

https://dspace.org/

De novo production of Taxol intermediates by Saccharomyces Cerevisiae

Rabie, Renée

2022

https://hdl.handle.net/10539/37032

Downloaded from a DSpace repository. https://dspace.org/



DE NOVO PRODUCTION OF TAXOL INTERMEDIATES BY *SACCHAROMYCES CEREVISIAE*

by

Renée Rabie

(2006170)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

Supervisor: Professor Karl Rumbold

(Karl.Rumbold@wits.ac.za)

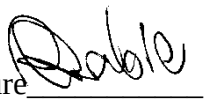
Advisor: Professor Marianne J Cronje

(marianne.cronje@wits.ac.za)

September 2022

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in Molecular and Cell Biology at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature

6th of September, 2022

ABSTRACT

Taxol is an invaluable anticancer molecule produced by Yew trees and their endophytic fungi. Harvesting taxol is difficult and often has low yields. For these reasons, a method to produce taxol heterologously in a fast-growing, well-studied, safe microbe is desirable. In this study, artificial genes were designed for the expression of two taxol biosynthesis pathway enzymes, as well as an assisting NADPH-cytochrome P450 reductase. The genes were designed to be compatible with the pCut transformation technique, which allows genomic integration into *Saccharomyces cerevisiae* strain BY4742. The genes were then divided into seven fragments. Two additional DNA fragments were amplified directly from the yeast genome because their complexity made them difficult and expensive to synthesise. These nine DNA fragments were designed to be assembled into three linear fragments of equal length for transformation of *S. cerevisiae*. Attempts at assembling these nine fragments into three inserts failed for various reasons, which largely came down to the complexity and integrity of the DNA, as well as the size of the fragments.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Professor Karl Rumbold, for all the assistance and knowledge imparted during my degree. I also wish to thank Michael Tobin for the much-needed technical advice.

CONTENTS

LIST OF FIGURES **1**

LIST OF TABLES **4**

ABBREVIATIONS **5**

INTRODUCTION **8**

Overview of taxol **8**

The utility of integrating the genes of taxol biosynthesis enzymes directly into the *S. cerevisiae* genome **10**

LITERATURE REVIEW **14**

The taxol biosynthesis pathway **14**

The components of the artificial genes designed to be integrated into the yeast genome **18**

The history, evolutionary purpose, and research utility of CRISPR-Cas9 **21**

The pCut transformation method **23**

***Saccharomyces cerevisiae* as host organism** **27**

DNA assembly methods **28**

Gibson assembly *28*

PCR splicing by overlap extension (SOE-PCR) *29*

Comparisons between DNA polymerases for SOE-PCR *31*

MATERIALS AND METHODS **33**

DNA Design **33**

Choosing the integration locus *33*

Design of the artificial genes *33*

Design of gBlocks for Gibson assembly *34*

DNA preparation **35**

DNA synthesised to order *35*

DNA extraction from yeast *36*

Primer design and PCR amplification of flanking regions *36*

Gibson assembly **37**

Gibson assembly of linear fragments *37*

Primer design for PCR amplification of synthetic gene fragments and pGEM®-T Easy vector *38*

PCR amplification of fragments and gel purification *40*

Gibson assembly of plasmid vectors *42*

Transformation of competent E. coli with pGEM®-T Easy plasmid *43*

PCR amplification of DNA fragments from assembled pGEM®-T Easy plasmid *44*

SOE-PCR **44**

Initial 3-fragment attempt *44*

Asymmetric PCR for SOE-PCR *45*

Splicing without primers *45*

Testing the integrity of Phusion-amplified DNA fragments after persistent SOE-PCR failure **47**

pCUT plasmids **47**

Growing the E. coli host bacteria *47*

Plasmid extraction *47*

The growth and preparation of the frozen stock of *S. cerevisiae* BY4742 . **48**

RESULTS AND DISCUSSION **50**

Yeast DNA extraction and PCR to amplify the flanking regions A and B **50**

PCR of designer fragments **50**

Gel purification of fragments **51**

Gibson assembly **54**

Linear Gibson assembly *54*

Gibson assembly into a plasmid vector *54*

Gibson assembly of Inserts 1 and 2 into the same vector *57*

PCR amplification of Gibson assembly product *57*

Transformation of competent E. coli with assembled plasmid vectors *59*

SOE-PCR **59**

Initial attempt *59*

SOE-PCR using asymmetric amplification *60*

SOE-PCR with no primers *61*

Reamplification of fragments to determine DNA integrity **63**

Growth of yeast **64**

CONCLUSION AND FUTURE RESEARCH **66**

REFERENCES **71**

ADDENDUM 1: DNA CONSTRUCTION **80**

Components of DNA construction **80**

Protein-encoding sequence *80*

Linker between protein and tag (18 bp) *81*

Solubility tag *81*

Promoter *82*

Terminator *82*

Artificial gene sequences **82**

Taxadiene synthase (4445 bp) *82*

Taxadiene 5α-hydroxylase (3356 bp) *83*

NADPH-cytochrome P450 reductase (4010 bp) *84*

DNA between cassettes (Cas9 target loci and attached PAM sequence) 85

DNA fragments for Gibson assembly 85

A (amplified from genomic DNA of S. cerevisiae BY4742) (949 bp) 85

R1 (1875 bp) 86

R2 (1875 bp) 86

R3 (1580 bp) 87

R4 (1580 bp) 87

R5 (1580 bp) 88

R6 (1875 bp) 88

R7 (1875 bp) 89

B (amplified from genomic DNA of S. cerevisiae BY4742) (920 bp) 89

V1 (Amplified from pGEM®-T Easy plasmid vector) (3064 bp) 89

V2 (Amplified from pGEM®-T Easy plasmid vector) (3063 bp) [Not used for final method] 90

V3 (Amplified from pGEM®-T Easy plasmid vector) (3062 bp) 91

V1.5 (Amplified from pGEM®-T Easy plasmid vector) (3060 bp) 92

1000 bp homologous arms for the three inserted target sequences 92

Inserted target sequence 805a 92

Inserted target sequence 1206a 93

Inserted target sequence 607c 94

Final DNA sequence 95

Legend 95

DNA sequence 95

ADDENDUM 2: PCR TABLES 100

LIST OF FIGURES

Figure	Page	Description
1	15	A potential taxol biosynthesis pathway constructed from the KEGG database (Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa et al, 2021) McElroy and Jennewein (2018), and Wang et al (2021), and including a putative intermediate (Walker & Croteau, 2000; Croteau et al, 2006; Wang et al, 2021; KEGG database [Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa et al, 2021]) in red. Enzymes that catalyse the formation of other branches of taxoid molecules have been excluded. Characterised enzymes and their reactions are shown in green. Uncharacterised enzymes are indicated in red. Dotted arrows indicate that numerous enzymatic reactions are taking place between two intermediates, and that the order of these reactions is unknown. Chemical structures were taken from the KEGG database (Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa et al, 2021).
2	17	Biosynthesis pathways of substrates used in the taxol biosynthesis pathway. A) Benzoate-CoA ligase catalyses the replacement of the terminal hydroxyl group of benzoate with CoA (KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). B) Phenylalanine aminomutase moves the amine group from the α carbon atom to the β carbon atom, and β-phenylalanyl-CoA ligase replaces the terminal hydroxyl group with CoA (Wang *et al*, 2021).
3	23	A schematic representation of the pCut transformation method developed by Apel et al (2017). A) The yeast is simultaneously transformed with the donor DNA and the pCut plasmid. B) Once Cas9 and its sgRNA are expressed, Cas9 cleaves the yeast DNA at one of the 23 predetermined loci. Cas9 cleaves between the 17th and 18th bp of the target site. The donor DNA has ~1000 bp homology with the yeast genomic DNA on each side. C) The yeast uses its mechanisms for HDR to insert the heterologous DNA into the yeast genome.
4	26	Recycling of pCut sequences at target loci with optimum integration efficiencies. A) The 20 bp DNA sequence of a pCut target with low integration efficiency is equipped with a PAM and added to the donor DNA sequence that is transformed into the yeast along with the appropriate pCut plasmid. B) The genomic DNA is cleaved by Cas9, and the new DNA is integrated through HDR.

