were taken to visualize a larger area and processed with ZEN lite 2.5 lite and FIJI softwares.

4.4. Firefly and renilla luciferase luminescence quantification

Firefly lucirerase (Fluc) and renilla luciferase (Rluc) activities were determined using the DualGlo® Luciferase Assay System (Promega) manufacturer's protocol with minor modifications. Homogenized leaf disc samples were extracted with 375µl of 'Passive Lysis Buffer,' vortexed and then centrifuged for 10 min (14,000×g) at 4 °C. Then, 7.5 µl of supernatant were transferred to a 96-well plate and 30 µl of LARII added to quantify the Fluc activity. Finally, 30 µl of Stop&Glow Reagent were used to measure the Rluc activity. Measurements were made using a GloMax 96 Microplate Luminometer (Promega) with a 2s delay and a 10s measurement. Fluc/Rluc ratios were determined as the mean value of three biological replicates coming from three independent agroinfiltrated leaves of the same plant.

4.5. Generation of *N. benthamiana* transgenic plants

Fully expanded leaves were sterilized and used to obtain 0.5 cm diameter leaf-discs with a cork-borer. After an overnight incubation in co-culture medium (MS supplemented with vitamins (Duchefa), 3% sucrose (Sigma-Aldrich), 0.8% Phytoagar (Duchefa), 1 µg/mL BAP (Sigma-Aldrich), 0.1 µg/mL NAA (Sigma-Aldrich)) the leaf-discs were inoculated for 15 min with the *A. tumefaciens* strain LBA4404 carrying the register module construct (Fig. S2, Table S1). Then, the discs were returned to the co-cultivation

medium and incubated for 2 more days in darkness. Next, discs were placed in the selection medium (MS supplemented with vitamins (Duchefa), 3% sucrose (Sigma-Aldrich), 0.8% Phytoagar (Duchefa), 1 µg/mL BAP (Sigma-Aldrich), 0.1 µg/mL NAA (Sigma-Aldrich), 500 µg/mL carbenicillin, 100 µg/mL kanamycin. Discs were transferred to fresh medium periodically until the callus and then the explants were formed (5-8 weeks). Explants were excised and planted in rooting medium (1/2 MS supplemented with vitamins (Duchefa), 3% sucrose (Sigma-Aldrich), 0.8% Phytoagar (Duchefa), 0.1 µg/mL NAA (Sigma-Aldrich), 100 µg/mL kanamycin) until enrooted. In vitro cultivation were carried out in a long day growth chamber (16 h light/ 8 h dark, 25ºC, 60-70% humidity, 250 $\mu mol \cdot m^{-2} \cdot s^{-1}$ photons). Samples for phenotyping were collected once the plants were sufficiently developed to harvest 20 mg of tissue to perform the Fluc and YFP quantifications. T1 seeds of Fluc and YFP positive plants were selected for a segregation analysis of the transgene in kanamycin plates. Those scored as a single copy by the chi-square test were used for subsequent experiments.

4.6. Generation of *N. benthamiana* transgenic hairy roots

WT and PB LUC:YFP T1-2 *N. benthamiana* plants were used for transformation with *Agrobacterium rhyzogenes* strain 15834 transformed with the EI LUC or EI PhiC31 constructs (Fig. 5A, Table S1). Fully expanded leaves were used to obtain sterilized leaf-discs which were cultivated for one day in co-cultivation medium without hormones. Saturated agrobacterium cultures were sub-cultivated and grown overnight. The culture medium was removed by centrifugation at 4.500 rpm for 15 min and then the bacterial pellet resuspended and adjusted with MS to an OD600 = 0.3. Those cell cultures were employed for inoculation of the sterilized leaf-discs for 30 min under agitation. Next, the leaf-discs were blotted, placed on in co-cultivation plates and chilled in darkness for 2 days. After that, the discs were transferred

to selection plates without kanamycin nor hormones. The medium was renewed periodically until roots emerged. Transformed roots were identified using a Leica MacroFluo MZZ16F with a DsRED filter. See "Generation of *N. benthamiana* transgenic plants" for the culture medium and growth conditions.

4.7. Estradiol induction experiments

Individual DsRED(+) hairy roots transformed with the EI LUC and EI PhiC31 constructs were isolated and divided in two with a scalpel and then separated in MS medium supplemented or not with 20 µM β-estradiol (Sigma-Aldrich). After 3 days, bright field and red fluorescence images were acquired using a FujiFilm LAS-3000 imager. Then, 100 µM luciferin diluted in MS was added and roots incubated for 30 min after the acquisition of the chemiluminescence images with the same device. Finally, all the roots were transferred to new MS medium without estradiol and grown for 7 days to repeat the imaging process. For the chemiluminescent imaging, the "Ultra" mode was employed with 1 sec of exposition time. Samples of induced roots were collected and stored at -80ºC to perform Fluc/Rluc quantification and PCR-analysis. The growing conditions of "Generation of N. benthamiana transgenic plants" applies for this section too.

4.8. YFP quantification

Samples of T0 *N. benthamiana* transgenic plants were used for the YFP quantification using a VICTOR X5 2030 (PerkinElmer), excitation filter F485 and emission filter F535-40. Grinded samples were added with 150 µL of PBS 1X, vortexed and then centrifuged at 13.000 rpm for 10 min at a 4ºC. Finally, 10 µL of the extract were pipetted to a 96-well plate and measured with the spectrofluorometer. The YFP values are expressed as the mean of the measure of three independent leaves from the same plant.

4.9. Genomic DNA extraction and PCR amplification

Genomic DNA was extracted following the CTAB protocol [48]. The obtained gDNA was used for amplification of the register module using MyTaqTM DNA Polymerase (Bioline) and a pair of specific primers for the PB or RL configuration and the PhiC31 recombinase (Table S2). DNA amplification was confirmed on 1% agarose gel electrophoresis.

5. References

1. Liu W, Neal Stewart C: **Plant synthetic biology**. *Trends in Plant Science* 2015, **20**(5):309-317.
2. Baltes NJ, Voytas DF: **Enabling plant synthetic biology through genome engineering**. *Trends in biotechnology* 2015, **33**(2):120-131.
3. Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I: **Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm**. *Science* 2000, **287**(5451):303-305.
4. South PF, Cavanagh AP, Liu HW, Ort DR: **Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field**. *Science* 2019, **363**(6422).
5. Rogers C, Oldroyd GED: **Synthetic biology approaches to engineering the nitrogen symbiosis in cereals**. *J Exp Bot* 2014, **65**(8):1939-1946.
6. Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP: **Improving photosynthesis and crop productivity by accelerating recovery from photoprotection**. *Science* 2016, **354**(6314):857-861.
7. Caemmerer Sv, von Caemmerer S, Quick WP, Furbank RT: **The Development of C4 Rice: Current Progress and Future Challenges**. *Science* 2012, **336**(6089):1671-1672.
8. Xie M, Fussenegger M: **Designing cell function: assembly of synthetic gene circuits for cell biology applications**. *Nature Reviews Molecular Cell Biology* 2018, **19**(8):507-525.
9. Kassaw TK, Donayre-Torres AJ, Antunes MS, Morey KJ, Medford JI: **Engineering synthetic regulatory circuits in plants**. *Plant Sci* 2018, **273**:13-22.
10. Vazquez-Vilar M, Quijano-Rubio A, Fernandez-Del-Carmen A, Sarrion-Perdigones A, Ochoa-Fernandez R, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data**. *Nucleic Acids Res* 2017, **45**(4):2196-2209.
11. Vazquez-Vilar M, Orzaez D, Patron N: **DNA assembly standards: Setting the low-level programming code for plant biotechnology**. *Plant Sci* 2018, **273**:33-41.

12. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli**. *Nature* 2000, **403**(6767):339-342.
13. Chiu T-Y, Jiang J-HR: **Logic Synthesis of Recombinase-Based Genetic Circuits**. *Sci Rep* 2017, **7**(1):12873.
14. Roslan HA, Salter MG, Wood CD, White MR, Croft KP, Robson F, Coupland G, Doonan J, Laufs P, Tomsett AB *et al*: **Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana**. *Plant J* 2001, **28**(2):225-235.
15. Zuo J, Niu QW, Chua NH: **Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants**. *Plant J* 2000, **24**(2):265-273.
16. Aoyama T, Chua NH: **A glucocorticoid-mediated transcriptional induction system in transgenic plants**. *Plant J* 1997, **11**(3):605-612.
17. Mett VL, Podivinsky E, Tennant AM, Lochhead LP, Jones WT, Reynolds PH: **A system for tissue-specific copper-controllable gene expression in transgenic plants: nodule-specific antisense of aspartate aminotransferase-P2**. *Transgenic Res* 1996, **5**(2):105-113.
18. Saijo T, Nagasawa A: **Development of a tightly regulated and highly responsive copper-inducible gene expression system and its application to control of flowering time**. *Plant Cell Rep* 2014, **33**(1):47-59.
19. Padidam M, Gore M, Lu DL, Smirnova O: **Chemical-inducible, ecdysone receptor-based gene expression system for plants**. *Transgenic Res* 2003, **12**(1):101-109.
20. Müller K, Siegel D, Rodriguez Jahnke F, Gerrer K, Wend S, Decker EL, Reski R, Weber W, Zurbriggen MD: **A red light-controlled synthetic gene expression switch for plant systems**. *Mol Biosyst* 2014, **10**(7):1679-1688.
21. Inniss MC, Silver PA: **Building Synthetic Memory**. *Current Biology* 2013, **23**(17):R812-R816.
22. Roquet N, Lu TK: **Digital and analog gene circuits for biotechnology**. *Biotechnol J* 2014, **9**(5):597-608.
23. Smith MCM, Brown WRA, McEwan AR, Rowley PA: **Site-specific recombination by phiC31 integrase and other large serine recombinases**. *Biochem Soc Trans* 2010, **38**(2):388-394.
24. Grindley NDF, Whiteson KL, Rice PA: **Mechanisms of site-specific recombination**. *Annu Rev Biochem* 2006, **75**:567-605.

25. Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ: **Synthetic gene networks that count**. *Science* 2009, **324**(5931):1199-1202.
26. Bonnet J, Subsoontorn P, Endy D: **Rewritable digital data storage in live cells via engineered control of recombination directionality**. *Proc Natl Acad Sci U S A* 2012, **109**(23):8884-8889.
27. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D: **Amplifying genetic logic gates**. *Science* 2013, **340**(6132):599-603.
28. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK: **Synthetic recombinase-based state machines in living cells**. *Science* 2016, **353**(6297):aad8559.
29. Siuti P, Yazbek J, Lu TK: **Synthetic circuits integrating logic and memory in living cells**. *Nat Biotechnol* 2013, **31**(5):448-452.
30. Daniel R, Rubens JR, Sarpeshkar R, Lu TK: **Synthetic analog computation in living cells**. *Nature* 2013, **497**(7451):619-623.
31. Lyznik LA, Gordon-Kamm WJ, Tao Y: **Site-specific recombination for genetic engineering in plants**. *Plant Cell Reports* 2003, **21**(10):925-932.
32. Wang Y, Yau Y-Y, Perkins-Balding D, Thomson JG: **Recombinase technology: applications and possibilities**. *Plant Cell Rep* 2011, **30**(3):267-285.
33. Petolino JF, Kumar S: **Transgenic trait deployment using designed nucleases**. *Plant Biotechnol J* 2016, **14**(2):503-509.
34. Srivastava V, Thomson J: **Gene stacking by recombinases**. *Plant Biotechnol J* 2016, **14**(2):471-482.
35. Lutz KA, Corneille S, Azhagiri AK, Svab Z, Maliga P: **A novel approach to plastid transformation utilizes the phiC31 phage integrase**. *Plant J* 2004, **37**(6):906-913.
36. Khaleel T, Younger E, McEwan AR, Varghese AS, Smith MCM: **A phage protein that binds ϕC31 integrase to switch its directionality**. *Mol Microbiol* 2011, **80**(6):1450-1463.
37. Sheth RU, Wang HH: **DNA-based memory devices for recording cellular events**. *Nat Rev Genet* 2018, **19**(11):718-732.
38. Merrick CA, Zhao J, Rosser SJ: **Serine Integrases: Advancing Synthetic Biology**. *ACS Synth Biol* 2018, **7**(2):299-310.
39. Ghosh P, Wasil LR, Hatfull GF: **Control of phage Bxb1 excision by a novel recombination directionality factor**. *PLoS Biol* 2006, **4**(6):e186.
40. Vazquez-Vilar M, Bernabé-Orts JM, Fernandez-Del-Carmen A, Ziarsolo P, Blanca J, Granell A, Orzaez D: **A modular toolbox for**

gRNA-Cas9 genome engineering in plants based on the GoldenBraid standard. *Plant Methods* 2016, **12**:10.

41. Olorunniji FJ, McPherson AL, Rosser SJ, Smith MCM, Colloms SD, Stark WM: **Control of serine integrase recombination directionality by fusion with the directionality factor**. *Nucleic Acids Res* 2017, **45**(14):8635-8645.
42. Thomson JG, Chan R, Smith J, Thilmony R, Yau Y-Y, Wang Y, Ow DW: **The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in Arabidopsis**. *BMC Biotechnol* 2012, **12**:9.
43. Thomson JG, Chan R, Thilmony R, Yau Y-Y, Ow DW: **PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome**. *BMC Biotechnol* 2010, **10**:17.
44. Zuo J, Niu QW, Møller SG, Chua NH: **Chemical-regulated, site-specific DNA excision in transgenic plants**. *Nat Biotechnol* 2001, **19**(2):157-161.
45. Hoff T, Schnorr KM, Mundy J: **A recombinase-mediated transcriptional induction system in transgenic plants**. *Plant Mol Biol* 2001, **45**(1):41-49.
46. Olorunniji FJ, Lawson-Williams M, McPherson AL, Paget JE, Stark M, Rosser SJ: **Control of φC31 integrase-mediated site-specific recombination by protein trans splicing**.
47. Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology**. *Plant Physiol* 2013, **162**(3):1618-1631.
48. Murray MG, Thompson WF: **Rapid isolation of high molecular weight plant DNA**. *Nucleic Acids Research* 1980, **8**(19):4321-4326.

6. Supplementary figures

Figure S1: design and cloning details of a custom-made toggle switch. **(A)** Detailed view of the phytobricks comprising the plant toggle switch. The GOIs are cloned as individual parts in Level 0 using the designated overhangs and then assembled with the PB or RL promoter elements to create a construct for the bistable regulation of each CDS. **(B-D)** Assembly procedure of custom PB and RL register modules using GoldenBraid (GB) cloning system. **(B)** Preparation of the direct and reverse genes of interest (GOIs) for the combination with the PB or RL registers. This involves a BsaI-mediated assembly of Level 1 transcriptional units using Level 0 phytobricks of a promoter, a CDS of the GOI, and a terminator which can be cloned from scratch or reused from the GB's collection. **(C)** Next, these plasmids are used as a PCR-templates to add a couple of overhangs (OH) which define each GOI as reverse or direct CDS and allow its subsequent cloning in a Level 0 plasmid, through a BsmbI-mediated reaction. Standard primers listed in the Table S2 can be used for this amplification. **(D)** Finally, both GOIs cloned as Level 0 parts can be combined with the register PB (GB1494) or the register RL (GB1506) in the Level 1 to conform the PB or RL register modules for the regulated expression of the two GOIs.

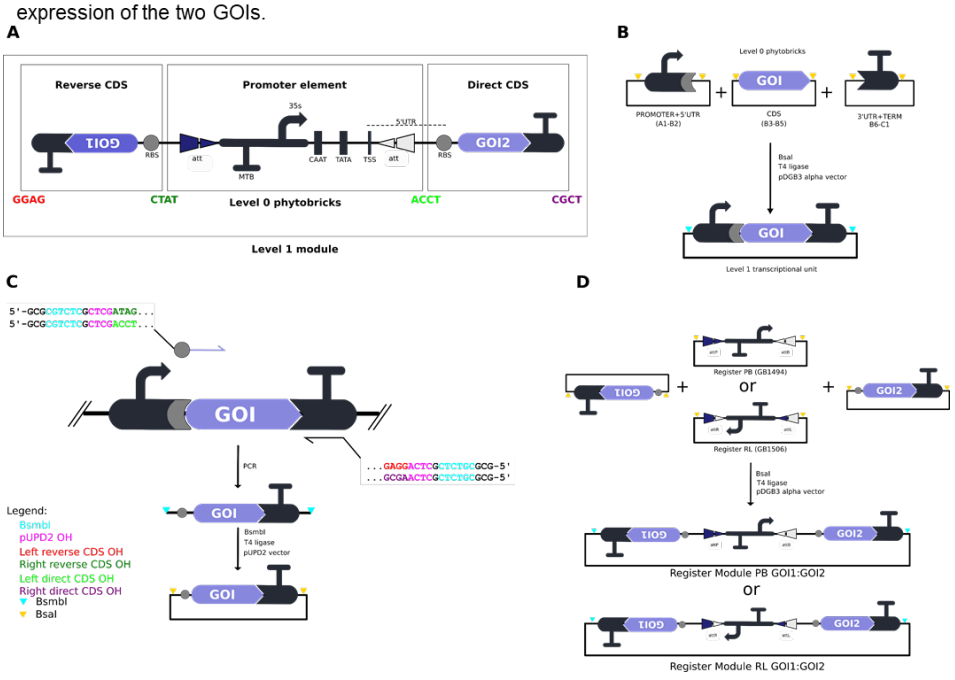


Figure S2: architecture of the register modules (RMs) used in the transient expression experiments (A) and for the generation of transgenic lines (B).

