Production of Novel Medicinal Cannabis Using Genome Editing Technology

Submitted by
Lennon Matchett-Oates
Bachelor of Science with Honours (Biotechnology)

A thesis submitted in total fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Applied Systems Biology

LaTrobe University
Victoria, Australia

January 2022

TABLE OF CONTENTS

TABLE OF CONTENTS ii

LIST OF TABLES vi

LIST OF FIGURES vii

SUMMARY viii

STATEMENT OF AUTHORSHIP ix

SCHOLARSHIP ACKNOWLEDGEMENT x

ACKNOWLEDGEMENTS xi

THESIS PREFACE xii

LIST OF ABBREVIATIONS xiii

CHAPTER 1 1

Introduction 1

1.1 Cannabis history 1

1.2 Botany and taxonomy 2

1.3 Horticulture 4

1.4 Medicinal cannabis industry 6

1.4.1 Applications of medicinal cannabis 7

1.5 Cannabinoid biosynthesis 8

1.5.1 Hexanoate pathway 12

1.5.2 MEP pathway 13

1.5.3 GPP pathway 15

1.5.3 Cannabinoid pathway 15

1.6 Cannabis genome, tools and resources 18

1.7 Double strand breaks in plants 20
1.7.1 Homologous recombination 20
1.7.2 Non-homologous end joining 21
1.8 Targeting DSBs with specific endonucleases 22
1.8.1 Zinc finger nucleases 23
1.8.2 TALENs 23
1.8.3 CRISPR/Cas9 24
1.9 Delivery methods for gene editing and gene suppression constructs 26
1.9.1 Electroporation-mediated transformation 26
1.9.2 PEG-mediated transformation 27
1.9.3 Agrobacterium-mediated transformation 27
1.9.4 Particle bombardment 28
1.10 RNA interference in gene suppression 29
1.10.1 RNAi mechanisms in plants 30
1.10.2 RNAi genome engineering 32
1.11 Development of cannabis *in vitro* culture 34
1.11.1 Micropropagation 34
1.11.2 Protoplast isolation 35
1.11.3 Callogenesis 37
1.11.4 Regeneration 38
1.11.5 Germplasm conservation 39
1.12 Research plan 41
1.13 Aims 42

CHAPTER 2 43

In silico analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation 43

2.1 Chapter preface 43

2.2 Publication details 43

2.3 Statement of contributions of joint authorship 43

2.4 Statement from co-author confirming the authorship contribution of the PhD candidate 43

2.5 Manuscript 44

CHAPTER 3 61

Development of a robust transient expression screening system in protoplasts of Cannabis 61

3.1 Chapter preface 61

3.2 Publication details 61

3.3 Statement of contributions of joint authorship 61

3.4 Statement from co-author confirming the authorship contribution of the PhD candidate 61

3.5 Manuscript 62

CHAPTER 4 74

Manipulation of cannabinoid biosynthesis via transient RNAi expression 74

4.1 Chapter preface 74

4.2 Publication details 74

4.3 Statement of contributions of joint authorship 74

4.4 Statement from co-author confirming the authorship contribution of the PhD candidate 74

4.5 Manuscript 75

CHAPTER 5 87

Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification........ 87

5.1 Chapter preface........ 87

5.2 Publication details........ 87

5.3 Statement of contributions of joint authorship........ 87

5.4 Statement from co-author confirming the authorship contribution of the PhD candidate........ 88

5.5 Manuscript........ 89

CHAPTER 6 109

6.1 General discussion........109

6.2 Background and context of the research........ 109

6.3 Overview of the research........111

6.4 Future directions........116

APPENDICIES 117

Appendix 1........117

Chapter 5 supplementary material........117

REFERENCES 121

LIST OF TABLES

Table 1.1 Taxonomy of cannabis including commonly referred species.. 2

Table 1.2 Representative chemical profiles of cannabis phenotypes... 3

Table 1.3 Issued license numbers granted under the Narcotics Drug Act 1967
in Australia... 6

Table 1.4 Cannbio-2 genome assembly compared to other widely used published
genome assemblies.. 19

LIST OF FIGURES

Figure 1.1 Morphological differences between subspecies of cannabis 3

Figure 1.2 Indoor and outdoor cultivation of cannabis 5

Figure 1.3 Chemical structure of major cannabinoids 8

Figure 1.4 Different types of trichomes 9

Figure 1.5 B locus genetic model determining cannabinoid composition in cannabis 10

Figure 1.6 Interconnected cannabinoid biosynthesis pathways 11

Figure 1.7 Catalysed reactions by *THCAS* and *CBDAS* to produce THCA and CBDA 17

Figure 1.8 DSB repair by HR And NHEJ 22

Figure 1.9 ZFN architecture to induce DSB 23

Figure 1.10 TALENs architecture to induce DSB 24

Figure 1.11 CRISPR architecture to induce DSB 25

Figure 1.12 Model for *Agrobacterium*-mediated transformation depicting steps involved through the process within the bacterium and plant host 28

Figure 1.13 Visual representation of gene delivery of DNA coated particles into plant cells 29

Figure 1.14 Different RNAi pathways in plants that regulate expression of genes 31

Figure 1.15 Micropropagation of cannabis through the different stages from donor plant to introduction into tissue culture to acclimatization into the greenhouse 35

Figure 1.16 Isolated cannabis protoplasts through enzymatic digestion.. 36

SUMMARY

Cannabis sativa L. (cannabis) use dates back as far as 6000 years for its medicinal properties, source of bast fibre and seed oil. However, due to its prohibited status for many years, scientific advancement has not been possible, leaving the germplasm in a rudimentary state. This thesis develops protocols and important resources for the genetic modification of cannabis using molecular genetics and tissue culture approaches. A comprehensive analysis of the genome structure of related cannabinoid biosynthesis genes from multiple enzymatic pathways, copy number variance and nucleotide polymorphisms allows for the development of a CRISPR/Cas9 single-guide RNA (sgRNA) and RNA interference (RNAi) catalogue with intelligent design to avoid possible nucleotide variants. This is achieved by analysing all related genes within a large pan-genome and identifying regions of structural variance to which online tools are used to design highly effective sgRNA and RNAi constructs for genome editing. Multiple tissue culture protocols are developed to enable transformation of cannabis. Protoplast isolation and transformation is explored using statistical methods to increase viable protoplast yields and transformation efficiency. Agroinfiltration protocol modification for in-house genetics is developed to transiently express developed RNAi vectors within leaf explants of cannabis, targeting for the first time the cannabinoid biosynthesis genes. Development of callogenesis and regeneration protocols for the development of stably transformed cannabis using the designed RNAi genetic resources is also explored. The resources and protocols developed will allow researchers to apply this knowledge to improve crop traits in cannabis as it has been developed on a large pan-genome, and in the application of tissue culture, the protocols verified on multiple cultivars.

STATEMENT OF AUTHORSHIP

I confirm that the work presented in this thesis has been primarily performed by myself unless explicitly identified within the text. This work is submitted in full for Doctor of Philosophy at LaTrobe University. These works have not been submitted elsewhere nor is this work submitted for an award or any other degree at another institution. Except where reference is made, this thesis contains no material published previously or written by anyone else.

Lennon Matchett-Oates

Date 10/1/2022

Scholarship Acknowledgement

This work was supported by Agriculture Victoria Research for Lennon Matchett-Oates. An additional complement to this scholarship was provided by Agriculture Victoria Research, Department of Jobs, Precincts and Regions.

This work was supported by an Australian Government Research Training Program Scholarship.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Noel Cogan for allowing me to explore my work and offering excellent advice, support and contacts to achieve my optimistic goals.
I want to thank my co-worker Ehab Mohamaden for hands on training and feedback throughout the developing project.
I would also like to thank Shivraj Braich for her bioinformatics help from beginning to end.
A huge thank you to the whole Molecular Genetics and Plant Functional Genomics teams who have helped me with support and advice. Anyone else assisted me throughout this journey, I thank you.
Finally, I want to thank my partner, Leonie, and dog, Chilli, for keeping me positive and feeling supported.

THESIS PREFACE

This thesis contains 6 chapters, with chapters 2 to 5 containing experimental research in the form of peer reviewed articles. Chapter 1 contains a literature review of the relevant research. The peer reviewed articles contain their own respective introduction, methodology, results and discussion in the respective journals formatting. Each peer reviewed article is preceded with a brief summary of the work performed with relevant publication details and contribution of co-authors. Each peer review article is curated in the relevant journal's citation and formatting requirements; this causes some redundancy between introductions and materials and methods. Chapter 6 provides a discussion integrating all the research topics published and future directions of this research. Chapters 1 and 6 will employ the same citation style and the bibliography will be provided at the end of this thesis. All supplementary material for peer reviewed articles can be found at the respective journal's website. All other relevant supplementary material will be provided as an Appendix.

LIST OF ABBREVIATIONS

2-AG	2-Arachidonoylglycerol
AAE	Acyl-Activating Enzymes
AEA	Anandamide
AGL5	Agamous-Like5
AGO	Argonaute Protein
ATP	Adenosine Triphosphate
BAP	6-Benzylaminopurine
CB1	Cannabinoid 1 Receptor
CB2	Cannabinoid 2 Receptor
CBD	Cannabidiol
CBDA	Cannabidiolic Acid
CBDAS	Cannabidiolic Acid Synthase
CBE	Cannabielsoin
CBG	Cannabigerol
CBGA	Cannabigerolic Acid
CBN	Cannabinol
cDNA	Complementary DNA
CDP-ME	4-Diphosphocytidyl-2c-Methyl-D-Erythritol
CDP-ME2P	4-Diphosphocytidyl-2c-Methyl-D-Erythritol-2-Phosphate
CDS	Coding Sequence
CMK	4-Diphosphocytidyl-2-C-Methyl-D-Erythritol Kinase
CO2	Carbon Dioxide
CRISPR/Cas9	Clustered Regularly Interspaced Short Palendromic Repeats
CrRNA	Crispr RNA
DABB	Dimeric A+B Barrel
DCL	Dicer-Like Protein
DMAPP	Dimethylallyl
DNA	Deoxyribonucleic Acid
DoH	Department Of Health
DSB	Double Strand Break
dsRNA	Double Stranded RNA
DXP	1-Deoxy-D-Xyulose 5-Phosphate
DXR	1-Deoxy-D-Xyulose 5-Phosphate Reductoisomerase
DXS	Deoxyxylulose 5-Phosphate Synthase
ER	Endoplasmic Reticulum
FACS	Fluorescence-Activated Cell Sorting
FAD2	Fatty Acid Desaturase
GC-FID	Gas Chromatography-Flame Ionization Detection
GFP	Green Fluorescent Protein
GOT	Geranylpyrophosphate:Olivetolate Geranyltransferase
GPP	Gernyl Diphosphate
GPP	Geranyl Pyrophosphate Synthase
GPP	Geranylpyrophosphate
GPR55	G Protein-Coupled Receptor 55

GUS	Beta-Glucuronidase
HDR	1-Hydroxy-2-Methyl-2-(E)-Butenyl 4-Diphosphate Reductase
HDS	4-Hydroxy-3-Methylbut-2-En-1-Yl Diphosphate Synthase
HMBPP	1-Hydroxy-2-Methyl-2-(E)-Butenyl 4-Diphosphate
HPL	Hydroperoxide Lyase
hpRNA	Hairpin RNA
HR	Homologous Recombination
IBA	Indole-3-Butyric Acid
IPP	Isopentenyl Diphosphate
IPP	Isopentyl Pyrophosphate
IPP	Isopentenyl Diphosphate Isomerase
ISSR	Inter Simple Sequence Repeat
LC-MS/MS	Liquid Chromatography - Tandem Mass Spectrometry
LOX	Lipoxygenase
MADC1	Male Associated DNA Sequence
MCT	4-Diphosphocytidyl-2c-Methyl-D-Erythritol Synthase
MDS	2c-Methyl-D-Erythritol 2,4-Cyclodiphosphate Synthase
MEP	Methylerythritol 4-Phosphate
MEP	2-C-Methyl-D-Erythritol 4-Phosphate
MES	2-Ethanesulfonic Acid
miRNA	MicroRNA
MS	Murashige And Skoog
mT	Meta-Topolin
NAA	1-Naphthaleneacetic Acid
NHEJ	Nonhomologous End Joining
OAC	OLA Cyclase
ODC	Office Of Drug Control
OLA	Olivetolic Acid
OLS	Olivetol Synthase
ORF	Open Reading Frame
PAM	Protospacer-Adjacent Motif
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PGR	Plant Growth Regulator
pha-siRNA	Phased-SiRNA
PT	Prenyltrasnferase
PTGS	Post Transcriptional Gene Silencing
RAPD	Random Amplified Polymorphic DNA
RdDM	RNA-Directed DNA Methylation
RISC	RNA-Induced Silencing Complex
RNAi	RNA Interference
RVD	Repeat Variable Di-Residue
sgRNA	Single-Guide RNA
siRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
sRNA	Small RNA

SVM	Support Vector Machine
TALENs	Tal Effector Nucleases
ta-siRNA	Trans-Acting-SiRNA
T-DNA	Transfer DNA
TDZ	Thidiazuron
THC	Tetrahydrocannabinol
THCA	Tetrahydrocannabinolic Acid
THCAS	Tetrahydrocannabinolic Acid Synthase
THCV	Tetrahydrocannabivarin
ThDP	Thiamine Diphosphate
VIGS	Virus Induced Gene Silencing
ZFN	Zinc Finger Nuclease

CHAPTER 1

Introduction

1.1 Cannabis history

Cannabis sativa is suspected to be one of the earliest plants cultivated in China, where evidence of its first use dates back to 4000 B.C (Zuardi, 2006). Traditionally, cannabis has been domesticated for the fibre out of the stems, which could be used to manufacture paper, ropes and textiles (Ren et al., 2021). Throughout history, cannabis domestication varied geographically to increase the yield and quality of these products. Due to varying agricultural techniques, vast phenotypic variation exists within the species today (Clarke, 1981).

Cannabis has been used for its medicinal benefits in many cultures around the world for hundreds of years (Grotenhermen & Müller-Vahl, 2016). Different preparations of cannabis have historically been used for the treatment of pain, inflammation, as an analgesic, sedative and to improve appetite (de Médicis Sajous, 1918; Mary Lynn Mathre, 2012). The first well documented record of cannabis for medicinal purposes was by Theodor Friedrich Ludwig Ness von Esenbeck in the 19th century (von Esenbeck, 1831), with the introduction of cannabis into western culture by Sir William Brooke O'Shaughnessy, publishing clinical data of Indian hemp in 1839 (Gorman, 1969). It was this study of Indian hemp, which contained considerably more active compounds compared to traditional hemp varieties, that sparked the western world's curiosity for the use of cannabis for its therapeutic benefits (McGeeney, 2013).

Since its introduction into the western world, cannabis has faced strict and detrimental reforms for the cultivation and use as a medicinal therapeutic agent. From as early as the 14th century (Johnson, 2010) up until present day, prohibition of cannabis has been implemented to varying degrees. The motive behind prohibition in the western world has historically been political and the stance certain political parties took on minority populations and communities of lower socioeconomic status (Robinson & Scherlen, 2014). Today, cannabis is still scheduled as a narcotic drug in many countries, however that is slowly changing with additions such as Canada, Spain, and Uruguay decimalising cannabis. The total consumers of cannabis worldwide is estimated at 200 million users (United Nations Office on Drugs and Crime, 2021) and that number is predicted to rise with increased legalisation across the world.

1.2 Botany and taxonomy

Cannabis is an annual, wind pollinated herb and the only genus of the Cannabaceae family (Table 1.1) (Kriese et al., 2004). The number of species in the genus Cannabaceae is currently debateable with reports suggesting a polytypic genus (Emboden, 1974; Hillig, 2005) or as a monotypic, highly polymorphic species (Doorenbos et al., 1971). Recently, the classification of cannabis in regards to its cannabinoid and terpenoid profiling has been discussed (Piomelli & Russo, 2016). However, three subspecies of cannabis are generally accepted: *Cannabis sativa*, *Cannabis indica* and *Cannabis ruderalis* (Piomelli & Russo, 2016; Small & Cronquist, 1976). The taxonomic classification of *Cannabis sativa* follows in Table 1.1.

Table 1.1: Taxonomy of cannabis including commonly referred species.

Kingdom:	**Plantae**
Subkingdom:	*Tracheobionta*
Superdivision:	*Spermatophyta*
Division:	*Magnoliophyta*
Class:	*Magnoliopsida*
Subclass:	*Hamamelididae*
Order:	*Urticales*
Family:	*Cannabaceae*
Genus:	*Cannabis*
(Species):	(*sativa, indica, ruderalis*)

Cannabis is generally dioecious (sex specific flowers residing on separate plants) with monoecious (both flowers are present on the same plant) forms also existing (Menzel, 1964). Male plants generally are taller and narrower compared to the female plant, however it is still difficult to distinguish between sexual morphology until flowering. Molecular techniques have been developed to differentiate at an early stage with varying results (Mandolino et al., 1999).

Morphological differences exist between subspecies: *C. sativa* is tall, narrow leaved with loose branches, *C. indica* is shorter, has wider darker green leaves, often with a purple tinge and *C. ruderalis* is the smallest of the species producing very few branches or leaves (Sawler et al., 2015; Thomas & ElSohly, 2015) (Figure 1.1).

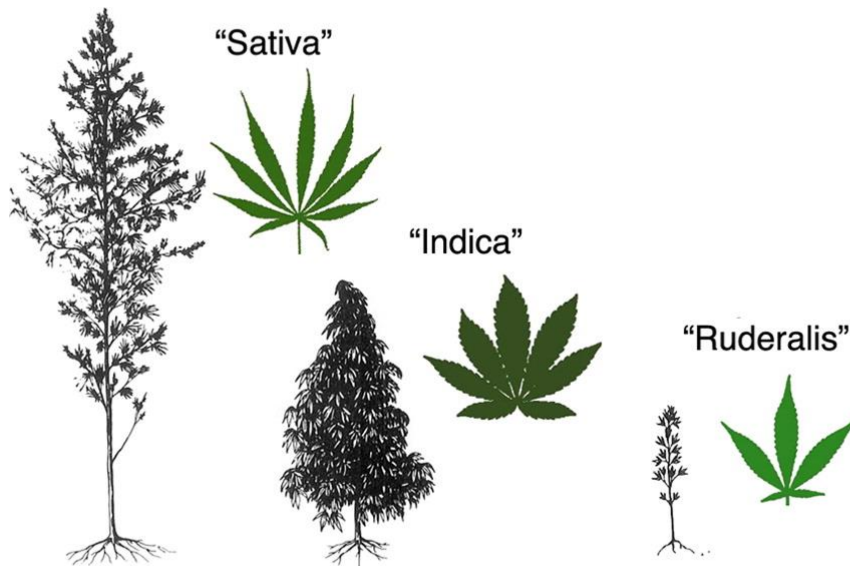


Figure 1.1: Morphological differences between subspecies of Cannabis. Image taken from McPartland (2018).

The distribution ratios of Tetrahydrocannabinol (THC): Cannabidiol (CBD) are under genetic control, with hemp containing a majority of non-psychoactive compounds compared to drug type phenotypes containing significantly higher levels of psychoactive THC (De Backer et al., 2009; Galal et al., 2009; Grassa et al., 2021; Van Bakel et al., 2011). The three phenotypes of cannabis are outlined in Table 1.2:

Table 1.2: Representative chemical profiles of Cannabis phenotypes

Drug type:	THC > 0.5% and CBD < 0.5%
Intermediate type:	CBD major cannabinoid with varying levels of THC
Hemp or fibre type:	Low THC levels (<0.3%)

1.3 Horticulture

Cannabis, being an annual plant, can be effectively grown indoors and outdoors (Figure 1.2). It is a short-day plant, flowering towards the end of summer and setting seeds before winter. When day length decreases to a critical point (12 hours), flowering occurs. This change is due to a response of the phytochrome proteins during darkness (Potter, 2014). The critical day length varies between strains and subspecies, which in turn will contribute to the cultivation option. However, both cultivations come with their own sets of advantages and limitations.

The lifecycle of the plant grown outdoors is typically complete in 5-7 months depending on environmental conditions e.g. time of planting, rain etc. Alternatively, growing cannabis indoors has the advantages of being able to control photoperiod, water availability, temperature and plant spacing, reducing the cultivation period to approximately 3 months.

Indoor cultivation allows for total control of the aforementioned variables during the life cycle of the plant. Vigorous growth of cannabis is achieved with optimum photo periods, CO_2 concentrations and temperatures, which are all controllable to achieve maximum biomass and avoid pests and disease (Cure, 1985; Idso & Idso, 1994; Sage & Sharkey, 1987; Strain & Cure, 1985). With the increasing interest in cannabis as a plant with medicinal and biotechnological applications, businesses are opting to grow cannabis indoors, building highly regulated facilities capable of producing large quantities of medical grade cannabis. Current projections on the annual revenue of manufacturing cannabis is approx. $4.8 million per acre (Higgins et al., 2016).

Comparatively, outdoor cultivation of cannabis begins at the turn of seasons, from spring to summer and can last until mid-winter. Beginning from seeds, healthy seedlings can be selected and moved to the field. The allogamous nature of cannabis can make it troublesome to maintain chemical profiles of genotypes when grown outdoors from seeds. Additionally, environmental factors need to be considered when outdoor cultivation is chosen. Wind, rain and fungal contamination, e.g. *Fusarium.* can all be detrimental to crop development and can contain harmful neurotoxins whilst also considerably affect biomass at harvesting time (Punja, 2021).

Figure 1.2: Indoor and outdoor cultivation of Cannabis. Images taken from Systemsnspace (2018) and Medium (2019).

The issue of genotypic variation can be overcome with the use of vegetative cuttings from a donor mother plant, creating genetically uniform crops with a consistent chemical profile. Cannabis is heterozygous in nature and because of this valuable traits may be lost with seed generation, thus the use of cuttings can offer control of genotypes or experiments designed to quantify effects of the environment (Clarke, 1981).

1.4 Medicinal cannabis industry

Medicinal cannabis is either dried floral buds of cannabis, or a derived extraction, prescribed to relieve symptoms of medical conditions e.g. epilepsy or cancer (Luckett et al., 2016; Stockings et al., 2018). Since 1961, laws set out by the United Nations defined cannabis as 'any plant of the genus Cannabis' with accompanying international treaties, such as in Europe, that the entire plant be controlled under national drug laws (Hughes, 2017).

The legal status of cannabis has changed dramatically for many countries, beginning in 1992 with Israel and subsequently Argentina and South Africa amongst others. The current state of medicinal cannabis in Australia is governed by the federal government with legislation to allow for cultivation of cannabis for medical or scientific purposes through a national licensing scheme, in an amendment to the Narcotics Drugs Act 1967, now referred to as the Narcotic Drugs Amendment (Medicinal Cannabis) Bill 2021. The manufacturing or research of medicinal cannabis is a responsibility for the Commonwealth, states and territories, with supply controlled under the Therapeutic Goods Act 1989. Currently, Victoria, Western Australia, South Australia and New South Wales states provide patients with access to medicinal cannabis.

The current number of licences granted under the Narcotics Drugs Act 1967 are outlined in Table 1.3 (Office of Drug Control, 2021):

Table 1.3: Issued license numbers granted under the Narcotic Drugs Amendment (Medicinal Cannabis) Bill 2021

Licence type	Number of licenses granted
Medicinal Cannabis License (cultivation and production)	42
Cannabis Research Licence (cultivation and production)	17

Countries such as The United States currently have 37 states with access to medicinal cannabis (ProCon, 2021), all of which are generating jobs, decreasing government spending on prohibition and helping improve social programs including the education system. In states such as Colorado, as of 2015, eighteen thousand jobs were created with $2.4 billion generated in state output (Kleiman et al., 2016).

The future of cannabis in Australia is projected to generate over $3.5 billion from legalisation (Di Natale, 2018). Employment and wealth creation opportunities from access to medical cannabis, as shown from international examples, is a large incentive for state and territories to work towards safe and efficient access to medical cannabis.

1.4.1 Applications of medicinal cannabis

Cannabinoid receptors are G-protein coupled receptors inhibiting adenylate cyclase whilst activating mitogen-activated protein kinase (Howlett, 2005). Two main cannabinoid receptors, CB1 and CB2, are located in the terminals of central and peripheral neurons mediating the release of neurotransmitters (Pertwee, 2005). CB1 receptors have also been found in immune cells and the tissues of the heart. Similarly, CB2 receptors have been found in immune cells (Cascio et al., 2017). CB1 receptors are the major cannabinoid receptor responsible for the psychoactive effects of THC, whereas CB2 receptors are responsible for the immunosuppressive effects of cannabinoids. CBD has shown to have a higher affinity for CB2 receptors reducing the psychotropic effects. A possible third type of cannabinoid receptor has been reported, GPR55, which shows high homology with current known receptors, however it is currently unknown how the mechanism mediates THC effects (Pertwee, 2007).

The endocannabinoid system is comprised of endogenous cannabinoid (endocannabinoids) receptors and metabolic enzymes that play a crucial role in homeostasis, regulating temperature, neurotransmission and inflammation. To date, seven endocannabinoids have been identified: Anandamide (AEA), 2-arachidonoylglycerol (2-AG), 2-AG ether, O-arachidonoyl ethanolamine, N-arachidonoyl dopamine, docosahexaenoyl ethanolamide and eicospentanoyl ethanolamide (Brown et al., 2010; Devane et al., 1992; Hanuš et al., 2001; Huang et al., 2002; Mechoulam et al., 1995; Porter et al., 2002). Endocannabinoid synthesis is stimulated by insulin, cytokines and neuronal activity (e.g. pain) and secreted throughout the body, including the brain and immune cells (Gui et al., 2015).

Medicinal cannabis is gaining interest as a therapeutic agent for the treatment of conditions including pain, multiple sclerosis, epilepsy and cancer. Currently, there is limited literature on the effects produced by cannabinoids in clinical trials (Haleem & Wright, 2020). Historically, cannabis has been administered orally as oils, inhaled or infused into foods and beverages. As a result of the increased knowledge and application of medicinal cannabis many of the cannabinoids are now isolated and purified for sublingual, oral or administration through inhalation. Despite the increases in medicinal applications, there are currently limited available legal medicinal cannabis preparations with Sativex® being the most well-known (for an extensive review, refer to Lim, See and Lee, 2017 (Lim et al., 2017)). Sativex® products have been used in treatment of multiple sclerosis with 2.7mg THC and 2.5mg CBD administered orally (Novotna et al., 2011). The use of CBD in clinical trials for the treatment of epilepsy has been examined since the late 70's (Mechoulam & Carlini, 1978). A recent clinical trial using CBD for epilepsy was conducted by Devinsky et al. (2016) using 214 patients with severe treatment-resistant epilepsy (Devinsky et al., 2016). Epidiolex®, a CBD extract, was

administered orally at 2-5mg/kg/day initially until a maximum dose of 50mg/kg/day was achieved. Thirty six percent of users reported a reduction in seizures, with only 3% of total users ceasing treatment.

1.5 Cannabinoid biosynthesis

Cannabinoids represent a diverse group of C_{21} terpenophenolic compounds predominantly present in *C. sativa* L. including their analogs and decarboxylated products (Andre et al., 2016; ElSohly & Slade, 2005). Plants from the *Radula* and *Helichrysum* genus have also been reported to contain cannabinoids (Appendino et al., 2008). Cannabinoids are stored in their acidic forms and through heat or long-term storage are decarboxylated into their neutral forms. To date, a total of over 120 cannabinoids have been reported from cannabis (ElSohly et al., 2017) with a further 450 other chemical constituents being identified (Leghissa et al., 2018) (a comprehensive list of cannabinoids can be found at Radwan et al.,2017). The term 'phytocannabinoids' is used now to refer to the natural production of cannabinoids due to the discovery of synthetic cannabinoids and chemically different endogenous cannabinoid receptor ligands.

The major cannabinoids studied for their pharmacological and therapeutic properties are tetrahydrocannabinol (THC), tetrahydrocannabivarin (THCV), cannabidiol (CBD), cannabigerol (CBG) and cannabinol (CBN) (Figure 1.3). CBN was the first phytocannabinoid to be identified from an oil extract of cannabis in the 19th century, with its structure and chemical synthesis achieved in the early 20th century (Pertwee, 2006). Shortly following the chemical synthesis of CBN, CBD and THC were also achieved (Adams et al., 1940; Wollner et al., 1942).

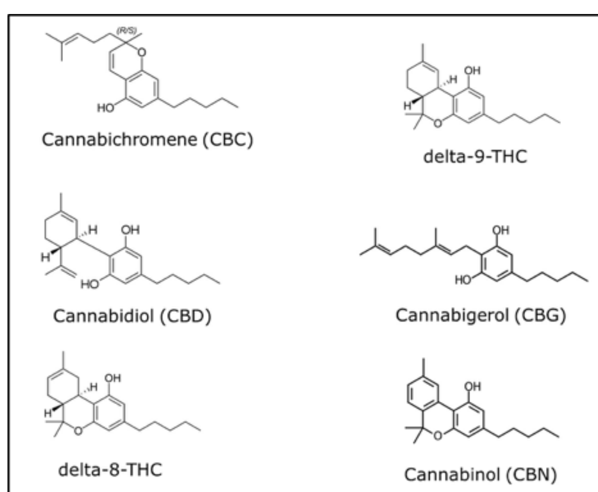


Figure 1.3: Chemical structure of major cannabinoids. Image taken from She (2016).

Phytocannabinoids accumulate in the glandular trichomes on the floral buds of the female plant as well as leaves, seeds and roots, however in considerably lower concentrations (Mahlberg & Kim, 1992; Ross et al., 2005). Three types of glands cover the surface of both pistilate and staminate plants: bulbous, capitate-sessile and capitate-stalked (Mahlberg & Kim, 2004). A gland consists of a spherical head attached to the stalk on flowers, leaves and bracts (Figure 1.4). Mahlberg and Kim (1992) demonstrated that cannabinoids were up to 20x more concentrated in sessile-stalked glands when compared to leaf vein, reporting their specialised structures for high levels of cannabinoid synthesis.

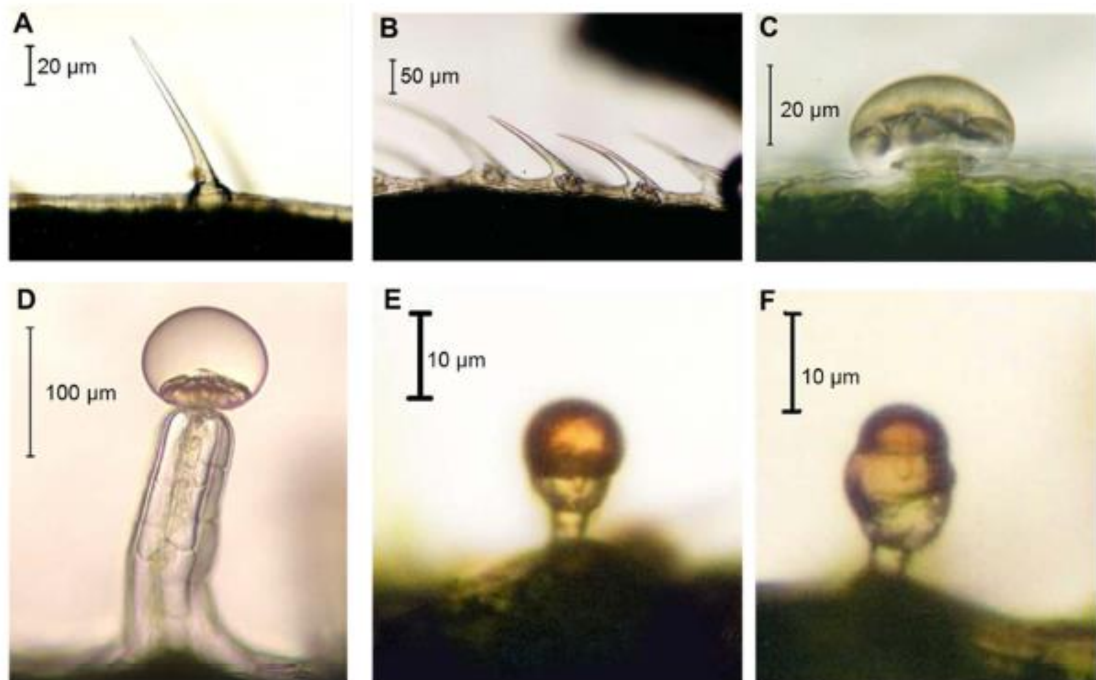


Figure 1.4: Different types of trichomes (A) Non-glandular; (B) cystolythic; (C) capitate sessile; (D) capitate-stalked; (E) bulbous; (F) complex. Images taken from Andre et al. (2016).

Cannabinoid quantity is controlled by several genes with external influences from the environment contributing to total yield. Particular genes involved in cannabinoid production influence the plants chemotype, trichome density, resin head size, as well as gender as female plants produce more cannabinoids (Grassi & McPartland, 2017). Due to large variety of cannabis strains containing varying levels of chemical variants, cannabinoid quality is referred to as chemotypes. Initially, chemotypes of cannabis were classed as “drug-types” and “fiber-types” (Fetterman et al., 1971) representing THC+CBN/CBD quotient >1 or <1 respectively. The following years saw discussions regarding the determining factors to a plants chemotypes (Fairbairn & Liebmann, 1974; Fournier, 1981; Hemphill et al., 1980; Small & Beckstead, 1973; Turner et al., 1979) with now three major chemical chemotypes and two minor chemotypes being accepted as the current model (Pacifico et al., 2006).

de Meijer et al. (2003) proposed the genetic determination for chemotypes as two alleles at a single gene locus, termed the B locus (de Meijer et al., 2003) (Figure 1.5).

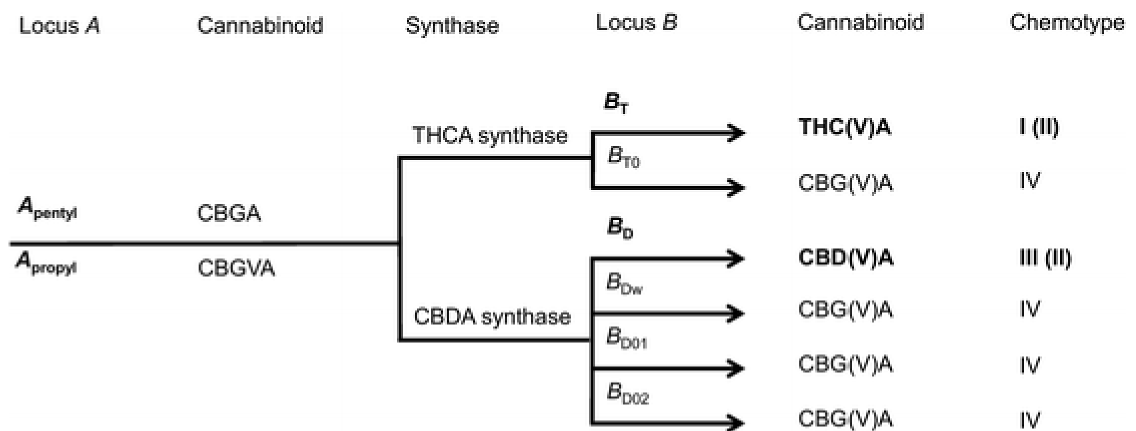


Figure 1.5: B locus genetic model determining cannabinoid composition in cannabis. Image taken from Welling et al. (2016).

Classical Mendelian behaviour for B_T and B_D alleles does occur (Figure 1.5), however co-dominance is inherent. The model from de Meijer et al. (2011) expresses *THCAS* and *CBDAS* as co-dominant inheritance, which has also been suggested in previous breeding experiments (Yotoriyama et al., 1980). There are discrepancies when it comes to monogenic inheritance and cannabis' variation not segregating into 100% THC, 50:50 or 100% CBD populations. Recent sequencing of the cannabis genome challenges this model further revealing multiple genes for *THCAS* and *CBDAS*, including multiple pseudo genes (Grassa et al., 2018; Van Bakel et al., 2011). Currently, the model to determine chemotype is derived from Grassa et al. (2018) that the cannabinoid biosynthesis genes are not located at a single locus but are pericentromeric, nested in repeats leading to low levels of recombination.

Biosynthesis of cannabinoids is a complex mechanism with numerous enzymatic steps and collaboration of multiple pathways. Fatty acids and isoprenoid precursors are synthesised via the hexanoate, methylerythritol 4-phosphate (MEP) and geranyl diphosphate (GPP) pathways (Figure 1.6).

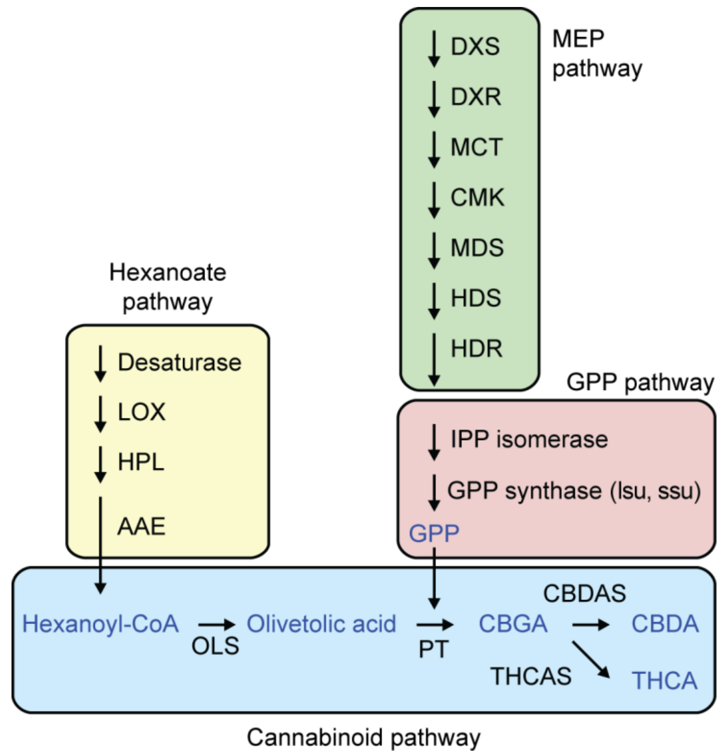


Figure 1.6: Interconnected cannabinoid biosynthesis pathways.
Image taken from van Bakel et al. (2011).

Via the Hexanoate pathway, Hexanoyl-CoA is produced, which acts as substrate for the polyketide synthase enzyme, *OLS*, yielding olivetolic acid (OLA) (Gagne et al., 2012; Galal et al., 2009). Via the MEP pathway, prenyl sidechains are synthesised for the substrate for geranyl diphosphate synthesis (Phillips et al*.*, 2008). GPP and OLA are added by an aromatic prenyltransferase (*PT*) creating Cannabigerolic acid (CBGA) (Taura et al., 2007). Finally, catalysation of THC and CBD oxidocyclases produce Tetrahydrocannabinolic acid (THCA) and Cannabidiolic acid (CBDA) (Fellermeier et al., 2001; Van Bakel et al., 2011). A more in-depth review of each enzymatic step from each pathway is provided below.

1.5.1 Hexanoate pathway

Desaturase

Desaturase, or more specifically fatty acid desaturase (*FAD2*), is the first identified step in the hexanoate pathway. Located in the endoplasmic reticulum (ER), *FAD2* introduces a double bond in oleate creating linoleate (McConn et al., 1993). More specifically, *FAD2* introduces the double bond 6 carbon atoms away from the methyl end of the fatty acid and like other desaturases, utilizes cytochrome b5 as the electron donor (Ohlrogge & Browse, 1995).

LOX

Lipoxygenase (*LOX*) is a nonheme iron containing enzyme catalysing deoxygenation of fatty acids, inserting oxygen at carbon position 9 or 13 in linoleate. *LOXs* are encoded by multiple gene families, with as many as 13 LOX genes recently discovered in rice (Ogunola et al., 2017).

HPL

Hydroperoxide lyase (*HPL*) is a cytochrome P450 enzyme cleaving C-C bonds in hydroperoxides of fatty acids, converting hydroperoxide intermediates into aldehydes and oxoacids (Grechkin et al., 2006). *HPL* is classified into two groups according to substrate specificity: *13-HPL* and *9-HPL* (Noordermeer et al., 2001), though the exact activation of *HPL* is yet to be determined.

AAE

Acyl-activating enzymes (*AAE*) are a superfamily forming Acyl-CoA thioesters, including hexanoyl-CoA, activating carboxylic acids through adenylate intermediates (Schmelz & Naismith, 2009). An in depth discussion of different *AAE*s is available from Shockey and Browse (2011) (Shockey & Browse, 2011). In cannabis, hexanoyl-CoA is a precursor for cannabinoid production requiring the localisation of this *AAE* to be in the glandular trichomes. Stout et al. (2012) analysed the transcriptome library of glandular trichomes, identifying 11 presumed *AAE* genes. LC-MS/MS analysis found 15.5pmol per gram of hexanoyl-CoA present in female flowers, with further *in vitro* assays showing *CsAAE1* with high transcript levels in glandular trichomes, suggesting that this *AAE* is responsible for synthesis of hexanoyl-CoA from hexanoate (Stout et al., 2012). Further evidence found by Stout et al. (2012) strengthening this opinion that *CsAAE1* is involved in the cannabinoid biosynthesis is the lack of peroxisomal signal, PTS1. Fluorescent protein fusion assays showing *AAE1* localised to the cytosol and the high expression of olivetol synthase in trichomes provides further evidence for its role as the *AAE* involved in cannabinoid biosynthesis.

1.5.2 MEP pathway

DXS

Deoxyxylulose 5-phosphate synthase (*DXS*) is a thiamine diphosphate (ThDP) dependant enzyme catalysing the reaction condensing pyruvate and D-glyceraldehyde-3-phosphate to form 1-deoxy-D-xyulose 5-phosphate (DXP) (Julsing et al., 2007). *DXS* binds ThDP at the active site forming a covalent intermediate between enzyme bound ThDP and pyruvate, followed by glyceraldehyde-3-phosphate stimulated decarboxylation of pyruvate forming *DXP* (Banerjee et al., 2016). This first step in the MEP pathway has been identified as rate limiting in the biosynthesis of isoprenoids (Estévez et al., 2001).

DXR

1-deoxy-D-xyulose 5-phosphate reductoisomerase (*DXR*) is a NADPH- dependent enzyme catalysing the formation of DXP into 2-C-methyl-D-erythritol 4-phosphate (MEP). Two modes of action have been proposed for this catalysed reaction. The first proposal, the α-ketol rearrangement mechanism begins with deprotonation of the C-3 hydroxyl group, after which a 1,2-migration giving methylerythrose phosphate followed by further reduction to MEP by NADPH (Munos et al., 2009). In the second proposed mode of action, the retro-aldol/aldol mechanism, *DXR* cleaves the C3-C4 bond of DXP in a retro-aldol manner generating a three carbon and two carbon phosphate bimolecular intermediate. Both these products are then reunited by an aldol reaction forming a new carbon bond. Munos et al. (2009) investigated both these mechanisms preparing [3-^{2}H]- and [4-^{2}H]-DXP carrying out a kinetic isotope effect study of the *DXR* reaction. The normal 2° kinetic isotope effect provided evidence supporting the retro-aldol/aldol mechanism, though it is stated that further investigation is required.

MCT

4-diphosphocytidyl-2C-methyl-D-erythritol synthase (*MCT*) catalyses the formation of 4-diphosphocytidyl-2C-methyl-D-erythritol (CDP-ME) from CTP and 2C-methyl-D-erythritol 4-phosphate (Rohdich et al., 2000).

CMK

4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (*CMK*) catalyses ATP-dependent phosphorylation of CDP-ME to 4-diphosphocytidyl-2C-methyl-d-erythritol-2-phosphate (CDP-ME2P) (Shan et al., 2011). *CMK* belongs to a large gene family including galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase (GHMP family) (Lange et al., 2000). Homologues of *CMK* have been identified in eubacteria, however, in plant plastids only a singular gene has been identified.

MDS

2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*MDS*) is a zinc dependent enzyme participating in the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl (DMAPP) (Miallau et al., 2003). This reaction in the non-mevalonate pathway is the only ATP-dependent step catalysing the transfer of the γ-phosphoryl moiety of ATP to *MDS*. The product undergoes two enzymatic steps, reduction and elimination, forming 1-hydroxy-2-methyl2-E-butenyl-5-diphosphate and further reduced to IPP and DMAPP (Miallau et al., 2003).

HDS

4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (*HDS*/*GcpE*) catalyses the reduction of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate into (*E*)-4-hydroxy-3-methylbut-2-en-1yl diphosphate via two one-electron transfers in the MEP pathway (Seemann et al., 2005). *HDS* has been shown to require intact 4Fe-4S clusters (iron-sulphur molecular ensembles) for maximum activity, similar to the bacterial homologue, *GcpE.* Unlike the bacterial enzyme, *HDS* in plants has not been shown to utilize NADPH/flavodoxin reductase as a reducing shuttle system. Two additional structural domains, N-terminal extension and a central domain, have been discovered in *Arabidopsis* that are absent in *E.coli* (Querol et al., 2002). The N-terminal region was found to target the protein to the chloroplast, consistent with its role in the MEP pathway.

HDR

1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (*HDR*) is the last step in the MEP pathway catalysing the reduction of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (HMBPP) producing isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Kim et al., 2008). *HDR* converts HMBPP into a 6:1 mix of IPP to DMAPP, which is further adjusted to 3:7 by IPP isomerase (Rohdich et al., 2003). Like *HDS*, *HDR* is required to have either a [4Fe-4S] or [3Fe-4S] cluster for catalytic activity (Xiao et al., 2009).

1.5.3 GPP pathway

IPP isomerase

Isopentenyl diphosphate isomerase (*IPP*) catalyses the reversible isomerisation of IPP to DMAPP, the common building blocks which all isoprenoids derive (Rodríguez-Concepción et al., 2000). IPP isomerase requires two divalent Zn^{+} cations located in a His_3Glu_2 pocket for activity, while the enzyme-substrate complex contains Mg^{2+} facilitating substrate binding (Lee & Poulter, 2006). The catalytic mechanism involving protonation/deprotonation of C67 and E116 has been suggested analysing the crystal structure with transition state analogues or irreversible inhibitors (Wu et al., 2005).

GPP synthase

Geranyl pyrophosphate synthase (*GPP*) is responsible for the C_{10} skeletons of monoterpenes crucial for defence mechanisms against pathogens and pollination (Langenheim, 1994). IPP and DMAPP are condensed into immediate precursors, geranyl diphosphate and farnesyl diphosphate, formed by geranyl pyrophosphate by the GPP synthase heterodimer. The large subunit of GPP synthase contains 2 aspartate rich motifs, which are important in prenyl-substrate binding. However, the small subunit lacks this motif and is inactive by itself (Wang & Dixon, 2009).

1.5.3 Cannabinoid pathway

OLS

Olivetol synthase (*OLS*) is a novel type III polyketide synthase forming olivetolic acid (OLA), an alkylresorcinolic acid forming the polyketide nucleus of cannabinoids, and the first step in cannabinoid biosynthesis pathway (Gagne et al., 2012). *OLS* catalyses the condensation of hexanoyl-CoA with three molecules of malonyl-CoA yielding OLA (Fellermeier & Zenk, 1998). This reaction, C2 to C7 aldol cyclisation, retains the carboxylate moiety, a rare event in plant polyketides. The large superfamily of polyketide synthases has made previous efforts into identifying the correct synthase responsible for OLA synthesis difficult. Currently, it has been hypothesised that the inability of the previously identified synthases to produce OLA is due to the absence of a crucial protein, OLA cyclase (*OAC*). Gagne et al. (2012) analysed the transcriptome of cannabis trichome cells to identify *OAC* functioning in concert with *OLS* to form OLA. *OAC* catalyses the C2-C7 intramolecular aldol condensation with carboxylate retention to form olivetolic acid. *OAC* is a dimeric α+β barrel (DABB), structurally homologous to polyketide cyclases from *Streptomyces*. *OAC* transcriptome profile found *OAC* to be concentrated in glandular trichomes matching other cannabinoid enzymes. *OAC* was expressed in *E.coli* with the purified protein assayed for hexanoyl-CoA and malonyl-CoA synthesis. OLA was produced, having no intrinsic polyketide synthase activity and with OLA only produced in the presence of *OLS*. It is hypothesised that *OAC* functions as an enzyme acting on the intermediate produced by

OLS, or it may alter the catalytic properties of *OLS* through allosteric regulation. Further investigations into the relationship between *OAC* and *OLS* is needed.

Prenyltransferase

Prenyltransferase (*PT*), also known as geranylpyrophosphate:olivetolate geranyltransferase (*GOT*) or aromatic prenyltransferase, catalyses the condensation of OLA with geranylpyrophosphate (GPP) to form cannabigerolic acid (CBGA) (Fellermeier & Zenk, 1998).

CBDAS

CBDAS catalyses the stereoselective oxidative cyclisation of CBGA, producing cannabidiolic acid (CBDA) (Figure 1.7). CBDA is the dominant constituent of fibre-type cannabis plants compared to drug-type plants (Shoyama et al., 1975). Initial studies were conducted in rapidly expanding leaves of cannabis with *CBDAS* being isolated from cytosolic fractions (Taura et al., 1996). Purification of *CBDAS* by a four-step procedure found the enzyme consisting of a single polypeptide with a molecular mass of 74kDa and a pI of 6.1, monomeric and containing a FAD cofactor. In a later study, the gene encoding *CBDAS* was cloned by reverse transcription and polymerase chain reactions (PCR) with degenerate primers (Taura et al., 2007). Amplification of cDNA ends produced a 1632 nucleotide ORF in size encoding a 544-amino acid polypeptide. The primary structure has 85% identity with *THCAS* (Sirikantaramas et al., 2004), which is expected as *THCAS* and *CBDAS* both catalyse stereoselective oxidocyclisation of CBGA in a fatty acid desaturase mechanism. Initial investigation proposed codominant inheritance, determined by relative cannabinoid concentration, describing the synthase genes were encoded by two alleles at the same loci (de Meijer et al., 2003). Though, evidence currently supports the two synthases are located at separate, genetically linked loci (Grassa et al., 2021).

Heterologous expression of *CBDAS* in insect cell cultures was attempted using a baculovirus-insect cell expression system (Taura et al., 2007). Infected insect cells were analysed after 3 days of incubation with baculovirus, harbouring *CBDAS*, for *CBDAS* activity, however activity was recorded in the culture medium indicating the enzyme was excreted from the cells. Purification of the recombinant enzyme by column chromatography produced a homogenous protein with a molecular mass of 62 kDa, slightly smaller than expected. However, the N-terminal amino acid sequence was identical to the native enzyme suggesting correct cleavage in insect cells.

CBDAS mode of action is currently unclear as the enzyme does not require coenzymes or cofactors for the oxidation of CBGA (Taura et al., 1996), though, sequence analysis indicates the possibility of *CBDAS* possessing a flavin acting as a coenzyme. Stereoscopic analyses of the

recombinant enzyme gave yellow colouration at 366nm on SDS-PAGE and an absorbance maxima at 365nm and 450nm, consistent with those of flavoproteins (Taura et al., 2007).

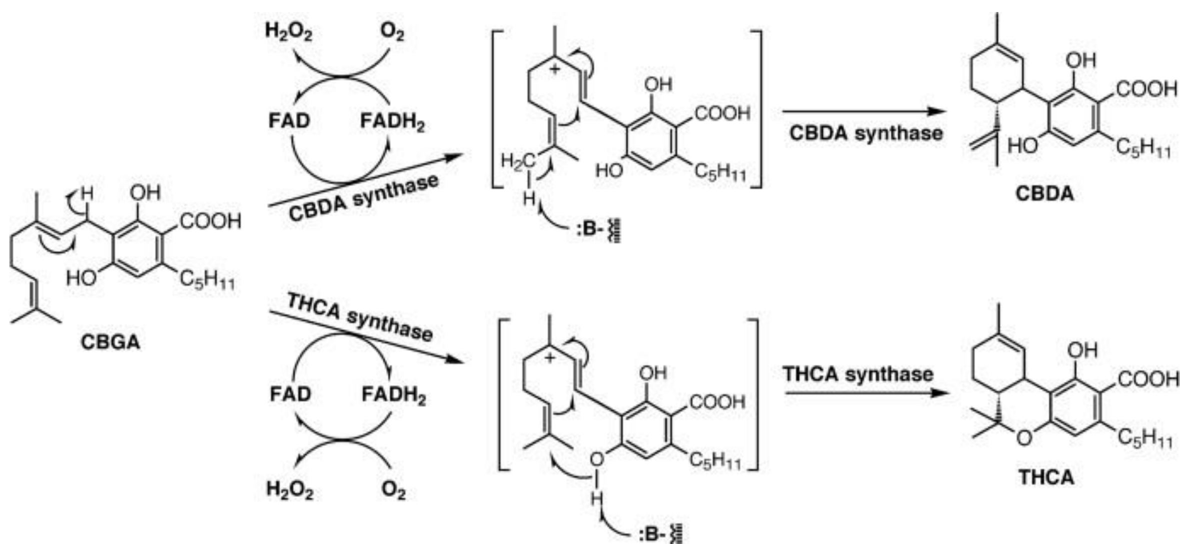


Figure 1.7: Catalysed reactions by *THCAS* and *CBDAS* to produce THCA and CBDA, respectively. Image taken from Taura et al. (2007).

THCAS

THCAS catalyses oxidative cyclisation of CBGA into THCA, the precursor of THC and dominant cannabinoid in drug-type plants (Sirikantaramas et al., 2004) (Figure 1.7). THCA was initially thought to be synthesized by the isomerisation of CBDA, though the pure drug-type Mexican strains did not contain CBDA (Shoyama et al., 1975). The first attempt at isolating *THCAS* from crude enzyme extracts of drug-type strains did not identify the enzyme regardless of using multiple techniques (Taura et al., 1995). Production of THCA was present in the soluble fraction of leaf bud tissue with the addition of CBGA as the substrate. This confirmed the biosynthesis of THCA from CBGA in a stereoselective oxidative cyclisation action from *THCAS*.

Purification of *THCAS* by column chromatography, producing a homogeneous protein was used for partial amino acid sequence determination (Sirikantaramas et al., 2004). Degenerative PCR and amplification of cDNA ends were used to clone *THCAS*. Slightly larger than *CBDAS*, the *THCAS* gene consists of a 1635 nucleotide ORF encoding a 545-amino acid polypeptide, this was the first cannabinoid biosynthesis gene to be cloned.

The deduced primary structure of *THCAS* was found to not be similar to other monoterpene cyclases due to the unique reaction coupled with a two electron oxidation (Croteau, 1987). *THCAS* does, however, have high homology to berberine bridge enzyme, a FAD oxidase involved in secondary metabolism (Dijkman et al., 2013). Supporting the deduced primary

structure of *THCAS*, high expression of recombinant protein by baculovirus allowed biochemical studies of the enzyme. The results drew parallels between the reaction similarities of *THCAS* and FAD oxidases (Kutchan & Dittrich, 1995). Crystals of the recombinant *THCAS* were prepared to determine the tertiary structure by X-ray crystallography (Shoyama et al., 2012). Further structural similarity of *THCAS* with FAD oxidases were seen with the covalent attachment of the FAD coenzyme at His112 and Cys176.

Hydride transfer from C-1 of CBGA to the FAD isoalloxazine ring initiates the reaction mechanism of *THCAS* (Sirikantaramas & Taura, 2017). THCA is formed by the removal of a single proton from the hydroxyl group of CBGA followed by the closure of the stereoselective ring. Transfer of the single hydride from the reduced flavin to molecular oxygen, creates H_20_2, and reactivates the flavin for the following reaction cycle (Sirikantaramas & Taura, 2017).

1.6 Cannabis genome, tools and resources

Cannabis contains a diploid genome (2n=20) containing a karyotype of a pair of X and Y sex chromosomes and 9 autosomes (Van Bakel et al., 2011). Within the cannabis species, female and male plants are homogametic (XX) and heterogametic (XY), respectively (Nicolas et al., 2004). Hermaphrodite populations of cannabis also exist with male and female flowers containing anthers and carpels. Cannabis is also capable of switching from female to male from changes in the environment or being chemically induced (Ram & Jaiswal, 1970). The size of the haploid genome was determined to be 818Mb for female plants and slightly larger, at 843Mb for male plants (Sakamoto et al., 1998). The difference in genome size is attributed to the slightly larger Y chromosome, which contains highly heterochromatic regions in the long arm. Sex determination has also been linked to the long arm of the Y chromosome, which contains copies of LINE-like retrotransposon repetitive sequences (Sakamoto et al., 2000).

An initial draft of the cannabis genome was published by van Bakel et al. (2011) using next generation sequencing (NGS) Illumina pair-end library construction on Purple Kush (PK), a sativa strain of cannabis. A total of 532Mb (with gaps) of the genome was assembled as well as a transcriptome analysis from six selected tissues, identifying more than 30,000 genes by RNA-Sequencing. Comparison of the cannabinoid pathway genes between drug strain PK and hemp (Finola) strain demonstrated the higher transcription levels of cannabinoid synthesis genes in the drug strain, PK compared to hemp strain Finola, was due to gene expression and not morphological differences (e.g. larger trichomes in drug variety). An online deposit of the genome scaffolds is freely available containing the Finola and PK genome (http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway).

Following the publication of the draft genome, thirteen other genome sequences have been deposited into NCBI (https://www.ncbi.nlm.nih.gov/genome/genomes/11681) containing varying levels of genome coverage, from 285Mb right up to 1333Mb. The most comprehensive of the genome assembly has been conducted by Braich *et. al.* (2020) with the most nucleotides assembled and genes annotated when compared to other available genome assemblies (Braich et al. 2020) (Table 1.4).

Table 1.4: Cannbio-2 genome assembly compared to other widely used published genome assemblies (Braich et al., 2020)

Data Type	**Cb-2[d]**	**Cb-2[r]**	**cs10**	**JL**	**Finola**	**PK**
Number of contigs/scaffolds	8,477	10	10	10	10	10
Assembly size with Ns (Mb)	914	904	854	798	785	640
Assembly size without Ns (Mb)	914	903	714	797	784	639
Largest contig/scaffold (Mb)	1.7	106	105	93	101	79
N50 (Mb)	0.2	91	92	83	87	72
N90 (Mb)	0.05	72	65	69	50	51

[d] Draft Cb-2 genome assembly. [r] RaGOO assigned Cb-2 genome assembly using cs-10 as the reference

Molecular diagnostics for the problematic sex determination have been developed. As female plants are required for commercial production, and any males present within the production environment will significantly reduce floral bud yield by undesired pollination. Predominantly, male sex-linked molecular markers are used for determination, however occasionally specific random amplified polymorphic DNA (RAPD) markers in females plants are used (Shao et al., 2003). Initial isolation of DNA fragments exclusive to males were discovered by Sakamoto et al. (2005) who developed the term Male Associated DNA Sequence (MADC1), a 729bp fragment obtained by RAPD encoding a reverse transcriptase homologous to the LINE-like retrotransposons on the Y chromosome. Further comprehensive studies identified 3 more MADC regions, MADC2, MADC3 and MADC4 (Mandolino et al., 1999; Sakamoto et al., 2005), all of which can be used for sex determination through simple PCR experiments described by Mandolino et al. 1999. However, despite the simplicity of these PCR based assays, correct determination can be problematic as the amplified DNA needs accurate resolution of the size in all instances, which can sometime cause challenges, the assay also lacks the capability of identifying hermaphrodites.

Another important consideration on large commercial scale is the prediction of chemotypes within a population. The single gene model of chemotype inheritance has led to the development of associated markers to assist in determining chemotype. The first marker, *B190/B200*, was developed to identify *THCAS* and *CBDAS* homozygous and heterozygous plants (de Meijer et al., 2003). Soon after the sequencing of *THCAS*, a new multiple PCR marker, *B1080/B1192*, was developed (Pacifico et al., 2006) giving an additional tool to help in selection of plants for production.

1.7 Double strand breaks in plants

Genome engineering aims to precisely alter nucleic acids in living cells to generate desired modifications. This is achieved by taking advantage of the DNA repair pathways and the imperfect nature of DNA repair by the introduction of a double strand break (DSB). DSBs result from a break in the phosphodiester backbone in both strands of the DNA. A DSB can occur from ionizing radiation, being chemically induced, spliced by a specific nuclease or occur during DNA replication (Pastwa & Błasiak, 2003). When a DSB occurs, two primary DNA repair mechanisms are used to repair the DSB: homologous recombination (HR) and nonhomologous end joining (NHEJ) (Figure 1.8).