		C) The yeast genome now contains a heterologous gene and a new target site that can be utilised in subsequent transformations with a different pCut plasmid.
5	28	The principle of Gibson assembly. 5’ ends of DNA fragments are degraded by T5 exonuclease, leaving long 3’ sticky ends. Homologous sticky ends anneal to each other. The gaps are then filled in 5’ to 3’ by Phusion DNA polymerase and the gaps in the phosphodiester backbone of the DNA are covalently closed by Taq DNA ligase. The 3’ sticky ends of the assembled fragment are not repaired in a linear assembly, but will disappear in a circular assembly, where each sticky end is matched with another that can act as a primer for 5’ to 3’ amplification.
6	30	DNA fragments priming each other during SOE-PCR. Fragment A can be amplified into the assembled product. Fragment B cannot be amplified, because there is no way to proceed with amplification in the 5’ to 3’ direction. Fragment A is amplified into the assembled product.
7	40	The two possible clones that could be designed when assembling theoretical DNA fragment XYZ into a vector. Vector A shows the fragment on the plus strand of the vector and vector B shows the fragment on the minus strand of the vector. Images not drawn to scale for pGEM®-T Easy vector or any of the inserts used in this study. For each of the three initial vectors of this study, as well as the fourth unplanned vector, the DNA sequence of interest was present on the minus strand.
8	54	Graphs comparing the size of the purified fragment with the final yield of the fragment. A) When all fragments of the first purification were included in the calculation, there was no correlation between fragment size and the final yield. B) When outlier fragment A of the first purification was removed, there was only a very weak correlation between the fragment size and the final yield. C) Fragments amplified with Phusion and purified in the second purification attempt showed no correlation between fragment size and yield, with the outlier fragment R4 as neither the longest nor the shortest.
9	56	Partially successful Gibson assembly of Phusion-amplified fragments after 1 hour of incubation at 50°C. Lane 1 contains GeneRuler 1kb DNA ladder. Lane 2 contains the products for the Gibson assembly of Vector 1. Lane 3 contains the products for the Gibson assembly of Vector 3. DNA migration was slower nearer to the centre of the gel than at the edges, producing a slanted appearance.

10	57	Figure 10: Partially successful Gibson assembly after 4 hours of incubation at 50°C. Lanes 1 and 7 contain GeneRuler 1 kb DNA ladder. Lane 2 contains the Gibson assembly products for the assembly of Vector 1. Lane 3 contains the equivalent amount of vector fragment V1 to compare the fluorescence and travel distance of the unassembled vector fragment. Lane 5 contains unassembled vector fragment V3 at an equal amount to the assembly reaction mix of Vector 3. Lane 6 contains the Gibson assembly products for the assembly of Vector 3.
11	59	Gel electrophoresis of PCR products amplified directly from diluted unpurified Gibson assembly product. Lane 1 contains DNA ladder. Lanes 2-4 contain PCR results from Insert 1. Lanes 5-7 contain PCR results from Insert 3. Lane 8 contains a Gibson assembly of Vector 1.
12	62	Gel electrophoresis results of the 100-cycle primer-free SOE-PCR. Lane 1 contains GeneRuler 1 kb DNA ladder. Lane 2 and lane 6 have a faint band just over 4000 bp. The other lanes have smears from the top. No smear was brightest around where the expected assembled fragment would be.
13	63	Results of SOE-PCR without primers and the subsequent attempt at reamplification. The contrast of both images was increased by 40% and the brightness reduced by 20% to show the very faint bands more clearly. On both gels, lane 4 contains GeneRuler 1 kb DNA ladder and lanes 3 and 5 contain the PCR results, split into two volumes of 30 µL each including loading buffer. The gel on the right shows the initial results for overlap extension PCR of fragment R3+4.

LIST OF TABLES

Table	Page	Description
1	24	Integration efficiencies and target sequences of different loci as determined by Apel *et al* (2017) in descending order.
2	36	Molecular weight of DNA fragments.
3	41	All primers used in this study
4	54	DNA yield by method.
A2.1	101	PCR mix for all fragments (OneTaq)
A2.2	101	PCR mix for all fragments (Phusion)
A2.3	102	Thermal cycler parameters for PCR amplification of gBlocks (OneTaq)
A2.4	103	Thermal cycler parameters for PCR amplification of gBlocks (Phusion)
A2.5	104	PCR primers for gBlocks
A2.6	104	PCR primers for flanking regions
A2.7	105	PCR primers for vectors
A2.8	106	Gibson isothermal buffer mix
A2.9	106	Gibson 2X master mix
A2.10	106	PCR mix for all fragments (Kapa)
A2.11	107	Thermal cycler parameters for PCR amplification of gBlocks (Kapa)

ABBREVIATIONS

dH_2O	Distilled water
E. coli	*Escherichia coli*
NCBI	National Centre for Biotechnology Information
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
Spp.	Species
CO_2	Carbon dioxide
M	Molar
m	Milli-
µ	Micro-
n	Nano-
p	Pico-
T. cuspidata	*Taxus cuspidata*
bp	Base pair(s)
kbp	Kilo-base pair
Amp	Ampicillin
YPD	Yeast extract peptone dextrose
CRISPR	Clustered Regularly-Interspaced Short Palindromic Repeats
DTT	Dithiothreitol
SOE-PCR	PCR splicing by overlap assembly
LB	Lysogeny broth
U/µL	Units per microlitre
ssDNA	Single-stranded DNA
dsDNA	Double-stranded DNA
NHEJ	Non-homologous end-joining
HDR	Homology-directed repair
ICL	Isocitrate lyase
PRM	Pheromone-Regulated Membrane protein
PAM	Protospacer-adjacent motif

CoA	Coenzyme A
MBP	Maltose-binding protein
YIps	Yeast integrating plasmids
YEps	Yeast centromere plasmids
s	Second(s)

Chapter 1:
Introduction

Overview of taxol

Taxol, or paclitaxel, is a terpenoid produced predominantly in *Taxus* spp. and is an invaluable drug used in cancer treatment (El-Sayed *et al*, 2020). The scarcity of taxol sources, the difficulty of extracting taxol, and the low yield of taxol extracted from natural sources makes it an attractive candidate for heterologous production in a fast-growing, high-yield microorganism.

Cancer is a form of tissue growth characterised by, among other important attributes, uncontrolled mitosis of cells, a resistance of these cancer cells to apoptosis, and a lack of differentiation of cancer cells or dedifferentiation of the cancer cell of origin (Hanahan, 2022). These features are relevant to using taxol and other chemotherapeutics as an anticancer treatment. Taxol is a cytotoxic anticancer drug that prevents mitosis of human cells (Arbuck *et al*, 1993; Gallego-Jara *et al*, 2020; Lim *et al*, 2022) that was isolated and characterised by Wani *et al* in 1971, after clinical trials were conducted in the 1960s to determine its effect on various forms of cancer (Martin, 1993; Gallego-Jara *et al*, 2020; Lim *et al*, 2022). Taxol inhibits mitosis by binding to tubulin and thereby stabilising the cell's microtubules, preventing chromosomes from attaching to microtubules and preventing the cell from exiting the G2 phase of mitosis (Gallego-Jara *et al*, 2020). The progenitor cells of differentiated tissues have comparatively slower growth or no growth depending on the tissue (Hanahan, 2022), which allows the mitosis-arresting taxol treatment to affect cancerous tissue more than normal tissue. Taxol is also believed to play a role in apoptosis of cancer cells by inactivating proteins that resist apoptosis and activating proteins that facilitate apoptosis (Gallego-Jara *et al*, 2020).

