

IMPERIAL

Synthetic biology for engineering *Arabidopsis thaliana* protoplasts



Elif Gediz Kocaoğlu

Supervisor: Dr. Naomi Nakayama

2024 | Submitted to the Imperial College London for the Degree of Doctor of
Philosophy in Bioengineering at the Centre for Synthetic Biology

Abstract

For over 10,000 years, humans have domesticated plants, but plant synthetic biology has ushered in a new era by facilitating the engineering of development. Plant development shows remarkable plasticity, enabling plants to adapt to the ever-changing environment. This plasticity even extends to the cell fates, for example, in protoplast regeneration. Upon removal of the cell wall, fully differentiated plant cells can revert to a stem-cell-like state, giving rise to protoplasts capable of regenerating cells, tissues, and even whole plants. Engineering this process has great potential for creating transgenic plants for agriculture and new cell culture lines for sustainable bioproduction. However, engineering protoplast regeneration requires tools that offer precise spatiotemporal control over gene expression, ensuring stable expression levels without mutations across generations. Efforts to identify the target genes regulating protoplast regeneration are equally crucial. Yet, protoplasts currently lack both the genetic tools and knowledge of target genes. To address these gaps, I aimed to establish *Arabidopsis thaliana* protoplasts as a new synthetic biology chassis. This involved characterising basic DNA parts which revealed combinatorial interactions in gene expression. With RNA folding analysis, I proposed physical interactions among the parts as one of the mechanisms underlying these interactions. Additionally, I have re-designed vectors with enhanced stability for assembling DNA constructs. These constructs have been used to create stable *A. thaliana* marker lines as stage-specific reporters of protoplast regeneration. Fluorescent proteins were screened for strongest signals after a protoplast transformation protocol was optimised. Furthermore, I also established a platform for regenerating protoplasts into microcallus. Plant growth and cell wall digestion conditions were shown to be crucial in this process. The tools and platforms developed in this study will facilitate precision engineering in plant synthetic biology. Additionally, they will enhance our understanding of molecular pathways underlying protoplast regeneration and further biotechnology driven by protoplasts.

Keywords: Plant synthetic biology, context dependency, developmental engineering, plant regeneration, regeneration engineering, protoplast, protoplast regeneration

Declaration

I hereby certify that everything included in this thesis is my own work, and in the instances where work has been used from other sources, it is acknowledged appropriately. The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

There are so many people I want to thank for so many things, that it is hard to know where to start. Firstly, a huge thank you to my supervisor, **Naomi Nakayama**, for not only guiding me through experiments but also encouraging me to grow as a person. I owe much of my development as a scientist to you. I am grateful to all the PIs in the SynBio centre for their support and guidance, especially for **Tom Ellis**, for your scientific and career advice.

BFF lab, you have been incredible! **Dhanya**, I can't thank you enough for all you helped me through. Whether it was regeneration or mental breakdown, you were my go-to advisor. **Oli**, I will miss having you as my bench and office desk neighbour, who does spontaneous push-ups and has a sick playlist. **Johnsson 2000**, you were the best protohelper, and I look forward to seeing you succeed in anything you decide to take on. **Alice**, can you ship croziflette to me, please? I wouldn't mind if you wanted to deliver it to me personally, every month. **Kyle**, I already miss having you in the lab, but I guess other people should be allowed to enjoy your company as well. **Sam, Sibi, Ritesh, Oscar** thank you all for supporting me! Although we didn't work in the lab at the same time, **Andreas**, you also deserve a special thank you. You have been tolerating my questions since I was a master's student in Edinburgh. Without you, there wouldn't be Mobius!

The morning gang, **Alicia** and **Alice Boo**, I wouldn't have survived without you. Our morning catch-ups were the favourite part of my days. **Davina**, you are my sarcastic rock that I know I can always rely on, I couldn't have reached the Acknowledgements section without you. Looking forward to the next Youtube link you will send from Dodo to brighten my day. **Duncan**, on top of your mental support, it was great having you as a flatmate and my personal running coach. One day I shall be as fast as you. **Klaudia**, I'm coming to visit you in Oxford as soon as I am done with this. Battersea Park is not the same without you! **Albert, Kyler, Andreas, Diego**, honestly, I can't imagine Imperial without you.

Moving on to non-imperial gratitude, **Işıl** thanks to you, my life in London started off great and even though you can never fully remember the title of my thesis, I couldn't have asked for a better quarantine buddy. **İdil**, you've been there for me since I was a baby, supporting me through all my significant milestones, including this PhD. **Burcu** and **Sara**, my forever cheerleaders, I appreciate your support despite the distance between us.

Last but actually, the most important acknowledgement goes to my family. **Anne** and **baba**, you have been the best parents anyone can ask for. I hope one day I can provide my children with the same level of care and support you've given me. **Aras**, I am who I am today because of you; words can never be enough to thank you. **Teyze**, seeing you as a professor inspired me to pursue my career in academia. **Kiyan**, tu es celui qui a vu tout les hauts et les bas de ce doctorat. Rentrer à la maison auprès de toi après une bonne ou une mauvaise journée est tout ce que je puisse demander. Merci beaucoup. **Elmo**, as a dog, you can't read this, but your persistent nudging to play didn't exactly help my thesis writing. Nonetheless, I still adore you.

CONTENTS

List of figures.....	i
List of acronyms	iii
1. Introduction.....	1
1.1. Engineering plant development.....	3
1.1.1. DNA assembly frameworks and toolkits	4
1.1.2. Tools for gene regulation	6
1.1.2.1. Generic promoter and terminator parts.....	7
1.1.2.2. Cell- or tissue-type specific promoters	9
1.1.2.3. Inducible systems.....	10
1.1.2.4. Logic gates.....	12
1.1.2.5. Multiplex regulation.....	12
1.1.3. Case studies.....	13
1.1.3.1. Shoot architecture engineering with hormone-activated Cas9-repressors (HACRs) .	15
1.1.3.2. Controlling growth via stomata opening with synthetic ion channels	15
1.1.3.3. Logic gate-based engineering of root architecture.....	16
1.1.4. Opportunities and challenges with engineering development	19
1.2. Engineering regeneration	21
1.1.1. Types of regeneration.....	21
1.2.2. Protoplast regeneration.....	23
1.2.3. Dedifferentiation.....	24
1.2.4. Cell wall regeneration	25
1.2.5. Cell cycle progression and cell division	25
1.2.6. Callus formation and whole plant regeneration.....	26
1.2.7. Opportunities and challenges of protoplast regeneration	26
1.2.8. Cell wall and dedifferentiation.....	29

1.3.	Aims and objectives	30
1.4.	Publications from this study	32
2.	Promoters, coding sequences, and terminators interact in gene regulation	33
2.1.	Introduction	34
2.1.1.	Transcription in plants	35
2.1.2.	Promoters controlling gene expression	36
2.1.3.	Terminators controlling gene expression	37
2.1.4.	5' and 3' UTRs controlling gene expression	39
2.1.5.	DNA and RNA structure controlling gene expression	41
2.1.6.	Mobius Assembly for Plant Systems (MAPS) shows interactions between genetic parts..	43
2.2.	Results.....	43
2.2.1.	Promoter characterisation reveals that <i>UBQ10</i> and <i>MAS</i> promoters are consistently strong	43
2.2.2.	TSS structure of the <i>UBQ10</i> promoter affects gene expression levels.....	46
2.2.3.	Terminators affect gene expression in inducible systems.....	47
2.2.4.	Terminators affect gene expression of constitutive promoters	50
2.2.5.	Changing the coding sequence changes the strength of promoter-terminator pairs	50
2.2.6.	Terminator functional dissection	52
2.2.7.	Promoter-coding sequence-terminator interactions in gene regulation	53
2.2.8.	Transcriptional and post-transcriptional regulation of reporter expression.....	55
2.2.9.	Local drop in GC content around dominant Poly-A site for consistently strong terminators	57
2.2.10.	<i>a priori</i> identification of novel motifs in consistent strong terminators	57
2.2.11.	Factors governing gene expression strength in outlier terminators	58
2.2.12.	Terminator structure and terminator strength	60
2.2.13.	Promoter-terminator interactions and gene expression consistency	62

2.3.	Discussion.....	64
2.3.1.	TSS structure of the <i>UBQ10</i> promoter affects gene expression.....	64
2.3.2.	Terminators affect gene expression in inducible systems.....	64
2.3.3.	Terminators affect gene expression of constitutive promoters.....	65
2.3.4.	Promoter-coding sequence-terminator interactions in gene regulation.....	65
2.3.5.	Factors governing terminator strength.....	66
2.3.6.	Terminator structure and terminator strength.....	66
2.3.7.	Promoter-terminator interactions and gene expression consistency.....	67
2.4.	Conclusion.....	68
3.	Mobius Assembly improvements: Developing stable vectors.....	69
3.1.	Introduction.....	70
3.1.1.	Transposon insertion.....	71
3.1.2.	Plasmid dimerisation.....	73
3.1.3.	<i>cer</i> domain for dimer resolution.....	74
3.2.	Results.....	77
3.2.1.	Transposon insertion in <i>spisPINK</i> expressing vectors.....	77
3.2.2.	<i>cer</i> domain addition improves growth rate of cells with MAPS vectors.....	79
3.2.3.	Plasmid dimerisation is prevented by the addition of the <i>cer</i> domain.....	81
3.3.	Discussion.....	85
3.3.1.	Transposon insertion in <i>spisPINK</i> expressing vectors.....	85
3.3.2.	Plasmid dimerisation and the addition of the <i>cer</i> domain.....	86
3.4.	Conclusion.....	89
4.	<i>Arabidopsis thaliana</i> marker lines for protoplast regeneration.....	90
4.1.	Introduction.....	91
4.1.1.	WOUND-INDUCED DEDIFFERENTIATION1 (WIND1).....	92

4.1.2.	WUSCHEL RELATED HOMEBOX 5 (WOX5).....	93
4.1.3.	WUSCHEL RELATED HOMEBOX 13 (WOX13).....	96
4.1.4.	INDOLEACETIC ACID-INDUCED PROTEIN 8 (IAA8).....	98
4.2.	Results.....	101
4.2.1.	Optimizing the protoplast transformation protocol for transient gene expression.....	101
4.2.2.	Fluorescent protein (FP) screening.....	103
4.2.3.	Testing <i>NLS150</i> and <i>VirD2</i> as alternative nuclear localisation signals.....	104
4.2.4.	Making constructs for fluorescent marker lines.....	106
4.2.5.	Confirming functionality of constructs with <i>A. thaliana</i> transient expression assay.....	109
4.2.6.	Generating stable lines of the <i>A. thaliana</i> protoplast regeneration markers.....	111
4.3.	Discussion.....	114
4.3.1.	FP screening.....	114
4.3.2.	Building constructs and confirming their functionality.....	115
4.3.3.	Stable <i>A.thaliana</i> marker lines for protoplast regeneration.....	116
4.3.4.	Stage-specific tools for monitoring and engineering regeneration.....	116
4.4.	Conclusion.....	117
5.	Establishing a protoplast regeneration platform	118
5.1.	Introduction	119
5.2.	Results.....	125
5.2.1.	Physiological markers show consistent culture progression for protoplast regeneration 125	
5.2.2.	Asymmetric and budding-like cell divisions are observed in the protoplast culture	128
5.2.3.	Contamination in protoplast culture and its prevention	129
5.2.4.	Cell wall digestion enzyme concentration affects culture progression	131
5.2.5.	Plants grown under fluctuating day/night temperature show symmetric cell division and microcallus formation	132

5.2.6.	Marker lines confirm cell division and microcallus formation.....	135
5.2.7.	Cell wall digestion enzymes and plant growth conditions deregulate marker genes	136
5.2.8.	Plant growth conditions deregulate marker genes on Day 0	140
5.2.9.	Preliminary data shows day/night temperature fluctuations deregulate <i>WOX5</i> expression 144	
5.3.	Discussion.....	145
5.3.1.	Monitoring protoplast culture progression	145
5.3.2.	Enabling protoplasts to divide symmetrically and form microcallus.....	146
5.3.3.	Preliminary data suggests that day/night temperature fluctuations deregulate <i>WOX5</i> expression	149
5.4.	Conclusion.....	150
6.	Discussion & Future perspectives.....	151
6.1.	Summary	152
6.2.	Gene regulation	152
6.2.1.	Improving the ethanol-inducible system	153
6.2.2.	Designing amplifiers for gene expression reporters.....	154
6.2.3.	Context dependency in gene regulation.....	156
6.3.	Construct stability	158
6.4.	Studying protoplast regeneration.....	161
6.4.1.	Waddington's landscape.....	161
6.4.2.	Protoplast regeneration in Waddington's landscape	162
6.4.3.	Reproducibility in protoplast regeneration with Waddington's landscape.....	162
6.5.	Engineering plant development.....	165
6.6.	Conclusion.....	168
7.	Materials & Methods.....	169
7.1.	Molecular biology	170

7.1.1.	Bacterial strains, growth conditions and growth curve measurements.....	170
7.1.2.	Bacterial transformations	170
7.1.3.	Plasmid isolation	171
7.1.4.	<i>A. thaliana</i> genomic DNA extraction.....	171
7.1.5.	PCR amplification	171
7.1.6.	Plasmid construction.....	172
7.1.7.	Mobius Assembly cloning	172
7.1.8.	Construct validation	173
7.1.9.	RNA extraction from protoplasts for part characterisation assays	173
7.1.10.	RNA extraction for cell wall digestion assays.....	173
7.1.11.	Reverse transcription and cDNA synthesis	174
7.1.12.	qRT-PCR for part characterisation assays	174
7.1.13.	qRT-PCR for cell wall digestion assays	174
7.2.	Cellular assays	175
7.2.1.	Protoplast isolation and transformation.....	175
7.2.2.	Luciferase assay	176
7.2.3.	Fluorescence assay.....	176
7.2.4.	Fluorescent protein screening	177
7.2.5.	Intracellular nitric oxide (Nox) measurements	177
7.3.	Whole-plant assays	178
7.3.1.	Agrobacterium floral dipping.....	178
7.3.2.	Microscopy.....	178
7.4.	Regeneration assays	179
7.4.1.	Plant growth conditions.....	179
7.4.2.	Protoplast isolation and culture.....	180
7.4.3.	Cell wall digestion assays	181

7.4.4.	Staining and microscopy	181
7.5.	<i>in silico</i> assays	182
7.5.1.	Functional dissection of FAD2 and HSP terminators	182
7.5.2.	RNA folding predictions	182
7.5.3.	<i>a priori</i> identification of novel terminator motifs.....	183
Appendix.....		184
8.1.	Protoplast transformation protocol for transient expression	187
8.1.1.	Media preparation protocol	187
8.1.2.	Transformation protocol.....	187
8.1.3.	Notes	188
Bibliography.....		190

List of figures

Figure 1. Major types of plant synthetic biology tools for developmental reprogramming.	14
Figure 2. Case studies showcasing the engineering of plant development using synthetic biology.....	18
Figure 3. Different types of regeneration.....	23
Figure 4. Protoplast regeneration.....	27
Figure 5. Simplified illustration of the plant cell wall.	30
Figure 6. Promoter and terminator's role in gene expression.	40
Figure 7. Dual-reporter characterisation of promoters.....	45
Figure 8. TSS structure affects gene expression levels.....	47
Figure 9. Dual-reporter characterisation of inducible promoters.....	49
Figure 10. Dual-reporter characterisation of terminators show that terminators affect gene expression levels..	52
Figure 11. Functional dissection of consistently strong <i>HSP</i> and <i>FAD2</i> terminators..	53
Figure 12. Gene expression analysis for luciferase and fluorescent protein reporters in PEG-transformed leaf mesophyll protoplasts.	56
Figure 13. Factors contributing to terminator strength..	59
Figure 14. Structural features of terminators.....	61
Figure 15. Combinatorial interactions between promoters and terminators at RNA level..	63
Figure 16. Example of transposition mechanisms observed in <i>E. coli</i>	72
Figure 17. Dimer resolution by the <i>cer</i> system in <i>E. coli</i>	75
Figure 18. Improving pMAPS vectors by changing the ori and the negative selection marker..	81
Figure 19. Preventing dimerisation of pMAPS vectors.....	82
Figure 20. Stability of pMAPS plasmid with <i>cer</i> domain over time.	85
Figure 21. Molecular mechanisms of <i>WIND1</i> regulation of regeneration..	94
Figure 22. Molecular mechanisms of <i>WOX5</i> action during plant regeneration.....	95
Figure 23. Molecular mechanisms of <i>WOX13</i> function in regeneration..	97
Figure 24. Molecular mechanisms of <i>IAA8</i> in plant development..	99
Figure 25. Deregulated gene expression patterns of the chosen TFs during protoplast regeneration. ...	101
Figure 26. Protoplast transformation with the optimised protocol.....	103
Figure 27. Fluorescent protein screening.....	106

Figure 28. Testing new nuclear localisation signals with transient expression in <i>A. thaliana</i> mesophyll protoplasts.	106
Figure 29. Construct designs for tri-colour fluorescent protoplast regeneration marker lines.....	108
Figure 30. Confirming functionality of the regeneration stage markers using protoplast transient expression assay.	110
Figure 31. Confirming the marker expression of IAA8:WIND1:WOX5:HYG construct using protoplast transient expression assay.....	111
Figure 32.. Reporter expression in the stable fluorescent regeneration stage marker lines.....	113
Figure 33. Issues observed during <i>A. thaliana</i> marker line screening.	114
Figure 34. Physiological and molecular markers of protoplast regeneration.	124
Figure 35. Physiological markers showing consistent of protoplast regeneration between biological replicates.....	126
Figure 36. Examples of different cell shapes observed on Day 6 are shown with Calcofluor White staining.....	127
Figure 37. Viability of the protoplast culture with FDA/PI staining.....	128
Figure 38. Timelapse images showing different cell division/growth patterns observed during protoplast culture.....	129
Figure 39. Efforts to reduce contamination in the protoplast culture..	130
Figure 40. Protoplast culture with 1X enzymes.....	133
Figure 41. Protoplast culture progression for plants grown in Condition 2.....	135
Figure 42. Marker lines showing cell division.	137
Figure 43. Gene expression analysis of isolated protoplasts (Day 0, dedifferentiation) from plants grown in Condition 1 and 2.....	141
Figure 44. Gene expression analysis of isolated protoplasts (Day 0) from plants grown in Conditions 2, 3 and 4..	144
Figure 45. Preliminary data showing the effect of day and night temperature fluctuations on transcription factors on Day 0 and 5.....	145
Figure 46. Incorporating suicide exon regulation into the ethanol-inducible system..	155
Figure 47. Suggested gene amplifier designs with feedforward loops..	156
Figure 48. Perspectives on context dependency in gene regulation.....	160
Figure 49. Waddington's epigenetic landscape to understand protoplast regeneration.	165
Figure 50. Future of engineering plant development..	167

List of acronyms

3'UTR	3' Untranslated region
5'UTR	5' Untranslated region
bp	Base pair
CDS	Coding sequence
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
FP	Fluorescent protein
GFP	Green fluorescent protein
gRNA	Guide RNA
kb	Kilobase pair
MES	2-(N-morpholino)ethanesulfonic acid
MGG	Maceration Glycine Glucose media
MMM	Mannitol Magnesium MES media
mRNA	Messenger RNA
mScarlet-I	Monomeric scarlet fluorescent protein
NOX	Nitric oxide
OD	Optical density
Ori	Origin of replication
PEG	Polyethylene glycol
PIM	Protoplast induction media
RFP	Red fluorescent protein
RNA	Ribonucleic acid
s.d.	Standard deviation
sfGFP	Superfolder green fluorescent protein
TSS	Transcription start site

CHAPTER 1

Introduction

Declaration of published work

The content on pages 3 – 21 of this chapter has been previously published in Journal of Experimental Botany and reproduced here under the terms of Creative Commons Attribution (CC BY) license. The original publication is as follows:

Kocaoglan, E.G., Radhakrishnan, D., Nakayama, N. Synthetic developmental biology: Molecular tools to re-design plant shoots and roots. Journal of Experimental Botany (2023).

Any additional work, figures, or tables presented in this chapter that were not part of the original publication are my own contributions and are clearly indicated.

1.1. Engineering plant development

Since the dawn of agriculture, we humans have sought to alter plant shape and architecture. Agricultural yield potential is tightly linked to plant growth and development. Harvest directly increases if we can grow larger crops¹. More fruits or grains would grow on a plant with more branches^{2,3}. Extensive root growth is likely to help with better soil anchorage, as well as nutrient and water absorption^{4,5}. A major breakthrough in the original Green Revolution was the development of semi-dwarf varieties of cereal crops, which increased the yield by reducing the damage due to structural failure⁶. When shoots fall to the ground, excessive moisture can impair fruit development, while harvesting becomes challenging⁷. As the world population continues to rise rapidly, and with the imminent threats of climate change, we are in dire need of new technologies to achieve sustainable agricultural improvement. Reprogramming plant growth and development is a key to the next Green Revolution, just like the first time.

The development of plant tissues or organs is wired through a complex molecular network involving parallel and converging pathways^{8–10}, interplays with hormonal regulations¹¹, and feedback/feedforward loops^{12,13}. Developmental processes progress with high spatial and temporal specificity, with different stages expressing distinct sets of genes^{14–17}. Furthermore, plant developmental programmes are environmentally plastic, and thus the underlying molecular mechanisms have different versions¹⁸. Genetic interventions towards desirable traits have occurred throughout the history of crop domestication via breeding efforts and, in the past three decades, also with genetic engineering. However, precision engineering of pre-defined morphology and architecture remains a challenge, presumably because of the complexity of the pathways requiring tight spatiotemporal control and a high degree of robustness.

Synthetic biology refers to the rational engineering of gene regulatory networks to generate predictable outcomes in a biological host (chassis). It is also called “engineering biology” for its application of engineering principles, such as formalized design and design-led, iterative cycles of trials to bring the outcome closer to the purpose. Over the past two decades of synthetic biology, much effort has been made to create engineering resources for the precision genetic engineering of molecular pathways complicated enough to act like a circuit¹⁹. In other words, synthetic biology opens new doors

to a whole new level of genetic modification of biological development. Synthetic developmental biology is a rapidly moving frontier of both synthetic biology and developmental biology in animal systems^{20,21}.

The time is also opportune for synthetic biological engineering of plant development. In this review, we first provide an overview of synthetic biology tools that are likely to be useful for developmental reprogramming of plants. Three case studies follow to illustrate the power of synthetic biology tools and strategies in rational modification of plant morphology. We conclude by highlighting future challenges and perspectives.

1.1.1. DNA assembly frameworks and toolkits

In engineering, formalized, modular and well-characterised building parts are used to construct a device. Similarly, in biology, how individual DNA parts function needs to be known to rationally design or re-design biological systems. With this goal, Synthetic Biology promoted formalization and characterisation of DNA parts adhering to the BioBricks standard of DNA part construction²². BioBricks are basic DNA parts with pre-defined flanking prefix and suffix sequences. These prefix and suffix sequences encode restriction enzyme recognition sequences, allowing users to create compatible sticky ends that self-assemble in specific combinations. For an extensive review on molecular cloning, readers are referred to²³. Multiple BioBricks can be combined in a fast and semi-scarless manner with Golden Gate technology²⁴. Golden Gate frameworks use Type IIS restriction enzymes which cut DNA downstream of their recognition sequence, allowing users to design their 4-bp overhangs and assign order to how the DNA parts will be assembled. Unique 4-bp overhang sequences were given to different types of DNA parts with Phytobrick standard, which facilitated community-wide part exchange²⁵. Although the plant community initiated Phytobrick standard, the overhang sequences comply with the commonly used selections for other eukaryotic organisms²⁶. Phytobrick has been adopted widely in non-plant communities (e.g., iGEM) as a universal overhang format (https://2016.igem.org/Resources/Plant_Synthetic_Biology/PhytoBricks).

Over the years, multiple variants of Golden Gate-based cloning toolkits have been developed for plant systems, which are summarized in Table 1. One of the earliest toolkits is MoClo Assembly, and it implemented a linear strategy to Golden Gate, where there is a hierarchy of Levels 0, 1 and 2²⁷. In MoClo, basic DNA parts (Phytobricks) form Level 0. When basic DNA parts such as a promoter,

coding sequence and terminator are combined to form a functional transcription unit (TU) at Level 1. Multiple TUs can be combined at Level 2 to form a multigene construct (Figure 1.A). MoClo has intermediate modules that provide end-linkers for expanding cloning capacity beyond Level 2. DNA assembly toolkits for plant gene expression control, like Plant MoClo toolkit, uses *Escherichia coli* as a propagation host. *E. coli* grows faster, and plasmid purification yields are higher compared to other bacterial species such as *Agrobacterium tumefaciens*, which is the most commonly used bacterium for gene transfer into plant cells. *A. tumefaciens* can insert the DNA fragment located between the Left Border (LB) and Right Border (RB) T-DNA insertion sequences, into the host plant genomes. As expression in plant tissue is the goal, vectors in plant DNA assembly toolkits accept TUs cloned between LB and RB sequences. The natural hosts of *A. tumefaciens* are dicots, and not monocots. Thus, plasmids having elements based on *A. tumefaciens* transformation works well for dicots, but they are challenging to use in monocots²⁸. Recently, Joint Modular Cloning (JMC) Toolkit has been developed which follows the same linear strategy as MoClo, but with higher capacity²⁹. JMC toolkit uses pMIN vectors which can be used to transform both *Nicotiana benthamiana* (dicot) and *Setaria viridis* (monocot).

Different from the linear cloning strategy of MoClo and JMC; GoldenBraid and Loop Assembly operates in a cyclical fashion, switching back and forth between two levels^{30,31}. Both GoldenBraid and Loop Assembly have 2 type of vectors: one for even levels (Level 0, 2, 4,...) and one for odd levels (Level 1, 3, 5, ...). Because they can switch back and forth between the two levels, they have unlimited cloning capacity (Figure 1.A). Unlimited cloning capacity is a common feature of the cloning systems that adopt cyclical-strategy. However cloning capacity is often limited by the metabolic burden imposed by plasmid size and stability³². Plant multi-gene constructs tend to be large (>20 kb) and together with a plasmid having large backbone, such as pCambia (~6.5 kb) used by GoldenBraid and Loop, cloning capacity could be reduced. Alternatively, these toolkits use pGreen-type vectors (~5 kb) which are smaller, but they require an additional plasmid pSoup (~9.2 kb) to replicate in *A. tumefaciens*^{30,31}. An advantage of using Loop Assembly is that vectors and parts in the toolkit are under OpenMTA, which enables unrestricted use.

Mobius Assembly is another framework for plant systems that combine linear and cyclical strategies. Mobius Assembly has only one vector (mUAV) to store basic parts, and it follows linear strategy until Level 2. For higher level constructs, users can switch back and forth between Level 2 and Level 1

vectors (cyclical cloning)³³. Mobius Assembly additionally uses a unique Type 2SII restriction enzyme, *AarI*, which has 7-bp recognition sequence. The use of *AarI* reduces domestication (removal of internal restriction enzyme sites) by 58.3% compared to systems using three 6-bp cutters such as GoldenBraid and Loop Assembly³⁴. Mobius Assembly also offers MethylAble feature where Phytobricks are flanked by two *BsaI* recognition sites. The divergent *BsaI* sites are prone to methylation and protected from digestion while the convergent sites are not. This allows Level 0 parts to be combined with pre-made Level 1 constructs or to be fused with a Level 2 TU, simplifying combinatorial DNA library construction.

Table 1. Golden Gate-based cloning toolkits for plant systems.

Toolkit	Vector	Strategy	Capacity	Part types	Highlights	Reference
MoClo Assembly	pGreen+pSoup	Linear	7 transcriptional units (TUs), can be increased by auxiliary modules	Natural and synthetic promoters, integrase-based toolbox, recombinase-based logic gates	Widely used but has 53 acceptor vectors to choose from, which makes cloning complicated	Weber et al. (2011)
Joint Modular Cloning (JMC)	pMIN	Linear	7 TUs, can be increased by expansion modules	Monocot and dicot promoters, terminators, enhancers	Shown to work in both monocots and dicots, uses an isoschizomer of <i>AarI</i> (<i>PagCI</i>) and uses expansion linkers to simplify combinatorial DNA assembly	Chamness et al. (2023)
GoldenBraid	pGreenII+pSoup or pCambia	Cyclical	Unlimited	Synthetic promoters	Initial assembly was pairwise which increases the number of cloning steps to achieve the final construct, fixed in GB2.0 and GB3.0	Sarrion-Perdigones et al. (2011)
Loop Assembly	pGreenII+pSoup pCambia	Cyclical	Unlimited	<i>Marchantia polymorpha</i> parts	Open MTA licence for unrestricted use but pCambia vectors have large backbone (~6.5 kb) which might cause stability issues	Pollak et al. (2020)
Mobius Assembly for Plant Systems (MAPS)	pMAPS	Cyclical	Unlimited	Small promoters, terminators, reporter genes	MethylAble for combinatorial DNA assembly, use of <i>AarI</i> as Type IIS restriction enzyme which reduces domestication	Andreou and Nakayama (2018) ; Andreou et al. (2021, Preprint)

Development of Golden Gate-based cloning toolkit for plants enabled simple and fast multi-gene assembly. Each toolkit possesses distinctive features, and the decision to use a specific toolkit depends on the planned molecular construction. JMC would be preferred when working with both monocots and dicots; and Mobius may be most efficient in building a series of constructs that differ in one part, especially for part characterisation thanks to the MethylAble feature.

1.1.2. Tools for gene regulation

Characterisation of DNA parts typically involves quantitation of the output gene expression, when DNA parts are combined with each other. Well-characterised DNA parts enable users to choose gene

expression levels by changing the promoters, terminators, and other regulatory sequences (e.g., enhancers), in a predictable manner³⁵. Thorough characterisation of such basic parts is crucial because expression strength of gene of interest varies depending on nature of the protein, application and chassis types^{29,36}. Desired range for the selected organism can be achieved by changing the basic and regulatory part combinations. Having well-characterised parts allows for more predictive construct design.

1.1.2.1. Generic promoter and terminator parts

DNA assembly toolkits included in Table 1 (MoClo, JMC, GoldenBraid, Loop Assembly and Mobius Assembly) adhere to the Phytobrick design standard; thus, they are compatible with each other. Collectively, these toolkits offer 41 constitutive promoters and 41 terminators that have been characterised in *A. thaliana* and *N. benthamiana*, which are both dicot plants^{29–31,33,34,37}. JMC contains 6 monocot promoters characterised in *S. viridis*. Regulatory parts can behave distinctly in monocots compared to dicots; for example, the most widely used sati in dicots, 35S and NOS showed weak expression in *S. viridis*. It has been known that 35S promoter has lower activity in monocot species compared to dicots, and JMC toolkit further highlights that expression strength of DNA parts can be species-dependent, and part characterisation should be conducted in each species of interest without the assumption of orthogonality^{29,36}. Orthogonality is the ability of DNA parts to operate in a defined way, independent of the contexts and backgrounds; this is one of the most sought-after traits in synthetic biology. Furthermore, it is not only promoters that control gene expression; terminators are receiving increasingly more attention for characterisation. The part characterisation for Mobius Assembly for Plant Systems (MAPS) showed that there is a synergistic interaction between promoters and terminators. Promoter strength can change 5-fold amongst 14 different terminators tested³⁴.

In addition to directly utilizing the promoter or terminator sequences of genes existing in plants or plant pathogens, one can rationally design synthetic parts. Rationally designing synthetic parts can facilitate orthogonality, allowing to minimize the presence of unknown functional sequences and the size of the plasmid. A synthetic promoter typically has a core promoter region with TATA box for transcription initiation and has *cis*-acting elements (the binding sites of transcriptional regulators, i.e., *trans*-elements). TATA-box location, GC content and *cis*-acting elements can be changed to build synthetic promoters with predicted strength^{26,38,39}. There are 59 synthetic promoters characterised for use in plants built with DNA assembly frameworks mentioned in this work. Synthetic terminators

for plants are yet to be developed and characterised but they are available for yeast *Saccharomyces cerevisiae*⁴⁰.

Two most used chassis for part characterisation are the transient expression system in leaf (commonly in *N. benthamiana*) and protoplasts. For leaf infiltration, a construct harbouring the part to be characterised is inserted into *A. tumefaciens*⁴¹. *A. tumefaciens* is cultured and used to infiltrate the leaves using a syringe without a needle. 4-5 days after infiltration, leaves are either imaged for the visible reporter activity, or the reporter molecules are extracted for spectrophotometer quantification. The second method, protoplast transient expression, requires protoplast isolation⁴². Plant leaves are collected, and cell walls are enzymatically digested to generate cells without cell walls (protoplasts). Then the construct with the part to be characterised are transformed into protoplasts by either electroporation or PEG treatment. 12 h to 24 h after transformation, the protoplasts produce enough reporter molecules for quantification. Protoplast transient expression can be used for both monocots and dicots; however, protoplast isolation and transformation are challenging due to the fragile nature of protoplasts. On the other hand, leaf infiltration is limited to dicot chassis so far and may not be suitable for characterising monocot parts.

Both characterisation platforms are prone to experiment-to-experiment variations. In the leaf infiltration system, *A. tumefaciens* distributes unevenly within a leaf and plasmid numbers vary from cell to cell⁴³. Protoplast preparation stresses the cells greatly, and the transformation efficiency and transgene expression can be inconsistent between batches. To overcome such experimental variations, the standard part characterisation protocol recommends a dual reporter system³⁵. Reporter systems are usually luciferase-based (e.g., Nluc and Fluc) or fluorescent protein-based (e.g., GFP and RFP). The first reporter acts as the normalising construct (Fluc), which is the same for each sample. The second reporter contains the part to be characterised (Nluc), which is expressed under the control of the part to be characterised. By calculating the fluorescence ratios of the two reporters (Nluc/Fluc), the normalised strength of a promoter or terminator part can be determined. Even though this method is mostly used to characterise constitutively expressed parts, the same strategy can be applied when characterising a standard part activity within a specific cell or tissue context.

1.1.2.2. Cell- or tissue-type specific promoters

Developmental engineering requires spatiotemporal control over gene expression. Therefore, one of the essential tools for engineering plant development is cell- or tissue-type specific promoters. Development is a temporally and spatially dynamic process rich in cell-type and tissue-type-specific gene activities. Constitutive expression of developmental genes throughout a multicellular organism can be deleterious, and it is advisable to confine the expression to specific cell types or tissues, such as the meristems (SAM, RAM or cambium), roots, shoots, flowers or leaves and distinct cell types therein ⁴⁴.

There are well-characterised promoters that demarcate a specific cell- or tissue-type in the root, in which the radial cell layers have distinct identity (Figure 2B). Marquès-Bueno et al., identified and characterised 28 different promoters varying in expression in different cell-types in the *A. thaliana* root, such as the stele, cortex, and quiescent centre⁴⁵. Even though their expression can be limited to a specific cell-type in the root, some of these promoters such as *PIN4* in columella root cap and stele, is also expressed in the leaves and flowers of mature plants; and it is involved in embryogenesis (Figure 1B)⁴⁶. In addition to the cell types in the primary root⁴⁷, have developed an integrase-based MoClo toolbox to modulate gene expression during lateral root formation. They have used promoters of *AUXIN RESPONSE FACTOR 7 (ARF7)*, *AUXIN RESPONSE FACTOR 19 (ARF19)*, *LATERAL ORGAN BOUNDARIES DOMAIN 16 (LBD16)*, and *GATA TRANSCRIPTION FACTOR 23 (GATA23)* transcription factors (Figure 1B). Similarly, these transcription factors show expression in mature leaves and flowers, although weaker compared to *PIN4*⁴⁶. The basal level expression might be negligible if a threshold gene expression is needed for a developmental switch (e.g., *LEAFY* gene in reproductive development); but if the presence or absence of a gene activity determines the shift, then the basal level expression would be detrimental (like *TOPELESS* in embryogenesis)^{48,49}.

One strategy to overcome the limitations of endogenous cell- or tissue-type specific promoters is to rationally design synthetic promoters. Rationally designed tissue-specific promoters are summarized and underlined in Figure 1B. Light-responsive *cis*-acting elements are incorporated into the synthetic promoters to confer green-tissue specificity as shown with *GSSP1,3,5* promoters ⁵⁰. As the leaf, sheath, panicle and stem are all green, chlorophyll-containing tissues, the transgene expression cannot be restricted to one type of tissue yet; however, expression strength can be varied between the green tissues with *BiGSSP2,3,6,7* ⁵¹. Similarly, *SynR2* synthetic promoter is built for root expression, but its

expression cannot be limited to one cell-type, and there is a basal level of expression in the leaf, stem and seed ⁵². Developments in single-cell transcriptional profiling technologies and identification of new, *cis*-acting elements would facilitate rational design of cell-type-specific promoters in the future.

1.1.2.3. Inducible systems

In addition to tissue-type specific promoters, inducible systems can provide spatiotemporal regulation. A typical inducible system will have an inducible promoter, which will drive the expression of the target gene in the presence of a trigger input, such as inducer chemical, heat-shock, or wounding.

Ethanol ⁵³, β -estradiol ⁵⁴ and dexamethasone or DEX ⁵⁵ are the most used chemically inducible systems in plants, each of which has a distinct method of gene activation. The ethanol inducible system was isolated from the fungus *Aspergillus nidulans*; where AlcR protein binds to *pAlcA* promoter in the presence of ethanol ⁵³. Originally developed more than twenty years ago, the ethanol inducible system was recently shown to cause growth defects in plants ⁵⁶. The β -estradiol and DEX systems action through a chimeric receptor. The chimeric receptor for the β -estradiol system is composed of the DNA binding domain of a bacterial transcription factor, VP16 activator domain and the regulatory region of human oestrogen receptor ⁵⁴. In the presence of β -estradiol, the chimeric receptor changes its confirmation, moves to the nucleus and binds to its target promoter. DEX inducible system works similarly. The transcription activator LhGR is retained in an inactive form by HSP90 protein. When DEX is present, HSP90 releases LhGR, which moves from the cytosol to the nucleus, binds to its *cis*-element *pOp6* and activates transcription of the target gene (Figure 1.C).

Chemically inducible systems are difficult to precisely control due to diffusion through cell layers, and they lack cell and tissue-type specificity by default. However, chemically inducible systems can be combined with cell or tissue-type specific promoters. Schürholz et al., created 19 driver lines, which have a tissue-specific promoter driving the expression of LhGR in different cell types in the root, shoot, SAM, procambium and xylem/phloem precursors⁵⁷. In the absence of the inducer, LhGR is retained in the cytoplasm in inactive state. Effector lines express effectors under the control of *pOp6* promoter. Only if the driver and effector lines are present where dexamethasone is applied, the expression of the effector is activated. A downside of this system is that crossing and generation of F1 lines is necessary and time-consuming.

In addition to chemically inducible systems, optogenetic systems allow induction of gene expression on light cue, combining tight spatiotemporal control and reversibility. It has been challenging to implement optogenetic switches in plants, as they are naturally responsive to a broad spectrum of light. In fact, most optogenetic light sensors originate from plant genes. PULSE system has overcome this by using two engineered photoreceptors: the blue-light module repressing gene expression in the presence of blue light and the red-light module activating expression only in the presence of monochromatic red light⁵⁸. The blue-light module is based on dimerization of photoreceptors. Without blue light, the photoreceptor is in its monomer form and cannot bind to its target domain. When the blue light is present, the photoreceptor dimerizes and binds to its target (Figure 1.C). The red-light module is based on attachment of a different photoreceptor to its DNA binding domain when red-light is present. When the blue-light module has a repressor domain and the redlight module has an activator domain, the blue-light module dominates and represses the target gene expression. Thus, when plants grow under white light (or in darkness), the target gene is not expressed. The PULSE system, in combination with CRISPR-Cas9, was demonstrated to control transient gene expression in *N. benthamiana* leaves and *LEAFY* gene in *A. thaliana* protoplasts⁵⁸.

An ideal inducible system for developmental engineering would be free of any basal level activity when the inducer is absent, while being able to induce gene expression on cue in a specific tissue or cell type. Optogenetic inducible systems outperform chemical systems in terms of leakiness and spatiotemporal specificity^{59,60}. Optogenetic switches tend to be binary (on/off), and the target tissue or cell types need to be accessible to light (i.e., close to the surface). A binary mode might be preferable in cases where a low level of expression of a gene can trigger a developmental switch. If a dynamic range is needed, chemically inducible systems may be preferable. A dynamic range is useful in modifying dose-dependent pathways (e.g., plant hormones such as auxin controls developmental switches throughout the plant life cycle, from embryogenesis to root and shoot patterning, through dynamically changing local accumulation^{61,62}). Another feature to consider when choosing an inducible system is reversibility. Some developmental genes need to be only transiently active (e.g., in regenerating *Arabidopsis* roots, *PLT2* needs to be expressed only at a specific time frame to allow shoot initiation from callus, but continuous expression inhibits shoot formation)⁶³.

1.1.2.4. Logic gates

Logic gates are genetic regulatory networks that mimic electronic circuits; they perform logical calculations—based on the presence or absence of certain inputs and give a binary output (on/off) (Boolean algebra). Inducible systems can be engineered to function as logic gates, and logic gates can be used to build genetic circuits. Recently logic gates for plant systems were developed⁶⁴. A simple logic gate is ON when its designated input is present, and it is OFF when the input is absent. Logic gates can be built to perform more complicated calculations such as AND, OR (demonstrated in Figure 1.D), and IMPLY and NIMPLY. Functionality of these logic gates have been tested in transient assays with *A. thaliana* protoplasts or *N. benthamiana* leaves or stable transformants of *A. thaliana*.

Epigenetic regulation underlies establishment and maintenance of cell fates and reflection of the environmental inputs in developmental reprogramming. Logic gates based on transient expression of inducers or target genes lack memory, as the gate will reverse its output when the inducer availability changes. Recombinase-based gene circuits provide an alternative to engineering cell fate inheritably and for lineage tracking. A recently developed recombinase-based logic gate works by adding a terminator in between the promoter and coding sequence, thereby repressing activation. This terminator is flanked by recombination signal sequences, and when the recombinase is active, it removes the terminator and thus initiates gene expression⁶⁵. This system uses AND/OR gates (Figure 2.D) with two recombinases: one is expressed under tissue-specific promoter as Input [A], (the cortex-specific *CO2* promoter, Figure 2.B) and the second one under a chemically inducible promoter as Input [B], (DEX inducible *pOp6* promoter, Figure 1.C). Therefore, gene expression will take place only in the cortex cells that are supplied with DEX. This system was built by MoClo assembly and shown to work in *A. thaliana* roots with the expression of fluorescent proteins⁶⁵.

1.1.2.5. Multiplex regulation

In addition to modulating the expression of one gene at a specific time and space, plant development often requires simultaneous regulation of multiple genes. Deactivated CRISPR (dCas)-based multiplex gene regulation systems have been developed and proven to be effective for plant systems. Different from nuclease-active CRISPR-Cas9 systems used for genome editing, dCas systems enable gene regulation without changing native DNA sequence, as dCas9 protein still retains the ability to bind its gRNA and target DNA sequence without target DNA cleavage. When dCas is fused to an activation domain (e.g., VP16, ERF2, EDLL), dCas9 carries the activation domain to its target sequence and

upregulate gene expression. Similar to activation, gene repression can be achieved by fusing a repression domain (e.g., BRD, TPL, SRDX) to dCas. Alternatively, by simply using the dCas-gRNA pair alone genes can be repressed, because dCas9-gRNA makes the target promoter inaccessible to hosts' transcriptional machinery⁶⁶. The recently developed CRISPR-Act3.0 system could simultaneously activate six genes in rice⁶⁷. Multiplex activation was demonstrated in *A. thaliana* by targeting two developmental regulator genes *FLOWERING LOCUS T (FT)* and *CL1*. For multiplex repression, three gRNAs repressing exogenous NOS promoter has been reported using dCas12a fused to BRD repression domain (GoldenBraid 4.0)⁶⁸.

In plant systems, combinatorial activation and repression (mixed-mode multiplex) has been limited to targeting one gene for activation and one gene for repression⁶⁹. However, in the yeast *S. cerevisiae*, recently 11 genes could be targeted, upregulating nine and downregulating two simultaneously (Figure 2.E). This system uses 24 gRNAs on a single array, which was assembled by Golden Gate cloning⁷⁰. Multiplex repression seems to be more challenging in plant systems than in *S. cerevisiae*.

1.1.3. Case studies

In the past 5 years, we are seeing more rational engineering of plant growth and development. Below we spotlight three showcase examples where synthetic biology technologies, especially the types of molecular tools mentioned above, were effectively used to modify plant morphology.

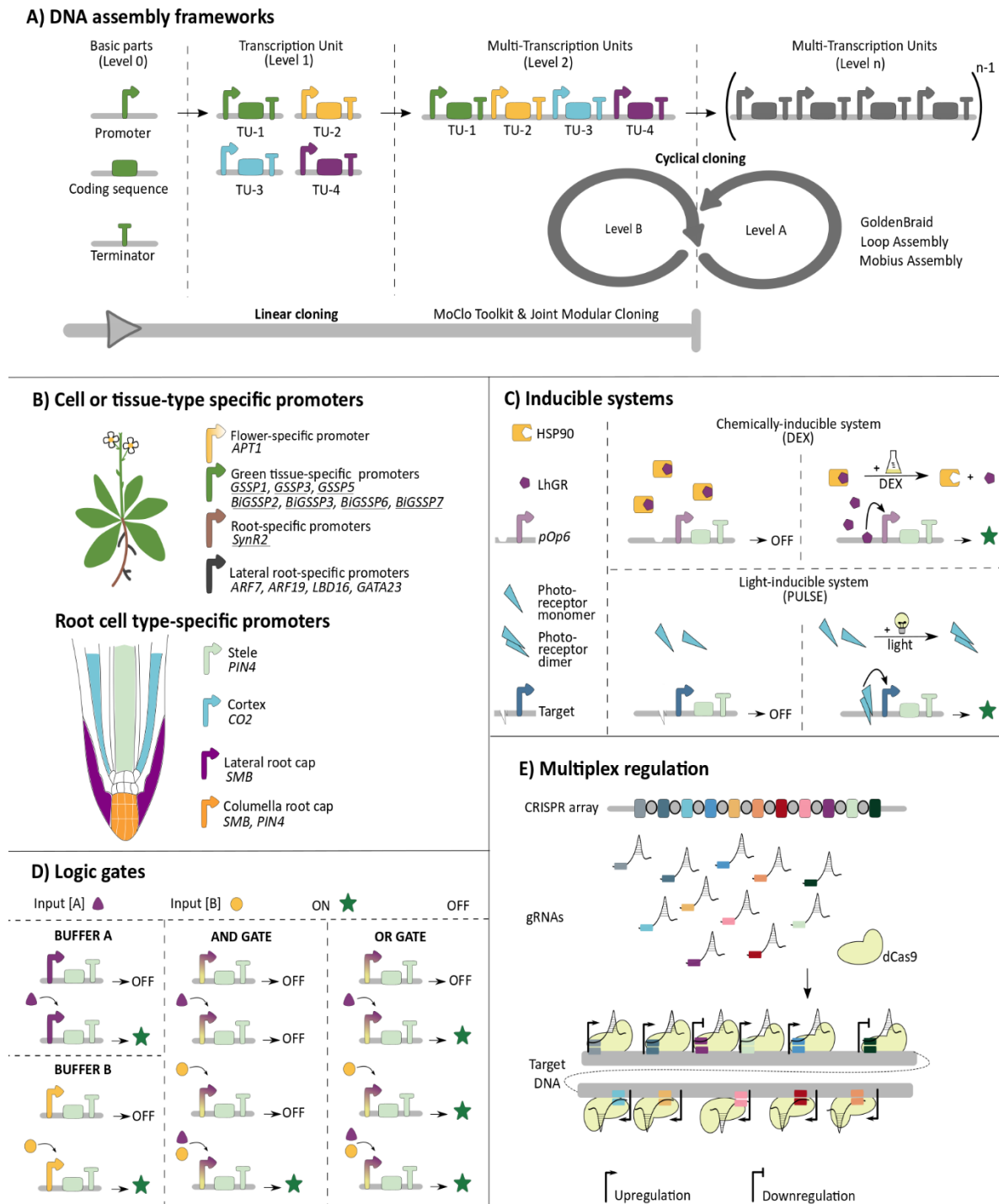


Figure 1. Major types of plant synthetic biology tools for developmental reprogramming. **A)** DNA Assembly frameworks. Basic parts form Level 0, and a transcription unit (TU) can be formed by combining a promoter, coding sequence and terminator at Level 1. Multiple TUs can be combined at Level 2. Linear cloning systems (e.g., MoClo³⁷ and Joint Modular Cloning²⁹) stop at Level 2. Cloning toolkits that follow cyclical cloning (e.g., GoldenBraid³⁰, Loop³¹ and Mobius Assemblies^{33,34}) can switch back and forth between Level A and B, allowing

infinite expansion. Mobius follows hierarchical assembly until Level 2, then joins cyclical cloning. **B)** Cell or tissue-type specific promoters. Promoters mentioned in this review. Rationally designed synthetic promoters are underlined. **C)** Inducible systems. Left panel shows a chemically inducible system based on 15examethasone (DEX). Without DEX, LhGR is bound to HSP90 away from the nucleus. When DEX is added, LhGR is released and can bind to its binding site *pOp6*, inducing gene expression⁵⁵. The right panel shows a light-inducible system, blue-light module from PULSE. Without blue light, the photoreceptor is a monomer and cannot bind to its target. When blue light is present the photoreceptor forms dimer and binds to its target, driving gene expression. In the PULSE system, the blue-light module is later engineered to have repressor domain for gene repression⁵⁸. **D)** Logic gates. AND gate based on two inputs [A] and [B] is at its 'ON' state when both [A] and [B] are present. OR gate is at its 'ON' state when [A] or [B] present, and 'OFF' only when both inputs are absent. **E)** Multiplex regulation. Multiple gRNAs can be expressed in a single CRISPR array and with the presence of dCas9-I systems multiple genes can be up or down regulated simultaneously. The diagram shows 9 upregulation and 2 downregulation reported in *S. cerevisiae*⁷⁰.

1.1.3.1. Shoot architecture engineering with hormone-activated Cas9-repressors (HACRs)

One of the first examples of synthetic biology applied to modulate plant development was HACRs⁷¹. HACRs combine Cas9, an auxin-induced-degron domain and TPL repressor domain to regulate gene expression. Cas9 is used to target HACRs to the promoter of interest, where the TPL domain represses gene expression. Auxin-induced-degron domain enables ubiquitination of HACRs upon auxin accumulation, leading to protein degradation. Therefore, where auxin accumulates, TPL is degraded, and the target gene expression is relieved from repression (Figure 2.A). As the proof-of-concept, HACRs was used to regulate *PIN-FORMED 1 (PIN1)* promoter, which mediates the polar transport of auxin within the cell and accumulation dynamics at the tissue scale. In line with the prediction of computational shoot architecture models^{72,73}, Arabidopsis plants with *PIN1*-targetted HACRs displayed a reduction in the strength of auxin-PIN1 feedback loop, generating a reduced number of branches. Lowering the transcriptional feedback strength was also effective in reducing the irregularity in the phyllotactic pattern that arises stochastically from anomalous distribution of auxin in the SAM.

HACRs have modular hormone-induced-degron domains, which can be swapped to respond to jasmonate and gibberellin, alternatively to auxin. Modular and conserved nature of HACRs make them useful tools to rewire connections between a hormone circuitry and developmental regulators.

1.1.3.2. Controlling growth via stomata opening with synthetic ion channels

Another example of synthetic proteins developed to alter plant morphology is the blue light-inducible potassium (K⁺) channel 1 (BLINK1), which is based on light-inducible control of ion channels⁷⁴. The

light-inducibility is built by fusing a plant photoreceptor module (LOV domain) to a viral potassium channel. In the darkness, the photoreceptor module binds to the catalytic domain of the potassium channel and blocks potassium export. When illuminated with blue light (or white light, which contains blue-range wavelengths), the photoreceptor module changes its conformation, frees the catalytic domain and triggers potassium transport out of the cell (Figure 2.B). As optogenetic systems are reversible, when illumination is off, the photoreceptor module recovers and switches back to its initial confirmation, inhibiting the export function. When BLINK1 was expressed under guard-cell-specific promoter (*pMYB60*), its light responsive function was restricted to stomata, while the potassium channel activity in the rest of the plants functioned normally. As stomatal aperture can be controlled by the polarization of guard cell membrane, BLINK1 can control stomatal opening kinetics. Under fluctuating white light (mimicking light fluctuations in nature), stomatal opening of the BLINK1 plants was faster, and their biomass was 2.2 times more than the wild type plants (Figure 2.B). The increase in their biomass was attributed to more efficient carbon assimilation without compromising water use⁷⁴.

This study presents a novel strategy to enhance carbon assimilation and plant growth under changing environmental conditions. Implementation of BLINK1 to crop species would likely improve agricultural production. The BLINK1 system and similar light-induced synthetic ion channels for other signalling molecules, such as calcium ions, could provide new strategies for engineering plant growth and development, as well as sensing and responding to environmental stressors.

1.1.3.3. Logic gate-based engineering of root architecture

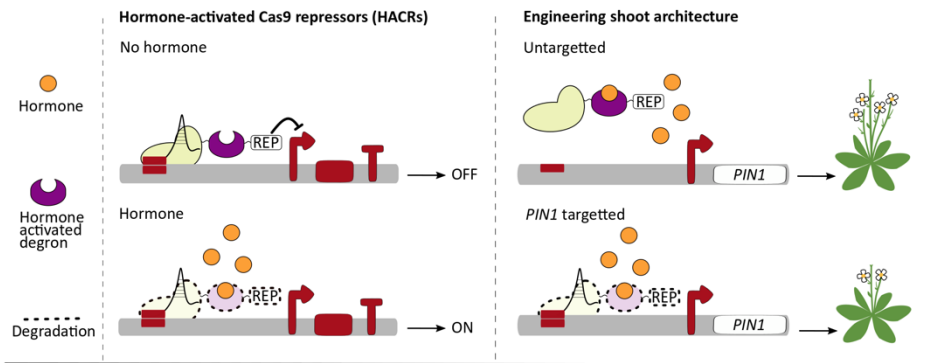
Another example that integrated multiple synthetic biology tools to engineer plant development is the logic gates study by⁶⁴. They combined synthetic activators, repressors and their target promoters to develop ten synthetic transcription factors. After characterising these synthetic transcription factors, the best performing AmtR-VP16 transcription factor (which contains VP16 transcriptional activator domain) is used to build a ANIMPLYB gate (Figure 2.C). In ANIMPLYB gate, gene expression will be turned ON when only Input [A] is present and will be OFF when both Inputs [A] and [B] are present. Using *SOMBRERO* (*SMB*, specific to the lateral and columella root cap) expression as Input [A] and *PIN4* (specific to the lateral and columella root cap) expression as Input [B], they were able to

confine GFP expression to only the lateral root cap (Figure 1.B and Figure 2.C). Using logic gates in this way presented a novel strategy for developing tissue-specific promoters.

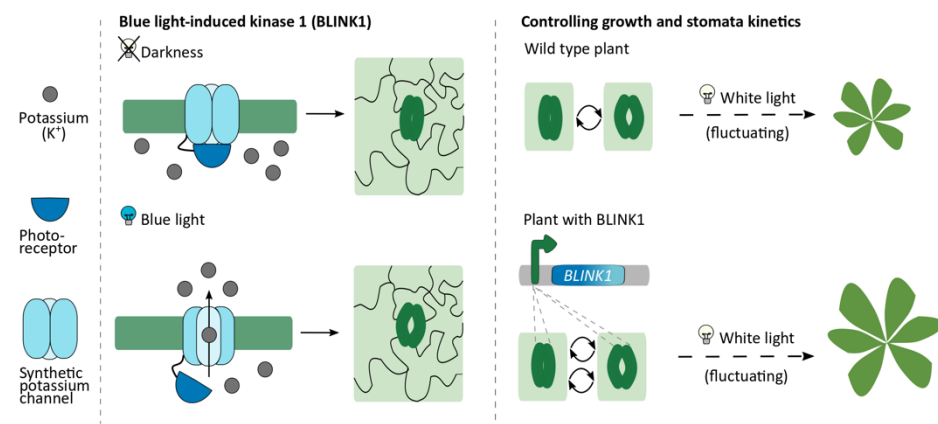
As a pilot study of logic-gate enabled plant developmental engineering, these logic gates were used to engineer root architecture. *GATA23* promoter is active predominantly in the lateral root stem cells (Figure 1B). *GATA23* promoter was designed to drive the expression of Amtr-VP16 activator, which is the input to the BUFFER gate. Amtr-VP16 then switches on the expression of *IAA14* mutant gene (*slr-1*), which blocks root branching. The strength of the synthetic promoter was tuned by changing the copy number of *AmtR* operator (BUFFER 2X gate) and mutating nucleotide sequences of Amtr binding site (BUFFER M3 and BUFFER M1) in synthetic promoters. With this construct, *Arabidopsis* plants with varying levels of lateral root density could be generated.

The study demonstrated the systematic application of engineering principles in modifying plant development by using multiple rounds of design-build-test-learn cycles. It also addressed how fine-tuning of gene regulatory tools is required, especially in the different species and when shifting between transient leaf infiltration-based expression and stable expression mode *in planta*. This study opens possibilities of modifying plant root architecture (the length and branching pattern) to improve soil anchorage and water and nutrient absorption, thus enabling the plants to tolerate drought stress and prevent lodging.

A) Shoot architecture engineering with HACRs



B) Controlling growth and stomata kinetics with BLINK1



C) Logic gate-based engineering of root architecture

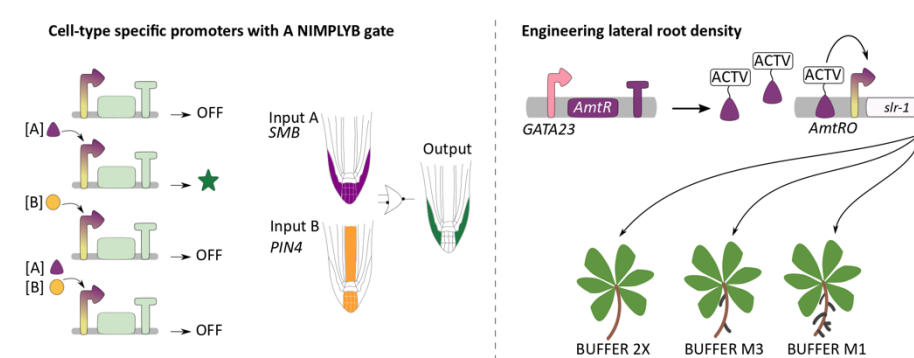


Figure 2. Case studies showcasing the engineering of plant development using synthetic biology. **A)** Shoot engineering using HACRs system ⁷¹. The guide RNA targets HACRs to the promoter of interest, where the TPL represses gene expression. Auxin-induced-degron enables ubiquitination of HACRs upon auxin accumulation, and this leads to degradation of the HACR complex. When auxin accumulates, the target gene is activated. When HACRs are targeted to *PIN-FORMED 1* (*PIN1*), which disperse auxin, they dampened the *PIN1*-auxin feedback loop, causing plants to be shorter and have less branching. **B)** Controlling growth and stomata kinetics with BLINK1 ⁷⁴. BLINK1 is a synthetic potassium channel fused to a plant-derived photoreceptor. In darkness or in the absence of blue light, the photoreceptor inhibits the channel function. When blue light is present, the photoreceptor changes its conformation and triggers channel function. This induces potassium efflux out of the cell, hyperpolarizing the cell membrane and opening up the stomata aperture. When BLINK1 is expressed under a guard cell-specific promoter, the potassium channel function in stomata can be controlled with blue light

illumination. BLINK1 was shown to improve stomata kinetics and therefore carbon assimilation under fluctuating white light. As a result, BLINK1 plants have 2.2 times more biomass than wild-type plants. **c)** Root architecture engineering with logic gates⁶⁴. Cell-type specific promoters were constructed by using ANIMPLYB gate to confine gene expression to one cell type. ANIMPLYB gate activates the target gene when only [A] is present and represses when both [A] and [B] are present. Designing the logic gate to respond to *SOMBRERO* (*SMB*) and *PIN4* expression to drive the target gene expression confined to the lateral root cap. This system was used to tune the lateral root number via the expression of the mutant *IAA14* gene (*slr-1*), which blocks root branching. Synthetic promoter strength could be tuned in order to get a spectrum of lateral root number; the copy number and sequence of *AmtR* operator changes the promoter strength and therefore the lateral root number.

1.1.4. Opportunities and challenges with engineering development

Plant developmental engineering is an emerging field that requires tight spatiotemporal control over target genes. Recent advances in synthetic biology and development of tools, such as DNA assembly toolkits, cell or tissue-type specific promoters, inducible systems, logic gates and CRISPR-based multiplex regulation, are paving the way for precise and predictable control over developmental processes. The proof-of-concept studies show that the technology works and is promisingly widely applicable; however, complications are observed in predictive genetic engineering, pointing to certain improvements and further innovations which are needed in the future.

One recurring problem is interactions among parts; biology is full of emergent behaviours, which are recognized even among genetic parts and the host cellular environment. As more standard parts (e.g., promoters and terminators) are made, characterised and shared in the community, cloning of multigene constructs becomes increasingly feasible and accessible. However, a strong promoter was not always found to be strong; its activity changes depending on the pairing terminator sequence³⁴. Transgene activity is dependent on chassis type; species-to-species variations (e.g., monocot V.s. dicot) are commonly recognized, but even for the same species, a construct behaves differently depending on the cellular context^{75,76}. Such interactions were also observed intergenically among the different transcriptional units within a construct⁷⁷. Together these interactions and context-dependency challenge orthogonality assumption in rational designing of genetic modifications. We recommend validation of part activity in each construct, host chassis and environmental condition used, without assuming the part characterisation data in another study will hold generally.

In order to circumvent the interaction problem and enhance the orthogonality of genetic parts, synthetic promoters and terminators may hold the key. Currently, the standard parts library is expanding for promoters and terminators, which consist of numerous known and unknown *cis*-

elements. If a suite of reliably consistent promoters and terminators, deprived of any interactive DNA sequences, can be developed, they will be instrumental for predictive genetic engineering. Another highly useful future resource is DNA element library, in which we enlist a collection of functional short DNA sequences – for example, non-promiscuous binding site for a cell or tissue type-specific expression. These elements can be used to build developmental logic gates, as in Brophy *et al.*, (2022), to induce gene expression in specific cell types, as well as to prevent basal level expression. The recently developed GB_SynP synthetic promoters are rationally designed, and they can act as synthetic transcription factors in the dCasEV2.1 system³⁹. Inducible systems or logic gates with GB_SynP promoters will be useful in modulating developmental pathways.

Besides the interaction issues, synthetic developmental biology is challenged by transgene instability. Genome-integrated transgenes are often unstable, being genetically rearranged or epigenetically silenced over generations^{78,79}. This problem is not unique to plants and poses a serious problem in multigene modifications common in metabolic or developmental engineering in both plants and animals^{80,81}. As we deepen our understanding of how the transgenes are recognized to trigger the suppression, we can develop and implement new strategies to reduce such transgene instability. For example, the gene regulatory sequences (promoters, and probably terminators and specific *cis*-elements) seem to affect this, and thus they could be characterised for transgene stability. The genomic integration sites also matter; if a transgene construct lands within or near a heterochromatin region, it is more likely to be silenced. Specific landing pads for ‘high stability’ transgene integration would be beneficial^{82,83}. Transgene suppression is intimately linked to defense against pathogens, and its mechanisms are manifold. Some reduction may be possible, but complete suppression of the instability is probably unattainable.

The prime target of plant developmental engineering is root and shoot architecture, as it is thought to increase biomass yield and structural stability directly. Plant body plans may be rationally re-designed through rational engineering of developmental master regulators. *IDEAL PLANT ARCHITECTURE 1 (IPA1)* gene is one such candidate that has been successfully modified to improve grain yield⁸⁴, tiller number⁸⁵, plant height and lodging resistance^{86,87}. However, rational engineering of plant development is not straightforward as it may result in counter-productive outcomes. Dwarf statues intended for better structural stability may reduce the number of tillers and grains^{88,89}. While one may expect more branching would lead to more yield, some branches could develop low-quality

fruit and seeds, reducing the actual yield^{90,91}. In rational engineering of plant development, especially for agronomic traits, structural-functional analysis and normalized design of the new morphological features through multiple cycles of design-build-test-learn cycles are anticipated.

1.2. Engineering regeneration

In addition to modifying shoot and root architecture, engineering regeneration is also a mode of development that offers solutions for sustainability goals. Regeneration, broadly defined, involves the replacement of damaged or lost tissue and organs and it is the basis of horticultural practices and plant biotechnology. For instance, grafting is a commonly employed technique to join different plant tissues, resulting in a chimeric plant that shows increased resistance to both biotic and abiotic stresses⁹². Regeneration is also integral to transgenics, as after foreign DNA is inserted into a cell, whole plants need to be regenerated to obtain transgenic lines for species where alternative transformation techniques such as *Agrobacterium* floral dipping or leaf infiltration is not available⁹³. Furthermore, for conservation purposes, plants facing extinction can be clonally propagated by regenerating whole plants from tissue cultures⁹⁴. However clonal propagation should be performed strategically. Having genetically identical clones is not always desirable due to reduction in genetic diversity and susceptibility of the whole population to same disease with possible risk of extinction, as it was seen with Panama disease and Cavendish bananas⁹⁵.

Types of regeneration

Regeneration involves different mechanisms depending on the tissue or organ being regenerated, each following distinct molecular and cellular pathways. Stem cell activity is important for most types of regeneration, and plants maintain this activity in their shoots, roots and vasculature, in the special population of cells known as meristems. When part of the shoot and root meristems are removed, the remaining meristem cells can re-establish the wounded meristem^{96,97} (Fig. 3a). However, differentiated cells outside the meristem do not contribute to this process.

Plants can regenerate complete organs, in other words, undergo organogenesis. Shoots and roots can be directly induced from leaf or root explants^{98–100}. For example, root explants of *Citrus jambhiri* can produce shoot primordia and shoots upon cytokinin treatment⁹⁸ (Fig. 3b). Alternatively, tissue

explants can generate a callus—a mass of undifferentiated cells before organ regeneration. Factors like injury, stress, and the application of external hormones like auxin and cytokinin trigger this process. When organogenesis involves callus formation, it is called indirect organogenesis since the formation of a callus is an intermediate step (Fig. 3b).

Nevertheless, only specific cell subsets are involved in the callus formation in tissue culture. In *A. thaliana* leaves, the cambium (tissue located between the xylem and phloem) and vascular parenchyma cells are responsible for callus formation¹⁰¹. In the roots it is pericycle cells (located between the endodermis and vasculature)¹⁰¹. This suggests that stem cell-like cells give regenerative response, and somatic cells usually do not contribute to the regeneration process. This concept explains why mature rice plant leaves fail to give a regenerative response in tissue culture, as the stem cell-like cells surrounding the vasculature are depleted during plant development¹⁰¹. However, young rice plants still retain some cells around the vasculature at the leaf base, enabling regenerative response (Fig. 3c). In rice roots, phloem-pole pericycle cells are responsible for callus formation, contrasting with *A. thaliana*, where xylem-pole pericycle cells are involved¹⁰¹. This distinction can be attributed to developmental differences between the two species: in *A. thaliana*, xylem-pole pericycle cells give rise to lateral roots¹⁰², whereas in rice, phloem-pole pericycle cells give rise to lateral roots¹⁰³. Callus formation shows similarity to lateral root development¹⁰⁴, and the regeneration mechanisms evolved to accommodate differences in anatomy and development between the two species.

While plants generally possess much higher organogenic potential compared to the animal kingdom, they exhibit an especially remarkable mode of regeneration: somatic embryogenesis (Fig. 3d). This process occurs when somatic cells lose their cellular identities and revert to a stem-cell-like state through dedifferentiation. Subsequently, these cells undergo embryonic development, leading to the formation of entire plants. Hence, plants are considered capable of re-gaining totipotency¹⁰⁵. This somatic embryogenesis is a key technique for creating transgenic lines and clonal propagation. Similar to regenerative organogenesis, somatic embryogenesis can occur through direct or indirect pathways. Direct somatic embryogenesis, although uncommon and not well-understood, involves the overexpression of embryogenic or meristem-specific transcription factors such as *BABY BOOM (BBM)* and *WUSCHEL (WUS)*¹⁰⁶ (Fig. 3d). In contrast, indirect somatic embryogenesis, which is mediated by callus formation, typically yields more prolific plant regeneration¹⁰⁷ (Fig. 3d).