1.7.1 Homologous recombination

Homologous recombination is the repair of DNA DSBs by copying homologous DNA templates into the break point (Hanin & Paszkowski, 2003). HR repair involves the interaction between the snipped DNA molecule and the sister chromatid, with proteins encoded by *Rad52,* an epistatic gene in complimentary DNA annealing, mediating the interaction. A 3' single stranded overhang is produced and serves to direct the complementary DNA strand leading to resynthesis and crossover resolution (Thompson & Schild, 2001). This DNA repair mechanism ensures genome integrity, however it may affect the genome structure by the creation of new alleles or combinations in meiotic recombination events. HR allows for the introduction of foreign DNA sequences into the genome as a recombination event. To allow foreign DNA to be incorporated, the inserted gene fragment must contain flanking sequences identical up and down stream of the target location, thus allowing foreign DNA to replace the endogenous DNA. Often, DNA integrates into the genome through the NHEJ pathway making HR an infrequent event in plants. However, the HR pathway is a precise process that can lead to highly accurate repairs and specific modifications, with fewer mutations compared to NHEJ, which is intrinsically mutagenic (Mao et al., 2008). The frequency of HR in plants has been reported to range from 10^{-4} to 10^{-6}, meaning that a large number of transformed cells need to be generated and screened to identify the rare recombination event (Voytas, 2013). The first work of DSB repair in plants was carried

out in the early 90's by Puchta et al. (1993). Puchta's team integrated a defective reporter gene, *GUS,* into the tobacco genome carrying a rare-cutting restriction enzyme site, *I-SceI*. The cleavage of the *GUS* gene by *I-SecI* and repair by HR restored the function of *GUS*, demonstrating the first documented HR event within plants. Soon after the first HR event, Kempin et al. (1997) reported the first HR recombination event in *Arabidopsis* (Kempin et al., 1997). The *AGAMOUS-LIKE5* (*AGL5*) MADS-box gene was knocked out using *Agrobacterium*-mediated transformation by HR. Highlighting the scarcity of HR events, only 1 successful transformation event was recovered from 750 transformants.

1.7.2 Non-homologous end joining

In plants, NHEJ is the main mechanism for DSB repair leading to genomic changes such as deletions, insertions or rearrangements to the break site (Gorbunova & Levy, 1997; Pipiras et al., 1998). Classic NHEJ repair involves the dimerization of KU70 and KU80 proteins to form KU protein complexes that bind directly to the DSB initiating repair. Different classes of NHEJ repairs in plants have been characterised (Lehman et al., 1994; Nicolas et al., 1995): repair of the break is accompanied by incorporation of "filler" DNA, or break ends are joined with or without deletions (Puchta, 2004). If imprecise repair occurs, mutations can cause knock outs or protein misfolding altering gene function (Curtin et al., 2011). Frequency of imprecise to precise NHEJ of the DSB is undefined, however, imprecise repairs occur frequently enough that mutations can be recovered (Voytas, 2013).

Salomon and Puchta (1998) first observed NHEJ in the tobacco gene, cytosine deaminase, with a restriction site between the promotor and ORF of the gene, pushing it out of frame (Salomon & Puchta, 1998). A large population of 5-fluorocytosine resistant clones were found to have unique genomic sequences inserted into the DSB. More recently, Qi et al. (2013) found the most common mutation, from DSBs, found in *Arabidopsis* were short bp deletions, accounting for approximately 75% of NHEJ mutations (Qi et al., 2013). The NHEJ mechanism is also shown to play a large role in stable transformations of rice. Saika et al. (2014) discuss their recent finding in stably transformed rice calli, demonstrating that NHEJ knock-down lines of Ku70 and Ku80 genes decreased the transformation frequencies up to 70%, highlighting the importance of the NHEJ in stable transformation (Saika et al., 2014).

Understanding and harnessing the repairing mechanism now offers the ability to carry out gene function studies in the pathway of choice. The application of specific endonucleases (discussed in detail below) removes the difficulty in creating DSB using previous methods described. The addition of the genome editing tools, harnessing the repair mechanisms, will allow for the study of practically any organism.

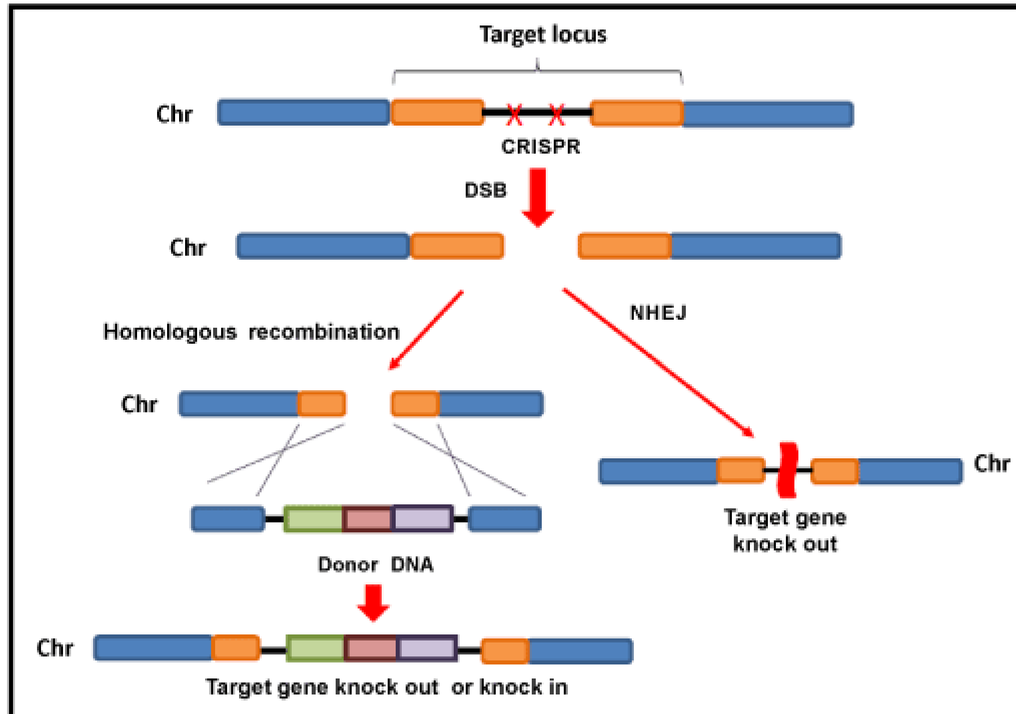


Figure 1.8: DSB repair by HR and NHEJ. Image taken from Shafie et al. (2014).

1.8 Targeting DSBs with specific endonucleases

Traditionally, only a select few model organisms have been used to investigate roles of genes in a biological system. The ability to add or delete genetic information has been rather limited, consequently leading to such detailed exacting knowledge of gene function and underlying biological mechanisms in only a few organisms. The discovery of targeting endonucleases is promising for the study of gene function in practically any organism. The ability to induce a DSB at a targeted locus, followed by the desired modification during DNA repair has enabled precise genetic modification in a growing number of organisms. Three engineered endonucleases are regularly used to introduce targeted modifications at endogenous loci leading to gene disruption, gene expression correction and gene addition: Zinc finger nucleases (ZFNs), TAL effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR/Cas). These nucleases are programmable, sequence specific modules linked to a DNA cleaving domain (Gaj et al., 2013). Here, a review of these nucleases and applications for genetic analysis is discussed.

1.8.1 Zinc finger nucleases

ZFNs were the first engineered nucleases to cleave site-specific DNA in the model organism *Drosphilia melanogaster* (Bibikova et al., 2003). A ZFN contains a tandem array of Cys2-Hys2 finger domains linked to the *FokI* catalytic domain, with the finger domains each recognising 3bp of DNA (Wolfe et al., 2000) (Figure 1.9). A single zinc-finger consists of approximately 30 amino acids in a conserved ββ configuration (Beerli & Barbas III, 2002; Gaj et al., 2013). The finger arrays are fused to the catalytic domain of *FokI* functioning as a dimer. Binding of the zinc-fingers to the target loci brings the two *FokI* monomers into close proximity causing them to dimerise, creating a DSB (Kim et al., 1996). Initial studies using ZFNs, 3 fingers were used to bind a 9-bp target, enabling a DNA cleaving site of 18bp. More recently, the addition of more fingers (up to 6 per ZFN) allows for more specificity.

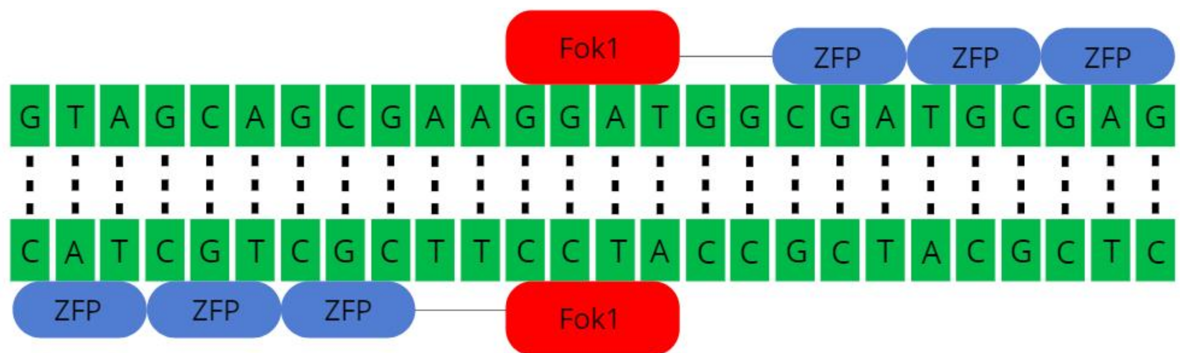


Figure 1.9: ZFN architecture to induce DSBs.

The *FokI* domain is the critical aspect of ZFNs popularity as it possesses many characteristics supporting the goal of targeted DNA breaks. To cleave DNA, *FokI* must dimerise with an adjacent *FokI* domain on both orientations with appropriate spacing to permit the dimer formation (Vanamee et al., 2001). Cleavage domains have also been designed to only cleave as a heterodimer as well as novel spacing requirements for binding events enhancing the specificity, reducing homodimers and improved cleaving activity (Miller et al., 2007; Shimizu et al., 2009).

1.8.2 TALENs

Shortly after the discovery of ZFNs, transcription activator-like effector nucleases (TALENs) were added to the toolkit of specific nucleases. TALE proteins are naturally occurring from the pathogenic bacteria genus *Xanthomanas*, containing DNA-binding domains consisting of ~34 aa repeat domains each recognising a single base pair (Gaj et al., 2013) (Figure 1.10). Within the ~34 aa, a repeat variable di-residue (RVD) at positions 12 and 13 dictate the specificity of binding to the DNA helix (Chen & Gao, 2013). Four TAL effectors with specific binding

domains have been identified binding to the four amino acids A,T,C & G. TALEs are injected into the host by the bacteria through the type III secretion system, interfering with the normal transcriptional activities, up regulating specific genes (Cermak et al., 2011). TALEs contain specific structural features such as secretion and translocation signals, nuclear localization and acidic transcription activation domains allowing binding to DNA. Like ZFNs, TALENs contain an engineered specific DNA binding domain linked to a *Fok*1 cleavage domain. The customisable nature of the DNA binding domain can target any sequence according to the RVD (Bogdanove & Voytas, 2011). Although the single base recognition of the DNA binding repeats allows for more specificity in design compared to the triplet finger design on ZFNs, cloning of repeat TALE arrays are a challenge due to the extensive identical sequences (Cermak et al., 2011).

The cleaving mechanism creating DSBs is similar to ZFNs. Both contain the *Fok*1, which functions as a dimer, with TALENs designed in pairs to allow both monomers to bind at adjacent locations on the DNA separated by a DNA spacer. This allows the *Fok*1 to form dimers, cleaving DNA and activating the DNA repair mechanisms.

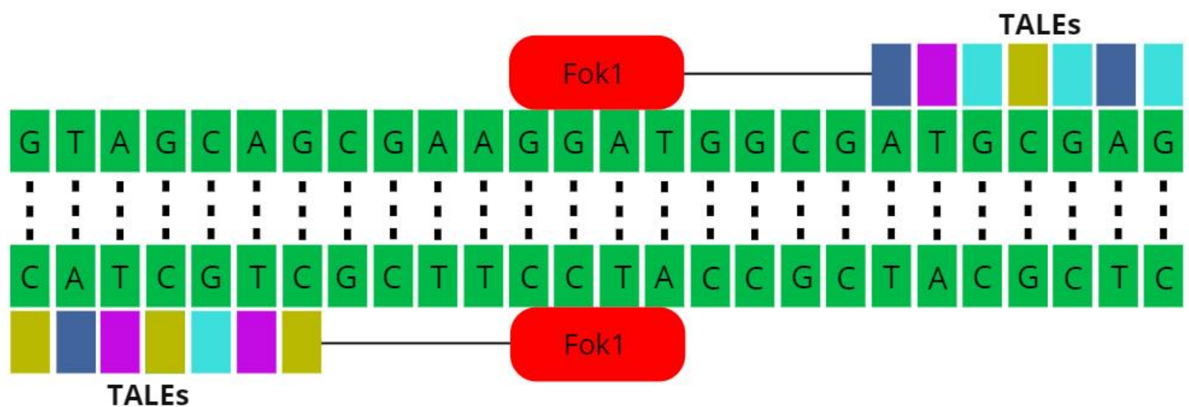


Figure 1.10: TALENs architecture to induce DSBs

1.8.3 CRISPR/Cas9

CRISPR is the adaptive immunity mechanism found in many bacteria and archaea providing resistance from invading DNA. The newest of the genome editing technologies, CRISPR has provided the simplest form of genome editing compared to ZFNs and TALENs. Through evolution, bacteria and archaea have developed a defence mechanism against viral and plasmid infections by copying short segments of exogenous DNA into the host genome at a locus of CRISPRs (Terns & Terns, 2011). This locus acts like a genomic memory of the invading DNA allowing for quick transcription of crispr RNA (crRNA). Foreign DNA will be recognised by the

crRNA through base-pairing, recruiting CRISPR-associated endonuclease (Cas) which will then introduce a DSB into the exogenous DNA stopping integration and replication (Figure 1.11). Unlike TALENs with tandem repeats, CRISPR repeat clusters are separated by non-repeating sequences known as spacers, with these elements adjacent to Cas genes (Jansen et al., 2002). Another highly important factor controlling Cas binding is the presence of a protospacer-adjacent motif (PAM) site, which is an essential targeting component giving the Cas the ability to distinguish self from non-self. The CRISPR system has the advantage of being reprogrammable by simply changing the single-guide RNA (sgRNA), also with the ability of multiplexing, targeting multiple genes at once (Cong et al., 2013).

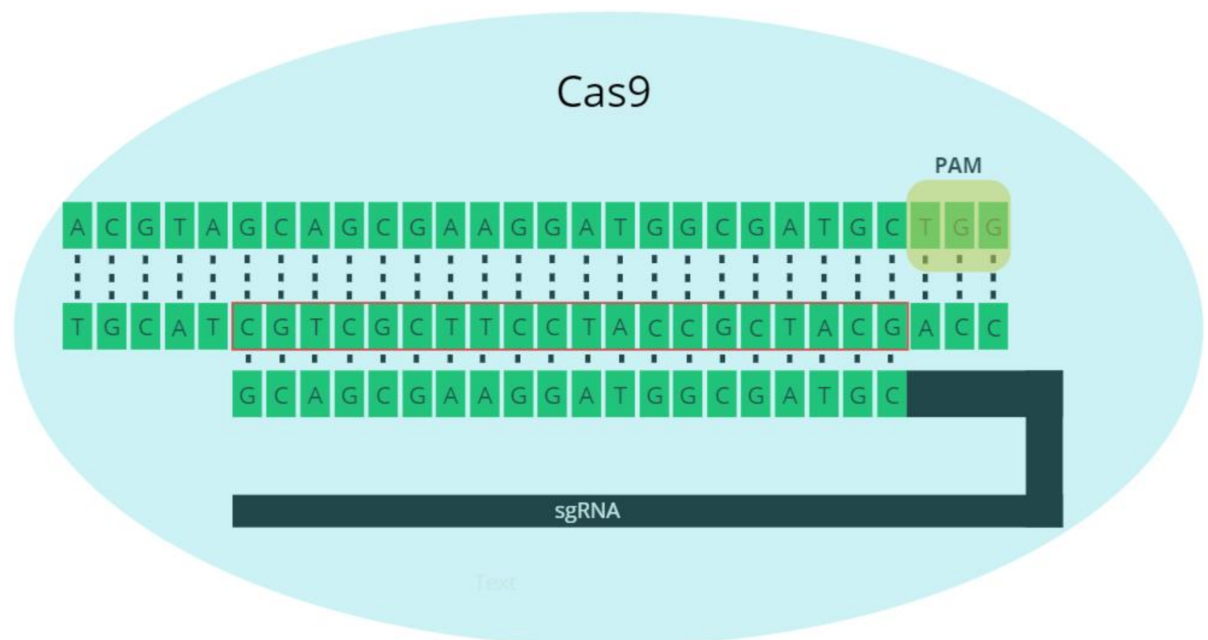


Figure 1.11: CRISPR architecture to induce DSBs

Designing CRISPR modules for sequence specific DSB is comparatively simpler than ZFNs or TALENs due to the sgRNA/Cas complex. Only the 5' 20nt sequence sgRNA needs to be designed complementary to the endogenous loci. Most importantly, offsite cleaving needs to be considered when designing sgRNAs, which is caused by mismatching in the specific sgRNA. Computational tools have been developed for designing sgRNAs with high specificity and efficiency. An extensive list of computational tools is discussed by Cui et al. (2018). To predict on-target efficiency, Doench et al. (2014) have developed rules, which are employed by the most common online tools, such as CHOPCHOP (Montague et al., 2014), E-CRISP (Heigwer et al., 2014) and PROTOSPACER (MacPherson & Scherf, 2015). The rules stipulate for greatest on-target efficiency, a support vector machine (SVM) chooses sgRNAs with a logistic regression classifier giving a score between 0 and 1 (Pei et al., 2018), with 1 being most effective. The approach was improved again by Root's lab by the inclusion of counts of position-independent

nucleotide, location of target site and melting temperature (Cui, Xu, Cheng, Liao, & Peng, 2018; Doench et al., 2016). Further models for on-site prediction, such as Elastic-Net (Zou & Hastie, 2005), are discussed in detail by Cui et al. (2018).

To assist with the prediction of off-target effects from mismatched sgRNAs, a variety of tools have also been developed. A common online tool used, http://crispr.mit.edu, was develop by the Zhang laboratory (Hsu et al., 2013) through investigation of several hundred sgRNA variants and their accompanying off-target loci with indels mutations more than 100bp. Mismatch tolerance was found to be influenced by number and position of mismatches, leading to the development of a penalty matrix, with scores given between 0 and 1, with 1 having a bigger effect on cleavage. Each position in the sgRNA can be assigned a score according to its potential off-target potential. This approach has also been adopted by several other online tools, including CHOPCHOP.

Even though the concept of genome editing has been around for a considerable amount of time, targeted genome editing in plants is still relatively young. Many more studies are needed to prove the efficiency of on-site targeting and reduction in off-site targeting, which is always a risk when engineering specific sequencing nucleases. Regarding the engineering of specific sequencing nucleases, much can be learned from the ever-growing publications in genome editing and development of online tools for ZFNs, TALENs and sgRNA construction. These newer technologies enabling transcriptional and translational regulation should be used in conjunction with more traditional methods to gain greater control of gene expression. The ability to cause targeted mutagenesis has great potential in more than just the model plants mentioned. With the increasing interest, new approaches to modify biosynthetic pathways, increase herbicide resistance and countless other benefits can be achieved in shorter time frames and with more precision.

1.9 Delivery methods for gene editing and gene suppression constructs

1.9.1 Electroporation-mediated transformation

Electroporation uses a physical transfection method using electrical pulses to temporarily create holes in cell membranes where gene editing vectors can easily pass into the cytoplasm. The introduced DNA can be expressed transiently and shortly after, stably incorporated into the genome of the host.

This technique is widely used in plant protoplasts to create transgenic plants to study transient gene expression. This technique has also been used to regenerate stably transformed plants, including rice (Toriyama et al., 1988) and sugarcane (Arencibia et al., 1995). The first recorded use of electroporation gene therapy was performed in 1982 with the delivery of plasmid DNA into mouse L cells (Neumann et al., 1982). Since, electroporation has been adapted to be used in

the delivery of genome editing constructs into plant protoplasts, including CRISPR/Cas9 (Bhowmik et al., 2018) and ZFNs (Wright et al., 2005).

1.9.2 PEG mediated transformation

Polyethylene Glycol (PEG) is a polycation used in the fusion of protoplasts, and subsequently the transfer of DNA across membranes, however the mode of action is not fully understood. DNA, such as plasmids, is incubated with protoplasts with the addition of PEG and a divalent ion to allow the transfer of the DNA through the membrane. It is hypothesised that PEG compacts the DNA allowing it to associate with the membrane by means of the neutralisation of charges in the DNA and the membrane (Shillito, 1999). The high osmotic pressure of PEG causes water to be released from the protoplast, this is where it is supposed the DNA enters the cell. As early as 1985, the uptake of plasmid DNA was observed in wheat with the selectable marker NPT II (Lörz et al., 1985). Similar to electroporation, this method has been used for transient expression and stable transformation studies using genome editing constructs in protoplasts, including banana (Wu et al., 2020) and corn (Sant'Ana et al., 2020).

1.9.3 Agrobacterium-mediated transformation

Agrobacterium tumefaciens is a soil based phytopathogen infecting plant wound sites causing crown gall disease through the delivery of T-DNA into the host plant cells. The T-DNA is nestled on a Ti plasmid between a left and right border, which can be modified using multiple cloning techniques to introduce any DNA sequence of choice. On the Ti plasmid reside *vir* genes, which play an essential role in the genetic transformation process. The Vir proteins function as members of sensory signal transduction genetic regulatory systems and are induced by the plant wound phenolic compounds (Gelvin, 2003). Once the virulence system is activated within the bacteria, the T-DNA complex is generated and transferred into the host nucleus where it is integrated into the plant genome where it is expressed (Figure 1.12). The exact mode of action through all the integrated processes is still under some debate, however this method of plant transformation has been extensively used across many plant species. Genome editing constructs for transient and stable transformation in many explant types have been achieved in wheat (Zhang et al., 2018) and tobacco (Chen et al., 2018) among many other species.

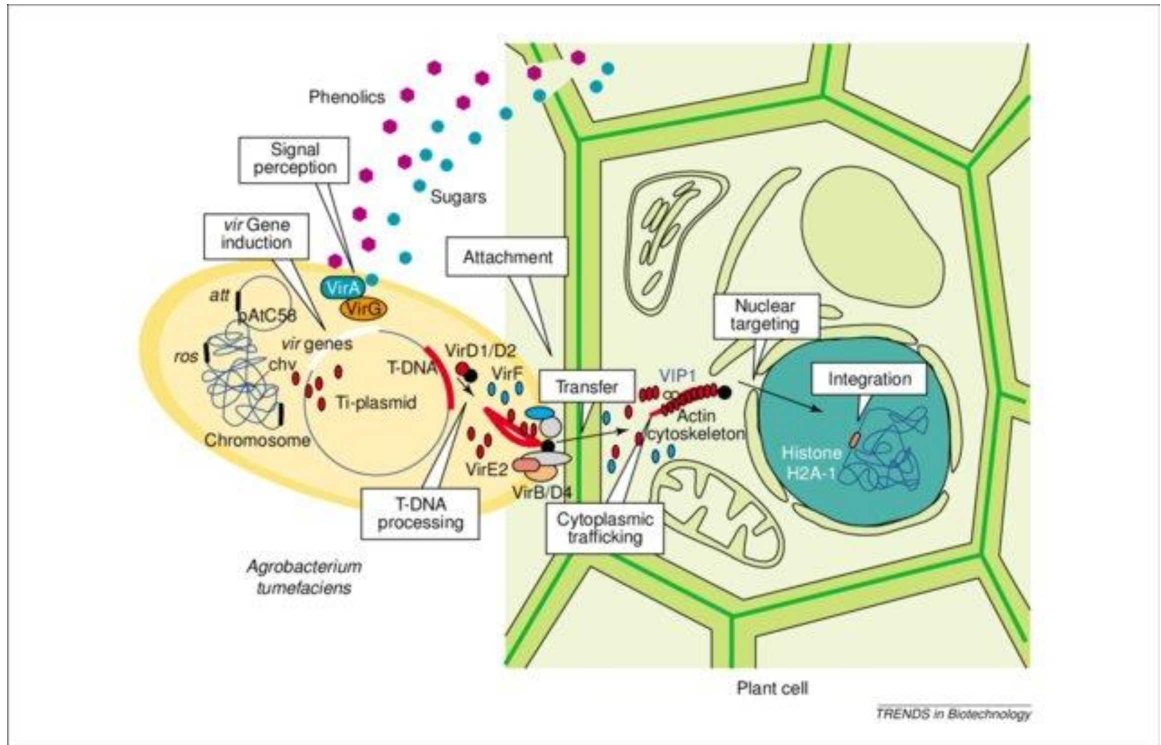


Figure 1.12: Model for *Agrobacterium*-mediated transformation depicting steps involved through the process within the bacterium and plant host. Image taken from Gelvin (2003b).

1.9.4 Particle bombardment

Particle bombardment, or 'biolistics' or 'gene gun', is a direct transfer method for plant transformation by directly delivering DNA into the plant cell nucleus (Figure 1.13). Heavy metal particles, such as gold, are coated with exogenous DNA and fired into cells using a helium vacuum. Particle bombardment was the primary method for plant transformation prior to *Agrobacterium*-mediated transformation but is still widely used today due to the advantage of not having a biological limitation. The widely applicable use of this transformation to produce stably transformed plants was first performed across a large variety of rice genotypes (Christou et al., 1991). This method is still used for delivery of genome editing and gene suppression constructs in sugarcane (Jung et al., 2012), wheat (Liang et al., 2018) and soybean (Bonawitz et al., 2019).

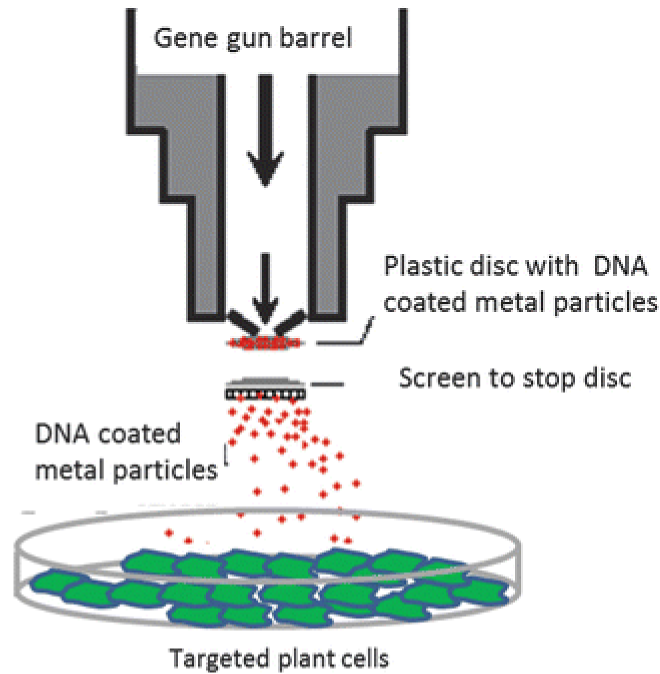


Figure 1.13: Visual representation of gene delivery of DNA coated particles into plant cells. Image taken from Hurst and Finley (2018).

1.10 RNA interference in gene suppression

RNA interference (RNAi) is an evolutionary RNA-dependent gene silencing process in eukaryotes induced by double stranded RNA (dsRNA) and hairpin RNA (hpRNA) (Guo et al., 2016). The discovery of the noncoding RNA in viral infections and their virus-specific responses date back to as early as 1928 (Wingard, 1928). The defence mechanism of RNAi against invading plant viruses and bacterial pathogens is highly specific, targeting and cleaving the intruding RNA through multiple pathways. In general, all pathways include Dicer-like (DCL) and Argonaute (AGO) family proteins (Baulcombe, 2004). dsRNA, or hpRNA, is cleaved by RNAse III polymerase DCL proteins, producing 20-25bp double stranded nucleotide small interfering RNAs (siRNAs) with two nucleotide overhangs at the 3' end of each strand (Majumdar et al., 2017). Each siRNA consists of a guide strand (antisense) and passenger strand (sense), with the guide strand being incorporated into the RNA-induced silencing complex (RISC) and the passenger strand being degraded in the cytoplasm. The siRNA-RISC complex is guided to the messenger RNA (mRNA) target by complementary base pairing initiating cleavage by AGO proteins and preventing transcript translation.

1.10.1 RNAi mechanisms in plants

Multiple RNAi pathways in plants exist, with each pathway being highly specific for the intended purpose of the small RNAs (sRNA). For translational studies in plants, siRNA and microRNA (miRNA) are most commonly used, with trans-acting-siRNA (ta-siRNA), phased-siRNA (pha-siRNA) and RNA-directed DNA Methylation (RdDM) pathways also being used (Figure 1.14). All these pathways utilise DCL and AGO complexes to achieve mRNA cleavage, DNA methylation, translational repression or chromatin modification (Vaucheret, 2008).

siRNA are generated from long dsRNAs introduced into the cytoplasm from viral RNA replication, endogenous dsRNA and single stranded RNAs from hairpin loop secondary structures (Carthew & Sontheimer, 2009). dsRNA is recognised and cleaved by different DCL proteins (DCL-1-4) yielding siRNAs of different lengths. The DCL-1 protein is responsible for cleaving dsRNA into pre-miRNA and processing pre-miRNA into mature-miRNA (Kurihara & Watanabe, 2004). DCL-2 cleaves dsRNA into 22-nt siRNA and co-ordinates with DCL-4 in antiviral defence. DCL-2 is also capable of generating sufficient antiviral siRNA in the absence of DCL-4 (Garcia-Ruiz et al., 2010). DCL-3 creates 24-nt siRNAs responsible for silencing transposons, repetitive elements and plays a role in viral defence (Xie et al., 2004). Finally, DCL-4 generates 21-nt siRNA involved in antiviral defence and generating ta-siRNA in gene regulation (Gasciolli et al., 2005). Post processing of dsRNA into siRNA by DCL proteins, the siRNAs are incorporated into the RISC complex and the guide strand targets the specific mRNA by base pairing, where the AGO protein cleaves the intended target (Zhang et al., 2015).

miRNA are 20-24-nt sRNA, transcribed, capped and polyadenylated by RNA polymerase II, from endogenous MIR genes (Bartel, 2009; Kim, 2005). This results in a pri-miRNA with 3’ and 5’ overhangs, forming an imperfect fold-back stem loop due to sequence compatibility. This complex is then trimmed by DCL-1 in the nucleus into mature miRNA comprised of a guide strand and miRNA* (passenger strand). Similar to siRNA, the mature miRNA is loaded into the RISC complex and one strand is selected as the guide (Meijer et al., 2014).

ta-siRNA are 21-nt sRNAs originating from non-coding transcripts, named TAS genes, processed by RNA polymerase II (Vazquez et al., 2004). ta-siRNAs are synthesised by miRNA cleaving fragments of the TAS transcript and converted to dsRNA by RDR6, followed by processing of DCL-4 into 21-nt siRNA (Xie et al., 2005). ta-siRNA production is highly specific, with only 22-nt miRNA able to initiate production. Similar to the miRNAs, ta-siRNA are methylated by HEN1 and subsequently interact with AGO1 and AGO7 to degrade target mRNA.

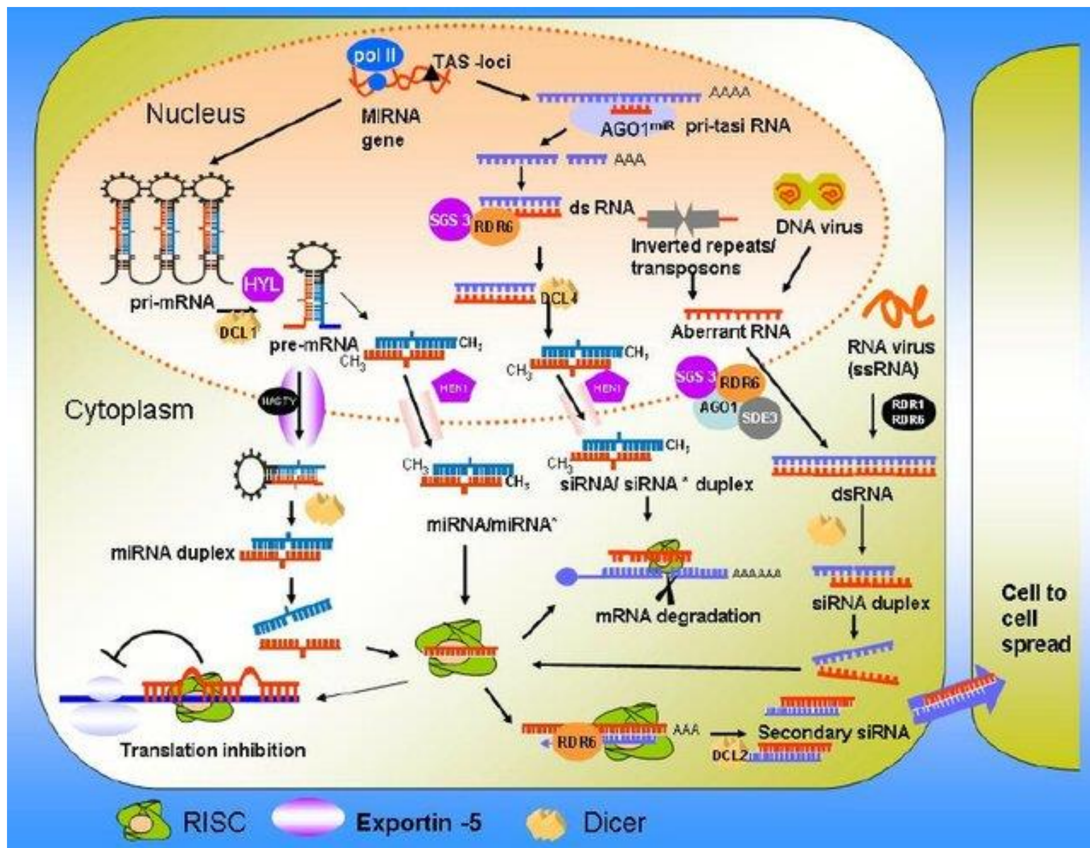


Figure 1.14: Different RNAi pathway in plants that regulate expression of genes. Image taken from Sanghera et al. (2010).

The RdDM pathway is found only in plants and is responsible for DNA methylation and transcriptional silencing in the nucleus and playing an integral role in silencing transcriptional elements and repetitive DNA to maintain genome stability (Xie et al., 2005). 24-nt siRNAs are generated by RNA Polymerase IV, RDR2 and DCL-3 direct RdDM. The mode of action is initiated with Polymerase IV transcribing highly repetitive DNA generating single stranded RNA, which is then converted to dsRNA by RDR2 and subsequently processed by DCL-3 into the 24-nt siRNA, which are methylated at the 3' hydroxyl group at the terminal nucleotides by HEN1 (Li et al., 2005). AGO4 loads the 24-nt siRNA to form RISC, which interacts with non-coding RNA from Polymerase V, recruiting DRM2 causing *de novo* DNA cytosine methylation.

AGO proteins bind the sRNAs to form RISC for transcriptional and post transcriptional gene silencing. The understanding of AGO proteins and their respective sRNA binding preferences has evolved through mutational studies mainly in *Arabidopsis*, rice and sorghum (Kapoor et al., 2008; Liu et al., 2014). sRNA interacts with each domain in AGO proteins, more specifically the PAZ domain containing RNA 3' terminus binding used to bind the guide strand of dsRNA, the 5' prime end of the guide strand interacts with the Mid domain while the rest of the guide strand

tracks along the positively charged surface. It has been shown the PIWI domain folds, catalysing the endonucleolytic cleavage of mRNA through base pairing, beginning the first step in mRNA silencing events (Parker et al., 2004).

1.10.2 RNAi genome engineering

Using the RNAi pathways described, gene silencing, DNA methylation and chromatin modification is relatively easy to achieve through developed technologies. RNAi genome engineering involves delivering constructs into the plant cells to generate siRNA or miRNA to be processed by DCL proteins, with each construct specifically designed to initiate the different pathways.

The delivery techniques of RNAi vector constructs is similar to other genome engineering techniques, with dsRNA being introduced by projectile bombardment (Schweizer et al., 2000), *Agrobacterium* infiltration (Johansen & Carrington, 2001) and virus induced gene silencing (VIGS) (Kumagai et al., 1995).

In plant genome engineering, most commonly siRNAs are generated through the expression of hpRNA, which fold back to create dsRNA, acting as substrates for DCL proteins to create a large quantity of siRNAs. The first works on generating siRNAs by hpRNA showed that gene constructs containing intron-spliced RNA induce post transcriptional gene silencing (PTGS) with almost 100% efficiency (Smith et al., 2000). This discovery of sense and antisense copies of the target gene separated by an intron has led to the development of easy, sequence interchangeable plasmids for the modification of any chosen gene. The intron, also known as a spacer, contains a non-complementary sequence, often containing bacterial resistance genes for selection, which provides stability for the hpRNA construct as repetitive inverted repeat DNA is unstable in bacteria. The sense and antisense arms create a dsRNA arm, which does not require RNA dependent RNA Polymerases to generate the dsRNA. The hpRNA is processed by DCL-4 to create 21-nt siRNAs, though DCL-2 and 3 are also known to process hpRNA into 22 and 24-nt siRNA (Guo et al., 2016). The length of the dsRNA used for gene silencing is also a considering aspect of effective gene silencing with previous reports suggest that the length of the dsRNA plays a significant role in the effectiveness of the vector (He et al., 2020; Wang & Carmichael, 2004).

Initially hpRNA constructs were difficult and tedious to construct due to the inverted repeat structure. However, several commercially available constructs are now available with only a few cloning steps required to complete vector construction. pHannibal and pKANNIBAL have been widely used allowing the insertion of a PCR product by conventional restriction enzyme digestions and DNA ligation steps (Helliwell & Waterhouse, 2003; Wesley et al., 2001).

Improving on the pH/KANNIBAL system, incorporating the simple cloning steps using the LR Clonase technique, pHELLSGATE has been developed which utilises the Gateway unidirectional *in vitro* cloning system (Helliwell et al., 2002). PCR products (of the gene construct of choice) flanked with *att*L1+2 sites can easily be cloned into vectors containing *att*R1+2 sites through modular cloning, which decrease the length of inverted repeat sequences increasing stability in *E.coli.*

Improving on the simplicity to construct hpRNA vectors, the use of Golden Gate Cloning has allowed for the single step cloning of multiple PCR products into vectors such as RNAi-GG (Yan et al., 2012). Using type IIs restriction enzymes, which cut outside of the recognition site to create sticky ends, allow for seamless sequence insertion in both orientations in a single step. Unlike the previous vectors discussed, the orientation of the insert, in regards to the promoter, using Golden Gate Cloning does not influence the functionality of RNA silencing as both orientations are incorporated.

As with other genome engineering tools, off targeting using RNAi is a consideration that needs to be addressed and minimised through careful design. With millions of endogenous sRNAs present in plants, off-target effects could severely affect normal plant function in plant pathogen defence. As with CRISPR and ZFNs, online tools exist which allow the user to design and asses potential off-targets in an attempt to minimise the off-targeting effects (Naito & Ui-Tei, 2012). As few as seven nucleotides of sequence homology between siRNA and the target mRNA can lead to the inhibition of expression (Birmingham et al., 2006).

1.11 Development of cannabis *in vitro* culture

By far the most common method for growing cannabis is using seeds, with their use as starting material in studies having been conducted looking at physiological responses to photo periods (Lisson et al., 2000) to elicitation studies in cannabis cell suspension cultures (Flores-Sanchez et al., 2009). Different methods are available for seed germination including the use of traditional moist soil (Chandra et al., 2013) or moist filter paper, using DARIA induction media (Wielgus et al., 2008) and germination in the dark (Wahby et al., 2013). Although seed propagation is the most common technique, it is not possible to maintain cultivars using seed as this results in large genomic and phenotypic variation, with the potential of male populations being introduced. This limitation requires different techniques to initiate propagation studies for *in vitro* culture since it conserves genetic homogeneities among clones.

The use of maintaining cannabis in tissue culture has the advantage over conventional propagation techniques, allowing for higher multiplication rates, elimination of disease and overcoming of heterozygosity due to the allogamous nature. However, tissue culture systems for cannabis regeneration have only been accomplished in a few instances. A summary of *in vitro* protocols developed for cannabis is covered by Lata et al. (2017) and Monthony et al. (2021). From the limited results in regenerating cannabis, the most successful propagation method is through direct or indirect organogenesis.

1.11.1 Micropropagation

In vitro propagation is advantageous over classic seed propagation due to the high multiplication rate, ability to eliminate disease and overcoming heterozygosity and the ability to maintain a specific heterozygous genotype of an outcrossing species alive, effectively in perpetuity. Nodal segments are removed from a mother donor plant, sterilised through washing with ethanol and bleach and placed into semi-solid medium most commonly containing MS basal salts and vitamins and a mix of auxins and cytokinins (Figure 1.15). This process allows for a high volume of plants to be maintained in a small space, decreasing the required space needed and reducing the costs of operation overall. The earliest studies showed hemp was capable of micropropagation using nodal segments with the use of Indole-3-butyric acid (IBA) and 6-Benzylaminopurine (BAP) hormones (Richez-Dumanois et al., 1986). Since, numerous studies have explored the use of different basal salts, such as MS (Lata et al., 2009b) and DARIA (Wielgus et al., 2008) effect to encourage shoot and root initiation. Exploration and optimisation into the different combinations of auxins and cytokinins on shoot and root initiation have also been explored, including Thidiazuron (TDZ) (Lata et al., 2010) BAP and 1-Naphthaleneacetic acid (NAA) (Wielgus et al., 2008). An extensive tabulation of the different media and hormonal combinations to initiate shoot and root initiation has recently been previously reported (Monthony et al., 2021).

Figure 1.15: Micropropagation of cannabis through the different stages from donor plant (A) to introduction into tissue culture (B+C) to acclimatization into the green house (D-G). Image taken from Thomas and ElSohly (2016).

1.11.2 Protoplast isolation

Using protoplasts as an explant source for genetic transformation offers advantages over other techniques due many transformational events occurring at the same time and the ability to avoid chimerism during regeneration. Protoplasts are plant cells which have had the cell wall removed either mechanically, or most commonly, enzymatically (Figure 1.16). Transforming protoplasts is most commonly achieved through electroporation or PEG-mediated transformation, but to date there is limited available information in applying protoplast isolation and transformational protocols to cannabis. Previously, only two resources existed which, describe in varying detail the method for protoplast isolation with the protoplast concentration and viability collected either not reported or not in significant detail. Similar to other plant species, the combinations of enzymes used to digest the plant cell wall are the same. Jones (1979) in his thesis explored the effect of different enzyme combinations of varying enzymes and varying levels of osmotic solutions (Jones, 1979). The only data collected states that protoplasts were isolated from young and old leaves ranging in protoplast concentrations from 10^3 - 10^5/mL, but provides no exact information, nor the experimental data collection process. Protoplast isolation from a Mexican strain of cannabis have been reported. Morimoto et al. (2007) describes the osmotic solution and

enzymes used to isolate protoplasts but no data is given on protoplast concentration or viability (Morimoto et al., 2007). More recently, three studies have been released exploring the variables involved in protoplast isolation and transfection (Beard et al., 2021) and using protoplasts for sgRNA construct verification (Zhang et al., 2021).

The use of protoplasts as a viable explant for transformational experiments requires further analysis of the variables involved to increase protoplast isolation and viability from a wider gene-pool before this approach can become a viable approach as it is in other plant species.

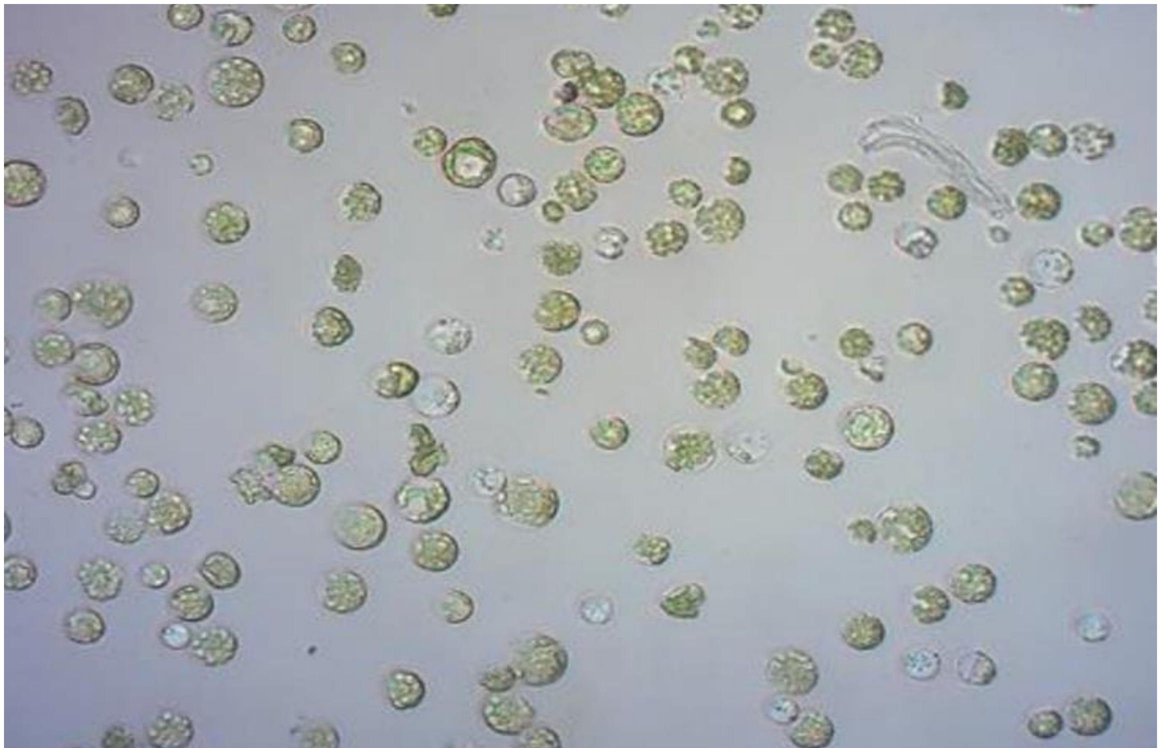


Figure 1.16: Isolated cannabis protoplasts through enzymatic digestion. Image modified from Beard et al. (2021).

1.11.3 Callogenesis

Callus is a mass of unorganised undifferentiated cells that have grown in response to stresses and hormone combinations. Callus can be pluripotent and totipotent; either programmed to become predetermined organs or the ability to regenerate into a whole plant. Direct organogenesis has previously been achieved by using nodal explants containing axillary buds, cotyledons and shoot tips (Lata et al., 2009b). Successful multiplication of shoots from explants were achieved by Lata et al. (2009b). Fourteen shoots per explant was achieved from nodal explants using Murashige and Skoog (MS) with 0.5µM TDZ. TDZ was found to be a better cytokinin than BAP or kinetin. Elongated shoots were transferred to half strength MS supplemented with 500mg of activated charcoal and 2.5µM IBA resulting in a 95% success rate in rooting (Lata et al., 2009b). In a later study, Lata et al., (2016) developed a mass propagation technique in one step for rapid shoot proliferation and invitro rooting from direct organogenesis (Lata et al., 2016). Instead of TDZ, meta-topolin (mT) was used at differing concentration in MS media with roots being produced within 4-6 weeks. One hundred percent of regenerated plantlets survived once being acclimatized in a controlled grow room. Inter simple sequence repeat (ISSR) markers were used to test genetic fidelity in the propagated plantlets, with all plantlets being monomorphic and comparable to the donor mother. Cannabinoid profile was also found to be similar in donor plant and regenerated plantlets by gas chromatography-flame ionization detection (GC-FID).

Indirect organogenesis has also been used to accomplish plant regeneration in hemp. Five different cultivars of hemp were chosen by Slusarkiewicz-Jarzina's laboratory as the source material for leaf explants and petioles (Slusarkiewicz-Jarzina et al., 2005). MS media was used, supplemented with various concentrations of 2,4-dichlorophenoxyacetic acid (2,4-D), Kinetin and NAA. Callus was induced from 83% of petiole explants with 2mg DICAMBA/L and 65% of leaf explants using the same PGR. Regeneration of plantlets from petiole and leaf callus was approximately 2%. Feeny and Punja (2003) also obtained callus formation from leaves and petiole, however could only achieve root production in leaf explants (Feeney & Punja, 2003). In depth meta-analysis of attempted direct and indirect organogenesis and embryogenesis in cannabis has been reviewed (Monthony et al., 2021).

The production of cannabinoids from callus cultures has previously been shown to be unsuccessful (Pacifico et al., 2008). Cell suspension culture is a further *in vitro* technique used in metabolite production, with the initial cannabinoid production studies dating back to the 1980's. Loh (1983) investigated the production of secondary metabolites from cannabis in suspension cultures from embryo, leaf and stem explants with varying concentrations of PGRs (Loh et al., 1983). Olivetol and CBD were converted to cannabielsoin (CBE) when cell suspensions were inoculated with CBD, and an unidentified cannabinoid was produced when inoculated with olivetol. However, the cannabinoid production was inefficient and unstable due to the need for

the addition of an exogenic precursor. More recently, Flores-Sanchez et al. (2009) treated cannabis cell cultures with biotic and abiotic elicitors, monitoring any cannabinoid production (Flores-Sanchez et al., 2009). H-NMR was used to analyse the metabolic profiles, showing that various levels of metabolites were produced, however no cannabinoids were detected in the cell cultures. Interestingly, THCAS expression in these cell cultures showed no expression in control or elicitor treated cell cultures. Taking these results into consideration, it suggests that in cell suspension cultures, cannabinoid synthesis could not be induced as a defence response. However, due to the specific cells in which cannabinoids accumulate, the most common cells to start a cell suspension culture are derived from parenchyma cells, which could be the source for the failed attempts. It has been reported that sometimes a state of differentiation in the culture is required for the production of secondary metabolites (Ramawat & Mathur, 2007). As it stands, more research is needed to understand the complex requirements for cannabinoid production from cell suspension cultures.

1.11.4 Regeneration

An efficient regeneration protocol is critical for clonal propagation. As mentioned previously, few reports have successfully regenerated cannabis from direct organogenesis (Monthony et al., 2021). Difficulty in regeneration comes partly from the responses to PGRs and media compositions, all of which need to be tailored to the specific strain of cannabis used. Not only are PGRs and media compositions highly variable, explants chosen contribute to the success of regeneration due to their interactions between endogenous growth substances and PGRs (Jones et al., 2007). The first report by Hemphill (1978) successfully obtained root formation from callus, but failed to develop any shoots (Hemphill et al., 1978). Since Hemphill's attempt, different explant material has been used including apical and axillary buds (Richez-Dumanois et al., 1986), leaf callus (Mandolino & Ranalli, 1999) and calli regenerated from differing explants (Slusarkiewicz-Jarzina et al., 2005) in an attempt to regenerate whole plants, with limited results. The highest level of success, in regard to regeneration efficiency, was reported by Lata et al. (2010) with over 90% of leaf callus regenerating. However, when this study was replicated, they failed to recover any regenerates (Monthony et al., 2021) leaving the original results reported speculative.

TDZ has shown to be a popular cytokinin for shoot induction due to the ability to stimulate better shoot proliferation (Parveen & Shahzad, 2010). Lata et al. (2009) developed a direct organogenesis protocol using TDZ from nodal explants, with superiority demonstrated when compared to BAP or kinetin. Shoot regeneration was achieved on MS media containing 0.05-5μM TDZ from nodal explants, with shoots proliferating after 14 days. The optimal concentration of TDZ was found to be 0.5μM, with 100% of explants responding with an average

of 13 shoots per culture. Increasing concentrations of TDZ resulted in suppressed shoot formation. An interesting observation came from the difficulty in root induction from shoots culture on half strength MS without activated charcoal, which has also been reported in different plant species. Shoots rooted on media with differing auxins with the presence of IBA resulted in a twofold increase in roots and root length, with up to 95% success in half strength MS containing 2.5µM IBA and activated charcoal (Lata et al., 2009b).

Shoot regeneration from lateral buds using TDZ and NAA on MS media was investigated by Bing et al. (2007) (Bing et al., 2007) and auxiliary bud induction using TDZ was also investigated by Wang et al. (2009) (Wang et al., 2009). Further examples of using TDZ as the preferred cytokinin are reviewed by Lata et al. (2017).

Issues still exist for regeneration within cannabis, with reproducibility of published protocols on the same genotype questionable (Monthony et al., 2021). The development of a robust, cross cultivar regeneration protocol is needed to avoid the investment required for individualised media compositions.

A simplified approach using hypocotyls for regeneration in cannabis with a lack of media hormones has recently been developed (Galán-Ávila et al., 2021). The pericycle cells within the hypocotyl are programmed to generate whole plants and serve as an excellent option for genetic transformation. The lack of hormones within the media should help reduce cultivar variance to regeneration attempts in the future.

1.11.5 Germplasm conservation

Tissue culture techniques are being used for clonal propagation, conserving high yielding elite clones. Many other plant species germplasm are conserved using techniques such as cryopreservation and synthetic seed technology used in crops such as sweet potato (Pennycooke & Towill, 2000) and conifers (Attree & Fowke, 1993). Currently, there are few studies available on germplasm conservation of cannabis. The first record of hemp cryopreservation dates back to 1989 as a means to preserve suspension cultures (Jekkel et al., 1989). Cryoprotectants applied were DMSO, glycerol, proline and PEG with increasing concentrations with a cooling rate of 2 °C/min. A maximum viability of 58% was achieved using 10% DMSO and -10°C temperature transfer. Synthetic seed technology has been applied to axillary buds isolated from shoot cultures encapsulated in calcium alginate beads (Lata et al., 2009a). Gel complexation was achieved using 5% sodium alginate with 50mM $CaCl_2.2H_20$. Encapsulated explants responded with greatest regrowth to MS supplemented with 0.5µM TDZ and 0.0075% PPM under *in vitro* conditions. *In*

vivo conditions obtained 100% conversion of encapsulated explants when potted 1:1 in potting mix.

1.12 Research plan

This PhD project is designed to develop necessary genomic resources and tissue culture protocols using genome editing targeting the cannabinoid biosynthesis genes. This PhD project is in two components, an initial evaluation of the cannabis genome for the evaluation of polymorphisms and copy number variance in cannabinoid biosynthesis related genes of a large pan-genome, followed by tissue culture protocol development and implementation of RNA interference vector constructs for the modulation of cannabinoid content.

The initial evaluation of the cannabis genome copy number variance and nucleotide polymorphisms is necessary due to the genetic diversity that exists in the outbreeding cannabis species and to design resources that are of more value broadly. The need to establish a database on variance in the cannabinoid biosynthesis genes from a large pan-genome is imperative to accurately design genome editing constructs to ensure minimal off-targeting within the highly homologous gene sequences. This evaluation of gene variance will be performed using genome wide sequencing and the subset of relevant sequence data will be analysed through sequence alignment and SNP variant calling. This approach will allow for genome editing approaches, such as CRISPR/Cas-9, Zinc Finger Nucleases, TALENs and RNAi to be implemented through intelligent design of genome editing constructs to create novel chemotypic profiles for medicinal purposes.

Limited tissue culture protocols exist for hemp, and fewer reports for drug-type cannabis' response in tissue culture exists, with only protocols for specific cultivars being developed. Protocols in this PhD will be developed for a range of cultivars including micropropagation, callus induction, regeneration, protoplast isolation and transformation. These protocols will be invaluable for developmental studies, viral elimination and genome editing attempts to improve the cannabis germplasm.

This PhD will implement these two components to target cannabinoid biosynthesis genes, allowing for the development of tailored cannabis cultivars with novel cannabinoid profiles. The extensive genome editing catalogue targeting all genes, specifically cannabinoid biosynthesis genes, for use in different genome editing approaches where sequence variation knowledge is vital, the development of RNA interference vectors targeting the cannabinoid biosynthesis genes, along with the developed tissue culture protocols will provide indispensable knowledge and understanding in these areas of cannabis research, which are currently lacking verified robust protocols and genetic resources. Using these methods, the creation of novel chemovars for medical conditions, such as epilepsy, can be achieved and will allow further expansion into other areas of medical research.

1.13 Aims

- Chapter 2: Analyse a large pan-genome, consisting of 660 cannabis genomes, to develop detailed locations of SNPs and CNVs in all related enzymatic pathways leading to cannabinoid production for use in genome editing construct design. Design a comprehensive catalogue of single-guide RNA (sgRNA) for CRISPR/Cas-9 targeting unique PAM sites across gene sequences that are universal, by targeting conserved regions, and specific, by targeting regions where consensus sequence variation exists
- Chapter 3: Develop a robust protocol for the isolation and transfection of cannabis protoplasts by exploring the variables affecting yield, viability, and transfection efficiency through heterologous expression of *GFP*
- Chapter 4: Design and construct RNA interference vectors correlating to Cannbio-2 cannabinoid biosynthesis genes to modulate cannabinoid concentration. Assess multiple RNAi constructs silencing efficacy on cannabinoid biosynthesis gene transcription, using Quantitative Real-Time PCR, through transient leaf disc agroinfiltration
- Chapter 5: Develop and evaluate a robust callus induction and regeneration protocol through variable manipulation to successfully generate a stably transformed cannabis plant with a highly efficient RNAi construct previously developed targeting cannabinoid biosynthesis genes

CHAPTER 2

In silico analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation

2.1 Chapter preface

To date, no genome editing attempts targeting the medicinally important cannabinoid biosynthesis genes have been attempted. The sequence information of these highly homologous genes varies greatly between genotypes and in-depth knowledge of sequence variance is vital for genome editing attempts. This chapter details the analysis of publicly available genomes and a large pan-genome to identify and characterise SNPs and CNVs that exist in the genes collaborating in cannabinoid biosynthesis. The development of a consensus sequence from the pan-genome provides in-depth knowledge on SNP location allowing for the development of a sgRNA catalogue targeting unique, and universal, PAM sites. From this information effective sgRNA can be assessed through multiple online tools evaluating sgRNA efficacy and used to improve the cannabis germplasm by modifying cannabinoid accumulation to create novel chemotypic cultivars.

2.2 Publication details

Title: *In silico* analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation

Journal: PLOS ONE Journal. DOI: 10.1371/journal.pone.0257413

Stage of publication: Published

Authors: Lennon Matchett-Oates, Shivraj Braich, German Spangenberg, Simone Rochfort, Noel Cogan

2.3 Statement of contribution of joint authorship

LMO analysed consensus sequence to identify CNV, characterised SNPs and designed sgRNA catalogue for all genes. SB generated consensus sequence containing SNP locations. LMO drafted the entirety of the manuscript with all accompanying figures. GS, SR and NC helped conceive project direction and assisted in editing the manuscript. All authors approved the final manuscript.

2.4 Co-author statement confirming authorship of PhD candidate

As co-author of the manuscript 'Matchett-Oates L, Braich S, Spangenberg GC, Rochfort S, Cogan NOI. In silico analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation. Plos one. 2021 ;16(9) :e0257413.', I confirm that Lennon Matchett-Oates has made the following contributions,

- Identification of all gene sequences from the pathway
- Assistance in the identification and characterisation of all associated variants
- Design of all genome edit constructs
- Generation of all figures and tables
- Writing the manuscript, critical appraisal of the content and response to reviewers

Dr Noel O.I. Cogan

Date: 20/12/2021

PLOS ONE

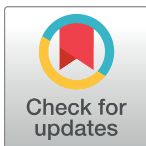
RESEARCH ARTICLE

In silico analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation

L. Matchett-Oates[1,2]*, **S. Braich**[1,2], **G. C. Spangenberg**[1,2], **S. Rochfort**[1,2], **N. O. I. Cogan**[1,2]

1 Agriculture Victoria, AgriBio, The Centre for AgriBioscience, Bundoora, Victoria, Australia, **2** School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia

* lennon.matchett-oates@agriculture.vic.gov.au



OPEN ACCESS

Citation: Matchett-Oates L, Braich S, Spangenberg GC, Rochfort S, Cogan NOI (2021) *In silico* analysis enabling informed design for genome editing in medicinal cannabis; gene families and variant characterisation. PLoS ONE 16(9): e0257413. https://doi.org/10.1371/journal.pone.0257413

Editor: Shawky M. Aboelhadid, Beni Suef University Faculty of Veterinary Medicine, EGYPT

Received: April 7, 2021

Accepted: August 31, 2021

Published: September 22, 2021

Copyright: © 2021 Matchett-Oates et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its files.

Funding: Funding for research was provided by Agriculture Victoria and Agriculture Victoria Services.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

Cannabis has been used worldwide for centuries for industrial, recreational and medicinal use, however, to date no successful attempts at editing genes involved in cannabinoid biosynthesis have been reported. This study proposes and develops an *in silico* best practices approach for the design and implementation of genome editing technologies in cannabis to target all genes involved in cannabinoid biosynthesis.

Results

A large dataset of reference genomes was accessed and mined to determine copy number variation and associated SNP variants for optimum target edit sites for genotype independent editing. Copy number variance and highly polymorphic gene sequences exist in the genome making genome editing using CRISPR, Zinc Fingers and TALENs technically difficult. Evaluation of allele or additional gene copies was determined through nucleotide and amino acid alignments with comparative sequence analysis performed. From determined gene copy number and presence of SNPs, multiple online CRISPR design tools were used to design sgRNA targeting every gene, accompanying allele and homologs throughout all involved pathways to create knockouts for further investigation. Universal sgRNA were designed for highly homologous sequences using MultiTargeter and visualised using Sequencher, creating unique sgRNA avoiding SNP and shared nucleotide locations targeting optimal edit sites.

Conclusions

Using this framework, the approach has wider applications to all plant species regardless of ploidy number or highly homologous gene sequences.

Significance statement

Using this framework, a best-practice approach to genome editing is possible in all plant species, including cannabis, delivering a comprehensive *in silico* evaluation of the

cannabinoid pathway diversity from a large set of whole genome sequences. Identification of SNP variants across all genes could improve genome editing potentially leading to novel applications across multiple disciplines, including agriculture and medicine.