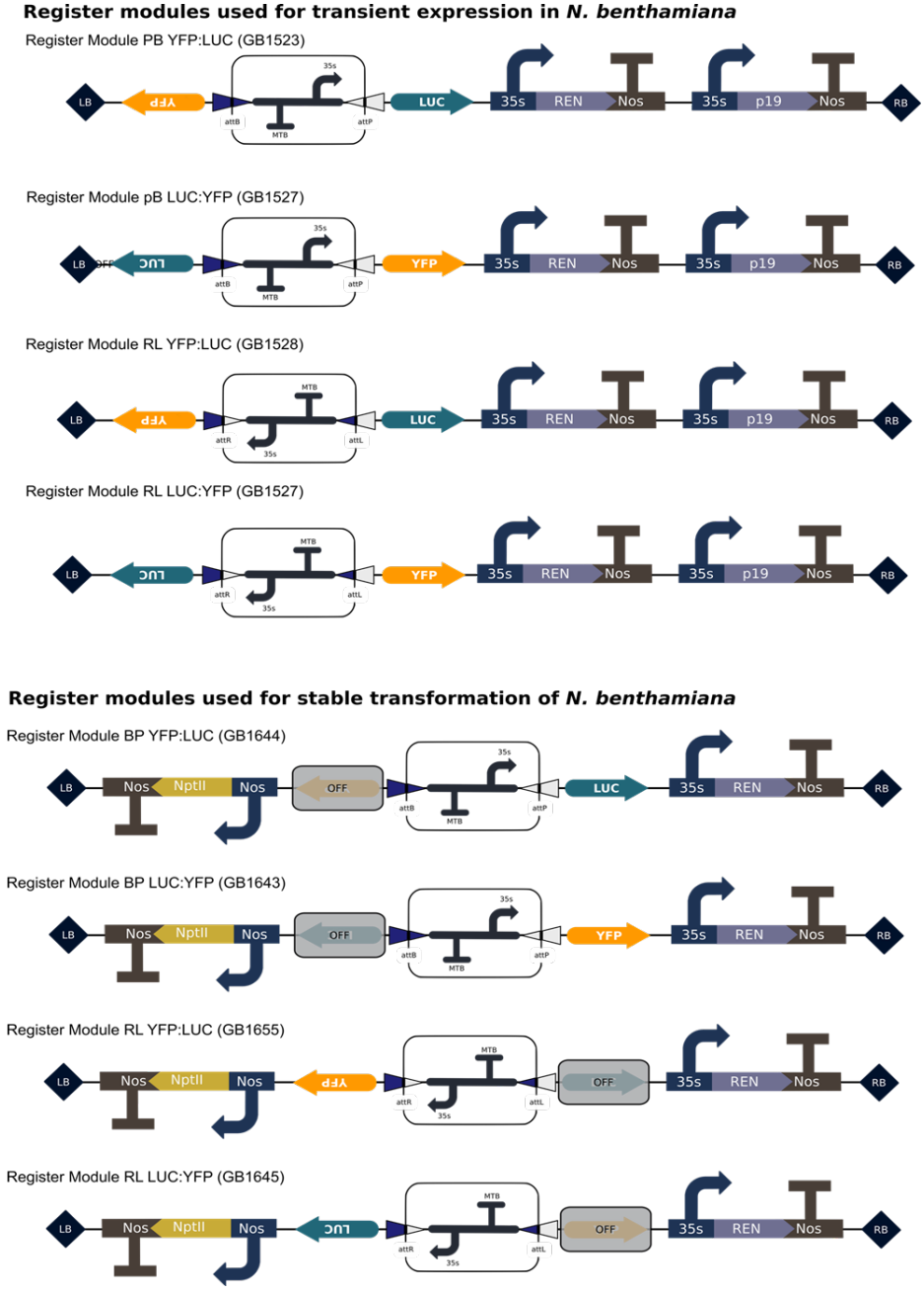


Figure S3: phenotyping results of the transgenic lines obtained with the transformation with the RMs. Both, the luminescence (Fluc) and the fluorescence intensity (FI) was measured for each line and are indicated in arbitrary units (a.u.). Marked lines were selected for further experiments in the T1 generation. The bars shows the mean values ± SD of three replicates.

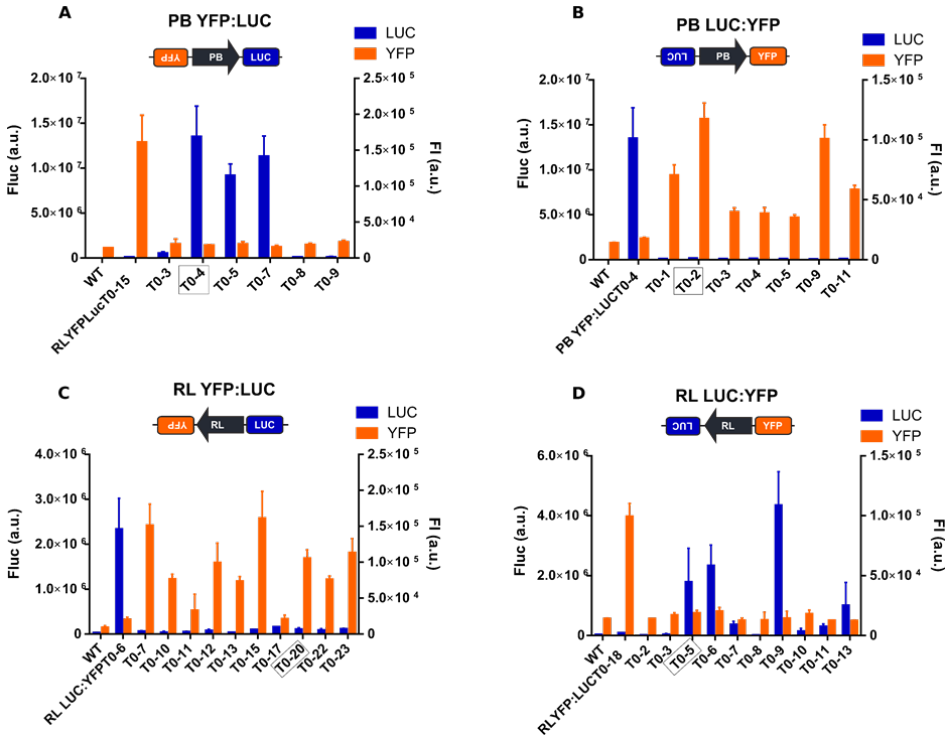


Figure S4: influence of the optical density (OD) of the *A. tumefaciens* cultures encoding the SET or RESET operators on the recombination of RMs in *N. benthamiana* stable lines. **A)** Effect of the OD of PhiC31 cultures in the SET process of the PB LUC:YFP T1-2 line. **B)** Effect of the OD of PhiC31 cultures in the SET process of the PB YFP: T1-4 line. **C)** Effect of the ratio of PhiC31:RDF on the reset of the RL LUC:YFP T1-5 line. Ratios are obtained modifying the OD of PhiC31 and RDF cultures. **D)** Effect of the ratio of PhiC31:RDF on the reset of the RL YFP:LUC T1-20 line. Ratios are obtained modifying the OD of PhiC31 and RDF cultures. Bars indicate the fold change between the Fluc/Rluc ratios of treated (PhiC31 or PhiC31-RDF) and the untreated sample (P19), expressed as mean ± SD of three agroinfiltrated leaves.

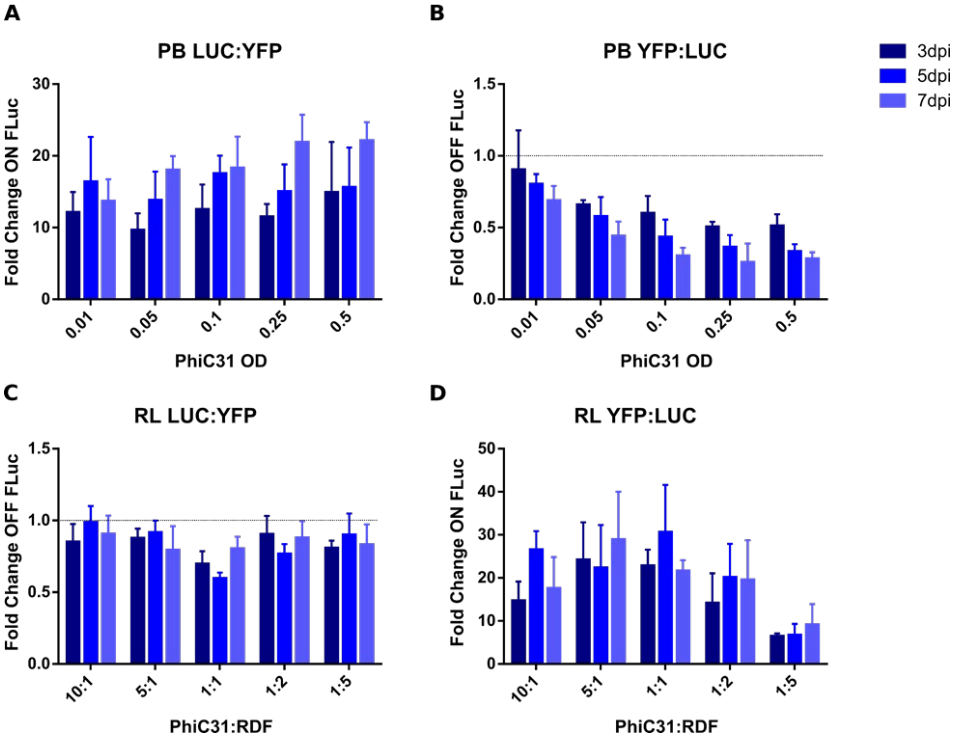


Figure S5: Quantification of the fluorescence intensity (FI) of the confocal images showed in Fig. 3. **(A)** FI of the YFP switch on of recombined (+) or not recombined samples (-) for the SET and RESET operations of PB YFP:LUC and RL LUC:YFP register modules. **(B)** Quantification of the YFP switch off of recombined (+) or not recombined samples (-) for the SET and RESET operations of PB LUC:YFP and RL YFP:LUC register modules. Bars shows the mean FI ± SD of nine images.