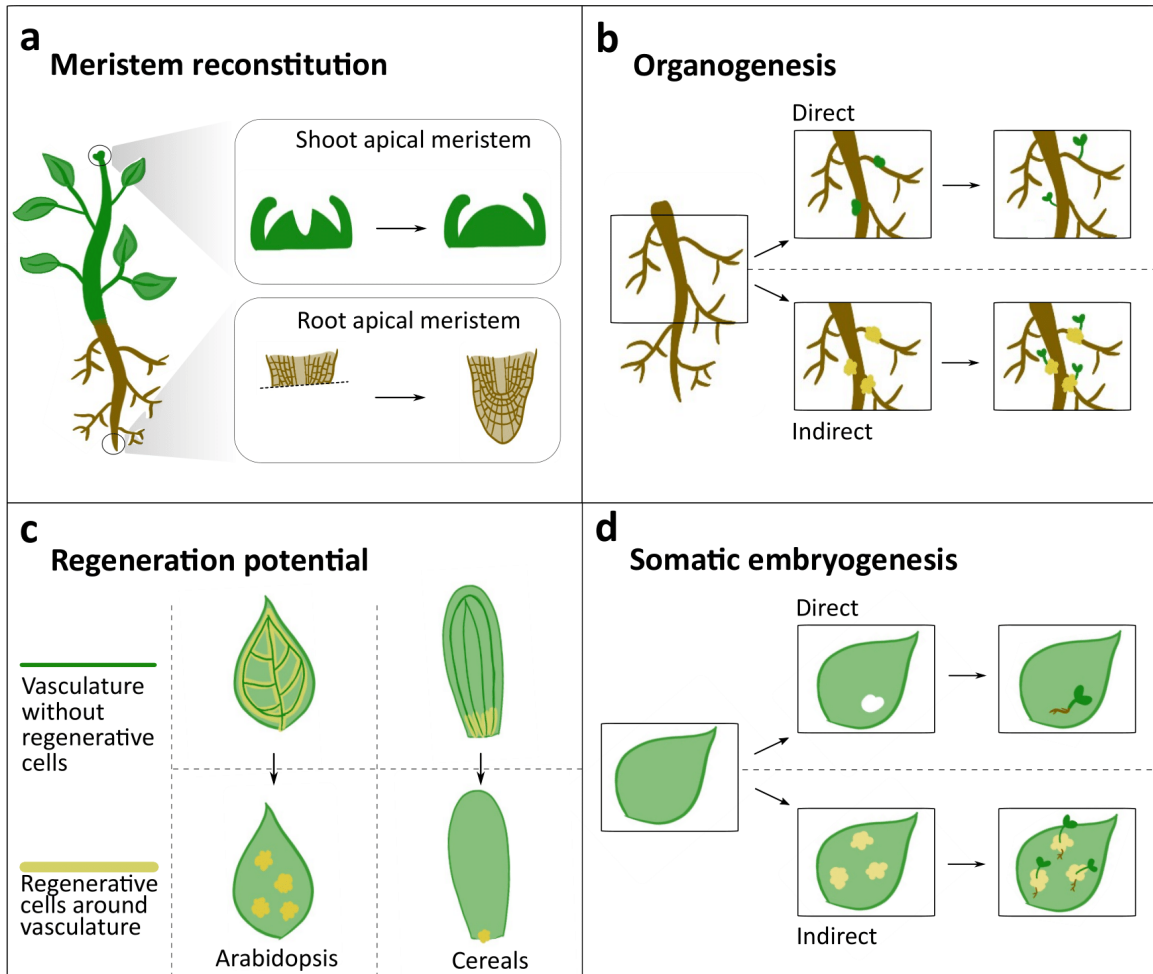


Figure 3. Different types of regeneration. **a)** Meristem reconstitution after laser ablation. **b)** Organogenesis. Direct organogenesis involves organ primordia formation followed by organ regeneration. Indirect organogenesis involves callus formation (yellow mass) prior to organ regeneration. Root is shown as an example but shoot can also go through direct and indirect organogenesis. **c)** Regeneration potential of Arabidopsis and cereals. Arabidopsis has more regenerative cells around its vasculature, whereas in cereals they are only present at the base of young leaves, making Arabidopsis more regenerative than cereals. **d)** Somatic embryogenesis. Direct embryogenesis involves embryo formation (white mass) while indirect embryogenesis involves callus formation (yellow mass).

1.2.2. Protoplast regeneration

While the regeneration potential of plants is already remarkable in tissue culture, as mentioned above, it is further highlighted by protoplast regeneration. Protoplasts are plant cells without cell walls, and they are generated through protoplasting, which is the enzymatic digestion and removal of the cell wall. Protoplasting triggers dedifferentiation and reprogram the cell identities to a stem-cell-like state (Fig. 4a). Then, the cells can re-differentiate into a specific fate, ultimately regenerating whole plants,

with the controlled application of phytohormones and callus formation¹⁰⁸ (indirect somatic embryogenesis, Fig. 4b).

Protoplasting can be performed using multiple types of tissues such as leaves or roots. Obtaining root protoplasts is generally more challenging due to the highly heterogeneous cell types present in roots, each with different cell wall compositions, making enzymatic digestion difficult¹⁰⁹. In contrast, leaf protoplasts consist of more homogeneous cell types and offer a larger amount of tissue per plant, making them the more commonly used source for protoplasts¹⁰⁹. Upon cell wall digestion of leaf tissue, all cell types within the leaf, including those with higher regeneration potential around the vasculature, are released (Fig. 4c). However, studies have shown that approximately 98% of dividing cells originate from protoplasts expressing mesophyll cell identity markers right after isolation, which are subsequently lost with dedifferentiation¹¹⁰.

1.2.3. Dedifferentiation

Dedifferentiation of protoplasts likely begins during the initial stages of cell wall digestion¹¹⁰. A microarray-based transcriptomic analysis revealed that around 3,507 genes were deregulated in freshly isolated protoplasts compared to the intact leaves¹⁰⁸. Notably, among these genes is *WOUND-INDUCED DEDIFFERENTIATION 1 (WIND1)*, an early induced transcription factor known for triggering dedifferentiation and conferring pluripotency in tissue culture^{108,111} (Fig. 4d). Over the course of Day 0 to Day 1, 2,635 genes undergo deregulation¹⁰⁸. Among the deregulated genes, photosynthesis-related genes and chlorophyll synthase genes were downregulated, in line with, chloroplast degradation expected for mesophyll cell dedifferentiation. On the other hand, dehydration (e.g. *EARLY RESPONSIVE TO DEHYDRATION1*) and hypoxic response genes (hypoxia responsive gene, *AT5G27760*, and alcohol dehydrogenases *ADH*, *ATA1*) were found upregulated. However, among the downregulated genes, those associated with nuclear and RNA processes are reactivated starting from Day 1 onward, suggesting that the dedifferentiation process is largely completed by Day 1¹⁰⁸.

These transcriptional changes are likely induced by epigenetic modifications. Histone acetylation has been demonstrated to increase chromatin accessibility in freshly isolated protoplasts compared to intact leaf cells, driving cellular reprogramming^{110,112,113}. Treating protoplasts with Trichostatin A, a histone deacetylase inhibitor, directly after their isolation, enhances their

regeneration potential¹¹⁴. This further highlights the role of histone acetylation and chromatin accessibility in protoplast regeneration.

1.2.4. Cell wall regeneration

Following dedifferentiation, the subsequent stage of protoplast regeneration involves re-differentiation. When exposed to phytohormones, protoplasts can regenerate their cell wall. Initially, around half of cell-wall-related genes (112 out of 217) are downregulated in freshly isolated protoplasts, compared to their expression in leaves, and gradually reactivated in culture¹⁰⁸. Certain cell wall-related genes such as *CELLULOSE SYNTHASEs* and *CELLULOSE SYNTHASE-LIKEs* are upregulated by Day 2, coinciding with the completion of cell wall regeneration^{108,110,115}. *WOX13*, a member of the WUSHEL-related homeobox (WOX) gene family, known for involvement in cell division and inhibition of premature differentiation¹¹⁶, is also upregulated from Day 2 to Day 4¹⁰⁸ (Fig. 4d). *WOX13* plays a critical role in regeneration processes in moss *Physcomitrella patens*, where it facilitates cell growth during stem cell formation by upregulating cell wall loosening enzymes¹¹⁷. Although *WOX13* has been linked to callus formation during grafting¹¹⁸ and shoot regeneration in tissue culture¹¹⁹, its precise role in protoplast regeneration remains unclear. It is probable that it contributes to cell wall remodelling during this process.

1.2.5. Cell cycle progression and cell division

The biosynthesis and regulation of auxin and cytokinin undergo significant changes during protoplast regeneration, which are expected given their crucial roles in plant development. On Day 1, several cytokinin regulatory genes, response regulators and factors were activated¹⁰⁸. The response to auxin is more complex; on Day 0 auxin biosynthesis genes are upregulated due to the wound response, followed by downregulation¹⁰⁸. Additionally on Day 0, the *IAA8* gene, which is associated with lateral root formation¹²⁰, shows a decrease in expression, followed by a gradual increase¹⁰⁸ (Fig. 4d). This pattern aligns with the behaviour of AUX/IAAs, which act as negative regulators of auxin-response factors and typically exhibit an inverse trend to that of auxin-response genes. Endogenously produced auxin (IAA) plays an essential role in enhancing the auxin response of protoplasts. This reinitiates cell cycle by activating of genes controlling G2/M phase¹¹⁰. Importantly, while auxin biosynthesis is crucial for the initial cell division, its importance diminishes in subsequent divisions¹¹⁰.

During the initial week of culture, there is a surge in the activation of cell cycle-related genes, coinciding with the resumption of cell division, peaking at Day 4, when initial divisions are seen¹⁰⁸. On Day 4, *WOX5*, known as one of the master regulators of callus formation¹²¹, is also upregulated, followed by downregulation at Day 7¹⁰⁸ (Fig. 4d). This expression profile might suggest that *WOX5* is required for initial cell division and microcallus formation, but not as much for the later stages of protoplast regeneration. This is similar – and probably linked – to the role of endogenously produced auxin. Auxin application was shown to increase *WOX5* expression, which, in turn, triggers endogenous auxin production during callus formation in tissue culture¹²¹.

1.2.6. Callus formation and whole plant regeneration

After undergoing multiple rounds of cell division, protoplasts develop into microcallus (minimum four cells, 70-500 µm diameter), and subsequently progress to form callus¹⁰⁸ (larger than 500 µm diameter) (Fig. 4b,d). While genes such as *WUS* and *ESR1* (downstream targets of *WIND1*) are expressed in microcalli, key morphogenic and embryonic factors like *WOX5*, *WOX2*, and *BBM* are not expressed during this stage¹¹⁴. Any callus that formed is transferred to shoot-inducing medium for shoot regeneration and subsequently to root-inducing medium for root regeneration¹⁰⁸. The main distinction between these media is in their concentrations of auxin and cytokinin. The entire process of regenerating whole plants from single protoplasts takes approximately four months (Fig. 4b).

1.2.7. Opportunities and challenges of protoplast regeneration

Beyond being a model of dedifferentiation, protoplast regeneration stands out as a promising technique for generating transgenic lines. The integration of foreign DNA into plants remains a challenge, where common methods other than protoplast transformation include *Agrobacterium*-mediated transformation, biolistics, and more recently carbon nanoparticles. These methods each offer distinct advantages and drawbacks, as outlined by Su et al. (2023)¹²². Among these, *Agrobacterium*-mediated and protoplast transformations are the most frequently used to date. While *Agrobacterium* transformation offers simplicity and efficiency, it causes biotic stress and faces limitations in monocots, which are not natural hosts for *Agrobacterium* and thus cannot be transformed¹²². Conversely, protoplast transformation offers universality across monocots and dicots, leveraging polyethylene glycol (PEG), like bacterial transformations¹²².

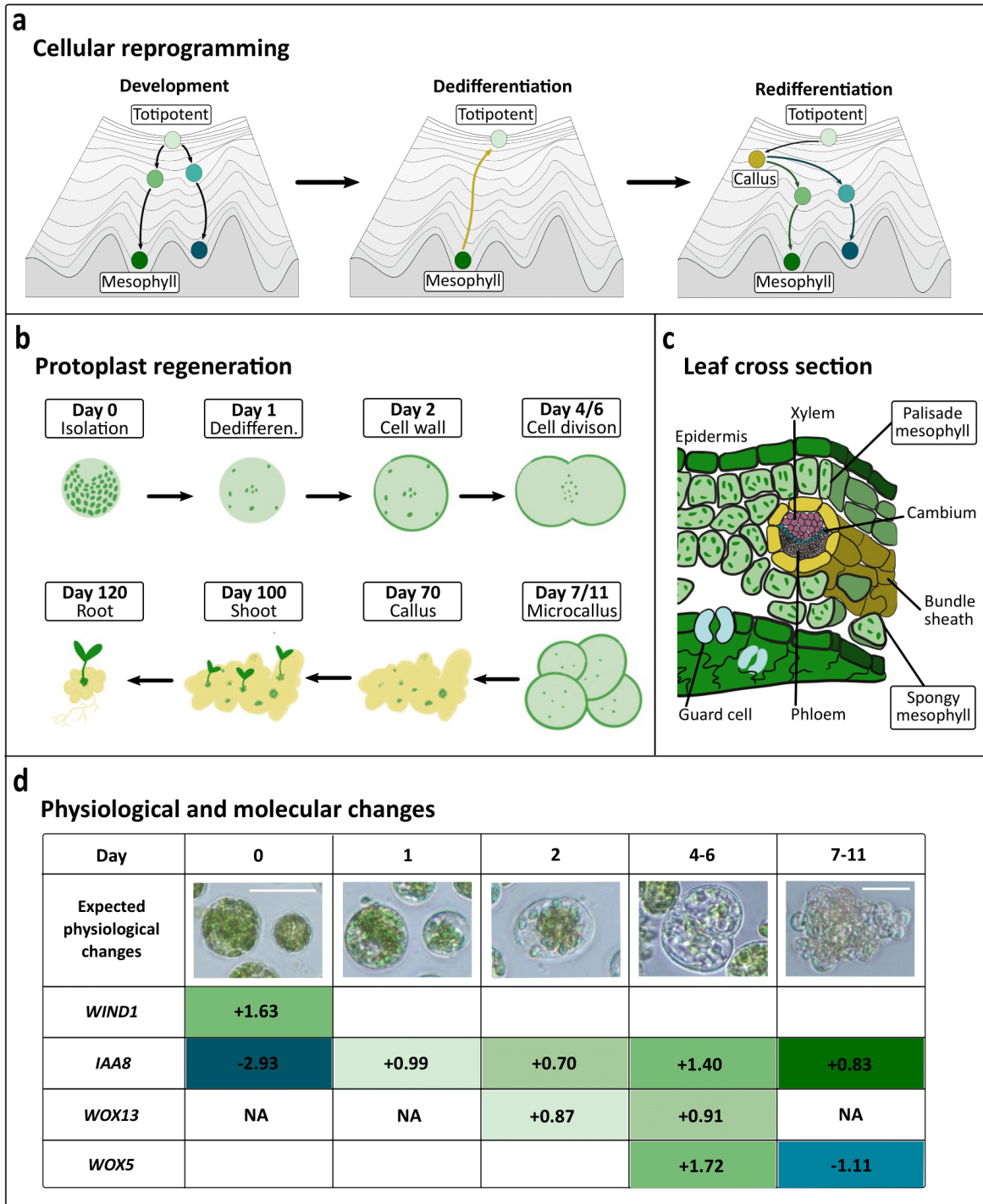


Figure 4. Protoplast regeneration. **a)** Cellular reprogramming during protoplast regeneration. During normal development cells differentiate from totipotent cells (embryo). Protoplasting triggers dedifferentiation where mesophyll cells lose their identities and gain totipotency. During protoplast culturing totipotent cells re-differentiate through callus formation. **b)** Illustration of protoplast regeneration. After protoplast isolation (Day 0), cells dedifferentiate which is completed around Day 1. Then, by Day 2 protoplasts regenerate their cell walls. First cell division occurs between Day 4 and 6. Then after several rounds of cell division microcallus is observed

between Day 7/11. Around Day 70 callus forms, which then regenerates shoots and roots, respectively. Whole process takes around four months. **c)** Simplified illustration of *A. thaliana* leaf cross section. Palisade and spongy mesophyll cells are highlighted with a box as they are the sources of regenerating protoplasts. **d)** Physiological and molecular changes during protoplast regeneration. *WIND1*, *IAA8*, *WOX13*, *WOX5* transcription factors are chosen for their deregulated gene expression and their roles in plant development and regeneration, explained more in Chapter 4. Green colour shows upregulation and blue colour shows downregulation based on RNAseq data reported by Chupeau et al., (2013)¹⁰⁸. Changes in transcript amount is relative to the previous time point. For Day 0 it is relative to living plant. Besides that it is (Day 1/Day 0); (Day 2/Day 1); (Day 4/Day 2); (Day 7/Day 4). Scale bar shows 50 μm .

Following transformation, protoplasts must be regenerated to establish transgenic lines. Protoplast regeneration relies on somatic embryogenesis, a process that does not require the presence of regeneration-competent cells surrounding the vasculature, unlike in cereals where regeneration potential in tissue culture is limited for this reason^{101,110}. Since regenerating protoplasts originate from mesophyll cells, theoretically, all plants possessing mesophyll cells should be amenable to transformation and regeneration. Nevertheless, the regeneration potential of protoplasts, and of regeneration in general, varies among species and genera¹²³. Protoplasts are predominantly derived from families such as Solanaceae and Brassicaceae, which are both dicots¹²⁴. Although there are instances of successful protoplast regeneration in monocots crucial for agriculture, such as the Poaceae family, containing rice¹²⁵, maize¹²⁶, and barley¹²⁷, regenerating wheat protoplasts appears to be challenging¹²⁸. Moreover, for the monocots, protocols typically involve obtaining protoplasts from embryogenic callus rather than mesophyll cells of the leaf. This extra step *in vitro* to obtain callus extends the time spent in tissue culture, consequently raising the chances of somaclonal variation¹²⁹. This is undesirable since it can introduce genetic variation and can impact the efficiency of regeneration. Studying protoplast regeneration from dedifferentiated mesophyll protoplasts, therefore, has the potential to increase the genetic stability of transgenic crops.

One possible explanation for the difficulty in protoplast regeneration in cereal crops is the divergence in cellular and molecular regeneration mechanisms from the dicots, such as *A. thaliana* or *N. benthamiana*¹⁰¹, which are the focus of most established protocols. Similarly, monocotyledonous crops have undergone extensive domestication and selection for desirable agronomic traits, potentially leading to a reduction in genetic diversity related to resilience¹³⁰. Regeneration potential may be one of the traits that has been compromised during this domestication process. Therefore, it is imperative to investigate the mechanisms underlying protoplast regeneration to enhance the regeneration potential of recalcitrant crop species.

As previously mentioned, molecular mechanisms differ among the different types of regeneration, and the majority of our understanding of the molecular mechanisms of regeneration stems from tissue culture studies. The regulatory mechanisms of protoplast regeneration and tissue culture are likely different, as evidenced by distinct gene expression profiles between deregulated genes in protoplasts and callus formed from leaf tissue culture¹¹⁴. This discrepancy suggests that the molecular pathways governing callus formation in protoplast regeneration diverge from those previously identified in tissue culture, introducing an additional layer of complexity that remains incompletely elucidated. Furthermore, there appears to be no common developmental program consistently enriched during protoplast regeneration. Instead, different tissue-specific-genes, including leaf axil, lateral root cap, root quiescent centre, and premature stomata cells, show enrichment at different stages during microcalli formation¹¹⁴. This observation implies that multiple meristem pathways potentially converge, highlighting the necessity of having comprehensive understanding of protoplast regeneration mechanisms before we can engineer them.

1.2.8.

Cell wall and dedifferentiation

One approach to start unravelling dedifferentiation and the subsequent regeneration mechanisms is investigating the potential triggers of the process, which is the removal of the cell wall. This raises the question: what does the plant cell wall do and what does its removal mean? Answering these questions itself is a challenge because the plant cell wall has a great deal of function and its removal can initiate a myriad of cellular responses. The plant cell wall plays a pivotal role in crucial cellular processes such as growth, division, development, signalling, and defence. In line with its diverse functions, plant cell wall has a complex structure consisting of a juxtaposition of sugar polymers (e.g. cellulose, hemicellulose, pectins), proteins (e.g. glycoproteins), phenolic polymers (e.g. lignins) and water¹³¹(Fig. 5). Monitoring cell wall integrity serves as a mechanism for plants to assess their well-being and respond to biotic and abiotic stressors¹³². Thus, the removal of this integrity sensor may initiate cellular programming through various pathways, including chemical, physical, or common signalling mediators.

In terms of chemical triggers, enzymatic cell wall removal generates fragments primarily consisting of sugar polymers, proteins, and phenolic polymers, which can elicit stress responses^{133–135} and potentially trigger dedifferentiation. From a physical point of view, changes in cell shape induced during protoplasting can impact microtubule alignments¹³⁶, potentially initiating transcriptional

responses and stimulating dedifferentiation. Additionally, mechanical cues from the cell wall influence cell fate determination¹³⁷ and organ initiation¹³⁸, highlighting the importance of mechanical cues in development. These triggers lead to the activation of common secondary signalling molecules such as ROS, essential for protoplast division and subsequent regeneration^{139,140}. This highlights the necessity of an optimal redox environment and membrane depolarization for initiating dedifferentiation.

Altering the composition of the cell wall may allow for the modification of cell'' chemical and physiological properties, thereby eliciting downstream developmental signals to change their behaviour. However, this requires a sophisticated level of precision engineering and an understanding of the role of the plant cell wall in dedifferentiation and regeneration, which we currently lack.

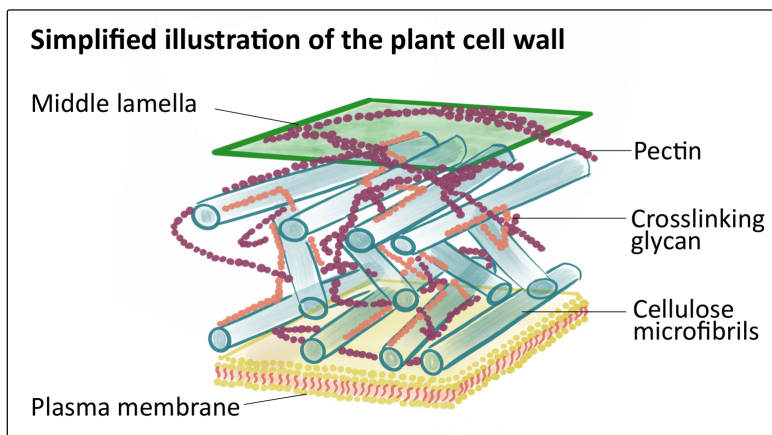


Figure 5. Simplified illustration of the plant cell wall.

1.3. Aims and objectives

To engineer plant development, particularly regeneration, precise spatiotemporal control over gene regulation is imperative. Furthermore, identifying the target genes to regulate remains a challenge, particularly in the context of protoplast regeneration where molecular mechanisms are largely unelucidated. Nonetheless, investigating protoplasts is important as they serve as valuable tools for cell and developmental biology and in generating transgenic plant lines. Furthermore, a promising application of protoplast regeneration involves directing cell differentiation in cell culture to establish sustainable bioproduction platforms. To effectively engineer protoplasts and potentially enhance the regeneration capacity, synthetic biology tools for gene regulation, robust protocols for consistent protoplast regeneration and molecular pathway identification are required. To fill the gaps in current

technologies, my PhD research aims to establish *A. thaliana* mesophyll protoplasts as a synthetic biology chassis.

Objective 1 (Chapter 2): Investigation of the combinatorial interactions between DNA parts to improve predictability of gene expression in *A. thaliana* protoplasts.

I studied the factors contributing to promoter and terminator strength and consistency, and whether coding sequence changes the strength of promoter-terminator pairs. I explored the role of basic part and transcript RNA secondary structure in the combinatorial interactions.

Objective 2 (Chapter 3); Stabilisation of Mobius Assembly vector performance for plant systems.

I identified the plasmid instability issues associated with pMAPS plasmids in *E. coli*, both observed by myself and external users of the Mobius Assembly kit. I modified the pMAPS plasmid backbone to fix the instability issues.

Objective 3 (Chapter 4); Generation of *A. thaliana* stable marker lines that could be used as stage-specific reporters of protoplast regeneration.

I optimised the protoplast transformation protocol and performed a fluorescent protein screening to identify best-performing fluorescent proteins that can overcome autofluorescence. I made promoter cassettes that can be used as stage-specific gene regulation tools and confirmed that the newly made parts function as expected and that the *Agrobacterium*-transformed *A. thaliana* lines are stable over three generations.

Objective 4 (Chapter 5); Establishment of a protoplast regeneration platform.

I monitored physiological markers of protoplast regeneration and used the stable lines developed in the previous chapter to monitor stage-specific marker gene expression. I improved protoplast regeneration by changing the plant growth conditions and cell wall digestion protocols. I examined the effect of these changes on protoplast regeneration competence through qRT-PCR analysis.

1.4. Publications from this study

- **Kocaoglan, E.G.**, Radhakrishnan, D., Nakayama, N. Synthetic developmental biology: Molecular tools to re-design plant shoots and roots. *Journal of Experimental Botany* (2023).
- **Kocaoglan, E.G.**, Andreou, A., Nirkko, J., Ochoa-Villarreal, M., Nakayama, N. Mobius Assembly for Plant Systems uncovers combinatorial interactions among promoters, coding sequences and terminators in gene regulation. *Manuscript under preparation*

CHAPTER 2

**Promoters, coding sequences and
terminators interact
in gene regulation**

2.1. Introduction

Around 11,000 BC, humans transitioned from a nomadic hunter-gatherer lifestyle to permanent settlements and started to rely on agriculture. This shift led to the emergence of early forms of domestication, which involved selection of desired traits¹⁴¹. During this era, many species, primarily cereals, were domesticated independently across the region, from the Levant to Turkey. In fact, the era between 11,000 – 3,000 BC marks the global transition to agriculture, which fundamentally changed the structure of human societies¹⁴¹. However, early forms of domestication, such as seed retention, can be viewed as the outcome of unintentional selection rather than a controlled selection of traits¹⁴¹. Fast-forward to the 21st century, we can now do sophisticated forms of domestication with breeding and genetic engineering methods. This includes the integration of traits such as disease resistance¹⁴², drought tolerance¹⁴³, and biofortification¹⁴⁴ into crops. However, plant synthetic biology aspires to shift towards achieving precise spatiotemporal and dynamic control over target genes in specific cell types. Yet, achieving precision engineering poses its challenges, as it is not always as straightforward as identifying and modulating a single gene target. Plant development often includes intersecting molecular pathways^{145–147}, hormonal interplay¹⁴⁸, and feedback/feedforward loops^{149,150}. Further complicating matters is that gene expression patterns fluctuate based on developmental stages^{151,152} and environmental conditions¹⁵³. Consequently, attaining precision multigene engineering remains an ongoing challenge in plant synthetic biology.

This challenge can be simplified by making standardized DNA building blocks known as Phytobricks¹⁵⁴. Characterising these Phytobricks facilitates predictability at an elemental level and, therefore, in the output gene expression. Phytobricks can be any part of a transcription unit, including promoters, 5' untranslated regions (UTRs), coding sequences, 3' UTRs, and terminators. Promoters and terminators have seen more extensive development in these part collections than UTRs. This might be because some promoters and terminators are designed to contain UTRs rather than separating them as different basic parts¹⁵⁴.

Beyond predictability, construct stability represents another significant challenge in synthetic biology. A stable construct maintains its designed vector and parts (structure and function) over time, under different conditions. In plant synthetic biology, in particular, constructs should be stable in three distinct chassis: *Escherichia coli* (propagation host), *Agrobacterium tumefaciens* (commonly used for

gene delivery), and the target plant species. Consequently, vector stability necessitates confirmation in *E. coli* and *A. tumefaciens*, while parts require plant characterisation.

Existing DNA cloning toolkits, such as MoClo, JMC, GoldenBraid, Loop Assembly, and Mobius Assembly, offer a repertoire of 41 constitutive promoters and 41 terminators characterised in *A. thaliana* and *Nicotiana benthamiana*, both dicot plants^{27,76,155–158}. JMC specifically includes six monocot promoters characterised in *Setaria viridis*⁷⁶. Notably, regulatory parts may exhibit distinct behaviour in monocots compared to dicots; for instance, widely used promoters in dicots like 35S and NOS display weak expression in *S. viridis*. The toolkit underscores that the strength of DNA part'' expression can be species-dependent, necessitating part characterisation in each species of interest without assuming orthogonality^{76,159}. Orthogonality, the ability of DNA parts to function independently of the context and the host background, is a coveted trait in synthetic biology. However, the observed variability in gene expression across species raises the question of whether ultimate orthogonality in synthetic biology is achievable.

The rational design of synthetic parts promotes orthogonality, ormalizedg the presence of unknown functional sequences and reducing plasmid size. Synthetic DNA parts, developed using a bottom-up approach, provide well-defined and minimal units, enhancing our understanding of factors controlling gene regulation. The collection of synthetic promoters is the most comprehensive catalog among DNA parts, which is unsurprising given that the primary emphasis in part characterisation has also been on promoters¹⁶⁰. Although synthetic terminators for plants remain to be developed and characterised, counterparts for the yeast *Saccharomyces cerevisiae* are already available¹⁶¹.

Gene regulation is a classic subject of molecular biology research, which continues to reveal new mechanisms. Engineering with synthetic biology helps to understand how gene regulation works bottom up, as opposed to molecular biology, which focuses on regulatory framework within a natural context. Synthetic biology is thus catalysing our understanding of gene regulation.

2.1.1. Transcription in plants

Gene expression starts with transcription, and in plant systems, it begins with the binding of RNA polymerase II to the promoter region of a gene. Promoters have consensus sequences where transcription factors bind, facilitating the pre-initiation comple''s formation. With the formation of

pre-initiation complex, The DNA helix unwinds, creating a transcription bubble, and RNA polymerase II begins synthesizing a complementary RNA strand from the DNA template¹⁶².

After initiation, the next phase is elongation, where RNA polymerase II traverses the gene that is being expressed and adds ribonucleotides to the growing RNA strand in the 5' to 3' direction¹⁶². Transcription terminates when RNA polymerase II encounters specific termination signals within the gene. These signals trigger the release of the newly synthesized RNA transcript and disassociation of the RNA polymerase II from the DNA template. In plants, termination involves cleavage and polyadenylation signals that direct the addition of a poly(A) tail to the 3' end of the RNA transcript, a process facilitated by a complex of proteins modulated by genetic parts known as terminators¹⁶².

After the transcripts are generated, they are subjected to post-transcriptional processing to produce a mature mRNA ready for translation. This processing includes 5' capping, splicing, and 3' polyadenylation¹⁶². Together, these steps ensure that the genetic information encoded in plant DNA is accurately and efficiently transcribed into functional mRNA, which can then be translated into proteins that execute various cellular functions essential for plant cellular functions.

2.1.2. Promoters controlling gene expression

It is understandable that promoters have garnered significant attention, given their pivotal role in controlling gene expression within plants. Positioned upstream of the genes they regulate, plant promoters serve as essential regulatory domains, influencing the initiation and extent of transcription (Fig. 6a). With synthetic biology, the constituent parts of promoters have been categorized into distinct types: distal, proximal, core, and downstream elements¹⁶³. Distal elements typically harbour enhancers or repressors and engage with transcription factors to govern promoter function. Proximal element contains *cis*-acting elements which function locally on the same DNA strand to facilitate the binding of the transcriptional machinery, including RNA polymerase II and help formation of the pre-initiation complex¹⁶⁴. Core promoter includes B recognition element (BRE) has TATA box immediately next to it. BRE helps TATA-binding protein to bind to TATA box, which together recruit other transcription factors and help form pre-initiation complex. The downstream promoter element is after the transcription start site (TSS), which is sometimes referred to as 5'UTR and can also interact with regulatory proteins to modulate transcription activity (Fig. 6a).

One of the main determinants of promoter strength is the TATA box (TATATAA). The presence of TATA boxes tend to increase gene expression strength, but there are also promoters without TATA boxes¹⁶⁵. In the dataset of Jores et al., (2021); promoters harbouring a TATA box were up to fourfold stronger than TATA-less ones¹⁶⁴. Promoter strength also increases when a canonical TATA box was inserted into a TATA-less promoter. In addition to the presence and absence, the location of the TATA boxes can affect gene expression where the strongest gene expression effect is seen when the TATA box is located at 23 to 59 bp upstream of the TSS¹⁶⁴.

Another significant factor affecting gene expression strength in promoters is the GC content, which is species-dependent¹⁶⁴. Monocot genomes exhibit a higher GC content compared to dicot genomes, a pattern that extends to their promoter sequences (e.g., Rice promoter GC content 43%, Arabidopsis 36%)¹⁶⁴. GC content can affect gene expression in multiple ways, such as the binding affinity of transcription factors and stability of transcription initiation complex.

In addition to GC content, presence of *cis-acting* elements can also affect promoter strength. For example, promoters with an Initiator (Inr) element at the TSS were generally stronger than those without it¹⁶⁴.

2.1.3. Terminators controlling gene expression

Historically, researchers have relied on promoters to control gene expression and neglect the influence of terminators. However, termination signals located at the end of a gene play crucial roles in stopping transcription and ensuring the proper cleavage and processing of RNA. Similar to promoters, the constituent parts of terminators have been categorized as far upstream element, near upstream element and the cleavage element (Fig. 6b). Far upstream element is involved in the efficiency of mRNA 3' end processing. Near upstream element is probably the best defined, which has roles in the efficiency of transcriptional termination and contains the polyadenylation signal (PolyA signal) to signal transcription termination. Cleavage element is typically located downstream of PolyA, and defines the precise site where the pre-mRNA will be cleaved¹⁶⁶.

There are two common theories on how terminators can stop gene expression: Allosteric model and Torpedo models. In Allosteric model, the RNA polymerase changes confirmation upon recognizing a terminator and the Poly-A site on the terminator. Consequently, the polymerase slows down and dissociates from the DNA template that it is attached to (Fig. 6b)¹⁶⁷. However, even after the pre-

mRNA is cleaved, the polymerase is not immediately released from the template DNA, and the polymerase continues to synthesize nascent RNA. Torpedo model hypothesizes that the nascent RNA that is still attached to the elongation complex is degraded by a 5'→3' exonuclease. The exonuclease pursues the elongating polymerase, eventually reaching the polymerase and triggering termination (Fig. 1b)¹⁶⁷. More recently, the unified Allosteric/Torpedo Model describes both means of transcription termination, the conformation change and exonuclease activity, operates simultaneously (Fig.1b)¹⁶⁸.

Termination efficiency has been recently shown to impact gene expression levels¹⁶⁹. Presence of motifs such as the canonical poly-A signal AAUAAA tend to make terminators strong and therefore increase gene expression, where the motifs AAUAAG/U perform nearly as well¹⁶⁹. Poly-A signal functions by signaling cleavage of pre-mRNA, increasing mRNA stability, protection against degradation and transports of RNA from nucleus to cytoplasm¹⁷⁰. Even though strong terminators tend to have one-dominant Poly-A site^{169,171} terminator strength is not just about the polyadenylation. Felippes et al., 2022 have demonstrated that the first 32 bp of *HSP* terminator contains a transferable element (does not contain polyadenylation signal) that increases the strength of other terminators in a PolyA-independent manner (except for only *NOS* terminator, it is dependent on Poly-A)¹⁷².

In addition to polyadenylation, other motifs such as the presence of UGUA around 30-40 bp upstream of cleavage site is thought to enhance cleavage and thus increase terminator strength¹⁶⁹. Presence of UUUUU motif is more complicated, where it decreases terminator strength in maize protoplasts, especially if they surround a UGUA motif. In contrast, U-rich sequences increase terminator strength in tobacco leaves.

Apart from the presence of motifs, the GC content of terminators can influence on gene expression. The specific impact of GC content on terminator strength varies based on the context, the organism, and the regulatory elements at play. Each organism possesses species-specific cellular machinery for pre-mRNA cleavage and processing, contributing to this variability. Terminators with a moderate GC content tend to be stronger compared to those with either high or low GC content, reflecting the nuanced nature of terminator function¹⁶⁹. For instance, the average GC content of terminators in *Arabidopsis* is 32.5%, whereas it is 40.9% for maize terminators. Notably, in tobacco leaves, terminators with a GC content around 30–35% exhibit the highest strength, whereas in maize protoplasts, terminators with a GC content of approximately 35–40% demonstrate superior strength. Interestingly, when experiments are repeated using terminators with random sequences, then in

tobacco leaves those with 30% GC content are the strongest, while in maize protoplasts, random sequences with a GC content of 40% show the highest strength¹⁶⁹ This aligns with the observation relative GC contents of monocot and dicot genomes; and the effect of terminator GC content is context dependent.

In the context of transgene stability, the choice of terminator emerges as a predominant factor influencing whether the transgene transcript undergoes the RDR6-silencing pathway and maintains stable expression¹⁷¹. Transgene transcripts are susceptible to targeting by RNA-dependent-RNA polymerase 6 (RDR6), a mechanism that transforms single-stranded RNA into double-stranded RNA. Subsequently, this double-stranded RNA is cleaved into 21-nucleotide-long fragments, functioning as small-interfering RNAs (siRNAs) through DICER LIKE (DCL4). These siRNAs can bind to complementary regions on the DNA, inducing methylation and gene silencing¹⁷¹. SiRNAs likely bind to the promoter regions of these transgenes. It was shown that constructs lacking terminators produce unpolyadenylated readthrough mRNAs, subject to silencing through RDR-dependent pathway¹⁷³.

2.1.4. 5' and 3' UTRs controlling gene expression

Although sometimes considered as parts of promoters and terminators, 5'UTRs and 3'UTRs also regulate gene expression (Fig. 6). In plant systems, 5' and 3' regulatory components are commonly derived from wild-type or minimally altered plant viruses are utilized. In the evolutionary dynamics of plant RNA viruses, UTRs play dual roles: they must facilitate viral RNA replication (in the 3' to 5' direction along the positive-stranded viral RNA) and ensure efficient translation of viral proteins (in the 5' to 3' direction along the same positive-stranded RNA)¹⁷⁴. Balancing these replication and translation functions is crucial for optimizing the viral replication cycle. However, in plant transgene constructs, the objective is the stable accumulation of the desired protein at a predetermined level¹⁷⁴. Consequently, investigating factors that govern gene expression and designing synthetic UTRs with known functions is essential for predictability and orthogonality in precision engineering.

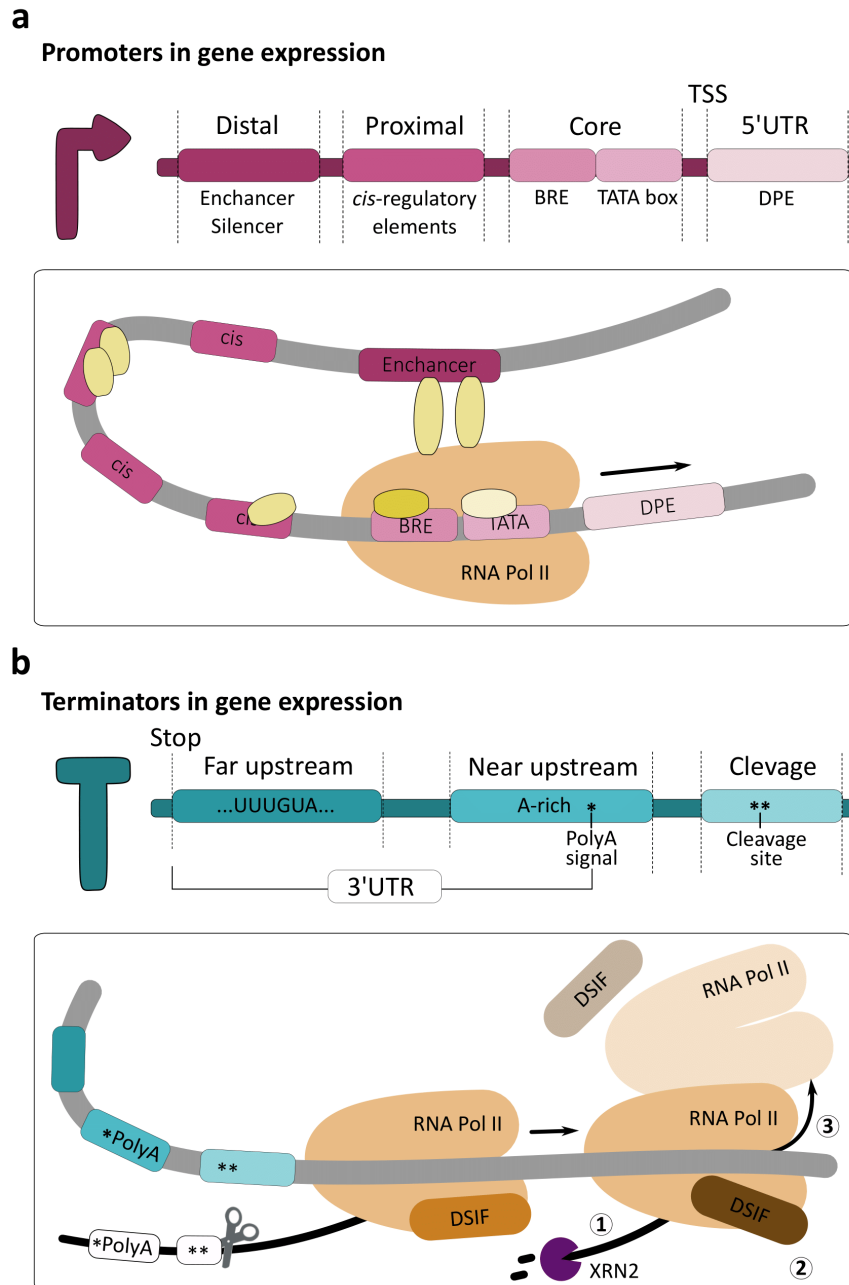


Figure 6. Promoter and terminator's role in gene expression. 5'UTR and 3'UTRs are sometimes considered as part of promoters and terminators, respectively. **a)** Promoters controlling gene expression. Promoter domains can be separated into distal, proximal, core, and 5'UTR (or downstream promoter element DPE). BRE and TATA box help recruit transcription factors and RNA polymerase II and forms pre-initiation complex. Distal, proximal and downstream promoter elements can interact with transcription factors to regulate promoter activity. Yellow ellipse indicate transcription factors. **b)** Terminators in gene expression. Terminator domains can be categorized as far upstream, near upstream and cleavage elements. Number 1 represents the Torpedo model; and Numbers 2 and 3 represent the Allosteric model. Unified model suggests 1,2 and 3 happens simultaneously. DRB sensitivity-inducing factor (DSIF), exonuclease that degrades RNA in a 5'→3' direction (XRN2)

It has been proposed that in plant systems the 5'UTR enhances translation efficiency, while the 3'UTR increases mRNA accumulation¹⁷⁵. To investigate gene regulatory factors Peyret et al., (2019)¹⁷⁴ have tested 4 modified 5'UTR and 9 modified 3'UTR. Modified 5'UTRs vary in length (60-70 bp), GC content 30%-41%, number of AAC motif and they avoided AUG start codons, which are known to have inhibitory effects and are present in around 30% of plant mRNAs¹⁷⁶. 5'UTRs that are shorter (and thus have less secondary structure) and lack AUG performed better than the wild type. Additionally, it was demonstrated that introducing two AUG codons in 5'UTR reduces expression of *AtABI1* and *OsBRI1* genes between 38%-78%, with the repression magnitude depending on the location and length of the resulting transcript¹⁷⁷. This reduction is at post-transcriptional level since mRNA levels were similar, in line with Meshcheriakova et al. (2014)¹⁷⁵.

Similar to modified 5'UTRs in Peyret et al., (2019)¹⁷⁴, modified 3'UTRs vary in length 170–200 bp, GC content (33-42%), number of AAUAAA poly-A signal sequences upstream and potentially UUUU sequences downstream of the CA sequences. Contrary to modified 5'UTRs, modified 3'UTRs have weaker gene expression than the wild type sequence. While the studies referenced here do not utilize truly "synthetic" UTRs which involves creating a large dataset of randomized sequences and incorporating rationally designed features, nevertheless they are starting to shed light on important regulatory features.

2.1.5. DNA and RNA structure controlling gene expression

Beyond the discussed sequence features of Phytobricks, structural part characteristics can also play a role in gene expression. The most stable DNA structure is the canonical right-handed double helix, known as B DNA. However, certain environments and sequence motifs can promote alternative conformations, referred to as non-canonical secondary structures. These non-canonical structures have the potential to impact gene expression, particularly at promoter regions, by serving as landing sites for transcription factors¹⁷⁸. Additionally, they may act as physical barriers for nucleosome formation, enhancing accessibility¹⁷⁹. Regions with high secondary structures can induce RNA polymerase pausing¹⁸⁰, and DNA-looping through secondary structures can bring distant regions into physical contact, influencing gene expression, for instance, between promoters and terminators¹⁸¹.

Apart from DNA, RNA secondary structures also have effects in regulating gene expression. Lower mRNA folding energies, typically associated with greater structural stability, have been found

to impact ribosome function and translation in organisms such as *E. coli* and *S. cerevisiae*^{182,183}. Beyond global structures, local secondary structures can also play a pivotal role in shaping gene expression. Notably, pseudoknots, hairpins, loops, and the presence of hydrogen bonding patterns upstream of the transcription start site have been shown to reduce ribosome sequestration and translation rates^{184,185}. A design-of-experiments study with 224,000 synthetic sequences in *E. coli* has highlighted the critical role of secondary structures at the 30 bp region both upstream and downstream of the start codon (referred to as STR-30:+30), STR+1:+60 and STR+31:+90¹⁸⁶. Furthermore, it revealed that even distal regions to the STR can exert substantial influence over transcript and protein abundance by outcompeting initiation-limiting structures¹⁸⁶. While comprehensive studies in plants are still pending, *in vivo* genome-wide profiling of *A. thaliana* RNA secondary structures revealed that unstructured regions at 5'UTRs are enriched in high translation efficiency mRNAs¹⁸⁷. RNA structure has been suggested to take part in polyadenylation processes as well. Mutating the canonical polyadenylation signal AAUAAA to AAGAAA retains the wild-type structure and maintains original polyadenylation efficiency, while a 3-nt mutation (UAGAAU) disrupts the structure, reducing polyadenylation by 50%¹⁸⁸.

RNA structures are inherently dynamic rather than static, and their conformation can undergo dynamic changes in response to changing environmental conditions. For example, at higher temperatures bHLH TF PHYTOCHROME INTERACTING FACTOR 7 (PIF7) translation is enhanced due to its RNA taking a less structured conformation¹⁸⁹. Similar mechanism is in place for the light-responsive PHOTOSYSTEM II REACTION CENTER PROTEIN A (psbA) whose 5'UTR of RNA becomes less structured after exposure to high light intensity, increasing its expression¹⁹⁰.

The study of RNA structure at the 3'UTR appears more extensive in the context of plant viruses, as many viruses utilize 3'Cap-Independent Translation Enhancer (3'CITE)¹⁹¹. These are sequences exposed on the surface of loops in the 3'UTR, which base-pair with complementary sequences exposed on loops in the 5'UTRs. This establishes long-distance RNA:RNA interactions and recruits translation initiation factors and regulates protein expression¹⁹¹. There are seven classes of 3'CITEs, which have different secondary structures, interactions with 5'UTRs and regulatory features¹⁹¹. It is highly likely that this regulatory potential exists in non-viral 3'UTRs as well.

While how individual basic parts participate in gene regulation through sequence and, to a lesser extent, structural features are under study, the interactions between these parts remain largely

unexplored. Therefore, this study addresses the combinatorial interactions between promoters, coding sequences, and terminators, revealing context dependency within a transcription unit.

2.1.6. Mobius Assembly for Plant sSystems (MAPS) shows interactions between genetic parts

This study demonstrated that terminators can alter the gene expression strength of a promoter by up to sevenfold. Additionally, it was shown that modifying the coding sequence results in varying strengths among promoter-terminator pairs. Certain terminators, such as the *NOS* terminator, exhibit changing relative expression strength, being either weak or strong depending on the specific promoter-coding sequence combination. Notably, there was only two consistently strong terminators, *HSP* and *FAD2*. On the other hand, the *Rbcs2b* and *E9-RbcS* terminators were consistently weak. The observed differences in gene regulation appear to predominantly impact abundance of transcript levels in luciferase samples, which is not the case for fluorescence samples. RNA-structures of both the basic parts and the transcription unit, are suggested as one of the factors participating post-transcriptional gene regulation.

2.2. Results

2.2.1. Promoter characterisation reveals that *UBQ10* and *MAS* promoters are consistently strong

For the initial characterisation experiments, luciferase system was used since it gives stronger and more sensitive signal. To evaluate promoters, the *NLuc* transcription was terminated either by *HSP* and *NDUFA8* terminators, while 17 different promoters drove the transcription initiation. For more description on the selection of promoters used in this section, readers are referred to Andreou et al., 2021³⁴. Notably, these promoters are constitutive, meaning that they are active in most plant tissues, although likely at different levels. *UBQ10* exhibited by far the highest expression activity among the promoters, followed by *MAS* and *UBQ11* when combined either with *HSP* or *NDUFA8* terminator (Fig. 7a,b). The *HSP* terminator resulted in elevated gene expression compared to *NDUFA8* terminator for all the promoters, except *TUB9*. The lowest expression was found for *FAD2* and *NDUFA8* promoters.

After characterising promoters with luciferase, the reporter was swapped to fluorescence as it is more commonly used and does not require addition of substrate. However, using fluorescent proteins in protoplasts can be challenging as protoplasts emit red and green range autofluorescence from their chloroplasts; and blue range autofluorescence from their cytoplasm¹⁹². Over the years, fluorescent proteins with improved brightness and dynamics have been developed; but their expression haven't been characterised in protoplasts. Fluorescent protein screening was done in protoplasts (details in Chapter 3) and the brightest proteins two from different spectrum (sfGFP and mScarlet-I) were identified. *sfGFP* was used as the main reporter, while *mScarlet-I* was used for the normalising construct, being expressed under *UBQ10* promoter and with *UBQ5* terminator.

Similar to the luciferase system, promoters were evaluated with *HSP* terminator for *sfGFP* expression. Unlike luciferase reporters, not all the promoters drove strong enough expression that can be detected with a plate reader (Fig. 7c). The highest expression was from *UBQ10* (6.0 RFU) and *MAS* promoters (6.3 RFU), followed by *35S* promoter (2.2 RFU). Signals from the rest of the promoters could not be distinguished from the background autofluorescence.

It was interesting to see that gene expression strength rankings of promoters were changing when part combinations were changing (*NLuc:HSP*, *NLuc:NDUFA8*, *sfGFP:HSP*). *UBQ10* and *MAS* promoters were the only ones resulting in consistent strong expression in all three combinations tested. Even though *UBQ11* and *OCS* promoters were amongst the highest expression in *NLuc* system, with *sfGFP* their expression wasn't strong enough to be detected (*OCS* promoter 2.0 RFU, *UBQ11* promoter 1.8 RFU, NC 1.7 RFU) (Fig.6).

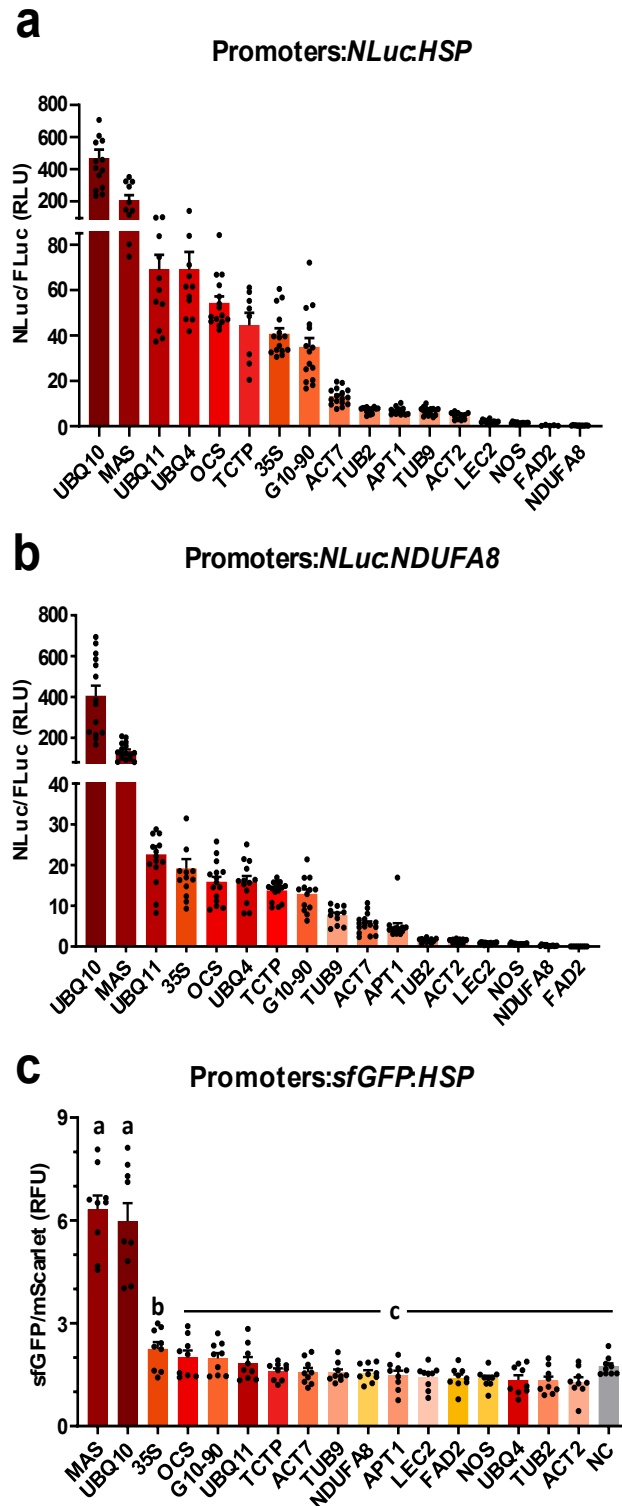


Figure 7. Dual-reporter characterisation of promoters. The promoter library was characterised with **a)** Promoters:*NLuc:HSP* **b)** Promoters:*NLuc:NDUFA8* or **c)** Promoters:*sfGFP:HSP*. *A. thaliana* leaf mesophyll protoplasts were transformed with PEG, and the expression strength of the promoters was assayed in 96-well

plates. For luciferase reporters (a and b) Nano luciferase activity normalized by Firefly luciferase (*UBQ10::fLuc::UBQ5*). For fluorescence reporters (c) superfolder GFP level normalized to mScarlet-I level (*UBQ10::mScarlet-I::UBQ5*). Bar graphs show mean reporter activity values $\pm$ SE. Negative control (NC) was untransformed protoplasts. For luciferase reporters, NC values were subtracted from measured sample values during deconvolution, making negative control values 0. For each sample, three biological replicates (protoplasts originating from different plants) are transformed which have three technical replicates each (protoplasts originating from the same plant, transformed in separate wells but simultaneously). All the data points are shown on the graph as individual dots. One-way ANOVA with Tukey's HSD was done individually for each terminator category. Letters indicate statistical significance, where different letters indicate statistically significant differences in the data. (NLuc= Nano and Fluc=Firefly luciferase. sfGFP=superfolder green fluorescent protein). Results in panels a and b are obtained by previous lab members¹⁵⁸.

2.2.2. TSS structure of the *UBQ10* promoter affects gene expression levels

Given that *UBQ10* is the strongest promoter in the toolkit, I aimed to investigate the characteristics contributing to its strength. *UBQ10* TSS was predicted by comparing TSSPlant online tool output and the available transcript data. Multiple TSS were predicted, but the TSS that is experimentally shown to constitute the majority of the transcripts (66 bp before the ATG codon) were chosen for the following experiments. Notably, its TSS structure is predicted to be unstructured and single-stranded (Fig. 8a,b), which might be affecting the promoter's strength. Consequently, three versions of the wild type (WT) TSS were designed. The first version, termed Mutated Control (MUTC), involves point mutations that preserve the predicted structure. Additionally, two other versions, Mutated 1 (MUT1) and a slightly more structured Mutated 2 (MUT2), were designed with point mutations altering the TSS structure (Fig. 3a,b). TSSPlant and Softberry software were used to confirm the mutated versions do not interfere with the identified motifs or introduce new motifs compared to the WT. When the TSS variants were used to express *sfGFP* with five different terminators (*HSP*, *FAD2*, *35S*, *NOS*, *RbcS2b*), the results revealed no statistically significant difference in expression between WT and MUTC (Fig. 8c). However, MUT2 exhibited a reduction in expression by approximately 20% for the *HSP* and *FAD2* terminators. Additionally, MUT1 resulted in a 60% decrease in *35S* expression. In contrast, MUT1 led to a 30% increase in *sfGFP* expression with the *NOS* terminator. There was no difference between the three mutated versions and the WT for the *RbcS2b* samples, where the WT expression is already weak (Fig. 8c).

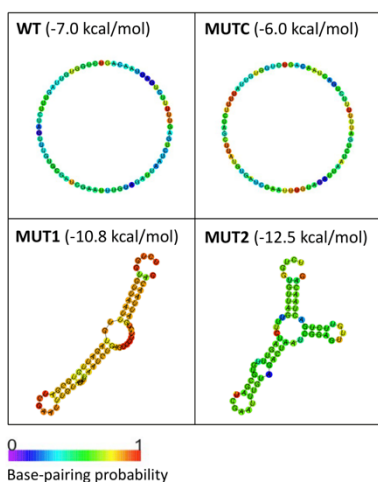
a

Designed mutations at *UBQ10* transcription start site (TSS)

Wild Type (WT)	TCTGGTGTAGTTTCTAGTTTGTGCGATCGAATTTGTAGATTAATCTGAGTTTTTCTGATTACAGAATG
Mutated control (MUTC)	T C A T T C
Mutated 1 (MUT1)	A G A C G G C
Mutated 2 (MUT2)	

b

Predicted structures of *UBQ10* TSS



c

Effect of TSS structure on reporter expression

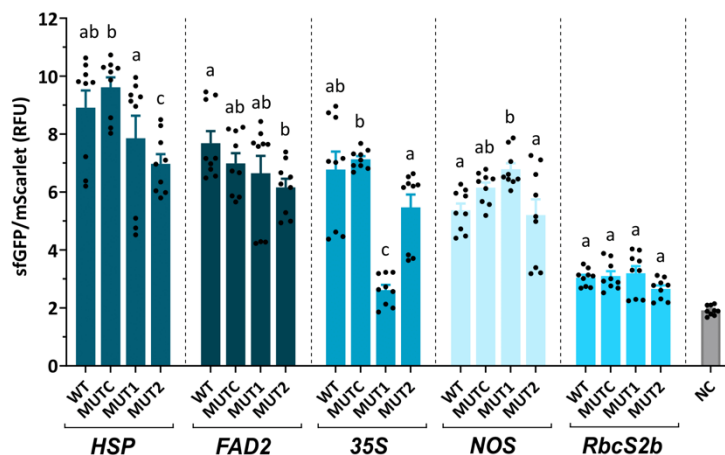


Figure 8. TSS structure affects gene expression levels. **a)** Nucleotide sequences showing the point mutations made for the Wild Type (WT), Mutated control (MUTC), Mutated 1 (MUT1) and Mutated 2 (MUT2) sequences. **b)** Predicted structures for the WT, MUTC, MUT1 and MUT2 sequences. **c)** Effect of changing the TSS structure on reporter expression. *sfGFP* was expressed with *UBQ10* containing WT, MUTC, MUT1 or MUT2 type TSS. Five different terminators were chosen to assess the effect: *HSP*, *FAD2*, *35S*, *NOS* and *RbcS2b*. One-way ANOVA with Tukey's HSD was done individually for each terminator category. Letters indicate statistical significance, where different letters indicate statistically significant differences in the data. Normalisation construct was *UBQ10:mScarlet-I:UBQ5* for all samples. Bar graphs show mean reporter activity values $\pm$ SE. Negative control (NC) was untransformed protoplasts. For each sample, three biological replicates with three technical replicates are shown on the graph as individual dots.

2.2.3. Terminators affect gene expression in inducible systems

The fact that changing the part pairings altered promoter strength of constitutive promoters, prompted inducible promoter characterisation. Three chemically inducible systems are commonly used in plant sciences: dexamethasone, estradiol and ethanol inducible systems^{193–195}. They are all based on two-component mechanisms involving at least two transcriptional units, in which the exogenously applied chemical activates the transcription factor that further 'transacti'ate the downstream target genes. The target genes are driven by the specific promoters that contain binding

sites for the transactivator (*pOp6*, *lexA* and *alcSynth*, respectively), and hence promoters cannot be changed. Terminators, on the other hand, can be altered. Interestingly, terminators combined with *pOp6-35S* and *lexA-35S* inducible promoters resulted in narrow and more uniform *NLuc* expression, spanning at around 2.2 and 2.9 orders of magnitude, respectively (Fig. 9a,b). For *lexA-35S*, the expression is spread between 37.9 and 111.6 RLU when combined with *UBQ5* and *35S* terminators, respectively. 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 a wide array of expression activity similar to the constitutive promoters; the expression with *HSP* terminator (337.9 RLU) is 7.2 times higher the expression that *LEC2* terminator provides (46.7 RLU) (Fig. 9c). Both *pOp6-35S* and *lexA-35S* have a basal expression of around 10 RLU (when combined with *HSP*); however, *lexA-35S* has demonstrated ~11-fold activation amplitude (111.6 RLU), while *pOp6-35S* had only three times activation amplitude (25.1 RLU). Combined with *HSP* terminator, *alcSynth* basal expression was very high (250.3 RLU), leading to only 1.3 times of activation (337.9 RLU) upon treatment with ethanol, demonstrating the leakiness of the ethanol-inducible system.

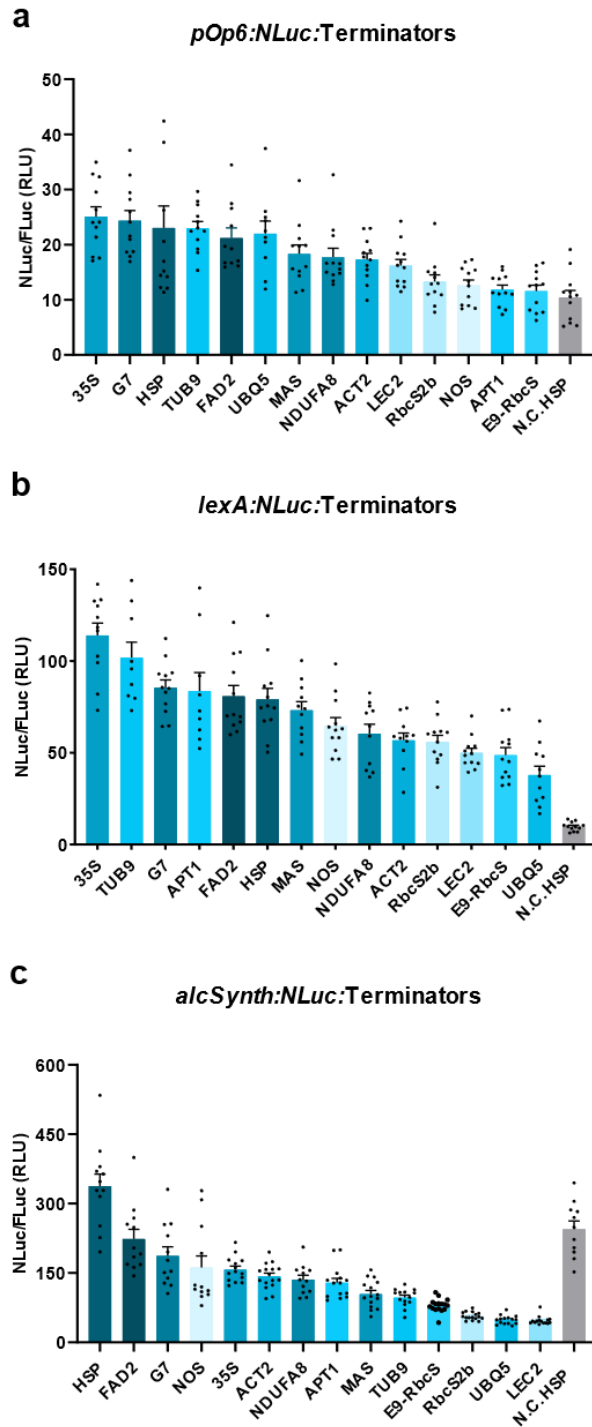


Figure 9. Dual-reporter characterisation of inducible promoters. Terminators were combined with the inducible promoters in following combinations: **a)** *pOp6-35S*; **b)** *lexA-35S* and **c)** *alcSynth-34S*. *A. thaliana* leaf mesophyll protoplasts were transformed with PEG, and the expression strength of the promoters was assayed in 96-well plates. The construct used for characterisation was [UBQ10:Transactivator:HSP] [Promoter:NLuc:HSP] with [UBQ10:FLuc:UBQ5] as normalising construct. 'Transactivator' corresponds to a) *LhGR*; b) *sXVE* and c) *Al'R*. 'Promoter' to a) *pOp6-35S*; b) *LexA-35S* and c) *alcSynth-35S*. Chemical induction was achieved with a) 2.5μM

DEX; b) 5 μ M β -estradiol and c) 0.1% EtOH for. NLuc = Nano and FLuc = Firefly luciferase activity. Bars normalized luciferase activity values in mean $\pm$ SE. NC = negative control which is uninduced cells used to evaluate leakiness of the inducible systems. Results in this figure are obtained by previous lab members¹⁵⁸. For each sample, three biological replicates with three technical replicates are shown on the graph as individual dots.

2.2.4. Terminators affect gene expression of constitutive promoters

All of the 14 terminators in Mobius Assembly Toolkit were then characterised with the strong *UBQ10* (Fig. 10a) and the weak *NDUFA8* promoter (Fig. 10b), both driving *NLuc* expression. Changing the terminators resulted in a wide range of expression spanning 5.3 and 6.3 orders of magnitude for *UBQ10* and *NDUFA8* promoters, respectively (Fig. 10a,b). For the strong *UBQ10* promoter, the *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, the *HSP* terminator led to the highest expression with 0.327 RLU and *APT1* to the lowest at 0.052 RLU.

2.2.5. Changing the coding sequence changes the strength of promoter-terminator pairs

After seeing the combinatorial effect of promoter-terminator pairs, the next step was to investigate whether changing the coding sequence changes the strength of the same pairs. So, fluorescence reporters were used instead of luciferase. Since *UBQ10* and *MAS* promoters were the only ones that could be reliably detected with sfGFP, they were evaluated with all the terminators (Fig. 10c,d). *HSP* terminator was the strongest with both promoters (9.0 RFU with *UBQ10* and 12.1 RFU with *MAS*). The *FAD2* terminator was on the high-expression side for both promoters, along with *UBQ10*pro-35Sterm and *MAS*pro-*NOS*ter combinations. *Rbcs2b* terminator gave the lowest expression with both *UBQ10* (3.1 RFU) and *MAS* (2.3 RFU) promoters. Overall, a 3-fold expression range is observed by using different terminators with *UBQ10* promoter; and a 6-fold expression range is observed with *MAS* promoter.

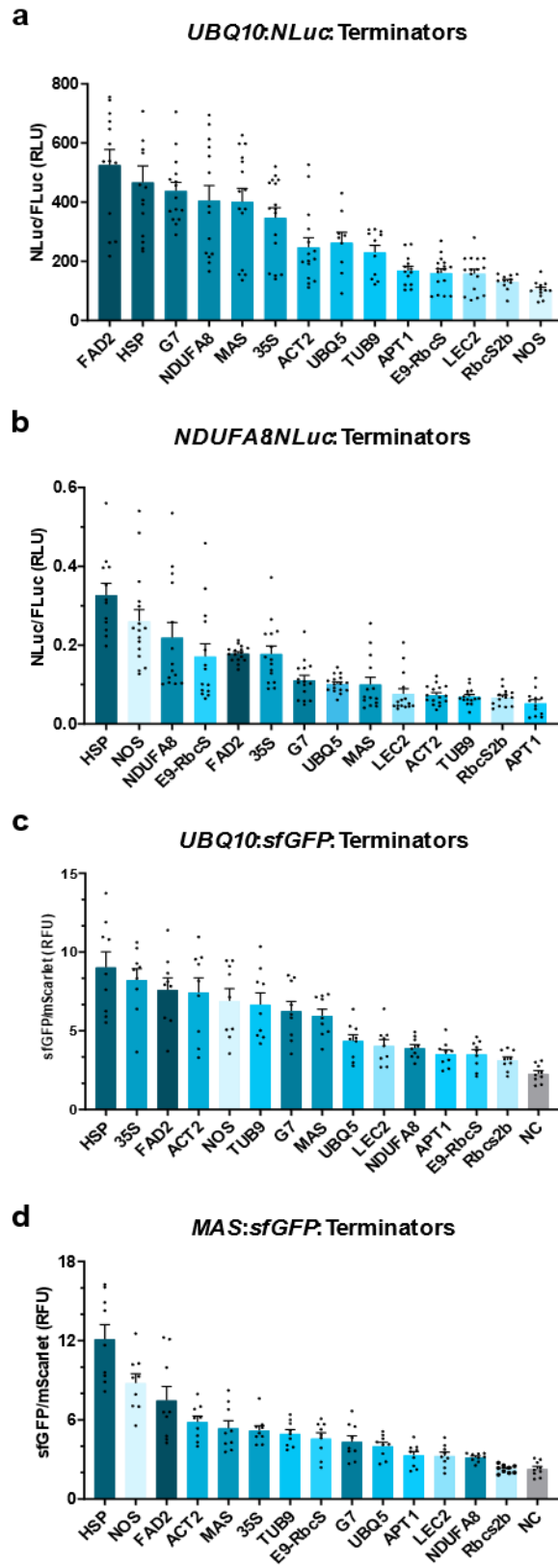


Figure 10. Dual-reporter characterisation of terminators show that terminators affect gene expression levels. Terminators were combined with the promoters in following combinations: **a)** *UBQ10:NLuc*; **b)** *NDUFA8:NLuc* and **c)** *UBQ10:sfGFP*; and **d)** *MAS:sfGFP*. *A. thaliana* leaf mesophyll protoplasts were transformed with PEG, and the expression strength of the terminators was assayed in 96-well plates. [*UBQ10:FLuc:UBQ5*] and [*UBQ10:mScarlet-I:UBQ5*] as normalising construct. NLuc = Nano and Fluc = Firefly luciferase activity, sfGFP= superfolder green fluorescent protein. Bars show normalized expression strength in mean $\pm$ SE. NC = negative control (uninduced cells). Results in panels a and b are obtained by previous lab members¹⁵⁸. For each sample, three biological replicates with three technical replicates are shown on the graph as individual dots.

2.2.6. Terminator functional dissection

To investigate terminator effects on gene expression further, five different terminator variations were generated for two strong terminators (*HSP* and *FAD2*), sequentially deleting Poly-A signals, putative destabilization signals and Musashi binding elements. Statistical analysis of the results showed only a significant difference between *FAD2* Sequence 5, which is the shortest part with no Poly-A site, and the intact *FAD2* as well as *FAD2* Sequence 1 (200 bp, longest terminator from the deletion series (Fig 10a,b). Concerning the rest of the *HSP* series as well the *FAD2* series, there was no significant difference in the reporter expression level in *Arabidopsis* protoplast (Fig. 11a,b).