Introduction

Cannabis sativa L. belongs to the *Cannabaceae* family and is one of the earliest domesticated plant species with archaeological evidence of cultivation beginning in China as early as 5000 B.C [1]. Cannabis has since been used throughout the world for its fibre in textiles, protein-rich seeds and therapeutic properties. The medicinal benefits of cannabis have been explored by many cultures around the world for centuries, with different preparations used to treat pain, inflammation and to improve appetite [2]. Today, cannabis is classed as an illicit drug in many countries, however, the consumption of cannabis for its psychoactive properties is estimated to be in excess of 190 million users worldwide [3].

Cannabis is an annual, wind pollinated herb, mainly dioceous but monoecious plants do exist. The number of species in the genus *Cannabis* is currently debated with reports suggesting a polytypic genus [4, 5] or as a monotypic, highly polymorphic species [6, 7]. The classification of cannabis has recently been suggested to follow its cannabinoid and terpene profile [8], however, three species of cannabis are generally accepted: *Cannabis sativa*, *Cannabis indica* and *Cannabis ruderalis* [7].

Cannabis contains a group of unique pharmacologically active chemical compounds called cannabinoids primarily produced in the glandular trichomes on female flowers. Phytocannabinoids represent a diverse group of C_{21} terpenophenolic compounds with a total of 120 cannabinoids currently reported [9].

The mammalian endocannabinoid system is comprised of endogenous cannabinoid receptors and metabolic enzymes that play a crucial role in homeostasis. The therapeutic potential for medicinal cannabis to aid in regulating physiological, immunological and behavioural conditions is of great interest. Reported *in vivo* effects in human and animal models indicate therapeutic applications in conditions such as multiple sclerosis [10], cancer [11], pain management [12] and epilepsy [13]. The highly polymorphic nature of cannabis is currently a limiting factor in reliable dosing quantities of cannabinoids, creating uncertainty in product efficiency.

Δ^9-Tetrahydrocannabinol (THC), responsible for the psychoactive properties in cannabis, and Cannabidiol (CBD), non-psychoactive with diverse pharmacological properties, are the most abundant cannabinoids found in cannabis with their therapeutic properties being extensively reviewed [14]. Phytocannabinoids are synthesised in their acidic forms and undergo decarboxylation into their active neutral forms with heat or time [15, 16]. Due to the large variation of cannabis strains containing different levels of chemical variants, cannabinoid fractions are referred to as chemotypes. Initially, chemotypes of cannabis were classed as "drug-types" and "fibre-types" [17] representing THC+CBN/CBD quotient >1 or <1 respectively. It was later agreed that a plants chemotype was broken down into three major and two minor chemotypes as the current model [18]. Biosynthesis of the major cannabinoids, THC and CBD, from the common precursor cannabigerol (CBG) is performed by tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS) [19, 20]. de Meijer *et. al* (2003) proposed the genetic determination for chemotypes as two alleles at a single gene locus, termed the B locus. The B_T allele encodes THCAS, and with the B_D allele encoding CBDAS. Those

with high THC and low CBD have B_T/B_T and B_D/B_D genotypes respectively and contain high levels of CBD with little to no THC, and B_T/B_D genotypes similar concentrations of THC and CBD. More recently, Grassa *et. al* (2018) completed the chromosome genome sequence assembly of cannabis finding that cannabinoid biosynthesis genes are not located at a single locus but are pericentromeric, nested in repeats leading to low levels of recombination.

Biosynthesis of cannabinoids is complex with numerous enzymatic steps and interactions. Fatty acids and isoprenoid precursors are synthesised via the hexanoate, methylerythritol 4-phosphate (MEP) and geranyl diphosphate (GPP) pathways. Hexanoyl-CoA is produced via the hexanoate pathway, acting as the substrate for olivetolic acid synthase (OLS) yielding OLA [21]. Prenyl sidechains are synthesised via the MEP pathway for the substrate for geranyl diphosphate synthesis. GPP and OLA are added by an aromatic prenyltransferase (PT) creating CBGA [22]. Finally, catalysation of THC and CBD oxidocyclases produce THCA and CBDA [23, 24] (Fig 1). Identification of all genes encoding biosynthetic enzymes now allows biotechnological approaches to control cannabinoid content by allowing genomically informed decisions on molecular breeding with tools such as genome editing.

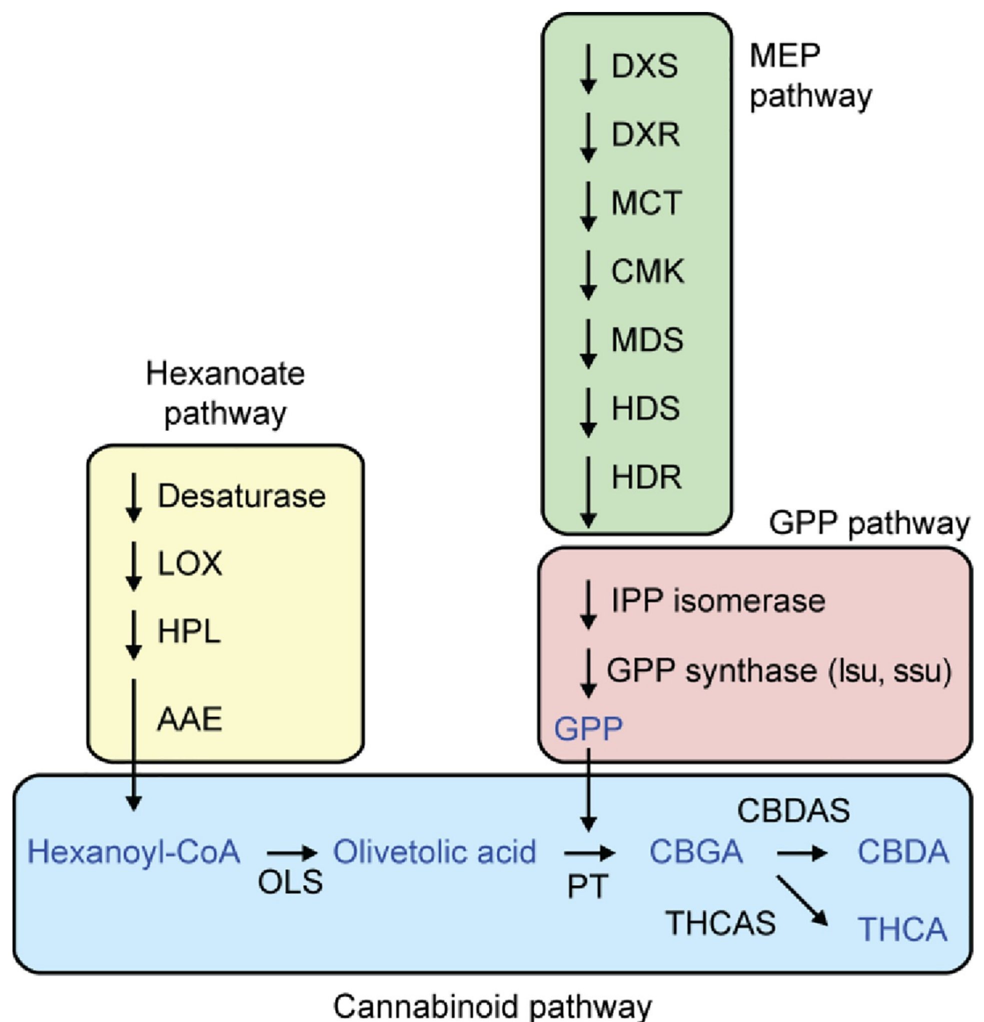


Fig 1. Overview of the cannabinoid biosynthesis pathways. Modified from van Bakel *et al.* (2011) [24].

https://doi.org/10.1371/journal.pone.0257413.g001

The development of genome editing technologies, such as Clustered Regularly Interspaced Palindromic Repeats (CRISPR/Cas9), Zinc Finger Nucleases (ZFNs) and Transcription activator-like effector nucleases (TALENs) utilise sequence specific nucleases to induce a double strand break (DSB) at a specific genomic location through homologous binding of guide proteins [25]. Plants' predominant repair pathway mechanism is through non-homologous end joining (NHEJ), and less often through homologous recombination (HR) [26]. NHEJ repairs the cut DNA without a homologous DNA template, however NHEJ can be error-prone, causing mutations such as base pair deletions, insertions or rearrangements [26, 27]. HR requires the provision of a DNA template, with homologous flanking regions used as a guide, to repair the break either correctly or by incorporating alterations that are desired into the DNA break point [28]. The use of genome editing techniques to manipulate gene function in a range of plant species has allowed for the generation of improved crop varieties, improved resistance and increased yield [29–31]. Within the CRISPR/Cas 9 system, the single-guide RNA (sgRNA), a 20nt oligo complementary to the gene of interest, guides the Cas9 endonuclease to the protospacer-adjacent motif (PAM) site, where Cas9 binds and cleaves the DNA strand [32]. Online tools available for sgRNA design and plasmid construction have been extensively reviewed [33] with CRISPR/Cas9 being broadly implemented in plants such as Arabidopsis, tobacco, rice and sorghum [34, 35]. ZFNs contain a tandem array of Cys2-Hys2 finger domains linked to the *Fok*I catalytic domain, with the finger domains each recognising 3bp of DNA [36]. The finger arrays are fused to the catalytic domain of *FokI* functioning as a dimer. Binding of the zinc-fingers to the target loci brings the two *FokI* monomers into close proximity causing them to dimerise, creating a DSB [37]. Similar in the mode of action to ZFNs, TALENs are comprised of a nonspecific *Fok*I nuclease domain fused to a DNA binding domain containing highly conserved repeats from the transcription activator-like effectors (TALEs) secreted by *Xanthomonas* spp. [38].

Off-target mutations caused by inefficient guide design and *FokI* monomer dimerisation could disrupt the functions of unintended genes, causing genetic instability and unintended cytotoxic effects. Single nucleotide polymorphisms (SNP) in genomic DNA across large, diverse populations will disrupt the homology-based binding of sgRNA, ZFs and TALEs with CRISPR/Cas9, ZFNs and TALENs. Target specificity is tightly controlled by sequence homology, with an increasing number of mismatches, off-target cleavage also increases [39]. Avoiding off-target effects is critically important for effective and efficient genome editing, with the need for genomically informed designs based on thorough deep-read genome sequencing being more important than ever. If these tools are to be regulated and used in product design, absolute confidence in design based on homology is needed.

In this study we outline the best practice workflow for identifying target sequences and their corresponding design using sgRNA in cannabis for the manipulation of the entire pathway of THC and CBD synthesis. Through genomically informed decisions based on previously published cannabis pangenome, generic and specific sgRNA can be designed using online tools to successfully target genes of interest with no *in silico* detected off-targets. The workflow here can help make informed decisions on gene targeting in cannabis, leading to novel cannabinoid production by targeting cannabis biosynthesis genes, accelerating the understanding of the relationships of genes in cannabinoid production.

Materials and methods

Genome sequence assembly

Genome sequence assembly of the Cannbio-2 genotype were performed by Braich *et al.* [40] (https://doi.org/10.46471/gigabyte.10) with a brief summary given here. RaGOO [41] was used to scaffold the draft genome of Cannbio-2 to chromosome scale pseudomolecules with CBDRx

genome assembly as the reference. Access to CBDrx genome (known as cs10 in NCBI) is available through The European Nucleotide Archive (PRJEB29284) (https://www.ebi.ac.uk/ena/data/view/PRJEB29284). PK and Finola genome assemblies were accessed through the NCBI BioProject database (PRJNA73819) (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA73819).

Cannbio-2 and pangenome gene analysis

Cannabinoid biosynthesis genes were accessed from a variety of sources and public databases (Table 1) to annotate Cannbio-2. Sequences were downloaded and used as a query for BLAST analysis against the Cannbio-2 genome assembly with an e-value threshold set at $<10^{-10}$. Identified regions of interest from the reference genome were annotated using NCBI nBLAST to confirm sequence identity and MEGANTE [42] and coding sequences (cds) visualised using FGENESH [43]. Sequences are available in S1 Table in S1 Data.

Publicly available cannabis genomes were downloaded (as described above) and were BLAST analysed using Cannbio-2 gene sequences described here with an e-value threshold set at $<10^{-10}$ to determine copy numbers within each respective genome (Table 1).

SNP discovery

SNP discovery was performed by Braich *et al.* [44], with a brief summary given here. Genomic DNA was extracted from fresh leaf material from a range of 660 mixed cultivars (high CBD, high THC, balanced THC:CBD, male and female plants) using DNeasy 96 Plant Kit (QIAGEN,

Table 1. Source of gene query/NCBI accession number and gene copy and homolog number for available genomes discovered using BLAST.

Gene		Pathway	Cannbio-2	CBDrx	Finola	PK V2
	NCBI Accession Number/Source of Query		Copy number/homologs			
DXS1	KY014576.1	MEP	1	1	-	1
DXS2	KY014577.1	MEP	1	1	-	1
DXR	KY014568	MEP	1	1	1	2
MCT	KY014578	MEP	1	1	1	1
CMK	KY014575	MEP	1	1	1	1
MDS	HQ734721.1	MEP	1	1	1	1
HDS	KY014570.1	MEP	1	1	1	1
HDR	KY014579.1	MEP	1	1	1	1
IPP/IPI	KY014569.1	GPP	1	-	1	1
GPP LSU	KY014573.1	GPP	1	1	1	1
GPP SSU	KY014567.1	GPP	1	1	-	1
FAD2	PK genome, scaffold71447:2,827–3,852	Hexanoate	4	5	7	3
LOX	PK genome, scaffold53609:3,286–7,284	Hexanoate	1	1	1	1
HPL	PK genome, scaffold14797:30,184–30,623	Hexanoate	1	1	1	1
AAE1	JN717233	Hexanoate	1	1	-	1
OLS	EU551162.1	Cannabinoid	1	1	1	2
OAC	JN679224.1	Cannabinoid	2	1	1	2
GOT	Publication number: US20120144523A1	Cannabinoid	1	1	1	1
CBDAS	AB292682	Cannabinoid	9[1]	11 total	9 total	14 total
THCAS	AB057805	Cannabinoid	1			
CBCAS	Publication number: WO/2015/196275	Cannabinoid	3[2]			

[1]2 genes and 7 homologs.

[2]2 genes and 1 homolog.

https://doi.org/10.1371/journal.pone.0257413.t001

Hilden, Germany) according to the manufacturer's instructions. Each library was prepared using enzymatic shearing using MspJI (NEB, MA, USA) in-house library prep protocol and sequenced on a HiSeq3000 instrument (Illumina Inc. San Diego, CA, USA)The resulting sequence data was reference aligned to the Cannbio-2 genome assembly previously described, using the BWA MEM algorithm [45]. Variants were identified using SAMtools [46] and a bed file with scaffold regions of interest matching to gene sequences of cannabinoid biosynthesis genes was created. Alignments were sorted and used for variant calling with an adjusted mapping quality (-C 50) and minimum read depth of 5 generating a consensus sequence. Consensus sequences for CDS sequences of genes of interest are available in S2 Table in S1 Data.

Determination between allele or gene

Presence of an allele, or extra copies of a gene, were determined based on genomic nucleotide multiple sequence alignments using MUSCLE [47]. Sequences of similar length with alignment similarity between 80–98%, which produced identical translated proteins were determined as alleles. Where large variation existed between genomic nucleotide sequence length or content, or where nucleotide sequences were <1000bp, predicted mRNA sequences were used from FGENESH [44] for alignment. Alleles were determined if similarity equalled >98%. Additional gene copies were determined if greater than two haplotypes were found with similarities >90% but <98%, due to cannabis being an outbreeding species and the Cannbio-2 genome sequence assembly is based off a heterozygous plant.

sgRNA design and confirmation

CHOPCHOP [48], CRISPR MultiTargeter [49], Crispor [50] and ZiFit [51] were used for the selection of sgRNAs for use with CRISPR-Cas9. Entire CDS region, calculated by FGENESH [43] and MEGANTE [42], were used as search queries. sgRNA on and off-target parameters suggested by each online tool was used. For visual confirmation of SNP avoidance, sgRNAs were manually aligned to Cannbio-2 and consensus sequences using Sequencher [52]. sgRNA designs are available in S3 Table in S1 Data.

Results

Genome mining for cannabinoid biosynthesis genes

To locate all the genes involved in cannabinoid biosynthesis, query references were downloaded from publicly available databases (Table 1) and BLAST analyses was performed against the Cannbio-2 genome assembly.

All genes in the MEP, GPP, Hexanoate and Cannabinoid pathway were identified (Table 1). Two 1-deoxy-D-xylulose 6-phosphate synthase (DXS) genes were discovered in the MEP pathway alongside single copies of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS) and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (HDR). Single genes of isopentenyl diphosphate isomerase (IPP/IPI), geranyl pyrophosphate synthase (GPP), small and large subunits, were identified in the GPP pathway. In the hexanoate pathway, four copies of fatty-acid desaturase (FAD2) were identified using the Purple Kush (PK) desaturase gene sequence as the query. Translated proteins from all FAD2 homologs were tBLASTn analysed for confirmation of correct annotation and all are believed to be involved in cannabinoid biosynthesis. Lipoxygenase (LOX) and hydroperoxide lyase (HPL) were identified using the associated PK gene

sequences as the queries with very low (<1%) sequence variation existing. Acyl-activating enzyme (AAE1) was found using previously published sequences (Table 1) amongst the AAE superfamily, containing 15 AAE homologs. Translated AAE1 annotation was confirmed using tBLASTn and isolated from the large superfamily of highly homologous gene sequences. In the cannabinoid pathway a single copy of olivetol synthase (OLS) was discovered with >98% identity to deposited OLS sequences in NCBI. Two copies of olivetolic acid cyclase (OAC) were discovered. The CDS of the set of alleles and a single copy of OAC were aligned and 14 SNPs exist between the set. All OAC sequences were correctly annotated using MEGANTE and tBLASTn to confirm copy number. Two complete identical, functional CBDAS-like genes were discovered (CBDAS-like#1 and #2) with three closely related homologs also existing (CBDAS-like#3–5). CBDAS-like homologs contain several SNPs causing sequence variation in translated protein sequences. Four truncated CBDAS homologs were also discovered (CBDAS-truncated#1–4), with each containing stop codons resulting in truncated protein sequences. Two complete copies of cannabichromenic acid synthase (CBCAS) were found (CBCAS#1 +#2) with identical sequences except at base pair 662 with a SNP of C to T, though identical proteins are predicted. One closely related truncated homolog of CBCAS was also discovered (CBCAS-truncated) producing a substantially shorter predicted protein sequence. One single copy of THCAS was also discovered.

Pan-genome copy number variance comparison

Within the publicly available cannabis genome sequences, the assembled gene set was then used to query gene copy number and identify potential homologs. Differences exist between the datasets in terms of gene copy number due to the resolution of the sequence data, genetic mapping, scaffolding technologies and natural variation in different genomes. Variations in gene presence and copy number, using the assembled reference gene list, exist for DXS1, DXS2, DXR, IPP/IPI, GPP_SSU, FAD2, AAE1, OLS, OAC, CBDAS, THCAS and CBCAS (Table 1). Within the Finola genome, DXS1, DXS2, GPP_SSU and AAE1 were not discovered, with copy number variation existing for FAD2, OLS and OAC when compared to Cannbio-2 (Table 1). Within the CBDrx genome, no copy of IPP/IPI was discovered, which is confirmed by the most recent release of the CBDRx genome. Copy number variations exist for FAD2 compared to Cannbio-2, with 4 FAD2 genes being discovered in Cb-2 and 5 in CBDRx. The updated PK genome had at least one copy of each gene, with variations in copy number existing for DXR, FAD2, OLS and OAC and synthase genes compared to Cannbio-2.

Analysis of SNPs and informed sgRNA design

To assess gene variation, the six hundred and sixty whole genomes that were sequenced were used to establish a resource of SNP locations (consensus sequence) (S3 Table in S1 Data), which were then overlayed onto the identified genes integral to the cannabinoid biosynthesis. With the exception of FAD2, which belongs to a large, diverse family of desaturases, the cannabinoid biosynthesis genes are highly conserved with little variation within their sequences (Table 2). Each consensus sequence containing SNP locations was then used for intelligent guide designs to avoid all known nucleotide variations, creating universal sgRNA which can be broadly used on any cannabis genotype, and in the instance of highly similar gene sequences, unique sgRNA designed to target only a specific gene of interest (Fig 2). Sequences from the reference genome were entered into the online design tools CHOPCHOP, CRISPR MultiTargeter, Crispor and ZiFit to generate sgRNA based on their preferred scoring matrixes followed by manual and visual comparison. Taking the highest-ranking scores from each online tool, which predict off-targeting potential and greatest binding affinity, each sgRNA was visualised, using Sequencer, to

Table 2. Gene length defined as genome base-pair length including introns, and location within Cannbio-2 with accompanying consensus SNP data from pangenome.

Gene	Gene Length	CB-2 Genome Location	# SNPs in Pangenome
DXS1	3601	Chr:9 14103141–14103512	6
DXS2	2892	Chr:4 79846660–79849546	71
DXR	3689	Chr:3 10253558–10257868	68
MCT	4242	Chr:4 36521598–36525845	155
CMK	4031	Chr:2 12810228–12814256	103
MDS	1946	Chr:5 86405983–86407926	70
HDS	5383	Chr:2 100426265–100431627	211
HDR	2309	Chr:X 7955658–7957964	76
IPP/IPI	2921	Chr:2 13601913–13604831	50
GPP_LSU	1281	Chr:4 91310699–91311977	31
GPP_SSU	1061	Chr:6 55780334–55781392	19
FAD2#1	1123	Chr:2 104383871–104384992	57
FAD2#2	1085	Chr:2 104394699–104395781	52
FAD2#3	1091	Chr:2 104401152–104402226	53
FAD2#4	1084	Chr:2 104420931–104422048	25
LOX	4162	Chr:2 102127730–102131887	133
HPL	7201	Chr:8 53062338–53070863	200
AAE1	6688	Chr:3 50354410–50361096	220
OLS	1418	Chr:8 61667472–61668887	35
OAC	692[1]	Chr:9 5793422–5794110	17
OAC#2	548[1]	Chr:9 6925677–6926172	15
GOT	7350	Chr:X 65676960–65684340	264
THCAS	1868	Chr:7 29533343–29535211	37
CBCAS#1	1635	Chr:7 29465648–29467283	2
CBCAS#2	1635	Chr:7 29577848–29579483	2
CBCAS-truncated	1420	Chr:7 29518627–29519784	5
CBDAS-like#1	1900	Chr:7 33131612–33133245	3
CBDAS-like#2	1628	Chr:7 33199940–33201573	0
CBDAS-like#3	1700	Chr:7 33234459–33236068	12
CBDAS-like#4	1500	Chr:7 33275773–33277406	24
CBDAS-like#5	1704	Chr:7 33371944–33373577	13
CBDAS-truncated#1	449	Chr:7 33122514–33123491	9
CBDAS-truncated#2	1839	Chr:7 33341204–33342916	14
CBDAS-truncated#3	990	Chr:7 34564480–34566132	46
CBDAS-truncated#4	1113	Chr:7 34569433–34570813	40

[1]complete CDS only.

https://doi.org/10.1371/journal.pone.0257413.t002

identify regions the sgRNA would target, whether that be in regions of sequence homology across the pan-genome or in regions consisting of SNPs. A total of 145 sgRNAs were designed targeting every gene in the combined pathways (S1 Table in S1 Data). The sgRNA generated consists largely of a pool of universal sequences, which regardless of cultivar used, can target each gene in the combined pathways through the use of the consensus sequence generated. Multiple Cannbio-2 specific sgRNA were also designed in regions where sequence heterogeneity towards the 5' translated regions dictated universal sgRNA design was not possible. All sgRNA were re-BLAST analysed against the reference genome for detection of off-site targeting, with results confirming no complete 20-nt sgRNA had potential off targets outside their respective gene sets.

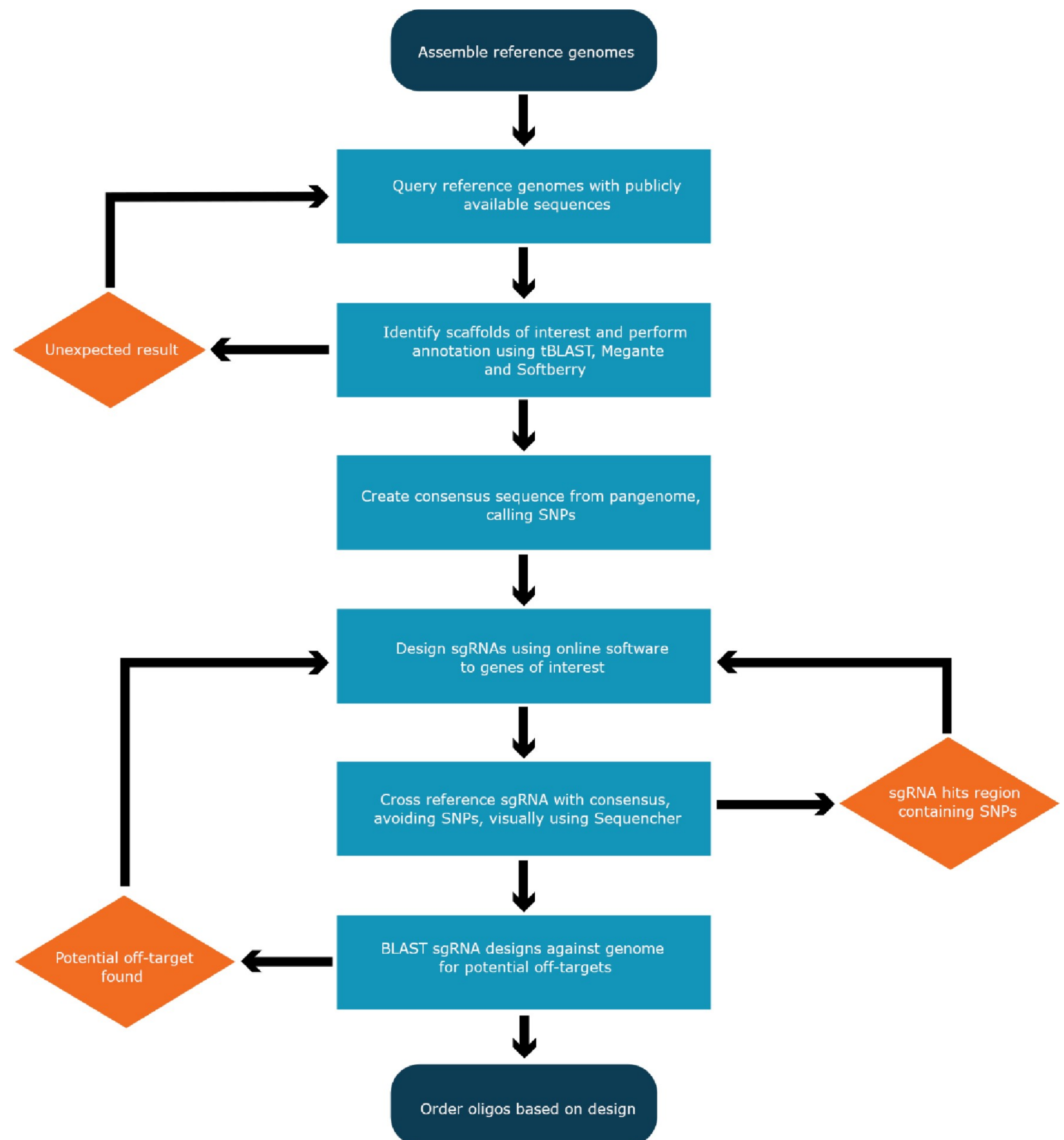


Fig 2. Informed genome editing pipeline for intelligent design of sgRNA.

https://doi.org/10.1371/journal.pone.0257413.g002

Discussion

Phytocannabinoids are of particular interest for their pharmacological applications in a growing number of medical conditions. Knowledge and understanding of the gene interactions and their relationship to final cannabinoid concentration can facilitate improved cannabis strains with desired novel cannabinoid levels. Creating a pangenome consensus of each gene in the contributing pathways allows for genomically informed decisions, based on known SNP location and frequency as well as presence absence variations (PAV), for crop improvement by means of genome editing. Using publicly available sequence information, at least one full length transcript for all genes involved in cannabinoid biosynthesis were found agreeing with

previous genome sequencing and genome mining reports [24, 53]. Gene copy number in the MEP pathway also agrees with previously published analysis [53]. Two DXS genes were discovered, with previous reports showing DXS1 having elevated expression levels in photosynthetic tissues, underlining its importance in isoprenoid production [54]. DXS2 accumulates in the roots with expression patterns suggesting synthesis of specific isoprenoids, however, it's role in cannabinoid biosynthesis is yet to be determined. Multiple genes for DXR [55], HDR [56] and IPI/IPP [57] have been previously reported, however, only singular copies of these genes were discovered in the Cannbio-2 genome. It is possible that multiple copies of these genes could be responsible for the accumulation of cannabinoid precursors, leading to novel cannabinoid levels. Fatty acid desaturase enzymes belong to two large multifunctional classes, either membrane bound, or soluble. The desaturase of interest in cannabinoid production, FAD2, is involved in the hexanoate pathway, leading to the production of hexanoyl-CoA, the first precursor in the cannabinoid pathway. Despite the complexity of the number of FAD2 gene sequences, it is believed that the correct version was identified, although our data shows four copies of this gene, where previous comparative studies discovered seven gene copies in the Finola genome [58] and only 2 copies in the CBDRx genome [59]. Further evidence of gene copy number variance, across published genomes, exists for OLS, and OAC posing the question if gene copy number directly influences chemovar determination. Previous studies have utilised short read sequence data in the identification of gene sequences and due to the anticipated degree of sequence similarity from the duplicate gene copies, taking a reference-aligning approach would be inaccurate to use the data generated to infer CNVs. However, with the availability of long read sequencing technology that can generate sequence data through extended repetitive regions, describing genome architecture and gene sequence and structure at a much higher level, makes it a reliable platform to use for the determination of CNVs. THC-rich PK cultivar has two copies of OLS and OAC, whereas CBD-rich cultivar, CBDrx, has just one copy of each from our BLAST search results, though 2 copies of OLS and no copies of OAC are reported. The presence of OAC is a polyketide synthase enzyme catalyses olivetolic acid, which forms the polyketide nucleus of cannabinoids [21]. This suggests that this particular polyketide was not included in the CBDRx genome, though it is considered essential for cannabinoid biosynthesis. The Cannbio-2 cultivar, with relatively equal (1.8:1) THC and CBD cannabinoid concentrations contains a single copy of OLS and 2 copies of OAC.

The exact relationship between gene copy number and cannabinoid production needs to be further studied through metabolic engineering in heterologous hosts or through genome editing. Using the discovered synthase genes from the Cannbio-2 genome sequence as the query against CBDrx, Finola and PK genomes, the total number of synthase genes varies considerably between the cultivars. In the CBDrx genome [59] 16 synthase genes are reported, however only 11 were discovered in CBDrx using sequences from Cannbio-2 as queries. Identification of which synthase genes were not identified is difficult due to the nested repeating nature of synthase genes around the centromere.

As long read sequencing is error prone, the correct assembly of *CBDAS* in the Cannbio-2 assembly has proven problematic, potentially exacerbated due to the hybrid nature of the genotype. It is therefore likely that the CBDAS gene has been incorrectly assembled and either a chimeric version of the functional and non-functional gene alleles, or that the non-functional allele only has been assembled, most likely as the gene that is referred to as CBDAS-truncated#3. The Cannbio-2 genome clearly has a functional CBDAS allele as a 100% identity sequence has been identified from the transcriptome data set [60] (Cannbio_016865).

Grassa *et.al* (2021) has identified the total number of potential synthase genes in reference to a sequence alignment to *THCAS* mRNA >82%. The variation in synthase genes is most likely due to PAV across different cultivars, which in the case of maize is common [61]. Total

synthase gene number for Finola and PK is not given in the original genome [62], however 9 and 14 genes were found when querying with Cannbio-2 sequences. Grassa *et.al* (2021) has identified 5 and 16 synthase genes within the PK and Finola from their respective approach to discovering copy numbers.

THCAS and CBDAS CNV have recently been reported from multiple cannabis cultivars with similar findings that this CNV partially explains variation in cannabinoid content [63, 64]. Multiple gene copies is a known method to increase production of secondary metabolites [65] which could lead to the understanding that increased copy number of synthase genes would in turn increase cannabinoid production. However, possibly a greater explanation of increased cannabinoid potency was discussed by Grassa *et al.* (2018) with the discovery that separate QTLs, not linked to synthase gene clusters, were responsible for up to 17% variation in cannabinoid quantity. This could possibly help explain the current gene copy number variation in the observed genes mentioned.

Complete absence of sequence data is present for specific genes in the CBDrx, Finola and PK genomes posing the question whether genome assembly, or actual PAV mechanisms are responsible. Within the Finola genome, 4 genes could not be identified. Both forms of DXS are not present and with previous studies demonstrating DXS knock down lines produce reduced levels of isoprenoids and contain more severe phenotypic characterisations [66, 67], suggesting the fragmented genome failed to identify and assemble the specific genes of interest. GPP SSU and AAE1 were also not identified, however, from previous reports both these genes are critical for isoprenoid and cannabinoid production indicating they are missed in the genome assembly. AAE1 was found to be the gene which synthesises hexanoyl-CoA from hexanoate supplying the cannabinoid pathway [68] and since Finola still produces cannabinoids, it is concluded that it was also an assembly error. GPP is a heterodimer requiring both subunits, large and small, for optimum activity. GPP activity has shown to still be active but at lower levels when the small subunit was inactive [69], however both subunits were still present, suggesting the absence of GPP SSU in the Finola genome is also due to assembly error. The absence of IPP/IPI in the CBDrx genome is also strongly suggested to be due to assembly error, since previous studies on *Arabidopsis* double mutant knockdown of IPP/IPI produced dwarfism and male sterility [70].

The SNP location resource revealed some genes are more highly conserved than others. The variable conservative nature of genes was observed indicating a continuing evolution of recombination and divergence. Comparative analysis of SNPs present in genes of variable copy number in Cannbio-2, CBDrx, Finola and PK genomes was performed (excluding results of no gene presence). Through multiple sequence alignments of coding sequences, it was observed that the presence of SNP's occurred in the extra gene copy where the presence of homozygous alleles exists. This suggests that either sequencing error has occurred, or in fact there is an extra copy of the gene and a set of alleles. Within the Cannbio-2 genome, OAC produced three sequence similarity matches with two sequences determined as alleles with an extra copy of the gene existing as a truncated version of the gene. When gene sequences were aligned, SNPs occurred in all genes and when translated, nearly identical protein sequences (>99%) were produced confirming that an extra copy of the gene was present, potentially in a hemizygous condition. Within the PK genome, copy number variation exists for OLS and OAC. In a similar way to OAC in the Cannbio-2 genome, OLS produced three hits, two of which were determined to be alleles and one to be an extra copy. SNPs existed in all three sequences when coding regions were aligned with similar results obtained from protein sequence alignment. Initial alignment of both OAC hits, in PK, found a 98.5% similarity in genomic sequences, however no gene prediction was possible on one of the sequences, possibly due to a premature stop codon from a SNP rendering this gene inactive potentially indicating that it exists as a pseudogene.

How this copy number variation contributes to differential cannabinoid production is yet to be fully elucidated, however using the known SNP location for each extra copy gene in Cannbio-2, sgRNA could be designed to help understand this relationship. Using multiple online tools for the design of sgRNA ensured that all possible guide designs could be assessed for *in silico* off-targeting. Each tool implements different scoring rules based on off-targets, mismatches, efficiency score, existence of self-complimentary regions, GC content, location of guide and multiple sequence alignments [48, 49]. Due to the diversity in gene content and sequence variation and the absence of a well characterised pan-genome for cannabis, analysis by these multiple tools was necessary and essential. The presence of a PAM site is necessary for sgRNA binding and even though these tools scanned the gene sequence for the PAM sites, results occasionally varied between the online tools. Visualisation of sgRNAs was clear using CHOPCHOP compared to the other tools and regularly provided the best guide designs. However, when highly homologous sequences were used MultiTargeter was able to perform sequence alignments and produce unique sgRNA for each sequence, a feature not possible within the other tools. Designing the sgRNA for the unique synthases were first run using MultiTargeter and further verified using CHOPCHOP for visualisation. sgRNA designed were targeted to the earliest possible exon for maximum likelihood of a frame shift mutation. The error prone nature of NHEJ often occurs with small deletions, or insertions, occurring at the DSB leading to protein misfolding and thus production of a knockout gene. Each identified gene, with accompanying allele where applicable, were analysed and sgRNAs were designed to be either universal, inactivating both related genes, or if sequence heterozygosity exists, specific sgRNA were designed (S1 Table in S1 Data). Mutational studies identifying differential expression in isoprenoid biosynthesis genes, including DXS [67], DXR [71], IPP/IPI [70] and MDS [72] have previously been reported. Mutational studies on the unique synthase genes are yet to be reported, potentially due to the high homology between enzymes. Using genome editing, sequence homogeneity between synthase genes could potentially lead to off-target editing, with targets suggested to have at least several nucleotides different for discrimination [73]. Where possible, each synthase gene, and accompanying homologs, had universal and specific sgRNA designed that could be used regardless of cultivar, strain or population chosen as the target. The reported sequence similarity between THCAS, CBDAS and CBCAS, up to 95% [62], requires precise, intelligent design, using multiple online tools and a large consensus population to improve the likelihood of correct gene knock down. Potential off targeting predictions given by sgRNA online tools currently use the previously fragmented genome of PK [24]. To circumvent this, each sgRNA was used as a query to BLAST against the Cannbio-2 genome for potential off-targets. From the BLAST results no sgRNA had an unexpected sequence match elsewhere in the genome, however singular nucleotide mismatches do occur. How these mismatches are tolerated during directed genome editing is yet to be determined, however it is expected that off-targeting will be more prevalent with more highly homologous gene sets.

Applying this logical workflow *in silico* is the benchmark standard, essential to ensure that correct genes and associated SNPs are identified before genome editing can begin. This approach has wider applications in all genome editing efforts within species that have paleopolyploidy, large PAV gene populations or crop species with high levels of variations within the genome. This workflow explains each step taken and the tools to use to obtain universal or specific sgRNA to any gene of choice quickly and effectively, where each step can encounter issues and how to correct them making this approach critical for effective genome editing with minimal off-targeting. This same approach can easily be applied to the more recent CRISPR--Cas12a system which has been gaining popularity with editing plant genomes. The availability of fully sequenced genomes, pangenomes and the ability to accurately predict potential off-target effects and edits makes this method applicable to all plant gene editing applications

regardless of species. Only recently the ability to analyse the cannabis genome has become available showing that using this approach, with current technologies available, this method can be used quickly and effectively. Even with the limited literature and resources available for completed cannabis genomes, quick, intelligent design for genome editing in cannabis is now possible. Understanding the effect of gene copy number, PAV and SNP location and density on cannabinoid production can help create unique cannabinoid profiles for medicinal purposes.

Supporting information

S1 Data.
(XLSX)

Author Contributions

Conceptualization: L. Matchett-Oates.

Data curation: S. Braich.

Formal analysis: L. Matchett-Oates.

Supervision: G. C. Spangenberg, S. Rochfort, N. O. I. Cogan.

Writing – original draft: L. Matchett-Oates.

Writing – review & editing: L. Matchett-Oates.

References

1. Li H-L. An archaeological and historical account of cannabis in China. Econ Bot. 1973; 28(4):437–48.
2. Kalant H. Medicinal use of cannabis: history and current status. Pain Res Manag. 2001; 6(2):80–91. https://doi.org/10.1155/2001/469629 PMID: 11854770
3. Office on Drugs and Crime UN. World Drug Report 2017. Executive Summary—Conclusions and Policy Implications. 2017 [cited 2018 Jun 25]; Available from: https://www.unodc.org/wdr2017/field/Booklet_1_EXSUM.pdf
4. Emboden WA. Cannabis—a polytypic genus. Econ Bot. 1974; 28(3):304.
5. Hillig KW. Genetic evidence for speciation in Cannabis (Cannabaceae). Genet Resour Crop Evol. 2005; 52(2):161–80.
6. Doorenbos NJ, Fetterman PS, Quimby MW, Turner CE. Cultivation, extraction, and analysis of Cannabis sativa L. Ann N Y Acad Sci. 1971; 191(1):3–14.
7. Small E, Cronquist A. A practical and natural taxonomy for Cannabis. Taxon. 1976;405–35.
8. Piomelli D, Russo EB. The Cannabis sativa versus Cannabis indica debate: an interview with Ethan Russo, MD. Cannabis cannabinoid Res. 2016; 1(1):44–6. https://doi.org/10.1089/can.2015.29003.ebr PMID: 28861479
9. ElSohly MA, Radwan MM, Gul W, Chandra S, Galal A. Phytochemistry of Cannabis sativa L. In: Phytocannabinoids. Springer; 2017. p. 1–36.
10. Alexander SPH. Therapeutic potential of cannabis-related drugs. Prog Neuro-Psychopharmacology Biol Psychiatry. 2016; 64:157–66. https://doi.org/10.1016/j.pnpbp.2015.07.001 PMID: 26216862
11. Ramer R, Hinz B. Inhibition of cancer cell invasion by cannabinoids via increased expression of tissue inhibitor of matrix metalloproteinases-1. JNCI J Natl Cancer Inst. 2008; 100(1):59–69. https://doi.org/10.1093/jnci/djm268 PMID: 18159069
12. Perez J, Ribera MV. Managing neuropathic pain with Sativex®: a review of its pros and cons. Expert Opin Pharmacother. 2008; 9(7):1189–95. https://doi.org/10.1517/14656566.9.7.1189 PMID: 18422475
13. Devinsky O, Marsh E, Friedman D, Thiele E, Laux L, Sullivan J, et al. Cannabidiol in patients with treatment-resistant epilepsy: an open-label interventional trial. Lancet Neurol. 2016; 15(3):270–8. https://doi.org/10.1016/S1474-4422(15)00379-8 PMID: 26724101

14. Cascio MG, Pertwee RG, Marini P. The Pharmacology and Therapeutic Potential of Plant Cannabinoids. In: Cannabis sativa L-Botany and Biotechnology. Springer; 2017. p. 207–25.
15. Flores-Sanchez IJ, Verpoorte R. Secondary metabolism in cannabis. Phytochem Rev. 2008; 7(3):615–39. https://doi.org/10.1093/pcp/pcn150 PMID: 18854334
16. Trofin IG, Dabija G, Vaireanu D, Filipescu L. Long-term storage and cannabis oil stability. Rev Chim. 2012; 63(3):293–7.
17. Fetterman PS, Keith ES, Waller CW, Guerrero O, Doorenbos NJ, Quimby MW. Mississippi-grown cannabis sativa L.: Preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part. J Pharm Sci. 1971; 60(8):1246–9. https://doi.org/10.1002/jps.2600600832 PMID: 5127101
18. Pacifico D, Miselli F, Micheler M, Carboni A, Ranalli P, Mandolino G. Genetics and Marker-assisted Selection of the Chemotype in Cannabis sativa L. Mol Breed. 2006; 17(3):257–68.
19. Taura F, Morimoto S, Shoyama Y, Mechoulam R. First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis. J Am Chem Soc. 1995; 117(38):9766–7.
20. Taura F, Morimoto S, Shoyama Y. Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. J Biol Chem. 1996; 271(29):17411–6. https://doi.org/10.1074/jbc.271.29.17411 PMID: 8663284
21. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE. Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. Proc Natl Acad Sci. 2012; 109 (31):12811–6. https://doi.org/10.1073/pnas.1200330109 PMID: 22802619
22. Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. FEBS Lett. 2007; 581 (16):2929–34. https://doi.org/10.1016/j.febslet.2007.05.043 PMID: 17544411
23. Fellermeier M, Eisenreich W, Bacher A, Zenk MH. Biosynthesis of cannabinoids: Incorporation experiments with 13C-labeled glucoses. Eur J Biochem. 2001; 268(6):1596–604. https://doi.org/10.1046/j.1432-1033.2001.02030.x PMID: 11248677
24. Van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, et al. The draft genome and transcriptome of Cannabis sativa. Genome Biol. 2011; 12(10):R102. https://doi.org/10.1186/gb-2011-12-10-r102 PMID: 22014239
25. Voytas DF. Plant genome engineering with sequence-specific nucleases. Annu Rev Plant Biol. 2013;64. https://doi.org/10.1146/annurev-arplant-042811-105552 PMID: 23451779
26. Gorbunova V, Levy AA. Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res. 1997; 25(22):4650–7. https://doi.org/10.1093/nar/25.22.4650 PMID: 9358178
27. Van Gent DC, Van Der Burg M. Non-homologous end-joining, a sticky affair. Oncogene. 2007; 26 (56):7731. https://doi.org/10.1038/sj.onc.1210871 PMID: 18066085
28. Siebert R, Puchta H. Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. Plant Cell. 2002; 14(5):1121–31. https://doi.org/10.1105/tpc.001727 PMID: 12034901
29. Kumar S, AlAbed D, Worden A, Novak S, Wu H, Ausmus C, et al. A modular gene targeting system for sequential transgene stacking in plants. J Biotechnol. 2015; 207:12–20. https://doi.org/10.1016/j.jbiotec.2015.04.006 PMID: 25913173
30. Ainley WM, Sastry-Dent L, Welter ME, Murray MG, Zeitler B, Amora R, et al. Trait stacking via targeted genome editing. Plant Biotechnol J. 2013; 11(9):1126–34. https://doi.org/10.1111/pbi.12107 PMID: 23953646
31. Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, et al. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature. 2009; 459(7245):437. https://doi.org/10.1038/nature07992 PMID: 19404259
32. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science (80-). 2013; 339(6121):823–6. https://doi.org/10.1126/science.1232033 PMID: 23287722
33. Cui Y, Xu J, Cheng M, Liao X, Peng S. Review of CRISPR/Cas9 sgRNA Design Tools. Interdiscip Sci Comput Life Sci. 2018;1–11. https://doi.org/10.1007/s12539-018-0298-z PMID: 29644494
34. Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods. 2013; 9(1):39. https://doi.org/10.1186/1746-4811-9-39 PMID: 24112467

35. Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res. 2013; 41(20):e188–e188. https://doi.org/10.1093/nar/gkt780 PMID: 23999092

36. Wolfe SA, Nekludova L, Pabo CO. DNA recognition by Cys2His2 zinc finger proteins. Annu Rev Biophys Biomol Struct. 2000; 29(1):183–212. https://doi.org/10.1146/annurev.biophys.29.1.183 PMID: 10940247

37. Kim Y-G, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci. 1996; 93(3):1156–60. https://doi.org/10.1073/pnas.93.3.1156 PMID: 8577732

38. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol cell Biol. 2013; 14(1):49. https://doi.org/10.1038/nrm3486 PMID: 23169466

39. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013; 31(9):822. https://doi.org/10.1038/nbt.2623 PMID: 23792628

40. Braich S, Baillie RC, Spangenberg GC, Cogan NOI. A New and Improved Genome Sequence of Cannabis sativa. Gigabyte. 2020;1.

41. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019; 20(1):1–17. https://doi.org/10.1186/s13059-018-1612-0 PMID: 30606230

42. Numa H, Itoh T. MEGANTE: a web-based system for integrated plant genome annotation. Plant Cell Physiol. 2013; 55(1):e2–e2. https://doi.org/10.1093/pcp/pct157 PMID: 24253915

43. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 2006; 7(1):S10. https://doi.org/10.1186/gb-2006-7-s1-s10 PMID: 16925832

44. Braich S, Matchett-Oates L, Bailie B, Mohammaen E, Fulgueras K, Rochfort S, et al. Reference Genome and Whole Genome Resequencing in Medicinal Cannabis for Genomic Selection and Genome Editing, Enabling Precision Breeding. In San Diego; 2019.

45. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr arXiv13033997. 2013;

46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

48. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. 2016; 44(W1):W272–6. https://doi.org/10.1093/nar/gkw398 PMID: 27185894

49. Prykhozhij S V, Rajan V, Gaston D, Berman JN. CRISPR multitargeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. PLoS One. 2015; 10(3): e0119372. https://doi.org/10.1371/journal.pone.0119372 PMID: 25742428

50. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud J-B, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016; 17(1):148. https://doi.org/10.1186/s13059-016-1012-2 PMID: 27380939

51. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 2013; 31(3):227. https://doi.org/10.1038/nbt.2501 PMID: 23360964

52. Corporation GC. Sequencher verson 5.4.6 DNA sequence analysis software. Ann Arbor: Gene Codes Corporation; 2011.

53. Booth JK, Page JE, Bohlmann J. Terpene synthases from Cannabis sativa. PLoS One. 2017; 12(3): e0173911. https://doi.org/10.1371/journal.pone.0173911 PMID: 28355238

54. Walter MH, Hans J, Strack D. Two distantly related genes encoding 1-deoxy-d-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. Plant J. 2002; 31(3):243–54. https://doi.org/10.1046/j.1365-313x.2002.01352.x PMID: 12164805

55. Seetang-Nun Y, Sharkey TD, Suvachittanont W. Molecular cloning and characterization of two cDNAs encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase from Hevea brasiliensis. J Plant Physiol. 2008; 165(9):991–1002. https://doi.org/10.1016/j.jplph.2007.06.014 PMID: 17936410

56. Kim S-M, Kuzuyama T, Kobayashi A, Sando T, Chang Y-J, Kim S-U. 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms Ginkgo biloba and

Pinus taeda. Planta. 2008; 227(2):287–98. https://doi.org/10.1007/s00425-007-0616-x PMID: 17763867

57. Phillips MA, D'Auria JC, Gershenzon J, Pichersky E. The Arabidopsis thaliana type I isopentenyl diphosphate isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. Plant Cell. 2008; 20(3):677–96. https://doi.org/10.1105/tpc.107.053926 PMID: 18319397

58. Bielecka M, Kaminski F, Adams I, Poulson H, Sloan R, Li Y, et al. Targeted mutation of Δ12 and Δ15 desaturase genes in hemp produce major alterations in seed fatty acid composition including a high oleic hemp oil. Plant Biotechnol J. 2014; 12(5):613–23. https://doi.org/10.1111/pbi.12167 PMID: 24506492

59. Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Timothy Motley S, et al. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. New Phytol. 2021; https://doi.org/10.1111/nph.17243 PMID: 33521943

60. Braich S, Baillie RC, Jewell LS, Spangenberg GC, Cogan NOI. Generation of a comprehensive transcriptome Atlas and transcriptome Dynamics in Medicinal cannabis. Sci Rep. 2019; 9(1):1–12. https://doi.org/10.1038/s41598-018-37186-2 PMID: 30626917

61. Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. 2009; 5(11): e1000734. https://doi.org/10.1371/journal.pgen.1000734 PMID: 19956538

62. Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, et al. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Res. 2019; 29(1):146–56.

63. Vergara D, Huscher EL, Keepers KG, Givens RM, Cizek CG, Torres A, et al. Gene copy number is associated with phytochemistry in Cannabis sativa. AoB Plants. 2019; 11(6):plz074. https://doi.org/10.1093/aobpla/plz074 PMID: 32010439

64. McKernan KJ, Helbert Y, Kane LT, Ebling H, Zhang L, Liu B, et al. Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. bioRxiv. 2020;

65. Lavania UC. Genomic and ploidy manipulation for enhanced production of phyto-pharmaceuticals. Plant Genet Resour. 2005; 3(2):170–7.

66. Araki N, Kusumi K, Masamoto K, Niwa Y, Iba K. Temperature-sensitive Arabidopsis mutant defective in 1-deoxy-d-xylulose 5-phosphate synthase within the plastid non-mevalonate pathway of isoprenoid biosynthesis. Physiol Plant. 2000; 108(1):19–24.

67. Estévez JM, Cantero A, Reindl A, Reichler S, León P. 1-Deoxy-D-xylulose 5-phosphate, a limiting enzyme for plastidicIsoprenoid biosynthesis in plants. J Biol Chem. 2001; https://doi.org/10.1074/jbc.M100854200 PMID: 11264287

68. Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. Plant J. 2012; 71 (3):353–65. https://doi.org/10.1111/j.1365-313X.2012.04949.x PMID: 22353623

69. Wang G, Dixon RA. Heterodimeric geranyl (geranyl) diphosphate synthase from hop (Humulus lupulus) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci. 2009; 106(24):9914–9. https://doi.org/10.1073/pnas.0904069106 PMID: 19482937

70. Okada K, Kasahara H, Yamaguchi S, Kawaide H, Kamiya Y, Nojiri H, et al. Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in Arabidopsis. Plant Cell Physiol. 2008; 49(4):604–16. https://doi.org/10.1093/pcp/pcn032 PMID: 18303110

71. Xu C, Li H, Yang X, Gu C, Mu H, Yue Y, et al. Cloning and Expression Analysis of MEP Pathway Enzyme-encoding Genes in Osmanthus fragrans. Genes (Basel). 2016; 7(10):78. https://doi.org/10.3390/genes7100078 PMID: 27690108

72. Hsieh M-H, Goodman HM. Functional evidence for the involvement of Arabidopsis IspF homolog in the nonmevalonate pathway of plastid isoprenoid biosynthesis. Planta. 2006; 223(4):779–84. https://doi.org/10.1007/s00425-005-0140-9 PMID: 16231155

73. Soyars CL, Peterson BA, Burr CA, Nimchuk ZL. Cutting Edge Genetics: CRISPR/Cas9 Editing of Plant Genomes. Plant Cell Physiol. 2018; https://doi.org/10.1093/pcp/pcy079 PMID: 29912402

Chapter 3

Development of a robust transient expression screening system in protoplasts of *Cannabis*

3.1 Chapter preface

The recalcitrant nature of cannabis to transformational efforts has led to a lack of robust protocols to enhance the germplasm. Protoplast isolation and transformation is one such protocol enabling singular transformational events to be utilised for homogenous genetic modifications events whilst avoiding the troublesome chimeric events using other transformational techniques. To approach this issue, this chapter presents a thorough investigation into the variables involved in viable protoplast isolation and heterologous expression of GFP. Analysis of each variable's impact on yield, viability and transfection efficacy, and their respective significance, is determined with calculated optimal variable parameters determined from orthogonal arrays. This is one of the first reported protocols to efficiently isolate a significant number of viable protoplasts and also the first report of transient expression of GFP in cannabis protoplasts, providing an invaluable tool for protoplast transformational experiments in cannabis.

3.2 Publication details

Title: Development of a robust transient expression screening system in protoplasts of *Cannabis*

Journal: *In Vitro* Cellular & Developmental Biology – Plant. doi: 10.1007/s11627-021-10178-0

Stage of Publication: Published

Authors: Lennon Matchett-Oates, Ehab Mohamaden, German Spangenberg, Noel Cogan

3.3 Statement of contribution of joint authorship

LMO and EM developed protocols to isolate and transform protoplasts. LMO drafted entire manuscript and performed all experiments. GS and NC helped conceive project direction and assisted in editing the manuscript. All authors approved the final manuscript.

3.4 Co-author statement confirming authorship of PhD candidate

As co-author of the manuscript 'Matchett-Oates L, Mohamaden E, Spangenberg GC, Cogan NOI. Development of a robust transient expression screening system in protoplasts of Cannabis. In Vitro Cellular & Developmental Biology. doi: 10.1007/s11627-021-10178-0.' I confirm that Lennon Matchett-Oates has made the following contributions,

- Participation in the development of protocols employed
- Generation of all biologicals used in the study
- Generation of all figures and tables

Chapter 3

- Writing the manuscript, critical appraisal of the content and response to reviewers

Dr Noel O.I. Cogan

Date: 20/12/2021

In Vitro Cellular & Developmental Biology - Plant
https://doi.org/10.1007/s11627-021-10178-0

PROTOPLAST CULTURE AND/OR FUSION

Development of a robust transient expression screening system in protoplasts of *Cannabis*

L. Matchett-Oates[1,2] · **Ehab Mohamaden**[1] · **G. C. Spangenberg**[1,2] · **N. O. I. Cogan**[1,2]

Received: 8 November 2020 / Accepted: 23 March 2021 / Editor: Ted Klein
© The Author(s) 2021

Abstract
Transient expression systems in mesophyll protoplasts have been utilised in many plant species as an indispensable tool for gene function analysis and efficacious genome editing constructs. However, such a system has not been developed in *Cannabis* due to the recalcitrant nature of the plant to tissue culture as well as its illegal status for many years. In this study, young expanding leaves from aseptic *in vitro Cannabis* explants were used for protoplast isolation. Factorial designs were used to optimise variables in viable protoplast isolation and transient expression of GFP, with a range analyses performed to determine, and quantify, significantly impacting variables. Viable protoplast yields as high as 5.7×10^6 were achieved with 2.5% (*w/v*) Cellulase R-10, 0.3% (*w/v*) Macerozyme R-10 and 0.7 M mannitol, incubated for 16 h. As indicated by the transient expression of GFP, efficiency reached 23.2% with 30 μg plasmid, 50% PEG, 1×10^6 protoplasts and a transfection duration of 20 min. Application of the optimised protocol for protoplast isolation was successfully evaluated on three subsequent unrelated genotypes to highlight the robustness and broad applicability of the developed technique.

Keywords Cannabis · Protoplasts · Transient · Transformation

Introduction

Cannabis sativa L. (*Cannabis*) is a highly polymorphic, wind pollinated herb originating in China where evidence of its cultivation dates back to 4000 BC (Zuardi 2006). Recent interest in the medicinal properties of phytocannabinoids produced by *Cannabis* has led to increased legalisation around the world, along with a growing medicinal industry (ProCon.org 2021). Recently, chromosomal resolution of the *Cannabis* genome was published along with other genetic resources for identification of important cannabinoid biosynthesis genes (Grassa *et al.* 2018; Laverty *et al.* 2019). Next generation sequencing has also started to unravel the complexity of the *Cannabis* genome and transcriptome atlas (Braich *et al.* 2019). Progress in the genetic tools for manipulation and analysis of genes have prompted research into *Cannabis*, but currently, *Cannabis* remains a recalcitrant species to deliver biotechnology tools to. Transgenic hairy root cultures of *Cannabis* have been performed previously using agroinfiltration of vectors with the GUS reporter gene (Wahby *et al.* 2013). However, hairy root cultures only offer the accumulation of metabolites within the root structures (Gurunani *et al.* 2015) making this transformational technique unsuitable for biotechnological applications seeking phytocannabinoids, which accumulate in the female floral tissues. Transformation of hemp callus cultures with a foreign *Escherichia coli* gene, *manA*, has previously been successful using *Agrobacterium tumefaciens* with a success rate of 31% (Feeney and Punja 2003). PCR analysis of callus cultures confirmed stable gene integration with up to four T-DNA copies being integrated into the genome; however, plantlet regeneration was unsuccessful.

With the recent release of the chromosomal assembly and high-density linkage map of the *Cannabis* genome (Grassa *et al.* 2018; Laverty *et al.* 2019), identification of genes involved in cannabinoid biosynthesis is now relatively straightforward, with benchmark standards being developed for genetic engineering of these genes (Matchett-Oates *et al.* 2020.). However, to date, no procedures for the evaluation of genes

✉ L. Matchett-Oates
lennon.matchett-oates@agriculture.vic.gov.au

[1] Agriculture Victoria, the Centre for AgriBioscience, AgriBio, 5 Ring Road, Bundoora, Victoria 3083, Australia

[2] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3086, Australia

Published online: 17 May 2021

through functional screening exist, due to the recalcitrant nature of *Cannabis in vitro* and the difficulty in obtaining transgenic explants (Feeney and Punja 2017). In addition, the illegal status of cannabis during the time in which biotechnology has emerged and has caused *Cannabis* to fall behind in genetic improvement studies. Genetic transformation is used for the study of gene function and genetic improvement in plants. Protocols for many non-model species, in which regeneration of transgenic plants has not been achieved, have been developed for screening genome editing constructs. To date, only brief protocol outlines for protoplast isolation have been reported (Jones 1979; Morimoto *et al.* 2007), with no protoplast transient expression protocols existing for *Cannabis*. Creating stable transformants is expensive and time consuming making this approach for large scale evaluation of cannabinoid biosynthesis genes limiting. Utilising protoplasts ability to transiently express DNA constructs has allowed for high-throughput transient screening of genes in *Arabidopsis* (Marion *et al.* 2008), gene-silencing in barley (Douchkov *et al.* 2005) and functional analysis of newly isolated genes in tobacco (Fischer and Hain 1995). Such studies are yet to be conducted on *Cannabis* making the need for an efficient transient expression system in protoplasts of high importance.

Protoplasts are osmotically fragile due to the lack of a cell wall from enzymatic digestion, allowing for the transfer of DNA constructs through the plasma membrane using common methods such as PEG-mediated transfection (Yoo *et al.* 2007), *Agrobacterium* infiltration (Clough and Bent 1998) and biolistic bombardment (Vain *et al.* 1993). Transient gene expression from plant protoplasts has widely been used to study cell death related processes (Chen *et al.* 2015), developmental studies (Sheen 2001), subcellular localisation of proteins (Su *et al.* 2010) and expression of foreign genes (Zhang *et al.* 2011). During isolation, optimum conditions such as osmotic balance, enzyme concentration and digestion time need to be established by optimising all variables contributing to yield and viability. The use of reporter genes in transient assays is often used due to the non-toxic nature, stability and ease of detection allowing for protein localisation and interactions studies (Leffel *et al.* 1997) making protoplasts an important tool in gene analysis.

