THE UNIVERSITY
of EDINBURGH

Development of synthetic biology tools and methods for plant systems

Andreas Andreou

THE UNIVERSITY *of* EDINBURGH

Thesis submitted for the degree of

Doctor of Philosophy

University of Edinburgh

Biological Sciences

Year of Submission 2019

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for any other degree. Except where otherwise stated by reference, acknowledgement or contribution, the work presented is entirely my own.

The work presented in Chapter 1 was previously published in PLoS One. 2018 Jan 2;13(1):e0189892. doi: 10.1371/journal.pone.0189892. eCollection 2018 as “Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly” by Andreas Andreou and Naomi Nakayama.

The protocol presented in Chapter 5 was accepted for publication in “Methods in Molecular Biology - Springer as “Mobius Assembly” by Andreas Andreou and Naomi Nakayama.

Andreas Andreou,

August 2019

Acknowledgements

I want to express my thanks to my supervisors. Firstly, to Dr Naomi Nakayama for helping me to secure the PhD scholarship, for her continuous support during my PhD research and for her motivation. She has always given me the encouragement and freedom to pursue my own ideas, and her guidance has helped my research and writing of this thesis. My sincere gratitude also goes to Prof. Chris French, who provided me with the opportunity to work on my side project in his lab and his invaluable advice on all my biology work. Without their precious support, it would not have been possible to be here writing this thesis.

I thank my fellow labmates for the stimulating discussions, for their constant support and encouragement and for all the fun we have had over the previous four years. A special thanks to Samuel, Efrain, Chris, Marcos, Felippe, Dariusz, Pedro, Stuart, Nahuel, Andri, Beatrice, Rist, Savvas, Laura, Louis, Tara, Cathal, Maddy, Niki, Sophie, Paulina, Panupon, Alejandra, Stamatis, Anton and Uriel, who helped me in different stages of my PhD studies in so many different ways.

Special thanks go two people who actively involved in my project. To Marisol who was the first to test my cloning system and who was really supportive in many ways for the completion of my thesis and Jess for her collaboration and contribution in my project.

Eric, a good friend and business partner, you always got my back! I am incredibly grateful for your multifold support in my PhD and the collaboration in the enamel project.

Many thanks to my friends in Edinburgh Chris P., Chris M, Chris V., Irene, Ioradnis, Ioannis S., Ioannis P., Niki, Lito, Yiota and Argyris, for the making my life more colourful. Argyri, thanks for the incredible time we had in the flat, you were more like a brother to me than a flatmate!

A massive thanks also go to Eric, Sally and Jess for proofreading my thesis.

Most importantly, I would like to thank my family: My mother Skevi and father Ioannis for the sacrifices that they have made for me to be here and their constant support. My brothers Antonis and Iakovos and my sister Irene for being amazing to me! My cousin and four-month flatmate Andreas for the fantastic time we had in Edinburgh. My aunt and teacher Xenia, a real fighter, who is no longer with us; I will always follow your advice.

Finally, to Pigi. Thank you for making my life beautiful, for listening to all my problems, for your company, for the emotional support, for tolerating my stubbornness and for your unconditional love. Because of you, I can finish this journey. Sorry for not dedicating more time to you because of my PhD. Σε αγαπώ!

Abstract

Plant synthetic biology is a fast-evolving field in which standardised tools and methods are developed to empower research and commercial bioproduction in plant systems. Nevertheless, in the whole synthetic biology landscape, plant systems lag compared to microbial and mammalian systems. Plant cell cultures are becoming a popular chassis for bioproduction, which can combine the metabolic diversity and capacity of the plants with the benefits of cell cultures as in microbial systems, such as fast growth, resource- and energy-efficient cultivation, and secure containment. My PhD was focused on the establishment of multiple enabling tools and methods to confer complex yet predictive genetic engineering in plant systems. These plant synthetic biology resources are being implemented to development cell-type-specific plant cell biofactories particularly competent in the biosynthesis of specific compounds.

Firstly, I have developed Mobius Assembly, a user-friendly Golden Gate Assembly system for fast and easy generation of complex DNA constructs. Mobius Assembly toolkit was adapted for Plant Systems (MAPS). I devised a new category of compact binary vectors, which can be employed for the whole plant, plant cell culture, and plant protoplast transformations. I found that the combination of binary vectors and Agrobacterial strains determine the efficiency of plant cell culture transformation.

Secondly, I developed a high-throughput, protoplast expression protocol. Using Mobius Assembly and the new protoplast expression analysis platform, I characterised a library of short promoters and terminators for their regulatory effect on reporter gene expression. The results highlighted the strong influence of terminators in gene expression depending on different promoters; this observation implies a synergistic interaction between terminators and promoters. I further made and optimised standardised parts of three commonly used chemically

inducible gene expression systems in plant science and tested them for possible crosstalks when used in conjunction.

Lastly, I employed MAPS to create a sequential plant expression system that can be used to activate multiple gene expression at differential levels and timings. As a pilot study, this sequential system was used to express master transcription factors for the differentiation of plant cell cultures into xylem vessels, a cell type that produces a high level of phenolic compounds and thus may serve as biofactories for production of anthocyanin and other phenolics.

Lay Summary

Plants provide us with oxygen, food, medicines and materials. As humans have started to switch to a more natural and sustainable lifestyle in recent years, there is an even higher need for plant products. Biological research can further improve plants and offer alternative ways to produce different valuable compounds, such as food and medicines. The way to do that is by changing the DNA, the molecule which carries information about the plant's characteristics, and therefore changing what and how they can produce these compounds. However, there are not many tools available for simple DNA changes, and in this work, I have therefore developed such tools. First, I created a method to glue together different DNA molecules and put them back in the plant. I identified various problems that can occur during this prosses and how to successfully fix them. I also showed that some start and end parts that are used to read the DNA in plant cells are more significant than others in controlling some characteristics of the plants. I also used plant cell cultures, which are single cells derived from plants, that grow and divide in flasks and can be one solution to the overuse of different plants. I could show that they have potential in replacing whole plants in the production of compounds and that we can further improve them using the methods I have developed.

Table of Contents

General Introduction

The emergence of plant cell cultures

In recent years, there has been an emerging trend among people towards a more natural and sustainable lifestyle. The global population has been increasing, and consequently, the human demand for resources and products are ever-growing. Therefore, the demands for plant-derived products such as plant-based food, natural ingredients for cosmetics, herbal medicines, personal care products, food supplements, eco-friendly dyes for clothes and hair, and active ingredients for drugs will be under increasing demand. Natural plant harvest cannot account for these market needs sustainably. Overexploitation of plants has already endangered more than 15,000 medicinal species out of 28,187 (State of the World's Plants 2017 - Royal Botanic Garden Kew - (stateoftheworldsplants.org).

An alternative solution to natural harvest is the use of plant cell cultures, which embody the versatile secondary metabolism of plants in a fast-growing chassis. Plant cell cultures are a vital tool in both basic and applied research in plant biology, and they are useful for commercial applications (1). They are the source of a plethora of natural products for the pharmaceutical, agriculture, cosmetic and food industries (2–4). They constitute powerful biofactories as they are independent of seasonal parameters. They are well defined, thus providing standardised batches to comply with current good manufacturing practice (2, 4–6). They have lower water demands, smaller carbon footprint and no requirements for pesticides, and can also grow faster, produce higher biomass and more metabolites compared to whole plants. These factors, together with less complex media compositions, make plant cultures more cost-effective and suitable for large scale production. The acceptance of the general public is further improved by the fact that they do not risk the transmission of any human or animal pathogens.

History of plant cell cultures

Plant cell cultures derive from dedifferentiated plant cells which grow on artificial media under sterile conditions (1). Kotte was the first to successfully culture plant tissue in 1922, from root tips (7). Using a similar technique in 1934, White managed to create an immortal line from tomato root tips in liquid medium (8). The same year Gautheret created the first “true” plant tissue culture from cambial tissue of *Acer pseudoplatanus* (9). The transition from tissue to single-cell culture was achieved in 1958 by Muir et al. He was shaking callus cultures to isolate cells and place them into a liquid medium (10). Tulecke and Nickell demonstrated the first scale-up plant cell culture in 20 litter carboys in 1959, for different plant species such as Lolium, holly, ginkgo and rose (11). Thirty years later, in 1990, Sijmons achieved the first recombinant protein production (human serum albumin) in plant cell suspension cultures (12).

Several available cell suspension cultures are used as bioproduction platforms; however, the three most popular are those from tobacco (Nicotiana tabacum), rice (Oryza sativa), and carrot (Daucus carota) (13). Bright Yellow 2 (BY-2) is the most commonly used tobacco cell line. It was developed by Kato et al. in 1968, and it has a doubling time of 16–24 h, meaning that it can multiply up to 100 times over a week (14). BY-2 can be engineered by Agrobacterium-mediated transformation, and many biopharmaceuticals have been produced using this system such as the Hepatitis B Surface Antigen and interleukins 4, 10 and 12 (13).

Applications and limitations of plant cell cultures

Plant cell cultures have mainly three applications in plant biology. They are exploited as experimental tools to study plant biological processes. For example, they have been used to investigate the early transcriptional defence responses in Arabidopsis (15) and to monitor changes in free nuclear calcium in response to elicitors in Tobacco (16). They are also crucial for the preservation of endangered species and their reintroduction back to nature if they are extinct (17). Lastly and

probably most importantly plant cell cultures are used for metabolic engineering of fine chemicals (18).

Although plant cells have great potential as biofactories, they lag behind compared to the microbial and mammalian systems, mainly because they are less than ten years commercially available (13). First of all, there are limited tools available for plant cell cultures, as they are an emerging technology. Plant cell cultures also grow in clusters rather than single cells, which affects their oxygen and nutrient availability and makes tricky the generation of monoclonal transgenic lines. Polyclonal transgenic lines can also result from the transformation at different loci with Agrobacterium-mediated transformation (19). Furthermore, cells can spontaneously undergo somaclonal variations, and heterogenic cell populations can be formed, even with integration in specific sites using CRISPR technologies (20). For these reasons, there is a need for several cycles of screening and selection to identify hight yield clones. Growth medium optimisation is another challenge for increasing the productivity of cell suspension cultures (21). For example, the upsurge of nitrogen concentration in BY-2 culture increased their productivity 150 times (22). Lastly, proper cell banking is required to ensure production stability (23), considering that cryopreservation protocols in plant cell cultures should be optimised specifically for different species.

History of plant protoplasts

Protoplasts are spherical cells lacking their cell wall, which was removed by mechanical or/and enzymatic treatment. The name “protoplast” comes from the Greek word “πρωτόπλαστος” which means "first-formed, a term coined by Hanstein in 1880 (24). J. Klercker made the earliest recorded attempt for plant protoplast isolation in 1892 (25). He peeled off the epidermis of the leaf from *Stratiotes aloides* (water warrior) and plasmolysed the cells to detach them from the cell wall. He subsequently cut the edge of the cell wall with a microrazor and squeezed out the protoplasts. In 1927, E. Küster used ripe fruits of *Solatium*

nigram (Black nightshade) and *Lycopersicon esculenium* (tomato) for protoplast isolation (26). In these fruits, the native enzymes hydrolyse the cell walls, and free protoplasts can be released by slicing the fruit. In 1931, R. Chambers and K. Hofler immersed thin slices of onion (*Allium cepa*) epidermis in sucrose to plasmolyse the cells and released the protoplasts by cutting the epidermis with a razor (27).

E. C. Cocking developed the first reported enzymatic method for protoplast isolation in 1960 (28). Cocking used a concentrated solution of cellulase extracted from the fungus *Myrothecium verrucaria* to degrade the cell wall of *Lycopersicon esculentum* root tip cells, resulting in the releasing of many protoplasts. An improvement to the enzymatic isolation method was made by Takebe et al. in 1968 (29). A two-step technique involved treatment of tobacco leaves with pectinase (or macerozyme) first to separate the cells, followed by cellulase to degrade the cell walls. The procedure was further simplified by the simultaneous use (one-step method) of a mixture of cellulase and macerozyme (30). This method, with improvements and modifications, is being used nowadays.

Even though protoplasts could be easily isolated, several factors influence their isolation yield and viability. Among these factors are the type of the plant (which affects the cell wall composition), temperature, concentration and pH of the digestion enzymes and the nature of the osmoticum (31). For example, plasmolysis before isolation of apricot leaves using sorbitol decreased the cytoplasmic damage and fusion of the protoplasts (32). Furthermore, the use of glycine in the enzyme solution improved the protoplast release from melon (*Cucumis melo and C. metuliferus*) leaves in a species-dependent manner (33). Protoplast viability and yield can also be enhanced by mechanically treating the plant tissue (slicing or perforation) before the enzymatic lysis and by adjusting the osmolarity of the isolation medium (34).

The versatility of plant protoplasts

Plant protoplasts is a flexible experimental system, as they can be transformed with DNA, RNA and proteins with different methods such as PEG transformation, electroporation and microinjection (35). Protoplasts transient expression is faster and more convenient over stable gene expression (36). Especially in CRISPR technologies, protoplasts can be used to determine the gene-editing efficiency which can be subsequently regenerated into plants (37). Also, the direct delivery of the protein molecules of the genome-editing reagents into the protoplasts can lead to transgenes free from exogenous DNA. Nevertheless, plant regeneration from protoplast is difficult in most of the plant species.

Protoplasts have been used to study a plethora physiological processes in the plant cells (35, 38). More specifically, they have been used for the study of cell signaling pathways triggered by hormones, environmental signals, and pant pathogen elicitors (39). They have also been used to elucidate the plant circadian rhythms (40), for subcellular protein localisation and protein-protein interaction studies (41). They have been employed for high-throughput analysis of gene regulation (42) and to characterise genetic parts and circuits for plant expression (43). Protoplast transient expression systems are ideal for non-model plants whose stable transformation is not yet available or is tedious (41, 44). Several protoplast transformation protocols have been developed for different species, such as Arabidopsis, Tobacco, rice, maise, cucumber and orchid (35, 45–49).

Synthetic biology and standardisation in DNA assembly

Synthetic biology is a fast-expanding field at the interface between biology and engineering that facilitates predictive engineering of living organisms with novel functionalities (50). Engineering principles, such as standardisation, modularity, and simplicity, are implemented to reduce the unpredictability of complex and often non-linear living systems. Standardised DNA parts with consistent and well-characterised functionalities can be utilised just like human-made standardised

parts (e.g. metric or imperial screws and optic boards or circuitry components) in mechanical and electric engineering. This will enable the implementation of the engineering 'design-construct-test' cycle in biology, ultimately allowing expansion of potential designs and efficient construction and testing, while encouraging automation, community-wide part-exchange and resource establishment. In order to realise such universal standardisation in synthetic biology, there remain fundamental challenges to consolidate part compatibility and to simplify and streamline the construction process.

Rebatchouk et al. were the first to introduce the concept standards in plasmid assembly methods in 1996 with their assembly method NOMAD (51). NOMAD was using type IIS restriction endonucleases to assemble standardised "building blocks" and construct reusable modules in any desired order. However, this attempt was not recognised by the broad scientific community. The recognition of the important in standardisation came in 2003 with the introduction of the BioBrick Assembly, which is using four restriction enzymes to enable an idempotent assembly process (52). Several methods were further developed following the principle of BioBricks (53–55). Meanwhile, the Standard European Vector Architecture (SEVA) defined modular parts for plasmids so they can be applied in different bacterial species (56). SEVA was further updated to include more parts for construction and deployment of complex bacterial phenotypes and to adopt the SBOL (Synthetic Biology Open Language) format (57, 58).

Type IIS restriction endonucleases came to the spotlight with the development of Golden Gate Assembly (59, 60). In Golden Gate Assembly, the user sets the overhangs of the restriction sites while digestion and ligation co-occur in the same tube; however, the composite parts were not reusable. MoClo and Golden Braid methods solved this limitation using two Type IIS enzymes, enabling the hierarchical construction of multi transcriptional units. The popularity of the Golden Gate Assembly methods was increased in the last few years with the development of many variants (61–67). Nevertheless, the new assembly variants

are increasing the complexity of the assembly process. Mobius Assembly and Loop Assembly, on the contrary, focus on the simplicity and user-friendliness of the cloning (68, 69).

Part toolkits in plant synthetic biology

The establishment and adoption of a common syntax for the basic standard parts (Phytobrics) in Golden Gate methods was an advancement in plant synthetic biology (70). However, the available libraries of standard parts for different genetic functions are limited (71). More specifically, two complete collections are available on Addgene: the plant MoClo collection (72) and its expansion (73) and the Golden Braid 2.0 collection (74). There is also a developing collection of parts for Marchantia available from Haseloff Group (http://marpodb.io/query). Additionally, in Addgene, there are plant CRISPR resources and vectors kits available for cloning with limited or even without genetic parts (https://www.addgene.org/collections/plant/). In other cases, researches even though characterise parts they do not follow the adopted standards nor deposit their parts in plasmid repositories. Apart from the scarcity on part repositories, the integration of experimental specifications in the part databases is limited. Only the Golden Braid collection proceeded to the incorporation of functional descriptions of standard parts in their repository (71). The generation of more plant collections with a broad range of genetic elements accompanied by characterisation data is an immediate requirement to accelerate the research in plant community further.

Aim and scope of this thesis

Plant synthetic biology still lags in its development compared to microbial synthetic biology. Even though the potential of plant systems for the bioproduction has been long recognised (75), they are generally not chosen as synthetic biology chassis due to their slow growth, physiological and genomic complexity, as well as the lack of available tools and techniques for genetic manipulation (76). When it comes to multigene delivery to plants, the predictability of the outcome is highly

decreased as it depends on three different chassis: *E. coli*, Agrobacterium and plant species. The present work aimed to develop tools and methods that will accelerate plant synthetic biology, enable the use of plant cell cultures as the next generation synthetic biology chassis, and create a system to make plant cell cultures exceptionally competent in the biosynthesis of specific bio compounds.

Thesis Overview

The system, described in Chapter 4, is based on the unique property of plant secondary metabolism to be closely connected with the differentiated phase of the cells. Different cell types have potency in the synthesis of various metabolites. For example, petal cells produce anthocyanins, and xylem vessels and fibre cells produce flavonoids. On the contrary, the dedifferentiated state of cell cultures in the bioreactors likely suppresses the competency for producing valuable compounds. It was hypothesised that induced cell differentiation would be sufficient to activate genes involved in the biosynthesis of plant chemicals. To achieve that, a guided plant cell differentiation system was developed, which combines plant inducible systems and employs transcription factors to induce cell proliferation, activate the cell fate specification and stimulate the plant secondary metabolism.

During the design of the guided differentiation system, experimental challenges were encountered. The first one was the assembly of the system, which in its initial stage was comprised of 31 parts: Antibiotic selection and transformation verification cassettes with three parts each (promoter-CDS-terminator), three transcription activator units with three parts each (promoter-CDS-terminator) and three inducible promoter units with five parts each (inducible promoter-transcription factor-linker-fluorescent protein-terminator). It was attempted to build the system using MoClo system. Primers were designed to amplify the selected transcription factors for MoClo vectors. However, the restriction enzymes used by MoClo were abundant in the target amplification genes, and its cloning framework

was too complicated, with several vectors and assisting plasmids. To address these limitations, I developed a new DNA assembly method, called Mobius Assembly, which is described in Chapter 1.

The second problem was the unavailability of characterised parts to control the multigene expression in the system. Specifically, eight terminators and five constitutive promoters were needed. Searching through available part libraries, two issues were identified. The first one was the shortage of parts, only 4-5 promoters and terminators derived mainly from *Agrobacterium tumefaciens* have been used during the past 20 years. Secondly, the few characterised parts derived from Arabidopsis have a large size (1-4 kb). The need for new parts for the system led to the identification and characterisation of new elements for plant gene expression, which is presented in Chapter 3.

Additionally, a reliable and high-throughput system to characterise the genetic parts and test the constructs was required. Before the characterisation of the genetic elements, a high-throughput leaf protoplast assay was developed, along with a detailed protocol for the transformation process, which is discussed in Chapter 2. The part characterisation could also show the importance of the terminators in controlling gene expression (Chapter 3). Mobius Assembly was adjusted for plant transformation, and it was equipped with a new feature to ease the generation of combinatorial libraries, which are included in Chapter 3.

It was demonstrated that there is no crosstalk activity among the inducible systems, which were combined to create the guided differentiation system. Additionally, it was tested whether the expression of the inducible systems can be tuned by switching terminators. Those experiments are found in Chapter 4.

The next step was to build the guided differentiation system using the Mobius assembly. However, first, the transformation efficiency was optimised, since the TOP10 strain was not able to efficiently receive large constructs housed by the pGreen vectors. Switching to the JM109 strain could resolve the transformation

issues, but this led to plasmid instabilities, as the final constructs were large and had repetitive sequences. To tackle this problem, the stability of different backbones in different *E. coli and* Agrobacterium strains was tested, which is described in Chapter 3. After building the system, it was attempted to transform into the MM1 Arabidopsis cell line, but with no fruitful results. It was found that plant binary vectors have a significant impact on the plant cell transformations, and also the Agrobacteria strains affect this process. Therefore, the pGreen backbone was changed to pLX, and subsequently, the effects of the abovementioned factors for the plant cell culture transformation were studied (Chapter 3). Even though the pLX vectors were efficient for the cell culture transformations, they could not be used for applications with high demands for DNA. For this reason, a new set of binary vectors were devised, using a fusion origin of replication between pUC and WSK1 from *Paracoccus pantotrophus*. The specific vectors were employed to study how the functional dissection of two terminators influences the gene expression in both Arabidopsis protoplasts and MM1 cell lines (Chapter 3).

After the successful transformation of the guided differentiation system, as a proof concept, it was attempted to differentiate the MM1 cells to xylem vessels and induce the production of anthocyanins, which is described in Chapter 4.

The last chapter includes a detailed protocol for Mobius Assembly cloning. It incorporates all cloning problems encountered during this work, as well as the feedback received from groups on the broader plant community, who have already used the assembly method. It aims to provide step-by-step guidance for use and troubleshooting for the Mobius Assembly and will enable wide-reaching use by other plant scientists and the synthetic biology community.

References

1. Ogita,S. (2015) Plant Cell, Tissue and Organ Culture: The Most Flexible Foundations for Plant Metabolic Engineering Applications. *Nat. Prod.*

Commun., **10**, 1934578X1501000527.

2. Ochoa-Villarreal,M., Howat,S., Jang,M.O., Kim,I.S., Jin,Y.W., Lee,E.K. and Loake,G.J. (2015) Cambial meristematic cells: A platform for the production of plant natural products. *N. Biotechnol.*, **32**, 581–587.

3. Ochoa-Villarreal,M., Howat,S., Hong,S., Jang,M.O., Jin,Y.-W., Lee,E.-K. and Loake,G.J. (2016) Plant cell culture strategies for the production of natural products. *BMB Rep.*, **49**, 149–158.

4. Eibl,R., Meier,P., Stutz,I., Schildberger,D., Hühn,T. and Eibl,D. (2018) Plant cell culture technology in the cosmetics and food industries: current state and future trends. *Appl. Microbiol. Biotechnol.*, **102**, 8661–8675.

5. Tekoah,Y., Shulman,A., Kizhner,T., Ruderfer,I., Fux,L., Nataf,Y., Bartfeld,D., Ariel,T., Gingis–Velitski,S., Hanania,U., *et al.* (2015) Large-scale production of pharmaceutical proteins in plant cell culture—the protalix experience. *Plant Biotechnol. J.*, **13**, 1199–1208.

6. Georgiev,V., Slavov,A., Vasileva,I. and Pavlov,A. (2018) Plant cell culture as emerging technology for production of active cosmetic ingredients. *Eng. Life Sci.*, **18**, 779–798.

7. Kotte,W. (1922) Kulturversuche mit isolierten Wurzelspitzen Beitr. Allg. Bot.

8. White,P.R. (1934) Potentially Unlimited Growth of Excised Tomato Root Tips in A Liquid Medium. *Plant Physiol.*, **9**, 585–600.

9. Gautheret,R.-J. (1934) Culture du tissu cambial.

10. Muir,W.H., Hildebrandt,A.C. and Riker,A.J. (1958) The Preparation, Isolation, and Growth in Culture of Single Cells from Higher Plants. *Am. J. Bot.*, **45**, 589–597.

11. Tulecke,W. and Nickell,L.G. (1959) Production of Large Amounts of Plant Tissue by Submerged Culture. *Science (80-.).*, **130**, 863 LP – 864.

12. Sijmons,P.C., Dekker,B.M., Schrammeijer,B., Verwoerd,T.C., van den Elzen,P.J. and Hoekema,A. (1990) Production of correctly processed human serum albumin in transgenic plants. *Biotechnology. (N. Y).*, **8**, 217–221.

13. Santos,R.B., Abranches,R., Fischer,R., Sack,M. and Holland,T. (2016) Putting the Spotlight Back on Plant Suspension Cultures. *Front. Plant Sci.*, **7**, 297.

14. Kato,K., Matsumoto,T., Koiwai,S., Mizusaki,S., Nishida,K., Nogushi,M. and Tamaki,E. (1972) Ferment Technology Today.

15. González-Pérez,S., Gutiérrez,J., García-García,F., Osuna,D., Dopazo,J., Lorenzo,Ó., Revuelta,J.L. and Arellano,J.B. (2011) Early transcriptional defense responses in Arabidopsis cell suspension culture under high-light conditions. *Plant Physiol.*, **156**, 1439–1456.

16. Lecourieux,D., Lamotte,O., Bourque,S., Wendehenne,D., Mazars,C., Ranjeva,R. and Pugin,A. (2005) Proteinaceous and oligosaccharidic elicitors induce different calcium signatures in the nucleus of tobacco cells. *Cell Calcium*, **38**, 527–538.

17. Liao,Z., Chen,M., Sun,X. and Tang,K. (2006) Micropropagation of Endangered Plant Species BT - Plant Cell Culture Protocols. In Loyola-Vargas,V.M., Vázquez-Flota,F. (eds). Humana Press, Totowa, NJ, pp. 179–185.

18. Ochoa-Alejo,N. (2006) Capsaicin Accumulation in Capsicum spp. Suspension Cultures BT - Plant Cell Culture Protocols. In Loyola-Vargas,V.M., Vázquez-Flota,F. (eds). Humana Press, Totowa, NJ, pp. 327–334.

19. Nocarova,E. and Fischer,L. (2009) Cloning of transgenic tobacco BY-2 cells; an efficient method to analyse and reduce high natural heterogeneity of transgene expression. *BMC Plant Biol.*, **9**, 44.

20. James,E. and Lee,J.M. (2006) Loss and recovery of protein productivity in

genetically modified plant cell lines. *Plant Cell Rep.*, **25**, 723–727.

21. Schillberg,S., Raven,N., Fischer,R., Twyman,R.M. and Schiermeyer,A. (2013) Molecular farming of pharmaceutical proteins using plant suspension cell and tissue cultures. *Curr. Pharm. Des.*, **19**, 5531–5542.

22. Holland,T., Sack,M., Rademacher,T., Schmale,K., Altmann,F., Stadlmann,J., Fischer,R. and Hellwig,S. (2010) Optimal nitrogen supply as a key to increased and sustained production of a monoclonal full-size antibody in BY-2 suspension culture. *Biotechnol. Bioeng.*, **107**, 278–289.

23. Kirchhoff,J., Raven,N., Boes,A., Roberts,J.L., Russell,S., Treffenfeldt,W., Fischer,R., Schinkel,H., Schiermeyer,A. and Schillberg,S. (2012) Monoclonal tobacco cell lines with enhanced recombinant protein yields can be generated from heterogeneous cell suspension cultures by flow sorting. *Plant Biotechnol. J.*, **10**, 936–944.

24. Hanstein,J. von (1880) Das Protoplasma als Träger der pflanzlichen und thierischen Lebensverrichtungen für Laien und Fachgenossen. *Winter, Heidelb.*

25. Klercker,J. Von (1892) A method for isolating living protoplasts. *Ofvers Vetensk. Akad. Forh. Stock.*, **49**, 463.

26. Küster,E. (1927) Ober die Gewinnung nackter Protoptasten. *Protoplasma*, **3**, 223-.233.

27. Chambers,R. and Höfler,K. (1931) Micrurgical studies on the tonoplast of Allium cepa. *Protoplasma*, **12**, 338–355.

28. Cocking,E.C. (1960) A Method for the Isolation of Plant Protoplasts and Vacuoles. *Nature*, **187**, 962–963.

29. Takebe,I., Otsuki,Y. and Aoki,S. (1968) Isolation of tobacco mesophyll cells in intact and active state. *Plant Cell Physiol.*, **9**, 115–124.

30. Power,J.B. and Cocking,E.C. (1970) Isolation of Leaf Protoplasts: Macromolecule Uptake and Growth Substance Response. *J. Exp. Bot.*, **21**, 64–70.

31. Sinha,A., Wetten,A. and Caligari,P. (2003) Effect of biotic factors on the isolation of Lupinus albus protoplasts. *Aust. J. Bot. - AUST J BOT*, **51**.

32. Ortín-Párraga,F. and Burgos,L. (2003) Isolation and culture of mesophyll protoplast from apricot. *J. Hortic. Sci. Biotechnol.*, **78**, 624–628.

33. Sutiojono,E., Kantharajah,A.S., Nonhebel,H. and Ebert,G. (2002) The effect of glycine and explant source on the yield of viable protoplasts from Cucumis melo cvs. Green Delica and Fastoso and Cucumis metuliferus. *J. Appl. Bot.*, **76**, 168–171.

34. Davey,M.R., Anthony,P., Power,J.B. and Lowe,K.C. (2005) Plant protoplasts: status and biotechnological perspectives. *Biotechnol. Adv.*, **23**, 131–171.

35. Yoo,S.-D., Cho,Y.-H. and Sheen,J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.*, **2**, 1565.

36. Zhao,J., Xu,J., Chen,B., Cui,W., Zhou,Z., Song,X., Chen,Z., Zheng,H., Lin,L., Peng,J., *et al.* (2019) Characterization of Proteins Involved in Chloroplast Targeting Disturbed by Rice Stripe Virus by Novel Protoplast(-)Chloroplast Proteomics. *Int. J. Mol. Sci.*, **20**.

37. Woo,J.W., Kim,J., Kwon,S. Il, Corvalan,C., Cho,S.W., Kim,H., Kim,S.-G., Kim,S.-T., Choe,S. and Kim,J.-S. (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.*, **33**, 1162–1164.

38. Sheen,J. (2001) Signal Transduction in Maize and Arabidopsis Mesophyll Protoplasts. *Plant Physiol.*, **127**, 1466 LP – 1475.

39. Fraiture,M., Zheng,X. and Brunner,F. (2014) An Arabidopsis and Tomato Mesophyll Protoplast System for Fast Identification of Early MAMP-Triggered Immunity-Suppressing Effectors BT - Plant-Pathogen Interactions: Methods and Protocols. In Birch,P., Jones,J.T., Bos,J.I.B. (eds). Humana Press, Totowa, NJ, pp. 213–230.

40. Urquiza-García,U. and Millar,A.J. (2019) Expanding the bioluminescent reporter toolkit for plant science with NanoLUC. *Plant Methods*, **15**, 68.

41. Lin,H.-Y., Chen,J.-C. and Fang,S.-C. (2018) A Protoplast Transient Expression System to Enable Molecular, Cellular, and Functional Studies in Phalaenopsis orchids . *Front. Plant Sci.* , **9**, 843.

42. Basu,S., Krishnan,A., Ambavaram,M., Rahman,L., Ramegowda,V. and Pereira,A. (2014) Identification of genes directly regulated by a transcription factor in rice.

43. Schaumberg,K.A., Antunes,M.S., Kassaw,T.K., Xu,W., Zalewski,C.S., Medford,J.I. and Prasad,A. (2016) Quantitative characterization of genetic parts and circuits for plant synthetic biology. *Nat. Methods*, **13**, 94+.

44. Li,J., Liao,X., Zhou,S., Liu,S., Jiang,L. and Wang,G. (2018) Efficient protoplast isolation and transient gene expression system for Phalaenopsis hybrid cultivar ‘Ruili Beauty’. *Vitr. Cell. Dev. Biol. - Plant*, **54**, 87–93.

45. Wang,Z., Zhang,Z., Zhao,W., Wang,H., Cheng,J., Huang,H., Sui,X. and Li,X. (2012) An efficient cucumber (Cucumis sativus L.) protoplast isolation and transient expression system. *Sci. Hortic. (Amsterdam).*, **150**, 206–212.

46. Chen,S., Tao,L., Zeng,L., Vega-Sanchez,M.E., Umemura,K. and Wang,G.-L. (2006) A highly efficient transient protoplast system for analyzing defence gene expression and protein–protein interactions in rice. *Mol. Plant Pathol.*, **7**, 417–427.

47. Li,J.-F., Bush,J., Xiong,Y., Li,L. and McCormack,M. (2011) Large-Scale

Protein-Protein Interaction Analysis in Arabidopsis Mesophyll Protoplasts by Split Firefly Luciferase Complementation. *PLoS One*, **6**, e27364.

48. Faraco,M., Di Sansebastiano,G. Pietro, Spelt,K., Koes,R.E. and Quattrocchio,F.M. (2011) One Protoplast Is Not the Other! *Plant Physiol.*, **156**, 474 LP – 478.

49. Wu,F.-H., Shen,S.-C., Lee,L.-Y., Lee,S.-H., Chan,M.-T. and Lin,C.-S. (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods*, **5**, 16.

50. Andrianantoandro,E., Basu,S., Karig,D.K. and Weiss,R. (2006) Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, **2**.

51. Rebatchouk,D., Daraselia,N. and Narita,J.O. (1996) NOMAD: a versatile strategy for in vitro DNA manipulation applied to promoter analysis and vector design. *Proc. Natl. Acad. Sci.*, **93**, 10891 LP – 10896.

52. Shetty,R.P., Endy,D. and Knight,T.F. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, **2**, 5.

53. Anderson,J.A., Dueber,J.E., Leguia,M., Wu,G.C., Goler,J.A., Arkin,A.P. and Keasling,J.D. (2009) BglBricks Assembly Standard.

54. Li,S., Zhao,G. and Wang,J. (2016) C-Brick: A New Standard for Assembly of Biological Parts using Cpf1. *ACS Synth. Biol.*, 10.1021/acssynbio.6b00114.

55. Liu,J.K., Chen,W.H., Ren,S.X., Zhao,G.P. and Wang,J. (2014) IBrick: A new standard for iterative assembly of biological parts with homing endonucleases. *PLoS One*, **9**, 1–10.

56. Silva-Rocha,R., Martínez-García,E., Calles,B., Chavarría,M., Arce-Rodríguez,A., de Las Heras,A., Páez-Espino,A.D., Durante-Rodríguez,G., Kim,J., Nikel,P.I., *et al.* (2013) The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex

prokaryotic phenotypes. *Nucleic Acids Res.*, **41**, D666–D675.

57. Martínez-García,E., Aparicio,T., Goñi-Moreno,A., Fraile,S. and de Lorenzo,V. (2015) SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.*, **43**, D1183–D1189.

58. Martínez-García,E., Goñi-Moreno,A., Bartley,B., McLaughlin,J., Sánchez-Sampedro,L., Pascual del Pozo,H., Prieto Hernández,C., Marletta,A.S., De Lucrezia,D., Sánchez-Fernández,G., *et al.* (2019) SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Res.*, **48**, D1164–D1170.

59. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**.

60. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*, **4**, e5553.

61. Moore,S.J., Lai,H.E., Kelwick,R.J.R., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.*, **5**, 1059–1069.

62. Garrigues,S., Manzanares,P., Yenush,L., Orzaez,D., Gandía,M., Hernanz-Koers,M. and Marcos,J.F. (2018) FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. *Fungal Genet. Biol.*, **116**, 51–61.

63. Taylor,G.M., Mordaka,P.M. and Heap,J.T. (2018) Start-Stop Assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Res.*, **47**, e17–e17.

64. van Dolleweerd,C.J., Kessans,S.A., Van de Bittner,K.C., Bustamante,L.Y.,

Bundela,R., Scott,B., Nicholson,M.J. and Parker,E.J. (2018) MIDAS: A Modular DNA Assembly System for Synthetic Biology. *ACS Synth. Biol.*, **7**, 1018–1029.

65. Lin,D. and O'Callaghan,C.A. (2018) MetClo: methylase-assisted hierarchical DNA assembly using a single type IIS restriction enzyme. *Nucleic Acids Res.*, **46**, e113.

66. Iverson,S. V., Haddock,T.L., Beal,J. and Densmore,D.M. (2016) CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.*, **5**, 99–103.

67. De Paoli,H.C., Tuskan,G.A. and Yang,X. (2016) An innovative platform for quick and flexible joining of assorted DNA fragments. *Sci. Rep.*, **6**, 19278.

68. Andreou,A.I. and Nakayama,N. (2018) Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS One*, **13**, e0189892.

69. Pollak,B., Cerda,A., Delmans,M., Álamos,S., Moyano,T., West,A., Gutiérrez,R.A., Patron,N.J., Federici,F. and Haseloff,J. (2019) Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.*, **222**, 628–640.

70. Patron,N. and et al (2015) Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytol.*

71. Vazquez-Vilar,M., Quijano-Rubio,A., Fernandez-del-Carmen,A., Sarrion-Perdigones,A., Ochoa-Fernandez,R., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2017) GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Res.*, **45**, gkw1326.

72. Engler,C., Youles,M., Gruetzner,R., Ehnert,T.M., Werner,S., Jones,J.D.G., Patron,N.J. and Marillonnet,S. (2014) A Golden Gate modular cloning toolbox

for plants. *ACS Synth. Biol.*, **3**, 839–843.

73. Gantner,J., Ordon,J., Ilse,T., Kretschmer,C., Gruetzner,R., Löfke,C., Dagdas,Y., Bürstenbinder,K., Marillonnet,S. and Stuttmann,J. (2018) Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. *PLoS One*, **13**, e0197185.

74. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palaci,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiol.*, **162**, 1618–1631.

75. Fesenko,E. and Edwards,R. (2014) Plant synthetic biology: a new platform for industrial biotechnology. *J. Exp. Bot.*, **65**, 1927–1937.

76. Boehm,C.R. and Bock,R. (2019) Recent Advances and Current Challenges in Synthetic Biology of the Plastid Genetic System and Metabolism. *Plant Physiol.*, **179**, 794 LP – 802.

Chapter 1 Mobius Assembly: a versatile Golden-Gate framework towards universal DNA assembly

1.1. Abstract

Synthetic biology builds upon the foundation of engineering principles, prompting innovation and improvement in biotechnology via a design-build-test-learn cycle. A community-wide standard in DNA assembly would enable bio-molecular engineering at the levels of predictivity and universality in design and construction that are comparable to other engineering fields. Golden Gate Assembly technology, with its robust capability to unidirectionally assemble numerous DNA fragments in a one-tube reaction, has the potential to deliver a universal standard framework for DNA assembly. While current Golden Gate Assembly frameworks (e.g. MoClo and Golden Braid) render either high cloning capacity or vector toolkit simplicity, the technology can be made more versatile – simple, streamlined, and cost/labour-efficient, without compromising capacity. Here we report the development of a new Golden Gate Assembly framework named Mobius Assembly, which combines vector toolkit simplicity with high cloning capacity. It is based on a two-level, hierarchical approach and utilises a low-frequency cutter to reduce domestication requirements. Mobius Assembly embraces the standard overhang designs designated by MoClo, Golden Braid, and Phytobricks and is largely compatible with already available Golden Gate part libraries. In addition, dropout cassettes encoding chromogenic proteins were implemented for cost-free visible cloning screening that colour-code different cloning levels. As proofs of concept, we have successfully assembled up to 16 transcriptional units of various pigmentation genes in both operon and multigene arrangements. Taken together, Mobius Assembly delivers enhanced versatility and efficiency in DNA assembly, facilitating improved standardisation and automation.

1.2. Introduction

In typical synthetic biology construction, modular DNA parts (e.g. promoters, coding sequences and terminators) are assembled to build molecular devices (e.g. functional transcriptional units), which can be combined further to assemble genetic modules (e.g. biosynthetic pathways). In addition, standardisation sets rules on how these modular parts are designed and assembled (1). The use of widely accepted standard parts and assembly methods facilitates exchangeability among users, allowing the reusability of available constructs, as well as automation of construction. A simple design would aid the efficiency and versatility of molecular engineering.

DNA assembly thus is a pivotal technology in synthetic biology, since it materialises the construction of molecular modules, such as transcriptional units (TUs) and genetic circuits, from individual DNA parts, such as promoters, coding sequences and terminators (2). DNA synthesis technologies are rapidly improving to provide affordable *ex vivo* synthesis of large DNA sequences; however, DNA assembly is and will be the predominant strategy for the assembly of DNA fragments larger than 1kb for the foreseeable future (3). Currently, there are several DNA assembly methods used across the synthetic biology community (for an extensive review, see (4)). Those methods mainly fall into three categories: long-overlap assembly (e.g. Gibson Assembly (5)), site-specific recombination that exploits phage integrases (e.g. Gateway cloning (6)), and restriction endonuclease-based strategies. The endonuclease-based assembly methods are the most commonly used category that allows standardisation when following a specific framework and set of rules.

One of the first attempts to standardise a restriction enzyme-based DNA assembly method was BioBricks, which was developed more than a decade ago (7). The reusability and simplicity of BioBricks make them popular; for example, they became the standard DNA assembly framework for the iGEM (international

Genetic Engineered Machine) competition, which has played an instrumental role in nurturing new generations of synthetic biologists and synthetic biology tools and innovations (http://igem.org). There have been efforts to alleviate their drawbacks, such as the in-frame stop codon in the fusion scar and frequent need for 'domestication' (i.e. elimination of internal restriction sites) (8–10). However, the pairwise nature of BioBrick assembly makes the construction of multipart systems time-consuming. Cloning with BioBricks can be labour-intensive since the digestion and ligation take place in separate reactions.

Within a few years following BioBricks, a new generation of DNA assembly technology called Golden Gate Assembly was introduced (11, 12). This is based on Type IIS restriction endonucleases, which cleave double-stranded DNA outside their recognition sites. They leave a short single-stranded overhang, whose sequence can be defined by users. Golden Gate Assembly employs *BsaI* restriction sites, which are eliminated during subcloning, allowing simultaneous digestion and ligation in a one-pot/tube reaction. Additionally, the use of distinct 4bp overlaps allows unidirectional, scarless cloning of multiple parts, since they are molecular screws and screw holes that are specifically designated to find each other.

Nonetheless, the original Golden Gate Assembly framework lacked reusability, since the composite parts (e.g. TUs) could not be assembled further in multigene constructs. It also lacked standardisation, since no community-wide rule was defined for the 4bp overhangs. Major breakthroughs came when the MoClo and Golden Braid variants of Golden Gate Assembly were developed to enable the hierarchical construction of multi-TUs and the full reusability of composite parts (13–16).

The Golden Braid 2.0 framework uses a simple pairwise approach where multipartite expansion is achieved by switching between two levels, α and Ω (14). The standard parts feed the α and Ω levels, and the shift between the α and Ω levels doubles the number of TUs. The core vector toolkit of Golden Braid is

comprised of only five plasmids. To date, the Golden Braid toolkit is mainly targeted at plant systems, although it can be made compatible with other chassis with a few modifications. An updated version of Golden Braid, G.B 3.0, comes with an online list of reusable standard parts providing the characterisation data (17).

On the other hand, the MoClo framework uses a complex yet high capacity vector toolkit to achieve parallel assembly (15, 16). The standard parts feed Level 1, which is comprised of seven plasmids. Assembly of up to six TUs takes Level 1 to Level 2 or to Level M/P, each of which involves seven vectors and a suite of End-linkers. The first direction employs seven Level 2 vectors and 21 End Linkers and can assemble up to six TUs per round (15). The use of seven of the End Linkers results in non-modular constructs, and the cloning cannot further continue to the next level. The exploitation of the other fourteen End Linkers allows the augmentation of up to six TUs. The second direction uses two more levels of vectors, M and P levels, seven for each, each of which contains seven End Linkers (16). Furthermore, up to six TUs are fused in a Level M vector, while switching between the M and P levels allows multigene augmentation. The MoClo toolkit was initially released for general eukaryotic expression (15), which was then adapted for plants (18), mammalian cells (19) and yeast (20, 21), and it has just recently been extended for use in *Escherichia coli* (22–24) with modifications in the vectors and/or assembly standards. The mammalian MoClo (mMoClo) includes six parts positioning vectors to generate the standard parts, nine TU positioning vectors (Level 1), nine linker vectors and one destination vector (Level 2), which give the capacity of up to 9TU construct (19). Instead of distributing a set of vectors, MoClo Yeast Toolkit provides parts to construct the vectors and uses the Assembly Connectors for the generation of multi-TUs (20). With the five current Assembly Connectors in the toolkit, users can assemble up to 6TUs, but by constructing new Assembly Connectors they can expand the assembly capacity. The EcoFlex MoClo vector toolkit for *E. coli* contains two Level 0, six

Level 1, six Level 2, two Level 3 vectors, and two Secondary Modules (23). Level 2 can receive up to 5TUs from Level 1, and up to 20TUs can be assembled into Level 3. The CIDAR MoClo toolkit contains six Transcriptional Unit Vectors (DVK) and sixteen Basic Part/Device Vectors (DVA) which allows in two levels the generation of up to 4 part devices/circuits (24). In all cases, MoClo systems tend toward more complex vector toolkits, sometimes with capped cloning capacity in some.

Assembly speed, toolkit and protocol simplicity, and cloning capacity make a DNA-assembly method attractive to users; however, the two most popular Golden Gate variants compromise at least one of them. Golden Braid sacrifices capacity of multi-gene cloning in favour of simplicity, while MoClo emphasises cloning capacity and in so doing complicates its vector toolkit. A simple assembly method helps users assimilate and troubleshoot their cloning or vector toolkit problems; on the other hand, high capacity assembly methods are preferred because they are time- and cost-effective. Furthermore, since both methods implement Type IIS restriction enzymes that are frequent cutters, they are burdened with heavy requirements for domestication, which is labour-intensive.

In order to address the tradeoffs and limitations of these current methods, ultimately to encourage universal standardisation in synthetic biology construction, we developed Mobius Assembly, a new, highly versatile framework for hierarchical Golden Gate Assembly. Mobius Assembly embodies both simplicity and cloning capacity and thus allows exponential and theoretically unlimited augmentation of TUs. The two-level design, comprised of four Acceptor Vectors in each level and seven Auxiliary Plasmids, enables a quadruple assembly with a compact vector toolkit. Mobius Assembly also adopts the 4bp standard overhangs defined by MoClo and Golden Braid to promote the sharing of standard parts. Another new feature, the replacement of a frequent cutter with the rare cutter *AarI* reduces domestication needs. Furthermore, the vectors are demarcated with specific visible markers for cloning screening. As a proof of

concept, we have used Mobius Assembly to successfully reconstruct multi-gene biosynthetic clusters to produce protoviolaceinic acid and carotenoids. Additionally, to validate the capacity of the cloning system, we built a 16TU construct.

1.3. Results and discussion

1.3.1. Design features of the Mobius Assembly framework

The Mobius Assembly framework commences at Level 0, which represents the standard part library. It uses the Mobius Universal Acceptor Vector (mUAV), to convert amplified PCR fragments into standard, interchangeable parts (**Figure 1A**). mUAV has a backbone derived from pSB1C3 and thus confers chloramphenicol resistance. We introduced the chromoprotein amilCP (25) as a visible cloning screening marker, which imbues a purple colour to the colonies (**Figure 2B**) (see below for the choice of the visible reporter genes). This negative screening marker is flanked by *AarI* recognition sites. *AarI* cuts through the *BsaI* sequence, generating fusion sites (CTCT and TGAG) where a PCR fragment will be cloned in (**Figure 1A**). The insert should be amplified with a pair of primers each of which bears an *AarI* restriction site, a fusion site that matches with the mUAV overhangs, and a 4bp standard overhang, from 5' to 3'. *AarI* digestion releases the *amilCP* gene, which is replaced by a standard part, resulting in a Level 0 plasmid. It should be noted that users can use any backbone in all levels of Mobius Assembly if the backbone we provide does not meet specific experimental requirements. Mobius Assembly cassettes are flanked with *Eco*RI and *Pst*I restriction sites, and the backbone can be swapped with a simple digestion/ligation step.

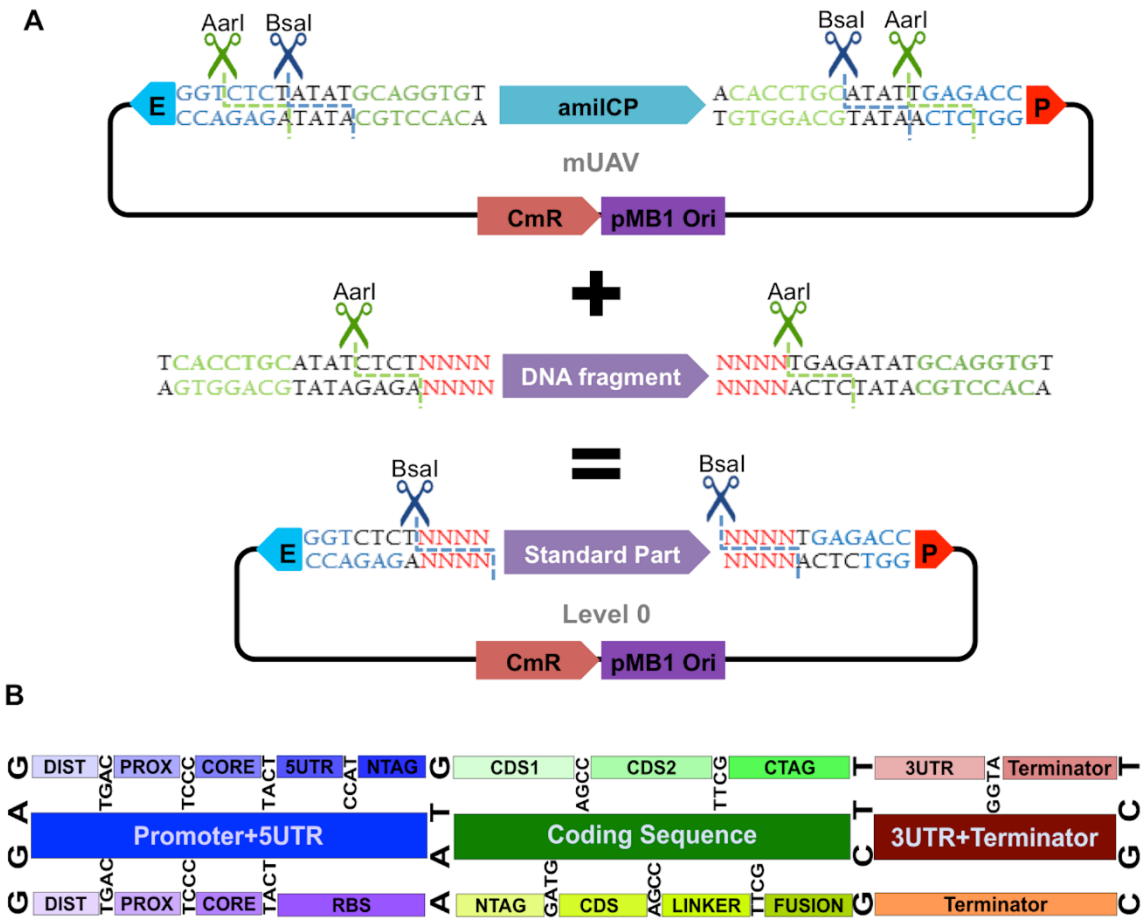


Figure 1. Mobius Assembly standard part generation. (**A**) Mobius Universal Acceptor Vector (mUAV) is the vector which converts and hosts DNA fragments as standard parts. mUAV is flanked by the Type IIS restriction enzymes *BsaI* and *AarI* and carries amilCP gene as visible cloning screening marker. The inserts are amplified with primers containing *AarI* recognition sites, the fusion sites with the mUAV, and the standard overhangs, and they replace amilCP cassette in a Golden Gate reaction. The standard parts are released by *BsaI* digestion. E: *EcoRI*; P: *PstI*. (**B**) Mobius Assembly embraces the 4bp standard part overhangs defined by MoClo, Golden Braid, and Phytobricks, to facilitate part sharing. The middle row illustrates the standard overhangs for major functional parts (promoter, coding sequence, and terminator); the top row shows the recommended overhangs for eukaryotic sub-functional parts, while the bottom row indicates ones for the prokaryotic counterparts.

Mobius Assembly was designed such that the standard parts are released by *BsaI* digestion, as in MoClo and Golden Braid, to facilitate exchangeability. At the same time, we introduced *AarI* as a second restriction enzyme to address the domestication issue (**Figure 1A**). We opted for *AarI* because it is a rare cutter that recognises the 7bp sequence CACCTGC(4/8)^ and leaves a 4bp overhang. Other

Type IIS rare cutters leave 2 or 3bp overhangs or contain a large (e.g. 20bp) space between the recognition and cut sites. Golden Braid 2.0 (14) employs three restriction enzymes *BsaI*, *Bsm*BI and *Btg*ZI, all of which recognise 6bp sequences. MoClo (15, 16) also uses 6bp cutters *BsaI*, *Bpi*I and sometimes *BsmB*I. The exchange of a 6bp cutter with *AarI* thus theoretically drops domestication requirements by 58.3% compared to the systems using three 6bp cutters (e.g. Golden Braid 2.0) and by 37.5% to the systems using two 6bp cutters (e.g. CIDAR MoClo), while maintaining assembly efficiency and a 4bp overhang.

Mobius Assembly embraces the standard 4bp overhangs used by MoClo and Golden Braid. These specific sets of overhangs are becoming more common; being also adopted for Phytobricks, the newly emerging standard part collection for the iGEM registry (**Figure 1B**). Phytobricks were developed to propose a unifying design for universally interchangeable DNA parts (26). Between MoClo and Golden Braid 2.0, there is partial compatibility of standard parts since they use the same sets of 4bp overhangs and use *BsaI* for the TU assembly. Full compatibility is possible when a sequence is free from all restricted recognition sites used by each assembly framework. However, because the additional enzymes they require are frequent cutters, direct compatibility is limited. The scarcity of *AarI* sites facilitates direct use of the available standard parts that have been generated by MoClo or Golden Braid, as well as Phytobricks, by reducing re-domestication requirements. We searched for the presence of *AarI* recognition sites through existing publically available standard parts compatible with Mobius Assembly, and we found none in great majority: for CIDAR MoClo only one out of 59 parts (cre_CD) contains an *AarI* site; from the MoClo Plant Parts Kit, four out of 95 contain *AarI* sites (pICSL80016, 43844, 75111, and 42222); in Golden Braid 2.0 Kit, 8 out of 56 parts need domestication for *AarI* (GB0082, 0145, 0096, 0208, 0575, 1041, 1079, and 0023); and all 20 Phytobricks in the iGEM distribution collection hold no *AarI* restriction sites. Therefore, Mobius Assembly is directly compatible with existing and future standard parts with other Golden Gate DNA

assembly systems. Less domestication also renders Mobius Assembly more efficient for the generation of new standard parts.

To enable theoretically infinite assembly with a simple vector toolkit, we establish a two-level cloning framework that undergoes cycled, two-tier hierarchical augmentation, hence the name “Mobius Assembly”. The single TU assembly takes place in Level 1, and multi-TU assembly can be further continued by switching back and forth between Level 1 and Level 2 vectors **(Figure 2A)**. Mobius Assembly enables the assembly or the addition of any number of composite parts as far as the vector or the chassis can handle.

There are four Level 1 Acceptor Vectors, all of which are equipped with a kanamycin resistance gene since the Mobius Assembly cassette is housed in the pSB1K3 backbone (**Figure 2A**). In Level 1 vectors, the *spisPink* chromoprotein gene (25) serves as a negative cloning screening marker, which colours the colonies pink (**Figure 2C**); it is released by *BsaI* digestion and marks colonies with successfully assembled constructs as white. All Level 1 Acceptor Vectors contain the standard fusion sites at the 5’ (GGAG), and 3’ (CGCT) ends to house a (multi-)TU, plus the additional 4bp fusion sites for the cloning of up to four TUs in a Level 2 Acceptor Vector (**Figure 3A**).

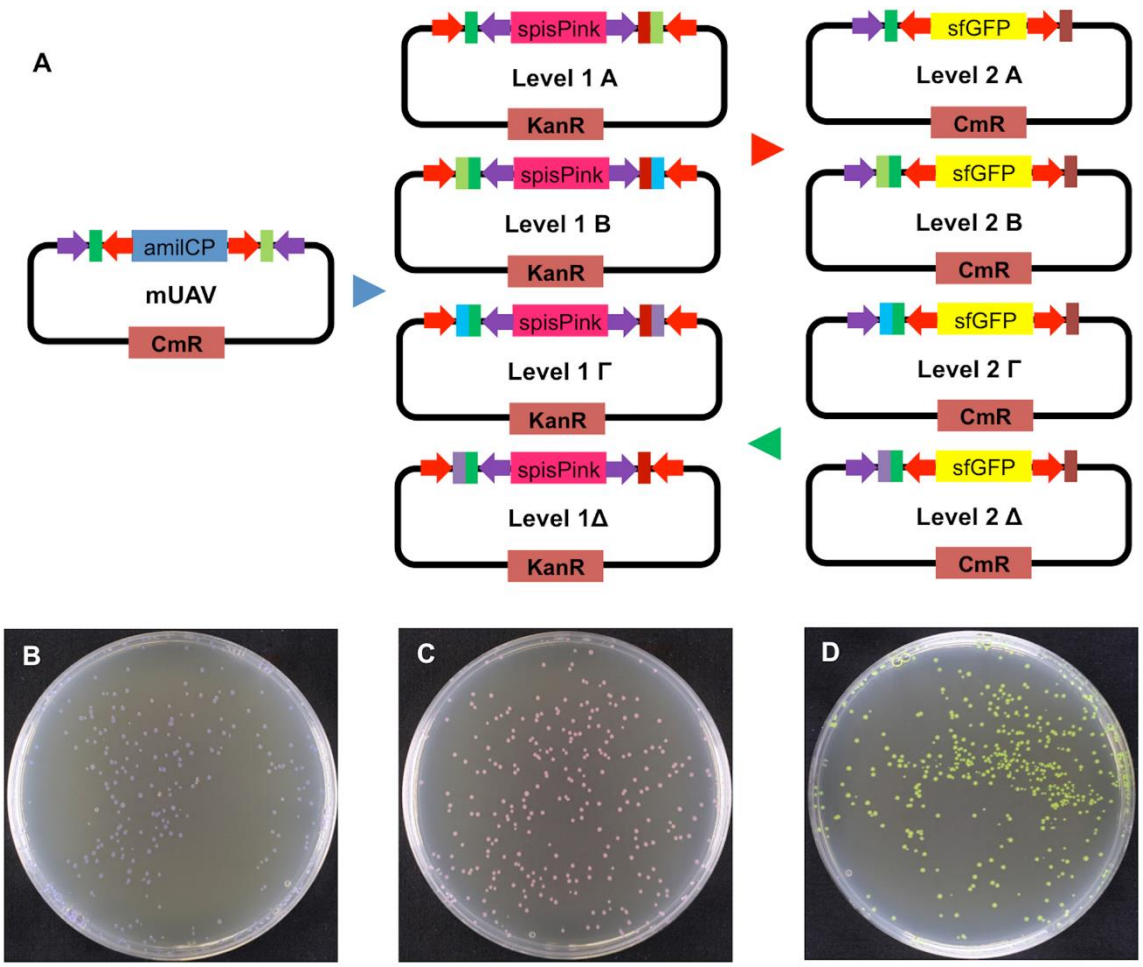


Figure 2. Mobius Assembly framework. (**A**) Mobius Assembly uses a two-level (Level 1 and 2) approach for the transcriptional unit (TU) and multi-TU augmentation. Each level is comprised of four Acceptor Vectors. The four Level 1 Acceptor Vectors (A, B, Γ, and Δ) carry *spisPink gene* as the visible cloning screening marker and confer Kanamycin resistance. The four Level 2 Acceptor Vectors (A, B, Γ, and Δ) carry *sfGFP gene* as the visible cloning screening marker and confer Chloramphenicol resistance. The standard parts stored in mUAVs are released and fused in a Level 1 reaction to form a TU. Up to four Level 1 TUs can be fused in a Level 2 reaction to form a multi-TU cassette. Switching back and forth between Level 1 and 2 leads to further expansion of multi-TUs according to the geometric sequence: 1, 4, 16, 64,…. Red arrows denote *AarI* restriction sites and Purple arrows *BsaI* restriction sites. (**B, C, and D**) *E. coli* colonies carrying mUAV (**B**), Level Acceptor 1 Vector A (**C**), and Level 2 Acceptor Vector A (**D**), which respectively exhibit purple, magenta and yellow colour after overnight incubation. Successful assembly produces white colonies.

Level 2 is comprised of four Acceptor Vectors, which have the pSB1C3 backbone that confers chloramphenicol resistance (**Figure 2A**). They all contain the *sfGFP* (superfolder GFP) gene (27) as a negative cloning screening marker, which makes the colonies yellow (**Figure 2D**); the screening marker is released by *AarI* digestion and successful assembly results in white colonies. Level 2 Acceptor Vectors have the same 5' overhangs as Level 1 Acceptor Vectors and a common 3' fusion site (ACCC), where the linkers from the Auxiliary Plasmids will anneal, providing the appropriate fusion sites to enable the assembly of up to four TUs in a Level 1 Acceptor Vector (**Figure 3C**).

In total, there are seven Auxiliary Plasmids that provide four End-to-End and three Middle-to-End 50bp linkers, which confer kanamycin resistance (**Figure 3B**). In the scenario where four Level 1 vectors are assembled into a Level 2 Acceptor Vector, an End-to-End Auxiliary Plasmid (4A, 4B, 4Γ, or 4Δ) is recruited to provide a linker containing three types of overhangs: i) the 5' End overhang CGCT, which anneals to the 3' overhang of Level 1 Vector Δ, ii) Level 2 End overhang depending on the type of Level 2 Acceptor Vector (4A=CAGA, 4B=GTCA, 4Γ=CTTG, or 4Δ=CGCT) used, and iii) the 3' End overhang ACCC (**Figure 3B and D**). The Auxiliary Plasmid 4Δ is also used when eight or twelve TUs are assembled into B and Γ Acceptor Vectors, respectively. When less than four TUs are fused together, a Middle-to-End Auxiliary Plasmid (1, 2, or 3) is used to provide a linker containing three types of overhangs: i) 5' end overhang depending on the number of TUs being combined (1=CAGA, 2=GTCA, or 3=CTTG), ii) the overhang CGCT necessary to continue assembly back to Level 1, and iii) the 3' end overhang ACCC (**Figure 3B and D**). Cloning from Level 2 to Level 1 does not require any Auxiliary Plasmids.