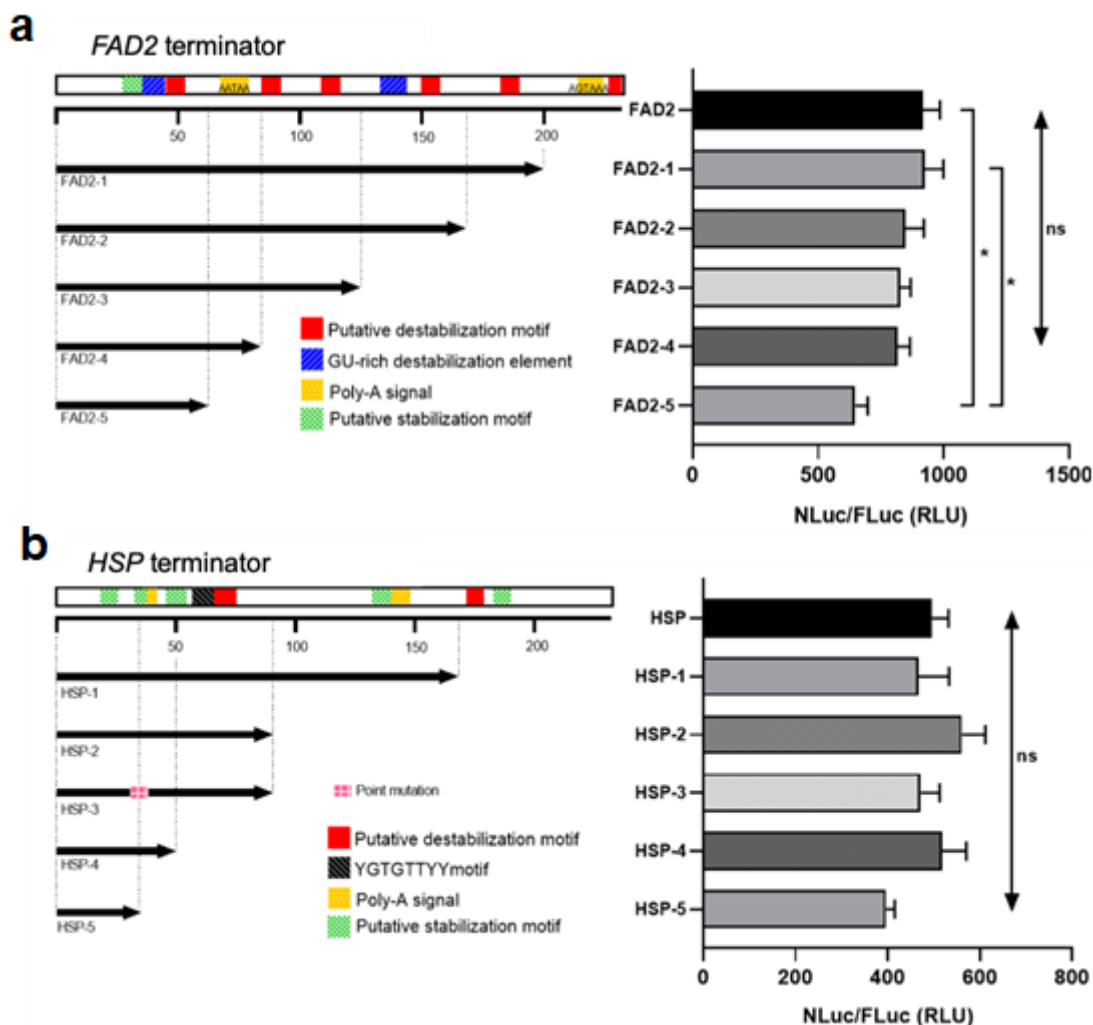


Figure 11. Functional dissection of consistently strong *HSP* and *FAD2* terminators. **a)** Deletion series showing *FAD2* terminator. **b)** Deletion series showing *HSP* terminator. NLuc= Nano and FLuc=Firefly luciferase activity, bar graphs show luciferase activity values in mean $\pm$ SE. $p < 0.05$, one-way ANOVA and Tukey's HSD test. Results in this figure are obtained by previous lab members¹⁵⁸.

2.2.7. Promoter-coding sequence-terminator interactions in gene regulation

Part characterisation data revealed that the gene activation function of a promoter, terminator or coding sequence is not independent or additive, but they have combinatorial effects. A clear example of such promoter-terminator interaction is seen with the *NOS* terminator, which in combination with *UBQ10* resulted in weak (Fig. 10a), *NDUFA8* strong (Fig. 10b), *pOp6* weak (Fig 7a), *lexA* medium (Fig. 9b) and *alcSynth* medium expression (Fig. 9c) in luciferase system. When we switched to a fluorescence-based reporter, then combination of *NOS* terminator with *UBQ10* and *MAS* promoters drove medium and strong expression respectively (Fig. 10c,d). The terminators paired with the

inducible promoters *pOp6-35S* and *lexA-35S* resulted in negligible interaction as they follow a specific pattern (strong/medium/weak), with except *UBQ5*, *APT1* and *NOS*. On the other hand, *AlcSynth-34S* does not follow this pattern. Among the 14 terminators characterised, only four exhibited stable behaviour with both coding sequences: *FAD2* and *HSP* were consistently on the strong side, while *LEC2* and *RbcS2b* tending towards the weak side (Table 2). Nevertheless, promoters remain the dominant factor influencing gene expression levels, as evident in Table 3. In this case, the heatmap is shaded based on the overall expression rather than the relative expression illustrated in Table 2.

Table 2. Heat map showing how combinatorial interactions between promoters, coding sequence and terminators.

	PROMOTERS	LUCIFERASE					FLUORESCENCE	
		<i>pOp6</i>	<i>lexA</i>	<i>alcSynth</i>	<i>NDUFA8</i>	<i>UBQ10</i>	<i>UBQ10</i>	<i>MAS</i>
TERMINATORS	G7	24.3	85.46	187	0.11	444.8	6.2	4.3
	NOS	12.6	64.7	161.8	0.26	103.9	6.9	8.8
	ACT2	17.3	57.4	142.6	0.073	247	7.4	5.8
	APT1	11.8	83.4	132.2	0.052	171.4	3.5	3.3
	35S	25.1	111.6	157.6	0.177	346.9	8.2	5.2
	LEC2	16.2	50	46.7	0.075	158.2	4	3.3
	TUB9	23	106.4	92.2	0.071	235	6.7	4.9
	MAS	18.3	71.7	104.5	0.098	409.3	5.9	5.4
	E9-RbcS	11.6	48.7	79.7	0.188	160.6	3.5	4.6
	RbcS2b	13.3	55.9	56.7	0.066	132.4	3.1	2.3
	HSP	23	79.1	337.9	0.327	464.2	9	12.1
	UBQ5	20.4	37.9	48.1	0.101	239.6	4.4	4
	NDUFA8	17.7	61.1	135.7	0.216	423.7	3.9	3.2
	FAD2	21.5	80.8	223.4	0.178	547.9	7.6	7.5

Expression level of *Nluc* normalised by *Fluc* for the luciferase system and *sfGFP* normalised by *mScarlet-I* for fluorescent system. The colour scale is set independently in each promoter series. Dark blue colour shows the lowest expression, light blue mid-level expression and dark red shows the highest expression.

Table 3. Heat map for reporter gene expression with global ranking.

		LUCIFERASE					FLUORESCENCE	
PROMOTERS		<i>pOp6</i>	<i>lexA</i>	<i>alcSynth</i>	<i>NDUFA8</i>	<i>UBQ10</i>	<i>UBQ10</i>	<i>MAS</i>
TERMINATORS	<i>G7</i>	24.3	85.4 6	187.0	0.110	444.8	6.2	4.3
	<i>NOS</i>	12.6	64.7	161.8	0.260	103.9	6.9	8.8
	<i>ACT2</i>	17.3	57.4	142.6	0.073	247	7.4	5.8
	<i>APT1</i>	11.8	83.4	132.2	0.052	171.4	3.5	3.3
	<i>35S</i>	25.1	111. 6	157.6	0.177	346.9	8.2	5.2
	<i>LEC2</i>	16.2	50.0	46.7	0.075	158.2	4.0	3.3
	<i>TUB9</i>	23.0	106. 4	92.2	0.071	235	6.7	4.9
	<i>MAS</i>	18.3	71.7	104.5	0.098	409.3	5.9	5.4
	<i>E9-RbcS</i>	11.6	48.7	79.7	0.188	160.6	3.5	4.6
	<i>RbcS2b</i>	13.3	55.9	56.7	0.066	132.4	3.1	2.3
	<i>HSP</i>	23.0	79.1	337.9	0.327	464.2	9.0	12.1
	<i>UBQ5</i>	20.4	37.9	48.1	0.101	239.6	4.4	4.0
	<i>NDUFA8</i>	17.7	61.1	135.7	0.216	423.7	3.9	3.2
	<i>FAD2</i>	21.5	80.8	223.4	0.178	547.9	7.6	7.5

Heat map showing how the combination of different promoters and terminators affects the expression level of NLuc 55normalizin by Fluc for the luciferase system and sfGFP 55normalizin by mScarlet-I for fluorescent system. The colour scale is set independently for each reporter. Dark blue colour shows the lowest expression, light blue mid-level expression and dark red shows the highest expression.

2.2.8. Transcriptional and post-transcriptional regulation of reporter expression

To see if the observed changes in gene expression were due to abundance of transcript levels, Qpcr was performed with protoplasts transformed with chosen constructs. Because transforming enough number of protoplasts for RNA extraction is laborious, number of constructs tested was reduced. *HSP* and *FAD2* terminators were chosen for strong expression of both of the reporter proteins. For weak expression *NOS* and *Rbcs2b* terminators were selected for *NLuc*; and *Rbcs2b*, *APT1* and *E9-RbcS* terminators were chosen for *sfGFP*. Because *NOS* and *35S* terminators are widely used in plant systems, protoplasts are transformed with these terminators for both reporters.

For *NLuc* constructs, *FAD2* and *HSP* terminators have significantly higher Mrna compared to *Rbcs2b* terminator; but there was no statistically significant difference between the rest of the terminators (Fig. 12a). Similarly, *sfGFP* constructs, *FAD2* and *HSP* terminators had significantly higher Mrna than rest of the terminators (Fig. 12a). *E9-Rbcs2b* terminator had significantly lower Mrna than the rest of the terminators (Fig. 12a). While *NLuc* displayed a linear correlation between the measured reporter expression levels and Mrna levels (Fig. 12b); *sfGFP* exhibited no such correlation (Fig. 12c). Surprisingly the 35S terminator yielded higher reporter expression in both systems despite Mrna levels (Fig. 12b,c). As transcript abundance did not directly align with reporter protein expression, an investigation into the predicted folding energies of the transcripts was done. However, no linear correlation was seen between predicted transcript folding energies and Mrna levels, or between the folding energies and reporter gene expression (Fig. 12d,e). There may be a logarithmic correlation for luciferase samples in Fig 12e. However, this needs to be validated with a larger data set.

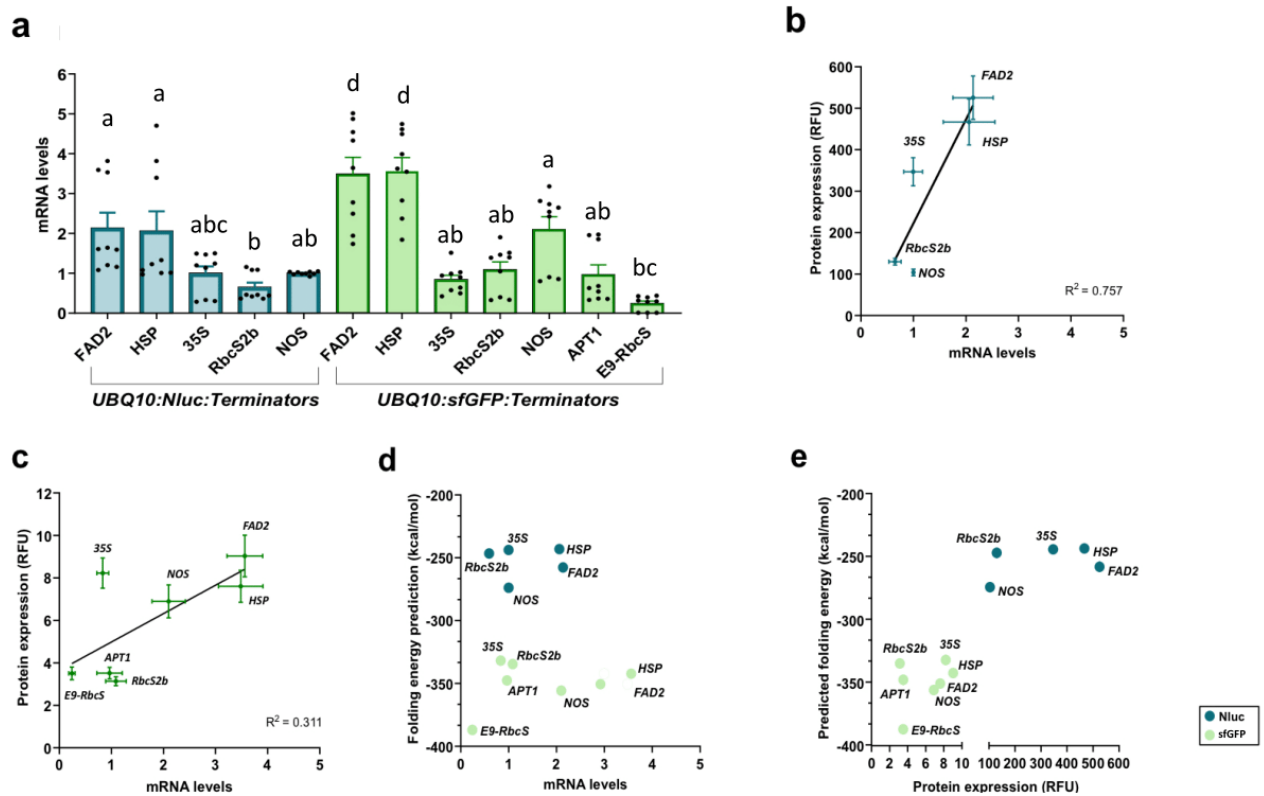


Figure 12. Gene expression analysis for luciferase and fluorescent protein reporters in PEG-transformed leaf mesophyll protoplasts. **a)** Qpcr analysis for *NLuc* and *sfGFP* expression with *UBQ10* promoter and different terminators. *UBQ10:Fluc:UBQ5* is used as the 56normalizing construct for *NLuc* and *UBQ10:mScarlet-I:UBQ5* is used as the 56normalizing construct. Statistical analysis was done using One-way ANOVA and Tukey's HSD. All the data points from three biological replicates are shown on the graph as dots. **b)** Correlation of NLuc expression to Mrna levels for different terminators. **c)** Correlation of sfGFP expression to Mrna levels for different

terminators. Lines represent $\pm$ SE. **d)** Correlation of predicted transcript folding energy to Mrna levels. **e)** Correlation of predicted transcript folding energy to reporter protein expression levels. Blue circles represent *NLuc* constructs, and green circles represent *sfGFP* constructs.

The Qpcr results and folding energy analysis unveiled distinctive post-transcriptional regulatory features. The 35S terminator for both *NLuc* and *sfGFP* has a post-transcriptional enhancer effect; while *NLuc:NOS*, together with *sfGFP:RbcS2b* and *sfGFP:APT1* constructs demonstrated post-transcriptional repression. Investigating known terminator strength factors revealed that the terminator GC content, presence of the canonical AAUAAA poly-A signal and the UGUA motifs might be involved (Fig. 13).

2.2.9. Local drop in GC content around dominant Poly-A site for consistently strong terminators

Consistently strong terminators have a wide range of GC content *HSP* (29.1%), *FAD2* (36.5%). Followed by 35S (32%) on the strong side. And weak terminators with *RbcS2b* (30%), *E9-RbcS* (32.8%), *APT1* (35%), *NOS* (38.7%). However, one feature that is predominant in *HSP*, *FAD2* and 35S terminators is that the GC content is lowest at a 50 bp window, which is located around their dominant Poly-A signal, and then the GC content goes back up (Fig. 13a).

2.2.10. *a priori* identification of novel motifs in consistent strong terminators

Looking further into consistent terminator strength, we search for the UGUA motif that increases strength of the 35S terminator, in *HSP* and *FAD2* terminators. The *HSP* terminator has two UGUA motifs at locations 119 bp and 206 bp, while *FAD2* lacks the motif.

Interestingly, XSTREME motif discovery tool was able to identify 4 motifs that possibly have functional roles. These motifs are present in *HSP* and *FAD2* terminators but are absent in other terminators in the toolkit (**Fig. 13b**). The first motif is CAAUGUUUUGUGUC that is found around 150 bp in both terminators, corresponding to the cleavage site. Other 3 motifs are CUCAUUAUGUUA, UUGUUGUGUUAUGAC and UUUUUCUAAUAUUA are found at similar locations in both of the terminators but around 10-20 bp apart. Only the CUCAUUAUGUUA motif is present in *UBQ5* terminator at 167 bp (**Fig. 13b**).

2.2.11. Factors governing gene expression strength in outlier terminators

After investigating consistently strong terminators, focus was then shifted to the outlier terminators, which are the terminators that perform above or below than expected compared to their mRNA levels. The GC content of outlier terminators are moderate and vary between 30.5%-35.0% considering average terminator GC content of *A. thaliana* (32.5%) (Fig. 13c).

The presence of the canonical AAUAAA poly-A signal tends to enhance terminator strength; and from the chosen terminators, *RbcS2b* is the only outlier terminator that does not contain the canonical Poly-A signal. The 35S (at 9 bp, 126 bp), *NOS* (19 bp, 100 bp) and *APT1* (185 bp) terminators, all have the signal (Fig. 13c)

Along with the Poly-A signal, the presence of the UGUA motif has an enhancer effect. 35S terminator has three repeats of UGUA motif (x3 starting from 96 bp separated by UU). On the other hand, the presence of UUUUU motif's role is context dependent and in *A. thaliana* its role is not well understood. The UUUUU motif is present in *NOS*, *RbcS2b* and *APT1* terminators. All the motifs and their location within the terminators are shown in Fig. 13c. No direct correlation emerged between the base-pairing of the AUG start codon and gene expression for the samples (Fig. 13d). Base-pairing of the AUG start codon reduces gene expression in bacteria¹⁹⁶. However, for our samples there was no correlation. 35S terminator is predicted to have base-pairing at AUG codon for *sfGFP*; and *APT1* terminator again with *sfGFP*. Other constructs were predicted to not have base-pairing at AUG (Fig. 13d).

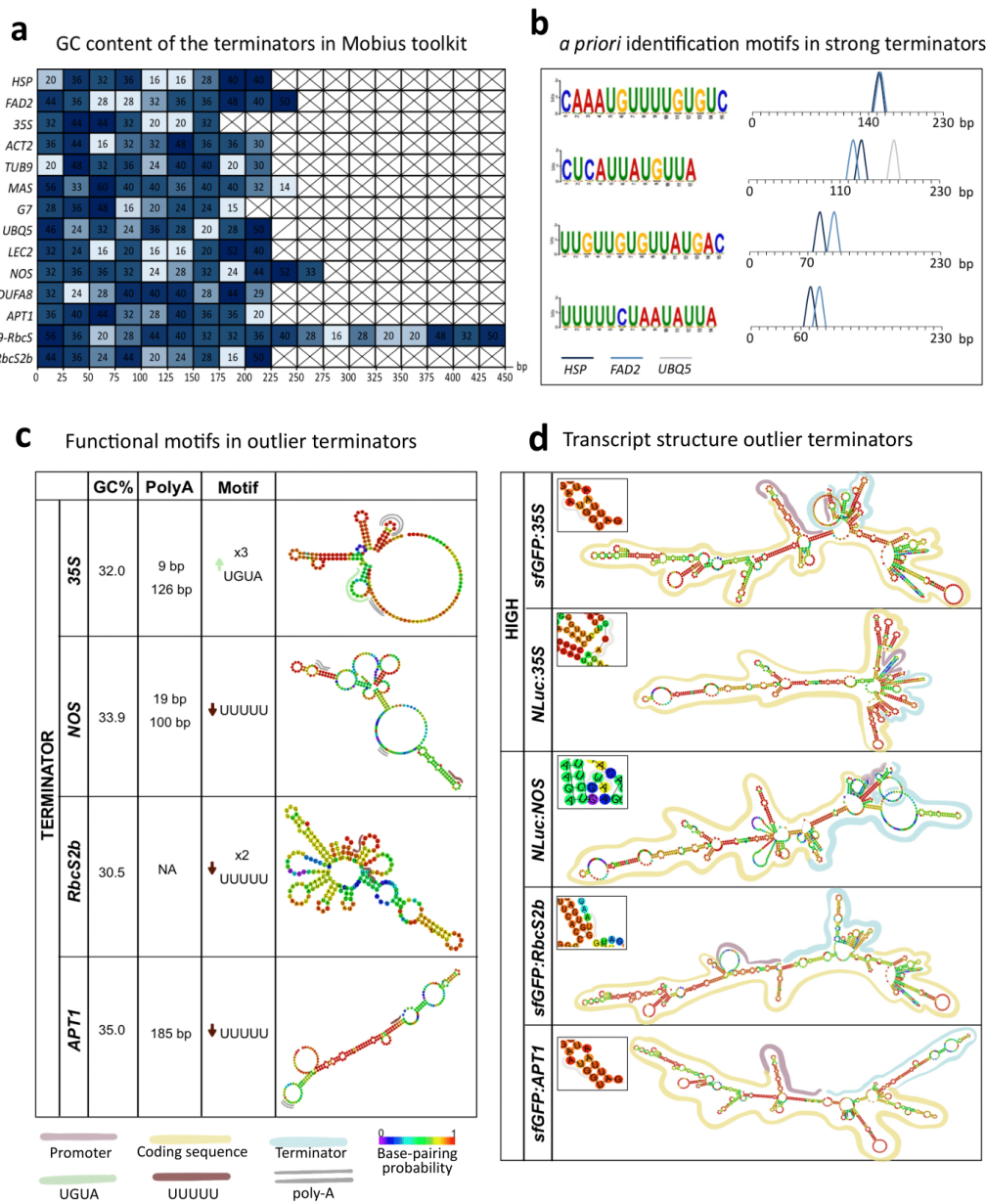


Figure 13. Factors contributing to terminator strength. **A)** GC content of the terminators in the toolkit with 25 bp windows. Light blue means low GC while dark blue means higher GC content. Consistently strong terminators *HSP*, *FAD2* and *35S* (weaker than the first two) show a local GC drop of 50-bp that correlates with their dominant Poly-A site. **b)** *a priori* identification of motifs in strong terminators. XSTREME software identified four motifs that are present in *HSP* and *FAD2* terminators and lack in other terminators in our toolkit. First of the identified motifs, CAAAUGUUUGUGUC at 145 bp and 146 bp for *HSP* and *FAD2*, respectively, possibly correlates with high cleave efficiency. **c)** Functional motifs in outlier terminators. *RbcS2b* lacks the canonical Poly-A signal, whereas *35S*, *NOS* and *APT1* do. The enhancer UGUA motif is present in *35S* terminator, while the UUUUU motif that is context dependent is present in *NOS*, *RbcS2b* and *APT1* terminators. **d)** Transcript structure of outlier terminators. AUG start codon seems to be base-paired for *sfGFP:35S* and *sfGFP:APT1* constructs, but not for others. *UBQ10* promoter from its predicted TSS is used as the ‘promoter’ region.

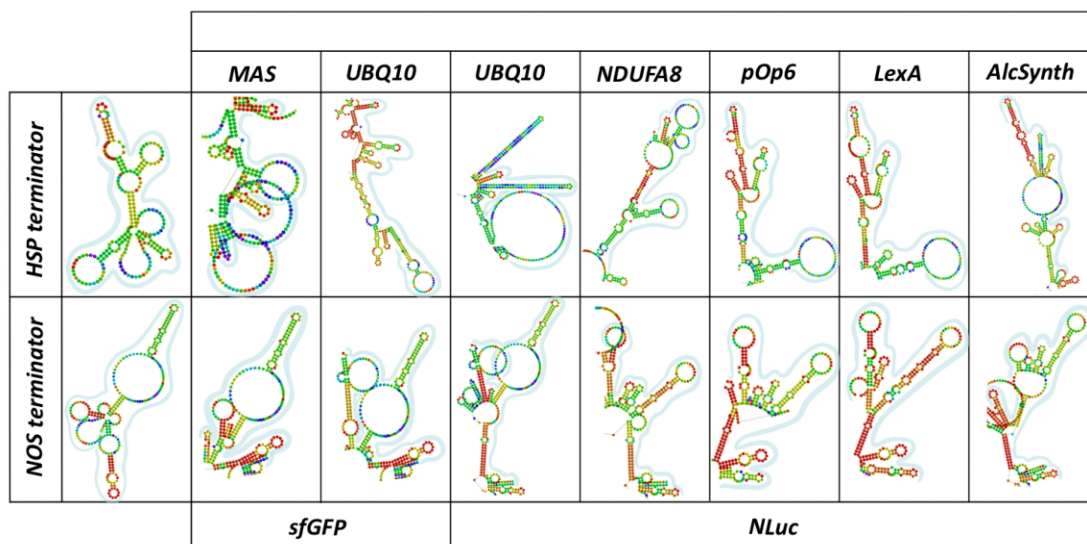
2.2.12. Terminator structure and terminator strength

While the features described partly explained observed reporter expression levels, they fell short in elucidating the transcriptional regulation variations observed in different constructs. For example, the UGUA motif can explain the 35S terminator performing above its transcriptional RNA levels for both constructs. However, how the *NOS* terminator is transcriptionally regulated for *sfGFP* and not for *NLuc*; or how *RbcS2b* is transcriptionally regulated for *NLuc* and not for *sfGFP* is still unknown. Additionally GC content or presence of motifs do not explain how relative strength of a part can change combinatorically.

Following investigations into nucleotide features, a more in-depth exploration of terminator strength was undertaken, focusing on RNA structure analysis that unveiled unique patterns (Fig. 14a). For strong and consistent terminators like *HSP*, in addition to the presence of canonical poly-A signal, we also see a loop bigger than 20 bp (Fig. 14a). Interestingly, when the *NOS* terminator is paired with strong promoters (*UBQ10* and *MAS*), it also has this big loop, which disappears when combined with other promoters (Fig. 14a). 35S terminator also has this loop when combined with *sfGFP*, and it does not with *NLuc* (Fig 12d). However, 35S terminator has a structure with high base pairing probability that seem to be conserved (Fig. 13d).

Expanding on these structural features, terminators from Felippes et al. (2022) that showed increased strength when combined with the first 32 bp of the HSP terminator were used as templates¹⁷². Terminator structure alone and within the transcript, with and without the addition of these 32 bp were investigated. It was revealed that the addition of the 32 bp increases the loop number (> 20 bp) or size for the *NOS* and *ACS2* terminator used in their study; and makes structures with higher base-pairing probability of *Rbcs1A* (Fig. 14b).

a NOS terminator has a conserved domain when paired with strong promoters



b Terminator structure analysis with *HSP'32* addition from Felippes et al., (2022)

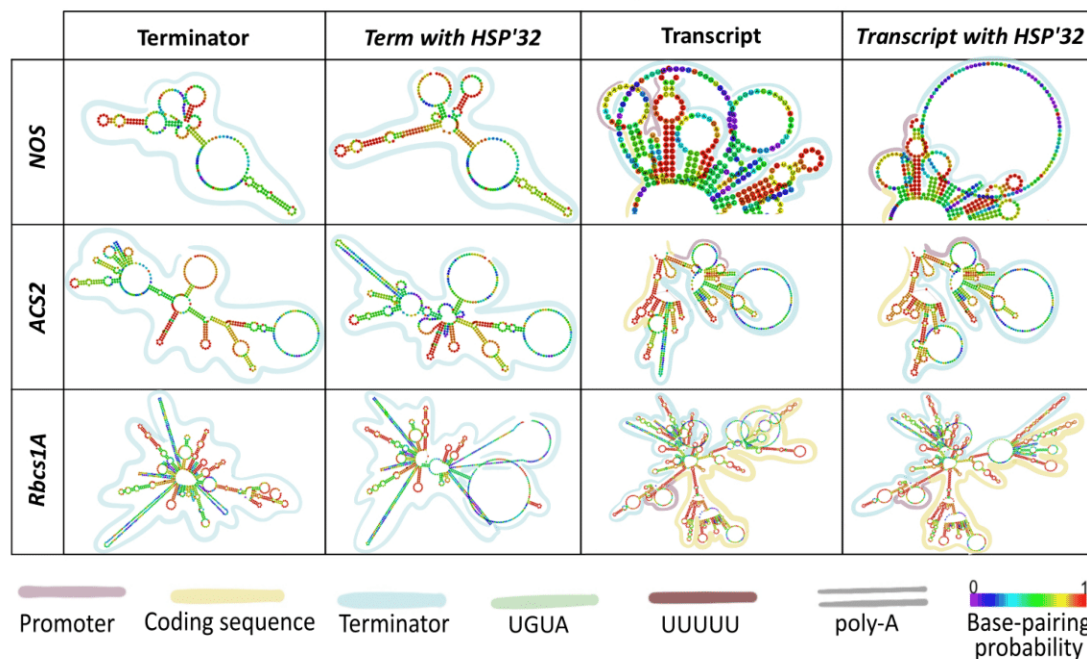


Figure 14. Structural features of terminators. **a)** *NOS* terminator conserved loop domain (> 20 bp) when paired with strong promoters like *UBQ10* and *MAS*. The loop domain changes conformation to stem-loop when it is combined with weaker promoters like *NDUFA8*, *pOp6*, *LexA* and *AlcSynth*. Predicted transcription start site of these promoters are used as promoter templates. **B)** Terminator structure analysis with *HSP'32* addition from Felippes et al., (2022). *NOS*, *ACS2* and *Rbcs1A* were used as terminators combined with *Rbcs1A* promoter and GiFiP reporter. Combining terminators with the initial 32 bp of *HSP* terminator increases their strength. From a structural point of view, addition of *HSP'32* increases the loop size or the number of loops bigger than 20 bp for *NOS* and *ACS2*. For *Rbcs1A* it makes the terminator more structured.

2.2.13. Promoter-terminator interactions and gene expression consistency

After investigating terminator features that might affect gene expression strength, part interactions at RNA level were investigated to explain the combinatorial effect. By using ViennaFold2.0 software, RNA folding predictions were made starting from the TSS of promoters, coding sequence and terminators. Promoter, coding sequence and terminator combinations were chosen among the ones experimental data exists. So, *MAS* and *UBQ10* promoters were chosen for *sfGFP* expression. For *NLuc*, *UBQ10* promoter was chosen for strong expression, *NDUFA8* for weak expression and all three inducible promoters (*pOp6*, *lexA* and *alcSynth*) were included. Combining selected components with terminators of varying strengths, including consistently robust ones (*HSP*, *FAD2*), a frequently employed terminator with slightly fluctuating strong expression (*35S*), the highly variable *NOS* terminator, and a consistently weak terminator (*RbcS2b*), revealed distinct interaction patterns. Specifically, *UBQ10* and *MAS* promoters exhibited no interaction with *HSP* and *RbcS2b* terminators and displayed minimal to medium interactions with the *FAD2* terminator (Fig. 15). This suggests a potential lack of interaction between components demonstrating consistent expression patterns. Notably, when the *NOS* terminator exhibited weakest performance, it demonstrated strong interactions with promoters such as *UBQ10:NLuc* and *pOp6:NLuc* (Fig. 15). Analysis of inducible promoters indicated varying degrees of interactions except *pOp6:FAD2* and *pOp6:RbcS2b* constructs (Fig. 15).

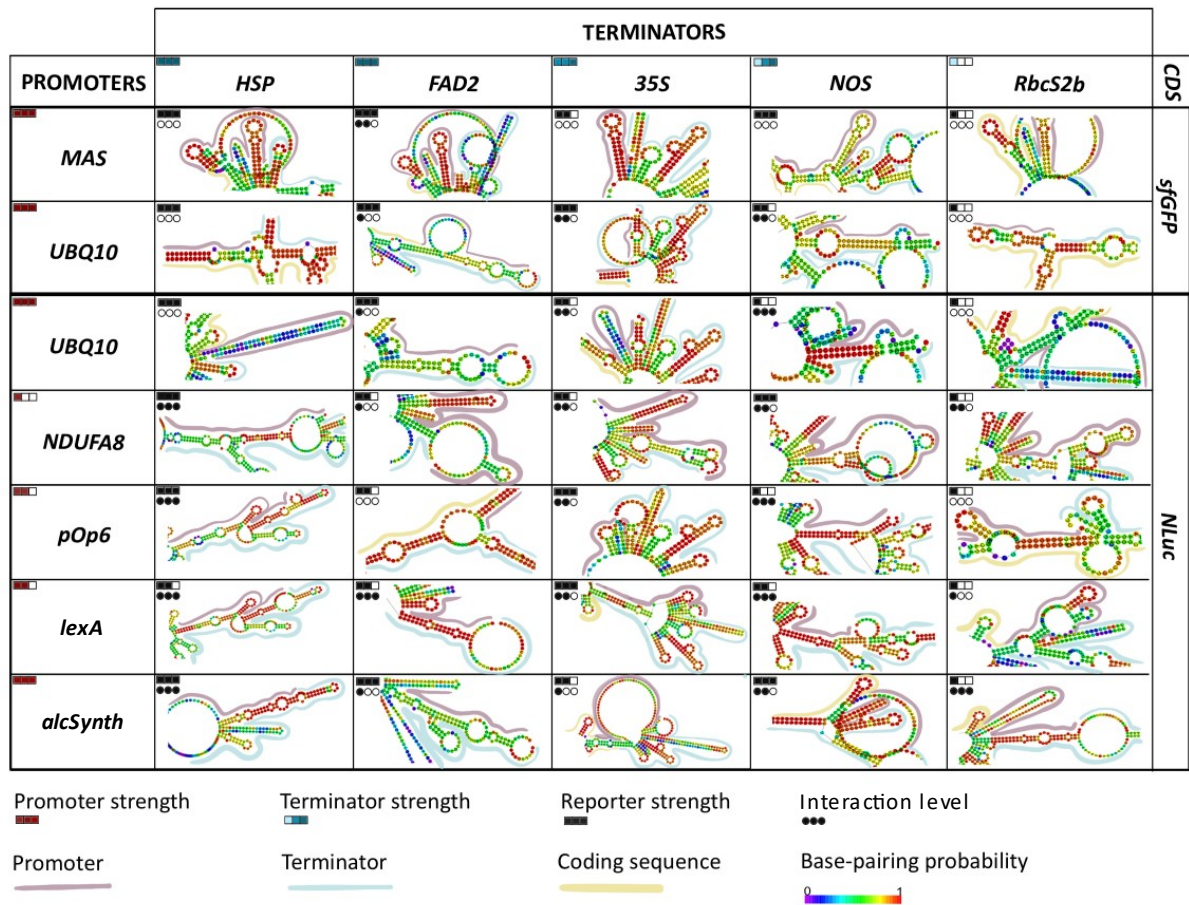


Figure 15. Combinatorial interactions between promoters and terminators at RNA level. *MAS*, *UBQ10*, *NDUFA8*, *pOp6*, *lexA* and *alcSynth* promoters from their predicted transcription start site were combined with the chosen terminators *HSP*, *FAD2*, *35S*, *NOS* and *RbcS2b*. *MAS* and *UBQ10* promoters were combined with *sfGFP*, in addition to NLuc. Terminator strength is represented as blue grid, where dark blue and full grid represent strong, and light blue and empty grid means weak. Full grid with changing colours (such as the one of *NOS*) means terminators expression is changing depending on which parts it is combined with. Promoter and reporter strength are represented similarly, where full grid means strong and empty grid means weak. Interaction levels are categorized as none (empty circles), weak (one circle full), medium (two circles full) and strong (three circles full). For none, there is no base-pairing predicted between promoter and terminator parts. For weak, there is only base-pairing with blue and green colours, which means it is < 0.5 probable. For medium, there is base pairing in yellow and orange colours which means , 0.75 probability; or 5 bp with red base pairing (around 1.0 probability). Strong interactions are defined as 5 or more red base-pairing. Consistently behaving parts do not show interactions with each other like the *UBQ10* and *MAS* promoters with *HSP*, *RbcS2b* and *FAD2* terminator (minimal interactions).

2.3. Discussion

The Mobius Assembly Toolkit introduces a catalogue of short promoters and terminators to meet the expanding need for well-characterised components in plant synthetic biology. Promoter characterisation revealed that *UBQ10* and *MAS* are consistently strongest promoters for luciferase and fluorescence systems in *A. thaliana* protoplasts (Fig. 7).

2.3.1. TSS structure of the *UBQ10* promoter affects gene expression

The *UBQ10* promoter was chosen as a case study and the role of its TSS (also called 5'UTR) structure was investigated further. Among the designed TSS variants, more structured TSS tend to decrease gene expression strength, aligning with expectations based on prior research (Fig. 8)^{184–186}.

An interesting exception was observed with the *NOS* terminator, where the MUT1 variant exhibited a 30% increase in strength (Fig. 8b,c). While this study specifically examined the influence of the isolated TSS structure, it is essential to consider the structure of parts when they form the transcript. For instance, in the case of MUT1 with the *NOS* terminator, the transcript could be adopting a conformation that either reduces the TSS structure or imparts features that enhance the strength of the *nos* terminator.

2.3.2. Terminators affect gene expression in inducible systems

In addition to constitutive promoters, chemically inducible systems are popular tools to control the timing and amplitude of transgene expression. β -estradiol, Dex, and ethanol-inducible systems were developed almost 20 years ago and have been used in many studies since then¹⁹³. However, no work has been reported about enhancing the target gene expression using different terminators. In this study, it has been shown that the gene expression ranges of the inducible promoters can be modulated by exchanging the terminators. It was shown that terminators can change the strength of *alcSynth* 7.2 fold, with more modest differences for the *lexA* and *pOp6* promoters (Fig. 9). *pOp6* and *lexA* are fully synthetic and comprised of a minimal promoter and a tandem repeat of the specific transcription factor binding sites; in contrast, the ethanol-inducible system¹⁹⁷ drives its promoter from the natural promoter found in *Aspergillus nidulans*. It is possible that *alcSynth* promoter is interacting

more with the terminators at the DNA level, although this interaction difference is not seen at the RNA level (Fig. 15). Furthermore, the Dex and Estradiol systems had similar basal expression levels (~10 RLU) whereas the ethanol-inducible system had significantly high basal expression (250.3 RLU) (Fig. 9). A possible explanation for the leakiness of the alcohol system might be TMV Ω ; as the addition of the TMV Ω in the *LhGR-pOp6* system exhibited slightly increased basal expression¹⁹⁴.

2.3.3. Terminators affect gene expression of constitutive promoters

Comprehensive characterisation of the terminators, using the high throughput protoplast transient assay described in Chapter 4 and 6, underscores the strong influence of the terminators on gene expression (Fig 9). It is common practice to rely on promoters to control gene expression level and neglect the influence of terminators. For example, to compensate the loss of expression after the domestication of the synthetic plant promoter *G10-90* promoter, researchers replaced it with the *AtRPL37Ac* promoter¹⁹⁵. However, they were using the *E9-RbcS* terminator, which we found to drive low gene expression; the gene expression could have been restored with a more activating terminator instead. Even though promoters have great influence on gene expression (Table. 2), terminators could alter the gene expression up to 7.3 times for the same promoter (Fig. 9,10)

In a recent study 2.4 fold difference was seen between *NOS* and *HSP18.2* terminators when combined with *CsVMV* promoter (Renilla-Firefly luciferase system, *N. benthamiana* infiltration)¹⁹⁸. In another study by Chamness et al. a statistically significant difference was not observed between 8 different terminators tested (Renilla-Firefly luciferase system, *N. benthamiana* infiltration)⁷⁶. Possible explanations to observing more modest differences in gene expression between terminators include using different DNA part sequences, different chassis, and different dual-luciferase system. Therefore, results obtained with one set-up might not represent another set-up well.

2.3.4. Promoter-coding sequence-terminator interactions in gene regulation

In addition to the role of terminators in gene regulation, which is mainly through processing of mRNAs¹⁶⁶, this study shows that the strength of promoter-terminator pairs also change when the reporter protein coding sequence is changed. This, implies that some of the regulatory motifs or structures are also context dependent (Fig 9, Table 2).

The coding sequence can impact gene expression through factors such as codon usage bias, gene length, GC content, correct 5' cap, and Mrna folding¹⁹⁹. Both *NLuc* (516 bp, GC 52.7%) and *sfGFP* (714 bp, 61.3%) have been codon optimised for *A. thaliana*. Any factors related to the factors listed above would affect the magnitudes of the reporter readout, but the relative strength of the promoter-terminator pairs would remain the same. However, the relative strengths of promoter-terminator pairs change when reporter proteins are changed (Fig. 10 and 11, Table 2), highlighting combinatorial interactions. Only four out of the 14 terminators exhibited stable behaviour with both reporters tested: *FAD2* and *HSP* were always strong while *RbcS2b* and E9-RbcS were always weak.

2.3.5. Factors governing terminator strength

Looking at terminators showing consistently strong behaviour, a local drop in GC content (50 bp) was identified which also contains the Poly-A signal and cleavage site (Fig. 13a). So, it is likely that this low GC region have structural features that aid recognition of transcription termination and cleavage. Additionally, *a priori* motif investigation revealed the CAAAUGUUU motif at location 145 bp for *HSP*; and 146 bp for *FAD2*, which possibly aids with cleavage site recognition and cleavage efficiency contributing to overall terminator strength (Fig 12b).

With our deletion series we have seen that only the shortest *FAD2* terminator (70 bp) has significantly reduced strength compared to the full length *FAD2*, while the 35 bp *HSP* terminator's strength is statistically insignificant compared to the full length *HSP* terminator (Fig. 11). Since Felippes et al., 2023 have also showed that the first 32 bp of the *HSP* terminator is a transferable element that enhances the strength of other terminators, it is likely that the *FAD2* terminator has similar functionality, but between 70-200 bp regions, which corresponds to the location where identified motifs reside (Fig. 13a,b).

2.3.6. Terminator structure and terminator strength

While the nucleotide compositions and presence of motifs can explain why a terminator is consistently strong or weak, they do not explain the combinatorial interactions that were observed. Interactions can be explained at DNA or RNA level, but since it is hard to do *in silico* folding predictions of long strands of double-stranded DNA, focus has been shifted to RNA secondary structures. The *HSP* terminator has a conserved loop structure (>20 bp), while the *NOS* terminator only has a similar

structure when combined with strong promoters (Fig 12a). Likewise, Felippes et al. (2022) terminators have larger loop size and higher base-pairing probability when they are in the stronger expression form, compared to their wild-type forms (Fig. 9b). Whether the loop-feature has functional roles need to be validated further as it could also be the sequences of these terminators are the cause of the structural features. This investigation can be done by introducing point mutations that alter/sustain the structure or by using synthetic terminators with randomized sequences.

2.3.7. Promoter-terminator interactions and gene expression consistency

As the combination of terminators with different promoters has the potential to change terminator structure, it was interesting to study whether this change resulted from physical interactions between the two components (Fig. 15). It is plausible that certain promoters and terminators possess *cis-acting* elements that make them strong and physically restrict their interactions with other parts, while other *cis-acting* elements may require physical interactions with other parts to function effectively. In mammalian systems for instance, *cis-acting* elements on terminators require specific secondary structures to interact between the polyadenylation complex and Mrna²⁰⁰. Similar mechanisms may be at play in plant systems, and the absence of interaction could serve as an enhancing feature for one part while acting as a repressive feature for another.

Taken together these results suggest that the interactions within a TU should be investigated as a whole. Gene expression research should not be limited to specific motifs and domains in promoters or terminators in order to reach true orthogonality and precision in engineering biology. More in-depth work such as direct RNA sequencing should be done to investigate how changing the genetic parts alter Mrna variants in gene regulation without introducing biases during reverse transcription. This would also be a method to distinguish transcriptional regulation from post-transcriptional regulation, as it will allow the detection of less stable transcripts as opposed to Qpcr which can only detect stable transcripts. Hi-C Interaction Frequency Inference (HIFI) algorithms for RNA/chromatin structure assay and *a priori* dissection of the promoter, coding, terminator and UTR sequences in influencing gene activity is required to verify the theory of looping or any other mechanisms on the roles of plant terminator genetic parts on transcriptional or translational regulation.

2.4. Conclusion

Plant synthetic biology lacks characterised parts to develop more predictable biosynthetic pathways and high-throughput methods to perform a systematic genetic part characterisation. There is also a need for characterised genetic elements in different plant chassis, for example, Arabidopsis, tobacco, and barley. Throughout this work, the power of terminators in controlling gene expression has been underlined, which is rather neglected, and the attention primarily focuses on the promoters. It was shown that terminators and the coding sequences act combinatorically with the promoters, and it is incorrect to assume that a promoter or a terminator can always be strong or weak. Further studies need to be conducted to elucidate the mechanisms of these interactions.

CHAPTER 3

Mobius Assembly toolkit improvements: Developing stable vectors

3.1. Introduction

Synthetic biologists are in a continuous arms race with evolution. Synthetic biologists design, build and test constructs and transform their host-of-interest with their constructs. However, since there are finite resources within a cell, the intrinsic gene expression and the construct start competing for the resources. As a result, the host organism tends to mutate the construct in favour of its endogenous genes, altering the construct performance²⁰¹. Synthetic biologists then try to identify and fix these mutations, and the arms race continues. Construct stability, therefore, presents a significant challenge in synthetic biology.

The resource competition between the intrinsic genes and the construct causes an unnatural load on the host gene expression system, termed 'burden'. When plasmids are used to express transgenes up to approximately 50% of total transcripts in *Escherichia coli* can be attributed to plasmid-related genes, thereby imposing transcriptional burden on the host²⁰¹. The magnitude of this burden depends upon several factors, including the strength of basic genetic parts²⁰², plasmid copy number and size²⁰³ and growth media composition²⁰⁴. Quantifying burden in bacteria involves assessing changes in growth rate and fluorescence reporters upon gene induction²⁰¹. Alternatively, burden can be quantified by methods such as quantitative RNAseq and ribosome profiling²⁰⁵. In organisms where measuring growth rate is challenging, dual-colour fluorescent reporters, known as "capacity monitors", can be used. These capacity monitors have a constitutive expression cassette with one fluorescent reporter and a test cassette (basic part library or inducible system) with a different colour fluorescent reporter²⁰⁶. Since the two cassettes rely on the same pool of cellular transcriptional and translational resources, the ratio of the two reporters is used as a proxy for burden. Such systems were used to quantify burden in mammalian cells²⁰⁶ but are yet to be implemented in plant systems. Whole plant-level burden measurements are complicated since different cell types have different metabolic flux distributions that accompany their functions²⁰⁷. Nonetheless, single-cell systems such as protoplasts or cell suspension cultures are promising platforms to quantify the burden in plants.

Influenced mainly by burden, the functionality of a transgene construct is seldom stable within a population and tends to diminish over time as mutations emerge to make the host more fit to survive^{208,209}. When plasmids are used to deliver transgene constructs, burden contributes to plasmid

instability. High burden hinders the cellular machinery's ability to sustain active metabolism for plasmid synthesis. Consequently, the plasmid maintenance systems trigger stress responses, exacerbating instability and generating mutations²¹⁰, such as deletions²¹¹ and point mutations²¹².

3.1.1.

Transposon insertion

Until recently, transposon insertions were not traditionally considered a direct manifestation of cellular burden even though they can contribute to transgene mutations as they are induced by various biotic and biotic stress^{213–216}. Transposons are DNA sequences that can move from one location to another in the genome, hence their nickname 'jumping genes'. Transposons often contain terminal repeat sequences, which transposase enzymes recognize to facilitate their movement²¹⁷. Transposons can not only insert themselves into different locations on the genome and plasmids, but they can also move non-transposon DNA and remove genes by internal deletions²¹⁷.

The changes transposons make can regulate gene expression by direct mutations or small non-coding RNAs²¹⁷. These can have positive or negative impact depending on whether they promote survival fitness or disrupt a vital gene. In addition to positive and negative impacts, most transposition events exhibit a neutral effect²¹⁸. These neutral transposition events are a mechanism to increase genomic diversity in all kingdoms of life²¹⁷, where pre-existing mutations can give a selective advantage under stress (or burden).

Mhiri et al. (2022)²¹⁹ offers an extensive review of plant transposons, while this study focuses on transposon families in *Escherichia coli*, the primary organism where plasmid stability issues may arise. Different family of transposons are found in *E. coli* including insertion sequence 1 (IS1), IS2, IS3, IS4, IS5, IS30 and IS150 families²²⁰. These families have different features for terminal repeats, DNA targets, and insertion mechanisms²²¹. There is still a lot to uncover about the mechanism of transposon action. Still, the first step includes the generation of 3'OH at the transposon termini and cleavage of the flanking 5'DNA, releasing the linear insertion sequence²²² (Fig. 16). The 3'OH end then acts as a nucleophile and either makes double-stranded breaks (simple insertion) or single-stranded nicks (co-integration) as demonstrated in Fig. 1. Even though it was thought that each transposase can operate by one mechanism or the other, IS903 transposase (from the IS5 family) is capable of both, albeit with co-integration occurring at a lower frequency than simple insertion.²²² Some transposases

use circular intermediates prior to the release of linear insertion sequence like some members of IS1²²³ and IS3 families²²⁴ (Fig. 16).

Although there are overlaps between transposon inducers and cellular burden, these have not been studied together. In order to make stable genetic constructs and cellular strains, it is important to investigate these two factors together. Otherwise, newly developed constructs potentially exhibit instability.

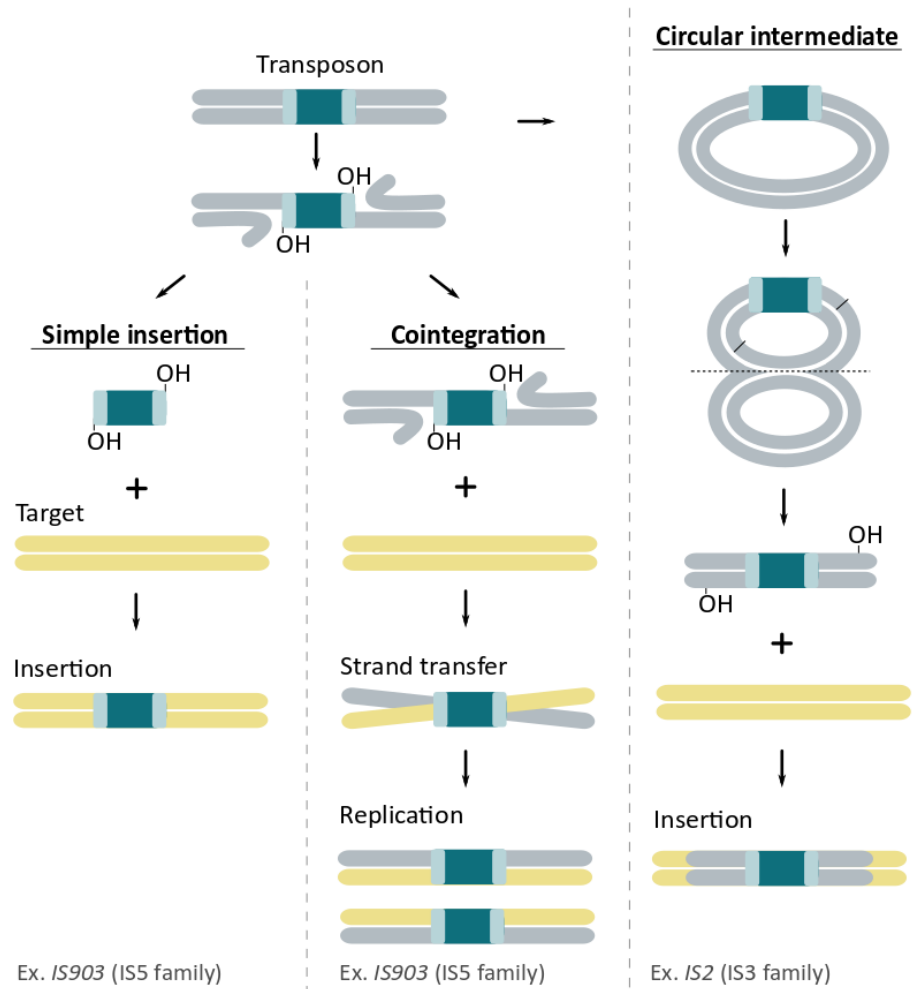


Figure 16. Example of transposition mechanisms observed in *E. coli*. The first step is facilitated by adding OH at the 3' end of transposons. Subsequently, there are two possible scenarios: either the 5' ends of the transposon undergo cleavage, resulting in simple insertion, or strand transfer and replication occur, leading to cointegration before insertion²²². *IS903*, belonging to the IS5 family, can switch between simple and co-integration mechanisms²²². Certain transposons, such as *IS2* from the IS3 family, utilize circular intermediates. During *IS2* transposition, a figure-eight structure forms on the strands containing the transposon, which is then cleaved at the 3' end through the generation of 3'OH²²⁴. The linear transposon subsequently proceeds with the insertion.

3.1.2.

Plasmid dimerisation

In addition to transposon insertion, another area of research that is more commonly associated with burden is plasmid dimerisation. Dimers form during plasmid replication when odd numbers of recombination exchanges occur, and two copies of the plasmid get fused in a head-to-tail confirmation²²⁵. Although dimers are the most common form of multimerization, a higher number of dimerisation can also be found, such as tetramers and hexamers²²⁶.

A dimer has two replication inhibitor systems. This system controls the copy number and does not control the number of plasmid molecules in a cell. Having two sets of replication inhibitors results in a halved copy number for dimers. However, a dimer also has two active ori and which ori will be the next for replication is chosen randomly²²⁷. Considering a cell with 18 monomers and 1 dimer, the probability that the next ori that will be replicated is a monomer is 1/20 while, a dimer will be replicated is 2/20. This means dimers will replicate twice as fast compared to monomers. So, the proportion of dimers within a cell will increase every generation, and eventually, dimer-only cells can arise in a culture, which is known as the 'Dimer Catastrophe'²²⁷.

If dimers are not resolved back to monomers, the dimer plasmids can fail to correctly segregate during cell division, causing plasmid instability and loss. For plasmids with a low copy number, active partitioning makes sure that each cell will inherit the plasmid²²⁸. On the other hand, high copy number plasmids appear to be randomly inherited, as no evidence of active partitioning is seen²²⁷. For multi-copy number plasmids, the probability that a cell will be plasmid-free after division is dependent on plasmid copy number (n)²²⁷:

$$p_{\text{plasmid-free}} = 2^{(1-n)} \quad (\text{Eq.1})$$

For example, assuming that colE1 ori has approximately 18 copies/cell (n=18, usually around 15-25)²²⁹. In a monomeric culture the probability of having a plasmid-free cell is 7.6×10^{-6} by the equation given above. However, for a dimer it is 3.9×10^{-3} . This means the probability of having a plasmid-free cell for colE1 ori is 500 times more likely for a dimer than a monomer. This change in probability is even more dramatic for pUC ori, which has approximately 61 copies/cell (n=61, around 55-65 copies and chosen as 61 for simplicity)²²⁹. Getting a plasmid-free cell probability is 8.7×10^{-19} for the monomeric form of pUC and 1.9×10^{-6} for a dimeric pUC plasmid. This is exemplified in plasmids with high burden (such as large and high copy number plasmids or toxic protein producing plasmids),

since the cells with the plasmid will grow slowly while cells without will grow rapidly²²⁷. As a result, plasmid-free cells can dominate the culture.

Although plasmid loss may represent a problem for colE1 ori, for higher copy number oris such as pUC, plasmid loss due to dimerisation is deemed unlikely as the number 1.9×10^{-6} is still significantly smaller than 3.9×10^{-3} . However, predicting plasmid-loss is not easy enough to be calculated by a single-order equation, and there are other factors contributing to their maintenance. One such factor is burden, and high copy number plasmids exert a larger burden on the cell, that's why they tend to be lost easily once selective conditions change²³⁰

Additionally, higher copy number plasmids have an inherent risk of dimerisation; and dimers increase metabolic burden compared to monomers mainly attributed to increased rate of replication and use of resources for plasmid replication^{231,232}. This increased burden of plasmid dimers generally manifests itself in growth defects, causing cells containing dimers to grow slower than their monomer counterparts²³².

Regardless of the plasmid copy number, features of dimers such as increased risk of plasmid loss, reduced copy number and growth retardation are industrially relevant problems. This means a bioproduction host containing dimers has increased burden, which is likely to create mutations, and you have an increased risk of plasmid loss, reducing the productivity²³³.

In contrast to the detrimental outcomes associated with plasmid dimers, cells harbouring these dimers can gain a competitive edge under specific selective conditions, such as exposure to antibiotics or environmental stressors²³⁴. Considering that environmental stress, including exposure to antibiotics, constitutes a form of cellular burden, it raises the possibility that other types of burden could similarly induce dimerisation. Investigating the casual relationship between burden and dimerisation will improve our understanding of plasmid stability and likely provide insights into the underlying reasons for dimer emergence.

3.1.3. *cer* domain for dimer resolution

The presence of plasmids increases burden and as outlined in the preceding section they have instability issues such as dimerisation. Dimerisation appears to exert a more adverse impact on the host organism than its positive effects. Nonetheless, plasmids serve as essential communication tools in bacteria, aiding in their swift evolution and adaptation²³⁵. This raises the question of whether there

are dimerisation control mechanisms present in nature. Indeed, several dimer resolution sites have identified in plasmids including *cer* (ColE1)²³¹, *psi* (pSC101)²³⁶ and *mrw* (pJHCMW1)²³⁷

The *cer* site is a non-coding, ~240 bp region that signals site-specific recombination (Fig. 17)²³¹. It has a 30 bp core region where recombination is catalyzed by tyrosine recombinases XerC/D, and rest of the sequence is for accessory protein binding for arginine-dependent DNA binding protein (ArgR, also called XerA) and aminopeptidase A (PepA, also called XerB)²³⁸. ArgR and PepA control the direction of recombination so that dimers can only be converted to monomers. They arrange two directly repeated *cer* sequences in an antiparallel position. This conformation brings XerC and XerD together, where they bind to 30 bp *cer* synapse region (Fig. 17). Then XerC and XerD interact with PepA and initiate recombination, where XerC catalyzes the strand exchange. Additionally, *cer* site encodes a 70 nt RNA (Rcd) that is only transcribed during dimer formation by the P_{cer} promoter²³⁸. Rcd binds to the enzyme tryptophanase and increases indole production within the cell. This induces a quiescent state where cell division and chromosomal replication arrest, but the cell can still stay metabolically active. This pausing is thought to be part of dimer formation checkpoint, allowing the *cer* system to resolve the dimer²³⁸.

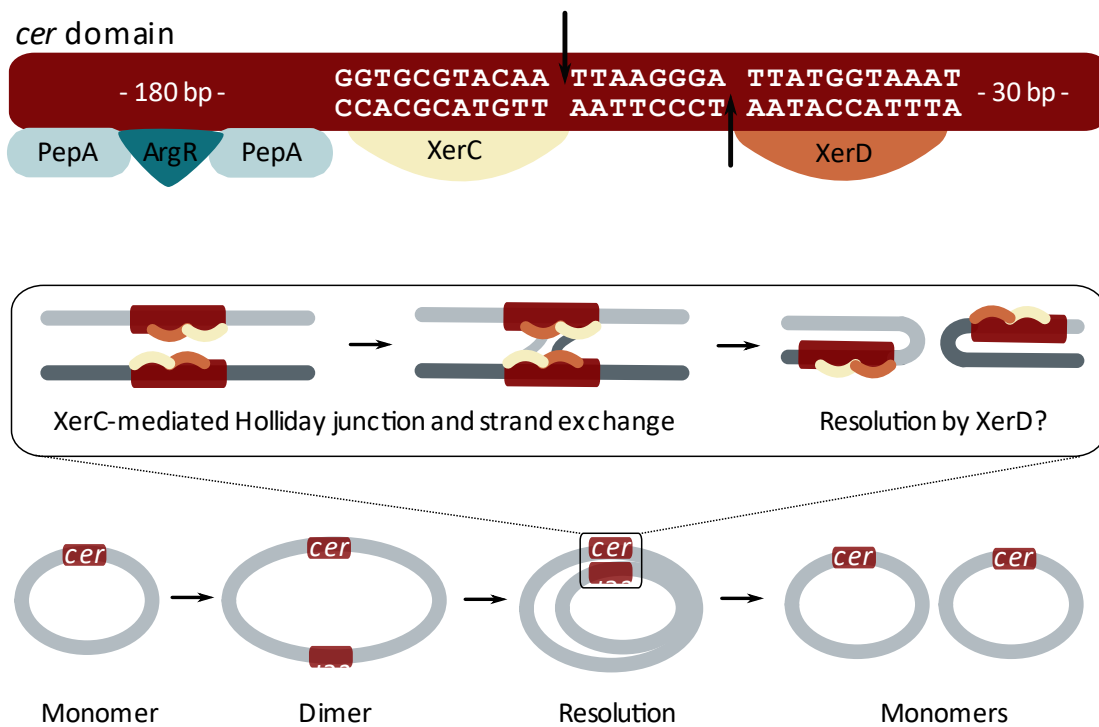


Figure 17. Dimer resolution by the *cer* system in *E. coli*. The *cer* domain is a ~240 bp region where 180 bp at the 5' end is accessory region where PepA and ArgR binds. Recombination specifically happens at the 30 bp synapse

region, where XerC binds to the left and XerD binds to the right side of the 8 bp recombination site. When a dimer forms ArgR and PepA proteins control the directionality of the and arrange the two *cer* sites in an antiparallel manner. This conformation brings XerC and XerD together, which then initiate recombination. XerC mediates Holliday junction formation and strand exchange. How Holliday junction is resolved is not clear, but XerD is hypothesized to be involved²³⁸.

The presence of the *cer* site in plasmids stabilises copy number, and, therefore, improves their stability and productivity in bioproduction^{233,239}. A case study based on citramalate production in *E. coli* and red fluorescent protein (RFP) reporter as an indicator of citramalate productivity demonstrated that after 20 h of continuous culture, the RFP signal begins to decline and reaches 10% of its initial value by 55 h²³³. The introduction of the *cer* site allows for a steady-state RFP production (and consequently citramalate) for 58 hours, resulting in a total citramalate yield increase of more than 50%²³³. Another investigation, spanning 120 h, revealed that even in the absence of antibiotics, the *cer* site promotes segregational stability. This, in turn, elevates transketolase production to 137 mg/L, a significant improvement compared to the 77 mg/L observed without the *cer* site²³⁹.

Plasmid, and thus construct, stability issues including transposon insertion and dimerisation can emerge due to burden, or their presence can escalate the burden. Studying plasmid stability is important because developing constructs and circuits for engineering biological systems loses its purpose if these circuits undergo mutations over time, failing to perform as anticipated. Construct stability issues become even more significant in plant systems where a heterologous construct typically needs to be stable in three chassis: *E. coli* as propagation host, *Agrobacterium tumefaciens* for gene transfer and the target plant species.

In this chapter, I outline my efforts to stabilise Mobius Assembly vector performance for plant systems. During my PhD, I encountered issues such as growth retardation of cells with pMAPS plasmids both on plates and in liquid culture, as well as inconsistent cloning efficiency. Moreover, we received inquiries from other users of the Mobius Assembly kit who encountered similar issues, particularly related to cell growth and chromoprotein expression. Consequently, I conducted comprehensive whole-plasmid sequencing across different Mobius vectors and bacterial generations, revealing that the problems were associated with transposon insertion and plasmid dimerisation. I have found that *spisPINK* chromoprotein triggers transposon insertion (*insB* and *IS2*). Even though these transposons belong to different insertion sequence families, they were both inserted within the *J23106* promoter used to express *spisPINK*. On the other hand, plasmid dimerisation was resolved by

the addition of the *cer* domain in the backbone. The findings of this study, coupled with investigation of similar phenomena in *A. tumefaciens* and in plant systems, will contribute to stability efforts in synthetic biology.

3.2. Results

3.2.1. Transposon insertion in *spisPINK* expressing vectors

While using the Mobius Assembly for Plant Systems (MAPS) Toolkit (Fig. 18a), I encountered issues such as the retarded growth of vectors or plasmids on plates, following storage at -80°C. Colonies only became visible after a 36-hour incubation at 37°C. Despite obtaining colonies, the growth of liquid cultures was also slow. Colony size and negative selection marker expression was heterogenous (Fig. 18b). Additionally, cloning efficiency varied depending on which colony was used to isolate the plasmid. The presence of heterogeneous colonies on the plate and the inconsistent cloning efficiency indicated that a mixture of normal and mutated cells in the cell stocks. Notably, these issues were absent with the Level 0 vector, were most pronounced with the Level 1 vector, and the Level 2 vector exhibited similar problems to a lesser extent.

In terms of MAPS vector architecture, Level 0 vectors are simpler than Level 1 and 2. Level 0 only has one ori for replication in *E. coli* (pUC) which makes it small (2.9 kb). In contrast, Level 1 and 2 vectors feature a fused ori (pUC) to WKS1 for replication in both *E. coli* and *A. tumefaciens*, making their total size 4.3 kb. Since I did not experience any issues with Level 0 vectors, which is expected considering their simple architecture, I focused my vector stability improvement on Level 1 and Level 2 vectors.

To identify the underlying problems, two colonies from each vector stock (Level 1 A-D; Level 2 A-D) were sent to whole-plasmid sequencing. For Level 1 vectors that showed severe growth defects, two different types of transposon insertion were identified using BLAST (Fig. 18c). The first one is 777 bp long *insB* insertion, which belongs to the IS1 family of transposons. The second one is 1331 bp long *IS2*, which belongs to the IS3-family transposons. Despite the transposases belonging to two different families, both were inserted at a similar location—within the middle of the J23106 promoter. During the *insB* insertion, the final 9 bp of the J23106 promoter is duplicated (AGTGCTAGC). Similarly, during

IS2 insertion, 5 bp in the middle region of the promoter is duplicated (CTAGC). Upon comparing the identified transposon sequences with their wild-type counterparts, it was noted that *insB* exhibited 8 point mutations, whereas *IS2* displayed 5 point mutations (Fig. 18d). As transposons typically employ inverted repeat sequences to facilitate their insertion, the TnCentral online tool was utilized to identify these repeats. The analysis revealed that *insB* utilizes a 23 bp inverted repeat, and *IS2* utilizes a 41 bp inverted repeat flanking themselves (Fig. 18d).

Identifying *IS1* and *IS3* family transposons was interesting, as *IS5* is the family that is previously associated with transposon insertion in widely used pGreenII and pCambia vectors^{240,241}. pGreenII has been observed to impose burden on *E. coli* and caused improper cell division and cell death, which triggered transposon insertion to the plasmids and caused instability issues²⁴⁰. Watson et al., 2016 used this burden as a selective pressure and after rounds of selection they have identified a transposon insertion that restored the growth defects. Sequencing of these cells has revealed that they have acquired *IS5* family transposase, a 1.2 kbp transposable element in the region between the *ColE1* origin of replication (*ori*) and the stop codon of neomycin resistance gene²⁴⁰. The *IS5* insertion was also recorded for pCambia vectors between the *ColE1* *ori* and kanamycin resistance gene for vectors distributed by Arabidopsis Biological Resource Center²⁴¹. These could be triggered by large size of the pCambia plasmids (~11 kb), plasmid-specific factor or another environmental factor causing cellular burden. The insertions recorded for pCambia do not cause any growth defects in *E. coli*; and they are outside of the left and right borders, so they do not get inserted to plant genome²⁴¹.

Different from the pGreenII and pCambia vectors where transposon insertions disrupt the *ori* and the antibiotic resistance genes, in pMAPS vectors transposons were seen in the promoter of the marker gene. Thus, I hypothesized that transposon insertion might be triggered by *spisPINK* vectors exerting an increased burden on their host. As a preliminary preventative measure, the *ori* was changed from pUC to *ColE1*, given that *ColE1* has a significantly lower copy number per cell. Indeed, it was shown that cells expressing *spisPINK* from the pUC vector had growth retardation (Fig. 18e).

Following *spisPINK* production exerting increased burden on the cells, the reporter protein was changed to mScarlet-I. Comparing mScarlet-I expression from pUC and *ColE1* showed that the growth rate both vectors were similar and there was no obvious growth retardation associated with increased plasmid copy number (Fig. 18f). Since I mainly use MAPS vectors for protoplast transformation, which necessitates a high amount of DNA (4 µg purified plasmid per 75 µl

transformation), I opted to express *mScarlet-I* from the high copy number ori for the following experiments.

3.2.2. ***cer* domain addition improves growth rate of cells with MAPS vectors**

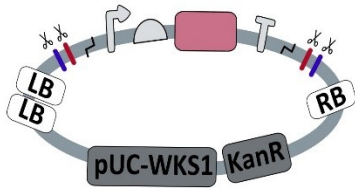
Even though it was shown that *spisPINK* triggered transposon insertion to the J23106 promoter, this only explains the growth retardation observed in Level 1 vectors. Level 2 vectors also showed growth retardation, however their cause was different. Whole-plasmid sequencing (Nanopore) has revealed that Level 2 vectors multimerize, and most commonly become dimers, even though trimers are sometimes observed (Fig. 19a). To resolve this issue, the plasmid made in the previous section (*mScarlet-I*:pUC) was used. When 6 colonies were randomly selected and grown, one colony showed growth retardation (Fig. 19c). This colony carrying dimerised plasmids was confirmed with whole-plasmid sequencing. This indicated that that probably the high copy number ori (pUC) was triggering dimerisation independent of selection marker since dimerisation was seen with both *mScarlet-I* and *sfGFP* reporters.

In order to maintain the high copy number ori for obtaining a high yield of purified plasmids essential for protoplast transformation, strategies to prevent dimer formation were explored. A 241 bp long *cer* domain, previously identified for resolving dimers in *E. coli*²³¹, was fused to the pUC-WKS1 ori, resulting in the creation of the new generation of MAPS vectors (Fig. 19b). The addition of the *cer* domain improved growth rate both for cells grown in liquid culture directly from the -80°C stocks (Fig. 19d) and from colonies (Fig. 19e). The impact was particularly notable for samples originating from the -80°C stocks. Although plasmids lacking the *cer* domain achieved higher stationary phase density when grown from colonies, cells without the *cer* domain exhibited greater heterogeneity and a higher standard deviation (Fig. 19e). The *cer* domain's effect was similarly confirmed for Level 2 plasmids expressing *sfGFP* (Fig. 19f-h). The addition of the *cer* domain resulted in a significant increase in growth rate for cells derived from -80°C stock (Fig. 19f) and from colonies (Fig. 19g).

The *cer* domain is not expected to be active in *A. tumefaciens*, suggesting that its addition should not have a noticeable impact. However, the increased size of the vectors might impose an additional burden on their host. To confirm that the addition of the *cer* domain does not adversely affect *A. tumefaciens* growth, an investigation was done for *A. tumefaciens* without a plasmid, MAPS plasmids, and MAPS plasmids with the *cer* domain (Fig. 19h). The results revealed no significant difference in growth rates, confirming that there is no increased burden associated with the increased vector size in *A. tumefaciens* (Fig. 19h).