The evaluation of transiently expressed shared terpene biosynthesis genes between *Cannabis*, and other species has been reported (Reed and Osbourn 2018; Smirnoff 2019). However, *Cannabis* contains a unique set of enzymes in the cannabinoid biosynthesis pathway, which synthesise phytocannabinoids that are only produced in this species. The elucidation of the cannabinoid biosynthesis genes in the draft *Cannabis* genome provided the first complete look into the multiple pathways involved (Van Bakel *et al.* 2011). Identification and characterisation of the first unique enzyme in the cannabinoid pathway, OLS, have evolved from identification of an unknown polyketide synthase (Raharjo *et al.* 2004). Initially characterised as a novel polyketide synthase-olivetol synthase (Taura *et al.* 2009), later, the mechanism for olivetol production requiring olivetol acid cyclase, a DABB protein (Gagne *et al.* 2012), was identified and correctly characterised. Similarly, identification and functional studies on the oxidocyclases, tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS) have evolved as technological advancements into molecular cloning techniques, and DNA sequencing has improved. Tetrahydrocannabinol (THC), one of the major constituents in *Cannabis*, was originally presumed to be formed by the isomerisation of cannabidiolic acid (CBDA) (Shoyama *et al.* 1975). Protein sequencing and PCR cloning of tetrahydrocannabinolic acid (THCA) confirmed that THC was produced from cannabigerolic acid (CBGA) by THCAS (Sirikantaramas *et al.* 2004). Identification of CBDAS (Taura *et al.* 1996) occurred shortly after with a similar strategy used for gene identification (Taura *et al.* 2007).

Genetic engineering of *Cannabis* offers the opportunity to produce higher levels of cannabinoids with tailor-made chemical profiles for medicinal applications. However, the recalcitrant nature of *Cannabis* and the absence of methods to efficiently generate transgenic plants currently makes transient expression systems important to understand the molecular regulatory mechanisms responsible for cannabinoid biosynthesis, which remain largely unknown. Prior to this report, a transient expression system in *Cannabis* protoplasts did not exist, making the characterisation of unique gene function speculative. In this present study, an efficient protoplast isolation and transient expression system using *Cannabis* leaf mesophyll protoplasts are reported. Optimisation to obtain a high yield of viable protoplasts and PEG-mediated transfection of the protoplasts using fluorescent reporter genes is also reported. Furthermore, due to the outbreeding highly heterozygous nature of *Cannabis*, strain-dependent response variation is expected; this protoplast isolation and transient expression systems is shown to be suitable to many strains, highlighting the applicability for future protoplast and transient expression studies.

Materials and Methods

Plant Material and Expression Vectors All research was performed under Medicinal Cannabis Research Licence (RL011/18) and Permit (RL01118P4) issued through the Department of Health (DoH), Office of Drug Control (ODC) Australia. A specific genotype (*C. sativa*) with high THC content (25%) was used for optimisation within this study. Aseptic plantlets were derived from apical meristems of mature vegetative mother plants, which were sterilised by a 1-min 80% ethanol wash, washed three times with sterile water, disinfected with 10% (*v/v*) sodium hypochlorite (White King, Melbourne,

Australia) solution containing 4.5% active chlorine for 15 min and finally washed three times with sterile water. Apical meristems were cultured in root induction media containing ½ Murashige and Skoog salts and vitamins (MS; Murashige and Skoog 1962) (Duchefa Biochemie, Haarlem, The Netherlands), 1% sucrose (*w/v*) (Sigma-Aldrich, St. Louis, MO), 1% agar (*w/v*) (Duchefa Biochemie) and 1 mg L^{-1} Indole-3-butyric acid (IBA) (Sigma-Aldrich) adjusted to pH 5.7 prior to autoclaving at 121°C for 15 min. The apical meristems were cultured at 26°C under lighting of 74 μmol m^{-2} s^{-1} supplied by fluorescent lamps for 18 h per day. The green fluorescent protein (GFP) expression vector pDONR221-GFP was constructed from pDONR221 (Invitrogen, Carlsbad, CA) through BP Clonase reaction by the insertion of the attB1-CaMV35D-p_turboGFP(D)_AtuNos-t-attB2 cassette (Supplementary Information).

Protoplast Isolation, Purification and Quantification Protoplasts were isolated from well rooted, 1- to 2 mo-old plantlets with young leaves cut into 0.5- to 1.0-mm-thin strips (Fig. 1*a*–*d*) and incubated in a Petri dish containing digestion media comprising of 1, 2 and 2.5% (*w/v*) Cellulase Onozuka R-10 (Yakult Honsha Co., Ltd., Tokyo, Japan); 0.3, 0.4 and 0.5% (*w/v*) Macerozyme R-10 (Yakult Honsha Co.); 0.3, 0.5 and 0.7 M mannitol (Sigma-Aldrich, St Louis, MO); 20 mM MES (Sigma-Aldrich, St Louis, MO); 20 mM KCl (Sigma-Aldrich, St Louis, MO); and 10 mM $CaCl_2$ (Sigma-Aldrich), pH adjusted to 5.8 and filter-sterilised using a 0.22-μm filter (Sigma-Aldrich). Leaf strips were incubated in the dark at 28°C without agitation for between 8 and 24 h with each digestion replicated three times. Following digestion, the cellular suspension was mechanically filtrated through a 70-μm mesh (Corning, NY) into a sterile 50-mL polypropylene centrifuge tube (Corning, NY) and centrifuged at 700×*g* for 10 min. Following centrifugation, the resulting supernatant was discarded, and the pellet resuspended with 3 mL of W5 (5 mM glucose, 5 mM KCl, 10 mM MES, 125 mM $CaCl_2$, 154 mM NaCl, pH 5.8), transferred to a 15-mL round bottomed tube and 3 mL 20% (*w/v*) sucrose added and centrifuged again. Protoplasts were collected from the interphase (Fig. 1*e*) into a fresh 15-mL round bottom tube and 3 mL W5 added and centrifuged again with the supernatant removed. Finally, the pellet was resuspended in 1 mL W5, and 100 μL of the resuspended protoplasts were diluted with 0.5 M Evans Blue, and approximately 50 to 200 protoplasts were counted using a haemocytometer under a light microscope. The viability of the protoplasts was calculated by (viable protoplasts/total number of protoplasts) × 100%.

PEG-Mediated Transfection Isolated protoplasts were divided into aliquots of 5×10^5 or 1×10^6 and centrifuged at 700×*g* for 10 min with the supernatant removed. PEG4000 solutions (20%, 30%, 40% and 50% (*w/v*)) (Sigma-Aldrich) were prepared by dissolving in ddH_20 containing 0.1 M $Ca(NO_3)_2$ $4H_2O$ and 0.4 M Mannitol (Sigma-Aldrich). A volume of 100 μL transformation buffer (15 mM $MgCl_2$, 0.5 M Mannitol, 0.1% (*w/v*) MES, pH 5.7) was added to the protoplasts and gently mixed. Plasmid concentrations (5, 10 and 30 μg) in 60 μL of ddH_2O were added to the protoplast solution and gently mixed followed by addition of 150 μL of warmed (42°C) PEG4000 solutions. The protoplast solution was mixed gently and incubated in the dark at room temperature (22 ± 1°C) for between 10 and 60 min. The transfection reaction was stopped by the addition of 5 mL W5 dropwise, followed by a further 5 mL in a gentle stream. The protoplast mixture was centrifuged once more at 700×*g* for 10 min, removing the supernatant and the addition of 150 μL W5 followed by incubation in the dark at room temperature for 48 h. The expression of GFP was observed under a fluorescence microscope (OLYMPUS CKX53, Tokyo, Japan) (excitation emission wavelengths 470 to 490 nm, 510 nm).

FACS Analysis The protoplasts were analysed using Influx™ FACS instrument (BD Biosciences, Franklin Lakes, NJ) fitted with a 200-μm nozzle using W5 buffer as a sheath fluid. The sheath pressure was set at 4 psi, and the sample pressure was set at 5 psi. A 466-nm Coherent Sapphire Solid state laser was used for excitation, and emission was measured using a 517/18 nm band-pass filter for GFP. The photomultiplier tube voltage was set at 16.41 V for forward scatter, 20.59 V for side scatter and 40.37 V for GFP. The threshold value for event detection was set at 0.3 on side scattering. For analysis, a gate was set using PEG-transfected control protoplasts. Frequency of GFP expressing cells was recorded, and data was processed using BD FACSTM Software v1.0.0.650 (BD Biosciences).

Statistical Analysis Experimental data was statistically analysed using Minitab 19 Statistical Software (Version 19, State College, PA) and R Studio (Version 1.1.453, RStudio, Inc., Boston, MA).

Results

Mesophyll Protoplast Isolation From *Cannabis* Leaf Tissue Initially, protoplasts were generated from rapidly expanding leaves from 1- to 2-mo-old *in vitro* plantlets that were used as the source material (Fig. 1*a*, *b*). Briefly, plantlets cultured at 26°C in ½ MS medium with 5 μM IBA were used for mesophyll protoplast isolation. One gram of young leaf material was cut into 0.5- to 1-mm strips and immediately transferred into digestion medium in the dark without agitation at 28°C for 8 to 24 h (Fig. 1*a*–*d*). Protoplasts were collected through a

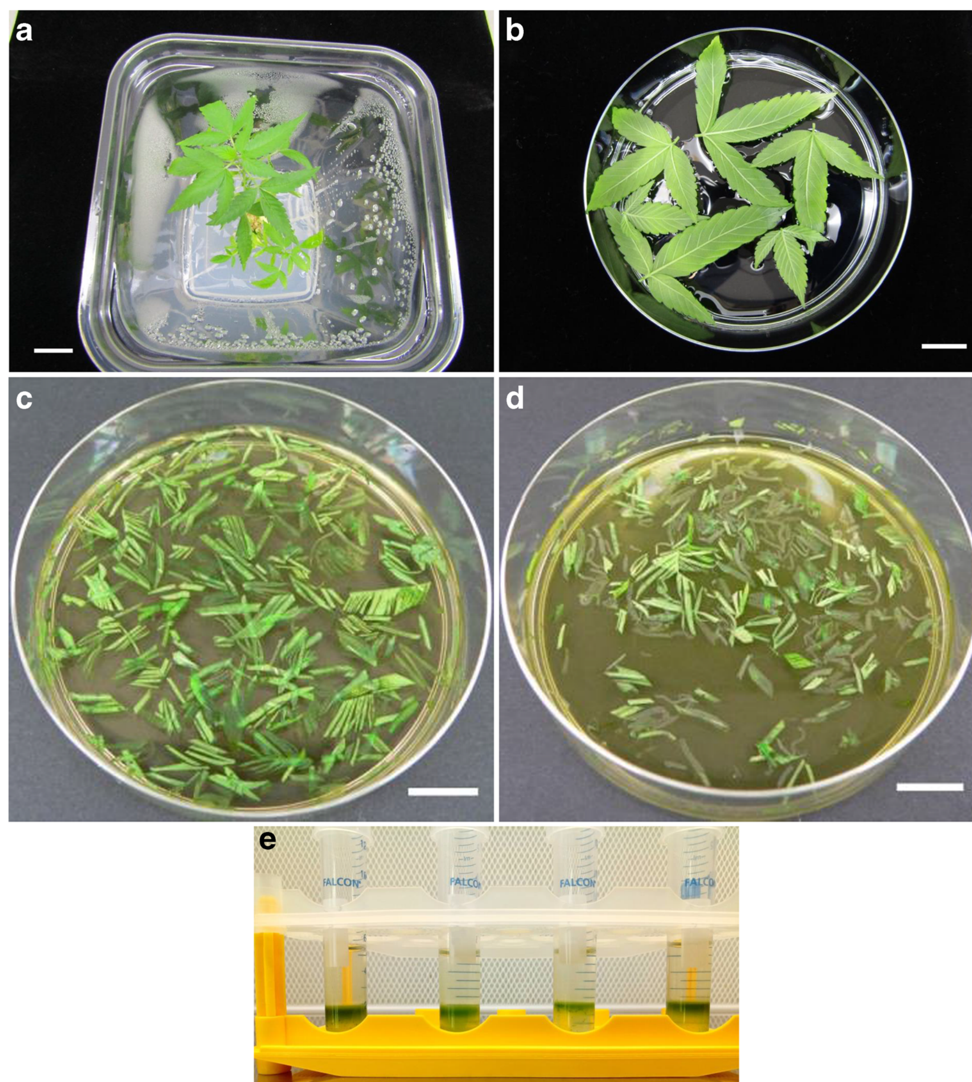


Figure 1. (*a*) *In vitro* 1- to 2-mo-old rooted explants grown on Murashige and Skoog medium containing 1 mg L^{-1} indole-3-butyric acid suitable for protoplast isolation. *Scale bar* = 2 cm. (*b*) Healthy *Cannabis sativa* L. leaves suitable for protoplast isolation. *Scale bar* = 2 cm. (*c*) *Cannabis sativa* L. leaves sliced into 0.5- to 1-mm-thin strips with a fresh razor blade and placed in media. *Scale bar* = 2 cm. (*d*) *Cannabis sativa* L. 16 h post digestion in the dark at 28°C without shaking. *Scale bar* = 2 cm. (*e*) Purified *Cannabis sativa* L. protoplast interphase after density gradient centrifugation.

70-µm nylon mesh, centrifuged and washed as previously described.

Single Factor Effect on Mesophyll Protoplast Isolation To optimise mesophyll protoplast isolation, enzyme and mannitol concentrations and enzymatic digestion time were adjusted. Mesophyll protoplasts were isolated from young, expanding leaves with variable digestion parameters, outlined in Table 1. Total protoplast yields ranged from 1.2×10^6 to 9.2×10^6 per gram of fresh leaf weight with the highest yield obtained from treatment 7. Protoplast viability ranged from 39 to 79%, with the highest viability obtained from treatment 9. An L9 (3^4) orthogonal test (Table 1) was designed to identify statistically significant variables to optimise protoplast isolation, with a range analysis performed to predict the optimal combination of the variables. The range analysis of the means between the factors and levels (Table 2), calculated Delta scores (mean score of the highest range minus the lowest mean range) for enzymolysis time (R 2.21) and Cellulase R-10 (R 1.56) significantly influencing protoplast isolation, with Macerozyme R-10 (R 1.44) and mannitol (R 1.23) proving less significant. Analysis of each level calculated response (k1–3) from the range analysis (Table 2), the Duncans multiple range test (Duncans) assigned the best possible combination of each factor as A3:B1:C3:D2. The calculated theoretical best combination was already performed in treatment 7, with an average yield of 5.7×10^6 viable protoplasts.

Increasing Cellulase-R10 concentration significantly increased protoplast yield within multilevel factors (Fig. 3), from 1.31 (1%) to 2.86 (2.5%) $\times 10^6$ viable protoplasts. This trend, of increasing concentrations greatly increasing yields, was not observed for the other variables, where applicable. Macerozyme R-10 effected protoplast yields significantly at 0.3% producing 2.93×10^6 viable protoplasts, whereas at 0.4% and 0.5%, this decreased to 1.50×10^6 and 2.14×10^6 viable protoplasts, respectively. A concentration of 0.7 M mannitol was more substantial in producing 2.89×10^6 viable protoplasts, whereas 0.3 M produced 2.01×10^6, which

Table 1. Results from the orthogonal L9 (3^4) array for *Cannabis sativa* L. protoplast isolation optimization

Treatment combination	Factors				Total yield	Yield of viable	Viability
	A	B	C	D	× 10^6 (g FW)	× 10^6 (g FW)	(%)
T1	1 (1.0%)	1 (0.3%)	1 (0.3 M)	1 (8 h)	1.8 ± 0.5	1.3 ± 0.3	71 ± 3.5
T2	1 (1.0%)	2 (0.4%)	2 (0.5 M)	2 (16 h)	2.5 ± 0.5	1.5 ± 0.2	63 ± 10.4
T3	1 (1.0%)	3 (0.5%)	3 (0.7 M)	3 (24 h)	2.0 ± 0.1	1.2 ± 0.1	59 ± 3.8
T4	2 (2.0%)	1 (0.3%)	2 (0.5 M)	3 (24 h)	3.8 ± 0.3	1.8 ± 0.1	47 ± 1.5
T5	2 (2.0%)	2 (0.4%)	3 (0.7 M)	1 (8 h)	2.6 ± 0.1	1.8 ± 0.2	68 ± 6.4
T6	2 (2.0%)	3 (0.5%)	1 (0.3 M)	2 (16 h)	5.3 ± 0.3	3.6 ± 0.2	68 ± 4.0
T7	3 (2.5%)	1 (0.3%)	3 (0.7 M)	2 (16 h)	7.8 ± 1.0	5.7 ± 1.0	72 ± 5.9
T8	3 (2.5%)	2 (0.4%)	1 (0.3 M)	3 (24 h)	3.1 ± 0.2	1.2 ± 0.01	39 ± 1.8
T9	3 (2.50%)	3 (0.5%)	2 (0.5 M)	1 (8 h)	2.2 ± 0.2	1.7 ± 0.1	79 ± 2.7

(*A*) Cellulase R-10 concentration (*w/v*), (*B*) Macerozyme R-10 concentration (*w/v*), (*C*) Mannitol concentration, (*D*) enzymolysis time, *T1-9* different treatment combinations, *1, 2, 3* three factor levels

performed greater than 0.5 M, which produced 1.67 × 10^6 viable protoplasts (Table 3). Enzymolysis time of 16 h produced a two-fold increase in viable protoplasts, with 3.61 × 10^6 compared to a shorter digestion time of 8 h (1.58 × 10^6) or a longer digestion time of 24 h (1.39 × 10^6).

Transient Transfection of *Cannabis* Mesophyll Protoplasts To establish the first transient expression system, pDONR221-GFP was used to study the effects of plasmid, PEG and protoplast concentration, including incubation time on transfection efficiency. The transfection efficiency reached 23.2% in treatment 12 (30 μg plasmid, 50% PEG (*w/v*), 20 min of incubation and 1 × 10^6 protoplasts), with the lowest recorded transfection rate of 5.78% in treatment 5 (10 μg plasmid, 20% PEG (*w/v*), 20 min incubation time and 1 × 10^6 protoplasts).

Single Factor Effect on Mesophyll Protoplast Transfection Factors with multiple levels across variables (Table 3) were designed to optimise plasmid concentration, PEG concentration (*w/v*), transfection time and protoplast concentration variables in transfection efficiency. Mesophyll protoplasts were transfected using the PEG-mediated transfection protocol described previously (Fig. 2) with transfection efficiency ranging from 5.78 to 23.20% with the highest average transfection rate, 23.20%, obtained from treatment 12. Within multilevel factors, increasing plasmid concentration greatly increased protoplast transfection within efficiencies between 10 and 30 μg (11.11 to 17.29%). Increasing PEG concentration from 20 to 50% saw increases in transfection efficiency, with 50% achieving 16.23% on average, compared to 20% PEG (*w/v*), which saw 10.93% transfection efficiency. Incubation time of 30 min gave the highest average transient efficiency of 15.82%. Similarly, a 20-min incubation resulted in 16.61% transfection efficiency. Incubation of 10 and 60 min saw a reduction of transfection efficiency, 12.02% and 13.25%, respectively. Protoplast density of 5 × 10^5 achieved higher transient efficiencies on average compared to 1 × 10^6, with 14.63% and 11.72% efficiency, respectively. From the range analysis of the means (Table 4), the calculated Delta values of plasmid concentration (R 6.18%) and PEG concentration (R 5.3%) greatly influence transfection efficiency, with transfection time (R 3.8%) and protoplasts density (R 2.91%) proving less significant.

Table 2. Range analysis of L9 *Cannabis sativa* L. protoplast isolation orthogonal array

	A	B	C	D
k1	1.31	2.93	2.01	1.58
k2	2.40	1.50	1.67	3.61
k3	2.86	2.14	2.89	1.39
Range	1.56	1.44	1.23	2.22
Optimum combination	A3	B1	C3	D2

k1, k2, and k3 indicate mean viable protoplast ×10^6 (g FW) at the 1, 2, and 3 levels. The larger the range value, the greater the influence of the factors on test results. All values expressed as ×10^6

(*A*) cellulase R-10 concentration (*w/v*), (*B*) Macerozyme R-10 concentration (*w/v*), (*C*) Mannitol concentration, (*D*) enzymolysis time

Discussion

Protoplasts offer a versatile experimental system, with transient expression systems widely applied in *Arabidopsis* (Yoo *et al.* 2007), rice (Wang *et al.* 1988) and tobacco (Töpfer *et al.* 1988) to analyse gene expression and function and to deliver genetic improvements in plants. Crucial for the development of an efficient transient expression system for genome editing constructs is the routine generation of high-quality protoplasts

Table 3. Results from the factorial array for *Cannabis sativa* L. protoplast transfection efficiency optimization

Treatment combination	Factors				Transfection efficiency
	A	B	C	D	(%)
T1	1 (5 μg)	1 (20%)	1 (10 m)	1 (0.5)	9.78 ± 2.59
T2	1 (5 μg)	2 (30%)	2 (20 m)	1 (0.5)	14.35 ± 2.48
T3	1 (5 μg)	3 (40%)	3 (30 m)	2 (1)	10.90 ± 0.35
T4	1 (5 μg)	4 (50%)	4 (60 m)	2 (1)	9.41 ± 0.48
T5	2 (10 μg)	1 (20%)	2 (20 m)	2 (1)	5.78 ± 1.22
T6	2 (10 μg)	2 (30%)	1 (10 m)	2 (1)	6.20 ± 0.47
T7	2 (10 μg)	3 (40%)	4 (60 m)	1 (0.5)	16.44 ± 2.09
T8	2 (10 μg)	4 (50%)	3 (30 m)	1 (0.5)	16.08 ± 1.35
T9	3 (30 μg)	1 (20%)	3 (30 m)	1 (0.5)	17.24 ± 1.99
T10	3 (30 μg)	2 (30%)	4 (60 m)	1 (0.5)	13.88 ± 0.33
T11	3 (30 μg)	3 (40%)	1 (10 m)	2 (1)	14.83 ± 0.66
T12	3 (30 μg)	4 (50%)	2 (20 m)	2 (1)	23.20 ± 0.61

(*A*) plasmid concentration, (*B*) PEG concentration (*w*/*v*), (*C*) experiment time, (*D*) protoplast density ($\times 10^6$), *T1-12* different treatment combinations, *1, 2, 3, 4* four factor levels

and robust transfection protocols. Only two reports of isolated *Cannabis* protoplasts exist; however, the reports briefly discuss the methods for protoplast isolation with no data on protoplast yield and viability, or the data collection was not described in detail. Morimoto *et al.* (2007) digested *Cannabis* leaves with 1% Cellulase R-10, 0.2% Macerozyme R-10, 0.1% pectolyase Y-23 and 0.4 M mannitol at 30°C for 4 h with gentle agitation; however, no yield or viability data was given. Jones (1979) explored a range of enzyme combinations to produce protoplast from young and old leaf tissue, as well as callus, with concentrations between 1×10^3 and 1×10^5 reported. Protoplast isolation from the closely related species *Humulus lupulus* (hops) has previously been reported from cell suspension cultures. Several cell wall digesting enzyme mixtures were trialled with varying Cellulysin and Driselase concentrations in 0.4 M mannitol osmoticum, with the highest yield of protoplasts (9.3 to 9.9×10^6 per gFW) obtained from 2% Cellulysin and 1% Driselase with 99 to 100% viability

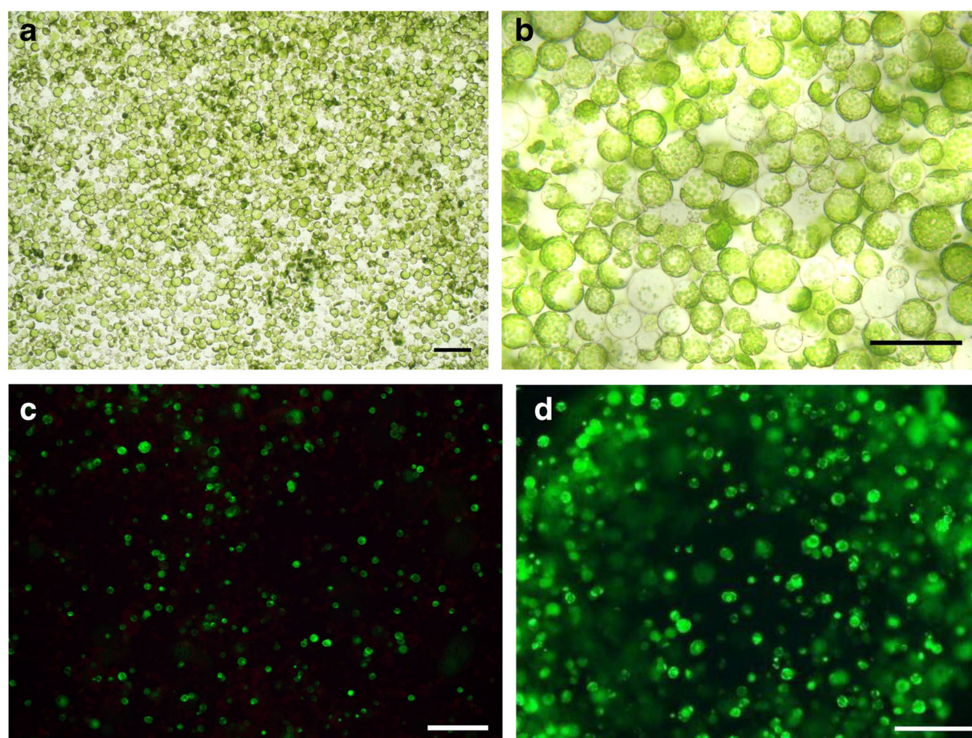


Figure 2. (*a*) Visualisation of *Cannabis sativa* L. protoplasts under brightfield microscopy. *Scale bar* = 200 μm. (*b*) Visualisation of *Cannabis sativa* L. protoplasts under fluorescence microscopy with GFP filter set. *Scale bar* = 50 μm. (*c*) Isolated protoplasts under fluorescence microscopy with GFP filter set. *Scale bar* = 200 μm. (*d*) Isolated *Cannabis sativa* L. protoplasts under fluorescence microscopy with GFP filter set. *Scale bar* = 100 μm.

Table 4. Range analysis of factorial design for *Cannabis sativa* L. protoplast transfection efficiency

	A	B	C	D
k1	11.11	10.93	12.02	14.63
k2	11.13	11.48	15.61	11.72
k3	17.29	14.06	15.82	
k4		16.23	13.25	
Range	6.18	5.30	3.80	2.91

k1, k2, k3, and k4 indicate mean transfection efficiency (%) at the 1, 2, 3, and 4 levels. All values expressed as percentages

(*A*) plasmid concentration, (*B*) PEG concentration (*w/v*), (*C*) experiment time, (*D*) protoplast density ($\times 10^6$)

reported (Furze *et al.* 1987). Heale *et al.* (Heale *et al.* 1989) reported leaf mesophyll isolation from the hop cultivar, Challenger, by firstly removing the epidermis followed by incubating in 4% Cellulase, 0.3% Macerozyme and 2% hemicellulose for 7 h on an orbital shaker protoplast yields between 4 and 7×10^5 mL^{-1} were achieved with viability ranging from 80 to 90% under optimal conditions. An efficient protocol for transient expression of LUC activity in rose (another member of the Order Rosales with *Cannabis*) leaves using *Agrobacterium* has also been reported (Lu *et al.* 2017). Recently, transient expression of β-glucuronidase (GUS) and GFP in a range of agroinfiltrated organs and tissues has been reported in *Cannabis*, with the protocol being optimised for hemp cultivars (Deguchi *et al.* 2020).

The development of an efficient protoplast transient expression system to screen genome editing constructs requires established protocols for the isolation of viable protoplasts and a competent transfection workflow, which until now have not been reported for *Cannabis*. With several factors affecting yield and viability of protoplasts and transfection efficiencies, this development of a routine system allows for significant advances to be made towards understanding and improving *Cannabis* cultivars. The method described here investigates the effect each variable has on protoplast isolation and transfection, with each being analysed for their significance.

To optimise protoplast isolation, Cellulase-R10, Macerozyme-R10, mannitol concentrations and digestion times were adjusted (Table 1). Under the optimum conditions, viable *Cannabis* leaf mesophyll protoplasts reached 5.7×10^6 protoplasts g FW^{-1} (Table 1). The increase in Cellulase concentration to 2.5% (*w/v*) showed significantly increased yields, 2.88×10^6, from the range analysis. Although 2.5% (*w/v*) Cellulase concentration was the highest tested, increasing levels past a saturation point has been shown to decrease protoplast yields in tobacco (Kuriakose *et al.* 2012) and *Magnolia* (Shen *et al.* 2017). The optimal Macerozyme concentration was determined to be 0.3% (*w/v*), whilst an increase to 0.4% saw more than a 50% drop in protoplast recovery to 1.50×10^6 calculated from the range analysis. For both enzymes, the increase in concentration at which viable protoplast yield decreases is presumably due to the influence those enzymes have on the membrane integrity. The application of the optimised protoplast isolation protocol to verify robustness on Cannbio-2, the alternative high THC strain and a high CBD strain produced a significant yield of viable protoplasts for subsequent transient expression experiments, demonstrating the robustness and versatility of the protocol described. The protocol produced similar protoplast viability, with Cannbio-2 averaging 79%, equalling to the best performing treatment 9, and the high CBD strain producing higher levels of viability with 82% (Table 5). Although viable protoplast yield from these cultivars is proportionately fewer, the protocol has been optimised for a specific high yielding THC strain, with the calculated optimal conditions proven to be an advisory starting point for further optimisation on any chosen cultivar (Table 5).

Protoplasts lack of cell walls that require a stabilized environment, which is controlled by the osmotic gradient for proper osmolarity to sustain viable protoplasts (Sain *et al.* 2017). Mannitol is frequently used for it is inert metabolically and slowly diffuses through the cellular membrane (Chawla 2011). Mannitol concentration of 0.7 M resulted in the highest yield, calculated by the range analysis, of 2.91×10^6, with lower concentrations of 0.3 M and 0.5 M yielding 2.02×10^6 and 1.67×10^6, respectively. Similar to enzyme concentration, increasing mannitol concentration, thus causing an imbalance in the osmoticum, decreases protoplast yield, as is seen with wheat (Jia *et al.* 2016) and pineapple (Priyadarshani *et al.* 2018). Enzymolysis time significantly affected viable protoplast yield, with digestions of 24-h yielding as low as 39% viability in treatment 8, which most certainly caused by over digestion of the cell walls. Comparatively, 8 and 16 h were determined to yield significantly higher levels of viability, with 16 h shown to result in a larger concentration of viable protoplasts (Fig. 3) and the best digestion time for the release of viable protoplasts without over digestion.

The uptake of DNA through the plasma membrane of protoplasts for transient expression studies requires considerable concentrations of PEG acting in tandem with divalent cations (Maas and Werr 1989). With increasing concentrations of PEG, DNA hydration reduces causing structural changes

Table 5. Results of the mean *Cannabis sativa* L. protoplast isolation on multiple cultivars using optimised variable conditions

Cultivar	Yield of viable protoplasts ($\times 10^6$)	Viability (%)
Cannbio-2	1.97	79
High THC	1.82	72
High CBD	2.46	82

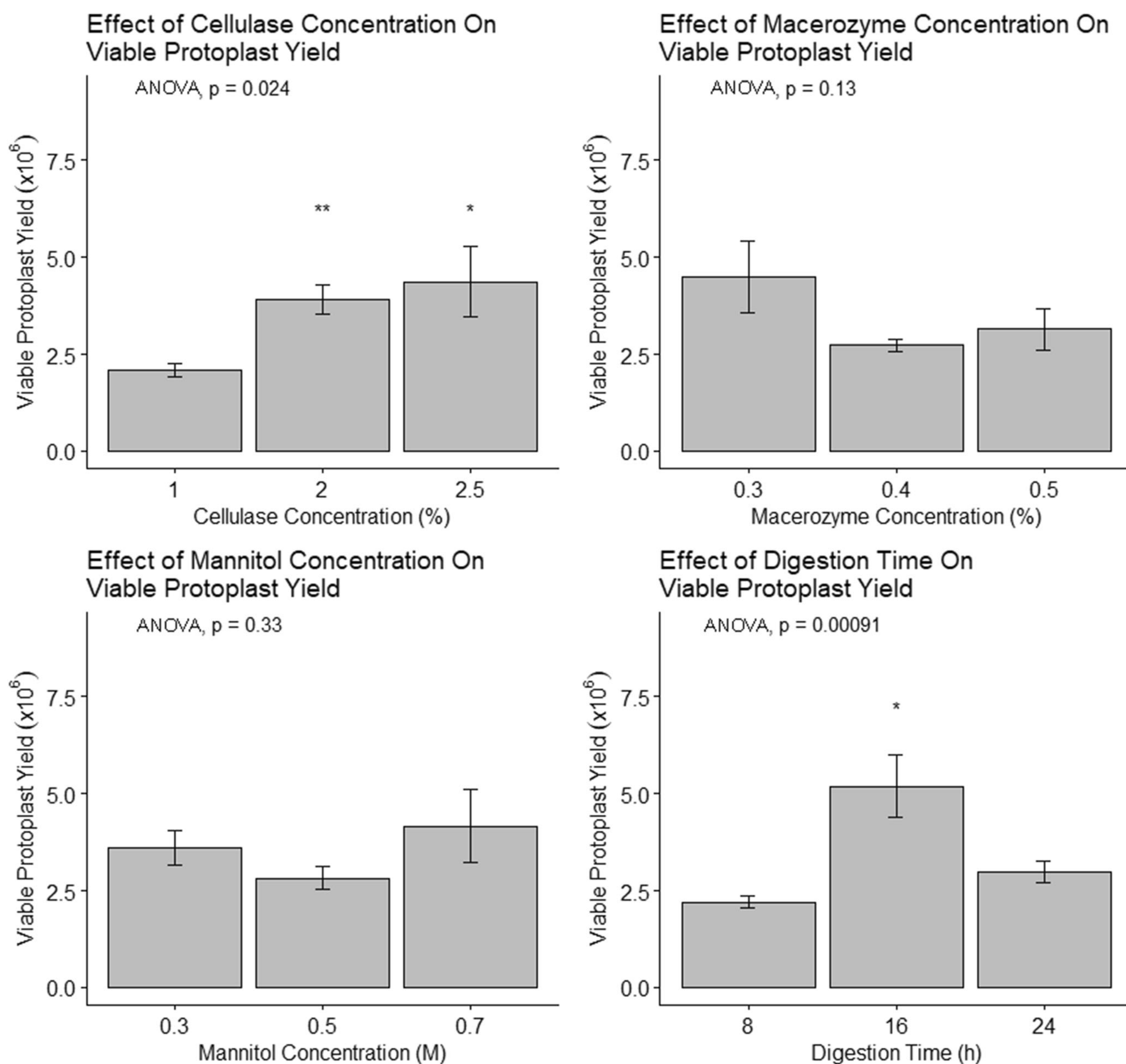


Figure 3. Effects of different factors on viable *Cannabis sativa* L. protoplast isolation yield. *Bars* represent standard errors (SE). Statistical significance was determined using a one-way ANOVA test (*$p < 0.05$, **$p < 0.01$).

and thus reducing the transfection efficiency (Saenger *et al.* 1986). In *Cannabis* protoplasts, transient expression efficiency at the highest investigated PEG concentration (50%) achieved 16.23% (Fig. 4), which is a significantly higher concentration of PEG than has been previously reported in rice (Page *et al.* 2019) or pea (Nicolaisen and Poulsen 1993). Results show that increasing plasmid concentration from 5 to 10 μg achieved similar transfection efficiencies, 11.11% and 11.13%, respectively (Table 3). These results are inconsistent with findings in pepper (Jeon *et al.* 2007), pineapple (Priyadarshani *et al.* 2018) and *Phaseolus* (Nanjareddy *et al.* 2016) in which there is an approximate doubling in relevant transient expression efficiencies between these two concentrations. This suggests that *Cannabis* protoplasts require higher concentrations of plasmid to achieve increased levels of expression. Intermediate exposure time to high concentrations on PEG was shown to increase transfection efficiencies (Fig. 4), with the optimal exposure time determined to be 30 min, achieving 15.82%. Increasing the incubation time to 60 min saw a sharp decrease in transfection efficiency, falling to 12.02%. The increased exposure to the high concentrations of PEG resulting in lower transfection efficiencies is expected due to DNA becoming less hydrated, with similar results found in carrot, rapeseed and soybean (Rasmussen and Rasmussen 1993). This protocol has been optimised for this particular cultivar (Fig. 4). This protocol provides a relevant

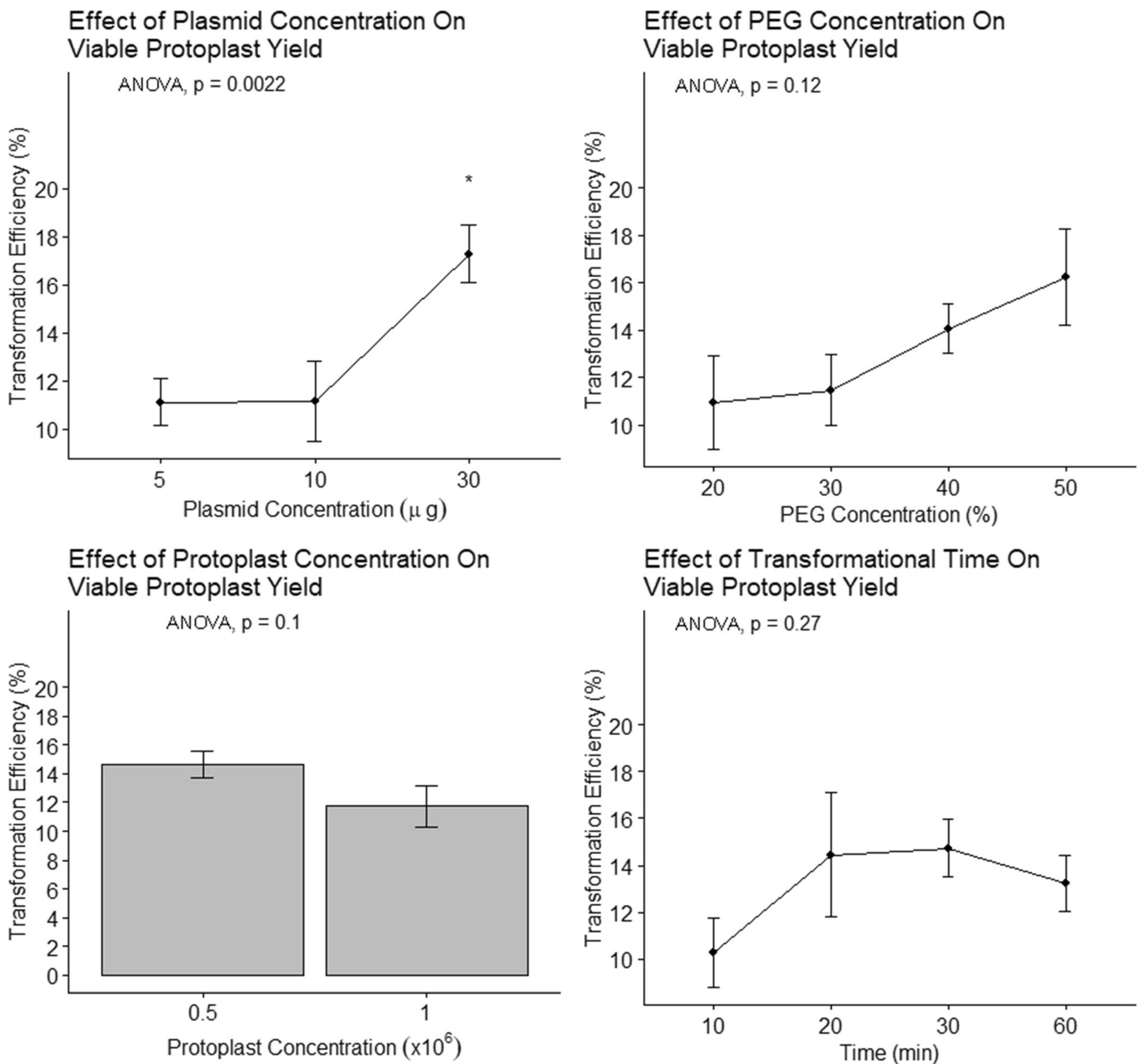


Figure 4. Effects of different factors on *Cannabis sativa* L. protoplast transfection efficiency. Bars represent standard errors (SE). Statistical significance was determined using a one-way ANOVA test (*$p < 0.05$).

starting point for optimisation regardless of genetics. This level of efficiency in transfection allows for cellular studies, including genome editing using CRISPR and ZFNs.

Conclusion

The method described here is the first reported for the transient expression of heterologous genes in *Cannabis* protoplasts. The variables involved in protoplast isolation were verified on three cultivars with varying cannabinoid content for protocol robustness. The transfection protocol was optimised for a high THC yielding strain within mesophyll protoplasts. This method can be easily adapted for transient expression studies using CRISPR/Cas-9, protein-protein interaction or other investigations in *Cannabis* where a transient gene expression system is desired.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11627-021-10178-0.

Acknowledgements Authors acknowledge Canopy Growth Corporation for access to plant material to evaluate the protocols, where applicable.

The authors also wish to acknowledge Navdeep Kaur for FACS operation and optimisation.

Author contribution LMO and EM in protocol development and optimisation. LMO assembled the manuscript.

Funding Funding was provided by Agriculture Victoria Services and Agriculture Victoria Research.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Braich S, Ballie R, Jewell L, Spangenberg G, Cogan N (2019) Generation of a comprehensive transcriptome atlas and transcriptome dynamics in medicinal cannabis. Sci Rep 9:1–12

Chawla HS (2011) Introduction to plant biotechnology, vol 3. Science Publishers, Enfield New Hampshire

Chen J, Yi Q, Song Q, Gu Y, Zhang J, Hu Y, Liu H, Liu Y, Yu G, Hang Y (2015) A highly efficient maize nucellus protoplast system for transient gene expression and studying programmed cell death-related processes. Plant Cell Rep 34:1239–1251

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16:735–743

Deguchi M, Bogush D, Weeden H, Spuhler Z, Potlakayala S, Kondo T, Zhang Z, Rudrabhatla S (2020) Establishment and optimization of a hemp (*Cannabis sativa* L.) agroinfiltration system for gene expression and silencing studies. Sci Rep 10:1–11

Douchkov D, Nowara D, Zierold U, Schweizer P (2005) A high-throughput gene-silencing system for the functional assessment of defense-related genes in barley epidermal cells. Mol Plant Microbe Interact 18:755–761

Feeney M, Punja ZK (2003) Tissue culture and *Agrobacterium*-mediated transformation of hemp (*Cannabis sativa* L.). *In Vitro* Cell Dev Biol - Plant 39:578–585

Feeney M, Punja ZK (2017) The role of *Agrobacterium*-mediated and other gene-transfer technologies in Cannabis research and product development, Chandra, S., Lata, H., ElSohly, M. (eds), *Cannabis sativa L.*-Botany and Biotechnology, vol 1, Springer, Cham. pp. 343–363

Fischer R, Hain R (1995) Tobacco protoplast transformation and use for functional analysis of newly isolated genes and gene constructs. In: Galbraith DW, Bourque DP, Bohnert HJ (eds) Methods in Cell Biology, vol 50. Elsvier, pp 401–410

Furze JM, Rhodes MJC, Robins RJ (1987) The use of agarose bead culture for the regeneration of single cell-derived colonies from protoplasts isolated from suspension cultures of *Humulus lupulus*. Plant Cell Tiss Org Cult 8:17–25

Gagne SJ, Stout J, Liu E, Boubakir Z, Clark S, Page J (2012) Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. Proc Natl Acad Sci 109:12811–12816

Grassa CJ, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, Schwartz CJ, Weiblen GD (2018) A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. bioRxiv:458083. https://doi.org/10.1101/458083

Gurunani SG, Gholse YN, Yeole MP, Chaple DR (2015) Hairy root culture: an experimental system for secondary metabolite production. Res J Pharm Tech 8:728–730

Heale JB, Legg T, Connell S (1989) *Humulus lupulus* L.(Hop): *In vitro* culture; Attempted production of bittering components and novel disease resistance. In: Bajaj YPS (ed) Medicinal and Aromatic Plants II, Biotechnology in Agriculture and Forestry, vol 7. Springer, Berlin, Heidelberg, pp 264–285

Jeon JM, Ahn NY, Son BH, Kim CY, Han C, Kim G, Gal SW, Lee S (2007) Efficient transient expression and transformation of PEG-mediated gene uptake into mesophyll protoplasts of pepper (*Capsicum annuum* L.). Plant Cell Tiss Org Cult 88:225–232

Jia X, Zhang X, Qu J, Han R (2016) Optimization conditions of wheat mesophyll protoplast isolation. Agri Sci 7:850–858

Jones RL (1979) Cell culture, protoplast isolation, and cell fusion of Cannabis Sativa L.: evaluation of chilling preventative chemicals and quality control of bananas in the tropics, University of Houston Central Campus, M. Sc. Thesis, DOI: 10.13140/RG.2.2.19104.43527

Kuriakose B, Du Toit ES, Jordaan A (2012) Transient gene expression assays in rose tissues using a Bio-Rad Helios hand-held gene gun. S Afr J Bot 78:307–311

Laverty KU, Stout J, Sullivan MJ, Shah H, Gill N, Holbrook L, Deikus G, Sebra R, Hughes TR, Page JE, van Bakel H (2019) A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Res 29:146–156

Leffel SM, Mabon SA, Stewart CN Jr (1997) Applications of green fluorescent protein in plants. Biotechniques 23:912–918

Lu J, Bai M, Ren H, Liu J, Wang C (2017) An efficient transient expression system for gene function analysis in rose. Plant Method 13:116

Maas C, Werr W (1989) Mechanism and optimized conditions for PEG mediated DNA transfection into plant protoplasts. Plant Cell Rep 8: 148–151

Marion J, Bach L, Bellec Y, Meyer C, Gissot L, Faure J (2008) Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of *Arabidopsis* seedlings. Plant J 56:169–179

Matchett-Oates L, Braich S, Spangenberg G, Rochfort S, Cogan N (2020) *In silico* enablinginformed design for genome editing in medicinal cannabis, gene families and variant characterisation. Plant Methods. IN REVIEW

Morimoto S, Tanaka Y, Sasaki K, Tanaka H, Fukamizu T, Shpyama Y, Shoyama Y, Taura F (2007) Identification and characterization of cannabinoids that induce cell death through mitochondrial permeability transition in Cannabis leaf cells. J Biol Chem 282:20739–20751

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473–497

Nanjareddy K, Arthikala M, Blanco L, Arellano E, Lara M (2016) Protoplast isolation, transient transformation of leaf mesophyll protoplasts and improved Agrobacterium-mediated leaf disc infiltration of Phaseolus vulgaris: tools for rapid gene expression analysis. BMC Biotechnol 16:53

Nicolaisen M, Poulsen GB (1993) Optimization of polyethylene glycol mediated transient gene expression in pea protoplasts. Plant Cell Tiss Org Cult 35:93–97

Page MT, Parry MAJ, Carmo-Silva E (2019) A high-throughput transient expression system for rice. Plant Cell Environ 42:2057–2064

Priyadarshani S, Hu B, Li W, Ali H, Jia H, Zhao L, Ojolo SP, Azam SM, Xiong J, Yan M, Rahman ZU, Wu Q, Qin Y (2018) Simple protoplast isolation system for gene expression and protein interaction studies in pineapple (*Ananas comosus* L.). Plant Method 14:95

ProCon.org (2021) Legal Medical Marijuana States and DC. Available at: https://medicalmarijuana.procon.org/legal-medical-marijuana-states-and-dc/ (Accessed: 24 March 2021).

Raharjo TJ, Cheng W, Verberne MC, Peltenburg-Looman AMG, Linthorst HJM, Verpoorte R (2004) Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. Plant Physiol and Biochem 42:291–297

Rasmussen JO, Rasmussen OS (1993) PEG mediated DNA uptake and transient GUS expression in carrot, rapeseed and soybean protoplasts. Plant Sci 89:199–207

Reed J, Osbourn A (2018) Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. Plant Cell Rep 37: 1431–1441

Saenger W, Hunter WN, Kennard O (1986) DNA conformation is determined by economics in the hydration of phosphate groups. Nature 324:385–388

Sain PM, Rodríguez S, Disalvo EA (2017) Effect of osmotic stress on the surface properties of protoplasts as measured by merocyanine 540. Soc Plant Res 30:75–80

Sheen J (2001) Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. Plant Physiol 127:1466–1475

Shen Y, Meng D, McGrouther K, Zhang J, Cheng L (2017) Efficient isolation of Magnolia protoplasts and the application to subcellular localization of MdeHSF1. Plant Method 13:44

Shoyama Y, Yagi M, Nishioka I, Yamauchi T (1975) Biosynthesis of cannabinoid acids. Phytochem 14:2189–2192

Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F (2004) The gene controlling marijuana psychoactivity molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic-acid synthase from Cannabis sativa. J Biol Chem 279:39767–39774

Smirnoff N (2019) Engineering of metabolic pathways using synthetic enzyme complexes. Plant Physiol 179:918–928

Su C, Wang Y, Hsieh T, Lu C, Tseng T, Yu S (2010) A novel MYBS3-dependent pathway confers cold tolerance in rice. Plant Physiol 153: 145–158

Taura F, Morimoto S, Shoyama Y (1996) Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. J Bio Chem 271:17411–17416

Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S (2007) Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. FEBS Lett 581:2929–2934

Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y, Morimoto S (2009) Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Lett 583:2061–2066

Töpfer R, Pröls M, Schell J, Steinbiß HH (1988) Transient gene expression in tobacco protoplasts: II. Comparison of the reporter gene systems for CAT, NPT II, and GUS. Plant Cell Rep 7:225–228

Vain P, McMullen MD, Finer JJ (1993) Osmotic treatment enhances particle bombardment-mediated transient and stable transformation of maize. Plant Cell Rep 12:84–88

Van Bakel H, Stout J, Cote A, Tallon C, Sharpe A, Hughes T, Page J (2011) The draft genome and transcriptome of *Cannabis sativa*. Genome Biol 12:R102

Wahby I, Caba JM, Ligero F (2013) *Agrobacterium* infection of hemp (*Cannabis sativa* L.): establishment of hairy root cultures. J Plant Interact 8:312–320

Wang YC, Klein TM, Fromm M, Cao J, Sanford JC, Wu R (1988) Transient expression of foreign genes in rice, wheat and soybean cells following particle bombardment. Plant Mol Biol 11:433–439

Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Prot 2:1565–1572

Zhang Y, Su J, Duan S, Ao Y, Dai J, Liu J, Wang P, Li Y, Liu B, Feng D, Wang J, Wang H (2011) A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. Plant Method 7:30–38

Zuardi AW (2006) History of cannabis as a medicine: a review. Braz J Psychiatry 28:153–157

Chapter 4

Manipulation of cannabinoid biosynthesis via transient RNAi expression

4.1 Chapter preface

Cannabinoid biosynthesis genes, responsible for the accumulation of phytocannabinoids, hold significant value in providing medicinally important benefits for a large array of medical conditions. These genes are highly homologous with their respective roles within transcriptional activity still being investigated. The accumulation of individual cannabinoids is due in part to the efficacy of each gene and possibly their respective copy number in the genome. Each produced cannabinoid has an array of listed medicinal benefits, with each cultivar of cannabis producing significantly variable levels of each of these medicinally important cannabinoids. The ability to create a cannabis cultivar producing an increased level of desired cannabinoids, whilst significantly limiting the accumulation of the undesirable cannabinoids, is of great medicinal importance. In this chapter, the use of RNAi vector constructs to target individual cannabinoid biosynthesis genes to significantly silence the transcription of these important genes using agroinfiltration is investigated. Using gene specific RNAi constructs, significant reduction in detectable transcript levels by qPCR are reported. This is the first report of modification in cannabinoid biosynthesis gene expression allowing for the use of such constructs to produce tailored cannabis chemovars. The data presented here also sheds light on the capability of these highly homologous enzyme's ability to synthesis different cannabinoids *in planta*.

4.2 Publication details

Title: Manipulation of cannabinoid biosynthesis via transient RNAi expression

Journal: Frontiers in Plant Science – Technical Advances in Plant Science

Stage of publication: Published https://doi.org/10.3389/fpls.2021.773474

Authors: Lennon Matchett-Oates, German Spangenberg, Noel Cogan

4.3 Statement of contribution of joint authorship

LMO designed all RNAi constructs and performed all experiments. LMO drafted entire manuscript. GS and NC helped conceive project direction and assisted in editing the manuscript. All authors approved the final manuscript.

4.4 Co-author statement confirming authorship of PhD candidate

As co-author of the manuscript 'Matchett-Oates L, Spangenberg GC, Cogan NOI. Manipulation of cannabinoid biosynthesis via transient RNAi expression. Frontiers in Plant Science, 2021,12,2766. doi.org/10.3389/fpls.2021.773474.' I confirm that Lennon Matchett-Oates has made the following contributions,

Chapter 4

- Development of protocols employed
- Generation of all biologicals used in the study
- Generation of all figures and tables
- Writing the manuscript, critical appraisal of the content and response to reviewers

Dr Noel O.I. Cogan

Date: 20/12/2021

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fpls.2021.773474

Manipulation of Cannabinoid Biosynthesis *via* Transient RNAi Expression

Lennon Matchett-Oates[1,2], German C. Spangenberg[1,2] and Noel O. I. Cogan[1,2]*

[1]Agriculture Victoria, AgriBio, The Centre for AgriBioscience, Bundoora, VIC, Australia, [2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia

Cannabis sativa L. produces unique phytocannabinoids, which are used for their pharmaceutical benefits. To date, there are no reports of *in vivo* engineering targeting the cannabinoid biosynthesis genes to greater elucidate the role each of these genes play in synthesis of these medically important compounds. Reported here is the first modulation of cannabinoid biosynthesis genes using RNAi *via* agroinfiltration. Vacuum infiltrated leaf segments of the Cannbio-2 *C. sativa* strain, transfected with different RNAi constructs corresponding to *THCAS, CBDAS,* and *CBCAS* gene sequences, showed significant downregulation of all cannabinoid biosynthesis genes using real-time quantitative PCR. Using RNAi, significant off-targeting occurs resulting in the downregulation of highly homologous transcripts. Significant ($p < 0.05$) downregulation was observed for *THCAS* (92%), *CBDAS* (97%), and *CBCAS* (70%) using pRNAi-GG-*CBDAS*-*UNIVERSAL*. Significant ($p < 0.05$) upregulation of *CBCAS* (76%) and non-significant upregulation of *THCAS* (13%) were observed when transfected with pRNAi-GG-*CBCAS*, suggesting the related gene's ability to synthesize multiple cannabinoids. Using this approach, increased understanding of the relationship between cannabinoid biosynthesis genes can be further elucidated. This RNAi approach enables functional genomics screens for further reverse genetic studies as well as the development of designer cannabis strains with over-expression and/or downregulation of targeted cannabinoid biosynthesis genes. Functional genomics screens, such as these, will further provide insights into gene regulation of cannabinoid biosynthesis in *Cannabis*.

Keywords: ***Cannabis sativa***, **RNAi, cannabinoid biosynthesis genes, *agrobacterium*, post-transcriptional gene silencing, THCAS, CBDAS, CBCAS**

OPEN ACCESS

Edited by:
Marcos Egea-Cortines, Universidad Politécnica de Cartagena, Spain

Reviewed by:
Nirit Bernstein, Agricultural Research Organization (ARO), Israel
Chen Junfeng, Shanghai University of Traditional Chinese Medicine, China

****Correspondence:***
Lennon Matchett-Oates
lennon.matchett-oates@agriculture.vic.gov.au

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

Received: *10 September 2021*
Accepted: *08 November 2021*
Published: *10 December 2021*

Citation:
Matchett-Oates L, Spangenberg GC and Cogan NOI (2021) Manipulation of Cannabinoid Biosynthesis via Transient RNAi Expression. Front. Plant Sci. 12:773474. doi: 10.3389/fpls.2021.773474

INTRODUCTION

Cannabis sativa L. is one of the earliest domesticated and cultivated plants with records of its use in central Asia dating back more than 6,000 years (Li, 1973). Cannabis belongs to the *Cannabaceae* family and has been used for millennia for its source of bast fiber, seed oil, food, and psychoactive constituents for recreational and medicinal purposes (Touw, 1981). Cannabis produces more than 120 cannabinoids, which are unique secondary metabolites found only in cannabis (ElSohly et al., 2017). Cannabis contains a unique Cannabinoid biosynthesis

pathway which produces biologically inactive compounds, such as Tetrahydrocannabinolic acid (THCA) and Cannabidiolic acid (CBDA; Matchett-Oates et al., 2021a) which when decarboxylated are converted to their biologically active forms Δ^9-tetrahydrocannabinol (THC) and Cannabidiol (CBD) displaying psychoactive and non-psychoactive properties, respectively (Kogel et al., 2018). Other major cannabinoids of interest produced are cannabigerol (CBG), cannabichromene (CBC), cannabinol (CBN), and tetrahydrocannabivarin (THCV). The pharmacological effects of these cannabinoids have been of great interest due to the affinity these chemical compounds have for the endogenous cannabinoid system receptors (Movahedi et al., 2015). The use of medicinal cannabis in the treatment of conditions, including pain management (Campbell et al., 2001), cancer (Machado Rocha et al., 2008), multiple sclerosis (Rog et al., 2005), and epilepsy (Russo, 2017), has been widely reviewed. THC has been the primary cannabinoid studied in cannabis research since its discovery (Gaoni and Mechoulam, 1964), but now considerable interest exists in understanding the activity of the other major cannabinoids and their possible therapeutic properties. More specifically, the common precursor of all cannabinoids is CBG, which is enzymatically synthesized into the unique phytocannabinoids, giving cannabis its therapeutic potential (Borrelli et al., 2013).

The dioecious, wind pollination nature of cannabis has created a highly diverse genetic pool in which strains are generated in clandestine breeding efforts, creating a highly diverse population with high levels of sequence and copy number variations affecting the drug content (Weiblen et al., 2015; Matchett-Oates et al., 2021a). Cannabis can be classified into different chemotypes according to their CBD:THC ratio (Pacifico et al., 2006). THCA synthase (THCAS) and CBDA synthase (CBDAS) are the competing enzymes for the common precursor, cannabigerolic acid (CBGA), which determines the chemotype of cannabis plants. The loci containing these synthase genes have recently been resolved showing that as many as 13 synthase gene copies reside within chromosome 7 (Grassa et al., 2018). Further comparison of publicly available cannabis genomes shows that there is significant variation in total synthase gene copy number with sequence homology between all genes being greater than 90% (Grassa et al., 2021; Matchett-Oates et al., 2021a). It is this variation and tightly linked regions that makes the cannabinoid biosynthesis pathway complex to engineer with the intent to create novel designed chemotypes of cannabis for therapeutic uses. Such examples to engineer the cannabinoid pathway within yeast to produce cannabinoids are already possible (Luo et al., 2019), though the adaptation of this approach toward medical applications is still yet to be addressed.

Development of new cannabis strains for medicinal purposes through traditional breeding efforts is a lengthy and expensive process. The use of targeted gene silencing tools to accurately and efficiently knockdown targeted gene expression will enable the generation of novel cannabinoid profiles. The development of genetically modified plants raises public concern for their potential consequences on human health. An alternative when using RNAi is the application of exogeneous dsRNA to induce gene silencing without risking societal acceptance. However, the majority of studies regarding exogenous application of dsRNA is rarely applied under open-field conditions assessing the environmental factors affecting RNAi efficacy, with such practices currently unperformed using cannabis. The use of RNAi is not considered genetically modified through some regulatory agencies (Office of the Gene Technology Regulator, 2018), which can improve the end point consumers opinions regarding novel chemotypes developed using RNAi technologies. Through genome-wide association studies on *THCAS* and *CBDAS* loci, it has been shown that a cannabis variety with a functional *THCAS* but a non-functional *CBDAS* locus is possible (Welling et al., 2020). Conversely, a cannabis variety with a non-functional *THCAS* locus has not been discovered, indicating trace levels of THC will always be produced, such is the case with hemp. Using gene silencing tools, designer strains with high levels of CBD producing zero THC are possible, as are strains with elevated levels of CBG, which contains anti-cancer properties (Borrelli et al., 2014), through the knockdown of the downstream enzymatic processes of *THCAS, CBDAS,* and *CBCAS*. The use of environmental pressures applied through varying nutrient concentrations (Saloner and Bernstein, 2021; Shiponi and Bernstein, 2021) or light spectrum and lighting source (Magagnini et al., 2018; Namdar et al., 2019) has previously demonstrated significant modulation of secondary metabolites, up to 300% in some instances (Shiponi and Bernstein, 2021). While this ability to variably control cannabinoid content in cannabis using environmental conditions is significant, the synergistic effects of all cannabinoids either increasing or decreasing make this approach incapable of producing a complete knockdown/significant downregulation of specific cannabinoids to create novel chemotypes. The generation of stably transformed lines is a lengthy process, requiring protocol development for transformation and regeneration. Transient expression systems are widely used as a valuable tool for vector construct evaluation, all the while being fast and inexpensive with specific protocols in cannabis already developed (Schachtsiek et al., 2019; Deguchi et al., 2020) exploring dsRNA and virus-induced gene silencing mechanisms, with significantly downregulated targeted gene expression levels observed. RNAi transient gene suppression is a well-characterized method for reverse genetics and can allow for rapid screening of RNAi constructs for later stable transformation using *Agrobacterium*. Intron-containing hairpin RNA (ihpRNA) are used to induce degradation of targeted genes using RNAi mechanisms. The generation of small interfering RNA (siRNA), from dsRNA by Dicer-like proteins (DCLs), binds to the RNA-induced silencing complex (RISC), with one strand of the siRNA acting as a guide, targeting mRNA which share a complementary sequence (Majumdar et al., 2017). Once base pairing occurs, Argonaute (AGO) proteins cleave the target mRNA thus preventing transcription translation. This RNAi mechanism was first shown to be highly effective (Waterhouse et al., 1998) and has since been widely used for silencing endogenous and viral RNA in many plant species (Younis et al., 2014).