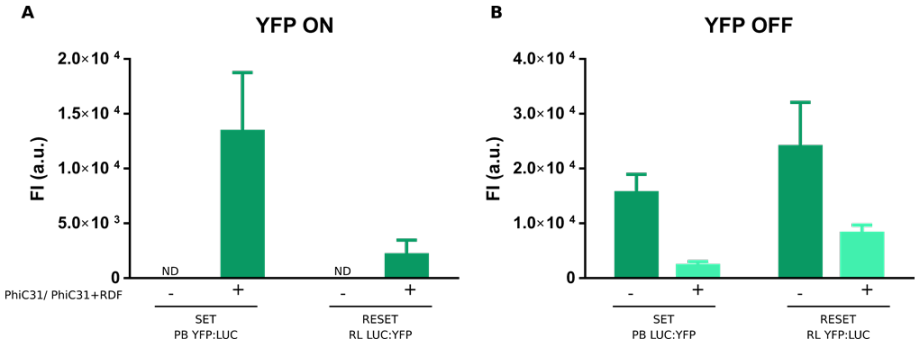
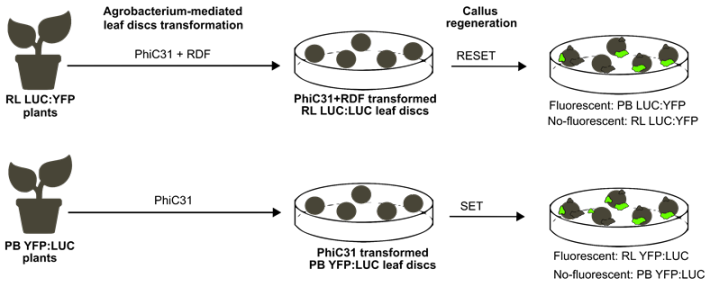
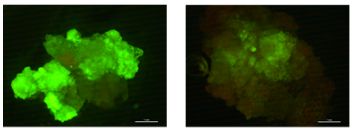


Figure S6: Recording the T-DNA expression with the PB and RL register modules. **(A)** Representation of the transformation experiment. Non-fluorescent transgenic RL LUC:YFP and PB YFP:LUC leaf discs were transformed with *A. tumefaciens* carrying a T-DNA with its appropriate actuator (PhiC31+RDF or PhiC31). Leaf discs were cultivated *in vitro* until the callus emerged. If the T-DNA was being expressed, then the presence of the effectors would be registered by the RESET and SET recombination processes which would turn on the YFP expression giving fluorescent callus. **(B)** SET of the PB YFP:LUC and RESET of the RL LUC:YFP generate fluorescent callus visible under the magnification microscope. Scale-bar represents 1 mm. **(C)** Efficiency of the SET and RESET recombination estimated as the number of fluorescent callus respect the total of leaf discs.

A



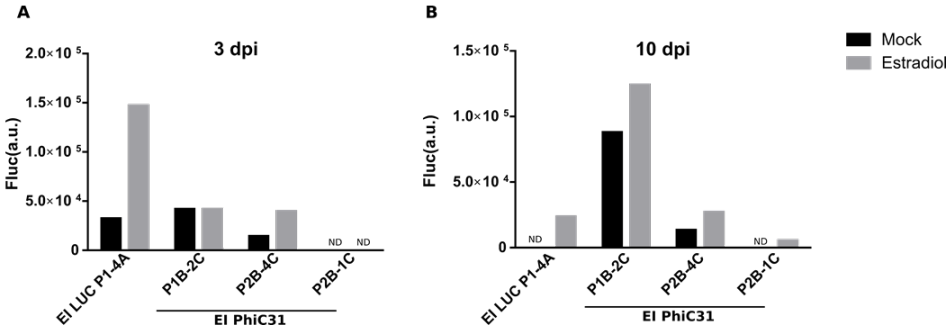
B



C

	SET	RESET
Total fluorescent callus	38	16
Total leaf discs	103	110
% Efficiency	**37**	**15**

Figure S7: Densitometry of luminescence signal showed in the Fig. 5C image. **(A)** Quantification of the luminescence signal displayed by estradiol-inducible (EI LUC) and PhiC31-inducible (EI PhiC31) roots incubated for 3 days in estradiol-free (Mock) or estradiol-containing (Estradiol) MS plates. **(B)** Quantification over the same roots 7 days after the estradiol removal. ND means no detected.



7. Supplementary tables

Table S1: Constructs generated in this study. Sequences are accessible at GB cloning website using the GB database ID.

Level 0 GB phytobricks		
GB database ID	**Name**	**Category**
1481	Assembly Register 1/3 Part- T35S:DsRed:5'UTR	Other
1483	Assembly Register 3/3 Part - 5'UTR:YFP:Tnos	Other
1494	Assembly Register 2/3 Part - PhiC31 PB (attP:TMtb:P35S:attB)	Other
1496	PhiC31 integrase (Plant Codon Optimized)	B3-B4-B5
1498	RDF (gp3)	B3-B4-B5
1499	Assembly Register 1/3 Part - TNos:Luciferase:5'UTR	Other
1500	Assembly Register 3/3 Part - 5'UTR:Luc:Tnos	Other
1506	Assembly Register 2/3 Part - PhiC31 RL (attR:P35S:TMtb:attL)	Other
1507	Register 1/3 Tnos:YFP:5'UTR	Other

Level ≥1 GB phytobricks		
GB database ID	**Name**	**Category**
460	P35s:DsRED:T35s-SF	Module
1129	35S:ER:LexADBD/Gal4AD:T35S	TU
1130	OpLexA:mini35S:Luciferase:T35s	TU
1131	35s:ER:LexABD/Gal4AD:T35s-OpLexA:mini35s:Luciferase:Tnos	Module
1495	Register PhiC31 PB (DsRed:YFP)	Other
1497	P35S:phiC31:T35S	TU
1508	P35S:RDF:T35S	TU
1510	Register phiC31 RL (DsRed:YFP)	Other
1513	Register phiC31 PB (Luc:YFP)	Other

1514	Register phiC31 RL (Tnos:Luc:attR:P35S:TMtb:attL:YFP:TNos)	Other
1517	Register phiC31 PB (YFP:Luc)	Other
1518	Register phiC31 RL (YFP:Luc)	Other
1523	Register PB phiC31 Luc:YFP-P35S:Renilla:TNos-P35S:p19:Tnos	Module
1524	Register RL phiC31 Luc:YFP-P35S:Renilla:TNos-P35S:p19:Tnos	Module
1527	Register PB phiC31 YFP:Luc-P35S:Renilla:TNos-P35S:p19:Tnos	Module
1528	Register RL phiC31 YFP:Luc-P35S:Renilla:TNos-P35S:p19:Tnos	Module
1529	OplexA:mini35S:phiC31:Tnos	TU
1531	Pnos:phiC31:Tnos	TU
1532	P35S:ER:lexABD/Gal4AD:T35S - OplexA:mini35S:phiC31:Tnos	Module
1601	Pnos:ER:LexABD:GAL4AD:Tnos	TU
1643	Tnos:NptII:Pnos-PB LUC:YFP-P35s:Rluc:T35s	Module
1644	Tnos:NptII:Pnos-PB YFP:LUC-P35s:Rluc:T35s	Module
1645	Tnos:NptII:Pnos-RL LUC:YFP-P35s:Rluc:T35s	Module
1655	Tnos:NptII:Pnos-RL YFP:LUC-P35s:Rluc:T35s	Module
1677	Pnos:ER:lexABD:GAL4AD:T35s-OplexA:mini35S:phi31:Tnos	Module
2060	NOS:PhiC31:TNOS - 35s:RDF:T35s	Module
2313	NOS:ER:LexABD:GAL4AD:T35s - OpLexA:mini35s:PhiC31:TNOS - 35s:DsRED:T35s - SF	Module
2388	35s:ER:LexABD:GAL4AD:T35s - OpLexA:mini35s:LUC:Tnos - 35s:DsRED:T35s - SF	Module

Table S2: oligonucleotides used in this study.

Name	Sequence (5'-3')
JO18SEP01 RL YFPLUC F	CAGAGCAGAGATCATGGTGTTAG
JO18SEP02 RL YFPLUCR	GCATACGACGATTCTGTGATTTG
JO18SEP05 BP YFPLUC F	TTGTGGCTGTTGTAGTTGTACTC
JO18SEP06 BP YFPLUCR	ATCATGGTGTTAGCCTTCTATGG
JO18SEP03 PhiC31 F	GTTGAATTAGACTGTGGACCGAT
JO18SEP04 PhiC31 R	ATCTTGTGCATCGTCTTCATCAT
ALF15EN04	GCGCGTCTCGACGAAAATATAGTTGAAACAGA
ALF15EN05	GCGCGTCTCGTCGTACTAGAGCCAAGCTGATCTC
ALF15NOV03	GCGCGTCTCGCTCGCTATAGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGTCGCAAAAACTATATGCTCT
ALF15NOV04	GCGCGTCTCGCTCAAGGTCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACTAGTAAATTGTAATGTTGTTTGTTG
ALF15DIC06	GCGCGTCTCGCTCGCTATAGTAGTGCCCCAACTGGG
ALF15DIC07	GCGCGTCTCGCTCAAGGTCGGTGCGGGTGCCA
ALF15DIC08	GCGCGTCTCGCTCGATAGAAACAACATTACAATTTACTATTCTAGTCGA
ALF15DIC09	GCGCGTCTCGCTCGACCTAAACAACATTACAATTTACTATTCTAGTCGA
ALF15DIC10	GCGCGTCTCGCTCAGGAGCGAGTCGGTCCCATT
ALF15DIC11	GCGCGTCTCGCTCAGGAGAGGTCACTGGATTTTGGTTTTAGG
ALF15DIC12	GCGCGTCTCGCTCAAGCGAGGTCACTGGATTTTGGTTTTAGG
ALF15DIC13	GCGCGTCTCGCTCAAGCGCGAGTCGGTCCCATT

Table S3: agroinfiltration cultures and their respective OD600 used in the experiments of Fig. 3 and Fig. S4.