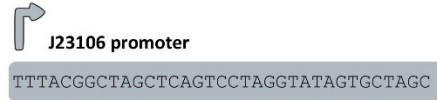
a

MAPS plasmid map



b

Transposon insertion to spisPINK plasmids



c

Colonies with growth defects



insB (IS1 family)
(777 bp)

TTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGC *insB* AGTGCTAGC

IS2 (IS3 family)
(1331 bp)

TTTACGGCTAGC *IS2* CTAGCTCAGTCCTAGGTATAGTGCTAGC

d

Sequences inserted to spisPINK plasmids

insAB

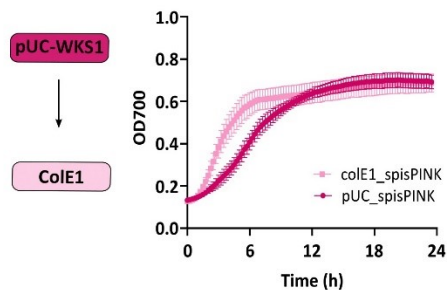
AGTGCTAGCGGTAATGACTCCAATTATTGATAGTGTATTTATGTTTCAGATAATGCCCGATGACTTTGTCATGCAGCTCCACCGATTTTGAGAACGACAGCGA
CTTCCGTCGCCCGGTGCCAGGTGCTGCCTCAGATTCAGGTTATGCCGCTCAATTGCTGCGTATATCGCTTGCTGATTACGTGCAGCTTTCCCTTCAGGGG
GGATTATACAGCGGCCAGCCATCCGTCATCATATCACCAGCTCAAAGGGTGACAGCAGGCTCATAAGACGCCCCAGCGTCGCCATAGTGCGTTACCGGAA
TACGTGCGCAACAACCGTCTTCCGGAGACTGTATACGCGTAAACAGCCAGCGCTGGCGCGATTAGCCCCGACATAGCCCCACTGTTGCTCCATTTCGCG
GCAGACGATGACGTCACCTGCCCGCTGTATGCGCGAGGTTACCGACTGCGGCCTGAGTTTTTAAAGTGAAGTAAATCGTGTGAGGCCAACGCCATAATG
CGGGCTGTTGCGCCGCGCATCCAACGCCATTATCGGCCATATCAATGATTTTCTGGTGCGTACCGGGTTGAGAAGCGGTGTAAGTGAACGTCAGTTGCCATGTT
TTACGGCAGTGAGAGCAGATAGCGCTGATGTCGGCGGTGCTTTTCCGCTTACGCACCCCGTCAGTAGCTGAACAGGAGGACAGCTGATAAAACAG
AAGCCACTGGAGCACCTCAAAAACACCATCATACATAAATCAGTAAGTTGGCAGCATCACC

IS2

TAGACTGGCCCCCTGAATCTCCAGACAACCAATATCACTTAATAAAGTGATAGTCTTAATACTAGTTTTTAGACTAGTCATTGGAGAACAGATGATTGATGT
CTTAGGGCCGGAGAAACGACAGCGGCTACCCACAGGAAAAGATCGCAATTGTTTCAGCAGAGCTTTGAACCGGGGATGACGGTCTCCCTCGTTCGCCCGCA
ACATGGTGTAGCAGCCAGCTGATTTCTTGGCGTAAGCAATACAGGAAGGAAGTCTTACTGCTGTCGCCCGCGGAGAACAGGTTGTTCTCGCTCTGA
ACTTGCTGCCGCCATGAAGCAGATTAAAGAACTCCAGCGCTGCTCGGCAAGAAACGATGGAAAATGAATCCTCAAGAAGCCGTTGAATATGGACGGGC
AAAAAAGTGATAGCGCACGCGCCCTTATTGCCCGGGATGGGAGTAAGCTTAGTCAGCGGTTGCTCCGGGTGTCGCGTGCAGCTTGCACGTCATTCTC
AGACGAACCGATGACTGGATGGATGGCCCGCGCAGTCGTCACTGATGATACGGATGTGCTTCTCCGTATACACCATGTTATCGGAGAGCTGCCAAGCTAT
GGTTATCGTGGGTATGGGCGCTGCTTCGACAGCAGGCAACTTGATGGTATGCTGCGATCAATGCCAAACGTTTACCGGATCATGCGCCAGAAATGCG
CTGTTGCTTGAGCGCAAAACCTGCTGTACCGCCATCGAAACGGGCACATACAGGCAGAGTGGCGGTGAAAGAAAGCAATCAGCGATGGTGTCTGACGGGTTT
GAGTTCTGCTGTGATAACGGAGAGAGACTGCGTGTACGTTGCGCTGGACTGCTGTGATCGTGAGGCTGCACCTGGGCGGTCACTACCGGCGGCTTCAACA
GTGAAACAGTACAGGACGTCATGCTGGGAGCGGTGGAACGCGCTTCGGCAACGATCTTCCGTGCTCCAGTGGAGTGGCTGACGGATAATGGTTTCATGCT
ACCGGGCTAATGAAACACGCGAGTTCGCCCGGATGTTGGGACTTGAACCGGAAGAACACGCGGTGCGGAGTCCGGAGAGTAACGGAATAGCAGAGAAGCTTCG
TGAAACGATAAAGCGTGACTACATCAGTATCATGCCAAACAGACGGGTTAACGGCAGCAAGAACCTTGCAGAGGCGTTTCGAGCATTATAACGAATGGC
ATCCGCATAGTGCCTGGGTTATGCTCGCCACGGGAATATCTCGGCGAGCGGCTTGAATGGGTTAAGTGATAACAGATGCTGGGAAATATAGGGGCAAA
TCCACTAGC

e

Changing origin of replication (ori)



f

Changing reporter protein and ori

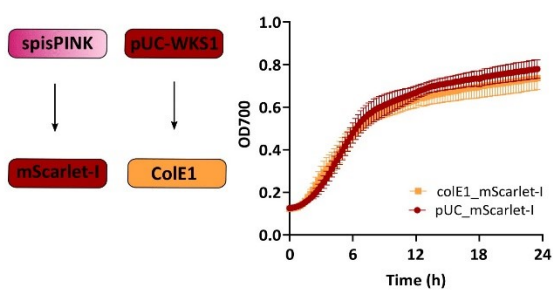


Figure 18. Improving pMAPS vectors by changing the ori and the negative selection marker. **a)** Level 1 MAPS plasmid map. There are two left borders (LB) and one right border (RB) T-DNA insertion sequences flanking *spisPINK* expression cassette as a negative selection marker. pUC-WKS1 is the fused ori and KanR is the kanamycin resistance cassette. Red lines show BsaI restriction enzyme recognition sequences and blue lines show AarI restriction enzyme recognition sequences. **b)** Colonies streaked directly from glycerol stock that show varying colours of negative selection markers *spisPINK* (Level 1) that grow slowly. **c)** Detected transposon insertion for Level 1 vectors that carry *spisPINK* marker. Two types of transposons were identified *insB* (IS1 family) and *IS2* (IS 3 family) within the J23106 promoter region that was used to express the negative selection marker *spisPINK*. Transposon insertions were highlighted in red and duplicated promoter regions are highlighted in pink. **d)** Sequences of identified transposons. Top panel *insB* and bottom panel *IS2*. Mutations identified from wild-type sequences are shown in red. Bolded text indicated inverted repeat sequences identified by TnCentral tool. **e)** Changing the ori from pUC (high copy number) to ColE1 (medium copy number) improves growth retardation of Level 1 plasmids, highlighting increased burden of associated with *spisPINK* production. **f)** Changing the negative selection marker from *spisPINK* to mScarlet-I. No growth retardation was seen with pUC plasmids expressing mScarlet-I compared to ColE1.

3.2.3. Plasmid dimerisation is prevented by the addition of the *cer* domain

Even though growth rate is improved by the addition of *cer* domain, this does not necessarily mean that plasmid stability is improved, and dimerisation is prevented. To investigate stability and dimerisation over a long period of time, cells with Auxiliary 2 plasmid (only pUC ori, size 2.3 kb) referred to as negative control, pMAPS and pMAPS with *cer* domain were cultured for 13 days. The growth curve and RFP expression per cell were measured on Day 1 and Day 7. Whole-plasmid sequencing was done on Day 1, 7 and 13 to correlate observed phenotypes with the possible mutations on the plasmids (Fig. 20a).

The addition of the *cer* domain initially resulted in an improved growth rate compared to the negative control and pMAPS on Day 1 (Fig. 20a). However, by Day 7, this effect diminished, and the negative control reached a higher stationary phase density with similar growth rates, as depicted in Fig. 20c. When the RFP signal per OD was calculated from plate reader data on Day 1 (Fig. 20d), plasmids with the *cer* domain exhibited slightly higher RFP production. By Day 7, pMAPS without the *cer* domain were expressing higher RFP than pMAPS with *cer* (Fig. 20e). Notably, the RFP strength was lower for both samples on Day 7 (Fig. 20d,e).

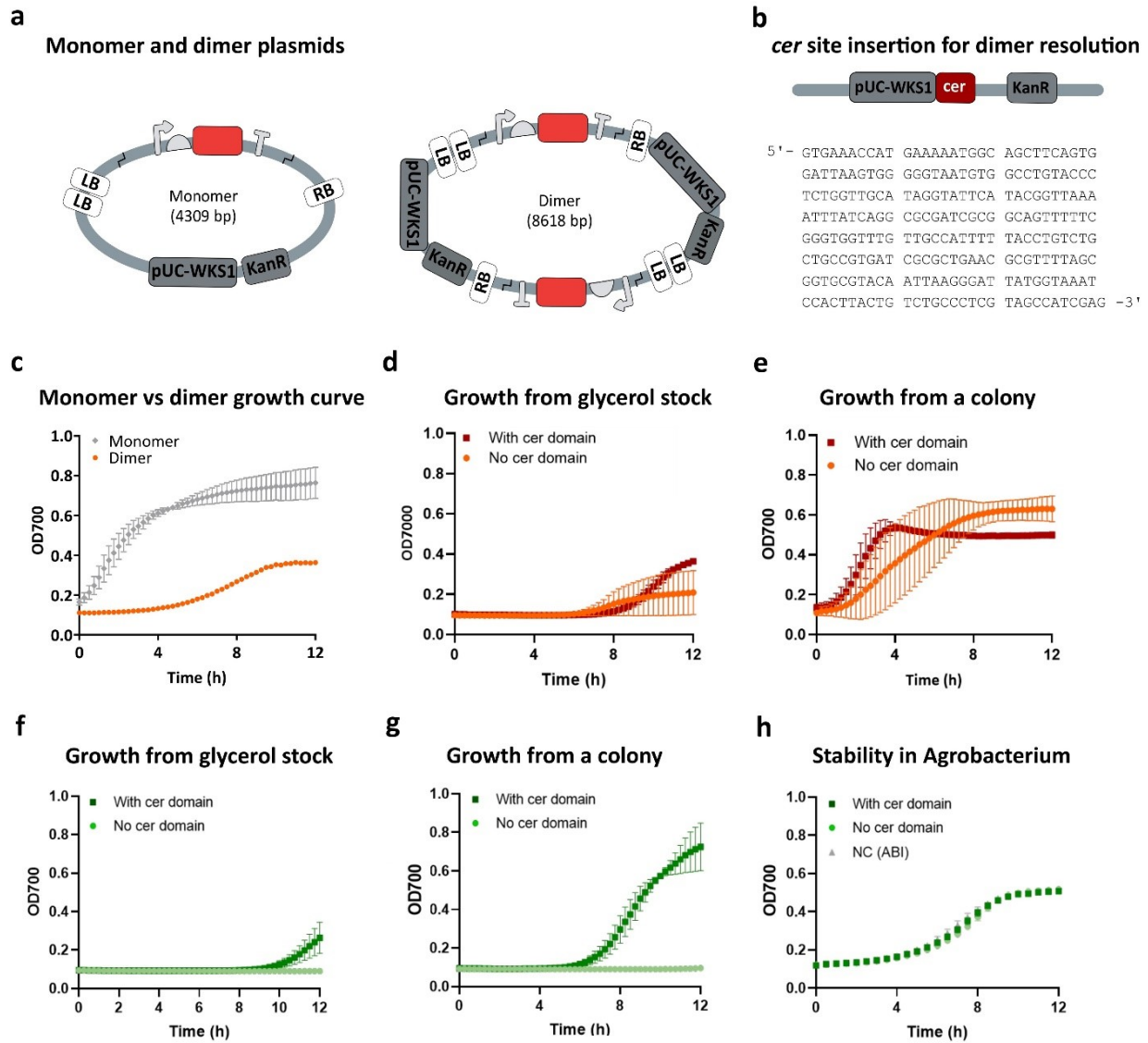
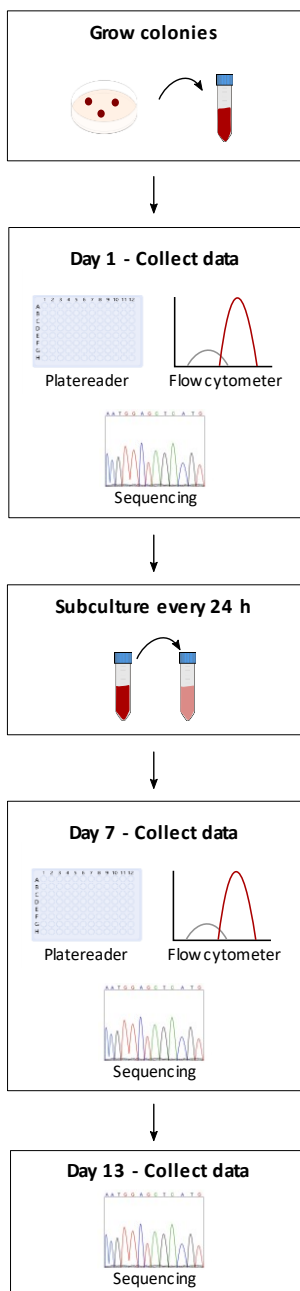


Figure 19. Preventing dimerisation of pMAPS vectors. **a)** Representation of monomeric and dimeric forms of pMAPS vectors. **b)** Sequence of 241 bp *cer* domain for dimer resolution and its location in pMAPS vectors. **c)** An example of cells with dimeric plasmid showing growth retardation. **d-g)** *cer* domain insertion improves growth kinetics of cells carrying pMAPS plasmids in *E. coli*. Cells with Level 1 *cer* domain plasmids grow faster than cells without *cer* domain plasmids when cells are grown from glycerol stock **d)** or **e)** colonies. Similar growth improvement was seen with Level 2 pMAPS plasmids that have *cer* domain when cells are grown from glycerol stock **f)** or **g)** colonies. **h)** Growth kinetics was also tested in *A. tumefaciens* ABI strain with Level 2 pMAPS plasmids, and no growth retardation was seen with or without *cer* domain.

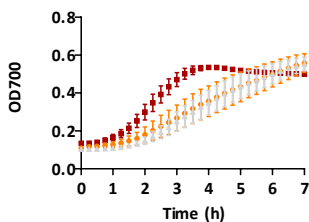
RFP production per OD is an indication of how much each cell produces RFP, though it may not be the most precise technique. To obtain accurate information on RFP production per cell, a flow cytometer was used. On Day 1, due to unexpectedly high fluorescence, the RFP signal approached the axis limit at the current voltage setting (200 V), making it challenging to compare pMAPS with and without *cer* (Fig. 20f). By Day 7, a decrease in fluorescence strength was observed in both plasmids with and without *cer*, but no cells with loss of fluorescence were observed (Fig. 20g).

Sequencing results revealed no mutations on Day 1 and Day 7; however, mutations were observed in samples taken on Day 13. While the negative control showed no mutations, plasmid dimers were observed in pMAPS plasmids. Plasmids with the *cer* domain exhibited a duplication of 200 bp at the 5' end of the pUC domain (Fig. 20h). This is likely an artifact of the *cer* dimer resolution system.

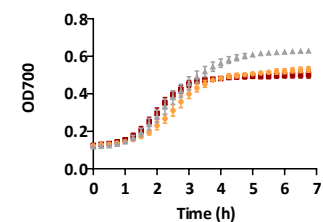
a Plasmid stability experiment set-up



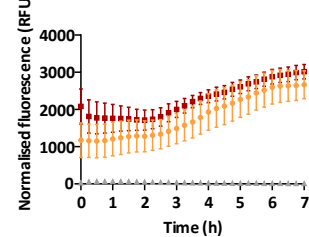
b Day 1 Growth curve



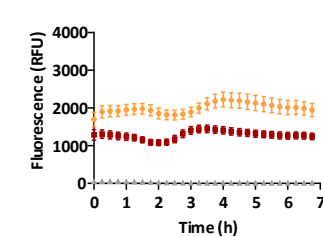
c Day 7 Growth curve



d RFP expression per OD

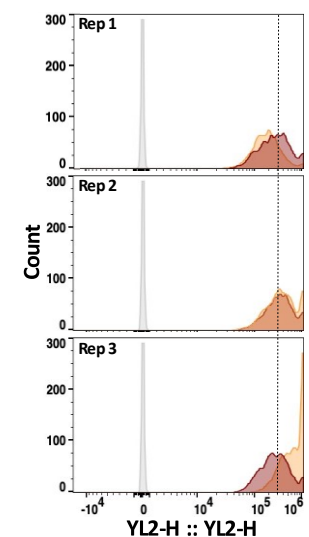


e RFP expression per OD



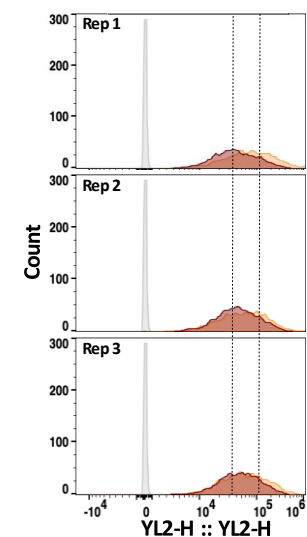
■ With cer domain ● No cer domain ▲ NC (empty vector)

f Day 1 RFP expression per cell



Sample	Count		
	Rep 1	Rep 2	Rep 3
With cer	3004	2646	3165
No cer	2680	3124	3172
NC	2206	2302	2675

g Day 7 RFP expression per cell



Sample	Count		
	Rep 1	Rep 2	Rep 3
With cer	2264	2807	2570
No cer	2422	2721	2669
NC	2784	3367	2677

h Sequencing results at Day 13

Mutation	Sample	NC-1	NC-2	NC-3	With cer-1	With cer-2	With cer-3	No cer-1	No cer-2	No cer-3
	Day 1	None	None	None	None	None	None	None	None	None
	Day 7	None	None	None	None	None	None	None	None	None
	Day 13	None	None	None	None	200 bp ins	200 bp ins	None	Dimer	Dimer

Figure 20. Stability of pMAPS plasmid with cer domain over time. **a)** Illustration of the experimental set-up. Colonies are grown directly from glycerol stock and saturated liquid culture has started. Then cells were diluted 1:100 and used as starter cultures for growth curve analysis in plate reader, RFP expression in flow cytometer and miniprep for whole-plasmid sequencing at Day 1 and Day 13. Cells are subcultured every day at the same time for 13 days, and then sent for whole-plasmid sequencing to identify any mutations that have occurred. **b)** Growth curve of cells on Day 1 and **c)** Day 7. **d)** RFP expression per OD700 at Day 1 and **e)** Day 7. RFP expression per cell as measured by flow cytometry at **f)** Day 1 and **g)** Day 7. A decrease in RFP expression per OD700 and per cell was seen between Day 1 and Day 7 for both plasmids with cer domain and plasmids without cer domain. No mutations were detected on Day 1 and Day 7. **h)** Sequencing results on Day 1, Day 7 and Day 13. On Day 13 plasmid dimers were identified for pMAPS without cer domain while partial pUC ori (around 200 bp) duplication was seen for plasmids with cer domain, effectively preventing plasmid *dimerisation*. Plasmid used for stability experiments with pMAPS are Level 1 with mScarlet-I as negative selection marker, with cer domain (dark red), without cer domain (orange) and negative control (empty Level 2 auxiliary plasmid).

3.3. Discussion

In synthetic biology, the intricate relationship between cell burden and plasmid stability plays an important role in construct functioning. Increased burden frequently leads to plasmid instability, as the host endeavors to alleviate the metabolic load associated with heterologous gene expression. While using MAPS vectors, issues such as transposon insertion and dimerisation have been observed, serving as manifestations of plasmid instability problems that may be triggered by an increased burden on the host.

3.3.1. Transposon insertion in spisPINK expressing vectors

While cells with both Level 1 and Level 2 vectors induced growth retardation in their *E. coli* hosts, transposon insertions were exclusively observed in Level 1 vectors. Since both levels use the same promoter (JM23106), ribosome-binding-site (B0034), and double terminator (rrnB T1 and T7Te), the key distinction between the two sets lies in their choice of negative selection markers: spisPINK chromogenic protein for Level 1 and sfGFP for Level 2. The improvement in growth retardation upon switching to a lower copy number ori (ColE1) strongly suggests that spisPINK chromoprotein imposes an increased burden on its host (Fig. 18e). Collectively, these findings lead to the suspicion that the trigger for transposon insertion in Level 1 vectors is associated with the toxicity of spisPINK, which originates from the coral *Stylophora pistillata*²⁴², chromoprotein to *E. coli*, a phenomenon commonly observed in chromoproteins and frequently attributed to protein aggregation²⁴².

The identified transposon insertions belonged to different IS families: IS1 family of insB and IS3 family for IS2. While the insertion mechanism of IS2 is well-characterised (Fig. 16), insB's insertion mechanism remains less understood. However, with both transposons, a duplication of the insertion site was observed: 9 bp duplication for insB insertion and 5 bp duplication for IS2 (Fig. 18c,c). Although belonging to different families, both transposons were inserted within the J23106 promoter. Recent findings in *Drosophila melanogaster* suggest that transposons exhibit a preference for the promoter region of genes during insertion²⁴³. This supports the notion that in bacteria, transposon insertion within the promoter region can be one of the gene regulatory mechanisms, particularly when a gene producing a toxic protein is involved.

Alternatively, the J23106 promoter might inherently attract transposons. The absence of transposon insertion with mScarlet-I expression, despite burden, suggests that the J23106 promoter may function as a hotspot under specific conditions. Further validation with a broader screening of colonies is essential to identify additional transposon families and their insertion locations.

Although the observed transposon insertions had a negative impact on cell growth and plasmid amounts, they did not seem to affect cloning efficiency since the insertions did not disrupt the restriction enzyme recognition site, overhangs, or left and right borders for *A. tumefaciens* gene insertion (Fig. 19b,c). However, their potential contribution to further instability issues over time cannot be ruled out.

It's worth noting that transposons should not always be considered as adversaries in synthetic biology; in fact, they can be harnessed to create novel tools, as seen in integrative and conjugative elements (ICE) for gene transfer to undomesticated bacteria²⁴⁴ and integrases for recording cell differentiation in plant development²⁴⁵. The continual pursuit of developing novel tools is a key aspect of synthetic biology. Studying transposons, potentially insB and IS2 observed in this study, holds the potential to inspire new tools, as demonstrated by the examples of ICE and integrase recorders.

3.3.2. Plasmid dimerisation and the addition of the *cer* domain

While transposon insertion was absent in Level 2 vectors, they faced another issue: plasmid dimers. Once these dimers form, they can rapidly dominate the population, a phenomenon referred to as 'Dimer Catastrophe'²²⁷. Dimers increase the burden on the host, as evident by the growth retardation (Fig. 20b). This burden may contribute to potential instability issues, making dimers undesirable.

Despite the risk associated with high copy number plasmids forming dimers, I opted to retain the pUC ori in the pMAPS backbone as it gives high plasmid yields, after ensuring there was no growth retardation compared to the ColE1 ori (Fig. 19e).

The next step involved preventing dimer formation, for which the *cer* dimer resolution system was employed. The 241 bp long *cer* domain was fused to the pUC-WKS1 ori at the 5' end, resulting in improved growth for both Level 1 and Level 2 vectors carrying mScarlet-I and sfGFP, respectively (Fig. 20). While the exact proportion of dimers is yet unknown at this timepoint, the increase in growth rate with the *cer* domain aligns with previous observations²³³.

In addition to enhanced growth rates, plasmids containing the *cer* domain were stable over 7 days of continuous subculturing, a process analogous to repeated batch fermentation in bioproduction. Interestingly, on Day 7, pMAPS without *cer* was also stable (Fig 20). However, by Day 13, plasmid dimers were observed for pMAPS without *cer*, and a 200 bp section of the pUC ori (~45% of the sequence) was duplicated at the 5' end for pMAPS with *cer* (Fig 20). This suggests that plasmid dimerisation was triggered, or dimers became predominant, between Day 7 and Day 13.

It would be interesting to conduct daily growth curve and flow cytometry measurements during Day 7 and Day 13 to detect when dimers become predominant and whether that could be associated with phenotypic changes. Furthermore, for a meaningful comparison of RFP production per cell between pMAPS with and without the *cer* domain, flow cytometry needs to be repeated on Day 1, given that the signal is currently too close to the axis limit (Fig. 20f). On that note, it is essential to emphasize that the inclusion of the *cer* domain was not intended to boost the RFP signal, as it is already easily visible to the naked eye; instead, its primary purpose was to enhance stability. Flow cytometry measurements on Day 13 could be incorporated to gather information on the proportion of cells that might have lost RFP expression, offering a more accurate indication of plasmid stability. Additionally, to address the observed 200 bp duplication in pMAPS with the *cer* domain, which is likely nonfunctional, the option of relocating the *cer* domain to a different location on the backbone is recommended.

Another common stability issue arises when constructing glycerol stocks with the assumption that they will remain viable indefinitely. Initially, the early versions of pMAPS did not exhibit any growth defects and thrived from glycerol stocks without issue. However, approximately 2 to 3 years

later, these problems became apparent. Although the solutions presented in this study were tested approximately 1 year after construct creation, their long-term stability should also be tested.

Although I utilised pMAPS vectors for plasmid stability experiments, this issue is commonly encountered by researchers using different backbones such as pGreenII and pCambia vectors^{240,241}. Additionally, plasmid instability issues are not limited to plant researchers. Those working with organisms like yeast and mammalian cells also rely on *E. coli* as a propagation host. Therefore, the stability framework employed in this study could prove beneficial for others in identifying transposon insertions and dimerisations.

It would also be interesting to test different strains of *E. coli* and their susceptibility to plasmid instability would be interesting. Different strains have different genetic backgrounds, metabolic activity and growth rates, participating in different plasmid maintenance systems²⁴⁶. Fast growing strains, such as DH5 α and Turbo were showing to be unstable compared to MG1655²⁴⁷. JM109 strain (recA⁻, endA⁻) was used in this study, and the results may vary in different strains. IS-free *E. coli* strains such as MDS42²⁴⁸ and BLK16²⁴⁹ can be used if transposon insertion is a particular concern, while how they cope with dimerisation is not investigated. In the future, it is anticipated that fast-growing, stable strains will be developed, despite these characteristics currently appearing to be mutually exclusive. Understanding of burden, transposon, and dimerisation phenomena at a deeper level will contribute to the development of the novel strains, backbones and stabilising elements.

3.4. Conclusion

This study aimed to address the plasmid stability issues observed in *E. coli* carrying pMAPS vectors, namely transposon insertion and plasmid dimerisation. The expression of spisPINK was found to trigger *insB* and *IS2* insertion into the J23106 promoter, likely to repress spisPINK production. Substituting spisPINK with mScarlet-I successfully resolved transposon insertion. To prevent dimer formation in the pMAPS vectors, the *cer* domain was incorporated into the pMAPS vectors. pMAPS:mScarlet vectors, with and without the *cer* domain, demonstrated stability over seven days of repeated culturing, with no significant changes in growth rates observed on Day 7. However, adding the *cer* domain substantially improved growth rates when cells were grown from -80°C stocks. Overall, this study contributes insights into the interplay between burden and plasmid stability, contributing to the predictability in synthetic biology.

CHAPTER 4

***Arabidopsis thaliana* fluorescent marker lines for protoplast regeneration**

4.1. Introduction

Emerging molecular tools, including synthetic DNA parts, high-capacity DNA cloning toolkits, logic gates, and inducible systems, are now available for engineering plant development²⁵⁰. Yet, given the complex molecular and hormonal dynamics in plant development, an iterative design-build-test cycle is likely necessary. Expanding the selection of basic DNA parts, understanding part interactions, and developing stable vectors collectively narrow the divide between the ‘prototype’ and the final construct. This is bringing us closer to engineering plant development.

Regeneration stands out as a domain with a broad range of impact in plant development. It is at the basis of agricultural practices such as tissue culture¹⁰⁷, grafting⁹², and transgenics²⁵¹. However, the term “regeneration” includes a variety of definitions and mechanisms. For effective engineering, it’s crucial to delineate the specific mode of regeneration and the target genes involved. Here I will be focusing on protoplast regeneration, and I define successful culture progression as regeneration. Protoplasts are of particular interest as they can be transformed relatively easy with PEG²⁵². If protoplasts can be regenerated into whole plants after transformation, it opens the possibility of creating new transgenic lines for species that are hard to transform, such as wheat¹²⁸. Protoplast regeneration also stands out from a developmental biology point of view since it is a model for dedifferentiation. This provides a promising avenue for discovering the triggers of totipotency. These insights may apply to animal systems, as the regeneration mechanisms in plants and animals often have analogous components due to convergent evolution²⁵³. Additionally, if protoplast regeneration can be controlled, then cells can be guided to specific cell types, which can be used as chassis for secondary metabolite production.

To study protoplast regeneration, I’ve identified my target transcription factors (TF) to be *WOUND-INDUCED DEDIFFERENTIATION1 (WIND1)*, *WUSCHEL RELATED HOMEBOX 5 (WOX5)*, *INDOLEACETIC ACID- WUSCHEL RELATED HOMEBOX 13 (WOX13)*, and *INDOLE ACETICACID INDUCED PROTEIN 8 (IAA8)*. This selection was based mainly on their deregulated expression patterns during protoplast regeneration, making them stage-specific markers. While *WIND1*, *WOX5*, and, more recently, *WOX13* are recognized regulators of regeneration, *IAA8* remains relatively unexplored. Our understanding of these selected targets primarily stems from studies in tissue culture and leaf wounding, with limited insights into protoplast regeneration due to tracking single cells to the

multicellular stage being challenging. Acknowledging that the same target may exert opposing effects depending on the regeneration mechanism is crucial^{118,119}. This underscores the significance of investigating transcription factors with established roles in tissue culture alongside novel candidates to uncover their potential roles in protoplast regeneration.

4.1.1. *WOUND-INDUCED DEDIFFERENTIATION1 (WIND1)*

WIND1 belongs to APETALA2/Ethylene Responsive Factor (AP2/ERF) family which regulate development and plant metabolism in response to biotic and abiotic stress²⁵⁴. Among the AP2/ERF family, *WIND1* is one of the first TFs rapidly induced after wounding and one of the master regulators of regeneration¹⁴⁷. The cells that express *WIND1* can undergo dedifferentiation, leading to callus formation. From this callus, new tissues and organs can be regenerated, establishing *WIND1* as a factor conferring pluripotency and regeneration competence¹⁴⁷.

Even though *WIND1* expression is a wound response, overexpression of *WIND1* can induce callus formation on leaves, roots, and hypocotyls from the epidermal cell layer, even without wounding¹⁴⁷. This effect is not limited to Arabidopsis, and it could be observed in rapeseed, tomato, and tobacco, suggesting that *WIND1* function is conserved across species^{255,256}. Although not every callus can regenerate organs, *WIND1* overexpression in Arabidopsis increases the number of somatic embryos on the callus, indicating an enhanced potential for organ regeneration¹²¹.

When the root is wounded, callus formation can occur at the wound and unwounded sites, where shoots regenerate from both²⁵⁷. These unwounded and wounded calli appear to have distinct modes of formation; unwounded callus originates from lateral root primordia and expresses root meristem markers, whereas callus at the wound site lacks these markers¹⁰⁴. Nevertheless, the suppression of *WIND1* disrupts shoot regeneration in both types of calli²⁵⁷. Interestingly, for root meristem regeneration, *ETHYLENE RESPONSE FACTOR 115 (ERF115)*, acting upstream of *WIND1*, is crucial for regenerating root meristem after wounding, while *WIND1* is not (Fig. 21)^{257–259}.

The molecular mechanisms of *WIND1* and its downstream targets are currently under investigation. It has been shown that *WIND1* promotes callus formation via cytokinin response and activates *ENHANCER OF SHOOT REGENERATION 1 (ESR1)* by directly binding to its promoter at a 14-bp VWRE motif (GAAAAGAAAATTTC), which is essential for callus formation and shoot regeneration²⁶⁰.

ESR1 can also be upregulated by exogenous auxin application, whereas the *WIND1* transcript does not appear auxin-dependent (Fig. 1)²⁶⁰.

To find additional downstream targets beyond *ESR1*, studies using an estrogen-inducible *WIND1* expression system have shown that *WIND1* induction upregulates 2,390 genes and downregulates 2,140 genes at least two-fold within 24 hours²⁶¹. These genes predominantly relate to cellular reprogramming, defense response, cell wall remodeling, and vascular formation. Notably, *WIND1* induction activates *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE19 (XTH19)* and *XTH20*, highlighting the potential role of *WIND1* in cell wall remodeling²⁶¹ (Fig. 21). Furthermore, among the vascular formation targets, *VASCULAR-RELATED NAC-DOMAIN 6 (VND6)*, *VND7*, and *LATERAL ORGAN BOUNDARIES DOMAIN 30 (LBD30)* were identified. This is probably one of the downstream pathways elucidating the role of *WIND1* in mesophyll cell differentiation into vascular cells observed in the VISUAL system and has functions in grafting (Fig. 21)^{261,262}.

Collectively, these findings position *WIND1* as a master regulator of regeneration, orchestrating complex functions across diverse cellular contexts. In the context of protoplast regeneration, our information is more limited. *WIND1* expression approximately doubles with protoplasting compared to the intact leaf, and it is expressed broadly in the protoplast population^{108,114}. The downstream target *ESR1* was essential for protoplast division, where its induction triggers division, and no viable cells can be detected with the loss of function mutants after 11 days of the culture¹¹⁴. However, investigating *WIND1* would provide intriguing insights into its specific role and might uncover *WIND1*-dependent or independent roles of *ESR1* in protoplast regeneration.

4.1.2. *WUSCHEL RELATED HOMEBOX 5 (WOX5)*

In addition to *WIND1*, another crucial transcription factor involved in callus formation is *WOX5*. *WOX5* belongs to the *WUSCHEL*-related homeobox (*WOX*) clade of genes many of which are closely associated with stem cell maintenance in plant development¹¹⁶. *WOX5* is required for quiescent center division rate control and maintenance, where the loss of *WOX5* causes terminal differentiation of root stem cells, and overexpression of *WOX5* blocks differentiation of distal stem cell descendants that would generally differentiate. *WOX5* expression is regulated by a complex interplay between *PLETHORA 1 (PLT1)*, *PLT3*, *SCARECROW (SCR)*, *TEOSINTE BRANCHED 1 (TCP20)*, and *TCP21* proteins during embryogenesis, primary root formation, secondary root development, and the root regeneration process^{263,264}.

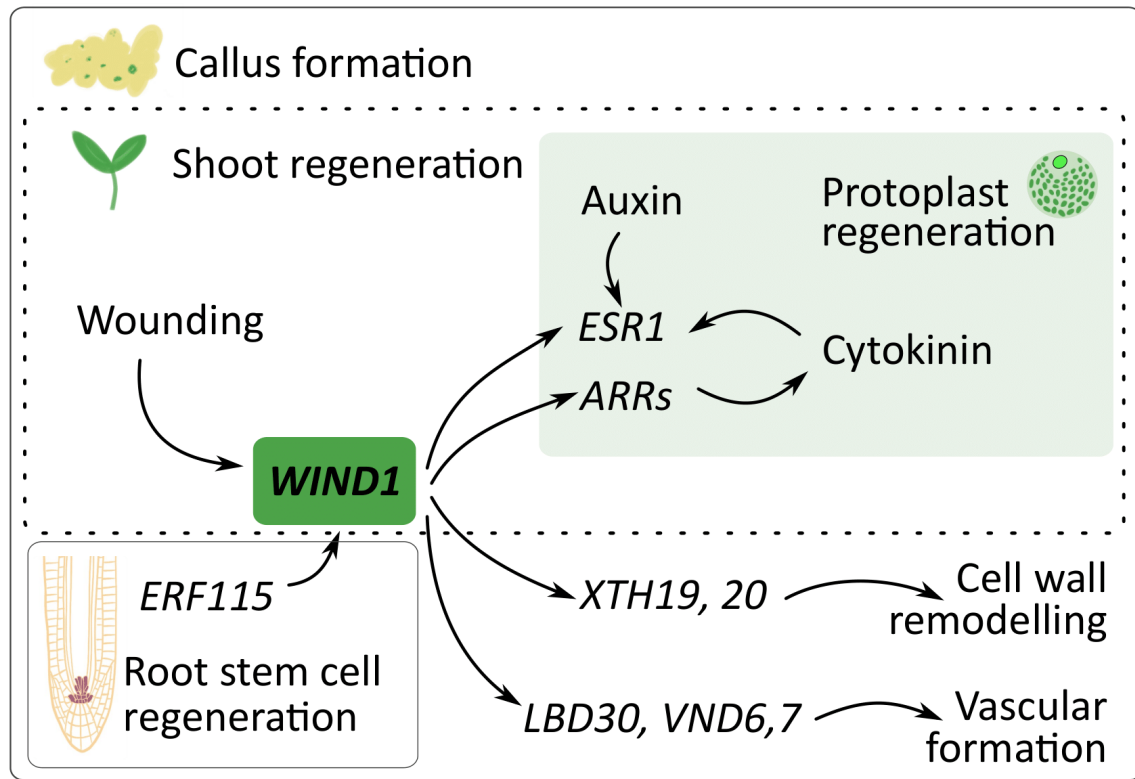


Figure 21. Molecular mechanisms of *WIND1* regulation of regeneration. *WIND1* can promote callus and shoot regeneration through the cytokinin pathway and activate *ESR1* and *ARRs*. In addition to cytokinin response, *XTH19,20* (cell wall modifiers) and *LBD30, VND6,7* (vascular differentiation regulators) are also expressed during callus regeneration. *ERF115* acts upstream of *WIND1* and is involved in callus formation. *ERF115* was shown to be essential for root stem cell regeneration, whereas *WIND1* is not. Information on protoplast regeneration is more limited. However, *ESR1* was shown to be crucial for protoplast division.

In tissue regeneration, *WOX5* is the key player for auxin-induced callus formation. It is important to emphasize that the wound-induced and *WIND1*-mediated callus formations follow different pathways. While *WIND1*-mediated callus formation involves dedifferentiation, *WOX5*-induced callus formation consists of the differentiation of pericycle-like cells already present in the tissue (in the root, cotyledon, or petal) toward a root meristem-like tissue¹⁰⁴. It was shown that upstream signals of *WOX5* activation involve auxin accumulation, which triggers *WOX11* activation, and then *WOX11* binds *WOX5* in tissue culture (Fig. 22)²⁶⁴. The pattern of *WOX11* and *WOX5* expression is similar to their dynamics in adventitious root formation, highlighting similarities in root development pathway and callus formation²⁶⁵.

It was shown that *WOX5* acts together with *PLT1* and *PLT2*, and together, they activate *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1 (TAA1)* for endogenous auxin biosynthesis (Fig. 22)²⁶⁶. *WOX5* is also involved in cytokinin signalling, where it represses negative regulators of cytokinin signalling (Type-A Arabidopsis response regulators *ARR5*, *ARR6*, and *ARR7*), thus promoting cytokinin sensitivity and stimulating shoot regeneration (Fig. 22)²⁶⁶. Until recently, it was suggested that *WOX5* expression is essential for pluripotency acquisition for callus formation and increased cytokinin sensitivity in callus-inducing-medium (CIM), and after transferring to cytokinin-rich shoot-inducing-medium (SIM) media, its expression should go down to allow shoot differentiation^{266,267}. Interestingly, in addition to transient expression in CIM, recently, it was shown that overexpression of *WOX5* in SIM improves shoot regeneration (Fig. 22). Therefore, continuous overexpression of *WOX5* promotes both callus and shoot regeneration in tissue culture not only in *Arabidopsis* but also in *Oryza sativa*²⁶⁸.

WOX5's role in protoplast regeneration is less clear, where the *wox5-1* mutant displays only slightly reduced regeneration capacity¹¹⁴. Additionally, regenerating microcalli does not express *WOX5* but expresses *ESR1* at Day 50¹¹⁴. *WOX5* gets transiently induced on Day 4; however, its expression decreases by Day 7. This suggests that *WOX5* might be necessary for cell division and at early phases of microcallus formation but not for the mature microcallus.

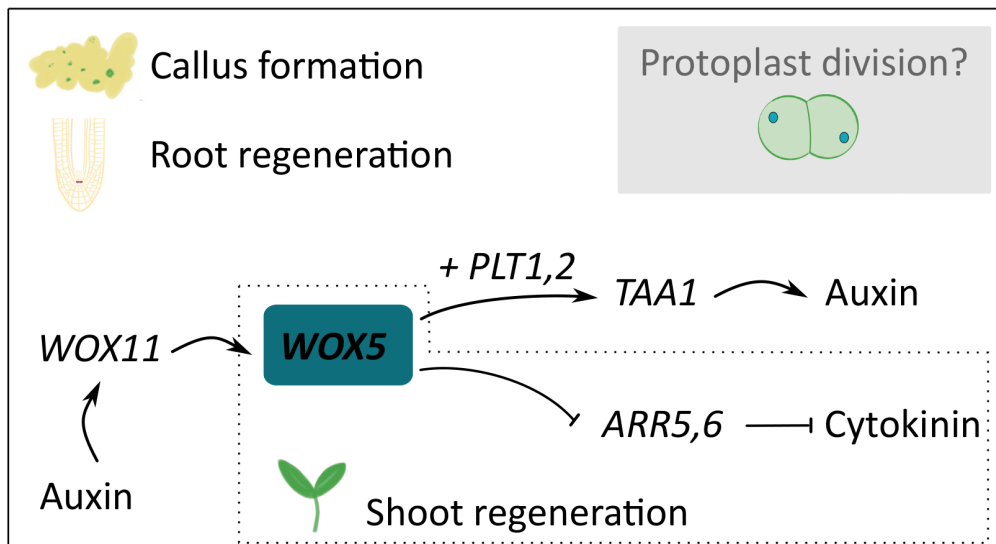


Figure 22. Molecular mechanisms of *WOX5* action during plant regeneration. *WOX5* is involved in callus, root, and shoot regeneration. Auxin triggers *WOX11*, which induces *WOX5*. Together with *PLT1,2*, *WOX5* activates *TAA1* for endogenous auxin synthesis. Recently, it was also shown that by suppressing *ARR5* and *ARR6*, which are negative regulators of cytokinin signaling, *WOX5* enhances shoot regeneration. *WOX5* might be involved in protoplast division.

4.1.3. **WUSCHEL RELATED HOMEODOMAIN 13 (WOX13)**

In addition to *WOX5*, *WOX13* is another member of the *WOX* family that is important for regeneration. At the beginning of my Ph.D. studies, the function of *WOX13* in the regeneration of seed plants was yet to be uncovered. However, proteomic studies have revealed that the *WOX13* clade is notably the most conserved within the *WOX* family³⁰. Furthermore, it has been shown that *WOX13* is essential for leaf regeneration after wounding in moss *Physcomitrella patens*¹¹⁷. *PpWOX13* functions by upregulating cell wall loosening enzymes, a necessary step for initiating cell growth during stem cell formation¹¹⁷. These were the initial motivations for selecting *WOX13* as a marker candidate.

Recent evidence has shown that *WOX13* is essential for callus growth and organ reconnection in *A. thaliana* grafting assays¹¹⁸. *WOX13* specifically promotes proliferation and the formation of highly vacuolated large cells on the callus surface, and the *wox13* mutant fails to generate callus and reconnect organs. Upon investigating the molecular mechanisms underlying *WOX13*'s involvement in callus growth, it was found that *WIND1* induces *WOX13* expression (Fig. 23)¹¹⁸.

There are 435 genes identified as direct targets of *WOX13*, including those related to cellular reprogramming (*WIND2,3*), cell wall loosening (*CELLULASE 1 (CEL1)*, *CEL3*, *GLYCOSYL HYDROLASE 9B8 (GH9B8)*, and *GH9B13*), and pectin degradation (*PECTATE LYASE LIKE 17-19 (PLL17-19)*, *PLL22*, and *PLL26*) (Fig. 23)¹¹⁸. The cell wall-modifying role of *WOX13* following wounding, coupled with the observation that *PpWOX13LB* can complement the *A. thaliana wox13* mutant, suggests that the function of *WOX13* is conserved across species, consistent with previous proteomic studies^{118,269}.

An additional function of *WOX13* during grafting was unveiled after observing that callus formation is asymmetric, originating from the top part of the graft²⁷⁰. This asymmetry is linked to auxin accumulation on the top part of the graft, inducing the expression of *WOX13* (Fig. 23). It is important to note that this auxin-dependent pathway is *WIND1*-independent, which explains why *WIND1*-repressed (*WIND1-SRDX*) plants can still respond to wounding¹¹⁸. Additionally, while wild-type plants successfully heal wounded stems within seven days, *wox13* mutants cannot do so²⁷⁰.

In contrast to its role in promoting callus formation during grafting, *WOX13* has been demonstrated to suppress the shoot formation from pluripotent callus in tissue culture¹¹⁹. It was revealed that *WOX13* downregulates factors related to shoot apical meristem (SAM) formation, such as *WUSCHEL 1 (WUS)*, *SHOOT MERISTEMLESS (STM)*, *ENHANCER OF SHOOT REGENERATION 2 (ESR2)*, and *CUP-SHAPED COTYLEDON1 (CUC1)*, while simultaneously upregulating cell wall modifiers like *EXPANSIN 17 (EXPA17)* and *ENDO-BETA-MANNANASE 7 (MAN7)* (Fig. 23)¹¹⁹. The downregulation of *WUS*

is particularly interesting, as it is a crucial regulator of SAM. Further examination has shown heterogeneous gene expression in the callus, with spatial separation of *WOX13* and *WUS* expression. It is suggested that there is a negative feedback loop between *WOX13* and *WUS* (Fig. 23), where *WOX13* promotes differentiation, and *WUS* promotes SAM identity¹¹⁹.

As evident from grafting and wounding assays, *WOX13* exhibits diverse roles, promoting or repressing different regenerative mechanisms depending on the context^{118,119,270}. This underscores the importance of investigating *WOX13* expression in protoplast regeneration, where its transcript can be detected first on Day 2 and is upregulated between Days 2-4. Then, its transcript becomes undetectable by Day 7¹⁰⁸. This dynamic pattern implies that *WOX13* could orchestrate various cellular functions during protoplast regeneration. *WOX13* could promote cell wall remodeling as indicated by its upregulation on Days 2-4 while potentially inhibiting division, as it is absent between Days 5-7, coinciding with the first round of division.

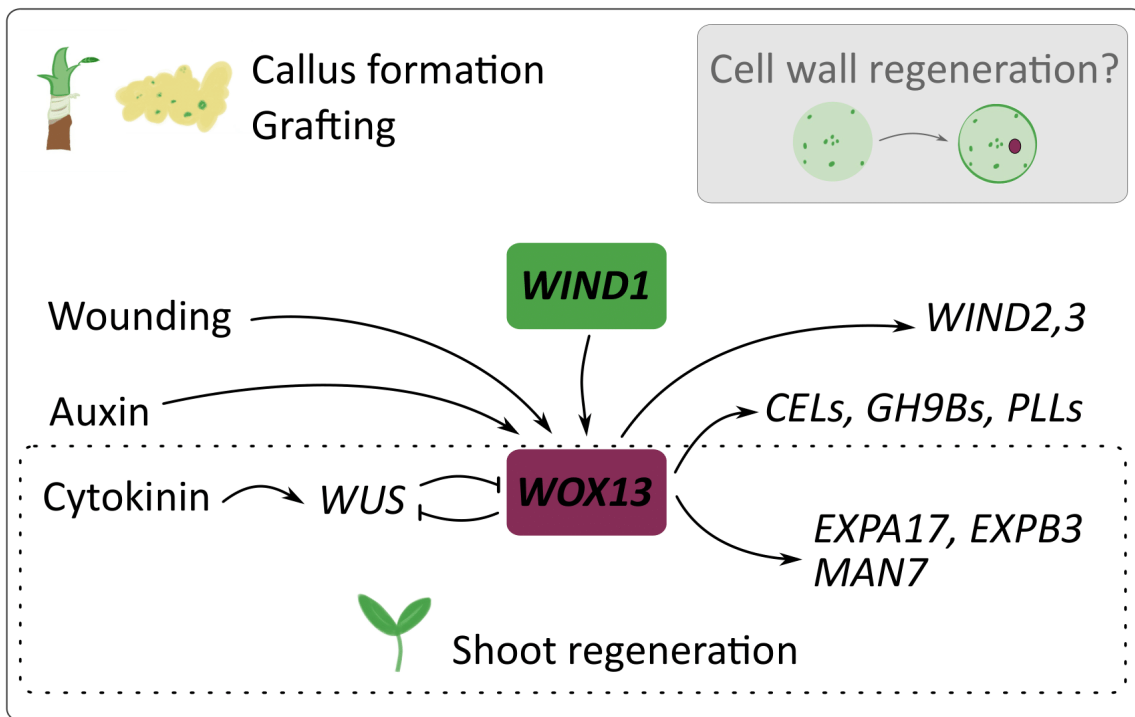


Figure 23. Molecular mechanisms of *WOX13* function in regeneration. During callus formation and grafting, *WOX13* can be induced by auxin, wounding, or *WIND1* action. Consequently, *WOX13* can upregulate *WIND1,2* for cellular reprogramming and *CELs, GH9Bs*, and *PLLs* for cell wall modifications. During shoot regeneration, however, *WOX13* acts as a negative regulator by repressing *WUS*. There is also a negative feedback loop between *WOX13* and *WUS*, controlling SAM regeneration. Also, during shoot regeneration, *WOX13* induces cell

wall modifiers such as *EXPA17*, *EXPB3*, and *MAN7*, contributing to cell differentiation. *WOX13* might be involved in protoplast cell wall regeneration, which is shown with nuclear localised red fluorescence signal.

4.1.4. *INDOLEACETIC ACID-INDUCED PROTEIN 8 (IAA8)*

Contrary to *WIND1*, *WOX5*, and *WOX13*, *IAA8* is a less studied factor in regeneration. It belongs to the Aux/IAA family of transcription factors participating in callus formation and initiation of lateral root development^{271,272}. They encode generally unstable nuclear proteins, a feature that enables dynamic changes in Aux/IAA abundance in response to auxin levels²⁷¹. At lower auxin levels, Aux/IAA proteins repress the activity of AUXIN RESPONSE FACTOR (ARF) transcription factors by forming a heterodimer with them. Conversely, when auxin levels rise, auxin binds to the TRANSPORT INHIBITOR RESPONSE 1 (TIR1) and AUXIN F BOX PROTEIN (AFB) receptors, collectively forming the SCF ubiquitin ligase complex. This SCF complex targets Aux/IAs for ubiquitination and subsequent degradation, thereby releasing the repression on ARF.

Aux/IAA proteins have a conserved mechanism of action due to four conserved domains²⁷¹. Domain I interacts with TOPLESS (TPL) and TPL-related (TPR) proteins to repress their ARF partners. Domain II interacts with the SCF ubiquitin ligase complex and regulates Aux/IAA abundance. Domains III and IV facilitate the homodimerization of Aux/IAA proteins and their heterodimerization with ARFs²⁷¹. Within the Arabidopsis genome are 29 encoded Aux/IAA proteins, and among them, *IAA8* stands out because it is more stable compared to other IAs²⁷¹. Its transcript levels do not change with one h auxin application, positioning *IAA8* as a late-response gene whose expression relies on *de novo* protein synthesis²⁷².

In the natural developmental context, *IAA8* is a negative regulator of lateral root development by repressing *IAA1*, *IAA19*, and *GH3.3* (Fig. 24)¹²⁰. Thus, *IAA8* induction suppresses lateral root formation, and loss of function mutants have more lateral roots than wild type 35. Ectopic overexpression of *IAA8* leads to abnormal gravitropism¹²⁰. On the other hand, *IAA8* functions as a positive regulator of seed germination, downregulating the transcription of *ABSCISIC ACID INSENSITIVE3 (ABI3)*. Notably, cold and reactive oxygen species (ROS) collaborate to enhance *IAA8* protein accumulation during germination (Fig. 24)²⁷³. In zinnia culture, mesophyll cells differentiating to tracheary elements express *IAA8* and the *zIAA8* transcript is expressed during vascular development²⁷⁴. *IAA8* further contributes to floral organ development by regulating jasmonic acid

levels, likely via ARF6/8 proteins (Fig. 24)²⁷⁵. As expected from its roles in development, *IAA8* is expressed in flowers, roots, stems, leaves, and siliques²⁷⁵.

The specific role of *IAA8* in protoplast regeneration remains undiscovered. However, recent findings contribute to a better understanding of the roles of Aux/IAA in regeneration. Notably, research has revealed that CALMODULIN 6 (CaM6), a calcium-sensing protein, interacts with *IAA19*, destabilizing *IAA19*'s interaction with ARF7. This interaction alleviates ARF7 repression, influencing auxin-induced callus formation²⁷⁶. It is possible that *IAA8* may harbour similar roles, yet to be unveiled.

Despite our current lack of understanding of how *IAA8* is involved in regeneration, it is essential to recognize its established roles in development, particularly in lateral root formation. Given the parallels between regeneration and lateral root formation, it is plausible that *IAA8* has a role in regeneration. Transcriptomic studies of regenerating protoplast revealed that the expression of *IAA8* is downregulated approximately threefold after protoplasting. Subsequently, there is a gradual increase in expression throughout one week¹⁰⁸. This makes *IAA8* a great marker candidate since the gradual increase in its expression levels could indicate successful progression in regeneration. Further connecting *IAA8* expression to physiological changes during protoplast culture will also contribute to a deeper understanding of its roles in regeneration.

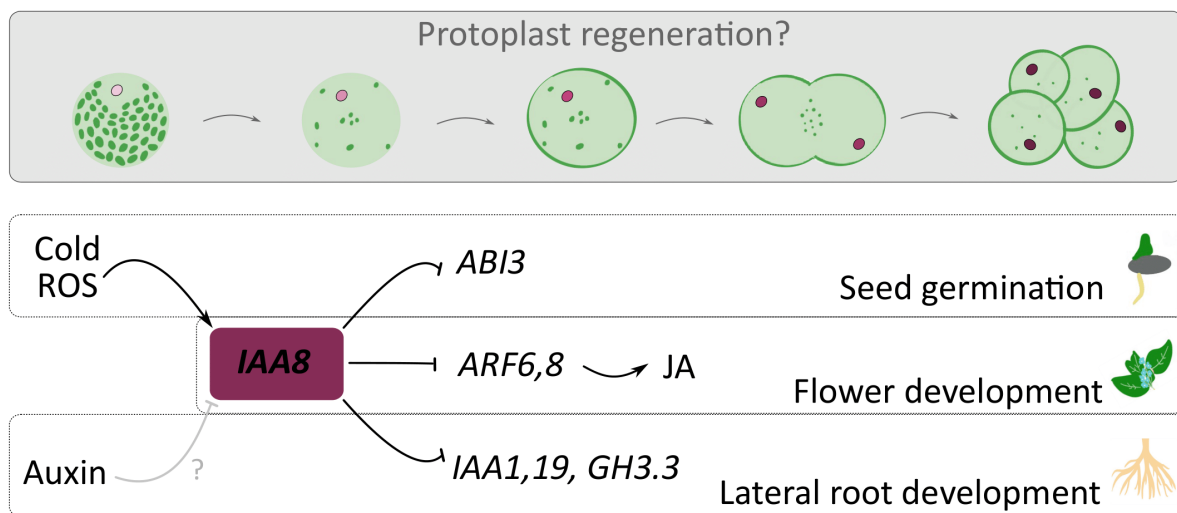


Figure 24. Molecular mechanisms of *IAA8* in plant development. *IAA8* belongs to the Aux/IAA family of proteins whose abundance changes dynamically depending on auxin levels. However, the *IAA8* transcript does not change with one h auxin application. Hence, *IAA8* regulation appears at a protein level, as indicated by a gray line. *IAA8* is a negative regulator of lateral root development and acts by downregulating *IAA1*, *IAA19*, and *GH3.3*. Additionally, *IAA8* interacts with ARF6 and ARF8 during flower development, which alters jasmonic acid (JA) signaling; during seed germination, *IAA8* represses *ABI3* and can trigger germination. Notably, cold and ROS

treatments increase *IAA8* accumulation and initiate germination. *IAA8*'s role in protoplast regeneration is yet to be determined but transcriptomic studies show its gradual upregulation during the culture, indicated by gradual increase in nuclear localised red fluorescent protein signal.

Even though transcriptomic data is available for the selected gene targets, the constraint of transcriptomic studies to specific time points might lead to omitting transient changes in gene expression between those time points. Continuous tracking of cells in the regenerative process is crucial to capture these dynamics, and thus, establishing fluorescent marker lines is helpful to correlate molecular target expression with physiological events. Additionally, these marker lines will be used as regeneration stage-specific reporters. *WIND1* will be the reporter of initiation of dedifferentiation, *IAA8* dedifferentiation and progression into regeneration, *WOX13* cell wall regeneration and *WOX5* for cell division (Fig. 25). In pursuit of this goal, I have optimised a transient expression protocol to facilitate fluorescent protein (FP) screening in *A. thaliana* mesophyll protoplasts. Such screens require a high efficiency protoplast transformation (>50%)²⁷⁷, and I have optimised the transformation protocol to the level of up to 70% efficiency. Subsequently, I selected the best FPs from each spectrum, creating robust marker lines for *WIND1*, *WOX5*, *WOX13*, and *IAA8*. Fluorescence reporter expression in these marker lines is strong enough to be distinguished from protoplast autofluorescence. Additionally, I have developed a tricolour marker line with *P_{WOX13}* expressing *mScarlet-I*, *P_{WIND1}* expressing *sfGFP*, *P_{WOX5}* expressing *mTagBFP2* (named *WOX13:WIND1:WOX5:HYG*). Overall, this Chapter aims to develop stage-specific tools to monitor and engineer protoplast regeneration, which can be used for studying plant development beyond this research.

Deregulated gene expression of TFs during protoplast regeneration

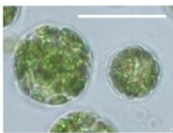
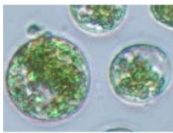
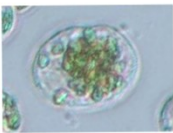
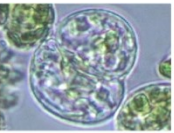
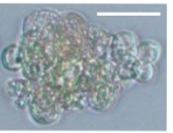
Day	0	1	2	4-6	7-11
Expected physiological changes					
<i>WIND1</i>	+1.63				
<i>IAA8</i>	-2.93	+0.99	+0.70	+1.40	+0.83
<i>WOX13</i>	NA	NA	+0.87	+0.91	NA
<i>WOX5</i>				+1.72	-1.11

Figure 25. Deregulated gene expression patterns of the chosen TFs during protoplast regeneration. *WIND1*, *IAA8*, *WOX13*, *WOX5* transcription factors are chosen for their deregulated gene expression and their roles in plant development and regeneration as explained above. Green colour shows upregulation and blue colour shows downregulation based on RNAseq data reported by Chupeau et al., (2013)¹⁰⁴. Changes in transcript amount is relative to the previous time point. For Day 0 it is relative to living plant. Besides that it is (Day 1/Day 0); (Day 2/Day 1); (Day 4/Day 2); (Day 7/Day 4). Scale bar shows 50 μ m.

4.2. Results

4.2.1. Optimizing the protoplast transformation protocol for transient gene expression

The selection of FPs as reporters is a critical decision, as their visibility depends on the expression levels of the marker. In the context of protoplasts, FPs pose a particular challenge, as they often lack sufficient brightness to overcome the autofluorescence in protoplasts, which originates from chloroplasts in the red and green ranges, as well as from the cytoplasm in the blue range¹⁹². Therefore, conducting FP screening before establishing stable lines allows for a rational choice of FPs. FP screening could be done with transient expression assays in protoplasts or with agroinfiltration in tobacco. As these FPs will be employed to monitor protoplast regeneration, a protoplast transient expression assay is used in this study.

A prerequisite for FP screening in the protoplast transient expression assay is achieving a consistently high transformation efficiency to generate a robust signal for quantification with a plate reader and or microscopy. To address this, I optimised a transformation protocol, the final version of which is detailed in Chapter 7 (Materials & Methods) and Appendix Section 8.1. Modifications were made based on the protocols by Chupeau et al., 2013¹⁰⁸, with reference to Faraco et al., 2011²⁵². In summary, the improvements were made by the changes as follows. The cell wall digestive enzyme concentrations were doubled as ONOZUKA R-10 (0.2%), MACEROZYME R-10 (0.06%) and Driselase (0.08%). The digestion duration was limited to 10 hours, with a recommendation against exceeding 12 hours. The resulting protoplasts were isolated and washed three times with MGG solution, using gravitational sedimentation instead of centrifugation during the washes.

With the new protocol, the protoplast number doubled to be 1×10^6 cells/mL and resuspended in MMM solution (0.4 M mannitol, 15 mM MgCl_2 , 0.1% w/v MES, pH unadjusted usually between 4.75-5.0). Protoplasts were transformed in 1 ml 96-well plates, with 8 μl DNA (500 ng/ μl) added to 75 μl of the protoplast suspension. This was followed by adding 83 μl PEG (0.4 M mannitol, 0.1 M $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$, 40% PEG4000, pH 8) and a 1-minute incubation. Subsequently, the tubes were filled with MGG solution and incubated in darkness for 1 hour at room temperature. After the incubation, the solution was removed, and the protoplasts were resuspended in 100 μl of fresh MGG, either placed in the plate reader at 23°C for continuous measurement or kept in darkness at room temperature for later fluorescence measurements. This optimised protocol achieved transformation efficiencies as high as 70% (Fig. 26), which could be in general as low as 20%. The transformation efficiency achieved in this study is similar to the 60-90% range reported by the Sheen Lab²⁷⁷, which has the highest reported transformation efficiency. Notably, they use only three leaves per plant, precisely those numbered five to seven. This approach demands a considerable number of plants and growth space for large-scale transformations, which is not needed with the protocol developed in this study.

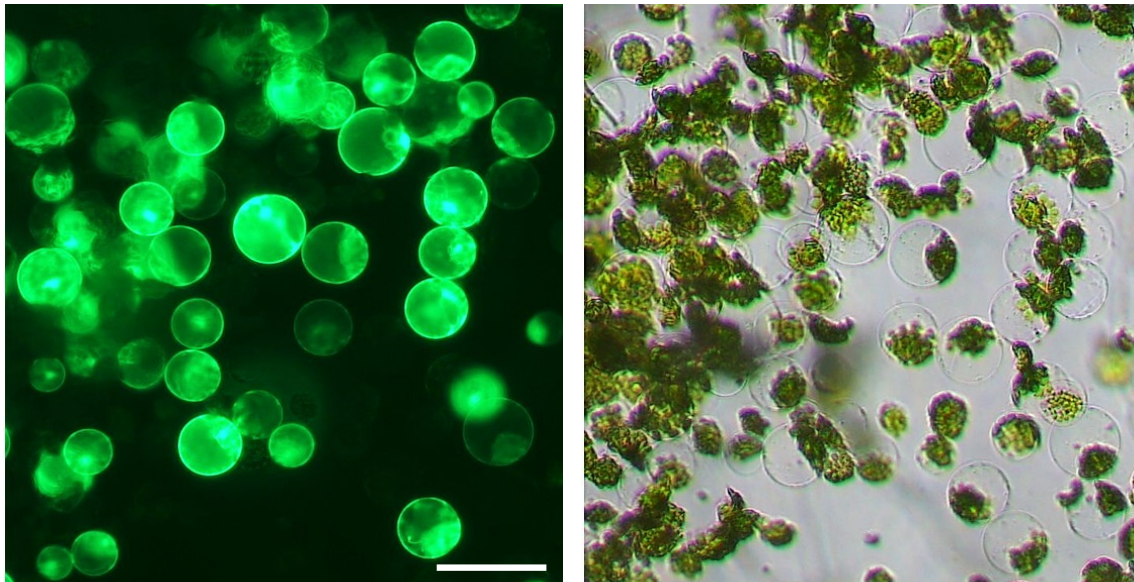


Figure 26. Protoplast transformation with the optimised protocol. *A. thaliana* mesophyll protoplasts were transformed with the *UBQ10:sfGFP:HSP* construct using PEG. The green channel is on the left, and the brightfield image is on the right. The majority of the intact protoplasts have *sfGFP* signals. Microscopy images were taken with Nikon Ti Eclipse T2; the scale bar is 50 μm .

4.2.2. Fluorescent protein (FP) screening

I then used the improved protoplast transformation protocol to screen FP proteins to identify FP variants with improved brightness and dynamics that can potentially overcome protoplast autofluorescence. All four spectra (green, red, yellow, and blue) were included to identify the best-performing FPs in protoplast systems. For green ranges, sfGFP and mGFPmut2 were selected; for red, mScarlet-I and mRuby2; for yellow, mVenusNB and eYFP; for blue, mTagBFP2, SCFP3A, and mTurquoise2 were used (Fig. 27). Nuclear localised versions of the fluorescent proteins (SV40) at C terminus were included as well to more clearly distinguish FP expression from background autofluorescence.

In general, the expression of FPs was detectable 6 h after transformation in plate readers (Fig. 27a, c, e, g). The strongest signal was seen with sfGFP, mScarlet-I, and eYFP_SV40 from green, red, and yellow spectra, respectively. Amongst them, sfGFP had the highest fold change, followed by mScarlet-I and eYFP_SV40. The blue range was hard to characterise in general due to strong blue range autofluorescence in protoplasts. However, mTagBFP2 performed better than the other proteins

tested. Fluorescence signal plateaued around 15 h after transformation for green (Fig. 7a) and red spectra (Fig. 27c). In contrast, expression continued to increase for the yellow spectrum (Fig. 7e) and the blue spectrum (Fig. 27g). Overall, nuclear-localised version of the FPs had weaker signal than the unlocalised versions in plate readers, except eYFP_SV40.

It is essential to highlight that this FP screening uses the *UBQ10* promoter and *HSP* terminator, one of the strongest constitutive expression combinations. However, when combined with TF promoters, and not expressed with the strongest constitutive promoter, the expression levels of FPs can be weaker, making it challenging to distinguish them from background autofluorescence in microscopy. Consequently, nuclear-localised versions of sfGFP, mScarlet-I, and mTagBFP2 will be employed, as they are easier to distinguish under the microscope, which will be the main method for capturing FP activity, even though the overall signal is weaker (Fig. 27b, d, f, h).

4.2.3. Testing *NLS150* and *VirD2* as alternative nuclear localisation signals

Once the choice of the FP and the preference for nuclear localisation are established, the tri-colour marker construct was designed. To avoid plasmid instability caused by repetitive sequence²⁷⁸, repeating the same SV40 nuclear localisation signal, which is an example of canonicalK-K/R-X-K/R motif, was not desirable. As a result, alternative nuclear localisation signals also belonging to K-K/R-X-K/R type signals, were explored at the C terminus, and *VirD2*²⁷⁹ and the *NLS150*²⁸⁰ signals were selected. Expression cassettes for *sfGFP* with these alternative nuclear localisation signals were constructed under the *UBQ10* promoter and *HSP* terminator. The functionality of these alternative nuclear localisation signals was confirmed using microscopy, as shown in Fig. 28.

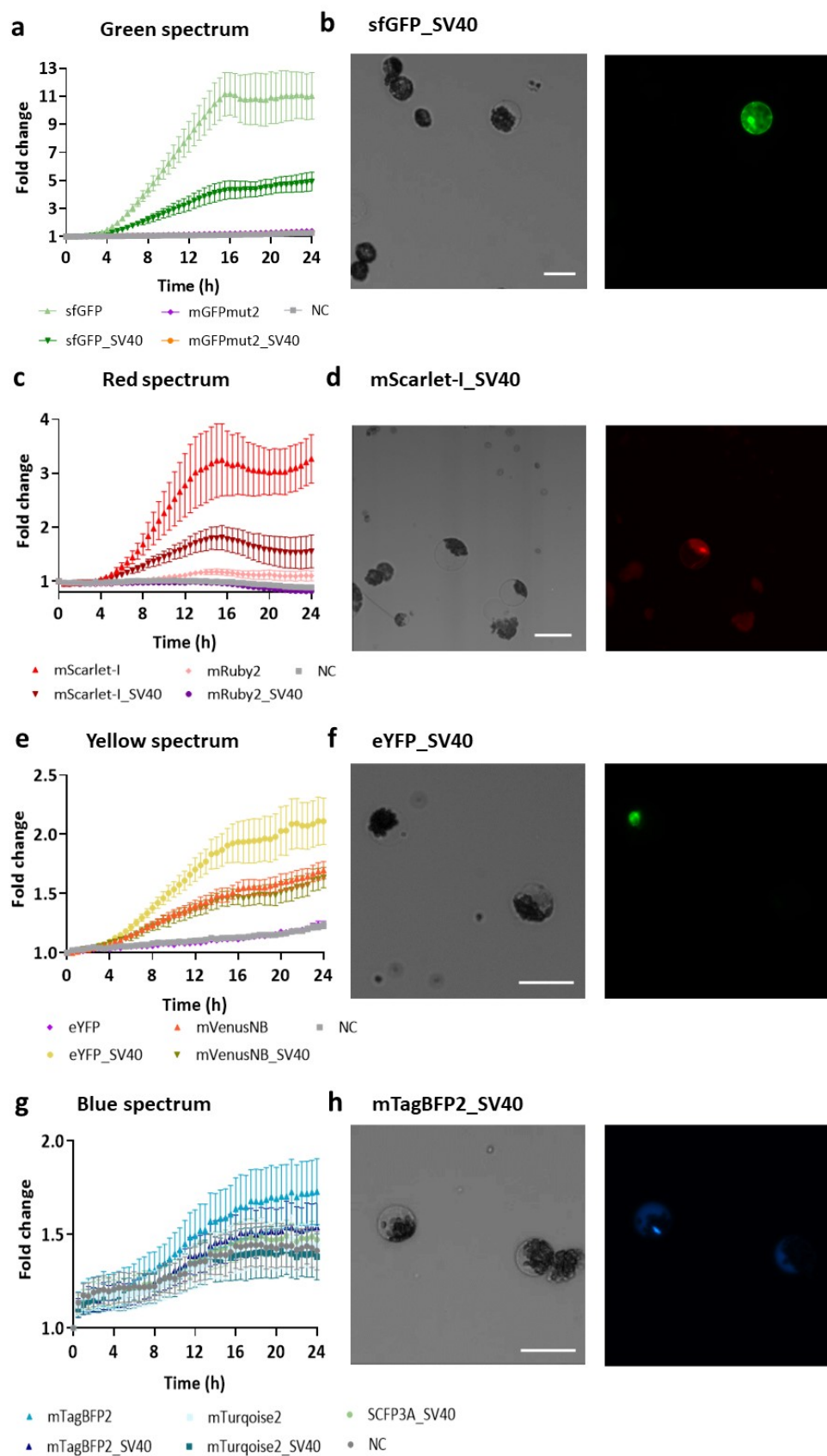


Figure 27. Fluorescent protein screening. All the fluorescence proteins are expressed under the *UBQ10* promoter and with the HSP terminator housed in a pMAPS vector. Constructs were transformed with PEG to *Arabidopsis* leaf mesophyll protoplasts in 96-well plates. **a)** Change in green fluorescence, **b)** sfGFP microscopy image. **c)** Change in red fluorescence. **d)** mScarlet-I microscopy image. **e)** Change in yellow fluorescence. **f)** eYFP_SV40 microscopy image. **g)** Change in blue fluorescence. **h)** mTagBFP2 microscopy image. Data points show a mean $\pm$ SE of 6 points (2 biological replicates with three technical replicates each). NC= Negative control, untransformed protoplasts. The scale bar is 50 μ m.