Limited reports of transient expression systems in cannabis exist. Recently, GFP has been transiently expressed in mesophyll

protoplasts of cannabis with over 20% transformation efficiency (Matchett-Oates et al., 2021b). *Agrobacterium*-mediated transformation protocols have previously been used for the stable transformation of hairy roots cultures to express β-glucuronidase (GUS; Wahby et al., 2013) and expression of phosphomannose isomerase (PMI) in friable callus (Feeney and Punja, 2003). More recently, transient RNAi *Agrobacterium*-mediated transformation of cannabis has been reported (Schachtsiek et al., 2019). Virus-induced gene silencing, utilizing *Cotton leaf crumple virus* (CLCrV), showed transcriptional silencing in virus affecting genes. Optimization of variables involved in transient *Agrobacterium*-mediated transformation has also been explored using heterologous expression of GUS and GFP in multiple tissue types (Deguchi et al., 2020). To our knowledge, this article is the first to report the use of transient expression RNAi constructs in cannabis to silence the medically important cannabinoid biosynthesis genes. The interaction between the highly homologous genes and the ability to silence all related genes using a single construct is also described. Successful silencing of the conserved homologous biosynthesis genes enables us to unravel gene function and their relationships within this important biosynthetic pathway.

MATERIALS AND METHODS

Plant Material and Growth Conditions

All research was performed under Medicinal Cannabis Research Licence (RL011/18) and Permit (RL01118P4) issued through the Department of Health (DoH), Office of Drug Control (ODC) Australia.

Leaf material from the *C. sativa* cultivar "Cannbio-2" (1,1.8, THC,CBD) was used for transient expression experiments. Cannbio-2 plants were propagated in 9-L plastic pots using coco-coir and grown using hydroponics nutrients coco A+B (THC®, Australia) as per manufacturer's recommended nutrient strength, in a controlled greenhouse environment at 25°C day time temperature, 20°C night time temperature, 50-60% humidity. Leaf explants were chosen from young, newly developing shoot apical meristems from the top half of the plant. Leaf explants were chosen from young, newly developing shoot apical meristems from the top half of the plant on approximately 2-month-old donor plants grown under high pressure sodium grow lights (Papillon, Holland), 500 $\mu mol\, m^{-2} s^{-1}$, with a photoperiod of 18-h light and 8-h dark regime.

Identification of Candidate Genes, siRNA Design, and Gene Amplification

Sequence data of the endogenous *THCAS*, *CBDAS,* and *CBCAS* genes were accessed from the Cannbio-2 genome assembly (Braich et al., 2020; https://doi.org/10.46471/gigabyte.10; BioProject: PRJNA667278). *THCAS*, *CBDAS,* and *CBCAS* gene sequences were determined by BLAST querying the Cannbio-2 genome assembly with an e-value threshold set at $<10^{-10}$. Exons from the gene sequences were predicted using FGENESH (Solovyev et al., 2006) and ExPASy (Gasteiger et al., 2003). Predicted gene sequences were viewed and aligned using Geneious Prime 2020.2.[1] siRNAs from amplified gene sequences were predicted using pssRNAit,[2] using the software's recommended parameters, to generate a library of siRNA fragments within the chosen gene sequences (**Supplementary Data**). The number of predicted off-target sites within the Cannbio-2 cannabinoid biosynthesis genes was performed by BLASTn analysis of each siRNA sequence, recording the total number of exact sequence homology matches, with off-targeting determined as an exact sequence residing within a different biosynthesis gene set. In the instance of pRNAi-GG-*CBDAS-UNIVERSAL* an off-target is defined as an exact match that does not reside within the CBDAS-truncated#4 homolog.

Primers were designed, using Primer3 (Untergasser et al., 2012), in gene regions of sequence variance and homology, with products between ~250 and ~600 base pairs for siRNA generation *in vivo* (**Supplementary Data**). Each forward and reverse primer had the 5' adapter sequences "acca ggtctc aggag" and "acca ggtctc atcgt," respectively. DNA fragments were PCR-amplified from Cannbio-2 genomic DNA, using Phusion polymerase (New England Biolabs, Ipswich, MA) with PCR cycling as follows: 98°C 30 s, 35 cycles of 98°C 10 s, 60°C 30 s, 72°C 30 s, and final extension 72°C 10 min.

Plasmid Construction, *Agrobacterium* Culture Conditions, and Vacuum Infiltration

For expression of siRNAs, pRNAi-GG vector was used within this study. pRNAi-GG was provided by The Arabidopsis Biological Resource Center (TAIR). The construction of the vectors containing gene sequences of interest was followed according to a previously published protocol (Yan et al., 2012). Briefly, 50 ng of purified PCR products was mixed with 200 ng of pRNAi-GG with 5 units of Bsal (New England Biolabs, Ipswich, MA) and 10 units of T4 Ligase (New England Biolabs, Ipswich, MA) in a total volume of 20 µl in T4 ligation buffer. Restriction-ligation was carried out at 37°C for 2 h followed by a final digestion at 50°C for 5 min and heat inactivation at 80°C for 5 min. *E. coli* DH5α competent cells were transformed with 5 µl of the mixture and plated on LB media containing 25 mg/L kanamycin and 5 mg/L chloramphenicol.

Recombinant bacterial colonies were PCR verified with primers flanking the PCR product insert, and bands were visualized using a TapeStation 2200 (Agilent, Santa Clara, CA) with colonies of expected band sizes sequence verified. Final constructs were labelled pRNAi-GG-*THCAS*, pRNAi-GG-*CBDAS*, pRNAi-GG-*CBCAS,* and pRNAi-GG-*CBDAS-UNIVERSAL* (**Supplementary Data**).

Recombinant *Agrobacterium tumefaciens* strains were generated *via* electroporation following a previously published protocol (Lin, 1995). *Agrobacterium* culture conditions and vacuum infiltration protocols were performed using a previously reported protocol (Deguchi et al., 2020) with slight modifications.

[1]https://www.geneious.com
[2]https://plantgrn.noble.org/pssRNAit/

In summary, for the expression of pRNAi-GG constructs, *A. tumefaciens* strain GV3101 was used for transient expression experiments. Recombinant *A. tumefaciens* were inoculated and grown in YM media (0.5 g/L K_2HPO_4, 0.2 g/L $MgSO_4$-$7H_2O$, 0.1 g/L NaCl, 10 g/L mannitol, 0.4 g/L yeast extract, PH 7; Sigma-Aldrich, St. Louis, MO) overnight at 220 rpm at 30°C. The culture was centrifuged at 4,000 *g* for 10 min and resuspended to an $OD_{600} = 0.5$ in infiltration media (10 mM MES, 1x MS and vitamins, 2% glucose, 200 μM acetosyringone, pH 5.6; Sigma-Aldrich, St. Louis, MO) and placed on a rotary shaker (Ratek, Australia) for 2 h prior to vacuum infiltration. Immediately before infiltration, 5 mM ascorbic acid, 0.05% (v/v) Pluronic F-68, and 0.015% (v/v) Silwet L-77 (Sigma-Aldrich, St. Louis, MO) was added to the *A. tumefaciens* culture.

Leaf segments (approx. 2 cm × 2 cm) were taken from young fully expanded leaves of *ca.* 2-month-old, donor Cannbio-2 plants and placed in a Petri dish (100 mm × 15 mm) containing *A. tumefaciens* suspension. The Petri dish was then placed in a desiccator (Tarsons, West Bengal, India) for 2 min at 400 mbar with vacuum pressure gently released. Vacuum was reapplied once more allowing thorough infiltration. Leaf material was washed with sterile water and transferred onto moist (ddH_2O) filter paper (Whatman, Maidstone, United Kingdom) in a Petri dish and placed in a controlled environment room at 24°C with an 18 h photoperiod for 4 days.

Quantitative Real-Time PCR Analysis of Agroinfiltrated Leaf Segments

Seventy-two hours post-vacuum agroinfiltration, leaf segments were snap frozen in liquid nitrogen and total RNA was extracted following manufacturer's instructions (RNeasy Plant Mini Kit, Qiagen, Hilden, Germany). cDNA synthesis and qPCR were carried out in one step with Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Ipswich, MA) following manufacturer's instructions. Quantitative PCR parameters used were as follows: 95°C for 60 s, 40 cycles at 95°C for 15 s, and 59°C for 15 s carried out with a CFX-96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA). Melting curves were measured, and gene expression levels were calculated from the cycle threshold according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Paired t test was performed ($p = 0.05$) to determine significance using RStudio (version 1.1.453, RStudio Inc., Boston, MA). The *UBQ5* gene was used as an internal reference (Deguchi et al., 2020), with three biological replicates used for all qPCR experiments with

TABLE 1 | Cannbio-2 analysis of cannabinoid biosynthesis genes with PCR amplification, copy number, and siRNA prediction information.

Cannabinoid biosynthesis gene	Accession number/Source of query	Copy number/ homologs	Primer pairs used for amplification	Product size	Predicted siRNA #
THCAS	AB057805	1	F: AACTATTTTATGCTCTAAGAAAGT R: TTTGTTATGAAGTGAGTCATGA	603 bp	93
CBDAS	AB292682	9	F: AAGTCCCATTTGTTATAGTAGA R: TTGACAAGCTCATGTATCTC	442 bp	70
CBCAS	Publication number: WO/2015/196275	3	F: GGCCAGTATTCTCTGCTC R: CTAGTTCTGAAGTGAGTCGTG	606 bp	95
CBDAS-UNIVERSAL	-	-	F: CCGGAGCTACCCTT R: GGCTATACGTGGTGG	247 bp	38

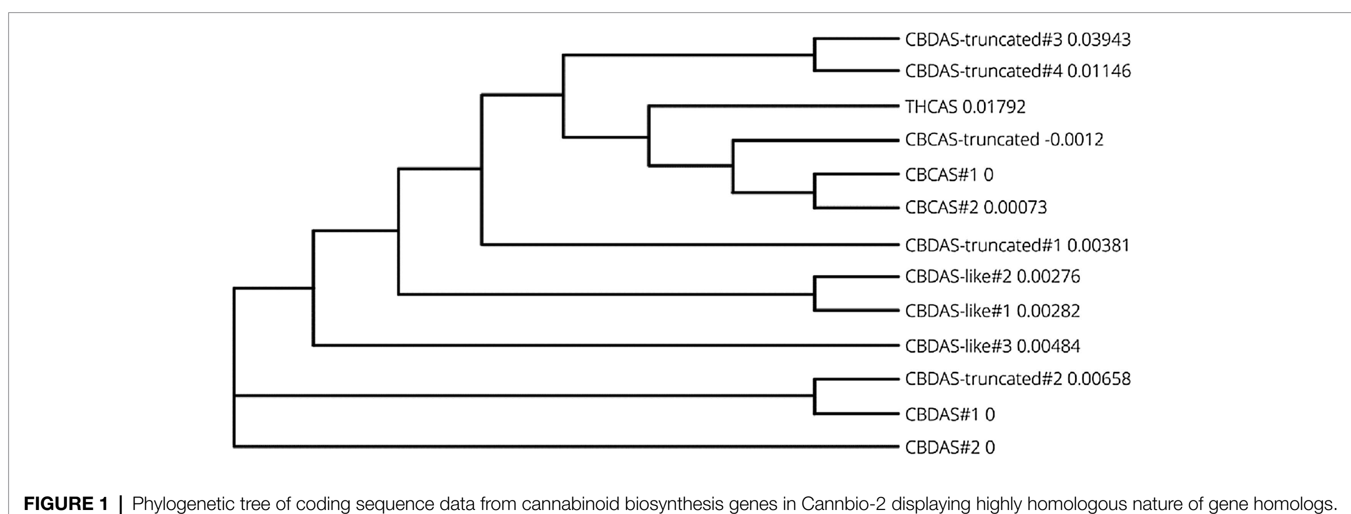


FIGURE 1 | Phylogenetic tree of coding sequence data from cannabinoid biosynthesis genes in Cannbio-2 displaying highly homologous nature of gene homologs.

TABLE 2 | Identity matrix of cannabinoid biosynthesis genes coding sequences in Cannbio-2 global alignment.

	THCAS	CBCAS#1	CBCAS#2	CBCAS-truncated	CBDAS-like#1	CBDAS-like#2	CBDAS-like#3	CBDAS-like#4	CBDAS-like#5	CBDAS-truncated#1	CBDAS-truncated#2	CBDAS-truncated#3	CBDAS-truncated#4
THCAS													
CBCAS#1	96.31												
CBCAS#2	96.23	99.93											
CBCAS-truncated	95.88	99.90	99.79										
CBDAS-like#1	93.10	92.38	92.29	91.83									
CBDAS-like#2	93.10	92.38	92.29	91.83	100.00								
CBDAS-like#3	92.98	92.28	92.20	91.71	99.51	99.51							
CBDAS-like#4	92.60	92.03	91.96	91.44	99.51	99.51	99.44						
CBDAS-like#5	92.37	91.64	91.56	91.03	99.49	99.49	99.26	99.11					
CBDAS-truncated#1	92.19	92.74	92.74	52.19	98.96	98.96	98.44	98.96	98.54				
CBDAS-truncated#2	72.86	71.35	71.28	61.84	74.38	74.38	74.80	76.62	75.72	98.44			
CBDAS-truncated#3	86.79	86.66	86.66	86.42	87.01	87.01	86.90	86.90	86.87	84.74	56.28		
CBDAS-truncated#4	67.78	66.70	66.60	62.03	66.14	66.14	67.17	67.59	67.39	90.00	82.72	57.83	

two technical replicates. All primer sequences are listed in **Supplementary Data**.

RESULTS

Identification of Cannabinoid Genes and siRNAs Prediction

To establish RNAi in *C. sativa*, *THCAS*, *CBDAS*, and *CBCAS* gene sequences were determined by BLAST querying the Cannbio-2 genome sequence assembly with publicly available sequences (**Table 1**). Each cannabinoid biosynthesis gene, and accompanying homologs, were analyzed for functionality using FGENESH and ExPASy and subsequently BLASTn analyzed for homology to publicly available sequences and pairwise aligned using MUSCLE to create a phylogenetic tree (**Figure 1**) and a matrix with identity percentages of coding sequences (**Table 2**).

Within the Cannbio-2 genome, a single functional copy of *THCAS* exists; however, *CBDAS* and *CBCAS* contain nine and three homologs/pseudogenes, respectively. Using FGENESH and ExPASy, two identical, full-length potentially functional *CBDAS* cannabinoid biosynthesis genes were discovered (CBDAS-like#1 and #2), and three homologs were identified containing several single nucleotide polymorphisms (SNPs) leading to differences in predicted protein translations (CBDAS-like#3-5), however full length and potentially functional, and four copies of *CBDAS* were found to be truncated when proteins were predicted (CBDAS-truncated#1-4). The coding sequences (CDS) of each *CBDAS* homologs were aligned, and non-truncated homologs are shown to be >86% homologous. The high levels of sequence similarity of the *CBDAS* homologs (**Table 2**) at the DNA level, and regardless of the size of the PCR insert for siRNA generation, sequence homology is too significant to identify one best-fit homolog for vector design, and thus, a single homolog of *CBDAS* was chosen, identified as CBDAS-like#1 within the Cannbio-2 genome (**Supplementary Data**), for pRNAi-GG-*CBDAS* vector construction.

Two full-length, potentially functional copies of *CBCAS* were found (CBCAS-like#1 and #2) having identical sequence homology, except for base pair 482, where a synonymous SNP occurs (T to C); however, this does not affect predicted translated proteins (**Supplementary Data**). A truncated *CBCAS* homolog was also discovered at only 969 bp designated CBCAS-truncated. CBCAS#2 was chosen within the Cannbio-2 genome for pRNAi-GG-*CBCAS* vector construction (**Supplementary Data**). A significantly smaller sequence (247 bp; **Supplementary Data**), homologous to the CBDAS-truncated#4 homolog, was chosen in a region of high homology from the sequence alignment of all cannabinoid synthesis genes CDS, however lower in homology (<90%) within the subset of *CBDAS* sequences, designated "*CBDAS-UNIVERSAL*" to determine whether a smaller gene sequence for RNAi containing lower homology could be more effective in gene silencing through off-targeting. A graphic representation for the alignment of cannabinoid biosynthesis genes, with the PCR products sizes, is shown in **Figure 2**.

The gene sequences selected for RNAi were analyzed using pssRNAit to assess the degree of off-targeting to the identified

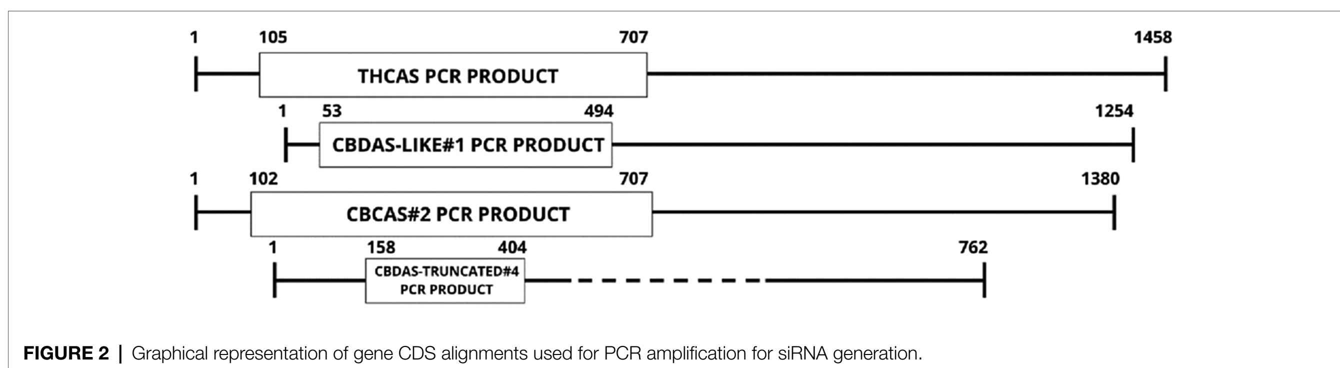


FIGURE 2 | Graphical representation of gene CDS alignments used for PCR amplification for siRNA generation.

cannabinoid gene sequences for each specific vector. Efficient gene silencing requires the formed siRNA to contain minimal off-targeting silencing effects. From the amplified *THCAS* sequence, 93 siRNA were predicted with 1,609 potential off-targets, *CBDAS* with 70 predicted siRNA and 1,609 potential off-targets, *CBCAS* with 95 predicted siRNA and 1,647 potential off-targets, and *CBDAS-UNIVERSAL* with 38 predicted siRNA with 630 potential off-targets (**Table 1**).

To filter out irrelevant off-target sites not residing within the cannabinoid genes, each siRNA was aligned to Cannbio-2 cannabinoid biosynthesis genes for sequence similarity to greater understand off-targeting potential within these highly homologous sequences. A total number of 369 exact targets for pRNAi-GG-*THCAS* exist within Cannbio-2 cannabinoid biosynthesis genes with 93 exact matches to *THCAS* and 276 off-targets existing within the other gene sets (**Table 3**). pRNAi-GG-*CBDAS* contained 447 total exact targets within all biosynthesis genes, with 381 targeting a minimum of 1 *CBDAS* homologs and containing considerably more off-targets tallying 64 sites not residing within *CBDAS* homologs (**Table 3**). pRNAi-GG-*CBCAS* contained a similar number of total targets, 428, with 276 targets within CBCAS homologs and contained substantially more off-targets, with 152 exact matches across other gene sets (**Table 3**). Within the pRNAi-GG-*CBDAS-UNIVERSAL* predicted siRNA, only 69 exact targets exist within all biosynthesis genes. A total of 38 siRNA sites exist within the predicted CBDAS-truncated#4 gene sequence, with the remaining 31 target sites residing within *CBDAS* homologs (**Table 3**).

Vector Construction, Generation of Recombinant *Agrobacterium,* and Vacuum Infiltration

To test the efficiency of silencing cannabinoid biosynthesis genes, recombinant expression vectors were made for the four target sequences. The vectors contained sense-antisense orientation separated by an intron and were cloned into an *E. coli* strain.

Eight recombinant colonies were chosen, for each treatment, for colony PCR using sequence-specific primers residing within the specific sequence and residing on the vector backbone. All clones showed the expected bands confirming the correct inserts, which were subsequently sequenced to confirm the correct sequences as expected.

Agrobacterium strain, GV3101, was chosen for *Agrobacterium*-mediated transient expression in leaf segments of Cannbio-2. Recombinant pRNAi-GG vectors were transformed into GV3101 with appropriate selection. Agroinfiltration was achieved using vacuum infiltration on the excised cannabis leaf segments optimized for use with Cannbio-2 leaf material.

Silencing of Cannabinoid Biosynthesis Genes

Leaf segments of *C. sativa* Cannbio-2 strain were infiltrated with recombinant *A. tumefaciens* and incubated in a climate-controlled environment. To investigate the extent of downregulation of the cannabinoid biosynthesis genes, quantification of the transcript levels of *THCAS*, *CBDAS,* and *CBCAS* was performed using qPCR. Each genes expression level was analyzed in three biological replicates and two technical replicates with gene primer pairs located upstream of the respective RNAi construct design.

Using the reference gene *UBQ5* for normalization in all qPCR experiments, infiltrated leaf segments saw varying levels of downregulation in all cannabinoid biosynthesis genes, and in one instance, upregulation of *THCAS* and *CBCAS* in response to RNAi transient expression compared to leaf segments infiltrated with disarmed *Agrobacterium* as negative controls.

Agroinfiltration with pRNAi-GG-*THCAS* successfully downregulated *THCAS*, *CBDAS,* and *CBCAS*. From the qPCR data, pRNAi-GG-*THCAS* saw a 57% reduction in *THCAS* transcript levels (**Figure 3A**). Interestingly, using the *THCAS* gene sequence for RNAi, between the vectors, was ranked the 3rd most effective for downregulating the targeted gene. Off-targeting of this vector construct caused downregulation of *CBDAS* with a 71% reduction (non-significant, $p = 0.48$) in transcript levels making this, also, the 3rd most effective in downregulating *CBDAS.* The highly homologous sequence of *CBCAS* saw a more conserved reduction of 39% (non-significant, $p = 0.45$) in transcript levels, with the off-targeting effect of this vector ranking it also third in silencing *CBCAS*.

Agroinfiltration with pRNAi-GG-*CBDAS* downregulated the three cannabinoid biosynthesis genes more effectively, comparatively. The pRNAi-GG-*CBDAS* vector saw a significant ($p < 0.05$) reduction of CBDAS with 92% downregulation (**Figure 3B**), making it the second most effective vector for downregulation of *CBDAS* behind

pRNAi-GG-*CBDAS-UNIVERSAL*. Significant ($p < 0.05$) off-target downregulation of *THCAS* was observed with a 77% reduction in transcript levels, making this more efficient in inadvertent downregulation of *THCAS* than using the gene-specific sequence of *THCAS* to produce siRNA. Increased downregulation of *CBCAS* was also observed, with a 53% reduction (non-significant, $p = 0.07$) in transcript levels compared to the control, making this vector the second most effective construct for silencing *CBCAS*.

Agroinfiltration with pRNAi-GG-*CBCAS* was least effective in silencing cannabinoid biosynthesis genes, which conversely saw non-significant upregulation of *THCAS* and *CBCAS* transcript levels (**Figure 3C**). pRNAi-GG-*CBCAS* saw a 39% downregulation (non-significant, $p = 0.22$) of *CBDAS*, making it the least effective vector for *CBDAS* silencing. Interestingly, off-targeting caused *THCAS* to be upregulated by 13% (non-significant, $p = 0.42$) compared to the control regardless of the >96% homology shared between the two genomic sequences. This increase makes this the least effective vector for *THCAS* silencing. *CBCAS* transcript levels were significantly ($p < 0.05$) upregulated 76% using the targets gene sequence, rendering it least effective of all vectors for gene silencing of *CBCAS*.

Agroinfiltration with pRNAi-GG-*CBDAS-UNIVERSAL* was significantly more efficient in downregulating *THCAS*, *CBDAS*, and *CBCAS*. The small construct, homologous to a highly conserved region of the aligned gene sequences, saw comparatively dramatic decreases of transcript levels compared to the other constructs (**Figure 3D**). A significant ($p < 0.05$) downregulation of *THCAS*, with a 92% reduction in transcript levels, was observed due to off-targeting, making this vector highly effective in targeting *THCAS*. A significant ($p < 0.05$) reduction in *CBDAS* was also observed, with a 97% reduction in transcript levels compared to the control. Like pRNAi-GG-*CBDAS*, pRNAi-GG-*CBDAS-UNIVERSAL* is most effective in silencing the targeted gene used to create the vector construct (CBDAS-truncated#4), also making this smaller construct the most effective in downregulating *CBDAS*. Significant ($p < 0.05$) reduction in *CBCAS* was also observed, with a 70% decrease in transcript levels compared to the control. This off-targeting effect makes this vector the most effective in silencing *CBCAS* compared to the other vectors.

TABLE 3 | Off-targeting frequency in each cannabinoid gene from generated siRNA in each vector.

RNAi vector	Cannabinoid biosynthesis genes												
	THCAS	**CBCAS#1**	**CBCAS#2**	**CBCAS-truncated**	**CBDAS-like#1**	**CBDAS-like#2**	**CBDAS-like#3**	**CBDAS-like#4**	**CBDAS-like#5**	**CBDAS-truncated#1**	**CBDAS-truncated#2**	**CBDAS-truncated#3**	**CBDAS-truncated#4**
pRNAi-GG-*THCAS*	93	53	53	51	18	18	16	23	17	0	19	0	6
pRNAi-GG-*CBDAS*	18	16	16	14	64	64	70	54	54	0	60	0	15
pRNAi-GG-*CBCAS*	53	95	95	86	16	16	16	18	12	0	16	0	5
pRNAi-GG-*CBDAS-UNIVERSAL*	0	0	0	0	5	5	5	5	5	0	5	1	38

DISCUSSION

Genetic transformation of cannabis has only recently been achieved using *Agrobacterium* (Schachtsiek et al., 2019; Deguchi et al., 2020). Induced RNA silencing by hairpin-loop RNAi constructs have previously been optimized through the exploration of variables involved in vacuum infiltration by measuring relative GUS expression (Deguchi et al., 2020). Building upon the approach developed by Deguchi et al. (2020), vacuum infiltration was achieved in leaf segments of Cannbio-2, a cultivar with a ratio 1:1.8 THC to CBD, to significantly reduce the relative expression of cannabinoid biosynthesis genes *THCAS*, *CBDAS,* and *CBCAS*. This work is the first successful downregulation of these cannabinoid biosynthetic genes, showing that the use of RNAi constructs

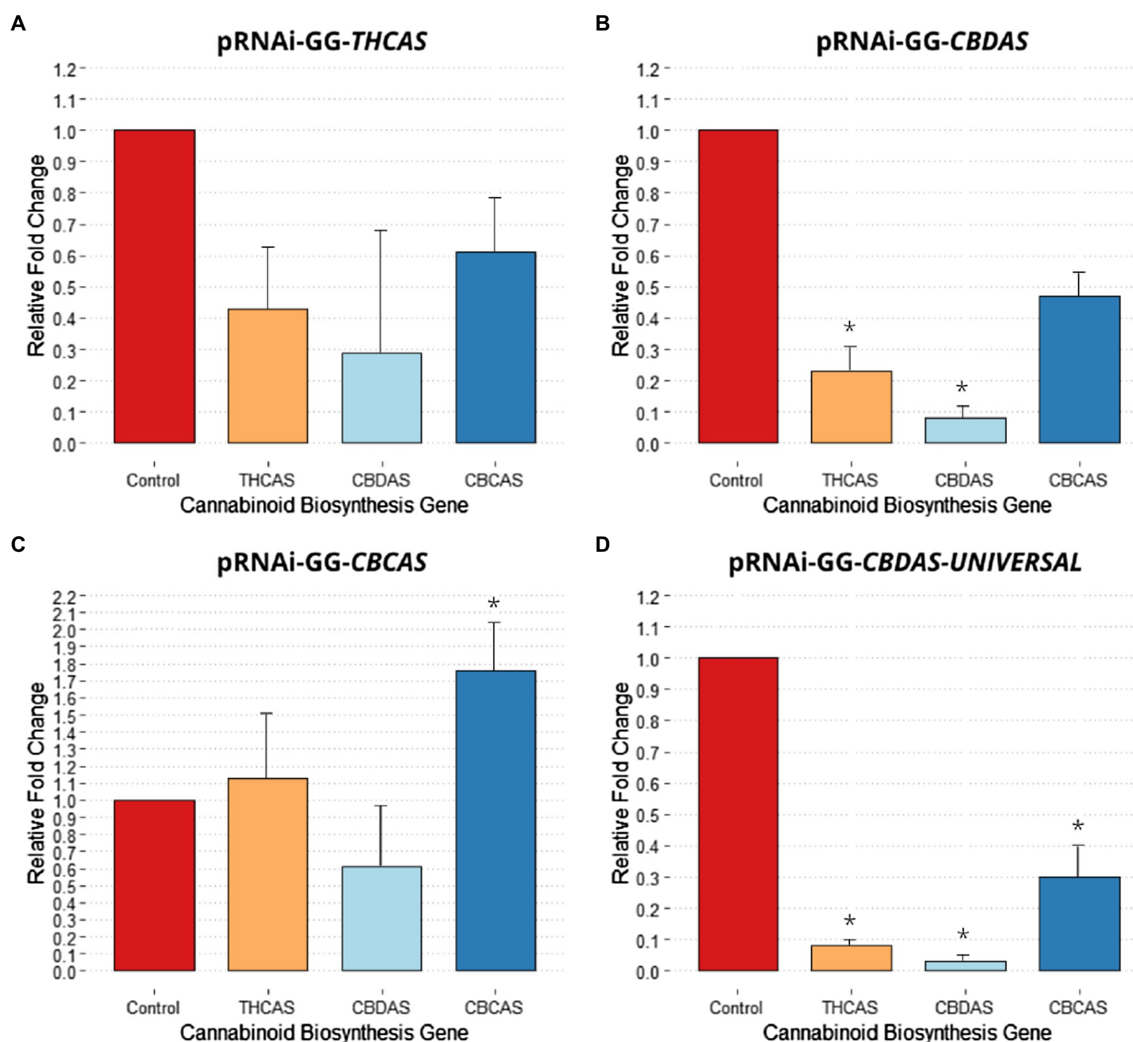


FIGURE 3 | Effect of different pRNAi-GG vectors on cannabinoid biosynthesis gene relative expression change. **(A)** Relative fold change post agroinfiltration with pRNAi-GG-*THCAS*. **(B)** Relative fold change post agroinfiltration with pRNAi-GG-*CBDAS*. **(C)** Relative fold change post agroinfiltration with pRNAi-GG-*CBCAS*. **(D)** Relative fold change post agroinfiltration with pRNAi-GG-*CBDAS-UNIVERSAL*. Significance is determined by paired *t*-test, ($p < 0.05$) is denoted by *. Error bars represented SE.

with the gene sequences of each gene, respectfully, results in varying levels of suppression.

In this paper, the downregulation of cannabinoid biosynthesis genes was evaluated using vacuum agroinfiltration. Using the common Golden Gate Cloning method to construct RNAi vectors, with sense and antisense sequence inserts, downregulation of *THCAS*, *CBDAS,* and *CBCAS* was observed to varying effectiveness. In this study, it was hypothesized that using large (400–600 bp) RNAi constructs to silence-specific cannabinoid biosynthesis genes would result in a downregulation of the other highly homologous gene sequences due to siRNA off-targeting. Observing the relative transient expression levels of the targeted genes 4 days post-agroinfiltration with pRNAi-GG-*THCAS* saw a downregulation of 57, 71 and 39% of *THCAS*, *CBDAS,* and *CBCAS*, respectfully (**Figure 3A**). The siRNA generated using pRNA-GG-*THCAS* targeted substantially more regions within *THCAS* and *CBCAS* compared to *CBDAS* (**Table 2**). While the results were all non-significant due to the variance between treated samples, off-targeting is still prevalent as demonstrated by the ability to downregulate non-specific targets. This confirmation of the hypothesis can be explained by the highly homologous (>90%) gene sequences, which when amplified and used in RNAi, will produce siRNA (**Table 1**) that will have significant off-targeting. siRNA predicted from the amplified *THCAS* sequence were more effective in downregulating the *CBDAS* transcripts, comparatively, to *THCAS* and *CBCAS*, which are more highly sequence homologous (>96%) than *CBDAS* is to *THCAS* (92%). The most likely explanation for this increased downregulation of *CBDAS* would be the fact that Cannbio-2 contains 5 potentially functional copies. Within the Cannbio-2 genome (Braich et al., 2020), a fully functional *CBDAS* gene

is absent due to assembly error within the retrotransposon regions in a hybrid genotype. Cannbio-2 does contain an identical *CBDAS* gene within the transcriptome (Braich et al., 2019; Cannbio_016865); however, this is not present within the genome. However, several full-length, potentially functional *CBDAS* homologs exist in which their function is yet to be determined. The increased copy number of *CBDAS* is due to the cannabinoid biosynthesis genes being arranged in tandem arrays in long terminal repeat retrotransposons on chromosome 7 (Grassa et al., 2021). The flanking long terminal repeats for *CBDAS* provide an explanation for the movement of the synthase cassette and possible illegitimate recombination resulting in increased synthase numbers. This increased copy numbers will greatly affect RNAi specificity and will result in a higher number of off-targeting sites.

pRNAi-GG-*CBDAS* agroinfiltration qPCR data show significant ($p < 0.05$) downregulation in *CBDAS*, with a reduction of 92% (**Figure 3B**). Increased downregulation, compared to pRNAi-GG-*THCAS*, was also observed for *THCAS* and *CBCAS*, with 77% ($p = 0.03$) and 53% ($p = 0.07$), respectfully. The presence of 3 *CBCAS* homologs results in a higher number of potential exact targets compared to *THCAS* (**Table 3**); however, downregulation is twice as effective in *THCAS* than *CBCAS*. Within the genomic sequences and alignment of these two genes and their high level of sequence similarity, it could be expected that the siRNA generated would not contain greater affinity for *THCAS,* but instead downregulate *CBCAS* further due to increased target sites. This, however, is not observed. The increased downregulation despite lower off-target site numbers could be due to the generation of more efficacious siRNAs, which regardless of off-targeting, demonstrate the capability of inhibiting transcription with target sequence variation.

Shorter PCR products for RNAi could also potentially explain higher siRNA efficacy in silencing cannabinoid biosynthesis genes compared to larger inserts. Support of this hypothesis is provided by the qPCR data from agroinfiltration of pRNAi-GG-*CBDAS-UNIVERSAL*, a 247 bp fragment, which produced significant ($p < 0.05$) reduction in *THCAS*, *CBDAS,* and *CBCAS* (**Figure 3D**). The smaller RNAi construct reduced *THCAS*, *CBDAS,* and *CBCAS* by 92, 97, and 70%, respectfully. Increased efficacy of shorter dsRNA fragments has previously been confirmed in potato (He et al., 2020), with evidence supporting shorter dsRNA length resulting in increased levels of insecticidal protection compared to the larger RNAi constructs investigated. On the contrary, within *Arabidopsis* plants expressing RNAi dsRNA constructs with varying length, there was no observed significant correlation between dsRNA length and reduction of *Fusarium graminearum* infection (Höfle et al., 2020). These studies suggest that within *Cannabis* the effect of dsRNA length and specific region of the gene targeted (e.g., earlier exons) could play a vital role in efficacy, though such assumptions require further investigation and testing.

An additional explanation for the higher efficacy of pRNAi-GG-*CBDAS-UNIVERSAL* is the concentration of more highly effective siRNA, within the shorter sequence, compared to larger fragments which could contain lower efficiency siRNAs. Despite the recent surge in cannabis genome sequencing efforts, the lack of detailed genome sequence annotations and tools to correctly assess the potential for off-targeting of predicted siRNA to the highly homologous cannabinoid biosynthesis gene sequences, as such with the prediction tool "pssRNAit," requires further investigation. Without the availability of a comprehensive *Cannabis* genome sequence resource to detect the potential off-targeting of these highly homologous genes, the exact sequences of each siRNA were aligned against the Cannbio-2 gene sets and analyzed for off-targeting potential. The limitation of this approach is the inability to correctly evaluate all possible off-targets when slight siRNA sequence variation exists due to the highly homologous nature of all the cannabinoid biosynthesis genes. However, regarding exact siRNA sequence matches residing outside of the intended target, a large number of predicted siRNA produced from pRNAi-GG-*THCAS, CBDAS,* and *CBCAS* exists. Though, interestingly, no exact matches outside of the *CBDAS* homologs are present within any of predicted pRNAi-GG-*CBDAS-UNIVERSAL* siRNA (**Table 2**). The lower concentration of exact siRNA targets could increase the efficacy of each siRNA, explained by the significant downregulation of *CBDAS*, but it does not explain how this construct is equally capable of significantly downregulating all the highly homologous genes. It is evident that significant off-targeting occurs; however, many base pair differences are tolerated in siRNA targeting is undetermined. Previous work has determined that it is not only the amount of mismatches but also the identity of the matched nucleotides that play an important role in unintended silencing (Du et al., 2005). It was discovered that adenine and cytosine, along with G:U wobble base pair mismatches are silenced with equal efficiency. With these gene sequences being so highly homologous (**Figure 1**), it is highly probable this would explain the success of pRNAi-GG-*CBDAS-UNIVERSAL.*

Interestingly, contradictory to the proposed hypothesis of collective downregulation of all targeted genes, pRNAi-GG-*CBCAS* agroinfiltration resulted in significant upregulation of *CBCAS* and an observed slight increase in *THCAS*. The 95 predicted siRNAs had a total of 329 exact matches between the CBCAS homologs and THCAS and only 94 matches within the CBDAS homologs resulting increase of 13% in transcript levels of *THCAS* and 76% increase of *CBCAS* transcript levels and a decrease of 39% in *CBDAS* (**Figure 3D**). An explanation for the upregulation could be the specific sequence containing inefficient siRNA or that the siRNA which did downregulate *CBDAS* triggers a biological response to upregulate the highly similar genes to assist in the enzymatic conversion of CBGA. Alternatively, it is possible that the siRNA generated failed to degrade the mRNA and instead interfered with the translation of THCAS and CBCAS, triggering a feedback loop mechanism leading to increased levels of transcription of these two genes. Examples of such a phenomenon have been observed in mammalian cells (Portnoy et al., 2011; Scacheri et al., 2004) and in wheat lines with RNAi resulting in a compensatory effect increasing total protein content (Gil-Humanes et al., 2008). To date, there are no examples of complete knockdown of

individual cannabinoid biosynthesis genes *in vivo* to confirm that specific enzymes can synthesize different cannabinoids. However, multiple cannabinoids have been produced from a single coding sequence of *CBCAS* in yeast through modulating yeast growth conditions (Peet et al., 2016).

Using RNAi to significantly downregulate the medicinally important cannabinoid biosynthesis genes can be achieved using *Agrobacterium*. Much like Deguchi et al. (2020) and Schachtsiek et al. (2019), the use of RNAi in Cannabis to significantly downregulate targeted genes is shown to be possible using different RNAi mechanisms, such as the introduction of dsRNA or virus-induced gene silencing. The drawback from using RNAi to target these genes, and the others previously explored, is the unintended off-targeting, resulting in silencing of the other highly homologous genes. To completely and specifically downregulate a specific enzyme, a sequence-specific genome editing approach, such as CRISPR/Cas-9, would be more applicable by making a large library of constructs and events and then screening for a targeted single gene for knock out (Matchett-Oates et al., 2021a). This approach will allow the investigation into site-specific genome editing events, resulting in a complete knockdown, and whether *in vivo* feedback loops result in gene regulation, through upregulation, in these cannabinoid biosynthesis genes. The use of this agroinfiltration RNAi approach, generating a transformational event resulting in a designer cannabis strain with significantly reduced THC, CBD, and CBC concentrations, is possible. The decreased gene expression will potentially lead to a dramatic increase in the precursor CBGA, which is currently found in minute concentrations, comparatively (Stack et al., 2021). The targeted manipulation of the cannabinoid pathway in this manner could enable the future development of novel genetically modified cannabinoid strains that could deliver new therapeutics pending consumer acceptance of its biotechnology approach. The production of a transgenic cannabis plant using RNAi, in some countries, is not considered genetically modified (Office of the Gene Technology Regulator, 2018), addressing consumer concerns regarding genetic modifications of consumed products.

CONCLUSION

Reported within this study is the first downregulation of cannabinoid biosynthesis genes in cannabis using transiently expressed RNAi constructs in leaf segments. This evaluation of RNA silencing efficiency will help further unravel the relationship each cannabinoid biosynthesis gene has through detailed functional genomic screens. This approach can also play an important role in producing stably transformed *C. sativa* designer strains with modulated expression profiles of the medically important cannabinoid biosynthesis genes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LM-O, NC, and GS provided manuscript conception. LM-O performed data collection and analysis. NC and GS edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by Agriculture Victoria Research and Agriculture Victoria Services.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.773474/full#supplementary-material

REFERENCES

Borrelli, F., Fasolino, I., Romano, B., Capasso, R., Maiello, F., Coppola, D., et al. (2013). Beneficial effect of the non-psychotropic plant cannabinoid cannabigerol on experimental inflammatory bowel disease. *Biochem. Pharmacol.* 85, 1306–1316. doi: 10.1016/j.bcp.2013.01.017

Borrelli, F., Pagano, E., Romano, B., Panzera, S., Maiello, F., Coppola, D., et al. (2014). Colon carcinogenesis is inhibited by the TRPM8 antagonist cannabigerol, a *Cannabis*-derived non-psychotropic cannabinoid. *Carcinogenesis* 35, 2787–2797. doi: 10.1093/carcin/bgu205

Braich, S., Baillie, R., Jewell, L., Spangenberg, G. C., and Cogan, N. O. I. (2019). Generation of a comprehensive transcriptome atlas and transcriptome dynamics in medicinal cannabis. *Sci. Rep.* 9:16583. doi: 10.1038/s41598-019-53023-6

Braich, S., Baillie, R. C., Spangenberg, G. C., and Cogan, N. O. I. (2020). A new and improved genome sequence of *Cannabis sativa*. *Gigabyte* 1. doi: 10.46471/gigabyte.10 [Epub ahead of print]

Campbell, F. A., Tramèr, M. R., Carroll, D., Reynolds, D. J. M., Moore, R. A., and McQuay, H. J. (2001). Are cannabinoids an effective and safe treatment option in the management of pain? A qualitative systematic review. *BMJ* 323, 13–16. doi: 10.1136/bmj.323.7303.13

Deguchi, M., Bogush, D., Weeden, H., Spuhler, Z., Potlakayala, S., Kondo, T., et al. (2020). Establishment and optimization of a hemp (*Cannabis sativa* L.) agroinfiltration system for gene expression and silencing studies. *Sci. Rep.* 10:3504. doi: 10.1038/s41598-020-60323-9

Du, Q., Thonberg, H., Wang, J., Wahlestedt, C., and Liang, Z. (2005). A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.* 33, 1671–1677. doi: 10.1093/nar/gki312

ElSohly, M. A., Radwan, M. M., Gul, W., Chandra, S., and Galal, A. (2017). Phytochemistry of *Cannabis sativa* L. *Prog. Chem. Org. Nat. Prod.* 103, 1–36. doi: 10.1007/978-3-319-45541-9_1

Feeney, M., and Punja, Z. K. (2003). Tissue culture and agrobacterium-mediated transformation of hemp (*Cannabis sativa* L.). *In Vitro Cell. Dev. Biol. Plant* 39, 578–585. doi: 10.1079/IVP2003454

Gaoni, Y., and Mechoulam, R. (1964). Isolation, structure, and partial synthesis of an active constituent of hashish. *J. Am. Chem. Soc.* 86, 1646–1647. doi: 10.1021/ja01062a046

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563

Gil-Humanes, J., Pistón, F., Hernando, A., Alvarez, J. B., Shewry, P. R., and Barro, F. (2008). Silencing of γ-gliadins by RNA interference (RNAi) in bread wheat. *J. Cereal Sci.* 48, 565–568. doi: 10.1016/j.jcs.2008.03.005

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., et al. (2021). A new cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* 230, 1665–1679. doi: 10.1111/nph.17243

Grassa, C. J., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., et al. (2018). A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. BioRxiv 458083 [Preprint]. doi: 10.1101/458083

He, W., Xu, W., Xu, L., Fu, K., Guo, W., Bock, R., et al. (2020). Length-dependent accumulation of double-stranded RNAs in plastids affects RNA interference efficiency in the Colorado potato beetle. *J. Exp. Bot.* 71, 2670–2677. doi: 10.1093/jxb/eraa001

Höfle, L., Biedenkopf, D., Werner, B. T., Shrestha, A., Jelonek, L., and Koch, A. (2020). Study on the efficiency of dsRNAs with increasing length in RNA-based silencing of the Fusarium CYP51 genes. *RNA Biol.* 17, 463–473. doi: 10.1080/15476286.2019.1700033

Kogel, C. C., Lopez-Pelayo, H., Balcells-Olivero, M. M., Colom, J., and Gual, A. (2018). Psychoactive constituents of cannabis and their clinical implications: a systematic review. *Adicciones* 30, 140–151. doi: 10.20882/adicciones

Li, H.-L. (1973). An archaeological and historical account of cannabis in China. *Econ. Bot.* 28, 437–448. doi: 10.1007/BF02862859

Lin, J.-J. (1995). "Electrotransformation of agrobacterium," in *Electroporation Protocols for Microbiology*. ed. J. A. Nickoloff (Humana Press), 171–178.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567, 123–126. doi: 10.1038/s41586-019-0978-9

Machado Rocha, F. C., Stefano, S. C., De Cassia Haiek, R., Rosa Oliveira, L. M. Q., and Da Silveira, D. X. (2008). Therapeutic use of C*annabis sativa* on chemotherapy-induced nausea and vomiting among cancer patients: systematic review and meta-analysis. *Eur. J. Cancer Caree* 17, 431–443. doi: 10.1111/j.1365-2354.2008.00917.x

Majumdar, R., Rajasekaran, K., and Cary, J. W. (2017). RNA interference (RNAi) as a potential tool for control of mycotoxin contamination in crop plants: concepts and considerations. *Front. Plant Sci.* 8:200. doi: 10.3389/fpls.2017.00200

Magagnini, G., Grassi, G., and Kotiranta, S. (2018). The effect of light spectrum on the morphology and cannabinoid content of *Cannabis sativa* L. *Med. Cannabis Cannabinoids* 1, 19–27. doi: 10.1159/000489030

Matchett-Oates, L., Braich, S., Spangenberg, G., Rochfort, S., and Cogan, N. O. I. (2021a). In silico analysis enabling genome editing in medicinal cannabis, gene families and variant characterisation. *PLoS One* 16:e0257413. doi: 10.1371/journal.pone.0257413

Matchett-Oates, L., Mohamaden, E., Spangenberg, G., and Cogan, N. (2021b). Development of a robust transient expression screening system in protoplasts of cannabis. *In Vitro Cell. Dev. Biol. Plant* 1–11. doi: 10.1007/s11627-021-10178-0

Movahedi, M., Ghasemi-Omran, V., and Torabi, S. (2015). The effect of different concentrations of TDZ and BA on *in vitro* regeneration of Iranian cannabis (*Cannabis sativa*) using cotyledon and epicotyl explants. *J. Plant Mol. Breed.* 3, 20–27. doi: 10.22058/jpmb.2015.15371

Namdar, D., Charuvi, D., Ajjampura, V., Mazuz, M., Ion, A., Kamara, I., et al. (2019). LED lighting affects the composition and biological activity of *Cannabis sativa* secondary metabolites. *Ind. Crop. Prod.* 132, 177–185.

Office of the Gene Technology Regulator (2018). Technical Review of the Gene Tchnology Regulations 2001. Available at: https://www.ogtr.gov.au/sites/default/files/files/2021-07/decision_ris.docx (Accessed October 19, 2021).

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., and Mandolino, G. (2006). Genetics and marker-assisted selection of the chemotype in *Cannabis sativa* L. *Mol. Breed.* 17, 257–268. doi: 10.1007/s11032-005-5681-x

Peet, R., Sun, M., Winnicki, R., and Donsky, M. (2016). US9394510B2.

Portnoy, V., Huang, V., Place, R. F., and Li, L. (2011). Small RNA and transcriptional upregulation. *Wiley Interdiscip. Rev. RNA* 2, 748–760. doi: 10.1002/wrna.90

Rog, D. J., Nurmikko, T. J., Friede, T., and Young, C. A. (2005). Randomized, controlled trial of cannabis-based medicine in central pain in multiple sclerosis. *Neurology* 65, 812–819. doi: 10.1212/01.wnl.0000176753.45410.8b

Russo, E. B. (2017). Cannabis and epilepsy: an ancient treatment returns to the fore. *Epilepsy Behav.* 70, 292–297. doi: 10.1016/j.yebeh.2016.09.040

Saloner, A., and Bernstein, N. (2021). Nitrogen supply affects cannabinoid and terpenoid profile in medical cannabis (*Cannabis sativa* L.). *Ind. Crop. Prod.* 167:113516

Scacheri, P. C., Rozenblatt-Rosen, O., Caplen, N. J., Wolfsberg, T. G., Umayam, L., Lee, J. C., et al. (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci.* 101, 1892–1897. doi: 10.1073/pnas.0308698100

Schachtsiek, J., Hussain, T., Azzouhri, K., Kayser, O., and Stehle, F. (2019). Virus-induced gene silencing (VIGS) in *Cannabis sativa* L. *Plant Methods* 15:157. doi: 10.1186/s13007-019-0542-5

Shiponi, S., and Bernstein, N. (2021). The highs and lows of P supply in medical cannabis: effects on cannabinoids, the ionome, and morpho-physiology. *Front. Plant Sci.* 12:657323. doi: 10.3389/fpls.2021.657323

Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7, S10.1–S10.12. doi: 10.1186/gb-2006-7-s1-s10

Stack, G. M., Toth, J. A., Carlson, C. H., Cala, A. R., Marrero-González, M. I., Wilk, R. L., et al. (2021). Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy* 13, 546–561. doi: 10.1111/gcbb.12793

Touw, M. (1981). The religious and medicinal uses of cannabis in China, India and Tibet. *J. Psychoactive Drugs* 13, 23–34. doi: 10.1080/02791072.1981.10471447

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115–e115. doi: 10.1093/nar/gks596

Wahby, I., Caba, J. M., and Ligero, F. (2013). Agrobacterium infection of hemp (*Cannabis sativa* L.): establishment of hairy root cultures. *J. Plant Interact.* 8, 312–320. doi: 10.1080/17429145.2012.746399

Waterhouse, P. M., Graham, M. W., and Wang, M.-B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci.* 95, 13959–13964. doi: 10.1073/pnas.95.23.13959

Weiblen, G. D., Wenger, J. P., Craft, K. J., ElSohly, M. A., Mehmedic, Z., Treiber, E. L., et al. (2015). Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytol.* 208, 1241–1250. doi: 10.1111/nph.13562

Welling, M. T., Liu, L., Kretzschmar, T., Mauleon, R., Ansari, O., and King, G. J. (2020). An extreme-phenotype genome-wide association study identifies candidate cannabinoid pathway genes in Cannabis. *Sci. Rep.* 10:18643. doi: 10.1038/s41598-020-75271-7

Yan, P., Shen, W., Gao, X., Li, X., Zhou, P., and Duan, J. (2012). High-throughput construction of intron-containing hairpin RNA vectors for RNAi in plants. *PLoS One* 7:e38186. doi: 10.1371/journal.pone.0051204

Younis, A., Siddique, M. I., Kim, C.-K., and Lim, K.-B. (2014). RNA interference (RNAi) induced gene silencing: a promising approach of hi-tech plant breeding. *Int. J. Biol. Sci.* 10:1150. doi: 10.7150/ijbs.10452

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Matchett-Oates, Spangenberg and Cogan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chapter 5

Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification

5.1 Chapter preface

The recalcitrant nature of cannabis explants to transformational attempts, and the lack of regeneration protocols for drug-type cultivars, has made previous attempts to improve the cannabis germplasm through genome editing for its potential medicinal purposes, elusive. The medicinally important cannabinoid biosynthesis genes are, as previously demonstrated in Chapter 4, amendable to downregulation using RNA interference constructs via agroinfiltration. In this chapter, the co-transformation of cannabis cotyledons and hypocotyls with a highly effective cannabinoid gene silencing construct and an accompanying GFP vector was investigated. Following transformation, indirect organogenesis of cotyledon callus and the effect of increasing concentrations of TDZ were investigated for the respective regeneration efficiency. Different transformed explant types demonstrate varying levels of stable gene suppression and upregulation of cannabinoid genes. The data presented here demonstrates the ability to successfully regenerate cannabis from a transformational event in hypocotyls with a stably integrated RNAi construct significantly upregulating the cannabinoid biosynthesis genes. The data also presents *Cannabis'* ability to upregulate these genes to maintain a level of homeostasis through biosynthesis of non-specific cannabinoids. This is the first reported event of a stably integrated genome editing construct in a regenerated plant targeting the medicinally important genes creating a novel chemovar.

5.2 Publication details

Title: Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification

Journal: Industrial Crops and Products

Stage of publication: Submitted

Authors: Lennon Matchett-Oates, Ehab Mohamaden, German Spangenberg, Noel Cogan

5.3 Statement of contribution of joint authorship

LMO designed all RNAi constructs and performed all experiments. EM helped in regeneration protocol development. LMO drafted entire manuscript. GS and NC helped conceive manuscript direction and assisted in editing the manuscript. All authors approved the final manuscript.

5.4 Co-author statement confirming authorship of PhD candidate

As co-author of the manuscript 'Matchett-Oates L, Mohamaden E, Spangenberg GC, Cogan NOI. Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification. Submitted to Plants.' I confirm that Lennon Matchett-Oates has made the following contributions,

- Development of protocols employed
- Generation of all biologicals used in the study
- Generation of all figures and tables
- Writing the manuscript, critical appraisal of the content and response to reviewers

Dr Noel O.I. Cogan

Date: 20/12/2021

Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification

Lennon Matchett-Oates[a,b*], Ehab Mohamaden[b], German Spangenberg[a,b], Noel Cogan[a,b]

[a] Agriculture Victoria, AgriBio, The Centre for AgriBioscience, Bundoora, VIC, Australia
[b] School of Applied Systems Biology, La Trobe University, Bundoora, VIC, Australia
* Correspondence: lennon.matchett-oates@agriculture.vic.gov.au

Abstract: *Cannabis sativa* L. contains numerous phytocannabinoids with great medical potential. The genes responsible for the accumulation of cannabinoids have only recently been identified and able to be targeted using RNA interference, demonstrating significant downregulation of these highly homologous gene sets is now possible, potentially leading to unique chemotypes. Reported here is the application of this developed approach to generate multiple stably transformed cannabis explants, with accompanying developed regeneration protocols from cotyledons and hypocotyls. Agroinfiltration with pRNAi-GG-*CBDAS*-UNIVERSAL within cotyledon-derived callus shows highly significant ($p < 0.05$) down and upregulation within tetrahydrocannabinolic acid synthase (*THCAS*), cannabidiolic acid synthase (*CBDAS*), and cannabichromenic acid synthase (*CBCAS*) transcripts. Regeneration of a stably transformed hypocotyl explant showed significant upregulation of cannabinoid genes indicating these genes' ability to synthesize non-specific cannabinoids *in planta*. The results from using RNA interference to target individual cannabinoid biosynthesis genes suggest that even base-pair specific approaches, such as CRISPR, to predictably manipulate cannabinoid production will prove to be more difficult due to *Cannabis'* innate ability to upregulate these genes to maintain homeostasis.

Keywords: *Cannabis* sativa, RNAi, *Agrobacterium,* THCAS, CBDAS, CBCAS, cannabinoid biosynthesis genes, PTGS

1. Introduction

Cannabis sativa L (cannabis). has been used as far back as 4000 BC in central Asia as a source of bast fibre, seed oil and for its psychoactive properties (Li, 1973; Touw, 1981). Cannabis produces unique phytocannabinoids, which contain therapeutic properties, with over 120 of these cannabinoids being discovered thus far (ElSohly et al., 2017). Phytocannabinoids are produced in their acidic, non-active forms, which become active through decarboxylation or enzymatic degradation. Of these phytocannabinoids, Δ^9-tetrahydrocannabinol (THC) and cannabidol (CBD) have gained great attention for their pharmacological benefits, including anti-inflammatory action (Klein, 2005), epileptic relief (Szaflarski and Bebin, 2014), and cancer pain management (Degenhardt et al., 2015). Several other phytocannabinoids also produce medicinal benefits, such as the upstream cannabinoid cannabigerol (CBG), which provides benefits for inflammatory bowel disease (Borrelli et al., 2013) and cannabichromene (CBC) which contains anti-bacterial and anti-fungal activity (Turner and Elsohly, 1981).

Due to the clandestine breeding of cannabis to produce cultivars with higher levels of psychoactive THC, the population has resulted in highly variable copy numbers of cannabinoid biosynthesis genes (Grassa et al., 2021; Matchett-Oates et al., 2021a). With the recent comprehensive update to the cannabis genome (Braich et al., 2020; Grassa et al., 2021) along with an accompanying transcriptome (Braich et al., 2019), greater understanding of the *THCAS* and *CBDAS* loci, with their pericentric nested repeats resulting in several pseudogenes, highlightes the complexities in cannabinoid biosynthesis *in planta* and gene specific designs for genome editing.

Cannabis is a dioecious, wind pollinated species with seed propagation being the primary method for elite cultivar selection for most of its history. The inherent genetic diversity that outbreeding seed propagation adds further amplifies the difficulty in applying biotechnological and tissue culture approaches to a recalcitrant species. Due to the status of cannabis' legality worldwide, only recently has research into *in vitro* propagation and regeneration occurred (Monthony et al., 2021b). Investigation into the factors influencing micropropagation success has been widely researched (Chandra et al., 2010; Chaohua et al., 2016; Lata et al., 2017; Wang et al., 2009), though due to the extensive outbreeding, this has resulted in variable micropropagation approaches being necessary. Clonal propagation through direct and indirect embryogenesis has had limited success in cannabis. Callogenesis has been achieved in several cannabis cultivars using basal mediums, such as Murashige and Skoog (MS) (Movahedi et al., 2015) and Driver

44 and Kuniyaki Walnut (DKW) (Page et al., 2020), and several plant growth regulators, such as thidiazuron (TDZ) (Lata et al., 2009)
45 and 1-naphthaleneacetic acid (NAA) (Slusarkiewicz-Jarzina et al., 2005) have been explored. Regeneration of cannabis from
46 differing explants has also been reported and collectively analysed (Monthony et al., 2021b). A significant bottleneck for achieving
47 highly efficient regeneration protocols in cannabis is due to the highly polymorphic nature of cannabis even within the same variety,
48 which often leads to results being irreplicable (Monthony, et al., 2021a). Multiple hemp cultivars have been investigated to determine
49 growth regulator combinations ability to regenerate (Slusarkiewicz-Jarzina et al., 2005). Most recently, successful direct
50 regeneration from cotyledons has been reported through several media combinations using an outbreeding seed propagation
51 population (Galán-Ávila et al., 2020). Hypocotyl regeneration through hormone-free media has also been reported with considerably
52 high transformation and regeneration efficiencies across multiple hemp cultivars using seed populations (Galán-Ávila et al., 2021),
53 greatly accelerating the potential for producing transformed cannabis cultivars avoiding the troublesome regeneration media
54 optimisation for individual genotypes from segregating seed populations.
55
56 Due to the recalcitrant nature of cannabis, limited biotechnological transformational approaches have been reported. Transient
57 expression of heterologous green fluorescent protein (GFP) in cannabis leaf mesophyll protoplasts have recently been reported
58 (Beard et al., 2021; Matchett-Oates et al., 2021c) as one potential for improving the germplasm. Using *Agrobacterium*-mediated
59 transformation, optimisation of variables involved in vacuum infiltration of leaf discs, via GUS expression, has been reported
60 (Deguchi et al., 2020). Using this approach, transient agroinfiltration using RNA interference (RNAi) constructs significantly
61 downregulating tetrahydrocannabinolic acid synthase (*THCAS*), cannabidiolic acid synthase (*CBDAS*), and cannabichromenic acid
62 synthase (*CBCAS*) has been reported (Matchett-Oates et al., 2021b). The use of CRISPR/Cas-9 has also recently been reported
63 targeting endogenous housekeeping genes within protoplast genomes (Zhang et al., 2021).
64
65 Generation of a stably transformed cannabis plant has recently been achieved through different explant types (Galán-Ávila et al.,
66 2021) providing an indispensable protocol for genome editing attempts. Using this approach, with specific gene silencing RNAi
67 constructs targeting cannabinoid biosynthesis genes, will allow for the production of an elite cultivar with modified cannabinoid
68 biosynthesis expression levels, as has been proven previously within transiently expressing leaf segments (Matchett-Oates et al.,
69 2021b). The RNAi mechanism is initiated by an introduced double-stranded RNA (dsRNA), a gene sequence sharing homology to
70 the target gene of choice, which is processed by Dicer-like Proteins (DCLs) to form small interfering RNA (siRNA). siRNA are
71 loaded into the RNA-induced Silencing Complex (RISC), with one strand of the siRNA acting as a guide targeting mRNA with
72 complementary sequences (Majumdar et al., 2017). Following base pairing between RISC and mRNA, Argonuate (AGO) proteins
73 cleave the target mRNA preventing translation. Using this approach, significant downregulation of the highly homologous
74 cannabinoid biosynthesis genes can allow for the accumulation of the medicinally important cannabinoid precursor, CBG, which
75 only occurs in minimal concentrations due to the highly effective nature of *THCAS, CBDAS,* and *CBCAS*. Conversely, plant
76 homeostasis could be triggered through miRNA, such is the case in *Arabidopsis* in the regulation of Phosphate (Chiou et al., 2006)
77 and other nutrients (Paul et al., 2015), to upregulate the cannabinoid biosynthesis genes further highlighting the possible capabilities
78 of these genes to synthesise non-specific cannabinoids.
79
80 In this study an RNAi construct, previously reported with transient expression to significantly down and upregulate the main
81 cannabinoid biosynthesis, was agroinfiltrated into cotyledons and hypocotyls of cannabis. Direct and indirect organogenesis was
82 attempted using previously demonstrated protocols in cotyledons and hypocotyls and an optimised cotyledon regeneration protocol
83 developed for an intermediate chemotype. Quantitative PCR (qPCR) was used to detect the modulation of the cannabinoid
84 biosynthesis genes within control and transformed samples. To the best of our knowledge, this is the first report of a stably
85 transformed, regenerated cannabis plant containing a construct targeting the medicinally important cannabinoid biosynthesis
86 pathway. This study also quantifies the transformation and regeneration limitations and constaints in a segregating cannabis

87 population, and defines the scale of effort required to ensure the desired outcomes, to produce the desired novel chemotypes in
88 stable regenerated plants.