Optimization of PhiC31		Optimization of PhiC31+RDF		PB register kinetics		RL register kinetics	
Culture	OD	Culture	OD	Culture	OD	Culture	OD
PhiC31	0.01	PhiC31	0.1	PhiC31	0.1	PhiC31	0.1
	0.05	RDF	0.01	NA	NA	RDF	0.1
	0.1		0.05	NA	NA		
	0.25		0.1	NA	NA		
	0.5		0.2	NA	NA		
			0.5	NA	NA		
P19	0.1	P19	0.1	2 vol. P19	0.1	P19	0.1

CHAPTER IV

General Discussion and Future Perspectives

In this thesis, we have provided a new set of tools for plant biotechnology expanding the available resources for genome engineering with CRISPR/Cas12a, and with a new switch based on the integration system of the bacteriophage PhiC31. Detailed discussions about the implementation of these tools in the GB3.0 cloning system and their functional characterization have already been provided in Chapters II and III of this thesis. As the final goal of this work is to provide additional tools for plant breeding, this General Discussion and Future Perspectives section will be focused on the contribution of CRISPR/Cas systems, and particularly Cas12 to precision plant breeding. We will also discuss how transgenic breeding can evolve hand in hand with Plant Synthetic Biology thanks to the development of more refined phytobricks such as the toggle switch presented in Chapter III. Finally, we will try to envision some future applications resulting from connecting CRISPR/Cas systems (Cas9 and Cas12a) to PhiC31-based toggle switch.

1. Uses of CRISPR/Cas12a in precision plant breeding and beyond

1.1. Targeting non-coding regulatory regions with Cas12a

In Chapter II, Cas12a was successfully targeted to non-coding regions such as the intron of *N. benthamiana XT1* gene, or the UTRs of the *A. thaliana PDS3*. As discussed earlier, targeting non-coding-T/A-rich regions is another of the main advantages of Cas12, as this broadness the range of breeding strategies. By targeting regulatory elements, a selectable range of transcriptional alleles can be generated, enabling precise fine-tuning of desirable traits (Fig. 1A). Recently, Lippman et al., [12] developed a genetic scheme for rapid generation and evaluation of novel transcriptional alleles by targeting promoter regions of tomato quantitative trait–related genes such as *SlCLV3*, *SlS*, and

SlSP, thus creating a continuum of variation and leading to the selection of mutated alleles with improved yield. Other well-known *cis*-regulatory elements are upstream open reading frames (uORFs) which are protein coding-regions with a transcriptional start codon before the primary open reading frame (pORF). As Fig. 1B depicts, these uORFs act as post-transcriptional inhibitors of the downstream primary pORF. The uORFs are widespread in plants. In *Arabidopsis* for example, more than 35% of the mRNAs contain uORFs in its 5'UTR. Caixia Gao´s laboratory demonstrated that CRISPR-mediated disruption of uORFs can be used for increasing the production of a specific protein by enhancing translation of the respective mRNA [13]. Targeting uORFs of two lettuce genes involved in the ascorbic acid production (*LsGGP1* and *LsGGP2*) increased the content of this antioxidant, enhancing the resistance to an herbicide known to produce oxidative stress [13].

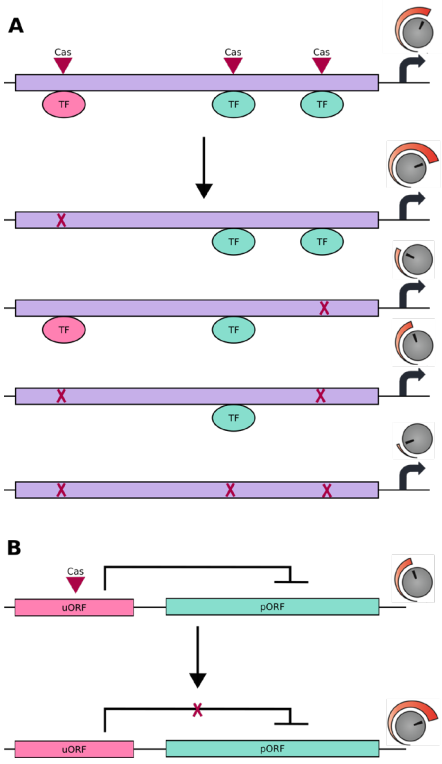


Figure 1: CRISPR/Cas-driven diversity through mutation of non-coding regulatory regions. A) Transcriptional output can be graduated when mutating binding sites of a promoter or enhancer region where activation (pale green) or repressor (pale pink) transcription factors bind. **B)** Similar output can be obtained if an upstream open reading frame (uORF) is mutated. Transcription of these regulatory elements acts as post-transcriptional inhibitors of the primary ORF (pORF) transcription, which encodes for the actual gene product.

1.2.New multiplexing capacities with Cas12a

In Chapter II of this thesis (subsection 2.1), we showed the multi-targeting ability of Cas12a using individual gRNA expression cassettes or tandemly arrayed constructs. Multiplexing is a key feature of editing systems as it enables fast pyramiding of beneficial traits into an elite background in just a few generations. Multiple examples demonstrating the robustness of CRISPR/Cas9-mediated knock-out of simultaneous coding sequences can be traced in the literature, impacting different crucial plant breeding traits such as yield or quality [1-4], biotic and abiotic stress resistance [5, 6] or clonal seed production [7].

Furthermore, multiplex editing could impact the biodiversity of cultivated species by enabling fast neo-domestication processes. Selective breeding during thousands of years led to the modern domesticated crops with improved yield, architecture and nutritional characteristics, but at the same time this process reduced the genetic variability of the resulting varieties thus forcing the breeders to seek that diversity in their wild relatives and introduce the desired traits into the elite germplasm by cross-breeding. Nowadays, most of the genes involved in the domestication process are well-known. Therefore, this introgression process can be skipped out by targeting these domestication genes directly in semi-domesticated or wild plants turning them rapidly into commercial crops. Hence, this strategy enables the use of the vast genetic diversity present in wild species as a source of allele-mining, widely expanding the crop germplasm pool [14]. Very recently this concept was shown by two studies featured in the same issue of Nature Biotechnology. Zsögön et al., [15] exploited the multiplexing ability of CRISPR/Cas9 by simultaneously disrupting six domestication-related genes important for yield and productivity in *Solanum pimpinellifolium*, an ancestor of the current tomato cultivars. Notably, the engineered *S. pimpinellifolium* displayed intermediate features between the ancestor and modern

tomatoes with altered morphology, size, number and nutritional value of the fruits. On the other side, Li et al., [16] extended this approximation and introduced desirable traits into four stress-tolerant wild-tomato accessions not only by editing of coding sequences, but also of *cis*-regulatory regions or uORFs of genes associated with morphology, flower and fruit production, and ascorbic acid synthesis.

Many of the multiplexing examples discussed above were achieved by constructing tandem assembled single-gRNA expression cassettes, which result in large constructs that may affect the stoichiometry and stability of the different gRNAs due to repetitive use of RNApolIII promoters. One of the best approaches to avoid these drawbacks is employing a unique promoter that drives a polycistronic gene construct containing several gRNAs. In this type of set-ups, the gRNA is interspersed with either (i) self-processing ribozyme sequences [8], (ii) Csy4 endoribonuclease recognition sites [9], or (iii) transfer RNA (tRNA) sequences recognized by endogenous plant ribonucleases [10]. All these strategies yield a polycistronic transcript that is latterly processed, releasing individual gRNAs. However, these approaches require accessory repetitive sequences or proteins which might complicate the cloning process. In this regard, the ability of CRISPR/Cas12a to process its own gRNA provides a simplified strategy for multiplex genome editing in plants, as demonstrated in this thesis and also recently by Wang et al., [11], who targeted four genes simultaneously in rice. Further work will be required to understand the limits of multiplexing with Cas12a, and to what extent the introduction of a large number of targets will influence their individual efficiency. In any case, the availability of Cas12a enzymes with additional PAM sites, easy multiplexing setups, and eventually with enhanced HR activities will be a valuable tool for fast-breeding, including the domestication of new crops.