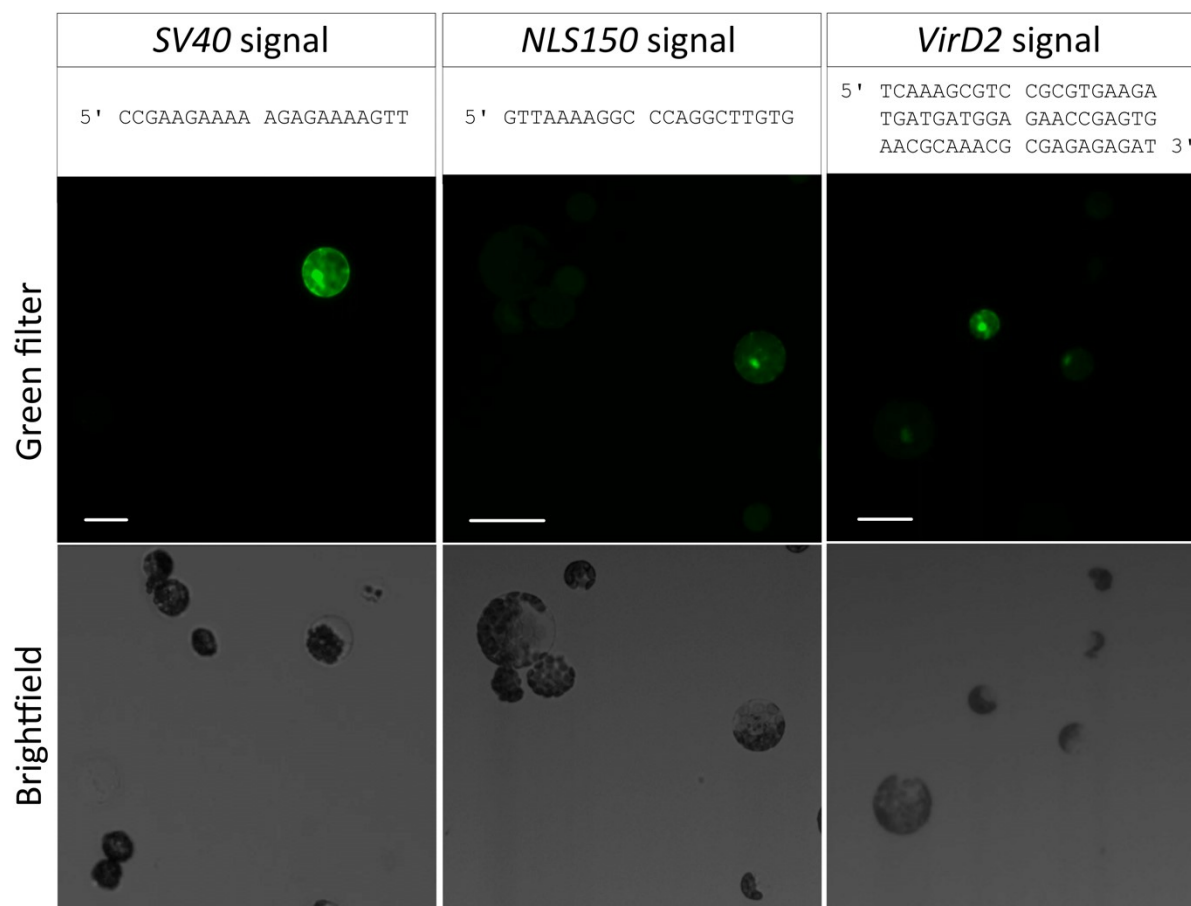


Figure 28. Testing new nuclear localisation signals with transient expression in *A. thaliana* mesophyll protoplasts. *UBQ10:sfGFP*: HSP construct was built with SV40, NLS150, or VirD2 location signal—green filter images on the top and brightfield images on the bottom panels. Scale bars show 50 μ m.

4.2.4. Making constructs for fluorescent marker lines

After choosing the FPs and nuclear localisation signals, it was time to design and make the Level 0 basic parts. Phytobrick parts for the *P_{WOX13}* (1899 bp)²⁶⁹, *P_{IAA8}* (2032 bp)²⁷⁵, *P_{WIND1}* (2992 bp)²⁶⁰, and *P_{WOX5}* (4801 bp)¹¹⁶ promoters were PCR amplified from *A. thaliana* genomic DNA with primers listed on Appendix Table 1. The length of the promoters was determined based on previous

studies^{116,260,269,275}. This involved incorporating partial 5'UTR and utilizing TSSPlant to ensure that any known motifs near the regions employed in prior studies are not overlooked. Furthermore, Phytobrick versions of *mScarlet-I_SV40*, *sfGFP_VirD2*, and *mTagBFP2-NLS150* were also made as nuclear localised FP reporter cassettes (Fig. 29).

The combination of promoters with FPs and terminators was decided as follows. P_{WOX5} was designated the colour blue, as explained later, to avoid exposing cells to UV-range light for excitation during continuous tracking, which could be toxic to the cells²⁸¹. P_{WIND1} was assigned the strongest FP (Fig. 27), *sfGFP*, to make confirmation of stable lines easier because it is expressed in the leaf, where autofluorescence is highest¹⁹². Both P_{WOX13} and P_{IAA8} are expressed in the root, where red autofluorescence is absent, facilitating easier screening of marker lines¹⁹². Regarding terminator selection, different strong terminators were preferred. Consequently, the *HSP* terminator was chosen for $P_{WOX13}:mScarlet-I_SV40$ and $P_{IAA8}:mScarlet-I_SV40$ as they will not be used simultaneously, along with the *FAD2* terminator for $P_{WIND1}:sfGFP$. At the time of this decision, only luciferase data on terminator characterisation was available; therefore, *NDUFA8* was selected as a strong terminator in the luciferase system and assigned to $P_{WOX5}:mTagBFP2_NLS150$ (Chapter 2, Fig. 9). All TUs were housed in pMAPS_1A vectors to create single-colour marker lines. Additionally, for tricolour lines, $P_{WIND1}:sfGFP_VirD2:FAD2$ was housed in pMAPS_Flip_1B, and $P_{WOX5}:mTagBFP2_NLS150:NDUFA8$ was housed in pMAPS_1C. Here and in the preceding text, 'Flip' means the vector is accepting the insert in the reversed orientation. I have built the Flip vectors to isolate individual transcription units (such as position B) from the adjacent units (positions A and C) and prevent context dependency, as discussed in Chapter 6. The antibiotic for the selection of plants was chosen to be hygromycin, and thus, the $P_{35S}:HYG:MAS$ construct occupied pMAPS_Flip_1B for single colour markers and pMAPS_Flip_1D for tri-colour markers. I constructed the Flip_1B and 1D vectors to isolate the TUs from each other, preventing any potential leaky co-activation of adjacent TUs.

Finally, at Level 2, single-colour TUs are combined with hygromycin expression cassette (Fig. 29). The tricolour markers $WOX13:WIND1:WOX5:HYG$ and $IAA8:WIND1:WOX5:HYG$ were more challenging to build as they are large constructs (~18 kb, Fig. 29). Therefore, instead of using one-pot Golden Gate assembly, separate restriction-digestion and overnight ligation steps are employed. All Level 2 constructs were sequenced before proceeding with protoplast or *Agrobacterium* transformation. Meanwhile, the tricolour construct containing *WOX13* was effectively transformed into *A. tumefaciens* ABI strain. The tricolour construct involving *IAA8* was not stable in *A. tumefaciens*.

Both ABI and GV3103 strains were tested with electro-competence and chemical competence for transforming the tricolour *IAA8* plasmid. Furthermore, growth temperatures of 24°C, 28°C, and 30°C, with suggested (2 h) and extended (4 h) recovery periods following transformation, did not produce positive *A. tumefaciens* colonies.

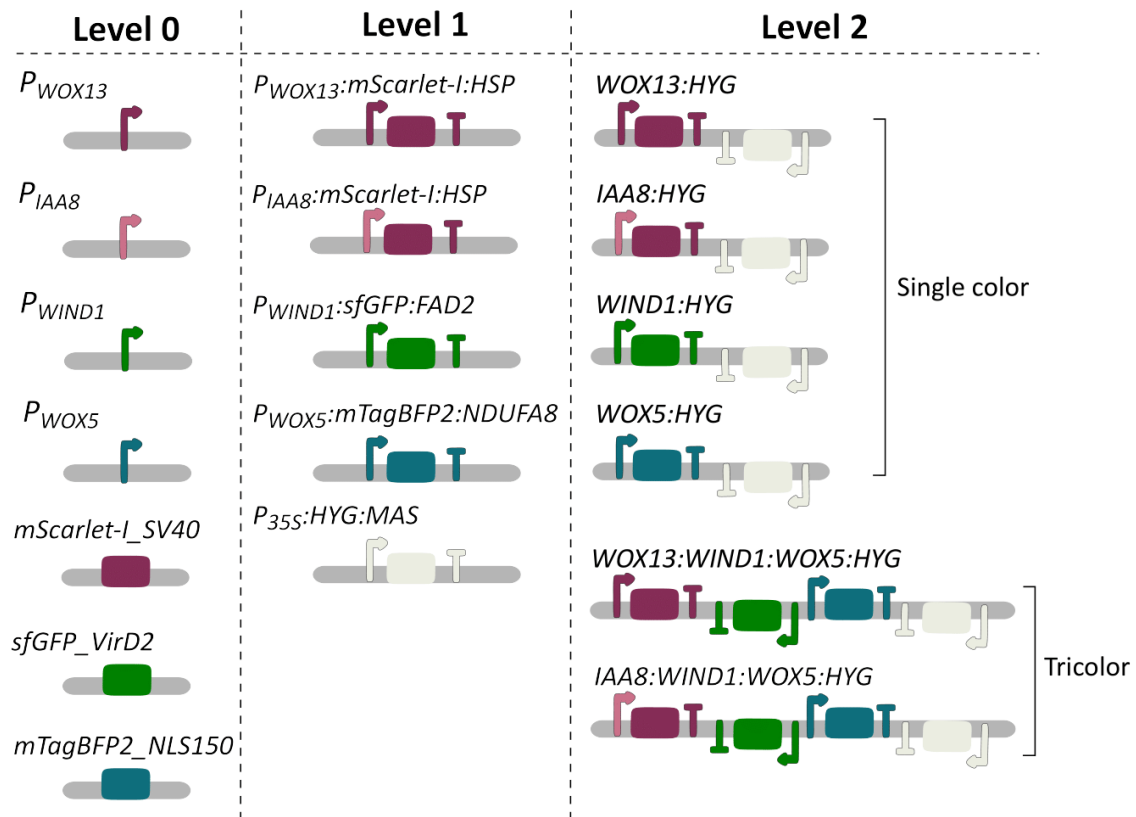


Figure 29. Construct designs for ttri-colour fluorescent protoplast regeneration marker lines. Level 0 (Phytobrick) promoter parts were made for *WOX13*, *IAA8*, *WIND1* and *WOX5*. Additionally, nuclear-localised versions of best-performing fluorescent proteins were made where mScarlet-I_SV40 was chosen for the red spectrum, sfGFP_VirD2 was selected for the green range, and mTagBFP2-NLS150 was chosen for the blue spectrum. For simplicity, nuclear localisation signals are not indicated at Level 1 and Level 2. Later, Level 1 parts are made where both *P_{WOX13}:mScarlet-I: HSP* and *P_{IAA8}:mScarlet-I: HSP* are housed in pMAPS_1A vectors. *P_{WIND1}:sfGFP: FAD2* is housed in pMAPS_1A and 1B, and *P_{WOX5}:mTagBFP2:NDUFA8* is housed in pMAPS_1A and 1C. Hygromycin selection marker (*P_{35S}:HYG: MAS*) is housed in pMAPS_1B and 1. Finally, single-colour fluorescent reporters, Level 1As, are combined with 1B_ P35S:HYG: MAS, making Level 2 single-colour constructs. Tricolour constructs are made with either *WOX13* or *IAA8* and combined with *WIND1:WOX5:HYG*. Newly-made flip vectors are used for positions 1B and 1D to isolate the TUs from each other.

4.2.5. Confirming functionality of constructs with *A. thaliana* transient expression assay

Given that the generation of stable marker lines for *A. thaliana* typically takes approximately a year, I tried to validate the functionality of the constructs concurrently through protoplast transient expression assays first. $P_{WIND1}:sfGFP_VirD2$ signal is detected around 6 hours post-transformation, and its intensity increases throughout the timelapse (Fig. 30a), consistent with FP screening results (Fig. 27). This suggests that *WIND1* was expressed in protoplasts during or before transformation, as expected from previous transcriptomic studies¹⁰⁸. $P_{WIND1}:sfGFP_VirD2$ is not expressed in all protoplasts, as not every protoplast undergoes transformation. Additionally, among the protoplasts that are transformed, not all are expected to express *WIND1*.

The remaining single-colour markers were imaged on Day 2, as their expression was not expected earlier. Fluorescent signals for $P_{WOX13}:mScarlet-I_SV40:HSP$ (Fig. 30b), $P_{IAA8}:mScarlet-I_SV40:HSP$ (Fig. 30c), and $P_{WOX5}:mTagBFP2_NLS150:NDUFA8$ (Fig. 30d) confirm that the promoters can drive nuclear-localised FP expression. *WOX5* signal was fainter compared to other FPs, which is expected to be upregulated on Day 4. However, contamination frequently appeared three days following protoplast transformation. As a result, a transient expression approach was employed rather than culturing transformed protoplasts and imaging was done at Day 2.

During the transient expression, the *WOX13* tricolour construct was not yet completed. Consequently, only the *IAA8* tricolour construct was utilized for the protoplast transient expression assay. Following a protocol similar to the single-colour constructs, imaging was conducted on Day 2 post-transformation, due to contamination appearing around three days post-transformation. The nuclear-localised FP signals were observed in the red (*IAA8*), green (*WIND1*), and blue (*WOX5*) spectra, with the blue signal (*WOX5*) being the faintest (Fig. 31).

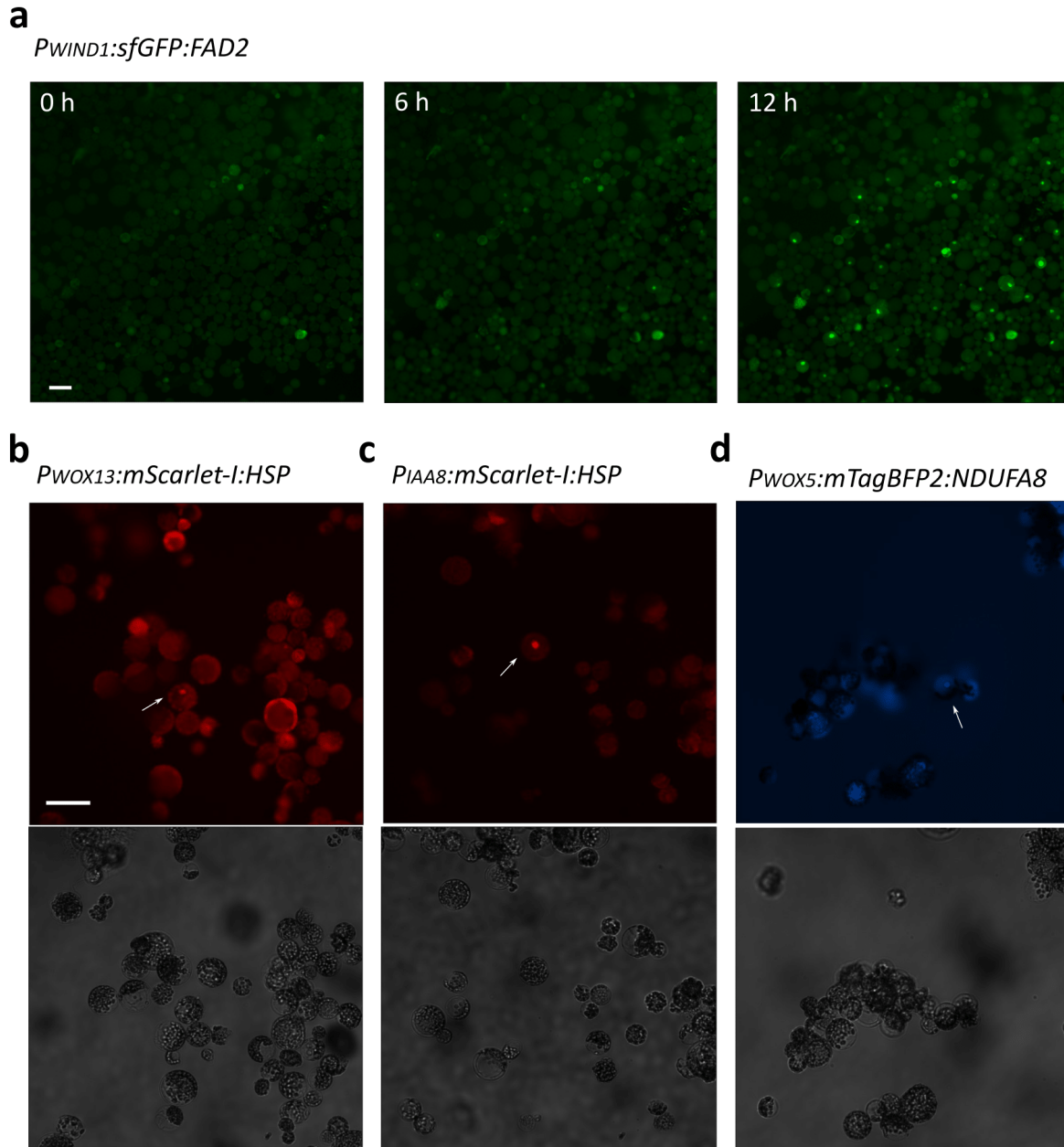


Figure 30. Confirming functionality of the regeneration stage markers using protoplast transient expression assay. **a)** *P_{WIND1}:sfGFP_VirD2:FAD2* timelapse images. The first image is right after transformation, making $t = 0$ h, the second image is taken at 6 h, and the third image at 12 h. *sfGFP_VirD2* signal appearing around 6 hours suggests that *WIND1* was expressed before the transformation process **b)** *P_{WOX13}:mScarlet-I_SV40:HSP* **c)** *P_{IAA8}:mScarlet-I_SV40:HSP* **d)** *P_{WOX5}:mTagBFP2-NLS150:NDUFA8*. Panels b-d were captured on Day 2 post-transformation. This timing was chosen because these signals were not expected immediately after transformation; instead, they were expected to activate during later stages of protoplast culture—Day 2 for *IAA8* and *WOX13* and Day 4 for *WOX5*. Signals confirm the functionality of promoters with the fluorescent signals of *WIND1*, *IAA8*, *WOX13*, and *WOX5*. Scale bars show 50 μm .

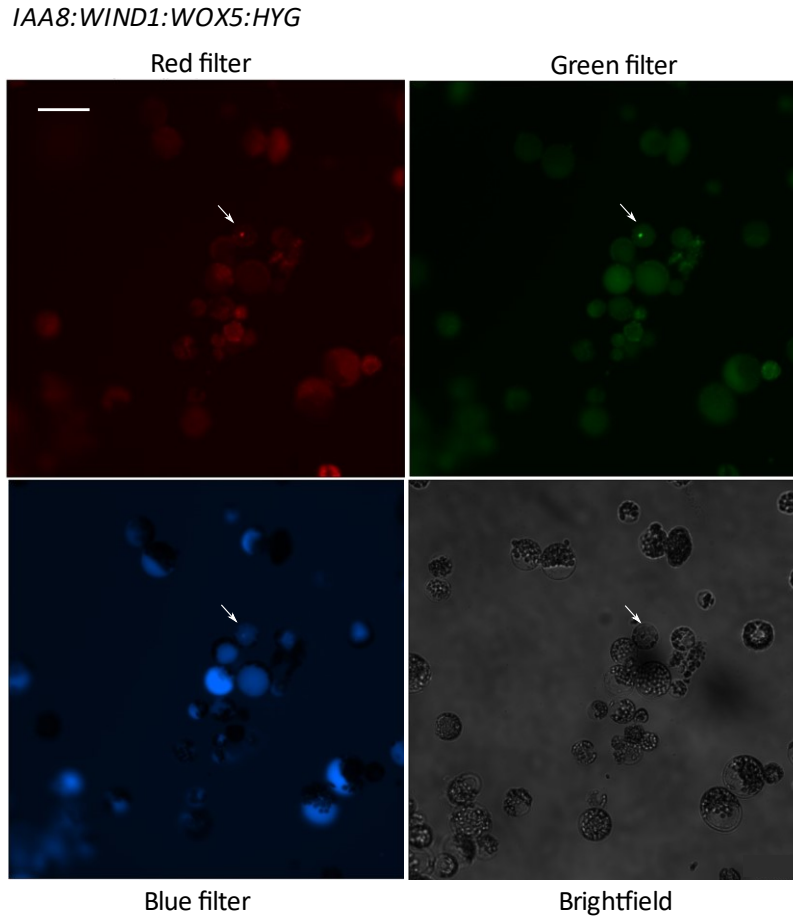


Figure 31. Conforming the marker expression of *IAA8:WIND1:WOX5:HYG* construct using protoplast transient expression assay. Image taken two days post-transformation. Red filter *mScarlet-I_SV40* is the reporter for *IAA8* expression. Green filter *sfGFP_VirD2* signal and blue *mTagBFP2-NLS150* are reporters of *WIND1* and *WOX5*, respectively. The scale bar is 50 μm .

4.2.6. Generating stable lines of the *A. thaliana* protoplast regeneration markers

Stable *A. thaliana* marker lines were generated by *Agrobacterium* floral dipping and selected on hygromycin as described in Chapter 7, Materials & Methods. Plants that grew on hygromycin were selected for microscopy, and expression of FPs in expected cell types was confirmed for three generations. Representative images for both single colour and the *WOX13* tricolour line (the *IAA8* tricolour was not stable in *A. tumefaciens*) are shown in Fig. 32. Routinely ten independent insertion lines are screening at first generation (T1), followed by five lines were screened for the preceding generations (T2, T3). It is advised to sequence the insertions and characterise the single insertion events for floral dipping in the future, even though they were not performed here.

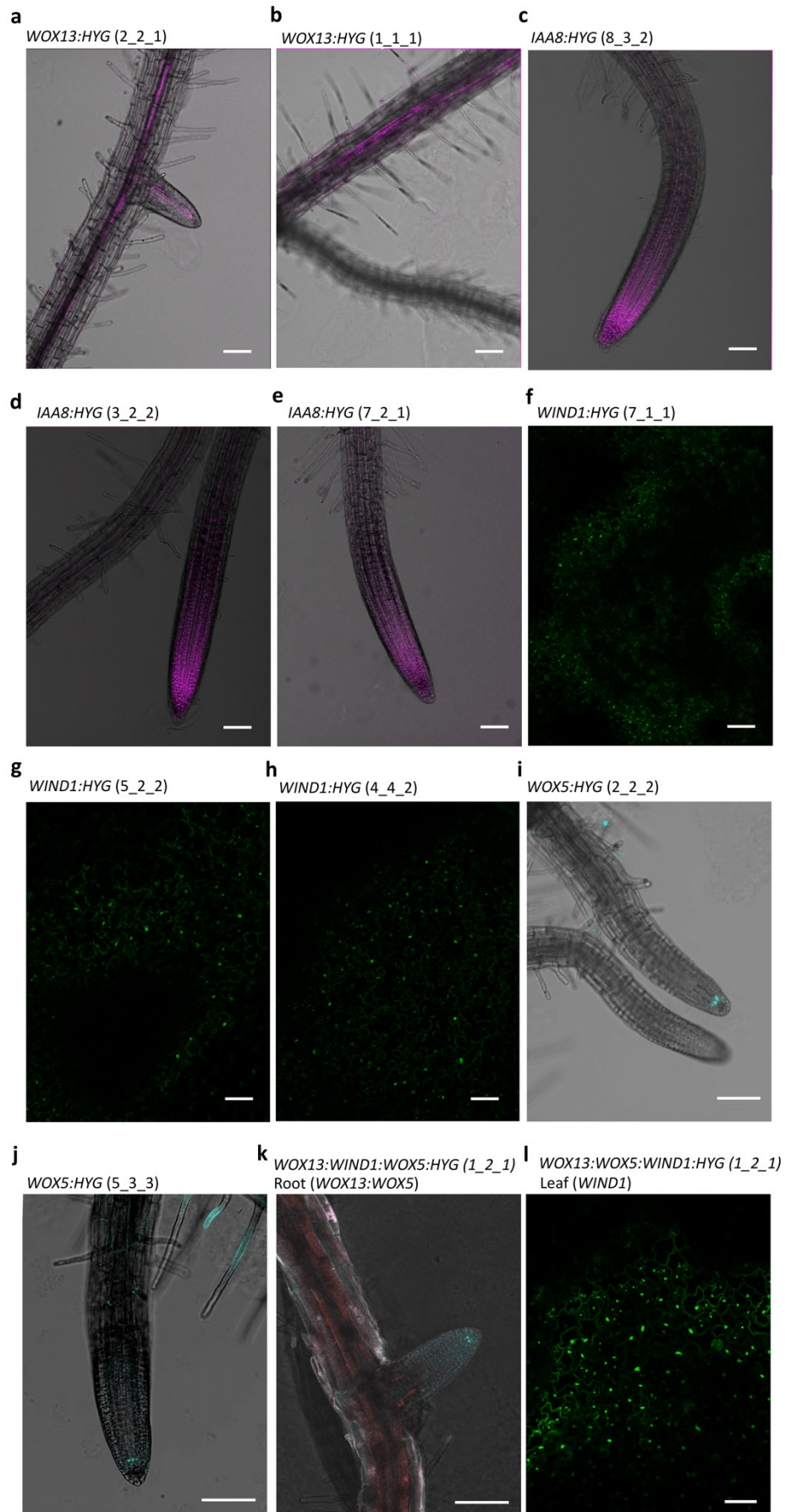


Figure 32.. Reporter expression in the stable fluorescent regeneration stage marker lines. All lines were transformed by *Agrobacterium* floral dipping and grown for three generations to confirm stable insertions. All images were taken with a Leica SP8 Confocal microscope. **a,b)** *WOX13:HYG* lines expressing *mScarlet-I_SV40* in the vasculature and lateral roots. **c,d,e)** *IAA8:HYG* lines expressing *mScarlet-I_SV40* in the root, expression higher at the root tip. **f,g,h)** *WIND1:HYG* lines expressing *sfGFP_VirD2* in the leaf. **i,j)** *WOX5:HYG* lines expressing *mTagBFP2-NLS150* in the quiescent center (QC) of a lateral root. **k,l)** Tricolour *WOX13:WIND1:WOX5:HYG* marker line expressing *mScarlet-I* (*WOX13*) and *mTagBFP2-NLS150* (*WOX5*) in the root (e) and *sfGFP_VirD2* (*WIND1*) in the leaf. Scale bars show 100 μ m.

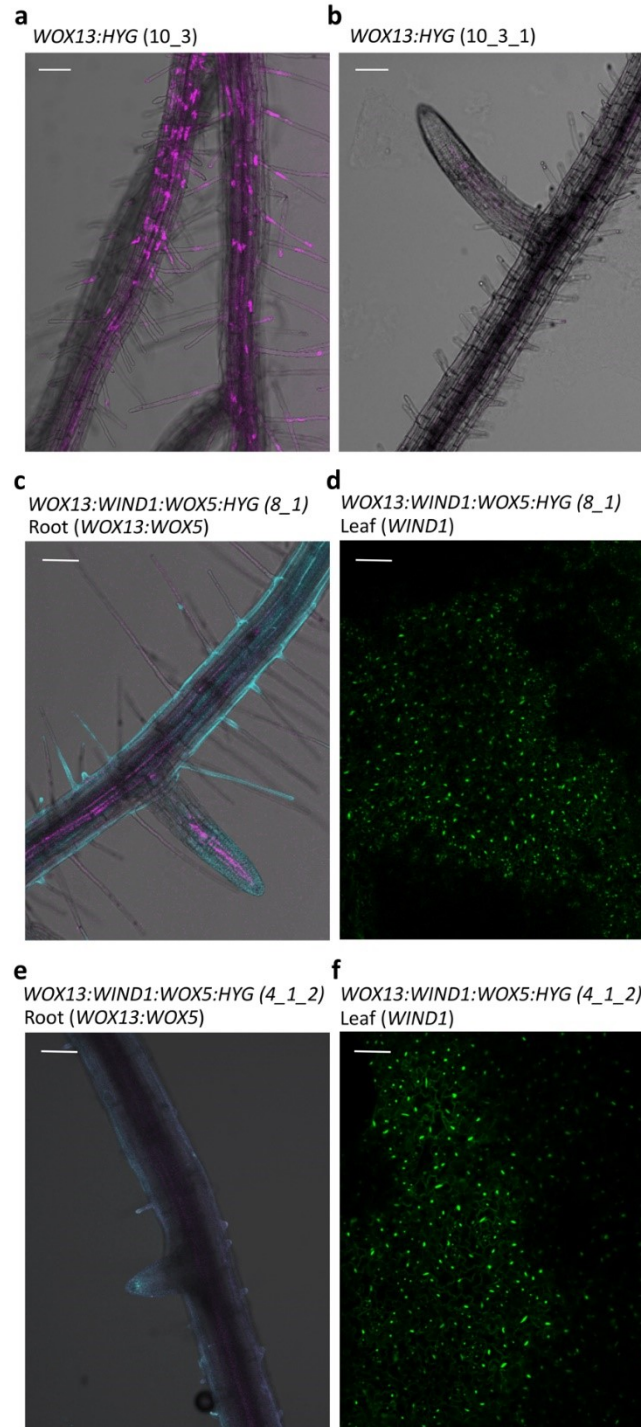


Figure 33. Issues observed during *A. thaliana* marker line screening. **a,b)** Decrease in fluorescent signal intensity and loss of *WOX13* expression in root hairs from T2 to T3. **c,d)** Tricolour marker line expressing strong *WOX13* and *WIND1* with weak and unspecific *WOX5* expression.. **e,f)** Tricolour marker line expressing strong *WOX5* and *WIND1* with weak *WOX13* expression. Scale bars show 100 μ m.

4.3. Discussion

While plants possess a remarkable ability for regeneration, it varies species to species. Engineering regeneration holds significant potential for advancing agricultural practices, including tissue culture, grafting, and transgenics. This is particularly relevant for species that are challenging to transform or regenerate. In my research, I have focused on protoplast regeneration, particularly on transcription factors *WIND1*, *WOX5*, *WOX13*, and *IAA8*. While their roles are well-established in tissue culture and plant development, their involvement in protoplast regeneration has yet to be determined. Additionally, developing fluorescent marker lines connecting their expression to protoplast physiology during regeneration presents an opportunity to make regeneration stage-specific markers for tracking culture progression and potentially engineering regeneration.

4.3.1. FP screening

FP screening using the optimised protoplast transformation protocol for consistent and high transformation efficiency (Fig. 26, described in Chapter 7, Materials and Methods) identified sfGFP, mScarlet-I, and mTagBFP2 as the best-performing fluorescent proteins from different spectra (Fig. 27). It is worth noting that signals from the blue spectrum were comparatively weaker than those from other spectra (Fig. 27). Developing novel blue-range FPs would be essential to facilitate the more widespread and simplified use of tricolour marker lines.

In this study a relative intensity approach was taken for comparing FPs. If precise quantification of the fluorescent protein signal is essential, normalisation should be considered to account for differences in well-to-well transformation efficiency:

$$\text{Normalised fluorescence} = \frac{\text{Fluorescence intensity at } t_n}{\text{Fluorescence intensity at } t_0 \times \text{transformation efficiency}}$$

(Eq. 2)

Transformation efficiency can be calculated from microscopy images. Equation 2 is recommended for FPs exhibiting more than a two-fold change, as lower signals become challenging to distinguish from background autofluorescence both in a plate reader and under a microscope.

In addition to the FP screening, alternative nuclear localisation signals, namely *NLS150* and *VirD2*, were validated as functional in protoplasts (Fig. 28). This eliminated the need to use the SV40 signal repeatedly. It would be interesting to characterise the nuclear localisation signals further and assess potential differences in FP intensity. Subsequently, the characterised FPs and nuclear localisation signals were incorporated into reporter constructs and transcription factor promoter parts built in this study (Fig. 29).

4.3.2. Building constructs and confirming their functionality

Despite the successful sequencing of the tricolour *IAA8:WIND1:WOX5:HYG* construct in *E. coli*, attempts to transform it into *A. tumefaciens* proved unsuccessful. In contrast, the tricolour *WOX13:WIND1:WOX5:HYG* construct was effectively transformed. The combination of IAA8 TU with other TUs appeared to pose issues, as the individual TUs were stable in *A. tumefaciens* on their own. Consequently, this specific construct was exclusively used for transient expression assays, and creating a stable marker line was not pursued (Fig. 31).

The functionality of the newly built TF promoters to activate nuclear-localised FP signals was initially validated through a protoplast transient expression assay (Fig. 30). Timelapse microscopy was employed for *WIND1*, validating the presence of *WIND1* at the time of or before transformation; the other markers were imaged on Day 2. According to RNAseq data, *WOX5* is expected to be upregulated 2-fold on Day 4, but the exact quantity of the *WOX5* transcript before Day 4 is not reported¹⁰⁸. Therefore, although observing *WOX5* expression on Day 2 initially seemed surprising, the heterogeneity of protoplast cultures suggests that each protoplast may exhibit different expression patterns. Recent findings indicate that approximately 3% of the protoplast population expresses *WOX5* after isolation, albeit at varying intensities¹¹⁴. Thus, it is plausible that a small fraction of the population might be expressing *WOX5* on Day 2. Nevertheless, timelapse imaging would be essential for all reporters to validate dynamic changes in FP expression and promoter function throughout the culture. Considering the stable marker lines are available, monitoring construct function through this method is preferable, as the additional transformation process could potentially impact regeneration.

4.3.3. Stable *A. thaliana* marker lines for protoplast regeneration

Stable *A. thaliana* whole plant marker lines were established for five constructs, providing strong expression at T3 of three lines for *IAA8* and *WIND1*, two lines for *WOX13* and *WOX5*, and one line for the tricolour (*WOX13:WIND1:WOX5:HYG*) (Fig. 32). Detecting the *WOX5* signal was challenging probably due to a combination of using a blue-range FP and the *NDUFA8* terminator being weak. At the time of making the constructs, only luciferase characterisation data was available for terminators, which showed the *NDUFA8* as a strong terminator. However, for the expression of fluorescent proteins, it is shown to be weak (Chapter 2). It would be interesting to change the terminator to for example *35S* or with *WOX5*'s native terminator and see whether they improve the signal. Similarly, it was difficult to detect strong signal for all FPs in three colour marker lines. This can be explained by the expression of a construct with a transgenic promoter differing from the expression of the endogenous gene due to context dependency, which is discussed further in Chapter 6.

4.3.4. Stage-specific tools for monitoring and engineering regeneration

The marker lines developed in this study serve multiple purposes. Firstly, they will facilitate investigating the roles of selected TFs in protoplast regeneration. While this can be accomplished using individual marker lines, the advantage of employing a tricolour, or multicolour, system arises when the molecular pathways affected by a specific treatment are unknown. For example, mechanical inputs from the plant cell wall are becoming promising regulators of development, however, the molecular pathways underlying cell wall-induced signaling remain largely unexplored^{282,283}. In this context, the tricolour marker line can help elucidate the effects of cell wall modifications on the activation or suppression of the target TF and its associated pathways, enabling the simultaneous evaluation of three targets. This reduces the workload by a third compared to individual marker lines.

Secondly, the TF promoters can serve as stage-specific tools for engineering regeneration. While the transcriptomic data suggests that they are expressed at specific stages of regeneration during the initial stages of protoplast culture (Day 0 – Day 11)¹⁰⁸, further investigation utilizing the marker lines will elucidate their roles in later stages of regeneration. Should their expression remain specific to the initial stages, they could potentially serve as stage-specific promoters, similar to cell-type-specific promoters that are indispensable for developmental engineering²⁵⁰. On the other hand, if these transcription factors (TFs) are involved in later stages of regeneration, such as callus, shoot, or root formation, they could be coupled with inducible systems to confer temporal specificity. In such a

setup, the TF promoters would express the transactivators (such as *LhGR*) and the inducible promoters (such as *pOp6*) would express the gene-of-interest. This inducible system would be active only when both the TF promoter is on and the inducer (such as DEX) is exogenously applied; a process that can be tailored according to the user's preference.

It is worth noting that I have chosen these TF to study their roles in regeneration along with their stage-specificity. However, if stage-specificity is the main aim, then alternative promoters can be used such as the promoters of *CELLULOSE SYNTHASE 1* and *3* (*CESA1*, *CESA3*) for cell wall regeneration; or the promoters of *CYCLIN-DEPENDENT KINASE B2;1* (*CDKB2;1*) and *MITOTIC-LIKE CYCLIN 3B* (*CYC3B*) for cell division. Given the likelihood of these genes being active across multiple stages of regeneration, incorporating them with inducible systems in a manner similar to the method described above would prove beneficial.

4.4. Conclusion

The ability to engineer and enhance protoplast regeneration holds immense value for plant biotechnology, yet much remains to be understood about its underlying molecular mechanisms. Therefore, in this study, I aimed to develop stage-specific tools to study protoplast regeneration. Overcoming protoplast autofluorescence proved to be a significant obstacle in achieving this objective, as it could mask FP signals across all spectra. To tackle this issue, I optimised a transient transformation protocol to ensure consistent and high transformation efficiency. This enabled the rapid screening of constructs with subsequent FP screening aimed at identifying best-performing reporters for marker line creation. Novel FPs, alternative nuclear localisation signals, and TF promoters were confirmed to be functional in whole plants, and FP signals were visualized in protoplasts. Five stable *A. thaliana* marker lines were successfully established. The tools and methods developed in this section are used throughout my thesis and will hopefully serve as valuable resources for engineering regeneration and development.

CHAPTER 5

Establishing a protoplast regeneration platform

5.1. Introduction

Reproducibility, the ability to obtain consistent results when an experiment is independently repeated, is arguably the most important aspect of scientific research²⁸⁴. In scientific research, the aim is to uncover or engineer rules or mechanisms, and reproducibility shows that the findings are reliable and relevant across a broader range of situations and contexts. However, a Nature survey involving 1,500 scientists revealed surprising statistics: 70% struggled to reproduce published results, and 60% encountered difficulties replicating their own²⁸⁵. The causes of this reproducibility crisis are multifaceted, ranging from lack of transparency regarding experimental protocols to potential misidentification of organisms or samples and the influence of cognitive biases²⁸⁵. Even when human error is eliminated, the complexity of biological systems, which respond to environmental conditions and other unknown factors not continuously controlled in the experiment, presents inherent challenges to reproducibility. In biology, achieving reproducibility demands meticulous standardisation, detailed documentation, and a keen awareness of environmental and experimental variables.

One such area grappling with reproducibility is protoplast regeneration. While instances of successful protoplast division and whole plant regeneration are documented, failures are also widespread^{108,114,286–290}. In the context of *Arabidopsis thaliana*, most protoplast regeneration protocols begin with an acknowledgement of the inability to replicate prior methods, followed by the assertion that a reproducible protocol has now been developed^{108,114,289,290}. Paradoxically, these newly developed protocols often prove challenging for other scientists to reproduce, leading to the emergence of even newer variations in the methods (Table 4).

One of the first documentation of protoplast regeneration was presented by Damm and Willmitzer et al. in 1988²⁸⁶, utilizing ½ MS medium and long-day conditions for plant growth for 3-4 weeks²⁸⁶. Protoplasting involved using 1% Cellulase Onozuka R10 and 0.25% Macerozyme R10 for up to 18 hours. Subsequently, alginate embedding was employed, with protoplasts kept in darkness at 4°C for 2 days, followed by culturing at 26°C²⁸⁶. Subsequent methods have built upon this foundation, introducing modifications to plant growth conditions, protoplasting procedures, and regeneration protocols summarized in Table 4. Plating efficiency is a common parameter to calculate regeneration efficiency, and it is defined as the ratio of callus with a size of at least 0.11 mm to a number of protoplasts plated. It was shown that plating efficiency is higher with ½ MS medium compared to B5

Gamborg medium, while B5 Gamborg media yields larger plants, providing more tissue for protoplasting²⁸⁷. Photoperiod variations during plant growth include dim continuous light, long day (16 hours light / 8 hours dark), and short day (10 hours light / 14 hours dark) (Table 4). A study investigating the impact of photoperiod revealed a fourfold increase in plating efficiency under short-day conditions compared to long day²⁸⁷. Notably, light intensity ranges widely across studies, with approximately 120- 250 lux for continuous light, 3,000 – 7,400 lux for long-day conditions, and 3,700 to 14,000 lux for short-day conditions (Table 4). The effect of light spectrum on protoplast regeneration was only tested between white and blue light. It was observed that 3-week-old plants displayed enhanced plating efficiency under white light, whereas 4-week-old plants exhibited higher plating efficiency when cultivated under blue light. Temperature for plant growth varies between 20-25°C, with one study reporting day/night temperature fluctuations (20/16°C day/night)²⁸⁷. Plant age also exhibited variability, ranging from 10 days-old for long-day plants to 4 weeks-old for short-day plants (Table 4).

Protoplasting enzymes primarily consist of Cellulase Onozuka R10 and Macerozyme R10, with concentrations ranging from 1% to 0.1% for Cellulase Onozuka R10 and 0.25% to 0.03% for Macerozyme R10, encompassing a tenfold variation (Table 4). Notably, Jeong et al. deviate from traditional enzymes, employing a mixture of Viscozyme, Celluclast, and Pectinex²⁹⁰. The duration of protoplasting spans a considerable range, typically lasting between 12-18 hours; however, certain studies only specify the duration as 'overnight,' leaving room for variability, ranging from 8 to 18 hours¹⁰⁸.

The initial regeneration conditions involve the use of alginate embedding, maintaining the culture at 4°C for 48 hours, followed by transferring the culture to 26°C in darkness (Table 1). Investigation into the impact of darkness and low light levels (~100-250 lux) revealed no observable differences²⁸⁹. Typically, the regeneration temperature aligns with the temperature for plant growth (e.g., 20°C for Chupeau et al., 2013¹⁰⁸) or is slightly higher (around 1-5 °C) compared to plant growth conditions in other protocols (Table 4). Chupeau et al. pioneered the development of liquid culture with a 50% initial cell division rate. However, it is noteworthy that Chupeau et al. did not provide information on plating efficiency.

Apart from variations in the plant growth, protoplasting, and regeneration conditions, differences in the media composition (both for plant growth and for protoplast culturing), including the type and concentration of phytohormones, are also reported, which are described in detail in Reed

& Bargmann., 2021¹²⁹. The reason why previous methods cannot be reproduced remains unclear, and it seems that all parameters (plant growth conditions, protoplasting, regeneration culture and media composition) influence culture progression collectively. Moreover, protoplasts are extremely fragile since they do not have the cell wall. Consequently, careful handling of protoplasts, including how the handler holds and shake the plates, likely plays a crucial role. Unfortunately, such nuances are frequently omitted in scientific reports and prove challenging to articulate in space constrained journal articles. This challenge is eloquently worded as “Plant tissue culture in general, and protoplast regeneration in particular, is often lightheartedly considered more of an art form than a science, requiring an experienced eye and instinctual decision making, as comprehensive systematic approaches are too vast in scope to be feasible.” (Reed and Bargmann., 2021¹²⁹)

While the specific molecular pathways impacted by environmental conditions in protoplast regeneration are unknown, more information is available on how these conditions affect tissue culture. Regarding temperature, it was found that among 17°C, 22°C and 27°C, 27°C exhibited the highest rates of callus and shoot formation from hypocotyl explants²⁹¹. This effect is attributed to the upregulation of regeneration and meristem-associated genes such as *CUP-SHAPED COTYLEDON 1* (*CUC1*) and *YUCCAs*²⁹¹. Notably, the histone variant *H2A.Z*, identified as a negative regulator of shoot organogenesis, is enriched at 17°C and decreases at higher temperatures²⁹¹. Additionally, although no direct correlation with regeneration has been established, temperature gradients influence *ERF115* levels, saw a transcription factor crucial for root regeneration^{259,292}.

In terms of light, exposure to 24 h of low-fluence red light or 24 h of darkness following root excision significantly increased the frequency of shoot regeneration²⁹³. However, continuous low-fluence red light during callus and shoot induction stages inhibited the explants' ability to regenerate shoots²⁹³. During the shoot regeneration process from barley callus, it was demonstrated that a light intensity of approximately 4500 lux during a 16 h photoperiod could either promote or inhibit shoot regeneration, depending on the cultivar²⁹⁴. Cultivars experiencing inhibited shoot regeneration also exhibited higher abscisic acid accumulation, whereas those displaying enhanced shoot regeneration showed lower abscisic acid levels. This variation is likely influenced by the light-sensitive expression of the *9-CIS-EPOXYCAROTENOID DIOXYGENASE* (*HvNCED1*) gene²⁹⁴.

Table 4. Summary of different protocols developed for protoplast culture.

Reference	Plant growth	Protoplasting	Regeneration	Efficiency
Damm & Willmitzer (1988) ²⁸²	B5 Gamborg medium 0.07% glucose Long-day (16 h light /8 h dark) 3000 lux 21°C 3 - 4 weeks	1% (w/v) Cellulase Onozuka 0.25 % (w/v) Macerozyme Darkness for 14 - 18 h	Alginate embedding Darkness for 48 h at 4°C Then darkness at 26°C	30 % initial division 0.4 - 0.6 % plating
Masson & Paszkowski (1992) ²⁸³	1/2 MS medium, 1% sucrose Short-day (10 h light /14 h dark) 9000 - 14000 lux 20/16°C (Day/Night) 3 - 4 weeks	1% (w/v) Cellulase Onozuka 0.25 % (w/v) Macerozyme Darkness, overnight	Alginate embedding Darkness for 48 h at 4°C Then darkness at 26°C	7 - 10 % plating
Siemens et al., (1993) ²⁸⁴	1/2 MS medium, 1% sucrose Long-day (16 h light /8 h dark) ~ 4500 - 6000 lux 25°C 3 - 4 weeks	1% (w/v) Cellulase Onozuka 0.5 % (w/v) Macerozyme Darkness for 12 - 16 h	Alginate embedding Darkness for 48 h at 4°C Then darkness at 26°C	1 - 4 % plating
Dovzhenko et al., (2003) ²⁸⁵	B5 Gamborg medium Continuous light ~ 120 - 250 lux 25°C 8 - 14 days	24 h, 4 °C dark pretreatment 0.5% (w/v) Cellulase Onozuka 0.5% (w/v) Macerozyme Darkness for 12 - 14 h	Alginate embedding Darkness for 24 h Then continuous light	5 - 6 % plating
Chupeau et al., (2013) ¹⁰⁴	1/2 MS medium, 1% sucrose Short-day (10 h light /14 h dark) Green box, ~ 3700 lux 20°C 18 - 21 days	0.1% (w/v) Cellulase Onozuka 0.03% (w/v) Macerozyme 0.04% (w/v) Driselase Darkness, overnight	Liquid culture Darkness, 20°C	50 % initial division Plating not reported
Jeong et al., (2021) ²⁸⁶	1/2 MS medium, 1% sucrose Long-day (16 h light /8 h dark) ~ 7400 lux 22 - 23 °C 10 days	2% (v/v) Viscozyme L 1% (v/v) Celluclast 1.5 1% (v/v) Pectinex ultra Darkness for 12 h 50 rpm shaking	Alginate embedding Darkness, 25°C	15 % initial division 6 % plating
Xu et al., (2021) ¹¹⁰	1/2 MS medium, 1% sucrose Long-day (16 h light /8 h dark) 21°C 11 days	0.1% (w/v) Cellulase Onozuka 0.03% (w/v) Macerozyme 0.04% (w/v) Driselase Darkness, overnight (~ 12 h)	Alginate embedding Darkness, 25°C	58 % initial division 0.5 % plating

The main differences are in their plant growth, protoplasting, and regeneration media and conditions. Dovzhenko et al. (2003) does not mention regeneration temperature, and Chupeau et al. (2013) does not mention plating efficiency; however, they report 3% of microcalli-formed shoots.

Diurnal environmental changes, including photoperiod, light intensity, and thermoperiods, collaborate with the circadian rhythm to orchestrate a myriad of biological responses, including growth, development, and stress responses^{295,296}. A study utilizing the Vascular Cell Induction Culture

System Using Arabidopsis Leaves (VISUAL) demonstrated that circadian rhythms are disrupted in dedifferentiated cells, and re-establishing circadian rhythms is essential for vascular cell differentiation²⁹⁷. Exposure to random light/dark cycles was found to disrupt circadian rhythms and subsequently hinder vascular cell differentiation²⁹⁷. The significance of circadian rhythms in regeneration was further highlighted through JUMONJI C DOMAIN-CONTAINING PROTEIN 30 (JMJ30), a histone demethylase regulating circadian rhythm length²⁹⁸. JMJ30 emerged as a pivotal player in promoting callus formation by removing the repressive H3K9me3 mark on *LOB-DOMAIN PROTEIN GENE 16 (LBD16)* and *LBD29* promoters, thereby activating them²⁹⁹. Notably, *jmj30-1* mutants exhibited reduced callus formation capacity²⁹⁹. Additionally, *JMJ30* is temperature sensitive, with cold treatment reducing its expression and elevated temperatures enhancing its expression, establishing another link between environmental inputs and coordinated circadian rhythm³⁰⁰.

In summary, evidence indicates that environmental conditions exert control over regeneration capacity, exhibiting variations among regeneration modes, species, and even different ecotypes within the same species. Given the significance of protoplast regeneration in the generation of transgenic lines and potentially for bioproduction as guided differentiation platforms, addressing the reproducibility issue becomes imperative. So, this chapter aims to establish a protoplast regeneration platform, under well-documented environmental conditions spanning from nonregenerative to ultimately regenerative states. This also includes identifying physiological markers of protoplast regeneration and using the stable lines developed in the previous chapter to monitor stage-specific marker gene expression (Fig. 34). This will hopefully facilitate the unravelling of the reported reproducibility challenges, identification of factors influencing each stage of the regeneration and enhance the regeneration capacity of plants.

I began by adapting the protocol by Chupeau et al. (2013)¹⁰⁸ as the foundational method for this study, primarily because it is the only protocol employing liquid culture and was the most recent one available at the time. Liquid culture would have allowed the embedding of protoplasts and calli in different gels with different mechanical properties and was one of the initial research ideas. This study demonstrates that the cell wall digestion enzyme concentration significantly influences protoplast regeneration. Using a combination of 0.2% (w/v) Cellulase Onozuka R10, 0.06% (w/v) Macerozyme R10, and 0.08% (w/v) Driselase exclusively yielded budding cells with asymmetric division while halving the concentration resulted in symmetric cell division and multiple rounds of cell divisions. Altering plant growth conditions from short-day (Condition 1) to long-day with diurnal temperature

fluctuations (Condition 2) further enhanced culture regeneration, leading to the formation of microcalli. Validation of cell division and microcallus formation was conducted using marker lines generated in Chapter 4 (*WIND1*, *IAA8*, *WOX13:WIND1:WOX5*). qPCR analysis after protoplast isolation has revealed that both mesophyll cell identity markers and transcription factors associated with regeneration can be deregulated by plant growth and cell wall digestion. The findings from this study will contribute to reproducibility initiatives in protoplast regeneration, offer insights into plant regeneration within a broader context and help establish *A. thaliana* mesophyll protoplasts as a synthetic biology chassis.

Physiological and molecular markers of protoplast regeneration

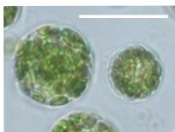
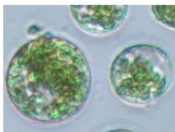
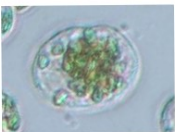
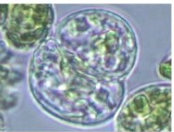
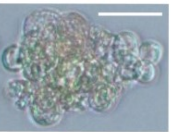
Day	0	1	2	4/6	7/ 11
Expected physiological changes					
Chloroplast autofluorescence	+	+	+		
Calcofluor-white signal	-	+	+	+	+
<i>WIND1</i>	+1.63				
<i>IAA8</i>	-2.93	+0.99	+0.70	+1.40	+0.83
<i>WOX13</i>	NA	NA	+0.87	+0.91	NA
<i>WOX5</i>				+1.72	-1.11

Figure 34. Physiological and molecular markers of protoplast regeneration. Chloroplast autofluorescence is used as a dedifferentiation marker, and after Day 2 it gets too dim to quantify. Calcofluor-white is used as a cell wall marker, useful for both in its absence on Day 0, and later for cell wall regeneration and remodelling. *WIND1*, *IAA8*, *WOX13*, *WOX5* transcription factors are chosen for their deregulated gene expression and their roles in plant development and regeneration, explained more in Chapter 4. Green colour shows upregulation and blue colour shows downregulation based on RNAseq data reported by Chupeau et al., (2013)¹⁰⁴. Changes in transcript amount is relative to the previous time point. For Day 0 it is relative to living plant. Besides that it is (Day 1/Day 0); (Day 2/Day 1); (Day 4/Day 2); (Day 7/Day 4). Scale bar shows 50 μ m.

5.2. Results

5.2.1. Physiological markers show consistent culture progression for protoplast regeneration

Developing a consistent protoplast regeneration protocol involves establishing a baseline that includes physiological markers to monitor culture progression. I followed Chupeau et al., 2013¹⁰⁸ to start protoplast cultures, which is described in Chapter 6. Plants grown *in vitro* for 18-21 days under Condition 1 (short day, 4000 lux, light spectrum 5200 K) exhibited 6-8 small leaves in growth chamber. To enhance the yield of isolated protoplasts, a 2X enzyme concentration (0.2% w/v Cellulase Onozuka R10, 0.06% w/v Macerozyme R10 and 0.08% w/v Driselase) was initially employed, proving effective for protoplast isolation for transient expression assays (Chapter 4). Under these conditions, the first physiological marker focus was on tracking cell wall regeneration using Calcofluor white stain which binds cellulose and chitin (Fig. 35a). Immediately after isolation (Day 0), there was no calcofluor white signal, with only cytoplasmic autofluorescence visible. Days 1 and 2 reveal partial cell wall regeneration, progressing to complete encirclement of the protoplast periphery with calcofluor signal on Day 3. By Day 6, the signal intensified, indicating ongoing cell wall deposition and a strong signal visible throughout the cell (Fig. 35a). The cell shape changes were noticeable from Day 1, coinciding with the onset of cell wall regeneration. Signal intensity was quantified using ImageJ to provide a more quantitative assessment (Fig. 35b), demonstrating a consistent trend across all three biological replicates. However, Day 6 exhibited a high standard deviation due to the presence of cells with strong signals, no signals, and partial signals within the culture (Fig. 35e).

The second physiological marker, chloroplast autofluorescence, was subsequently monitored. Chloroplasts emit strong autofluorescence at a wide range of wavelengths, between 520-710 nm, peaking at 680 nm³⁰¹. Through dedifferentiation, and potentially induced by culturing in darkness, chloroplasts undergo degradation, albeit partially, to preserve certain essential cellular functions¹⁰⁸. Therefore, chloroplast autofluorescence is expected to decrease and be a dedifferentiation marker. Strong chloroplast autofluorescence was observed immediately after isolation, sharply decreasing on Day 1 (Fig. 35c). Quantitative analysis confirms the steepest drop between Day 0 and Day 1, followed by a plateau in the signal (Fig. 35d). The microscope employed for this quantification has a 550 nm emission filter, which does not align with the peak emission of chloroplasts (680 nm). Consequently,

when chloroplast autofluorescence is weak, the resolution of the images drops, especially after Day 2.

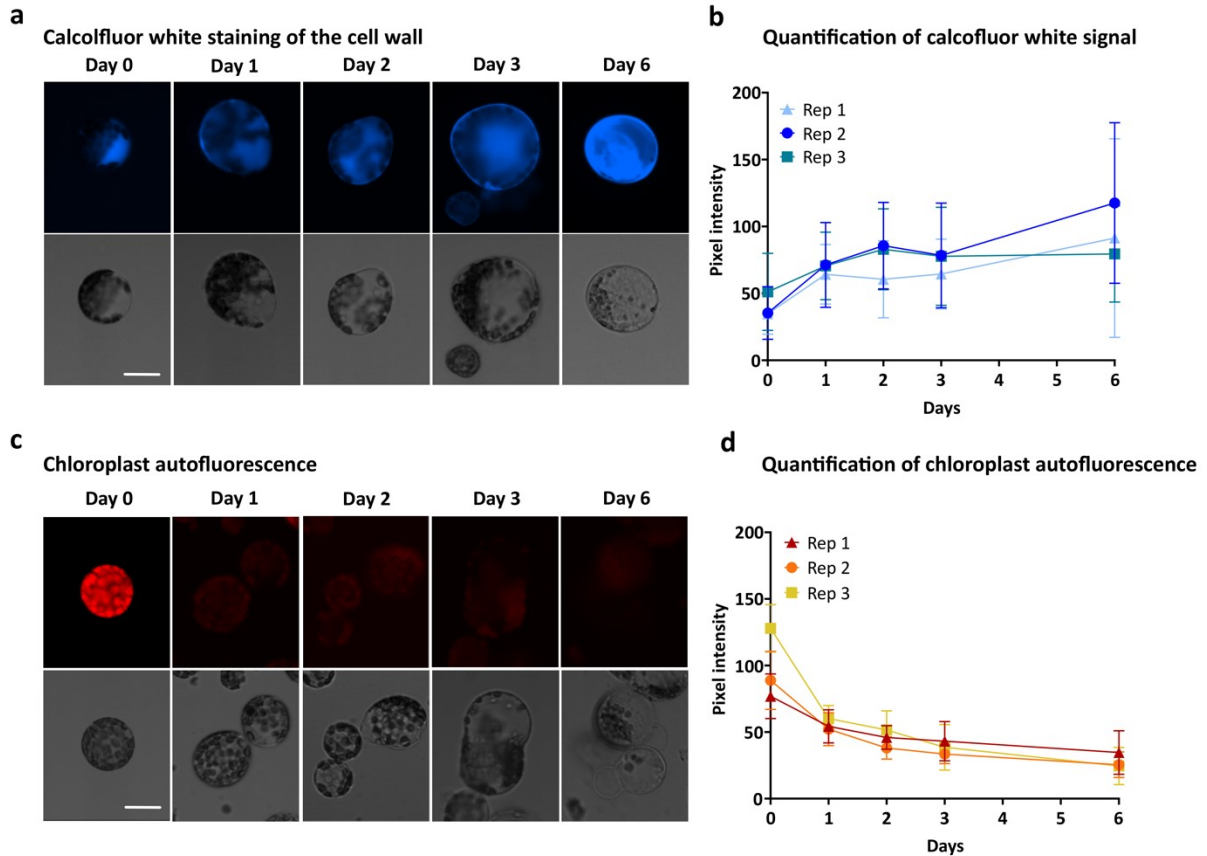


Figure 35. Physiological markers showing consistent of protoplast regeneration between biological replicates **a)** Calcofluor white staining of the cell wall on Days 0, 1, 2, 3, and 6. Blue filter images are presented on the top panel, and brightfield images are on the bottom panel. **b)** Quantification of calcofluor white signal with ImageJ. Continuous increase in the signal intensity is seen with a high standard deviation on Day 6. **c)** Change in chloroplast autofluorescence throughout the culture. The top panel shows images with a red filter, and the bottom panel shows bright field images. **d)** Quantification of chloroplast autofluorescence with ImageJ. A sharp decrease in signal is seen on Day 1. Images on Day 0-6 show different cells. Plants are grown in Condition 1. The scale bars show 25 μ m, and the error bars are mean $\pm$ SD.

It was interesting that cells with diverse shapes emerged from Day 1 onwards, which became more pronounced by Day 6 (Fig. 36). Examples include soft triangular cells, budding cells, elongated cells, and cells with incomplete and uneven cell walls. Cells with spherical or elliptical shapes tend to have even cell wall signal around the cell periphery (Fig. 35a) compared to cells with abnormal shapes such as budding or elongated cells (Fig. 36).

Different cell shapes observed on Day 6

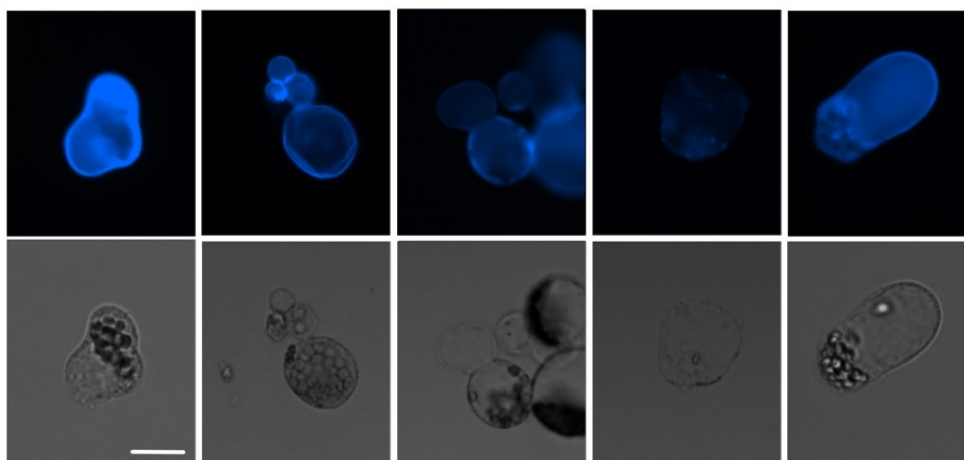


Figure 36. Examples of different cell shapes observed on Day 6 are shown with Calcofluor White staining. The scale bar shows 25 μm .

Cell viability was also monitored throughout the culture with fluorescein diacetate/propidium iodide (FDA/PI) dual staining. FDA is an acetylated derivative of the green fluorescent dye fluorescein, which is permeable to cells. Outside the cell, the acetyl groups render the dye non-fluorescent. Once the FDA enters the cytoplasm, nonspecific esterases remove the acetyl groups from the molecule, converting it to fluorescein, which emits a green signal³⁰². In contrast, PI is a cell-impermeant nucleic acid stain that emits red-range nuclear fluorescence. PI can penetrate cells only if the cell membrane is compromised (i.e., in dead cells)³⁰². Using FDA/PI dual staining for assessing viability may present limitations. Extracellular esterases can yield false FDA signals, and cell membrane damage does not invariably indicate cell death, leading to potential false PI signals³⁰². However, despite its limitations, dual staining remains the most widely used method for assessing cell viability in protoplast culture¹⁰⁸, and it is utilized in this study while considering these limitations. Around 80% of the protoplasts were alive on Day 0, which gradually decreased to around 50% by Day 6 (Fig. 37).

Overall, physiological markers (Calcofluor-White signal, chloroplast autofluorescence and viability staining) have shown that the biological replicates are consistent with each other.

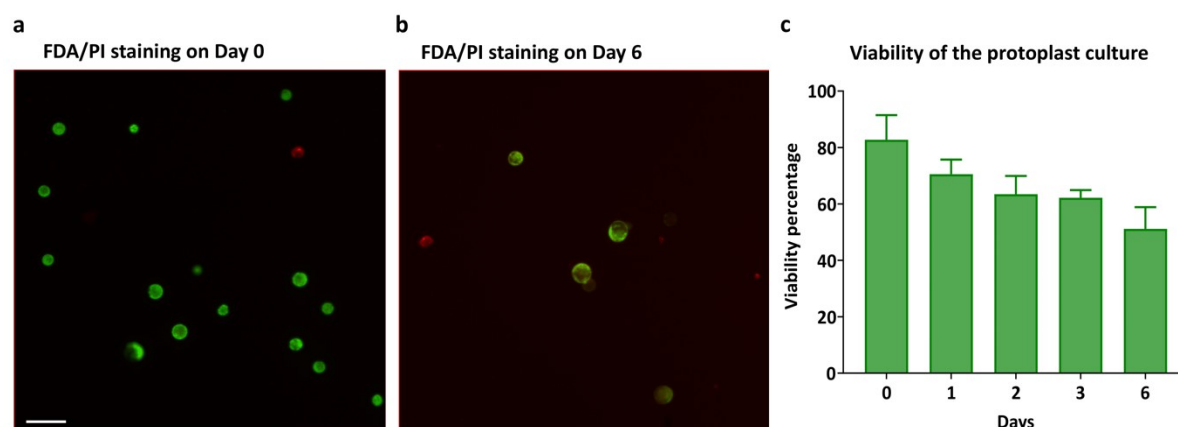


Figure 37. Viability of the protoplast culture with FDA/PI staining. **a)** Around 80% of protoplasts are alive on Day 0, after isolation. **b)** Cell viability decreases to around 50% by Day 6. **c)** Quantification of protoplast viability during culture progression. Green cells show FDA signal (alive) and dead cells show dead signal (red). The scale bars show 100 μm , and the error bars are mean $\pm$ SD.

5.2.2. Asymmetric and budding-like cell divisions are observed in the protoplast culture

The markers mentioned above, cell wall regeneration and chloroplast autofluorescence for dedifferentiation, are useful for the early stages of protoplast culture. However, a crucial phase in protoplast regeneration is cell division, occurring typically between Days 4 – 6¹⁰⁸. Through multiple rounds of cell division, protoplasts give rise to microcalli (minimum four cells, 70-500 μm diameter) and calli (larger than 500 μm diameter), as detailed in Chapters 1 and 4. Thus, the initial cell division is significant for the successful progression of regeneration.

Calcofluor White staining can serve as an indicator of cell division (via cell plate formation); however, its presence was not observed on Day 6 after staining (Fig. 35a,36). Therefore, timelapse microscopy was employed to investigate cell division. Notably, consistent budding-like cell division was observed on Day 5 (Fig. 38). However, this cell division exhibited uneven inheritance of chloroplasts and likely the cellular machinery and DNA between the two cells (Fig. 38a). Furthermore, frequent observations of budding-like protrusions were noted, raising uncertainty about whether these were indicative of cell division or cell growth, as brightfield images did not provide clear distinctions (Fig. 38b). Cells going through multiple rounds of cell division or microcallus was not observed under these conditions.

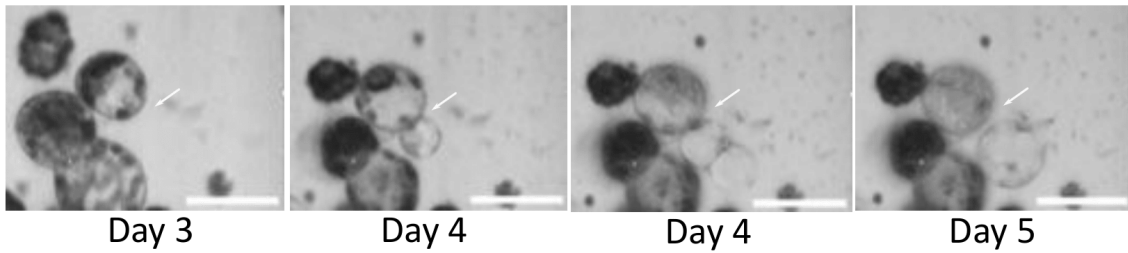
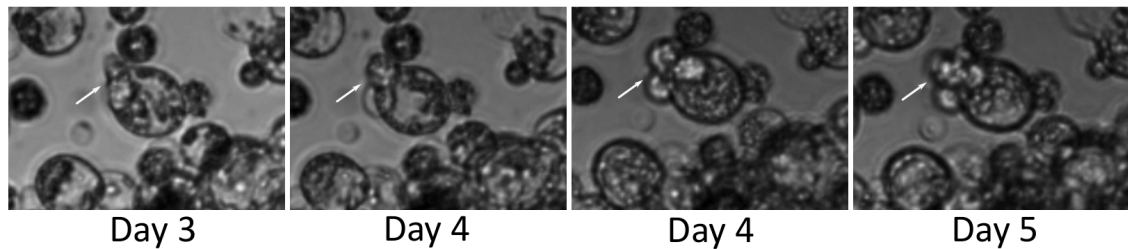
a**Budding-like cell division****b****Budding-like protrusions**

Figure 38. Timelapse images showing different cell division/growth patterns observed during protoplast culture. **a)** Budding-like cell division showing asymmetric inheritance of cell contents. **b)** Budding-like protrusions, which could be cell division or cell growth, were observed in the culture. Plants are grown in Condition 1 and the scale bar represents 50 μm .

5.2.3. Contamination in protoplast culture and its prevention

A challenge for establishing regenerating protoplast cultures was contamination, which was often noticeable from Day 3 onward. The source of the contamination was unclear, as plants were grown *in vitro*, all tools used were autoclaved, all media was filter sterilized and all handling was done under laminar flow hood. Initially, protoplast was cultured in 5 cm diameter borosilicate plates, one plate for each sampling day, sealed with surgical tape (3M). Approximately 300 μl was taken daily from a plate for imaging. However, this setup led to contamination, as evident in Fig. 6a, where no viable cells remained around Day 6.