89 **2. Materials and Methods**

90 **2.1 Plant material and Agrobacterium culture conditions**
91 All research was performed under Medicinal Cannabis Research Licence (RL011/18) and Permit (RL01118P4) issued through the
92 Department of Health (DoH), Office of Drug Control (ODC) Australia.
93 Cotyledon explants from a segregating F2 inter-cross between a high THC male and high CBD female cultivar were used for
94 transformation experiments. All seeds were initially surface sterilised by soaking in 80% (v/v) ethanol for 1 min with gentle shaking.
95 Seeds were then rinsed with sterile ddH_20 three times and soaked in 15% (v/v) bleach (4.25% active sodium hypochlorite) (White
96 King, Australia) for 10 min with gentile agitation. Seeds were rinsed three times again in ddH_20 and approximately 50 seeds were
97 placed in each 50 mL centrifuge tube (Corning, NY) half filled with sterile ddH_20 for 5-7 days in a controlled environment room at
98 24°C in the dark for germination to enable cotyledon excision. Sterile seeds used for hypocotyl excision were placed in germination
99 medium [1/2 MS and vitamins, 1.5% sucrose, 3.5g/L Gelrite, pH 5.8] (Sigma Aldrich, St. Louise, MO) in Stericon-8 culture vessels
100 (PhytoTech Labs, Lenexa, KS) for 7-14 days in a controlled environment room at 24 degrees with a 16/8 photoperiod, providing 74
101 $\mu mol\ m^{-2}\ s^{-1}$ light intensity supplied by fluorescent lamps.
102

103 For the expression of siRNA, the vector pRNAi-GG-*CBDAS-UNIVERSAL* (Supplementary Material) was used within the study.
104 Construction of the pRNAi-GG- *CBDAS-UNIVERSAL* has previously been described (Matchett-Oates et al., 2021b). Cotyledons
105 and hypocotyls were also co-transformed with the binary vector pDPI-13 (Supplementary Material), constructed from a pPZP200-
106 based binary vector with CaMV35S-p_turboGFP_nos-t and CsVMV-p_hph_CaMV35S-t cassettes and spectinomycin selection, to
107 express GFP. Recombinant *Agrobacterium* strain GV3101 was used for all transformation experiments. *Agrobacterium* was grown
108 overnight in YM media [0.5 g/L K_2HPO_4, 0.2 g/L $MgSO_4$-$7H_2O$, 0.1 g/L NaCl, 10 g/L Mannitol, 0.4 g/L yeast extract, PH 7] (Sigma
109 Aldrich, MO, USA) with appropriate selection at 200 rpm at 30°C. The cultures were centrifuged at 4000 *g* for 10 min and
110 resuspended to $OD_{600} = 0.5$ in infiltration media [10 mM MES, 1x MS, 2% glucose, pH 5.6] (Sigma Aldrich, St. Louise, MO) and
111 mixed prior to transformation.
112

113 **2.2 Cotyledon and hypocotyl excision, *Agrobacterium*-mediated transformation and co-cultivation**
114 Once seeds were germinated, for cotyledon transformation, the seed coat was pried open using sterile forceps and scalpel with the
115 cotyledons removed by cutting the two cotyledons apart and removing the radicle. The cotyledons were placed into the mixed
116 *Agrobacterium* cultures for 15 min, in a Petri dish, in the dark at room temperature (22°C) for static infection. Post infiltration,
117 cotyledons were placed on sterile filter paper (Whatman, Maidstone, UK) for 5 min to remove excess *Agrobacterium* culture.
118 Cotyledons were moved onto solid co-cultivation media [1x MS and vitamins, 3% sucrose, 0.5 g/L MES hydrate, 0.7% agar, 2 mg/L
119 2,4-D, 200 µM acetosyringone, pH 5.6] (Sigma Aldrich, St. Louise, MO) for 3 days in the dark at 24°C. Control cotyledons were
120 treated identically, however a disarmed *Agrobacterium* was used.
121

122 Whole hypocotyls, excised from below cotyledons and above radicle, from germinated seeds were placed in *Agrobacterium* culture
123 for 40 min in the dark at room temperature for static infection. Post infection, hypocotyls were placed on sterile filter paper
124 (Whatman, Maidstone, UK) for 5 min to remove excess *Agrobacterium* culture. Hypocotyls were then transferred to co-cultivation
125 media [1/2 MS and vitamins, 1.5% sucrose, 200 µM acetosyringone, 3.5 g/L Gelrite, pH 5.6] (Sigma Aldrich, St. Louise, MO) for
126 3 days in a controlled environment room as described above. Similar to cotyledons, control hypocotyls were used with a disarmed
127 *Agrobacterium*.

128 **2.3 Callus induction and regeneration**

129 Cotyledon callus induction was carried out on callus induction media (CIM) [1x MS and vitamins, 3% sucrose, 0.7% agar, 1 mg/L
130 Kinetin, 0.2 mg/L NAA, pH 5.8] (Sigma Aldrich, St. Louise, MO). Three days post co-cultivation, cotyledons were rinsed 3 times
131 with sterile ddH_2O and washed with 200 mg/L Timentin (PhytoTech Labs, Lenexa, KS) for 2 min. Post wash, cotyledons were
132 moved to filter paper for 5 mins to remove excess moisture and moved onto CIM, 12 per plate, and placed in an incubator at 24°C
133 in the dark for 5 weeks. Phenotypic development of callus was photographed weekly and images analysed using ImageJ (ImageJ
134 1.53e) (National Institute of Mental Health, MD, USA) using the free-hand tool. Callus mass on selected sub-set was performed on
135 an electronic balance (Mettler Toledo, Columbus, OH) (Model ML204T) after 14 days initiation and weighed every consecutive 7
136 days.
137
138 Cotyledon regeneration media [1x MS and vitamins, 2% maltose, 0.1 g/L myo-inositol, 10 mM MES, 0.7% agar, 120 mg/L
139 Timentin, pH 5.8] (Sigma Aldrich, St. Louise, MO) was prepared with three increasing concentrations of Thidiazuron (TDZ) (Sigma
140 Aldrich, St. Louise, MO) (2.5, 5 and 10 µM). Post callus induction, calli was transferred onto regeneration media in SteriCon-13
141 culture vessels (PhytoTech Labs, Lenexa, KS) for 6-8 weeks.
142
143 Post co-cultivation, hypocotyls were rinsed with sterile ddH_2O three times and placed in 200 mg/L Timentin wash for 5 mins with
144 occasional gentle agitation. Post Timentin wash, hypocotyls were placed on filter paper for 5 min and then transferred to regeneration
145 media [1/2 MS and vitamins, 1.5% sucrose, 3.5 g/L Gelrite, 120 mg/L Timentin, pH 5.8] for 4-8 weeks in Stericon-8 culture vessels
146 (PhytoTech Labs, Lenexa, KS) in a controlled environment room as described above. Regenerated shoots were excised and moved
147 to rooting media [1/2 MS and vitamins, 1% sucrose, 5 µM IBA, 1% agar, 120 mg/L Timentin, pH 5.8] (Sigma Aldrich, St. Louise,
148 MO).
149

150 **2.4 Quantitative Real-Time PCR (RT-qPCR) of cannabinoid genes**

151 For cotyledon callus RNA, approximately 100 mg callus mass was snap frozen in liquid nitrogen with total RNA extracted following
152 manufacturers instruction (RNeasy Plant Mini Kit, Qiagen, Hilden, Germany) and quantified using a Nanodrop 1000 (ThermoFisher
153 Scientific, Walthan, MA). Approximately 5 µg of RNA was treated with DNase I (New England Biolabs, Ipswich, MA) as per
154 manufacturer's instructions and used for subsequent RT-qPCR. For regenerated hypocotyl RNA, approximately 100 mg of leaf
155 material was snap frozen in liquid nitrogen and total RNA extracted, quantified and treated as described above. cDNA synthesis and
156 qPCR were carried out in one step with Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Ipswich, MA) following
157 manufacturer's instructions and using parameters as previously described (Matchett-Oates et al., 2021b). Briefly, quantitative PCR
158 parameters used were as follows: 95°C for 60 sec, 40 cycles at 95°C for 15 sec and 59°C for 15 sec carried out with a CFX-96
159 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA). Melting curves were measured and gene expression levels were
160 calculated from the cycle threshold according to the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Paired *t*-test was performed (p =
161 0.05) to determine significance using RStudio (Version 1.1.453, RStudio Inc., Boston, MA). The *UBQ5* gene was used as an internal
162 reference (Deguchi et al., 2020; Matchett-Oates et al., 2021b), with technical triplicates. Due to the numerous pseudogenes within
163 the *CBDAS* locus, two primer pairs were designed, one set targeting the known functional copy of *CBDAS*, and one set designed to
164 target *CBDAS*-like homologs (Matchett-Oates et al., 2021b) and mixed in equal molar concentrations to target all possible CBDAS-
165 homologs during RT-qPCR. All primers bind outside gene sequences targeted by generated siRNA to accurately assess
166 down/upregulation. For pooled control samples, each treated control was quantified using a Nanodrop 1000 (ThermoFisher
167 Scientific, Walthan, MA) and 500 ng of each control pooled for use as a combined control. All primer sequences are listed in
168 Supplementary Material.
169
170
171

2.5 GFP expression detection and analysis

Cotyledons, hypocotyls, callus, and leaf material were imaged under fluorescence with GFP filter set (Excitation 395-455 nm, emission 480 nm) using a Leica camera (CH-9435) with Leica Application Suite software (4.12.0). GFP fluorescence data for each callus was collected at week 5, prior to moving to regeneration media to determine gene stability. Hypocotyls were measured at week 5 on regeneration media. Data was collected as presence/absence in explant material.

3. Results

3.1 Cotyledon callus induction frequency, size and weight characteristics

In total, 321 co-transformed cotyledons were evaluated for callus induction and two-dimensional growth, with a subset of 84 cotyledons weighed across 4 weeks to measure growth parameters. A total of 60 control cotyledons were also run in parallel where innoculation with a disarmed *Agrobacterium* occurred. Seeds were germinated and cotyledons placed on callus induction media (CIM) over 4 weekly intervals (Groups 1-4; Table 1). Cotyledons were placed on CIM for 5 weeks, post 3-day co-cultivation with the *Agrobacterium* culture, with size, weight, and induction frequency being measured at week 2 (Table 1).

Table 1: Measured cotyledon characteristics during 4 measured weeks on CIM

	Callus induction	Calli size growth (avg. per week)	Calli weight increase (avg. per week)
Group 1 (84 samples)	88.10%	0.142 cm^2/week	0.033 g/week
Group 2 (59 samples)	94.92%	0.125 cm^2/week	
Group 3 (78 samples)	100.00%	0.193 cm^2/week	
Group 4 (120 samples)	100.00%	0.138 cm^2/week	
Average (Transformed)	95.76%	0.146 cm^2/week	
Control (60 samples)	93.33%	0.100 cm^2/week	0.035 g/week

*Dead calli not included in growth and weight average calculations

Exceedingly high callus induction rates were recorded across all 4 Groups, with 100% success measured in groups 3 and 4. Slightly lower callus induction frequencies, 88.1% and 94.9% were recorded for groups 1 and 2, respectively, with the average callus induction rate, for transformed callus, recorded at 95.76%. The control group had 93.3% success rate, which only Group 1 did not achieve this level of callus induction. Throughout the 5 weeks on CIM, 4 measurements for two-dimensional calli growth were recorded using ImageJ (Figure 1).

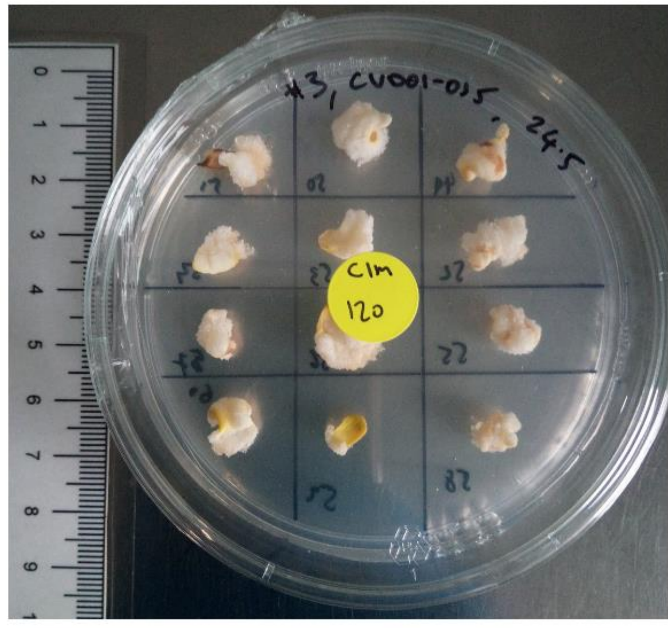
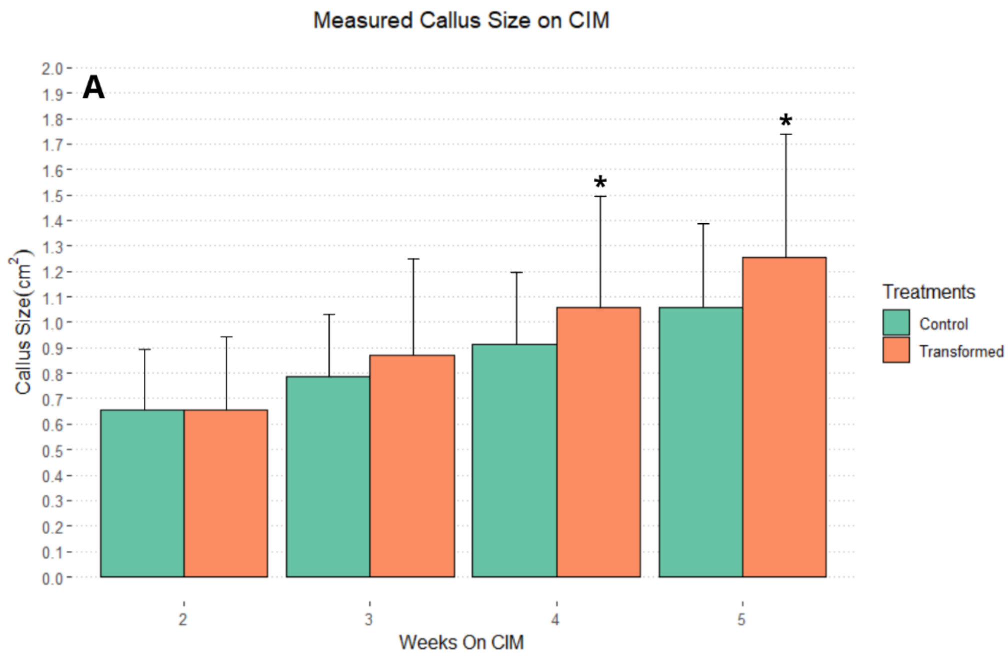
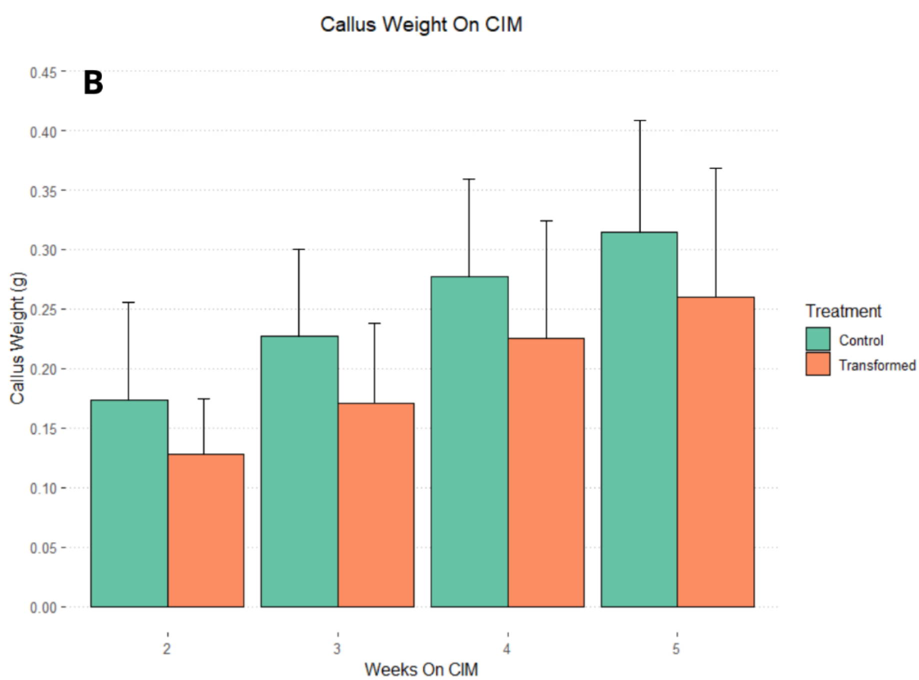


Figure 1: Cotyledon callus formation on CIM measured using ImageJ

195 Calli growth gradually increased across 4 weeks on CIM for trasnformed and control groups (Figure 2A). Regarding individual
196 groups, Group 2 calli size increased, on average, the least out of transformed cotyledons, with an increase of 0.125 cm^2/week. Group
197 3 had the largest average weekly increase, with 0.193 cm^2/week. Groups 1 and 4 performed similarly with 0.143 and 0.138 cm^2/week,
198 respectively, with the average callus growth for the transformed cotyledons, being 0.146 cm^2/week (Table 1). Measurements taken
199 at weeks 4 and 5 on CIM, saw the transformed cotyledons significantly ($p < 0.05$) out-grow the control cotyledons (Figure 2A).
200 Weeks 4 and 5 for the control group measured 0.915 cm^2 and 1.056 cm^2, respectively, with the transformed group measuring 1.082
201 cm^2 and 1.274 cm^2, respectively. Phenotypic response to callus induction (and transformation) saw the transformed cotyledons
202 produce white, less dense callus compared with the control group, which produced more compact friable callus.
203
204
205

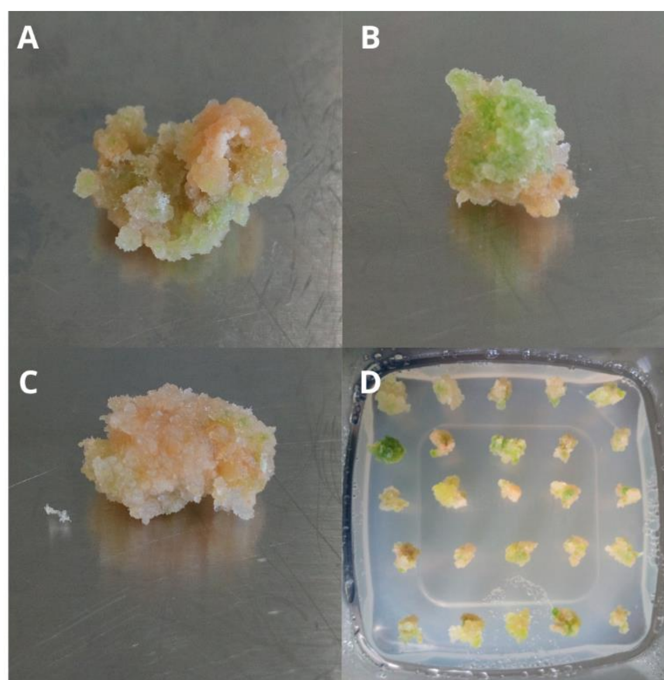


206



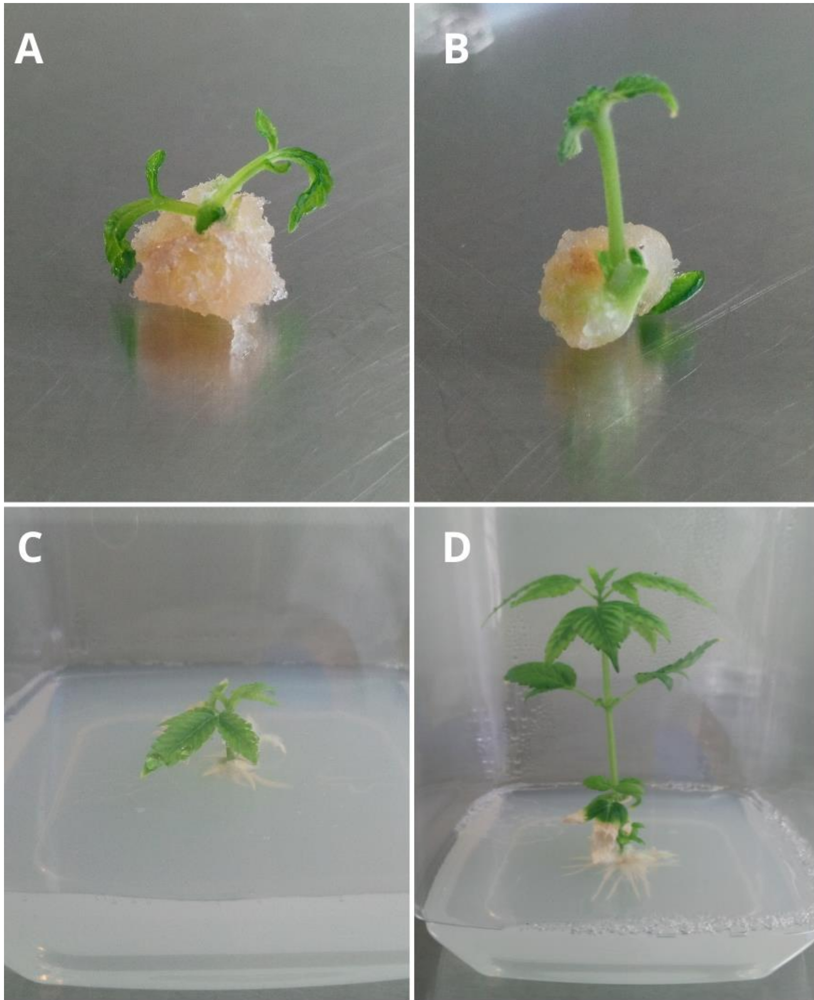
207

208 **Figure 2: A; Measured callus size for transformed and control callus over 4 weeks on callus induction media. Error bars**
209 **represent standard deviation. Asterix represent $p < 0.05$ using paired *t*-test. B; Average weight of the Group 1 transformed**
210 **callus compared with control over 5 weeks on CIM**
211
212
213 Group 1 (84 cotyledons) was used as a sub-set to measure calli weights from weeks 2-5 on CIM. The 84 cotyledons were weighed
214 and recorded with an average of 0.033 g/week increase (Table 1), with sixty control cotyledons also weighed simultaneously. The
215 control callus had greater mass than the transformed cotyledons throughout the entire experiment (Figure 2B), gaining on average
216 0.035 g/week (Table 1).
217
218 **3.2 Calli response to different regeneration treatments and transformation frequency**
219 Post 5 weeks on CIM, calli masses were transferred onto 2.5, 5 or 10 µM TDZ regeneration media to encourage shoot organogenesis.
220 Transformed and control calli were randomly assigned different TDZ treatments and placed within the controlled environment room
221 for 8 weeks. Phenotypic response to different TDZ concentrations on regeneration media of the transformed cotyledons showed no
222 significant variation, with size, colour and formation relatively uniform, suggesting that regardless of TDZ concentration,
223 unorganised friable callus was produced with eventual plastid development present in green photosynthetic cells (Figure 3A-D).
224 Similarly, non-responding callus within the control group showed eventual plastid development in unorganised cellular masses with
225 sizes and colour comparable to transformed callus.
226



227
228 **Figure 3: Comparison of callus phenotypic response to varying regeneration media compositions post 5 weeks. A: 2.5 µM**
229 **TDZ treatment (transformed callus); B: 5 µM TDZ treatment (transformed callus); C:10 µM TDZ treatment**
230 **(transformed callus); D: Control treatment on 5 µM TDZ treatment**
231
232 No regeneration was observed from transformed cotyledons treated with varying TDZ concentrations. However, leaf primordia
233 were present in two transformed calli mass with regenerated leaves present (Supplementary Figure 1), though organogenesis
234 progress was not observed from these two calli. The control callus did produce three regenerated shoots, with a single shoot produced

235 from each of the three regeneration media compositions (Figure 4A-D), resulting in a regeneration frequency within the control
236 group of 5%.



237

238 **Figure 4: Regeneration of control callus; A: Regenerating shoots from calli on 5 µM TDZ regeneration media; B:**
239 **Regenerating shoots from calli on 10 µM TDZ regeneration media; C: Regenerated shoot in rooting media post 7 days; D:**
240 **Regenerated shoot in rooting media post 28 days**

241

242 Transformation efficiency, defined as a single callus mass expressing GFP, was recorded at week 5 on regeneration media to evaluate
243 GFP stability in transformed callus, with transformation scored as present/absent from GFP excitation under observation with a GFP
244 filter. Similar GFP expression rates were observed across all 4 groups, with Group 2 having the lowest transformation efficiency of
245 57.6% and Group 3 achieving the highest efficiency with 69.2% (Table 2) (Figure 5). The average transformed cotyledons
246 expressing GFP post 5 weeks on regeneration media equalled 62.3%.

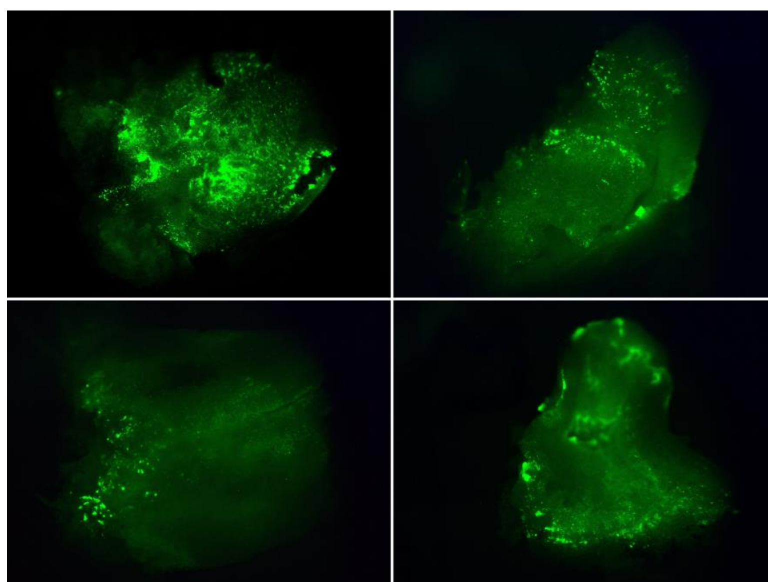
247

248 **Table 2: Transformation efficiency, defined as single callus mass expressing GFP, in cotyledonspost 5 weeks on CIM**

Cotyledon callus	Transformation efficiency
Group 1 (84 samples)	63.1%
Group 2 (59 samples)	57.6%
Group 3 (78 samples)	69.2%
Group 4 (120 samples)	59.2%
Average	62.3%

249

250



251 **Figure 5: GFP expressing calli masses after 5 weeks on regeneration media**

252

253 ### 3.3 Hypocotyl rooting, regeneration and transformation efficiency

254 In total, 304 transformed and 40 control (disarmed *Agrobacterium*) hypocotyls were evaluated for rooting, regeneration and 255 transformation efficiency on hormone-free regeneration media post *Agrobacterium* co-transformation. Hypocotyls were placed on 256 regeneration media for 4-8 weeks at weekly intervals (Groups 1-6) with efficiencies tabulated at week 4, respectively (Table 3).

257

258 **Table 3: Hypocotyl rooting, regeneration and transformation efficiencies across transformed and control**

	Average Rooting Induction %	Average Shoot Regeneration %	Transformation Efficiency %
Group 1 (56 samples)	8/56 (14.3%)	0.0	31/56 (55.4%)
Group 2 (47 samples)	9/47 (19.1%)	0.0	26/47 (55.3%)
Group 3 (45 samples)	10/45 (22.2%)	2/45 (4.4%)	18/45 (40.0%)
Group 4 (52 samples)	11/52 (21.2%)	0.0	24/52 (46.2%)
Group 5 (51 samples)	12/51 (23.5%)	1/51 (2.0%)	23/51 (45.1%)
Group 6 (43 samples)	10/43 (23.3%)	0.0	20/43 (46.5%)
Average (Transformed)	20.6%	1.1%	48.1%
Control (40 samples)	19/40 (47.5%)	8/40 (20.0%)	-

259

260 Spontaneous rooting efficiencies across transformed groups were remarkably lower compared with the control group. The lowest 261 rooting efficiency, Group 1, achieved just 14.3% with Groups 2-6 only performing slightly better, with 19.1%, 22.2%, 21.2%, 262 23.5%, and 23.3%, respectively. The average rooting efficiency for transformed hypocotyls was 20.6%. The control group, with 263 disarmed *Agrobacterium*, achieved 47.5% efficiency (Figure 6), more than double the transformed hypocotyls.

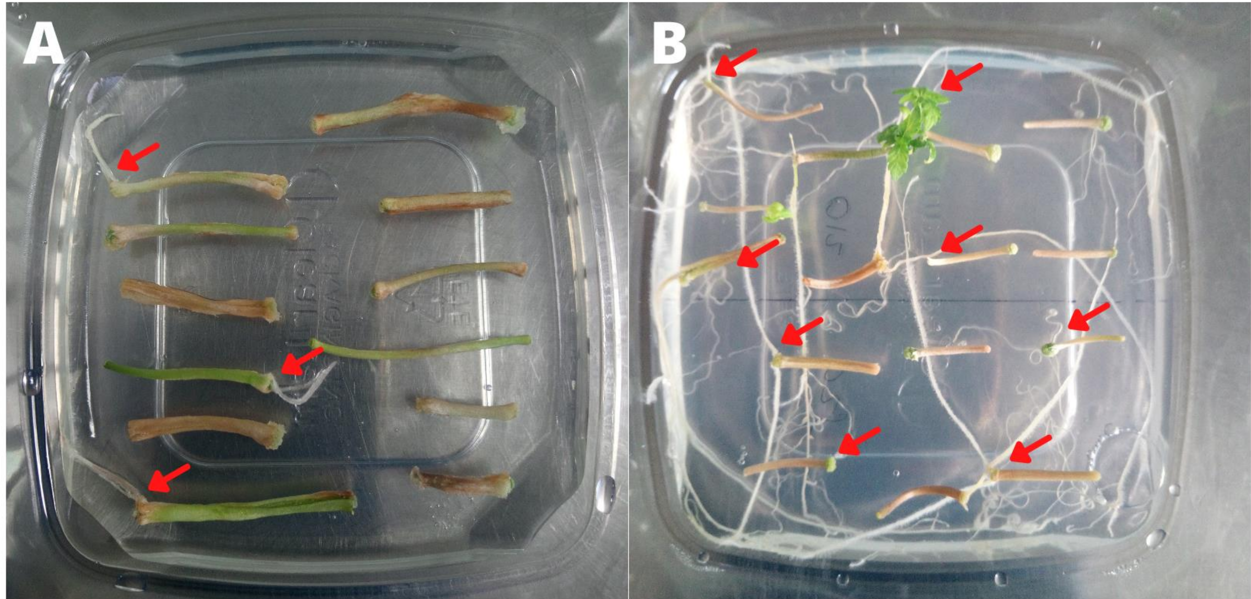
264

265

266

267

268



269 **Figure 6: Spontaneous rooting of regenerating hypocotyls; A: Co-transformed hypocotyls; B: Control hypocotyls. Arrows**
270 **identifying rooting hypocotyls. Images taken at week 3 on regeneration media**

271

272 Responding hypocotyls with regenerated shoots was not achieved in Groups 1, 2, 4, and 6. Whereas, within Group 3, 2 hypocotyls
273 responded with regenerated shoots (Figure 7B) and Group 5 produced 1 regeneration event, for an average of 1.1% regeneration
274 response across all transformed hypocotyls. Comparatively, the control group generated 8 regenerated shoots for a regeneration
275 frequency of 20% (Table 3) (Figure 7A).

276

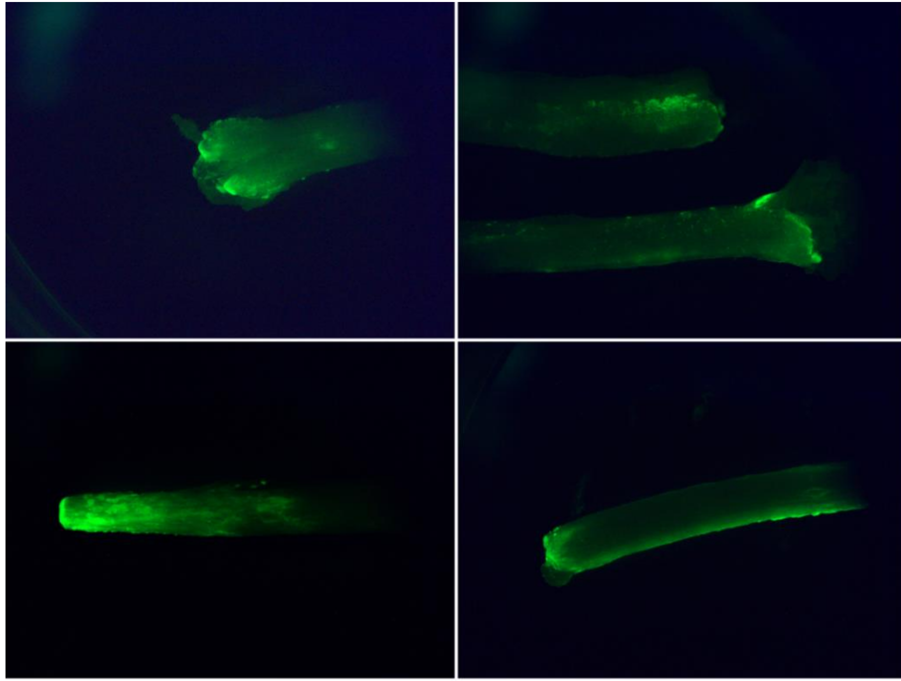
277

278 **Figure 7: Regenerated hypocotyls; A: Control hypocotyl; B: Transformed hypocotyl**

279

280 4 weeks post co-infiltration with *Agrobacterium*, the presence of GFP was recorded in transformed hypocotyls (Figure 8), with GFP
281 scored as present/absent from GFP excitation under observation with a GFP filter. Group 1 achieved the highest average rate of
282 transformation, with 55.4% efficiency, followed closely by Group 2 with 55.3%. Group 3 achieved the lowest efficiency, with
283 40.0%. Groups 4, 5, and 6 achieved very similar results, with 46.2%, 45.1%, and 46.5%, respectively, with the average
284 transformation efficiency across groups equalling 48.1% (Table 3).

285



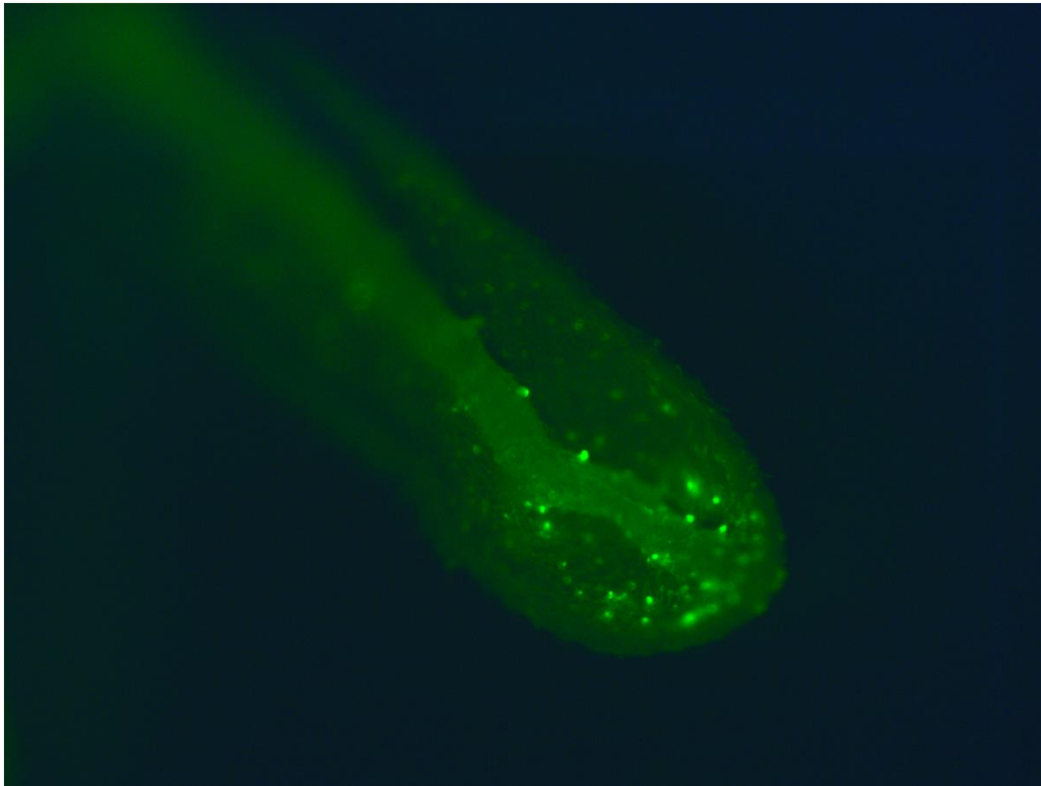
286

287 **Figure 8: GFP positive co-transformed hypocotyls post 4 weeks on regeneration media**

288

289 A total of 3 regenerated shoots from transformed hypocotyls were present, though only 1 acclimatised to rooting media. The single,
290 acclimatised regenerated hypocotyl was imaged for stable GFP fluorescence. Selected leaves contained GFP excitation within the
291 leaf tip apex and leaf midveins (Figure 9).

292



293

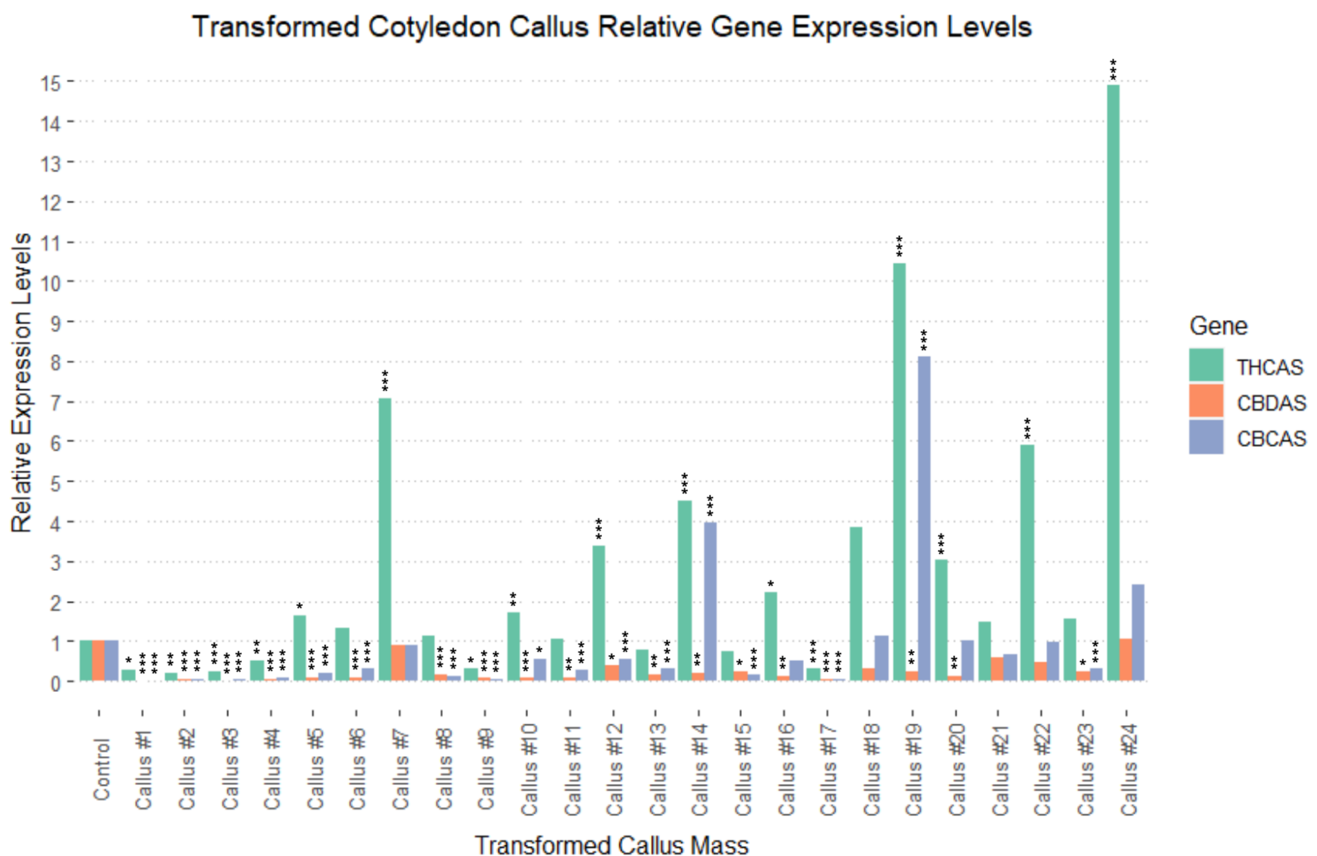
294 **Figure 9: Regenerated transformed hypocotyl leaf expressing GFP**

295

296 **3.4 RT-qPCR data of cotyledon calli and regenerated hypocotyl**

297 Cotyledon callus transcript levels of targeted genes from mRNA in control and transformed samples were analysed for transcript
298 level modification. In total, 12 control and 24 transformed cotyledon callus RNA was extracted and analysed using RT-qPCR with
299 specific primers (Supplementary Material) targeting *THCAS*, *CBDAS*, *CBCAS* using *UBQ5* as an internal housekeeping reference
300 gene for normalisation. Due to the outcrossing nature of cannabis, cotyledon calli are not genetically uniform with variance in gene
301 expression from targeted genes. To best normalise transcript levels, equal RNA concentrations of each control were mixed to
302 produce a pooled sample which were used for normalisation for relevant gene expression, which was run along with individual
303 control callus RNA. The pooled RNA expression levels were used for subsequent calculations in callus relative gene expression
304 levels of *THCAS*, *CBDAS,* and *CBCAS* using the $2^{-\Delta\Delta Ct}$ method. Extracted RNA from cotyledon masses, post 8 weeks on
305 regeneration media, has shown significant down and upregulation of transcript levels, within every transformed callus, with the
306 exception of callus #21. Significant downregulation of all three genes was recorded for callus masses 1-4, 9, and 17. Significant
307 upregulation in at least one of the three genes was recorded for callus masses 5, 7, 10, 12, 14, 16, 19, 20, 22, and 24. Significant
308 upregulation of *THCAS* was substantially more prevalent compared with *CBDAS* and *CBCAS*, with 10 calli masses producing
309 significant results. *CBCAS* also produced significant upregulation in 3 instances, with *CBDAS* not being significantly upregulated
310 in any callus mass. *CBDAS* was the most substantially affected with significant downregulation recorded in 20 callus masses (Figure
311 10). Co-transformation of cotyledons with a GFP vector, measured by a significant (p <0.05) modified expression profile of
312 cannabinoid genes and GFP excitation under fluorescence, saw a co-transformation rate of 75% (18/24).

313



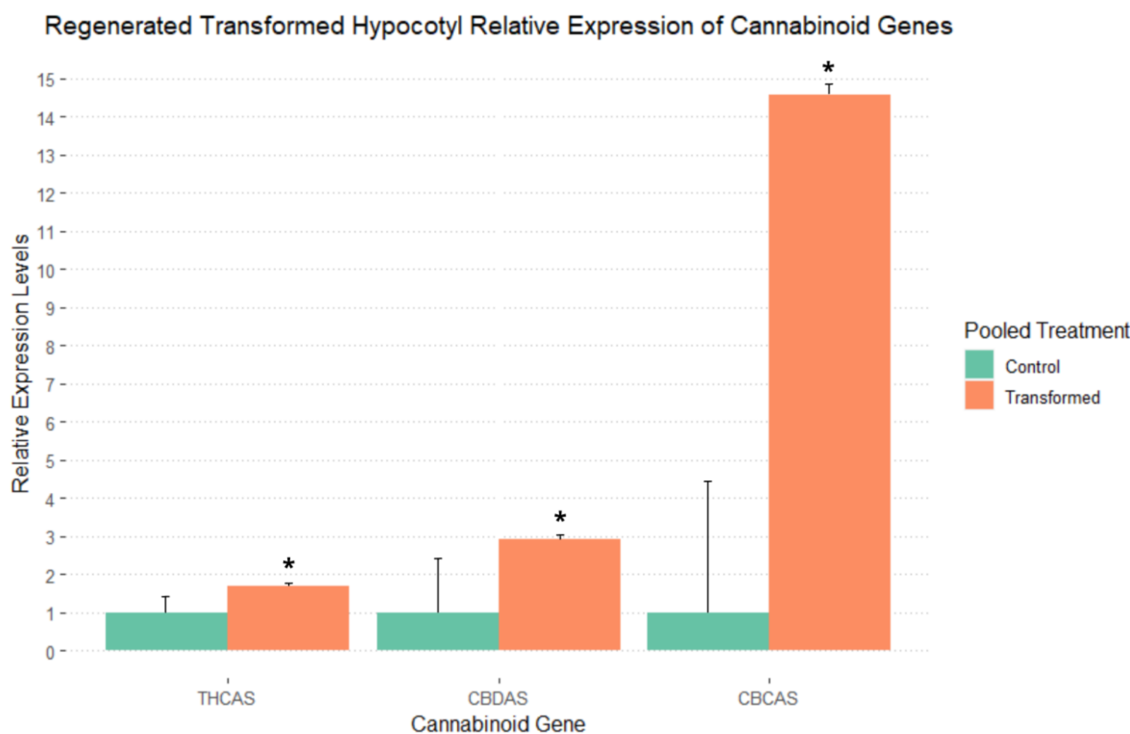
314

315 **Figure 10: Transformed callus relative expression levels for *THCAS*, *CBDAS* and *CBCAS*. Asterisks are representative of**
316 **significance using paired *t*-test; 1 asterisks p < 0.05; 2 asterisks p < 0.01; 3 asterisks p < 0.001. All p values and standard**
317 **errors are derived from technical triplicates and can be found in Supplementary Table 1**

318

319

320 A total of three regenerated shoots from transformed hypocotyls were produced. However, only one regenerated transformed
321 hypocotyl survived once transferred to rooting media, with necrosis occurring shortly after to the remaining two regenerated
322 plantlets. To analyse the relative gene expression in the single regenerated hypocotyl, the 8 control regenerated hypocotyls, like the
323 cotyledon callus, had equal RNA concentrations pooled for normalised gene expression analysis which was run alongside the
324 individual controls. Three biological replicates were taken from the regenerated hypocotyl for normalisation run with technical
325 triplicates. With these two normalised qPCR data sets, transcript levels for all cannabinoid biosynthesis genes were significantly (p
326 = 0.05) increased, with *THCAS, CBDAS,* and *CBCAS* having relative expression levels of 1.68, 2.9, and 14.57, respectively (Figure
327 11). Within the 8 control hypocotyls, large expression variation (Supplementary Table 2) exists as expected from a segregating F2
328 cross, with the largest variation existing for *CBCAS* with a standard error of 3.59, followed by *CBDAS* with 1.54 and *THCAS* with
329 0.55 (Figure 11).
330



331

332 **Figure 11: Regenerated transformed hypocotyl relative expression levels of cannabinoid biosynthesis genes vs. pooled**
333 **control regenerated hypocotyls. Error bars within pooled control represents standard error of relative expression levels**
334 **between 8 regenerated control hypocotyls. Error bars within pooled transformed represents standard error between pooled**
335 **biological replicates of transformed vs. pooled control samples from biological and technical triplicates. Asterix represent *p***
336 **< 0.05 using paired *t*-test. Data available in Supplementary Table 2**
337

338 **4. Discussion**

339 The induction of RNA interference in cannabis through *Agrobacterium*-mediated transformation has recently been accomplished
340 and optimised (Deguchi et al., 2020; Schachtsiek et al., 2019). The targeting of the specific cannabinoid biosynthesis genes using
341 RNAi has demonstrated the effects significant off-targeting of siRNA has on relative expression levels in leaf explants (Matchett-
342 Oates et al., 2021b). Cannabis regeneration has been attempted numerous times, with success varying greatly due to genotype and
343 regeneration media (Monthony et al., 2021b) as well as explant source (Galán-Ávila et al., 2021). Building upon previously
344 developed protocols and approaches to genetically transform and regenerate cannabis, *Agrobacterium*-mediated transformation of
345 cotyledons and hypocotyls using an RNAi vector targeting cannabinoid biosynthesis genes was attempted. This study demonstrates
346 the ability to stably transform multiple cannabis explant types with RNAi constructs to significantly modulate cannabinoid

347 biosynthesis, as well as elucidate the required investment required to produce desired chemotypic profiles using biotechnological
348 approaches due to regeneration bottlenecks.
349
350 The evaluation of cotyledon explant response to callus induction and regeneration was measured during the experiment. Callus
351 induction between the control (disarmed *Agrobacterium*) and transformed showed little variation, 93.33% and 95.76%, respectfully,
352 indicating little affect from plasmid integration on callogenesis efficiency. This level of callogenesis from cotyledons is higher than
353 previously published results (Chaohua et al., 2016; Mandolino and Ranalli, 1999; Wielgus et al., 2008), though falls in line with
354 previously published results in other cannabis explant sources (Lata et al., 2010; Monthony et al., 2021b). Interestingly, the
355 combination of PGR's used in this study have been previously attempted (Wielgus et al., 2008) though the rate for callogenesis is
356 significantly lower than reported here. This is likely due to the genotypes used in the studies, with cannabis displaying high levels
357 of recalcitrance to callogenesis and regeneration (Monthony et al., 2021b) even when the same genotypes are used (Monthony et
358 al., 2021a). Callus growth saw the co-transformed samples grow 2-dimensionally larger in comparison to the control, $0.146cm^2$ and
359 $0.100 cm^2$ respectively. The phenotypic response to infection with an armed Ti plasmid saw pale white, fluffy, non-compact friable
360 callus forming resulting in a larger 2-dimensional area when imaged. Comparatively, the control group produced pale white, more
361 compact callus which had a smaller 2-dimensional footprint. This phenotypic variation to treatment conditions suggests that the
362 incorporation of a binary vector results in different callus formation. Whether this phenotypic response explicity effects regeneration
363 efficacy needs further investigation as the regeneration frequencies achieved here are insufficient to accuretly analyse. Between
364 weeks 4 and 5, the pooled co-trasnformed callus was significantly ($p = 0.05$) larger (Figure 2A) as a result of the pale white less
365 compact callus. The surface area reported here is far larger for the pooled samples of control and co-trasnformed than previously
366 reported in other cannabis genotypes (Movahedi et al., 2015; Page et al., 2020). Although previous studies have explored multiple
367 media compositions to optimise callogenesis, the reported callus surface area measurements are smaller than reported here. Again,
368 plant genotype is a significant variable in response to callogenesis. Monitoring callus weight during callogenesis, the sub-group
369 used to compare transformed and control treatments weighed less during all time points in comparison. The production of the pale
370 white, fluffy callus saw a converse relationship between size and weight compared with the control. Unlike callus size, the data
371 produced within this study sees relatively similar callus weights at the end of callogenesis as in previous studies (Movahedi et al.,
372 2015; Page et al., 2020).
373
374 The use of TDZ for organogenesis in cannabis has previously been attempted with regeneration successful from leaf juvenile leaf
375 explants (Lata et al., 2010) though no regeneration was recorded when a replication study was attempted (Monthony et al., 2021a).
376 The argument that the genotype plays a considerable role in the ability for certain explants to regenerate is valid, however even in
377 replicated studies using the same genotype results can be variable. A wide range of hormonal treatments to encourage indirect
378 organogenesis, or direct organogenesis, in cannabis have been attempted and analysed (Monthony et al., 2021b) with the overall
379 consensus being that a high regeneration frequency is currently unobtainable due to the large genetic variation from outbreeding.
380 Regardless, successful attempts at regeneration prove that currently, whilst regeneration frequencies are low, it is still achievable
381 (Slusarkiewicz-Jarzina et al., 2005; Wielgus et al., 2008; Farag, 2014; Chaohua et al., 2016; Galán-Ávila et al., 2021). Within this
382 study, the regeneration frequency of the control cotyledon was approximately 5% (3/60), with one regenerated cotyledon achieved
383 in each TDZ treatment. The transformed cotyledons did not produce any regenerated shoots (though leaf primordia were present).
384 With a regeneration frequency of only 5% determining if the transfection with the armed Ti plasmid contributed to no regeneration
385 event occurring is speculative. Though, from the callogenesis response to armed Ti transfection, it could be assumed that since the
386 transfection significantly affected callus weight and size, it would also greatly affect regeneration capacity, though further
387 investigation is required. If such an approach was to be implemented to obtain genetically modified cannabis, the input volume of
388 cotyledon explants used for regenreation would need to be exceedingly high to obtain sufficient regeneration numbers to capture
389 trasnformational events. There is a need to overcome the regeneration bottleneck from cannabis, either through use of an industry

390 monoculture or extensive protocol development of in-house genotypes, should significant investment be made into improving the
391 germplasm.
392
393 Direct regeneration from hypocotyls has recently been achieved using hormone free media (Galán-Ávila et al., 2020, 2021), greatly
394 reducing the requirements of cultivar specific genetics for regeneration capabilities. The several hemp cultivars used by Galán-Ávila
395 et al. (2020 and 2021) coupled with the outcrossed cultivar used within this study demonstrates this approach as a significant
396 improvement over using other explant types. Within the hemp varieties previously studied, approximately 50% saw shoot formation,
397 where in this study, the control (disarmed innoculation) saw 20% and only 1.1% of the transformed hypocotyls responded (Table
398 3). Whilst examples of transformation decreasing regeneration from hypocotyls were recorded (Galán-Ávila et al., 2021), the
399 reduction recorded here was far more prominent. Three responding transformed hypocotyls were recorded, however only 1 rooted
400 *in vitro*, whereas all 8 responding control (disarmed *Agrobacterium* innoculation) rooted successfully demonstrating the significant
401 effect transfection has on shoot regeneration recovery.
402
403 The effect of siRNA off-targeting within the highly homologous cannabinoid gene sets was previously demonstrated within leaf
404 explants (Matchett-Oates et al., 2021b), with significant up and downregulation recorded highlighting the lack of tolerance to siRNA
405 base-pair differences. Transformed callus within this study show stable integration of the RNAi vector, which shows increased
406 levels of significant down and upregulation within all transformed callus, bar one (Figure 10). CBDAS-homologs were most
407 significantly affected, downregulation wise, which is to be expected due to the siRNA generated targeting only within this gene set
408 (Matchett-Oates et al., 2021b). Interestingly, from the previous transient expression profiles after agroinfiltration, all three
409 cannabinoid transcript levels were significantly reduced, where within this study, while this does occur in 6 callus, significant
410 upregulation was also recorded most greatly in *THCAS* and *CBCAS* within 10 calli. The explant tissue type between this and the
411 previous study could explain the different expression profiles (Supplementary Figure 2). The callus contained lower transcript levels
412 in comparison to the leaf explant, which can be explained by the lack of developed organelles and biological structures causing a
413 differential expression profile. The effectiveness of the RNAi however within the expression profiles of transformed vs. control
414 callus cannot be understated, with many callus presenting significantly ($p = 0.001$) lower, or higher, transcript levels.
415
416 The generation of a normalised expression profile, from the 8 regenerated control hypocotyls, saw significant upregulation of all
417 cannabinoid biosynthesis genes in the singular regenerated transformed hypocotyl. Significant upregulation of all three genes was
418 not observed within the callus expression profiles, though examples of all three genes being significantly affected were recorded 6
419 times, exhibiting the effect siRNA off-targeting has on the transcript levels of the highly homologous gene sets; whether that be
420 through possible non-specific upregulation (Zirpel et al., 2018; Fulvio et al., 2021) or through transcription-translation interference
421 causing upregulation, as is seen using microRNAs (Vasudevan, 2012) and also within siRNA of mammalian cells (Portnoy et al.,
422 2011). *CBCAS* saw the largest increases in relative gene expression profiles in the regenerated transformed hypocotyl (Figure 11)
423 reaching levels approximate to that of *THCAS* upregulation within cotyledon callus mass #24. Within chemotypic profiles,
424 cannabichromenic acid (CBCA) is accumulated in much lower concentrations in comparison to the chemical products produced by
425 more highly efficient synthase genes, *THCAS* and *CBDAS* (Jin et al., 2021). This could help explain the significantly higher relative
426 expression profile; with few copies of the *CBCAS* transcript present, non-specific upregulation to maintain homeostasis due to the
427 downregulation of more highly efficient cannabinoid genes could see significantly increased transcript levels. Although
428 downregulation of these genes were not recorded, cleaving of mRNA by siRNA is still occurring, resulting in the upregulation of
429 these genes due to their innate ability to synthesise non-specific cannabinoids as has been previously shown (Zirpel et al., 2018;
430 Fulvio et al., 2021; Matchett-Oates et al., 2021b). To assess whether cannabinoid accumulation is significantly affected during
431 flowering requires further investigation.
432

433 This is the first report of a stably regenerated cannabis plant targeting the medicinally important cannabinoid biosynthesis genes for
434 modulated expression profiles. The data presented here demonstrates the ability to alter cannabinoid expression profiles in multiple
435 tissue types using RNA interference. Targeting individual synthase genes using this approach is seemingly not suitable, however
436 this provides great insight into the innate ability of gene upregulation to help synthesise non-specific cannabinoids to maintain
437 homeostasis. This report also highlights the significant investment that is required to obtain regeneration frequencies of
438 trasnformational events in cannabis with the intention to obtain desired chemotypic profiles.
439

440 **5. Conclusions**

441 Reported here is the first description of stably transformed cannabis explants and a regenerated stably transformed cannabis plant
442 with modified cannabinoid transcript levels. This work helps understand the mechanism of homeostasis within cannabis to
443 maintain cannabinoid biosynthesis in the event of downregulation. This report will also play a vital role in the understanding and
444 development of designer cannabis strains with modified expression profiles using genome editing. Furthermore, this study has
445 quantified the efficiencies of transformation and regeneration with their inherent limitations, to obtain and detail the relevant scale
446 in which to deliver the necessary range of transformants for obtaining desired novel chemotypes that will inform all future studies
447 pursing that end.

448 **Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Figure S1: Regenerated leaf primordia from multiple
449 transformed cotyledon callus messes, Figure S2: Leaf explant and callus explant relative expression levels of cannabinoid biosynthesis genes,
450 Table S1: Standard error and p-values of transformed cotyledon callus masses from $2^{-\Delta\Delta Ct}$ method using pooled control for normalization. First
451 numerical value represents relative expression followed by *p* value from technical triplicates, Table S2: Relative expression levels of cannabinoid
452 genes from transformed regenerated hypocotyl compared with 8 control regenerated hypocotyls. RT-qPCR primers. Vectors used within study.