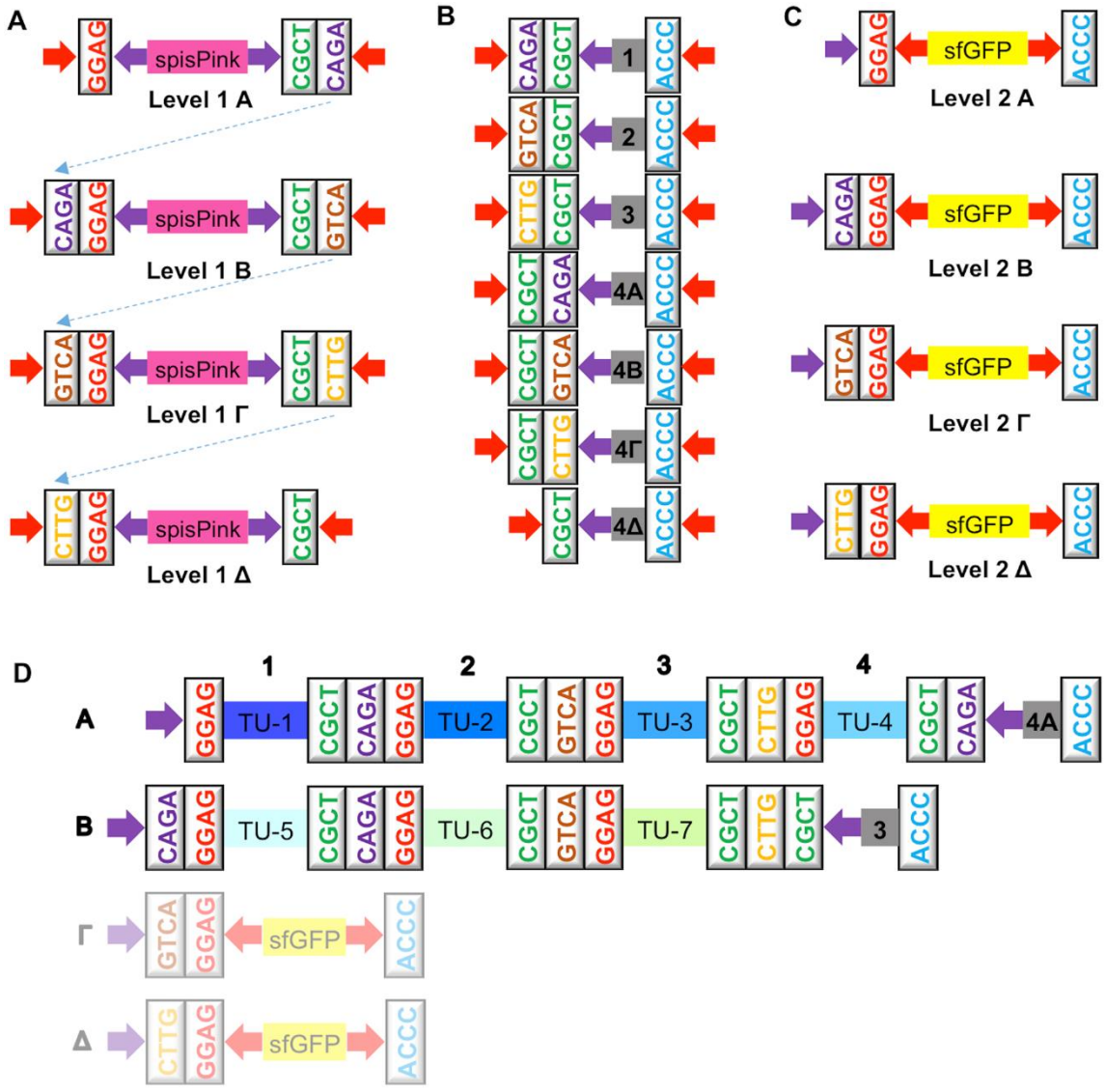


Figure 3 Mobius Assembly vector toolkit. (**A**) The overhangs of the four Level 1 Acceptor Vectors. *BsaI* digestion releases the *spisPink* gene upon digestion to expose GGAG and CGCT, between which a TU will be incorporated. Each type of vector has unique overhangs at the 3' end, which guides the assembly of up to four TUs in a Level 2 Acceptor Vector. (**B**) Seven Auxiliary Plasmids provide End-to-End linkers and Middle-to-End linkers to assist Level 2 cloning. (**C**) The overhangs of the four Level 2 Acceptor Vectors. Digestion with *AarI* releases the *sfGFP* gene and exposes GGAG and ACCC, between which up to four TUs will be fused into with the assistance of an Auxiliary Plasmid. 4A, 4B, 4Γ and 4Δ End-to-End linkers provide 5' and 3' overhangs and the missing Level 2 overhang when four Level 1 TUs are fused. Middle-to-End linkers 1, 2, and 3 are used when one, two or three Level 1 cassettes are fused in Level 2. They provide 5' and 3' overhangs and the CGCT overhang necessary for the cloning back to Level 1. (**D**) An example of how the Auxiliary Plasmids are used. A 7-TU construct is generated by combining the four TUs in the Level 2 Acceptor Vector A and the remaining three TUs in Vector B in a Level 1 reaction. Auxiliary Plasmid 4A is used for the four TUs

in Acceptor Vector A, and the Auxiliary Plasmid 3 for the three TUs in Vector B. Red arrows demarcate *AarI* restriction sites and purple arrows *BsaI* restriction sites.

The design and workflow of Mobius Assembly framework cater to both capacity and simplicity. Having only four Acceptor Vectors in each level and seven Auxiliary Plasmids, Mobius Assembly vector toolkit is simple. Mobius Assembly framework elevates the assembly capability, in the manner described by the exponential geometric sequence $a_n=a_1r^{n-}1$ where *a1*=1, *r*=4 and n = 1, 2, 3, 4, and so on. Moreover, the addition of new TU (single or multiple) in an already constructed multi-TU is possible by switching between the two cloning levels.

Since the protocol for Golden Gate Assembly with *AarI* was not optimised before, two sets of experiments were conducted to optimise the efficiency of Level 2 assembly (four TUs plus one linker). The optimisation was conducted for different buffers and DNA ligases. More specifically *AarI* unique buffer, 2x Tango Buffer (recommended for double digests by the supplier of *AarI*), and T4 DNA ligase buffer were tested in combination with T4 DNA ligase (Thermo Fisher Scientific) or T7 DNA ligase (NEB). The different sets of DNA ligase and buffer were tested for Level 2 assembly of four chromoprotein TUs (tsPurple, amilGFP, asCP, and aeBlue genes, each with J23103 promoter and T7Te terminator). The assembly efficiency was significantly higher when the reaction was carried out in T4 DNA ligase buffer with T4 DNA ligase (**Figure 4 and Table 1**), and thus we have chosen that combination for our construction.

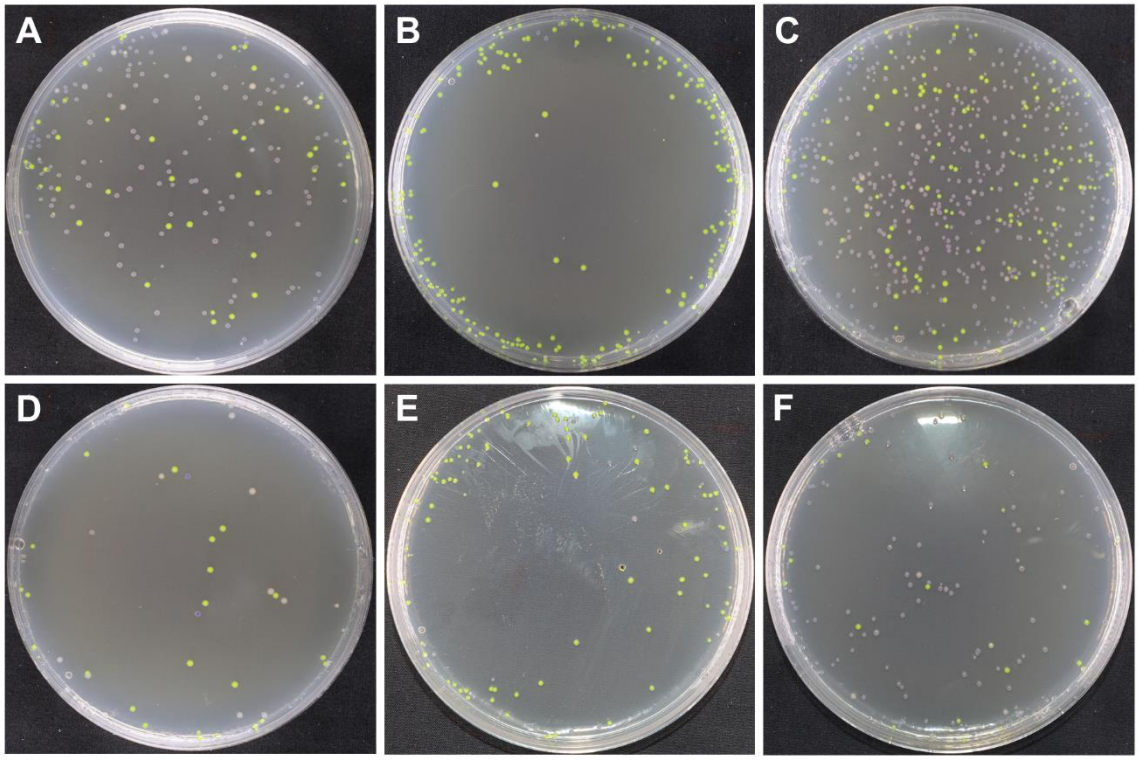


Figure 4. Optimisation of Level 2 assembly reagents. Level 2 cloning was optimised for the different reaction buffers and DNA ligases. The chromoprotein TUs *tsPurple, amilGFP, asCP, and aeBlue* were combined in Level 2 Acceptor Vectors, and the cells with successfully assembled constructs grew into blue colonies. Different sets of reagents tested were: (**A**) Buffer *AarI* + T4 DNA ligase, (**B**) 2x Tango Buffer + T4 DNA ligase, (**C**) T4 DNA Ligase Buffer + T4 DNA ligase, (**D**) Buffer *AarI* + T7 DNA ligase, (**E**) 2x Tango Buffer T7 DNA ligase, (**F**) T4 DNA Ligase Buffer + T7 DNA ligase. T4 DNA Ligase Buffer + T4 DNA ligase combination resulted in the optimum assembly efficiency.

Table 1. **Colony counts for the Level 2 assembly reagents.**

	Buffer-ligase combinations					
	T4 DNA Ligase Buffer-T4 DNA ligase	T4 DNA Ligase Buffer-T7 DNA ligase	2xTango Buffer-T4 DNA ligase	2xTango Buffer-T7 DNA ligase	AarI Buffer-T4 DNA ligase	AarI Buffer-T7 DNA ligase
Replicate 1						
# blue colonies	159	46	7	3	46	4
# yellow colonies	31	12	49	13	4	3
# white colonies	0	1	0	0	0	0
Replicate 2						
# blue colonies	484	86	5	8	124	10
# yellow colonies	117	15	124	75	43	18
# white colonies	8	3	2	1	9	4
Replicate 3						
# blue colonies	298	162	21	16	54	21
# yellow colonies	90	40	167	268	372	192
# white colonies	8	13	4	3	2	3

Furthermore, we tested thermocycling conditions with varied digestion and ligation times for Level 2 assembly. Assembly of the four red-class chromoprotein TUs (tsPurple, efforRed, asCP, or mRFP1 gene together with J23103 promoter and T7Te terminator) were used to test the effect of the thermocycling conditions on the assembly efficiency. Six different thermocycling conditions were tested, all of which were identical in total duration (1 hour and 30min) but varying the period of restriction digestion and ligation, the ratio between them, and the number of the cycles of digestion/ligation. No considerable difference was observed among the various thermocycling conditions for the assembly efficiency of the 3.2kb 4TU

construct (**Figure 5**). Assembly of much larger constructs, however, might be more efficient if many cycles of short digestion/ligation are implemented. The assembly efficiency dropped across different conditions when old reagents (e.g. four months since opening) were used (**Table 2**).

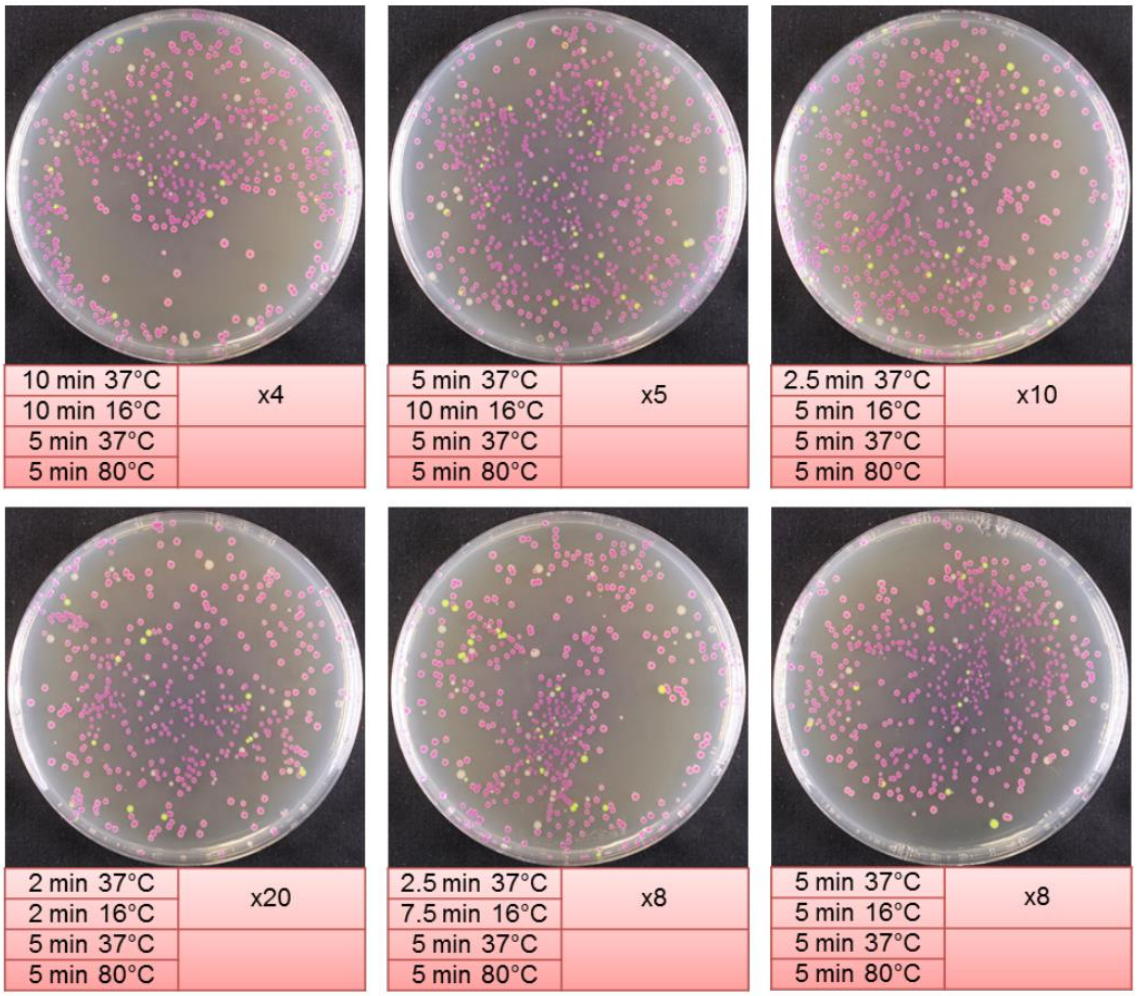


Figure 5. Optimisation of Level 2 assembly reaction conditions. Level 2 cloning was optimised for the different thermocycling conditions used for the assembly reactions. The chromoprotein TUs *tsPurple*, *eforRed*, *asCP* and *mRFP1* were cloned into Level 2 Acceptor Vectors under different thermocycling conditions (digestion/ligation). Cells with successfully assembled constructs grew into pink colonies.

Table 2 **Colony counts for the Level 2 assembly optimisation.**

	Replicate 1					
Digestion (min)-Ligation (min)	**10 - 10**	**5 - 10**	**2.5 - 5**	**2 - 2**	**2.5 - 7.5**	**5 - 5**
# pink colonies	268	438	323	279	289	368
# yellow colonies	9	28	14	9	26	13
# white colonies	13	20	13	9	15	16
Pink/total colonies	0.92	0.90	0.92	0.94	0.88	0.93
Yellow/total colonies	0.03	0.06	0.04	0.03	0.08	0.03
White/total colonies	0.04	0.04	0.04	0.03	0.05	0.04

	Replicate 2					
Digestion (min)-Ligation (min)	**10 - 10**	**5 - 10**	**2.5 - 5**	**2 - 2**	**2.5 - 7.5**	**5 - 5**
# pink colonies	179	104	254	184	171	163
# yellow colonies	22	13	30	22	20	19
# white colonies	1	1	3	2	2	2
Pink/total colonies	0.89	0.88	0.89	0.88	0.89	0.89
Yellow/total colonies	0.11	0.11	0.10	0.11	0.10	0.10
White/total colonies	0.00	0.01	0.01	0.01	0.01	0.01

	Replicate 3					
Digestion (min)-Ligation (min)	**10 - 10**	**5 - 10**	**2.5 - 5**	**2 - 2**	**2.5 - 7.5**	**5 - 5**
# pink colonies	188	135	140	202	186	264
# yellow colonies	64	55	43	81	78	60
# white colonies	4	3	5	3	3	3
Pink/total colonies	0.73	0.70	0.74	0.71	0.70	0.81
Yellow/total colonies	0.25	0.28	0.23	0.28	0.29	0.18
White/total colonies	0.02	0.02	0.03	0.01	0.01	0.01

The last feature we have introduced into Mobius Assembly is visible cloning screening by constitutively expressed chromogenic proteins, which replace Blue-White screening with the inducible lacZ operon. To identify effective visible markers, we screened eight chromoproteins (amilCP, amilGFP, spisPink, asPink, aeBlue, mRFP1, and tsPurple) and sfGFP (superfolder GFP), for strong and fast colour development during overnight incubation. Expression of these marker

genes is controlled by the Anderson promoter J23106 and the rrnBT1-T7TE terminator. *amilCP*, *spisPink*, and *sfGFP*, which, after overnight incubation, develop strong purple, magenta, and yellow colours, respectively, (**Figure 2B-D**), were selected as cloning screening markers for Level 0, 1, and 2. The strength and speed of colour development were strain-dependent. After overnight incubation expression for all the chromogenic genes was faster in TOP10 strain than in DH5α; while the colour was clearly identifiable in TOP10, in DH5α it was only possible after more prolonged incubation (e.g. 24 hours). The difference in speed of colour development is probably due to differences in plasmid copy number, as we could extract higher concentrations of the plasmid from TOP10.

The benefits of cloning screening with chromogenic proteins are multifold. By eliminating the need for two expensive chemicals – IPTG and chromogenic substrate (X-gal) – the cloning screening becomes less costly. In addition, cloning chassis are no longer confined to the *E. coli* strains harbouring the lacZΔM15 deletion mutation necessary for X-Gal screening. Furthermore, the use of distinct colours in each cloning level assists users in distinguishing between different cloning levels and can be exploited by automated assembly platforms.

1.3.2. Proof-of-concept experiments

To validate the Mobius DNA Assembly in functionally reconstructing multigene constructs, we assembled genes involved in the violacein and carotenoid biosynthesis pathways as well as chromoprotein TUs. For the proof-of-concept experiment, pigmentation genes were chosen as their colour development facilitated the identification of the correctly assembled constructs with the naked eye.

As the first proof-of-concept experiment for Mobius Assembly, to reconstruct biosynthetic pathways organised into clusters sharing regulatory sequences, four genes from the violacein operon were re-assembled. Violacein is a bisindole pigment mainly produced in bacteria of the genus *Chromobacterium* (28). The

amino acid L-tryptophan, which is colourless within the visible range, is converted to the purple pigment violacein by the sequential activities of five enzymes co-localized in an operon: VioA, VioB, VioC, VioD and VioE (**Figure 6A**). VioA converts L-tryptophan into indole-3-pyruvic acid imine, which is then dimerised by VioB. VioE catalyses the conversion of the dimer into protodeoxyviolaceinic acid, which is then converted to protoviolaceinic acid by VioD. The final product, violacein, results from conversion of protoviolaceinic acid via the action of VioC.

We first created the standard parts by PCR amplifying each of the four genes (*vioA*, *vioB*, *vioD* and *vioE*) using BBa_K598019 as the template and cloning it in the mUAV. *vioA* was amplified without a Ribosome Binding Site (RBS), while *vioB* and *vioD* were amplified with their RBS; for *vioE* an RBS was added via PCR primers. Next *vioA* coding sequence was fused to a weak promoter (Anderson Promoter J23103+B0034 RBS) in Level 1 Acceptor Vector A, *vioB* in Level 1 Acceptor Vector B, *vioD* in Level 1 Acceptor Vector Γ, and *vioE* with the rrnBT1-T7Te terminator in Level 1 Acceptor Vector Δ. Finally, the four genes were fused in Level 2 Acceptor Vector A (**Figure 6B**). The expression of the construct gave colonies with a deep green to black colour due to the production of protoviolaceinic acid, indicating successful reconstruction of the cluster (**Figure 6C**). Five colonies were selected for colony PCR, which resulted in products of the expected size (**Figure 6D**). The nature of the pigmentation was identified by spectrophotometry as spanning a wide range of emission wavelengths from UV to the visible spectrum, which is consistent with the dark blackish hues of the colonies (**Figure 6E**).

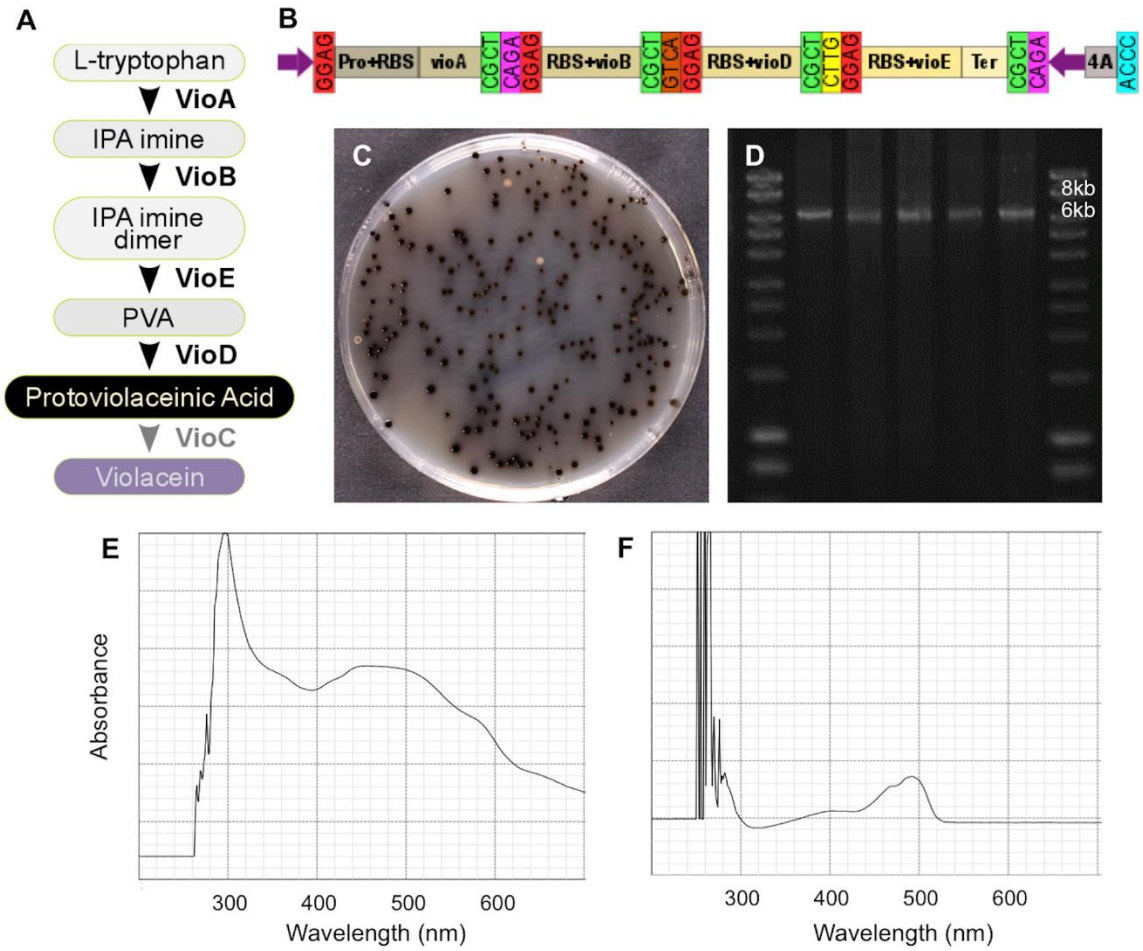


Figure 6. Reconstruction of the violacein biosynthesis operon. (**A**) Schematic of the violacein biosynthetic pathway showing enzymes mediating the conversion of the intermediates. (**B**) Diagram showing the assembly of *vioA*, *vioB*, *vioD* and *vioE* in a Level 2 Acceptor Vector A. (**C**) Cells transformed with the *vioABDE* operon formed dark green to black colonies due to the production of protoviolaceinic acid. (**D**) Agarose gel electrophoresis of PCR products from five dark green/black colonies verified the correct size of the construct (6.5kb). (**E**) UV-Visible range spectrophotometry of the ethanol extract from a dark green colony showed a wide spectrum of absorbance indicative of protoviolaceinic acid. (**F**) Spectrophotometry of sfGFP extract from a colony without the recombinant plasmid. IPA: Indole-3-pyruvic acid; PVA: Protodeoxyviolaceinic acid.

To test the functional reconstruction of a biosynthetic pathway comprised of different TUs, rather than in an operon arrangement, five genes involved in carotenoid biosynthesis were assembled in three different combinations. Carotenoids are a group of omnipresent pigments produced by a diverse range of living organisms, including plants, algae, and microbes (29). *E. coli* cannot

naturally synthesise carotenoids; however, introduction of the carotenoid biosynthetic genes results in accumulation of specific variants (30). The template used in this study is the carotenoid biosynthesis operon from *Pantoea ananatis* (31). Farnesyl pyrophosphate (FPP), which is a precursor of several isoprenoid compounds, naturally exists in *E. coli* and is the substrate for carotenoid biosynthesis. **Figure 7A** depicts the carotenoid biosynthesis pathway. In the first step the enzyme geranylgeranyl pyrophosphate (GGPP) synthase encoded by the *crtE* gene takes isopentenyl pyrophosphate (IPP) along with farnesyl pyrophosphate (FPP) to generate GGPP; GGPP is then converted to a carotenoid intermediate, phytoene, via the action of phytoene synthase encoded by *crtB* gene. The first carotenoid, lycopene, exhibits a pink colour and results from desaturation of phytoene by *crtI*. The lycopene cyclase from the *crtY* gene mediates conversion of lycopene to β-carotene, which is orange in colour. β-carotene is converted to yellow zeaxanthin by the enzyme beta-carotene hydroxylase encoded by *crtZ*.

Cloning of the carotenoid biosynthesis pathways proceeded as follows. Firstly, in Level 0, five genes involved in the carotenoid biosynthesis, *crtE, crtB, crtI, crtY* and *crtZ,* were cloned into the mUAV. Next, we assembled them into TUs (promoter+RBS:coding_sequence:terminator), each of which included the terminator rrnBT1-T7Te. J23110+B0034:*crtE* and J23103+B0034:*crtZ* were cloned in Level 1 Acceptor Vector A and J23103+B0034:*crtB*, J23103+B0034:*crtI*, J23103+B0034:*crtY* in Level 1 Acceptor Vectors B, Γ, and Δ, respectively. The weak promoter J23103 was chosen after we observed that strong overexpression of these carotenoid biosynthesis genes was lethal to the cells (data not shown). To synthesise lycopene, three TUs *crtE*, *crtB*, and *crtI* were assembled in a Level 2 Acceptor Vector A (**Figure 7B**), and successfully constructed cassettes resulted in pink-coloured colonies (**Figure 7C**). For biosynthesis of β-carotene, four TUs *crtE, crtB, crtI*, and *crtY* were assembled, resulting in orange coloured colonies (**Figure 7D**). To form the final expression cassette, TU *crtZ* was cloned in Level 2

Acceptor Vector B, which was fused with *crtEBIY* in a second assembly step back to Level 1 Acceptor Vector A. This five TU construct led to colonies with a yellow colour, consistent with zeaxanthin accumulation (**Figure 7E**). The size of the constructs was verified by colony PCR amplification of the insert (**Figure 7F, G and H**). The identity of the carotenoid variants produced by each construct was verified using spectrophotometry; the expected emission peaks (32) were observed for lycopene (**Figure 5I**), β-carotene (**Figure 7J**) and zeaxanthin (**Figure 7K**) extracts.

Lastly, we tested the hierarchical assembly capacity of Mobius Assembly by assembling a 16TU construct, which is 18.2kb in size (20.4kb with the vector). To create this high-level multi-TU construct, carotenoid, chromoprotein, and violacein TU modules were combined. The chromoproteins for this experiment were selected such that the individual colours would be detectable, even in combination. Eight chromoprotein genes were cloned into Level 1 Acceptor Vectors, each of which was combined with the weak promoter J23103+B0034 and weak transcription terminator for the *E.coli* RNA polymerase, T7Te. We chose to have the expression of each gene low to avoid any possible toxicity due to overproduction of pigments and/or competition for transcription/translation machinery (33). The chromoprotein TUs were grouped into two categories of four genes each, according to their colour ranges – yellow (*scOrange*, *amilGFP*, *amajLime*, *fwYellow*) and pink (*tsPurple*, *efforRed*, *asCP*, *mRFP1*) – and they were respectively assembled in Level 1 Acceptor Vectors A, B, Γ, and Δ.

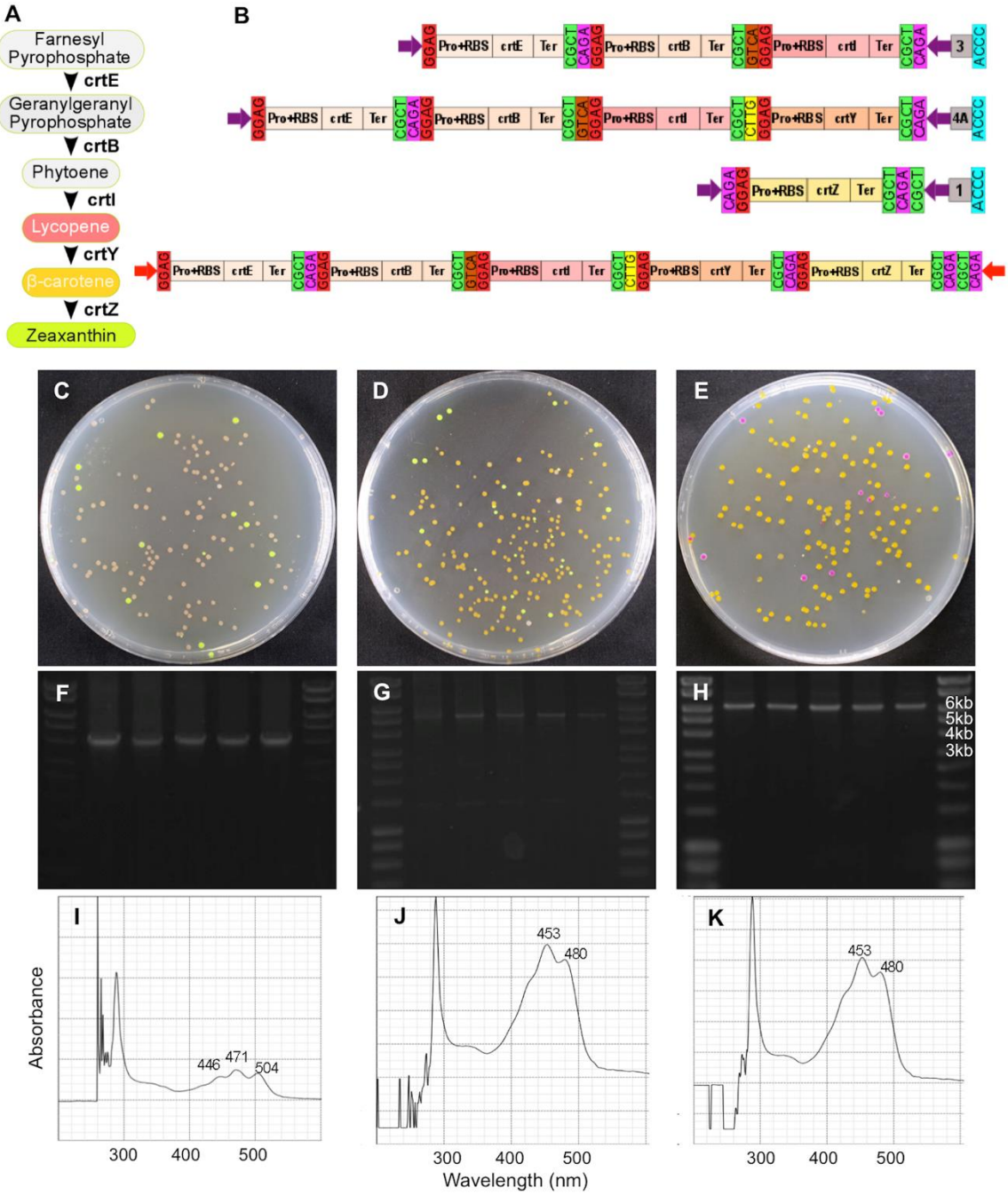


Figure 7. Reconstruction of the carotenoid biosynthesis transcriptional units. (**A**) A schematic of the carotenoid biosynthetic pathway showing the enzymes mediating the production of zeaxanthin as the final product. (**B**) The multi-TU constructs made by Mobius Assembly to produce lycopene (*crtEBI*) and β-carotene (*crtEBIY*) (in Level 2 Acceptor Vectors) and for zeaxanthin (*crtEBIYZ*) by assembling *crtEBIY* and *crtZ* back in a Level 1 A Vector. Colonies producing lycopene are pink (**C**), β-carotene orange (**D**), and zeaxanthin yellow (**E**). Cells carrying intact Level 2 Vectors produced bright yellow colonies, and Level 1 Vectors pink colonies. Gel electrophoresis of the PCR from five

colonies (pink, orange and yellow from each cloning, respectively) verified the correct size of the constructs; 4.3kb for lycopene (**F**), 5.7kb for β-carotene (**G**), and 6.7kb zeaxanthin (**H**). UV-Visible spectrophotometry showed expected peaks for lycopene (446nm, 472nm, and 503nm, **I**), β-carotene (450nm and 478nm, **J**) and zeaxanthin (450nm and 478nm, **K**).

The single chromoprotein TUs were then fused in Level 2 Acceptor Vectors to form 4TU constructs. More specifically, the yellow group was assembled in Level 2 Acceptor Vector B, and the pink group in Γ (**Figure 8A**). The successful constructs were identified firstly by their displayed composite colony colour (yellowish and pink) and secondly by colony PCR of the insert. The violacein operon was reconstructed as described above but in Level 2 Acceptor Vector Δ. In addition, the carotenoid biosynthesis module *crtEBIY*, as described above, was used in Level 2 Acceptor Vector A. In the final assembly step, the four Level 2 Vectors were fused back to Level 1 Acceptor Vector A to generate the 16TU construct (**Figure 8B**). Again, the correct assemblies were distinguished by the black colour of the colonies due to dominant pigmentation by protoviolaceinic acid (**Figure 8C**). Six colonies were selected for DNA plasmid isolation and double restriction digestion with *EcoRI* and *PstI* or *PstI* and *AleI*, which resulted in the anticipated patterns of DNA bands (**Figure 8D and 8E**). The presence of all the 16 TUs in the final construct was further verified by Sanger sequencing.

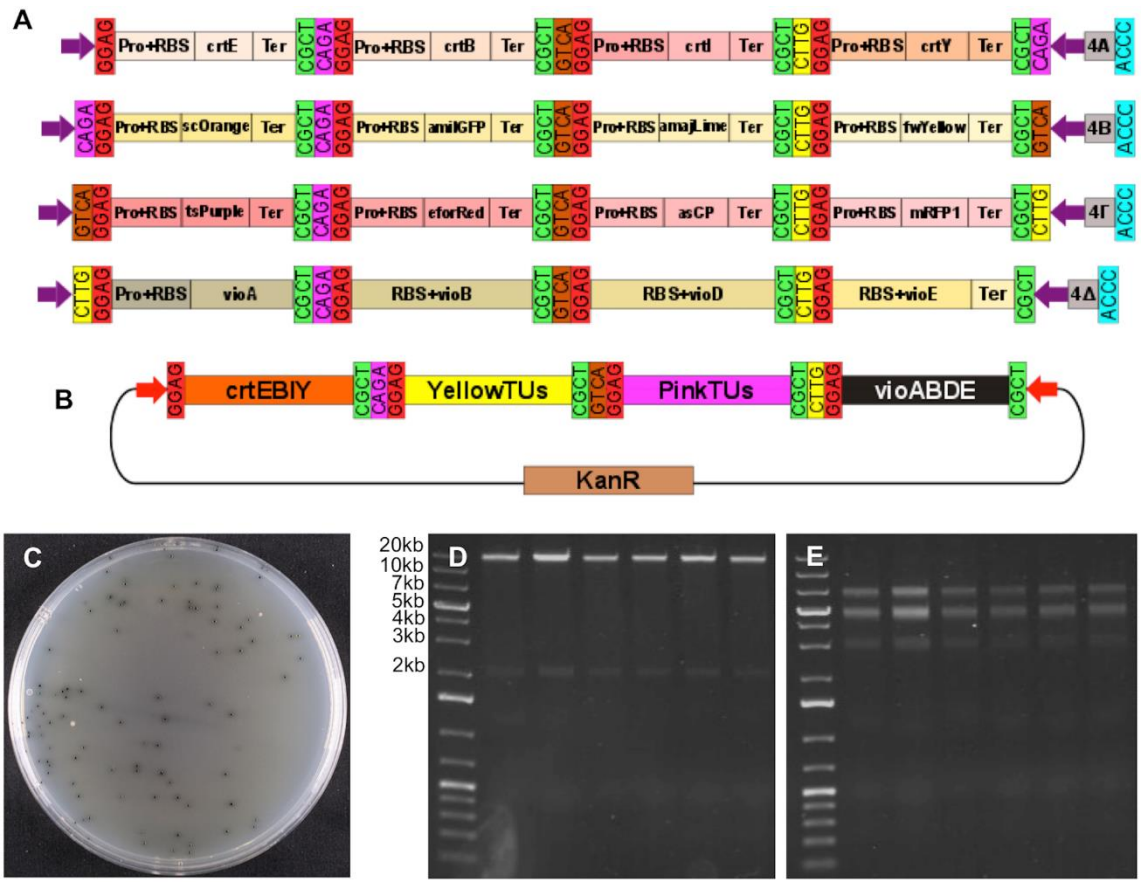


Figure 8. Proof-of-concept assembly of 16TU construct. (**A**) A schematic showing the four intermediate Level 2 constructs for the assembly of the 16-TU construct. The carotenoid biosynthesis genes crtE, crtB, crtI, and crtY assembled in the Vector A, the yellow chromoprotein genes scOrange, amilGFP, amajLime, and fwYellow in the Vector B, the pink chromoprotein genes tsPurple, eforRed, spisPink, and mRFP1 in the Vector Γ, and the violacein biosynthesis genes vioA, vioB, vioD and vioE in the Vector Δ. (**B**) A schematic of the 16TU construct derived from the assembly of the four Level 2 cassettes, each containing 4-TUs, in the Level 1 Acceptor Vector A. (**C**) Cells transformed with the successfully assembled 16TU construct grew into black colonies due to predominant colouring by protoviolaceinic acid. (**D**) Gel electrophoresis of six plasmids (isolated from the black colonies) digested with *PstI* and *EcoRI* resulting in bands of expected sizes - 18.2kb for the insert and 2.2kb for the vector. (**E**) The same plasmids were digested with *PstI* and *AleI* resulting in the bands of expected sizes -7.1kb, 5.1 and 4.9kb (appear merged on the gel), and 3.2kb.

1.4. Materials and methods

1.4.1. Cell culturing and plasmid preparation

E. coli cells (DH5α or TOP10) were grown in 5 ml LB growth medium supplemented with appropriate antibiotics for overnight incubation at 37 °C, 230 rpm. Plasmids were isolated using either Monarch (NEB) or PureYield™ (Promega) Plasmid Miniprep. The Inoue Method (34) was used to prepare home-made ultra-competent cells of DH5α (NEB) or TOP10 (Thermo Fisher Scientific), which were transformed with the constructs as follows: 5 µl of the DNA solution was incubated with 50 µl of the competent cells on ice for 10 min, followed by a heat shock at 42 °C for 40 s and re-cooled on ice for 10 min. SOC medium (400 µl) was added, and after 1 hr incubation at 37 °C, 200 rpm, 50 µl of the cell suspension was plated on LB agar plates with antibiotic selection. The plates were incubated overnight at 37 °C.

1.4.2. Part and vector generation

pSB1C3 (iGEM DNA distribution) was used as the backbone for mUAV and Level 2 Acceptor Vectors, while pSB1K3 (iGEM DNA distribution) was used for the construction of Level 1 Acceptor Vectors. In the preliminary system, pCR8 (Thermo Fisher Scientific) was used as the backbone, and some of our part standards are in the pCR8-based mUAV. Bacterial promoters and terminators, as well as the genes for the carotenoid and violacein biosynthetic pathways, were cloned from the iGEM DNA distribution kit. Chromoproteins were kind gifts from the Uppsala iGEM Association. All the standard parts were domesticated for *AarI* and *BsaI* when necessary and were cloned into mUAV.

To generate the Mobius Assembly TU cassettes for the vector toolkit, J23106 promoter, a chromoprotein gene (amilCP for mUAV, spisPink for Level 1 Acceptor Vectors, and sfGFP for Level 2 Acceptor Vectors), and the rrnBT1-T7Te terminator were combined using Golden Gate assembly. They were then

amplified using Q5 DNA polymerase (NEB) with primers bearing *EcoRI, AarI, BsaI,* and *PstI* restriction sites and 4 bp overhangs. Subsequently, the pSB1C3/pSB1K3 backbones and the Golden Gate cassettes were digested with *EcoRI*-HF and *PstI*-HF (NEB) for 20 min at 37 °C followed by purification with PCR Clean-up kit (Macherey Nagel). Ligation was mediated by T7 DNA ligase (NEB) to construct mUAV, Level 1 and Level 2 Acceptor Vectors. To construct the 50 bp linkers in the Auxiliary Plasmids, a short sequence was PCR amplified from *scOrange* gene using primers that contained the appropriate overhangs and the *AarI* and *BsaI* recognition sites. The PCR products were purified and then digested for 2 hrs at 37 ^{o}C with *AarI* (Thermo Fisher Scientific). A Level 1 Acceptor Vector A was also digested with *AarI* and ligated to the Auxiliary Plasmid cassettes with 1 µl T7 DNA ligase (NEB) for 20min at RT. The constructs were verified by Sanger sequencing (GATC-Biotech or Edinburgh Genomics).

1.4.3. Golden Gate assembly

The DNA assembly was carried out in 10 µl reaction comprised of ~50 ng Acceptor Vector and twice as many molars of the insert parts, in addition to 1 µl 1 mg/ml BSA (NEB), 1 µl T4 DNA ligase buffer (NEB or Thermo Fisher Scientific), 0.5 µl *AarI* (Thermo Fisher Scientific) for cloning in mUAV and Level 2 Acceptor Vectors or *Eco*31I (*BsaI*) (Thermo Fisher Scientific) for cloning in Level 1 Acceptor Vectors, and 0.5 µl T4 DNA ligase (NEB or Thermo Fisher Scientific). For reactions with *AarI*, extra 0.2 µl 50x oligos (0.025 mM) of the enzyme recognition sites were added. The one-tube reaction was incubated in a thermocycler for five times cycles of (37 °C for 5 min, 16 °C for 10 min) followed by 5 min digestion at 37 °C and 5 min deactivation at 80 °C. For the assembly of the 16 TU construct, the reaction was set in 20 µl with double amount of buffers and enzymes and the thermocycling conditions were altered to: 40 cycles of (37 °C for 2.5 min, 16 °C for 5 min) followed by 5 min digestion at 37 °C and 5 min deactivation at 80 °C.

1.4.4. Pigment spectrophotometry

Lycopene, β-carotene, zeaxanthin, and protoviolaceinic acid were extracted from 100 ml overnight cultures in LB. Cells were harvested by centrifugation at 2500 x g for 10 min and then resuspended in 2 ml 96% ethanol and 2 min vortexing for cell lysis. The lysate was then centrifuged for 5 min at 14000 x g, and the supernatant was filtered through a 0.22 µm filter (Millipore) and used for spectrophotometry for the UV-visible range wavelengths (200-780 nm) using Biowave II (Montreal-biotech).

1.5. References

1. Arkin,A. (2008) Setting the standard in synthetic biology. *Nat. Biotechnol.*, **26**, 771–4.

2. Ellis,T., Adie,T. and Baldwin,G.S. (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. (Camb).*, **3**, 109–118.

3. Hughes,R.A. and Ellington,A.D. (2017) Synthetic DNA Synthesis and Assembly : Putting the Synthetic in Synthetic Biology. 10.1101/CSHPERSPECT.A023812.

4. Casini,A., Storch,M., Baldwin,G.S. and Ellis,T. (2015) Bricks and blueprints: methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568–576.

5. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth*, **6**, 343–345.

6. Hartley,J.L., Temple,G.F. and Brasch,M.A. (2000) DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.*, **10**, 1788–1795.

7. Shetty,R.P., Endy,D. and Knight,T.F. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, **2**, 5.

8. Anderson,Jc., Dueber,J.E., Leguia,M., Wu,G.C., Goler,J.A., Arkin,A.P. and Keasling,J.D. (2010) BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng.*, **4**, 1.

9. Liu,J.K., Chen,W.H., Ren,S.X., Zhao,G.P. and Wang,J. (2014) IBrick: A new standard for iterative assembly of biological parts with homing endonucleases. *PLoS One*, **9**, 1–10.

10. Li,S., Zhao,G. and Wang,J. (2016) C-Brick: A New Standard for Assembly of Biological Parts using Cpf1. *ACS Synth. Biol.*, 10.1021/acssynbio.6b00114.

11. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**.

12. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*, **4**, e5553.

13. Sarrion-Perdigones,A., Falconi,E.E., Zandalinas,S.I., Juárez,P., Fernández-del-Carmen,A., Granell,A. and Orzaez,D. (2011) GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules. *PLoS One*, **6**, e21622.

14. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palaci,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiol.*, **162**, 1618–1631.

15. Weber,E., Engler,C., Gruetzner,R., Werner,S. and Marillonnet,S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One*, **6**, e16765.

16. Werner,S., Engler,C., Weber,E., Gruetzner,R. and Marillonnet,S. (2012) Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. *Bioeng. Bugs*, **3**, 38–43.

17. Vazquez-Vilar,M., Quijano-Rubio,A., Fernandez-del-Carmen,A., Sarrion-Perdigones,A., Ochoa-Fernandez,R., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2017) GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Res.*, **45**, gkw1326.

18. Engler,C., Youles,M., Gruetzner,R., Ehnert,T.M., Werner,S., Jones,J.D.G., Patron,N.J. and Marillonnet,S. (2014) A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.*, **3**, 839–843.

19. Duportet,X., Wroblewska,L., Guye,P., Li,Y., Eyquem,J., Rieders,J., Rimchala,T., Batt,G. and Weiss,R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.*, **42**, 13440–13451.

20. Lee,M.E., DeLoache,W.C., Cervantes,B. and Dueber,J.E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.*, **4**, 975–986.

21. Obst,U., Lu,T.K. and Sieber,V. (2017) A modular toolkit for generating Pichia pastoris secretion libraries. *ACS Synth. Biol.*, 10.1021/acssynbio.6b00337.

22. Schindler,D., Milbredt,S., Sperlea,T. and Waldminghaus,T. (2016) Design and assembly of DNA sequence libraries for chromosomal insertion in bacteria based on a set of modified MoClo vectors. *ACS Synth. Biol.*, 10.1021/acssynbio.6b00089.

23. Moore,S.J., Lai,H.E., Kelwick,R.J.R., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.*, **5**, 1059–1069.

24. Iverson,S. V., Haddock,T.L., Beal,J. and Densmore,D.M. (2016) CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.*, **5**, 99–103.

25. Alieva,N.O., Konzen,K. a., Field,S.F., Meleshkevitch,E. a., Hunt,M.E., Beltran-Ramirez,V., Miller,D.J., Wiedenmann,J., Salih,A. and Matz,M. V. (2008) Diversity and evolution of coral fluorescent proteins. *PLoS One*, **3**, 1–12.

26. Patron,N. and et al (2015) Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytol.*

27. Pedelacq,J.-D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotech*, **24**, 79–88.

28. August,P.R., Grossman,T.H., Minor,C., Draper,M.P., Macneil,I.A., Pemberton,J.M., Call,K.M., Holt,D. and Osburne,M.S. (2000) Biosynthetic Pathway from Chromobacterium violaceum 513 Sequence Analysis and Functional Characterization of the Violacein Biosynthetic Pathway from Chromobacterium violaceum JMMB Research Article. *J. Mol. Microbiol. Biotechnol*, **2**, 513–519.

29. Misawa,N. and Shimada,H. (1998) Metabolic engineering for the production of carotenoids in non- carotenogenic bacteria and yeasts. *J. Biotechnol.*, **59**, 169–181.

30. Misawa,N., Nakagawa,M., Kobayashi,K., Yamano,S., Izawa,Y., Nakamura,K. and Harashima,K. (1990) Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in Escherichia coli. *J. Bacteriol.*, **172**, 6704–6712.

31. Nishizaki,T., Tsuge,K., Itaya,M., Doi,N. and Yanagawa,H. (2007) Metabolic engineering of carotenoid biosynthesis in Escherichia coli by ordered gene assembly in Bacillus subtilis. *Appl. Environ. Microbiol.*, **73**, 1355–1361.

32. Rodriguez,D. (2001) A Guide to Carotenoid Analysis in Foods.

33. Ceroni,F., Algar,R., Stan,G.-B. and Ellis,T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods*, **12**, 1–8.

34. Sambrook,J. and Russell,D.W. (2006) The Inoue Method for Preparation and Transformation of Competent E. Coli: “Ultra-Competent” Cells. *Cold Spring Harb. Protoc.* , **2006**, pdb.prot3944.

Chapter 2 : Protoplast isolation and transformation protocol development

2.1. Introduction

Plant cells are fortified by the cell wall, which impedes the delivery of macromolecules (including DNA, RNA and proteins) for routine molecular biology techniques. The treatment of plant tissues or cells with cell wall digestion enzymes results in the release of spherical wall-less cells, called protoplasts. Plant protoplasts are a fast and versatile system to study a plethora of biological processes and constitute a high-throughput platform for identification and characterisation experiments. For example, plant protoplasts have been used for the study of signal transduction (1), high-throughput analysis of how transcription factors regulate well-defined target promoters (2), identification of genes that are regulated by transcription factors (3), to characterise genetic parts and circuits (4), to study the circadian clock in plants (5, 6) and to study protein-protein interactions (7). Protoplast transient expression systems are ideal tools for non-model plants whose stable genetic transformation is not yet available or is tedious and time-consuming (8, 9). Several protoplast transformation protocols have been developed for different species, such as Arabidopsis, rice, maise, cucumber and orchid (7, 10–14).

Most of the protoplast transformation protocols have their critical steps such as the protoplast isolation method, the amount of DNA or protoplasts used, the incubation times or the reagents they use; consequently, they usually rely on the expertise to be efficient and reproducible. For example, the first method used for isolating protoplasts, called Tape-Arabidopsis Sandwich (13), gave reduced transformation results. This could be due to the harsh protoplast isolation process by removing the epidermal cell layer using adhesive tape. A streamlined protocol with reliable functionality and reproducibility was required for the characterisation of a library of plant promoters and terminators (**Chapter 3**), and the crosstalk

experiments between three plant inducible systems were designed to be carried out with Arabidopsis protoplasts (**Chapter 4**). Therefore, a high-throughput transient transformation protocol for Arabidopsis protoplasts in a 1 ml well plate format was developed and optimised. Briefly, protoplasts were isolated from Arabidopsis leaves by mechanical and enzymatic treatment overnight. The protoplasts were harvested and cleared from tissue and cell debris, and then transformed using polyethylene glycol (PEG). PEG creates transient holes in the plasma membrane, thus allowing the DNA to enter the cell more efficiently (15).

2.2. Results and discussion

The protocol development was based on the petal transformation protocol from Faraco et al. (12). Although it is highly efficient, with ~60% yield, the isolation step requires large amounts of starting material and reagents. It is also work-intensive and susceptible to contaminations, as protoplasts float on the medium surface, and their recovery requires removal of the medium with a pump. The final protocol combines the Chupeau protoplast isolation protocol (16), which was is easy and reproducible, with the Faraco transformation protocol with various modifications. Specific parameters were optimised to improve the efficiency of the transformation and simplify the protocol.

For the protoplast isolation, the concentration of the cell wall digestion enzymes was doubled, resulting in almost twice as many protoplasts (~6,1x10^5 cells/ml over ~3.8x10^5 cells/ml). Protoplasts were transformed with a 35S:GFP:35S construct (gift from Alistair McCormick lab, University of Edinburgh) and successful transformants were identified by the presence of GFP emission **(Figure 1)**. Dilutions of the protoplast suspension showed that a starting concentration between ~6x10^4 to 1x10^5 cells/ml was optimal for the transformation efficiency (**Table 1** and **Figure 2**).

Number of protoplasts	$3.8x10^4$	$5.7x10^4$	$6x10^4$	$9.1x10^4$	$1.1x10^5$	$1.8x10^5$
Transformation efficiency	-	++	+++	+++	+++	++
	0%	30%±2.9%	50%±2.1%	50%±1.9%	50%±2.5%	30%±3.7%

Table 1. **Testing the protoplast transformation efficiency using GFP expression.** The percentage of successfully transformed protoplasts was observed in a haemocytometer (Construct used: pGW7.0 35S:GFP:UBQ5, 10 µg DNA, two replicates).

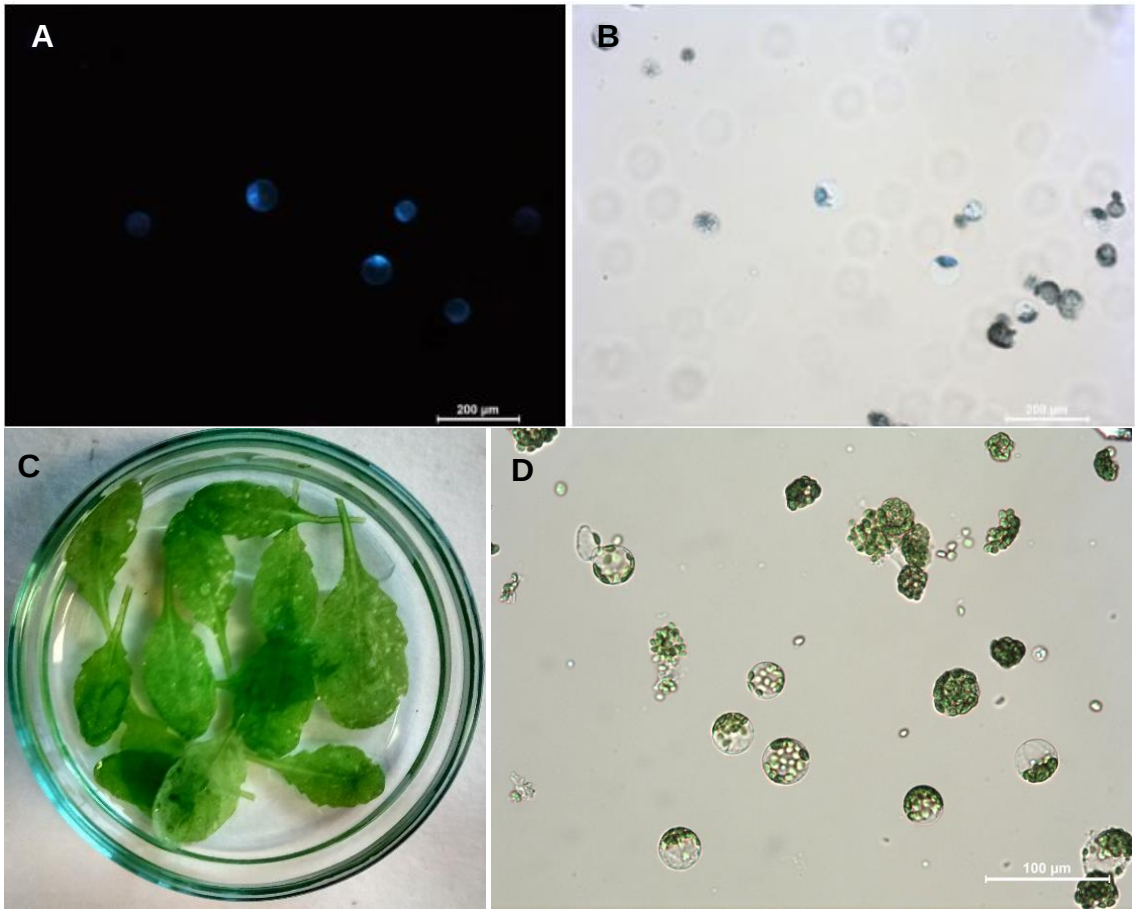


Figure 1. **A and B: Fluorescence microscopy of transformed protoplasts.** Successfully transformed protoplast show a GFP signal (left image), while non-transformed protoplasts are only visible with brightfield microscopy (right image).
C: Mechanical and enzymatic treatment of Arabidopsis leaves: Leaves are treated with a needle-bed and cell wall digestion enzymes to release the protoplasts. **D: Protoplast image showing protoplast cells and cell debris in brightfield.**

Different amounts of DNA, protoplasts and the number of washing cycles were tested for improved transformation efficiencies. A large amount of plasmid DNA recommended for the protoplast transformation, i.e. 20-50 µg per reaction (12), was reduced to make the transformations more cost-effective. The washing steps are the most time-consuming stages in the protocol (centrifugations and change of medium), which can also reduce the protoplast numbers. To further quantify the transformations, the commercial Nano luciferase gene and the Nano-Glo Luciferase Assay System was used, which additionally provides high sensitivity and low background noise compared to the GFP-based fluorescent reporter. Firefly and nano luciferases have been proved great systems for quantitative evaluation of transformation efficiency in plant protoplasts (2, 6).

The wash steps were required to wash out the cell wall digestion enzymes and cell debris before the transformation, and to harvest the cells before the assays. It was found that a reduction of the washing steps to two or one lowers the transformation efficiency, as described in **Figure 2A**. Observation of the samples with brightfield microscopy showed that they still contained tissue debris that could bind DNA and therefore interfered with the transformation (**Figure 1D**). The removal of the tissue and cell debris is the most critical part of the protocol due to the significant influence on the transformation efficiency, which should be assessed before each transformation using brightfield. Omitting the post-transformation centrifugation to harvest the cells before the assay did not affect the luciferase readouts as protoplasts were already settled at the bottom of the plates (**Figure 2C**).

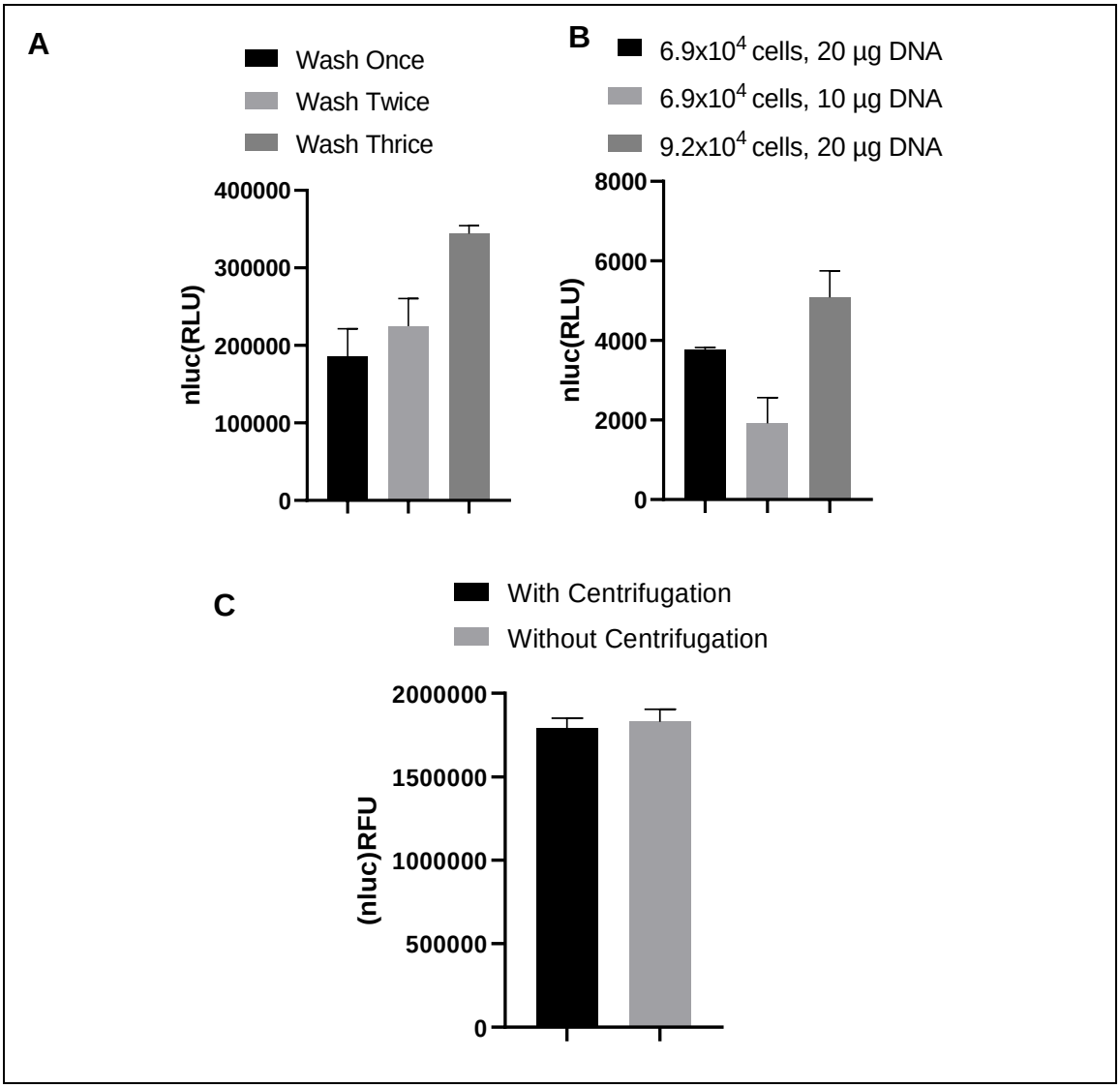


Figure 2. Parameters that affect the protoplast transformation efficiency. The Nano luciferase (nluc) gene was used for the protoplast transformation experiments in combination the Nano-Glo Luciferase Assay System. A. Influence of the pre-transformation centrifugation steps on transformation efficiency. Reduction of the washing steps to two or one lowers the efficiency (construct used, 35S:nluc:35S, $2.7x10^4$ cells, 20 µg DNA, with post-transformation centrifugation). B. Influence of the protoplast and DNA quantity on the transformation efficiency. Higher quantities of DNA or number of protoplasts resulted in higher luminescence readouts (construct used, 35S:nluc:UBQ5, two pre-transformation centrifugation steps, with post-transformation centrifugation). C. Influence of post-transformation centrifugation on the luminescent readouts (construct used, 35S:nluc:35S, $5.63x10^4$ cells, 10 µg DNA, two pre-transformation centrifugation steps). Omitting the washing steps does not change

the luciferase readouts, presumably because the protoplasts were already settled at the bottom of the plates.