Taxol is generally harvested from the bark of the pacific yew tree (*Taxus brevifolia*), and occasionally from other species of yew (*Taxus* spp.). Yew trees grow slowly and harvesting all of the bark from a live yew tree will kill it (Schepartz, 1993). The bark of the yew trees can be partially stripped from a living tree and allow the tree to recover and continue growing, and there are laws regulating the logging of yew trees to reduce the waste of valuable bark (Schepartz, 1993). The harvest process is

labour-intensive and expensive, and yields little taxol compared to the weight of bark processed (Martin, 1993; Schepartz, 1993). To produce only 1 kg of taxol requires the bark of around 2500 trees (Martin, 1993). Taxol intermediates are often also harvested to be used to synthesise the final product, but due to the portion of the taxol biosynthesis pathway enzymes that have not yet been fully elucidated, processes do not yet exist to synthesise taxol from certain intermediates further up the pathway from the final product (Schepartz, 1993; Mamadalieva & Mamedov, 2020). The needles of *Taxus* spp. also yield small amounts of taxol and taxol intermediates, from which taxol can be synthesised semi-synthetically (Schepartz, 1993; Liu *et al*, 2016).

Although the existence of taxol-producing endophytes used to be in dispute (Heinig *et al*, 2013), subsequent studies have not only shown taxol production, but have shown that taxol production in endophytic fungi can be increased by altering the growth medium of the fungi (Subban *et al*, 2019). Taxol-producing endophytic fungi even play a vital role in the protection of the host trees by weaponizing the antifungal properties of taxol against wood-decaying fungi that might enter through cracks in the bark (Soliman *et al*, 2015).

Although taxol production from plant cell cultures is possible and has been studied extensively, there are many drawbacks. The environment in which the cell cultures are grown is carefully managed, which helps prevent contamination and environmental damage to the cells and the taxol, but it is difficult to maintain and has a low and inconsistent yield (Liu *et al*, 2016). For this reason, the prospect of producing taxol in a hardy industrial microbe is very attractive.

This study is another attempt to use the biosynthetic potential of single cell organisms to build up toward the *de novo* synthesis of taxol. It builds on the work of previous researchers who have introduced taxol biosynthesis enzymes into both yeasts and bacteria and fine-tuned the performance of these enzymes in their new environment. While previous studies have focussed on the integration of taxol biosynthesis genes into the *S. cerevisiae* genome, none have used the pCut method developed by Apel *et al* (2017), apart from their successful attempt to produce taxadiene synthase when developing their method. The pCut method simplifies the

process of genomic integration by utilising one of 27 pre-made plasmids to express Cas9 and a guide RNA sequence targeting one of 23 predetermined loci, eliminating the need to clone genes into a plasmid. The pCut method achieves scarless integration of a gene into one of these 23 loci, which is ideal for the multi-gene pathway of taxol biosynthetic enzymes, although the suboptimal integration efficiency of certain target loci could present a problem.

The utility of integrating the genes of taxol biosynthesis enzymes directly into the *S. cerevisiae* genome

Numerous researchers have successfully integrated part of the taxol biosynthesis pathway into microorganisms. The host organisms are usually bacterial or yeast, to allow for rapid proliferation and easy growth conditions. Heterologous expression of taxol biosynthesis pathway enzymes is not only an important part of building towards total synthesis of taxol in microbial cell factories, but can play an important role in the characterisation of newly discovered potential pathway enzymes (Walker *et al*, 2004).

To this researcher's knowledge, the largest number of heterologous taxol biosynthesis genes ever expressed in *Saccharomyces cerevisiae* in one study is eight, which was achieved by DeJong *et al* (2005) using plasmid vectors; however, only five biosynthesis enzymes were expressed at one time within the same yeast. While DeJong *et al* (2005) successfully expressed up to five taxol biosynthesis pathway genes in *S. cerevisiae* at once and thereby gathered a great deal of information about the heterologous expression of these enzymes, their study also provides a glimpse into the unfeasibility of expressing the entire taxol biosynthesis pathway in *S. cerevisiae* using plasmids. DeJong *et al* (2005) cloned taxol biosynthesis genes into plasmid vectors, with either one or two genes per vector, and found that cloning two genes into the same vector could cause a decrease in the expression of that gene. This sacrifice would likely be necessary when reconstructing the whole pathway using plasmids, as cloning each gene into its own vector would result in over 20 distinct vectors, each with their own selection marker. Expression of these genes using plasmid vectors would also place a

considerable metabolic burden on the cell when compared to heterologous genes integrated into the *S. cerevisiae* genome (Zhang *et al*, 1996; Egger *et al*, 2020).

Another benefit of genomic integration lies in the relative stability of the heterologous gene. Plasmids in *S. cerevisiae* are often unstable and can be lost or significantly altered by the yeast cell (Zhang *et al*, 1996). The number of plasmid copies within any given cell is also uncertain, because plasmid vectors are not necessarily divided equally between mother and daughter cells (Zhang *et al*, 1996). Yeast integrating plasmids, linearised DNA fragments designed to be integrated into the yeast genome, often have a low integration efficiency and a small but uncertain copy number (Sikorski & Hieter, 1989; Zhang *et al*, 1996). The pCut method has the advantage of integrating only a single copy of the DNA of interest into one of 23 predetermined loci with >90% integration efficiency in 10 of these loci, which allows for a very controlled assembly of a new strain of *S. cerevisiae* (Apel *et al*, 2017).

Although plasmids remain a useful avenue for the heterologous expression of individual taxol biosynthesis genes, engineering a strain of *S. cerevisiae* to carry out preceding steps in the pathway could remove some of the labour and some of the cost involved in future studies on different taxol biosynthesis enzymes. A strain of *S. cerevisiae* with taxol biosynthesis genes in its genome could form the basis of future studies integrating more taxol biosynthesis genes, both known and putative, into the genome, and future studies aiming to increase the yield of taxol intermediates. If an *S. cerevisiae* strain with the genes for taxol biosynthesis enzymes within the genome exists, the study of putative taxol biosynthesis enzymes could also be carried out without the need for synthetic or semisynthetic pathway intermediates, or the need to transform the yeast with many heterologous genes.

This study does not aim to reproduce the entire taxol biosynthesis pathway into *S. cerevisiae* or to maximise the yield of the taxol biosynthesis intermediates of interest. This study aims to produce a strain of *S. cerevisiae* that efficiently carries out the first two steps of the taxol biosynthesis pathway by integrating the taxol biosynthesis genes directly into the genome of *S. cerevisiae* BY4742. This new strain could serve as a scaffold for future research regarding taxol biosynthesis in

S. cerevisiae. While this study will focus on the first two enzymes in the pathway, and the accompanying NADPH-cytochrome P450 reductase, future studies could build on it to eventually reproduce the entire taxol pathway in yeast for affordable mass-production of taxol. To achieve this goal, it is necessary to design artificial genes to be inserted into the genome and prepare cultures of competent *S. cerevisiae* BY4742 for transformation using the method set out by Apel *et al* (2017). Afterwards, the yeast must be co-transformed with the relevant pCut plasmid and the donor DNA. Successfully transformed strains will show on selective media, and from these strains successful integration can be determined through colony PCR.