453 **Author Contributions:** Conceptualization LMO, GS and NC. Methodology LMO and EM. Formal analysis LMO. Writing LMO. Review and
454 editing LMO, GS, NC. Supervision NC. All authors have read and agreed to the published version of the manuscript

455 **Funding:** This research was funded by Agriculture Victoria Research and Agriculture Victoria Services

456 **Data Availability Statement:** All available data can be found within manuscript and Supplementary Materials, with further enquiries being
457 directed to the corresponding author.

458 **Conflicts of Interest:** The authors declare no conflict of interest.

459 **References**

460 Beard, K. M., Boling, A. W. H., and Bargmann, B. O. R. (2021). Protoplast isolation, transient transformation, and flow-cytometric analysis of
461 reporter-gene activation in Cannabis sativa L. *Industrial Crops and Products*, *164*, 113360. https://doi.org/10.1016/j.indcrop.2021.113360
462 Borrelli, F., Fasolino, I., Romano, B., Capasso, R., Maiello, F., Coppola, D., … Di Marzo, V. (2013). Beneficial effect of the non-psychotropic
463 plant cannabinoid cannabigerol on experimental inflammatory bowel disease. *Biochemical Pharmacology*, *85*(9), 1306–1316.
464 https://doi.org/10.1016/j.bcp.2013.01.017
465 Braich, S., Baillie, R. C., Jewell, L. S., Spangenberg, G. C., and Cogan, N. O. I. (2019). Generation of a comprehensive transcriptome Atlas and
466 transcriptome Dynamics in Medicinal cannabis. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-53023-6
467 Braich, S., Baillie, R. C., Spangenberg, G. C., and Cogan, N. O. I. (2020). A New and Improved Genome Sequence of Cannabis sativa. *Gigabyte*,
468 *1*. https://doi.org/10.46471/gigabyte.10
469 Chandra, S., Lata, H., Khan, I. A., and ElSohly, M. A. (2010). Propagation of elite Cannabis sativa for the production of D9-Tetrahydrocannabinol
470 (THC) using biotechnological tools. *Medicinal Plant Biotechnology*, 98–114. https://doi.org/10.1055/s-0030-1251774
471 Chaohua, C., Gonggu, Z., Lining, Z., Chunsheng, G., Qing, T., Jianhua, C., … Jianguang, S. (2016). A rapid shoot regeneration protocol from the
472 cotyledons of hemp (Cannabis sativa L.). *Industrial Crops and Products*, *83*, 61–65. https://doi.org/10.1016/j.indcrop.2015.12.035
473 Chiou, T.-J., Aung, K., Lin, S.-I., Wu, C.-C., Chiang, S.-F., and Su, C. (2006). Regulation of phosphate homeostasis by microRNA in Arabidopsis.
474 *The Plant Cell*, *18*(2), 412–421. https://doi.org/10.1105/tpc.105.038943
475 Degenhardt, L., Lintzeris, N., Campbell, G., Bruno, R., Cohen, M., Farrell, M., and Hall, W. D. (2015). Experience of adjunctive cannabis use for
476 chronic non-cancer pain: findings from the Pain and Opioids IN Treatment (POINT) study. *Drug and Alcohol Dependence*, *147*, 144–150.
477 https://doi.org/10.1016/j.drugalcdep.2014.11.031
478 Deguchi, M., Bogush, D., Weeden, H., Spuhler, Z., Potlakayala, S., Kondo, T., … Rudrabhatla, S. (2020). Establishment and optimization of a

479 hemp (Cannabis sativa L.) agroinfiltration system for gene expression and silencing studies. *Scientific Reports*, *10*(1), 1–11.
480 https://doi.org/10.1038/s41598-020-60323-9

481 ElSohly, M. A., Radwan, M. M., Gul, W., Chandra, S., and Galal, A. (2017). Phytochemistry of Cannabis sativa L. In *Phytocannabinoids* (pp. 1–
482 36). Springer. https://doi.org/10.1038/s41598-020-60323-9

483 Farag, S. (2014). Cannabinoids production in Cannabis sativa L.: An in vitro approach.

484 Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., … Cannazza, G. (2021). Analysis of Sequence Variability and
485 Transcriptional Profile of Cannabinoid synthase Genes in Cannabis sativa L. Chemotypes with a Focus on Cannabichromenic acid synthase.
486 *Plants*, *10*(9), 1857. https://doi.org/10.3390/plants10091857

487 Galán-Ávila, A., García-Fortea, E., Prohens, J., and Herraiz, F. J. (2020). Development of a direct in vitro plant regeneration protocol from
488 Cannabis sativa L. seedling explants: developmental morphology of shoot regeneration and ploidy level of regenerated plants. *Frontiers in*
489 *Plant Science*, *11*, 645. https://doi.org/10.3389/fpls.2020.00645

490 Galán-Ávila, A., Gramazio, P., Ron, M., Prohens, J., and Herraiz, F. J. (2021). A novel and rapid method for Agrobacterium-mediated production
491 of stably transformed Cannabis sativa L. plants. *Industrial Crops and Products*, *170*, 113691. https://doi.org/10.1016/j.indcrop.2021.113691

492 Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., … Schwartz, C. J. (2021). A new Cannabis genome
493 assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytologist*.
494 https://doi.org/10.1111/nph.17243

495 Jin, D., Henry, P., Shan, J., and Chen, J. (2021). Identification of chemotypic markers in three chemotype categories of cannabis using secondary
496 metabolites profiled in inflorescences, leaves, stem bark, and roots. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.699530

497 Klein, T. W. (2005). Cannabinoid-based drugs as anti-inflammatory therapeutics. *Nature Reviews Immunology*, *5*(5), 400–411.
498 https://doi.org/10.1038/nri1602

499 Lata, H., Chandra, S., Khan, I. A., and ElSohly, M. A. (2010). High frequency plant regeneration from leaf derived callus of high Δ9-
500 tetrahydrocannabinol yielding Cannabis sativa L. *Planta Medica*, *76*(14), 1629–1633. https://doi.org/10.1055/s-0030-1249773

501 Lata, H., Chandra, S., Khan, I. A., and ElSohly, M. A. (2017). Micropropagation of Cannabis sativa L.—An Update. In *Cannabis sativa L.-Botany*
502 *and Biotechnology* (pp. 285–297). Springer. https://doi.org/10.1007/978-3-319-54564-6_13

503 Lata, H., Chandra, S., Khan, I., and ElSohly, M. A. (2009). Thidiazuron-induced high-frequency direct shoot organogenesis of Cannabis sativa L.
504 *In Vitro Cellular and Developmental Biology-Plant*, *45*(1), 12–19. https://doi.org/10.1007/s11627-008-9167-5

505 Li, H.-L. (1973). An archaeological and historical account of cannabis in China. *Economic Botany*, *28*(4), 437–448.

506 Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method.
507 *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

508 Majumdar, R., Rajasekaran, K., and Cary, J. W. (2017). RNA interference (RNAi) as a potential tool for control of mycotoxin contamination in
509 crop plants: concepts and considerations. *Frontiers in Plant Science*, *8*, 200. https://doi.org/10.3389/fpls.2017.00200

510 Mandolino, G., and Ranalli, P. (1999). Advances in biotechnological approaches for hemp breeding and industry. Haworth Press, Binghamton,
511 NY.

512 Matchett-Oates, L. Braich, S., Spangenberg, G.; Rochfort, S., and Cogan, N. (2021a). In silico analysis enabling genome editing in Medicinal
513 cannabis, gene families and variant characterisation. *PLoS One*. https://doi.org/10.1371/journal.pone.0257413

514 Matchett-Oates, L., Spangenberg, G., and Cogan, N. (2021b). Manipulation of cannabinoid biosynthesis via transient RNAi expression. *Frontiers*
515 *in Plant Science*. https://doi.org/10.3389/fpls.2021.773474

516 Matchett-Oates, L., Mohamaden, E., Spangenberg, G., and Cogan, N. (2021c). Development of a robust transient expression screening system in
517 protoplasts of cannabis. *In Vitro Cellular and Developmental Biology-Plant*. https://doi.org/10.1007/s11627-021-10178-0

518 Monthony, A. S., Kyne, S. T., Grainger, C. M., and Jones, A. M. P. (2021a). Recalcitrance of Cannabis sativa to de novo regeneration; a multi-
519 genotype replication study. *PloS One*, *16*(8). https://doi.org/10.1371/journal.pone.0235525

520 Monthony, A. S., Page, S. R., Hesami, M., and Jones, A. M. P. (2021b). The past, present and future of Cannabis sativa tissue culture. *Plants*,
521 *10*(1), 185. https://doi.org/10.3390/plants10010185

522 Movahedi, M., Ghasemi-Omran, V., and Torabi, S. (2015). The effect of different concentrations of TDZ and BA on in vitro regeneration of

523 Iranian cannabis (Cannabis sativa) using cotyledon and epicotyl explants. *Journal of Plant Molecular Breeding*, *3*(2), 20–27.
524 https://doi.org/10.22058/jpmb.2015.15371
525 Page, S. R. G., Monthony, A. S., and Jones, A. M. P. (2020). Basal media optimization for the micropropagation and callogenesis of Cannabis
526 sativa L. *BioRxiv*. https://doi.org/10.1101/2020.02.07.939181
527 Paul, S., Datta, S. K., and Datta, K. (2015). miRNA regulation of nutrient homeostasis in plants. *Frontiers in Plant Science*, *6*, 232.
528 https://doi.org/10.3389/fpls.2015.00232
529 Portnoy, V., Huang, V., Place, R. F., and Li, L. (2011). Small RNA and transcriptional upregulation. *Wiley Interdisciplinary Reviews: RNA*, *2*(5),
530 748–760. https://doi.org/10.1002/wrna.90
531 Schachtsiek, J., Hussain, T., Azzouhri, K., Kayser, O., and Stehle, F. (2019). Virus-induced gene silencing (VIGS) in Cannabis sativa L. *Plant*
532 *Methods*, *15*(1), 1–9. https://doi.org/10.1186/s13007-019-0542-5
533 Slusarkiewicz-Jarzina, A., Ponitka, A., and Kaczmarek, Z. (2005). Influence of cultivar, explant source and plant growth regulator on callus
534 induction and plant regeneration of Cannabis sativa L. *Acta Biologica Cracoviensia Series Botanica*, *47*(2), 145–151.
535 Szaflarski, J. P., and Bebin, E. M. (2014). Cannabis, cannabidiol, and epilepsy—from receptors to clinical response. *Epilepsy and Behavior*, *41*,
536 277–282. https://doi.org/10.1016/j.yebeh.2014.08.135
537 Touw, M. (1981). The religious and medicinal uses of Cannabis in China, India and Tibet. *Journal of Psychoactive Drugs*, *13*(1), 23–34.
538 https://doi.org/10.1080/02791072.1981.10471447
539 Turner, C. E., and Elsohly, M. A. (1981). Biological activity of cannabichromene, its homologs and isomers. *The Journal of Clinical*
540 *Pharmacology*, *21*(S1), 283S-291S. https://doi.org/10.1002/j.1552-4604.1981.tb02606.x
541 Vasudevan, S. (2012). Posttranscriptional upregulation by microRNAs. *Wiley Interdisciplinary Reviews: RNA*, *3*(3), 311–330.
542 https://doi.org/10.1002/wrna.121
543 Wang, R., He, L.-S., Xia, B., Tong, J.-F., Li, N., and Peng, F. (2009). A micropropagation system for cloning of hemp (Cannabis sativa L.) by
544 shoot tip culture. *Pak. J. Bot*, *41*(2), 603–608.
545 Wielgus, K., Luwanska, A., Lassocinski, W., and Kaczmarek, Z. (2008). Estimation of Cannabis sativa L. tissue culture conditions essential for
546 callus induction and plant regeneration. *Journal of Natural Fibers*, *5*(3), 199–207. https://doi.org/10.1080/15440470801976045
547 Zhang, X., Xu, G., Cheng, C., Lei, L., Sun, J., Xu, Y., … Chen, X. (2021). Establishment of an Agrobacterium-mediated genetic transformation
548 and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (Cannabis Sativa L.). *Plant Biotechnology Journal*.
549 https://doi.org/10.1111/pbi.13611
550 Zirpel, B., Kayser, O., and Stehle, F. (2018). Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L.
551 *Journal of Biotechnology*, *284*, 17–26. https://doi.org/10.1016/j.jbiotec.2018.07.031
552
553

CHAPTER 6

General Discussion

6.1 Background and context of the research

Cannabis is one of the earliest plants cultivated in China, dating back to 4000 BC for its use in textile making, fibre for paper (Zuardi, 2006) and its unique psychoactivity, which provides pain treatment (Martín-Sánchez et al., 2009), anti-inflammation (Pellati et al., 2018) and increased appetite (Hussain et al., 2015). Cannabis contains over 120 phytocannabinoids which target the endogenous cannabinoid receptors that play a crucial role in homeostasis (Pacher et al., 2006). It is these unique cannabinoids which have found recent medicinal purpose in conditions which have been previously untreatable with current medical and pharmaceutical approaches, such as grand mal epileptic seizures in children with Dravet Syndrome (Devinsky et al., 2018; McCoy et al., 2018). Due to the illegal status of the plant since as early as the 14th century (Johnson, 2019), it has been illicitly selected for higher THC content, whilst not bottlenecking the rest of the genome and preserving much of the underlying diversity present within the gene pool. Whilst THC contains medicinal properties that enables it to replace many opiate-based pharmaceuticals, there has also been a recent interest in CBD and its properties, leading to cross breeding of traditional hemp and drug-types. Whilst precision breeding in cannabis is beginning to improve the germplasm for desirable traits and chemotypic profiles (Naim-Feil et al., 2021), the time required between breeding cycles can take years before a desirable trait or chemovar is produced. Regarding the cannabinoid synthase genes, the exact mechanism in which it is regulated is still being debated. With many pseudogenes residing within the nested repeats of the *CBDAS* locus (Grassa et al., 2021), breeding to better regulate or silence all potential mechanisms of regulation for the accumulation of the upstream cannabinoid, CBGA, will be both highly challenging and costly in resources and time. Genome editing provides the ability to increase, introduce or reduce the presence of a trait in as short a time frame as a single plant generation (Cao et al., 2016). Genome editing can be used to accurately modify the important cannabinoid biosynthesis pathway to create unique chemovars of cannabis with tailored cannabinoid levels. Through genome editing, medical products can be more readily produced without the need for expensive and time-consuming laboratory-based interventions in postproduction via formulation or purification. The ability to provide optimal plant material directly without the need of manufacturing steps will greatly increase the availability of the product to the patient, whilst also reducing the overall costs.

For successful genome editing to occur, robust and repeatable tissue culture protocols are necessary to generate the range of events that then are accurately assessed. A precise evaluation of the edited event needs to be recorded and compared to pooled genotypes from identical transformational events. In cannabis, cultivars of hemp, such as Finola (Galán-Ávila et al., 2021) and Epsilon 68 (Wróbel et al.,

2020) are routinely used for tissue culture protocol development. Drug type cultivars are also commonly used (Lalge et al., 2016; Lata et al., 2009; Mestinšek-Mubi et al., 2020) however, significant variation exists between protocol approaches in regards to media and cultivars used. For example, regarding the first step of producing sterile explants for genome editing, micropropagation approaches in cannabis are widely varied with different basal media such as MS, B5, DKW, BABI and WPM being used (Page et al., 2020). The ability to transform multiple explant tissues improves the catalogue of transformational approaches that can be implemented to advance the germplasm. Unlike other agronomically important crops, cannabis lacks critically important robust protocols therefore limiting genome editing. Regeneration also remains a bottle neck for cannabis, with limited reports existing for specific cultivars (Galán-Ávila et al., 2020; Plawuszewski et al., 2005; Slusarkiewicz-Jarzina et al., 2005), demonstrating the lack of a set of robust protocols applicable to different cultivars. Re-assessment of published protocols has also shown that a lack of reproducibility using the same cultivars also exists (Monthony et al., 2020a), highlighting the recalcitrant and genetically diverse nature of cannabis.

Using protoplasts as the explant source for transformation contains unique advantages over other techniques, specifically the potential for thousands of transformational events from a single genotype and the ability to avoid chimerism through selection. Transformation of cannabis protoplasts has been reported with the expression of GFP at transformational efficiencies proficient enough for full-scale studies (Beard et al., 2021). Using protoplasts explants, CRISPR/CAS-9 genome editing has also been achieved targeting the *PDS* gene with different sgRNA designs to assess indel frequencies (Zhang et al., 2021). The protocols however are based within hemp lineages or are genotype specific, further highlighting the necessity to develop more robust protocols to isolate and transform protoplasts in cultivars regardless of chemotypic profiles.

As an alternative approach, *Agrobacterium*-mediated transformation is an efficient approach for genetic improvement that has had great success in important agronomical crops such as wheat (Zhang et al., 2018), corn (Ishida et al., 2007) and canola (Mashayekhi et al., 2008). *Agrobacterium*-mediated transformation in cannabis is possible, though the limiting step remains for regeneration. Such an example has been demonstrated in stem and leaf explants of hemp varieties that were recalcitrant to embryogenesis and organogenesis (Feeney & Punja, 2003). To optimise the efficacy of *Agrobacterium* infection and transgene integration, variables involved in transformation and choice of explant needs to be considered. Within cannabis, such variables involved in transformation efficiencies have been investigated (Deguchi et al., 2020), including explant response variation to *Agrobacterium* transformation efficiency (Galán-Ávila et al., 2021), however such results should be considered cultivar or genotype specific. The implementation of these newly developed protocols targeting the medicinally, and economically, important cannabinoid biosynthesis genes could help

greatly understand their mode of action, non-specific cannabinoid biosynthesis capabilities and applicably towards genome editing efforts to manipulate expression levels.

RNA interference (RNAi), a post-transcriptional gene silencing approach, offers an alternative approach to modern, more accurate genome editing techniques, such as CRISPR/Cas-9. A particular advantage RNAi offers is the ability to affect genes with high levels of homology due to off-targeting. Within the cannabinoid biosynthesis genes, significantly high levels of sequence homology exist with multiple homologs and pseudogenes present thus increasing the likelihood that the use of RNAi will result in significant modification of cannabinoids produced. Using RNAi still allows the ability to improve the cannabis germplasm whilst avoiding any potential licensing and permit restrictions relating to genome editing technologies. Recently, two different approaches to RNAi, Virus Induced Gene Silencing (VIGS) (Schachtsiek et al., 2019) and Hairpin RNAi (HpRNAi) (Deguchi et al., 2020) have been reported for cannabis. Similar to previous tissue culture and transformational protocols published, these RNAi studies have explored the variables involved in transformational efficiencies using reporter genes. The tools developed will be indispensable for future cannabis RNAi studies through implementation and modification to target cannabinoid biosynthesis genes.

The development of tissue culture protocols have allowed for transformational studies to begin to be published containing modified developed protocols from previously published reports (Galán-Ávila et al., 2021). Currently lacking within the area of transformational approaches for genetic improvement of cannabis, regardless of transformational or genetic engineering approach, is the targeting of the medically and economically important cannabinoid biosynthesis genes. Comprehensive genome analysis of the highly homologous cannabinoid genes can allow for intelligent design of genome editing vector constructs to help create unique chemovars, which will have great medical and economical potential. The ability to produce tailored chemotypic profiles in cannabis using readily available genetic engineering approaches, and the growing number of published protocols in tissue culture, should be explored to improve the medicinal potential of cannabis.

6.2 Overview of the research

A comprehensive, long-read sequenced genome of cannabis was required to begin developing genome editing constructs targeting all genes involved in cannabinoid biosynthesis. Mining the Cannbio-2 reference genome for cannabinoid biosynthesis genes using publicly available sequence data provides an invaluable genomic resource to which comparative analysis of publicly available genomes was possible, but also a large data-set pan-genome was probed for sequence variation. Using Cannbio-2 gene sequences (Braich et al.,2020) to probe Finola, PK and CBDRx, copy number variation between published cannabis genomes was reported. This analysis is of importance to determine the completeness and variation within the genome assemblies, but also as evidence supporting the chemotypic inheritance model proposed by Grassa et al. (2021). Prior to this model, the assumption

that chemotypic profiles were inherited through classical Mendelian inheritance of the B locus were accepted (de Meijer et al., 2003). It is not until recently that the presence of pericentric nested repeats of two locus containing several pseudogenes could be elucidated with confidence (Grassa et al., 2021). From the analysis, copy number variance in multiple genes are present between the publicly available genomes, with greatest variance residing within cannabinoid biosynthesis genes. The extent of copy number variance within these loci of the genome discovered further ads to the understanding of the recombination and replication mechanism during breeding.

Using the Cannbio-2 reference gene sequences, the pangenome was probed to produce consensus sequences for all genes involved in cannabinoid biosynthesis, thus creating an invaluable genetic resource with known SNP locations within a diverse gene pool. Though SNP data exists in cannabis (McKernan et al., 2020), in-depth knowledge on SNP location within each cannabinoid synthase gene is lacking, creating the issue of cultivar specific approaches to genome editing. This SNP data generated in this thesis is critical to design intelligent sgRNA capable of targeting individual cannabinoid synthase genes regardless of cultivar used by avoiding known sequence variation locations. It is currently unknown if knocking out of specific cannabinoid biosynthesis genes, through genome editing, results in modified cannabinoid accumulation. Further, it is unknown if knocking out upstream genes will result in variable cannabinoid concentrations, though evidence in certain genes involved in isoprenoid synthesis suggests abnormal plant development occurs (Araki et al., 2000; Estévez et al., 2001), though this has not been explored in cannabis. Development of a comprehensive sgRNA catalogue targeting each gene applicable across cultivars forwards the medicinal potential cannabis. The knowledge of CNVs and SNPs within the cannabis gene pool is crucial in implementing a broad-stroke-approach of genome editing in cannabis to improve the germplasm.

Development of multiple transformational approaches to improve the germplasm is crucial in a highly diverse gene pool. Domestication of cannabis throughout time, geared towards a higher THC content, has led to a highly polymorphic subspecies with significant phenotypic responses to growth conditions. Biotechnological approaches to improve the germplasm of cannabis is currently in its infancy requiring significant efforts into successful transformation protocols and increasing efficiency. Before expensive and time consuming full-scale transformational studies using genome editing approaches are attempted, an effective transient expression system should first be established to evaluate the effectiveness of the vector constructs in delivering the desirable genome editing event. The choice of protoplasts for transformational studies provide a significant level of quantification from genome editing events within a single genotype detailing efficacy, with such approaches widely used in tobacco (Bossche et al., 2013), maize and *Arabidopsis* (Sheen, 2001). Cannabis, however, is lacking the fundamental protocols for protoplast isolation and transformation applicable to a wider gene pool severely limiting the biotechnological approaches to improve the germplasm. The lack of a cell wall renders protoplasts incredibly sensitive to osmotic stress leaving little room for favourable

culture conditions. Statistical analysis of the variables involved in protoplast isolation were produced, providing crucial data on the significance that a range of treatment conditions responsible in viable protoplast collection provide. The development of a robust, multi-cultivar approach to address the favourable culture conditions to isolate healthy, viable protoplasts significantly accelerates cannabis' genetic improvement potential. Generating a substantial concentration of protoplasts allows for the transfer of vector constructs using an array of approaches such as PEG (Lazzeri et al., 1991), particle bombardment (Wang et al., 1988) and direct DNA microinjection (Masani et al., 2014). Similar to protoplast isolation, considerable variables are at play dictating the successful (and efficient) transfer of vector constructs into the protoplast genome. In regards to PEG, considerable concentrations are lethal, severely reducing transmissibility of genetic transformation (Wang et al., 2021). Common variables involved, such as plasmid concentrations and incubation period have been explored in several species (Cao et al., 2014; Li et al., 2018; Wang et al., 2021). Such investigations do not exist in cannabis protoplasts leaving the optimal conditions open to speculation. Proof of concept in protoplast transformation using common reporter genes, such as *GFP* or *GUS*, were lacking leaving the question of cannabis' protoplast as a potential for genome editing open-ended.

Isolation and transformation of cannabis protoplasts has now been achieved, producing a significant concentration of viable protoplasts with transformation efficiencies sufficient for stable transformation applications (Beard et al., 2021; Matchett-Oates et al., 2021c; Chapter 3). Much like micropropagation and regeneration attempts in cannabis, the chemotypic profile used within Beard *et al.* (2021) contains a hemp lineage, which is phenotypically and genetically significantly dissimilar to drug-type cultivars. Whilst extensive studies into the variation in phenotypic response to culture conditions between high CBD vs high THC cultivars have not yet been produced, it can be assumed such variable responses exists from meta-analyses into differing responses to tissue culture treatments (Monthony et al., 2021b). In this thesis, exploration into the variables involved in protoplast isolation and transformation in high THC cultivars provide an invaluable understanding into protocol optimisation for cannabis cultivars. Understanding the statistical significance each variable contributes to viable protoplast isolation and transformation is crucial as the first step towards developing robust and highly effective protocols for genetic improvement. The data produced also lays important groundwork into an educated foundation when developing cultivar specific approaches to this genetic improvement approach.

Highly successful cultivars used in genome editing are often adept to multiple biotechnological approaches. Development of variable methodologies towards improving the genome in new, emerging species, such as cannabis, is fundamental to address the inherent recalcitrant nature of cannabis to regeneration (Masani et al., 2014). Utilizing the previously reported protocol for *Agrobacterium* transformation in several explant types, and the demonstration of the effectiveness of RNAi mechanism in cannabis (Deguchi et al., 2020), targeting of the cannabinoid biosynthesis genes using

this protocol is relatively uncomplicated. RNAi as an approach to post-transcriptionally silence gene expression through the production of small-interfering RNA (siRNA) that share homology to the target gene sequence that has been used extensively before the introduction of CRISPR and TALENs (Senthil-Kumar & Mysore, 2010). The complexity in achieving downregulation using RNAi is due to the high homology of the synthase genes, requiring in-depth sequence information for the target cannabinoid biosynthesis genes. Approaches such as CRISPR/Cas-9 can exploit single base pair polymorphisms due to the high efficiency of the Cas protein (Zhou et al., 2015), with cleavage occurring 3bp upstream from the PAM site. However, RNAi does not share this specificity as the ~21-mer siRNA acts as the guide and location for cleavage from AGO proteins. Utilisation of RNAi to silence cannabinoid biosynthesis genes is then hypothesised to exhibit severe levels of off-targeting in the highly homologous gene sets. Interestingly, it is this high level of off-targeting that gives RNAi the advantage over more precise genome editing technologies for cannabinoid downregulation due to high sequence homology. This is precisely observed through the transient downregulation of closely related genes in leaf explants (Matchett-Oates et al., 2021b; Chapter 4). The off-targeting exhibited by the siRNA can only be speculatively predicted due to the lack of prediction software with cannabis genome sequence information but it is likely anticipated due the complexity of the system. Regardless, siRNA with several base pairs of differences is clearly tolerated using RNAi with significant silencing still occurring. Alternatively, a shortcoming of RNAi is its inability to specifically target a single cannabinoid biosynthesis gene, seriously limiting functional studies approaches to unravel the potential synthase activity of the many pseudogenes present in published datasets (Matchett-Oateset al., 2021a; Chapter 2). However, the ability to silence all the synthase genes simultaneously provides an opportunity to observe the effect such downregulation plays on upstream precursors, such as CBGA. This phytocannabinoid is reported to have great potential medicinal benefit (Anderson et al., 2021) and is the precursors to THCA, CBDA and CBCA. This phytocannabinoid is efficiently synthesized into the aforementioned acidic forms resulting in low levels of CBGA accumulating within female flowers. The ability to produce a novel chemovar with elevated levels of CBGA and significantly reduced levels of downstream cannabinoids will be valuable to greater understand the mechanisms of the biosynthesis pathway.

Interestingly, upregulation of transcript levels are observed in certain cannabinoid synthase genes when treated with sequence specific dsRNA constructs (Matchett-Oates et al., 2021b; Chapter 4; Matchett-Oates et al., 2021; Chapter 5). Upregulation, as a form of gene activation, can play a vital role in homeostasis of plants to ensure critical processes are maintained. Such an example could explain upregulation in cannabis, with examples of non-specific synthase gene’s ability in yeast to synthesise non-specific cannabinoids (Peet et al., 2016). siRNA interference in protein translation can attribute to feedback loops resulting in higher transcript levels accumulating (Gil-Humanes et al., 2008; Portnoy et al., 2011; Scacheri et al., 2004), potentially explaining such upregulation reported in

cannabis. As previously mentioned, though a more specific approach (such as CRISPR/Cas-9) could achieve single homolog knock outs leading to greater understanding of genes and pseudogenes roles in the cannabinoid biosynthesis pathway, RNAi provides the opportunity to study the effect of complete pathway downregulation through a single, effective vector construct. This will prove incredibly valuable in unravelling the importance this particular pathway plays in the plant's homeostasis.

Having demonstrated the effectiveness of RNAi to significantly silence cannabinoid biosynthesis genes, establishing an effective cultivar specific regeneration protocol to recover transformation events are significant advancements in cannabis' development, that have now been delivered. Notoriously difficult to encourage regeneration, due to the high levels of inter-species variation, cannabis can in some instances undergo direct and indirect organogenesis and embryogenesis (Flores-Sanchez et al., 2009; Galán-Ávila et al., 2020; Mandolino & Ranalli, 1999; Plawuszewski et al., 2005; Slusarkiewicz-Jarzina et al., 2005; Wielgus et al., 2008). However, regeneration approaches are cultivar specific, and in certain circumstances not reproducible when using the same cultivar (Monthony et al., 2020a). The importance of a regeneration protocol in cannabis for different biotechnological approaches, such as protoplast division and regeneration or direct embryogenesis from somatic tissue, is crucial to improve the ability to manipulate the cannabis genome. Tissue specificity responses to regeneration is seemingly cultivar specific (Monthony et al., 2020b), with rudimentary empirical data collection on tissue explant reaction to limited media compositions a reasonable approach to evaluating tissue regeneration capabilities. Ideally, non-meristematic somatic tissue, such as leaf or petiole, such as in corn (Ahmadabadi et al., 2007) and common bean (Veltcheva & Svetleva, 2005), could be used for transformational and regeneration in cannabis. However, currently, limited (in their nature) explant choices including cotyledons (Chaohua et al., 2016; Movahedi et al., 2015) and hypocotyls (Galán-Ávila et al., 2020), are commonly used. Hypocotyl explant choice, and cotyledons depending on age which they are excised, causes considerable difficulties as each seed used contains a different genetic composition.

Whilst transformation of cannabis has previously been achieved as previously discussed, regeneration from such an event still proves difficult (Zhang et al., 2021). Proof-of-concept gene editing cassettes, such as silencing *PDS* to produce albino phenotypes has been achieved (Galán-Ávila et al., 2021), opening the possibilities of using the developed protocol with genome editing constructs targeting the cannabinoid biosynthesis genes. Integrating the highly effective RNAi cassette via *Agrobacterium* in an attempt to regenerate stable transformants from the commonly used explants, cotyledon and hypocotyls, it is possible to produce the first novel cannabis strain with modified cannabinoid content. As with previous attempts at regenerating cotyledons post transformation (Mandolino & Ranalli, 1999; Movahedi et al., 2015), no successful shoot regeneration occurred, however evidence of the

RNAi cassettes ability to downregulate synthase transcript levels were evident in relative expression levels of most calli masses (Matchett-Oates et al., 2021; Chapter 5), producing a large array of different transcript profiles. Alternatively, using hypocotyls in a hormone free media should result in less recalcitrance to regeneration due to the pericycle cells' innate ability to regenerate. The presence of a selective media to suppress *Agrobacterium* growth has previously been shown to decrease the explant response to regeneration (Galán-Ávila et al., 2021), creating a slight hurdle to this simple regeneration protocol. The application of the highly effective RNAi cassette, using this proven regeneration protocol, has produced the first stably transformed cannabis genotype with modified cannabinoid transcripts, producing a significant advancement towards improving the germplasm and generation of novel cannabinoid profiles. The demonstration, as seen in the leaf explants (Matchett-Oates et al., 2021b; Chapter 4) and in callus transcript levels (Matchett-Oates et al., 2021; Chapter 5), that cannabis can upregulate these cannabinoid biosynthesis genes to maintain a level of homeostasis requires further considerations into biotechnological approaches to target either individual genes or the gene set with accuracy and efficiency to modulate expression levels. Although modifying an individual gene to repress specific cannabinoid accumulation may be more difficult than anticipated, the discovery of this homeostasis *in planta* confirms the previous discoveries of non-specific cannabinoid biosynthesis in yeast (Peet et al., 2016) and the investigations into THC accumulation in hemp varieties with an inactive *THCAS* gene (Fulvio et al., 2021).

6.3 Future Directions

With increasing published research in all aspects of cannabis tissue culture and genetic analysis, research has demonstrated the highly diverse culture conditions and genetic make-up individual cultivars require and contain (Matchett-Oates et al., 2021a,b; Chapters 2 and 3; Monthony et al., 2021). Developing multiple approaches to achieve the same goal (whether that be micropropagation, regeneration or transformation) isn't an ideal scenario for the genetic improvement of cannabis. Whilst producing a broader array of protocols, exploring the effects of plant growth regulators and explant types on regeneration is important. However, a more comprehensive approach with robust and generalised protocols applicable to a wider gene pool are necessary. With the advancements reported within this thesis, invaluable data has been generated to the continuing research to achieve these goals. An important aspect to consider, regarding the genetic improvement of cannabis using genome editing, is the underlining acceptance of genetically modified organisms. The challenging legal and licensing environment that surrounds genome editing technologies, such as CRISPR/Cas-9 and RNAi, for commercial deployment provides some barriers to its use in the delivery of improved genotypes. However, many of the challenges and obstacles identified in transgenics are not present in medicinal cannabis cultivation. For example, medicinal cannabis cultivation requires female plants for inflorescent production eliminating the potential for male pollen to escape strict containment. Further, without the presence of pollen, gene flow between plants is extremely unlikely. Finally, with cannabis

being highly regulated, vegetative propagation is used to maintain genetics to ensure consistent chemotypic profiles are preserved. These factors lead to the maintenance of a monoculture where the transformational event is contained, reducing the concern for social acceptance.

With the use of genome editing technologies, novel plant genetics can be produced that are similar to plants produced using traditional breeding techniques creating a cross-over of regulatory boundaries to GMO regulations and acceptance (Camacho et al., 2014). With the final GM (genetically modified) product being ingested, social acceptance in Europe is lower compared to North America (Lucht, 2015). Though survey participants were accepting that GM crops were beneficial, concerns were still present regarding undefined perceived risks. A degree of separation between GM product and human interaction, such as GM crop development for livestock feed, is seemingly more widely acceptable with the traceability of GM exposure to livestock unquantifiable (Nadal et al., 2018). Pharmaceutical production using GM microbes for the production of synthetic chemicals are extensively used (Petsas & Vagi, 2019). Such approaches using GM are crucial for human health with social acceptance for these medications scoring highly (Olynk et al., 2017). It seems that the social acceptance of GM crops involves direct consumption and whether it is classified as a ‘food’. For the case of cannabis, containing significant medical potential, the development of GM crops for pharmaceutical products, such as oils, caps and dried flower requires sensible regulations and communications to society on the benefits such modifications to the genome will contain.

Within this thesis, a critical in-depth analysis of the economically and medicinally important cannabinoid biosynthesis pathway has revealed further degrees of complexity regarding the implementation of genome editing. Regarding genetic analysis, the complexity of the cannabinoid biosynthesis pathway, due to long terminal repeat retrotransposons within the genome, needs to be further investigated more broadly across a much larger gene pool of cannabis cultivars. The presence of copy number variance within the synthase genes (Chaohua et al., 2016; Grassa et al., 2021; Matchett-Oates et al., 2021a; Chapter 2) discovered within this thesis increases the difficulty in developing genome editing constructs targeting specific homologs within the genome, though now significant progress has been made with the genetic resources developed described here.

Understanding the role unique synthase gene homologs and pseudogenes play in cannabinoid biosynthesis is yet to be addressed as genetic transformation studies in cannabis are currently limited. Due to the difficulty in correctly identifying synthase copy number within the genome due to the high levels of duplication from retrotransposons (Braich et al., 2020; Grassa et al., 2021) , and identifying regions of sequence variation where precise genome editing can be used (Matchett-Oates et al., 2021a; Chapter 2), elucidation into their respective roles is currently speculative. The sgRNA catalogue developed in this thesis, as well as the genetic resource of known SNP locations within all the genes involved in cannabinoid biosynthesis, will be invaluable for genome editing attempts in a

large diverse gene pool due to the large pangenome in which it is developed. *In vivo* experimental data into the characterisation of the developed sgRNA is required to choose highly effective sgRNA constructs. Using the genomic resources within this thesis to further comprehend the effect pseudogenes within the genome have on cannabinoid biosynthesis can be investigated. The presence of several THCAS-like and CBDAS-like homologs within Cannbio-2 (Braich et al., 2020), CBDRx (Grassa et al., 2021), Purple Kush and Finola (Van Bakel et al., 2011) requires investigation into whether or not such homologs effect biosynthesis. Analysis of defective signal peptides has previously been demonstrated with experimental data that some specific homologs have no synthase activity despite the high levels of sequence homology (Taura et al., 2007), though the effect such pseudogenes have is still being debated. Mining of the Cannbio-2 genome has revealed several potentially functional copies of CBDAS-like homologs, which at the nucleotide level are highly homologous to *CBDAS*, however translated proteins show divergence into a new clade (Matchett-Oates et al., 2021a; Chapter 2). Investigating these genes' role in cannabinoid biosynthesis using precise genome editing approaches, such as CRISPR/Cas-9, will be greatly beneficial in understanding if *CBDAS* sequence divergence is occurring resulting in functionally identical homologs with relatively large sequence variation.

This thesis has demonstrated the need of more precise genome editing approaches, such as CRISPR/Cas-9, to target individual homologs to fully elucidate all the gene interactions and feedback mechanisms at work in the final cannabinoid biosynthesis step. Using the sgRNA catalogue designed in this thesis, coupled with the developed transient expression protocol, different sgRNA constructs can be assessed for their ability to knock-out targeted genes. Although base pair variance is tolerated using sgRNA (Anderson et al., 2015), the significance of off-targeting should be considerably reduced compared to siRNA generated through dsRNA. The easy to assemble RNAi vector, using Golden Gate Cloning, can be easily modified to target any gene of choice for functional studies in cannabis. Whilst the medically important cannabinoid genes are of highest interest for silencing, the genes located within the adjacent pathways (MEP, GPP and Hexanoate) provide excellent opportunities to assess the effect of silencing these genes involved in isoprenoid production have on cannabinoid accumulation downstream. Through simple PCR amplification and restriction enzyme assembly, these questions can now be addressed in cannabis, which will also have scientific merit outside the cannabis species, as the aforementioned pathways exist in other species that create isoprenoids.

Even through the use of hormone-free regeneration media, as such used in this thesis and previously published reports (Galán-Ávila et al., 2021), large variation in efficiency exists between cultivars causing serious bottle-necks in improving cannabis' genome. However, using the cross-cultivar approach of hormone free media with hypocotyls allows the application of genome editing constructs, such as those used in this thesis and can be implemented in a high throughput manner. The use of cotyledons as an explant source across multiple cultivars has been previously demonstrated great

potential (Chaohua et al., 2016; Movahedi et al., 2015) as a source for genetic transformation. The highly effective transformation protocol used within this thesis indicated high levels of stable gene integration through GFP expression post several weeks transformation. Promising signs are evident through the regeneration of leaf primordia on limited calli masses, indicating favourable conditions of culture were partially met. Although cotyledons are restrictive in their nature, the ability to successfully subculture this explant indefinitely *in vitro* can potentially help overcome this shortfall. Whether or not such 'infinite' sub-culturing conditions can be achieved is yet to be determined within the cultivar used for this study. However, given optimal conditions, calli cultures can be maintained indefinitely (White, 1939). With the development of hormone-free hypocotyl regeneration seemingly possible across multiple cultivars (as is evident in this thesis and previous reports (Galán-Ávila et al., 2021)), this approach has great potential to regenerate transformational events in hemp and drug-like cultivars alike. Whilst the control frequency of regeneration observed is considerably less than reported, none-the-less, shoot regeneration does occur and serves as an excellent approach for genetic improvement. The application of this simple protocol coupled with the efficient RNAi construct in this thesis, demonstrates the straightforward approach to targeting the medicinally important cannabinoid genes without the need for complex media cultures and tailored hormone compositions. The removal of complex media compositions in other approaches previously summarised (Monthony et al., 2021) has significantly reduced the time needed to produce transformants. Coupling this new protocol with a single cloning step, and a highly efficient RNAi vector, generating stable transformants has become relatively straightforward. As such, the ease to which this approach can be modified to effectively target any gene of choice within cannabis cannot be understated. Gene functional studies into the several homologs within synthase genes, or the effect upstream gene regulation plays in plant homeostasis can now be relatively easily explored. With so little known about the regulatory mechanisms behind unique cannabis genes, this new simplified approach can offer great genetic resources for answering such questions.

From the results in this thesis, the use of RNAi can significantly silence synthase transcript levels. It has also been shown that significant upregulation can occur, suggesting the ability of synthase enzymes ability to synthesise non-specific cannabinoids *in vivo*, such is the case in yeast models (Peet et al., 2016). With this being the first report of specific targeting of this important pathway, the ability to expand the findings into increasing RNAi efficiency and the application in stable transformations is promising. The effectiveness of RNAi is arguably determined by the length of the dsRNA introduced through a vector (He et al., 2020), the position in which the siRNA targets (i.e. earlier exon regions) and the sheer number of siRNA generated (Majumdar et al., 2017). The strongest gene silencing within this thesis occurred using a shorter dsRNA construct, further investigations into the optimal length of dsRNA should be performed, as has been previously (Höfle et al., 2020). Due to the size of dsRNA used in hairpin RNAi constructs, and low levels of sequence variation within this targeted

gene set, accurately silencing a single homolog could prove difficult. Significant off-targeting was demonstrated when a more highly diverse dsRNA construct was used only containing precise targets within CBDAS homologs. Whilst this approach is seemingly inappropriate for targeting a single homolog, using such an approach is beneficial when investigating broad gene silencing effects on cannabinoid accumulation with a single construct increasing accumulation of the upstream cannabinoid, CBG. Investigations into whether this scenario occurs, or whether cannabinoid biosynthesis continues through non-specific cannabinoid synthesis can provide great insight into the current largely unknown mechanisms.

To facilitate the future genetic improvement of cannabis, the development of efficient pipelines in genome editing, such as the sgRNA catalogue developed within this thesis, can help progress the push for personalised medication. The potential using genome editing has in producing tailored medicinal cannabis cultivars for specific conditions is incredibly valuable (Mathur & Sutton, 2017). The current lack of published reports using genome editing to target any gene within cannabis opens this work produced here to the potential of using these genetic resources to begin tailoring genetic modification events with known medical conditions. A pharmacogenetic approach to understand cannabinoid interaction within the body, whether that be receptors, transporters or bioactivation proteins encoded by certain genes (Hryhorowicz et al., 2018) can be used to match individual patients conditions and genetic predispositions to medicinal cannabis. This approach will benefit from the resources developed here and could have far reaching possibilities to produce a cannabis plant with personalised cannabinoid levels.

Through the development of a protoplast isolation and transformation protocol in this thesis, developing a protoplast regeneration protocol to advance this research requires significant investment. With so few reports of protoplast isolation now reported (Beard et al., 2021; Morimoto et al., 2007; Zhang et al., 2021), no attempts at regenerating transformational events have been described. As with many important crop species, such as rice (Abdullah et al., 1986) and potato (Haberlach et al., 1985), the ability to recover numerous transformational events simultaneously greatly improves the likelihood of recovering a genome editing event with desirable characteristics, such as knock outs. Investigations into the applicability of this protocol on the widely used fibre-types in regeneration would be of great interest, with most regeneration protocols being developed with this chemotype of cannabis (Monthony et al., 2021). The first hurdle to overcome is the instigation of protoplasts to regenerate the cell wall in favourable culture conditions. As with the approach taken in this thesis, a highly effective and time saving approach would be using orthogonal arrays to explore the variables involved in regenerating the cell wall. This approach cuts down on time and can be statistically analysed for the presence of significant independent variables on the dependent variable. Once cell wall regeneration has been achieved and micro-calli is formed, applications of the already developed, cultivar specific, regeneration protocols can be applied. Alternatively, screening for positive tissue

culture responses from a diverse genetic pool and then subsequently using the best performing genetics for transformational events can create a work-around. This will allow for the breeding of the transgene, or edited event, to a more desirable cultivar avoiding the time-consuming approach of optimising protocols for a desirable cultivar at the beginning. The use of protoplasts for the characteristics of indels caused by sgRNA constructs has recently been reported (Zhang et al., 2021) demonstrating the possibility of using CRISPR within protoplasts that can be used within positively responding cultivars to regeneration.

The use of cell cultures to produce secondary metabolites is widely used in other species (Varma, 2010). Such elicitation experiments have previously been explored within cannabis with little success (Gorelick & Bernstein, 2017), though the premise of using large volumes of cells to biosynthesise cannabinoids, or other desirable secondary metabolites, is promising. The ability to produce millions of viable protoplasts from a single gram of leaf material as described in this thesis, coupled with the ability of protoplasts ability to infinitely divide given the optimal environmental conditions, can potentially lead to a larger accumulation of cannabinoids as an alternative to traditional horticulture approaches to cannabinoid accumulation within the floral tissue. Other important metabolites produced, such as terpenes and flavonoids can be mass produced from cannabis protoplasts in a similar approach. The use of cell cultures to produce pharmaceuticals are also possible within cannabis, which may be preferential to other plant species due to the large volume of biomass that can be produced with cannabis. This thesis could help researchers, with the identification of important variables in protoplast isolation, cut down on protocol development time with a level of confidence the protocols will be suitable as they are also tested on multiple cultivars within this thesis.

Protoplast fusion, where the fusion of two individual protoplasts occurs, is now seemingly possible with a robust protoplast isolation protocol being developed. This approach is an alternative to genome editing using nucleases to improve the genetic make-up within an individual. This approach can help develop novel chemotypes whilst avoiding any gene technology regulation through the absence of direct nucleotide modifications. Examples using this approach have been developed commercially in citrus (Grosser & Gmitter Jr, 1990) and breeding of new varieties such as in *Pleurotus* (Gwon et al., 2021) and in ornamental plants (Naing et al., 2021). The production of tetraploid cannabis has already been achieved (Parsons et al., 2019) with increases in CBD concentrations within flowers and fan leaf size. It is evidently possible to significantly increase cannabinoid content compared to diploid genetics, opening the doors to using tetraploidy through techniques, such as protoplast fusion, to modify any desired pathway, whether that be for increased terpene production or increased trichome densities. The development of an optimised protoplast isolation protocol developed within this thesis can help facilitate this area of research as an alternative to classical breeding and genome editing attempts.

Targeting genes other than cannabinoid biosynthesis using RNAi is now possible in cannabis, with the other important phenotypic characteristics of terpene families providing an excellent target for down or upregulation. There are claims of therapeutic potential that the terpene metabolites in cannabis contain (the entourage effect), with recent anti-seizure activity detected in brain slices of rodents (Walsh et al., 2021). With the expansion of patients, including the very young, people will undoubtedly prefer products that are more appealing, with terpenes such as Limonene and Myrcene providing a lemon or mango aroma and taste. With such a large potential global market and consumers looking beyond medicinal benefits of cannabis, the ability to down or upregulate terpenes in already developed cultivars has great economical potential. As further research into the medicinal potential terpenes continues, the ability to use the tools developed here to modifying concentrations within cannabis are invaluable.

Using genome editing targeting alternative aspects of productivity within cannabis, other important processes can be manipulated for benefit. Targeting endogenous miRNA for overexpression, such as miR397, can lead to significant increases of biomass (Patel et al., 2019) and could greatly increase yield thus increasing the output in medicinal cannabis crops. Targeting trichome development genes to increase density will greatly increase cannabinoid accumulation within female flowers thus increasing cannabinoid yield from every plant using in cultivation, greatly increasing profits requiring no extra input. Such an example of targeting trichome developmental genes in *Arabidopsis* using genome editing has previously been successful (Ryder et al., 2017). Using the RNAi approach described within this thesis will provide researchers helpful tools and protocols for targeting these terpene genes, further personalising cannabis for the consumer.

This thesis has designed comprehensive genetic resources and protocols specific for use in medicinal cannabis. There is a need for more robust approaches to genetically improve cannabis without the need for cultivar specific culture conditions. For example, the approach taken within this thesis in protoplast isolation and transformation are illustrations of protocol design to identify significant variables to design robust, inter-species protocols for genetic improvement. Having demonstrated the ability to target specific genes from large multigene families that have incredibly high sequence similarity using bioinformatical and tissue culture approaches, great groundwork has been achieved for further research. This thesis has also delivered transgenic approaches that can be further evaluated and biological resources that can be of significant benefit. Genetic transformation of cannabis prior to this thesis was rarely achieved, with the results targeting cannabinoid biosynthesis using the approach described here propelling cannabis forward within a species once deemed recalcitrant. Coupled with the regeneration of explant sources containing targeted, modified expression profiles, this thesis has opened the door for further research into this medical and economically crop for personalised medications. As such, the work presented in this thesis provides an excellent starting point in the improvement of the cannabis genome through the robust protocols and genetic resources developed.

APPENDICIES

Appendix 1

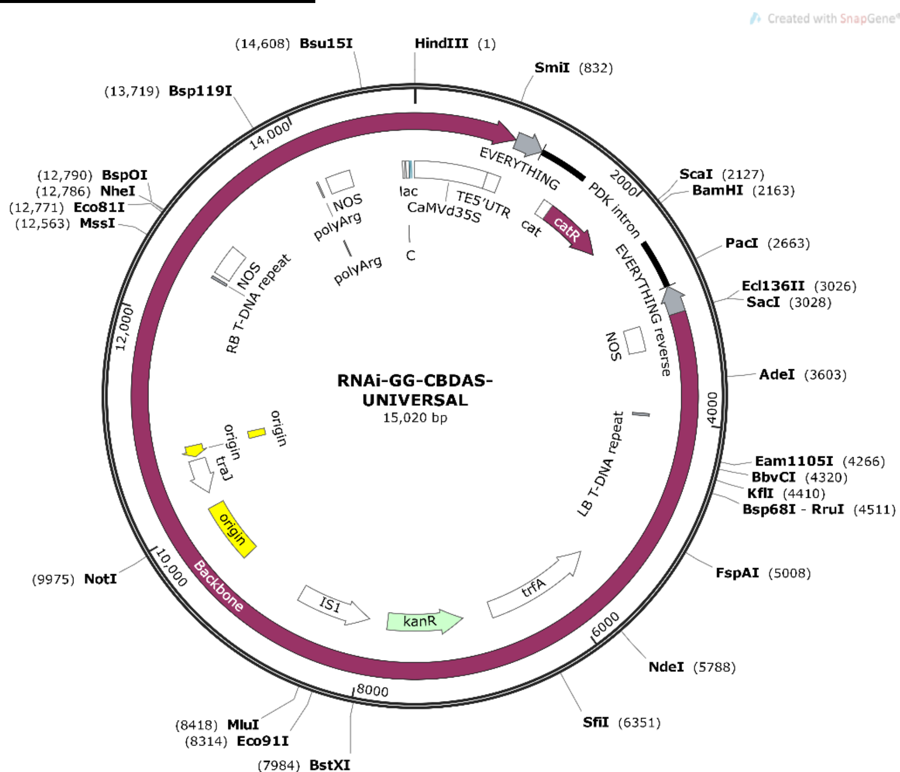
Supplementary material

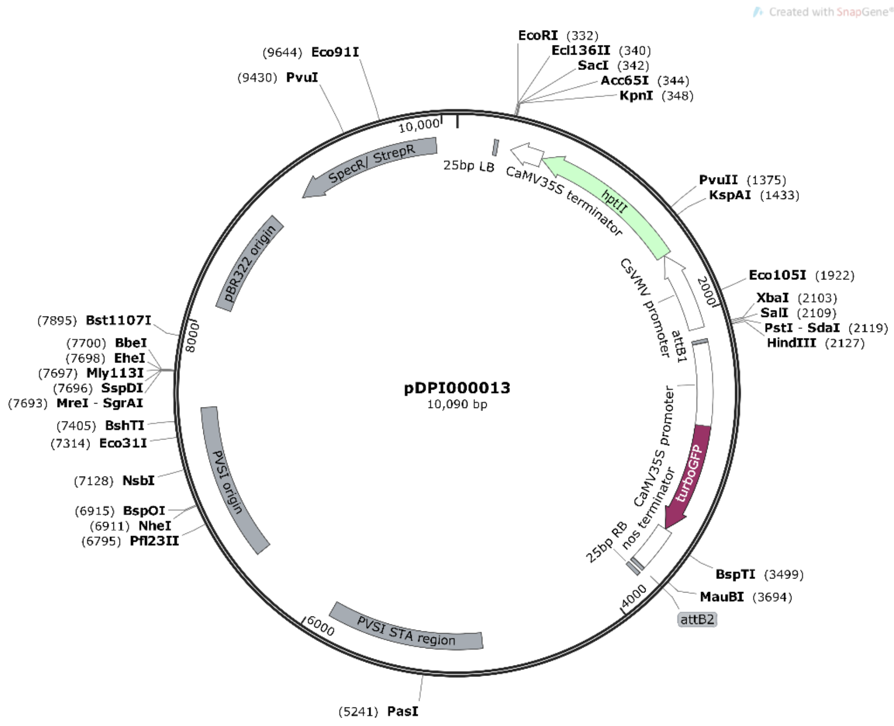
Lennon Matchett-Oates, Ehab Mohamaden, German Spangenberg, Noel Cogan. Transformation and regeneration of medicinal cannabis with an integrated RNAi vector for major cannabinoid modification. *Plants. In review.*

qPCR Primers used for cannabinoid biosynthesis gene amplification

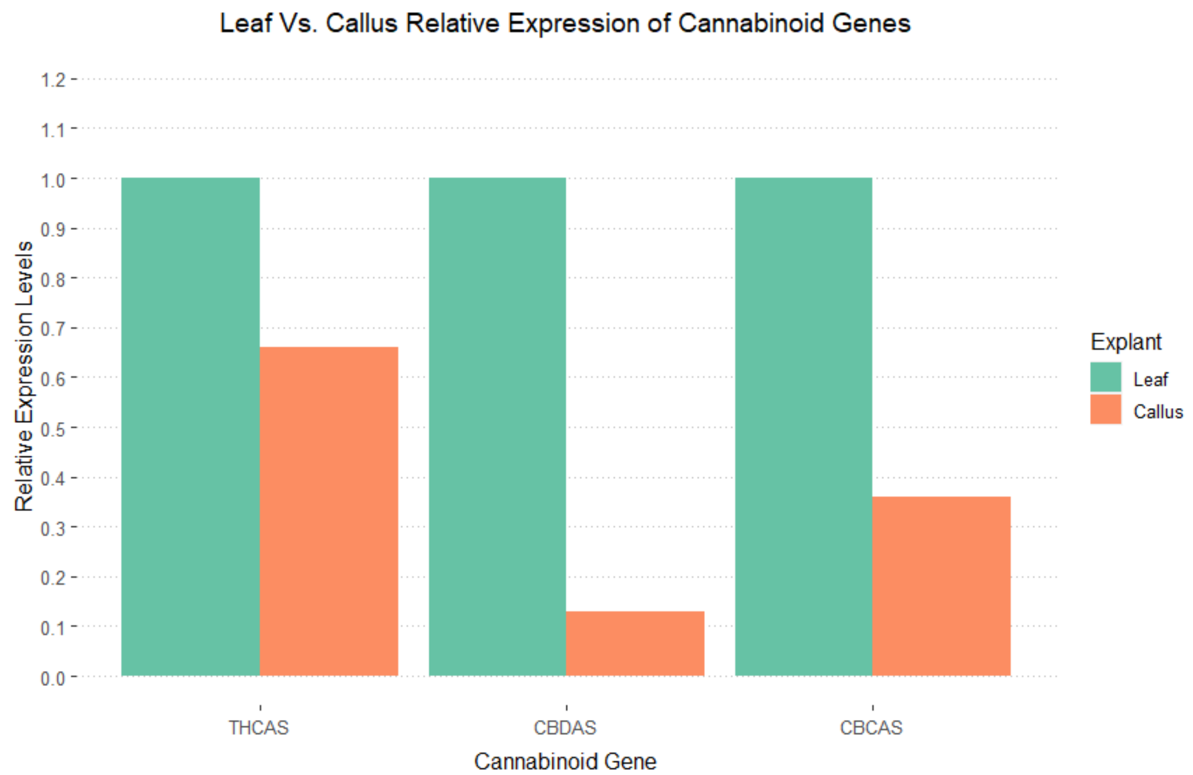
Target Gene	Forward Primer	Reverse Primer
THCAS	AATTTTTCATGGTGGAGTGGAT	AAAATTTACAACACCACTGTAGAA
CBDAS#1	CAGTTACTTCTCCTCCATTTTCC	ATAGTATCAATCCAGCTCAACT
CBDAS#2	GGAAAACTGAAGAGAAGTAAGTG	AATTGAGCTGGATTGATACTAT
CBCAS	ACCTGGGAGAAGCAAGAAG	GGATTAGTTTTTCCTAAATCAAGG

Vectors used within study





Supplementary Figure 1: Regenerated leaf primordia from multiple transformed cotyledon callus masses.



Supplementary Figure 2: Leaf explant and callus explant relative expression levels of cannabinoid biosynthesis genes

Supplementary Table 1: Standard error and p-values of transformed cotyledon callus masses from 2−ΔΔCt method using pooled control for normalisation. First numerical value represents relative expression followed by *p* value from technical triplicates.

Sample	*THCAS*	*CBDAS*	*CBCAS*
Control Pooled	0.31; NA	0.26; NA	0.34; NA
Callus#1	0.1; 0.01	0.005; 0.0005	0.02; 0.0001
Callus#2	0.07; 0.01	0.01; 0.003	0.01; 0.0007
Callus#3	0.02; 0.0002	0.001; 0.00003	0.01; 0.0008
Callus#4	0.08; 0.01	0.01; 0.001	0.02; 0.0002
Callus#5	0.2; 0.02	0.005; 0.003	0.03; 0.0002
Callus#6	0.2; 0.15	0.01; 0.001	0.03; 0.006
Callus#7	1.58; 0.001	0.06; 0.8	0.16; 0.6
Callus#8	0.19; 0.64	0.02; 0.002	0.03; 0.003
Callus#9	0.1; 0.01	0.01; 0.0003	0.01; 0.0002
Callus#10	0.18; 0.01	0.04; 0.006	0.08; 0.02
Callus#11	0.16; 0.89	0.04; 0.01	0.02; 0.001
Callus#12	0.63; 0.005	0.05; 0.02	0.06; 0.009
Callus#13	0.73; 0.25	0.04; 0.004	0.07; 0.004
Callus#14	0.81; 0.001	0.01; 0.002	0.3; 0.0004
Callus#15	0.35; 0.32	0.09; 0.04	0.03; 0.0004
Callus#16	0.45; 0.03	0.05; 0.008	0.13; 0.06
Callus#17	0.01; 0.00003	0.01; 0.001	0.01; 0.004
Callus#18	1.14; 0.2	0.16; 0.06	0.06; 0.2
Callus#19	1.22; 0.00005	0.05; 0.01	1.17; 0.0001
Callus#20	0.36; 0.001	0.04; 0.01	0.12; 0.96
Callus#21	0.73; 0.2	0.17; 0.2	0.13; 0.09
Callus#22	0.72; 0.0002	0.13; 0.07	0.32; 0.71
Callus#23	0.61; 0.8	0.15; 0.04	0.04; 0.001
Callus#24	1.56; 0.0002	0.18; 0.8	0.49; 0.02

Supplementary Table 2: Relative expression levels of cannabinoid genes from transformed regenerated hypocotyl compared to 8 control regenerated hypocotyls

Control regenerated hypocotyl	Relative expression of transformed hypocotyl		
	THCAS	*CBDAS*	*CBCAS*
#1	**1.40**	**1.63**	**11.20**
#2	**1.05**	**1.13**	**8.74**
#3	**1.05**	**5.39**	**8.97**
#4	**0.95**	**11.18**	**14.93**
#5	**2.72**	**0.70**	**19.49**
#6	**2.01**	**1.23**	**11.78**
#7	**3.60**	**4.79**	**19.70**
#8	**5.04**	**10.95**	**39.59**
Average	**1.68**	**2.90**	**14.57**

REFERENCES

Abdullah, R., Cocking, E. C., & Thompson, J. A. (1986). Efficient plant regeneration from rice protoplasts through somatic embryogenesis. Bio/Technology, 4(12), 1087–1090.

Adams, R., Hunt, M., & Clark, J. H. (1940). Structure of cannabidiol, a product isolated from the marihuana extract of Minnesota wild hemp. I. Journal of the American Chemical Society, 62(1), 196–200.

Ahmadabadi, M., Ruf, S., & Bock, R. (2007). A leaf-based regeneration and transformation system for maize (Zea mays L.). Transgenic Research, 16(4), 437–448.

Anderson, E. M., Haupt, A., Schiel, J. A., Chou, E., Machado, H. B., Strezoska, Ž., … Vermeulen, A. (2015). Systematic analysis of CRISPR–Cas9 mismatch tolerance reveals low levels of off-target activity. Journal of Biotechnology, 211, 56–65.

Anderson, L. L., Heblinski, M., Absalom, N. L., Hawkins, N. A., Bowen, M., Benson, M. J., … Chebib, M. (2021). Cannabigerolic acid, a major biosynthetic precursor molecule in cannabis, exhibits divergent effects on seizures in mouse models of epilepsy. British Journal of Pharmacology.