Additionally, the number of protoplasts and amount of DNA positively influences the transformation efficiency (**Figure 2B**). All centrifugation steps introduce variation in the luminescent intensities due to the different amounts of protoplast being discarded with the supernatant. To minimise such variations, the Nano luciferase was used as an experimental reporter, and firefly luciferase was introduced as control. The two luciferase genes were delivered to the protoplasts in the same constructs, and the luminescence of nano luciferase was normalised against the luminescence of the firefly luciferase.

Initially, the protocol was used for transient protoplast transformation of constructs in 15 ml glass tubes. However, for the requirements of the high-throughput characterisation experiments, the reaction volume was scaled down to operate in 2 ml and 1 ml 96-well plates. For the 2 ml reaction volume, the protoplast amount was reduced from 300 µl to 150 µl, the DNA from 20 µg to 8 µg, PEG solution from 300 µl to 166 µl and washing solution from 7 ml to 1600 µl. For the 1 ml reaction, the reagents were reduced to 75 µl protoplasts, 4 µg DNA, 83 µl PEG and 800 µl washing solution. The ratio of the nluc and fluc was the same, and absolute values in the 2 ml reaction were more than double the 1 ml reaction, as shown in **Figure 3**. The luminescence signal from the 1 ml reaction was not strong enough to be measured with low sensitivity of the light detection sensor.

Further decreasing of the volume to 300 µl well plates gave high Nano luciferase values (data not shown). Possibly, there was not enough volume to wash out the PEG solution, which affects the assay. To guarantee the transformation efficiency and to retain the low volume of reagents, 1 ml of the reaction was used, and the amount of DNA was doubled.

The final protocol was used for the characterisation experiments of plant promoters and terminators, as well as the crosstalk experiments for the inducible

systems. The detailed materials and method for the isolation and transient transformation of Arabidopsis mesophyll protoplasts can be found below.

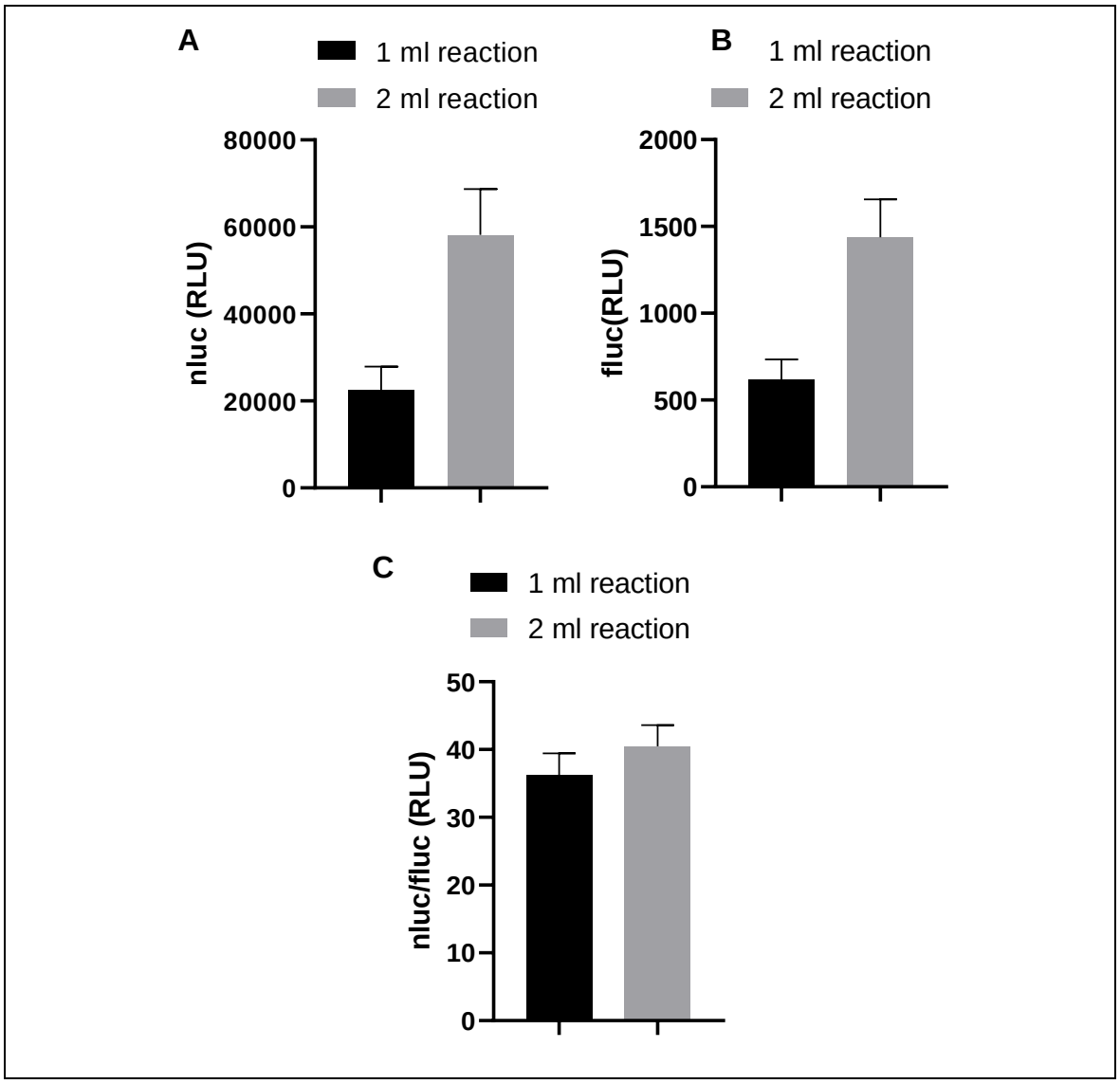


Figure 3. Scaling down protoplast transformation in 96-well plates. The transformation was carried out in 2 ml and 1 ml 96 well plates. The construct UBQ10:nluc:HSP-UBQ10:fluc:UBQ5 was used for the transformations. For the 2 ml well plates, the following reagents were used: 150 µl protoplasts ($5x10^5$ cells/ml), 8 µg plasmid DNA, 166 µl PEG and 1600 µl washing solution. For the 1 ml plates, all the reagents were used in half amounts. Nanoluciferase (**A**) and firefly (**B**) luminescence were almost three times higher in 2 ml than in 1 ml well plates. Even though the normalised values (nluc/fluc) do not have considerable difference (**C**), higher absolute values allow flexibility of the luminesce detector sensitivity.

2.3. Materials

2.3.1. Protoplast Isolation

Unless otherwise stated, all materials were autoclaved, and deionised water was used. Filter-sterilised MGG solution (Used fresh or kept at -80^{o}C). Filter-sterilised MGG solution with enzymes (Used fresh or kept at -80^{o}C). *A. thaliana plants,* 4-6 weeks old. Small glass Petri dishes. Metal tweezers. Scissors. Needle bed sterilised by 70% ethanol (Kenzan). 70% ethanol. 50 ml falcon tubes. 70 µm or 40 µm Millipore filters. 5 ml pipette tips or serological pipettes. Aluminium foil. 1 ml pipette tips with cut off tip. Laminar flow hood. Brightfield microscope.

2.3.2. Protoplast transformation

Filter-sterilised MGG solution (Used fresh or kept at -80 ^{o}C). Filter-sterilised MMM solution (kept at -80 ^{o}C). Freshly prepared PEG solution. Haemocytometer. 5 ml pipette tips or serological pipettes. 1 ml and 200 µl pipette tips with cut off tip. 1 ml 96 well plates. 20 µl, 200 µl and 1 ml multichannel pipettes.

2.3.3. Media

MGG Medium (1 l): 10ml Macrosalts (100× - 2.5 g KNO_3, 0.25 g $MgSO_4$, 0.134 g $(NH_4)_2SO_4$, 0.15 g $CaCl_2$, 0.15 g NaH_2PO_4), 1ml Microsalts A (1000× - 0.075 g KI, 0.3 g H_3BO_3, 1 g $MnSO_4$ 4*$H_2$0, 0.2 g, $ZnSO_4$), 100 µl Microsalts B (10000× -0.25 g Na_2MoO_4, 0.025 g $CuSO_4$, 0.025 g $CoCl_2$), 1 ml Vitamins (1000× -0.1 g Pantothenate-Ca, 0.1 g Niacin, 0.1 g Pyridoxin, 0.1 g Thiamin, 10 g Inositol), 100 µl Biotin (10000× 1 mg/ml), 45 g Glucose, 25 g Glycine, 0.7 g MES. Fill up to 1 l, adjust pH to 5.6 using 1 M NaOH and filter sterilised. MGG with enzymes (200 ml). In MGG medium add: 0.4 g, Onozuka R10 (Yakult), 0.12 g Macerozyme, 0.16 g, Driselase. Sterilise using 0.22 µm syringe filters.
MMM Medium (100 ml): 9.11 g Mannitol, 0.3 g MgCl2, 0.1 g MES.

PEG Solution (10 ml): 0.72 g Mannitol, 0.24 g $Ca(NO_3)_2$ 4*H_2O, 4 g PEG 4000. Adjust pH to 8 with 20 µl KOH (0.5 M).

2.3.4. Protoplast Isolation

Harvest the leaves using scissors and collect them in a beaker with water. Roughly, a half tray of Arabidopsis plants is enough for about 450 transformations. Wash the leaves by sequential submersion in four different beakers filled with water. Dry the leaves on Whatman filter paper and transfer them to small glass plates. Place enough leaves to cover the surface of the petri dish. Eight plates are enough for about 450 transformations. Put 7 ml of MGG (with 2× enzymes) in each glass petri dish and perforate the leaves by pressing them with the needle bed to make them easy to digest. Ensure leaves are evenly distributed across the surface of the petri dish. Wrap the glass dishes in aluminum foil to ensure digestion in darkness and incubate overnight at room temperature.

The next day gently shake the plate to release the protoplasts from the leaves. Check for the presence of protoplasts under the microscope. Place a 70 µm or 40 µm Millipore filter on a 50 ml falcon tube and wet it with MGG medium (with no enzymes). Gently apply the protoplast solution to the filter using tip-cut 1 ml pipette tip to remove any remaining plant debris (one falcon tube per petri dish). Rinse the plate with MGG to fully recover the protoplasts and again apply them to the filter. Centrifuge falcon tubes at 100 min^{-1} for 5 min at 20 ^oC (acceleration = 3 and deceleration = 3). Remove the supernatant using 5 ml pipette and discard. Add 5 ml of MGG to each tube and gently resuspend the protoplasts by rolling the tube and repeat steps 16-17 twice. Count the cells using a hemocytometer and repeat wash steps if there is still too much debris. Keep the protoplast in MGG until you proceed to the transformation section.

2.3.5. Protoplast transformation

Centrifuge falcon tubes at 100 min^{-1} for 5 min at 20 ^{o}C (acceleration = 3, deceleration = 3). Remove the supernatant and add enough volume of MMM in each tube, to achieve protoplast concentration ~$5x10^5$ cells/ml. Gently resuspend the protoplasts by rolling the tubes and combine them in one 50 ml falcon tube. Add 8 µl (500 ng/µl) plasmid DNA to 96 well plates. If you work with numerous constructs, prepare the DNA beforehand. Pour a volume of protoplasts enough for 96 well plates in a multichannel pipette reservoir. Transfer 75 µl protoplasts into each well using a 200 µl multichannel pipette and cut-tip pipette tips and mix by gently hitting the well plate. Add 83 µl PEG solution to each well and incubate for 1 min. Mix by gently hitting the well plate. Do half of the plate first and then continue with the other half. Add 800 µl MGG to each well immediately after 1 min incubation. Incubate plates in darkness for 1 h. Centrifuge for 10 min at 200 min^{-1} at 20 ^{o}C (acceleration=3 and deceleration=3). Carefully remove supernatant using multichannel pipette with slow suction and not disrupting the protoplasts and resuspend in 100 µl MGG. Add any reagents for the experiment (e.g. inducers) and incubate in darkness at room temperature overnight.

2.4. References

1. Sheen,J. (2001) Signal Transduction in Maize and Arabidopsis Mesophyll Protoplasts. *Plant Physiol.*, **127**, 1466 LP – 1475.

2. Wehner,N., Hartmann,L., Ehlert,A., Böttner,S., Oñate-Sánchez,L. and Dröge-Laser,W. (2011) High-throughput protoplast transactivation (PTA) system for the analysis of Arabidopsis transcription factor function. *Plant J.*, **68**, 560–569.

3. Basu,S., Krishnan,A., Ambavaram,M., Rahman,L., Ramegowda,V. and Pereira,A. (2014) Identification of genes directly regulated by a transcription factor in rice.

4. Schaumberg,K.A., Antunes,M.S., Kassaw,T.K., Xu,W., Zalewski,C.S., Medford,J.I. and Prasad,A. (2016) Quantitative characterization of genetic

parts and circuits for plant synthetic biology. *Nat. Methods*, **13**, 94+.

5. Kim,J. and Somers,D.E. (2010) Rapid Assessment of Gene Function in the Circadian Clock Using Artificial MicroRNA in Arabidopsis Mesophyll Protoplasts. *Plant Physiol.*, **154**, 611 LP – 621.

6. Urquiza-García,U. and Millar,A.J. (2019) Expanding the bioluminescent reporter toolkit for plant science with NanoLUC. *Plant Methods*, **15**, 68.

7. Li,J.-F., Bush,J., Xiong,Y., Li,L. and McCormack,M. (2011) Large-Scale Protein-Protein Interaction Analysis in Arabidopsis Mesophyll Protoplasts by Split Firefly Luciferase Complementation. *PLoS One*, **6**, e27364.

8. Lin,H.-Y., Chen,J.-C. and Fang,S.-C. (2018) A Protoplast Transient Expression System to Enable Molecular, Cellular, and Functional Studies in Phalaenopsis orchids . *Front. Plant Sci.* , **9**, 843.

9. Li,J., Liao,X., Zhou,S., Liu,S., Jiang,L. and Wang,G. (2018) Efficient protoplast isolation and transient gene expression system for Phalaenopsis hybrid cultivar 'Ruili Beauty'. *Vitr. Cell. Dev. Biol. - Plant*, **54**, 87–93.

10. Wang,Z., Zhang,Z., Zhao,W., Wang,H., Cheng,J., Huang,H., Sui,X. and Li,X. (2012) An efficient cucumber (Cucumis sativus L.) protoplast isolation and transient expression system. *Sci. Hortic. (Amsterdam).*, **150**, 206–212.

11. Chen,S., Tao,L., Zeng,L., Vega-Sanchez,M.E., Umemura,K. and Wang,G.-L. (2006) A highly efficient transient protoplast system for analyzing defence gene expression and protein–protein interactions in rice. *Mol. Plant Pathol.*, **7**, 417–427.

12. Faraco,M., Di Sansebastiano,G. Pietro, Spelt,K., Koes,R.E. and Quattrocchio,F.M. (2011) One Protoplast Is Not the Other! *Plant Physiol.*, **156**, 474 LP – 478.

13. Wu,F.-H., Shen,S.-C., Lee,L.-Y., Lee,S.-H., Chan,M.-T. and Lin,C.-S. (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods*, **5**, 16.

14. Yoo,S.-D., Cho,Y.-H. and Sheen,J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.*, **2**, 1565.

15. Kofer,W., Eibl,C., Steinmüller,K. and Koop,H.-U. (1998) PEG-mediated plastid transformation in higher plants. *Vitr. Cell. Dev. Biol. - Plant*, **34**, 303–309.

16. Chupeau,M.-C., Granier,F., Pichon,O., Renou,J.-P., Gaudin,V. and Chupeau,Y. (2013) Characterization of the Early Events Leading to Totipotency in an Arabidopsis Protoplast Liquid Culture by Temporal Transcript Profiling. *Plant Cell*, **25**, 2444–2463.

Chapter 3 A characterised toolbox for multigene assembly and delivery in plant systems

3.1. Introduction

Multigene transformation has become a common strategy in plant biotechnology. It is essential in the engineering of metabolic pathways and complex traits, in multiplex genome engineering, and the study of complex genetic circuits and regulatory networks (1–4). An example is the transfer of nine genes in a single construct encoding for fatty acid desaturases and elongases in the Indian mustard plant to increase the yield of long-chain polyunsaturated fatty acids (5). A second example is the simultaneous deletion of six genes in tomato by expressing 12 gRNAs from a single transcript (6). Multigene transformation requires a set of technologies for the successful generation of the constructs and their delivery to plant systems, for which there are many variations of characterised DNA elements that control gene expression.

DNA assembly is an essential method for the generation of recombinant constructs in plant biotechnology for both basic and applied research. The generation of construct libraries of complex biosynthetic pathways demands a fast and easy DNA assembly method. We previously developed a user-friendly and straightforward DNA assembly method, which combines high cloning capacity with a compact vector toolkit, called Mobius Assembly (7). Mobius Assembly is based on IIS restriction endonucleases, which cleave DNA outside their recognition sequence allowing the unidirectional fusion of DNA sequences in a one-tube reaction (8, 9). The popularity and the extensive use of DNA assembly methods based on IIS restriction endonucleases led to the development of new variants and improvements. It is now possible to predict the fidelity of the three or four base overhang ligations (10, 11), while new methods were developed to deal with domestication requirements (12), scars at coding sequence borders (13), the polarity of the TU growth (14) and backbone modularity (15). Also, already

established methods were adjusted for other organisms such as cyanobacteria (16, 17) and fungi (18, 19). Recent developments in the plant field include the Loop Assembly (20) which is an assembly method that shares the user-friendly philosophy of Mobius Assembly, and an extension of the MoClo part toolkit for plants (21).

The promoter is a DNA sequence located near the transcription initiation site (TSS) and is responsible for the commencement and regulation of gene transcription. How far a promoter extends is not precisely defined, and the only distinct element is the core promoter, which is located ±40–50 bp around the TSS (22). The functional elements of the core promoter direct the RNA polymerase II to start RNA synthesis (23, 24). The two most common core promoter elements are the TATA-box and the Initiator (Inr) (24). Highly expressed and tissue-specific genes usually have TATA-box in their core promoter, while ubiquitously expressed or “housekeeping” genes have TATA-less promoters with Inr (24, 25). TATA box consensus sequence is difficult to define; however, structural analysis results revealed a tendency towards TATAWAAR (W = A or T, R = G or A) (26). For higher plants, however, a study of 75 published genomic DNA sequences revealed TCACTATATATAG as the consensus sequence (27). The Inr element spans the TSS, is not conserved among species and is poorly defined (28). In Arabidopsis, the motifs TYA(+1)YYN and TYA(+1)GGG were identified as Inr consensus sequences (29) (Y = T or C and N = G or T or A or C).

Apart from the core promoter elements, several distal and proximal elements participate in transcription and its regulation (30). One of them, the Y-Patch (YP) is a pyrimidine-rich motif which is found in plants but not animals and is associated with both the TATA-box and the Inr motif (31, 32). There are also elements identified to drive tissue-specific or developmental stage-specific gene expression such as the cis-acting element GSE2 responsible for the gene expression in green tissues in rice (33).

Terminators are genetic elements, usually located at the end of a gene or operon, and they terminate transcription. As yet underestimated, the terminator is likely to play critical roles in gene expression by controlling transcription arrest and mRNA stability and tunes other transcription functions. Terminators in plants contain three main elements: a far-upstream U-rich element (FUE), a near-upstream A-rich element (NUE) and a U-rich element (CE), along with a variety of trans-acting polyadenylation factors, which regulate polyadenylation (34, 35). 3'UTR also participates in quality control in the post-transcriptional regulatory mechanism for eukaryotic genes, through Nonsense-mediated mRNA decay (36, 37). Genome-wide analysis of the transcript degradation profiles in Arabidopsis revealed specific sequence motifs at 3' UTR that stabilise or destabilise mRNA (38). Several motifs can be found in the 3' UTR, which might be involved in polyadenylation, or mRNA decay (39). Additionally, polyadenylation influences gene expression. Strong gene expression is linked with short poly(A) tails, while poly(A)-binding proteins can facilitate both the protection and degradation of mRNA (40). In Arabidopsis, the poly(A) tail was found to block RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) from converting aberrant mRNAs into substrates for degradation (41). Unpolyadenylated transcripts which derive from terminator-less constructs or readthrough mRNAs from transgenes with strong promoters are subjected to (RDR6) mediated silencing (42).

Currently, the most used promoters and terminators come from the cauliflower mosaic virus (35S), Agrobacterium opine genes (NOS, MAS, OCS) and recently from plant genome (e.g. UBQ10pro and HSPter from Arabidopsis). The shortage of well-characterised promoters and terminators results in the repeated use of the same parts within a multigene construct. This might negatively impact on plasmid stability due to repetitive sequences (43) or even raise the frequencies of homologous-dependent gene silencing (44). Consequently, a plant toolbox should be equipped with more promoters and terminators. Besides, there is a need for short parts, especially crucial for single vector multigene delivery; they will

substantially help reduce the already large size of multi-TU constructs. Large constructs are connected to structural plasmid instabilities (45), and they decrease the agrobacterium-mediated transformation efficiency in plants (46).

To improve plant transformation efficiency, new vectors need to be developed. Plant binary vectors are an essential tool for the delivery and incorporation of constructs into the plant genome. A typical plant transformation vector is comprised of 1) left and right T-DNA borders for the transfer of the DNA constructs into the host plant's nuclear DNA genome (47); 2) plasmid replication functions in *E. coli and Agrobacterium*; 3) markers for the selection and maintenance in *E. coli and Agrobacterium*. Currently, there are five categories of the vectors based on the origin of replication in Agrobacterium: 1) based on the RK2 origin of replication (e.g. pCB series (48)) or RK2 in combination with ColE1 (eg pORE (49)); 2) based on pVS1 + ColE1 origins of replication (e.g. pCambia (https://cambia.org)); 3) based on pSa + ColE1 origins of replication (e.g. pGreen (50)); 4) based on the pRi origin of replication in combination with ColE1 (e.g. pCGN (51)), or with F factor (eg pBIBAC (52)) or with Phage P1 (eg. pYLTAC (53)); 5) based on the BBR1 origin of replication with the pLX vectors (54). The pLX vectors revolutionised the construction of binary vectors as they shifted from the cut-and-paste strategy to a modular design with minimal functional parts and introduction of stability features.

Plasmid instability issues are one of the main challenges of recombinant DNA production, as it affects plasmid yield and quality (55). For plant binary vectors the challenge is doubled since they should be stable in two different chassis, *E. coli* and Agrobacteria. The plasmid size (45), plasmid copy number (56), direct repeats (43), and inverted repeats (57), among others (55), are the factors affecting structural plasmid stability. Some popular plant vectors (e.g. pCambia) are based on bulky plasmid backbones (6,2 kb), and thus they burden the size of the final constructs by default. Efforts have been made to generate binary

backbones with a smaller size, such as pGreen ((2.5 kb, (50)), pLSU ((4.6 kb (58)) and pLX ((3.3 kb (54)) for the prevention of instabilities.

Stable whole plant transformation is the ultimate goal in plant biology; however, it is laboursome and time-consuming. Transient assays by Agrobacterium-mediated DNA transfer in plant tissues (e.g. leaves) or transformation of isolated protoplasts directly with DNA are faster alternative systems (59). However, these need cultivated plants and are still preparation-intensive when it comes to testing libraries of constructs. Plant cell cultures combine the benefits of plant systems with those of microbial cultures, and they can be established as a next-generation synthetic biology chassis. They can be utilised for transient experiments or, with further improvement, for stable transgene integrations. This new wave of changes has brought plant cell cultures in the spotlight. As new concepts, such as ‘cellular agriculture’, promise to bring revolution to the food industry (60), plant cell cultures are being tested as possible alternatives food sources (61). For example, the first chocolate-based on *Theobroma cacao* suspension cells has been produced (62). Even though bioproduction from plant cell cultures has been traditionally based on extensive cell line screening (63) researchers have started to use metabolic engineering approaches with the characteristic example of anthocyanin production in Nicotiana tabacum cell cultures (64). Nonetheless, the methods and the part toolbox available for genetic engineering of plant cell cultures are limited, and more characterisation and standardisation are wanted.

In the present work, we established a plant engineering toolkit, Mobius Assembly for Plant Systems (MAPS), which is an adaptation of the Mobius Assembly for plant expression. We characterised a library of new promoters and terminators for plants, and we showed the importance of the terminators on regulating gene expression. Our results indicate a synergistic effect between the terminators and the promoters. Furthermore, based on the pLX architecture, we developed a new plant binary vector using the WKS1 origin of replication from a *Paracoccus pantotrophus* cryptic plasmid, suitable for transformation methods that demand

high yield plasmid DNA. We showed that plasmid stability is strain and vector-specific, and we stressed out the importance of testing the plasmid stability before cloning experiments. Finally, for the first time, we demonstrate that plant cell culture transformation is Agrobacterium strain and plasmid specific, paving the way to establish plant cell cultures as a next-generation synthetic biology chassis.

3.2. Results

3.2.1. Mobius Assembly for Plant Systems (MAPS)

Mobius Assembly for Plant Systems (MAPS) is an expansion of our cloning system Mobius Assembly to mediate transformation and gene expression in plants. It combines cloning stages in binary vectors for transformation into plant systems, either whole-plant or cell-based. Like the Mobius Assembly in *E. coli*, MAPS works as a circulation of two cloning levels with four vectors in each level, enabling quadruple augmentation of transcriptional units (**Figure 1**). The introduction of the rare cutter *AarI* minimises the need for removing internal restriction sites, and the exploitation of constitutively expressed chromogenic proteins for clonal screening eliminates the need for additives in the selection media. MAPS vector toolkit consists of a core set of pMAP cloning/destination vectors (Level 1 Acceptor Vectors A-Δ and Level 2 Acceptor Vectors A-Δ), which have a fusion origin of replication (WKS1+pUC) to replicate in *E. coli* and Agrobacteria. The kit also has pLX (BBR1 and RK2) based destination vectors (Level 1 and Level 2 Acceptor Vectors A), which are medium and low copy number vectors respectively, to deal with possible instabilities that occur in large constructs housing repetitive or similar sequences. The mUAV and the 7 Auxiliary plasmids are also included, as described in the original Mobius Assembly kit (7). Also, the MAPS part toolkit contains plant promoters, terminators, antibiotic resistance genes and reporter genes.

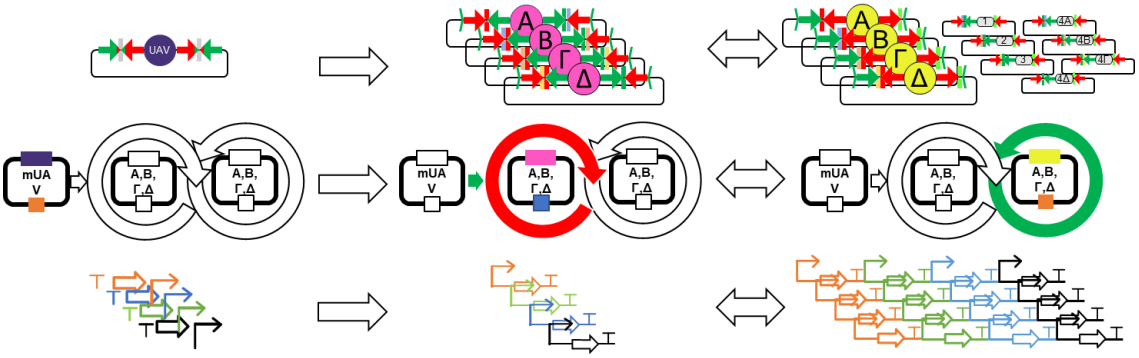


Figure 1. Mobius Assembly for Plant Systems (MAPS). MAPS is an adaptation of the Mobius Assembly for plant systems, based on small plant binary vectors. The core vector toolkit is comprised of one storage vector in Level 0 (mUAV), four Level 1 Acceptor Vectors, four Level 2 Acceptor Vectors and seven Auxiliary plasmids. It employs *BsaI* in Level 1 cloning and the rare cutter *AarI* in Level 0 and Level 2 cloning. In Level 0, functional sequences (e.g. promoters and terminators) are cloned into mUAV to convert them into Phytobricks. Phytobricks are then fused in Level 1 Acceptor Vectors to generated TUs. TUs are further combined in Level 2 Acceptor Vectors to make multi-TUs with the help of the Auxiliary plasmids. The cloning can further continue by switching back and forth between Level 2 and Level 1. The cloning levels are demarcated with chromogenic proteins which are served as negative screening markers: amilCP for Level 0, spisPink for Level 1 and sfGFP for Level 2. Purple, pink and yellow circles represent Level 0, Level 1 and Level 2 Vectors, respectively, the oval shapes are Auxiliary Plasmid, green and red arrows are *BsaI* and *AarI* respectively.

3.2.2. MethylAble feature

Combinatorial DNA library assembly is a powerful method for applications such as part characterisation and metabolic pathway optimisation. However, manual generation of a combinatorial DNA library takes time, effort and resources, especially when it involves several cloning steps. We developed a feature to propagate intact *BsaI* recognition sites during Level 1 cloning, so Level 2 constructs can directly receive Phytobricks (**Figure 2**). According to Rebase (http://rebase.neb.com), GGTCTC^{m5}/C$_{m5}$CAGAG methylation protects *BsaI* digestion. An *amilCP* expression cassette was designed to carry in each site divergent and convergent *BsaI* recognition sites bordering the four base pair overhangs. The overhangs correspond to the part that the MethylAble Feature replaces. The divergent *BsaI* recognition sites were designed to be prone to CpG

methylation (CGGTCTC^{m5}G/GC^{m5}CAGAGC), and consequently, *BsaI* digestion is blocked, while the convergent sites (TGGTCTC^{m5}T/AC^{m5}CAGAGA) are not.

Upon CpG methylation, the *amilCP* TU is cloned with intact divergent *BsaI* sites in a Level 1 reaction, and subsequently to a Level 2 reaction (**Figure 2**). Therefore, Level 0 parts can directly be inserted in a premade Level 1 TU or fused to a Level 2 TU. As *amilCP* chromoprotein will be expressed, the correct constructs will have a purple colour until the Level 0 parts replace the *amilCP* expression cassette. The MethylAble feature can be designed for any standard part, or combinations of parts, by setting the appropriate overhangs. It is also possible to use this feature to change the polarity of the TU growth in a Level 2 reaction by replacing the standard overhangs with the Level 1 overhangs. MethylAble plasmids are built with isothermal assembly using mUAV as a template and overlapping primers to introduce the *BsaI* recognition sites and the selected overhangs.

As a proof of concept, the MethylAble Feature was used to build the library of the three inducible promoters, each of which was combined with 14 new terminator parts. The MethylaAble cassette was designed to have the terminator overhangs (GCTT-CGCT). In Level 1, the constructs were made, as shown in (**Table 1**). Briefly, a normaliser unit was built in Vector A, the three trans-activator units in Vector B and three inducible promoters were combined with the MethylAble feature in Vector Γ. Without the MethylAble feature, the inducible promoters should have been combined with all terminators in this step. Then in a Level 2 reaction the normaliser, the trans-activators and the inducible promoters were fused to generate three constructs in total. Lastly, in a Level 1 reaction, the inducible constructs were combined with Level 0 terminators to give the final 14 constructs.

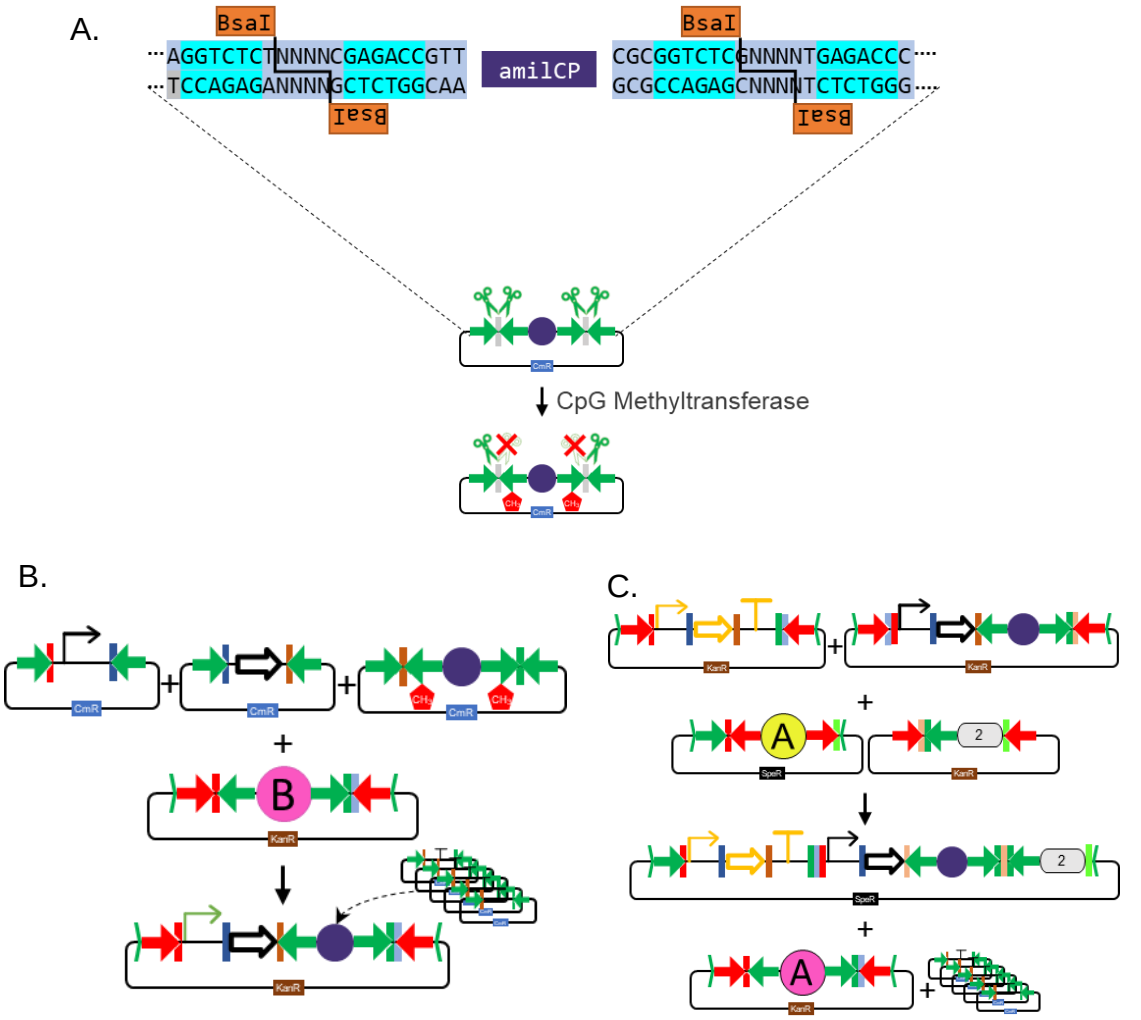


Figure 2. MethylAble feature. It is a new feature of the Mobius Assembly to facilitate the generation of combinatorial DNA libraries. **A**. MethylAble plasmid: It has amilCP gene flanked by inward and outward-facing *BsaI* recognition sites, which both cleave at the Phytobrick overhangs. The outward-facing *BsaI* sites are blocked with CpG methylation. In this example, the MethylAble plasmid replaces the terminator part, which it is fused with a promoter and a coding sequence in a Level 1 reaction (**B**). If desired, the Level 1 construct can directly receive Phytobricks. The Level 1 construct is then fused with a second TU, resulting in a 2-TU construct (**C**). This construct can be fused directly with a library of Phytobricks in a Level 1 reaction. Pink circles demarcate Level 1 and yellow circles Level 2 Acceptor Vectors, respectively; the purple circle shows the MethylAble cassette and oval numbered shape the Auxiliary Plasmid. Green and red arrows are *BsaI* and *AarI,* respectively.

			Number of constructs	
			With feature	W/out feature
Level 1 cloning				
A	**B**	**Γ**		
UBQ10:fluc:UBQ5	UBQ10:sXVE:HSP	lexA:nluc:^{Me}amilCP^{Me}	**7**	**46**
	UBQ10:LhGR:HSP	pOp6:nluc:^{Me}amilCP^{Me}		
	UBQ10:AlcA:HSP	alcSynth:nluc:^{Me}amilCP^{Me}		
Level 2 cloning				
A				
UBQ10:fluc:UBQ5-UBQ10:sXVE:HSP-lexA:nluc:^{Me}amilCP^{Me}			**3**	**42**
UBQ10:fluc:UBQ5-UBQ10:LhGR:HSP-pOp6:nluc:^{Me}amilCP^{Me}				
UBQ10:fluc:UBQ5-UBQ10:AlcA:HSP-alcSynth:nluc:^{Me}amilCP^{Me}				
Level 1 cloning				
A				
UBQ10:fluc:UBQ5-UBQ10:sXVE:HSP-lexA:nluc:Terminator$_{1-14}$			**42**	**-**
UBQ10:fluc:UBQ5-UBQ10:LhGR:HSP-pOp6:nluc: Terminator$_{1-14}$				
UBQ10:fluc:UBQ5-UBQ10:AlcA:HSP-alcSynth:nluc: Terminator$_{1-14}$				
Total			**52**	**88**

Table 1. Application of the MethylAble Feature to generate a library of three inducible systems where the promoters are combined with 14 different terminators. In Level 1 a normaliser unit was built in vector A, the three transactivator units in vector B and three inducible promoters were combined with the MethylAble Feature in vector Γ. Without the MethylAble feature, the inducible promoters should have been combined with all terminators in this step. Then in a Level 2 reaction the normaliser, the transactivators and the inducible promoters were fused to generate three constructs in total. Lastly, in a level 1 reaction, the inducible constructs were combined with the Level 0 terminators to give the final constructs. Letters represent the Acceptor Vectors.

3.2.3. Development of a new plant binary vector (pMAP)

Initially, we used pGreen based vectors for the transient transformation of Arabidopsis mesophyll protoplasts. pGreen is a small binary vector bearing the

pSa OriV for replication in Agrobacterium, colE1 origin of replication for *E. coli* and minimal synthetic Left and Right Borders of T-DNA sequences derived from the Agrobacterium plasmid pTiT37 (50). For the replication in Agrobacterium, pSoup, a helper plasmid carrying the replication protein RepA, is co-transformed with pGreen. Nevertheless, pGreen had transformation issues with large constructs and proved unsuitable for the transformation of plant cell cultures.

Consequently, we switched to pLX based vectors, which demonstrated excellent performance in the transformation of cell cultures. However, experiments which require a high amount of DNA (e.g. protoplast transformation) cannot utilise pLX vector due to the medium-copy-number origin of replication. I tried to increase the copy number of BBR1 (pLXBBR1mut) with a single mutation (65); however, the amount of plasmid DNA produced did not match the performance of ColE1 Ori, and plasmid instabilities resulted when introduced in several agrobacterial strains. Therefore, it was decided to build a new small plant binary vector based on the pLX architecture, which is suitable for protoplast transformation, the transient transformation of cell cultures, tobacco leaves, and whole-plant stable transformation.

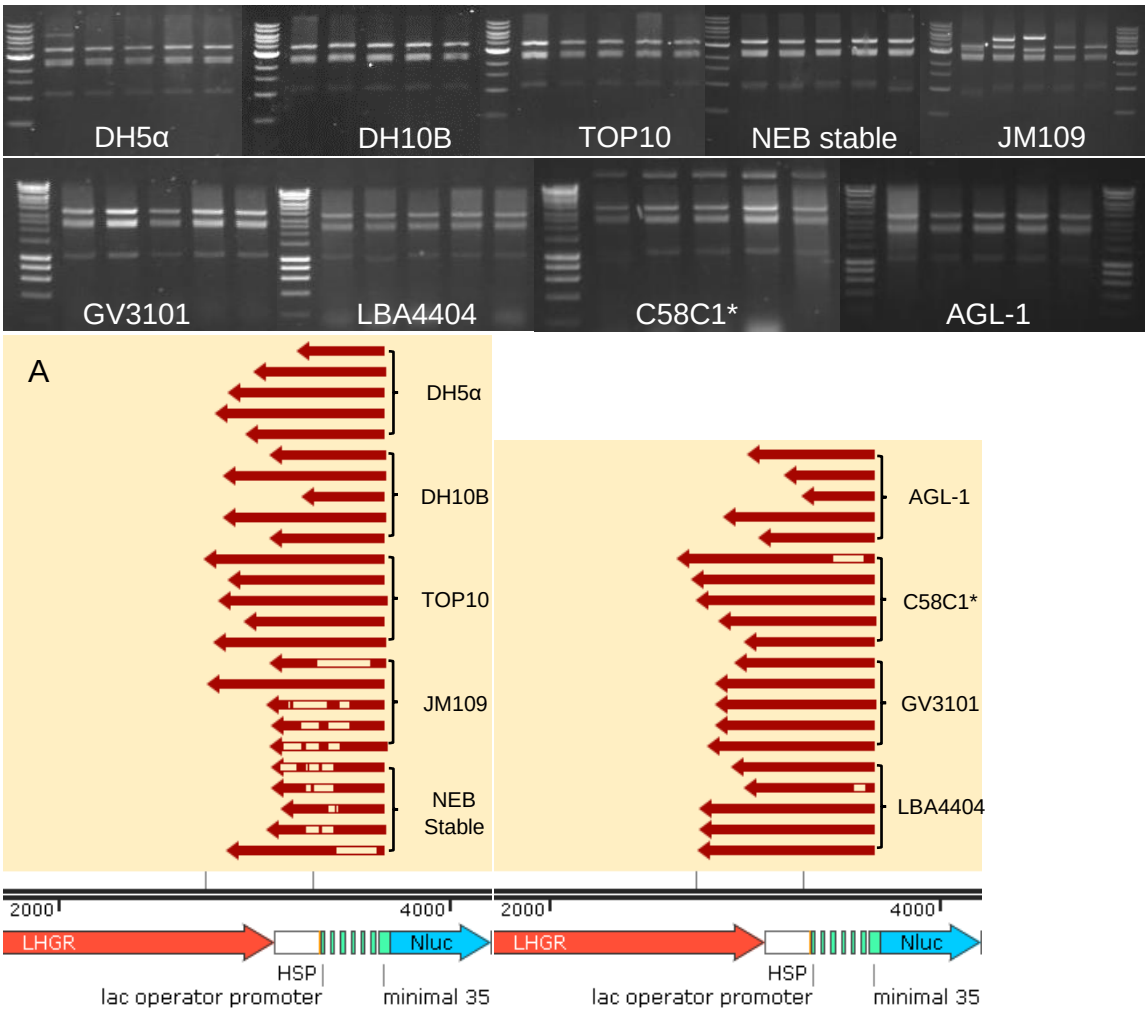
To achieve this, the BBR1 Ori of pLX vectors was replaced with a fusion of pWKS1 and pUC19 Ori in co-directional orientation. The new origin of replication is only 37.84% larger than BBR1 (1978 bp instead of 1435 bp), which is still smaller than the short version of pVS1 + colE1 in pLSU (2654 bp, (66) and pSa+colE1 in pGreen (2179 bp without the plasmid stability domains, (50). pWKS1 Ori derives from a small, multicopy and cryptic plasmid pWKS1 (2697 bp) of *Paracoccus pantotrophus* DSM 11072, which has only replication and mobility function domains (67). Cryptic plasmids usually encode proteins that are involved in plasmid replication and mobilisation, and they do not have obvious benefits to the host cells that carry them (68, 69). The replication domain consists of an origin of vegetative replication OriV (249 bp) and the gene encoding its cognate binding protein RepA (1020 bp). pWKS1 Ori was found to be functional for *Agrobacterium*

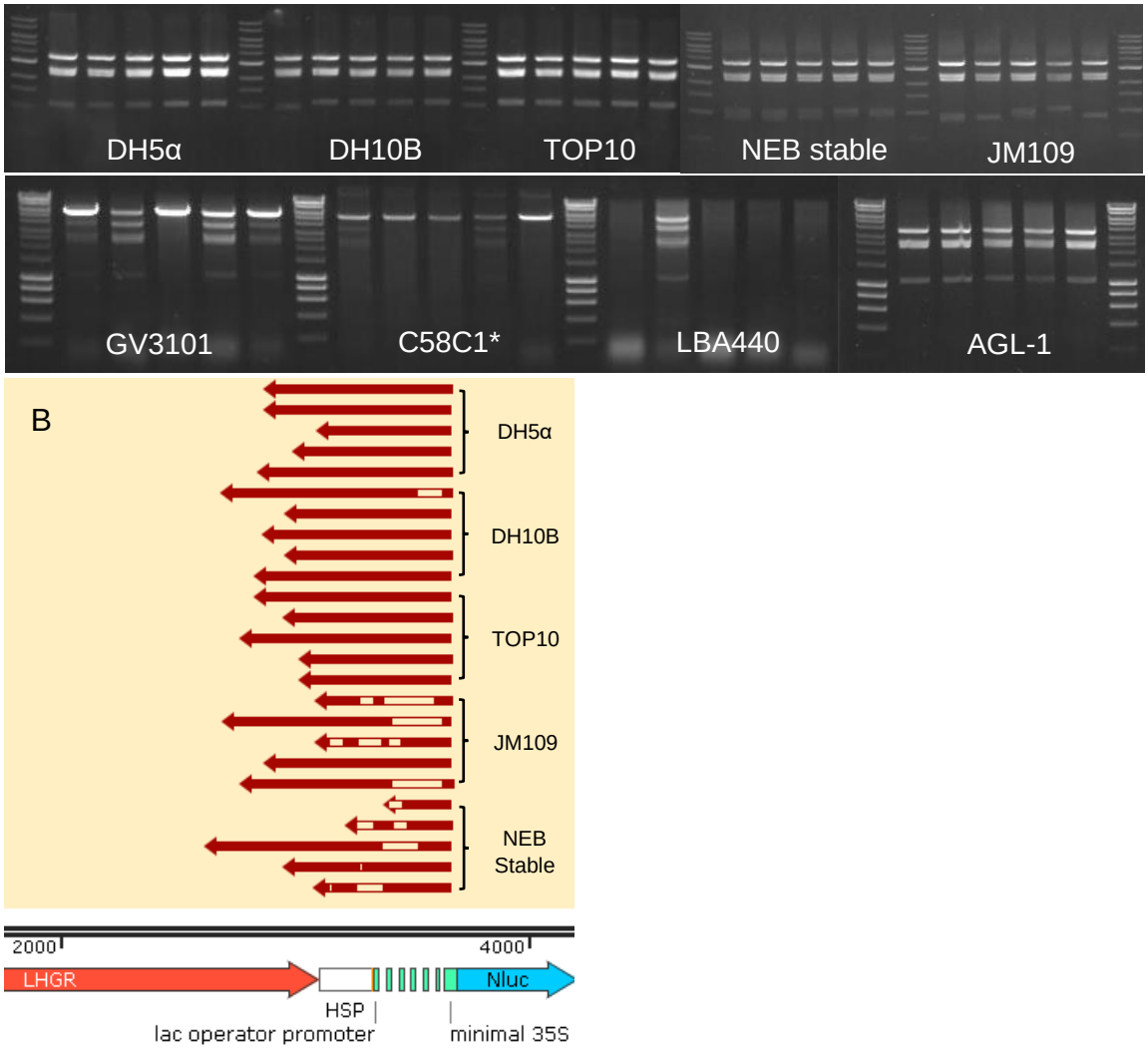
tumefaciens but not in *E. coli* (67). The minimal stable pUC Ori from pUC19 was amplified and fused in co-directional transcriptional orientation with pWKS1 Ori. The minimum sequence requirement of pUC for stable replication in *E. coli* was found to include RNA I/RNA II transcripts on 5′-side and dnaA/dnaA′ boxes on the 3′-side, while co-directional transcription of two different replicons in the same plasmid was shown to increase transformation efficiency and DNA yield (58).

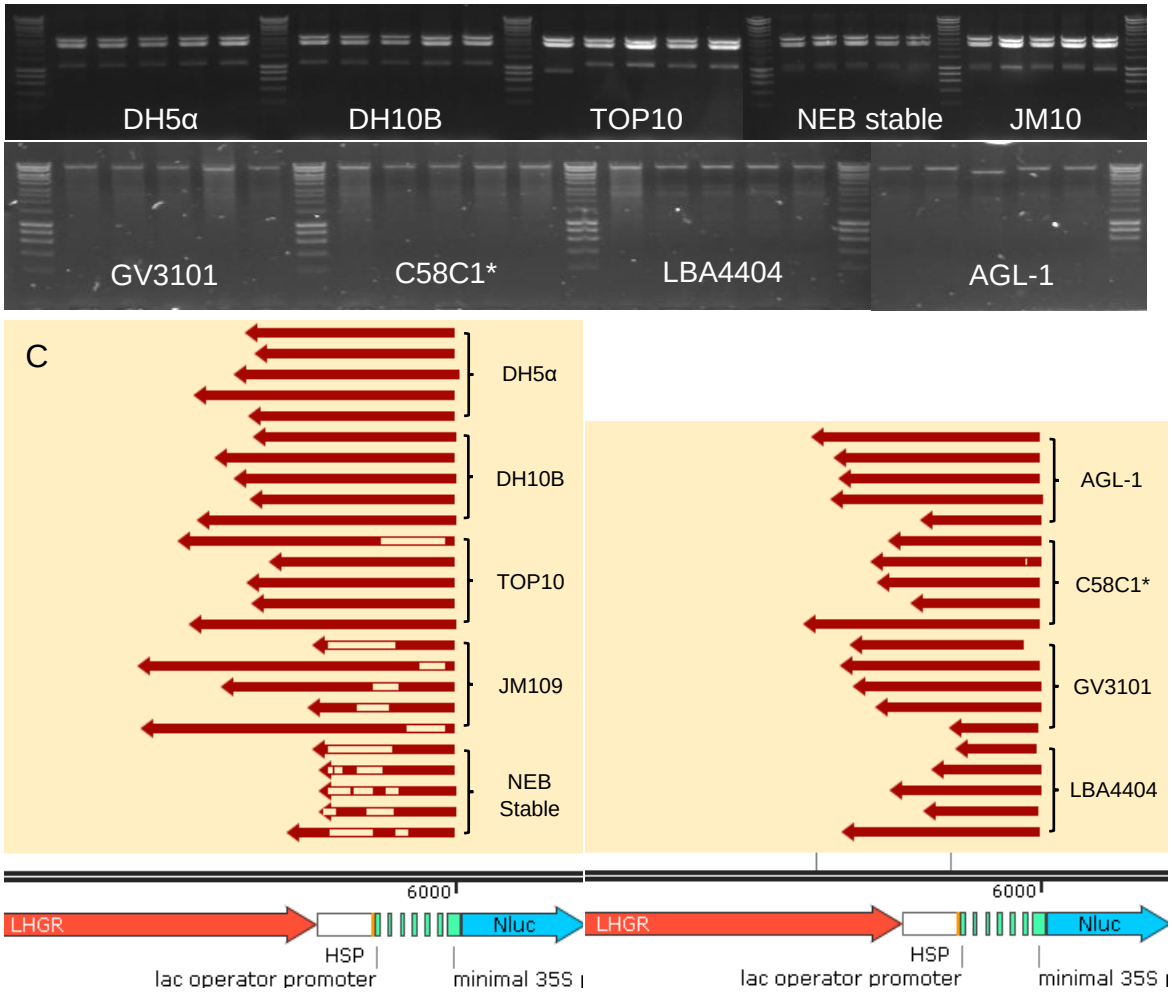
3.2.4. Structural plasmid stability

Structural plasmid stability was evaluated during propagation in *E. coli* and Agrobacterial strains. The 3-TU insert UBQ10:LhGR:HSP-pop6:nluc:HSP-UBQ10:fluc:UBQ5 was cloned either in pGreen, pLX, pLXBBR1mut or pMAP-based Mobius Assembly Vectors. pGreen based vectors exhibited good stability in DH10B (5/5), Dh5α (5/5), and TOP10 (4/5) but poor stability in JM109 (0/5) and NEB stable (0/5) strains according to the restriction digestion profiles and sequencing results (**Figure 3**). Instabilities were detected as random deletions in the pOp6 sequence. pLX based plasmids found stable in DH5α (5/5), DH10B (5/5) and TOP10 (5/5) strains; however, they also showed poor stability in NEB stable (0/0), and JM109 (0/0) strains with deletions in pOp6 domain. A similar pattern was observed for pLX BBR1mut version for *E. coli strains*, as they were stable in DH5α (5/5), DH10B (4/5) and TOP10 (5/5) strains, yet unstable in JM109 (1/4) and NEB stable (1/4) strains. Lastly, the pMAP vector was stable in DH5α (5/5), DH10B (5/5) and TOP10 (5/5) and also unstable in NEB stable (2/5) and JM109 (0/5). Regarding the Agrobacteria strains AGL1 (4/5), GV3101 (5/5), LBA4404 (5/5) and C58C1* (5/5), no major instability problems were detected in pGreen vector. pLX was also stable in Agrobacteria with AGL1 (5/5), GV3101 (5/5), LBA4404 (4/5) and C58C1* (4/5). pMAP was stable in Agrobacteria as well, having all the colonies correct (AGL1 (5/5), GV3101 (5/5), LBA4404 (5/5) and C58C1* (5/5). Nevertheless, only AGL1 strain was able to propagate stable plasmids for the pLXBBR1mut version, which has a higher plasmid copy number. Severe instability issues were observed in GV3101, LBA4404 and C58C1* strains

with aberrant restriction digestion patterns and even plasmid loss, rendering the high copy BBR1mut plasmids unsuitable for propagation in Agrobacteria.




DH5α
DH10B
TOP10
NEB stable
JM109
GV3101
C58C1*
LBA440
AGL-1
B
DH5α
DH10B
TOP10
JM109
NEB
Stable
2000
4000
LHGR
Nluc
HSP
lac operator promoter
minimal 35S


DH5α
DH10B
TOP10
NEB stable
JM10
GV3101
C58C1*
LBA4404
AGL-1
C
DH5α
DH10B
TOP10
JM109
NEB Stable
AGL-1
C58C1*
GV3101
LBA4404
6000
LHGR
Nluc
HSP
lac operator promoter
minimal 35S
6000
LHGR
Nluc
HSP
lac operator promoter
minimal 35S

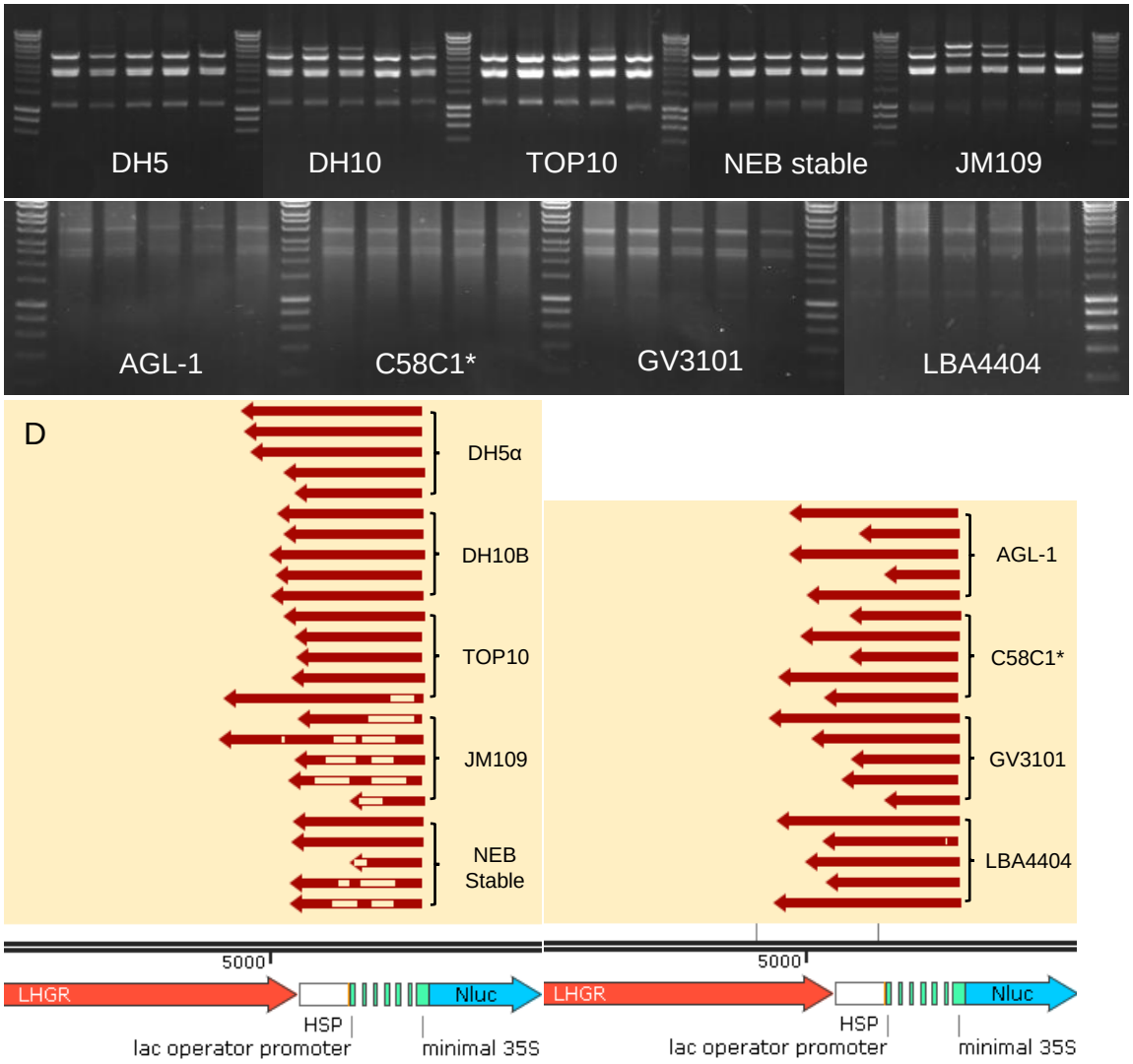


Figure 3. Structural plasmid stability. Stability was tested by electrophoresis in 1% agarose gel after digestion with *PstI*-HF and then by Sanger sequencing of the tandem short repeats of the pOp6 sequence. The stability of the construct UBQ10:LhGR:HSP-pOp6:nluc:HSP-UBQ10:fluc:UBQ5 was evaluated in both *E. coli* (DH5α, DH10B, TOP10, JM109, and NEB stable) and *A. tumefacience* strains (AGL-1, C58C1*, GV3101, and LBA4404) housed by the binary vectors: pLX (**A**), pLX-BBR1mut (**B**), pGreen (**C**) and pMAP (**D**). Alignment of the sequencing data was performed with Snapgene. The concentration of pGreen generated in Agrobacteria is low; consequently, we used high volume of plasmid DNA, but the digestion was incomplete. Low concertation of pMAP also resulted in low visibility of the second band.

3.2.5. Effect of Agrobacterium strains and binary vectors in the MM1 cell line transformation

Four widely used Agrobacterium strains (GV3101, AGL1, C58C1*, LBA4404) were evaluated for their capability to transform the rapidly dividing Arabidopsis cell suspension culture MM1 (70). We also tested the effect of the binary vector in the transformation efficiency. The evaluation was carried out by a luciferase assay using the construct NOS:BlpR:NOS-UBQ10:nluc:HSP, housed either in pGreen, pLX or pMAP based Mobius Assembly vectors. The results showed that all the Agrobacteria strains could infect the MM1 cell line (**Figure 4**). The LBA4404 strain exhibited the highest transformation efficiency giving the most top luciferase signal. Interestingly, LBA4404 infection also resulted in the formation of cellular clumps, which were not observed in the other strains. The second-best activity was found in the GV3101 and C58C1* strains. GV3101 was more efficient in transfection than C58C1*, harbouring the pMAP vector, while there was no significant difference when bearing the pLX vector. On the other hand, the AGL1 strain had the lowest transfection capacity for all the binary vectors tested. Regarding the binary vectors, pMAP performed better than pLX in GV3101 strain, leading to high luciferase values, while there was no statistical difference with C58C1*, LBA4404 and AGL1 strains. Strikingly, pGreen performed poorly in all Agrobacteria strains compared to the other vectors, and thus, it is not suitable for plant cell culture transformations.