To address this, Plant Preservative Mix (PPM, Plant Cell Technology), a broad-range biocide against microbes and fungi commonly used in tissue culture, was added to prevent contamination. While PPM successfully prevented contamination, it resulted in red autofluorescence from the protoplast nucleus and cell death starting from Day 1 (Fig. 39b). Subsequently, an alternative approach was attempted: culturing protoplasts in 1.5 ml Eppendorf tubes with 1 mL culture each, which was then used for imaging. Although contamination was not observed in the two trials, chloroplast

autofluorescence did not decrease as expected (Fig. 39c,d), and cell wall regeneration proved to be incomplete and asymmetric on Day 5 (Fig. 39e).

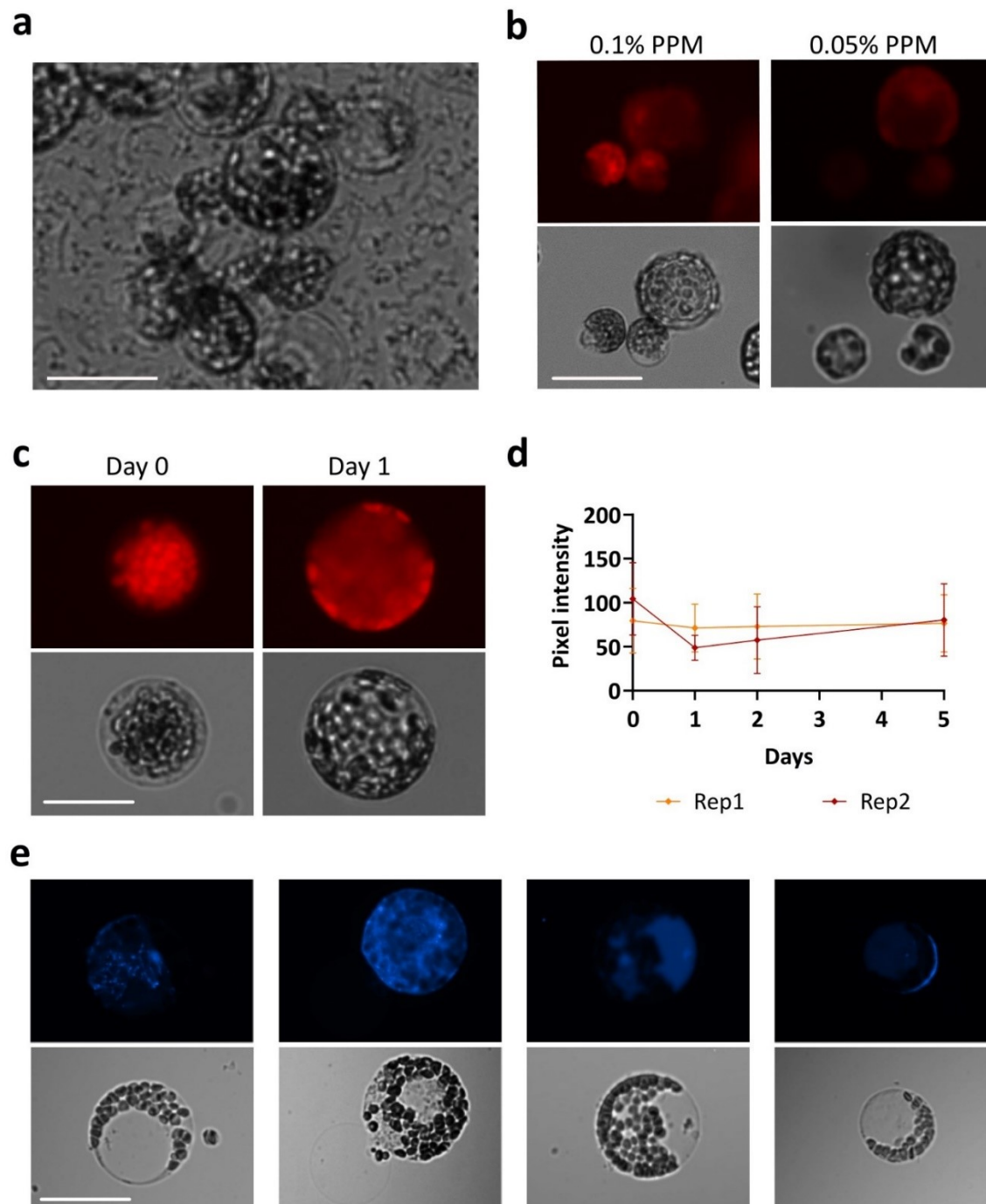


Figure 39. Efforts to reduce contamination in the protoplast culture. **a)** Example of contamination observed in the culture, whether bacterial or yeast, is unclear. **b)** Application of Plant Preservative Mix (PPM) to reduce contamination. PPM prevents contamination but causes death and nuclear autofluorescence in the red range

both at 0.1% (left panel) and 0.05% (right panel) v/v concentrations. Images taken on Day 1. **C)** Protoplasts cultured in Eppendorf do not show a decrease in autofluorescence between Day 0 and Day 1. **D)** quantified using ImageJ. **e.** Examples of incomplete and asymmetric cell wall of protoplasts cultured in Eppendorfs. Images were taken on Day 5. Plants for protoplasting are grown in Condition 1 and the scale bar represents 50 μm .

Contamination was effectively prevented by using 3.5 cm μ -dishes (ibidi, cat. no 81150) with a 4 mL culture volume, sealed using one layer of surgical tape and one layer of parafilm. Borosilicate plates (5 cm diameter) used previously have gaps between the lid and the plates, likely increasing the chances of getting contamination. On the other hand, the μ -dishes, have a more secure locking mechanism with no gaps between the plate and the lid. A dedicated dish was prepared for each sampling day to avoid unsealing the plates multiple times and decreasing the chances of contamination. Each tool used during protoplasting and culture is individually wrapped with aluminium foil and then put in an autoclavable box, which is again wrapped with aluminium foil.

5.2.4. Cell wall digestion enzyme concentration affects culture progression

When plants were grown under conditions reported by Chupeau et al., (2013)¹⁰⁸ (designated as Condition 1, Table 2), they were too small and difficult to handle. Therefore, a 2X concentration of cell wall digestion enzymes was employed to enhance the yield of protoplasts. Initially, no concerns were anticipated, considering that literature commonly reported using up to 10X concentration (Table 3). Considering the uneven cell wall regeneration observed around the periphery in budding-type cells (Fig. 36) and asymmetric pattern of cell division (Fig. 38), I concluded that gentler cell wall digestion conditions might be more suitable for protoplast regeneration. So, the enzyme concentration was adjusted back to the 1X concentration (0.1% w/v Cellulase Onozuka R10, 0.03% w/v Macerozyme R10, 0.04% w/v Driselase) as reported by Chupeau et al., 2013¹⁰⁸. Protoplast viability, initially around 75% after isolation, declined to approximately 35% by Day 6 (Fig. 40a).

To explore the potential impact of enzyme concentration on cell wall regeneration, the proportion of protoplasts with varying degrees of regeneration (none, partial, or complete) was quantified. Complete encirclement of Calcofluor White signal around the protoplast periphery was considered as complete regeneration (Fig. 37a, Day3 and Fig. 40e, Day 3), while any signal between no signal and complete was categorized as partial (Fig. 39e and Fig. 40e, Day 1). The analysis revealed that the proportion of protoplasts lacking a cell wall was lower with 2X enzyme concentration,

accompanied by a higher partial cell wall proportion with 2X enzymes on Days 1, 3, and 6 (Fig. 40b). However, the proportion of protoplasts with a complete cell wall was higher on Days 3 and 6 with 1X enzymes, almost 1.5 times that of the 2X enzyme condition (Fig. 40b). This indicates that the 1X condition resulted in better cell wall regeneration.

Furthermore, attention was also given to different cell shapes observed in the culture and their prevalence within the population. Cell shape categorization included spherical, elliptical, or budding forms. Spherical cells were identified when the circularity ratio ranged between 1 and 0.995, calculated using ImageJ, while elliptical cells had a circularity ratio lower than 0.995 (Fig. 40c). Budding cells were defined as those featuring buds or circular protrusions (Fig. 40c). The analysis of cell shape abundance revealed a gradual decrease in spherical cells, with elliptical cells constituting around 20% of the population, and budding cells prevailing as the majority from Day 4 onwards, comprising 45% of the population (Fig. 40d). Spherical and elliptical cell shapes typically exhibit more uniform cell wall deposition compared to budding-type cells (Fig. 40a, Fig 36). However, further quantification is required to validate this observation, and the uniform cell wall deposition could potentially be associated with symmetric cell division.

The improved setup revealed more cells undergoing symmetric cell division and multiple rounds of divisions for the first time (Fig. 40e). However, despite these improvements, microcallus formation was not observed under these conditions (Condition 1: short day, 21°C, 4000 lux).

5.2.5. Plants grown under fluctuating day/night temperature show symmetric cell division and microcallus formation

To enhance the progression of the culture to microcallus formation, a different plant growth condition was tried. This new condition, referred to as Condition 2, adopts a long-day setting with an average temperature of 20.2°C, decreasing the temperature in the night by approximately 5°C (Table 5). Light source used in Condition 2 is Philips Master (TL-D 36W/865) with 8400 lux intensity at 6500K, which is a colder-colour light source. Unlike Condition 1, which is a growth cabinet, Condition 2 operates in a growth room where temperature control is at the building level, leading to susceptibility to daily temperature variations. Intriguingly, despite these variations, successful observations of symmetric cell division (initial division rate approximately 5%) and microcallus formation have been achieved under Condition 2 (Fig. 41a).

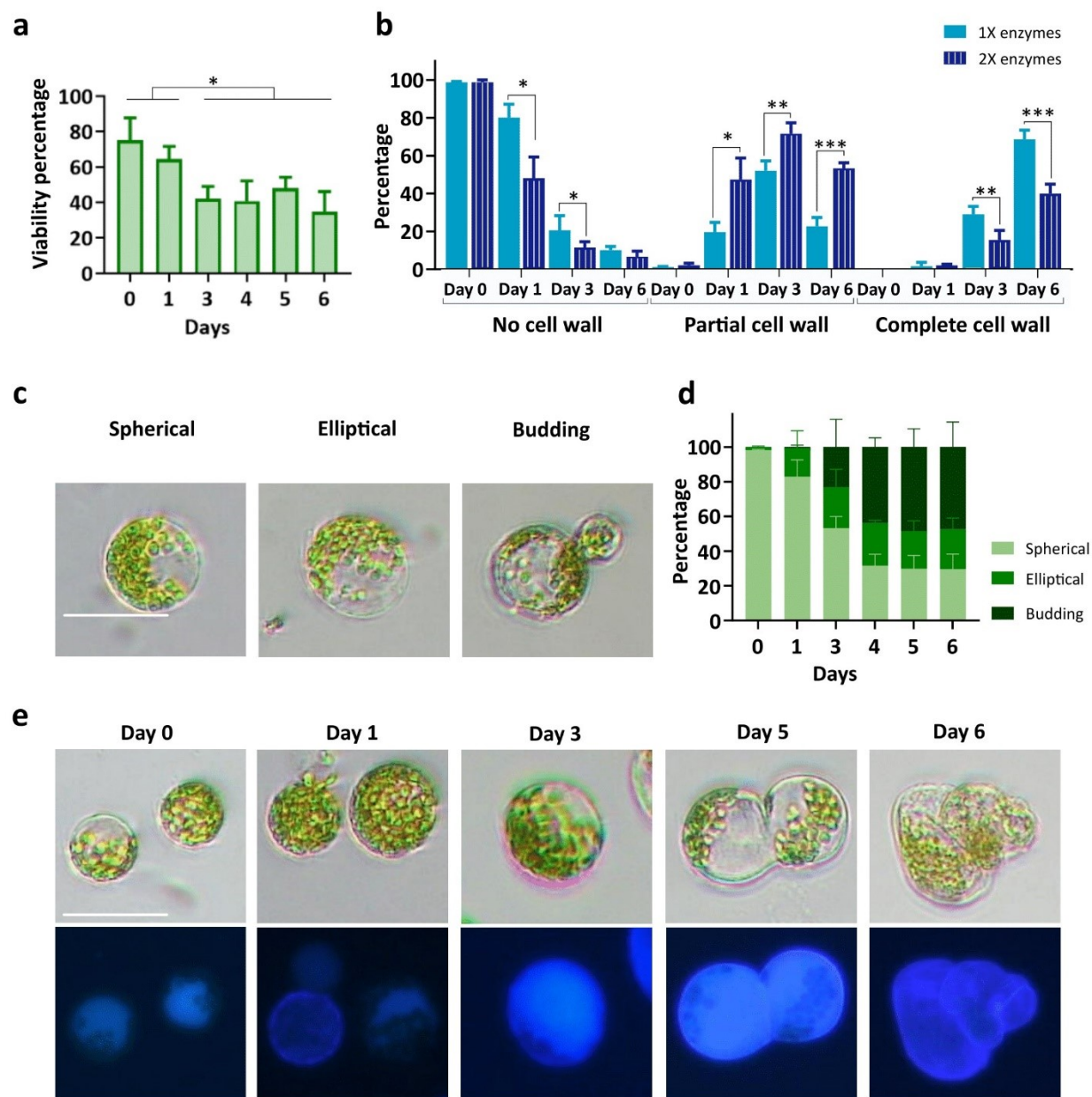


Figure 40. Protoplast culture with 1X enzymes. **a)** Protoplast viability, which is around 75% after isolation and goes down to around 35% by Day 6. One way ANOVA test was done to calculate statistical difference between the protoplast viability. **b)** Comparison of cell wall regeneration between 1X and 2X enzymes on Days 0, 1, 3, and 6. After isolation, protoplasts do not have a cell wall, which gradually regenerates. On Days 3 and 6, there are more protoplasts with complete cell walls when 1X enzymes are used compared to 2X enzymes. Light blue bars show 1X enzymes, while dark blue bars with stripes show 2X enzymes. Student's t test was used to calculate statistical significance between the two values. **c)** Protoplast cell shape categories. ImageJ was used to calculate circularity to distinguish spherical and elliptical cells. Protoplasts with circularity index between 1 and 0.995 are categorized as spherical (first image). Protoplasts with circularity index below 0.995 are categorized as elliptical (middle image). Protoplasts that have buds are categorized as budding (third image) **d)** After isolation, protoplast

is mostly spherical. As the culture progresses, elliptical and budding-shaped cells are observed. **e.** Images showing protoplast culture progression with 1X enzymes on Days 0, 1, 3, 5, and 6. Brightfield images are shown on the top panel and calcofluor staining is shown on the lower panel. Microscope images are taken with Nikon Ti2 Eclipse. Plants are grown in Condition 1, the scale bar represents 50 μm and the error bars show mean $\pm$ SD. Experiments in this figure were done by MEng student Tsung-Yen Lee (Johnsson) under my direct supervision.

Table 5. Plant growth conditions used in this study.

Condition	Photoperiod	Light source	Light intensity	Temperature	Humidity
#1	10 h light/ 14 h dark	Panasonic (FL40SS/ENW/37) 5200K	4000 lux	21.3°C	53%
#2	16 h light / 8 h dark	Philips, TL- (D/36W/865) 6500 K	8400 lux	20.2 $\pm$ 5.0 °C	44%
#3	16 h light / 8 h dark	Panasonic (FL40SS/ENW/37) 5200K	9810 lux	21.3°C	53%
#4	16 h light / 8 h dark	Philips, TL- (D/36W/865) 6500 K	7620 lux	21.2°C	51%

Upon staining protoplasts with calcofluor, cell plate formation is evident under a blue filter, in conjunction with brightfield images on Day 6 (Fig. 41b). Furthermore, calcofluor and FDA/PI staining revealed that microcallus was formed resulting from multiple rounds of cell division with most cells remaining viable in the microcallus (Fig. 41c).

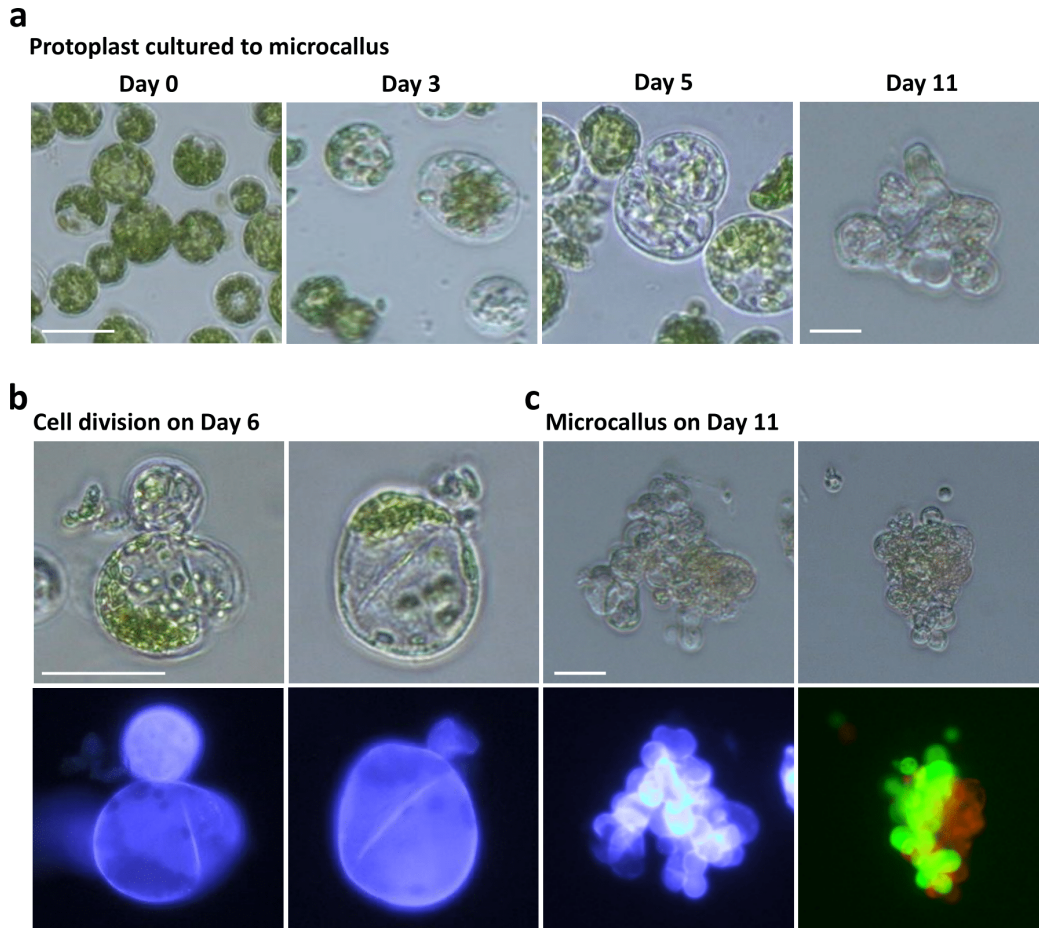


Figure 41. Protoplast culture progression for plants grown in Condition 2. **a)** Brightfield images showing cell shape changes between Day 0 and 3, symmetric cell division on Day 5, and microcallus formation on Day 11. **b)** Calcofluor staining showing cell plate formation during cell division on Day 6. **c)** Microcallus observed on Day 11. Calcofluor staining is on the left, and FDA/PI staining for viability is on the right. Images are taken with Nikon Ti2 Eclipse. The scale bars represent 50 μm .

5.2.6. Marker lines confirm cell division and microcallus formation

While cell plate formation was observed through brightfield and calcofluor white staining, definitive confirmation of cell division was achieved with nuclear marker lines. To achieve this, the marker lines developed in Chapter 4 were used. Under Condition 2, where cell wall digestion utilized 1X enzymes, timelapse images of the *P_{WIND1}:sfGFP_VirD2:FAD2* marker (expresses nuclear localised *sfGFP* under *WIND1* promoter) revealed two cell divisions on both Day 6 and Day 7 (Fig. 42a). Similarly, timelapse images of *P_{IAA8}:mScarlet-I_SV40:HSP* (expresses nuclear localised *mScarlet-I* under *IAA8*

promoter) showed two cell divisions (Fig. 42b). Notably, in both timelapses (P_{WIND1} and P_{IAA8} marker lines), just before cell division, the nuclear signal disappeared, and it reappeared as two nuclei. Distinguishing cell boundaries from the brightfield images were challenging due to the limited resolution of the Nikon Ti2 inverted microscope used for time-lapse imaging. Nevertheless, both of these lines validate the presence of *WIND1* and *IAA8* promoter activity during initial cell division, suggesting potential roles for these transcription factors during initial cell division.

To overcome the resolution limitations of Nikon Ti2, the Confocal SP8 microscope was utilized for the tricolour marker line (*WOX13:WIND1:WOX5:HYG*) on Day 6. Using the tricolour marker line offers advantages over single-colour markers since it enables the simultaneous screening of three colours. This is particularly beneficial when the roles of transcription factors in specific processes are unknown, streamlining the screening process. Strong nuclear signals were observed for *WOX13* (red) and *WIND1* (green), while the *WOX5* (blue) nuclear signal was absent (Fig. 42c). Unfortunately, due to the limited availability of the confocal microscope and the confirmation of the tricolour marker line towards the final months of my PhD research, further timelapse imaging could not be conducted for the tricolour line.

In addition to cell division, the $P_{WIND1}:sfGFP_VirD2:FAD2$ marker line was employed to confirm *WIND1* expression in a microcallus comprising at least six cells (Fig. 42d). This suggests that *WIND1* likely has roles in microcallus formation and it could be used as a potential genetic target for engineering regeneration.

5.2.7. Cell wall digestion enzymes and plant growth conditions deregulate marker genes

Despite the improvement in protoplast regeneration, growing callus from microcallus was challenging. I attempted to replicate the protocol outlined by Chupeau et al. (2013)¹⁰⁸, increasing the initial protoplast plating concentration twofold and fourfold. Additionally, I experimented with doubling the initial auxin in PIM and doubling the cytokinin concentrations in CIM. My rationale was that while plating efficiency remained consistent, increasing the plating protoplast concentration or the auxin in PIM, and cytokinin in CIM might increase the number of microcalli and subsequently calli. However, despite these adjustments, microcallus failed to progress into callus (data not shown). Furthermore, contamination was a recurrent issue observed during cell culturing on 9 cm borosilicate plates. Larger plates, instead of μ -dishes, were needed to accommodate the larger volume of cultures resulting from the transition from PIM to CIM.

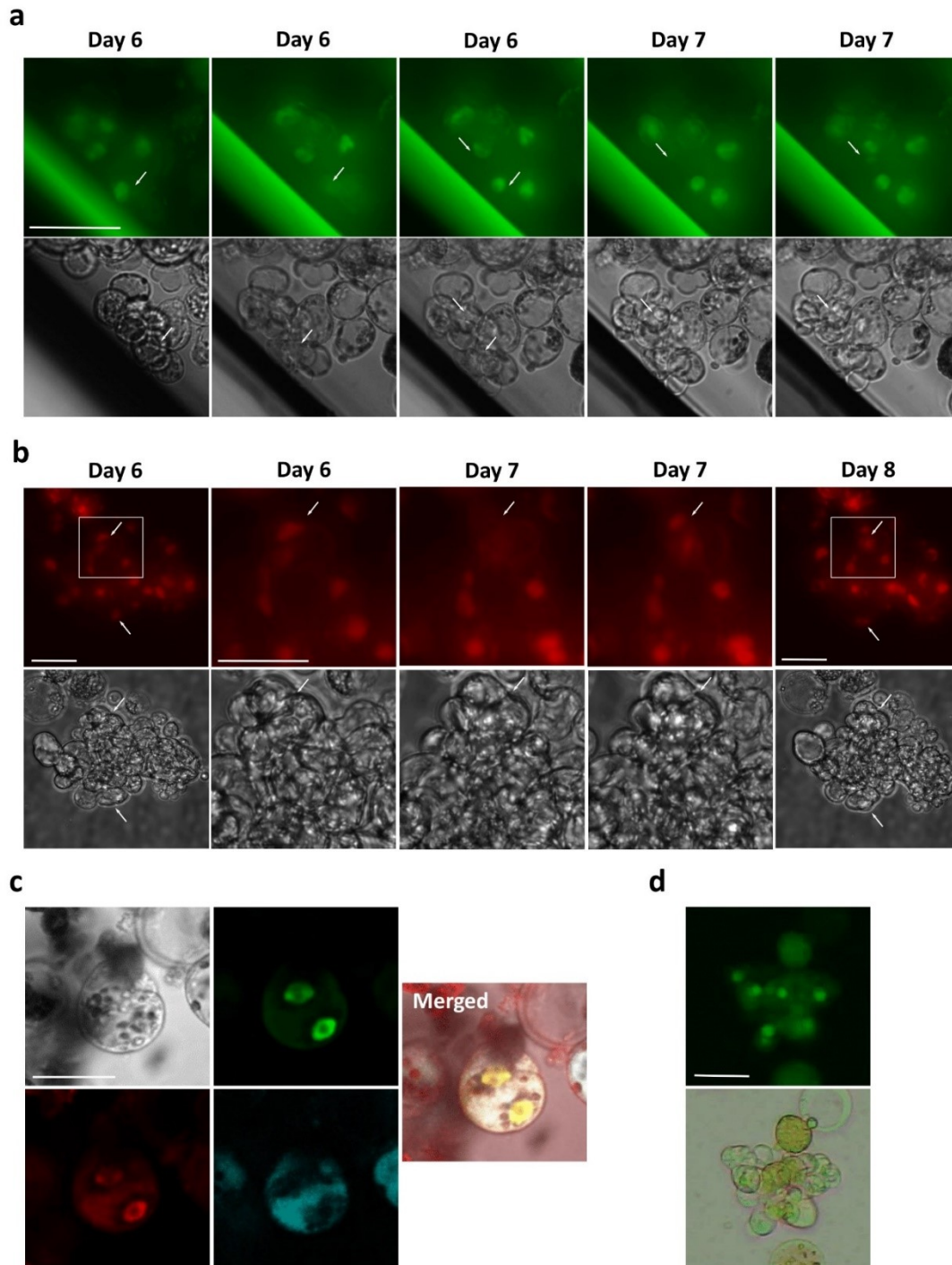


Figure 42. Marker lines showing cell division. **a)** *P_{WIND1}:sfGFP_VirD2:FAD2* marker line developed in this study showing cell division on Day 6 and 7. **b)** *P_{IAA8}:mScarlet-I_SV40:HSP* marker lines showing cell division. The first and last images show the complete microcallus cluster that is undergoing cell division. While panels in the middle are zoomed in images of the same microcallus to highlight cell division on Day 7. **c)** *P_{WOX13}:WIND1:WOX5:HYG* marker lines showing cell division on Day 6. The yellow colour indicates a merged green and red nuclear signal. **d)** Microcallus observed from *P_{WIND1}:sfGFP_VirD2:FAD2* marker line on Day 11. Images on a and b are taken with

Nikon Ti2, images on panel c are taken with SP8 confocal microscope and panel d is taken with Nikon Ti2 Eclipse. All marker lines are developed in this study and details are available in Chapter 4. Scale bars show 50 μm .

Table 6. Mesophyll identity markers chosen for this study. Phosphoenolpyruvate (PPT); Nitric oxide (NOX).

Gene	Acronym	Role	Expression	Reference
<i>LIGHT-HARVESTING CHLOROPHYLL B-BINDING 2</i>	<i>LHCB2.1</i>	Part of light harvesting complex and its phosphorylation drives state transitions during photosynthesis	Chloroplast, leaf mesophyll cell	Pietrzykowska et al., (2014) ²⁹⁹ Sawchuk et al., (2008) ³⁰⁰
<i>FERREDOXIN-NADP OXIDOREDUCTASE 1</i>	<i>FNR1</i>	NADPH dehydrogenase activity and electron transporter for photosynthesis	Chloroplast, leaf, stem, flower	Lintala et al., (2009) ³⁰¹
<i>BETA CARBONIC ANHYDRASE 2</i>	<i>CA2</i>	Zinc metalloenzymes interconvert CO_2 and HCO_3^-	Leaf	Di Mario et al., (2016) ³⁰²
<i>CAB UNDEREXPRESSED 1</i>	<i>CUE1</i>	Encodes PPT and phosphate translocator through plastids, NOX homeostasis	Chloroplast, golgi apparatus, mesophyll cells (leaf) and root	Li et al., (1995) ³⁰³ Lechon et al., (2020) ³⁰⁴

Different digestion mixtures were tested since the regeneration capacity is influenced by the concentration of cell wall digestion enzymes. These included a control with no enzymes ('No enzyme'), 0.1% Cellulase Onozuka R10 ('Cellulase'), and 0.1% Cellulase Onozuka combined with 0.06% Pectinase (Sigma, cat. no P2401-500U) named 'Cel+Pec', as well as a mixture of 1X of all enzymes (0.1% Cellulase Onozuka R10, 0.03% Macerozyme, 0.04% Driselase) named 'Protoplast' (Fig. 43h). All the treatments included making small incisions on the harvested leaves and incubating the leaves in the indicated solutions overnight in darkness. No enzyme samples did not yield protoplasts, as expected, and interestingly, neither did Cellulase only samples. Cel+Pec and Protoplast samples did yield fully released protoplasts. Overall, no discernible trend was found when increasing cell wall digestion enzyme components (Fig. 43). The addition of any cell wall digesting enzyme significantly reduced

LHCB2.1 transcript levels compared to the No enzyme samples for both Condition 1 and 2 (Fig. 43a). Which might indicate that *LHCB2.1* expression levels change with cellular reprogramming due to cell wall damage or modifications. Regarding *FNR1* and *CA2*, wounding and darkness appeared to be triggers for decreased transcript amounts, with no clear change in transcript levels with the addition of cell wall digestion enzymes (Fig. 43b,c). Cel+Pec and Protoplast samples exhibited higher *CUE1* than the other samples under Condition 1 (Fig. 43d), meaning that *CUE1* could be involved in cellular reprogramming in response to cell wall digestion.

Regarding transcription factors, *WOX5* is not detectable in the leaves of a living plant; therefore, it is not considered in this section. Under Condition 1, the *WIND1* transcript in protoplast samples shows a 2.4-fold upregulation compared to the living plant, representing the treatment with the lowest *WIND1* transcript (Fig. 43e). *IAA8* transcripts are consistently downregulated by approximately 5-fold for all samples, including protoplasts (Fig. 43f). While Chupeau et al., (2013)¹⁰⁸ did not detect *WOX13* transcript, my experiment resulted in increases of around 5-fold and 4-fold in transcript levels for No enzyme and Cellulase; while there is a 2-fold increase for Cel+Pec and protoplast samples under Condition 1, respectively (Fig. 43g). This could be a result of using different plant growth conditions affecting plant hormone balance (auxin/cytokinin) and therefore leading to different *WOX13* transcript levels.

When comparing RNA isolated from Condition 1 (characterised by asymmetric cell division and no microcallus) to Condition 2 (marked by symmetric cell division and microcallus formation), distinct expression patterns were observed. Details of the plant growth conditions 1 and 2 are available in Table 4. *LHCB2.1* expression did not show significant differences between the two conditions (Fig. 43a). However, *FNR1* expression was lower, while *CA2* and *CUE1* expression were higher in Condition 2 (Fig. 43b-d). Notably, *WIND1* expression was approximately threefold higher for protoplasts under Condition 2 than Condition 1, reaching 4-fold relative to the living plant, while the rest of the cell wall digestion treatments reached around 20-fold change (Fig. 43e). On the other hand, *IAA8* expression was roughly doubled, reaching 0.5-fold relative to the living plant (Fig. 43f). *WOX13* expression did not exhibit significant differences (Fig. 43g). This suggests that plant growth conditions can deregulate mesophyll identity markers and transcription factors involved in regeneration, each in distinct mechanisms.

These findings underscore the impact of cell wall digesting enzyme concentrations and plant growth conditions in dedifferentiation by deregulating mesophyll cell identity marker genes and the chosen transcription factors.

5.2.8. Plant growth conditions deregulate marker genes on Day 0

After observing the deregulation of mesophyll marker genes and selected transcription factors due to plant growth conditions, a more controlled experiment was conducted to compare three distinct growth conditions. Condition 2, characterised as the regenerative condition, remained unchanged, while Conditions 3 and 4 were introduced to replicate the specific conditions of Condition 2 in the well-controlled environments (cabinets) (Table 4). Condition 2 is not well-controlled and is subjected to diurnal temperature fluctuations beyond our influence, potentially posing reproducibility challenges in the future. So, the aim was to try to reproduce to regenerative plant growth conditions (Condition 2), which would help identify factors important for protoplast regeneration. It is hypothesized that the diurnal temperature fluctuations and/or the light intensity may have regeneration enhancing effects. So, Condition 3 featured higher light intensity (9810 lux) from a warmer-colour light source (Panasonic FL40SS/ENW/37 light source, 5200K) and Condition 4 had slightly lower light intensity (7620 lux) from a cooler-colour light source (Philips, TL-D/36W/865, 6500 K). While the temperature remained constant in Conditions 3 and 4 (Table 4, Fig. 44a). Plants grown under Condition 2 and 4 were larger than plants grown under Condition 3, suggesting that the cooler-colour light source (Philips, TL-D/36W/865, 6500 K) could be a preferable choice. However, no cell division or microcallus formation was observed under Conditions 3 and 4.

The expression levels of mesophyll marker genes *LHCB2.1* and *FNR1* were nearly half in Conditions 3 and 4 compared to Condition 2 (Fig. 44b,c). *CA2* expression was approximately 50% lower in Condition 2 than in any other conditions (Fig. 44d). However, there was no statistically significant difference in *CUE1* levels (Fig. 44e). Regarding the transcription factors, *WIND1* levels in Condition 2 were nearly double those in Condition 4 (Fig. 44f), while *IAA8* levels were roughly half (Fig. 44g). *WOX13* levels did not exhibit significant differences (Fig. 44h). These results indicated that the daily temperature fluctuations during plant growth contribute to higher regeneration competence. To establish a link between daily temperature fluctuations and regeneration, *JMJ30* (a temperature-sensitive histone demethylase involved in circadian rhythm and regeneration) transcript levels were examined. Indeed, *JMJ30* levels under Condition 2 were half of those under Conditions 3 and 4 (Fig. 44i), suggesting *JMJ30* has a role in protoplast regeneration.

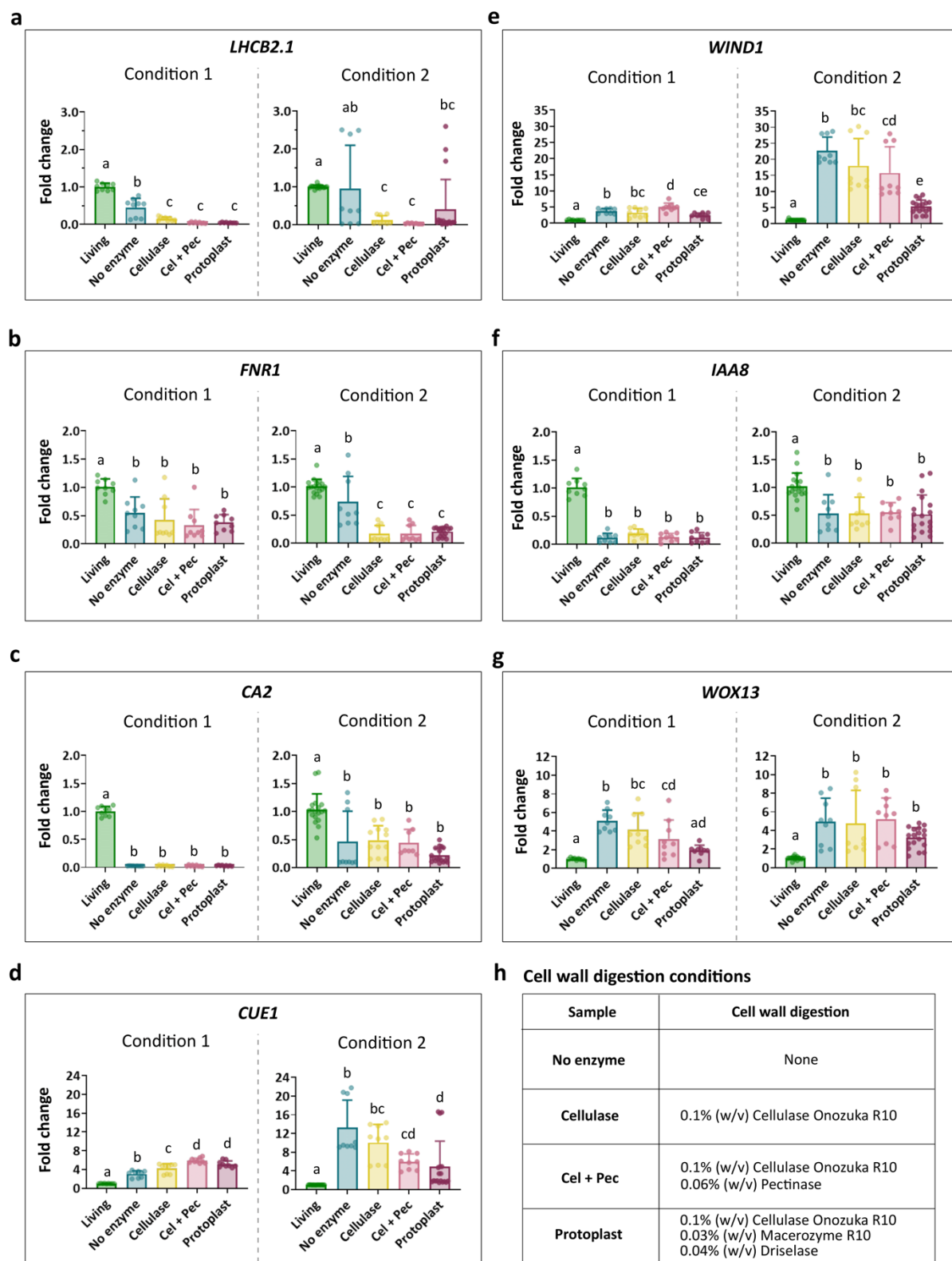


Figure 43. Gene expression analysis of isolated protoplasts (Day 0, dedifferentiation) from plants grown in Condition 1 and 2. **a)** *LHCB2.1* expression, **b)** *FNR1* expression, **c)** *CA2* expression, **d)** *CUE1* expression, **e)** *WIND1* expression, **f)** *IAA8* expression, **g)** *WOX13* expression, **h)** Description of cell wall digestion treatments. Panels

show a-d mesophyll cell fate marker genes, whereas e-g are transcription factors chosen in Chapter 4. Fold change is calculated by the $\Delta\Delta C_T$ method and extracting RNA from the leaves from a living plant the same age as other samples. No enzyme indicates leaves that are cut and placed in MGG solution. The cellulase sample only has 0.1% Cellulase Onozuka R10. The Cel + Pec sample has 0.1% Cellulase Onozuka R10 and 0.06% pectinase from *Rhizopus* species. The protoplast sample is 1X enzymes (0.1% Cellulase Onozuka R10, 0.03% Macerozyme and 0.04% Driselase. One-way ANOVA with Tukey's HSD test was done to show statistical significance and the error bars show mean $\pm$ SD.

Comparing three long-day conditions, including a cool-coloured light source (Condition 2 and 4) and a warmer-coloured light source at higher intensity (Condition 3) for protoplast samples at Day 0, the mesophyll identity markers were generally lowest in Condition 2, with the exception of *CUE1* expression (Fig. 11b-e). *CUE1* plays a role in maintaining NOX homeostasis during root growth, and elevated *CUE1* levels could potentially contribute to NOX homeostasis for protoplasts under Condition 2 (Fig. 44j)³⁰⁸. Surprisingly, *WIND1* expression was highest in Condition 2, and Conditions 2 and 3 exhibited similar levels of *IAA8* (Fig. 44f,g). Despite Conditions 2 and 4 being the most similar, Condition 4 displayed double the *IAA8* expression. This data implies that beyond light source and intensity, the temperature fluctuations in Condition 2 possibly induce both transcriptional and hormonal changes

Moreover, intracellular nitric oxide (NOX) levels were assessed using 4,5-diaminofluorescein-2 diacetate (DAF-2, A) stain. NOX can be indicator of cellular stress and is a regulator of plant development³⁰⁹. Additionally, *CUE1*, which exhibits deregulation between Condition 1 and 2 (Fig. 33d) but not between 2-4 (Fig. 44i), is implicated in NOX homeostasis. Although there was no statistically significant difference in intracellular NOX levels, the standard deviation was smallest under Condition 2, while it was the highest under Condition 4 (Fig. 44j). Unfortunately regeneration assays could not be conducted concurrently with qPCR assays due to the substantial quantity of protoplasts required for high-quality RNA extraction. Nevertheless, these results support that in Condition 2, where there are daily temperature fluctuations, a greater number of protoplasts might be maintained in a more homogenous state for regeneration.

a**Plant growth conditions**

Condition	Photoperiod	Light source	Light intensity	Temperature	Humidity
2	16 h light	6500 K	8400 lux	20.2 ± 5.0 °C	44%
3	16 h light	5200 K	9810 lux	21.3 °C	53%
4	16 h light	6500 K	7620 lux	21.2 °C	51%

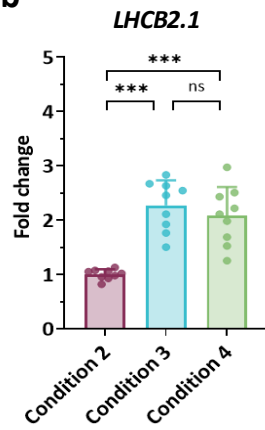
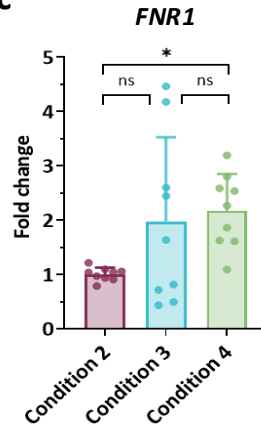
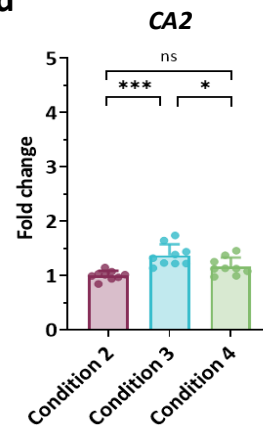
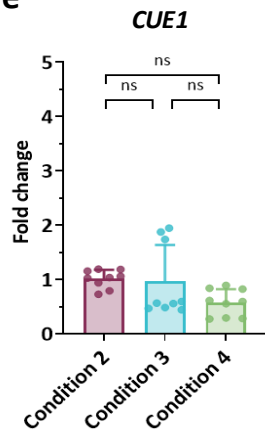
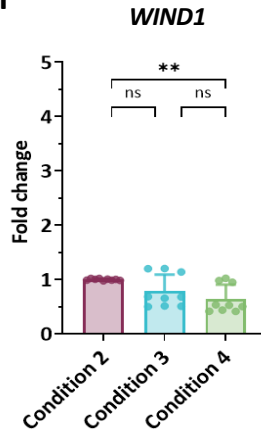
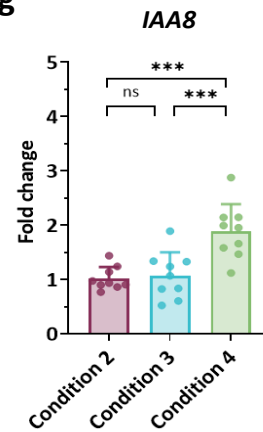
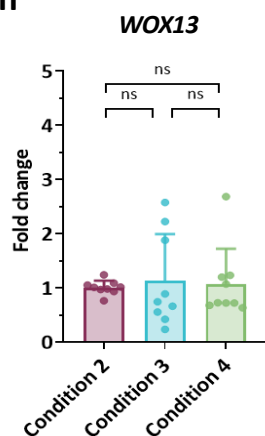
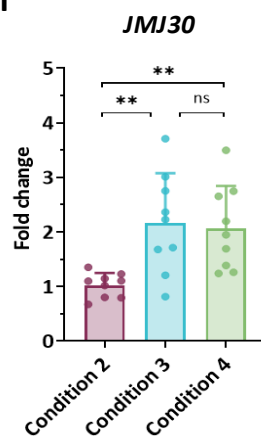
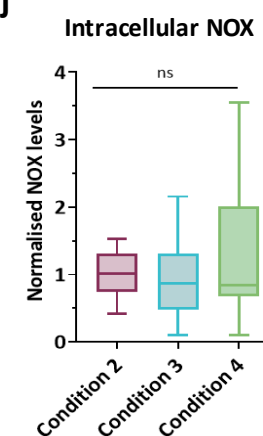
b**c****d****e****f****g****h****i****j**

Figure 44. Gene expression analysis of isolated protoplasts (Day 0) from plants grown in Conditions 2, 3 and 4. All conditions are long day, with changing light source, intensity and temperature. Condition 2 is a growth room, where Conditions 3 and 4 are growth chambers. **a)** Description of plant growth conditions 1 and 2. All conditions are long day, with changing light intensity and temperature. **b)** *LHCB2.1* expression, **c)** *FNR1* expression, **d)** *CA2* expression, **e)** *CUE1* expression, **f)** *WIND1* expression, **g)** *IAA8* expression, **h)** *WOX13* expression, **i)** *JMJ30* expression. **j)** Intracellular NOX measurements with DAF-2A dye. One-way ANOVA with Tukey's HSD test was done to determine statistical significance. The error bars show mean $\pm$ SD.

5.2.9. Preliminary data shows day/night temperature fluctuations deregulate *WOX5* expression

In response to the previous data suggesting the potential role of daily temperature fluctuations in regeneration competence, a controlled experiment was designed. Due to constraints in available plant growth space that could be set for varied environmental conditions, a single incubator with heating pads was used (Fig. 45a). Plants grown on heating pads maintained a constant temperature of 24°C (Condition 5, Cons), while those on racks in the incubator experienced a day-time temperature of 24°C and a night-time temperature of 18°C (Condition 6, Fluct, Fig. 45b). Acknowledging that short-day conditions were reported to be optimal for protoplast regeneration, the final trial was conducted under short-day conditions, in contrast to the previous long-day conditions, as a last attempt for successful regeneration.

Additionally, due to the significant condensation of plants on heating pads in the first trial, I positioned them on the vertical axes. Care was taken to avoid sowing seeds at the bottom of the plates, where water tends to accumulate. Regular removal of condensed water was carried out on a weekly basis. The limited surface area of heating pads restricted the simultaneous growth of only 8 plates, consequently limiting the number of protoplasts that could be isolated. Notably, short-day plants being smaller than long-day plants added another limitation, reducing the available tissue for protoplasting and, consequently, the number of isolated protoplasts. As a result, only enough RNA could be isolated for qPCR, with no protoplasts remaining for further culturing.

Preliminary findings regarding *WIND1* and *JMJ30* expression do not show a significant difference between constant and fluctuating temperatures (Fig. 45c,d). However, *WIND1* expression nearly doubles, and *JMJ30* expression almost triples on Day 5 compared to Day 0 (Fig. 45c,d). For further analysis, *WOX5* expression was included in this investigation, given its presence in protoplast samples while being undetected in leaf samples. The fold change comparison in this section was conducted by using protoplasts grown under constant temperature as the reference. The results revealed that fluctuating temperature samples exhibit nearly double the *WOX5* transcript levels of constant

temperature samples on both days (Fig. 45e). Furthermore, *WOX5* transcript is upregulated 20 times on Day 5 compared to Day 0 for both sets of samples (Fig. 45e), suggesting upregulation of *WOX5* may underpin the enhanced regeneration in the temperature fluctuating condition

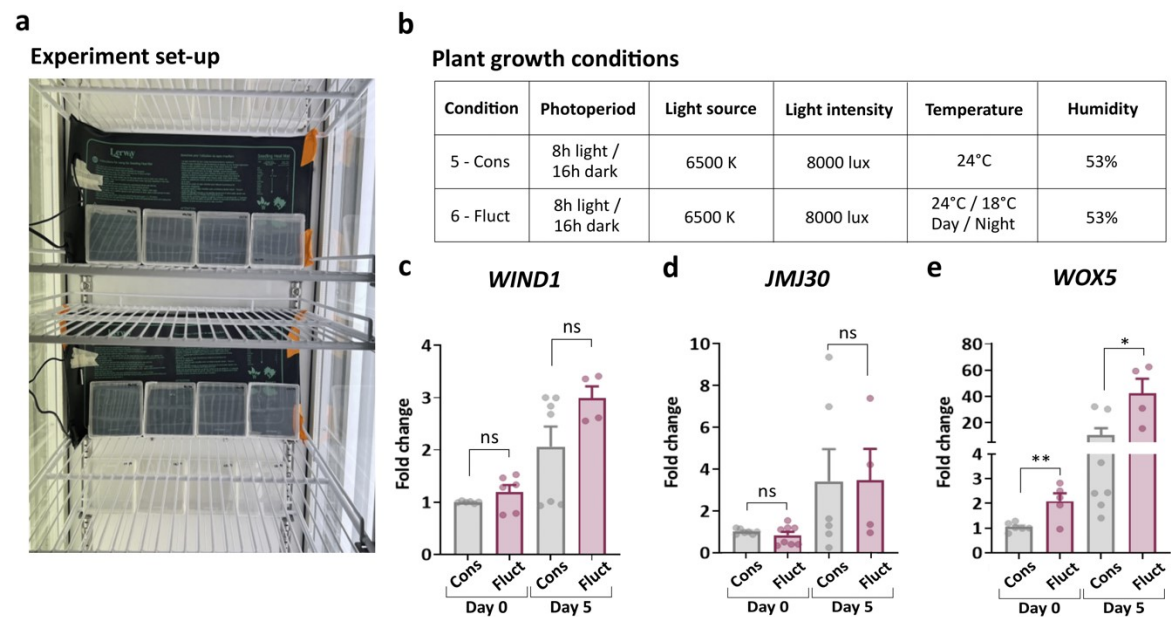


Figure 45. Preliminary data showing the effect of day and night temperature fluctuations on transcription factors on Day 0 and 5. Regeneration temperature of both conditions is 24°C. **a)** Picture of the experiment set-up. Plants on the heating pads are grown under constant 24°C (Cons) and plants in the incubators experience day/night temperature fluctuations (Fluct). **b)** Details of the plant growth conditions. **c)** *WIND1* expression, **d)** *JMJ30* expression **e.** *WOX5* expression. Fluct samples on Day 5 only have 2 biological replicate due to low quality of RNA extraction, while all other samples have 3 biological replicates. Student’s T- test was done to determine statistical significance between Cons and Fluct samples on Day 0 or 5. The error bars show mean ± SD.

5.3. Discussion

Protoplast regeneration, essential for applications in transgenics, agriculture, and developmental biology, faces challenges in terms of reproducibility¹²⁹. Each laboratory venturing into protoplast regeneration often encounters the need to optimise and adapt existing protocols due to difficulties in achieving successful regeneration^{108,114,287–290}.

5.3.1. Monitoring protoplast culture progression

Chupeau et al., (2013) method¹⁰⁸ was selected to establish a protoplast regeneration platform in our laboratory, and the initial step was to replicate this method with growth conditions designated as

Condition 1. The main goal of this study was to establish consistent culture physiology across each batch and unravel the reported variability in protoplast regeneration. Consistency of the culture was characterised by calcofluor staining for cell wall regeneration and tracking chloroplast autofluorescence (Fig. 35). Achieving consistency and reproducibility is crucial for using protoplasts as a chassis for research and innovation. While visual examination of images appeared consistent between Days 0-6 in each batch, a more precise analysis was performed using ImageJ to quantify cell wall and chloroplast signals (Fig. 35b,d), revealing consistent cell behaviours in my experiments. Uneven cell wall deposition is also frequently observed, likely due to higher cell wall digestion enzyme concentration (2X) used during protoplasting observed (Fig. 36). This might have been too harsh for the cells, causing membrane or cellular damage and preventing uniform cell wall regeneration. Additionally, contamination (Fig. 36) and asymmetric cell division (Fig. 38) were common issues observed in protoplast culture. Both using low concentrations of PPM and the use of Eppendorf tubes for cell culture, with a volume of 1 mL, did not yield successful regeneration (Fig. 39c,d). PPM can cause growth defects in *A. thaliana* seedlings³¹⁰, raising the possibility of similar effects on protoplasts. As for Eppendorfs, the outcome could be due to either the plastic surface of the Eppendorf tubes or the small culture volume, as volumes below 4 ml were observed to be detrimental²⁸⁹. A set-up employing a 4 ml volume in μ -dishes or a 10 ml volume in borosilicate plates, each covered with a layer of surgical tape and a layer of parafilm, demonstrated minimal occurrences of contamination. This set-up was utilized in the subsequent experiments.

5.3.2. Enabling protoplasts to divide symmetrically and form microcallus

Cell wall digestion concentration and plant growth conditions were modified to address the asymmetric cell division observed in the culture. Cell wall digestion enzyme concentration was reduced from 2X to 1X, which was initially increased to enhance the release of protoplasts from the leaf. This indeed resulted in protoplasts showing symmetrical division, as evidenced by calcofluor white staining of the cell wall and cell plate (Fig. 38). Incubation in 2X concentration might be too harsh on protoplasts, while 1X could be gentler. Additionally, the duration of incubation was limited to a maximum of 14 h, as longer digestion duration can also produce more fragile protoplasts.

Despite observing multiple rounds of cell division, microcallus formation was absent on Day 11. Consequently, an alternative plant growth condition, namely Condition 2, was explored (Table 5). Condition 2 experiences daily temperature variations of $20 \pm 5^\circ\text{C}$, occasional daytime overheating due

to building-level heating issues and features a cooler-coloured light source with a long-day photoperiod. When cultivated under Condition 2 with 1X enzymes, protoplasts demonstrated symmetrical division and exhibited microcallus formation by Day 11, ranging in size from 0.1 mm to 0.4 mm (Fig. 38,40). Cell division was also confirmed by marker lines showing nuclear-localised fluorescence (P_{WIND1} , P_{IAA8} and tricolour $WOX13:WIND1:WOX5:HYG$; Fig. 39). However, the cell division rate was approximately 5%, which remained significantly lower than 50% reported by Chupeau et al. (2013)¹⁰⁸. Notably, one of the initial reports on protoplast regeneration suggested a cell division rate of 0.005% in liquid culture, prompting the development of alginate embedding²⁸⁶. In addition to the low cell division rate, no callus was obtained from microcallus under these conditions. This could be due to the number of initially dividing cells being low, whose viability and regeneration progression will inevitably decrease further, causing unsuccessful callus regeneration. Therefore, there is a need to enhance the division rate. To address this, alginate embedding was tested twice, albeit under Condition 1. An additional trial should be conducted using plants cultivated under Condition 2, employing alginate embedding. The mechanochemical environment is expected to differ between liquid culture and alginate embedding, with the latter appearing to be more regenerative (Table 3). Culturing protoplasts isolated from plants grown under Condition 2, with alginate embedding, will help determine if the culture can successfully regenerate under conditions reported in other studies^{114,286–288,290}.

Moreover, owing to the contamination risk, cell imaging was not performed between the microcallus stage and the conclusion of the culture in Callus Inducing Media (CIM), a process that spans more than a month. Consequently, it remains uncertain at which point cell viability decreases in CIM. One approach to address this is to conduct characterisation studies similar to the ones employed in this study between Days 0-6, for microcalli cultured in CIM. Methods involving alginate embedding entail the replacement of media with fresh CIM every 10 days, which could be one of the reasons for increased regeneration efficiency. This approach couldn't be implemented in liquid media, as microcallus tends to both sediment and float, requiring centrifugation to settle them at the bottom – which is not part of the Chupeau et al., (2013)¹⁰⁸ protocol. However, attempts at centrifugation led to contamination in the culture, likely due to sharing a centrifuge with the research center where bacteria, yeast, and cyanobacteria are routinely cultured. The protoplast culture could significantly benefit from conducting these procedures in a clean room at a tissue-culture level, minimizing the risk of contamination.

Differences in regeneration potential were seen between plants grown under Condition 1 and Condition 2 (Fig. 40e vs Fig. 41a,b) even though the protoplasting and regeneration conditions were the same. To gain more information at the molecular level, qPCR analysis was performed at Day 0 to assess the degree of dedifferentiation. Effects of 1X and 2X enzymes on regeneration progression were seen by observing symmetric cell division, but still microcalli or calli was not seen. Therefore, to increase the regeneration capacity further, cell wall digestion enzyme treatments were tested. It is established that the removal of the cell wall serves as the trigger for dedifferentiation (and the subsequent regeneration), but it remains uncertain whether it is the digestion of specific cell wall compounds or the complete removal that initiates the process. There is a possibility that digesting only cellulose or a combination of cellulose and pectin could release protoplasts with enhanced regeneration potential.

Regarding mesophyll cell identity markers, the overall relative expression of *LHCB2.1*, *FNR1*, *CA2*, and *CUE1* increased in Condition 2 compared to Condition 1. This increase is likely attributed to the extended photoperiod and higher light intensity in Condition 2 (Figure 44a-d). Cel+Pec and Protoplast samples exhibited the lowest marker identity expression under Condition 2, which are the two treatments that produced protoplasts, aligning with the expected cellular reprogramming in protoplasts. This suggests that complete removal of the plant cell wall may be necessary for dedifferentiation.

In terms of selected transcription factors, particularly under Condition 2, relative *WIND1* levels increased threefold for Cel+Pec and Protoplast samples and fourfold for No enzyme and Cellulase samples. Given that *WIND1* overexpression is known to promote callus formation in tissue culture, this upregulation may represent one pathway contributing to the enhanced regeneration competence of plants in Condition 2²⁵⁷. Notably, relative *IAA8* levels were higher in plants grown under Condition 2, while there was no significant difference in *WOX13* transcripts. The environmental factors that are different between the two conditions—temperature, light spectrum, and intensity—have been documented to regulate auxin and cytokinin levels during plant development, which are likely affecting protoplast regeneration competence as well^{311–315}.

Considering that environmental factors affected regeneration in this study, the levels of *JMJ30*, a temperature-sensitive histone demethylase known to regulate circadian rhythm, were examined. Indeed, the levels were lowest in Condition 2 where temperature is the lowest owing to diurnal temperature fluctuations (Fig. 44i). The diurnal temperature fluctuations might be altering the levels

of temperature-sensitive epigenetic factors, such as *JMJ30*, which in return affect regeneration. Although *JMJ30* has not been extensively studied in the context of protoplast regeneration, it could potentially regulate the expression of genes crucial in the early stages of regeneration, through epigenetic modifications. In terms of epigenetics of protoplast regeneration, it is known that histone acetylation and increased chromatin accessibility accompanies dedifferentiation^{110,114}. Histone methylation/demethylation can reduce or increase accessibility, depending on the context³¹⁶. Hence, *JMJ30* might exert global epigenetic effects, possibly by decreasing chromatin accessibility. Consequently, low levels of *JMJ30* could lead to increased chromatin accessibility. Alternatively, *JMJ30* might have roles in differentiation, and thus reduced levels of *JMJ30* could enable dedifferentiation.

While NOX levels did not exhibit a significant difference in terms of normalised levels, Condition 2 displayed the least standard deviation, suggesting a potential role of NOX homeostasis in managing cellular stress (Fig. 44j). This parallels observations with reactive oxygen species levels in protoplasts, where an initial burst is necessary for dedifferentiation, but excessive levels or failure to reach homeostasis can abort regeneration^{139,140,317}.

5.3.3. Preliminary data suggests that day/night temperature fluctuations deregulate *WOX5* expression

To investigate the effects of temperature fluctuations further, an experiment was set-up with added heating pads in a growth cabinet (Fig. 45). By Day 5, both samples exhibited upregulated levels of *WIND1*, *JMJ30*, and *WOX5* compared to Day 0 (Fig. 45c-e). Although *WIND1* (Fig. 45c) and *JMJ30* (Fig. 12d) levels did not show a significant difference, preliminary findings indicate a twofold increase in *WOX5* levels in plants subjected to diurnal temperature fluctuations (Fig. 45e). *WOX5* is involved in auxin and cytokinin signaling in both callus and shoot regeneration²⁶⁶, and it is potentially involved in protoplast division through a similar mechanism given the twentyfold increase of its transcript levels on Day 5 (Fig. 45e). Inducible over-expression of *WOX5*, along with loss-of-function and complementation assays, to help identify *WOX5*'s role in protoplast regeneration.

However, the standard deviation is high for the samples in Figure 45, due to the limited amount of RNA that was extracted from the samples. It is recommended to repeat the experiment in two identical growth chambers, one with constant temperature, and the other with temperature fluctuations. This will enable increasing the number of plants and would yield a larger number of protoplasts for both RNA extraction and culturing. Thus, this approach would enable a more robust, functionally based comparison of regeneration competence and progression between the two

samples. Moreover, separate environmental chambers would mitigate issues related to condensation and potential developmental challenges associated with the bottom-up heating of the plates in the current setting. Furthermore, employing developed marker lines would facilitate tracking dynamic changes in the culture, as the optimisation of timelapse conditions was not possible to date.

5.4. Conclusion

In this study I aimed to establish a reliable and reproducible protoplast regeneration platform. Initial attempts proved nonregenerative; however, through modifications in cell wall digestion enzyme concentrations and plant growth conditions, a culturing protocol capable of symmetric division and microcallus formation was successfully developed. The physiological characteristics of the culture, along with molecular dynamics over the course of protoplast dedifferentiation and regeneration, were documented under various growth conditions, revealing deregulation in both mesophyll cell identity genes and transcription factors associated with regeneration. Notably, the most regenerative plant growth condition involved diurnal temperature fluctuations, with a preliminary experiment indicating a twofold upregulation of *WOX5* levels compared to constant temperature. Investigating the environmental conditions in protoplast regeneration could hold the key to addressing reproducibility issues in the field. With future research, the identification of crucial environmental factors is anticipated, paving the way for truly reproducible and consistent protoplast regeneration protocols.

CHAPTER 6

Discussion & Future perspectives

6.1. Summary

In my dissertation research, I have developed and applied synthetic biology tools to investigate gene regulation and protoplast regeneration. My research showed combinatorial interactions between promoters, coding sequences, and terminators in gene regulation, with RNA folding suggested as a potential mechanism to explain these interactions (Chapter 2). While making constructs for the gene regulation assays, I encountered plasmid stability issues in Mobius Assembly vectors. Consequently, I have identified and successfully addressed the underlying causes of the plasmid stability issues (Chapter 3). Moreover, during the gene regulation assays in protoplasts, I found achieving a consistently high transformation efficiency challenging. I, thus, devised a high-throughput protoplast transformation protocol for transient expression. I used this protocol to screen fluorescent proteins and identify top performers. I then used the top performers to create five stable *A. thaliana* lines reporting the expression dynamics of transcription factors involved in protoplast regeneration (Chapter 4). Protoplasts are notoriously heterogenous in their culture behaviour; therefore, I have established a regeneration platform to culture mesophyll protoplasts into microcallus (Chapter 5). While establishing this platform, I showed that the plant growth conditions and cell wall digestion enzyme concentrations affect regeneration competence.

In this chapter, I will explore future directions in plant synthetic biology, especially in the areas associated with my dissertation projects, including potential avenues for refining available gene regulation tools and context dependency. Additionally, I will share my perspectives on engineering plant development with a focus on protoplast regeneration.

6.2. Gene regulation

Two essential components are required to engineer plant development successfully. First, a clear understanding of the targets is necessary. Second, genetic tools for precision control are needed to engineer the desired biological behaviour. While a substantial amount of knowledge exists on organ development, such as root development in *A. thaliana*³¹⁸, a significant gap exists in our understanding of other developmental processes, including regeneration, in both *A. thaliana* and other species¹²³.

Regarding genetic tools, the development of tools to engineer complex programs like regeneration is an ongoing process. Frequently, analogous tools are initially established in bacteria, yeast, or mammalian cells and subsequently tailored for use in plant systems (such as the PULSE optogenetic system that uses engineered photoreceptors to reversibly induce gene expression with red light³¹⁹). Nevertheless, this adaptation from different organisms poses its own set of challenges.

While the activation or repression of certain genes above or below a specific threshold may suffice for a physiological response, genes often exhibit dynamic expression at different stages of development. For example, during the initial stages of callus formation, root stem cell markers such as *PLETHORA 2 (PLT2)* should be induced, and then, they must be repressed to allow shoot regeneration³²⁰. Constitutive overexpression of *PLT2* abolishes shoot formation from the callus³²⁰. Consequently, achieving precise control over target genes in a spatiotemporal manner becomes imperative for engineering plant regeneration and overall in plant development. .

6.2.1. Improving the ethanol-inducible system

Inducible systems are widely employed for temporal control of gene expression, and when combined with cell-type-specific promoters, they allow spatiotemporal control³²¹. As demonstrated in this study, the Estradiol and Dex systems exhibit effective performance with low basal expression and robust induction. Even though Estradiol and Dex systems might be useful in small-scale laboratory experiments, they are impractical for large-scale field applications due to their steroid hormonal nature³²². The copper-inducible system overcomes this limitation, as copper is an inexpensive microelement essential for plant growth³²². However, given the complexity of developmental pathways involving multiple genes, it is desirable to have a selection of inducible systems, ideally tailored for each target gene. Consequently, the need for additional inducible systems persists.

The ethanol-inducible system holds promise as a gene induction tool, with 0.1% ethanol proving sufficient for its induction³²³. Unfortunately, this study has shown that it is leaky and falls short of strong gene activation upon induction. To address these drawbacks, a circuitry strategy incorporating a suicide exon (*HyP5SM*) into the system is suggested. When the suicide exon cassette introduces the P5SM RNA element and an in-frame premature stop codon³²⁴. Therefore, when the suicide exon is expressed, it causes nonsense-mediated decay of the transcript and prevents translation. When its partner L5 is present, it binds to the P5SM RNA element and displaces the exon-defining splice factor, which results in exon skipping. Due to background levels of exon skipping and

gene expression caused by the endogenous L5 machinery in *A. thaliana*, an engineered version of the suicide exon and *O. sativa* L5 (*OsL5*) was developed and successfully demonstrated to function orthogonally³²⁴.

The *HyP5SM* and *OsL5* system can be incorporated in the ethanol-inducible system as illustrated in Fig. 46. The *HyP5SM* would be positioned within the reporter gene (*sfGFP*) that is constitutively expressed, and *OsL5* would be expressed under the control of the *P_{alcSynth}* promoter. The *alc* regulatory element would be constitutively expressed at a medium/low level based on the part characterisation done in this research. In the absence of ethanol, *OsL5* is not expressed, *alcR* is present at low levels, and *sfGFP* undergoes non-sense mediated decay. In the presence of ethanol, *OsL5* is expressed, leading to the skipping of the suicide exon. Through a feedforward loop, increased *AlcR* expression induces further *OsL5* expression (Fig. 46). A similar strategy was employed to reduce basal expression levels in the Dex system in *A. thaliana*³²⁵ Once it is confirmed that this system effectively prevents leakiness of the ethanol-inducible system and maintains optimal activation levels, it can be explored for diverse applications, such as developmental gene regulation or enhancing disease resistance in crops, possibly with the *A. thaliana* L5 system for orthogonality. Improvements to the existing inducible systems and the development of new systems, such as optogenetic technologies, will equip synthetic biology with the tools necessary for developmental engineering.

6.2.2. Designing amplifiers for gene expression reporters

In addition to their role in regulating gene expression timing, inducible systems have the potential to function as signal amplifiers, a concept already established in bacterial systems but lacking in plant systems³²⁶. Similar to the positive feedback loops in bacterial quorum sensing, transactivators of inducible systems for plants, such as *XVE* in the Estrogen-inducible system, can be placed under a cell-type-specific promoter, resulting in low levels of *XVE* expression. Without β -estradiol, a low amount of *XVE* will be present but will not activate gene expression. In the presence of β -estradiol, *P_{OlexA}* promoter will be induced, producing a reporter, *sfGFP*, along with more *XVE*, positively regulating gene expression (Fig. 47a). This could be valuable when investigating the presence or absence of a transcription factor in cases where the gene's expression is weak, or the signal is challenging to detect. For example, in the three-colour marker line (*WOX13:WIND1:WOX5*), I could not detect the *WOX5* promoter activity in dividing protoplasts, whereas I could detect the *WIND1* and *WOX13* signals. While it is possible that *P_{WOX5}* is not activated in our setup, transcriptomic studies suggest the presence of

WOX5 transcript during division¹⁰⁸. It is important to note that P_{WOX5} was expressing *mTagBFP2* as a reporter, and generally, blue fluorescent proteins are not as bright as green and red fluorescent proteins³²⁷. Despite the development of new versions of blue fluorescent proteins, such as Electra2³²⁷, signal amplification could still be helpful for the blue spectrum, especially when constructing multi-gene reporters or autofluorescence is a concern.

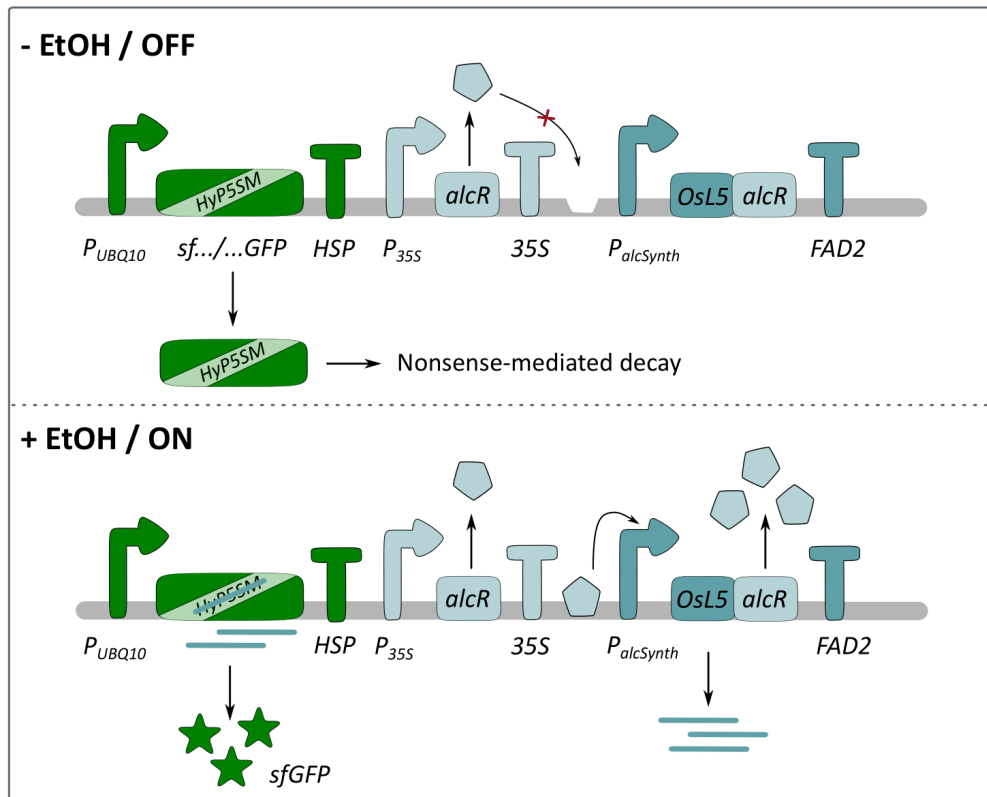


Figure 46. Incorporating suicide exon regulation into the ethanol-inducible system. *sfGFP* is expressed constitutively with the *HyP5SM* exon, introducing a premature stop codon. *alcR* element is expressed at a medium/low level constitutively. $P_{alcSynth}$ can control the expression of *OsL5* and *alcR*. In the absence of ethanol (EtOH), *sfGFP* is expressed with *HyP5SM*, causing nonsense-mediated decay (top panel). In the presence of ethanol (bottom panel), *OsL5* is expressed, binding to the *HyP5SM*, causing exon skipping and proper *sfGFP* translation. *alcR* expression from the $P_{alcSynth}$ creates a feed-forward loop for stronger $P_{alcSynth}$ induction.