Chapter 2:
Literature Review

LITERATURE REVIEW

The taxol biosynthesis pathway

Unfortunately, the pathway for taxol production has not yet been fully elucidated. While new pathway enzymes are discovered sporadically, it is a slow process, with many pathway dead ends and intersections to contend with (Kuang *et al*, 2019). As of now, 14 of over 20 pathway enzymes have been discovered (McElroy & Jennewein, 2018), and researchers are continuously working on elucidating the pathway further. Of particular note is that Kuang *et al* (2019) developed an innovative high-throughput sequencing method to analyse the mRNA transcripts of *Taxus cuspidata*. Kaung *et al* (2019) identified numerous enzymes that could potentially fill in the gaps in the pathway, and narrowed the candidates down to 9 cytochrome P450 enzymes and 7 acetyltransferase enzymes by comparing the transcription frequency of the candidates to the transcription frequency of known taxol biosynthesis genes. They published their findings on 16 potential taxol biosynthesis pathway enzymes, along with potential transcription factor enzymes. While every enzyme within the taxol biosynthesis pathway has not yet been identified, the types of enzymes required to carry out each step of the process have been determined (McElroy & Jennewein, 2018).

According to the Kegg database (Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021), the closest common precursor to taxol in *S. cerevisiae* is geranylgeranyl-diphosphate. The current putative taxol biosynthesis pathway is shown in Figure 1. The heterologous enzyme that irreversibly starts the taxol biosynthesis pathway is taxadiene synthase, a cyclase enzyme that converts the noncyclic molecule geranylgeranyl-diphosphate into the polycyclic molecule taxadiene (McElroy & Jennewein, 2018; Wang *et al*, 2021, KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). Afterwards, taxadiene 5α-hydroxylase adds a hydroxyl group to carbon atom C5 to form taxadiene-5α-ol, before the pathway branches into two variations: either the C5 hydroxyl group is acetylated by taxadiene-5α-ol O-acetyltransferase to form taxadiene 5α-yl acetate, which is in turn hydroxylated by taxane 10β-hydroxylase

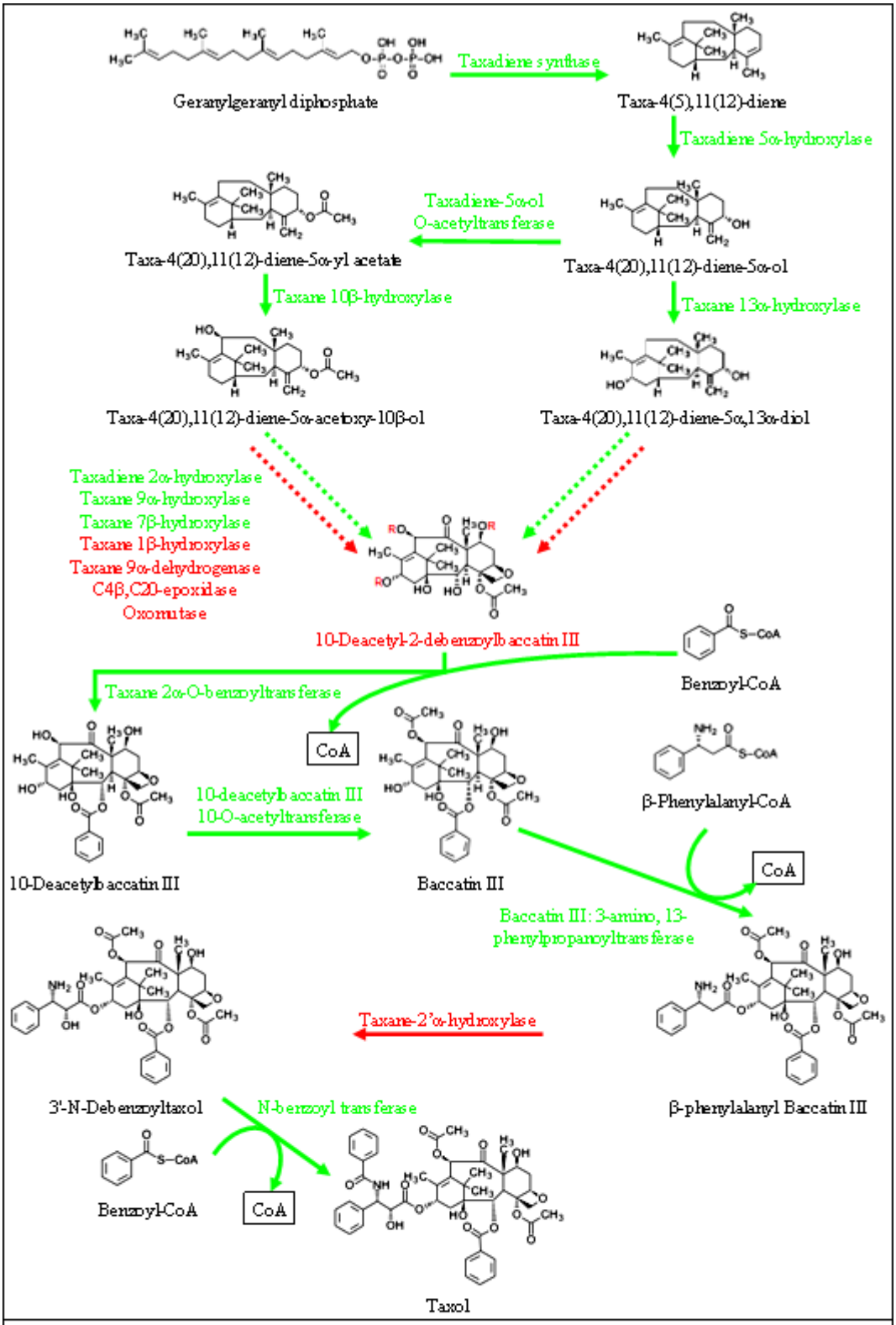


Figure 1: A potential taxol biosynthesis pathway constructed from the KEGG database (Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021) McElroy and Jennewein (2018), and Wang *et al* (2021), and including a putative intermediate (Walker & Croteau, 2000; Croteau *et al*, 2006; Wang *et al*, 2021; KEGG database [Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]) in red. Enzymes that catalyse the formation of other branches of taxoid molecules have been excluded. Characterised enzymes and their reactions are shown in green. Uncharacterised enzymes are indicated in red. Dotted arrows indicate that numerous enzymatic reactions are taking place between two intermediates, and that the order of these reactions is unknown. Chemical structures were taken from the KEGG database (Kanehisa & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021).

to form taxadiene-5α-acetoxy-10β-ol, or the carbon atom C13 is hydroxylated to form taxadiene 5α,13α-diol (McElroy & Jennewein, 2018; Wang *et al*, 2021; KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). Both precursors, taxadiene-5α-acetoxy-10β-ol and taxadiene 5α,13α-diol, then undergo a series of reactions for which the order is not yet fully known (McElroy & Jennewein, 2018; Wang *et al*, 2021). The molecule is hydroxylated at carbon atoms C1, C2, C7, and C9 by the enzymes taxadiene 2α-hydroxylase, taxane 7β-hydroxylase, and taxane 9α-hydroxylase, which have been discovered and characterised, and the enzyme taxane 1β-hydroxylase, which has not yet been discovered (McElroy & Jennewein, 2018; Wang *et al*, 2021). Taxane 9α-dehydrogenase, an enzyme that has not yet been characterised, oxidises the hydroxyl group at carbon atom C9 to form a ketone group (McElroy & Jennewein, 2018; Wang *et al*, 2021). Recently, Sanchez-Muñoz *et al* (2020) isolated a putative enzyme that was shown to make this modification. This period is also when the oxetane ring structure forms. It is not yet clear how the oxetane ring forms, and two potential enzymes have been suggested to be responsible but have not yet been discovered: C4β,C20-epoxidase and oxomutase (McElroy & Jennewein, 2018; Wang *et al*, 2021). This is unfortunate, because the oxetane ring is considered to play an important role in maintaining the molecular conformation of taxol (Boge *et al*, 1999) and in tubulin binding through noncovalent hydrogen-bonding (Samaranayake *et al*, 1991; Wang *et al*, 2000; Snyder *et al*, 2001); however, the significance of its role in the antimitotic activities of taxol differs between studies. These reactions produce the theoretical intermediate 10-deacetyl-2-debenzoylbaccatin III (Walker & Croteau, 2000; Croteau *et al*, 2006; Wang *et al*, 2021; KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]).