Andre, C. M., Hausman, J.-F., & Guerriero, G. (2016). Cannabis sativa: the plant of the thousand and one molecules. Frontiers in Plant Science, 7, 19.

Appendino, G., Gibbons, S., Giana, A., Pagani, A., Grassi, G., Stavri, M., & Rahman, M. M. (2008). Antibacterial cannabinoids from Cannabis sativa: a structure−activity study. Journal of Natural Products, 71(8), 1427–1430.

Araki, N., Kusumi, K., Masamoto, K., Niwa, Y., & Iba, K. (2000). Temperature-sensitive Arabidopsis mutant defective in 1-deoxy-d-xylulose 5-phosphate synthase within the plastid non-mevalonate pathway of isoprenoid biosynthesis. Physiologia Plantarum, 108(1), 19–24.

Arencibia, A., Molina, P. R., de la Riva, G., & Selman-Housein, G. (1995). Production of transgenic sugarcane (Saccharum officinarum L.) plants by intact cell electroporation. Plant Cell Reports, 14(5), 305–309.

Attree, S. M., & Fowke, L. C. (1993). Embryogeny of gymnosperms: advances in synthetic seed technology of conifers. Plant Cell, Tissue and Organ Culture, 35(1), 1–35.

Banerjee, A., Preiser, A. L., & Sharkey, T. D. (2016). Engineering of recombinant poplar deoxy-D-xylulose-5-phosphate synthase (PtDXS) by site-directed mutagenesis improves its activity. PloS One, 11(8), e0161534.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell, 136(2), 215–233.

Baulcombe, D. (2004). RNA silencing in plants. Nature, 431(7006), 356–363.

Beard, K. M., Boling, A. W. H., & Bargmann, B. O. R. (2021). Protoplast isolation, transient transformation, and flow-cytometric analysis of reporter-gene activation in Cannabis sativa L. Industrial Crops and Products, 164, 113360.

Beerli, R. R., & Barbas III, C. F. (2002). Engineering polydactyl zinc-finger transcription factors. Nature Biotechnology, 20(2), 135.

Bhowmik, P., Ellison, E., Polley, B., Bollina, V., Kulkarni, M., Ghanbarnia, K., & Kagale, S. (2018). Targeted mutagenesis in wheat microspores using CRISPR/Cas9. Scientific Reports, 8(1), 1–10.

Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science, 300(5620), 764.

References

Bing, X., Ning, L., Jinfeng, T., & Nan, G. (2007). Rapid tissue culture method of Cannabis sativa for industrial uses. CN, 1887043, 9.

Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., & Karpilow, J. (2006). 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nature Methods, 3(3), 199–204.

Bogdanove, A. J., & Voytas, D. F. (2011). TAL effectors: customizable proteins for DNA targeting. Science, 333(6051), 1843–1846.

Bonawitz, N. D., Ainley, W. M., Itaya, A., Chennareddy, S. R., Cicak, T., Effinger, K., & Samuel, J. P. (2019). Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. Plant Biotechnology Journal, 17(4), 750–761.

Bossche, R. Vanden, Demedts, B., Vanderhaeghen, R., & Goossens, A. (2013). Transient expression assays in tobacco protoplasts. In Jasmonate Signaling (pp. 227–239). Springer.

Braich, S., Baillie, R. C., Spangenberg, G. C., & Cogan, N. O. I. (2020). A New and Improved Genome Sequence of Cannabis sativa. Gigabyte, 1.

Brown, I., Cascio, M. G., Wahle, K. W. J., Smoum, R., Mechoulam, R., Ross, R. A., & Heys, S. D. (2010). Cannabinoid receptor-dependent and-independent anti-proliferative effects of omega-3 ethanolamides in androgen receptor-positive and-negative prostate cancer cell lines. Carcinogenesis, 31(9), 1584–1591.

Camacho, A., Van Deynze, A., Chi-Ham, C., & Bennett, A. B. (2014). Genetically engineered crops that fly under the US regulatory radar. Nature Biotechnology, 32(11), 1087–1091.

Cao, H. X., Wang, W., Le, H. T. T., & Vu, G. T. H. (2016). The power of CRISPR-Cas9-induced genome editing to speed up plant breeding. International Journal of Genomics, 2016.

Cao, J., Yao, D., Lin, F., & Jiang, M. (2014). PEG-mediated transient gene expression and silencing system in maize mesophyll protoplasts: a valuable tool for signal transduction study in maize. Acta Physiologiae Plantarum, 36(5), 1271–1281.

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. Cell, 136(4), 642–655.

Cascio, M. G., Pertwee, R. G., & Marini, P. (2017). The Pharmacology and Therapeutic Potential of Plant Cannabinoids. In Cannabis sativa L.-Botany and Biotechnology (pp. 207–225). Springer.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., & Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Research, 39(12), e82–e82.

Chandra, S., Lata, H., Khan, I. A., & ElSohly, M. A. (2013). The role of biotechnology in Cannabis sativa propagation for the production of phytocannabinoids. In Biotechnology for Medicinal Plants (pp. 123–148). Springer.

Chaohua, C., Gonggu, Z., Lining, Z., Chunsheng, G., Qing, T., Jianhua, C., … Jianguang, S. (2016). A rapid shoot regeneration protocol from the cotyledons of hemp (Cannabis sativa L.). Industrial Crops and Products, 83, 61–65.

Chen, K., & Gao, C. (2013). TALENs: customizable molecular DNA scissors for genome engineering of plants. Journal of Genetics and Genomics, 40(6), 271–279.

Chen, L., Li, W., Katin-Grazzini, L., Ding, J., Gu, X., Li, Y., & Deng, Z. (2018). A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants. Horticulture Research, 5(1), 1–12.

References

Christou, P., Ford, T. L., & Kofron, M. (1991). Genotype-independent stable transformation of rice (Oryza sativa) plants. Bio/Technology, 9, 957–962.

Clarke, R. C. (1981). Marijuana botany: An advanced study: The propagation and breeding of distinctive cannabis. Ronin publishing.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., & Marraffini, L. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science, 1231143.

Croteau, R. (1987). Biosynthesis and catabolism of monoterpenoids. Chemical Reviews, 87(5), 929–954.

Cui, Y., Xu, J., Cheng, M., Liao, X., & Peng, S. (2018). Review of CRISPR/Cas9 sgRNA Design Tools. Interdisciplinary Sciences: Computational Life Sciences, 1–11.

Cure, J. D. (1985). Carbon dioxide doubling responses: a crop survey. Direct Effects of Increasing Carbon Dioxide on Vegetation, 99–116.

Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., & Coffman, A. P. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc finger nucleases. Plant Physiology, pp-111.

De Backer, B., Debrus, B., Lebrun, P., Theunis, L., Dubois, N., Decock, L., & Charlier, C. (2009). Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. Journal of Chromatography B, 877(32), 4115–4124.

de Meijer, E. P. M., Bagatta, M., Carboni, A., Crucitti, P., Moliterni, V. M. C., Ranalli, P., & Mandolino, G. (2003). The inheritance of chemical phenotype in Cannabis sativa L. Genetics, 163(1), 335–346.

de Médicis Sajous, L. T. (1918). Sajous's Analytic Cyclopedia of Practical Medicine (Vol. 9). FA Davis Company.

Deguchi, M., Bogush, D., Weeden, H., Spuhler, Z., Potlakayala, S., Kondo, T., … Rudrabhatla, S. (2020). Establishment and optimization of a hemp (Cannabis sativa L.) agroinfiltration system for gene expression and silencing studies. Scientific Reports, 10(1), 1–11.

Devane, W. A., Hanus, L., Breuer, A., Pertwee, R. G., Stevenson, L. A., Griffin, G., & Mechoulam, R. (1992). Isolation and structure of a brain constituent that binds to the cannabinoid receptor. Science, 258(5090), 1946–1949.

Devinsky, O., Marsh, E., Friedman, D., Thiele, E., Laux, L., Sullivan, J., & Filloux, F. (2016). Cannabidiol in patients with treatment-resistant epilepsy: an open-label interventional trial. The Lancet Neurology, 15(3), 270–278.

Devinsky, O., Patel, A. D., Thiele, E. A., Wong, M. H., Appleton, R., Harden, C. L., … Group, G. P. A. S. (2018). Randomized, dose-ranging safety trial of cannabidiol in Dravet syndrome. Neurology, 90(14), e1204–e1211.

Di Natale, R. Tax and Regulate Cannabis (2018).Accessed Jan 2021 from : https://www.aph.gov.au/-/media/05_About_Parliament/54_Parliamentary_Depts/548_Parliamentary_Budget_Office/Costings/Publicly_released_costings/Tax_and_regulate_cannabis_DOC.docx?la=en&hash=0F230C451F3DCC0D5DCD545DB8275D43CB4CA9EB

Dijkman, W. P., de Gonzalo, G., Mattevi, A., & Fraaije, M. W. (2013). Flavoprotein oxidases: classification and applications. Applied Microbiology and Biotechnology, 97(12), 5177–5188.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., & Orchard, R. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology, 34(2), 184.

Doorenbos, N. J., Fetterman, P. S., Quimby, M. W., & Turner, C. E. (1971). Cultivation, extraction, and analysis of Cannabis sativa L. Annals of the New York Academy of Sciences, 191(1), 3–14.

ElSohly, M. A., & Slade, D. (2005). Chemical constituents of marijuana: the complex mixture of natural cannabinoids. Life Sciences, 78(5), 539–548.

ElSohly, M. A., Radwan, M. M., Gul, W., Chandra, S., & Galal, A. (2017). Phytochemistry of Cannabis sativa L. In Phytocannabinoids (pp. 1–36). Springer.

Emboden, W. A. (1974). Cannabis—a polytypic genus. Economic Botany, 28(3), 304.

Estévez, J. M., Cantero, A., Reindl, A., Reichler, S., & León, P. (2001). 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. Journal of Biological Chemistry, 276(25), 22901–22909.

Fairbairn, J. W., & Liebmann, J. A. (1974). The cannabinoid content of Cannabis sativa L grown in England. Journal of Pharmacy and Pharmacology, 26(6), 413–419.

Feeney, M., & Punja, Z. K. (2003). Tissue culture and Agrobacterium-mediated transformation of hemp (Cannabis sativa L.). In Vitro Cellular and Developmental Biology-Plant, 39(6), 578–585.

Fellermeier, M., & Zenk, M. H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Letters, 427(2), 283–285.

Fellermeier, M., Eisenreich, W., Bacher, A., & Zenk, M. H. (2001). Biosynthesis of cannabinoids: Incorporation experiments with 13C-labeled glucoses. European Journal of Biochemistry, 268(6), 1596–1604.

Fetterman, P. S., Keith, E. S., Waller, C. W., Guerrero, O., Doorenbos, N. J., & Quimby, M. W. (1971). Mississippi-grown cannabis sativa L.: Preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part. Journal of Pharmaceutical Sciences, 60(8), 1246–1249.

Flores-Sanchez, I. J., Peč, J., Fei, J., Choi, Y. H., Dušek, J., & Verpoorte, R. (2009). Elicitation studies in cell suspension cultures of Cannabis sativa L. Journal of Biotechnology, 143(2), 157–168.

Fournier, G. (1981). Les chimiotypes du chanvre (Cannabis sativa L.) Intérêt pour un programme de sélection. Agronomie, 1(8), 679–688.

Fulvio, F., Paris, R., Montanari, M., Citti, C., Cilento, V., Bassolino, L., … Cannazza, G. (2021). Analysis of Sequence Variability and Transcriptional Profile of Cannabinoid synthase Genes in Cannabis sativa L. Chemotypes with a Focus on Cannabichromenic acid synthase. Plants, 10(9), 1857.

Gagne, S. J., Stout, J. M., Liu, E., Boubakir, Z., Clark, S. M., & Page, J. E. (2012). Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. Proceedings of the National Academy of Sciences, 109(31), 12811–12816.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology, 31(7), 397–405.

Galal, A. M., Slade, D., Gul, W., El-Alfy, A. T., Ferreira, D., & Elsohly, M. A. (2009). Naturally occurring and related synthetic cannabinoids and their potential therapeutic applications. Recent Patents on CNS Drug Discovery, 4(2), 112–136.

Galán-Ávila, A., García-Fortea, E., Prohens, J., & Herraiz, F. J. (2020). Development of a direct in vitro plant regeneration protocol from Cannabis sativa L. seedling explants: developmental morphology of shoot regeneration and ploidy level of regenerated plants. Frontiers in Plant Science, 11, 645.

Galán-Ávila, A., Gramazio, P., Ron, M., Prohens, J., & Herraiz, F. J. (2021). A novel and rapid method for Agrobacterium-mediated production of stably transformed Cannabis sativa L. plants. Industrial Crops and Products, 170, 113691.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., & Carrington, J. C. (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. The Plant Cell, 22(2), 481–496.

Gasciolli, V., Mallory, A. C., Bartel, D. P., & Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Current Biology, 15(16), 1494–1500.

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiology and Molecular Biology Reviews, 67(1), 16–37.

Gil-Humanes, J., Pistón, F., Hernando, A., Alvarez, J. B., Shewry, P. R., & Barro, F. (2008). Silencing of γ-gliadins by RNA interference (RNAi) in bread wheat. Journal of Cereal Science, 48(3), 565–568.

Gorbunova, V., & Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Research, 25(22), 4650–4657.

Gorelick, J., & Bernstein, N. (2017). Chemical and physical elicitation for enhanced cannabinoid production in cannabis. In Cannabis sativa L.-botany and biotechnology (pp. 439–456). Springer.

Gorman, M. (1969). Sir William Brooke O'Shaughnessy: Pioneer chemist in a colonial environment. Journal of Chemical Education, 46(2), 99.

Grassa, C. J., Weiblen, G. D., Wenger, J. P., Dabney, C., Poplawski, S. G., Timothy Motley, S., … Schwartz, C. J. (2021). A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. New Phytologist.

Grassi, G., & McPartland, J. M. (2017). Chemical and Morphological Phenotypes in Breeding of Cannabis sativa L. In Cannabis sativa L.-Botany and Biotechnology (pp. 137–160). Springer.

Grechkin, A. N., Brühlmann, F., Mukhtarova, L. S., Gogolev, Y. V, & Hamberg, M. (2006). Hydroperoxide lyases (CYP74C and CYP74B) catalyze the homolytic isomerization of fatty acid hydroperoxides into hemiacetals. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids, 1761(12), 1419–1428.

Grosser, J. W., & Gmitter Jr, F. G. (1990). Protoplast fusion and citrus improvement. Plant Breeding Reviews, 8, 339–374.

Grotenhermen, F., & Müller-Vahl, K. (2016). Medicinal Uses of Marijuana and Cannabinoids. Critical Reviews in Plant Sciences, 35(5–6), 378–405. https://doi.org/10.1080/07352689.2016.1265360

Gui, H., Tong, Q., Qu, W., Mao, C.-M., & Dai, S.-M. (2015). The endocannabinoid system and its therapeutic implications in rheumatoid arthritis. International Immunopharmacology, 26(1), 86–91.

Guo, Q., Liu, Q., A Smith, N., Liang, G., & Wang, M.-B. (2016). RNA silencing in plants: mechanisms, technologies and applications in horticultural crops. Current Genomics, 17(6), 476–489.

Gwon, H.-M., Lee, Y.-H., Kim, J.-H., Baek, I.-S., Kang, H.-W., & Choi, J.-I. (2021). Breeding of new variety Pleurotus pulmonarius using protoplast fusion technique. Journal of Mushroom, 19(3), 166–175.

Haberlach, G. T., Cohen, B. A., Reichert, N. A., Baer, M. A., Towill, L. E., & Helgeson, J. P. (1985). Isolation, culture and regeneration of protoplasts from potato and several related Solanum species. Plant Science, 39(1), 67–74.

Haleem, R., & Wright, R. (2020). A scoping review on clinical trials of pain reduction with cannabis administration in adults. Journal of Clinical Medicine Research, 12(6), 344.

Hanin, M., & Paszkowski, J. (2003). Plant genome modification by homologous recombination. Current Opinion in Plant Biology, 6(2), 157–162.

Hanuš, L., Abu-Lafi, S., Fride, E., Breuer, A., Vogel, Z., Shalev, D. E., & Mechoulam, R. (2001). 2-Arachidonyl glyceryl ether, an endogenous agonist of the cannabinoid CB1 receptor. Proceedings of the National Academy of Sciences, 98(7), 3662–3665.

He, W., Xu, W., Xu, L., Fu, K., Guo, W., Bock, R., & Zhang, J. (2020). Length-dependent accumulation of double-stranded RNAs in plastids affects RNA interference efficiency in the Colorado potato beetle. Journal of Experimental Botany, 71(9), 2670–2677.

Heigwer, F., Kerr, G., & Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nature Methods, 11(2), 122.

Helliwell, C. A., Wesley, S. V., Wielopolska, A. J., & Waterhouse, P. M. (2002). High-throughput vectors for efficient gene silencing in plants. Functional Plant Biology, 29(10), 1217–1225.

Helliwell, C., & Waterhouse, P. (2003). Constructs and methods for high-throughput gene silencing in plants. Methods, 30(4), 289–295.

Hemphill, J. K., Turner, J. C., & Mahlberg, P. G. (1978). Studies on growth and cannabinoid composition of callus derived from different strains of Cannabis sativa. Lloydia.

Hemphill, J. K., Turner, J. C., & Mahlberg, P. G. (1980). Cannabinoid content of individual plant organs from different geographical strains of Cannabis sativa L. Journal of Natural Products, 43(1), 112–122.

Higgins, C., Ag News, U., Gordon-Smith, H., Schreiber, K., Blume, A., Grosbard, A., & Krieb, K. (2016). State of Indoor Farming State of Indoor Farming the Association for Vertical Farming. Retrieved from www.agrilyst.com

Hillig, K. W. (2005). Genetic evidence for speciation in Cannabis (Cannabaceae). Genetic Resources and Crop Evolution, 52(2), 161–180.

Howlett, A. C. (2005). Cannabinoid receptor signaling. In Cannabinoids (pp. 53–79).

Hryhorowicz, S., Walczak, M., Zakerska-Banaszak, O., Słomski, R., & Skrzypczak-Zielińska, M. (2018). Pharmacogenetics of cannabinoids. European Journal of Drug Metabolism and Pharmacokinetics, 43(1), 1–12.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., & Shalem, O. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology, 31(9), 827.

Huang, S. M., Bisogno, T., Trevisani, M., Al-Hayani, A., De Petrocellis, L., Fezza, F., & Chu, C. J. (2002). An endogenous capsaicin-like substance with high potency at recombinant and native

vanilloid VR1 receptors. Proceedings of the National Academy of Sciences, 99(12), 8400–8405.

Hughes, B. (2017). Legal notice. Publications Office of the European Union. https://doi.org/10.2810/566650

Hurst, W. J., & Finley, J. W. (2018). Genetically Modified Crops. In Principles of Food Chemistry (pp. 511–526). Springer.

Hussain, S. A., Zhou, R., Jacobson, C., Weng, J., Cheng, E., Lay, J., … Sankar, R. (2015). Perceived efficacy of cannabidiol-enriched cannabis extracts for treatment of pediatric epilepsy: a potential role for infantile spasms and Lennox–Gastaut syndrome. Epilepsy & Behavior, 47, 138–141.

Höfle, L., Biedenkopf, D., Werner, B. T., Shrestha, A., Jelonek, L., & Koch, A. (2020). Study on the efficiency of dsRNAs with increasing length in RNA-based silencing of the Fusarium CYP51 genes. RNA Biology, 17(4), 463–473.

Idso, K. E., & Idso, S. B. (1994). Plant responses to atmospheric CO2 enrichment in the face of environmental constraints: a review of the past 10 years' research. Agricultural and Forest Meteorology, 69(3–4), 153–203.

Ishida, Y., Hiei, Y., & Komari, T. (2007). Agrobacterium-mediated transformation of maize. Nature Protocols, 2(7), 1614–1621.

Jansen, R., Embden, J. D. A. van, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Molecular Microbiology, 43(6), 1565–1575.

Jekkel, Z., Heszky, L. E., & Ali, A. H. (1989). Effect of different cryoprotectants and transfer temperatures on the survival rate of hemp (Cannabis sativa L.) cell suspension in deep freezing. Acta Biologica Hungarica, 40(1–2), 127–136.

Johansen, L. K., & Carrington, J. C. (2001). Silencing on the spot. Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system. Plant Physiology, 126(3), 930–938.

Johnson, B. (2019). Addiction Medicine E-Book: Science and Practice. Elsevier Health Sciences.

Johnson, B. A. (2010). Medication treatment of different types of alcoholism. American Journal of Psychiatry, 167(6), 630–639.

Jones, M. P. A., Yi, Z., Murch, S. J., & Saxena, P. K. (2007). Thidiazuron-induced regeneration of Echinacea purpurea L.: micropropagation in solid and liquid culture systems. Plant Cell Reports, 26(1), 13–19.

Jones, R. L. (1979). Cell Culture, Protoplast Isolation, and Cell Fusion of Cannabis Sativa L.: Evaluation of Chilling Preventative Chemicals and Quality Control of Bananas in the Tropics. University of Houston Central Campus.

Julsing, M. K., Rijpkema, M., Woerdenbag, H. J., Quax, W. J., & Kayser, O. (2007). Functional analysis of genes involved in the biosynthesis of isoprene in Bacillus subtilis. Applied Microbiology and Biotechnology, 75(6), 1377–1384.

Jung, J. H., Fouad, W. M., Vermerris, W., Gallo, M., & Altpeter, F. (2012). RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. Plant Biotechnology Journal, 10(9), 1067–1076.

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J. P., Tyagi, A. K., & Kapoor, S. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics, 9(1), 451.

Kempin, S. A., Liljegren, S. J., Block, L. M., Rounsley, S. D., Yanofsky, M. F., & Lam, E. (1997). Targeted disruption in Arabidopsis. Nature, 389(6653), 802.

Kim, S.-M., Kuzuyama, T., Kobayashi, A., Sando, T., Chang, Y.-J., & Kim, S.-U. (2008). 1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase (IDS) is encoded by multicopy genes in gymnosperms Ginkgo biloba and Pinus taeda. Planta, 227(2), 287–298.

Kim, V. N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nature Reviews Molecular Cell Biology, 6(5), 376–385.

Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences, 93(3), 1156–1160.

Kleiman, M., Livingston, A., Heizer, D., Sederberg, C., Oglesby, P., Fox, S., & Saloga, C. W. (2016). The Economic Impact of Marijuana Legalization in Colorado Marijuana Policy Group. Retrieved from http://www.mjpolicygroup.com/pubs/MPG Impact of Marijuana on Colorado-Final.pdf

Kriese, U., Schumann, E., Weber, W. E., Beyer, M., & Brühl, L. (2004). Oil content, tocopherol composition and fatty acid patterns of the seeds of 51 Cannabis sativa L. genotypes. Euphytica, 137(3), 339–351.

Kumagai, M. H., Donson, J., Della-Cioppa, G., Harvey, D., Hanley, K., & Grill, L. K. (1995). Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proceedings of the National Academy of Sciences, 92(5), 1679–1683.

Kurihara, Y., & Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proceedings of the National Academy of Sciences, 101(34), 12753–12758.

Kutchan, T. M., & Dittrich, H. (1995). Characterization and mechanism of the berberine bridge enzyme, a covalently flavinylated oxidase of benzophenanthridine alkaloid biosynthesis in plants. Journal of Biological Chemistry, 270(41), 24475–24481.

Lalge, A. B., Mendel, P., Vyhnanek, T., Trojan, V., Kalousek, P., & Havel, L. (2016). Effects of different morphoregulators on growth and development of Cannabis Sativa L. In Proceedings of the International Ph. D. Students Conference on MendelNet (pp. 726–730).

Lange, B. M., Rujan, T., Martin, W., & Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proceedings of the National Academy of Sciences, 97(24), 13172–13177.

Langenheim, J. H. (1994). Higher plant terpenoids: a phytocentric overview of their ecological roles. Journal of Chemical Ecology, 20(6), 1223–1280.

Lata, H., Chandra, S., Khan, I. A., & ElSohly, M. A. (2009a). Propagation through alginate encapsulation of axillary buds of Cannabis sativa L.—an important medicinal plant. Physiology and Molecular Biology of Plants, 15(1), 79–86.

Lata, H., Chandra, S., Khan, I. A., & ElSohly, M. A. (2010). High frequency plant regeneration from leaf derived callus of high Δ9-tetrahydrocannabinol yielding Cannabis sativa L. Planta Medica, 76(14), 1629–1633.

Lata, H., Chandra, S., Khan, I., & ElSohly, M. A. (2009b). Thidiazuron-induced high-frequency direct shoot organogenesis of Cannabis sativa L. In Vitro Cellular & Developmental Biology-Plant, 45(1), 12–19.

Lata, H., Chandra, S., Techen, N., Khan, I. A., & ElSohly, M. A. (2016). In vitro mass propagation of Cannabis sativa L.: A protocol refinement using novel aromatic cytokinin meta-topolin and the assessment of eco-physiological, biochemical and genetic fidelity of micropropagated plants. Journal of Applied Research on Medicinal and Aromatic Plants, 3(1), 18–26.

Lazzeri, P. A., Brettschneider, R., Lührs, R., & Lörz, H. (1991). Stable transformation of barley via PEG-induced direct DNA uptake into protoplasts. Theoretical and Applied Genetics, 81(4), 437–444.

Lee, S., & Poulter, C. D. (2006). Escherichia coli type I isopentenyl diphosphate isomerase: structural and catalytic roles for divalent metals. Journal of the American Chemical Society, 128(35), 11545–11550.

Leghissa, A., Hildenbrand, Z. L., & Schug, K. A. (2018). A review of methods for the chemical characterization of cannabis natural products. Journal of Separation Science, 41(1), 398–415.

Lehman, C. W., Trautman, J. K., & Carroll, D. (1994). Illegitimate recombination in Xenopus: characterization of end-joined junctions. Nucleic Acids Research, 22(3), 434–442.

Li, J., Liao, X., Zhou, S., Liu, S., Jiang, L., & Wang, G. (2018). Efficient protoplast isolation and transient gene expression system for Phalaenopsis hybrid cultivar 'Ruili Beauty.' In Vitro Cellular & Developmental Biology-Plant, 54(1), 87–93.

Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Current Biology, 15(16), 1501–1507.

Liang, Z., Chen, K., Zhang, Y., Liu, J., Yin, K., Qiu, J.-L., & Gao, C. (2018). Genome editing of bread wheat using biolistic delivery of CRISPR/Cas9 in vitro transcripts or ribonucleoproteins. Nature Protocols, 13(3), 413.

Lim, K., See, Y. M., & Lee, J. (2017). A systematic review of the effectiveness of medical cannabis for psychiatric, movement and neurodegenerative disorders. Clinical Psychopharmacology and Neuroscience, 15(4), 301–312. https://doi.org/10.9758/cpn.2017.15.4.301

Lisson, S. N., Mendham, N. J., & Carberry, P. S. (2000). Development of a hemp (Cannabis sativa L.) simulation model 2. The flowering response of two hemp cultivars to photoperiod. Australian Journal of Experimental Agriculture, 40(3), 413–417.

Liu, X., Lu, T., Dou, Y., Yu, B., & Zhang, C. (2014). Identification of RNA silencing components in soybean and sorghum. BMC Bioinformatics, 15(1), 4.

Loh, W.-T., Hartsel, S. C., & Robertson, L. W. (1983). Tissue culture of Cannabis sativa L. and in vitro biotransformation of phenolics. Zeitschrift Fuer Pflanzenphysiologie, 111(5), 395–400.

Lucht, J. M. (2015). Public acceptance of plant biotechnology and GM crops. Viruses, 7(8), 4254–4281.

Luckett, T., Phillips, J., Lintzeris, N., Allsop, D., Lee, J., Solowij, N., & McCaffrey, N. (2016). Clinical trials of medicinal cannabis for appetite-related symptoms from advanced cancer: a survey of preferences, attitudes and beliefs among patients willing to consider participation. Internal Medicine Journal, 46(11), 1269–1275.

Lörz, H., Baker, B., & Schell, J. (1985). Gene transfer to cereal cells mediated by protoplast transformation. Molecular and General Genetics MGG, 199(2), 178–182.

MacPherson, C. R., & Scherf, A. (2015). Flexible guide-RNA design for CRISPR applications using Protospacer Workbench. Nature Biotechnology, 33(8), 805.

Mahlberg, P. G., & Kim, E. S. (2004). Accumulation of cannabinoids in glandular trichomes of Cannabis (Cannabaceae). Journal of Industrial Hemp, 9(1), 15–36.

Mahlberg, P. G., & Kim, E.-S. (1992). Secretory vesicle formation in glandular trichomes of Cannabis sativa (Cannabaceae). American Journal of Botany, 166–173.

References

Majumdar, R., Rajasekaran, K., & Cary, J. W. (2017). RNA interference (RNAi) as a potential tool for control of mycotoxin contamination in crop plants: concepts and considerations. Frontiers in Plant Science, 8, 200.

Mandolino, G, Carboni, A., Forapani, S., Faeti, V., & Ranalli, P. (1999). Identification of DNA markers linked to the male sex in dioecious hemp (Cannabis sativa L.). Theoretical and Applied Genetics, 98(1), 86–92.

Mandolino, Giuseppe, & Ranalli, P. (1999). Advances in biotechnological approaches for hemp breeding and industry. Haworth Press, Binghamton, NY.

Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair, 7(10), 1765–1771.

Martín-Sánchez, E., Furukawa, T. A., Taylor, J., & Martin, J. L. R. (2009). Systematic review and meta-analysis of cannabis treatment for chronic pain. Pain Medicine, 10(8), 1353–1368.

Mary Lynn Mathre. (2012). Cannabis in medical practice: A legal, historical and pharmacological overview of the therapeutic use of marijuana. McFarland.

Masani, M. Y. A., Noll, G. A., Parveez, G. K. A., Sambanthamurthi, R., & Prüfer, D. (2014). Efficient transformation of oil palm protoplasts by PEG-mediated transfection and DNA microinjection. PloS One, 9(5), e96831.

Mashayekhi, M., Shakib, A. M., Ahmad-Raji, M., & Bezdi, K. G. (2008). Gene transformation potential of commercial canola (Brassica napus L.) cultivars using cotyledon and hypocotyl explants. African Journal of Biotechnology, 7(24).

Matchett-Oates, L., Mohamaden, E., Spangenberg, G., & Cogan, N. (2021). Development of a robust transient expression screening system in protoplasts of cannabis. In Vitro Cellular & Developmental Biology-Plant.

Matchett-Oates, L., Braich, S., Spangenberg, G., Rochfort, S., Cogan, N. (2021). In silico analysis enabling genome editing in Medicinal cannabis, gene families and variant characterisation. PLoS One.

Matchett-Oates, L., Cogan, N., Spangenberg, G. (2021). Manipulation of cannabinoid biosynthesis via transient RNAi expression. Frontiers in Plant Science.

Mathur, S., & Sutton, J. (2017). Personalized medicine could transform healthcare. Biomedical Reports, 7(1), 3–5.

McConn, M., James, D., & Miquel, M. (1993). Mutants of Arabidopsis deficient in the synthesis of alpha-linolenate. Biochemical and genetic characterization of the endoplasmic reticulum linoleoyl desaturase. Journal of Biological Chemistry, 268(22), 16345–16351.

McCoy, B., Wang, L., Zak, M., Al-Mehmadi, S., Kabir, N., Alhadid, K., … Whitney, R. (2018). A prospective open-label trial of a CBD/THC cannabis oil in dravet syndrome. Annals of Clinical and Translational Neurology, 5(9), 1077–1088.

McGeeney, B. E. (2013). Cannabinoids and hallucinogens for headache. Headache: The Journal of Head and Face Pain, 53(3), 447–458.

McKernan, K. J., Helbert, Y., Kane, L. T., Ebling, H., Zhang, L., Liu, B., … Baybayan, P. (2020). Sequence and annotation of 42 cannabis genomes reveals extensive copy number variation in cannabinoid synthesis and pathogen resistance genes. BioRxiv.

McPartland, J. M. (2018). Cannabis systematics at the levels of family, genus, and species. Cannabis and Cannabinoid Research, 3(1), 203–212.

Mechoulam, R., & Carlini, E. A. (1978). Toward drugs derived from cannabis. Naturwissenschaften, 65(4), 174–179.

Mechoulam, R., Ben-Shabat, S., Hanus, L., Ligumsky, M., Kaminski, N. E., Schatz, A. R., & Compton, D. R. (1995). Identification of an endogenous 2-monoglyceride, present in canine gut, that binds to cannabinoid receptors. Biochemical Pharmacology, 50(1), 83–90.

Meijer, H. A., Smith, E. M., & Bushell, M. (2014). Regulation of miRNA strand selection: follow the leader? Portland Press Ltd.

Menzel, M. Y. (1964). Meiotic chromosomes of monoecious Kentucky hemp (Cannabis sativa). Bulletin of the Torrey Botanical Club, 193–205.

Mestinšek-Mubi, Š., Svetik, S., Flajšman, M., & Murovec, J. (2020). In vitro tissue culture and genetic analysis of two high-CBD medical cannabis (Cannabis sativa L.) breeding lines. Genetika, 52(3), 925–941.

Miallau, L., Alphey, M. S., Kemp, L. E., Leonard, G. A., McSweeney, S. M., Hecht, S., & Hunter, W. N. (2003). Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase. Proceedings of the National Academy of Sciences, 100(16), 9173–9178.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., & Kim, K. A. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology, 25(7), 778.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Research, 42(W1), W401–W407.

Monthony, A. S., Kyne, S. T., Grainger, C. M., & Jones, A. M. P. (2021). Recalcitrance of Cannabis sativa to de novo regeneration; a multi-genotype replication study. PloS One, 16(8), e0235525.

Monthony, A. S., Page, S. R., Hesami, M., & Jones, A. M. P. (2021). The past, present and future of Cannabis sativa tissue culture. Plants, 10(1), 185.

Morimoto, S., Tanaka, Y., Sasaki, K., Tanaka, H., Fukamizu, T., Shoyama, Y., … Taura, F. (2007). Identification and characterization of cannabinoids that induce cell death through mitochondrial permeability transition in Cannabis leaf cells. Journal of Biological Chemistry, 282(28), 20739–20751.

Movahedi, M., Ghasemi-Omran, V., & Torabi, S. (2015). The effect of different concentrations of TDZ and BA on in vitro regeneration of Iranian cannabis (Cannabis sativa) using cotyledon and epicotyl explants. Journal of Plant Molecular Breeding, 3(2), 20–27.

Munos, J. W., Pu, X., Mansoorabadi, S. O., Kim, H. J., & Liu, H. (2009). A secondary kinetic isotope effect study of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase-catalyzed reaction: evidence for a retroaldol-aldol rearrangement. Journal of the American Chemical Society, 131(6), 2048–2049.

Nadal, A., De Giacomo, M., Einspanier, R., Kleter, G., Kok, E., McFarland, S., … van Dijk, J. (2018). Exposure of livestock to GM feeds: detectability and measurement. Food and Chemical Toxicology, 117, 13–35.

Naim-Feil, E., Pembleton, L. W., Spooner, L. E., Malthouse, A. L., Miner, A., Quinn, M., … Cogan, N. O. I. (2021). The characterization of key physiological traits of medicinal cannabis (Cannabis sativa L.) as a tool for precision breeding. BMC Plant Biology, 21(1), 1–15.

Naing, A. H., Adedeji, O. S., & Kim, C. K. (2021). Protoplast technology in ornamental plants: Current progress and potential applications on genetic improvement. Scientia Horticulturae, 283, 110043.

Naito, Y., & Ui-Tei, K. (2012). siRNA design software for a target gene-specific RNA interference. Frontiers in Genetics, 3, 102.

Neumann, E., Schaefer-Ridder, M., Wang, Y., & Hofschneider, P. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. The EMBO Journal, 1(7), 841–845.

Nicolas, A. L., Munz, P. L., & Young, C. S. H. (1995). A modified single-strand annealing model best explains the joining of DNA double-strand breaks in mammalian cells and cell extracts. Nucleic Acids Research, 23(6), 1036–1043.

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., & Monéger, F. (2004). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biology, 3(1), e4.

Noordermeer, M. A., Veldink, G. A., & Vliegenthart, J. F. G. (2001). Fatty acid hydroperoxide lyase: a plant cytochrome P450 enzyme involved in wound healing and pest resistance. Chembiochem, 2(7-8), 494–504.

Novotna, A., Mares, J., Ratcliffe, S., Novakova, I., Vachova, M., Zapletalova, O., & Comi, G. (2011). A randomized, double-blind, placebo-controlled, parallel-group, enriched-design study of nabiximols*(Sativex®), as add-on therapy, in subjects with refractory spasticity caused by multiple sclerosis. European Journal of Neurology, 18(9), 1122–1131.

Office of Drug Control. (2021). Summary of licences granted. Retrieved June 27, 2021, from https://www.odc.gov.au/summary-licences-granted

Ogunola, O. F., Hawkins, L. K., Mylroie, E., Kolomiets, M. V, Borrego, E., Tang, J. D., & Warburton, M. L. (2017). Characterization of the maize lipoxygenase gene family in relation to aflatoxin accumulation resistance. PloS One, 12(7), e0181265.

Ohlrogge, J., & Browse, J. (1995). Lipid biosynthesis. The Plant Cell, 7(7), 957.

Olynk Widmar, N. J., Dominick, S. R., Tyner, W. E., & Ruple, A. (2017). When is genetic modification socially acceptable? When used to advance human health through avenues other than food. PloS One, 12(6), e0178227.

Pacher, P., Bátkai, S., & Kunos, G. (2006). The endocannabinoid system as an emerging target of pharmacotherapy. Pharmacological Reviews, 58(3), 389–462.

Pacifico, D., Miselli, F., Carboni, A., Moschella, A., & Mandolino, G. (2008). Time course of cannabinoid accumulation and chemotype development during the growth of Cannabis sativa L. Euphytica, 160(2), 231–240.

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., & Mandolino, G. (2006). Genetics and Marker-assisted Selection of the Chemotype in Cannabis sativa L. Molecular Breeding, 17(3), 257–268.

Page, S. R. G., Monthony, A. S., & Jones, A. M. P. (2020). Basal media optimization for the micropropagation and callogenesis of Cannabis sativa L. BioRxiv.

Parker, J. S., Roe, S. M., & Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. The EMBO Journal, 23(24), 4727–4737.

Parsons, J. L., Martin, S. L., James, T., Golenia, G., Boudko, E. A., & Hepworth, S. R. (2019). Polyploidization for the genetic improvement of Cannabis sativa. Frontiers in Plant Science, 10, 476.

Parveen, S., & Shahzad, A. (2010). TDZ-induced high frequency shoot regeneration in Cassia sophera Linn. via cotyledonary node explants. Physiology and Molecular Biology of Plants, 16(2), 201–206.

Pastwa, E., & Błasiak, J. (2003). Non-homologous DNA end joining. Acta Biochimica Polonica, 50(4), 891–908.

Patel, P., Yadav, K., Srivastava, A. K., Suprasanna, P., & Ganapathi, T. R. (2019). Overexpression of native Musa-miR397 enhances plant biomass without compromising abiotic stress tolerance in banana. Scientific Reports, 9(1), 1–15.

Peet, R., Sun, M., Winnicki, R. & Donsky, M. (2016). US9394510B2.

Pei, Z., Liu, J., Liu, M., Zhou, W., Yan, P., Wen, S., & Chen, Y. (2018). Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine. Interdisciplinary Sciences: Computational Life Sciences, 10(1), 126–130.

Pellati, F., Borgonetti, V., Brighenti, V., Biagi, M., Benvenuti, S., & Corsi, L. (2018). Cannabis sativa L. and nonpsychoactive cannabinoids: their chemistry and role against oxidative stress, inflammation, and cancer. BioMed Research International, 2018.

Pennycooke, J. C., & Towill, L. E. (2000). Cryopreservation of shoot tips from in vitro plants of sweet potato [Ipomoea batatas (L.) Lam.] by vitrification. Plant Cell Reports, 19(7), 733–737.

Pertwee, R G. (2005). Pharmacological actions of cannabinoids. In Cannabinoids (pp. 1–51).

Pertwee, R G. (2006). Cannabinoid pharmacology: the first 66 years. British Journal of Pharmacology, 147(S1).

Pertwee, R G. (2007). GPR55: a new member of the cannabinoid receptor clan? British Journal of Pharmacology, 152(7), 984–986.

Petsas, A. S., & Vagi, M. C. (2019). Trends in the bioremediation of pharmaceuticals and other organic contaminants using native or genetically modified microbial strains: a review. Current Pharmaceutical Biotechnology, 20(10), 787–824.

Phillips, M. A., León, P., Boronat, A., & Rodríguez-Concepción, M. (2008). The plastidial MEP pathway: unified nomenclature and resources. Trends in Plant Science, 13(12), 619–623.

Piomelli, D., & Russo, E. B. (2016). The Cannabis sativa versus Cannabis indica debate: an interview with Ethan Russo, MD. Cannabis and Cannabinoid Research, 1(1), 44–46.

Pipiras, E., Coquelle, A., Bieth, A., & Debatisse, M. (1998). Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. The EMBO Journal, 17(1), 325–333.

Plawuszewski, M., Lassocinski, W., & Wielgus, K. (2005). Regeneration of Polish cultivars of monoecious hemp [Cannabis sativa L.] grown in in vitro cultures. Biological Letters, 42(2 Spec. Vol.).

Porter, A. C., Sauer, J.-M., Knierman, M. D., Becker, G. W., Berna, M. J., Bao, J., & Leese, A. B. (2002). Characterization of a novel endocannabinoid, virodhamine, with antagonist activity at the CB1 receptor. Journal of Pharmacology and Experimental Therapeutics, 301(3), 1020–1024.

Portnoy, V., Huang, V., Place, R. F., & Li, L. (2011). Small RNA and transcriptional upregulation. Wiley Interdisciplinary Reviews: RNA, 2(5), 748–760.

Potter, D. J. (2014). Cannabis horticulture. Handbook of Cannabis.

ProCon. (2021). Legal Medical Marijuana States and DC.Accessed July 23. https://medicalmarijuana.procon.org/legal-medical-marijuana-states-and-dc/

Puchta, H. (2004). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. Journal of Experimental Botany, 56(409), 1–14.

Punja, Z. K. (2021). Epidemiology of Fusarium oxysporum causing root and crown rot of cannabis (Cannabis sativa L., marijuana) plants in commercial greenhouse production. Canadian Journal of Plant Pathology, 43(2), 216–235.

Qi, Y., Zhang, Y., Zhang, F., Baller, J. A., Cleland, S. C., Ryu, Y., & Voytas, D. F. (2013). Increasing frequencies of site-specific mutagenesis and gene targeting in Arabidopsis by manipulating DNA repair pathways. Genome Research, gr-145557.

Querol, J., Campos, N., Imperial, S., Boronat, A., & Rodríguez-Concepción, M. (2002). Functional analysis of the Arabidopsis thaliana GCPE protein involved in plastid isoprenoid biosynthesis. FEBS Letters, 514(2–3), 343–346.

Radwan, M. M., Wanas, A. S., Chandra, S., & ElSohly, M. A. (2017). Natural Cannabinoids of Cannabis and Methods of Analysis. In Cannabis sativa L.-Botany and Biotechnology (pp. 161–182). Springer.

Ram, H. Y. M., & Jaiswal, V. S. (1970). Induction of female flowers on male plants ofCannabis Sativa L. by 2-chloroethanephos-phonic acid. Cellular and Molecular Life Sciences, 26(2), 214–216.

Ramawat, K. G., & Mathur, M. (2007). Factors affecting the production of secondary metabolites. Biotechnology: Secondary Metabolites. Plants and Microbes, 59–102.

Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M. L., Yang, Y., & Mumtaz, A. S. (2021). Large-scale whole-genome resequencing unravels the domestication history of Cannabis sativa. Science Advances, 7(29), eabg2286.

Richez-Dumanois, C., Braut-Boucher, F., Cosson, L., & Paris, M. (1986). Multiplication végétative in vitro du chanvre (Cannabis sativa L.). Application à la conserva-tion des clones sélectionnés. Agronomie, 6(5), 487–495.

Robinson, M. B., & Scherlen, R. G. (2014). Lies, damned lies, and drug war statistics: a critical analysis of claims made by the office of National Drug Control Policy. SUNY Press.

Rodríguez-Concepción, M., Campos, N., María Lois, L., Maldonado, C., Hoeffler, J.-F., Grosdemange-Billiard, C., & Boronat, A. (2000). Genetic evidence of branching in the isoprenoid pathway for the production of isopentenyl diphosphate and dimethylallyl diphosphate in Escherichia coli. FEBS Letters, 473(3), 328–332.

Rohdich, F., Wungsintaweekul, J., Eisenreich, W., Richter, G., Schuhr, C. A., Hecht, S., & Bacher, A. (2000). Biosynthesis of terpenoids: 4-diphosphocytidyl-2C-methyl-D-erythritol synthase of Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 97(12), 6451–6456.

Rohdich, F., Zepeck, F., Adam, P., Hecht, S., Kaiser, J., Laupitz, R., & Bacher, A. (2003). The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. Proceedings of the National Academy of Sciences, 100(4), 1586–1591.

Ross, S. A., ElSohly, M. A., Sultana, G. N. N., Mehmedic, Z., Hossain, C. F., & Chandra, S. (2005). Flavonoid glycosides and cannabinoids from the pollen of Cannabis sativa L. Phytochemical Analysis: An International Journal of Plant Chemical and Biochemical Techniques, 16(1), 45–48.

Ryder, P., McHale, M., Fort, A., & Spillane, C. (2017). Generation of stable nulliplex autopolyploid lines of Arabidopsis thaliana using CRISPR/Cas9 genome editing. Plant Cell Reports, 36(6), 1005–1008.

Sage, R. F., & Sharkey, T. D. (1987). The effect of temperature on the occurrence of O2 and CO2 insensitive photosynthesis in field grown plants. Plant Physiology, 84(3), 658–664.

Saika, H., Nishizawa-Yokoi, A., & Toki, S. (2014). The non-homologous end-joining pathway is involved in stable transformation in rice. Frontiers in Plant Science, 5, 560.

Sakamoto, K., Abe, T., Matsuyama, T., Yoshida, S., Ohmido, N., Fukui, K., & Satoh, S. (2005). RAPD markers encoding retrotransposable elements are linked to the male sex in Cannabis sativa L. Genome, 48(5), 931–936.

Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H., & Satoh, S. (1998). Characterization; genome sizes and morphology of sex chromosomes in hemp (Cannabis sativa L.). Cytologia, 63(4), 459–464.

Sakamoto, K., Ohmido, N., Fukui, K., Kamada, H., & Satoh, S. (2000). Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant Cannabis sativa. Plant Molecular Biology, 44(6), 723–732.

Salomon, S., & Puchta, H. (1998). Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. The EMBO Journal, 17(20), 6086–6095.

Sanghera, G. S., Wani, S. H., Gill, M. S., Kashyap, P. L., & Gosal, S. S. (2010). RNA interference: Its concept and application in crop plants. Biotechnology Cracking New Pastures, MD Publishers, New Delhi, 33–78.

Sant'Ana, R. R. A., Caprestano, C. A., Nodari, R. O., & Agapito-Tenfen, S. Z. (2020). PEG-delivered CRISPR-Cas9 ribonucleoproteins system for gene-editing screening of maize protoplasts. Genes, 11(9), 1029.

Sawler, J., Stout, J. M., Gardner, K. M., Hudson, D., Vidmar, J., Butler, L., & Myles, S. (2015). The genetic structure of marijuana and hemp. PloS One, 10(8), e0133292.

Scacheri, P. C., Rozenblatt-Rosen, O., Caplen, N. J., Wolfsberg, T. G., Umayam, L., Lee, J. C., … Meyerson, M. (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. Proceedings of the National Academy of Sciences, 101(7), 1892–1897.

Schachtsiek, J., Hussain, T., Azzouhri, K., Kayser, O., & Stehle, F. (2019). Virus-induced gene silencing (VIGS) in Cannabis sativa L. Plant Methods, 15(1), 1–9.

Schmelz, S., & Naismith, J. H. (2009). Adenylate-forming enzymes. Current Opinion in Structural Biology, 19(6), 666–671.

Schweizer, P., Pokorny, J., Schulze-Lefert, P., & Dudler, R. (2000). Double-stranded RNA interferes with gene function at the single-cell level in cereals. The Plant Journal, 24(6), 895–903.

Seemann, M., Wegner, P., Schünemann, V., Bui, B. T. S., Wolff, M., Marquet, A., & Rohmer, M. (2005). Isoprenoid biosynthesis in chloroplasts via the methylerythritol phosphate pathway: the (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (GcpE) from Arabidopsis thaliana is a [4Fe–4S] protein. JBIC Journal of Biological Inorganic Chemistry, 10(2), 131–137.

Senthil-Kumar, M., & Mysore, K. S. (2010). RNAi in plants: recent developments and applications in agriculture. Gene Silencing: Theory, Techniques and Applications, 183–199.

Shafie, N. H., Saleem, M., Moses, E. J., Razak, S. R. A., & Yusoff, N. M. (2014). The CRISPR-Cas9 system: A new dawn in gene editing. Journal of Bioanalysis & Biomedicine, 6(6), 45.

Shan, S., Chen, X., Liu, T., Zhao, H., Rao, Z., & Lou, Z. (2011). Crystal structure of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE) from Mycobacterium tuberculosis. The FASEB Journal, 25(5), 1577–1584.

Shao, H., Song, S.-J., & Clarke, R. C. (2003). Female-associated DNA polymorphisms of hemp (Cannabis sativa L.). Journal of Industrial Hemp, 8(1), 5–9.

She, K. (2016). Cannabis Part 2: Cannabinoids and Endocannabinoids. Retrieved June 28, 2018, from http://segra-intl.com/segrapedia/cannabis-part-2-cannabinoids-and-endocannabinoids/

Sheen, J. (2001). Signal transduction in maize and Arabidopsis mesophyll protoplasts. Plant Physiology, 127(4), 1466–1475.

Shillito, R. (1999). Methods of genetic transformation: electroporation and polyethylene glycol treatment. In Molecular improvement of cereal crops (pp. 9–20).

Shimizu, Y., Bhakta, M. S., & Segal, D. J. (2009). Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. Bioorganic & Medicinal Chemistry Letters, 19(14), 3970–3972.

Shockey, J., & Browse, J. (2011). Genome-level and biochemical diversity of the acyl-activating enzyme superfamily in plants. The Plant Journal, 66(1), 143–160.

Shoyama, Yoshinari, Tamada, T., Kurihara, K., Takeuchi, A., Taura, F., Arai, S., & Kuroki, R. (2012). Structure and function ofΔ 1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of Cannabis sativa. Journal of Molecular Biology, 423(1), 96–105.

Shoyama, Yukihiro, Yagi, M., Nishioka, I., & Yamauchi, T. (1975). Biosynthesis of cannabinoid acids. Phytochemistry, 14(10), 2189–2192.

Sirikantaramas, S., & Taura, F. (2017). Cannabinoids: Biosynthesis and Biotechnological Applications. In Cannabis sativa L.-Botany and Biotechnology (pp. 183–206).

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., & Taura, F. (2004). The gene controlling marijuana psychoactivity molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic-acid synthase from Cannabis sativa. Journal of Biological Chemistry.

Slusarkiewicz-Jarzina, A., Ponitka, A., & Kaczmarek, Z. (2005). Influence of cultivar, explant source and plant growth regulator on callus induction and plant regeneration of Cannabis sativa L. Acta Biologica Cracoviensia Series Botanica, 47(2), 145–151.

Small, E, & Beckstead, H. D. (1973). Common cannabinoid phenotypes in 350 stocks of Cannabis. Lloydia.

Small, E, & Cronquist, A. (1976). A practical and natural taxonomy for Cannabis. Taxon, 405–435.

Smith, N. A., Singh, S. P., Wang, M.-B., Stoutjesdijk, P. A., Green, A. G., & Waterhouse, P. M. (2000). Total silencing by intron-spliced hairpin RNAs. Nature, 407(6802), 319–320.

Stockings, E., Zagic, D., Campbell, G., Weier, M., Hall, W. D., Nielsen, S., & Degenhardt, L. (2018). Evidence for cannabis and cannabinoids for epilepsy: a systematic review of controlled and observational evidence. Journal of Neurology, Neurosurgery & Psychiatry, 89(7), 741–753.

Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W., & Page, J. E. (2012). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. The Plant Journal, 71(3), 353–365.

Strain, B. R., & Cure, J. D. (1985). Direct effects of increasing carbon dioxide on vegetation. Duke Univ., Durham, NC (USA).

Taura, F., Morimoto, S., & Shoyama, Y. (1996). Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. Journal of Biological Chemistry, 271(29), 17411–17416.

Taura, F., Morimoto, S., Shoyama, Y., & Mechoulam, R. (1995). First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis. Journal of the American Chemical Society, 117(38), 9766–9767.

Taura, F., Sirikantaramas, S., Shoyama, Y., Shoyama, Y., & Morimoto, S. (2007). Phytocannabinoids in Cannabis sativa: recent studies on biosynthetic enzymes. Chemistry & Biodiversity, 4(8), 1649–1663.

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., & Morimoto, S. (2007). Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. FEBS Letters, 581(16), 2929–2934.

Terns, M. P., & Terns, R. M. (2011). CRISPR-based adaptive immune systems. Current Opinion in Microbiology, 14(3), 321–327.

Thomas, B. F., & ElSohly, M. (2015). The analytical chemistry of cannabis: Quality assessment, assurance, and regulation of medicinal marijuana and cannabinoid preparations.

Thomas, B. F., & ElSohly, M. A. (2016). The botany of Cannabis sativa L. The Analytical Chemistry of Cannabis, 1–26.

Thompson, L. H., & Schild, D. (2001). Homologous recombinational repair of DNA ensures mammalian chromosome stability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 477(1), 131–153.

Toriyama, K., Arimoto, Y., Uchimiya, H., & Hinata, K. (1988). Transgenic rice plants after direct gene transfer into protoplasts. Bio/Technology, 6(9), 1072–1074.

Turner, C. E., Elsohly, M. A., Cheng, P. C., & Lewis, G. (1979). Constituents of Cannabis sativa L., XIV: Intrinsic problems in classifying Cannabis based on a single cannabinoid analysis. Journal of Natural Products, 42(3), 317–319.

United Nations Office on Drugs and Crime. (2021). World Drug Report 2021: Drug Market Trends: Cannabis, Opioids. World Drug Report 2021. https://doi.org/10.18356/bdc264f4-en

Van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., & Page, J. E. (2011). The draft genome and transcriptome of Cannabis sativa. Genome Biology, 12(10), R102.

Vanamee, É. S., Santagata, S., & Aggarwal, A. K. (2001). FokI requires two specific DNA sites for cleavage1. Journal of Molecular Biology, 309(1), 69–78.

Varma, V. (2010). Advancements in the production of secondary metabolites. Journal of Natural Products, 3.

Vaucheret, H. (2008). Plant argonautes. Trends in Plant Science, 13(7), 350–358.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A. C., & Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Molecular Cell, 16(1), 69–79.

Veltcheva, M. R., & Lilova Svetleva, D. (2005). In vitro regeneration of Phaseolus vulgaris L. via organogenesis from petiole explants. Journal of Central European Agriculture, 6(1), 53–58.

von Esenbeck, T. F. L. N. (1831). Handbuch der medicinisch-pharmaceutischen Botanik (Vol. 2). Arnz & Comp.

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. Annual Review of Plant Biology, 64.

Wahby, I., Caba, J. M., & Ligero, F. (2013). Agrobacterium infection of hemp (Cannabis sativa L.): establishment of hairy root cultures. Journal of Plant Interactions, 8(4), 312–320.

Walsh, F. A., Sheridan, H., & Cunningham, M. (2021). Cannabis terpenes demonstrate acute anti-seizure activity in rodent brain slices in vitro. Planta Medica, 87(15), PC5-10.

Wang, G., & Dixon, R. A. (2009). Heterodimeric geranyl (geranyl) diphosphate synthase from hop (Humulus lupulus) and the evolution of monoterpene biosynthesis. Proceedings of the National Academy of Sciences, 106(24), 9914–9919.

Wang, Q., & Carmichael, G. G. (2004). Effects of length and location on the cellular response to double-stranded RNA. Microbiology and Molecular Biology Reviews, 68(3), 432–452.

Wang, Q., Yu, G., Chen, Z., Han, J., Hu, Y., & Wang, K. (2021). Optimization of protoplast isolation, transformation and its application in sugarcane (Saccharum spontaneum L). The Crop Journal, 9(1), 133–142.

Wang, R., He, L.-S., Xia, B., Tong, J.-F., Li, N., & Peng, F. (2009). A micropropagation system for cloning of hemp (Cannabis sativa L.) by shoot tip culture. Pak. J. Bot, 41(2), 603–608.

Wang, Y.-C., Klein, T. M., Fromm, M., Cao, J., Sanford, J. C., & Wu, R. (1988). Transient expression of foreign genes in rice, wheat and soybean cells following particle bombardment. Plant Molecular Biology, 11(4), 433–439.

Welling, M. T., Liu, L., Shapter, T., Raymond, C. A., & King, G. J. (2016). Characterisation of cannabinoid composition in a diverse Cannabis sativa L. germplasm collection. Euphytica, 208(3), 463–475.

Wesley, S. V., Helliwell, C. A., Smith, N. A., Wang, M., Rouse, D. T., Liu, Q., & Stoutjesdijk, P. A. (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. The Plant Journal, 27(6), 581–590.

White, P. R. (1939). Potentially unlimited growth of excised plant callus in an artificial nutrient. American Journal of Botany, 26(2), 59–64.

Wielgus, K., Luwanska, A., Lassocinski, W., & Kaczmarek, Z. (2008). Estimation of Cannabis sativa L. tissue culture conditions essential for callus induction and plant regeneration. Journal of Natural Fibers, 5(3), 199–207.

Wingard, S. A. (1928). Hosts and symptoms of ring spot, a virus disease of plants. Journal of Agricultural Research, 37(3), 127-154

Wolfe, S. A., Nekludova, L., & Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annual Review of Biophysics and Biomolecular Structure, 29(1), 183–212.

Wollner, H. J., Matchett, J. R., Levine, J., & Loewe, S. (1942). Isolation of a physiologically active tetrahydrocannabinol from Cannabis sativa resin. Journal of the American Chemical Society, 64(1), 26–29.

Wright, D. A., Townsend, J. A., Winfrey Jr, R. J., Irwin, P. A., Rajagopal, J., Lonosky, P. M., & Voytas, D. F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. The Plant Journal, 44(4), 693–705.

Wróbel, T., Dreger, M., Wielgus, K., & Słomski, R. (2020). Modified Nodal Cuttings and Shoot Tips Protocol for Rapid Regeneration of Cannabis sativa L. Journal of Natural Fibers, 1–10.

Wu, S., Zhu, H., Liu, J., Yang, Q., Shao, X., Bi, F., & Yi, G. (2020). Establishment of a PEG-mediated protoplast transformation system based on DNA and CRISPR/Cas9 ribonucleoprotein complexes for banana. BMC Plant Biology, 20(1), 1–10.

Wu, Z., Wouters, J., & Poulter, C. D. (2005). Isopentenyl diphosphate isomerase. Mechanism-based inhibition by diene analogues of isopentenyl diphosphate and dimethylallyl diphosphate. Journal of the American Chemical Society, 127(49), 17433–17438.

Xiao, Y., Chu, L., Sanakis, Y., & Liu, P. (2009). Revisiting the IspH catalytic system in the deoxyxylulose phosphate pathway: achieving high activity. Journal of the American Chemical Society, 131(29), 9931–9933.

Xie, Z., Allen, E., Wilken, A., & Carrington, J. C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proceedings of the National Academy of Sciences, 102(36), 12984–12989.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., & Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biology, 2(5).

Yan, P., Shen, W., Gao, X., Li, X., Zhou, P., & Duan, J. (2012). High-throughput construction of intron-containing hairpin RNA vectors for RNAi in plants. PLOS One, 7(5), e38186.

Yotoriyama, M., Ito, I., Takashima, D., Shoyama, Y., & Nishioka, I. (1980). Plant breeding of cannabis. Determination of cannabinoids by high-pressure liquid chromatography (author's transl). Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan, 100(6), 611.

Zhang, H., Xia, R., Meyers, B. C., & Walbot, V. (2015). Evolution, functions, and mysteries of plant ARGONAUTE proteins. Current Opinion in Plant Biology, 27, 84–90.

Zhang, S., Zhang, R., Song, G., Gao, J., Li, W., Han, X., & Li, G. (2018). Targeted mutagenesis using the Agrobacterium tumefaciens-mediated CRISPR-Cas9 system in common wheat. BMC Plant Biology, 18(1), 1–12.

Zhang, X., Xu, G., Cheng, C., Lei, L., Sun, J., Xu, Y., & Chen, X. (2021). Establishment of an Agrobacterium-mediated genetic transformation and CRISPR/Cas9-mediated targeted mutagenesis in Hemp (Cannabis Sativa L.). Plant Biotechnology Journal.

Zhou, X., Jacobs, T. B., Xue, L., Harding, S. A., & Tsai, C. (2015). Exploiting SNP s for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and redundancy. New Phytologist, 208(2), 298–301.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.

Zuardi, A. W. (2006). History of cannabis as a medicine: a review. Brazilian Journal of Psychiatry, 28(2), 153–157.