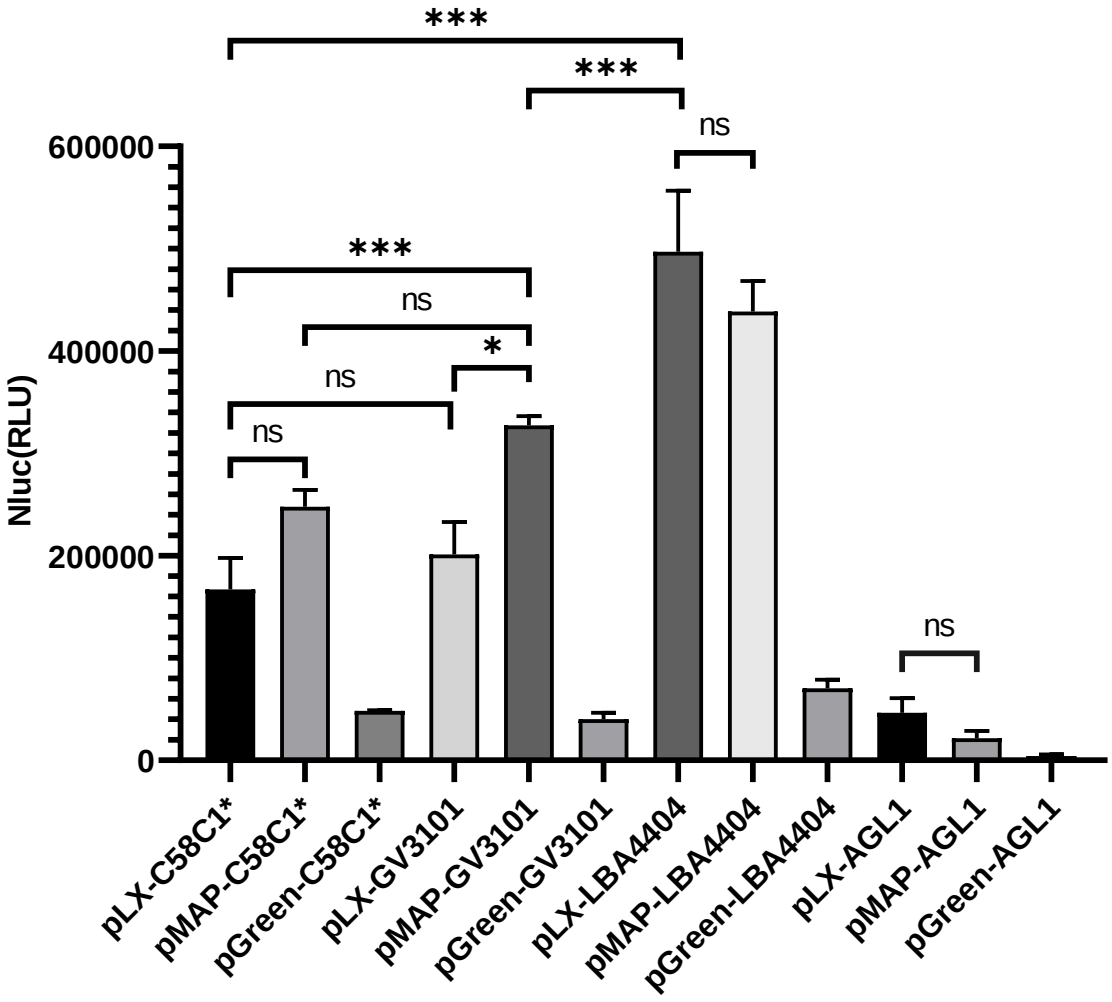


Figure 4. Effect of Agrobacterium strains and binary vectors in MM1 cell line transformation. The transformation efficiency was tested in transient Agrobacteria mediated transformation in MM1 Arabidopsis cell line using the construct NOS:BglR:NOS-UBQ10:nluc:HSP. The construct was cloned either in pLX, pGreen and pMAP Level 1 Vector and it was transformed in C58C1*, GV3101, LBA4404 and AGL-1 Agrobacterium strains. The Agrobacteria bearing the binary vectors were co-cultured with cells from the MM1 cell line in 6-well plates. After two days, Cefotaxime sodium salt was used to kill the Agrobacteria. Two days later, the cells were assayed for Nano luciferase activity in a plate reader. Nluc= Nano luciferase activity, bar graphs show luciferase activity values in mean±SE, $p < 0.05$, one-way ANOVA and Tukey's HSD test.

3.2.6. Selection of new promoters and terminators.

To select new promoter and terminator parts, we mainly opted for house-keeping genes, which are likely ubiquitously expressed (**Table 2**), so they can be used for different applications for plant research. Ubiquitously expressed genes used in qPCR were identified through literature search (71), and the most stably expressed were selected for promoter and terminator identification. LEC2 promoter was selected for expression in undifferentiated cells which cell cultures are based on (72). The first screening round included short promoters (~300bp) and short terminators (~200bp) from the genes ACT2, FAD2, TUB9, APT1, NDUFA8 and LEC2. The gene regulation level was assayed in Arabidopsis protoplasts. The new promoters were driving the expression of Nanoluciferase (NLuc) terminated by HSP, or the new terminators were placed downstream of the NLuc gene driven by the UBQ10 promoter. As the activity of the ~300 bp promoters was low, we designed new promoters derived from the genes TUB2, UBQ11, UBQ4, ACT7, as well as the more extended versions of the previous promoters (~500 kb) (**Figure 5**). The Ubiquitin promoters drove the highest expression levels, followed by TUB9 and ACT7. It is worth noting that TUB9 short and large version had similar expression levels, while APT1 and ACT2 short versions were stronger than the large versions. Short APT1 and ACT2 promoters drive the expression of a transcript variant of the genes, while the large version controls the expression of the main isoforms (**Figure 5 and Figure 6**). For TUB9, the larger version is an upstream extension of the short version, showing the absence of enhancer or suppressor elements in the extended sequence (**Figure 6**). Meanwhile, to our surprise, the selected 200bp terminators exhibited a wide range of NLuc activities, from the strongest FAD2 to the weakest LEC2 (**Figure 5)**. The promoters with higher activity and all the terminators were selected for further characterisation.

No	TAIR ID	Gene	Protein Function	Pro	Ter	Ref.
1	AT3G18780	**ACT2** (Actin 2)	Component of cell cytoskeleton	+	+	This study
2	AT3G12120	**FAD2** (Fatty Acid Desaturase 2)	Part of Lipid metabolism	+	+	This study
3	AT1G27450	**APT1** (Adenine PhosphoribosylTransferase 1)	Contributes to the recycling of adenine	+	+	This study
4	AT5G18800	**NDUFA8** (NADH:Ubiquinone oxidoreductase subunit A8)	Mitochondrial electron transport	+	+	This study
5	AT4G20890	**TUB9** (Tubulin 9)	A major constituent of microtubules	+	+	This study
6	AT5G62690	**TUB2** (Tubulin 2)	A major constituent of microtubules	+	-	This study
7	AT4G05050	**UBQ11** (Ubiquitin 11)	Ubiquitin-dependent protein catabolic process	+	-	This study
8	AT5G20620	**UBQ4** (Ubiquitin 4)	Ubiquitin-dependent protein catabolic process	+	-	This study
9	AT1G28300	**LEC2** (Leafy Cotyledon 1)	Embryo development	+	+	This study
10	AT5G09810	**ACT7** (Actin 7)	Component of cell cytoskeleton	+	-	This study
11	AT4G05320	**UBQ10** (Ubiquitin 10)	Ubiquitin-dependent protein catabolic process	+	-	(73)
12	AT3G16640	**TCTP** (Translationally Controlled Tumor Protein)	Guanine nucleotide exchange factor in the TOR signalling pathway	+	-	(74)

13	AT5G38420	**rbcS-2b** (Rubisco Small Subunit - 2b)	Member of the Rubisco small subunit	-	+	(75) Shorter versions in this study
14	AT5G59720	**HSP** (Heat shock protein)	Response to heat shock	-	+	
15	AT5G59720	**UBQ5** (Ubiquitin 5)	Ubiquitin-dependent protein catabolic process	-	+	
16	Agrobacterium gene	**NOS** (Nopaline synthase)	Nopaline synthesis	+	+	(76)
17	Agrobacterium gene	**MAS** (Mannopine synthase)	Mannopine Synthesis	+	+	
18	Agrobacterium gene	**OCS** (Octopine synthase)	Octopine synthesis	+	-	
19	Agrobacterium gene	**G7** (Gene 7)	Unknown function	-	+	
20	Cauliflower mosaic virus gene	**35S**	Function in genome replication	+	+	(77)
21	Synthetic	**G10-90**	10 G-box sequences fused to CaMV – 90/35S promoter	+	-	(78)
22	Synthetic	**lexA-35S**	lexA operator repeats fused to CaMV 35S core promoter	+	-	(79)
23	Synthetic	**pOp6-35S**	lacI operator repeats fused to CaMV 35S core promoter	+	-	(80)
24	Synthetic	**alcSynth-34S**	alcR binding sites fused to FMV 34S core promoter	+	-	(54)

Table 2. 1-10: Genes from which we amplify promoter or/and terminators for likely constitutive plant expression. We selected house-keeping genes, which tend to be ubiquitously expressed, and they are being used as reference genes in qPCR. LEC2 is being expressed in the developing embryo.11-23: Promoters and terminators isolated in other studies and used for characterisation in this work.

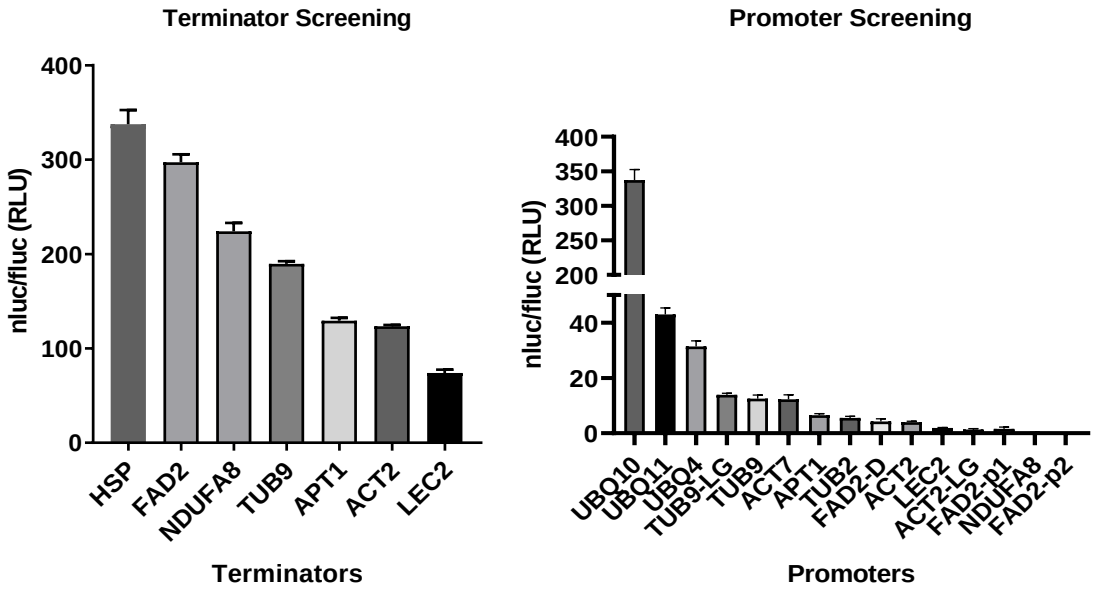


Figure 5. Characterisation of new promoters and terminators. Promoters and terminators were amplified from ubiquitously expressed housekeeping genes. For some of the promoters, there were short (~300bp) and long (~500bp, LG) versions. For FAD2, three promoters were selected p1, p2 and their fusion, FAD2-D. The expression level was assayed in Arabidopsis protoplasts. Nluc= Nano and Fluc=Firefly luciferase activity, bar graphs show luciferase activity values in mean±SE. All terminators are combined with UBQ10 promoter and all promoters with HSP terminator.

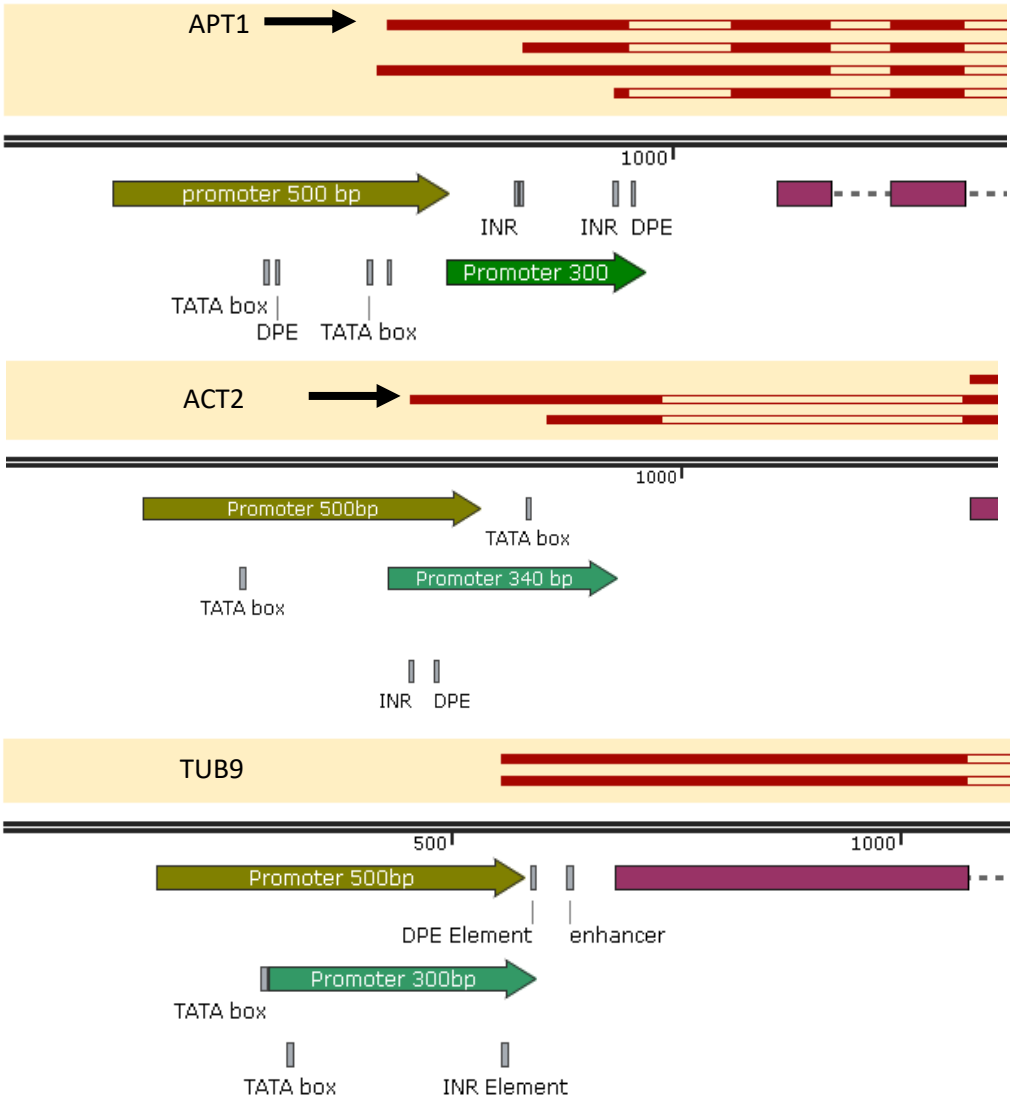


Figure 6. Two different versions of promoters from APT1, ACT2 and TUB9 genes. The black arrow shows the main isoform of the produced mRNA of the gene. INR=initiator elements, DPE= Downstream promoter element. cDNA isoforms were retrieved from TAIR (https://www.arabidopsis.org) and aligned to the genes using SnapGene software.

3.2.7. Characterisation of promoters and terminators

The library of the short promoters and terminators was characterised with the transient transformation of Arabidopsis mesophyll protoplast in 96-well plates. The Nano luciferase signal was normalised to a construct bearing the firefly luciferase flanked by UBQ10 promoter and UBQ5 terminator, to minimise the variation of the expression.

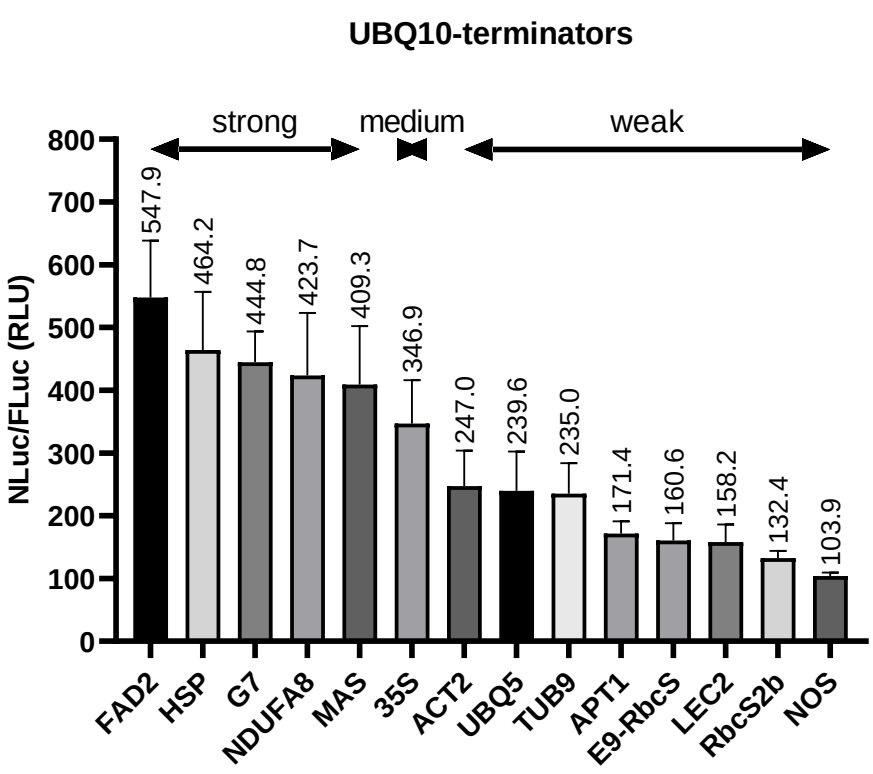
For the evaluation of the 14 terminators, the NLuc expression was driven either by the strong UBQ10 or the weak NDUFA8 promoter, while each terminator controlled the transcription termination. The different terminators resulted in a wide range of expression spanning 5.3 and 6.3 orders of magnitude for UBQ10 and NDUFA8 promoters, respectively (**Table 2**, **Figure 7A and B**). For the strong UBQ10 promoter, FAD2 terminator resulted in the highest activity with 547.9 RLU and NOS in the lowest with 103.9 RLU. For the weak NDUFA8 promoter, HSP terminator led to the highest expression with 0.327 RLU and APT1 to the lowest at 0.052 RLU. Notably, an interaction between promoters and terminators was observed, with the clear example of NOS terminator, which in combination with UBQ10 falls in the category of "weak", while in conjunction with NDUFA is categorised as "medium" to "strong" terminator. Two of the terminators isolated in this study (FAD2 and NDUFA8) were found to positively influence the expression, while four (APT1, LEC2, TUB9, ACT2) decreased the expression.

For the evaluation of the promoters, the NLuc transcription was terminated either by NDUFA8 or HSP terminator, while 17 different promoters drove the transcription initiation. Among the promoters, UBQ10 exhibited by far the highest expression activity, followed by MAS, when combined either with HSP or NDUFA8 terminator (**Table 2**, **Figure 7C and D**). With the exemption of the TUB9 promoter, HSP terminator resulted in more elevated expression than NDUFA8 to the various promoters. Two of the new isolated promoters, UBQ11 and UBQ4, were found to have a comparable or even better expression from the commonly used 35S and OCS promoters. UBQ11 had an expression of 68.3 RLU with HSP and 21.8 RLU with NDUFA8, and UBQ4 showed an expression of 66 RLU fused to HSP and 16 RLU combined with NDUFA8. The corresponding values of the 35S promoter were 41.7 RLU and 20.7 RLU, in combination with HSP and NDUFA8 terminators, respectively. Furthermore, the newly isolated promoters ACT7, TUB2, TUB9, APT1, ACT2 and LEC2 drove stronger expression from the NOS promoter. Lastly, the lowest expression was found for FAD2 and NDUFA8 promoters.

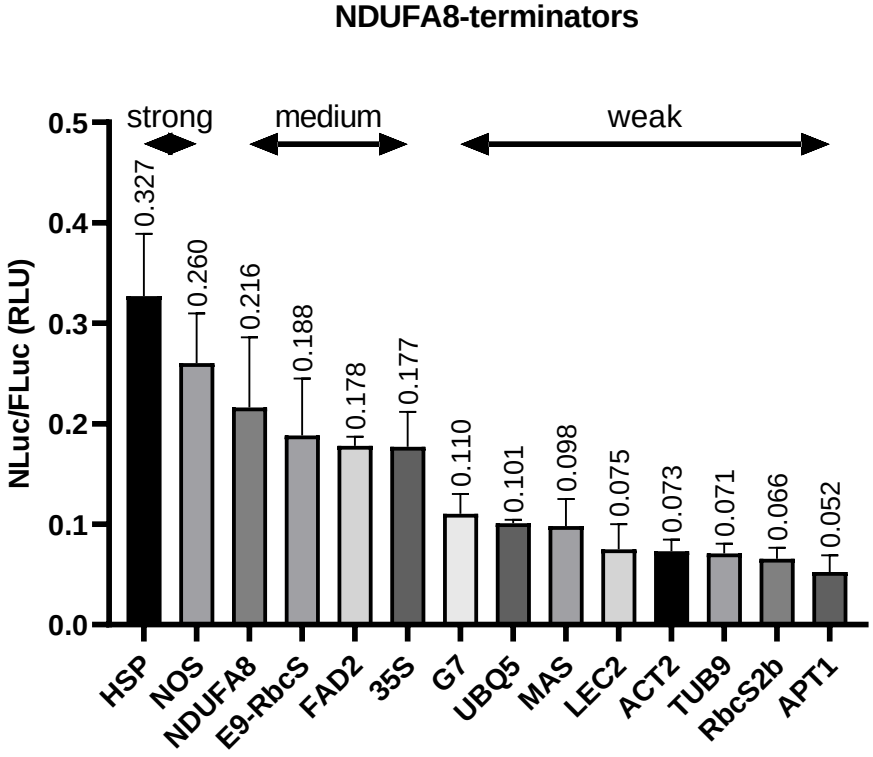
	TERM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PROM		lexA	pOp6	alcSynth	UBQ10	NDUFA8	35S	$G_{10\text{-}90}$	ACT2	APT1	FAD2	ACT7	TCTP	TUB2	TUB9	UBQ4	UBQ11	LEC2	NOS	MAS	OCS
1	G7	85.46	24.3	187	444.8	0.110															
2	NOS	64.7	12.6	161.8	103.9	0.260															
3	ACT2	57.4	17.3	142.6	247	0.073															
4	APT1	83.4	11.8	132.2	171.4	0.052															
5	35S	111.6	25.1	157.6	346.9	0.177															
7	LEC2	50	16.2	46.7	158.2	0.075															
8	TUB9	106.4	23	92.2	235	0.071															
9	MAS	71.7	18.3	104.5	409.3	0.098															
10	E9RbcS	48.7	11.6	79.7	160.6	0.188															
11	RbcS2b	55.9	13.3	56.7	132.4	0.066															
12	HSP	79.1	23	337.9	464.2	0.327	41.7	33.7	4.6	6.6	0.4	13	44.4	7.1	6.7	66	68.3	2.2	1.4	199	53.8
13	UBQ5	37.9	20.4	48.1	239.6	0.101															
14	NDUFA8	61.1	17.7	135.7	423.7	0.216	20.7	13.3	1.4	4.9	0.02	5.6	13.7	1.4	7.4	16	21.8	0.8	0.6	133.3	16
15	FAD2	80.8	21.5	223.4	547.9	0.178															

Table 2. High throughput characterisation of Arabidopsis promoters and terminators. Heat map showing how the combination of different promoters and terminators affects the expression levels of nano luciferase normalised by firefly luciferase. Green colour indicates high activity, while red low. Colour scale is applied independently for each promoter.

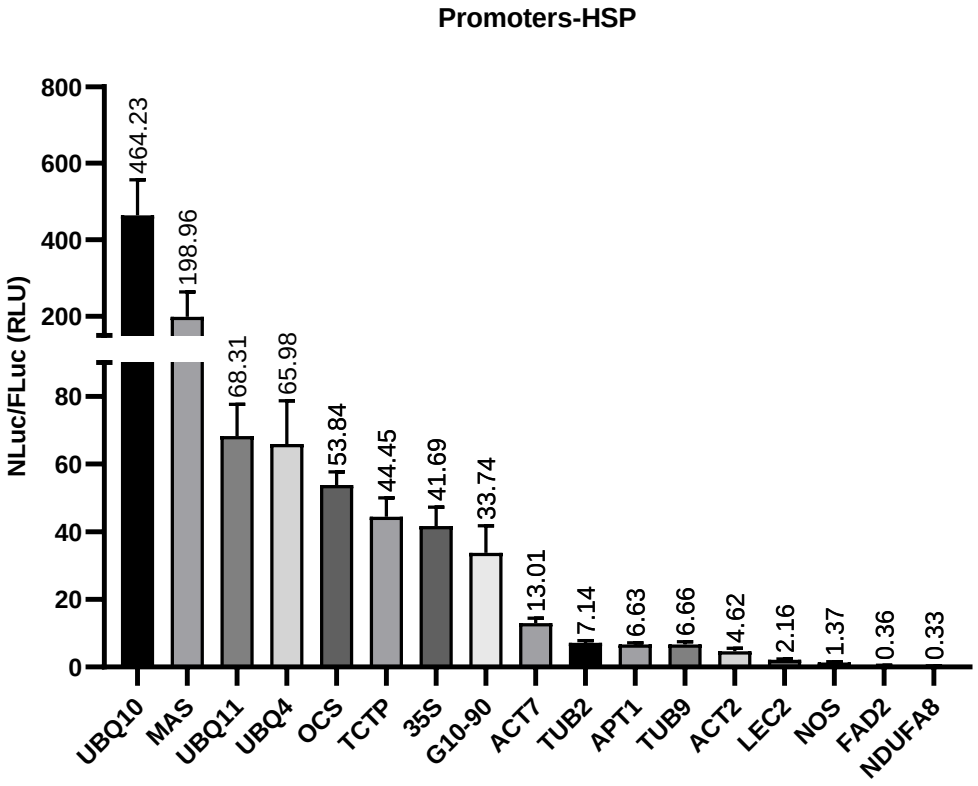
A


UBQ10-terminators
strong
medium
weak
NLuc/FLuc (RLU)
800
700
600
500
400
300
200
100
0
547.9
464.2
444.8
423.7
409.3
346.9
247.0
239.6
235.0
171.4
160.6
158.2
132.4
103.9
FAD2
HSP
G7
NDUFA8
MAS
35S
ACT2
UBQ5
TUB9
APT1
E9-RbcS
LEC2
RbcS2b
NOS

B


NDUFA8-terminators
strong
medium
weak
NLuc/FLuc (RLU)
0.5
0.4
0.3
0.2
0.1
0.0
0.327
0.260
0.216
0.188
0.178
0.177
0.110
0.101
0.098
0.075
0.073
0.071
0.066
0.052
HSP
NOS
NDUFA8
E9-RbcS
FAD2
35S
G7
UBQ5
MAS
LEC2
ACT2
TUB9
RbcS2b
APT1

C


Promoters-HSP
NLuc/FLuc (RLU)
800
600
400
200
80
60
40
20
0
464.23
198.96
68.31
65.98
53.84
44.45
41.69
33.74
13.01
7.14
6.63
6.66
4.62
2.16
1.37
0.36
0.33
UBQ10
MAS
UBQ11
UBQ4
OCS
TCTP
35S
G10-90
ACT7
TUB2
APT1
TUB9
ACT2
LEC2
NOS
FAD2
NDUFA8

D

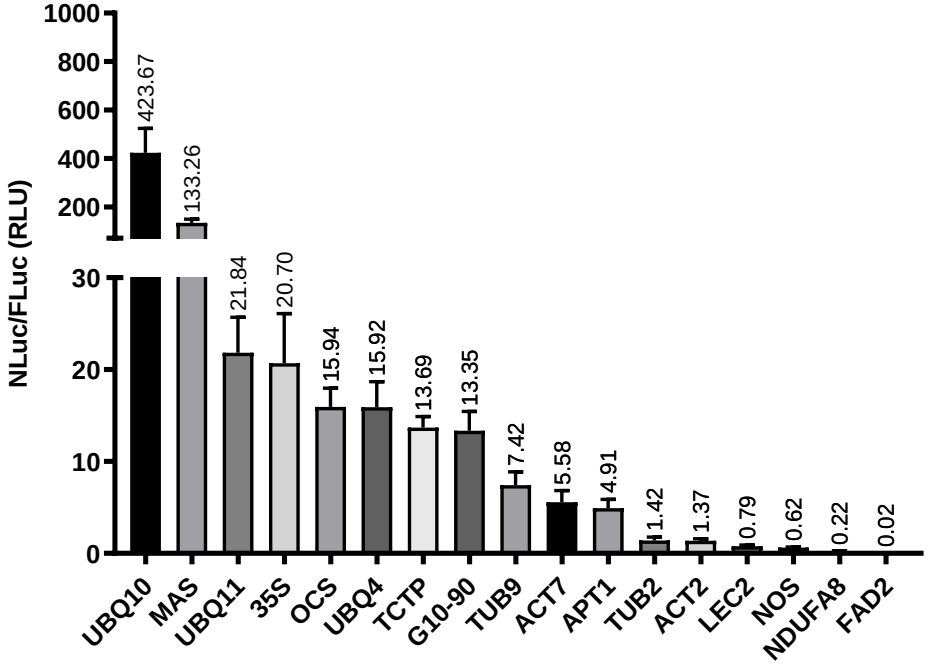
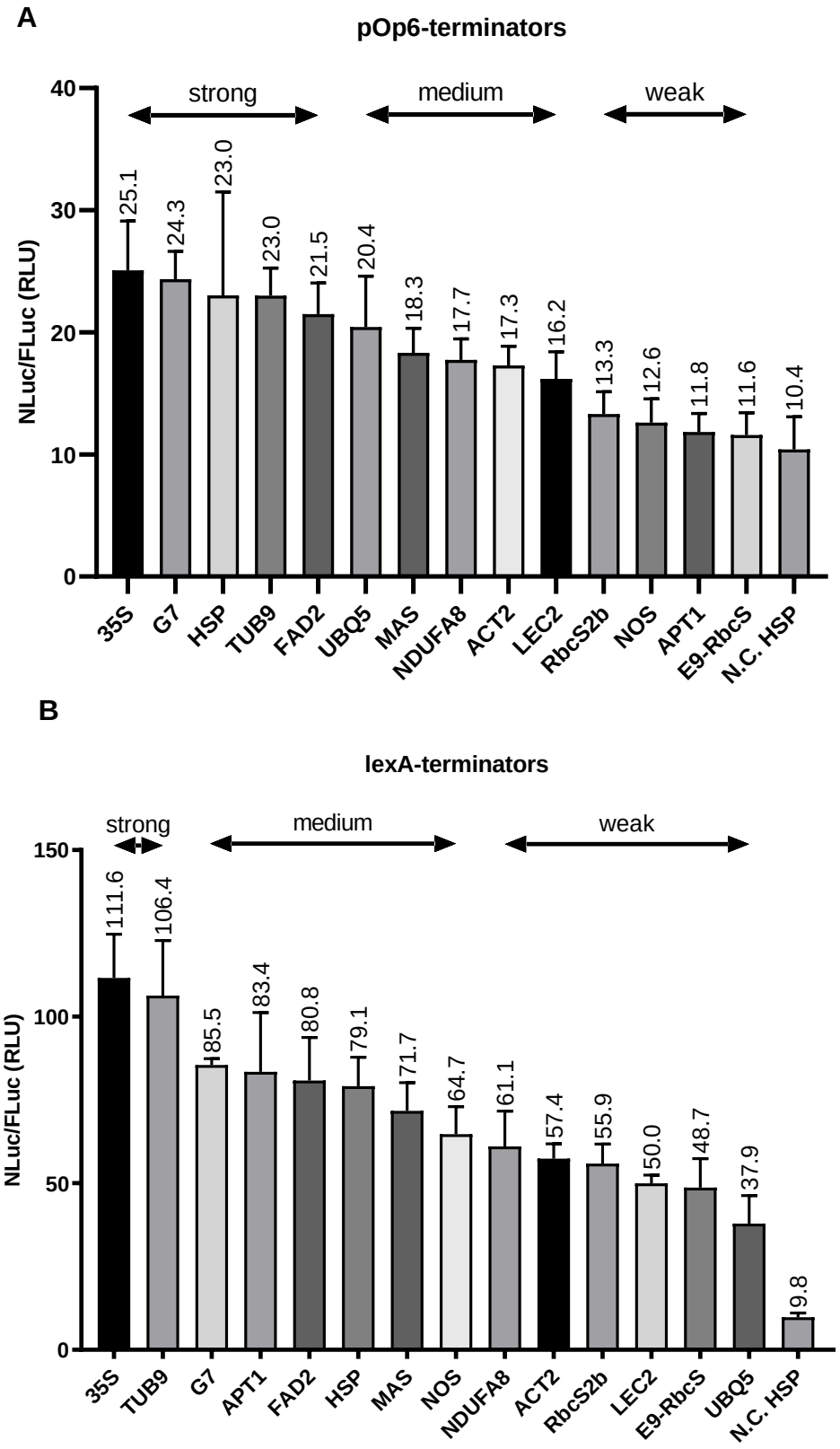

Promoters-NDUFA8
NLuc/FLuc (RLU)
1000
800
600
400
200
30
20
10
0
423.67
133.26
21.84
20.70
15.94
15.92
13.69
13.35
7.42
5.58
4.91
1.42
1.37
0.79
0.62
0.22
0.02
UBQ10
MAS
UBQ11
35S
OCS
UBQ4
TCTP
G10-90
TUB9
ACT7
APT1
TUB2
ACT2
LEC2
NOS
NDUFA8
FAD2

Figure 7. High throughput characterisation of Arabidopsis promoters and terminators. The characterisation was carried out with PEG-transformation of leaf mesophyll protoplasts from Arabidopsis in 96-well plates. The expression strength of the promoters and terminators was assayed in 96-well plates, in a plate reader measuring the Nano luciferase activity normalised by Firefly luciferase. Terminators were characterised combined either with UBQ10 (A), or NDUFA8 (B) promoters, while promoters were characterised either with HSP (C) or NDUFA8 (D) terminators. The construct for the promoter study was Promoter:nluc:HSP/NDUFA8:UBQ10:fluc:UBQ5 and for terminators, UBQ10/NDUFA8:nluc-Terminators:UBQ10:fluc:UBQ5 housed by pGreen based vector. Terminators were distributed in three arbitrary categories, by tripartition of their expression range: for UBQ10, Weak <251.9 RLU, medium =251.9-399.9 RLU and Strong >399.9 and for NDUFA8 Weak <0.143, medium 0.143-0.235 and strong >0.235. Nluc= Nano and Fluc=Firefly luciferase activity, bar graphs show luciferase activity values in mean±SE.

The inducible promoters also tested with the library of 14 terminators (**Table 2** and **Figure 8**). For pOp6-35S and lexA-35S, there is a uniform and gradual decrease of the luciferase expression spanning in both cases at around 2.2 and 2.9 orders of magnitude, respectively. For lexA-35S, the expression is spread between 37.9 and 111.6 RLU, when combined with UBQ5 and 35S terminators, respectively; and for pOp6-35S the expression ranges between 11.6 and 25.1 RLU with E9-RbcS and 35S terminators. In contrast, the alcSynth-34S promoter has shown an array of expression activity similar to the constitutive promoters. Notably, the expression with HSP terminator (337.9 RLU) is 7.2 times higher than the expression that LEC2 terminator provides (46.7 RLU).

For both the terminator series in conjunction with pOp6-35S and lexA-35S, the interactions between the promoters and the terminators appeared negligible, as they follow a specific pattern, with the exemption of UBQ5 and APT1. On the other hand, AlcSynth-34S does not follow this pattern. While pOp6-35S and lexA-35S have a basal expression of around 10 RLU (when combined with HSP), lexA-35S has demonstrated ~11-fold activation amplitude (111.6 RLU), while pOp6-35S had only three times activation amplitude (25.1 RLU). AlcSynth basal expression combined with HSP terminator was very high (250.3 RLU) leading to only 1.3 times of activation (337.9 RLU) upon treatment with ethanol.


A
pOp6-terminators
strong
medium
weak
NLuc/FLuc (RLU)
40
30
20
10
0
25.1
24.3
23.0
23.0
21.5
20.4
18.3
17.7
17.3
16.2
13.3
12.6
11.8
11.6
10.4
35S
G7
HSP
TUB9
FAD2
UBQ5
MAS
NDUFA8
ACT2
LEC2
RbcS2b
NOS
APT1
E9-RbcS
N.C. HSP
B
lexA-terminators
strong
medium
weak
NLuc/FLuc (RLU)
150
100
50
0
111.6
106.4
85.5
83.4
80.8
79.1
71.7
64.7
61.1
57.4
55.9
50.0
48.7
37.9
9.8
35S
TUB9
G7
APT1
FAD2
HSP
MAS
NOS
NDUFA8
ACT2
RbcS2b
LEC2
E9-RbcS
UBQ5
N.C. HSP

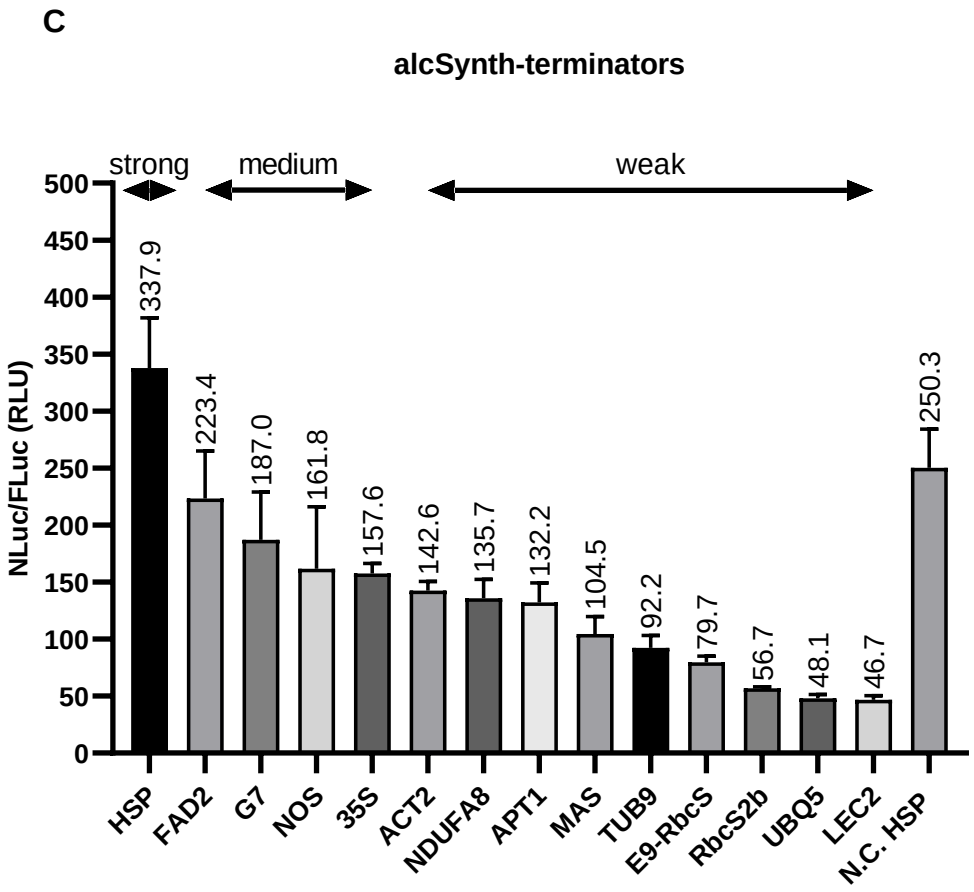


Figure 8. High throughput characterisation of three plant inducible promoters with the library of terminators. The characterisation was carried out with PEG-transformation of leaf mesophyll protoplasts from Arabidopsis in 96-well plates. The expression strength was assayed in 96-well plates, in a plate reader measuring the Nano luciferase activity normalised by Firefly luciferase. Terminators were characterised combined with either pOp6-35S (A), or lexA-35S (B), or alcSynth-34S (C) inducible promoters. The construct we used was UBQ10:fluc:UBQ5- UBQ10:Transactivator:HSP-Promoter:nluc:HSP in pGreen based vector, where Transactivator corresponds to LhGR (A), sXVE (B) and AlcR (C) and Promoter to pOp6-35S (A), lexA-35S (B) and alcSytn-34S (C). Terminators were distributed in three arbitrary categories, by tripartition of their expression range: for pOp6-35S, Weak <16.1 RLU, medium =16.1-20.4 RLU and Strong >20.4 RLU, for lexA-35S Weak <62.5 RLU, medium 62.5-87 RLU and strong >87 RLU and for alcSynth-34S, Weak <143.8 RLU medium =143.8-240.8 RLU and Strong >240.8 RLU. Chemical induction was achieved with 2.5μM DEX for (A), 5μM β-estradiol for (B) and 0.1% EtOH for (C). Nluc=Nano and Fluc=Firefly luciferase activity, bars show normalised luciferase activity values in mean±SE. N.C=negative control, uninduced cells.

3.2.8. Terminator functional dissections

To investigate the function of terminators, two strong terminators were selected, FAD2 and HSP, and dissected putative functional sequences by creating a deletion series. In total, six different versions were generated for each terminator. For the FAD2 terminator, Sequence 1 is the full length (200bp) terminator, while Sequence 2 removed the second Poly-A signal and a putative destabilisation signal (PDS) (**Figure 9A**). Sequence 3 resulted from an additional deletion of a PDS and a Musashi binding element (MBE). Sequence 4 removed another PDS from Sequence 3. Further shortening of the terminator by the removal of two more PDS gave rise to Sequence 5, and finally, elimination of the first Poly-A resulted in Sequence 6. Concerning the HSP terminator, the first dissection removed two overlapping putative stabilisation signals (PSS) and two overlapping PDS (**Figure 9B**). The second deletion of a 73bp segment removed the second Poly-A signal and a PSS to generate the second part. The HSP Sequence 3 had a mutation in the Poly-A site converting AATAAA to AAgcAA and removal of a segment with a YGTGTTYY motif and two overlappings PDS created part four. Finally, the last part results from the deletion of the Poly-A signal and two PSS.

Terminator activity was evaluated transiently in Arabidopsis mesophyll protoplasts and the MM1 cell line. Statistical analysis of the results showed only a significant difference between FAD2 part five, which is the shortest part with no Poly-A site, and the intact FAD2 as well as FAD2 part 1 (**Figure 10)**. Concerning the HSP series in Arabidopsis protoplasts and the MM1 cell line, as well the FAD2 series in the MM1 cell line there was no considerable difference among different dissected parts.

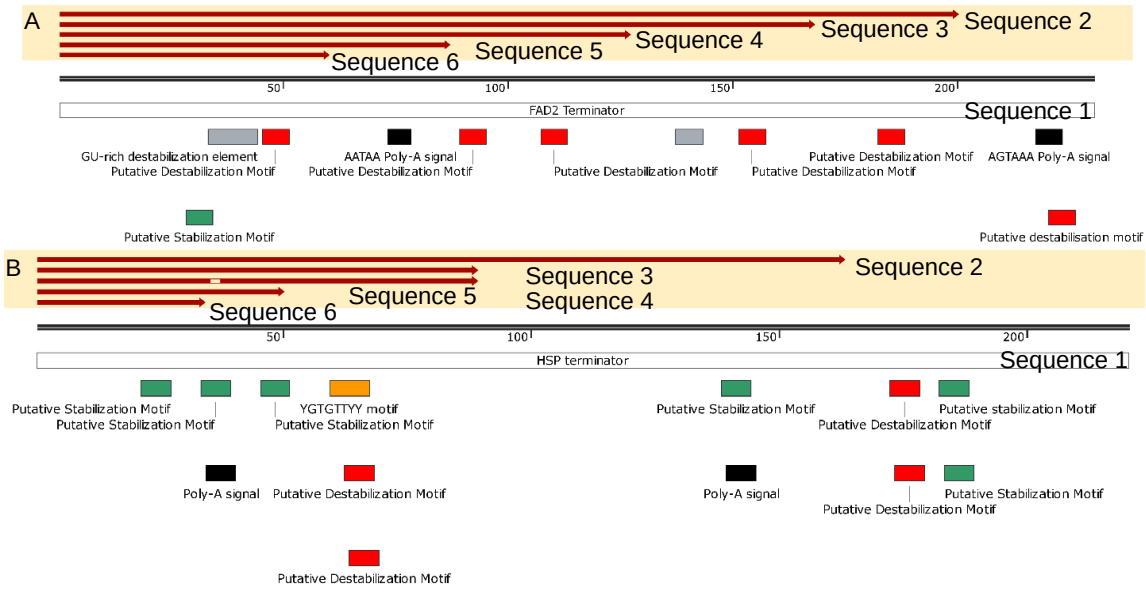
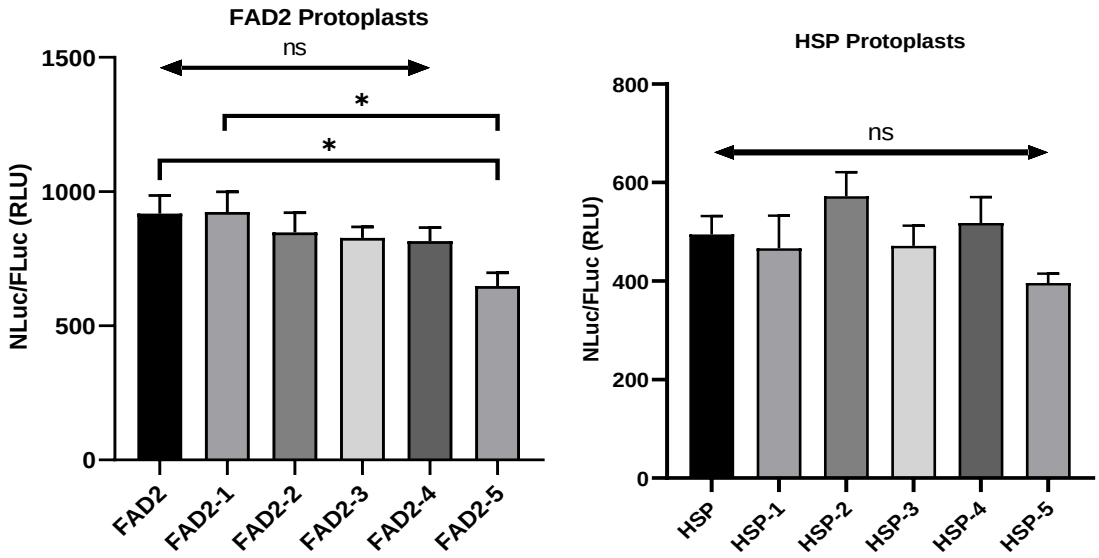


Figure 9. Deletion series of HSP and FAD2 terminators for functional dissections and expression activity. Five versions of HSP and FAD2 terminators were constructed by dissections of their full-length sequences. The dissections removed putative motifs that could affect their functions. Putative Stabilization/Destabilization Motifs were retrieved form Narsai et al. 2007, while the rest were identified using RegRNA2.0 and PASPA online tools.



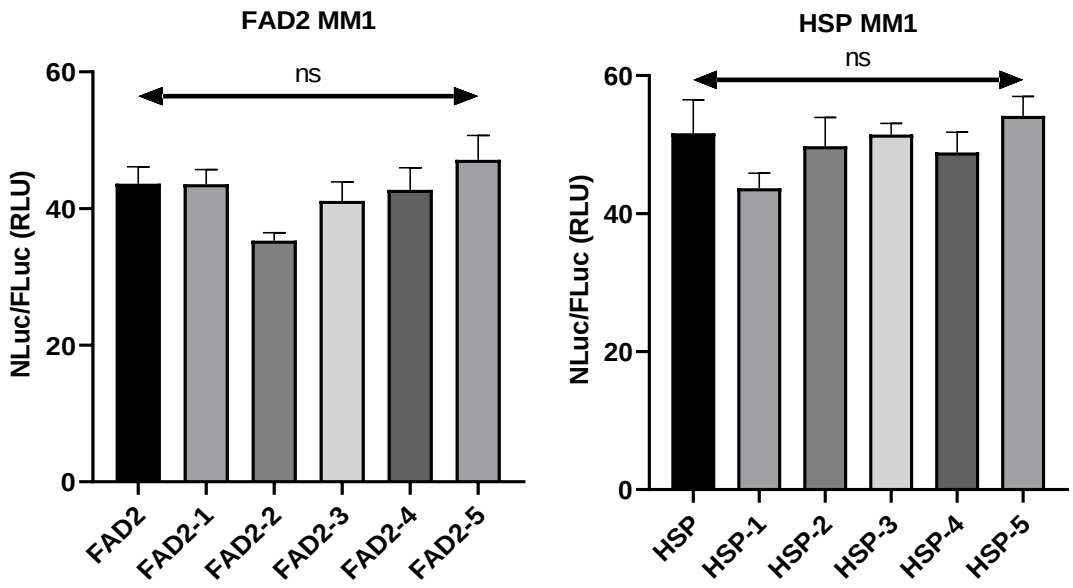


Figure 10. Expression activity of HSP and FAD2 terminator deletion series in Arabidopsis protoplast and MM1 cell line. Transformation of the construct was performed either with PEG-transformation of leaf mesophyll protoplasts from Arabidopsis in 96-well plates, or Agrobacteria mediated transformation of the MM1 cell line in 12-well plates. The expression activity of the constructs was assayed using a plate reader, in 96-well plates, measuring the Nano luciferase activity normalised by Firefly luciferase. The construct we used was UBQ10:nluc:Terminator Part:UBQ10:fluc:UBQ5, housed in a pMAP vector. Nluc= Nano and Fluc=Firefly luciferase activity, bar graphs show luciferase activity values in mean±SE. $p < 0.05$, one-way ANOVA and Tukey's HSD test.

3.3. Discussion

3.3.1. MAPS: A new platform to engineer plant systems

MAPS is a fast and user-friendly DNA assembly platform based on plant binary vectors, which comes with a characterised part toolkit for plant expression, as well as a new feature to facilitate combinatorial DNA library assemblies. The new feature – MethyAble – exploits *in vitro* DNA methylation to directly feed Phytobricks in the Level 2 constructs.

Initially, we used pGreen as a backbone to have a small and high copy cloning vector, which was used for protoplast transformations. However, pGreen had some transformation issues in *E. coli* (TOP10 cells) with large constructs and

proved not suitable for transformation of plant cell cultures. Consequently, we switched to the newly developed pLX vectors, which worked efficiently in plant cell culture transformations. Nevertheless, for some experiments that demand both high DNA yield (e.g. protoplast transformation) and transformation efficiency (e.g. cell cultures), neither of the backbones are suitable. For this reason, a new small binary vector (pMAP) was devised based on the pLX architecture using a fused replication of origin of pUC with pWKS1. This vector was highly efficient for cell culture transformations and was used to characterise dissected terminator parts in both Arabidopsis protoplasts and the MM1 cell culture.

3.3.2. The effect of vectors and bacterial strains on plasmid stability

Our results indicate that structural plasmid stability is an essential factor to consider when working with constructs susceptible to instabilities. The insert employed to test structural plasmid stability, (UBQ10:LhGR:HSP-pOp6:nluc:HSP-UBQ10:fluc:UBQ5) is prone to instabilities, since it is large (~6.7 kb) and has two direct repeats (two UBQ10 promoters and two HSP terminators), and the pOp6 promoter constitutes of six inverted repeats of the lac operator and five inverted spacer repeats. Most of the observed instabilities were deletions in the pOp6 sequence. Structural plasmid analysis revealed that the stability was strain-dependent; The constructs housed by pGreen, pLX and pMAP vectors were generally stable in DH10B, DH5a, TOP10 strains but unstable in JM109, and NEB stable strains.

Interestingly, even though all the *E. coli* strains we used are recA1^{-} to minimise recombination, they do not perform the same in terms of plasmid stability. Our results demonstrated that the bacterial strains are of crucial importance for construct stability. Similar conclusions were drawn by Moore *et al*. when they switched to JM109 from DH10B to deal with instability issues in the engineering of violacein pathway (56). All the vectors mentioned above were generally stable in the Agrobacteria strains (AGL1, GV3101, LBA4404, C58C1); however, the domains vulnerable to instabilities should be sequenced before any application.

Increasing the copy number of pLX based vectors by mutating the BBR1 origin of replication maintained the stability of the vectors in the same *E. coli* strains, which are recA$^-$, but abolished it in Agrobacteria strains, except AGL1 (**Figure 3B**). AGL1 is recA$^-$ strain, which stabilises the recombinant plasmids, hence justifying the results (81). Additionally, all the Agrobacteria strains had a slower growth rate on agar plates (3-4 instead of 2-3 days). Even though BBR1mut was tested before in C58 Agrobacterium strain for the production of β-carotene (82), we do not consider a suitable origin of replication for binary vectors because it could not stably maintain large and complex constructs in commonly used Agrobacteria strains. Overall, a selection of an origin of replication with a bacterium strain that favour plasmid stability should be considered, before transforming a construct into plants or plant cells,

3.3.3. The effect of binary vectors and Agrobacteria strains on plant cell culture transformation

Most of the *Agrobacterium* strains commonly used in plant biotechnology are derived from two wild isolates, C58 and Ach5 (83), and they differ in their Ti plasmids. C58 contains two plasmids, the nopaline type pTiC58 and a cryptic one pAtC58, which assists the virulence (84). Temperature-induced loss of pTiC58 plasmid generated C58C1 strain (85). Introduction of the disarmed pTiC58 plasmid (pMP90) led to the generation of GV3101:pMP90 (86). AGL1 also has the C58C1 background and has been engineered to carry a supervirulent, disarmed succinoamopine-type plasmid, pTiBo542, from the *Agrobacterium* strain A281 (81). C58C1(pTiB6S3ΔT)^H derives again from C58C1 strain and contains a disarmed octopine-type Ti plasmid pTiB6S3 (87) and harbours the pCH32 plasmid which overexpresses the virulence genes virE and virG (88). LBA4404 has Ach5 as its chromosomal background and contains a disarmed octopine-type plasmid pTiAch5, named pAL4404 (89).

It was found that Agrobacterium chromosomal background plays a vital role in the infection of the MM1 Arabidopsis cell line. LBA4404 (Ach5 background) outperformed all the strains with C58 background, and it is recommended for the cell culture transformations. On the contrary, in Arabidopsis floral dip transformation, GV3101 was shown to have higher transformation frequencies than LBA4404 (90), while in transient assays, C58C1* found to be the best strain either for leaves (83) or seedlings (91). Among the strains with C58 background, AGL1 provided the lowest infection showing that the disarmed virulent vector is another factor that affects the transformation efficiency. The other two C58 strains (GV3101 and C58C1*) exhibited strong transformation efficiency, with GV3101 being slightly more efficient when harbouring the pMAP vectors. It was also shown that the selection of the vector is of crucial importance for the cell culture transformation, with pLX based and pMAP vectors outperforming pGreen vectors. In fact, pGreen demonstrated the worst transformation efficiency, and it is not recommended for the transformation of the MM1 cell line. Among pLX and pMAP there was no considerable difference in the strains C58C1*, LBA4404 and AGL1, while pMAP demonstrated better activity in GV3101.

Three factors may explain these results: the copy number of the vector, the segregational plasmid stability, and the T-DNA borders. pLX and pGreen vectors have a medium and low copy number origin of replication for Agrobacteria, respectively. pMAP, on the other hand, resulted in plasmid concentrations between pLX and pGreen. The higher the copy number, the more copies of a construct can be transferred to the cells. pGreen's origin of replication, pSa, does not include the stability regions necessary for the maintenance of the plasmid (50, 92). On the other hand, pLX and pMAP origins of replication (BBR1 and WKS1) were derived from cryptic plasmids which are stable without the need of stability regions (67, 93). As there is no selective pressure during Agrobacterium transformation, possibly, pGreen plasmids are not 100% maintained in the agrobacteria strains. Also, even though WKS1 is thought to replicate at a lower

copy number than BBR1, pMAP exhibited similar or stronger transformation efficiency than pLX, which might be attributed to its better segregational plasmid stability. A third factor that can affect the transfection is the T-DNA borders of the binary vector. pGreen has minimal synthetic LB and RB sequences, derived from pTiT37 and an Overdrive from the LBA4417 plasmid (50), while pLX and pMAP have extended LB and RD T-DNA borders, including the overdrive from the plasmid pTiA6 (54).

3.3.4. New short promoters and terminators

We generated and characterised a new collection of short promoter and terminator standard parts for plant expression. The library is comprised of 17 constitutive and three inducible promoters, and 14 terminators. Ten of the promoters and six of the terminators were newly isolated in the present study, and they have a sequence length between 300bp-500bp (promoters) and 200bp (terminators). Until now, little attention has been paid to the size of the plant parts that have been isolated and characterised. Available repositories (e.g. MoClo plants, GB plants and GreenGate) have promoters and terminators up to 4kb. The large size of the expression elements can burden the multigene delivery in plants as the large size of the binary vectors will lead to plasmid instabilities and low transformation efficiency, or incomplete/truncated transformation (45, 46).

There are indications that short plant terminators are adequate for transcription termination. Notably, it was shown that transcripts with 3′ UTRs longer than 300 bp do not improve the mRNA stability (38). Also, in another study, ten randomly selected mRNA sequences from Arabidopsis showed an average 3’UTR length of 209bp (39). Our short terminators were found to have matching or surpassing expression activity compared to the commonly used terminators (e.g. NOS, 35S). Additionally, there are already examples of short, strong constitutive promoters from Arabidopsis, such as AtTCTP - 0.3 bp (74) and AtSCPL30 - 0.45bp (94). In this study, the plant toolbox was further equipped with short promoters with a range of expression levels. An Actin 2 promoter (787bp) was characterised in

MoClo and GB2.0 toolboxes, showing a comparable activity to NOS promoter (76, 95). The shorter version of this study (340bp) exhibited similar expression activity. More specifically, in combination with HSP, NOS had 1.4 RLU and ACT2 4.6 RLU, while the corresponding expression from the GB2.0 versions were 2.61 for NOS and 4.36 for ACT2.

New promoters were isolated, which had a variety of expression levels. UBQ11 and UBQ4 promoters showed strong expression, and they performed better than the commonly used standard promoter parts 35S and OCS, as well as the short promoter AtTCTP. Sequence analysis through the online programs PLACE and PlantCARE showed that the UBQ11 promoter contains 11 CAAT-box and 6 DOFCOREZM, while the UBQ4 promoter harbours 9 CAAT-box and 7 DOFCOREZM elements (**Figure 11**) DOFCOREZM elements are the binding sites of Dof1 and Dof2 TFs, which were found to enhance gene expression (96). CCAT-box also influences expression efficiency (97). Considering that the UBQ10 promoter is a well-known and commonly used constitutive promoter, and also exhibited the highest expression in our assay, we speculate that genes of the polyubiquitin family are good candidates for strong promoters.

The UBQ10 promoter in Arabidopsis protoplasts showed ~20 and ~11 times higher activity than the 35S promoter, combined with NDUFA8 and HSP terminators, respectively. On the contrary, in the part characterisation in *Nicotiana benthamiana* leaves, the 35S promoter had the highest activity as shown by the MoClo and GB2.0 papers (76, 95). Similarly, the OCS promoter had strong activity in our study, but low activity in the characterisation in tobacco leaves (76). Consequently, some genetic parts behave differently in different chassis, and we should not generalise the results to other plant species or expression systems (whole plants, transient leaf expression systems, protoplasts, or cell cultures).

UBQ11 promoter

```
>PlantCARE_21575
+ AACCTCACTT GTGAACCTTA CACAAAGGGT TGTACGCAGA CTATCATTAG CCAAATTCCA AGCTTCTTGC
- TTGGAGTGAA CACTTGGAAT GTGTTTCCCA ACATGCGTCT GATAGTAATC GGTTTAAGGT TCGAAGAACG

+ CTCAATTCCG GAGGTGTTTC TAGTGTTCAA CATGACAAAC AAAACCCATC TCTTTCAGTA TATGTCTCTC
- GAGTTAAGGC CTCCACAAAG ATCACAAGTT GTACTGTTTG TTTTGGGTAG AGAAAGTCAT ATACAGAGAG

+ AGTTGTGCTT AATTCAAATT TCAACTCAGA GAACTTCTTG GCATACTTAT CCAGATTATC TAATGATCTC
- TCAACACGAA TTAAGTTTAA AGTTGAGTCT CTTGAAGAAC CGTATGAATA GGTCTAATAG ATTACTAGAG

+ ATCTAATGGT AATTCAACTT TCAGTATATG TCTCGCAGCA AACTATCTTT ACATCAAATT TTTAACAACT
- TAGATTACCA TTAAGTTGAA AGTCATATAC AGAGCGTCGT TTGATAGAAA TGTAGTTTAA AAATTGTTGA

+ CAATGCACAA AATACTTTTC CTCAACCTAA AAATAAGGCA ATTAGCCAAA AACAACTTTG CGTGTGAACA
- GTTACGTGTT TTATGAAAAG GAGTTGGATT TTTATTCCGT TAATCGGTTT TTGTTGAAAC GCACACTTGT

+ ACGCGTACAC GTCCCTACAC ATACGTGTCA ATTTATAATT GGCTATTGCT TCCACGCCTT AGCTTTCTCG
- TGCGCATGTG CAGGGATGTG TATGCACAGT TAAATATTAA CCGATAACGA AGGTGCGGAA TCGAAAGAGC

+ TGACCGACCG AGTCGTCCTC GTCTTTTTTG CTTCTATAAA TCAAATACCC AAAGAGCTCT TCTTCTTCAC
- ACTGGCTGGC TCAGCAGGAG CAGAAAAAAC GAAGATATTT AGTTTATGGG TTTCTCGAGA AGAAGAAGTG

+ AATTCAGATT C
- TTAAGTCTAA G
```

UBQ4 promoter

```
>PlantCARE_6060
+ CTTAAGGGTG TGCTTGGGGG AACTAATTAA GCTTCTCTTA GTTTCATTGT TCATTTTTTT AAAAATGTGG
- GAATTCCCAC ACGAACCCCC TTGATTAATT CGAAGAGAAT CAAAGTAACA AGTAAAAAAA TTTTTACACC

+ TTGTGCGTTT TGTTATTTCG AGTTCCTTTA GCATTTTGAG TTTGTTTCTT CTAATGATTT GGTTTTTCTT
- AACACGCAAA ACAATAAAGC TCAAGGAAAT CGTAAAACTC AAACAAAGAA GATTACTAAA CCAAAAGAA

+ TATGAGTTGT TATTGTTCTC AATGTTTCAT TTTATAGTCT CTTTGGCTTC AAATTGTAAT AACTAATAAG
- ATACTCAACA ATAACAAGAG TTACAAAGTA AAATATCAGA GAAACCGAAG TTTAACATTA TTGATTATTC

+ CATTATATTG CAGATTTCTA AGTAGCTTGT CAACTATATC TCGACATTTG CACTAATAAG CAACTATTTA
- GTAATATAAC GTCTAAAGAT TCATCGAACA GTTGATATAG AGCTGTAAAC GTGATTATTC GTTGATAAAT

+ GCCGATGTTC AAAAAAATTT AAAATAAATA ATGTTTTGGG CCTTTGGGCC TTTTGGGCCA ATTAATTAAA
- CGGCTACAAG TTTTTTTAAA TTTTATTTAT TACAAAACCC GGAAACCCGG AAAACCCGGT TAATTAATTT

+ ACTCGGTTTT GGTCACCATT ATATAAGAAG GCTATTTTCT AGCGAAGTAT TATCTTCTGC TCTGATTCTT
- TGAGCCAAAA CCAGTGGTAA TATATTCTTC CGATAAAAGA TCGCTTCATA ATAGAAGACG AGACTAAGAA

+ CATCTCTCTC ACAGAAAAAA GTAAAAAAAA ACCTAGATCG CTCTTCACAT CTCTTCGAAC GATTCTCTCT
- GTAGAGAGAG TGTCTTTTTT CATTTTTTTT TGGATCTAGC GAGAAGTGTA GAGAAGCTTG CTAAGAGAGA

+ TTCAAG
- AAGTTC
```

Figure 11. Potential sequence elements in the UBQ11 and UBQ4 promoters linked to increased gene expression. Sequence analysis performed with the online programs PLACE and PlantCARE. Yellow highlighted sequences show the DOFCOREZM elements, pink the CAAT-boxes in the UBQ11 promoter and blue the CAAT-boxes in the UBQ4 promoter.