The next steps of the pathway require the assembly of two substrates, benzoyl-CoA and β-phenylalanyl-CoA. The assembly of benzoyl-CoA from benzoate and coenzyme A (CoA) can be catalysed by benzoate-CoA ligase in a reaction that utilises ATP and forms AMP (KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). Although no NCBI database entries or other literature for a *Taxus* spp. or *S. cerevisiae* benzoate-CoA ligase could be found, there are

numerous entries for green plants (79) and ascomycete fungi (6); therefore, if necessary, it might be possible to substitute one of these enzymes for one from *Taxus* spp. The other substrate, β-phenylalanyl-CoA, is the result of two modifications made to α-phenylalanine: first, phenylalanine aminomutase converts α-phenylalanine to β-phenylalanine, then β-phenylalanyl-CoA ligase replaces the terminal hydroxyl group with CoA (McElroy & Jennewein, 2018; Wang *et al*, 2021). These reactions are shown in Figure 2.

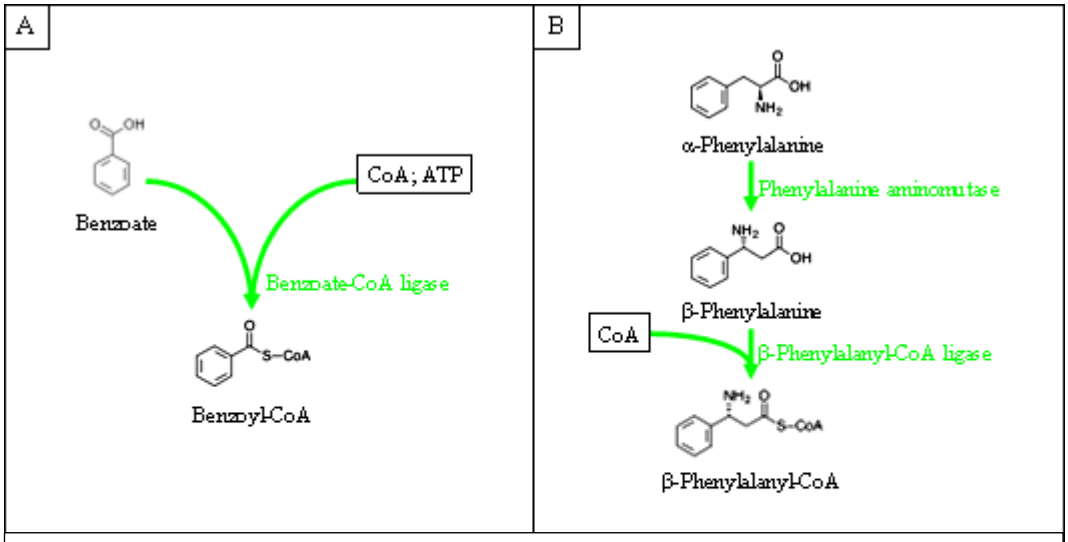


Figure 2: Biosynthesis pathways of substrates used in the taxol biosynthesis pathway. A) Benzoate-CoA ligase catalyses the replacement of the terminal hydroxyl group of benzoate with CoA (KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). B) Phenylalanine aminomutase moves the amine group from the α carbon atom to the β carbon atom, and β-phenylalanyl-CoA ligase replaces the terminal hydroxyl group with CoA (Wang *et al*, 2021).

Taxane 2α-O-benzoyltransferase catalyses the benzoylation of the hydroxyl group at carbon atom C2 to form 10-deacetylbaccatin III, using benzoyl-CoA as a substrate (McElroy & Jennewein, 2018; Wang *et al*, 2021; KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]). The hydroxyl group on carbon atom 10-deacetylbaccatin III is then acetylated by 10-deacetylbaccatin III 10-O-acetyltransferase to form baccatin III (McElroy & Jennewein, 2018; Wang *et al*, 2021; KEGG database [Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021]).

Baccatin III: 3-amino, 13-phenylpropanoyltransferase attaches a β-phenylalanyl group to baccatin 3 to form β-phenylalanyl Baccatin III. Taxane-2'α-hydroxylase, which has not yet been characterised, hydroxylases the carbon molecule C2' on this β-phenylalanyl group to form 3'-N-Debenzoyltaxol. Finally, N-benzoyl transferase benzoylates the amine group on carbon atom C3' to form taxol.

Jennewein *et al* (2005) found that an NADPH-cytochrome P450 reductase from *T. cuspidata* co-expressed with the cytochrome P450 enzyme taxane 10β-hydroxylase dramatically improved the activity of the latter enzyme through improved redox coupling between the *T. cuspidata* enzymes when compared to the redox coupling between native yeast reductases and taxane 10β-hydroxylase.

Notably, when DeJong *et al* (2005) assembled their yeast with five taxol biosynthesis genes, they included geranylgeranyl-diphosphate synthase as the first enzyme of the pathway, even though *S. cerevisiae* produces geranylgeranyl-diphosphate, albeit in low concentrations (Jiang *et al*, 1995; Song *et al*, 2017). While this study does not include geranylgeranyl-diphosphate synthase, it could be a good target for heterologous expression in future studies that aim to increase the yield of taxol intermediates.

The components of the artificial genes designed to be integrated into the yeast genome

This study utilised artificial genes. Artificial genes are DNA sequences not found in nature that can be used to express heterologous proteins in suitable hosts (Hughes & Ellington, 2017). Artificial genes can eliminate certain problems that may be present when dealing with naturally occurring genes, including introns. *S. cerevisiae* naturally only has 0.05 introns per gene, a massive decrease from intron-rich ancestors (Stajich *et al*, 2007). The removal of introns creates a smaller DNA cassette, which provides an advantage during the homologous recombination step inside the yeast (Apel *et al*, 2017). DNA can be designed with high-yield promoters and terminators from the host organism, and protein tags can be genetically grafted onto heterologous proteins to improve their function in an environment they are not naturally suited towards.

Artificial genes can also be codon-optimised for the host organism. Every organism shows a bias towards certain codons, which could exist for many reasons, including the relative abundance of certain tRNA molecules (Sharp & Li, 1987; Plotkin & Kudla, 2011), the optimal GC-content of the genome (Sharp & Li, 1987), and the RNA secondary structures caused by the nucleotide sequence, which can influence the level of expression of heterologous proteins (Kudla *et al*, 2009; Plotkin & Kudla, 2011). Optimising the sequence to fit within the codon bias of the host organism can vastly increase the yield of heterologous proteins (Sharp & Li, 1987; Plotkin & Kudla, 2011). It should be mentioned that Moriyama and Powell (1998) found that the bias of *S. cerevisiae* towards certain codons decreased as the length of the gene increased, but increased as the expression level of the gene increased.