1.3. Expanding the toolbox of transcriptional regulators

The genome engineering abilities of the CRISPR/Cas technology are beyond the introduction of DSB. Cas9 and Cas12a effector proteins can be deactivated by point mutations on their catalytic domains conserving the programmable targeting capacity [17, 18]. Deactivated versions of Cas proteins (dCas9 and dCas12a) can be harnessed as a protein scaffolds for the fusion of effector domains for biotechnological applications such chromatin imaging, transcriptional regulation, epigenetic regulation and base editing (Fig. 2). In our laboratory, we showed that dCas9 can be used as a programmable transcriptional factor (PTF) and provide strong and robust gene activation with genome-wide specificity in *N. benthamiana* [19], in agreement with other reports in several plant species [20-22]. Modest repression has been also achieved using Cas9 [22-24]. Currently, our group is exploring the use of dCas12a for transcriptional regulation (data not published) which has been confirmed in *A. thaliana* by Tang et al., [25].

Cas-based PTFs are valuable tools to build synthetic regulatory gene circuits, which often require the combination of activation and repressor activities targeting a different set of genes to achieve robust regulation. However, it is not possible to activate and repress simultaneously different targets in the same cell when utilizing a single Cas effector protein. The reason is that PTFs sharing the same PAM cannot discriminate between targets that need to be up or downregulated. A possible solution for this could be the use of two different PTFs with distinct PAM specificities, such as dCas9 and dCas12a. The simultaneous deployment of Cas9 and Cas12a has been successfully applied in mammalian cells using a fusion guide RNAs, which combine the scaffold of both Cas effector proteins and a shared protospacer sequence [27]. However, to this date no regulatory circuits using

combinations of dCas9 and dCas12a have been reported in plants.

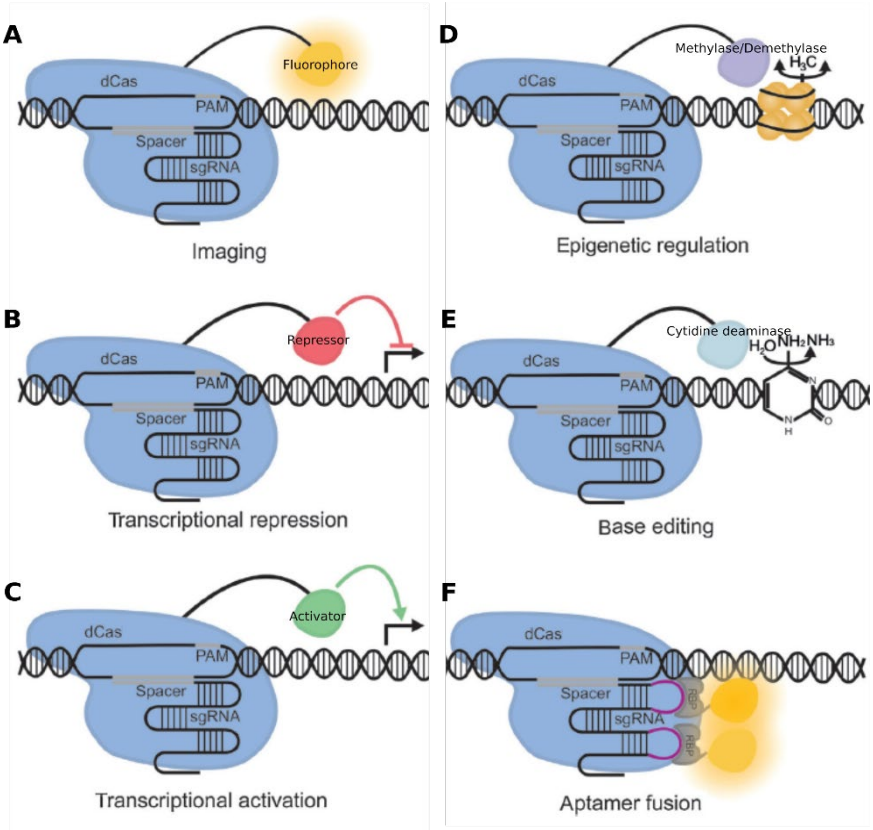


Figure 2: biotechnological applications of the deactivated Cas (dCas) effector proteins. The dCas proteins conserve its programmable targeting capabilities, which turns them in a perfect scaffold for fusing different protein domains. For example, dCas can be fused to a fluorophore for chromatin imaging or labeling determined sequences **(A)**; the direct fusion to an activator **(B)** or a repressor domains **(C)** can be used to modulate the transcriptional levels of a certain gene when the chimeric dCas protein is directed to a promoter region; the epigenetic state can be also changed when a methylase or demethylase is fused **(D)**; punctual base conversions can be performed when a cytidine deaminase domain is fused **(E)**; the sgRNA scaffold can be also engineered with RNA aptamers to recruit multiple copies of an RNA-binding proteins **(F)** thus enabling a magnification of the activation or the fluorescence intensity. Adapted from [26].

In conclusion, CRISPR/Cas systems offer a myriad of possibilities to accelerate the plant breeding process and enrich its genetic diversity. In Chapter II of this thesis, we have described additional tools for precision plant breeding based on CRISPR/Cas12a, a new CRISPR/Cas variant with different PAM specificity and attractive multiplexing features that could be an interesting complement to Cas9. We also provide a toolbox of Cas12a-related standardized phytobricks that have been functionally characterized in transient and stable experiments in various plant species.

2. Applications of the phage PhiC31-based toggle switch

The second main input of this thesis, the PhiC31 toggle switch, is by itself a basic circuit component that in combination with other gene elements, can be used to provide plants with innovative traits. Next, we will develop some examples of possible new traits enabled by PhiC31 toggle switch, and their potential applications.

2.1. Control of the production of toxic or detrimental compounds.

Plants are ideal biofactories to produce valuable compounds. These compounds can be originated by plant metabolism or incorporated by transgenic breeding techniques. Often the production of recombinant compounds in plants has detrimental effects on its development. For example, in our experience, the transient expression of human tumoral necrosis factor produces severe necrosis of leaves, which hinders the final yields (data not published). In cases like this, generating transgenic plants constitutively expressing the recombinant protein would not be possible, forcing to use *Agrobacterium*-mediated transient expression, which rise the price of the final product. A possible solution for this situation would be to temporally regulate the expression of the toxic transgene with the phage PhiC31-based

toggle switch. Once the plant would be fully developed the expression of the deleterious transgene would be switched on by the inducible PhiC31 recombinase, avoiding thus harmful developmentally associated effects, maximizing yield and reducing costs. The induction of the integrase could be achieved by a chemical agent as has been showed in subsection 2.4 of Chapter III or coupled to a developmental signal that is triggered in the final stages of plant growth such as flowering. Alternatively, when using perennial plant biofactories, a reversible toggle switch would be highly convenient as it would allow restricting transgene expression to short harvesting periods, freeing the plant from deleterious effects of recombinant gene expression during the rest of the season.

2.2. Regulation of transgene expression in field crops

One of the key features that make plants good biofactories for molecular pharming is its cheap and easy scalability process. In contrast to bacteria or mammalian cells, which needs of expensive culture media and sterile reactors to grow properly, plants only require a greenhouse, or a field supplemented with light, water and fertilizers. However, as we discussed in the previous section, when the transgene product is toxic or detrimental for plant growth it needs to be induced to avoid harmful effects. This is easier in the confined environment of a culture vessel, but in large settings such as a crop field it could rise some cost and safety issues, especially if the inducer is a toxic molecule and must be applied continuously, such in the conventional induction systems reviewed in Chapter I (section 3.4 Gene switches). However, these problems could be solved by exploiting the memory properties of a toggle switch. As mentioned in the previous section, just a brief pulse of inducer would be enough to promote the permanent induction of the transgene expression avoiding long and costly exposures that could raise safety issues.

2.3.A DNA-based memory device for recording cellular events

Data storage is becoming a pressing challenge to our modern society. Storing all the information generated to date and the one we create in our daily bases require a huge amount of space and resources. Because DNA is a high-density storage medium, can be exponentially amplified and preserved for long periods of time, synthetic biology companies such as Catalog are envisioning new DNA-based storage systems to, for example, compile all the Wikipedia's information in a bacterial pellet just the size of a sugar cube. This is only to exemplify the power of synthetic biology and DNA-based methods to solve practical problems such as data storage. Recombinases can be used for writing onto DNA. For example, 11 pairs of orthogonal recombinase systems have been used to store 1.375 bytes of information in the genome of *E. coli* [28]. Although more modestly, our PhiC31-based toggle switch could be repurposed as a DNA-based memory device for recording cellular events in two different states (PB or RL). DNA-modifying enzymes have been used in bacteria and mammalian cells for many purposes such as the detection of metabolites, expression of a gene, pH or ion concentration among others [29]. Those utilities could be expanded to plant systems for basic or applied perspectives, by coupling the expression of the integrase with any stimulus wanted to be registered.

In basic science, the advent of single-cell sequencing technologies is revolutionizing research studies with an unprecedented resolution of the biological processes. In *A. thaliana* this technique has been used to distinguish more than 3,000 root cells by capturing the gene expression dynamics [30]. This system could be used in conjunction with our toggle switch for the elucidation at unicellular resolution of, for example, the infection pathway that a virus follows across the plant tissues. The viral infection would trigger the expression of the integrase which

would turn on the expression of a tracking gene that, associated with a cell-type specific marker would allow following the viral course. The same principle would apply for any biological process to be studied, increasing the spatiotemporal resolution: a stimulus triggers the recombinase expression which registers it at the DNA level and then this is read.