Alternative to chemically inducible systems, the transactivators/promoters can be replaced with homoserine lactone variants, such as coumaroyl-homoserine lactone (pC-HSL), and its synthetic inducible promoter, which was recently shown to function in plants without causing growth defects³²⁸. Furthermore, dCas-9-regulated synthetic promoters can serve a similar amplifying purpose, where gRNA1 is expressed from the promoter of the gene-of-interest, activating the GB_SynP promoter³²⁹.

The GB_SynP promoter would express a fluorescent reporter (like *sfGFP*) and gRNA2, leading to stronger gene expression (Fig. 47b). Apart from serving as reporters; these amplifier systems could prove valuable in the future for building logic gates using genes expressed at low levels. Amplification of signals from the gene-of-interests (inputs) may be necessary for the gate to execute logic functions (outputs) while engineering developmental pathways.

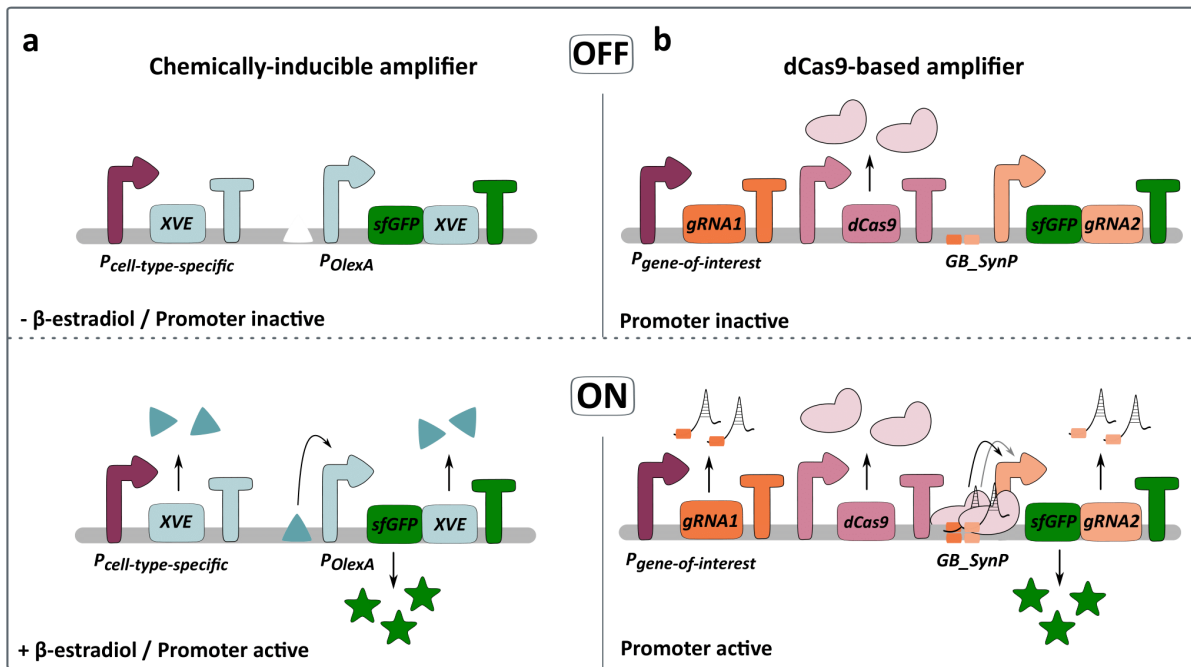


Figure 47. Suggested gene amplifier designs with feedforward loops. **a)** Chemically inducible amplifier, which uses components from an Estrogen-inducible system. XVE would be produced under a cell-type-specific promoter or the promoter of the gene of interest in low amounts. In the presence of β -estradiol, the *POlexA* promoter would be activated, producing *sfGFP* as a reporter and more XVE. **b)** dCas9-based amplifier which uses GB_SynP promoters. The promoter of the gene-of-interest is used to express *gRNA1* when the gene-of-interest is activated, which activates the GB_SynP promoter weakly. The activated GB_SynP promoter produces *sfGFP* as a reporter and *gRNA2*, resulting in stronger gene expression.

6.2.3. Context dependency in gene regulation

The suggested tools, including the modified ethanol-inducible system and gene amplifiers, would need to be characterised, akin to assessing any genetic part or tool. So far, most of the part characterisation is done in either *A. thaliana* or *N. benthamiana*, both dicot plants. However, parts can behave differently in monocot systems, such as the 35S and NOS promoters having lower expression in monocot *S. viridis*^{76,330}. The same part can have different expression levels, even within dicots like the

35S promoter. Similarly, the *OCS* promoter had strong activity in this study but low activity in *N. benthamiana*³³⁰. Consequently, genetic parts can behave differently in different chassis, and we should not generalise characterisation results among different plant species or expression systems (whole plants, transient leaf expression systems, protoplasts, or cell cultures). Yet, manually characterising all genetic parts—promoters, genes, terminators and UTRs—in all species of interest and across all expression systems is impractical. However, with the ongoing development and widespread use of automation systems, this task might become achievable in the future (Fig. 48).

This raises the question: Can we ever achieve true orthogonality in biology? Achieving orthogonality, defined as a construct displaying the same gene expression level independent of the organism's genetic background or growth conditions, appears unlikely due to variations in the cellular machinery of each organism. Nevertheless, reaching a satisfactory level is possible if we can understand gene regulation through a bottom-up approach, enabling us to foresee and offset the gene expression variations in different systems. Synthetic DNA parts are excellent tools for studying gene regulation, as exemplified by the study of synthetic promoters and the determination of TATA box locations and numbers on gene expression^{160,164}. Synthetic promoters have already been used to construct computational models to predict promoter strength in tobacco and maize protoplasts¹⁶⁴. With the continued development of synthetic UTRs and terminators in plant systems, similar models are expected to be created for these genetic parts.

Even though I have chosen the DNA parts following part characterisation data, I still had issues with stable gene expression while making *A. thaliana* marker lines. I have observed gene silencing, particularly evident between the T2 and T3 generations for the *P_{WOX13}:mScarlet-I_{SV40}:HSP* construct. This construct harbours the *HSP* terminator, which is shown to be consistently strong in this study in different contexts. Additionally, it has been demonstrated that the *HSP* terminator induces less promoter methylation, resulting in reduced silencing and elevated reporter gene expression compared to the 35S terminator³³¹. This suggests that having a strong terminator alone is not sufficient to maintain stable transgene expression across generations. The *P_{WOX13}* promoter used in the construct may be mismatched with the *HSP* terminator, the *mScarlet-I* gene might be prone to silencing, or the genetic parts might interact to promote silencing. In general, transgenes in plants are susceptible to silencing through the RDR6-mediated pathway³³². The same pathway is involved in terminator-dependent gene silencing, and it is suggested that the mismatch between promoters and terminator strength triggers this process (Gene Harmony Theory)¹⁶⁶. However, I propose that stable transgene

expression relies not only on the compatibility between the promoter and terminator but also on potential interactions among genetic parts based on the part characterisation results described in Chapter 2. Therefore, it is essential to consider the entire transcription unit rather than focusing solely on individual parts (Fig. 48a).

Besides the single transcription unit constructs, an additional layer of complexity emerges in multi-gene constructs. In *E. coli*, it has been demonstrated that the orientation and the relative transcription unit order can impact gene expression levels³³³. In the context of the three-colour marker lines (*WOX13:WIND1:WOX5:HYG*), *P_{WIND1}:sfGFP* consistently exhibited strong expression. Only one line displayed strong expression of both *P_{WOX13}:mScarlet-I* and *P_{WOX5}:mTagBFP2*, while the majority either strongly expressed *P_{WOX13}:mScarlet-I* or *P_{WOX5}:mTagBFP2*. *WIND1* reporter unit being consistently strong can be attributed to factors such as sfGFP being the brightest fluorescent protein (as shown in this study), the *P_{WIND1}* promoter being strong, or the positioning of the *WIND1* reporter unit, which is Flipped_1B. On the other hand, the strong expression of either *WOX13* or *WOX5* reporters indicates the local and the global architecture might influence gene expression levels (Fig. 48a).

Context dependency will likely pose a challenge until targeted gene insertion can be achieved in plant systems instead of random insertions. The evidence indicating that the insertion location can impact gene expression in yeast^{334–336}, a phenomenon likely applicable to plants, highlights this. In summary, strategies to prevent or offset context dependency involve three key levels: characterisation of synthetic DNA parts combinatorically within a transcription unit, characterisation of the relative order of transcription units, and characterisation of insertion location (Fig. 48a).

6.3. Construct stability

Context dependency in combinatorial part interactions, transcription unit order, and insertion location clearly influence construct stability as they are involved in gene silencing. However, transgene burden is another well-studied concept that impacts construct stability in bacteria and yeast, which has not yet been explored in plant systems. This could be because research on burden is predominantly focused on bioproduction and maximising yield, with plant systems only recently entering the

bioproduction scene. As a result, there has been a delay in studying these concepts in plants. By drawing from examples in bacteria and yeast, implementing capacity sensors (fluorescence-based reporters for quantifying burden) in plant systems is expected to enhance yields for bioproduction.

From another perspective, plants may possess a burden homeostasis system in which transgene silencing acts as a regulatory mechanism, maintaining transcriptional and translational burden at a specific level. While a high transcript level resulting from multiple copies of reporter genes is known to trigger RNA silencing, whether this effect is due to burden or repetitive sequences remains unclear³³⁷. Strongly expressed genes may be more susceptible to silencing due to increased burden, whereas medium-to-low strength genetic parts might be more stable across generations. If that is the case, rather than using strong parts that get silenced, the same (or higher) level of protein expression in subsequent generations (T3) can be accomplished by switching to medium-strength parts, potentially enhancing long-term stability. Integrating burden research into plant systems will allow for stable transgene expression and fine-tuning of transcription and translation processes to prevent excessive transcription initiation, as observed in *E. coli*²⁰¹.

Although much of my discussion has focused on construct stability in plants, my PhD research has also highlighted the significance of construct stability in bacteria for plant systems, given the prevalent use of *E. coli* as a propagation host. Until cloning can be performed directly in plants and high-yield plasmid isolation can be achieved from plants, we will probably rely on *E. coli*. Despite plant expression genes typically not being active in bacteria, the larger size of plant genes can impose a burden on *E. coli*, affecting stability²⁰³. Consequently, plasmid stability studies in bacteria, as conducted in this study, hold importance for plant scientists, as constructs become unusable if they mutate over time.

It is anticipated that online tools for construct design will be developed that incorporate statistical modelling or, as experimental datasets expand, incorporate predictive modelling. This will be enabled by incorporating high-throughput characterisation data across diverse species and expression systems. In this setup, users could select specific parts, such as a particular promoter or gene, and the online tool would provide suggestions for combining these parts to achieve different expression levels. Furthermore, the tool would incorporate DNA and RNA folding predictions to anticipate possible interactions between genetic parts, which was demonstrated in this study, and their effects on gene expression. Additionally, transcriptional and translational burdens can be considered for both plant species of interest and bacterial storage hosts. These functionalities prove

valuable not only for predicting expression levels but also for ensuring the stability of transgene during storage and expression over generations (Fig. 48b).

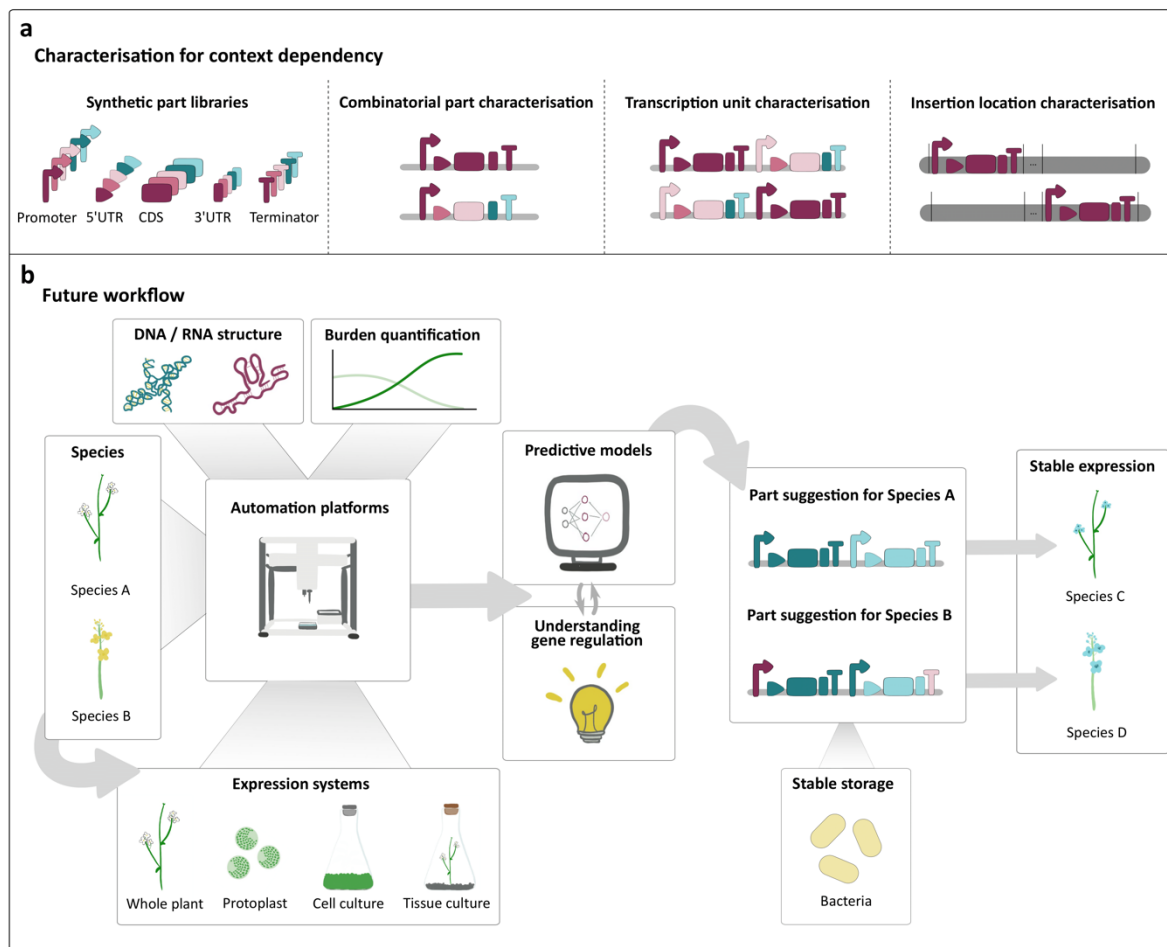


Figure 48. Perspectives on context dependency in gene regulation. **a)** Characterisation. First, synthetic part libraries should be made, which should be used to analyse reporter expression combinatorially. Additionally, the relative order of transcription units and genomic insertion location characterisation need to be characterised. **b)** Future workflow. All modes of characterisation outlined in (a) will be conducted across different species and expression systems, complemented by DNA and RNA structure prediction and *in vivo* confirmations. This necessitates high-throughput experiments, which are made feasible by advancements in automation platforms. Combining experimental data and modelling will enhance our understanding of gene regulation, facilitating the development of predictive models. Ultimately, these models aim to accurately predict the expression levels of selected genes in chosen species and may suggest part combinations. These parts can be stored stably in bacteria or plants with enhanced plasmid stability and burden knowledge. While true orthogonality may not be achieved, predictive models can anticipate variations in gene expression levels across different species or expression systems and adjust part recommendations accordingly.

6.4. Studying protoplast regeneration

The development of genetic tools providing stable spatiotemporal control over gene expression and an enhanced understanding of gene regulation will help unravel the molecular pathways involved in highly dynamic biological processes such as plant development. Developmental engineering includes a wide range of applications, like engineering plant architecture, as elaborated in Chapter 1. However, in terms of engineering development, this section focuses on protoplast regeneration. Protoplast regeneration is not limited to generating whole plants; instead, protoplasts can be regenerated into microcallus (or callus) and subsequently differentiate into different cell types, a process referred here as guided differentiation. This focus on guided differentiation arises as it is one of the direct applications of the platforms I have developed in this study.

For protoplast regeneration, the progress in genetic tool development and understanding gene regulation is particularly significant, given that our current knowledge of molecular pathways primarily stems from tissue culture. Recognising that the same molecular players may assume different roles in different regeneration modes is crucial. For instance, *WOX13* promotes callus formation in grafting, demonstrating a ‘pro-regeneration’ effect¹¹⁸. Conversely, during shoot regeneration in tissue culture, it inhibits shoot regeneration, taking an ‘anti-regeneration’ effect¹¹⁹. This makes it unlikely that a target's constitutive overexpression will enhance all regeneration modes. Therefore, in the context of protoplast regeneration, elucidating the roles of transcription factors previously identified as regulators of regeneration in tissue culture, including *WOX5* and *WOX13*, is essential. Hopefully, the stable marker lines developed in this study will help unravel these unknowns.

Establishing a protoplast regeneration platform is the initial step in investigating the molecular pathways involved in protoplast regeneration, enhancing the regeneration capacity of plants and achieving guided differentiation in cell culture. As I have encountered, reproducibility challenges often arise at this stage, leading different laboratories to modify existing protocols^{108,114,287,289,290}.

6.4.1. Waddington's landscape

The underlying reason for these reproducibility issues can be better understood by revisiting Waddington's landscape. This conceptual model illustrates how embryonic cells or tissues navigate diverse developmental choices, driving them toward distinct cell fates³³⁸. In this landscape, an abstract

set of constraints form the valleys and hills and an embryonic cell or tissue, represented by a rolling ball, makes its way down from the top of the hill (Fig. 49a). At a bifurcation point, epigenetic, genetic, and environmental factors collectively determine the direction the ball will take. While not explicitly labelled in the original diagram, previously mentioned factors significantly influence this decision-making process³³⁸. The Waddington landscape was initially drawn to conceptualise organismal development as a unidirectional process, and new arrows were added to depict alternative routes that can be taken by the ball, showing dedifferentiation and trans-differentiation processes. This landscape is also helpful in describing the heterogeneous fates mesophyll cells take after protoplasting: some adopt totipotency, some adopt pluripotency, and some retain their original identities. (Fig. 49b).

6.4.2. Protoplast regeneration in Waddington's landscape

Incorporating protoplast regeneration into this diagram involves a two-step process: dedifferentiation (Fig. 49c) followed by redifferentiation (Fig. 49d). Assuming the fixed topology (height, number, and positions of hills and valleys) in the diagram for a given organism, external forces are required to propel the ball (representing mesophyll cells) uphill towards a totipotent state. I propose that these forces originate from plant growth conditions and protoplasting conditions. Each factor within plant growth conditions—temperature, light, and media—can be viewed as independent vectors with varying magnitudes and directions. Similarly, the enzymes, concentrations, and digestion durations during protoplasting represent vectors with distinct magnitudes and directions. When the sum of these vectors (depicted as the yellow line) aligns optimally in magnitude and direction, it can propel the ball to the totipotency position. Likewise, during the redifferentiation step, the combined vectors from the previous step and culture conditions guide the ball downhill, requiring specific magnitudes and directions to achieve redifferentiation through a particular route (representing specific cell fate) (Fig. 49d).

6.4.3. Reproducibility in protoplast regeneration with Waddington's landscape

Plant growth, protoplasting, and culture conditions become essential determinants of the cell trajectory when aiming to reproduce previously published protocols for protoplast regeneration. Variation in these conditions can influence whether a protoplast will successfully regenerate. For

instance, one growth condition may not provide the optimal light and temperature, resulting in insufficient force to propel the ball forward (Fig. 49e). Conversely, another condition might offer excessive light intensity and temperature, causing the ball to veer off the regeneration trajectory and into an alternative path (Fig. 49e). I have demonstrated that altering the protoplasting conditions shifted cells from asymmetric to symmetric division. Furthermore, maintaining the same protoplasting conditions while changing the plant growth conditions resulted in the formation of microcallus, which was not observed under the original plant growth conditions. Nevertheless, it is crucial to highlight that despite the adjustments made to the conditions, I have not observed either callus formation or complete plant regeneration. This suggests that the combined force vectors were sufficient to guide the ball to a hill with a microcallus route, yet not the entire regeneration pathway. Alternatively, it is possible that only a small number of cells reach totipotency hill under the conditions identified here, while the majority of the cells may follow alternative paths because of heterogeneous cell fate changes (Fig. 49b). Protocol development, therefore, involves refining the vectors to guide most of the cells along the complete regeneration trajectory. Multiple combinations can yield the optimal sum of vectors, as each factor contributes uniquely, and a reduction in one factor may be compensated by increasing another.

It is possible that species like the *A. thaliana* Col-0 accession employed in this study may possess narrow and steep pathways, complicating vector adjustments and leading to reproducibility issues. Conversely, more regenerative species or ecotypes, such as *A. thaliana* WS accession¹⁰⁸, may feature broader and gentler inclines and potentially multiple converging paths for regeneration, which are influenced by the epigenetic and genetic background of each species (Fig. 49f). Therefore, it is essential to characterise and report protoplast morphology under well-defined conditions, as demonstrated in this study. This could serve as a map to understand the cells' trajectories under specific conditions, enabling us to modify them to attain the desired cell fates.

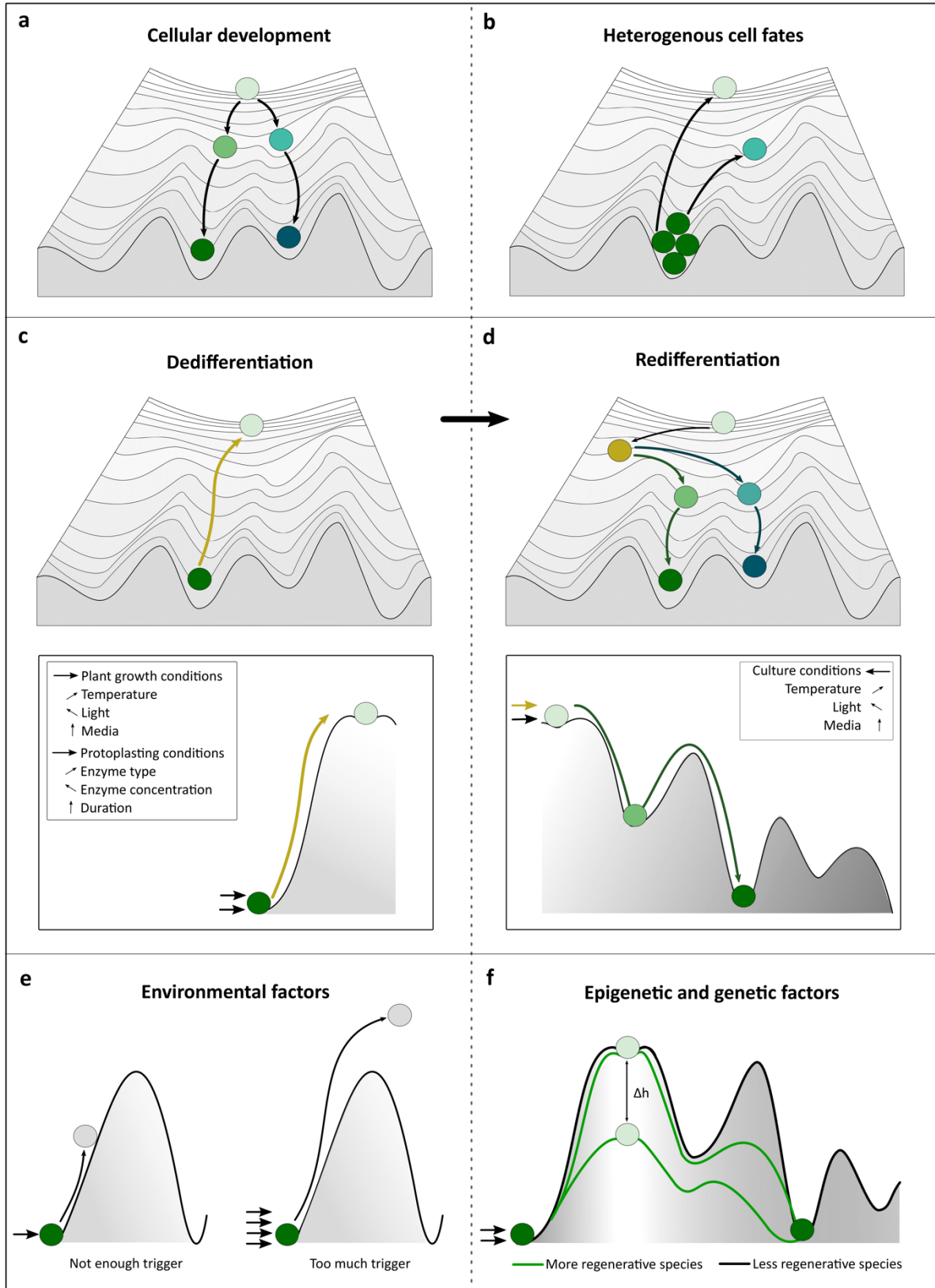


Figure 49. Waddington's epigenetic landscape to understand protoplast regeneration. **a)** Cellular development illustrating normal cell differentiation³³⁸. **b)** Heterogenous cell fates are adopted by mesophyll cells after protoplasting¹¹⁴. **c-d)** Two-step regeneration process of protoplasts. The first step is dedifferentiation (**c**), where mesophyll cells must be 'pushed uphill.' Each environmental variable during plant growth and protoplasting can be considered independent vectors acting from different directions and with different magnitudes. When the sum of the vectors is at the optimum direction and magnitude (yellow line), the mesophyll cells can reach a dedifferentiation state. The second step is redifferentiation (**d**), where the vectors from the previous step and the regeneration conditions act together to 'push the ball downhill.' **e)** Effect of environmental factors in a non-regenerative condition. The cells can't cross the hill if the trigger magnitude (or direction) is insufficient. If the trigger magnitude (or direction) is too much, then cells would overshoot the regenerative path and take an alternative one. This hill represents both dedifferentiation and any hill during redifferentiation. **f.** Effect of epigenetic and genetic factors. The background of the organism can dictate the height of the hills (Δh). More regenerative species (green line) will likely feature smaller hills, requiring fewer triggers for crossing the hills. Less regenerative species (black line) will likely have taller hills, requiring more triggers for a cell to cross them. Additionally, multiple routes for regeneration probably exist for more regenerative species, providing various combinations of plant growth, protoplast, and culture conditions for successful regeneration.

6.5. Engineering plant development

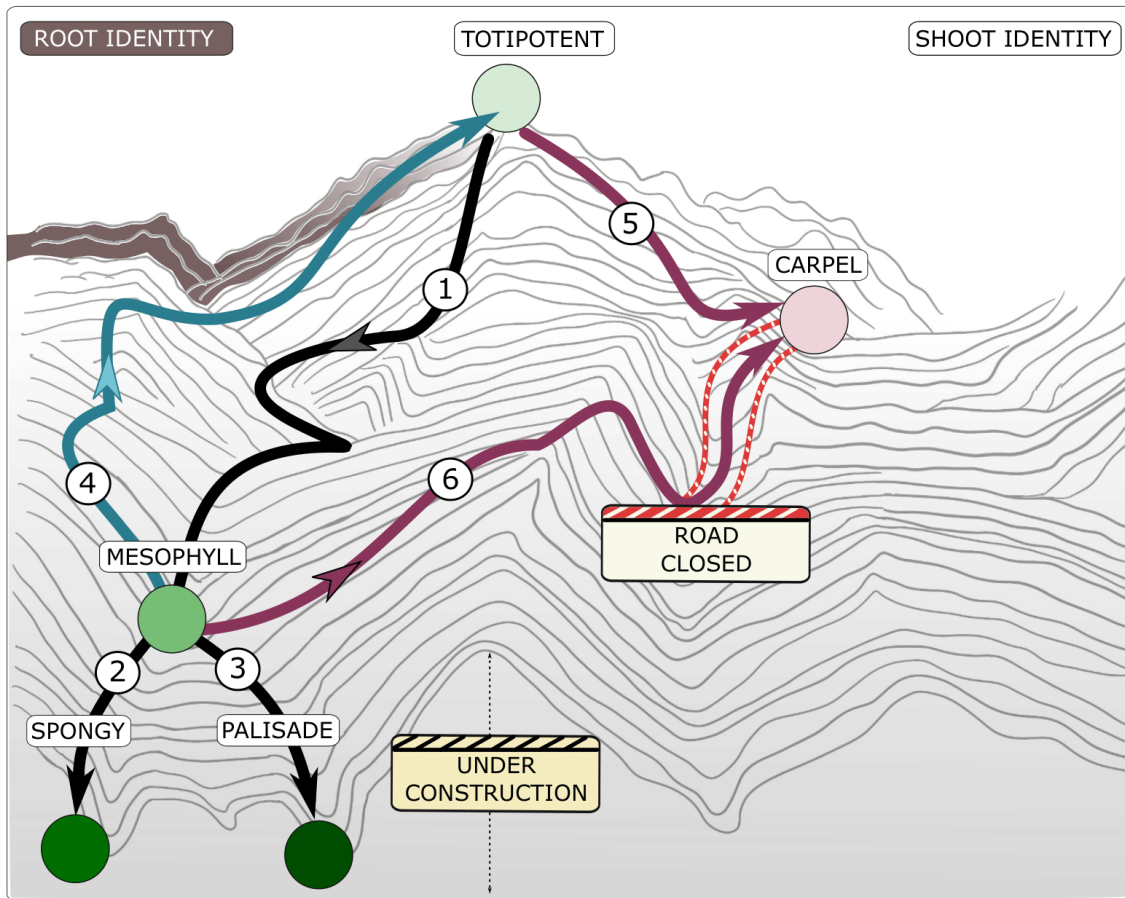
The categorisation of factors into either topology or vector and their classification as fixed or dynamic poses philosophical questions. Although the previous section assumed a fixed topology, the topology is likely dynamic, adapting to changes in epigenetics, genetics, and environmental stimuli. In such cases, overexpression of pro-regeneration genes could be linked to reducing elevation of the hills. Similarly, environmental factors, such as light or temperature, may influence the topology. Comparative transcriptomic studies under different environmental conditions will prove invaluable in uncovering epigenetic and genetic variances between highly regenerative and less regenerative species. This will help us understand the requisites for distinct modes of regeneration (routes) and mapping out the topology (Fig. 50a).

Once this map is established, it becomes feasible to redesign the topology or introduce specific vectors to guide cellular trajectories—essentially engineering different stages of development, not only regeneration. Current synthetic biology tools for engineering development include mostly gene regulation tools (Fig. 50b). These include well-characterised basic and cell-type specific parts, inducible systems, and DNA assembly methods employed and refined in my dissertation research. Innovative tools are emerging that merge CRISPR-dCas9 with activation and repression domains³³⁹, offering new strategies for engineering plant development alongside logic gates⁷⁵ and predictive modelling³⁴⁰. On the other hand, riboswitches³⁴¹, DNA origami³⁴², and sender-receiver systems³²⁸ have been used to control gene expression in plants, and they are predicted to be integrated into engineering

development. Popular synthetic biology tools like unnatural amino acids and artificial cells have not yet found their niche in regulating gene expression in plant systems. However, they hold potential as orthogonal signalling molecules and substitutes for bacteria in delivering developmental regulators. It is plausible that next-generation tools, such as control systems capable of auto-regulating gene expression at different stages, will emerge. Metaphorically speaking, these tools adjust the terrain, block specific pathways, or set up barriers to guide cellular trajectories in Waddington's diagram (Fig. 50b).

Looking ahead, with the mapped-out topology and gene regulation tools, we can guide the cells toward their desired fates (guided differentiation). For instance, we can direct totipotent cells to follow Route-1 towards mesophyll cell fate and subsequently guide them to follow Route-2 for spongy or Route-3 for palisade mesophyll fates (Fig. 50a). Engineering their abundance and ratios might help with optimising CO₂ exchange and photosynthesis^{343,344}. This approach holds promise for designing plants with enhanced metabolism and biomass production. Moreover, we can potentially redirect different cell types, such as mesophyll cells, abundant in leaf tissue, back to a totipotent state or guide their transdifferentiation to specific cell fates to produce secondary metabolites. For instance, sustainable vanillin production could be achieved by cultivating carpel cell cultures of *Vanilla planifolia* either by dedifferentiating mesophyll cells and then guiding redifferentiation (taking Routes-4 and Route-5, Fig. 50a) or directly from mesophyll cells through Route-6 (Fig. 50a). Developmental engineering in the form of guided differentiation in cell culture, highlights one of the main application of the protoplast regeneration platform established in this study. Additionally, this approach offers a promising alternative to the unsustainable whole-plant bio-production or petroleum-based extraction methods. Essentially, cell-culture-based guided differentiation presents an alternative or complementary approach to metabolic engineering³⁴⁵. It focuses on cultivating plant cells that are inherently competent to produce target secondary compounds rather than relying on complex metabolic engineering to engineer bacteria or yeast for the same purpose.

a Uncovering and re-designing cellular trajectories



b Tools for re-designing cellular trajectories

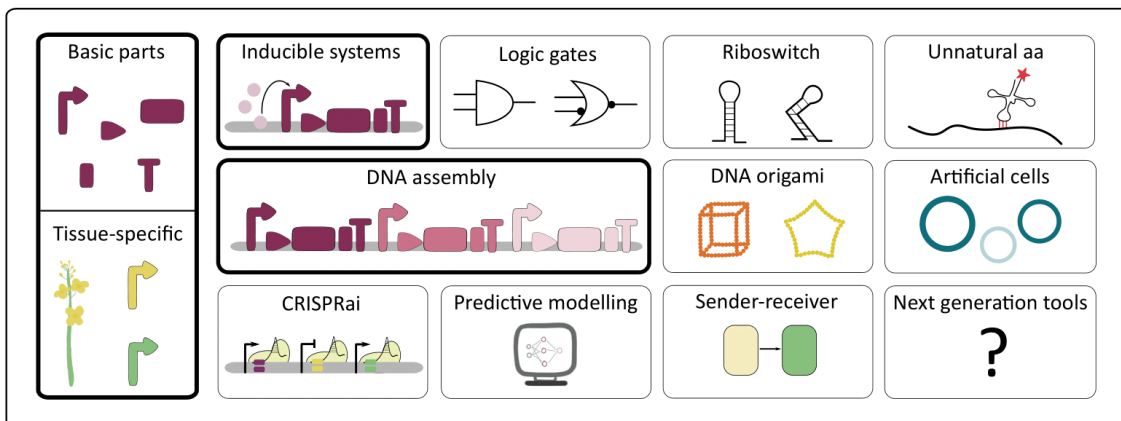


Figure 50. Future of engineering plant development. **a)** Uncovering and re-designing cellular trajectories. The front of the mountain represents shoot identity, while the back side represents root identity. Plant development and regeneration studies will help us map the topology and possible developmental routes. Existing trajectories can guide the differentiation of totipotent cells to mesophyll cells (Route-1). Route-2 can be taken for spongy mesophyll fate, and Route-3 for palisade mesophyll fate. Dedifferentiation studies will help identify existing or

novel routes from mesophyll fate to totipotent state (Route-4). In the future, guided differentiation (Route-5) or transdifferentiation (Route-6) can be used to create novel cell culture lines for secondary metabolite production. Synthetic biology tools help guide this cell trajectory step by modifying the topology 'Under Construction' or preventing alternative routes 'Road Closed' and directing the cells towards desired routes. **b)** Tools for re-designing cellular trajectories. Thick boxes for basic parts, tissue-specific parts, inducible systems, and DNA assembly methods highlight the tools used or refined in this study.

6.6. Conclusion

My dissertation research has contributed to mapping out the genetic landscape of protoplast regeneration, as well as to developing tools to engineer plant development, as highlighted in the gene regulation section. It has been exciting to witness the progress in plant synthetic biology, which is enabling the engineering of plant development ever more each day, bringing hope to creating solutions for urgent sustainability problems defining our times.

CHAPTER 7

Materials & Methods

7.1. Molecular biology

7.1.1. Bacterial strains, growth conditions and growth curve measurements

Escherichia coli strains Turbo, DH10 β and JM109 were used; and *Agrobacterium tumefaciens* strain ABI and GV3103 were used. *E. coli* cells were incubated at 37°C with Luria-Bertani (LB) medium, and *A. tumefaciens* was incubated at 28°C with Yeast Extract Peptone (YEP) medium. Growth curves are measured in Spark (TECAN) plate reader with orbital shaking Thermo Scientific Nunc F96 MicroWell (black, cat no. 10000631). Optical density (OD) measurements were performed at 700 nm, as most constructs have red fluorescence.

7.1.2. Bacterial transformations

Chemically competent *E. coli* were prepared as described by Chung & Miller³⁴⁶. *A. tumefaciens* were prepared mainly as electrocompetent³⁴⁷, and only for the tricolour reporter construct containing *IAA8* (Chapter 4), an additional freeze-thaw³⁴⁸ method was used. Modifications were employed for the electrocompetent protocol. Mainly, cells were washed twice with cold, sterile water and twice with 20% glycerol, followed by resuspension in 25% glycerol. For *E. coli* transformation, 5 μ l of the plasmid DNA (after Golden Gate, and lower amounts can be used for routine plasmid amplification) was incubated with 100 μ l of the competent cells on ice for 30 min, followed by the addition of S.O.C medium (900 μ l) and 1 h recovery at 37°C. The heat-shock step was not necessary for *E. coli* transformation. Following recovery, cells were centrifuged at $\sim$ 5000 RCF for 1 minute. The supernatant was then discarded, and the cells were resuspended in the residual media remaining at the bottom of the tube. Then, cell suspension was plated on LB agar plates using antibiotic selection. For *A. tumefaciens*, 1 μ l ($\sim$ 100 ng) of recombinant plasmid was added to the 50 μ l of competent cells after thawing. Cells were mixed with the plasmid by gentle pipetting, and the mixture was transferred to pre-chilled electroporation cuvettes and kept on ice. Bio-Rad (MicroPulser, 411BR 10266) was used at 2000V manual setting. The time constant was noted, and the aim was > 5.8 mS for successful transformation. Immediately after transformation, 800 μ l of YEP was added to the electroporation cuvette. Then, for better aeration, the transformation mixture was pipetted to 2 ml tubes. Recovery was performed at 28°C for 3-4 hours. After recovery, 100 μ l of the cell suspension was plated on YEP plates with antibiotics. Plates were incubated at 28°C for 2-3 days. For the freeze-thaw method, 1 μ l

plasmid DNA was added on thawed *A. tumefaciens*, which were frozen on dry ice and then thawed again at 37° for 5 minutes. For recovery, 500 µl YEP was added and same steps were followed as electrocompetent transformation.

7.1.3. Plasmid isolation

E. coli cells were incubated at 37°C and 200 min⁻¹ either in 5 ml (miniprep) or 100 ml (for midi prep) in LB growth medium supplemented with antibiotics. Cells bearing the LhGR-pOp6 and sXVE-lexA inducible constructs were grown for 24 h instead due to their slower growth rate. Plasmids were isolated using Monarch Plasmid Miniprep kit (NEB, cat. no T1010L) and with GeneJET Plasmid Midiprep kit (Thermo Fisher Scientific, cat. no K0482) for higher yield midiprep samples.

7.1.4. *A. thaliana* genomic DNA extraction

Around 100 mg of leaf tissue, usually from plants grown *in vitro* (described in section 7.4.1), was added in a 1.5 mL Eppendorf tube. Tissue was grounded with an autoclaved mini pestle. 400 µl Edward's genomic extraction buffer³⁴⁹ (1 M Tris.HCl at pH 7.5; 1M EDTA at pH 8; 1M NaCl; 0.5% (w/v) SDS; Milli-Q water, filter sterilized) was added on the grounded tissue. Sample was vortexed for 5 seconds, and then centrifuged for 5 minutes at max speed (usually 13800 RCF). 300 µl of the supernatant is transferred to a fresh Eppendorf. 300 µl isopropanol was added to the supernatant and the mixture was centrifuged for 10 minutes at max speed. Supernatant is removed and the DNA should be as a transparent pellet at the bottom of the tube. The DNA pellet was dried under a laminar flow hood for ~ 2 h or with a concentrator (Eppendorf, Concentrator Plus)) at 37/42°C for 20 minutes. After the pellet is dried, 100 µl nuclease-free water was added and the pellet was resuspended. DNA concentration was measured with a NanoDrop, and the 260/280 ratio will probably be less than 1.8 as this sample is not extensively purified but good enough for PCR amplification. Purified DNA is stored at -20° C.

7.1.5. PCR amplification

All PCR amplifications for plasmid construction and cloning were performed using Q5 High- Fidelity DNA Polymerase (NEB, cat. no M0492L). Where PCR amplification with Q5 High Fidelity Polymerase was unsuccessful (while amplifying promoters from *A. thaliana* genomic DNA), Phusion High Fidelity

Polymerase (NEB, cat. no E0553S) was used. PCR amplification was followed by purification with Monarch PCR & DNA Cleanup Kit (NEB, cat. no T1030L).

7.1.6. Plasmid construction

ColE origin of replication (ori) and *mScarlet-I* for bacterial expression were PCR amplified from plasmids (pAB344 and pAB546) kindly provided by Dr. Alice Boo (Guy-Bart-Stan Lab). pMAPS plasmids with new oris were built by PCR amplifying new oris, which are used as 'inserts'. Then, the pMAPS plasmids except the pUC ori were PCR amplified and used as 'backbone' and digested with 1 µl DpnI (NEB) for 1 h at 37°C, to remove any non-denatured DNA. The insert and the backbone were combined with Gibson Assembly, using NEBuilder kit (NEB, cat. no E2621L), following manufacturer's recommendations. To make pMAP vectors with mScarlet-I instead of spsPINK, similar strategy with Gibson Assembly was used. New vectors were first verified with double restriction digestion with EcoRI-HF (NEB) and PstI- HF (NEB). The new vectors were confirmed by whole-plasmid sequencing (Full Circle Labs).

7.1.7. Mobius Assembly cloning

A detailed protocol can be found in Andreou & Nakayama., (2018)³⁵⁰. Briefly, the assembly was performed in a one-tube reaction with a total volume of 10 µl, with 20 fmol Acceptor Vectors and 40 fmol inserts. It is important to note that for Level 2 cloning, auxiliary vectors are treated as inserts, for 40 fmol were added. Reaction mix for all the levels of cloning had 1 µl of 1 mg/ml BSA (diluted from 20 mg/ml - NEB), 1 µl T4 DNA ligase buffer (NEB) and 0.5 µl T4 DNA ligase (NEB). For Level 0 and 2 cloning, 0.5 µl AarI (ThermoFisher Scientific) and 0.2 µl 50 x oligos (AarI buffer) was used. For Level 1 cloning, Eco31I/BsaI-HFv2 (ThermoFisher Scientific/NEB) was used. For Level 0 cloning, reactions were incubated in a thermocycler for 5 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. For Level 1 and Level 2 cloning, 10 cycles were used instead of 5.

7.1.8. Construct validation

All plasmids were verified first by colony PCR using GoTaq Green Master Mix (Promega, cat. no M7122) or with double restriction digestion with EcoRI-HF (NEB) and PstI- HF (NEB). The constructs were further verified by Sanger sequencing (GATC Biotech-Eurofins, Edinburgh Genomics or Source BioScience) or by whole-plasmid sequencing (initially by Plasmidsaurus and later Full Circle Labs). After plasmids are sequenced, they are transformed into *A. tumefaciens* as described above. *A. tumefaciens* colonies formed on transformation plate were re-streaked on a fresh agar plate. Colonies that grew on the fresh plate are chosen for either colony PCR or restriction-digestion to confirm the presence of the plasmids. For restriction-digestion mini prep protocols are followed with x1.5 times solutions.

7.1.9. RNA extraction from protoplasts for part characterisation assays

For isolating RNA, around 1×10^6 protoplasts were used per transformation. Isolation and transformation protocols are described in Section 7.2.1, with reaction and solution volumes upscaled by 26 times. Transformations were performed in 50-ml conical tubes (Falcon) until after the addition of MGG solution (following PEG incubation). After that, each transformation culture was separated in three 14-ml round bottom culture tubes (Greiner, cat no. 191160) to help with sedimentation. 24 h after transformation, the supernatant was discarded and sedimented protoplasts were used for RNA extraction with GeneJET Plant RNA Purification Kit (Thermo Scientific, cat no. K0801). Extracted RNA was treated with DNA-free DNA Removal Kit (Invitrogen, cat no. AM1906). RNA quality was confirmed by using NanoDrop.

7.1.10. RNA extraction for cell wall digestion assays

Around 300 mg of leaves from plants grown in vitro, or for leaves that went through cell wall digestion and were still intact, was collected and grounded with a micro pestle in 500 μ l lysis solution from GeneJET Plant RNA Purification Kit. For samples that produced protoplasts and isolated right after protoplasting, 500 μ l of isolated protoplasts was used without the need for centrifugation as protoplasts sediment easily to the bottom of tubes. For samples that are taken at Day 5, 10 ml culture was spined in 50 ml centrifugation tubes (Falcon) at 10 RCF for 5 minutes (acceleration 2, deceleration 2 setting). After the supernatant is removed, 500 μ l lysis solution was added on the protoplast pellet and cells were lysed by vortexing. Manufacturer's suggestions were followed for the rest of the RNA

isolation process. Extracted RNA was treated with DNA-free DNA Removal Kit (Invitrogen). Integrity of RNA was tested by running the samples on 1.2% agarose gel that shows 25S and 18S ribosomal bands without smears, followed by NanoDrop.

7.1.11. Reverse transcription and cDNA synthesis

Reverse transcription was performed with UltraScript 2.0 cDNA Synthesis Kit (PCR Biosystems, cat no. PB30.31-02). This cDNA synthesis kit uses optimum amounts of oligo (dT) and random hexamers for unbiased amplification of RNA variants. It is advised to use 2 μ M of both oligo (dT) and random hexamers for cDNA synthesis when using a different kit.

7.1.12. qRT-PCR for part characterisation assays

qPCR reaction was set up with primers and probe sequences listed in Appendix Table 2 and all were synthesized by Integrated DNA Technologies (IDT). PCR efficiency of the designed primers and probes were confirmed with standard curve method to be at least 99%. SsoAdvanced Universal Probes Supermix (Bio-Rad, 4311971) was used following manufacturer's instructions. Applied Biosystems StepOnePlus instrument was used with MicroAmp Endura Plates (Applied Biosystems, cat no. 4483485) and MicroAmp Optical Adhesive films (Applied Biosystems, cat no. 4311971). A standard curve was included on the plate for each gene target tested, and that standard curve was used to calculate cDNA copy number of the samples on the same plate. Fluc and mScarlet-I cDNA copy numbers were used as normalising for luciferase and fluorescence constructs, respectively. After that all the constructs were normalised to the cDNA copy number of the *UBQ10:NLUC:NOS* terminator in the same transformation batch, to account for any batch-to-batch variations that might occur. One way ANOVA with Tukey's HSD test was done by GraphPad (Prism) to calculate the statistical significance between samples.

7.1.13. qRT-PCR for cell wall digestion assays

qRT-PCR reactions were set up as described above. Comparative C_T method was employed in this section where *ACT2* gene was the internal control for all samples³⁵¹. Samples collected immediately after isolation are normalised based on the transcript levels of living plants of the same age. When investigating the impact of plant growth conditions on regeneration, samples are normalised to the

transcript levels of protoplasts derived Condition #2. One way ANOVA with Tukey's HSD test was done by GraphPad (Prism) to calculate the statistical significance between samples.

7.2. Cellular assays

7.2.1. Protoplast isolation and transformation

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 ~7400 lux humidity ~65%) for 4-8 weeks. The protoplast isolation and transformation protocol was developed based on Chupeau et al., (2013)¹⁰⁸ and Faraco et al., (2011)²⁵². Briefly, leaves were digested with Maceration Glycine Glucose (MGG) media containing: 0.2% (w/v) Cellulase ONOZUKA R-10 (Yakult, supplied by SERVA, cat. no 16419.03), 0.06% (w/v) MACEROZYME R-10 (Yakult, supplied by SERVA, cat. no 28302.03) and 0.08% (w/v) Driselase (Sigma-Aldrich, cat. no D8037-5G) for 8-10 hours. It is advised not to exceed 12 h digestion, as longer digestion can cause fragile protoplasts and excess cell debris. The protoplasts were filtered and washed 3 times with MGG without enzymes in 50 ml tubes (Falcon, cat no. 352070). Gravitational sedimentation was used for washes, centrifugation was avoided, and tubes were covered with aluminum foil to keep protoplasts in dark. It usually takes 1 h for sedimentation for the first wash, and then 45 mins for the following two washes. While waiting for the sedimentation fresh PEG solution is prepared as follows: 0.4 M mannitol and 0.1 M $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ was dissolved in Mili-Q water (PEG mass subtracted from the final solution volume). Then, 40% PEG4000 is added gradually while vortexing, as PEG is difficult to dissolve. The pH was adjusted to 8, and the solution was kept at 30°C until use to avoid solidification.

After three washes are completed, the MGG solution was removed, and cells were resuspended in MMM solution (0.4 M mannitol, 15 mM MgCl_2 , 0.1% w/v MES, pH unadjusted, usually between 4.75 - 5) to reach the concentration of 6×10^5 cells/ml with a hemocytometer. Then, 96-well plates were prepared where 8 μl DNA (500 ng/ μl) was added to each well. 75 μl of the protoplast suspension was then added on the wells containing DNA. 83 μl PEG solution was added in the wells and protoplasts were incubated with PEG solution for 1 min. Subsequently, the tubes were filled with 800 μl MGG solution and incubated for 1 h at room temperature. Following the incubation, protoplasts should have sedimented to the bottom of the wells. If not, plate containing protoplasts can be

centrifuged at 10 RCF (2 acceleration and 2 declaration). Transformation solution was removed, and the protoplasts were resuspended in 100 µl of fresh MGG and incubated in darkness at room temperature. Protoplasts were always handled gently using wide orifice tips (Starlab, cat no. E1011-8400).

7.2.2. Luciferase assay

For characterisation experiments, a dual-Luciferase reporter assay system was used. Specific parameters were optimised to improve the transformation efficiency and reproducibility, simplify the protocol and reduce the reagents. A nano-Glo reporter assay system (Promega) was used for the the part characterisation experiments, as it has better sensitivity at low expression levels and improved reagent stability. Additionally, Nano luciferase (NLuc) is compatible with established instrumentation and protocols for firefly luciferase (FLuc) and suitable for plant expression³⁵².

The plates containing the transformed protoplasts were let to sediment 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. 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 and background signal was applied using the software developed by Mauri et al³⁵³. Then *NLuc* signal was normalised to a construct bearing the *FLuc* flanked by *UBQ10* promoter and *UBQ5* terminator to minimise the expression variations.

7.2.3. Fluorescence assay

For fluorescent protein assays, Spark (TECAN) plate reader was used with either CELLSTAR 96 well plates (Greiner, black, cat no. 655090) or Nunc F96 MicroWell (Thermo Scientific, black, cat no. 10000631). Excitation and emission for superfolder green fluorescent protein (sfGFP) was 485 nm and 520 nm, respectively. For mScarlet-I, excitation and emission wavelengths were 560nm and 620nm. Measurements were taken at two different gains to make sure signals do not overshoot. sfGFP values obtained from the plate reader was normalised to mScarlet-I values. Since fluorescence is not always on, unlike luminescence, and black walled plates are used to prevent bleed-through, deconvolution correction is not done for fluorescent samples.³⁵³

7.2.4. **Fluorescent protein screening**

For fluorescent protein screening, Spark (TECAN) plate reader and CELLSTAR 96 well plates were used. For green spectrum (excitation wavelength 485 nm, emission wavelength 520 nm), sfGFP and mGFPmut2 were used. For yellow spectrum (excitation wavelength 485 nm, emission wavelength 520 nm), mVenusNB and eYFP were chosen. For red spectrum (excitation wavelength 560nm, emission wavelength 620nm), mScarlet-I and mRuby2 were used. Blue spectrum screening (excitation wavelength 400nm, emission wavelength 465nm) was done with mTagBFP2, mTurquoise2 and SCFP3A. In addition to unspecific expression, a nuclear-targeted version was built for each fluorescent protein with SV40 nuclear localisation signal. Except for sfGFP and eYFP, all the fluorescent proteins were taken from the Yeast Toolkit (courtesy of the Tom Ellis Lab). Even though YTK parts were codon-optimised initially for *S. cerevisiae*, they are also found codon-optimised for *A. thaliana*. YTK parts were used as PCR templates to build Mobius Assembly Level 0 parts. All the fluorescent proteins were expressed with the *UBQ10* promoter and *HSP* terminator. Measurements were taken at two different gains (to make sure signals do not overshoot) every 30 minutes for 24 hours. Plate-reader data was confirmed with microscopy, images were acquired with Nikon Eclipse Ti and NIS-Elements program. For green and yellow channels, 460nm filter with 2 seconds exposure was used. For red and blue channels 550 nm filter and 370 nm filters were used respectively. Exposure was set to 2 seconds for the red channel; and it was set to 3 seconds for the blue channel.

7.2.5. **Intracellular nitric oxide (NOx) measurements**

After protoplasts were washed three times, incubated in the fridge for 1 h and resuspended in PIM, 200 µl of the culture is taken and put in Nunc F96 plates (Thermo Scientific). Three technical replicates are prepared for each sample. Then, 5,6-Diaminofluorescein diacetate (DAF, 2A; Cambridge Bioscience cat. no CAY85165-500) at a final 8 µM concentration was added to each well. Plate was sealed, covered with aluminum foil and kept at room temperature (20-24°C) for 30 minutes. Then NOx signal was measured with Spark (TECAN) plate reader with green range filter (Excitation: 485 nm; Emission: 520 nm). OD700 measurements are also taken for normalisation. Then the green range filter values are divided by OD700, to normalise any errors that might arise from uneven pipetting. Next, each value is divided by the corresponding values of protoplasts from Condition 2, establishing it as the baseline for normalisation.

7.3. Whole-plant assays

7.3.1. *Agrobacterium* floral dipping

Agrobacterium mediated whole plant transformations are performed by following Zhang et al³⁵⁴. Briefly, three pots of *A. thaliana* plants were grown under short-day conditions (10 h light, 14 h darkness, ~7000 lux) at 22°C for 6-8 weeks for each construct that were transformed. Each pots have five plants. When plants started flowering, first bolts of inflorescence were clipped, to increase the number of inflorescences. Floral dipping was performed 5-7 days after clipping. *A. tumefaciens* harbouring the plasmid for transformation were grown in 5 mL YEP medium with antibiotics. For ABI strain, 50 µg/ml kanamycin and antibiotic for plasmid selection; and for the GV3103 15 µg/ml rifampicin, 25 µg/ml gentamicin and antibiotic for plasmid selection were used. The next day, 200 mL of YEP was inoculated with the saturated culture from yesterday and grown till stationary phase with OD between 1.5–2.0. Cells were pelleted by centrifugation at 8000 RCF for 15 minutes at room temperature. Then cells were resuspended in 200 ml sucrose solution (5% w/v sterile). Right before the transformation 0.04% Silwet L-77 (PhytoTech Labs) was added to the resuspended cells. While working with Silwet, extra care is taken to avoid skin contact as it is toxic. The culture is transferred to a leak-proof plastic bag and *A. thaliana* was dipped in the culture for 10 seconds while gently shaking. Then the plants are removed from the culture, and a thin coat should be visible on the floral buds. Transformed plants are covered with opaque covers and plants are kept at darkness for one day in the growth chamber. After one day, cover is removed, and plants are grown as usual until majority of the siliques turn brown.

After the siliques turn brown, seeds are collected and selected on ½ Murashige and Skoog (MS) media (MP Biomedicals, cat no. 17317479) with 0.08% agar (Merck, cat no. A7921) supplemented with 25 µg/ml hygromycin for three generations. Plants that grow normally on hygromycin are used for microscopy to further validate the functionality of the constructs in whole plants.

7.3.2. Microscopy

Expression of the constructs in correct cell types was confirmed with a Leica SP8 Confocal microscope. For *mScarlet-I* expressing constructs (*P_{WOX13}* or *P_{IAA8}*) excitation wavelength was 570 nm and emission wavelength was 590-600 nm. The *P_{WIND1}:sfGFP* construct was excited at 485 nm and emission

wavelength in 505-510 nm range. The *P_{WOX5}:mTagBFP2* construct's emission and excitation wavelengths were 400 nm and 450-460 nm, respectively. Laser intensity was set to 70%, and gain was 200 for all images. Sequential scanning was employed to capture images of the tricolour construct, minimizing the risk of signal interference across different spectra. Brightfield images were acquired alongside all fluorescent images and overlaid on the fluorescent counterparts with ImageJ, resulting in the final presented images.

7.4. Regeneration assays

7.4.1. Plant growth conditions

A. thaliana Col-0 seeds were sterilized with 70% ethanol for 4 minutes and then washed 4 times with sterile MiliQ water. After sterilization, around 25 seeds were sown on petri plates (Starsted, cat no. 82.9923.422) containing ½ MS medium (MP Biomedicals) and 0.8% agar (Merck). Plates are sealed with two layers of micropore surgical tape (3M, cat no. 1533-1). Plates are covered with aluminum foil and put in the fridge for at least 2 days for stratification. After that, plants are placed in one of the conditions that are described in Table 7 for 18-21 days. Conditions 1,3 and 4 are growth chambers (PHCb, cat. no MLR-352H-PA) while Condition 2 is a built as a plant growth room in a dark room in the building.

Table 7. Plant growth conditions used in this study.

Condition	Photoperiod	Light source	Light intensity	Temperature	Humidity
#1	10 h light/ 14 h dark	Panasonic (FL40SS/ENW/37) 5200K	4000 lux	21.3°C	53%
#2	16 h light / 8 h dark	Philips, TL- (D/36W/865) 6500 K	8400 lux	20.2 ± 5.0 °C	44%
#3	16 h light / 8 h dark	Panasonic (FL40SS/ENW/37) 5200K	9810 lux	21.3°C	53%
#4	16 h light / 8 h dark	Philips, TL- (D/36W/865) 6500 K	7620 lux	21.2°C	51%

7.4.2.

Protoplast isolation and culture

Protoplast isolation was performed following Chupeau et al. with minor modifications¹⁰⁸. All the tools used for protoplast and culture are wiped with 70% ethanol and individually wrapped with aluminium foil and put in an autoclave box, which is again wrapped with aluminium foil and autoclaved. All steps are performed until sterile conditions using laminar flow hood (Thermo Fisher, Heraguard™ ECO Clean Bench, cat. no 15158902) with the bench undergoing UV sterilization 30 mins before starting protoplasting. 3-week-old plants grown *in vitro* were harvested on a borosilicate glass plate containing 10 ml MGG medium to prevent drying of the leaves. Approximately 3 longitudinal and 3 diagonal cuts were made on each leaf with a 11-mm sterile scalpel (Swann-Morton, cat. no 0503) on the same plate containing MGG. Leaves were then placed in a borosilicate glass plate (6 cm or 9 cm diameter) containing cell wall digesting solution containing 0.1% w/v Cellulase Onozuka R10 (Yakult), 0.03% (w/v) Macerozyme R10 (Yakult) and 0.04% (w/v) Driselase (Sigma) in 10 ml MGG for a maximum of 14 h. After digestion is completed, the plate is gently shaken (around 60 rpm, with circular movements) to release the protoplasts from the digested leaves. The solution turning green indicates cell wall digestion was successful. Then, 10 ml MGG was put in a 50 ml tube (Falcon), and protoplasts were filtered through a 100 µm sterile cell strainer (Fisherbrand™, cat. no 11517532) to remove debris. Instead of centrifugation, gravitational sedimentation is used during isolation for three washes with MGG, which takes around 2.5 – 3 h to complete.

After 3 washes, protoplast number was counted with a hemocytometer under microscope with 10X magnification. Then MGG media was removed, and protoplasts were kept in the fridge (4°C for 1 h). Then, protoplast concentration was adjusted to 4×10^5 cells/ml in Protoplast Induction Medium (PIM), which is half of the Chupeau et al. protocol, due to the observed the high level of osmotic shock observed with the previously reported concentration. 400 µl of protoplast culture was added to 3.6 ml of PIM in µ-dishes (ibidi cat. no 81156), sealed with one layer of surgical tape and one layer of parafilm, making the culture cell concentration 4×10^4 cells/ml. One µ-dish was prepared per sampling day for daily imaging. If protoplasts are not imaged every day, then they are cultured in larger volume, 1 ml of protoplast solution was added on 9 ml of PIM media on a borosilicate glass plate (9 cm diameter). Protoplasts were cultured in darkness at 24°C. Protoplasts were always handled gently and wide orifice tips (Starlab, cat no. E1011-8400) were used when they need to be sampled.

All media for isolation and culture involved making micro and macro salts solutions, which are filter sterilized and stored in the fridge for one month before being replaced with freshly prepared

ones. It is crucial to dissolve folic acid in 1M NaOH (2 mg/ml) before its addition to the PIM, as it does not dissolve effectively otherwise.

7.4.3. Cell wall digestion assays

Leaves from *A. thaliana* plants grown in vitro for 3-weeks were harvested and cut in MGG solution with a 11-mm scalpel. Then they are placed in one of the four conditions:

- 1) only MGG, no enzymes
- 2) 0.01% (w/v) Cellulase Onozuka R10 (Yakult)
- 3) 0.01% (w/v) Cellulase Onozuka R10 (Yakult), 0.06% (w/v) Pectinase (Sigma, cat. no P2401-500U)
- 4) 0.01% (w/v) Cellulase Onozuka R10 (Yakult), 0.03% (w/v) Macerozyme (Yakult), 0.04% (w/v) Driselase (Sigma)

Leaves were digested for 12 hours in darkness at 20°C, and then digested leaves or released protoplasts are used for RNA extraction as described in Section 7.1.11.

7.4.4. Staining and microscopy

Viability staining was conducted with fluorescein diacetate (FDA, 50 µg/ml final concentration) and propidium iodide (PI, 5 µg/ml final concentration) dual staining. Cell wall staining was done with 0.5 µg/mL calcofluor white (Calcofluor White/Fluorescent Brightener 28, Sigma). Stains were kept as 100X stock solutions in the fridge, FDA is kept in the freezer. 100 µl of protoplast culture was taken for each stain, and following 10 minutes of incubation in darkness, samples were imaged using an inverted microscope.

Protoplast culture was imaged using Nikon Eclipse Ti using 10X or 20X magnification unless otherwise stated. A green filter (460 nm filter, 50 ms exposure), red filter (550 nm filter 1 s exposure) or blue filter (370 nm filter, 2 s exposure) was used to image autofluorescence and staining.

For quantitative analysis of cell wall signal with calcofluor stain and chloroplast autofluorescence, images were converted to 8-bit RGB on ImageJ. Then at least 30 protoplasts were chosen for both analyses. For calcofluor stain, protoplast periphery was selected for pixel intensity

quantification; and for chloroplast autofluorescence 10 points on 10 different chloroplasts in a protoplast was chosen for quantification.

For cell shape categorization protoplast periphery was selected on ImageJ. Circularity ratio between 1 and 0.995 is considered spherical; and circularity ratio below 0.995 is considered as elliptical. Protoplasts are considered budding when they have one or more buds or protrusions on the main protoplast.

7.5. *in silico* assays

7.5.1. Functional dissection of FAD2 and HSP terminators

Five length deletion series were constructed for *HSP* and *FAD2* terminators. The deletions removed putative motifs that could affect their function to influence the gene expression level. Putative Stabilisation/Destabilisation Motifs³⁵⁵ while the rest were identified using RegRNA2.0³⁵⁶ and PASPA³⁵⁷ online tools. 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 the *HSP* part3, while the same method was used to build *HSP* part 5.

Transformation of the construct was performed with PEG-transformation of *A. thaliana* leaf mesophyll protoplasts in 96-well plates. The expression activity of the constructs was assayed using a plate reader, 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.

7.5.2. RNA folding predictions

For mRNA folding predictions, combinations that were chosen for RNA isolation were also used for predicting the transcript folding. First transcription start site is predicted to be at 570 bp (-66 bp before the start codon) for *UBQ10* promoter by using TSSPlant³⁵⁸ and nucleotide sequence from 570 bp to

the end of the terminators are used as mRNA template. mRNA folding is predicted by using RNAFold software³⁵⁹.

7.5.3. *a priori* identification of novel terminator motifs

MEME suite³⁶⁰ as used for identifying motifs present in *HSP* and *FAD2* terminators, that lack in other terminators in our toolkit. EXSTREME motif discovery function was chosen from MEME suite, for identifying motifs of any length, that can be present in anywhere in the sequence. RNA motif query was made by choosing both *A. thaliana* and prokaryote RNA databases.

Appendix

Table 1. Primers used in this study to PCR amplify promoters.

Primer	Sequence (5' -> 3')
WOX13_F	TCACCTGCATATCTCTGGAGTGTGTCTCTATATTGTCCCGGTGG
WOX13_R	GCACCTGCATACTCACATTCATCTCTTTTGTGAGAATTTTCG
IAA8_F	TCACCTGCATATCTCTGGAGCCATTTTGACACCCCTAGTAACAAC
IAA8_R	ACACCTGCATATCTCACATTCAGATCTTCAATATGCAGGACTTG
WIND1_1F	TCACCTGCATATCTCTGGAGTGGAGTTCTCGGCCAATCC
WIND1_1R	ACACCTGCTTGGTGTCTACTCTTATTGGTTAATGGTTATAATCAT
WIND1_2F	TCACCTGCGTATGACACCAATAAAAGAGAGATGATTAAACACA
WIND1_2R	ACACCTGCCCTCAGCTGCTTGGCTCTTGTTTATATAG
WIND1_3F	GCACCTGCAAGCAGCTGAGGTTAGTTCATCGGTGA
WIND1_3R	GCACCTGCATATCTCACATTTAAAATAAAAAATTGAATCCGAAAAAATTGAACCTGAGAC
WOX5_1F	TCACCTGCATATCTCTGGAGCGGTTTGTGTTGACGAAGAGTAG
WOX5_1R	GGCGCACCTGCATTTGCTCTCTGATTAAGTTTTTTTT
WOX5_2F	TCACCTGCCAGAGAGCAAATTATTTGGTTATATGGT
WOX5_2R	GCACCTGCGTCTGACATCGTAATGTCCACTAACATTC
WOX5_3F	TATTACACCTGCACGATGTCAGACCGCCTTCAACTTTTTTGAC
WOX5_3R	ACACCTGCATATCTCACATTGTTTCAGATGTAAAGTCCTCAAC

Table 2. List of primer and probe sequences used for qRT-PCR experiments.

Name	Gene	Sequence (5' -> 3')
qPCR_nLUC_F	<i>nLUC</i>	AGGTGGTGTACCCTGTGGATGAT
qPCR_nLUC_R	<i>nLUC</i>	CATACGGCCGTCCGAAATA
qPCR_nLUC_probe	<i>nLUC</i>	TTAAGGTGATCCTGCACTATGGCACA (FAM/ZEN/BQ)
qPCR_fLUC_F	<i>fLUC</i>	GGAAAGACGATGACGGAAA
qPCR_fLUC_R	<i>fLUC</i>	CGGTACTTCGTCCACAAACA
qPCR_fLUC_probe	<i>fLUC</i>	CGTGGATTACGTCGCCAGTCAAGT (SUN/ZEN/BQ)
qPCR_sfGFP_F	<i>sfGFP</i>	AGCAGAAGAACGGCATCAA
qPCR_sfGFP_R	<i>sfGFP</i>	GGGTGTTCTGCTGGTAGTGGT
qPCR_sfGFP_probe	<i>sfGFP</i>	AACTTCAAGATCCGCCACAACGTT (FAM/ZEN/BQ)
qPCR_mScarlet_F	<i>mScarlet-I</i>	GATTGTACCCAGAAGATGGTGT
qPCR_mScarlet_R	<i>mScarlet-I</i>	CTGTTTCTTAGCCTTGATGTAGTC
qPCR_mScarlet_probe	<i>mScarlet-I</i>	TGGCATTGAGATTGAAGGATGGTGGT (SUN/ZEN/BQ)

qPCR_ACT2_F	<i>ACT2</i>	CGCTCTTTCTTTCCAAGCTCA
qPCR_ACT2_R	<i>ACT2</i>	TGGCCTACCAACAACACTGG
qPCR_ACT2_probe	<i>ACT2</i>	TGGCTGAGGCTGATGATATTCAACCAAT (SUN/ZEN/BQ)
qPCR_WIND1_F	<i>WIND1</i>	AGTCGCTGGATCTTGTTCG
qPCR_WIND1_R	<i>WIND1</i>	TCAAAAGTCCCAAGCCAGAG
qPCR_WIND1_probe	<i>WIND1</i>	CAAACGGATCTCAGCCACCCATTTTC (FAM/ZEN/BQ)
qPCR_IAA8_F	<i>IAA8</i>	TGAGTTCTGGGAACGATAAGATAAA
qPCR_IAA8_R	<i>IAA8</i>	ACTCAGGAAGACCTAGCCTCA
qPCR_IAA8_probe	<i>IAA8</i>	TAGCTTTGAAATTGAGATTGCTCTTCCCA (FAM/ZEN/BQ)
qPCR_WOX13_F	<i>WOX13</i>	CCGGAGGAGGAATGTACGTG
qPCR_WOX13_R	<i>WOX13</i>	CATTCTCCCTCCTGCAAGATC
qPCR_WOX13_probe	<i>WOX13</i>	TGATGACCGATGAGCAGTATGAACT (FAM/ZEN/BQ)
qPCR_WOX5_F	<i>WOX5</i>	CTCCAACCTCAAGGTGGACA
qPCR_WOX5_R	<i>WOX5</i>	TCGAAGATCTAATGGCGGTG
qPCR_WOX5_probe	<i>WOX5</i>	AGAGCTAGAGGCAATAACCAGTACCGTGA (FAM/ZEN/BQ)
qPCR_FNR1_F	<i>FNR1</i>	TACTACCGAGGCACCACCAG
qPCR_FNR1_R	<i>FNR1</i>	ATACGGAACCTCACCTTCGG
qPCR_FNR1_probe	<i>FNR1</i>	TCAAGGAGTCTAAGAAACAGGAAGAAGGGA (FAM/ZEN/BQ)
qPCR_CUE1_F	<i>CUE1</i>	GCCTGTTAATGCTTTCGGAAC
qPCR_CUE1_R	<i>CUE1</i>	AAGCAGTCTTTGGCTTTGGC
qPCR_CUE1_probe	<i>CUE1</i>	CGCCCTTGCGGGTGTCTTCTTATAC (FAM/ZEN/BQ)
qPCR_CA2_F	<i>CA2</i>	TCCTCTTGACGGAAACAACCTCT
qPCR_CA2_R	<i>CA2</i>	GACACATTCACTGCCTCCCTTTC
qPCR_CA2_probe	<i>CA2</i>	CGGCCACATTGGTCTTCAAATGC (FAM/ZEN/BQ)
qPCR_LHCB2.1_F	<i>LHCB2.1</i>	AGCCATTGTTACTGATGCAGGT
qPCR_LHCB2.1_R	<i>LHCB2.1</i>	TCCGTGTCCAGGCCTACTTT
qPCR_LHCB2.1_probe	<i>LHCB2.1</i>	ACAGAGAACATAACAGGGTGGTGTGGTT (FAM/ZEN/BQ)
qPCR_JMJ30_F	<i>JMJ30</i>	CGTCAATGAGGTGAGACATGTTT
qPCR_JMJ30_R	<i>JMJ30</i>	CAATGAGCCATAGAATTGGTTATAACA
qPCR_JMJ30_probe	<i>JMJ30</i>	TGTACCTGGCAGATAATAATCACGCAGA (FAM/ZEN/BQ)

8.1. Protoplast transformation protocol for transient expression

8.1.1.