For the purposes of this study, the first two enzymes of the pathway are of interest. Taxadiene synthase is a cyclase enzyme with poor solubility in the cytosol which is likely localised in the plastids of *Taxus* spp. (McElroy & Jennewein, 2018; Wang *et al*, 2021). Apel *et al* (2017) solved the problem of the solubility of taxadiene synthase using protein tags. They genetically grafted a protein onto the N- and/or C-terminal of a protein of interest to imbue the protein of interest with a certain physical characteristic, like solubility or increased stability, to improve their function outside of their native environment (Apel *et al*, 2017). When Apel *et al* (2017) used taxadiene synthase as one of the model enzymes to design the protein tag library for the pCut transformation method, they found that the lack of solubility of taxadiene synthase led to deformities in the yeast and a low but significant titre of taxadiene that could not be improved regardless of the strength of the promoters they tested; however, the taxadiene titre could be improved significantly by tagging the protein with the soluble maltose-binding protein (MBP) from *Escherichia coli* on either the N-terminal or the C-terminal, with the C-terminal-tagged enzyme yielding a titre more than 15-fold as high as the untagged protein (Apel *et al*, 2017). Taxadiene 5α-hydroxylase and the NADPH-cytochrome P450 reductase are also hydrophobic and are likely localised to the endoplasmic reticulum in *Taxus* spp. cells (McElroy & Jennewein, 2018; Wang *et al*, 2021). For this study, the C-terminal MBP tag was chosen to improve solubility for all three heterologous enzymes.

The *ICL1* promoter is a high-yield glucose-repressed promoter for which the activity and yield compared to promoters commonly utilised for heterologous protein expression was examined by Maury *et al* (2018). In *S. cerevisiae*, it promotes the transcription of the *ICL1* gene, which encodes isocitrate lyase (ICL1) (Fernández *et al*, 1992; Lazarow, 2016). The isocitrate lyase enzyme was characterised in detail by Fernández *et al* (1992). Isocitrate lyase breaks the substrate isocitrate into the products glyoxylate and succinate, which forms an essential part of the glyoxylate cycle (Chew *et al*, 2019). The glyoxylate cycle allows the yeast to utilise two-carbon (C_2) molecules, including ethanol, to produce four-carbon (C_4) molecules for numerous downstream applications including gluconeogenesis when glucose levels in the growth medium become low (Fernández *et al*, 1992; Lazarow, 2016; Chew *et al*, 2019). The *ICL1* promoter becomes briefly active when the concentration of glucose in the growth medium becomes limiting, which allows for exponential growth under fermentative conditions during the log phase, followed by a burst of rapid expression of the heterologous gene when glucose becomes limiting, after which it will show minimal activity for the next 10 hours (Maury *et al*, 2018). The *ICL1* promoter therefore allows controlled expression of heterologous proteins without the need for additional inducers, and shifts the metabolic burden of heterologous protein expression until after the exponential growth phase of *S. cerevisiae*. Maury *et al* (2018) showed that this short period of expression also did not negatively impact the yield of the heterologous protein, which was expressed more than heterologous proteins with a promoter for constitutive expression. The short period of activity might in fact be beneficial to the survival of the yeast by reducing the metabolic burden of expressing heterologous genes. Reducing the metabolic burden of taxol biosynthesis genes is extremely important when considering the sheer number of enzymes involved in taxol biosynthesis. For these reasons, *ICL1* was chosen as the promoter for all three heterologous genes.

A high-capacity terminator is a terminator that leads to the increased expression of the gene preceding it (Curran *et al*, 2013). A terminator determines the length and character of the 3’ untranslated region and the extent to which the transcript is polyadenylated, which influence the half-life of the mRNA transcript; the longer

the half-life of the transcript, the more times it may be translated (Curran *et al*, 2013). A longer mRNA half-life reduces the number of times that a gene needs to be transcribed to attain a high yield of the heterologous protein, which reduces the metabolic burden of expressing the heterologous gene relative to the yield of the heterologous protein (Curran *et al*, 2013).

The *PRM9* gene in *S. cerevisiae* encodes one of the yeast's pheromone-regulated membrane proteins (Heiman & Walter, 2000). The *PRM9* terminator is a high-capacity terminator characterised by Curran *et al* (2013) during a study where they determined that a high-capacity terminator can compensate for a low-yield promoter, and that the yield of even a high-yield promoter could be significantly improved by pairing it with a high-capacity terminator. While the exact size of this terminator is not yet known, Curran *et al* (2013) utilised the first 250 bp after the end of the coding DNA of the *PRM9* gene. The *PRM9* terminator was chosen as the terminator for all three heterologous genes.

Lastly, the stop codon UAA (TAA in DNA), which is the most efficient stop codon in yeast (Bonetti *et al*, 1995), was chosen for all sequences to minimise the chances of the ribosomes translating the mRNA past the stop codon and creating a mutant protein.

The history, evolutionary purpose, and research utility of CRISPR-Cas9

The 29-32 bp clustered regularly-interspaced short palindromic repeats (CRISPR) present in a wide range of prokaryotic cells were first discovered in *Escherichia coli* by Ishino *et al* (1987), where they mentioned it in the final paragraph of their discussion and theorised that they could play a role in the stabilisation of RNA, but ultimately concluded that their purpose remained unknown. This discovery was soon validated by Nakata *et al* (1989), who found these repeats in *E. coli* and numerous other bacterial species, proving that the phenomenon was not unique to *E. coli*. These sequences were studied further by Francisco J.M. Mojica, who published an unrelated paper regarding transcription in *Haloferax mediterranei* where he also noted the presence of these repeats (Mojica *et al*, 1993), and then went on to characterise these repeats in another article on two different *Haloferax*

species, proving that they could also be present in archaea (Mojica *et al*, 1995). Mojica *et al* (1995) theorised that the repeats could be used in replicon partitioning and continued to study this hypothesis for the next few years. Tang *et al* (2002) were the first to identify that these repeated sequences code for RNA that would be edited to a stable non-messenger RNA.

Cas9 forms part of CRISPR, which likely evolved as a bacterial defence against viral DNA (Hsu *et al*, 2014). In nature, CRISPR is usually guided by a complex of two pieces of RNA; however, when utilised for biotechnological research, this complex is generally simplified into a single guide RNA, or sgRNA (Hsu *et al*, 2014; Kennedy *et al*, 2015). The RNA complex that guides Cas9 is processed from an RNA transcript of CRISPR, which contains short sequences of DNA (protospacers) between its short palindromic repeats (Hsu *et al*, 2014). The protospacer sequences are often taken from viral or other potentially threatening foreign DNA, which means that Cas9 equipped with its guide RNA complex can cleave invasive DNA to fend off infection (Rath *et al*, 2015).

Cas9 is a programmable nuclease that targets a specific 20 bp DNA sequence complementary to the guide RNA, which is followed by a protospacer-adjacent motif (PAM) with the sequence NGG (Hsu *et al*, 2014, Peng *et al*, 2016; Liu *et al*, 2019). It has a bilobed crystal structure, two DNA nicking domains, and an RNA complex that guides it to its target DNA sequence (Ran *et al*, 2013, Nishimasu *et al*, 2014). The nicking domain HNH nicks the DNA strand complementary to the target sequence of the guide RNA complex, and the nicking domain RuvC nicks the noncomplementary strand to form a double-strand break with blunt ends.

Cas9 holds great potential for genetic engineering by introducing double-strand breaks into the target DNA, allowing for repair by non-homologues end-joining (NHEJ) in order to introduce mutations, or by homology-directed repair (HDR), which allows the researcher to insert new DNA into a target locus (Hsu *et al*, 2014, Peng *et al*, 2016). Alternatively, the endonuclease domains can be deactivated or replaced with different protein domains, allowing these domains to be targeted towards a specific DNA sequence in a way that was never possible before (Liu *et al*, 2019).