Even though ACT7, TUB2, TUB9, APT1, ACT2 and LEC2 promoters had low expression activity, they performed better than NOS, the commonly used

promoter to drive the antibiotic/herbicide resistance for transgene selection. Therefore, they can be used for the expression of selection markers, or for any experiments that do not require high expression levels of the transgenes of interests.

3.3.5. Terminator effects on gene expression

These results showcase the strong influence of the terminators on gene expression. It is a common strategy to rely on promoters to control the level of gene expression and neglect the significance of terminators. For example, to compensate the loss of expression activity after the domestication of the G10-90 promoter, researchers replaced it with the AtRPL37aC promoter (79). However, they were using the psE9-RbcS terminator, which was found in this study to decrease the gene expression. The expression level could have been restored with a stronger terminator instead. The range of expression was independent of the promoter strength, as the terminators in combination with UBQ10 (max 547.9 RLU), NDUFA8 (max 0.3 RLU) and AlcSynth (max 337.9 RLU) showed expression ranges of 5.3, 6.3 and 7.3 times, respectively.

There are indications for interaction between promoters and terminators, as the terminators drove different expression levels when combined with different promoters. NOS terminator is a case in point, as it behaves as a weak terminator in conjunction with the UBQ10 promoter, yet as a strong terminator with the NDUFA8 promoter, and as a medium terminator with the AlcSynth promoter. On the other hand, terminators perform differently combined with the inducible promoters pOp6-35S and lexA-35S. They resulted in a more uniform expression range, ~2.2 times for pOp6-35S and ~2.9 times for lexA-35S, while the interactions with the promoters are negligible. LexA-35S and pOp6-35S inducible promoters are synthetic promoters comprised only of the minimal 35S promoter and DNA binding sites; in contrast, AlcSynth includes the intact alcA promoter and consequently behaves similarly to the other promoters characterised and contains more regulatory elements.

The synergistic regulation between natural promoters and terminators can be explained by the direct interaction of the terminator with the promoter (looping), which can further influence gene expression. Terminator function is not limited in the 3'UTR. Gene looping is the physical connection of the terminator with the promoter region of a gene (98), mediated by nucleic acid-binding proteins (99). It can influence gene expression through transcriptional memory (100), intron-mediated enhancement of transcription (101), transcription directionality (102), reinitiation of transcription (103), and transcription termination (104, 105). Moreover, it was recently shown that the absence of a terminator caused higher DNA methylation on the promoter region and reduction of the transgene expression compared to constructs with a terminator, indicating an alternative role in transcriptional gene silencing (106).

Therefore, a terminator can have a different impact on different promoters. On the other hand, the synthetic inducible promoters do not have binding sites for TFs; hence they do not interact with the terminators, and their function solely depends on the elements that are present on their sequence (e.g. Poly-A signal). For this reason, they do not have a wide range of the gene expression level, since the transcription mainly depends on the promoter. It was shown previously that double terminators increase expression (107, 108). Building on the double terminator observations, Diamo and Mason showed that there is a synergistic enhancement in the double terminator combinations (109). The effect of the double terminators was partially attributed to a reduction of RDR6-mediated silencing. mRNA read through from "leaky" terminators under the influence of a strong promoter triggers RDR6-mediated silencing. However, this explanation cannot elucidate the synergistic effects the terminators cast, especially combined with weak promoters, or between the double terminators. A possible interaction of the terminators with the promoters and the transcription initiation machinery could explain those observations. However, more in-depth work is required to

verify and establish the theory of looping or any other mechanisms on the roles of plant terminators on transcriptional or translational regulation.

As a start, it was attempted to study HSP and FAD2 terminators further and link their activity to different sequence elements and known motifs. To do that, we first mapped putative stabilisation and destabilisation hexamers, Poly-A signals and other regulatory motifs on the terminators, and constructed a deletion series to isolate the effects of these motifs (**Figure 9**). The putative stabilisation and destabilisation hexamers were found to correlate with the mRNA stability (38). The YGTGTTYY motif was found to be necessary for efficient formation of mRNA 3' termini (110). Musashi binding element represses the translation of mRNAs in mammalian stem cells or activates translation in maturing Xenopus oocytes (111).

Only removal of the FAD2 Poly-A site was able to reduce the expression considerably in Arabidopsis protoplasts, revealing that it is an essential part of the terminator activity (**Figure 10**). Regarding the other tested motifs, no conclusive correlations to the expression activity were made, as there were no statistically significant effects on the readout of the luciferase activity. Possibly, the strong UBQ10 promoter used could drive strong expression even without the presence of the terminator elements. Additionally, comparison of the luciferase activity in the two different plant systems revealed around ten times weaker expression in the MM1 cell line than the Arabidopsis protoplasts. The MM1 cell line has an embryonic identity, while protoplasts are mainly mesophyll parenchyma cells (though they start to de-differentiate within the first 24 hours of protoplast isolation) (112), which can explain the expression differences.

3.3.6. Terminator effects on inducible gene expression

Chemically inducible systems are essential for precise control of the timing and amplitude of transgene expression. Even though β-estradiol, Dex, and ethanol inducible systems were developed almost 20 years ago, and have been very popular since then (113); no work has been reported about manipulating their

expression using different terminators. In this study, it was attempted to modulate the gene expression ranges of the inducible promoters by exchanging the terminators. It was found that lexA has a 2.9 times difference in the expression level when combined with different terminators (37.9-111.6 RLU), while pOp6 showed 2.2 times change in the expression levels depending on the different terminators (11.6-25.1 RLU) (**Figure 8A and B**).

Both the Dex and Estradiol systems had similar basal expression levels of ~10 RLU (with the HSP terminator); however, lexA-35S has demonstrated ~10-fold activation amplitude, while pOp6-35S only three times (**Figure 8A and B**). The AlcSynth promoter, on the other hand, had a high basal expression (250.3 RLU with the HSP terminator) with 1.3 times activation (**Figure 8C**). AlcSynth is a fusion of the alcohol dehydrogenase promoter sequence (alcA) and six more copies of the AlcR binding sites derived from the promoters of the alcM, alcR and aldA with the minimal Figwort Mosaic Virus (FMV) 34S and TMV Ω enhancer sequence (54) (**Figure 12**). A possible explanation for the leakiness of the ethanol system might be TMV Ω; addition of the TMV Ω in the LhGR-pOp6 system exhibited slightly increased basal expression (80). Also, as the whole alcA promoter is directly fused to the minimal 35S or 34S promoter, constitutive transcription elements present in the sequence might be responsible for the leakiness. That could explain why alcSynth behaves like a constitutive promoter against the different terminators, showing a wide range of activity (46.6 to 321.8 RLU) and synergy. Synthetic design of the ethanol inducible promoter, in which only AlcR binding sites are present separated by random spacer sequences and fused to a minimal plant promoter, could reduce the leakiness and improve the stringency of the gene regulation. Taken together, only the Estradiol and Dex inducible systems worked well in cell-based systems, and more inducible systems, such as optogenetics or other chemical inducible systems are required to enrich the synthetic biology toolkit to build circuitry gene expression control.

Figure 12. AlcSynth ethanol inducible promoter. It is a fusion of the alcohol dehydrogenase promoter sequence (alcA) and elements of the alcM (pink), alcR (Orange) and aldA (yellow) with the minimal Figwort Mosaic Virus (FMV) 34S and the translational enhancer from the tobacco mosaic virus 5'-leader sequence (TMV Ω). Deep green is 5'➔3' and pale green 3'➔5' AlcR binding sites. Black box is TATA box.

3.3.7. Shareable kit for Addgene

A selection of cloning and transformation vectors and well as standard parts will be sent to Addgene, to be accessible to plant scientists. A table with all parts is shown in Appendix 3. The kit will have the mUAV and the 7 Auxiliary plasmids as described in Mobius Assembly (7). The main plasmids for cloning and plant transformation will be the four pMAP Level 1 and Level 2 vectors, which are suitable for transformation of several plant systems (e.g. protoplasts, cell cultures, whole plants). For plant transformation, the pLX based vectors with BBR1 and RK2 Ori (medium and low copy number, respectively) are included to deal with unstable constructs. From the promoters, we will exclude LEC2, FAD2 and NDUFA8, which have the lowest strength and will not have any application. From the terminators, we will exclude just E9-RbcS which double the size from the rest and can be replaced with others with the same activity. Only a few where consistently "high" or "low" in the characterisations conducted (HSP, FAD2, 35S, RbcS2b) (**Table 2**), however even those could act differently if they are combined with different promoters and genes. Consequently, all of them will be distributed as will be useful for different experimental requirements. The kit will also include the standard parts of the three inducible systems we characterised and a selection of fluorescent proteins with different characteristics, as well as the nano and firefly luciferases. Finally, the five agrobacterium strains AGL1, GV3101, LBA4404, C58C1* will be included as they will be useful for transforming different plant species or tissues/cells.

3.4. Materials and Methods

3.4.1. Bacterial strains and growth conditions

E. coli strains DH5α, DH10B, TOP10, JM109, NEB stable and *A. tumefaciens* strains AGL-1, C58C1*, GV3101, LBA4404 were used. *E. coli* chemically competent cells were made using the TSS preparation, described by Chung & Miller (114), except TOP10, which were bought from ThermoFisher Scientific. A. tumefaciens electrocompetent cells were prepared as follows: Overnight seed culture was grown in S.O.C medium (0.5 ml) at 28˚C, 200 min^{-1} with the appropriate antibiotics, was diluted in 500 ml of the same fresh medium. After overnight incubation at the same conditions, and when the culture reached OD_{600} 0.6-1.0, cells were harvested and washed three times with ice-cold 10% sterile glycerol. Finally, cells were resuspended in 1.5 ml 10% cold glycerol, aliquoted in 40 µl batches and snap-frozen in liquid nitrogen.

E. coli cells were incubated at 37°C (30°C for NEB stable), 200 min^{-1} either in 5 ml (for a high copy) or 10 ml (for a low copy), or 100 ml (for midi prep) LB growth medium supplemented with antibiotics. Cells bearing the LhGR-pOp6 and sXVE-lexA inducible systems were grown for 24 h instead due to the slower growth rate. Agrobacterium cells were cultured in 10 ml YEP medium at 28°C, 200 min^{-1} for two days.

Agrobacterium strain	Antibiotic Resistance	Antibiotic Concentration
AGL1	Rifampicin	50 µg/ml
	Carbenicillin	100 µg/ml
GV3101	Rifampicin	50 µg/ml
	Gentamycin	25 µg/ml

LBA4404	Rifampicin	50 µg/ml
C58C1*	Rifampicin Tetracycline Carbenicillin	50 µg/ml 15 µg/ml 100 µg/ml

3.4.2. Plasmid isolation

Plasmids were isolated using Monarch (NEB) or PureYield (Promega) or GeneJET (Thermo Fisher Scientific) Plasmid Miniprep Kits. For higher yields, the Promega PureYield Plasmid Midiprep System (Promega) with vacuum manifold and GeneJET Plasmid Midiprep Kit (Thermo Fisher Scientific) was used. Plasmid isolation from Agrobacteria was carried out with 1.5 times the recommended reagent volumes

3.4.3. Bacterial transformations

For *E. coli* transformation, 5 µl of the plasmid DNA was incubated with 100 µl of the competent cells (25 µl for TOP10) on ice for 30 min, followed by a heat shock at 42°C for 90 s (30 s for TOP10) and re-cooled on ice for 5 min. S.O.C medium (400 µl) was added, and after 1 h incubation at 37°C, 100 µl of the cell suspension was plated on LB agar plates with antibiotic selection. For Agrobacterium transformation, 1 µl of plasmid DNA was added into an aliquot of competent cells and incubated on ice for 5 min. Subsequently, the cells were transferred into an ice-cold electroporation cuvette, 0.2 cm gap (Bio-Rad) and pulsed twice in a Gene Pulser Xcell Electroporation System (Bio-Rad) using pre-set settings for Agrobacteria (25 µF, 200 Ω, 2400 V, 2 mm). S.O.C medium was added immediately (1 ml), and after 2h of shaking at room temperature, 20 µl was plated on YEP plates with the correct antibiotics and incubated at 30°C for 2-3 days until colonies had formed.

3.4.4. Molecular biology techniques

All PCR amplifications for plasmid construction and cloning were performed using Q5® High-Fidelity DNA Polymerase (NEB), followed by purification with Monarch® PCR & DNA Cleanup Kit (NEB). The Mobius Assemblies were verified first by colony PCR using GoTaq® Green Master Mix (Promega) and then with double restriction digestion with *EcoRI*-HF (NEB) and *PstI*-HF (NEB). The constructs were further verified by Sanger sequencing (GATC Biotech-Eurofins or Edinburgh Genomics).

3.4.5. Plasmid construction

Mobius Assembly Level 1 and Level 2, Acceptor Vectors for plants, were built using Gibson Assembly, while mUAV and Auxiliary Plasmids were kept the same as in (7). More specifically, the Mobius Assembly cassettes were amplified from the corresponding *E. coli* plasmids in the Mobius Assembly Vector toolkit and fused to the plant binary vectors. pGreen based Level 1 vectors were constructed using pGreen0029 (50) acquired from the John Innes Centre. NptI gene was replaced with a spectinomycin resistance gene amplified from pCR8 vector (ThermoFisher Scientific) to generate Level 2 vectors. pLX based vectors Level 1 and Level 2 were built using pLX-B3 and pLX-B2 (54), respectively, which were acquired from Centro Nacional de Biotecnología (CNB-CSIC). For pLX-BBR1mut, primers were designed to introduce a point mutation (C_{299} to T_{299}) into the Rep gene of pBBR1 origin of replication to increase the copy number (82). pLX Level 1A and Level 2A plasmids were PCR amplified into two parts using the primers carrying the mutation and the Assembly Linker 1 primers. Gibson assembly was employed to reconstruct the plasmids, and after transformation, colonies with bright pink and yellow colour were, selected for the pLX BBR1mut version. MethylAble modules were again devised with Gibson Assembly using mUAV as a template. Overlapping primers bearing two outward-facing *BsaI* sites prone to CpG methylation and the suitable standard overhangs were used to amplify

mUAV into two parts. The resulting parts were purified, digested with *DpnI* (*ThermoFisher Scientific) to* eliminate the template DNA, and they were fused in an isothermal reaction. WKS1 Ori was synthesised by Twist Bioscience into two parts (due to repetitive sequences) and pUC Ori was amplified from pUC19, both flanked by *BsaI* recognition sites. A pLX Level 1 A vector with the construct NOS:BglR:NOS-UBQ10:nluc:HSP was amplified outside the BBR1 Ori with primers harbouring *BsaI* recognition sites and fused with pUC and WKS1 Ori. The resulting plasmid was used as a template to amplify the pUC-WKS1 fused Ori, which was used to replace the BBR1 Ori from pLX Level 1A and Level 2A vectors with Gibson Assembly, resulting in the pMAP Level 1A and Level 2A vectors. The rest of the pMAP vectors were constructed again using isothermal assembly and as a template for the Mobius cloning cassettes the plasmids from the original Mobius Assembly kit (7).

3.4.6. Structural plasmid stability assessment

The construct UBQ10:fluc:UBQ5-UBQ10:LhGR:HSP-pop6:nluc:HSP was used for the stability studies and cloned either in pGreen based, pLX BBR1 based, pLX BBR1mut or pMAP based Mobius Assembly vectors and transformed in DH5α, DH10B, TOP10, JM109 or NEB stable E. coli strains. LB cultures, after 24h of incubation, were used for plasmid isolation. Structural plasmid stability was tested first electrophoretically in 1% agarose gel after digestion with PstI-HF (NEB) and then by Sanger sequencing of the tandem short repeats of the pOp6 sequence. Next, the Agrobacteria strains AGL-1, C58C1*, GV3101, LBA4404 were transformed with a construct from each plasmid with correct digestion pattern and sequence, and the same process was followed to determine the structural plasmid stability.

3.4.7. Mobius Assembly cloning

Mobius Assemblies were performed in a one-tube reaction with a total volume of 10 µl, with ~50 ng Acceptor Vectors and double amounts of inserts. Reagents added were 1 µl of 1 mg/ml BSA (diluted from 20 mg/ml - NEB), 1 µl T4 DNA ligase buffer (NEB or Thermo Fisher Scientific), 0.5 µl *AarI* (ThermoFisher Scientific) and 0.2 µl 50x oligos (*AarI* recognition site) for Level 0 and Level 2 cloning or Eco31I/*BsaI*-HFv2 (ThermoFisher Scientific/NEB) for Level 1 cloning, and 0.5 µl T4 DNA ligase (NEB or Thermo Fisher Scientific). The reactions were incubated in a thermocycler for 5-10 cycles of 5 min at 37°C and 10 min at 16°C, followed by 5 min digestion at 37°C and 5 min deactivation at 80°C. (5 cycles for Level 0 and the first round of Level 1 cloning – 10 cycles for Level 2 and large constructs >10 kb).

3.4.8. CpG Methylation

Midi-prep plasmid DNA of the MethylAble modules was used. The reaction was carried out in 20 µl total volume according to the NEB protocol. Briefly, up to 2 µg of plasmid DNA was incubated for 4 h at 37°C with 2 µl of CpG Methyltransferase from M.SssI (NEB) in 1× Methyltransferase Reaction Buffer supplemented with 2 µl of diluted SAM (6.4 mM). The reaction was stopped by heating to 65°C for 20 min and was purified by column chromatography.

3.4.9. Selection and amplification of Arabidopsis promoters and terminators

Ubiquitously expressed genes used in qPCR or genes expressed in the plant embryos were identified through a literature search. The most stably expressed were selected for promoter and terminator identification. Their sequences were retrieved from the TAIR webpage http://www.arabidopsis.org/index.jsp and blasted in NCBI to find the untranslated regions flanking the genes. A 1.5 kb sequence upstream of the start codon was run through the online prediction

software and TSSPlant (http://www.softberry.com) to identify TATA and TATA-less promoters or enhancer sites. In the promoter selection it was also considered, when possible, to include the initiator (INR) elements (YYA(+1)NWYY-TYA(+1)YYN-TYA(+1)GGG)) and downstream promoter (DPE) element (RGWYV). Finally, two promoter versions were created based on the sequence size, one ~300 bp and one ~500 bp.

Potential elements linked to increased gene expression were identified using PlantCare (115) and PLACE (116) software.

Terminators were selected with PASPA, a web server for poly(A) site prediction in plants and algae (http://bmi.xmu.edu.cn/paspa/interface/run_PASPA.php). A sequence 300bp downstream of the stop codon was run into PASPA, and the selection of the terminator was set 10bp after the second polyadenylation site, resulting in a sequence of around 200bp. They were then analysed in RegRNA2.0 for the identification of further RNA functional motifs (http://regrna2.mbc.nctu.edu.tw). Appropriate primers were designed for both promoters and terminators for cloning into mUAV.

3.4.10. Functional dissection of FAD2 and HSP terminators

Primers were designed to gradually remove functional sequence elements from the 3' end of each terminator through PCR and subsequently cloned to mUAV. Site-directed mutagenesis by Gibson Assembly was employed to mutate the poly-A signal of HSP part3, while the same method was used to build HSP part 5.

3.4.11. Plant growth

Arabidopsis thaliana (Wildtype Col-0) seeds were sown on the soil. After a 2-day pre-treatment at 4°C in darkness, they were grown under long-day conditions (21°C; 16h light / 8h dark cycles; light intensity ~100µmol/m^2s^{-1}; humidity ~65%) for two weeks. Then, seedlings were transplanted and moved to short-day

conditions (21°C; 9h light / 15h dark cycles; light intensity ~110µmol/m2s-1) and grown for 4-6 weeks until harvest.

3.4.12. Plant-cell suspension culture growth and Agrobacterium Mediated transformation

MM1 cell suspension culture of *Arabidopsis thaliana* ecotype Landsberg erecta (117) was subcultured weekly by 1:10 dilution in 1× MS medium (Sigma) supplemented with 2% w/v glucose (Sigma), 0.5 mg/L NAA (Sigma), 0.05 mg/L kinetin (Sigma), with the pH adjusted to 5.8 with KOH. Agrobacterium-mediated cell culture transformations were performed according to a modified protocol (118) under sterile conditions. Freshly prepared Agrobacterium plates (2-3 day-old maximum) and Arabidopsis cells 5-7 days post-sub-culturing were used. 25 ml of cell culture was transferred to a 50 ml falcon tube and centrifuged at 100 min^{-1}, at 22^{o}C for 10 min. The supernatant was removed with a 25 ml serological pipette and the cells were resuspended in 40 ml of transformation medium at pH 5.8 (1× MS plant salt mixture without NH_4NO_3 (Sigma-Aldrich), 1% w/v sucrose, 1% v/v B5 vitamin stock solution and 1 µg/ml 2,4-dichlorophenoxyacetic acid (Sigma-Aldrich)) and 25 µl of 100 mg/mL freshly prepared acetosyringone (Sigma-Aldrich) was added. B5 vitamin stock solution was prepared with 0.4 g/l nicotinic acid (Sigma-Aldrich), 0.4 g/l pyridoxine-HCl (Supelco/Sigma-Aldrich), 4 g/l thiamine-HCl (Merck), and 40 g/l myo-inositol (Merck). After gentle resuspension, 4 ml of plant cells were transferred in each well of a 6-well plate, and inoculated with a small amount of agrobacterium directly from the petri dish with the help of a 200 µl pipette. After 2 days of incubation on a rotary shaker, plant cells were supplemented with 5 ml of fresh culture medium and 50 µl of cefotaxime (Arcos) at 100 mg/mL which was added to eliminate agrobacteria.

3.4.13. Arabidopsis protoplast isolation and transformation

This protocol was developed based on (119) and (120). Briefly, leaves from 6-8 week old plants were digested with MGG digestion solution containing: Cellulase

ONOZUKA R-10 (0.2%), MACEROZYME R-10 (0.06%) (Yakult Pharmaceutical) and Driselase (0.08%) (Sigma-Aldrich). The following day, the protoplasts were filtered and washed 3-5 times with MGG without enzymes and resuspended in MMM solution (0.4 M mannitol, 15 mM $MgCl_2$, 0.1% w/v MES, pH 8) in a concentration of $5x10^5$ cells/ml. Protoplasts were transformed into 1 ml 96-well plates. 8 µl DNA (500 ng/µl) was added to 75 µl of the protoplast suspension, followed by addition of 83 µl PEG (0.4 M mannitol, 0.1 M $Ca(NO_3)_2$-$4H_2O$, 40% PEG4000, pH8) and 1 min incubation. Subsequently, the tubes were filled with MGG solution and incubated for 1 h at RT. After the incubation, the solution was removed, and the protoplasts were resuspended in 100 µl of fresh MGG and incubated overnight in darkness at RT. For the inducible system, 0.1% EtOH, 2.5 µM DEX (Acros) or 5 µM β-estradiol (LKT laboratories) were added.

3.4.14. Luciferase assay

The plates containing the transformed protoplasts were centrifuged at 200 min^{-1} (acc/decc. = 3) for 10 min, and 60 µl of supernatant was discarded. The protoplasts were resuspended, and 40 µl was transferred to white optical plates in a grid pattern with empty spaces between wells to reduce luminescence bleed-through. For the MM1 transformed cells, 40 µl were transferred with a cut-tip directly from the 6/12-well plates to 96-well plates in triplicate. Luciferase activity was assayed in an Omega luminescence plate-reader (Fluostar) with four different gains following the instructions of the Nano Dual-Luciferase® Reporter kit (Promega). A further correction for luminescence bleed-through was applied using the software developed by Mauri *et al*. (121).

3.5. References

1. Liu,W., Yuan,J.S. and Stewart Jr,C.N. (2013) Advanced genetic tools for plant biotechnology. *Nat. Rev. Genet.*, **14**, 781.
2. Naqvi,S., Farré,G., Sanahuja,G., Capell,T., Zhu,C. and Christou,P. (2010)

When more is better: multigene engineering in plants. *Trends Plant Sci.*, **15**, 48–56.

3. Que,Q., Chilton,M.-D.M., de Fontes,C.M., He,C., Nuccio,M., Zhu,T., Wu,Y., Chen,J.S. and Shi,L. (2010) Trait stacking in transgenic crops: Challenges and opportunities. *GM Crops*, **1**, 220–229.

4. Townson,J. (2017) Recent developments in genome editing for potential use in plants. *Biosci. Horizons Int. J. Student Res.*, **10**.

5. Wu,G., Truksa,M., Datla,N., Vrinten,P., Bauer,J., Zank,T., Cirpus,P., Heinz,E. and Qiu,X. (2005) Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.*, **23**, 1013–1017.

6. Čermák,T., Curtin,S.J., Gil-Humanes,J., Čegan,R., Kono,T.J.Y., Konečná,E., Belanto,J.J., Starker,C.G., Mathre,J.W., Greenstein,R.L., *et al.* (2017) A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. *Plant Cell*, **29**, 1196 LP – 1217.

7. Andreou,A.I. and Nakayama,N. (2018) Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS One*, **13**, e0189892.

8. Weber,E., Engler,C., Gruetzner,R., Werner,S. and Marillonnet,S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One*, **6**, e16765.

9. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*, **4**.

10. Ong,J.L., Bilotti,K., Evans Jr,T.C., Potapov,V., Lohman,G.J.S., Langhorst,B.W., Canton,B., Cahoon,D. and Knight,T.F. (2018) A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Res.*, **46**, e79–e79.

11. Potapov,V., Ong,J.L., Kucera,R.B., Langhorst,B.W., Bilotti,K., Pryor,J.M., Cantor,E.J., Canton,B., Knight,T.F., Evans,T.C., *et al.* (2018) Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth. Biol.*, **7**, 2665–2674.

12. Lin,D. and O'Callaghan,C.A. (2018) MetClo: methylase-assisted hierarchical DNA assembly using a single type IIS restriction enzyme. *Nucleic Acids Res.*, **46**, e113.

13. Taylor,G.M., Mordaka,P.M. and Heap,J.T. (2018) Start-Stop Assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Res.*, **47**, e17–e17.

14. van Dolleweerd,C.J., Kessans,S.A., Van de Bittner,K.C., Bustamante,L.Y., Bundela,R., Scott,B., Nicholson,M.J. and Parker,E.J. (2018) MIDAS: A

Modular DNA Assembly System for Synthetic Biology. *ACS Synth. Biol.*, **7**, 1018–1029.

15. Valenzuela-Ortega,M. and French,C. (2019) Joint Universal Modular Plasmids (JUMP): A flexible and comprehensive platform for synthetic biology. *bioRxiv*, 10.1101/799585.

16. Vasudevan,R., Gale,G.A.R., Schiavon,A.A., Puzorjov,A., Malin,J., Gillespie,M.D., Vavitsas,K., Zulkower,V., Wang,B., Howe,C.J., *et al.* (2019) CyanoGate: A modular cloning suite for engineering cyanobacteria based on the plant MoClo syntax. *Plant Physiol.*, 10.1104/pp.18.01401.

17. Crozet,P., Navarro,F.J., Willmund,F., Mehrshahi,P., Bakowski,K., Lauersen,K.J., Pérez-Pérez,M.-E., Auroy,P., Gorchs Rovira,A., Sauret-Gueto,S., *et al.* (2018) Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga Chlamydomonas reinhardtii. *ACS Synth. Biol.*, **7**, 2074–2086.

18. Garrigues,S., Manzanares,P., Yenush,L., Orzaez,D., Gandía,M., Hernanz-Koers,M. and Marcos,J.F. (2018) FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. *Fungal Genet. Biol.*, **116**, 51–61.

19. Marx,H., Egermeier,M. and Sauer,M. (2019) Golden Gate based metabolic engineering strategy for wild-type strains of Yarrowia lipolytica. 10.1093/femsle/fnz022.

20. Pollak,B., Cerda,A., Delmans,M., Álamos,S., Moyano,T., West,A., Gutiérrez,R.A., Patron,N.J., Federici,F. and Haseloff,J. (2019) Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.*, **222**, 628–640.

21. Gantner,J., Ordon,J., Ilse,T., Kretschmer,C., Gruetzner,R., Löfke,C., Dagdas,Y., Bürstenbinder,K., Marillonnet,S. and Stuttmann,J. (2018) Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. *PLoS One*, **13**, e0197185.

22. Mora,A., Sandve,G.K., Gabrielsen,O.S. and Eskeland,R. (2016) In the loop: promoter-enhancer interactions and bioinformatics. *Brief. Bioinform.*, **17**, 980–995.

23. Mithra,S.V.A., Kulkarni,K. and Srinivasan,R. (2017) Plant Promoters: Characterization and Applications in Transgenic Technology BT - Plant Biotechnology: Principles and Applications. In Abdin,M.Z., Kiran,U., Kamaluddin, Ali,A. (eds). Springer Singapore, Singapore, pp. 117–172.

24. Roy,A.L. and Singer,D.S. (2015) Core promoters in transcription: old problem, new insights. *Trends Biochem. Sci.*, **40**, 165–171.

25. Shahmuradov,I.A., Umarov,R.K. and Solovyev,V. V (2017) TSSPlant: a new tool for prediction of plant Pol II promoters. *Nucleic Acids Res.*, **45**, e65–e65.

26. Patikoglou,G.A., Kim,J.L., Sun,L., Yang,S., Kodadek,T. and Burley,S.K. (1999) TATA element recognition by the TATA box-binding protein has been conserved throughout evolution.

27. Joshi,C.P. (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.*, **15**, 6643–6653.

28. Molina,C. and Grotewold,E. (2005) Genome wide analysis of Arabidopsis core promoters. *BMC Genomics*, **6**, 25.

29. Hetzel,J., Duttke,S.H., Benner,C. and Chory,J. (2016) Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci.*, **113**, 12316 LP – 12321.

30. Lemon,B. (2002) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.*, **14**, 2551–2569.

31. Yamamoto,Y.Y., Ichida,H., Matsui,M., Obokata,J., Sakurai,T., Satou,M., Seki,M., Shinozaki,K. and Abe,T. (2007) Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics*, **8**, 67.

32. Yamamoto,Y.Y., Yoshitsugu,T., Sakurai,T., Seki,M., Shinozaki,K. and Obokata,J. (2009) Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J.*, **60**, 350–362.

33. Ye,R., Zhou,F. and Lin,Y. (2012) Two novel positive cis-regulatory elements involved in green tissue-specific promoter activity in rice (Oryza sativa L ssp.). *Plant Cell Rep.*, **31**, 1159–1172.

34. Hunt,A.G., Xing,D. and Li,Q.Q. (2012) Plant polyadenylation factors: conservation and variety in the polyadenylation complex in plants. *BMC Genomics*, **13**, 641.

35. Millevoi,S. and Vagner,S. (2010) Molecular mechanisms of eukaryotic pre-mRNA 3’ end processing regulation. *Nucleic Acids Res.*, **38**, 2757–2774.

36. Karamyshev,A.L. and Karamysheva,Z.N. (2018) Lost in Translation: Ribosome-Associated mRNA and Protein Quality Controls . *Front. Genet.* , **9**, 431.

37. He,F. and Jacobson,A. (2015) Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu. Rev. Genet.*, **49**, 339–366.

38. Narsai,R., Howell,K.A., Millar,A.H., O’Toole,N., Small,I. and Whelan,J. (2007)

Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell*, **19**, 3418–3436.

39. Li,X.-Q. and Du,D. (2014) Motif types, motif locations and base composition patterns around the RNA polyadenylation site in microorganisms, plants and animals. *BMC Evol. Biol.*, **14**, 162.

40. Nicholson,A.L. and Pasquinelli,A.E. (2018) Tales of Detailed Poly(A) Tails. *Trends Cell Biol.*, https://doi.org/10.1016/j.tcb.2018.11.002.

41. Baeg,K., Iwakawa,H. and Tomari,Y. (2017) The poly(A) tail blocks RDR6 from converting self mRNAs into substrates for gene silencing. *Nat. Plants*, **3**, 17036.

42. Luo,Z. and Chen,Z. (2007) Improperly Terminated, Unpolyadenylated mRNA of Sense Transgenes Is Targeted by RDR6-Mediated RNA Silencing in Arabidopsis *Plant Cell*, **19**, 943 LP – 958.

43. Oliveira,P.H., Prather,K.J., Prazeres,D.M.F. and Monteiro,G.A. (2010) Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. *Appl. Microbiol. Biotechnol.*, **87**, 2157–2167.

44. Peremarti,A., Twyman,R.M., Gomez-Galera,S., Naqvi,S., Farre,G., Sabalza,M., Miralpeix,B., Dashevskaya,S., Yuan,D., Ramessar,K., *et al.* (2010) Promoter diversity in multigene transformation. *Plant Mol. Biol.*, **73**, 363–378.

45. Ertl,P.F. and Thomsen,L.L. (2003) Technical issues in construction of nucleic acid vaccines. *Methods*, **31**, 199–206.

46. Park,S.H., Lee,B.-M., Salas,M.G., Srivatanakul,M. and Smith,R.H. (2000) Shorter T-DNA or additional virulence genes improve Agrobactrium-mediated transformation. *Theor. Appl. Genet.*, **101**, 1015–1020.

47. Komori,T., Imayama,T., Kato,N., Ishida,Y., Ueki,J. and Komari,T. (2007) Current Status of Binary Vectors and Superbinary Vectors. *Plant Physiol.*, **145**, 1155 LP – 1160.

48. Xiang,C., Han,P., Lutziger,I., Wang,K. and Oliver,D.J. (1999) A mini binary vector series for plant transformation. *Plant Mol. Biol.*, **40**, 711–717.

49. Coutu,C., Brandle,J., Brown,D., Brown,K., Miki,B., Simmonds,J. and Hegedus,D.D. (2007) pORE: a modular binary vector series suited for both monocot and dicot plant transformation. *Transgenic Res.*, **16**, 771–781.

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

51. McBride,K.E. and Summerfelt,K.R. (1990) Improved binary vectors for

Agrobacterium-mediated plant transformation. *Plant Mol. Biol.*, **14**, 269–276.

52. Hamilton,C.M. (1997) A binary-BAC system for plant transformation with high-molecular-weight DNA. *Gene*, **200**, 107–116.

53. Liu,Y.G., Shirano,Y., Fukaki,H., Yanai,Y., Tasaka,M., Tabata,S. and Shibata,D. (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 6535–6540.

54. Pasin,F., Bedoya,L.C., Bernabé-Orts,J.M., Gallo,A., Simón-Mateo,C., Orzaez,D. and García,J.A. (2017) Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synth. Biol.*, **6**, 1962–1968.

55. Silva,F., Queiroz,J.A. and Domingues,F.C. (2012) Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnol. Adv.*, **30**, 691–708.

56. Moore,S.J., Lai,H.E., Kelwick,R.J.R., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.*, **5**, 1059–1069.

57. Bi,X. and Liu,L.F. (1996) DNA rearrangement mediated by inverted repeats. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 819–823.

58. Lee,S., Su,G., Lasserre,E., Aghazadeh,M.A. and Murai,N. (2012) Small high-yielding binary Ti vectors pLSU with co-directional replicons for Agrobacterium tumefaciens-mediated transformation of higher plants. *Plant Sci.*, **187**, 49–58.

59. Jones,H.D., Doherty,A. and Sparks,C.A. (2009) Transient Transformation of Plants BT - Plant Genomics: Methods and Protocols. In Gustafson,J.P., Langridge,P., Somers,D.J. (eds). Humana Press, Totowa, NJ, pp. 131–152.

60. Mattick,C.S. (2018) Cellular agriculture: The coming revolution in food production. *Bull. At. Sci.*, **74**, 32–35.

61. Nordlund,E., Lille,M., Silventoinen,P., Nygren,H., Seppänen-Laakso,T., Mikkelson,A., Aura,A.-M., Heiniö,R.-L., Nohynek,L., Puupponen-Pimiä,R., *et al.* (2018) Plant cells as food – A concept taking shape. *Food Res. Int.*, **107**, 297–305.

62. Eibl,R., Meier,P., Stutz,I., Schildberger,D., Hühn,T. and Eibl,D. (2018) Plant cell culture technology in the cosmetics and food industries: current state and future trends. *Appl. Microbiol. Biotechnol.*, **102**, 8661–8675.

63. Ochoa-Villarreal,M., Howat,S., Jang,M.O., Kim,I.S., Jin,Y.W., Lee,E.K. and Loake,G.J. (2015) Cambial meristematic cells: A platform for the production of plant natural products. *N. Biotechnol.*, 10.1016/j.nbt.2015.02.003.

64. Appelhagen,I., Wulff-Vester,A.K., Wendell,M., Hvoslef-Eide,A.K., Russell,J., Oertel,A., Martens,S., Mock,H.P., Martin,C. and Matros,A. (2018) Colour bio-factories: Towards scale-up production of anthocyanins in plant cell cultures. *Metab. Eng.*, **48**, 218–232.

65. Tao,L., Jackson,R.E. and Cheng,Q. (2005) Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metab. Eng.*, **7**, 10–17.

66. Lee,S., Su,G., Lasserre,E., Aghazadeh,M.A. and Murai,N. (2012) Small high-yielding binary Ti vectors pLSU with co-directional replicons for Agrobacterium tumefaciens-mediated transformation of higher plants. *Plant Sci.*, **187**, 49–58.

67. Bartosik,D., Baj,J., Sochacka,M., Piechucka,E. and Wlodarczyk,M. (2002) Molecular characterization of functional modules of plasmid pWKS1 of Paracoccus pantotrophus DSM 11072. *Microbiology*, **148**, 2847–2856.

68. Challacombe,J.F., Pillai,S. and Kuske,C.R. (2017) Shared features of cryptic plasmids from environmental and pathogenic Francisella species. *PLoS One*, **12**, e0183554.

69. Zaleski,P., Wawrzyniak,P., Sobolewska,A., Łukasiewicz,N., Baran,P., Romańczuk,K., Daniszewska,K., Kierył,P., Płucienniczak,G. and Płucienniczak,A. (2015) pIGWZ12 – A cryptic plasmid with a modular structure. *Plasmid*, **79**, 37–47.

70. Menges,M. and Murray,J.A.H. (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. *Plant J.*, **30**, 203–212.

71. Guénin,S., Mauriat,M., Pelloux,J., Van Wuytswinkel,O., Bellini,C. and Gutierrez,L. (2009) Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J. Exp. Bot.*, **60**, 487–493.

72. Harada,J.J. (2001) Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *J. Plant Physiol.*, **158**, 405–409.

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

74. Han,Y.-J., Kim,Y.-M., Hwang,O.-J. and Kim,J.-I. (2015) Characterization of a small constitutive promoter from Arabidopsis translationally controlled tumor protein (AtTCTP) gene for plant transformation. *Plant Cell Rep.*, **34**, 265–275.

75. Nagaya,S., Kawamura,K., Shinmyo,A. and Kato,K. (2010) The HSP terminator of arabidopsis thaliana increases gene expression in plant cells.

Plant Cell Physiol., **51**, 328–332.

76. Engler,C., Youles,M., Gruetzner,R., Ehnert,T.M., Werner,S., Jones,J.D.G., Patron,N.J. and Marillonnet,S. (2014) A Golden Gate modular cloning toolbox for plants. *ACS Synth. Biol.*, **3**, 839–843.

77. Benfey,P.N. and Chua,N.H. (1990) The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. *Science*, **250**, 959–966.

78. Ishige,F., Takaichi,M., Foster,R., Chua,N.-H. and Oeda,K. (1999) A G-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants. *Plant J.*, **18**, 443–448.

79. Schlücking,K., Edel,K.H., Köster,P., Drerup,M.M., Eckert,C., Steinhorst,L., Waadt,R., Batistič,O. and Kudla,J. (2013) A New β-Estradiol-Inducible Vector Set that Facilitates Easy Construction and Efficient Expression of Transgenes Reveals CBL3-Dependent Cytoplasm to Tonoplast Translocation of CIPK5. *Mol. Plant*, **6**, 1814–1829.

80. Craft,J., Samalova,M., Baroux,C., Townley,H., Martinez,A., Jepson,I., Tsiantis,M. and Moore,I. (2005) New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *Plant J.*, **41**, 899–918.

81. Lazo,G.R., Stein,P.A. and Ludwig,R.A. (1991) A DNA Transformation–Competent Arabidopsis Genomic Library in Agrobacterium. *Bio/Technology*, **9**, 963–967.

82. Tao,L., Jackson,R.E. and Cheng,Q. (2005) Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metab. Eng.*, **7**, 10–17.

83. Wroblewski,T., Tomczak,A. and Michelmore,R. (2005) Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnol. J.*, **3**, 259–273.

84. Slater,S.C., Goodner,B.W., Setubal,J.C., Goldman,B.S., Wood,D.W. and Nester,E.W. (2008) The Agrobacterium Tumefaciens C58 Genome BT - Agrobacterium: From Biology to Biotechnology. In Tzfira,T., Citovsky,V. (eds). Springer New York, New York, NY, pp. 149–181.

85. Van Larebeke,N., Engler,G., Holsters,M., Van Den Elsacker,S., Zaenen,I., Schilperoort,R.A. and Schell,J. (1974) Large plasmid in agrobacterium tumefaciens essential for crown gall-inducing ability. *Nature*, **252**, 169–170.

86. Koncz,C. and Schell,J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol. Gen. Genet. MGG*, **204**, 383–396.

87. Van Larebeke,N., Engler,G., Holsters,M., Van Den Elsacker,S., Zaenen,I., Schilperoort,R.A. and Schell,J. (1974) Large plasmid in agrobacterium tumefaciens essential for crown gall-inducing ability. *Nature*, **252**, 169–170.

88. Hamilton,C.M. (1997) A binary-BAC system for plant transformation with high-molecular-weight DNA. *Gene*, **200**, 107–116.

89. Hoekema,A., Hirsch,P.R., Hooykaas,P.J.J. and Schilperoort,R.A. (1983) A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. *Nature*, **303**, 179–180.

90. Ghedira,R., De Buck,S., Nolf,J. and Depicker,A. (2013) The efficiency of Arabidopsis thaliana floral dip transformation is determined not only by the Agrobacterium strain used but also by the physiology and the ecotype of the dipped plant. *Mol. Plant. Microbe. Interact.*, **26**, 823–832.

91. Wu,H.-Y., Liu,K.-H., Wang,Y.-C., Wu,J.-F., Chiu,W.-L., Chen,C.-Y., Wu,S.-H., Sheen,J. and Lai,E.-M. (2014) AGROBEST: an efficient Agrobacterium-mediated transient expression method for versatile gene function analyses in Arabidopsis seedlings. *Plant Methods*, **10**, 19.

92. Tait,R.C., Close,T.J., Rodriguez,R.L. and Kado,C.I. (1982) Isolation of the origin of replication of the IncW-group plasmid pSa. *Gene*, **20**, 39–49.

93. Antoine,R. and Locht,C. (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from gram-positive organisms. *Mol. Microbiol.*, **6**, 1785–1799.

94. Jiang,P., Zhang,K., Ding,Z., He,Q., Li,W., Zhu,S., Cheng,W., Zhang,K. and Li,K. (2018) Characterization of a strong and constitutive promoter from the Arabidopsis serine carboxypeptidase-like gene AtSCPL30 as a potential tool for crop transgenic breeding. *BMC Biotechnol.*, **18**, 59.

95. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palaci,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiol.*, **162**, 1618–1631.

96. Yanagisawa,S. (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant J.*, **21**, 281–288.

97. Biłas,R., Szafran,K., Hnatuszko-Konka,K. and Kononowicz,A.K. (2016) Cis-regulatory elements used to control gene expression in plants. *Plant Cell. Tissue Organ Cult.*, **127**, 269–287.

98. Ansari,A. and Hampsey,M. (2005) A role for the CPF 3′ -end processing machinery in RNAP II-dependent gene looping. *Genes Dev.*, **19**, 2969–2978.

99. Calvo,O. and Manley,J.L. (2003) Strange bedfellows: Polyadenylation factors at the promoter. *Genes Dev.*, **17**, 1321–1327.

100. Hampsey,M., Singh,B.N., Ansari,A., Lainé,J.P. and Krishnamurthy,S. (2011) Control of eukaryotic gene expression: Gene loops and transcriptional memory. *Adv. Enzyme Regul.*, 10.1016/j.advenzreg.2010.10.001.

101. Moabbi,A.M., Agarwal,N., El Kaderi,B. and Ansari,A. (2012) Role for gene looping in intron-mediated enhancement of transcription. *Proc. Natl. Acad. Sci.*, **109**, 8505 LP – 8510.

102. Tan-Wong,S.M., Zaugg,J.B., Camblong,J., Xu,Z., Zhang,D.W., Mischo,H.E., Ansari,A.Z., Luscombe,N.M., Steinmetz,L.M. and Proudfoot,N.J. (2012) Gene loops enhance transcriptional directionality. *Science*, **338**, 671–675.

103. Al Husini,N., Kudla,P. and Ansari,A. (2013) A Role for CF1A 3′ End Processing Complex in Promoter-Associated Transcription. *PLOS Genet.*, **9**, e1003722.

104. Mukundan,B. and Ansari,A. (2013) Srb5/Med18-mediated termination of transcription is dependent on gene looping. *J. Biol. Chem.*, **288**, 11384–11394.

105. Medler,S. and Ansari,A. (2015) Gene looping facilitates TFIIH kinase-mediated termination of transcription. *Sci. Rep.*, **5**, 12586.

106. Pérez-González,A. and Caro,E. (2018) Effect of transcription terminator usage on the establishment of transgene transcriptional gene silencing. *BMC Res. Notes*, **11**, 511.

107. Beyene,G., Buenrostro-Nava,M.T., Damaj,M.B., Gao,S.J., Molina,J. and Mirkov,E.E. (2011) Unprecedented enhancement of transient gene expression from minimal cassettes using a double terminator. *Plant Cell Rep.*, **30**, 13–25.

108. Yamamoto,T., Hoshikawa,K., Ezura,K., Okazawa,R., Fujita,S., Takaoka,M., Mason,H.S., Ezura,H. and Miura,K. (2018) Improvement of the transient expression system for production of recombinant proteins in plants. *Sci. Rep.*, **8**, 4755.

109. Diamos,A.G. and Mason,H.S. (2018) Chimeric 3’ flanking regions strongly enhance gene expression in plants. *Plant Biotechnol. J.*, **16**, 1971–1982.

110. McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3’ termini. *Nucleic Acids Res.*, **13**, 1347–1368.

111. MacNicol,M.C., Cragle,C.E. and MacNicol,A.M. (2011) Context-dependent regulation of Musashi-mediated mRNA translation and cell cycle regulation.

Cell Cycle, **10**, 39–44.

112. Chupeau,M.-C., Granier,F., Pichon,O., Renou,J.-P., Gaudin,V. and Chupeau,Y. (2013) Characterization of the Early Events Leading to Totipotency in an Arabidopsis Protoplast Liquid Culture by Temporal Transcript Profiling. *Plant Cell*, **25**, 2444–2463.

113. Borghi,L. (2010) Inducible Gene Expression Systems for Plants. *Methods Mol. Biol.*, **655**, 65–75.

114. Chung,C.T. and Miller,R.H. (1993) Preparation and storage of competent Escherichia coli cells. *Methods Enzymol.*, **218**, 621–627.

115. Lescot,M., Dehais,P., Thijs,G., Marchal,K., Moreau,Y., Van de Peer,Y., Rouze,P. and Rombauts,S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, **30**, 325–327.

116. Higo,K., Ugawa,Y., Iwamoto,M. and Korenaga,T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.*, **27**, 297–300.

117. Menges,M. and Murray,J.A.H. (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. *Plant J.*, **30**, 203–212.

118. Oda,Y. (2017) VND6-induced Xylem Cell Differentiation in Arabidopsis Cell Cultures. *Methods Mol. Biol.*, **1544**, 67–73.

119. Chupeau,M.-C., Granier,F., Pichon,O., Renou,J.-P., Gaudin,V. and Chupeau,Y. (2013) Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. *Plant Cell*, **25**, 2444–63.

120. Faraco,M., Di Sansebastiano,G. Pietro, Spelt,K., Koes,R.E. and Quattrocchio,F.M. (2011) One Protoplast Is Not the Other! *Plant Physiol.*, **156**, 474 LP – 478.

121. Mauri,M., Vecchione,S. and Fritz,G. (2019) Deconvolution of Luminescence Cross-Talk in High-Throughput Gene Expression Profiling. *ACS Synth. Biol.*, **8**, 1361–1370.

Chapter 4 Guided differentiation system for cultured plant cells

4.1. Introduction

Plants are excellent sources of secondary metabolites which are highly attractive to the food and drug industries. For example, the anti-cancer drug paclitaxel, and the malaria treatment drug artemisinin are naturally made in the yew tree and sweet wormwood respectively, and each has an estimated market size of billions of dollars annually (1, 2). Nevertheless, the natural supply is not adequate to sustainably meet ever-growing demands. A well-established strategy to produce plant bio-compounds is the engineering of the plant biosynthetic pathways in bacteria or yeast (3–5). A recent new success story was the complete biosynthesis of cannabinoids in yeast (6). However, the transfer of plant biosynthetic pathways into microbial chassis has its shortcomings. Typically, not all genes involved in complex biosynthetic pathways are known, in some cases, post-translational modifications in microbes are not adequate to produce fully functional proteins, and plant enzymes often exhibit low activity in microbial chassis (7–9).

Plant cell cultures can be cultivated like bacterial or yeast cell cultures, yet offer a unique set of advantages over the microbial systems because of their plant origin (10–12). For example, they are more likely to have endogenous molecular toolsets for the biosynthesis of macromolecules and metabolites that are naturally made in plants, and possess the capacity for intricate post-transcriptional modifications such as protein glycosylation. Also, they have low susceptibility to human pathogens, and thus plant systems make safe biofactories for pharmaceutical compounds. Production using plant cell cultures is traditionally based on extensive cell line screening without intensive genetic modifications and breeding efforts, which may be the reasons why plant cell cultures have not yet been proven as successful biofactories, despite their promise.

Plant secondary metabolism is closely connected with cellular differentiation; differentiated cells are generally more competent in secondary metabolite synthesis (13, 14). Thus, the undifferentiated state of cell cultures likely suppresses their competency. On the other hand, cell proliferation rate, a crucial parameter to scale the bioproduction in industrial biotechnology, is tightly linked to the cells' undifferentiated status (13). Separating the stages of growth and production is, therefore, an essential requirement for biomanufacturing with cell cultures. Currently, the production of metabolites in plant cells is controlled using costly chemicals (e.g. hormones or elicitors) in the culture medium (14, 15). In regenerative medicine, vigorous research on the regulation of cell fate specification and differentiation led to the development of systems that can transform stem cells into specific specialised cell types, intending to create human tissues and organs (16, 17). However, in the plant field, such systems are only starting to emerge.

Transcription factors orchestrate cellular events by activating or suppressing an entire network of downstream genes, therefore acting as molecular switches (18). Transcription factors are also key players in cell fate specification and maintenance, and play a pivotal role in the differentiation of stem cells into specific cell types (19). For example, VND6 and VND7, a pair of closely related NAC-domain containing transcription factors, can induce differentiation of plant cells into metaxylem and protoxylem vessels in Arabidopsis cell suspensions and poplar leaves (20). Xylem vessels are elongated cells with a thick, lignified cell wall, responsible for water transport. Protoxylem develops before the plant organ has completed its growth, while the metaxylem develops after. To achieve cell wall lignification, immature xylem vessels activate the phenylpropanoid pathway for lignin synthesis. Lignin is a class of organic polymers responsible for the support tissues of vascular plants. The biosynthesis begins with phenylalanine, which is sequentially converted into cinnamic acid, *p*-coumaric acid and *p*-

coumaroyl-CoA (21). *P*-coumaroyl-CoA is the common intermediate for lignins, flavonoids, stilbenoids, coumarins and phenolic acids.

Transcription factors regulate secondary metabolite biosynthesis in plants (22, 23). For example, MYB75 is a sufficient positive regulator of anthocyanin production in Arabidopsis (24). Anthocyanins are important natural compounds as they have diverse pharmacological and industrial applications (1). Expression of MYB75 diverts the phenylpropanoid synthesis from lignin to flavonoids, more specifically to anthocyanins. MYB75 negatively regulates monolignol biosynthesis, upregulates the expression of anthocyanin biosynthetic genes and further elevates the expression of enzyme genes involved in *p*-coumaroyl-CoA synthesis (21). Transcription factors are, therefore, perfect candidate genes to target in manipulating cellular decision-making and secondary metabolism production.

Different plant cell types have the competency to produce specific bio-compounds in high concentrations. For example, immature xylem vessels and fibre cells have enhanced activity in the phenolic biosynthetic pathways (en route to producing lignin), and thus are good candidates for enhancing phenylpropanoid production (25, 26). The microenvironment in differentiated cells determines the transcriptional activity of genes that encode for enzymes related to the physiology and function of the specific cell type.

Taking these factors into consideration, a new approach for biomanufacturing in plant cell cultures was designed. It is based on a sequential cell differentiation system, in which cells are guided to differentiate into specialised cell types with identity-specific competency to produce specific bio-compounds. The approach is based on the sequential induction of key transcription factors which will control cell differentiation and bio-compound synthesis. The expression of the transcription factors should be activated at different time points and is therefore driven by multiple inducible systems which are combined in the same construct to ensure simultaneous delivery into plant cells.

There are three inducible systems commonly used in plant systems, which are all activated by membrane-permeable chemicals, i.e. β-estradiol, dexamethasone (Dex), and ethanol. These three systems have a common architecture consisting of two transcriptional units. One of them encodes a transactivator, a fusion transcription factor that contains a DNA binding, and a transcriptional activation domain. A transactivator binds to its binding sequence (cis-element) only when the inducer chemical is present. The transactivator is often expressed constitutively. The second transcriptional unit is driven by the inducible promoter containing the basal promoter sequence with the multimer repeats of the cis-element, which the trans-activator recognises. The β-estradiol inducible promoter is comprised of eight repeats of the lexA operator elements fused with the minimal 35S CMV promoter, and its transactivator (**XVE**) is the fusion among the DNA-binding domain of the bacterial repressor LexA (**X**), the activation domain VP16 (**V**), and the carboxyl region of the human estrogen receptor (**E**R) (27). The Dex inducible promoter consists of six copies of the Lac operator fused upstream of the minimal 35S CMV promoter. Its transactivator (LhGR) consist of a lac repressor mutant, **L**acI^{his} (tyrosine 17 is replaced by histidine, resulting in higher affinity than wild-type lac repressor), the Gal4 transcription-activation-domain-II (**G**) and the ligand-binding domain of a rat glucocorticoid receptor (G**R**) (28, 29). In both cases, the binding of the hormone to the receptor allows the transactivator to enter the nucleus and bind the inducible promoter. In the ethanol inducible system, the transcription factor AlcR from *Aspergillus nidulans* is activated by acetaldehyde, which is produced in plants by the ethanol metabolism, and the AlcR active form binds and activates the inducible promoter alcA (30, 31).

The current plant inducible systems have not undergone substantial improvements since they were first introduced. Their parts (e.g. promoters and terminators) were not systematically selected, and more importantly, their functions have not been tested in combination. A test was developed to verify the individual and combinatorial functionality of the three most common plant

inducible systems (estradiol, dexamethasone and ethanol inducible). Informed by these tests, a sequential cell differentiation system was built to guide the differentiation of the MM1 Arabidopsis suspension cell line into xylem vessels, which are further modulated to produce anthocyanins as a proof of concept. MM1 is a fast-growing cell line with a ploidy level of 6C DNA content, which was created in 2002 by Menges and Murray (32).

4.2. Results

4.2.1. Characterisation and crosstalk experiments for the plant inducible systems

In order to sequentially manipulate cell differentiation, a set of genes had to be controlled at different time points. Inducible systems that activate target gene expression upon exposure to a specific inducer chemical can serve this purpose. The initial plan was to create a triple inducible system to separate three different phases of cell fate, i.e. proliferation, differentiation and metabolite production (**Figure 1A**). The LhGR0-pOp6, sXVE-lexA and AlcR-alcA inducible systems would be fused in a single plant binary vector (**Figure 1B**). T-DNA transfer is possible with standard binary vectors as large as ~30 kb (33). Also, the single-vector multigene transfer can prevent the random incorporation of several T-DNAs, which could increase the probabilities for disruption of genes or sequence elements vital for plant growth, development and productivity (34).

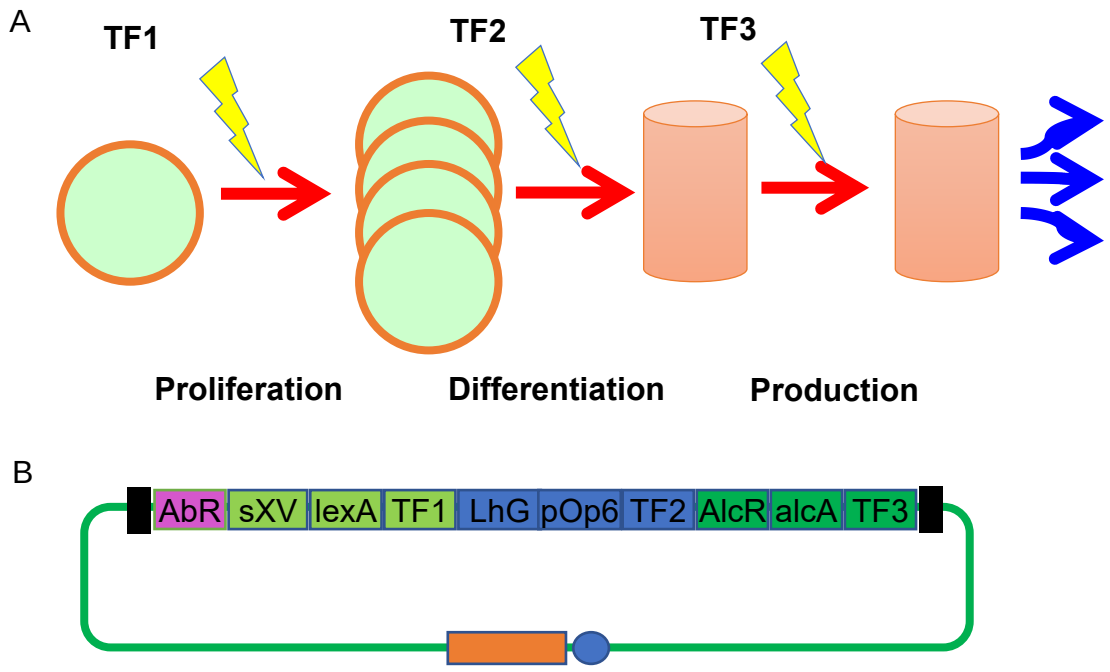


Figure 1. A. The concept design of the guided differentiation system of a plant cell culture biofactory. The designed guided differentiation system will control three phases of the plant cells. The first one is the proliferation, which is a desirable parameter in the bioreactors. The second step is the cell fate specification. Different cell types have the potency to produce some secondary metabolites in high yields. The last step is the stimulation of the differentiated cells to produce biocompounds. All the processes will be controlled by the expression of key transcription factors. **B. Triple inducible system.** It is a combination of three plant inducible systems (sXVE-lexA, LhGR-pOp6 and AlcR-alcA) in the same binary vector for the sequential expression of genes at different time points.

The first step towards the generation of the sequential inducible system was to clone all the transactivators and inducible promoters into mUAV used for Phytobricks. The inducible system was then reconstructed with the Mobius Assembly. Firstly, the same promoter and terminator for all three systems were directly compared with their activation amplitude in the transient transformation of Arabidopsis leaf protoplasts. Estradiol and Dex inducible systems exhibited similar activation amplitude (~40 RLU) and a minimal basal (un-induced) expression (**Figure 2**). Both systems were activated with a low concentration of inducers, 0.5 μM for β-estradiol and 2.5 μM for dexamethasone, and they could not induce more for higher concentrations. On the other hand, the ethanol

inducible system had a high expression even in non-induced samples and was not able to show substantial activation. For this reason, a synthetic ethanol promoter was devised as described by Pasin *et al*. (35). Even though the combination of the AlcR with the synthetic ethanol promoter was driving high expression levels, it was leaky (**Chapter 3, Figure 8C**). Consequently, a double differentiation system was built with a focus on cell differentiation and compound production.