Due to the high stability of the DNA, in applied science recombinase-mediated DNA storage could be used in de development of sentinel plants capable to detect chemicals, contaminants or pathogens. Antunes et al., [31] developed a visual system for the detection of trinitrotoluene in soil based on chlorophyll degradation. However, this type of visually monitored approximations require of constant surveillance of the plant condition which, on the other side, can be also affected by field conditions (e.g. temperature, wind, light) that largely differs from the lab-controlled environment and can lead to wrong interpretations. In contrast, if the detection of the compound creates a change in the DNA this could be easily tracked by exponential PCR amplification avoiding biases in user interpretation. Same approximation could be employed to diagnose a variety of viral infection in the crop field using PCR instead of other expensive procedures such as RT-PCR or ELISA [32]. Finally, this surveillance ability could be extended to prevent losses in the crops originated by unpredictable weather changes. For example, cold temperatures could trigger the expression of antifreeze proteins of Antarctic bacteria [33] through the recombinase action.

Summarizing, our recombination-based toggle switch offers a myriad of possibilities to detect, store and trigger cellular events by changing the structure of the DNA. The challenge here is coupling recombination activity with the external cues through diverse signaling processes. Fortunately, there is a large collection of genetic devices developed by synthetic biologists to detect a vast array of compounds and this toolbox is expected to

keep growing in the future, making easier to connect cellular signals with recombination.

3. Combined uses of toggle switches and Cas12/Cas9-based transcriptional regulators

We mentioned earlier that dCas proteins can be used as transcriptional regulators and that the incorporation of Cas12a to the toolbox would enable the combination of activation /repression elements in the same cell.

In this section we will discuss how the combination of a PhiC31-based toggle switch with CRISPR-Cas9/12a PTFs would configure powerful regulatory circuits, which could be used, among other purposes, to control and readjust metabolic fluxes. Metabolic pathways are articulated by enzymes that catalyze the conversion of initial compounds in final products through a collaborative effort. Usually, each enzyme is responsible for a single step. Therefore, metabolic pathways can be adjusted by influencing their enzymatic actuators. This can be achieved by increasing the amount of an enzyme (activation) which is acting as a bottleneck in the route. Alternatively, the metabolic flux to undesirable compounds can be avoided by decreasing the content of a branching enzyme (repression). This latter case can be accomplished by knocking out the coding sequence of the enzyme. However, in some cases this solution is not desirable since the irreversible disruption of certain genes could cause harmful effects on the cellular function or lead to unexpected changes in the metabolic pathways.

Hence, an easy way to simultaneously activate and repress various enzymes in the same cell would be to attach dCas9 and dCas12a to different transcriptional effector domains, and constitutively express them together with their gRNAs. Despite this simple design would enrich the current transcriptional regulation approaches, based on using a single deactivated

nuclease, it still lacks the necessary flexibility to readjust the metabolic fluxes. That is why regulating either, the expression of dCas proteins or the gRNAs with the PhiC31-toggle switch would integrate both PTFs in the same regulatory circuit allowing fine and adjustable control over the plant metabolism.

To illustrate the different possibilities that this design could offer, an example of a metabolic pathway involving four enzymes (E1-E4) will be used (Fig. 3). This pathway is divided in two branches, which lead to product A (mediated by E1 and E2) and product B (mediated by E3 and E4) respectively. Regulating these enzymes through the PTFs action would redirect the metabolism towards one product or the other, depending on the design of the genetic circuit. Thus, the regulation of one element of the PTF (either dCas or gRNAs) by the toggle switch and the constitutive expression of the remaining element, would direct the metabolic flux to the increased synthesis of product A or B depending on the configuration (SET or RESET) of the central regulatory element, which could be effectively operated by the recombination actuators. Compared to the design envisioned in the previous paragraph, in which both elements of the PTFs are constitutively expressed leading to a constant output, this design integrating both PTFs and PhiC31 switch allow the user to select and change between product A or B. However, these configurations (Fig. 3A-B) only permit regulation of one metabolic branch at a time. The complexity of this binary output could be scaled-up by regulating both branches at the same time, maximizing the production of one metabolite while the other it is being suppressed (Fig. 3C). To this end, would be necessary to place the gRNAs under the control of the switch and create a new construct including in the same array gRNAs from both deactivated endonucleases. Although this type of heterogeneous arrays mixing Cas9 and Cas12a gRNAs has not been reported yet, could be feasible to implement it adapting the existing strategies such as the tRNA processing system [10] or even exploiting the intrinsic RNase activity of Cas12a [34].

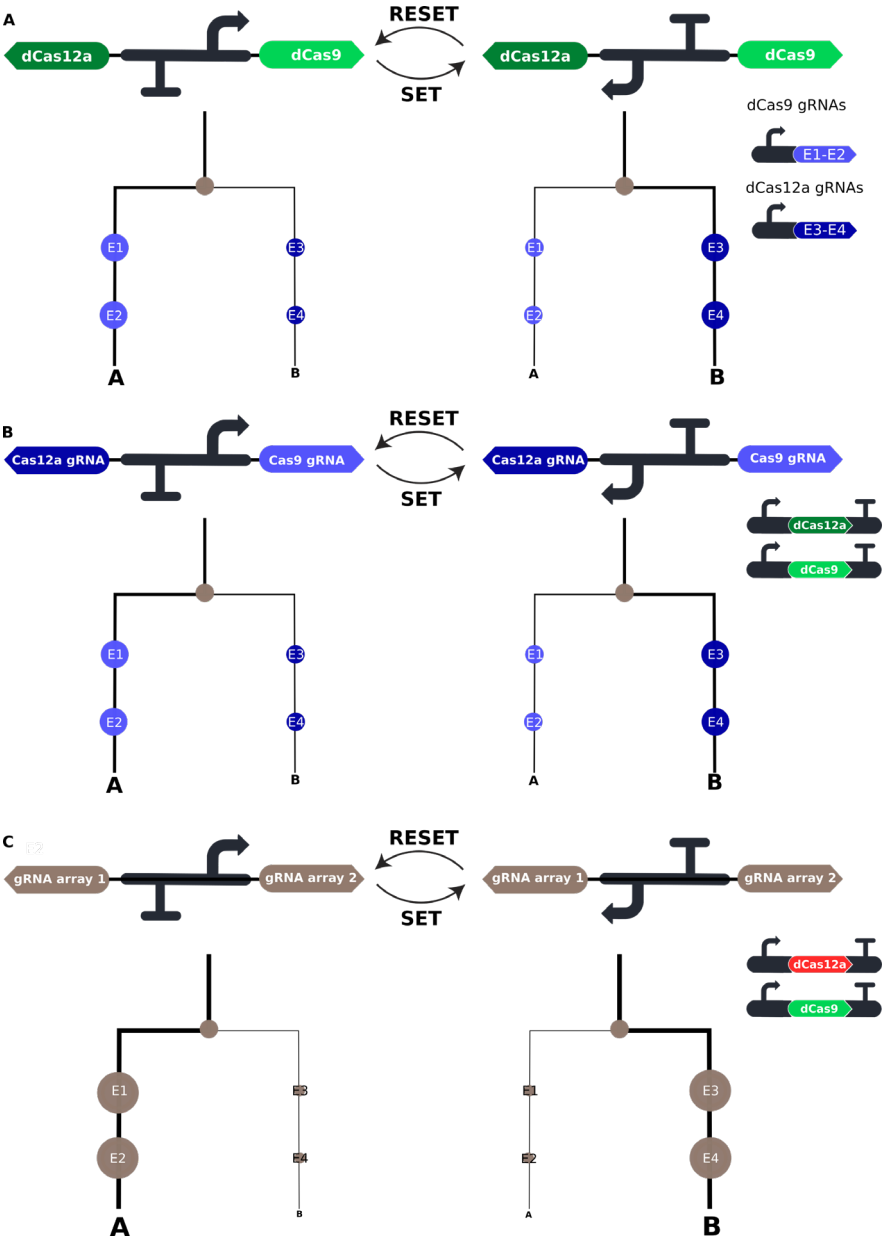


Figure 3: integrating phage PhiC31-based toggle switch and PTFs for modulating metabolic fluxes. A) Deactivated endonucleases (dCas9 and dCas12a) fused to activator domains (green color) could be subjected to bistable regulation while its gRNAs are constitutively expressed. **(Continued).**

(Figure 3 continued). The dCas9 gRNAs are directed towards the enzymes E1 and E2 (light blue) and dCas12 gRNAs to E3 and E4 (deep blue). Switching between both dCas activators would allow deviating the metabolic flux towards the production of A or B (bold line) by increasing the expression of the targeted enzymes (enlarged circles). **B)** same output could be achieved by regulating the expression of the gRNAs while constitutively expressing the PTFs. **C)** Regulation of mixed arrays containing gRNAs of both dCas12a and dCas9 (array 1 and array 2) would permit simultaneous regulation of both metabolic branches. Note that dCas9 is fused to an activator domain (green) while dCas12a is attached to a repressor one (red). As a result, the enzymes of one branch would be activated and the others repressed at the same time, maximizing thus the channeling of metabolites to the desired final product.