The pCut transformation method

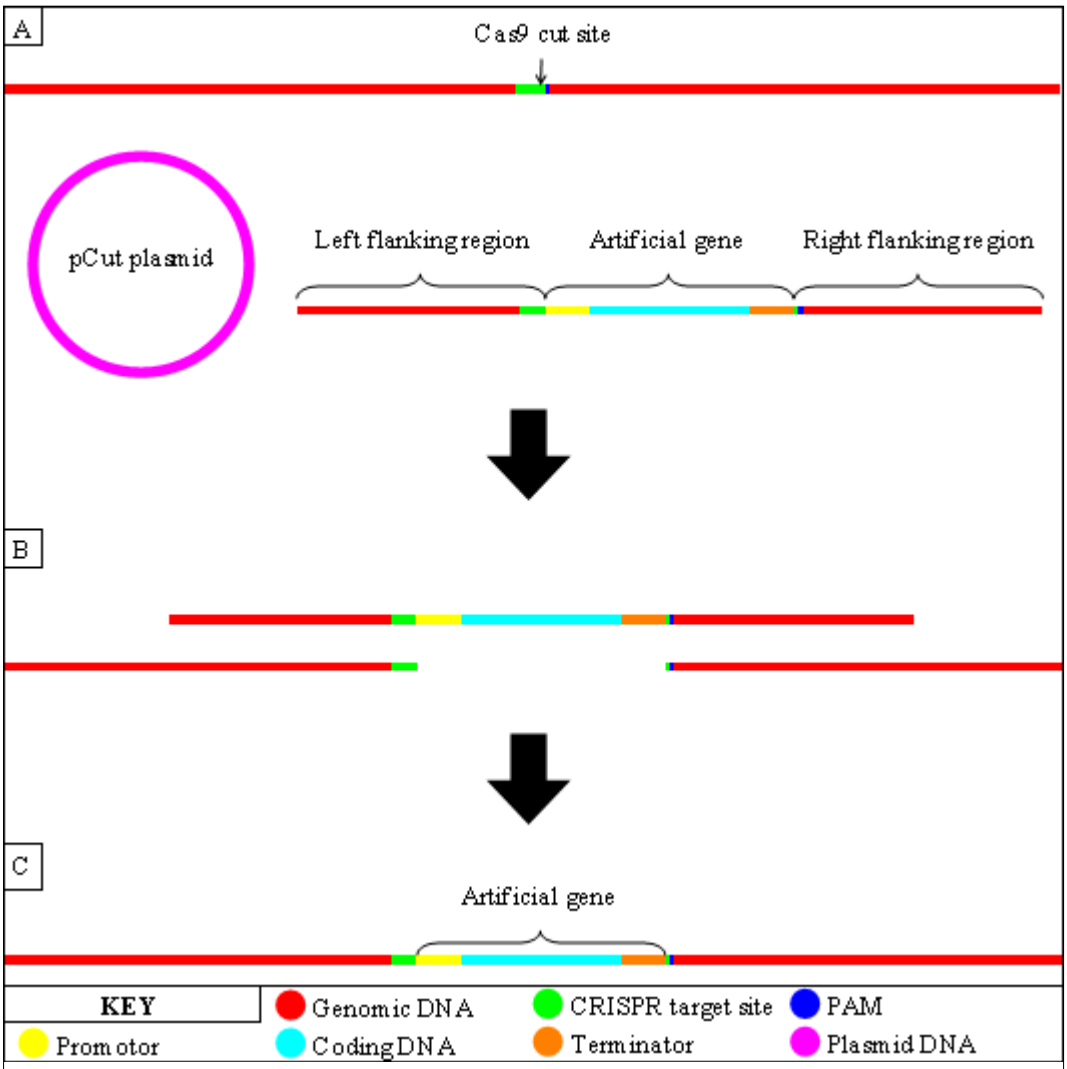


Figure 3: A schematic representation of the pCut transformation method developed by Apel *et al* (2017). A) The yeast is simultaneously transformed with the donor DNA and the pCut plasmid. B) Once Cas9 and its sgRNA are expressed, Cas9 cleaves the yeast DNA at one of the 23 predetermined loci. Cas9 cleaves between the 17th and 18th bp of the target site. The donor DNA has ~1000 bp homology with the yeast genomic DNA on each side. C) The yeast uses its mechanisms for HDR to insert the heterologous DNA into the yeast genome.

The pCUT method is a transformation method devised by Apel *et al* (2017) to seamlessly integrate donor DNA into one of 23 predetermined loci in the BY4742 strain of *S. cerevisiae*, a deletion strain of the haploid *S. cerevisiae* S288C created by Brachmann *et al* (1998). No cloning is required for this method; only donor DNA, *S. cerevisiae* BY4742, and one of the 27 pCUT plasmids. The pCUT plasmid contains the gene for Cas9 and its guide RNA for one of the 23 target sites. Because

S. cerevisiae BY4742 cannot produce leucine, lysine, histidine, or uracil, the plasmid also contains a selection marker gene – either *LEU2* or *URA3* – to be used in combination with either a leucine or a uracil dropout medium to select for yeast cells that have been successfully transformed with the plasmid, although colony PCR is still required to find transformants with the correct insert in the genomic DNA. Apel *et al* (2017) determined the transformation efficiency of each target locus using a GFP-encoding DNA insert, which allowed them to determine the integration efficiency of each locus through flow cytometry. When the yeast is simultaneously transformed with the plasmid and the linear donor DNA, the yeast will transcribe and translate Cas9 with its guide RNA, which will cleave the genomic DNA at the chosen target site. The yeast will then attempt to repair the cut by using (HDR) to insert the donor DNA into the target site (Apel *et al*, 2017).

Table 1: Integration efficiencies and target sequences of different loci as determined by Apel *et al* (2017) in descending order.

Locus	Integration efficiency (%)	Target sequence
1622b	100	TAAAGCCACCACATCGCAAA
911b	100	GTAATATTGTCTTGTTTCCC
106a	100	ATACGGTCAGGGTAGCGCCC
208a	99	GTCCGCTAAACAAAAGATCT
416d	99	TAGTGCACTTACCCCACGTT
1309a	99	CCTGTGGTGACTACGTATCC
308a	98	CACTTGTCAAACAGAATATA
HIS3b	98	AATATAGAGTGTACTAGAGG
720a	98	CAACAATTGTTACAATAGTA
511b	96	CAGTGTATGCCAGTCAGCCA
YPRCδ15c	87	AATCCGAACAACAGAGCATA
SAP155b	83	GGTTTTCATACTGGGGCCGC
1414a	82	GCGCCACAGTTTCAAGGGTC
1021b	72	CCTCTGTGTGGTGGTAATTG
SAP155c	51	ATGAAAGACAACTATAGGGC
1114a	46	CTTGTGAAACAAATAATTGG
1014a	42	TTATGTGCGTATTGCTTTCA
CAN1y	40	GATACGTTCTCTATGGAGGA
YOLCd1b	33	CTAGAATTTCCATTTTGCGT
RDS1a	16	ATTCAATACGAAATGTGTGC
607c	2	CTATTTTTGCTTTCTGCACA
805a	1	TTATTTGAATGATATTTAGT
1206a	1	CGAACATTTTTCCATGCGCT

The donor DNA must be carefully designed for this purpose. The heterologous DNA cassette must contain "flanking regions" on both sides. Each flanking region will have roughly 1000 bp of sequence homology with the yeast genomic DNA; the left flanking region will be homologous to the ~1000 bp upstream of the Cas9 cut site and the right flanking region will be homologous to the ~1000 bp downstream of the Cas9 cut site (Apel *et al*, 2017). If there is more than one fragment of DNA being transformed into the same locus, adjacent fragments must have at least 30-80 bp sequence homology between them (Apel *et al*, 2017). When the yeast is transformed with the donor DNA fragments and pCut plasmid and the yeast DNA is cleaved by Cas9, the yeast will use its own internal repair mechanisms to incorporate the pieces of heterologous DNA into its genomic DNA using HDR (Apel *et al*, 2017). Each of the 23 sites were initially found to have a different integration success rate, depicted in Table 1. The target site for each of the plasmids is also depicted in Table 1.