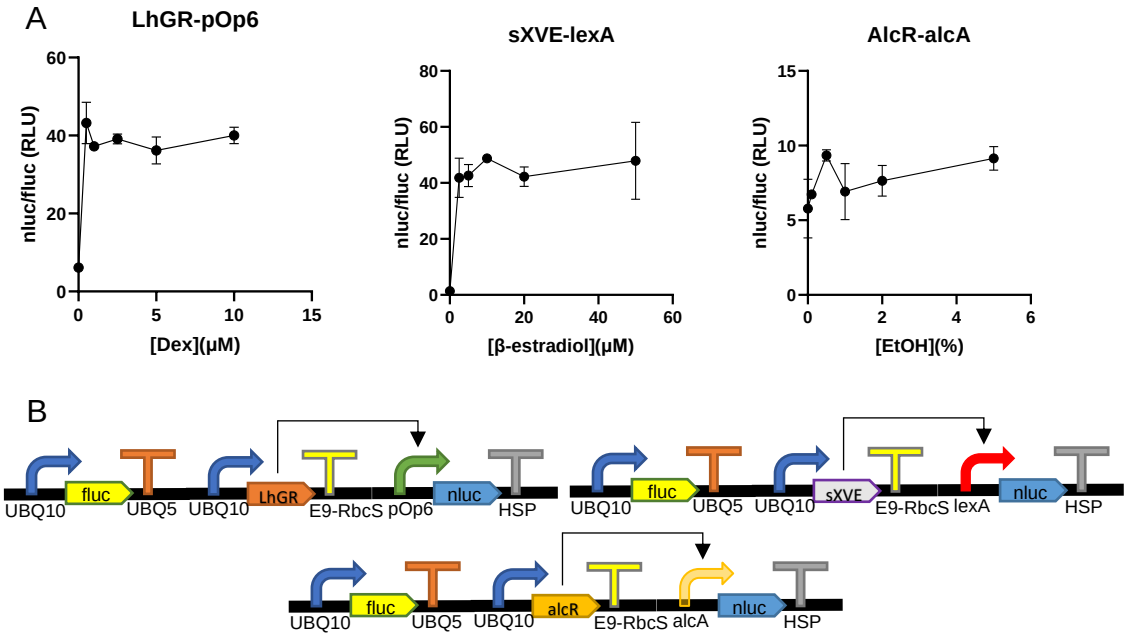


Figure 2. Induction curves of estradiol, Dex and ethanol inducible systems. A. The induction curves were generated by the normalised expression of Nano luciferase, against increasing concentration of the inducer. The selected concentrations were based on bibliographic data. The Dex and estradiol systems exhibited an activation upon addition of 0.5 μM and 2.5 μM of their cognate inducers respectively, reaching saturation at ~40 RLU. The ethanol system did not show substantial activation. Nluc= Nano and Fluc=Firefly luciferase activity, bars show normalised luciferase activity values in mean±SE. **B.** The induction assay was carried out in transiently transformed Arabidopsis leaf protoplasts with the following constructs: UBQ10:fluc:UBQ5-UBQ10:LhGR:E9-RbcS-pOp6-35S:nluc:HSP, UBQ10:fluc:UBQ5-UBQ10:sXVE:E9-RbcS:lexA-35S:nluc:HSP and UBQ10:fluc:UBQ5-UBQ10:AlcR:E9RbcS-alcA-35S:nluc:HSP.

Since in the multiple-inducible system, three individual inducible systems will coexist and co-function, it was essential to demonstrate that they do not interfere with one another. First, it was tested if the inducer molecules have an affinity to,

and activate their cognate receptors by adding β-estradiol, Dex or ethanol to the other inducible systems. It was found that β-estradiol activates the sXVE-lexA inducible system and Dex the LhGR-pOp6 inducible system mainly specifically (**Figure 3**). Ethanol had no impact on both inducible systems, and Dex could not affect the estradiol system. β-Estradiol gave a slightly higher value than the basal activity of the Dex system (9.8 RLU over 6.1 RLU), which is negligible compared to the activation level by Dex (39.1 RLU). Next, it was tested whether the transactivators act specifically to their cognate inducible promoters. All the possible combinations between the inducible promoters and the transactivators were cloned into a single construct. In each combination, the induction was carried out by the cognate inducer (ethanol for AlcR, β-estradiol for sXVE and Dex for LhGR). As shown in **Figure 4**, LhGR is specific for the pOp6 promoter and sXVE for lexA, and there is no crosstalk.

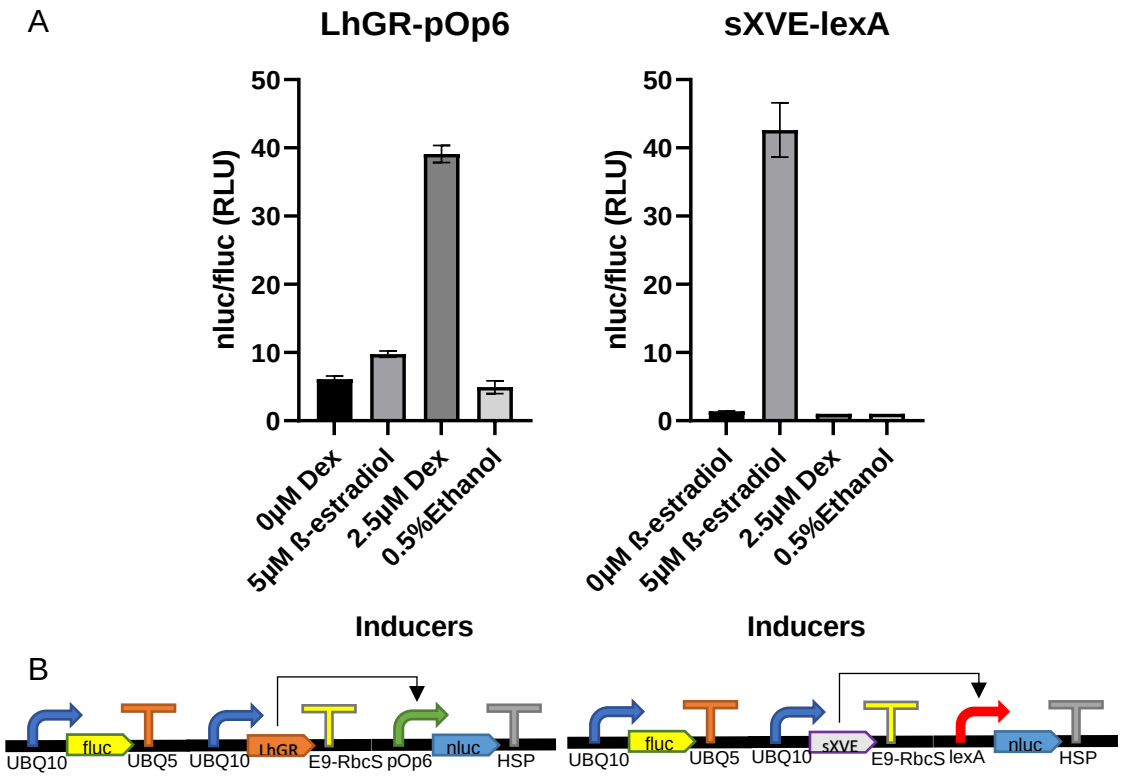


Figure 3. Effect of the inducer molecules on the inducible systems. A. Possible crosstalk of Dex, β-estradiol and ethanol was evaluated on the activation of Dex and

Estradiol systems. Both systems were fully activated only by their cognate inducers. Nluc= Nano and Fluc=Firefly luciferase activity, bars show normalised luciferase activity values in mean±SE. **B.** The assay was carried out in transiently transformed Arabidopsis protoplasts with the following constructs: UBQ10:fluc:UBQ5-UBQ10:LhGR:E9-RbcS-pOp6-35S:nluc:HSP and UBQ10:fluc:UBQ5-UBQ10:sXVE:E9RbcS-lexA-35S:nluc:HSP.

The influence of the transactivator expression on the inducible system activity was examined. To control the expression levels of the transactivator, three constructs with the same promoter UBQ10 and different terminators were built, i.e. one strong (HSP), one medium (NDUFA8) and one weak (E9-RbcS). Low expression of the transactivator had a high impact on the inducible system's activity, lowering the expression levels more than four times (**Figure 5**). On the other hand, medium and high transactivator expression led to a similar activity of the estradiol inducible system, suggesting that beyond a threshold the concentration of the transactivator does not affect the efficiency of the inducible system.

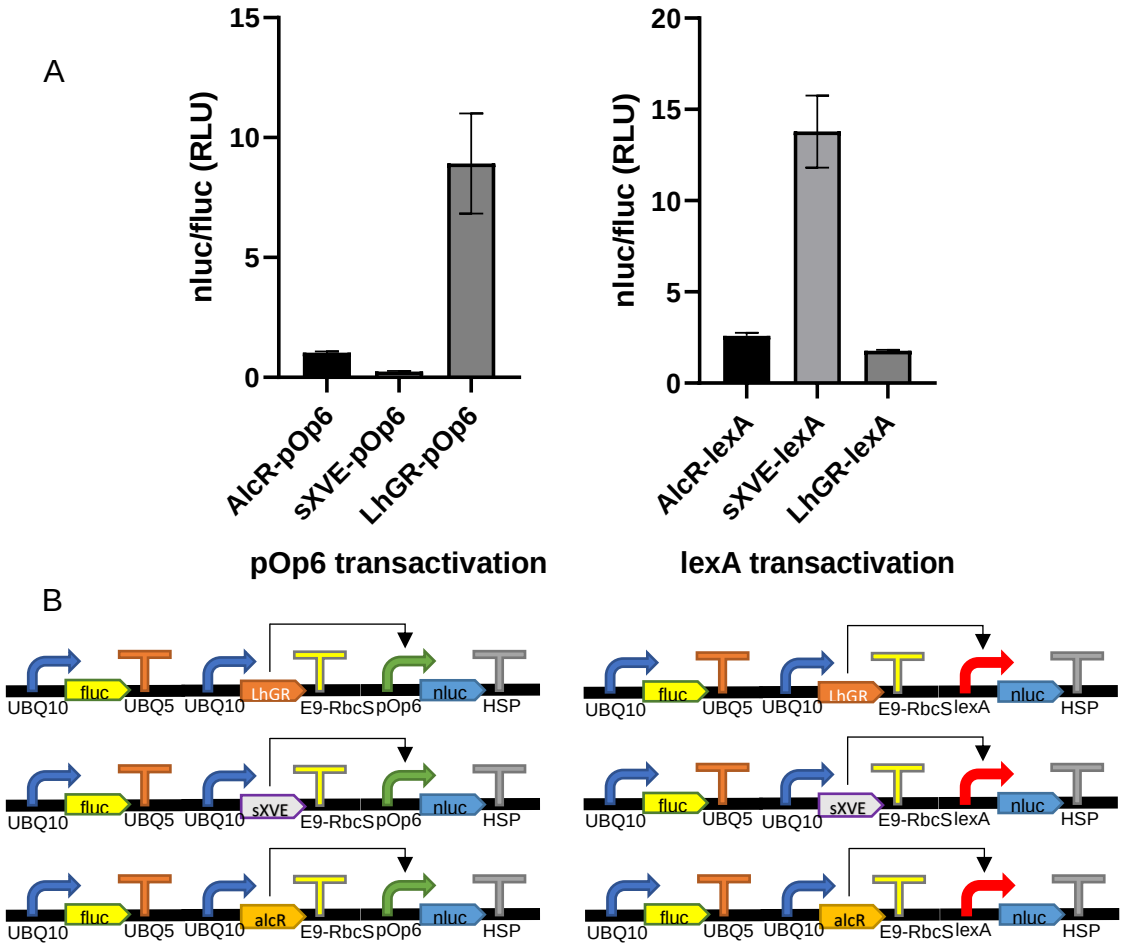


Figure 4. Crosstalk test between the transactivators and inducible promoters. A. It was tested whether AlcR, sXVE and LhGR can bind and activate pOp6 and lexA inducible promoters. The activation each time was triggered by the cognate inducer of the transactivator. No crosstalk was found as activation of the inducible promoter was only possible with the corresponding transactivator (LhGR for pOp6 and sXVE for lexA). Nluc= Nano and Fluc=Firefly luciferase activity, bars show normalised luciferase activity values in mean±SE. **B.** The assay was carried out in transiently transformed Arabidopsis protoplasts with the following constructs: UBQ10:fluc:UBQ5-UBQ10:Transactivator:E9-RbcS-pOp6-35S:HSP and UBQ10:fluc:UBQ5-UBQ10:Transactivator-E9-RbcS:lexA-35S:HSP (Transactivator=AlcR, sXVE and LhGR).

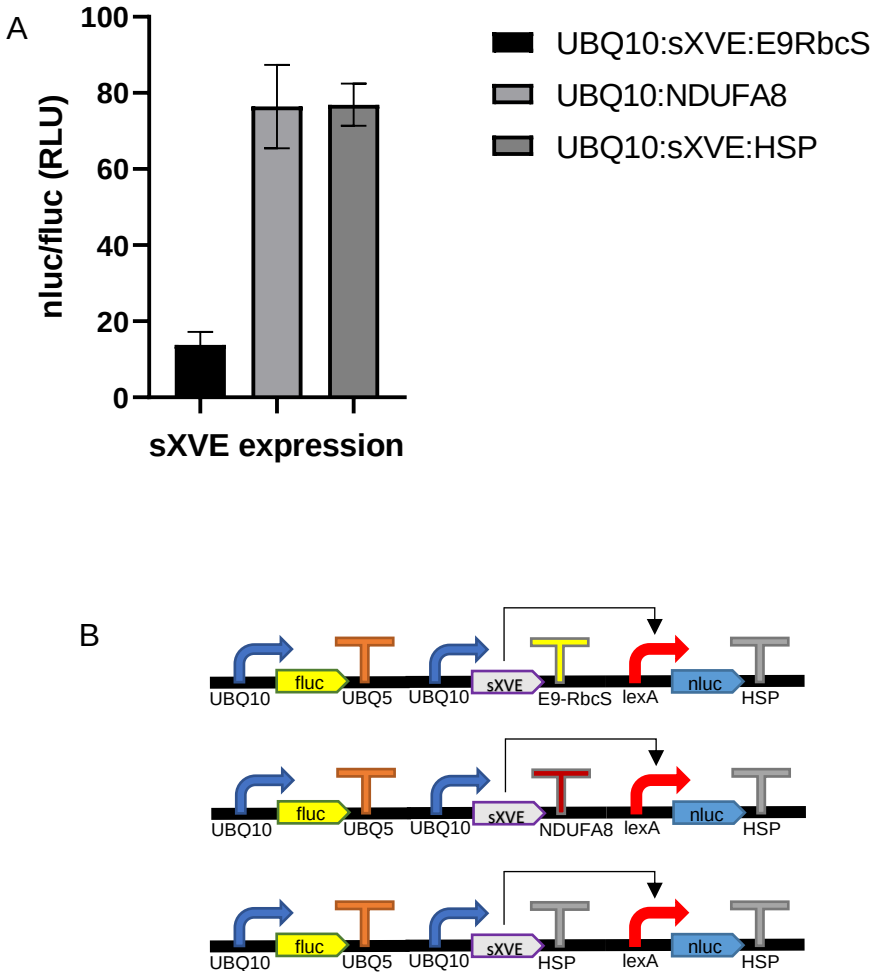


Figure 5. The effect of the transactivator expression on the activity of the estradiol inducible system. A. The expression levels of the sXVE were varied by changing its terminator, and it was tested how it affects the expression activity of the inducible system. E9 RbcS was used for low, NDUFA8 for medium and HSP for high expression. Medium and high expression of sXVE resulted in an inducible activity maximum of ~80 RLU. On the contrary, low sXVE expression reduced the inducible activity by a four-fold. Nluc= Nano and Fluc=Firefly luciferase activity, bars show normalised luciferase activity values

in mean±SE. **B.** The assay was carried out in transiently transformed Arabidopsis protoplasts with the following constructs: UBQ10:fluc:UBQ5-UBQ10:sXVE:Terminator-lexA-35S:HSP, where the terminator is: E9 RbcS, NDUFA8 or HSP.

4.2.2. Selection of the key transcription factors

Transcription factors (TFs) are key instructive components of the guided plant cell differentiation platform. They are perfect candidate genes for manipulating cellular behaviours, decision-making and stimulation of plant secondary metabolism. Suitable transcription factors were identified in the literature to be implemented in the plant cell lines, with focus on overexpression phenotypes (i.e. if the expression of the gene was sufficient to cause ectopic conversion of the cell differentiation or metabolic activity). **Table1** summarises the selection of TFs that can be classified into three categories, which correspond to three different timings of activation.

The first category of TFs was related to cell proliferation and dedifferentiation, which are important parameters for the production in the bioreactors. Homeo-domain containing TFs were selected from the WUSCHEL family, which play roles in the maintenance and proliferation of different meristems (stem cell niches) in the plant root and shoot, and are therefore considered stem cell markers.

The second category was about cell differentiation. TFs were chosen, which are considered master regulators of cell types with enhanced phenylpropanoid metabolism, such as xylem vessels and fibres (36). TFs have also selected that convert to cell types with high protein content, which can be used to produce antibodies and pharmaceutical peptides. Examples of these cell types are synergid and tapetum cells. Synergid cells are two specialised cells that control several steps of the angiosperm fertilisation process and are enriched in ribosomes, mitochondria, endoplasmic reticulum and Golgi stacks (37). Tapetum cells are specialised nutritive cells, which have more than one nucleus and help in the growth and development of the pollen grains (38).

For the production phase, two transcription factors were identified, MYB75, which promotes anthocyanin synthesis (39) and MYB12, which is a flavonol-specific activator of flavonoid biosynthesis (40). For the proof of concept experiments, a combination of VND6 transcription factor and MYB75 were selected. All selected TFs were PCR amplified from Arabidopsis cDNA and cloned into mUAV to form Phytobricks.

TFs related to cell proliferation/dedifferentiation			
	TAIR ID	**Isolation**	**Function**
WUSCHEL (WUS)	**AT2G17950**	Stem	Regulates the maintenance of shoot apical meristem (41)
WUSCHEL RELATED HOMEOBOX 4 (WOX4)	**AT1G46480**	Stem	Promotes vascular cambium activity, especially cell division (42)
WUSCHEL RELATED HOMEOBOX 5 (WOX5)	**AT3G11260**	Roots	Maintains root apical meristem activity (43)
TFs related to cell fate specification/differentiation			
VASCULAR-RELATED NAC-DOMAIN 6 (VND6)	**AT5G62380**	Stem	Plays a crucial role in cell specification into xylem vessels (44)
SECONDARY WALL-ASSOCIATED NAC DOMAIN 1 (SND1)	**AT1G32770**	Stem	Differentiates non-vascular cells into fibre cells (36)
NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1 (NST1)	**AT2G46770**	Stem	Differentiates non-vascular cells into fibre cells (36)

TAPETUM DETERMINANT 1 (TPD1)	**AT4G24972**	Flowers	Function in tapetum cell differentiation (45, 46)
DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1 (TDF1)	**AT3G28470**	Flowers	Function in tapetum cell differentiation (45, 46)
MYB98	**AT4G18770**	Flowers	Required for synergid cell differentiation (47)
TFs related to cell metabolism			
MYB75	**AT1G56650**	Stem	Promotes anthocyanin synthesis in Arabidopsis (21)
MYB12	**AT2G47460**	Stem	Activates phenylpropanoid biosynthesis (40)

Table 1. Transcription factors for the guided differentiation system. Three categories of TFs were selected: one for cell proliferation, one for cell differentiation and one related to secondary metabolism.

4.2.3. Sequential differentiation system development

The two-step sequential differentiation system was comprised of six transcriptional units (**Figure 6**). The first TU was the plant selection marker. The *Streptomyces hygroscopicus* bar gene, encoding for phosphinothricin acetyltransferase, which confers resistance to bialophos or phosphinothricin (glufosinate ammonium), was used as a plant selection marker. For the second TU, the Nano luciferase gene was selected as an indicator for the successful transformation of the plant cells. After every cell culture transformation, a luciferase assay was carried out to verify the successful Agrobacterium-mediated transformation of the cells. The third and fourth TUs were the estradiol inducible system expressing a TF directing the cell fate specification. The TF was fused to

a plant codon optimised sfGFP with a flexible (Gly)4Ser linker. Flexible linkers provide a certain degree of movement between two fusion proteins to maintain their bioactivities (48). The last two TUs were the Dex inducible system driving the expression of a TF modulating the secondary metabolism or biosynthesis. (Gly)4Ser linked this TF to a plant codon-optimised fluorescent protein mTFP1. For the constructs expressing TF for protein expression cell lines, the Dex system was used for the induction of mTFP1.

sfGFP and mTFP1 were selected to monitor the expression of the TFs in the cells. They have high brightness levels and were chosen to be able to compensate above the autofluorescence of the MM1 cell line. Stressed MM1 cells were found to emit a wide range of fluorescent signals in blue, green and red. Additionally, mTFP1 was chosen as an indicator for the protein expression in the protein expression lines.

For all the TUs, different promoters and terminators were used to reduce the structural plasmid instabilities of the final construct. For the same reason, short sequence elements were selected to keep the size of the final constructs as small as possible. Initially, the final constructs were cloned in the pGreen backbone (49), which is a commonly used small-sized binary vector for plant systems. However, after the unsuccessful transformation of the MM1 cell line using pGreen, the newly available plant binary vector pLX (50) could be used successfully.

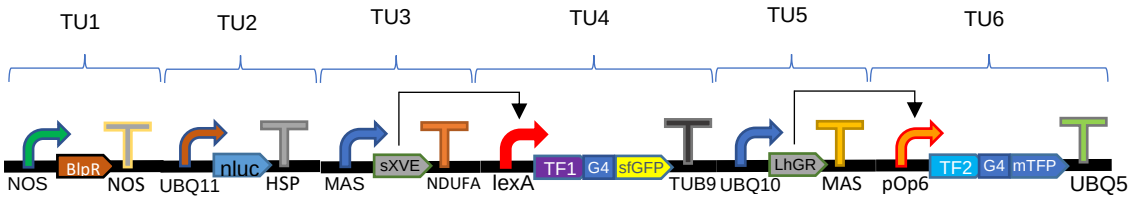


Figure 6. Structure of the constructs constitute the guided differentiation system. First, each TU was assembled in a Level 1 Acceptor vector: TU1, TU2, TU3 or TU4 in Mobius Assembly vectors A, B, Γ and Δ, respectively. TU5 and TU6 were assembled in vectors A and B, respectively. The final construct was built in a Level 2 Acceptor Vector A, harboring TU1-TU2-TU3-TU4-TU5-TU6. TU1 has the plant resistance cassette, TU2 a transformation screening marker (Nluc), TU3 and TU4 the estradiol inducible system driving TFs for cell differentiation, TU5 and TU6 the Dex inducible system driving the expression for TFs for secondary metabolism. TF1: VND6, SND1, NST3, TPD1, TDF1, MYB98 and TF2: MYB75, MYB12.

Transformation of MM1 cell line and transgene selection

Before transferring the guided differentiation system to plant cells, an efficient transformation protocol for plant cell cultures was needed. A re-culturing and herbicide selection was developed for the Arabidopsis MM1 cell line with a Nano luciferase expression construct over one month. The construct has a Basta (glufosinate-ammonium) resistance, and it was housed either in a pGreen or pLX vector and delivered to the MM1 cell line with Agrobacterium strain AGL1. Agrobacterium can transfer sequences flanked by two conserved 25 bp repeats, called T-DNA borders, from bacterial cells into the host plant cells (51).

The transformed cells were subcultured every seven days by diluting the cells 1:10 into a fresh culture medium, starting one week after the infection of the cells with AGL1. Two weeks after the transformation, the cells were split into two flasks, where one was supplemented with 15 µg/ml glufosinate-ammonium to select for stably transformed cells, and the other was kept without herbicide. There was a dependency of the transformation efficiencies on the types of binary vectors (**Figure 7**). One week after the transformation, the pLX construct showed a seven times higher signal than the pGreen construct. With the first subculture, the signal of pLX was reduced three times, while the signal of pGreen 6.7 times. After the second subculture, the pLX signal had a small reduction by 1.3 times even with the addition of herbicide, where the reduction was slightly higher (1.6 times).

The signals of the pGreen constructs were close to the signal of the control. At the fourth week, the signal of the cell lines transformed with the pLX construct

started to recover. More specifically, there were 2.7 times increase with the cell line using herbicide, and a 1.6 times increase for the cell line without selection. It was determined that the pLX is a more efficient vector background for the MM1 transformation.

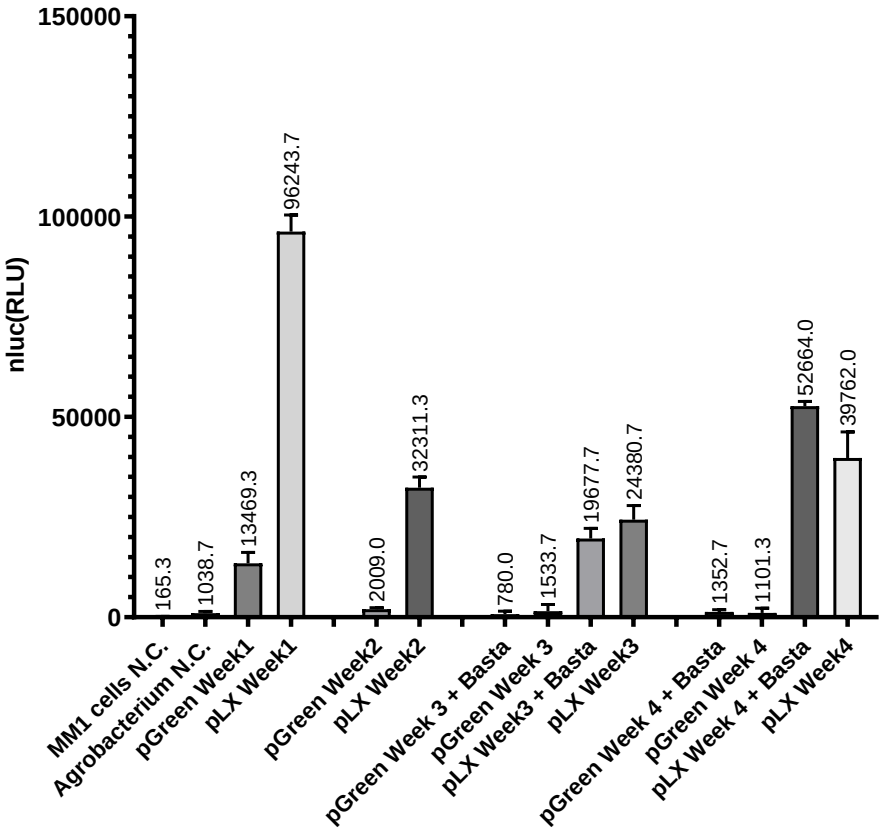


Figure 7. Transformation and herbicide selection of the MM1 Arabidopsis cell line. Arabidopsis cell line MM1 was transformed with the construct NOS:BglR:NOS-UBQ10:nluc:HSP housed either in pGreen or pLX vectors, followed by weekly subculture. The AGL1 strain was used for Agrobacterium-mediated transformation. Two weeks after the transformation, 15 µg/ml glufosinate-ammonium (Basta) was used to select for stably transformed cells, alongside non-selective conditions. Nano luciferase activity was an indicator for the transformation efficiencies and the propagation of the construct during the cell proliferation. Untransformed cells, and cells transformed with the AGL1 strain, carrying an empty vector, were used as controls.

4.2.4. Sequential differentiation experiments

The sequential differentiation experiments were conducted with the combination of the two TFs instructing the xylem vessel specification and anthocyanin biosynthesis, VND6 and MYB75. The construct was made in the pLX-based vector and transferred to the MM1 cell line with AGL1. After the transformation, the cells were treated with Dex, β-estradiol or a combination of both. Before induction of the TFs, the presence of the construct in the MM1 cell line was first verified with the luciferase assay.

The expression of the TFs (i.e. the fluorescent protein activity) and the morphology of the cells were examined by confocal microscopy. The blue channel was used to detect mTFP1 expression, which represents the level of MYB75, the green channel sfGFP expression, which reports the level of VND6, and the red channel the tracheary element development, which were stained with wheat germ agglutinin (WGA, Alexa Fluor™ 594 Conjugate). WGA binds to the secondary cell wall, more specifically to three β-(1-4)-linked *N*-acetyl-D-glucosamine residues, which are explicitly found in hemicellulosic components of xylem vessels or tracheary elements (52). As a control, untransformed cells and transformed but un-induced cells were used.

In the untransformed MM1 cell line, there were two populations of cells: the majority (~95%) were cells in which no signal was observed in any of the detection channels (**Figure 8)**, and stressed cells were observed, which produced a signal in all the channels (**Figure 9**). These two populations were also found among the transformed but uninduced cells, with a higher ratio of stressed cells than in the untransformed cells (**Figure 10**). In some cases, leaky expression of sfGFP was observed (**Figure 11**). Non-stressed, untransformed cells had small round dark spots equally distributed within the cell volume (**Figure 8**). On the contrary, stressed cells were irregularly shaped, had a concentrated dark area, and the whole cell was surrounded by a white zone (**Figure 10**). Transformed cells, which had been induced were changing into an oval shape, and the dark spots were

concentrated on the periphery of the cells (**Figure 12, Figure 14**). Colocalisation of the darks spots with the fluorescent signal indicates that they are the nucleus and the cytoplasm surrounded by the vacuole. Additionally, no secondary thickening was detected as there was no red signal bordering the cells. Xylem vessels have thick lignified cell walls, which would have been detected by WGA.

The cells treated with β-estradiol were found to express sfGFP; however, recognising the sfGFP signal was difficult due to the strong background autofluorescence in stressed cells. The sfGFP signal was monitored via the comparison of the blue, red and green channels and by counting signals only in the green channel (**Figure 12**). Even when the expression of VND6::sfGFP was verified, the cells did not completely differentiate into xylem vessels. A few xylem element-like cells were found in the induced samples, but they also showed a stress signal, and it was not possible to conclude whether they were transformed or not. mTFP1 would give a signal in the blue and green channels but not in the red channel (**Figure 13**). mTFP1 and sfGFP have overlapping areas in the emission and excitation spectrum. Simultaneous expression of mTFP1 and sfGFP was challenging to distinguish since it required the determination of differential signal intensity in the blue and green channels (**Figure 14 and Figure 15)**.

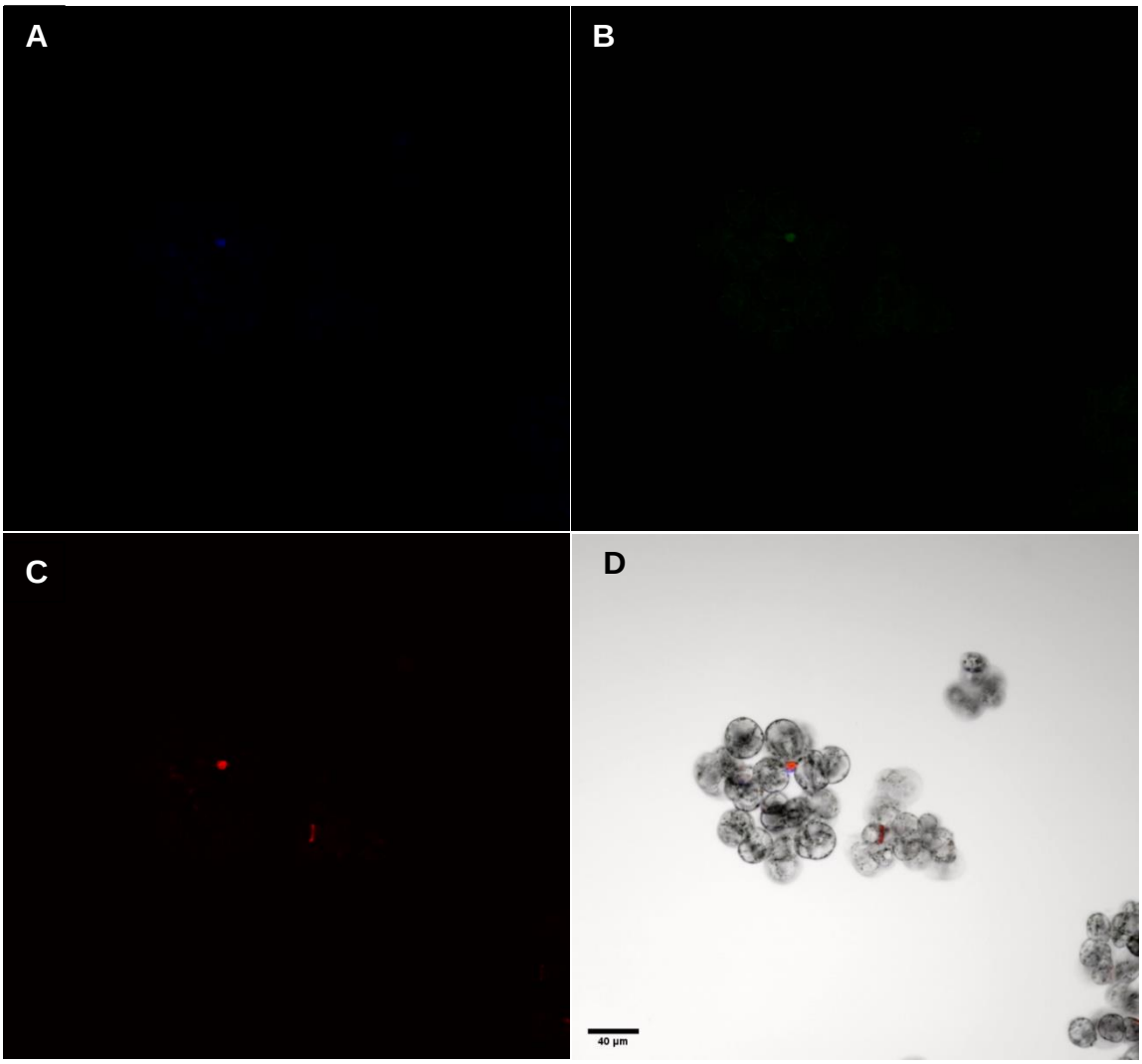


Figure 8. Confocal image of healthy untransformed MM1 cells. A. Blue channel, excitation at 405 nm and detection at 48 5 nm – 505 nm. **B.** Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. **C.** Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. **D.** Overlay channel of blue, green, red and brightfield. Non-stressed, untransformed cells do not have any signal in the selected channels. (Scale bar 40 µm).

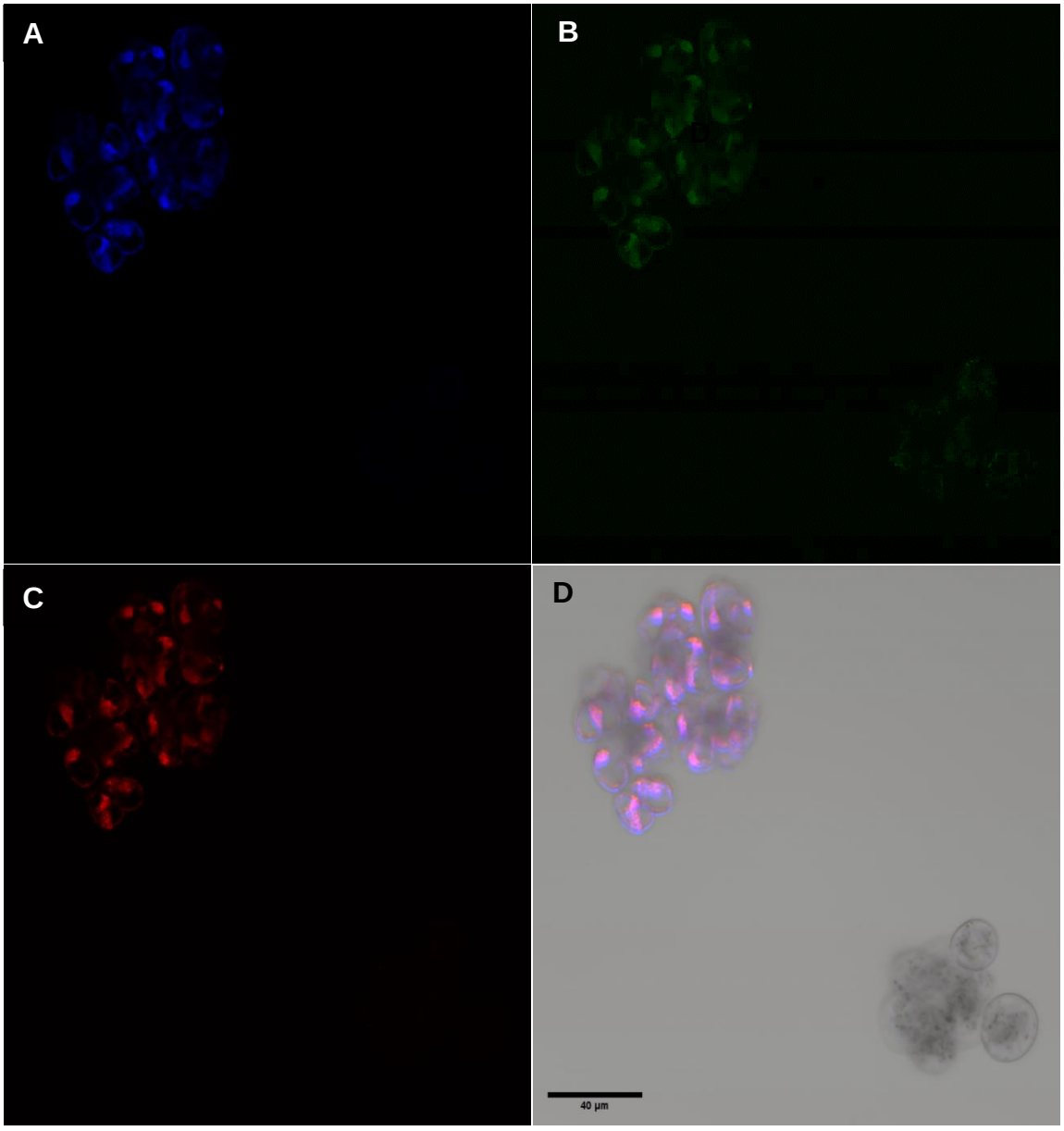


Figure 9. Confocal image of stressed untransformed MM1 cells. A. Blue channel, excitation at 405 nm and detection at 485 nm – 505 nm. B. Green channel, excitation at 488 nm and detection at 506nm – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. Stressed cells produced signal in all three channels. (Scale bar 40 µm).

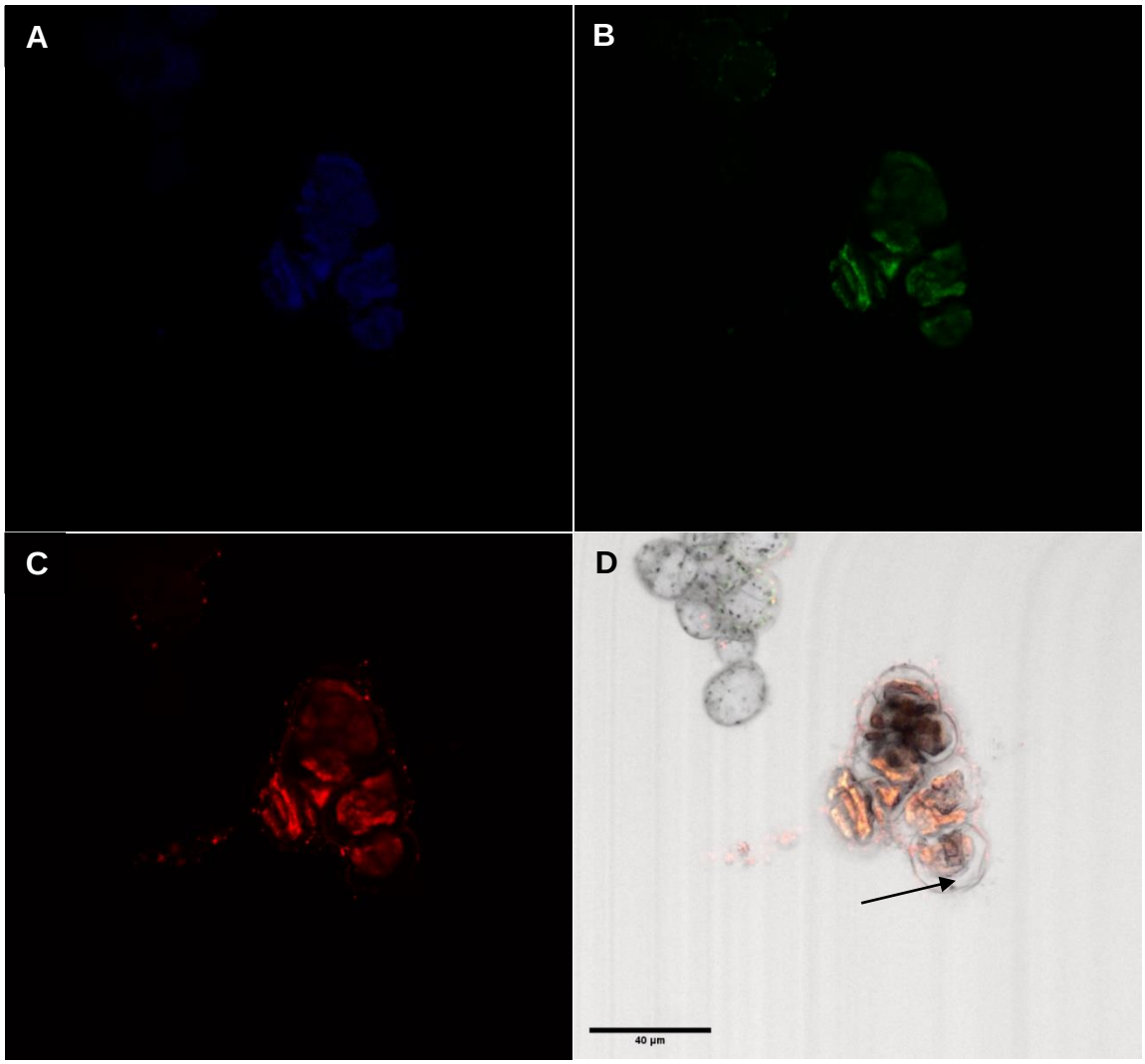


Figure 10. Confocal image of transformed, showing stressed and non-stressed uninduced MM1 cells. A. Blue channel, excitation at 405 nm and detection at 485 nm – 505 nm. B. Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm. Stressed cells have different cell shapes compared to non-stressed cells, and they produce fluorescent signals in all channels (blue, green, red). The arrow shows the white zone. (Scale bar 40 µm).

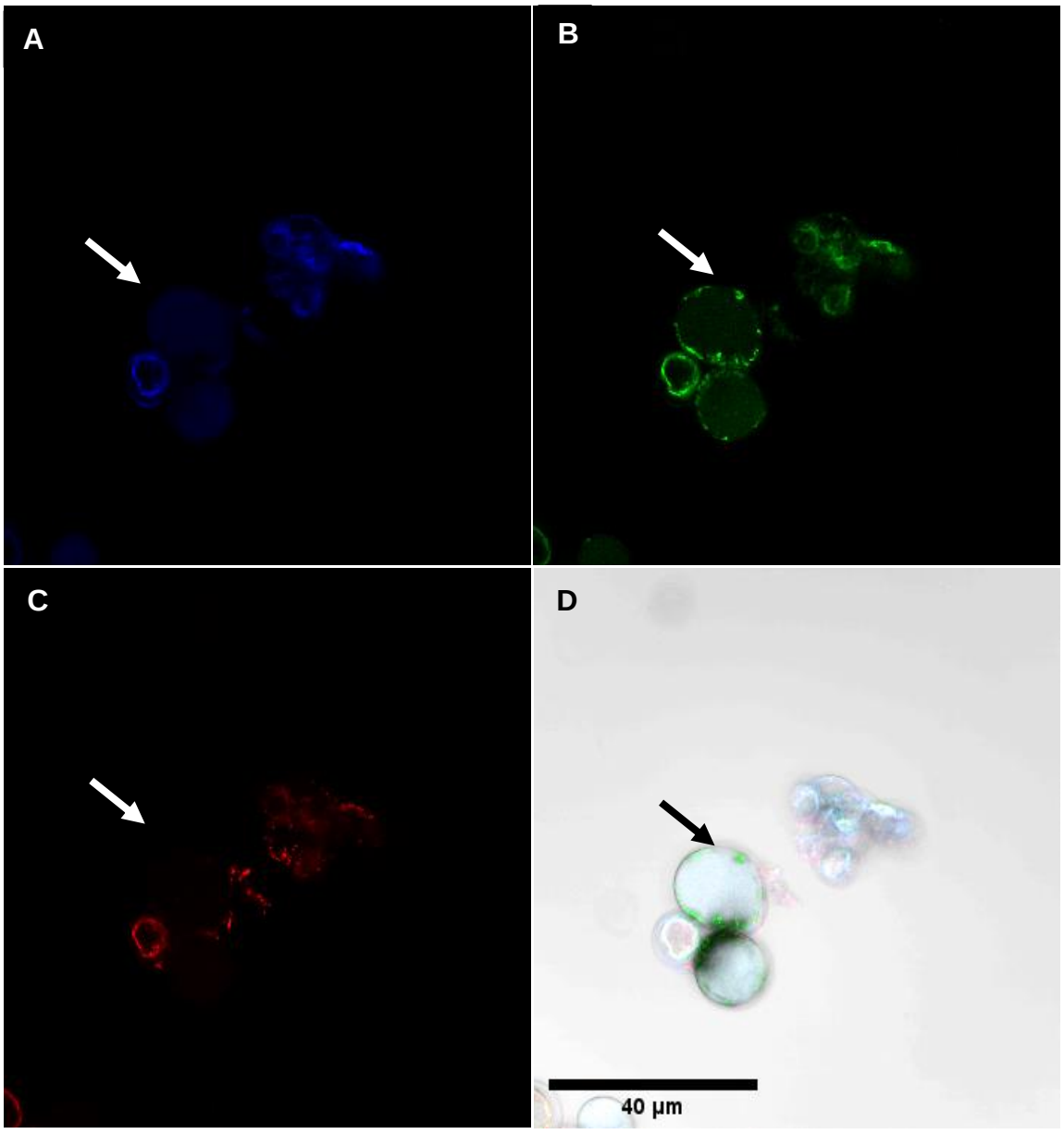


Figure 11. Confocal image of transformed, uninduced MM1 cells with leaky expression. A. Blue channel, excitation at 405 nm and detection at 485nm - 505nm. B. Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. The arrow points at a cell with leaky sfGFP expression, giving a characteristic expression domain in the green channel. (Scale bar 40 µm).

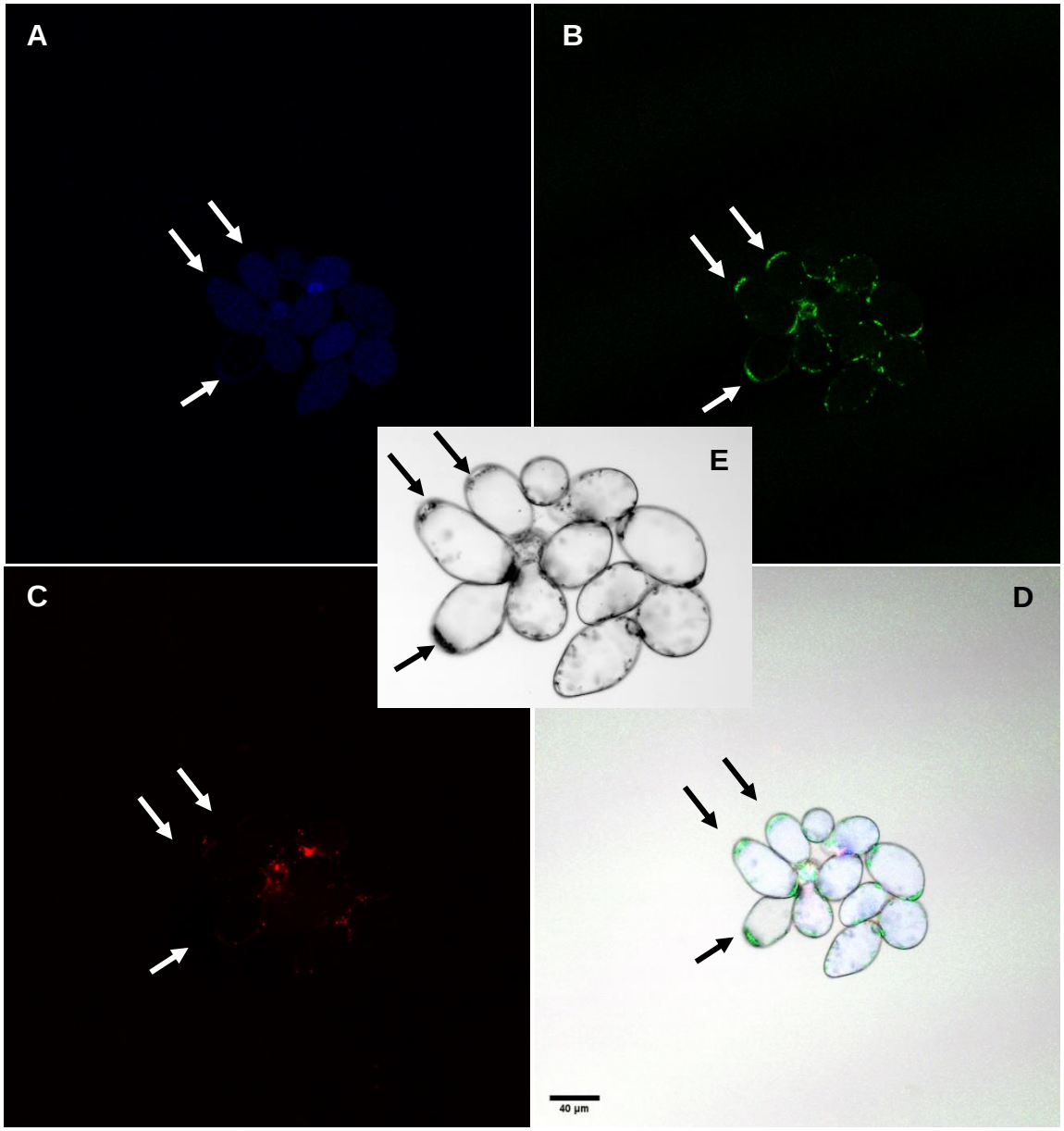


Figure 12. Confocal image of transformed, induced MM1 cells with β-estradiol. A. Blue channel, excitation at 405 nm and detection at 485 nm – 505 nm. B. Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. E. Brightfield. The arrows point at cells with sfGFP expression, having a clear signal only in the green channel. The cells are elongated and oval shaped, and their nucleus and cytoplasm are at the periphery of the cell. (Scale bar 40 µm).

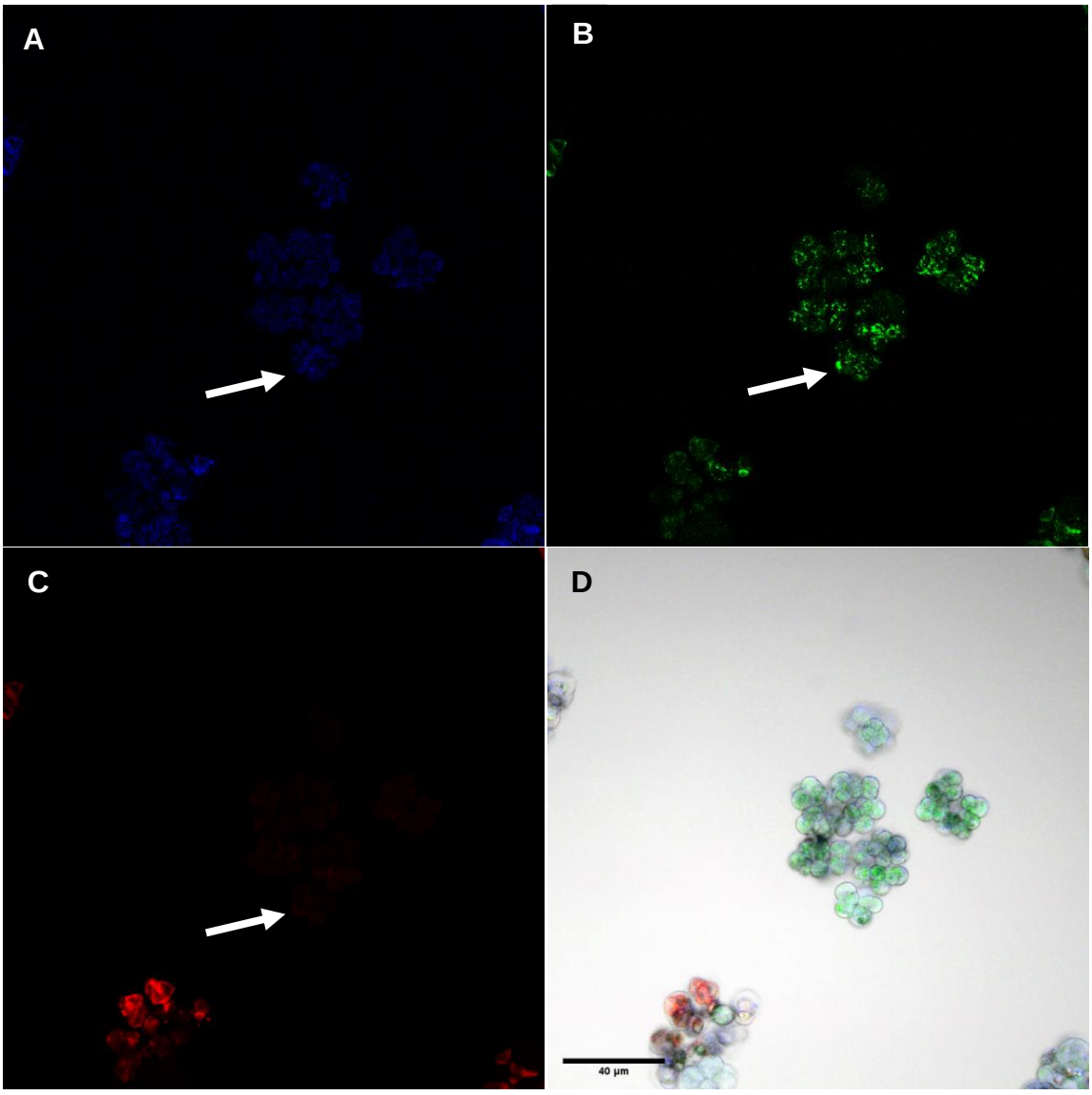


Figure 13. Confocal image of transformed MM1 cells induced with Dex. A. Blue channel, excitation at 405 nm and detection at 380 nm-400 nm. B. Green channel, excitation at 488 nm and detection at 506 nm-529 nm. C. Red channel, excitation at 552 nm-529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. Induced cells are visible in the blue and green channels, but not in red. The arrow indicates a cell with leaky sfGFP expression, which has a stronger signal in the green channel. (Scale bar 40 µm)

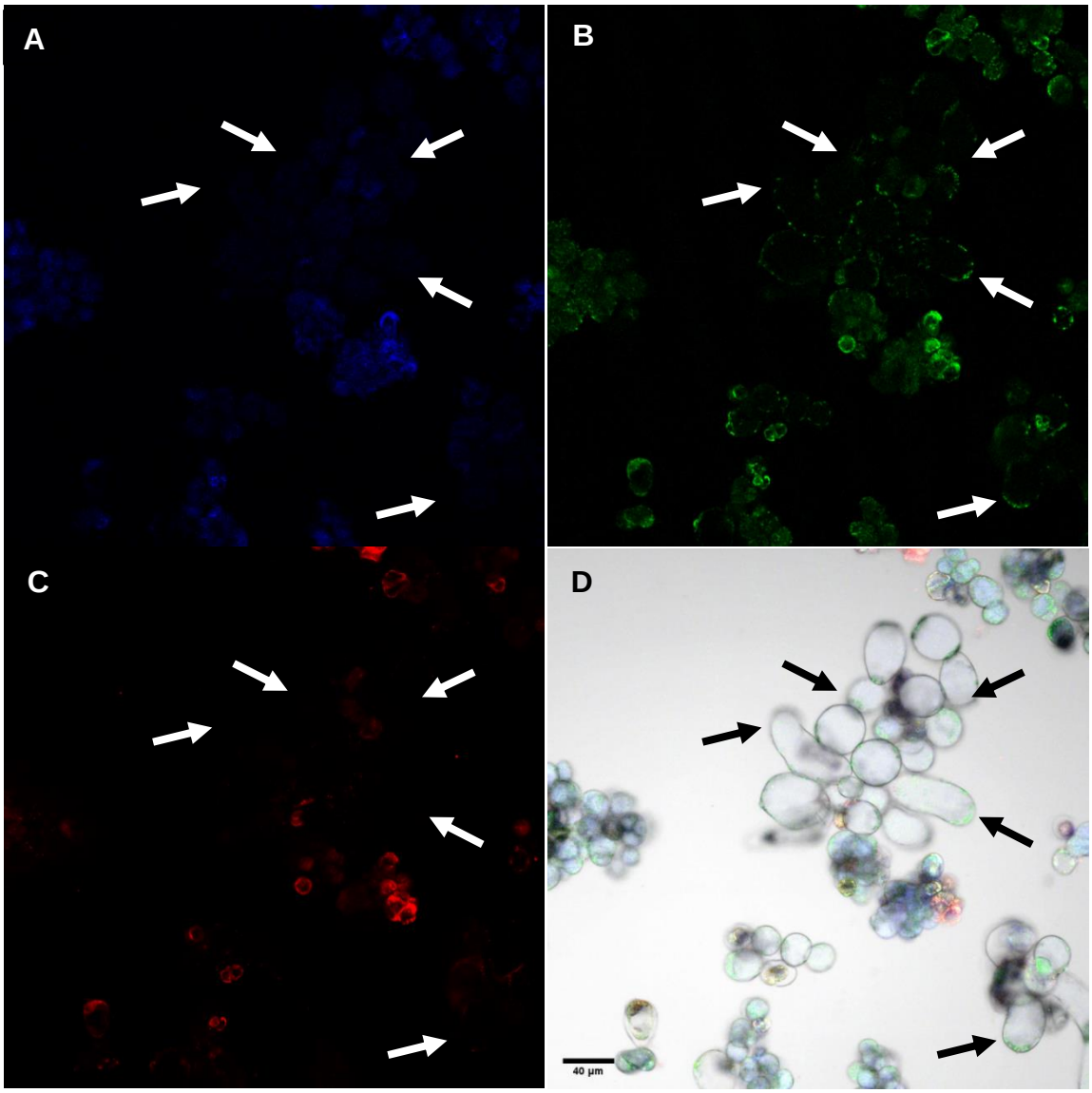


Figure 14. Confocal image of transformed MM1 cells, induced with β-estradiol and Dex (expression of sfGFP). A. Blue channel, excitation at 405 nm and detection at 380 nm – 400 nm. B. Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. C. Red channel, excitation at 552 – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. The arrows show transformed cells expressing sfGFP. Expression of mTFP1 is not detected. (Scale bar 40 µm).

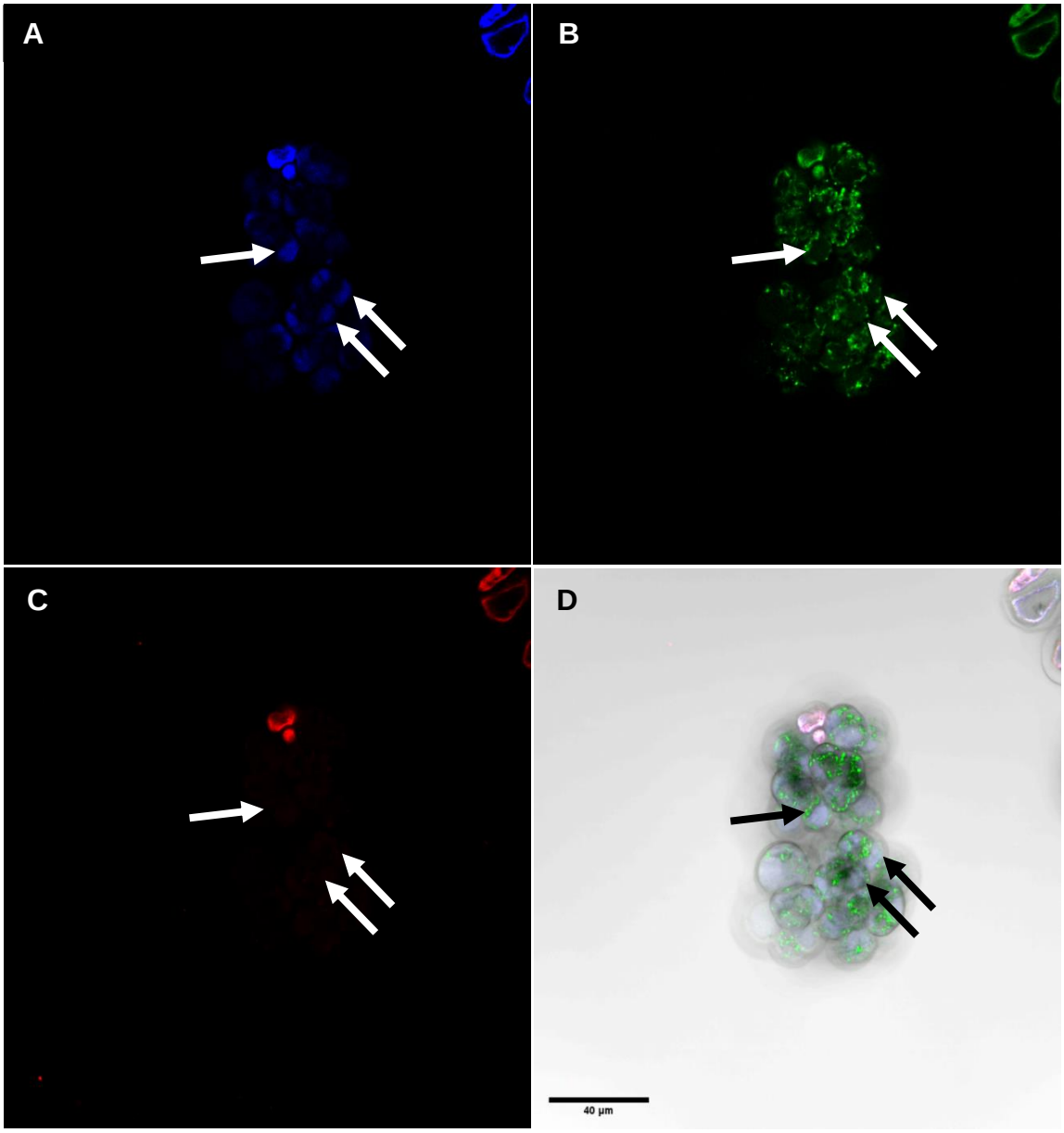


Figure 15. Confocal image of transformed MM1 cells induced with β-estradiol and Dex (expression of sfGFP and mTFP1). A. Blue channel, excitation at 405 nm and detection at 380 nm – 400 nm. B. Green channel, excitation at 488 nm and detection at 506 nm – 529 nm. C. Red channel, excitation at 552 nm and detection at 605 nm – 635 nm. D. Overlay channel of blue, green, red and brightfield. The arrows show transformed cells expressing sfGFP and mTFP1. (Scale bar 40 µm).