4. References

1. Shen L, Hua Y, Fu Y, Li J, Liu Q, Jiao X, Xin G, Wang J, Wang X, Yan C *et al*: **Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice**. *Science China Life Sciences* 2017, **60**(5):506-515.
2. Xu R, Yang Y, Qin R, Li H, Qiu C, Li L, Wei P, Yang J: **Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice**. *Journal of Genetics and Genomics* 2016, **43**(8):529-532.
3. Zhou J, Xin X, He Y, Chen H, Li Q, Tang X, Zhong Z, Deng K, Zheng X, Akher SA *et al*: **Multiplex QTL editing of grain-related genes improves yield in elite rice varieties**. *Plant Cell Rep* 2019, **38**(4):475-485.
4. Sánchez-León S, Gil-Humanes J, Ozuna CV, Giménez MJ, Sousa C, Voytas DF, Barro F: **Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9**. *Plant Biotechnology Journal* 2018, **16**(4):902-910.
5. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L: **Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew**. *Nature Biotechnology* 2014, **32**(9):947-951.
6. Miao C, Xiao L, Hua K, Zou C, Zhao Y, Bressan RA, Zhu J-K: **Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity**. *Proceedings of the National Academy of Sciences* 2018, **115**(23):6058-6063.
7. Wang C, Liu Q, Shen Y, Hua Y, Wang J, Lin J, Wu M, Sun T, Cheng Z, Mercier R *et al*: **Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes**. *Nat Biotechnol* 2019, **37**(3):283-286.
8. Gao Y, Zhao Y: **Self-processing of ribozyme-flanked RNAs into guide RNAsin vitroandin vivofor CRISPR-mediated genome editing**. *Journal of Integrative Plant Biology* 2014, **56**(4):343-349.
9. Cermak T, Curtin SJ, Gil-Humanes J, Čegan R, Kono TJY, Konečná E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL *et al*: **A multi-purpose toolkit to enable advanced genome engineering in plants**. *The Plant Cell* 2017:tpc.00922.02016.
10. Xie K, Minkenberg B, Yang Y: **Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing**

system. *Proceedings of the National Academy of Sciences* 2015, **112**(11):3570-3575.

11. Wang M, Mao Y, Lu Y, Tao X, Zhu J-K: **Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System**. *Mol Plant* 2017, **10**(7):1011-1013.
12. Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB: **Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing**. *Cell* 2017, **171**(2):470-480.e478.
13. Zhang H, Si X, Ji X, Fan R, Liu J, Chen K, Wang D, Gao C: **Genome editing of upstream open reading frames enables translational control in plants**. *Nat Biotechnol* 2018, **36**(9):894-898.
14. Massawe F, Mayes S, Cheng A: **Crop Diversity: An Unexploited Treasure Trove for Food Security**. *Trends Plant Sci* 2016, **21**(5):365-368.
15. Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, Freschi L, Voytas DF, Kudla J, Peres LEP: **De novo domestication of wild tomato using genome editing**. *Nat Biotechnol* 2018.
16. Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C: **Domestication of wild tomato is accelerated by genome editing**. *Nat Biotechnol* 2018.
17. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**(6096):816-821.
18. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A *et al*: **Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system**. *Cell* 2015, **163**(3):759-771.
19. Selma S, Bernabé-Orts JM, Vazquez-Vilar M, Diego-Martin B, Ajenjo M, Garcia-Carpintero V, Granell A, Orzaez D: **Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator**. *Plant Biotechnol J* 2019.
20. Li Z, Zhang D, Xiong X, Yan B, Xie W, Sheen J, Li J-F: **A potent Cas9-derived gene activator for plant and mammalian cells**. *Nature Plants* 2017, **3**(12):930-936.
21. Lowder LG, Zhou J, Zhang Y, Malzahn A, Zhong Z, Hsieh T-F, Voytas DF, Zhang Y, Qi Y: **Robust Transcriptional Activation in Plants Using Multiplexed CRISPR-Act2.0 and mTALE-Act Systems**. *Mol Plant* 2018, **11**(2):245-256.

22. Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, Aouida M, Mahfouz MM: **RNA-guided transcriptional regulation in planta via synthetic dCas9-based transcription factors**. *Plant Biotechnol J* 2015, **13**(4):578-589.
23. Lowder LG, Zhang D, Baltes NJ, Paul JW, 3rd, Tang X, Zheng X, Voytas DF, Hsieh T-F, Zhang Y, Qi Y: **A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation**. *Plant Physiol* 2015, **169**(2):971-985.
24. Khakhar A, Leydon AR, Lemmex AC, Klavins E, Nemhauser JL: **Synthetic hormone-responsive transcription factors can monitor and re-program plant development**. *Elife* 2018, **7**.
25. Tang X, Lowder LG, Zhang T, Malzahn AA, Zheng X, Voytas DF, Zhong Z, Chen Y, Ren Q, Li Q *et al*: **A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants**. *Nature Plants* 2017, **3**(3).
26. Kumlehn J, Pietralla J, Hensel G, Pacher M, Puchta H: **The CRISPR/Cas revolution continues: From efficient gene editing for crop breeding to plant synthetic biology**. *J Integr Plant Biol* 2018, **60**(12):1127-1153.
27. Kweon J, Jang A-H, Kim D-E, Yang JW, Yoon M, Shin HR, Kim J-S, Kim Y: **Fusion guide RNAs for orthogonal gene manipulation with Cas9 and Cpf1**. *Nature Communications* 2017, **8**(1).
28. Yang L, Nielsen AA, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA: **Permanent genetic memory with> 1-byte capacity**. *Nature methods* 2014, **11**(12):1261.
29. Sheth RU, Wang HH: **DNA-based memory devices for recording cellular events**. *Nat Rev Genet* 2018, **19**(11):718-732.
30. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT: **Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana**. *The Plant Cell* 2019, **31**(5):993-1011.
31. Antunes MS, Ha S-B, Tewari-Singh N, Morey KJ, Trofka AM, Kugrens P, Deyholos M, Medford JI: **A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity**. *Plant Biotechnol J* 2006, **4**(6):605-622.
32. Jeong J-J, Ju H-J, Noh J: **A review of detection methods for the plant viruses**. *Research in Plant Disease* 2014, **20**(3):173-181.
33. Muñoz PA, Márquez SL, González-Nilo FD, Márquez-Miranda V, Blamey JM: **Structure and application of antifreeze proteins from Antarctic bacteria**. *Microbial cell factories* 2017, **16**(1):138.

34. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E: **The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA**. *Nature* 2016, **532**(7600):517-521.

CONCLUSIONS

C1. The GoldenBraid (GB) biorepository, which offers a collection of DNA parts and webtools for the assembly of constructs for plant engineering, was expanded with new molecular tools for genome engineering and plant synthetic biology. More precisely, the CRISPR/Cas12a system was included to perform site-specific mutagenesis by introducing double-strand breaks at precise genome locations. This CRISPR/Cas12a toolset included two orthologs of the Cas12a endonuclease (AsCas12a and LbCas12a) and two guide RNA (gRNA) expression cassettes for targeting a single gene or multiple loci in the plant genome. In addition, a new reversible and bistable memory switch inspired in the bacteriophage PhiC31 recombination system was included in GB platform, to regulate the expression status (ON or OFF) of two genes of interest in reverse and forward orientations by inversion of a central DNA regulatory element through site-specific recombination.

C2. All the required elements to perform site-specific mutagenesis with the CRISPR/Cas12a system were extensively validated through transient and stable transformation experiments in *N. benthamiana*, *S. lycopersicum* and *A. thaliana* and comprehensively compared with the alternative CRISPR/Cas9 gene editing system. Concerning CRISPR/Cas12a system, the LbCas12a endonuclease was significantly more efficient than AsCas12a, and provided higher average mutagenesis rates when compared to SpCas9 in the analyzed loci. Mutagenesis rates of Cas12a were greatly influenced by the sequence of the target locus, probably due to the high variability in the secondary structure of the gRNA as the assays with shortened synthetic gRNAs revealed.

C3. The CRISPR/Cas12a system was found to have a deletion-prone mutagenesis profile, which was harnessed to produce germline-associated deletions in *A. thaliana* by targeting 5′ and 3′ untranslated regions of the *PDS3* gene. Genome-wide analysis of Cas12a-free mutants of the *PDS3* gene, discarded the presence

of off-targets mutations in the *A. thaliana* genome. One line though showed large chromosome rearrangements surrounding the targeted site.

C4. A phage PhiC31-based toggle switch was designed to be operated by SET and RESET recombination reactions, which inverted the DNA central regulatory element modifying the expression status of the flanking transgenes. Individual SET and RESET recombination reactions proved to be fully functional in *N. benthamiana* transient and stable experiments, in which the switch was delivered in the agroinfiltration mixture or stably integrated in the plant genome. SET and RESET transitions were evaluated using four different configurations of the toggle switch, with firefly luciferase and YFP as reporter genes. SET operation showed to be more efficient and stable than RESET as expected for a one-component reaction.

C5. The reversibility of the toggle switch was demonstrated by successfully operating a full SET-RESET recombination cycle through an infiltration and *in vitro* regeneration experiment.

C6. The inducibility and memory of the switch was demonstrated by operating a SET reaction in transgenic *N. benthamiana* hairy roots carrying a switch with an inactive luciferase reporter gene and transformed with an estradiol-inducible integrase. The PhiC31 integrase expression was chemically induced as evidenced by the increment in the luciferase expression triggered by the SET recombination. Luciferase signal increased in the course of the time indicating the system remained in the switched state maintaining a memory of the recombination.