Media preparation protocol:

1. Prepare MGG according to Chupeau et al., (2013)¹⁰⁴.
2. For MMM media follow the recipe on Table 3, pH unadjusted usually between 4.8-5.0.
3. For PEG media follow the recipe on Table 4. Dissolve Mannitol and $\text{Ca}(\text{NO}_3)_2 \cdot \text{H}_2\text{O}$ first in 6 ml Milli-Q water (or total volume – PEG mass directly used as volume). Add PEG gradually while mixing the solution. This way the solution will be ready in 15 minutes, as opposed to 2 hours when you combine everything at the same time. Adjust pH to 8 and confirm it with a pH strip as electrode-based pHmeters can be off due to high viscosity.

Table 3. MMM solution (100 ml)

Chemical	Concentration	Amount (g)
Mannitol	0.5 M	9.11
MgCl_2	15 mM	0.3
MES	0.1% (w/v)	0.1

pH unadjusted, between 4.8 - 5.0

Table 4. PEG solution (10 ml)

Chemical	Concentration	Amount (g)
Mannitol	0.4 M	0.72
$\text{Ca}(\text{NO}_3)_2 \cdot \text{H}_2\text{O}$	0.1 M	0.24
PEG 4000	40% (w/v)	4

pH 8

8.1.2.

Transformation protocol:

1. Leaves from plants grown in soil or *in vitro* for 4-8 weeks were collected in a 50 ml tube containing 20 ml MGG to prevent leaves from drying.
2. In a borosilicate plate containing 10 ml MGG transfer the leaves.
3. Cut the leaves with a scalpel. To make cutting easy, Keep the main vasculature intact and make cuts on both sides of the vasculature. Aim for 2-3 mm between cuts. When the whole leaf is cut, then make cuts on the vasculature. This way you can transfer the leaf as one piece while having cuts everywhere.
4. Transfer the leaf to the digestion plate (borosilicate, 20 cm) containing 50 ml MGG with cell wall digestion enzymes. Make sure the leaf surface is covered with the solution.
5. Repeat the same procedure for all the leaves, until the digestion plate surface is covered.
6. Wrap the digestion plate with a layer of surgical tape and a layer of parafilm. Wrap the plate with aluminium foil. Leave for digestion for 8-12 h at room temperature (20-24° C).

7. The solution should turn green with the release of protoplasts. Shake the digestion plate gently by moving it in circles and vertically (around 50 rpm).
8. Transfer the digestion solution containing protoplasts to 50 ml tube already containing 10 ml MGG. This is for cushioning the protoplasts. Use a 25 or 50 ml stripette.
9. Let protoplasts sediment for around 1 h. Protoplasts should be at the bottom and be dark green, the solution should be turning yellow towards the top.
10. Remove the solution at the top and keep the protoplasts at the bottom untouched.
11. Add 10 ml MGG on the protoplasts and gently resuspend protoplasts by tilting the tube.
12. Repeat steps 9-11 2 more times. Resuspending protoplasts should take less time, around 45 minutes for each wash.
13. While waiting for the protoplasts to sediment you can prepare the PEG solution and keep it at 30° C. You can prepare the DNA for the transformation. Put 8 µl DNA per well/tube for one transformation and repeat for the desired amount of transformations. Use a deep 96-well plate for transformation if you will be using a 96-well plate for measurements as the transformation volume per well is around 1.5 ml and the measurement plate will not be deep enough.
14. After the third wash, calculate the protoplast concentration using a haemocytometer.
15. Let the protoplasts sediment to change the solution. Should take around 40 minutes.
16. Remove the MGG solution and add MMM enough to make the protoplast concentration to reach 1×10^6 protoplasts/ml.
17. Put 75 µl protoplasts per well/tubes containing DNA. Use wide orifice tips (cat. No) or cut tips to handle protoplasts gently and not have pressure at the tip edge.
18. Add 83 µl PEG solution on protoplasts, incubate 1 minute.
19. Add 800 µl MGG on the protoplasts.
20. Cover the plate with xxx seal if using a plate. Wrap the plate/tubes with aluminium foil and incubate for 1 h at room temperature (20-24° C). If protoplasts stick to the side of the tube, it means they are mostly burst.
21. Protoplasts should be at the bottom of the plate/tube. If they are not you can use centrifuge at XXX RCF for 5 minutes (acceleration 2, decelerations 2).
22. Remove the supernatant, be careful not to disturb the protoplasts. It is okay to have some amount of liquid remaining.
23. Add 200 µl MGG to the protoplasts. If you are using a plate, cover it with a xx seal.
24. If you are doing time course measurements, put the 96-well plate in a plate reader directly, or you can measure the final timepoint after 24 h of incubation in darkness and at room temperature. Signals should be detectable around 5-6 h post-transformation and gradually increasing.

8.1.3.

Notes:

- Only use leaves from healthy plants. You can use all the true leaves.
- Usually, 3 large plants (6-8 weeks old) are enough for a 96-well transformation.
- This protocol is for small-scale transformation between 1-96 transformations for plate-reader experiments. For RNA extraction, you need to upscale, see Chapter 7.xx
- Three technical replicates per construct are recommended.

- You might need to do midiprep or multiple minipreps per construct. It is recommended to prepare the DNA for transformation and adjust the concentration to 500 ng/μl before the day of the transformation.
- Media preparation for MGG and MMM is recommended the day before the transformation. In contrast, for PEG, it is recommended on the day of transformation while waiting for protoplast sedimentation during protoplast washing.
- Transformation efficiency should be around 60-80%. If it is below 60%, you need to handle protoplasts gentler during transformation.
- You might start getting fungal contamination around 3 days post-transformation. If you need to measure the signal from protoplast on later days, you will need to add anti-fungal reagents.

Bibliography

1. Si, L. *et al.* OsSPL13 controls grain size in cultivated rice. *Nat Genet* **48**, 447–456 (2016).
2. Guo, W. *et al.* Overexpression of GmWRI1b in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. *Plant Biotechnology Journal* **18**, 1639–1641 (2020).
3. Zheng, M. *et al.* Knockout of two BnaMAX1 homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnology Journal* **18**, 644–654 (2020).
4. Uga, Y. *et al.* Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nat Genet* **45**, 1097–1102 (2013).
5. Li, Z. *et al.* Enhancing auxin accumulation in maize root tips improves root growth and dwarfs plant height. *Plant Biotechnology Journal* **16**, 86–99 (2018).
6. Khush, G. S. Green revolution: the way forward. *Nat Rev Genet* **2**, 815–822 (2001).
7. Shah, A. N. *et al.* Lodging stress in cereal—effects and management: an overview. *Environ Sci Pollut Res* **24**, 5222–5237 (2017).
8. Li, H., Johnson, P., Stepanova, A., Alonso, J. M. & Ecker, J. R. Convergence of Signaling Pathways in the Control of Differential Cell Growth in Arabidopsis. *Developmental Cell* **7**, 193–204 (2004).
9. Iwase, A. *et al.* The AP2/ERF transcription factor WIND1 controls cell dedifferentiation in Arabidopsis. *Curr Biol* **21**, 508–514 (2011).
10. Casanova-Sáez, R., Mateo-Bonmatí, E. & Ljung, K. Auxin Metabolism in Plants. *Cold Spring Harb Perspect Biol* **13**, a039867 (2021).
11. Durbak, A., Yao, H. & McSteen, P. Hormone signaling in plant development. *Current Opinion in Plant Biology* **15**, 92–96 (2012).

12. Benjamins, R. & Scheres, B. Auxin: The Looping Star in Plant Development. *Annual Review of Plant Biology* **59**, 443–465 (2008).
13. Weijers, D. & Wagner, D. Transcriptional Responses to the Auxin Hormone. *Annual Review of Plant Biology* **67**, 539–574 (2016).
14. Koltunow, A. M., Truettner, J., H.Cox, K., Wallroth, M. & B. Goldberg, R. Different Temporal and Spatial Gene Expression Patterns Occur during Anther Development. *Plant Cell* **2**, 1201–1224 (1990).
15. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**, 501–506 (2005).
16. Sparks, E., Wachsman, G. & Benfey, P. N. Spatiotemporal signalling in plant development. *Nat Rev Genet* **14**, 631–644 (2013).
17. Palovaara, J. *et al.* Transcriptome dynamics revealed by a gene expression atlas of the early *Arabidopsis* embryo. *Nature Plants* **3**, 894–904 (2017).
18. Shao, H.-B. *et al.* Understanding molecular mechanism of higher plant plasticity under abiotic stress. *Colloids and Surfaces B: Biointerfaces* **54**, 37–45 (2007).
19. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508–520 (2014).
20. Ebrahimkhani, M. R. & Ebisuya, M. Synthetic developmental biology: build and control multicellular systems. *Current Opinion in Chemical Biology* **52**, 9–15 (2019).
21. Zarkesh, I. *et al.* Synthetic developmental biology: Engineering approaches to guide multicellular organization. *Stem Cell Reports* **17**, 715–733 (2022).
22. Knight, T. *Idempotent Vector Design for Standard Assembly of Biobricks*:
<http://www.dtic.mil/docs/citations/ADA457791> (2003) doi:10.21236/ADA457791.
23. Casini, A., Storch, M., Baldwin, G. S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol* **16**, 568–576 (2015).

24. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE* **3**, e3647 (2008).
25. Patron, N. J. *et al.* Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist* **208**, 13–19 (2015).
26. Cai, Y.-M. *et al.* Rational design of minimal synthetic promoters for plants. *Nucleic Acids Research* **48**, 11845–11856 (2020).
27. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardised Assembly of Multigene Constructs. *PLoS ONE* **6**, e16765 (2011).
28. Sood, P., Bhattacharya, A. & Sood, A. Problems and possibilities of monocot transformation. *Biologia plant.* **55**, 1–15 (2011).
29. Chamness, J. C. *et al.* An Extensible Vector Toolkit and Parts Library for Advanced Engineering of Plant Genomes. <http://biorxiv.org/lookup/doi/10.1101/2022.10.15.511792> (2022)
doi:10.1101/2022.10.15.511792.
30. Sarrion-Perdigones, A. *et al.* GoldenBraid: An Iterative Cloning System for Standardised Assembly of Reusable Genetic Modules. *PLoS ONE* **6**, e21622 (2011).
31. Pollak, B. *et al.* Universal loop assembly: open, efficient and cross-kingdom DNA fabrication. *Synthetic Biology* **5**, ysaa001 (2020).
32. Rouches, M. V., Xu, Y., Cortes, L. B. G. & Lambert, G. A plasmid system with tunable copy number. *Nat Commun* **13**, 3908 (2022).
33. Andreou, A. I. & Nakayama, N. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS ONE* **13**, e0189892 (2018).
34. Andreou, A. I., Nirkko, J., Ochoa-Villarreal, M. & Nakayama, N. *Mobius Assembly for Plant Systems Highlights Promoter-Terminator Interaction in Gene Regulation.*

<http://biorxiv.org/lookup/doi/10.1101/2021.03.31.437819> (2021)

doi:10.1101/2021.03.31.437819.

35. Schaumberg, K. A. *et al.* Quantitative characterisation of genetic parts and circuits for plant synthetic biology. *Nat Methods* **13**, 94–100 (2016).
36. Wilmink, A., van de Ven, B. C. E. & Dons, J. J. M. Activity of constitutive promoters in various species from the Liliaceae. *Plant Mol Biol* **28**, 949–955 (1995).
37. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardised Assembly of Multigene Constructs. *PLoS ONE* **6**, e16765 (2011).
38. Jores, T. *et al.* Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat. Plants* **7**, 842–855 (2021).
39. Moreno-Giménez, E., Selma, S., Calvache, C. & Orzáez, D. GB_SynP: a modular dCas9-regulated synthetic promoter collection for fine-tuned recombinant gene expression in plants. (2022).
40. Curran, K. A. *et al.* Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol.* **4**, 824–832 (2015).
41. Li, X. Infiltration of *Nicotiana benthamiana* Protocol for Transient Expression via *Agrobacterium*. *BIO-PROTOCOL* **1**, (2011).
42. Yoo, S.-D., Cho, Y.-H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* **2**, 1565–1572 (2007).
43. Bashandy, H., Jalkanen, S. & Teeri, T. H. Within leaf variation is the largest source of variation in agroinfiltration of *Nicotiana benthamiana*. *Plant Methods* **11**, 47 (2015).
44. Nalapalli, S., Tunc-Ozdemir, M., Sun, Y., Elumalai, S. & Que, Q. Morphogenic Regulators and Their Application in Improving Plant Transformation. in *Rice Genome Engineering and Gene Editing: Methods and Protocols* (eds. Bandyopadhyay, A. & Thilmony, R.) 37–61 (Springer US, New York, NY, 2021). doi:10.1007/978-1-0716-1068-8_3.

45. Marquès-Bueno, M. M. *et al.* A versatile Multisite Gateway-compatible promoter and transgenic line collection for cell type-specific functional genomics in *Arabidopsis*. *The Plant Journal* **85**, 320–333 (2016).
46. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J* **88**, 1058–1070 (2016).
47. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. *An Integrase Toolbox to Record Gene-Expression during Plant Development*. <http://biorxiv.org/lookup/doi/10.1101/2022.09.16.508262> (2022) doi:10.1101/2022.09.16.508262.
48. Blazquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. LEAFY expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).
49. Long, J. A., Ohno, C., Smith, Z. R. & Meyerowitz, E. M. TOPLESS Regulates Apical Embryonic Fate in *Arabidopsis*. *Science* **312**, 1520–1523 (2006).
50. Wang, R. *et al.* Novel green tissue-specific synthetic promoters and cis-regulatory elements in rice. *Sci Rep* **5**, 18256 (2015).
51. Bai, J. *et al.* Comprehensive construction strategy of bidirectional green tissue-specific synthetic promoters. *Plant Biotechnology Journal* **18**, 668–678 (2020).
52. Mohan, C., Jayanarayanan, A. N. & Narayanan, S. Construction of a novel synthetic root-specific promoter and its characterisation in transgenic tobacco plants. *3 Biotech* **7**, 234 (2017).
53. Caddick, M. X. *et al.* An ethanol inducible gene switch for plants. **16**, (1998).
54. Zuo, J., Niu, Q.-W. & Chua, N.-H. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant Journal* **24**, 265–273 (2000).

55. Samalova, M., Brzobohaty, B. & Moore, I. pOp6/LhGR: a stringently regulated and highly responsive dexamethasone-inducible gene expression system for tobacco. *The Plant Journal* **41**, 919–935 (2005).
56. Randall, R. S. The plant AlcR-pAlcA ethanol-inducible system displays gross growth artefacts independently of downstream pAlcA-regulated inducible constructs. *Sci Rep* **11**, 2142 (2021).
57. Schürholz, A.-K. *et al.* A Comprehensive Toolkit for Inducible, Cell Type-Specific Gene Expression in Arabidopsis. *Plant Physiol.* **178**, 40–53 (2018).
58. Ochoa-Fernandez, R. *et al.* Optogenetic control of gene expression in plants in the presence of ambient white light. *Nat Methods* **17**, 717–725 (2020).
59. de Mena, L., Rizk, P. & Rincon-Limas, D. E. Bringing Light to Transcription: The Optogenetics Repertoire. *Frontiers in Genetics* **9**, (2018).
60. Pérez, A. L. A. *et al.* Optogenetic strategies for the control of gene expression in yeasts. *Biotechnology Advances* **54**, 107839 (2022).
61. Moller, B. & Weijers, D. Auxin Control of Embryo Patterning. *Cold Spring Harbor Perspectives in Biology* **1**, a001545–a001545 (2009).
62. Kieffer, M., Neve, J. & Kepinski, S. Defining auxin response contexts in plant development. *Current Opinion in Plant Biology* **13**, 12–20 (2010).
63. Kareem, A. *et al.* PLETHORA Genes Control Regeneration by a Two-Step Mechanism. *Current Biology* **25**, 1017–1030 (2015).
64. Brophy, J. A. N. *et al.* Synthetic genetic circuits as a means of reprogramming plant roots. *Science* **377**, 747–751 (2022).
65. Lloyd, J. P. B. *et al.* *Synthetic Memory Circuits for Programmable Cell Reconfiguration in Plants*. <http://biorxiv.org/lookup/doi/10.1101/2022.02.11.480167> (2022)
doi:10.1101/2022.02.11.480167.

66. Moradpour, M. & Abdulah, S. N. A. CRISPR/dCas9 platforms in plants: strategies and applications beyond genome editing. *Plant Biotechnology Journal* **18**, 32–44 (2020).
67. Pan, C. *et al.* CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. *Nat. Plants* **7**, 942–953 (2021).
68. Vazquez-Vilar, M. *et al.* The GB4.0 Platform, an All-In-One Tool for CRISPR/Cas-Based Multiplex Genome Engineering in Plants. *Front Plant Sci* **12**, 689937 (2021).
69. Lowder, L. G., Paul, J. W. & Qi, Y. Multiplexed Transcriptional Activation or Repression in Plants Using CRISPR–dCas9-Based Systems. in *Plant Gene Regulatory Networks: Methods and Protocols* (eds. Kaufmann, K. & Mueller-Roeber, B.) 167–184 (Springer, New York, NY, 2017).
doi:10.1007/978-1-4939-7125-1_12.
70. Shaw, W. M. *et al.* Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nat Commun* **13**, 4984 (2022).
71. Khakhar, A., Leydon, A. R., Lemmex, A. C., Klavins, E. & Nemhauser, J. L. Synthetic hormone-responsive transcription factors can monitor and re-program plant development. *eLife* **7**, e34702 (2018).
72. Prusinkiewicz, P. *et al.* Control of bud activation by an auxin transport switch. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17431–17436 (2009).
73. Refahi, Y. *et al.* A stochastic multicellular model identifies biological watermarks from disorders in self-organized patterns of phyllotaxis. *eLife* **5**, e14093 (2016).
74. Papanatsiou, M. *et al.* Optogenetic manipulation of stomatal kinetics improves carbon assimilation, water use, and growth. *Science* **363**, 1456–1459 (2019).
75. Brophy, J. A. N. *et al.* Synthetic genetic circuits as a means of reprogramming plant roots. *Science* **377**, 747–751 (2022).

76. Chamness, J. C. *et al.* *An Extensible Vector Toolkit and Parts Library for Advanced Engineering of Plant Genomes*. <http://biorxiv.org/lookup/doi/10.1101/2022.10.15.511792> (2022)
doi:10.1101/2022.10.15.511792.
77. Kallam, K. *et al.* *Tunable Control of Insect Pheromone Biosynthesis in Nicotiana Benthiana*.
<http://biorxiv.org/lookup/doi/10.1101/2022.06.15.496242> (2022)
doi:10.1101/2022.06.15.496242.
78. Finnegan, J. & McElroy, D. Transgene Inactivation: Plants Fight Back! *Nat Biotechnol* **12**, 883–888 (1994).
79. Dietz-Pfeilstetter, A. Stability of transgene expression as a challenge for genetic engineering. *Plant Science* **179**, 164–167 (2010).
80. Gelvin, S. B. Integration of Agrobacterium T-DNA into the Plant Genome. *Annual Review of Genetics* **51**, 195–217 (2017).
81. Cabrera, A. *et al.* The sound of silence: Transgene silencing in mammalian cell engineering. *Cell Systems* **13**, 950–973 (2022).
82. Danilo, B., Perrot, L., Botton, E., Nogu  , F. & Mazier, M. The DFR locus: A smart landing pad for targeted transgene insertion in tomato. *PLOS ONE* **13**, e0208395 (2018).
83. Dhiman, H., Campbell, M., Melcher, M., Smith, K. D. & Borth, N. Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. *Comput Struct Biotechnol J* **18**, 3632–3648 (2020).
84. Miura, K. *et al.* OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat Genet* **42**, 545–549 (2010).
85. Jiao, Y. *et al.* Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat Genet* **42**, 541–544 (2010).

86. Lu, Z. *et al.* Genome-Wide Binding Analysis of the Transcription Activator IDEAL PLANT ARCHITECTURE1 Reveals a Complex Network Regulating Rice Plant Architecture. *The Plant Cell* **25**, 3743–3759 (2013).
87. Zhang, L. *et al.* A natural tandem array alleviates epigenetic repression of IPA1 and leads to superior yielding rice. *Nat Commun* **8**, 14789 (2017).
88. Wang, Y. *et al.* Comparing the effects of GA-responsive dwarfing genes Rht13 and Rht8 on plant height and some agronomic traits in common wheat. *Field Crops Research* **179**, 35–43 (2015).
89. Jatayev, S. *et al.* Green revolution ‘stumbles’ in a dry environment: Dwarf wheat with Rht genes fails to produce higher grain yield than taller plants under drought. *Plant, Cell & Environment* **43**, 2355–2364 (2020).
90. Soyk, S. *et al.* Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene. *Cell* **169**, 1142–1155.e12 (2017).
91. Lemmon, Z. H. *et al.* Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature Plants* **4**, 766–770 (2018).
92. Melnyk, C. W. Plant grafting: insights into tissue regeneration. *Regeneration* **4**, 3–14 (2017).
93. Ritchie, S. & Hodges, T. Cell culture and Regeneration of Transgenic Plants. in *Transgenic Plants: Engineering and Utilization* vol. 1 147–148 (1993).
94. Kulak, V., Longboat, S., Brunet, N. D., Shukla, M. & Saxena, P. In Vitro Technology in Plant Conservation: Relevance to Biocultural Diversity. *Plants* **11**, 503 (2022).
95. Drenth, A. & Kema, G. The Vulnerability of Bananas to Globally Emerging Disease Threats. *Phytopathology*® **111**, 2146–2161 (2021).
96. Efroni, I. *et al.* Root Regeneration Triggers an Embryo-like Sequence Guided by Hormonal Interactions. *Cell* **165**, 1721–1733 (2016).

97. Reinhardt, D., Frenz, M., Mandel, T. & Kuhlemeier, C. Microsurgical and laser ablation analysis of interactions between the zones and layers of the tomato shoot apical meristem. *Development* **130**, 4073–4083 (2003).
98. Devi, T. R., Dasgupta, M., Sahoo, M. R., Kole, P. C. & Prakash, N. High efficient de novo root-to-shoot organogenesis in Citrus jambhiri Lush.: Gene expression, genetic stability and virus indexing. *PLoS ONE* **16**, e0246971 (2021).
99. Kareem, A. *et al.* Protocol: a method to study the direct reprogramming of lateral root primordia to fertile shoots. *Plant Methods* **12**, 27 (2016).
100. Liang, H. *et al.* Shoot organogenesis and somatic embryogenesis from leaf and root explants of *Scaevola sericea*. *Sci Rep* **10**, 11343 (2020).
101. Hu, B. *et al.* Divergent regeneration-competent cells adopt a common mechanism for callus initiation in angiosperms. *Regeneration* **4**, 132–139 (2017).
102. Casimiro, I. *et al.* Auxin Transport Promotes Arabidopsis Lateral Root Initiation.
103. Zhu, Z.-X. *et al.* A Gain-of-Function Mutation in OsIAA11 Affects Lateral Root Development in Rice. *Molecular Plant* **5**, 154–161 (2012).
104. Sugimoto, K., Jiao, Y. & Meyerowitz, E. M. Arabidopsis Regeneration from Multiple Tissues Occurs via a Root Development Pathway. *Developmental Cell* **18**, 463–471 (2010).
105. Grafi, G. How cells dedifferentiate: a lesson from plants. *Developmental Biology* **268**, 1–6 (2004).
106. Lowe, K. *et al.* Rapid genotype “independent” Zea mays L. (maize) transformation via direct somatic embryogenesis. *In Vitro Cell.Dev.Biol.-Plant* **54**, 240–252 (2018).
107. Long, Y., Yang, Y., Pan, G. & Shen, Y. New Insights Into Tissue Culture Plant-Regeneration Mechanisms. *Front. Plant Sci.* **13**, 926752 (2022).
108. Chupeau, M.-C. *et al.* Characterisation of the Early Events Leading to Totipotency in an *Arabidopsis* Protoplast Liquid Culture by Temporal Transcript Profiling. *The Plant Cell* **25**, 2444–2463 (2013).

109. Pasternak, T., Paponov, I. A. & Kondratenko, S. Optimizing Protocols for Arabidopsis Shoot and Root Protoplast Cultivation. *Plants* **10**, 375 (2021).
110. Sakamoto, Y. *et al.* Transcriptional activation of auxin biosynthesis drives developmental reprogramming of differentiated cells. *The Plant Cell* **34**, 4348–4365 (2022).
111. Iwase, A., Ohme-Takagi, M. & Sugimoto, K. WIND1: A key molecular switch for plant cell dedifferentiation. *Plant Signaling & Behavior* **6**, 1943–1945 (2011).
112. Zhao, J. *et al.* Two Phases of Chromatin Decondensation during Dedifferentiation of Plant Cells. *Journal of Biological Chemistry* **276**, 22772–22778 (2001).
113. Williams, L. *et al.* Chromatin reorganization accompanying cellular dedifferentiation is associated with modifications of histone H3, redistribution of HP1, and activation of E2F-target genes. *Developmental Dynamics* **228**, 113–120 (2003).
114. Xu, M., Du, Q., Tian, C., Wang, Y. & Jiao, Y. Stochastic gene expression drives mesophyll protoplast regeneration. *Sci. Adv.* **7**, eabg8466 (2021).
115. Kuki, H. *et al.* Quantitative confocal imaging method for analysing cellulose dynamics during cell wall regeneration in Arabidopsis mesophyll protoplasts. *Plant Direct* **1**, e00021 (2017).
116. Sarkar, A. K. *et al.* Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature* **446**, 811–814 (2007).
117. Sakakibara, K. *et al.* WOX13 like genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*. *Development* **141**, 1660–1670 (2014).
118. Ikeuchi, M. *et al.* Wound-inducible WUSCHEL-RELATED HOMEODOMAIN 13 is required for callus growth and organ reconnection. *Plant Physiology* **188**, 425–441 (2022).
119. Ogura, N. *et al.* WUSCHEL-RELATED HOMEODOMAIN 13 suppresses de novo shoot regeneration via cell fate control of pluripotent callus. *Sci. Adv.* **9**, eadg6983 (2023).

120. Arase, F. *et al.* IAA8 Involved in Lateral Root Formation Interacts with the TIR1 Auxin Receptor and ARF Transcription Factors in Arabidopsis. *PLoS ONE* **7**, e43414 (2012).
121. Ikeuchi, M., Sugimoto, K. & Iwase, A. Plant Callus: Mechanisms of Induction and Repression. *Plant Cell* **25**, 3159–3173 (2013).
122. Su, W., Xu, M., Radani, Y. & Yang, L. Technological Development and Application of Plant Genetic Transformation. *IJMS* **24**, 10646 (2023).
123. Ikeuchi, M. *et al.* Molecular Mechanisms of Plant Regeneration. *Annu. Rev. Plant Biol.* **70**, 377–406 (2019).
124. Yue, J.-J. *et al.* Protoplasts: From Isolation to CRISPR/Cas Genome Editing Application. *Front. Genome Ed.* **3**, 717017 (2021).
125. Pattanayak', A. & Sarma, C. M. 'PLANT REGENERATION FROM PROTOPLASTS OF RICE: RELATIVE EFFICIENCY OF DIFFERENT SOURCES OF PROTOPLASTS.
126. Rhodes, C. & Gray, D. Transformation and regeneration of maize protoplasts. in *Plant Tissue Culture Manual* vol. 1 79–91 (Springer, 1992).
127. Kihara, M. & Funatsuki', H. Fertile plant regeneration from barley (*Hordeum vuzdgate* L.) protoplasts isolated from primary calluses.
128. Brandt, K. M., Gunn, H., Moretti, N. & Zemetra, R. S. A Streamlined Protocol for Wheat (*Triticum aestivum*) Protoplast Isolation and Transformation With CRISPR-Cas Ribonucleoprotein Complexes. *Front. Plant Sci.* **11**, 769 (2020).
129. Reed, K. M. & Bargmann, B. O. R. Protoplast Regeneration and Its Use in New Plant Breeding Technologies. *Front. Genome Ed.* **3**, 734951 (2021).
130. Flint-Garcia, S. A. Genetics and Consequences of Crop Domestication. *J. Agric. Food Chem.* **61**, 8267–8276 (2013).

131. Houston, K., Tucker, M. R., Chowdhury, J., Shirley, N. & Little, A. The Plant Cell Wall: A Complex and Dynamic Structure As Revealed by the Responses of Genes under Stress Conditions. *Front. Plant Sci.* **7**, (2016).
132. Rui, Y. & Dinneny, J. R. A wall with integrity: surveillance and maintenance of the plant cell wall under stress. *New Phytologist* **225**, 1428–1439 (2020).
133. Ferrari, S. Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development. *Front. Plant Sci.* **4**, (2013).
134. González-Pérez, L. *et al.* In tobacco BY-2 cells xyloglucan oligosaccharides alter the expression of genes involved in cell wall metabolism, signalling, stress responses, cell division and transcriptional control. *Mol Biol Rep* **41**, 6803–6816 (2014).
135. Souza, C. D. A. *et al.* Cellulose-Derived Oligomers Act as Damage-Associated Molecular Patterns and Trigger Defense-Like Responses. *Plant Physiol.* **173**, 2383–2398 (2017).
136. Durand-Smet, P., Spelman, T. A., Meyerowitz, E. M. & Jönsson, H. Cytoskeletal organization in isolated plant cells under geometry control. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17399–17408 (2020).
137. Landrein, B. *et al.* Mechanical stress contributes to the expression of the STM homeobox gene in Arabidopsis shoot meristems. *eLife* **4**, e07811 (2015).
138. Peaucelle, A. *et al.* Pectin-Induced Changes in Cell Wall Mechanics Underlie Organ Initiation in Arabidopsis. *Current Biology* **21**, 1720–1726 (2011).
139. Papadakis, A. K. & Roubelakis-Angelakis, K. A. Oxidative stress could be responsible for the recalcitrance of plant protoplasts. *Plant Physiology and Biochemistry* **40**, 549–559 (2002).
140. Tiew, T. W.-Y., Sheahan, M. B. & Rose, R. J. Peroxisomes contribute to reactive oxygen species homeostasis and cell division induction in Arabidopsis protoplasts. *Front. Plant Sci.* **6**, (2015).
141. Allard, R. *Principles of Plant Breeding*. (John Wiley & Sons, 1999).

142. Paul, N. C. *et al.* Plant and Fungal Genome Editing to Enhance Plant Disease Resistance Using the CRISPR/Cas9 System. *Front. Plant Sci.* **12**, 700925 (2021).
143. Ahmed, R. F., Irfan, M., Shakir, H. A., Khan, M. & Chen, L. Engineering drought tolerance in plants by modification of transcription and signalling factors. *Biotechnology & Biotechnological Equipment* **34**, 781–789 (2020).
144. Zhu, Q. *et al.* Development of “Purple Endosperm Rice” by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System. *Molecular Plant* **10**, 918–929 (2017).
145. Li, H., Johnson, P., Stepanova, A., Alonso, J. & Ecker, J. Convergence of signaling pathways in the control of differential cell growth in Arabidopsis. *Developmental Cell* **7**, 193–204 (2004).
146. Casanova-Saez, R., Mateo-Bonmati, E. & Ljung, K. Auxin metabolism in plants. *Cold Spring Harbor Perspectives in Biology* **13**,.
147. Iwase, A. *et al.* The AP2/ERF Transcription Factor WIND1 Controls Cell Dedifferentiation in Arabidopsis. *Current Biology* **21**, 508–514 (2011).
148. Durbak, A., Yao, H. & McSteen, P. Hormone signalling in plant development. *Current opinion in Plant Biology* **15**, 92–96 (2012).
149. Benjamins, R. & Scheres, B. Auxin: the looping star in plant development. *Annual Review of Plant Biology* **59**, 443–465 (2008).
150. Weijers, D. & Wagner, D. Transcriptional responses to the auxin hormone. *Annual Review of Plant Biology* **67**, 539–574 (2016).
151. Koltunow, A., Truettner, J., Cox, K., Wallroth, M. & Goldberg, R. Different temporal and spatial gene expression patterns occur during anther development. *The Plant Cell* **2**, 1201–1224 (1990).
152. Sparks, E., Wachsman, G. & Benfey, P. Spatiotemporal signalling in plant development. *Nature Reviews Genetics* **14**, 631–644.

153. Shao, H.-B. *et al.* Understanding molecular mechanism of higher plant plasticity under abiotic stress. *Colloids and Surfaces B:Biointerfaces* **54**, 37–45 (2007).
154. Cai, Y.-M., Lopez, J. C. & Patron, N. Phytobricks: Manual and Automated Assembly of Constructs for Engineering Plants. *Methods in Molecular Biology* **2205**, 179–199 (2020).
155. Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY* **162**, 1618–1631 (2013).
156. Andreou, A. & Nakayama, N. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* **13**, (2018).
157. Pollak, B. *et al.* Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol* **222**, 628–640 (2019).
158. Andreou, A. I., Nirkko, J., Ochoa-Villarreal, M. & Nakayama, N. *Mobius Assembly for Plant Systems Highlights Promoter-Terminator Interaction in Gene Regulation*.
<http://biorxiv.org/lookup/doi/10.1101/2021.03.31.437819> (2021)
[doi:10.1101/2021.03.31.437819](https://doi.org/10.1101/2021.03.31.437819).
159. Wilmlink, A., van de Ven, B. C. E. & Dons, J. J. M. Activity of constitutive promoters in various species from the Liliaceae. *Plant Mol Biol* **28**, 949–955 (1995).
160. Cai, Y.-M. *et al.* Rational design of minimal synthetic promoters for plants. *Nucleic Acids Research* **48**, 11845–11856 (2020).
161. Curran, K. A. *et al.* Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol.* **4**, 824–832 (2015).
162. Strader, L., Weijers, D. & Wagner, D. Plant transcription factors — being in the right place with the right company. *Current Opinion in Plant Biology* **65**, 102136 (2022).

163. Mithra, A., Kulkarni, K. & Srinivasan, R. Plant Promoters: Characterisation and Applications in Transgenic Technology. in *Planr Biotechnology: Principles and Applications* (Springer Nature Singapore Pte Ltd., 2017).
164. Jores, T. *et al.* Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat. Plants* **7**, 842–855 (2021).
165. Kumari, S. & Ware, D. Genome-Wide Computational Prediction and Analysis of Core Promoter Elements across Plant Monocots and Dicots. *PLoS One* **8**, (2013).
166. De Felippes, F. F. & Waterhouse, P. M. Plant terminators: the unsung heroes of gene expression. *Journal of Experimental Botany* **74**, 2239–2250 (2023).
167. Porrua, O., Boudvillian, M. & Libri, D. Transcription Termination: Variations on Common Themes. *Trends in Genetics* **32**, (2016).
168. Eaton, J. D., Francis, L., Davidson, L. & West, S. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145 (2020).
169. Gorjifard, S. *et al.* *Features That Govern Terminator Strength in Plants*.
<http://biorxiv.org/lookup/doi/10.1101/2023.06.16.545379> (2023)
doi:10.1101/2023.06.16.545379.
170. Passmore, L. & Collier, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nature reviews molecular cell biology* **23**, (2022).
171. Felippes, F. *et al.* The key role of terminators on the expression and post-transcriptional gene silencing of transgenes. *The Plant Journal* **104**, 96–112 (2020).
172. De Felippes, F. F., Shand, K. & Waterhouse, P. M. Identification of a Transferrable Terminator Element That Inhibits Small RNA Production and Improves Transgene Expression Levels. *Front. Plant Sci.* **13**, 877793 (2022).

173. Luo, Z. & Chen, Z. Improperly Terminated, Unpolyadenylated mRNA of Sense Transgenes Is Targeted by RDR6-Mediated RNA Silencing in *Arabidopsis*. *The Plant Cell* **19**, 943–958 (2007).
174. Peyret, H., Brown, J. K. M. & Lomonossoff, G. P. Improving plant transient expression through the rational design of synthetic 5' and 3' untranslated regions. *Plant Methods* **15**, 108 (2019).
175. Meshcheriakova, Y. A., Saxena, P. & Lomonossoff, G. P. Fine-tuning levels of heterologous gene expression in plants by orthogonal variation of the untranslated regions of a nonreplicating transient expression system. *Plant Biotechnology Journal* **12**, 718–727 (2014).
176. Zhang, T., Wu, A., Yue, Y. & Zhao, Y. uORFs: Important Cis-Regulatory Elements in Plants. *IJMS* **21**, 6238 (2020).
177. Xue, C. *et al.* Tuning plant phenotypes by precise, graded downregulation of gene expression. *Nat Biotechnol* **41**, 1758–1764 (2023).
178. Zuber, P. K. *et al.* The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *eLife* **7**, e36349 (2018).
179. Westin, L., Blomquist, P., Milligan, J. F. & Wrangé, Ö. Triple helix DNA alters nucleosomal histone-DNA interactions and acts as a nucleosome barrier. *Nucl Acids Res* **23**, 2184–2191 (1995).
180. Chen, L. *et al.* R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Molecular Cell* **68**, 745-757.e5 (2017).
181. Li, L. *et al.* YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **17**, 161–168 (2021).
182. Faure, G., Ogurtsov, A. Y., Shabalina, S. A. & Koonin, E. V. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* **44**, 10898–10911 (2016).
183. Tuller, T., Waldman, Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *PNAS* **107**, 3645–3650 (2010).

184. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
185. Gu, W., Zhou, T. & Wilke, C. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLoS Computational Biology* **6**, (2010).
186. Cambray, G., Guimaraes, J. C. & Arkin, A. P. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat Biotechnol* **36**, 1005–1015 (2018).
187. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
188. Loke, J. C. *et al.* Compilation of mRNA Polyadenylation Signals in Arabidopsis Revealed a New Signal Element and Potential Secondary Structures. *Plant Physiology* **138**, 1457–1468 (2005).
189. Chung, B. Y. W. *et al.* An RNA thermoswitch regulates daytime growth in Arabidopsis. *Nat. Plants* **6**, 522–532 (2020).
190. Gawroński, P. *et al.* Light-Dependent Translation Change of Arabidopsis psbA Correlates with RNA Structure Alterations at the Translation Initiation Region. *Cells* **10**, 322 (2021).
191. Miras, M., Miller, W. A., Truniger, V. & Aranda, M. A. Non-canonical Translation in Plant RNA Viruses. *Front. Plant Sci.* **8**, (2017).
192. Donaldson, L. Autofluorescence in Plants. *MDPI, Molecules* **25**, (2020).
193. Borghi, L. Inducible Gene Expression Systems for Plants. *Plant Developmental Biology* **655**, 65–75 (2010).
194. Craft, J. *et al.* New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *The Plant Journal* **41**, 899–918 (2005).
195. Schlücking, K. *et al.* 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. *Molecular Plant* **6**, 1814–1829 (2013).

196. Chełkowska-Pauszek, A. *et al.* The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria. *IJMS* **22**, 7845 (2021).
197. Pasin, F. *et al.* Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synth. Biol.* **6**, 1962–1968 (2017).
198. Tian, C., Zhang, Y., Li, J. & Wang, Y. Benchmarking Intrinsic Promoters and Terminators for Plant Synthetic Biology Research. *BioDesign Research* **2022**, 1–12 (2022).
199. Indian Institute of Integrative Medicine, CSIR, Canal Road, Jammu-180001 & Singh, R. A report on DNA sequence determinants in gene expression. *Bioinformation* **16**, 422 (2020).
200. Zarudnaya, M. I. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Research* **31**, 1375–1386 (2003).
201. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat Methods* **15**, 387–393 (2018).
202. Bienick, M. S. *et al.* The Interrelationship between Promoter Strength, Gene Expression, and Growth Rate. *PLoS ONE* **9**, e109105 (2014).
203. Karim, A. S., Curran, K. A. & Alper, H. S. Characterisation of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res* **13**, 107–116 (2013).
204. Li, Z. & Rinas, U. Recombinant protein production-associated metabolic burden reflects anabolic constraints and reveals similarities to a carbon overfeeding response. *Biotechnology and Bioengineering* **118**, (2021).
205. Gorochowski, T. E. *et al.* Absolute quantification of translational regulation and burden using combined sequencing approaches. *Molecular Systems Biology* **15**, e8719 (2019).
206. Di Blasi, R. *et al.* Resource-aware construct design in mammalian cells. *Nat Commun* **14**, 3576 (2023).

207. Dalos, D., Morais, E. G., Oliveira e Silva, K. F. & Rhys Williams, T. C. Cell-type-specific metabolism in plants. *The Plant Journal* **114**, 1093–1114 (2023).
208. Sleight, S. C. & Sauro, H. M. Visualization of Evolutionary Stability Dynamics and Competitive Fitness of Escherichia coli Engineered with Randomized Multigene Circuits. *ACS Synth. Biol.* **2**, 519–528 (2013).
209. Castle, S. D., Grierson, C. S. & Gorochowski, T. E. Towards an engineering theory of evolution. *Nat Commun* **12**, 3326 (2021).
210. Silva, F., Queiroz, J. A. & Domingues, F. C. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnology Advances* **30**, 691–708 (2012).
211. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J Biol Eng* **4**, 12 (2010).
212. Brödel, A. K., Rodrigues, R., Jaramillo, A. & Isalan, M. Accelerated evolution of a minimal 63–amino acid dual transcription factor. *Sci. Adv.* **6**, eaba2728 (2020).
213. Zeller, G. *et al.* Stress-induced changes in the Arabidopsis thaliana transcriptome analysed using whole-genome tiling arrays. *The Plant Journal* **58**, 068–1082 (2009).
214. Tittel-Elmer, M. *et al.* Stress-Induced Activation of Heterochromatic Transcription. *PLoS Genet* **6**, e1001175 (2010).
215. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* **21**, 526–540 (2020).
216. Heieck, K., Arnold, N. D. & Brück, T. B. Metabolic stress constrains microbial L-cysteine production in Escherichia coli by accelerating transposition through mobile genetic elements. *Microb Cell Fact* **22**, 10 (2023).
217. Munoz-Lopez, M. & Garcia-Perez, J. DNA Transposons: Nature and Applications in Genomics. *CG* **11**, 115–128 (2010).

218. Arkhipova, I. R. Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. *Molecular Biology and Evolution* **35**, 1332–1337 (2018).
219. Mhiri, C., Borges, F. & Grandbastien, M.-A. Specificities and Dynamics of Transposable Elements in Land Plants. *Biology* **11**, 488 (2022).
220. Birkenbihl, R. P. & Vielmetter, W. Complete maps of IS1, IS2, IS3, IS4, IS5, IS30 and IS150 locations in *Escherichia coli* K12. *Molec. Gen. Genet.* **220**, 147–153 (1989).
221. Siguier, P., Gournay, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**, 865–891 (2014).
222. Tavakoli, N. P. Tipping the balance between replicative and simple transposition. *The EMBO Journal* **20**, 2923–2930 (2001).
223. Shiga, Y., Sekine, Y. & Ohtsubo, E. Transposition of IS 1 circles. *Genes to Cells* **4**, 551–561 (1999).
224. Lewis, L. A. *et al.* The Left End of IS2 : a Compromise between Transpositional Activity and an Essential Promoter Function That Regulates the Transposition Pathway. *J Bacteriol* **186**, 858–865 (2004).
225. Crozat, E., Fournes, F., Cornet, F., Hallet, B. & Rousseau, P. Resolution of Multimeric Forms of Circular Plasmids and Chromosomes. *Microbiol Spectr* **2**, 2.5.37 (2014).
226. Goebel, W. & Helinski, D. GENERATION OF HIGHER MULTIPLE CIRCULAR DNA FORMS IN BACTERIA. *PNAS* **61**, (1968).
227. Summers, D. K., Beton, C. W. H. & Withers, H. L. Multicopy plasmid instability: the dimer catastrophe hypothesis. *Molecular Microbiology* **8**, 1031–1038 (1993).
228. Williams, D. R. & Thomas, C. M. Active partitioning of bacterial plasmids. *Journal of General Microbiology* **138**, 1–16 (1992).
229. Shao, B. *et al.* Single-cell measurement of plasmid copy number and promoter activity. *Nat Commun* **12**, 1475 (2021).

230. Million-Weaver, S. & Camps, M. Mechanisms of plasmid segregation: Have multicopy plasmids been overlooked? *Plasmid* **75**, 27–36 (2014).
231. Summers, D. & Sherratt, D. Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *EMBO* **7**, 851–858 (1988).
232. Field, C. & Summers, D. Multicopy plasmid stability: revisiting the dimer catastrophe. *Journal of Theoretical Biology* **291**, 119–127 (2011).
233. Allen, J. R., Torres-Acosta, M. A., Mohan, N., Lye, G. J. & Ward, J. M. Segregationally stabilised plasmids improve production of commodity chemicals in glucose-limited continuous fermentation. *Microb Cell Fact* **21**, 229 (2022).
234. Mazin, A., Timchenko, T., Saparbaev, M. & Mazina, O. Dimerization of plasmid DNA accelerates selection for antibiotic resistance. *Molecular microbiology* **20**, 101–108 (1996).
235. Aminov, R. I. Horizontal Gene Exchange in Environmental Microbiota. *Front. Microbio.* **2**, (2011).
236. Cornet, F., Mortier, I., Patte, J. & Louarn, J. M. Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif*. *J Bacteriol* **176**, 3188–3195 (1994).
237. Dery, K. J. *et al.* Characterisation of the Replication and Mobilization Regions of the Multiresistance *Klebsiella pneumoniae* Plasmid pJHCMW1. *Plasmid* **38**, 97–105 (1997).
238. Blaby, I. K. & Summers, D. K. The role of FIS in the Rcd checkpoint and stable maintenance of plasmid ColE1. *Microbiology* **155**, 2676–2682 (2009).
239. French, C. & Ward, J. M. Improved production and stability of *E. coli* recombinants expressing transketolase for large scale biotransformation. *Biotechnol Lett* **17**, 247–252 (1995).
240. Watson, M. R. *et al.* An Improved Binary Vector and *Escherichia coli* Strain for *Agrobacterium tumefaciens* -Mediated Plant Transformation. *G3 Genes/Genomes/Genetics* **6**, 2195–2201 (2016).

241. Brkljacic, J. *et al.* Frequency, composition and mobility of *Escherichia coli* derived transposable elements in holdings of plasmid repositories. *Microbial Biotechnology* **15**, 455–468 (2022).
242. Bao, L., Menon, P. N. K., Liljeruhm, J. & Forster, A. C. Overcoming chromoprotein limitations by engineering a red fluorescent protein. *Analytical Biochemistry* **611**, 113936 (2020).
243. Cao, J. *et al.* Epigenetic and chromosomal features drive transposon insertion in *Drosophila melanogaster*. *Nucleic Acids Research* **51**, 2066–2086 (2023).
244. Brophy, J. A. N. *et al.* Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat Microbiol* **3**, 1043–1053 (2018).
245. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record gene-expression during plant development. *Nat Commun* **14**, 1844 (2023).
246. Shiloach, J., Reshamwala, S., Noronha, S. B. & Negrete, A. Analysing metabolic variations in different bacterial strains, historical perspectives and current trends – example *E. coli*. *Current Opinion in Biotechnology* (2010).
247. Tack, D. S., Tonner, P. D., Musteata, E., Paralanov, V. & Ross, D. *Improved Stability of an Engineered Function Using Adapted Bacterial Strains*.
<http://biorxiv.org/lookup/doi/10.1101/2020.03.05.979385> (2020)
doi:10.1101/2020.03.05.979385.
248. Pósfai, G. *et al.* Emergent Properties of Reduced-Genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006).
249. Umenhoffer, K. *et al.* Genome-Wide Abolishment of Mobile Genetic Elements Using Genome Shuffling and CRISPR/Cas-Assisted MAGE Allows the Efficient Stabilization of a Bacterial Chassis. *ACS Synth. Biol.* **6**, 1471–1483 (2017).
250. Kocaoglan, E. G., Radhakrishnan, D. & Nakayama, N. Synthetic developmental biology: molecular tools to re-design plant shoots and roots. *Journal of Experimental Botany* **74**, 3864–3876 (2023).

251. Su, W., Xu, M., Radani, Y. & Yang, L. Technological Development and Application of Plant Genetic Transformation. *IJMS* **24**, 10646 (2023).
252. Faraco, M., Di Sansebastiano, G. P., Spelt, K., Koes, R. E. & Quattrocchio, F. M. One Protoplast Is Not the Other! *Plant Physiology* **156**, 474–478 (2011).
253. Birnbaum, K. D. & Alvarado, A. S. Slicing across Kingdoms: Regeneration in Plants and Animals. *Cell* **132**, 697–710 (2008).
254. Licausi, F., Ohme-Takagi, M. & Perata, P. APETALA 2/Ethylene Responsive Factor (AP 2/ ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist* **199**, 639–649 (2013).
255. Zhou, C., Guo, J., Feng, Z., Cui, X. & Zhu, J. Molecular characterisation of a novel AP2 transcription factor ThWIND1-L from *Thellungiella halophila*. *Plant Cell Tiss Organ Cult* **110**, 423–433 (2012).
256. Iwase, A. *et al.* Arabidopsis WIND1 induces callus formation in rapeseed, tomato, and tobacco. *Plant Signaling & Behavior* **8**, e27432 (2013).
257. Iwase, A. *et al.* WIND1-based acquisition of regeneration competency in Arabidopsis and rapeseed. *J Plant Res* **128**, 389–397 (2015).
258. Marhava, P. *et al.* Re-activation of Stem Cell Pathways for Pattern Restoration in Plant Wound Healing. *Cell* **177**, 957-969.e13 (2019).
259. Heyman, J. *et al.* ERF115 Controls Root Quiescent Center Cell Division and Stem Cell Replenishment. *Science* **342**, 860–863 (2013).
260. Iwase, A. *et al.* WIND1 Promotes Shoot Regeneration through Transcriptional Activation of *ENHANCER OF SHOOT REGENERATION1* in Arabidopsis. *Plant Cell* **29**, 54–69 (2017).
261. Iwase, A. *et al.* WIND transcription factors orchestrate wound-induced callus formation, vascular reconnection and defense response in Arabidopsis. *New Phytologist* **232**, 734–752 (2021).

262. Kondo, Y. *et al.* Vascular Cell Induction Culture System Using Arabidopsis Leaves (VISUAL) Reveals the Sequential Differentiation of Sieve Element-Like Cells. *The Plant Cell* **28**, 1250–1262 (2016).
263. Shimotohno, A., Heidstra, R., Blilou, I. & Scheres, B. Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. *Genes Dev.* **32**, 1085–1100 (2018).
264. Hu, X. & Xu, L. Transcription Factors WOX11/12 Directly Activate WOX5/7 to Promote Root Primordia Initiation and Organogenesis. *Plant Physiol.* **172**, 2363–2373 (2016).
265. Liu, J. *et al.* WOX1 and 12 Are Involved in the First-Step Cell Fate Transition during de Novo Root Organogenesis in Arabidopsis. *Plant Cell* **26**, 1081–1093 (2014).
266. Zhai, N. & Xu, L. Pluripotency acquisition in the middle cell layer of callus is required for organ regeneration. *Nat. Plants* **7**, 1453–1460 (2021).
267. Pernisova, M. *et al.* Cytokinin signalling regulates organ identity via AHK4 receptor in *Arabidopsis*. *Development* dev.163907 (2018) doi:10.1242/dev.163907.
268. Lee, K., Kim, J. H., Park, O.-S., Jung, Y. J. & Seo, P. J. Ectopic expression of WOX5 promotes cytokinin signaling and de novo shoot regeneration. *Plant Cell Rep* **41**, 2415–2422 (2022).
269. Deveaux, Y. *et al.* Genes of the most conserved WOX clade in plants affect root and flower development in Arabidopsis. *BMC Evol Biol* **8**, 291 (2008).
270. Tanaka, H. *et al.* Auxin-Induced WUSCHEL-RELATED HOMEODOMAIN13 Mediates Asymmetric Activity of Callus Formation upon Cutting. *Plant and Cell Physiology* **64**, 305–316 (2023).
271. Dreher, K. A., Brown, J., Saw, R. E. & Callis, J. The *Arabidopsis* Aux/IAA Protein Family Has Diversified in Degradation and Auxin Responsiveness. *The Plant Cell* **18**, 699–714 (2006).
272. Abel, S., Nguyen, M. D. & Theologis, A. The PS-IAA4/5-like Family of Early Auxin-inducible mRNAs in *Arabidopsis thaliana*. *Journal of Molecular Biology* **251**, 533–549 (1995).

273. Hussain, S. *et al.* The Auxin Signaling Repressor IAA8 Promotes Seed Germination Through Down-Regulation of ABI3 Transcription in Arabidopsis. *Front. Plant Sci.* **11**, 111 (2020).
274. Groover, A. T., Pattishall, A. & Jones, A. M. IAA8 expression during vascular cell differentiation.
275. Wang, J., Yan, D.-W., Yuan, T.-T., Gao, X. & Lu, Y.-T. A gain-of-function mutation in IAA8 alters Arabidopsis floral organ development by change of jasmonic acid level. *Plant Mol Biol* **82**, 71–83 (2013).
276. Zhang, S. *et al.* The calcium signaling module CaM–IQM destabilizes IAA–ARF interaction to regulate callus and lateral root formation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202669119 (2022).
277. Yoo, S.-D., Cho, Y.-H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* **2**, 1565–1572 (2007).
278. Oliveira, P. H., Prather, K. J., Prazeres, D. M. F. & Monteiro, G. A. Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. *Appl Microbiol Biotechnol* **87**, 2157–2167 (2010).
279. Tinland, B., Koukolíková-Nicola, Z., Hall, M. N. & Hohn, B. The T-DNA-linked VirD2 protein contains two distinct functional nuclear localisation signals. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7442–7446 (1992).
280. Chang, C.-W., Couñago, R. M., Williams, S. J., Boden, M. & Kobe, B. The distribution of different classes of nuclear localisation signals (NLSs) in diverse organisms and the utilization of the minor NLS-binding site in plant nuclear import factor importin- α . *Plant Signaling & Behavior* **8**, e25976 (2013).
281. Jablonski, A. E. *et al.* Optically Modulatable Blue Fluorescent Proteins. *J. Am. Chem. Soc.* **135**, 16410–16417 (2013).
282. Jonsson, K., Hamant, O. & Bhalerao, R. P. Plant cell walls as mechanical signaling hubs for morphogenesis. *Current Biology* **32**, R334–R340 (2022).

283. Varapparambath, V. *et al.* Mechanical conflict caused by a cell-wall-loosening enzyme activates de novo shoot regeneration. *Developmental Cell* **57**, 2063–2080.e10 (2022).
284. Barba, L. A. Terminologies for Reproducible Research. Preprint at <http://arxiv.org/abs/1802.03311> (2018).
285. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
286. Damm, B. & Willmitzer, L. Regeneration of fertile plants from protoplasts of different *Arabidopsis thaliana* genotypes. *Molec. Gen. Genet.* **213**, 15–20 (1988).
287. Masson, J. & Paszkowski, J. The culture response of *Arabidopsis thaliana* protoplasts is determined by the growth conditions of donor plants. *The Plant Journal* **2**, 829–833 (1992).
288. Siemens, J., Torres, M., Morgner, M. & Sacristin, M. D. Plant regeneration from mesophyll-protoplasts of four different ecotypes and two marker lines from *Arabidopsis thaliana* using a unique protocol.
289. Dovzhenko, A., Dal Bosco, C., Meurer, J. & Koop, H. U. Efficient regeneration from cotyledon protoplasts in *Arabidopsis thaliana*. *Protoplasma* **222**, 107–111 (2003).
290. Jeong, Y. Y., Lee, H.-Y., Kim, S. W., Noh, Y.-S. & Seo, P. J. Optimization of protoplast regeneration in the model plant *Arabidopsis thaliana*. *Plant Methods* **17**, 21 (2021).
291. Lambolez, A. *et al.* Warm Temperature Promotes Shoot Regeneration in *Arabidopsis thaliana*. *Plant and Cell Physiology* **63**, 618–634 (2022).
292. Canher, B. *et al.* The regeneration factors ERF114 and ERF115 regulate auxin-mediated lateral root development in response to mechanical cues. *Molecular Plant* **15**, 1543–1557 (2022).
293. Wei, X., Ding, Y., Wang, Y., Li, F. & Ge, X. Early Low-Fluence Red Light or Darkness Modulates the Shoot Regeneration Capacity of Excised *Arabidopsis* Roots. *Plants* **9**, 1378 (2020).

294. Rikiishi, K., Matsuura, T., Ikeda, Y. & Maekawa, M. Light Inhibition of Shoot Regeneration Is Regulated by Endogenous Abscissic Acid Level in Calli Derived from Immature Barley Embryos. *PLoS ONE* **10**, e0145242 (2015).
295. Shim, J. S., Kubota, A. & Imaizumi, T. Circadian Clock and Photoperiodic Flowering in Arabidopsis: CONSTANS Is a Hub for Signal Integration. *Plant Physiol.* **173**, 5–15 (2017).
296. Michael, T. P. *et al.* Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. *PLoS Genet* **4**, e14 (2008).
297. Torii, K. *et al.* A guiding role of the Arabidopsis circadian clock in cell differentiation revealed by time-series single-cell RNA sequencing. *Cell Reports* **40**, 111059 (2022).
298. Lu, S. X. *et al.* The Jumonji C Domain-Containing Protein JMJ30 Regulates Period Length in the Arabidopsis Circadian Clock. *Plant Physiology* **155**, 906–915 (2011).
299. Lee, K., Park, O.-S. & Seo, P. J. JMJ30-mediated demethylation of H3K9me3 drives tissue identity changes to promote callus formation in Arabidopsis. *The Plant Journal* **95**, 961–975 (2018).
300. Gan, E.-S. *et al.* Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. *Nat Commun* **5**, 5098 (2014).
301. Kodama, Y. Time Gating of Chloroplast Autofluorescence Allows Clearer Fluorescence Imaging In Planta. *PLoS ONE* **11**, e0152484 (2016).
302. Boyd, V., Cholewa, O. M. & Papas, K. K. Limitations in the Use of Fluorescein Diacetate/Propidium Iodide (FDA/PI) and Cell Permeable Nucleic Acid Stains for Viability Measurements of Isolated Islets of Langerhans. (2010).
303. Pietrzykowska, M. *et al.* The Light-Harvesting Chlorophyll a/b Binding Proteins Lhcb1 and Lhcb2 Play Complementary Roles during State Transitions in Arabidopsis. *The Plant Cell* **26**, 3646–3660 (2014).

304. Sawchuk, M. G., Donner, T. J., Head, P. & Scarpella, E. Unique and Overlapping Expression Patterns among Members of Photosynthesis-Associated Nuclear Gene Families in Arabidopsis. *Plant Physiology* **148**, 1908–1924 (2008).
305. Lintala, M. *et al.* Comparative analysis of leaf-type ferredoxin-NADP⁺ oxidoreductase isoforms in Arabidopsis thaliana. *The Plant Journal* **57**, 1103–1115 (2009).
306. DiMario, R. J. *et al.* The Cytoplasmic Carbonic Anhydrases β CA2 and β CA4 Are Required for Optimal Plant Growth at Low CO₂. *Plant Physiol.* **171**, 280–293 (2016).
307. Li, H. CUE1: A Mesophyll Cell-Specific Positive Regulator of Light-Controlled Gene Expression in Arabidopsis. *The Plant Cell* (1995).
308. Lechón, T., Sanz, L., Sánchez-Vicente, I. & Lorenzo, O. Nitric Oxide Overproduction by cue1 Mutants Differs on Developmental Stages and Growth Conditions. *Plants* **9**, 1484 (2020).
309. Khan, M., Ali, S., Al Azzawi, T. N. I. & Yun, B.-W. Nitric Oxide Acts as a Key Signaling Molecule in Plant Development under Stressful Conditions. *IJMS* **24**, 4782 (2023).
310. Paul, A.-L., Semer, C., Kucharek, T. & Ferl, R. J. The fungicidal and phytotoxic properties of benomyl and PPM in supplemented agar media supporting transgenic arabidopsis plants for a Space Shuttle flight experiment. *Applied Microbiology and Biotechnology* **55**, 480–485 (2001).
311. Tucker, D. J. Effects of Far-red Light on the Hormonal Control of Side Shoot Growth in the Tomato. *Annals of Botany* **40**, 1033–1042 (1976).
312. Xia, J., Zhao, H., Liu, W., Li, L. & He, Y. Role of cytokinin and salicylic acid in plant growth at low temperatures. *Plant Growth Regul* **57**, 211–221 (2009).
313. Yoshida, S., Mandel, T. & Kuhlemeier, C. Stem cell activation by light guides plant organogenesis. *Genes Dev.* **25**, 1439–1450 (2011).
314. Halliday, K. J., Martinez-Garcia, J. F. & Josse, E.-M. Integration of Light and Auxin Signaling. *Cold Spring Harbor Perspectives in Biology* **1**, a001586–a001586 (2009).

315. Bianchimano, L., De Luca, M. B., Borniego, M. B., Iglesias, M. J. & Casal, J. J. Temperature regulation of auxin-related gene expression and its implications for plant growth. *Journal of Experimental Botany* **74**, 7015–7033 (2023).
316. Youn, H.-D. Methylation and demethylation of DNA and histones in chromatin: the most complicated epigenetic marker. *Exp Mol Med* **49**, e321–e321 (2017).
317. Papadakis, A. K., Siminis, C. I. & Roubelakis-Angelakis, K. A. Reduced Activity of Antioxidant Machinery Is Correlated with Suppression of Totipotency in Plant Protoplasts. *Plant Physiology* **126**, 434–444 (2001).
318. Petricka, J. J., Winter, C. M. & Benfey, P. N. Control of *Arabidopsis* Root Development. *Annu. Rev. Plant Biol.* **63**, 563–590 (2012).
319. Ochoa-Fernandez, R. *et al.* Optogenetic control of gene expression in plants in the presence of ambient white light. *Nat Methods* **17**, 717–725 (2020).
320. Kareem, A. *et al.* PLETHORA Genes Control Regeneration by a Two-Step Mechanism. *Current Biology* **25**, 1017–1030 (2015).
321. Machin, F. Q. *et al.* Inducible reporter/driver lines for the Arabidopsis root with intrinsic reporting of activity state. *The Plant Journal* **98**, 153–164 (2019).
322. Saijo, T. & Nagasawa, A. Development of a tightly regulated and highly responsive copper-inducible gene expression system and its application to control of flowering time. *Plant Cell Rep* **33**, 47–59 (2014).
323. Roberts, G. R. *et al.* The *alc-GR* System. A Modified *alc* Gene Switch Designed for Use in Plant Tissue Culture. *Plant Physiology* **138**, 1259–1267 (2005).
324. Hickey, S. F. *et al.* Transgene regulation in plants by alternative splicing of a suicide exon. *Nucleic Acids Research* **40**, 4701–4710 (2012).

325. Gonzalez, T. L. *et al.* Tight regulation of plant immune responses by combining promoter and suicide exon elements. *Nucleic Acids Res* **43**, 7152–7161 (2015).
326. Wang, B., Barahona, M. & Buck, M. Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks. *Nucleic Acids Res* **42**, 9484–9492 (2014).
327. Papadaki, S. *et al.* Dual-expression system for blue fluorescent protein optimization. *Sci Rep* **12**, 10190 (2022).
328. Boo, A. *et al.* Synthetic microbe-to-plant communication channels. *Nat Commun* **15**, 1817 (2024).
329. Moreno-Giménez, E., Selma, S., Calvache, C. & Orzáez, D. GB_SynP: A Modular dCas9-Regulated Synthetic Promoter Collection for Fine-Tuned Recombinant Gene Expression in Plants. *ACS Synth. Biol.* **11**, 3037–3048 (2022).
330. Engler, C. *et al.* A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
331. Pérez-González, A. & Caro, E. Effect of transcription terminator usage on the establishment of transgene transcriptional gene silencing. *BMC Res Notes* **11**, 511 (2018).
332. De Felippes, F. F. & Waterhouse, P. M. The Whys and Wherefores of Transitivity in Plants. *Front. Plant Sci.* **11**, 579376 (2020).
333. Nagy-Staron, A. *et al.* Local genetic context shapes the function of a gene regulatory network. *eLife* **10**, e65993 (2021).
334. Wu, X.-L. *et al.* Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **10**, 189 (2017).
335. Gui, Q. *et al.* Transcriptome Analysis in Yeast Reveals the Externality of Position Effects. *Molecular Biology and Evolution* **38**, 3294–3307 (2021).
336. Kong, S., Yu, W., Gao, N., Zhai, X. & Zhou, Y. J. Expanding the neutral sites for integrated gene expression in *Saccharomyces cerevisiae*. *FEMS Microbiology Letters* **369**, fnac081 (2022).

337. Schubert, D. *et al.* Silencing in Arabidopsis T-DNA Transformants: The Predominant Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. *Plant Cell* **16**, 2561–2572 (2004).
338. Waddington, C. H. *The Strategy of the Genes*. (Routledge Library Editions, 1957).
339. Khakhar, A., Leydon, A. R., Lemmex, A. C., Klavins, E. & Nemhauser, J. L. Synthetic hormone-responsive transcription factors can monitor and re-program plant development. *eLife* **7**, e34702 (2018).
340. Alamos, S. & Shih, P. M. How to engineer the unknown: Advancing a quantitative and predictive understanding of plant and soil biology to address climate change. *PLoS Biol* **21**, e3002190 (2023).
341. Andres, J., Blomeier, T. & Zurbriggen, M. D. Synthetic Switches and Regulatory Circuits in Plants. *Plant Physiol.* **179**, 862–884 (2019).
342. Zhang, H. *et al.* DNA nanostructures coordinate gene silencing in mature plants. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7543–7548 (2019).
343. Borsuk, A. M., Roddy, A. B., Thérout-Rancourt, G. & Brodersen, C. R. Structural organization of the spongy mesophyll. *New Phytologist* **234**, 946–960 (2022).
344. Lehmeier, C. *et al.* Cell density and airspace patterning in the leaf can be manipulated to increase leaf photosynthetic capacity. *The Plant Journal* **92**, 981–994 (2017).
345. Liu, Y., Sun, L., Huo, Y.-X. & Guo, S. Strategies for improving the production of bio-based vanillin. *Microb Cell Fact* **22**, 147 (2023).
346. Chung, C. & Miller, R. Preparation and storage of competent Escherichia coli cells. *Methods in Enzymology* **218**, 621–627 (1993).
347. Weigel, D. & Glazebrook, J. Transformation of agrobacterium using electroporation. *Cold Spring Harbor Protocols* (2006) doi:10.1101/pdb.prot4665.
348. Weigel, D. & Glazebrook, J. Transformation of agrobacterium using the freeze-thaw method. *Cold Spring Harbor Protocols* (2006) doi:10.1101/pdb.prot4666.

349. Edwards, K., Johnstone, C. & Thompson, C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucl Acids Res* **19**, 1349–1349 (1991).
350. Andreou, A. & Nakayama, N. Mobius Assembly. in *DNA Cloning and Assembly: Methods and Protocols* vol. 1116.
351. Schmittgen, T. D. & Livak, K. J. Analysing real-time PCR data by the comparative CT method. *Nat Protoc* **3**, 1101–1108 (2008).
352. Urquiza-Garcia, U. & Millar, A. Expanding the bioluminescent reporter toolkit for plant science with NanoLUC. *Plant Methods* **15**, (2019).
353. Mauri, M., Vecchione, S. & Fritz, G. Deconvolution of Luminescence Cross-Talk in High-Throughput Gene Expression Profiling. *ACS Synth. Biol.* **8**, 1361–1370 (2019).
354. Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* **1**, 641–646 (2006).
355. Andrianantoandro, E., Basu, S., Karig, D. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2**, (2006).
356. Liu, W., Yuan, J. & Stewart, N. Advanced genetic tools for plant biotechnology. *Nature reviews genetics* **14**, 781–793 (2013).
357. Ji, G. *et al.* PASPA: a web server for mRNA poly(A) site predictions in plants and algae. *Bioinformatics* **31**, 1671–1673 (2015).
358. Shahmuradov, I. A., Umarov, R. Kh. & Solovyev, V. V. TSSPlant: a new tool for prediction of plant Pol II promoters. *Nucleic Acids Res* gkw1353 (2017) doi:10.1093/nar/gkw1353.
359. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
360. Quang, D. & Xie, X. EXTREME: An online em algorithm for motif discovery. *Bioinformatics* **30**,