4.2.5. Anthocyanin analysis

Anthocyanin concentration was analysed using a spectrophotometer (wavelength = 535 nm), after acid-methanol extraction. The results showed no elevated anthocyanin levels in any of the samples regardless of induction with *VND6* and/or *MYB75* **(Table 2)**.

Treatment	**OD 535 nm**
Control-untransformed	0.013
Control-uninduced	0.015
Dex (VND6)	0.010
β-estradiol (MYB75)	0.014
β-estradiol + Dex (VND6 + MYB75)	0.012

Table 2. Analysis of anthocyanins. Anthocyanins were analysed at 535 nm using a spectrophotometer, after acidic methanol extraction.

4.3. Discussion

The sequential manipulation of plant cell differentiation will facilitate the development of plant biofactories to produce desired bio-compounds. Firstly, it will allow exploitation of unique metabolic profiles of specific cell types, bypassing the need to overexpress enzymes to produce metabolites of interest. Secondly, it will separate proliferation and differentiation phases, and switch to the latter only after adequate cells are obtained. It can also be used to mimic tissue, in the case that different cell types cooperate to produce a bio-compound.

4.3.1. Combinatorial compatibility and expression optimisation of three plant inducible systems

The choices for inducible systems in plants are limited, and their activation amplitude and kinetics have not been directly compared. However, this characterisation can address the possibility of crosstalk and allows their use in combination. For example, the estradiol inducible system was first constructed in 2000, and limited changes were introduced in 2013 when some of the restriction

sites were replaced, and the codon was optimised for Arabidopsis (27, 53). Since then, new DNA binding domains, activation domains, and hormone receptors have become available, and redesigning of the chimeric protein might improve gene activation and reduce the size of the system.

Recently, there have been attempts to enrich the inducible system toolbox in plants. Optogenetic tools, such as a red/far-red light switch, were developed, based on the plant photoreceptor phytochrome B and one of its interacting factors (PIF6) (54). Similarly, a green light switch based on the light-sensitive transcription factor CarH and its operator sequence CarO from *Thermus thermophilus* has been fabricated (55). However, as plants actively interact with light, these systems should undergo further crosstalk testing and modification to eliminate unwanted side effects on plant physiology. A temperature-controlled inducible system was developed based on a thermosensitive N-terminal degradation signal (lt-degron) (56). Furthermore, components that can be used with the inducible system were characterised *in planta*, such as the CRISPR-associated endoribonuclease Csy4 and its recognition sequence (57).

In this work, the three most commonly used gene induction systems were tested for their induction efficiency and combinatorial compatibility. Only two of them were shown to be specific and inducible in plant cell cultures. In the future, new inducible systems should be developed and characterised for their suitability to be used in multi-step gene induction.

Since the inducible systems will coexist and cofunction in the sequential differentiation system, it was important to demonstrate that they do not interfere with one another. The estradiol inducible system employs a human estrogen receptor which binds to β-estradiol (27), while the Dex inducible system uses a rat glucocorticoid receptor that recognises dexamethasone (58). These inducers are both steroids and have similar structures, thus raising concern for cross-induction. A study in HeLa cells showed that their receptors have antagonistic actions (59). Our work showed that the inducer molecules have an affinity to, and

activate, their receptors specifically, and there is no crosstalk between Dex and the estradiol inducible systems in plants (**Figure 4**). The estradiol inducible system employs the lexA operator sequence CTGT motif (**CTGT**acatataaccactggttttatat**ACAG)**. However, the lexA DNA binding domain can recognise variations of the CTGT motif and also non-canonical binding motifs (60, 61). The Dex inducible system utilises the synthetic and symmetric Lac operator (**AATTGTGAG**cg**CTCACAATT**). LacI also binds to the wild type sequences (AATTGTGAGCGGATAACAATT, AAATTGTAGCGAGTAACAACC and GGCAGTGAGCGCAACGCAATT), and some variations of them (62). The AlcR transcription factor from the ethanol inducible system binds the sequence WGCGG (63). Our results confirmed that there is no crosstalk between the DNA binding domains of the transactivators and the operator sequences of the inducible promoters (**Figure 4)**.

It was shown that the expression levels of the transactivator have a significant impact on the inducible systems' activity (**Figure 5)**. Low expression of the sXVE lowered the expression levels of the system, while medium and strong expression resulted in similar (presumably maximum) induced expression. Possibly, the low abundance of the transactivator molecules is not adequate to fill all the DNA binding sites of the inducible promoter, resulting in lower activation levels. However, after a threshold, where most of the DNA operators are occupied, the inducible system may reach its maximum expression capacity. Before this study, the Pea E9-RbcS terminator has been used in the estradiol inducible system for the expression of the sXVE (27). Here, the results indicate that the use of the E9-RbcS terminator neutralises the strong promoter, reducing the sXVE expression, and consequently the activity of the inducible system. Possibly lower expression of the sXVE is not adequate to fill all the operator sequences on the inducible promoter, resulting in lower expression levels.. In the optimisation of the estradiol inducible system, domestication of the G10-90 promoter resulted in a 20% decrease in the transactivator expression level (53). In order to improve this

reduction, the authors chose a different promoter instead of replacing the E9-RbcS terminator.

The ethanol inducible system was excluded from the sequential differentiation system. The first version of the ethanol inducible promoter (alcA), was showing neither strong activity nor inducibility (**Figure 2**). The second version (a synthetic promoter, alcSynth) showed strong expression, but it was highly leaky. Therefore, it was not suitable for tight controlled gene expression (**Chapter 3, Figure 8**). Compared to the alcA promoter, alcSynth has six more copies of alcR binding sites derived from the alcM, alcR and aldA promoters, the minimal Figwort mosaic virus 34S promoter and the translational enhancer from the tobacco mosaic virus 5'-leader sequence (TMV Ω).

4.3.2. Construction and transformation of a sequential differentiation system in the MM1 cell line

The large size and the repetitive elements of the sequential inducible systems render the constructs prone to rearrangements in *E. coli* and Agrobacteria. The main issues are the large size of the transactivators. The β-estradiol receptor (sXVE) is around 1.5 kb, and the Dex receptor (LhGR) is around 3.5 kb. The repetitive operator elements on the inducible promoters are also a likely cause. To minimise the probability of structural rearrangements on the plasmids, a small binary vector was chosen, and repeating sequence elements were avoided (i.e. using same or similar promoter and terminators).

Plant transformation vectors are typically over 10 kb. To keep the size of the backbone to a minimum, initially, the compact binary plasmid pGreenII (~3 kb) (49) was selected to house the double inducible system. However, pGreen showed transformation issues with TOP10 cells, which could be overcome by switching to the JM109 strain. It was then found that the JM109 strain was causing plasmid instabilities, probably due to the deletions of the lac operator sequence on the pOp6-35S inducible promoter. After testing the Dh5α and Dh10B strains,

it was determined that Dh10B is the best strain for stably propagating the sequential inducible system.

Nonetheless, several attempts to infect the MM1 Arabidopsis cell line using the pGreen vector resulted in no success. Therefore, the pLX backbone was used instead of pGreen, which is a newly developed small binary vector with a broad range origin of replication (BBR1) capable of propagation in *E. coli* and Agrobacteria (50). Regarding the Agrobacterium strain, AGL1 was the only one able to propagate the guided differentiation system. C58C1, GV3101 and LBA4404 resulted in no colonies after transformation, or there were plasmid deletions. AGL1 stability is due to the $recA^-$ genotype. Our results highlight that cloning and transformation protocols require sampling of variations of combinations, as not all combinations work equally well.

Not all binary vectors worked well for the transformation of Arabidopsis cell cultures. The pGreen binary vector was not efficient in the transformation of the MM1 cell line. The luciferase signal was low after the MM1 transformation, and with the subcultures and herbicide selection, the signal was lost. On the other hand, the pLX backbone showed a robust initial signal, which was maintained after the subcultures and selection. This could be due to the medium copy number of pLX compared to the low copy number of pGreen, which can transfer more copies of the construct to the plant cells.

The plasmid stability under non-selective conditions can also affect transformation. The pGreen origin of replication, pSa, does not include the stability regions necessary for the maintenance of the plasmid (49, 64). On the other hand, the pLX origin of replication, BBR1, derives from a cryptic plasmid, which is stable with and without stability regions (65). As the transfection of the plant cells is carried out without antibiotic pressure, it is possible that the pGreen plasmids are not 100% maintained in the agrobacteria strains. Another factor that can affect the transfection is the T-DNA borders of the binary vectors. pGreen has minimal synthetic LB and RB sequences, derived from pTiT37 and an overdrive

from the LBA4417 plasmid (49), while pLX has extended LB and RD T-DNA borders, including the overdrive from the plasmid pTiA6 (50).

When the transgene presence and activity level was monitored in the transformed MM1 cell lines over a month, there were time-dependent fluctuations in the luciferase signal. The strong signal at the beginning was due to multiple copies of the Nano luciferase gene, which was transiently delivered to the cells. In the subcultures, the luciferase signal probably drops because the constructs that are not incorporated in the plant cell genome cannot further propagate to the daughter cells. The signal becomes higher with the herbicide selection as the cells which do not have the resistance gene do not survive, leading to an increased population of transformed cells carrying the Nano luciferase gene. The small-signal increase that was observed in the cells under no selection pressure may be attributed to recovery from the stress which they were exposed to during the transformation. The recovery leads to denser cells populations and consequently to higher Nano luciferase signal.

4.3.3. Guided differentiation of the MM1 cells into xylem vessels for anthocyanin production

The sequential differentiation system was successfully constructed and delivered to the Arabidopsis MM1 cell line, mediated with co-culture of the Agrobacterium strain AGL1. DNA sequence has been verified prior to the cell culture transformation. The delivery of the system was initially verified with the luciferase assay. The switching on of the VND6 expression was shown by confocal microscopy (sfGFP). VND6 expression did not lead to complete differentiation of the MM1 cells into xylem vessels. The cells gained an elongated shape, which is a known phenotype associated with xylem vessels; however, the characteristic cell wall structure of the cell type - helical secondary cell walls - was not visible. There was also no clear signal detected with the cell wall dye WGA- Alexa Fluor 594.

There are multiple possible explanations for the unsuccessful differentiation into xylem vessels. First, the C-terminal tagging of VND6 might have hindered the function of the TF. VND6 is a NAC domain protein, and the NAC domain family have been previously characterised as transcriptional activators (66). Their N-terminal regions have DNA binding properties, and the C-terminal regions are responsible for the transcriptional activation. Deletion experiments of the VND7 C-terminal region confirmed its role in transcriptional activation (67). In this study, the researchers used N-tagging of VND7 with successful differentiation of tobacco BY-2 cells into xylem vessels. Future experiments should, therefore, include untagged and N-tagged constructs.

Two more factors affecting the differentiation into tracheary elements are the cell type and the transcription factors. The closely related VND6 and VND7 are considered functionally redundant, but may have distinct functions, and are likely species or cell line dependent. Yamaguchi *et al*. fused VND6 or VND7 to the potent activation domain of VP16 from the herpes virus protein and the GR hormone receptor, and used them in two different cell lines (68). In the Arabidopsis T87 line (69), expression of VND7 induced differentiation of approximately 10% in cultured cells. They did not mention if they used VND6 with the T87 line. VND7-VP16 expression in the tobacco BY-2 cell line resulted in more than 90% of cell differentiation. However, differentiation into xylem vessel was not observed, in the transgenic BY-2 cells expressing VND6-VP16.

Two more studies using a different Arabidopsis Col-0 suspension cell culture developed by Mathur *et al*. (70) showed ~50-60% (71) and 80% (72) differentiation of the cultured cells into xylem cells with secondary walls, by expression of native VND6. In future, VND7 should be tested in the MM1 cell line, as well as both VND6/7 in different cell lines. The fusion of the VP16 activation domain with VND6/7 is also a good strategy for further enhancement of the xylem vessel specification pathway.

In addition to the changes in the TF design and cellular background, further optimisation experiments should be carried out regarding the induction time, cell concentration and media additives (e.g. sugar content). In this work, the transformed cell culture did not undergo herbicide selection to acquire a homogenous population of transgenic cells due to time constraints. The future generation of a homogenous population of transgenic cells will reduce the mixed cell signals and phenotypes observed. Furthermore, fusing of the GR domain directly with the TF of interest (in this case, VND6/7) could be a new solution to decrease the size and the complexity of the sequential inducible system. If the hormone receptor is fused directly to the TF, they would not need the DNA binding domain of the transactivator or the inducible promoter sequence.

Regarding the expression of MYB75 and induction of anthocyanin production, similar strategies can be followed. The herbicide selection is necessary, as most of the cells are not transformed. The transformation protocol that was used suggests 1-10% success, and thus transgene-containing cells are diluted in the population of mostly untransformed cells. Moreover, MYB75 has a C-terminal transcriptional activation domain and an N-Terminal DNA binding domain, similarly to VND6/7, as well as many other TFs in general (73, 74). C-terminal tagging might reduce its transcription activation properties. Finally, the selection of sfGFP and mTFP1 as reporter tags was not an optimal combination as they have overlapping excitation (Ex) and emission (Em) spectra (mTFP1: Ex λ=272 and 462 and Em λ=492; sfGFP: Ex λ=485 and Em λ=510). Excitation with the UV laser at 405 nm results in 24.8% of mTFP1 and 14.7% of sfGFP and detection at 485-505 nm included an optimal emission area for mTFP1 and suboptimal for sfGFP. The 488 nm laser resulted in 99.6% excitation for sfGFP and only 17.4% for mTFP1, while the detection at 506-529 nm was optimal for sfGFP but suboptimal for mTFP1. With the selected settings, sfGFP expression could be monitored, but the mTFP1 expression was in both the blue and the green channels. Therefore, a clear distinction of sfGFP and mTFP1 was difficult. In the

future, one of the two fluorescent proteins should be replaced with a different one without spectral overlap, for example, Venus (Ex λ=515nm, Em λ=528nm).

4.4. Materials and Methods

4.4.1. Plant growth conditions

Wildtype (Col-0) *A. thaliana* seeds were planted in soil and grown in long-day conditions (21 °C, 16 h light intensity of ~100 µmol m^{-2} s^{-1}, humidity ~65%). After two weeks, they were transplanted and transferred to short-day conditions (21 °C, 9 h light intensity of ~110 µmol m^{-2} s^{-1}) and grown for 4-6 weeks until the protoplast isolation and transformation.

4.4.2. Nucleic acid extraction and cDNA synthesis

Genomic DNA was extracted from snap-frozen inflorescence tissues from Col-0 Arabidopsis plants using a plant DNA extraction kit (Sigma-Aldrich). Root, stem, and inflorescence tissues from mature plants were snap-frozen with liquid nitrogen, and total mRNA was isolated using a plant RNA extraction kit (Macherey Nagel). The cDNA library was constructed using SuperScriptIII Reverse Transcriptase (Life Technologies) and an Oligo(dT)12-18 primer.

4.4.3. Amplification of transcription factor genes and regulatory sequences

The sequences of the transcription factor genes were retrieved from the TAIR website (www.arabidopsis.org) and amplified from the cDNA made from the Arabidopsis tissues in which the gene was shown transcribed (see **Table**). Subsequently, Mobius Assembly primers were used for a second amplification and cloning into mUAV.

Parts for the dexamethasone inducible system were amplified from pL0M-SC-ncoLHGRN-17007 and pL0M-PU-pOP6-17005 vectors provided by Prof. Jane Langdale (Department of Plant Sciences, University of Oxford, UK). The estradiol

inducible system was cloned from pGPTVII vector from the AG Kudla Lab (Institute of Plant Biology and Biotechnology, Münster, Germany), and the ethanol inducible system from the Laboratoire de Biologie Cellulaire (INRA Versailles, France). Synthetic Ethanol promoter was synthesised from IDT, according to Pasin *et al.* (50).

4.4.4. Mobius Assembly

Mobius Assembly was carried out in 10 µl total reaction volume. ~50 µg Acceptor Vector and twice as many molars of the insert parts were used, in addition to 1 µl 10 mM ATP (NEB), 1 µl enzyme buffer, 0.5 µl *Aar*I (cloning in mUAV and Level 2 Acceptor Vectors) or *Bsa*I (cloning in Level 2 Acceptor Vectors) (Life Technologies), 0.5 µl T4 ligase (NEB), and for *Aar*I an extra 0.2 µl oligos of the enzyme recognition sites. The reaction was incubated in a thermocycler with 5-10 cycles of 37 °C for 10 min, 16 °C for 10 min, and 37 °C for 10 min, followed by deactivation at 80 °C for 20 min.

4.4.5. Protoplast isolation and transformation

The protocol was described extensively in Chapter 3. Briefly, leaves from 6-8 weeks old plants were digested with MGG solution (75) with twice the concentration of enzymes, i.e. Cellulase ONOZUKA R-10 and Macerozyme R-10 (Yakult Pharmaceutical) and Driselase (Sigma-Aldrich). The following day the protoplasts were filtered and washed with MGG without enzymes and resuspended in MMM solution (0.4 M mannitol, 15 mM $MgCl_2$, 0.1% (w/v) MES, pH 8) in a concentration of $5x10^5$ cells/mL. Protoplasts were transformed in 2 ml Eppendorf tubes or 1 ml 96 well plates. 8 µl DNA (500 ng/µl) was added to 75 µl of the protoplast suspension followed by the addition of 83 µl PEG (0.4 M mannitol, 0.1 M $Ca(NO_3)_2$, 40% PEG 4000, pH 8) and 1 min incubation. Subsequently, the tubes were filled with MGG solution and incubated for 1 h at room temperature. After the incubation the solution was removed, the protoplasts

were resuspended in 100 µl of fresh MGG and were incubated at room temperature in darkness overnight.

4.4.6. Plant cell suspension culture growth and transformation

MM1 cell suspension culture of *A.s thaliana* ecotype *Landsberg erecta* (32) was subcultured weekly by 1:10 dilution in 1× MS medium (Sigma) supplemented with 2% (w/v) glucose (Sigma), 0.5 mg/L naphthaleneacetic acid (Sigma), 0.05 mg/L kinetin (Sigma), with pH adjusted to 5.8 with KOH.

For transformation, 7-day-old cells in transformation medium were co-cultured with the AGL1 Agrobacterium strain (directly from a fresh colony) for two days at 29^{o}C. Agrobacteria were eliminated with 0.5 mg/L Cefotaxime (Sigma), and the cells were cultured for two more days, followed by a 1:1 dilution with the culture medium. After two days, they were adjusted to the weekly routine subculture, and two weeks later, they were subjected to herbicide selection with 15 mg/L glufosinate-ammonium (Sigma) (**Diagram 1**).

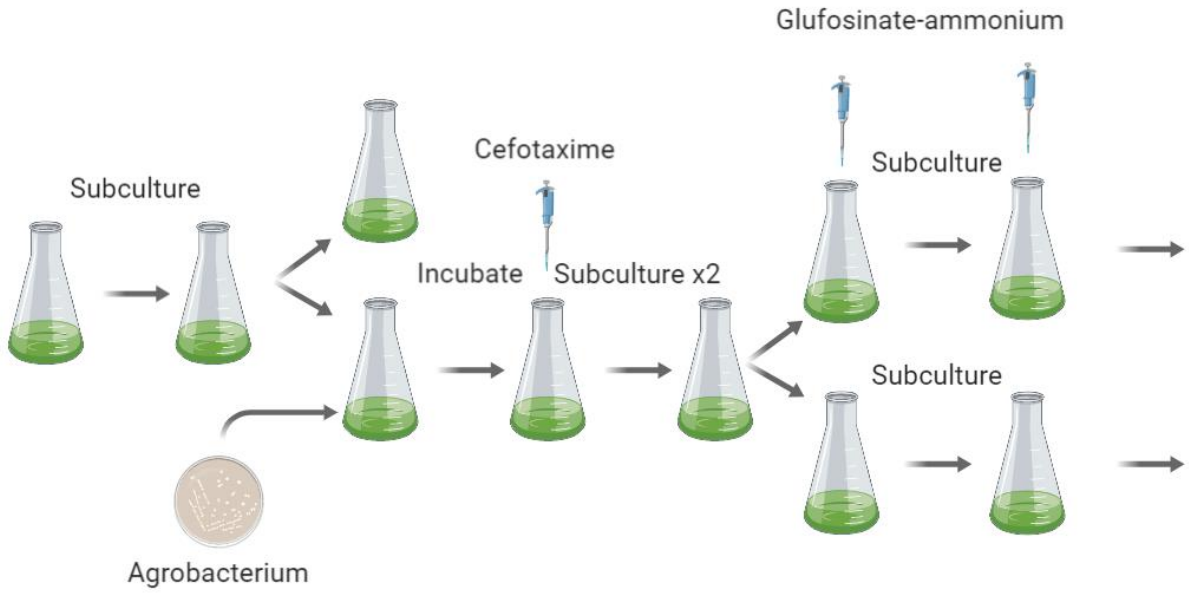


Diagram 1. Plant cell suspension culture growth and transformation. Seven-day old cells in transformation medium were co-cultured with the AGL1 Agrobacterium strain (directly from a fresh colony) for two days at 29^{o}C. Subsequently, Agrobacteria were

eliminated with 0.5 mg/L Cefotaxime (Sigma), and the cells were cultured for two more days, followed by a 1:1 dilution with the culture medium. After two days, they were adjusted to the weekly routine subculture, and two weeks later they were subjected to herbicide selection with 15 mg/L glufosinate-ammonium

4.4.7. Agrobacterium transformation and culturing conditions.

AGL1 Agrobacterium strain was cultured in 10 ml YEP medium at 28°C, 200 min^{-1} for two days. For Agrobacterium transformation, 1 µl of plasmid DNA was added into 40 µg of electrocompetent cells and incubated on ice for 5 min. Subsequently, the cells were transferred into an ice-cold electroporation cuvette, 0.2 cm gap (Bio-Rad) and pulsed twice in a Gene Pulser Xcell Electroporation System (Bio-Rad) using pre-set settings for Agrobacteria (25 µF, 200 Ω, 2400 V, 2 mm). S.O.C medium was added immediately (1 ml), and after 2 h of shaking at room temperature, 20 µl was plated on YEP plates with the correct antibiotics and incubated at 30°C for 2-3 days until colonies had formed.

4.4.8. Activation of the differentiation system

A 0.5 mL aliquot of 7-day-old culture was transferred to 6-well plates (Greiner Bio-One) containing 4.5 ml of culture medium without hormones. 2.5 µM DEX (Acros) or 5 µM β-estradiol (LKT laboratories) and 2 µM brassinolide (Sigma) were added to the medium and the cells were transferred to a rotary shaker (set at 120 min^{-1}) at 29°C. After 24 h hour, the induction was spiked with more hormones, and at 48 h the cells were used for microscopy.

4.4.9. Luciferase assay

Luciferase activity for both protoplasts and MM1 cells was assayed in an Omega luminescence plate-reader (Fluostar) following the instructions of the Nano-GloLuciferase or Nano-Glo Dual-Luciferase Reporter Assay System (Promega).

4.4.10. Epifluorescence and confocal microscopy and image analysis

Protoplasts were examined with a Nikon E600 fluorescence microscope, and the images were captured using the Leica Application Suite software. MM1 cells were

observed with an upright confocal laser scanning microscopy system (Leica SP8) using the Leica SPX software. Before observation, the cells were stained for 10 min with 1 mg/L Alexa Fluor 594-conjugated wheat germ agglutinin (Thermo Fisher Scientific) to visualise secondary cell walls (76). mTFP1, sGFP and Alexa Fluor 594 were excited with lasers set at 405 nm, 488 nm and 552 nm, respectively. Fluorescence was detected at 485 nm – 505 nm for mTFP1, at 506 nm – 529 nm for sfGFP and at 605nm - 635nm for Alexa Fluor 594 nm. Acquired images were analysed with Leica SPX or Image J.

4.4.11. Analysis of anthocyanins

10 ml of MM1 cell suspension harbouring the sequential inducible system was induced twice in four days with Dex, β-estradiol, or a combination of both hormones. Cells were harvested and resuspended in 5 volumes of acidic methanol (1% HCI) in 50 ml falcon tubes. Glass beads were added, and the tubes were then vortexed for 3 min followed by 1 h incubation on ice and sporadic vortexing. 2 ml of the extract was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 16,000 min^{-1} for 10 min. The supernatant was measured at 535 nm for detection of anthocyanins using the Biowave II spectrophotometer (Montreal Biotech).

4.5. References

1. Atanasov,A.G., Waltenberger,B., Pferschy-Wenzig,E.M., Linder,T., Wawrosch,C., Uhrin,P., Temml,V., Wang,L., Schwaiger,S., Heiss,E.H., *et al.* (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol. Adv.*, **33**, 1582–1614.
2. Harvey,A.L., Edrada-Ebel,R. and Quinn,R.J. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.*, **14**, 111–129.
3. Liu,X., Ding,W. and Jiang,H. (2017) Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale

production. *Microb. Cell Fact.*, **16**, 125.

4. Zebec,Z. and Scrutton,N.S. (2018) Genome Editing for the Production of Natural Products in Escherichia coli. *Adv. Biosyst.*, **2**, 1800056.

5. Pyne,M.E., Narcross,L. and Martin,V.J.J. (2019) Engineering Plant Secondary Metabolism in Microbial Systems. *Plant Physiol.*, **179**, pp.01291.2018.

6. Luo,X., Reiter,M.A., d'Espaux,L., Wong,J., Denby,C.M., Lechner,A., Zhang,Y., Grzybowski,A.T., Harth,S., Lin,W., *et al.* (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*, **567**, 123–126.

7. Li,S., Li,Y. and Smolke,C.D. (2018) Strategies for microbial synthesis of high-value phytochemicals. *Nat. Chem.*, **10**, 395–404.

8. Warzecha,H. (2008) Biopharmaceuticals from plants: a multitude of options for posttranslational modifications. *Biotechnol. Genet. Eng. Rev.*, **25**, 315–30.

9. Rosano,G.L. and Ceccarelli,E. a. (2014) Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.*, **5**, 1–17.

10. Georgiev,V., Slavov,A., Vasileva,I. and Pavlov,A. (2018) Plant cell culture as emerging technology for production of active cosmetic ingredients. *Eng. Life Sci.*, **18**, 779–798.

11. Marisol Ochoa-Villarreal, Howat,S., Hong,S.M., Jang,M.O., Jin,Y.W., Lee,E.K. and Loake,G.J. (2016) Plant cell culture strategies for the production of natural products. *BMB Rep.*, 10.5483/BMBRep.2016.49.3.264.

12. Tekoah,Y., Shulman,A., Kizhner,T., Ruderfer,I., Fux,L., Nataf,Y., Bartfeld,D., Ariel,T., Gingis–Velitski,S., Hanania,U., *et al.* (2015) Large-scale production of pharmaceutical proteins in plant cell culture—the protalix experience. *Plant Biotechnol. J.*, **13**, 1199–1208.

13. Deroles,S. (2009) Anthocyanin Biosynthesis in Plant Cell Cultures: A Potential Source of Natural Colourants BT - Anthocyanins: Biosynthesis, Functions, and Applications. In Winefield,C., Davies,K., Gould,K. (eds). Springer New York, New York, NY, pp. 108–167.

14. Diamond,A. and Desgagné-Penix,I. (2016) Metabolic engineering for the production of plant isoquinoline alkaloids. *Plant Biotechnol. J.*, **14**, 1319–1328.

15. Giri,C.C. and Zaheer,M. (2016) Chemical elicitors versus secondary metabolite production in vitro using plant cell, tissue and organ cultures: recent trends and a sky eye view appraisal. *Plant Cell, Tissue Organ Cult.*, **126**, 1–18.

16. Bai,F., Ho Lim,C., Jia,J., Santostefano,K., Simmons,C., Kasahara,H., Wu,W., Terada,N. and Jin,S. (2015) Directed Differentiation of Embryonic Stem Cells

Into Cardiomyocytes by Bacterial Injection of Defined Transcription Factors. *Sci. Rep.*, **5**, 15014.

17. Bian,J., Zheng,J., Li,S., Luo,L. and Ding,F. (2016) Sequential Differentiation of Embryonic Stem Cells into Neural Epithelial-Like Stem Cells and Oligodendrocyte Progenitor Cells. *PLoS One*, **11**, e0155227.

18. Lemon,B. and Tjian,R. (2000) Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.*, **14**, 2551–2569.

19. Wang,S. and Schiefelbein,J. (2014) Regulation of cell fate determination in plants. *Front. Plant Sci.*, **5**, 368.

20. Kubo,M., Udagawa,M., Nishikubo,N., Horiguchi,G., Yamaguchi,M., Ito,J., Mimura,T., Fukuda,H. and Demura,T. (2005) Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.*, **19**, 1855–60.

21. Deng,Y. and Lu,S. (2017) Biosynthesis and Regulation of Phenylpropanoids in Plants. *CRC. Crit. Rev. Plant Sci.*, **36**, 257–290.

22. Patra,B., Schluttenhofer,C., Wu,Y., Pattanaik,S. and Yuan,L. (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta*, **1829**, 1236–1247.

23. Vom Endt,D., Kijne,J.W. and Memelink,J. (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry*, **61**, 107–114.

24. Borevitz,J.O., Xia,Y., Blount,J., Dixon,R.A. and Lamb,C. (2000) Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. *Plant Cell*, **12**, 2383 LP – 2393.

25. Liu,J., Osbourn,A. and Ma,P. (2015) MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Mol. Plant*, **8**, 689–708.

26. Ohtani,M., Kurata,T., Sano,R., Demura,T., Morisaki,K., Uy,A.L.T., Sawada,Y., Matsuda,M., Hirai,M.Y., Suzuki,S., *et al.* (2016) Primary Metabolism during Biosynthesis of Secondary Wall Polymers of Protoxylem Vessel Elements. *Plant Physiol.*, **172**, 1612–1624.

27. Zuo,J., Niu,Q.W. and Chua,N.H. (2000) Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.*, **24**, 265–273.

28. Moore,I., Gälweiler,L., Grosskopf,D., Schell,J. and Palme,K. (1998) A transcription activation system for regulated gene expression in transgenic plants. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 376–381.

29. Craft,J., Samalova,M., Baroux,C., Townley,H., Martinez,A., Jepson,I., Tsiantis,M. and Moore,I. (2005) New pOp/LhG4 vectors for stringent

glucocorticoid-dependent transgene expression in Arabidopsis. *Plant J.*, **41**, 899–918.

30. Caddick,M.X., Greenland,A.J., Jepson, lan, Krause,K.-P., Qu,N., Riddell,K. V, Salter,M.G., Schuch,W., Sonnewald,U. and Tomsett,A.B. (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nat. Biotechnol.*, **16**, 177–180.

31. Roslan,H.A., Salter,M.G., Wood,C.D., White,M.R., Croft,K.P., Robson,F., Coupland,G., Doonan,J., Laufs,P., Tomsett,A.B., *et al.* (2001) Characterization of the ethanol-inducible alc gene-expression system in Arabidopsis thaliana. *Plant J.*, **28**, 225–235.

32. Menges,M. and Murray,J.A.H. (2002) Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. *Plant J.*, **30**, 203–212.

33. Naqvi,S., Farre,G., Sanahuja,G., Capell,T., Zhu,C. and Christou,P. (2010) When more is better: multigene engineering in plants. *Trends Plant Sci.*, **15**, 48–56.

34. Gelvin,S.B. (2017) Integration of Agrobacterium T-DNA into the Plant Genome . *Annu. Rev. Genet.*, **51**, 195–217.

35. Pasin,F., Tseng,X.-A., Bedoya,L.C., Heydarnejad,J., Deng,T.-C., Garcia,J.A. and Chen,Y.-R. (2018) Streamlined generation of plant virus infectious clones using the pLX mini binary vectors. *J. Virol. Methods*, **262**, 48–55.

36. Schuetz,M., Smith,R. and Ellis,B. (2012) Xylem tissue specification, patterning, and differentiation mechanisms. *J. Exp. Bot.*, **64**, 11–31.

37. Punwani,J.A. and Drews,G.N. (2008) Development and function of the synergid cell. *Sex. Plant Reprod.*, **21**, 7–15.

38. Wang,S., Zhang,G., Song,Q., Zhang,Y., Li,Z., Guo,J., Niu,N., Ma,S. and Wang,J. (2015) Abnormal Development of Tapetum and Microspores Induced by Chemical Hybridization Agent SQ-1 in Wheat. *PLoS One*, **10**, e0119557.

39. Shin,D.H., Cho,M., Choi,M.G., Das,P.K., Lee,S.-K., Choi,S.-B. and Park,Y.-I. (2015) Identification of genes that may regulate the expression of the transcription factor production of anthocyanin pigment 1 (PAP1)/MYB75 involved in Arabidopsis anthocyanin biosynthesis. *Plant Cell Rep.*, **34**, 805–815.

40. Mehrtens,F., Kranz,H., Bednarek,P. and Weisshaar,B. (2005) The Arabidopsis Transcription Factor MYB12 Is a Flavonol-Specific Regulator of Phenylpropanoid Biosynthesis. *Plant Physiol.*, **138**, 1083 LP – 1096.

41. Ikeda,M., Mitsuda,N. and Ohme-Takagi,M. (2009) Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell

regulation and as an activator in floral patterning. *Plant Cell*, **21**, 3493–3505.

42. Etchells,J.P., Provost,C.M., Mishra,L. and Turner,S.R. (2013) WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. *Development*, **140**, 2224–2234.

43. Pi,L., Aichinger,E., van der Graaff,E., Llavata-Peris,C.I., Weijers,D., Hennig,L., Groot,E. and Laux,T. (2015) Organizer-Derived WOX5 Signal Maintains Root Columella Stem Cells through Chromatin-Mediated Repression of CDF4 Expression. *Dev. Cell*, **33**, 576–588.

44. Kubo,M., Nakano,Y., Ohtani,M., Goue,N., Yamaguchi,M., Igarashi,H., Katayama,Y., Nishikubo,N., Demura,T., Kakegawa,K., *et al.* (2010) VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 Effectively Induce Transdifferentiation into Xylem Vessel Elements under Control of an Induction System. *PLANT Physiol.*, **153**, 906–914.

45. Jia,G., Liu,X., Owen,H.A. and Zhao,D. (2008) Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 2220–2225.

46. Zhu,J., Chen,H., Li,H., Gao,J.-F., Jiang,H., Wang,C., Guan,Y.-F. and Yang,Z.-N. (2008) Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. *Plant J.*, **55**, 266–277.

47. Kasahara,R.D., Portereiko,M.F., Sandaklie-Nikolova,L., Rabiger,D.S. and Drews,G.N. (2005) MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. *Plant Cell*, **17**, 2981–2992.

48. Chen,X., L Zaro,J. and Shen, wei-chiang (2012) Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.*, **65**.

49. Hellens,R.P., Edwards,E.A., Leyland,N.R., Bean,S. and Mullineaux,P.M. (2000) pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.*, **42**, 819–832.

50. Pasin,F., Bedoya,L.C., Bernabé-Orts,J.M., Gallo,A., Simón-Mateo,C., Orzaez,D. and García,J.A. (2017) Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synth. Biol.*, **6**, 1962–1968.

51. Hwang,H.-H., Yu,M. and Lai,E.-M. (2017) *Agrobacterium*-Mediated Plant Transformation: Biology and Applications. *Arab. B.*, **2017**.

52. Hogetsu,T. (1990) Detection of hemicelluloses specific to the cell wall of tracheary elements and phloem cells by fluorescein-conjugated lectins.

Protoplasma, **156**, 67–73.

53. Schlücking,K., Edel,K.H., Köster,P., Drerup,M.M., Eckert,C., Steinhorst,L., Waadt,R., Batistič,O. and Kudla,J. (2013) A New β-Estradiol-Inducible Vector Set that Facilitates Easy Construction and Efficient Expression of Transgenes Reveals CBL3-Dependent Cytoplasm to Tonoplast Translocation of CIPK5. *Mol. Plant*, **6**, 1814–1829.

54. Müller,K., Siegel,D., Rodriguez Jahnke,F., Gerrer,K., Wend,S., Decker,E.L., Reski,R., Weber,W. and Zurbriggen,M.D. (2014) A red light-controlled synthetic gene expression switch for plant systems. *Mol. Biosyst.*, **10**, 1679–1688.

55. Chatelle,C., Ochoa-Fernandez,R., Engesser,R., Schneider,N., Beyer,H.M., Jones,A.R., Timmer,J., Zurbriggen,M.D. and Weber,W. (2018) A Green-Light-Responsive System for the Control of Transgene Expression in Mammalian and Plant Cells. *ACS Synth. Biol.*, **7**, 1349–1358.

56. Faden,F., Ramezani,T., Mielke,S., Almudi,I., Nairz,K., Froehlich,M.S., Höckendorff,J., Brandt,W., Hoehenwarter,W., Dohmen,R.J., *et al.* (2016) Phenotypes on demand via switchable target protein degradation in multicellular organisms. *Nat. Commun.*, **7**, 12202.

57. Liang,Y., Richardson,S., Yan,J., Benites,V.T., Cheng-Yue,C., Tran,T., Mortimer,J., Mukhopadhyay,A., Keasling,J.D., Scheller,H. V, *et al.* (2017) Endoribonuclease-Based Two-Component Repressor Systems for Tight Gene Expression Control in Plants. *ACS Synth. Biol.*, **6**, 806–816.

58. Craft,J., Samalova,M., Baroux,C., Townley,H., Martinez,A., Jepson,I., Tsiantis,M. and Moore,I. (2005) New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *Plant J.*, **41**, 899–918.

59. Karmakar,S., Jin,Y. and Nagaich,A.K. (2013) Interaction of glucocorticoid receptor (GR) with estrogen receptor (ER) alpha and activator protein 1 (AP1) in dexamethasone-mediated interference of ERalpha activity. *J. Biol. Chem.*, **288**, 24020–24034.

60. Sanchez-Alberola,N., Campoy,S., Emerson,D., Barbé,J. and Erill,I. (2015) An SOS Regulon under Control of a Noncanonical LexA-Binding Motif in the Betaproteobacteria *J. Bacteriol.*, **197**, 2622 LP – 2630.

61. Fernández de Henestrosa,A.R., Ogi,T., Aoyagi,S., Chafin,D., Hayes,J.J., Ohmori,H. and Woodgate,R. (2000) Identification of additional genes belonging to the LexA regulon in Escherichia coli. *Mol. Microbiol.*, **35**, 1560–1572.

62. Zuo,Z. and Stormo,G.D. (2014) High-resolution specificity from DNA sequencing highlights alternative modes of Lac repressor binding. *Genetics*, **198**, 1329–1343.

63. Nikolaev,I., Lenouvel,F. and Felenbok,B. (1999) Unique DNA binding specificity of the binuclear zinc AlcR activator of the ethanol utilization pathway in Aspergillus nidulans. *J. Biol. Chem.*, **274**, 9795–9802.

64. Tait,R.C., Close,T.J., Rodriguez,R.L. and Kado,C.I. (1982) Isolation of the origin of replication of the IncW-group plasmid pSa. *Gene*, **20**, 39–49.

65. Antoine,R. and Locht,C. (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from gram-positive organisms. *Mol. Microbiol.*, **6**, 1785–1799.

66. Olsen,A.N., Ernst,H.A., Leggio,L. Lo and Skriver,K. (2005) NAC transcription factors: Structurally distinct, functionally diverse. *Trends Plant Sci.*, **10**, 79–87.

67. Yamaguchi,M., Kubo,M., Fukuda,H. and Demura,T. (2008) VASCULAR-RELATED NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in Arabidopsis roots and shoots. *Plant J.*, **55**, 652–664.

68. Yamaguchi,M., Goué,N., Igarashi,H., Ohtani,M., Nakano,Y., Mortimer,J.C., Nishikubo,N., Kubo,M., Katayama,Y., Kakegawa,K., *et al.* (2010) VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 Effectively Induce Transdifferentiation into Xylem Vessel Elements under Control of an Induction System. *Plant Physiol.*, **153**, 906 LP – 914.

69. Axelos,M., Curie,C., Mazzolini,L., Bardet,C. and Lescure,B. (1992) A protocol for transient gene expression in Arabidopsis thaliana protoplasts isolated from cell suspension cultures. *Plant Physiol Biochem*, **30**, 123–128.

70. Mathur,J., Szabados,L., Schaefer,S., Grunenberg,B., Lossow,A., Jonas-Straube,E., Schell,J., Koncz,C. and Koncz-Kálmán,Z. (1998) Gene identification with sequenced T-DNA tags generated by transformation of Arabidopsis cell suspension. *Plant J.*, **13**, 707–716.

71. Ohashi-Ito,K., Oda,Y. and Fukuda,H. (2010) Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation. *Plant Cell*, **22**, 3461–3473.

72. Oda,Y., Iida,Y., Kondo,Y. and Fukuda,H. (2010) Wood cell-wall structure requires local 2D-microtubule disassembly by a novel plasma membrane-anchored protein. *Curr. Biol.*, **20**, 1197–1202.

73. Teng,S., Keurentjes,J., Bentsink,L., Koornneef,M. and Smeekens,S. (2005) Sucrose-Specific Induction of Anthocyanin Biosynthesis in Arabidopsis Requires the MYB75/PAP1 Gene. *Plant Physiol.*, **139**, 1840 LP – 1852.

74. Pireyre,M. and Burow,M. (2015) Regulation of MYB and bHLH Transcription Factors: A Glance at the Protein Level. *Mol. Plant*, **8**, 378–388.

75. Chupeau,M.-C., Granier,F., Pichon,O., Renou,J.-P., Gaudin,V. and Chupeau,Y. (2013) Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. *Plant Cell*, **25**, 2444–63.

76. Oda,Y. (2017) VND6-induced Xylem Cell Differentiation in Arabidopsis Cell Cultures. *Methods Mol. Biol.*, **1544**, 67–73.

Overall Conclusions

The work described in this thesis created several methods and tools for the development of the field of plant synthetic biology.

DNA assembly is a pivotal technology for the generation of complex multigene constructs in plant biology, and a limiting step for the creation of transgenic plants. I have developed a simple yet versatile new Golden Gate Assembly framework, which was named Mobius Assembly for its (theoretically) unlimited cloning scheme that shifts between two tiers. Mobius Assembly combines a simple vector toolkit while conferring high cloning capacity. It is also equipped with visible cloning screening based on chromogenic proteins that demarcate different levels with distinct colours – a feature which can also be used as a security measure to ensure correct cloning steps in automated or manual cloning. I also opted for a low-frequency cutter *Aar*I to reduce domestication needs. The assembly was highly efficient; in all reactions attempted, cloning was successful in almost all colonies tested (a few exceptions were found when similar genes were combined, having undergone recombination events). Multiple genes were functionally cloned in both operon and individual transcriptional unit arrangements, and thus Mobius Assembly is compatible with bioengineering of both prokaryotic and eukaryotic systems. The streamlined DNA assembly enhances the efficiency and the user-friendliness of molecular cloning using Golden Gate Assembly. It is hoped that the Mobius Assembly will contribute to the establishment of universal standards in synthetic biology constructions. To this end, Mobius Assembly embraced Phytobricks, which are shared among a growing number of synthetic biologists using Golden Gate DNA assembly.

Plant synthetic biology lacks characterised parts for the development of more predictable biosynthetic pathways, as well as high-throughput methods to perform systematic genetic part characterisation. Part characterisation has not been widely accepted as necessary among plant scientists. There are currently only

two existing characterised libraries for higher plant expression: MoClo Plant and G.B3.0. These libraries are also only characterised in tobacco, and they lack diversity in available parts.

To address these issues, I developed a high-throughput gene expression platform based on mesophyll protoplasts, which was then used to characterise new standard parts for Arabidopsis expression. It was shown that the different genetic elements do not behave the same in different plant species (e.g. Arabidopsis and tobacco) or different plant cells from the same species (e.g. mesophyll protoplasts and Arabidopsis cell cultures). There is also a need for the plant community to generate and characterise parts in different plant chassis as, for example, Arabidopsis, tobacco, and barley. Therefore, when a part is being used from a library characterised in one plant species, it should not be assumed that it will behave the same in a different plant.

Concerning the plant part repositories, the sequence size was not considered during the isolation of new parts. Even though most of the transcriptional regulatory elements are confined within 200-300 bp sequences, there is a tendency to use large promoters and terminators of up to 4 kb. That might be tolerable for a single transcriptional unit delivery, but it constitutes an issue for multigene construction and transformation in plants because it creates plasmid instabilities and reduces the plant transformation efficiency. I, therefore, focused on the selection, isolation and characterisation of short genetic elements to ease the construction of complex pathways.

Another conclusion from this thesis is the power of terminators in the control of gene expression. In fact, terminators are neglected, and the attention is primarily focused on promoters. It was demonstrated that terminators have the power to change gene expression considerably. Without this knowledge, researches can unintentionally use a terminator which diminishes gene expression levels. Additionally, in the case of inducible promoters, terminators are among the few tools to manipulate gene expression. We also showed that terminators could act

in synergy with the promoters, and it is incorrect to assume that a terminator can always be strong or weak. In order to collect comprehensive information in gene expression, there should always be a characterisation of promoters and terminators in combination.

In addition to the general characterisation and standardisation of genetic elements and methods in plant biology, further tests and optimisation should be performed. This is particularly important for plant transformation where there should be different levels of quality control, starting from *E. coli*, continuing with Agrobacteria and finishing with the plants, to ensure the proper generation and delivery of the constructs. In the plant community, there is also the tendency to use personal favourites *E. coli* strains, Agrobacterium strains, binary vectors, cloning methods and transformation methods which can reduce the overall efficiency of an experiment. However, there is no gold standard, and in every experiment, all the conditions should be tested beforehand unless the same conditions have been tested by somebody else. Even a single new variation introduced in the system can cause perturbations. For example, for the cloning of complex multigene constructs for plant transformation, it was shown in this study that two main factors should be considered. Firstly, cell strains (*E. coli* and Agrobacteria) can be responsible for low transformation efficiency and plasmid instabilities. Secondly, binary vectors can also contribute to plasmid instabilities and transformation efficacy.

Consequently, the plant toolbox should be equipped with vectors with different origins of replication, as well as with different bacteria strains. To address that, I generated Mobius Assembly vectors with low, medium and high copy numbers to meet with the different experimental requirements. Regions that are prone to instabilities should always be sequenced after every cloning step to prevent the propagation of sequence rearrangements. The plant binary vectors are also responsible for successful plant transformation. Depending on the experimental chassis (plant species, tissues, cell or developmental stage), a different

combination of Agrobacteria and binary vectors might be more suitable in each case, as was shown in the experiments for the MM1 cell line transformation. I also equipped the plant toolbox with a new small binary vector, which can be used for protoplasts, cell cultures and whole plant transformation.

Plant cell cultures will have increasing importance in the following years due to socioeconomic changes in society, which rapidly increases the demand for plant-based products and active compounds. However, there are limited tools and standardised protocols for the genetic manipulation of plant cell cultures, and genetic parts have never been characterised for plant cell culture expression. In this work, we are enabling the easy genetic manipulation of plant cell cultures by showing the impact of the Agrobacteria and binary vectors in cell culture transformation and starting the characterisation of genetic elements in the MM1 cell line.

Notably, it was shown that specific binary vectors (e.g. pGreen) are not suitable for plant cell culture transformation, while some Agrobacterium strains can outperform others (e.g. LBA4404) in their ability to deliver constructs in plant cell cultures. In the future, the library of standard parts that were created should also be characterised in the MM1 cell line, as initial results showed different behaviour from Arabidopsis protoplasts.

I also constructed a guided differentiation system, which aims to exploit the power of differentiation by activating the secondary metabolism for the production of bio compounds in plant cell cultures. For this, I combined two plant inducible systems after ruling out any possibilities for crosstalk. For the development of such guided differentiation platforms, and even more complex ones, there is a need for new, well designed, more compact and tightly regulated inducible systems, which should be characterised on plant cell cultures. Even though the initial results from the proof of concept experiment were not conclusive mainly due to time constraints, the system, together with further optimisation and modifications, will become a powerful tool for the cell culture industry. In future experiments, different

transcription factors (with untagged activation domains) could also be tested with different cell culture types to find the conditions for complete differentiation of the cell into xylem vessels. Replacement of the native activation domains with stronger ones could also favour complete differentiation. Possible generation of synthetic transcription factors, which could bind specific genes related to differentiation, could be another solution, however, that requires excellent knowledge of the genes activated during differentiation. The system could also be expanded for the differentiation of cell cultures into other cell types than xylem vessels to produce other compounds (e.g. synergic or tapetal cells for protein production). Lastly, the sequential system should be tested after the stable transformation of the cell cultures to have a homogeneous population of cells. This will facilitate the observations and quantification of cell differentiation.

Apart from their research and industrial importance, plant cell cultures can become a powerful platform for teaching in biology practical lessons and biology competitions (e.g. iGEM), where time is a limiting factor. Genetic manipulation of plant cell cultures is much faster compared to whole plant stable manipulation, which usually takes months to complete. Consequently, plant cell cultures are becoming an alternative chassis to boost plant synthetic biology.

Appendix 1 Mobius Assembly Protocol

1.1. Summary/Abstract

Mobius Assembly is a versatile and user-friendly DNA Assembly method, which facilitates the rapid and simple generation of DNA constructs. Mobius Assembly combines high cloning capacity and vector toolkit simplicity to streamline combinatorial assemblies. It is a two-level hierarchical modular cloning system that enables quadruple assembly augmentation. It adopts the 4 bp standard overhangs defined by Phytobricks to promote standard part sharing, and it can be made compatible with different chassis. Furthermore, Mobius Assembly reduces domestication requirements and uses chromogenic proteins to facilitate the identification of positive assemblies.

1.2. Introduction

DNA assembly is a fundamental technology in molecular biology to link DNA fragments physically. Golden Gate Assembly exploits Type IIS restriction endonucleases to cut double-stranded DNA outside their recognition sites (1, 2). Consequently, the recognition sites are removed during cloning, while the generated overhangs do not depend on the restriction enzyme. Ligation and digestion are carried out simultaneously, and the unique overhangs are set by users allowing unidirectional, scarless assembly of multiple DNA fragments in a one-tube reaction.

Several variations of Golden Gate Assembly have been developed over the last decade, each addressing specific limitations of the method (3–6). However, most of them compromise at least at one of the following: 1) cloning capacity, 2) vector toolkit simplicity, 3) elimination of internal restriction sites (domestication), 4) standardisation, 5) dependence on specific strains for propagation, and/or 6) negative screening requires additives in the selection medium.

We developed Mobius Assembly to tackle the tradeoffs of the preexisting systems, ensuring user-friendliness (7). Mobius Assembly has a high cloning capacity (exponential assembly) even though the vector toolkit is minimalistic. There is a part storage level (Level 0) and two cloning levels (Level 1 and Level 2). Level 0 has one plasmid (mUAV), which converts a functional DNA fragment (e.g. promoter, coding sequence, tag) into a standard part. In Level 1, there are four Acceptor Vectors (A-Δ), with which up to four transcriptional units (TUs) can be formed in the first cloning round. In Level 2, there are four Acceptor Vectors (A-Δ), as well as seven Auxiliary Plasmids that contain a unique design that switches the overhangs and allows the continuation of the cloning back to Level 1. In Level 2, up to four TUs can be assembled. Cloning can further continue by switching back and forth between the Level 1 and Level 2, leading to an expansion of multi-TUs according to the geometric sequence: 1, 4, 16, 64, … Mobius Assembly uses two Type IIS restriction endonucleases: *Bsa*I (from Level 0 to Level 1 and from Level 2 to Level 1) and *Aar*I (in Level 0 and from Level 1 to Level 2). We used *AarI* since it is a rare cutter (7-bp recognition sequence instead of popularly used 6-bp cutters to reduce domestication requirements).

Mobius Assembly further simplifies cloning protocols via negative selection screening carried out by constitutively expressed chromogenic proteins. Each level is demarcated by a specific colour (purple: amilCP-Level 0, pink: spisPink-Level 1, and yellow: sfGFP-Level 2). This feature eliminates the need for IPTG and X-Gal and expensive blue/white screening.

In order to promote standard part sharing, Mobius Assembly adopts the Phytobrick common syntax, a predefined set of 4 bp overhangs flanking the different types of standard parts (**Figure 1**,(8)). Recently, to simplify the generation of combinatorial assembly libraries, we introduced a new feature based on DNA methylation. Mobius Assembly toolkits are available for *E. coli* and plant systems so far; however, it can be easily adapted to different chassis by transferring the Mobius Assembly cassettes to different backbones.

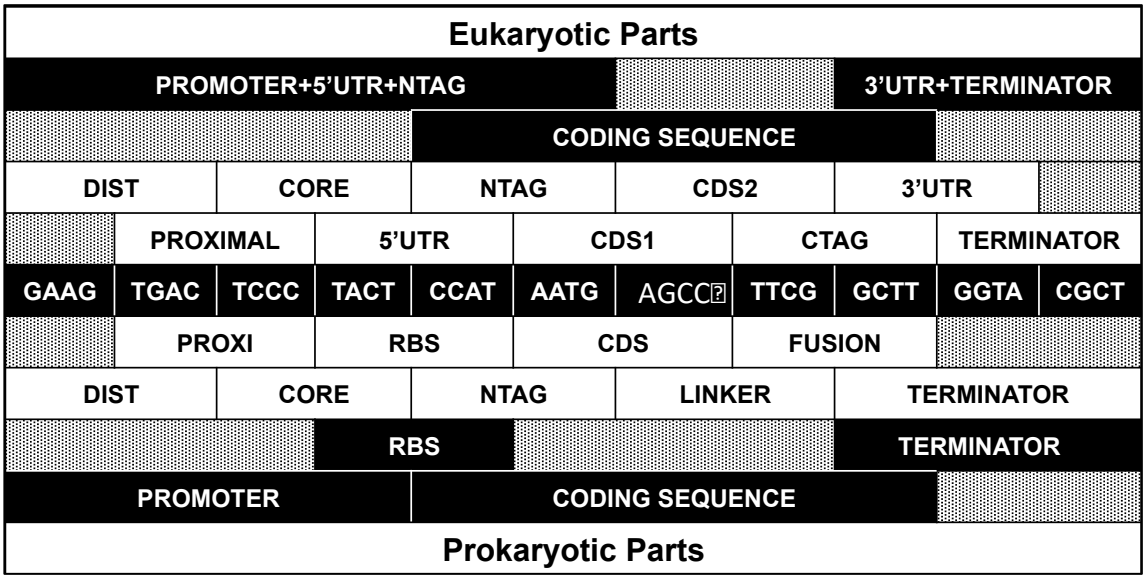


Figure 1. Phytobrick overhangs. Phytobrick are standard DNA parts which contain a common genetic syntax for their flanking overhangs (NNNN). The standardized syntax facilitates part sharing within the community. Top half depicts the major and minor functional parts of Phytobricks for eukaryotic genes, while the bottom half denotes the ones for prokaryotic genes.

1.3. Materials

1.3.1. Primer design/DNA synthesis

1. DNA construct design tool (e.g. SnapGene; Benchling, which is free for academic use).
2. DNA synthesis services (e.g. Twist Bioscience; Integrated DNA Technologies) (Optional).
3. Oligo synthesis service (e.g. Integrated DNA Technologies; Sigma-Aldrich).

1.3.2. Molecular Biology Techniques

1. Milli-Q sterilised water.

2. DNA Template (genomic DNA; complementary DNA; cDNA; plasmid DNA).
3. High-Fidelity DNA Polymerase (e.g. Q5® High-Fidelity DNA Polymerase, NEB).
4. 200 µl PCR tubes.
5. 1.5 ml microcentrifuge tubes.
6. Agarose gel: 1% agarose in 1xTAE (40mM Tris-Base, 20mM acetic acid and 1mM EDTA).
7. Gel Imaging System.
8. PCR Purification Kit (e.g. Monarch® Nucleic Acid Purification Kits, NEB).
9. (Optional) Gel Purification Kit (e.g. Monarch® DNA Gel Extraction Kit, NEB).
10. Plasmid Miniprep Kit (e.g. Monarch® Miniprep Kit, NEB; PureYield™ Plasmid Miniprep System, Promega).
11. Plasmid Midiprep Kit (e.g. PureYield™ Plasmid Midiprep System, Promega).
12. Micro-Volume Spectrophotometer.
13. DNA ladder (e.g. HyperLadder™ 1kb, Bioline)
14. Lysogeny Broth (LB) Medium: 1% tryptone, 0.5 %yeast extract, and 1% NaCl adjusted to pH 7.0. For plates, 1.5% agar is added.
15. Antibiotics (level and backbone-specific).
16. Home-Made competent cells (e.g. DH10B, DH5α, JM109, TOP10) prepared with the TSS method (9).
17. Super Optimal broth with Catabolite repression (S.O.C) medium: 2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 10mM MgSO4, and 20mM glucose.
18. Incubator (plate or shaking).
19. *PstI*-HF and *EcoRI*-HF (NEB).
20. Primers for insert verification and sequencing:
    VF2 (TGCCACCTGACGTCTAAGAA)

and VR (ATTACCGCCTTTGAGTGAGC)

21. Sequencing service.

1.3.3. Mobius Assembly cloning

1. Milli-Q sterilised water.
2. Thermocycler.
3. BSA, Molecular Biology Grade (NEB).
4. T4 DNA Ligase and 10X T4 DNA Ligase Buffer (e.g. Thermo Fisher Scientific, NEB).
5. For Level 0 and Level 2 cloning: OPTIZYME™ *AarI*, Fisher BioReagents™; Thermo Scientific™ *AarI* (both supplied with 50x Oligos).
6. For Level 1 cloning: *Eco31I,* Thermo Fisher Scientific; *BsaI*-HF®v2, NEB.

Plasmid construction and methylation

1. Isothermal Assembly kit (e.g. Gibson Assembly®, NEB; NEBuilder® HiFi DNA Assembly, NEB).
2. *DpnI* (e.g. NEB, Life Technologies).
3. Thermocycler.
4. CpG Methyltransferase (*M.SssI*) (NEB).

1.4. Methods

1.4.1. Primer design/DNA synthesis

DNA fragments should be designed to carry the selected standard overhangs and *AarI* restriction sites for cloning into the Mobius Universal Acceptor Vector (mUAV) (**Figure 2**). They can either be synthesised from a DNA synthesis company, or PCR amplified from genomic (see Note 1) or cDNA. For primer design or DNA synthesis, follow the guidelines below:

1. Select your sequence and open it in a DNA visualiser/editor software.

2. Add the sequence 5'-T**CACCTGC**ATATCTCT*NNNN*-3' at the beginning (5' end) of your DNA fragment and 3'-*NNNN*TGAGATAT**GCAGGTG**T-5' at the end (3' end). **CACCTGC** and **GCAGGTG** are the *AarI* recognition sites, CTCT and TGAG (the 5' and 3' overhangs) which pair with the overhangs of mUAV and *NNNN,* the Phytobrick Overhangs, respectively (Table 1).
3. For primer design, Step 2 should be modified to Forward Primer: 5'-T**CACCTGC**ATATCTCT*NNNN*+(17-25 bp) -3' and Reverse Primer: 3'-(17-25 bp) +*NNNN*TGAGATAT**GCAGGTG**T-5'.
4. For non-coding sequences, check the Phytobrick specification (Table 1) to select the overhangs (*NNNN)* according to the type of the part (e.g. GGAG-AATG for a promoter).
5. For coding sequences, in addition to the overhangs, further rules should be considered:
    a) If the 5' overhang starts with AATG, add one A next to the native ATG start codon to form the overhang.
    b) If the 5' overhang starts with CCAT remove the native AT from the sequence to form the overhang.
    c) If the 3' overhang stops with GCTT, be sure there is a stop codon right before.
    d) If the 3' overhang stops with AGCC or TTCG, remove the stop codon right before. GCC and TCG encode for alanine and serine, respectively, so A or T from the overhang should be the last amino acid of a triplet to preserve the reading frame. This can be done by a) completely removing the stop codon and adding two extra bases (GG, GC, TC, or AG), which encode for small amino acids (GGA, GGT == glycine; GCA, GCT == alanine; TCA, TCT, AGT == serine). b) completely removing the stop codon and the last base pair from the codon before, if the two remaining base pairs combined with A (for AGCC) or T (for TTCG) do not change the amino acid.

6. Check the sequence for illegal recognition sites (**CACCTGC == *AarI*, GGTCTC == *BsaI***). If the part is a coding sequence, remove the illegal recognition site with a silent point mutation (domesticate) (see Note 2).
7. If the part is a noncoding sequence (e.g. promoter), try to avoid changing known functional elements (e.g. transcription factor binding sites). If it is not possible, ensure that the mutation does not affect the function.
8. If you PCR amplify the part, you should introduce the point mutation by splitting the sequence, forming overhangs (see Note 3) and adding *AarI* recognition sites. The point mutation should be introduced in the overhangs (see **Figure 2**).
9. If you PCR amplify the sequence, run 2 µl of the reaction in gel electrophoresis to verify the presence and the correct size of the fragment.
10. Purify the PCR products using a column-based kit or carry out gel purification if there are multiple bands.

1.4.2. Level 0 cloning

In Level 0 cloning, DNA fragments generated as above are cloned into mUAV to form Phytobricks (**Figure 2**). Negative cloning screening is performed by dropout of amilCP (see Note 4); the successfully cloned constructs will result in white colonies, the unsuccessful in purple. Set up the Level 0 reaction as follows (see Note 5):

1. Total reaction volume is 10 µl (see Note 6). Calculate the volumes of the reagents and add the sterile distilled water first.
2. Add 50 ng of mUAV. We usually add 0.5 µl of 100 ng/µl stock.
3. Add the insert(s). Follow the 2:1 insert:vector ratio.
4. Sequentially add: 1 µl BSA, 1 µl T4 DNA ligase Buffer (see Note 7), 0.5 µl T4 DNA ligase, 0.5 µl *AarI*, and 0.2 µl 50X *AarI* oligonucleotides (0.025 mM).
5. Gently mix well and spin down in a microcentrifuge.

6. Incubate in a thermal cycler with the program: 5x (5 min at 37ºC + 10 min at 16 ºC), 5 min at 37ºC, 5 min at 80ºC and hold at 10ºC
7. Transform 5 µl of the reaction to 100 µl of competent cells. Incubate for 30 min on ice, heat-shock in a 42ºC heat block or water bath for 90 sec, cool for 3 min on ice, add 400 µl SOC media and incubate in a 37ºC shaker for 1 hour.
8. Plate 100 µl of the transformation culture on each LB agar plate containing 25 µg/ml chloramphenicol and incubate overnight at 37ºC.
9. Run colony PCR for 3 white colonies to confirm the size of the part (Optional).
10. Pick one colony for LB culture containing 25 µg/ml chloramphenicol and incubate overnight in a 37ºC shaker.
11. Miniprep the cultures and verify the construct by restriction digestion analysis with *PstI*-HF and *EcoRI*-HF.
12. Send the part for sequencing using the primers VF2 and VR.

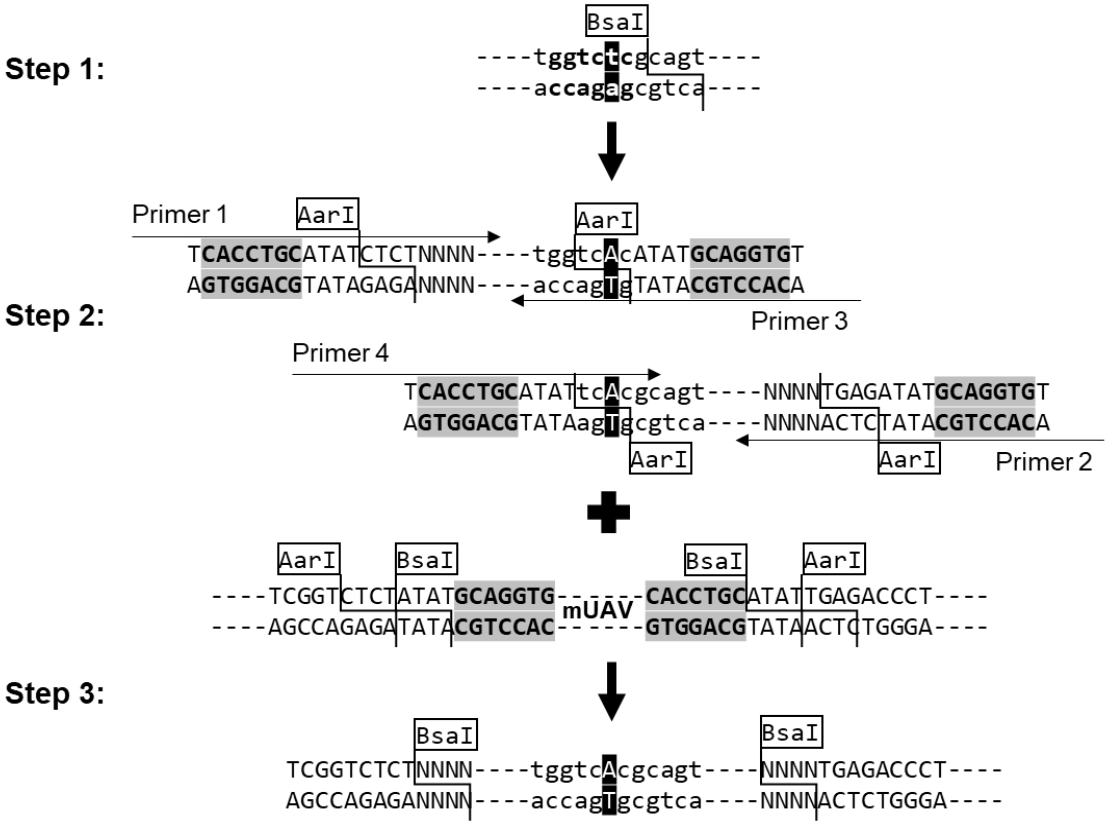


Figure 2. Sequence domestication and standard part (Phytobrick) cloning into mUAV. The coding sequence shown here bears an internal *BsaI* restriction site, which needs to be removed (i.e. domesticated). In Step 1, remove *in silico* the recognition site by changing A to T, resulting in a silence mutation of serine (TCT to TCA). In Step 2, split the sequence into two parts, both of which share an overhang that brings together the mutation (tcAc). At the split point add the *AarI* recognition sequence (GCAGGTG). Also, at the 5' of the first fragment and the 3' of the second fragment, add the *AarI* recognition sequence, the overhangs for cloning into mUAV (CTCT-TGAG), and the Phytobrick overhangss (NNNN). We design primers and PCR amplify the sequence in two parts (Primer 1 and 3 for the first fragrment and Primer 4 and 2 for the second fragment. In the last step (Step 3), the two fragments are fused together and cloned into the mUAV in one tube-reaction, generating a Phytobrick.

1.4.3. Level 1 cloning

In the first round of Level 1 cloning, Phytobricks from Level 0 are combined in a Level 1 Acceptor Vector to form a Transcriptional Unit (TU) (**Figure 3**). The second round of Level 1 cloning can be done by assembling multi-TUs back from

Level 2. There are four Level 1 Acceptor Vectors, A, B, Γ and Δ. We always start with the first acceptor vector (A) and sequentially use the other ones when we need to assemble multiple TUs. The successful assembly will result in white colonies because of the replacement of the pink chromoprotein spisPink, which is the negative screening marker. The backbones for Level 1 cloning cannot have chloramphenicol resistance, and they should have different antibiotic resistance from Level 2 Vectors. Set the Level 1 reaction as follows:

1. Total reaction volume is 10 µl. Calculate the volumes of the reagents and add the sterile distilled water first.
2. Add 20 fmol of Level 1 Acceptor Vector.
3. Add 40 fmol of the Phytobricks.
4. Add sequentially: 1 µl BSA, 1 µl T4 DNA ligase Buffer, 0.5 µl T4 DNA ligase and 0.5 µl *BsaI*.
5. Gently mix well and spin down in a microcentrifuge.
6. Incubate in a thermal cycler with the program: 5-10x (5 min at 37ºC + 10 min at 16ºC) (see Note 8), 5 min at 37ºC, 5 min at 80ºC and hold at 10ºC.
7. Transform 5 µl of the reaction to 100 µl of competent cells. Incubate for 30 min on ice, heat shock in a 42ºC heat block or water bath for 90 sec, cool for 3 min on ice, add 400 µl SOC media and incubate in a 37ºC shaker for 1 hour.
8. Plate 100 µl of the transformation on LB agar plates containing the appropriate antibiotic and incubate overnight at 37ºC.
9. Run colony PCR for 3 white colonies to confirm the size of the part (Optional).
10. Pick one colony for LB culture containing the appropriate and incubate overnight at 37ºC.
11. Miniprep the cultures and verify the construct by restriction digestion analysis with *PstI*-HF and *EcoRI*-HF.
12. Submit samples to sequencing only if:
    a. There are tandem repeats on the construct, and you need to check possible deletions.

b. The cloning in Level 2 is failing due to the rare scenario in which there are mutations in the overhangs or the *AarI* recognition sites.

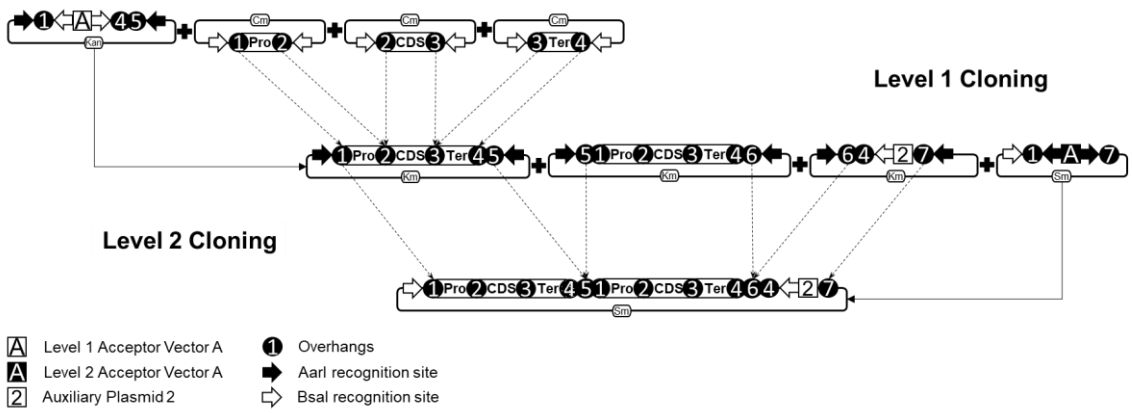


Figure 3. Level 1 and Level 2 cloning. Phytobricks are released by *BsaI* digestion and assembled in the Level 1 reaction to form a TU. The resulting TUs are then released by *AarI* and assembled in the Level 2 reaction to form multi-TU constructs. Level 2 cloning is assisted by Auxiliary Plasmids, which provide the overhangs for cloning back to Level 1. In this example two transcriptional units are assembled. When necessary, the assembly can further continue by switching back and forth between the two levels, quadrupling the number of TUs every time.

1.4.4. Level 2 cloning

In Level 2 cloning, multiple TUs from Level 1 are fused in a Level 2 Acceptor Vector with the help of the Auxiliary Plasmids (**Figure 3**). There are four Level 2 Acceptor Vectors, A, B, Γ and Δ. Every Acceptor Vector in Level 2 can take up to four TUs from Level 1. The backbones for Level 2 cloning should have different antibiotic resistance from the one for Level 1, and it should not be Kanamycin resistance because Auxiliary plasmids confer Kanamycin resistance. We always start with the Acceptor Vector A and sequentially use the other ones. The successful assembly will result in white colonies because of the inserts replace the yellow pigmentation by *sfGFP* (see Note 9). The Auxiliary Plasmids provide missing overhangs and help the Level 2 cloning (See examples below). Set the Level 2 reaction as follows:

1. Total volume reaction is 10 µl (see Note 10). Calculate the volumes of the reagents and add the sterile distilled water first.
2. Add 20 fmol of Level 2 Acceptor Vector.
3. Add 40 fmol of Level 1 TUs.
4. Add 40 fmol of the appropriate Auxiliary plasmid.
5. Add sequentially: 1 µl BSA, 1 µl T4 DNA ligase Buffer, 0.5 µl T4 DNA ligase, 0.5 µl *AarI*, and 0.2 µl 50X *AarI* oligonucleotide (0.025 mM).
6. Gently mix well and spin down in a microcentrifuge.
7. Incubate in a thermal cycler with the program: 5-10x (5 min at 37ºC + 10 min at 16 ºC), 5 min at 37ºC, 5 min at 80ºC and hold at 10ºC
8. Transform 5 µl of the reaction to 100 µl of competent cells. Incubate for 30 min on ice, heat shock in a 42ºC heat block or water bath for 90 sec, cool for 3 min on ice, add 400 µl SOC media and incubate in a 37ºC shaker for 1 hour.
9. Plate 100 µl of the transformation on LB agar plates containing the appropriate antibiotic and incubate overnight at 37ºC.
10. Run colony PCR for 3 white colonies to confirm the size of the part (Optional).
11. Pick one colony (see Note 11) for LB culture containing the appropriate and incubate overnight at 37ºC.
12. Miniprep the cultures and verify the construct by restriction digestion analysis with *PstI*-HF and *EcoRI*-HF.
13. Sequencing is necessary only if:
    a. There are tandem repeats on the construct, and you need to check possible deletions.
    b. The cloning in Level 2 is failing due to the rare scenario in which there are mutations in the overhangs or the *BsaI* recognition sites.

1.4.5. Examples for how to use the Auxiliary Plasmids

The Auxiliary Plasmids 4A, 4B, 4Γ and 4Δ are used when 4 TUs are assembled in Level 2, and they correspond to the four Level 2 Acceptor Vectors. For example, if you want to make a 4-TU construct comprising from the TUs ***a****, **b**, **c** and **d***, first you assemble your TUs in Level 1 Acceptor Vectors as follows:

TU *a*	Level 1 Vector A
TU *b*	Level 1 Vector B
TU *c*	Level 1 Vector Γ
TU *d*	Level 1 Vector Δ

Then, TUs ***a****, **b**, **c** and **d*** will be cloned in the Level 2 Vector A to form the 4-TU ***abcd,*** and in the reaction, you will use **Auxiliary Plasmid 4A** (as you assemble 4-TUs in Level 2 Acceptor Vector A).

TU a + TU b + TU c + TU d + Aux 4A	Level 2 Vector A

When you clone fewer than four TUs in your Level 2 Acceptor Vector, you use the **Auxiliary Plasmids 1, 2 and 3** corresponding to the number of TUs you clone: one, two or three.

Let us say you want to assemble 3 TUs ***a, b and c***. Again, you will first clone them in Level 1 Acceptor Vectors:

TU *a*	Level 1 Vector A
TU *b*	Level 1 Vector B
TU *c*	Level 1 Vector Γ

As before TUs ***a****, **b and c*** will be cloned in the Level 2 Vector A to create the 3-TU construct ***abc,*** but this time you will add the Auxiliary plasmid 3 in the reaction (as you are assembling 3 TUs).

TU a + TU b + TU c + Aux 3	Level 2 Vector A

A more complex example: Let us say you want to make a 10-TU construct comprising of the TUs ***a, b, c, d, e, f, g, h, i, j***. In the first cloning round you will assemble each of your TUs in a Level 1 Acceptor Vector:

TU *a*	Level 1 Vector A
TU *b*	Level 1 Vector B

TU *c*	Level 1 Vector Γ
TU *d*	Level 1 Vector Δ
TU *e*	Level 1 Vector A
TU *f*	Level 1 Vector B
TU *g*	Level 1 Vector Γ
TU *h*	Level 1 Vector Δ
TU *i*	Level 1 Vector A
TU *j*	Level 1 Vector B

In the second cloning round, TUs ***a, b, c, d*** will be cloned in Level 2 Acceptor Vector A (+ Auxiliary 4A, as you assemble 4-TUs in Level 2 Acceptor Vector A), TUs ***e, f, g, h*** in Level 2 Acceptor Vector B (+ Auxiliary 4B, as you assemble 4-TUs in Level 2 Acceptor Vector B) and TUs ***i, j*** in Level 2 Acceptor Vector Γ (+ Auxiliary 2, as you assemble 2 TUs):

TU a + TU b + TU c + TU d + Aux 4A	Level 2 Vector A
TU e + TU f + TU g + TU h + Aux 4B	Level 2 Vector B
TU i + TU j + Aux 2	Level 2 Vector Γ

Lastly, in the third cloning round the multi-TUs ***abcd***, ***efgh***, and ***ij*** will be assembled in a Level 1 Vector A to form the 10-TU construct.

4-TU abcd + 4-TU efgh + 2-TU ij	Level 1 Vector A

Exemption to this rule applies when we make constructs of **eight** or **twelve** TUs, where in the last Level 2 Acceptor Vector, we **always** add **Auxiliary Plasmid 4Δ,** which provides the necessary overhangs for the cloning back to Level 1. Cloning in Level 1 is skipped below and directly showing the assembly in Level 2.

Eight-TU construct:

TU a + TU b + TU c + TU d + Aux 4A	Level 2 Vector A
TU e + TU f + TU g + TU h + Aux 4Δ	Level 2 Vector B

Which back in Level 1 will form the 8-TU ***abcdefgh***:

4-TU abcd + 4-TU efgh	Level 1 Vector A

And twelve-TU construct:

TU a + TU b + TU c + TU d + Aux 4A	Level 2 Vector A
TU e + TU f + TU g + TU h + Aux 4B	Level 2 Vector B

TU i + TU j + TU k + TU l Aux 4Δ	Level 2 Vector Γ

Which back in Level 1 will form the 12-TU ***abcdefghijkl***

4-TU abcd + 4-TU efgh + 4-TU ijkl	Level 1 Vector A

1.4.6. MethylAble feature

The assembly of combinatorial libraries is a tedious task when it is not fully automated. Numerous constructs need to be built for each level. We developed an additional feature based on the methylation of the *BsaI* recognition site, to bypass the Level 1 cloning and directly feed into Level 2 with standard parts (**Figure 4**). According to Rebase, CGGTCTC^{m5}G/GC$_{m5}$CAGAGC methylation provides strong protection against *BsaI* digestion. We exploited this property to design an *amilCP* gene flanked by standard overhangs, which are bordered by inward and outward-facing *BsaI* restriction sites. The outward restriction sites are designed to be susceptible to CpG methylation (CGGTCTC^{m5}G/GC$_{m5}$CAGAGC) and thus protected from *BsaI* digestion while the inward-facing sites are not (TGGTCTC^{m5}T/AC$_{m5}$CAGAGA). Consequently, the *amilCP* gene is propagated to Level 2 with the intact outward-facing *BsaI* sites and overhangs. The Level 2 construct can then be fused with the Level 0 library of Phytobricks in a Level 1 reaction. The MethylAble feature can be designed for any standard part or combination of parts (see Note 12).

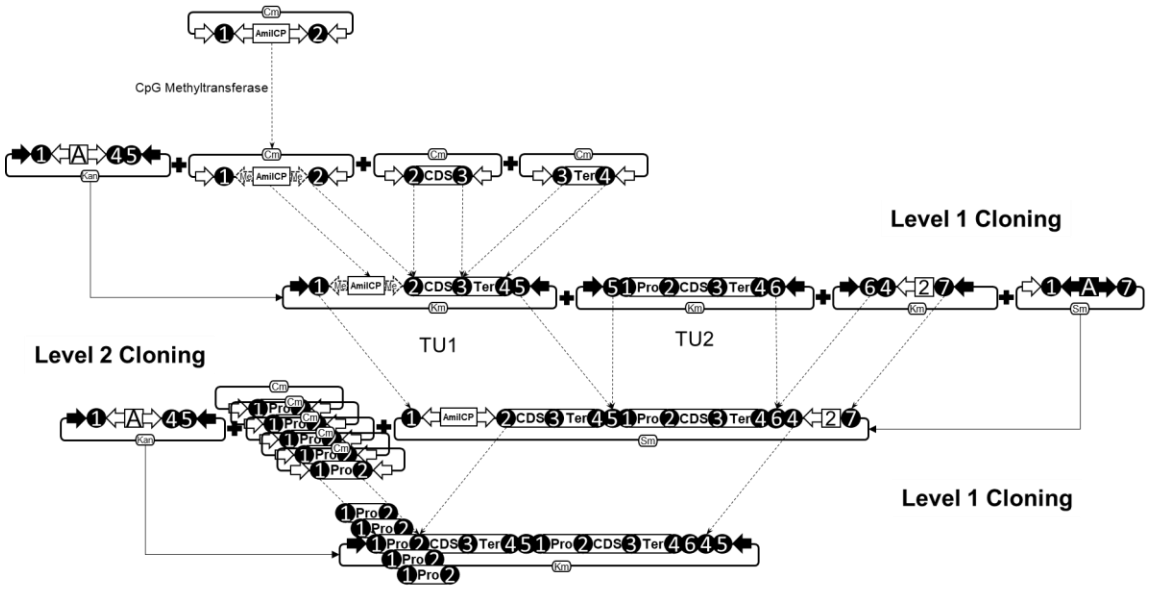


Figure 4. MethylAble feature. It is a feature that ease the generation of combinatorial libraries and enables direct assembly of standard parts with Level 2 constructs. The MethylAble plasmid has outward facing BsaI sites that are designed to be blocked with CpG methylation and are propagated to Level 2 constructs. It has the cognate overhangs of the part that replaces and amiCP as screening marker. In this way, a library of standard parts can be fused with a multi-TU in Level 2 to create a library of TUs, bypassing the generation Level 1 constructs. In this example the MethylAble plasmid takes the place of the promoter, which is assembled with a coding sequence and a terminator to form the TU1 in Level 1. In Level 2, the TU1 is fused with the TU2 generating a 2-TU construct. This construct is finally combined with a library of

MethylAble plasmids are built via isothermal assembly:

1. PCR amplify the *amilCP* cassette to insert the appropriate standard overhangs and the *BsaI* recognition sites. Use mUAV as the template and the primers: Forward:
   5’-TCAGGTCTCT**NNNN**CGAGACCGTTTACGGCTAGCTCAGTC-3’
   and Reverse:
   3’- GGGTGGGCCTTTCTGCGCGGTCTCG**NNNN**TGAGACCCTG-5’.
2. PCR amplify the backbone using mUAV as the template and the primers: Forward:
3. 5’-GCGCGGTCTCG**NNNN**TGAGACCCTGCAGTCCGGCAAAAA-3’

and Reverse:

3’-AAGGATGATTTCTGGAATTCAGGTCTCT**NNNN**CGAGACCGTT-5’.

4. Analyze 2 µl of the reaction by gel electrophoresis, to verify the correct sizes of the fragments.
5. Mix 8 µl of PCR reaction with 1 µl of 10X Cutsmart Buffer and 1 µl of *DpnI*.
6. Incubate at 37°C for 30 min and deactivate at 80°C for 20 min.
7. PCR purify the reaction.
8. Proceed with the isothermal assembly according to the manufacturer’s instructions.
9. Transform 5 µl of the reaction to 100 µl of competent cells. Incubate for 30 min on ice, heat shock in a 42^{o}C heat block or water bath for 90 sec, cool for 3 min on ice, add 400 µl SOC media, and incubate in a 37^{o}C shaker for 1 hour.
10. Plate 200 µl of the transformation on LB agar plates containing 25 µg/ml chloramphenicol and incubate overnight at 37^{o}C.
11. Pick a purple colony for LB culture containing 25 µg/mL chloramphenicol and incubate overnight at 37^{o}C.
12. Miniprep the cultures.
13. Send the plasmid for sequencing using VF2 and VF primers to verify the restriction sites and overhangs.

Before applying the MethylAble feature, you need to *in vitro* methylate the plasmid.

1. Isolate plasmid DNA using midi-prep kit and following the manufacturer’s instructions.
2. In a PCR tube add nuclease-free water to make the final volume 20 µl.
3. Add 2 µl methyltransferase Reaction Buffer (10X).
4. Add 2 µl of 6.4mM SAM (diluted from the stock).
5. Add 1-2 µg of plasmid DNA (no more than 5 µl).

6. Add 2 µl of CpG methyltransferase.
7. Incubate for 4h at 37°C and stop the reaction by heating at 65°C for 20 min.
8. PCR purify the reaction.

The methylated plasmid is ready to be used in Mobius Assembly cloning. Let us say you are building a promoter library to test the strengths of 20 previously uncharacterised promoters in plant systems. Promoters will be driving the expression of GFP, which is normalised by the expression of the reference protein RFP. The final constructs will be PromoterX:GFP:HSP-UBQ10:RFP:MAS, which is comprised of two TUs. Without MethylAble, you will need to make 20 constructs of PromoterX:GFP:HSPter in the Level 1 Acceptor Vector A and one construct, UBQ10pro:RFP:MASter, in the Level 1 Acceptor Vector B. Then, in 20 more Level 2 reactions, the two TUs will be combined. Instead, the methylated plasmid will replace the promoter in the Level 1 reaction, resulting in just one construct, amilCP:nluc:HSPter. You will then select for purple colonies (due to the expression of amilCP) and isolate the plasmid DNA, which will be used for the Level 2 reaction to form one multi-TU construct, amilCP:nluc:HSPter:UBQ10pro:fluc:UBQ5ter. Lastly, the assembly will finish in a Level 1 reaction. You can make a master mix of all the reagents for 20 reactions, along with a Level 1 Acceptor Vector A and amilCP:nluc:HSPter-UBQ10pro:fluc:UBQ5ter. Aliquot the master mix and add one of the 20 Level 0 promoters in each PCR tube.

1.4.7. Plasmid construction

For different chassis or experimental requirements, Mobius Assembly cassettes can be adapted to any backbone. They can be either PCR amplified or synthesised. This can be done for the whole toolkit or just for the final destination vectors. The easiest way to swap backbones is by the isothermal assembly which joins overlapping DNA fragments.

1. Retrieve the sequences of Mobius Assembly cassettes from https://www.addgene.org/kits/nakayama-mobius-assembly-toolkit/.
2. Select your destination vector(s) and open it in your DNA editing software.
3. Copy the Mobius Assembly cassette sequence and paste it into the destination vector. A Level 1 cassette is the sequence between 5'-GGAATTCCACCTGCATAT-3' and 3'- ATATGCAGGTGCTGCAG -5'; A Level 2 cassette is the sequence between 5'-TTCTGGAATTCGGTCTCA-3' and 3'-CTGCATATACCCCTGCAG-5'.
4. Design primers to PCR amplify the Mobius Assembly cassettes. For Level 1 vectors: Forward: 5'-(~10bp)+GGAATTCCACCTGCATAT and Reverse: 3'-ATATGCAGGTGCTGCAG+(~10bp)-5'. For Level 2 vectors: Forward: 5'-(~10bp)+TTCTGGAATTCGGTCTCA-3' and Reverse: 3'-CTGCATATACCCCTGCAG+(~10bp)-5'. The 10bp sequences come from the destination vector backbone, and they will form the 20bp overlapping DNA fragments.
5. Design primers to PCR amplify the selected backbone. For Level 1 vectors: Forward: 5'- GGTGCTGCAG+(17-25bp)-3' and Reverse 3'-(17-25bp)+GGAATTCCAC-5'. For Level 2 vectors: Forward: 5'-ACCCCTGCAG+(~17-25bp)-3' and Reverse (17-25bp)+TTCTGGAATT.
6. Set the PCR reaction following the manufacturer's instructions. Usually, an annealing temperature of 55ºC and 30 cycles work well for the amplification.
7. Analyse 2 µl of the reaction by gel electrophoresis to verify the correct size of the products.
8. If the destination vector shares the same antibiotic resistance with the donor plasmids, perform *DpnI* digestion:
    a. Mix 8 µl of PCR reaction with 1 µl of 10X Cutsmart Buffer and 1 µl of *DpnI*
    b. Incubate at 37°C for 30 min and deactivate at 80°C for 20 min.
9. PCR purify the reaction (optional).

10. Proceed with the isothermal assembly following the manufacturer's instructions.
11. Transform 5 µl of the reaction to 100 µl of competent cells. Incubate for 30 min on ice, heat shock in a 42^{o}C heat block or water bath for 90 sec, cool for 3 min on ice, add 400 µl SOC medium and incubate in a 37^{o}C shaker for 1 hour.
12. Plate 100-400 µl of the transformation on LB agar plates containing the appropriate antibiotic and incubate overnight at 37^{o}C (see Note 13). If the reaction is not efficient, you will need to plate the entire content of the tube.
13. Pick a pink (Level 1 Vectors) or yellow (Level 2 Vectors) colony for LB culture containing the appropriate antibiotics and incubate overnight at 37^{o}C.
14. Miniprep the cultures and verify the plasmid by restriction digestion analysis with *PstI*-HF and *EcoRI*-HF.
15. Send the plasmid for sequencing to verify the Mobius Assembly restriction sites and overhangs.

1.5. Mobius Assembly Troubleshooting

The successful assembly will result in mostly white colonies. Bear in mind that generation of complex and large constructs will reduce the number of white colonies. When you face problems with your cloning, follow the steps below to troubleshoot.

1. If there are no or very few colonies on the plates:
    a. Check that you used the correct antibiotic for selection.
    b. Check the competency of your cells.
    c. Change the *E. coli* strain (see Note 14)
    d. Use fresh T4 DNA ligase and/or T4 DNA buffer.
    e. Check if there is a mutation in the overhangs of the vector or the enzyme recognition site (a rare scenario).

2. If there are many negative colonies and few positives, use fresh restriction enzymes.
3. For 1 and 2 and especially for the assembly of large constructs:
    a. Check that you added the right plasmids, enzymes, and buffers for each reaction.
    b. Use more digestion/ligation cycles, which can be combined with halving digestion/ligation times.
    c. Be sure that you use the recommended molar ratios.
    d. Be sure about the DNA quantification (see Note 15).
    e. Double the amount of the enzymes.
    f. Increase the volume of the reaction to dilute possible reaction inhibitors.
    g. Avoid old and possibly degraded plasmid DNA.
4. Toxic or unstable inserts might cause mutations in the assembled constructs.
    a. Screen more colonies.
    b. Try different *E. coli* strains.
    c. Use a backbone with a low copy number origin of replication.
    d. Incubate at lower temperatures.
    e. Place toxic coding sequences under an inducible promoter.

1.6. Notes

1. Sometimes amplification from genomic DNA might be tricky, and you will need to first amplify the fragment with primers without the Mobius Assembly extension sequence.
2. Avoid using rare codons. Refer to the codon usage table of the organism you are working with.
3. Avoid using the same overhangs as Phytobricks or palindromic sequences.
4. To isolate plasmid DNA from any Acceptor Vector, use a single colony from a streaked agar plate to inoculate a liquid culture. If glycerol stock is used

directly as inoculum, the colour of the negative selection marker sometimes does not properly develop.

5. Before proceeding to an in vitro experiment in every level, always perform an *in-silico* simulation using your vector mapping software to check that you use the correct parts, vectors and overhangs.
6. If the concentration of the DNA is low or the parts are several, you might need to increase the volume of the reaction to 15 or 20 μl.
7. Always aliquot the T4 DNA ligase buffer to avoid repetitive thaw-freeze cycles which degrade ATP.
8. For the first round of Level 1 reaction, 5 cycles are adequate unless several parts with considerable size differences are being assembled, where the number of the cycles should be increased (up to 10). For the further rounds of Level 1 reaction, the number of the cycles should be increased as the constructs become larger.
9. For the isolation of Level 2 Acceptor Vectors (especially with high copy number backbones), the use of Monarch® Plasmid Miniprep Kit or PureYield™ Plasmid Miniprep (Promega) is recommended. We tested GeneJET Plasmid Miniprep Kit (Thermofisher) and QIAprep® Spin Miniprep Kit (Qiagen), but the quantity and purity of the DNA were low since sfGFP binds on the membrane of the columns and co-elute with the plasmid DNA. For midi-prep, we use the PureYield™ Plasmid Midiprep System (Promega) without any issues.
10. If the concentration of the DNA is low or you are cloning several parts, you might need to increase the volume of the reaction to 15 or 20 μl.
11. For large and complex constructs, it is recommended to screen at least two colonies.
12. The only limitation of this feature is that MethylAble parts starting with GGAG (promoter) should be placed at the beginning of the construct (in the first TU) and parts ending with CGCT (terminator) should be placed at the end (in the fourth TU). This rule prevents incorporating extra GGAG or

CGCT overhangs in the constructs, as all the Mobius Assembly Acceptor vectors already have those overhangs at the beginning (GGAG) and the end (CGCT) of the cloning cassettes.

13. Sometimes, high copy number vectors expressing spisPink do not grow well at 37ºC right after the isothermal assembly. Incubating at 30ºC or different *E. coli* strains may help.
14. We noticed that some backbones give no or very few colonies when having large inserts and transformed into specific strains. (eg. pGreen with TOP10 cells).
15. RNA and genomic DNA contaminations lead to overestimation of the plasmid concentration. They can be spotted on agarose gel electrophoresis as low molecular weight bands for the former and smear for the latter. RNA contamination occurs when RNase in the buffer is not working, and genomic DNA contamination may result from the shearing of host chromosomal DNA.

1.7. References

1. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*, **4**, e5553.

2. Engler,C., Kandzia,R. and Marillonnet,S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3**.

3. Sarrion-Perdigones,A., Vazquez-Vilar,M., Palaci,J., Castelijns,B., Forment,J., Ziarsolo,P., Blanca,J., Granell,A. and Orzaez,D. (2013) GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *Plant Physiol.*, **162**, 1618–1631.

4. Weber,E., Engler,C., Gruetzner,R., Werner,S. and Marillonnet,S. (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One*, **6**, e16765.

5. Moore,S.J., Lai,H.E., Kelwick,R.J.R., Chee,S.M., Bell,D.J., Polizzi,K.M. and Freemont,P.S. (2016) EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.*, **5**, 1059–1069.

6. Iverson,S. V., Haddock,T.L., Beal,J. and Densmore,D.M. (2016) CIDAR MoClo:

Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. *ACS Synth. Biol.*, **5**, 99–103.

7. Andreou,A.I. and Nakayama,N. (2018) Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS One*, **13**, e0189892.

8. Patron,N. and et al (2015) Standards for Plant Synthetic Biology: A Common Syntax for Exchange of DNA Parts. *New Phytol.*

9. Chung,C.T. and Miller,R.H. (1993) Preparation and storage of competent Escherichia coli cells. *Methods Enzymol.*, **218**, 621–627.

Appendix 2 Oligonucleotide sequences

Chromogenic Proteins	
amilCP FW	TCACCTGCATATCTCTaatgagtgtgatcgctaaac
amilCP RV	ACACCTGCATATCTCAAAGCttattaggcgaccacagg
spisPink FW	TCACCTGCATATCTCTaatgtcgcactcaaaacaag
spisPink RV	ACACCTGCATATCTCAAAGCttattacacttccagcacac
amilGFP FW	TCACCTGCATATCTCTaatgtcttattcaaagcatgg
amilGFP RV	ACACCTGCATATCTCAAAGCttattatttaaccttcaaaggg
mRFP1 FW	TCACCTGCATATCTCTaatggcttcctccgaagac
mRFP1 RV	ACACCTGCATATCTCAAAGCttaagcaccggtggagtg
tsPurple FW1	TCACCTGCATATCTCTAatggcgagcttggttaa
tsPurple RV1	ACACCTGCATATcagAtggctggtcagatgattg
tsPurple FW 2	TCACCTGCATATTCtgcgcacgacgtatcgt
tsPurple RV2	ACACCTGCATATCTCAAAGCttattacgtcgctttttcc
efforRed FW	TCACCTGCATATCTCTaatgtcagtgattaagcagg
efforRed RV	ACACCTGCATATCTCAAAGCttatgggagagccttcggcag
scOrange FW1	TCACCTGCATATCTCTaatgagcaaaatcagcgac
scOrange RV1	ACACCTGCATATgAtggcaggtcagatggtt
scOrange FW2	ACACCTGCATATcaTctgcgtaccacgtatcg
scOrange RV2	ACACCTGCATATCTCAAAGCttattaatggtgacccagt
amajLime FW	TCACCTGCATATCTCTaatggcactgagcaacaa
amajLime RV	ACACCTGCATATCTCAAAGCttattagaacggcacgac
asCP FW	TCACCTGCATATCTCTaatggcgagcttcctgaa
asCP RV	ACACCTGCATATCTCAAAGCttattaattatgacccagcttg
aeBlue FW1	TCACCTGCATATCTCTaatggcttcactggttaaa
aeBlue RV1	ACACCTGCATATagAtgggacgtcagatgattg
aeBlue FW2	ACACCTGCATATaTctgcgcactacctatcg
aeBlue RV2	ACACCTGCATATCTCAAAGCttattagtgatgccctaatt
sfGFP FW	TCACCTGCATATCTCTaatgcgtaaaggcgaaga
sfGFP RV	ACACCTGCATATCTCAAGCGtcattacttatacagctcg
Promoters + RBS	
J23110 FW	TCACCTGCATATCTCTGGAGTTTACGGCTAGCTCAGTC
J23110+ B0034 RV	ACACCTGCATATCTCACATTTAGTATTTCTCCTCTTTCACTAG
J23106 FW	TCACCTGCATATCTCTGGAGTTTACGGCTAGCTCAGTC
J23106 + B0034 RBS RV	ACACCTGCATATCTCAcatTTAGTATTTCTCCTCTTTCTC
J23103 FW	TCACCTGCATATCTCTGGAGctgatagctagctcagtc
J23103 RV	ACACCTGCATATCTCAcaTTTAGTATTTCTCCTCTTTctctagtagctagcataatc

J23103+B0034 RV	TTTCTCCTCTTTCTCTAGTAgctagcataatccctag
Terminators	
rrnBT1 FW	TCACCTGCATATCTCTGCTTCAAATAAAACGAAAGGCTC
T7Te RV	ACACCTGCATATCTCAAGCGCGCAGAAAGGCCCACCC
T7Te FW	TCACCTGCATATCTCTGCTTGGCTCACCTTCGGGTGG
Violacein genes	
VioA fw1	TCACCTGCATATCTCTAatgaaacattcttccgatatc
VioA rv1	ACACCTGCATATgcGcctgccgccttaacctt
VioA fw2	TCACCTGCATATgcgcgcgttttagcctggg
VioA rv2	ACACCTGCATATccGgccagttgcagcaggta
VioA fw3	TCACCTGCATATccggtgacgacggctggaa
VioA rv3	ACACCTGCATATCTCAAGCGtcacgcggcgatacgct
VioB fw1	TCACCTGCATATCTCTGGAGttaaggaggtaaaaaaaatgagc
VioB rv1	ACACCTGCATATcagAtgacggtggaactccgt
VioB fw2	TCACCTGCATATTctgcgctccctgggtcc
VioB rv2	ACACCTGCATATCTCAAGCGttaggcctcgcggctca
VioD fw1	TCACCTGCATATCTCTGGAGttaaggaggtaaaaaaaatgaag
VioD rv1	ACACCTGCATATtcAacctgcgggaccagagc
VioD fw2	TCACCTGCATATttgactacggccgcaataa
VioD rv2	ACACCTGCATATCTCAAGCGtcagcgctgcaaagcat
VioE fw0	ttaaggaggtaaaaaaaatggagaaccgtgagcc
VioE fw1	TCACCTGCATATCTCTGGAGttaaggaggtaaaaaaaatgg
VioE rv1	ACACCTGCATATCTCAAAGCttagcgcttggccgcga
Carotenoid genes	
crtB fw	TCACCTGCATATCTCTaatgaataatccgtcgttac
crtB rv	ACACCTGCATATCTCAAAGCttattagagcgggcgct
crtY fw	TCACCTGCATATCTCTaatgcaaccgcattatga
crtYrv	ACACCTGCATATCTCAAAGCttaacgatgagtcgtcataat
crtI fw	TCACCTGCATATCTCTaatgaaaccaactacggt
crtI rv	ACACCTGCATATCTCAAAGCttatatcagatcctccagc
crtE fw	TCACCTGCATATCTCTaatggattacgcgaacat
crtE rv	ACACCTGCATATCTCAAAGCtcacagagggatatcgg
crtZ fw	TCACCTGCATATCTCTaatgttgtggatttggaatg
crtZ rv	ACACCTGCATATCTCAAAGCttattacttcccggatgc
Universal Acceptor Vector	
mUAV FW	tttaattaagaattcGGTCTCTATATGCAGGTGTTTACGGCTAGCT CAGT
mUAV RV	atactagtctgcagGGTCTCAATATGCAGGTGCGCAGAAAGGCC CACCC

Level 1 and 2 Acceptor Vectors	
A lvl1 BioBrick FW	tttaattaagaattcCACCTGCATATGGAGAGAGACCTTTACGGCTAGCTCAGT
A lvl1 BioBrick RV	atactagtctgcagCACCTGCATATTCTGAGCGTGAGACCCGCAGAAAGGCCCACCC
B lvl1 BioBrick FW	tttaattaagaattcCACCTGCATATCAGAGGAGAGAGACCTTTACGGCTAGCTCAGT
B lvl1 BioBrick RV	atactagtctgcagCACCTGCATATTGACAGCGTGAGACCCGCAGAAAGGCccaccc
Γ lvl1 BioBrick FW	tttaattaagaattcCACCTGCATATGTCAGGAGAGAGACCTTTACGGCTAGCTCAGT
Γ lvl1 BioBrick RV	atactagtctgcagCACCTGCATATCAAGAGCGTGAGACCCGCAGAAAGGCccaccc
Δ lvl1 BioBrick FW	tttaattaagaattcCACCTGCATATCTTGGGAGAGAGACCTTTACGGCTAGCTCAGT
Δ lvl1 BioBrick RV	atactagtctgcagCACCTGCATATAGCGTGAGACCCGCAGAAAGGCccaccc
A lvl2 BioBrick FW	tttaattaagaattcGGTCTCAGGAGATATGCAGGTGTTTACGGCTAGCTCAGT
B lvl2 BioBrick FW	tttaattaagaattcGGTCTCACAGAGGAGATATGCAGGTGTTTACGGCTAGCTCAGT
Γ lvl2 BioBrick FW	tttaattaagaattcGGTCTCAGTCAGGAGATATGCAGGTGTTTACGGCTAGCTCAGT
Δ lvl2 BioBrick FW	tttaattaagaattcGGTCTCACTTGGGAGATATGCAGGTGTTTACGGCTAGCTCAGT
A-Δ lvl2 BioBrick RV	atactagtctgcagGGGTATATGCAGGTGCGCAGAAAGGCCCACCC
Auxiliary Plasmids	
aux A-Δ plasmid FW	aacccATATGCAGGTGctgcag
aux A-Δ plasmid RV	GAGCGATATGCAGGTGgaattcc
aux A-Δ seq RV	gCACCTGCATATgggtttggccacgcagc
aux A seq FW	cCACCTGCATATCGCTCAGAAGAGACCgtatgaggtggatggtg
aux B seq FW	cCACCTGCATATCGCTGTCAAGAGACCgtatgaggtggatggtg
aux Γ seq FW	cCACCTGCATATCGCTCTTGAGAGACCgtatgaggtggatggtg
aux Δ seq FW	cCACCTGCATATCGCTAGAGACCttctgtatgaggtggatgg
aux 1-3 FW	ACGCTAGAGACCgtatgaggtgg
aux 1 RV	cGGTCTCTAGCGTCTGATATGCAGGTGgaattcc
aux 2 RV	cGGTCTCTAGCGTGACATATGCAGGTGgaattcc
aux 3 RV	cGGTCTCTAGCGCAAGATATGCAGGTGgaattcc

Promoters	
UBQ10 FW	TCACCTGCATATCTCTGGAGgtcgacgagtcagtaat
UBQ10 RV	ACACCTGCATATCTCACATTctgttaatcagaaaaactcagat
TCTP FW	TCACCTGCATATCTCTGGAGCCAACACTCGAATCCCC
TCTP RV	ACACCTGCATATCTCACATTGGTCGCTTATTGATTGTT
OCS FW	TCACCTGCATATCTCTggagctgaaagcgacgttggatg
OCS RV	ACACCTGCATATCTCAcatTttggtagattgcaaatataatgg
G10-90 FW	TCACCTGCATATCTCTGGAGtagtttaaactgaaggcgg
G10-90 RV	ACACCTGCATATCTCACATTTGTCCTCTCCAAATGAAAT
NOS FW	TCACCTGCATATCTCTGGAGAGCGGAGAATTAAGGGAG
NOS RV	ACACCTGCATATCTCAcatTAATTGGATACCGAGGGG
CaMV 35S FW	TCACCTGCATATCTCTGGAGGTCAACATGGTGGAGCAC
CaMV 35S RV	ACACCTGCATATCTCACATTTGTCCTCTCCAAATGAAAT
ACT7 FW	TCACCTGCATATCTCTGGAGgtgtgaagtttgaattatgaaagacg
ACT7 RV	ACACCTGCATATCTCACATTgagaaagatagagaaatggaggag
FAD2 FW	TCACCTGCATATCTCTggagatccgcacgaatctacc
FAD2 RV	ACACCTGCATATCTCAcatTcacggcccgtgggtcga
APT1 FW	TCACCTGCATATCTCTggagggatgagttataggtttggt
APT1 RV	ACACCTGCATATCTCACATTgggaaatagaaggtggtg
ACT2 FW	TCACCTGCATATCTCTggagcacagtcatgaagccatc
ACT2 RV	ACACCTGCATATCTCACATTgattatctcttactttctctcttag
LEC2 FW	TCACCTGCATATCTCTggagcagaatacgcaaaaacgac
LEC2 RV	ACACCTGCATATCTCAcatTatggcacagaagaccacg
TUB9 FW	TCACCTGCATATCTCTGGAGgtaaaattgcctattattggtgtgaa
TUB9 RV	ACACCTGCATATCTCACATTgtttattgtttatgattgaggggtataaa
TUB2 FW	TCACCTGCATATCTCTGGAGcttggatcgttaagtagattttacatac
TUB2 RV	ACACCTGCATATCTCACATTccgtttctttctctactgtgtaag
NDUFA8 FW	TCACCTGCATATCTCTggagtcccgtagcgtccaccg
NDUFA8 RV	ACACCTGCATATCTCAcatTggttttgttgttctctttc
Terminators	
HSP FW	TCACCTGCATATCTCTGCTTATATGAAGATGAAGATGAAATATTTG
HSP RV	ACACCTGCATATCTCAAGCGCCATAGCACATACAGTAGTT
psE9-RbcS FW	TCACCTGCATATCTCTGCTTAGCTCCCTGGCCAATTC
psE9-RbcS RV	ACACCTGCATATCTCAAGCGGTTTGGGATGTTTTACTCC
UBQ5 FW	TCACCTGCATATCTCTGCTTTCTGTTGTAGCGGTAGAT
UBQ5 RV	ACACCTGCATATCTCAAGCGGCAACCAATGAATATCAAAG
ACTIN2 FW	TCACCTGCATATCTCTGCTTGCTCTCAAGATCAAAGGC
ACTIN2 RV	ACACCTGCATATCTCAAGCGACAAAACAATGGGACTAAAAC
APT1 FW	TCACCTGCATATCTCTGCTTCAAGAAACTGGAAGAGAAG

APT1 RV	ACACCTGCATATCTCAAGCGTTGATTCTATAAATAAAACCCATAAC
NDUFA8 FW	TCACCTGCATATCTCTGCTTGAATCACAAGTTCTTGTCAT
NDUFA8 RV	ACACCTGCATATCTCAAGCGAAGCAGATAAAAACAATTGTTC
FAD2 FW	TCACCTGCATATCTCTGCTTGGATGATGGTGAAGAAATTG
FAD2 RV	ACACCTGCATATCTCAAGCGCAAGAGTTTTACTGTCCGA
Gene7 FW	TCACCTGCATATCTCTGCTTGCTAGCTATATCATCAATTTATG
Gene7 RV	ACACCTGCATATCTCAAGCGATCTTGAAAGAAATATAGTTTAAAT
LEC2 FW	TCACCTGCATATCTCTGCTTcgttcatagctaacttgt
LEC2 RV	ACACCTGCATATCTCAAGCGtgtttatcgcgtgtaaggc
MAS FW	TCACCTGCATATCTCTGCTTCTTGGACTCCCATGTTG
MAS RV	ACACCTGCATATCTCAAGCGGATAATTTATTTGAAAATTCATAAG
NOS FW	TCACCTGCATATCTCTGCTTgatcgttcaaacatttggc
NOS RV	ACACCTGCATATCTCAAGCGgatctagtaacatagatgacacc
RbcS2b FW	TCACCTGCATATCTCTGCTTTCCCCTTTCTGGAATATTC
RbcS2b RV	ACACCTGCATATCTCAAGCGGAGCTAAACGAAAAATAAGATTT
TUB9 FW	TCACCTGCATATCTCTGCTTagaaaagatgattgattattacttttg
TUB9 RV	ACACCTGCATATCTCAAGCGttgtataggcttttgacattg
CaMV 35S FW	TCACCTGCATATCTCTGCTTTTTCTCCATAATAATGTGTG
CaMV 35S RV	ACACCTGCATATCTCAAGCGGATCTGGATTTTAGTACTGG
Coding Sequences	
flucCDS FW	TCACCTGCATATCTCTAATGGAAGACGCCAAAAAC
flucCDS RV	ACACCTGCATATCTCAAAGCTACACGGCGATCTTTCC
nlucCDS FW	TCACCTGCATATCTCTAATGGTCTTCACACTCGAA
nlucCDS RV	ACACCTGCATATCTCAAAGCTTACGCCAGAATGCGTTC
sXVECDS FW	TCACCTGCATATCTCTAATGAAGGCTCTTACTGC
sXVECDS RV	ACACCTGCATATCTCAAAGCTCAAACAGTAGCAGGAAATC
mVenusCDS FW	TCACCTGCATATCTCTAATGGTGAGCAAGGGCGA
mVenusCDS RV	ACACCTGCATATCTCAAAGCTTACTTGTACAGCTCGTCC
LhGRCDS FW	TCACCTGCATATCTCTAATGGCTAGTGAAGCTCG
LhGRCDS RV	ACACCTGCATATCTCAAAGCTTACTCTTTTTTTGGGTTTGG
AlcRCDS FW	TCACCTGCATATCTCTaatggcagatacgcgccg
AlcRCDS RV	ACACCTGCATATCTCAAAGCctacaaaaagctgtcaact
BlpR FW1	TCACCTGCATATCTCTAATGAGCCCAGAACGACGC
BlpR RV1	ACACCTGCATATAGATGGGTGTAGAGCGTGGAG
BlpR FW2	TCACCTGCATATATCTGCTGAAGTCCCTGGAG

BlpR RV2	ACACCTGCATATCTCAAAGCTAAATCTCGGTGACGGGCAGGAC
Hyg FW	tcacctgcatatctctaatgaaaaagcctgaactcaccgcgacgtctg
Hyg RV	acacctgcatatctcaaagcctattcctttgccctcggacgagtgctg
eG/C/YFP CDS FW	TCACCTGCATATCTCTaatggtgagcaagggcga
eG/C/YFP CDS RV	ACACCTGCATATCTCAAAGCttacttgtacagctcgtcc
eG/C/YFP CTAG FW	TCACCTGCATATCTCTTTCGgtgagcaagggcgagga
Plasmid Construction	
pGII spec RV2	GTAGTCGGCAAATAATCACTAGACCAATGTTACAC
pGII spec FW 1	ACATTGGTCTAGTGATTATTTGCCGACTACCTTGG
pGII spec FW2	TCACCGCTTCCCTCATAACACCCCTTGTATTACTG
pGII spec RV1	GTAATACAAGGGGTGTTATGAGGGAAGCGGTGATC
pGII AlvI1 FW1	tcgaattcCACCTGCATATGGAGAGAGACCTTTACG
pGII AlvI1 RV	ATATGCAGGTGgaattcGACCGGCATGCAAGCTG
pGII AlvI1 FW	ATATGCAGGTGctgcagtgacaggatatattggcgg
pGII AlvI1 RV 2	tcactgcagCACCTGCATATTCTGAGCGTGAGACCCGCAGAAAGGCCCACCC
pGII BlvI1 FW	TCgaattcCACCTGCATATCAGAGGAGAGAGACCTTTACG
pGII BlvI1 RV	tcactgcagCACCTGCATATTGACAGCGTGAGACCCGCAGAAAGGCccaccc
pGII ClvI1 FW	CGGTCgaattcCACCTGCATATGTCAGGAGAGAGACCTTTACGGCTAGCTCAGT
pGII ClvI1 RV	tcactgcagCACCTGCATATCAAGAGCGTGAGACCCGCAGAAAGGCccaccc
pGII DlvI1 FW	CGGTCgaattcCACCTGCATATCTTGGGAGAGAGACCTTTACGGCTAGCTCAGT
pGII DlvI1 RV	tcactgcagCACCTGCATATAGCGTGAGACCCGCAGAAAGGCccaccc
pGII AlvI2 FW1	CCGGTCgaattcGGTCTCAGGAGATATGCAGGTGTTTACGGCTAGCTCAGT
pGII AlvI2 RV	TGAGACCgaattcGACCGGCATGCAAGCTG
pGII AlvI2 FW	TGCGCACCTGCATATACCCctgcagTAAGTCGCTGTATGTGTTTG

pGII Alvl2 RV 2	CGACTTActgcagGGGTATATGCAGGTGCGCAGAAAGGCCCACCC
PLX Level 1A FW1	aacaatggagaaaaagagaattCACCTGCATATGG
PLX Level 1A RV2	TGCAGGTGaattctctttttctccattgtttacac
PLX Level 1A FW2	GCTCAGAATATGCAGGTGcctgcaggaattgttga
PLX Level 1A RV1	tcaacaattcctgcaggCACCTGCATATTCTGAGC
PLX Level 2A FW1	aaacaatggagaaaaagagaattcGGTCTCAGGAG
PLX Level 2A RV2	TGAGACCgaattctctttttctccattgtttacac
PLX Level 2A FW2	ACCCctgcagttgtctaaatttctgtatttgtttg
PLX Level 2A RV1	atacagaaatttagacaactgcagGGGTATATGCA
pMAP plasmid FW	GGGTCTCACTGAccatcatcagttcggtgg
pMAP plasmid RV	GGGTCTCTaaccgcataaccgccaa
pMPA pUC FW1	GGGTCTCAggttcccgtagaaaagatcaaa
pMAP pUC RV1	GGGTCTCAcggcagctcactcaaaggcgg
pMAP Level 1 cassette FW	ttGGaattcCACCTGCATAT
pMAP Level 1 plasmid RV	ATATGCAGGTGgaattCCaattctctttttctccattgt
pMAP Level 1 plasmid FW	ATATGCAGGTGctgcaGgaattgttgattttgtgatgac
pMAP Level 1 cassette RV	gtcatcacaaaatcaacaattcCtgcagCACCTGCATAT
Synthesised gene fragments	
AlcA-FMV34S	TCACCTGCATATCTCTGGAGgacccatgcggagccgcattggaggccatgcggagccgcacgcgttaacaagagcggctccgcttgacctctcgggatagttccgacctaggattggatgcatgcggaaccgcacgagggcggggcggaaattgacacaccactcctctccacgcaccgttcaagaggtacgcgtatagagccgtatagagcagagacggagcactttctggtactgtccgcacgggatgtccgcacggagagccacaaacgaggcggggccccgtacgtgctctcctaccccaggatcgcatccccgcatagctgaacatctatataagaaggcattcattcccatttgaaggatcatcagatactgaaccaatattatttttacaacaattaccaacaacaacaaacaacaaacaacattacaattactatttacaattacaATGTGAGATATGCAGGTGT
pWKS1 p1	GACCCGCCGCGCCGGTCGCTCGATACCTGTGGTCTCGCTCCCTCTGCGGCTACCGTCAGCGCCTCGCCTGCATCGCCGCCCTTTCCGATCCTCATCCCGCCCCAGCCTTATGGGGGGTCTCAGCCGATCCGCCGCCAGTTCAGGCCGTCCGGCCCCGGATTCTGACCATAATTTCATCGAAAAAAGGGGCGCAGCCCTTCTTGTTCTAATAGTTCTATAAGTTCAGGCGAAAATCGTGCAGCAATTACAAAAGGTTGCGCGTCTATAAGTGGGGAATCCAGCCGCAAAAGTGGAGAGACC

pWKS1 p2	GGTCTCAGTGGGGAATCCAGCCGCAAAAGTGGGGAATCC AGCCGAAATCGCGGATTGACGAGTGGGATTTCCCGCCCAA TAAATCCACCATGGGAAAGACACTCGACGTTGCCCGCGAC CGGGCCTTTGACCAGACCGCGACCGTGCTGCCCGCCGAA ATGGCGCGGGGGGTCTATATGCGCAACGCCCCCAGCCTC GCGGCCCTGAAGCTGATGCATCTGATGATCGCCACGGCG GGCGGGCGCATGGCCGATGACGTGCGCCACGAAATGCG GCTGGCCGACATCCGCAAGATCGACGGTATGGATAACCAC ACCCGGGCCAGCCTGACCCCGCTCTTTGCGGAACTGGCG GCGGCGGTGCTGACCCACGACGACCCGGAAAAGCGGGTC GTGACCATCGGTGGCCTGCTGGACGAAGCCCGGATCGAT TACCGGCACGAGGTCAGCGGCGATCTTCTGGTGTCGTGG ACCTTCCGCAGCATGTTCCGCCGCATGGCGGCGGAATCG AACCATTGGGCGATTCTCGACCGGCAGACCGTGTTCCACC TCGGCAGCAAGTATTCCGTGCTGTTGTTCCAGCACATCGC CAGCTTCAAGGAATACGACCACATTACCGGCAAGACCTTT ACCGTGCCGGAGTTGCGGGCTGTGTTTGGTATCCCCGAG GGCAAAATCAAGCGTTTCGCAGACCTCAACAGAGACGTGC TGACGCCCGCCATTGCCGAAATCAACCAGCTTTCCCGCCT GACTCTGACCGCCACGCCGAACAAGATCGGGCGCACCGT GGCCAGCGTGACGATTGCTTGGGAAGAAAAGCCCCTCGA AGGCAAGCGCTCGACCAAGGCCGAACTGGACCGCCCGAA GGTGGGCCGGAAGGCCCGGCGCGACGGCACCGCCGAGA CGGTGGCACGGGCCTTCCCGGCATCGGGCGGGATCGAGT TCGACCAGCATTGGCGCGACCTCAAGCGGGCGGCGGGCT GCAACATGGACAACACCATGATCGCCGACAAATTCCGGGC ATGGTGCGCCGGGAAGGGCCTCGCTCTCGATGCCCGGAA CATCGAACAGGCGTTCAGCAGCTTCTGCACCAAGGTGGG CCGGGTCTGAAGAGACC

Appendix 3 Mobius Assembly Toolkit for Plants

A/A	**Name**	**Resistance**
	Vector toolkit	
1	mUAV	chl
2	pMAP L1A	kan
3	pMAP L1B	kan
4	pMAP L1Γ	kan
5	pMAP L1Δ	kan
6	pMAP L2A	spec
7	pMAP L2B	spec
8	pMAP L2Γ	spec
9	pMAP L2Δ	spec
10	Auxiliary 1	kan
11	Auxiliary 2	kan
12	Auxiliary 3	kan
13	Auxiliary A	kan
14	Auxiliary B	kan
15	Auxiliary Γ	kan
16	Auxiliary Δ	kan
17	pLX(BBR1) L1A	kan
18	pLX(BBR1) L2A	spec
19	pLX(RK2) L1A	kan
20	pLX(RK2) L2A	spec
21	lvl0 MethylAble terminator	Chlo
	Promoters	
22	lvl0 UBQ10	Chlo
23	lvl0 MAS	Chlo
24	lvl0 UBQ11	Chlo
25	lvl0 UBQ4	Chlo
26	lvl0 OCS	Chlo
27	lvl0 TCTP	Chl
28	lvl0 35S	Chlo
29	lvl0 G10-90	Chlo
30	lvl0 ACT7	Chlo
31	lvl0 APT1	Chlo
32	lvl0 TUB9	Chlo
33	lvl0 ACT2	Chlo

34	lvl0 NOS	Chlo
	Terminators	
35	Lvl0 LEC2	Chlo
36	lvl0 FAD2	Chlo
37	lvl0 HSP	Chlo
38	lvl0 NDUFA8	Chlo
39	lvl0 G7	Chlo
40	lvl0 MAS	Chlo
41	lvl0 35S	Chlo
42	lvl0 UBQ5	Chlo
43	lvl0 ACT2	Chlo
44	lvl0 TUB9	Chlo
45	lvl0 APT1	Chlo
46	lvl0 RbcS2b	Chlo
47	lvl0 NOS	Chlo
	Antibiotic Resistance	
48	lvl0 NptII	Chlo
49	lvl0 BlpR	Chlo
50	lvl0 HygR	Chlo
	Fluorescent proteins	
51	lvl0 turboRFP CDS	Chlo
52	lvl0 turboRFP CTAG	Chlo
53	lvl0 mKate2 CDS	Chlo
54	lvl0 eGFP CDS	Chlo
55	lvl0 eYFP CDS	Chlo
56	lvl0 eCFP CDS	Chlo
57	lvl0 sfGFP CDS	Chlo
58	lvl0 sfGFP CTAG	Chlo
59	lvl0 mTFP1-FLAG CDS	Chlo
60	lvl0 mTFP1-FLAG CTAG	Chlo
61	lvl0 TagRFP CTAG	Chlo
62	lvl0 YPet CTAG	Chlo
63	lvl0 mVenus CDS	Chlo
	Luciferases	
64	lvl0 nluc CDS	Chlo
65	lvl0 fluc CDS	Chlo
	Inducible Systems	
66	lvl0 sXVE	Chlo

67	lvl0 lexA-35S	Chlo
68	lvl0 LhGR	Chlo
69	lvl0 pOp6-35S	Chlo
70	lvl0 AlcR	Chlo
71	lvl0 alcSynth	Chlo

Contributions

Chapter 1: Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly

Andreas Andreou conceived, designed, performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

Chapter 2: Protoplast isolation and transformation protocol development

Protoplast isolation and transformation: Andreas Andreou designed and performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

High throughput protoplast transformation: Andreas Andreou designed the experiments and analysed the results. Andreas Andreou, Marisol Ochoa-Villarreal and Jessica Nirkko performed the experiments. Naomi Nakayama supervised the process.

Chapter 3: A characterized toolbox for multigene assembly and delivery in plant systems

Mobius Assembly for Plant Systems (MAPS): Andreas Andreou designed and performed the experiments, and analysed the results, under the supervision of Naomi Nakayama.

MethylAble feature: Andreas Andreou conceived, designed, and performed the experiments, and analysed the results.

New plant binary vector (pMAP): Andreas Andreou designed and performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

Structural plasmid stability: Andreas Andreou designed and performed the experiments and analysed the results.

Effect of Agrobacterium strains and binary vectors in the MM1 cell line transformation: Andreas Andreou designed and performed the experiments and analysed the results.

Selection of new promoters and terminators. Andreas Andreou designed and cloned the terminators and the 300bp promoters and co-designed the 500 bp promoters with Marisol Ochoa-Villarreal. She cloned the 500 bp promoters and

performed the luciferase assays. Andreas Andreou analysed the results. Naomi Nakayama supervised.

Characterization of promoters and terminators: Andreas Andreou and Jessica Nirkko designed and performed the experiments. Andreas Andreou analysed the present version of the results in this thesis with the deconvolution of luminescence cross-talk, Jessica Nirkko analysed the previous version of the data. Eric Thorand wrote the MATLAB scripts for file format conversion. Eric Thorand, Iordanis Chatzinikolaidis and Jiayi Wang helped in the understanding of the Deconvolution software. Naomi Nakayama supervised.

Terminator functional dissections: Andreas Andreou and Marisol Ochoa-Villarreal designed the experiments. Andreas Andreou performed the MM1 and Jessica Nirkko the protoplast transformations and luciferase assays. Andreas Andreou analysed the results. Naomi Nakayama supervised.

Chapter 4: Guided differentiation system for cultured plant cells

Characterization and crosstalk experiments for the plant inducible systems: Andreas Andreou designed the experiments. Andreas Andreou performed the cloning of the constructs and Marisol Ochoa-Villarreal the luciferase assays. Naomi Nakayama supervised.

Selection of the key transcription factors: Andreas Andreou designed, performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

Sequential differentiation system: Naomi Nakayama conceived, Andreas Andreou designed and performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

Transformation of MM1 cell line and transgene selection: Andreas Andreou designed and performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

Sequential differentiation experiments - Anthocyanin analysis: Andreas Andreou designed and performed the experiments and analysed the results, under the supervision of Naomi Nakayama.

.