One of the challenges of integrating the genes for the entire metabolic pathway of Taxol into the genome of S. cerevisiae is the complexity of the Taxol biosynthesis pathway, which involves upwards of 20 enzymes. When developing the pCut transformation method, Apel *et al* (2017) identified 23 viable integration loci with unique 20 bp target sequences followed by a PAM with the sequence NGG. When Apel *et al* (2017) examined the integration efficiency of each locus after transformation with the pCut plasmids, three of the 23 loci yielded a 100% integration efficiency, three more yielded a 99% integration efficiency, and three more yielded a 98% integration efficiency. For each transformation, Apel *et al* (2017) transformed *S. cerevisiae* BY4742 with three DNA fragments: the DNA cassette containing the gene of interest, and two flanking regions with 1000 bp homology with the insertion locus and 30-60 bp homology with the DNA cassette; however, in the manual they wrote for their CASdesigner software, they specify that the integration efficiency decreases as the number of DNA fragments increase, and therefore the number of HDR reactions the cell needs to carry out, increase. For that reason, while the 100% integration efficiency loci are ideal, the loci with 99% and 98% integration efficiencies could be improved by transforming the cell with only a single fragment of donor DNA.

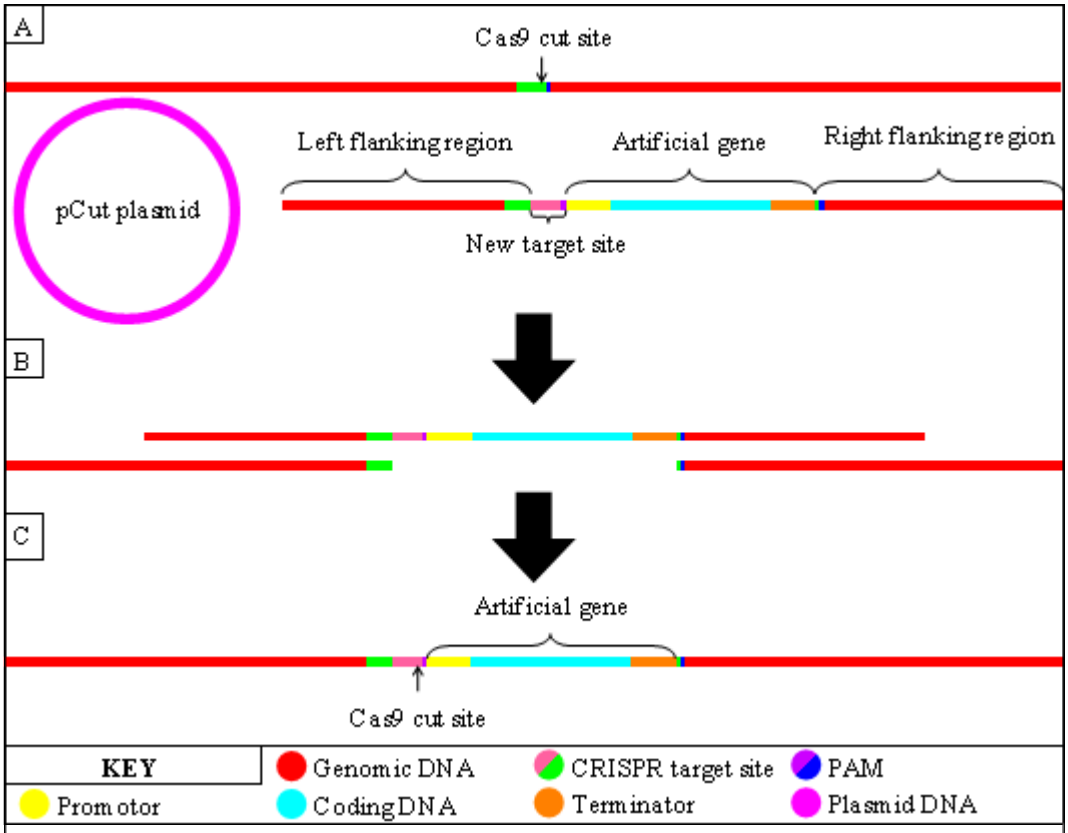


Figure 4: Recycling of pCut sequences at target loci with optimum integration efficiencies. A) The 20 bp DNA sequence of a pCut target with low integration efficiency is equipped with a PAM and added to the donor DNA sequence that is transformed into the yeast along with the appropriate pCut plasmid. B) The genomic DNA is cleaved by Cas9, and the new DNA is integrated through HDR. C) The yeast genome now contains a heterologous gene and a new target site that can be utilised in subsequent transformations with a different pCut plasmid.

The respective integration efficiencies for the three most efficient loci characterised by Apel *et al* (2017) were all 100%, and the integration efficiencies for the least efficient loci were 1%, 1%, and 2%, respectively. The latter three loci are unlikely to be of use for most genomic integration studies due to the relative difficulty in successfully integrating heterologous DNA at these loci compared to the other loci identified by Apel *et al* (2017). However, they each contain a unique sequence of 20 nucleotides that is not found elsewhere in the *S. cerevisiae* BY4742 genome and therefore represents a safe target for Cas9. Theoretically, if these 20-nucleotide sequences were each to receive a PAM, they could be added to a donor DNA sequence and inserted into an optimal integration locus. Once integrated into the yeast genome, this 23 bp sequence serves as a target for a pCut plasmid made to target one of the three loci with a low integration efficiency (607c, 805a, and

1206a). As shown in Figure 3, the original target sequence is lost after a successful integration and cannot be targeted again for further integrations. Integrating new target sequences with each integration of donor DNA allows further integrations at loci with optimum integration efficiencies. It should be noted that the recycling of the target sequences should only be performed if multiple integrations of a gene into the yeast genome are not a problem, because, while the probability of integration in the know-efficiency locus is low, it is not impossible. For this study, multiple integrations do not represent a problem. The basic principle is shown in Figure 4.

S. cerevisiae as host organism

S. cerevisiae is a yeast that propagates through budding, leaving a daughter cell slightly smaller than the mother cell (Herskowitz, 1988). The history of human use of *S. cerevisiae* is so vast that it was considered important enough to become one of the first microorganisms whose genome was fully sequenced (Baghban *et al*, 2019; Lahue *et al*, 2020). Due to its ancient purpose in baking, brewing, and winemaking, and how well it has been studied in an academic context, it is an excellent model organism for use in experimental and industrial microbiology (Foland *et al*, 2005; Karathia *et al*, 2011).

While *S. cerevisiae* may be used to study the tubulin-stabilising effects and eventual apoptosis caused by taxol, the yeast is not normally sensitive to taxol (Foland *et al*, 2005). To induce sensitivity to taxol, Foland *et al* (2005) used an *S. cerevisiae* strain with a mutated β-tubulin gene and a diminished number of ABC transporter genes to allow taxol to halt cell division and induce apoptosis. For the purpose of this study, this is a good indication that taxol intermediates will not have a cytotoxic effect on *S. cerevisiae* BY4742, which lacks these mutations, and will not only fail to effectively bind to the microtubules, but will be safely transported out of the cell by the wild-type ABC transporters (Foland *et al*, 2005). The lack of the oxetane ring structure in the taxol intermediates this study further mitigates the risk of cytotoxicity to the yeast, even if the importance of the oxetane ring structure varies between studies (Samaranayake *et al*, 1991; Wang *et al*, 2000; Snyder *et al*, 2001).

DNA assembly methods

Gibson assembly

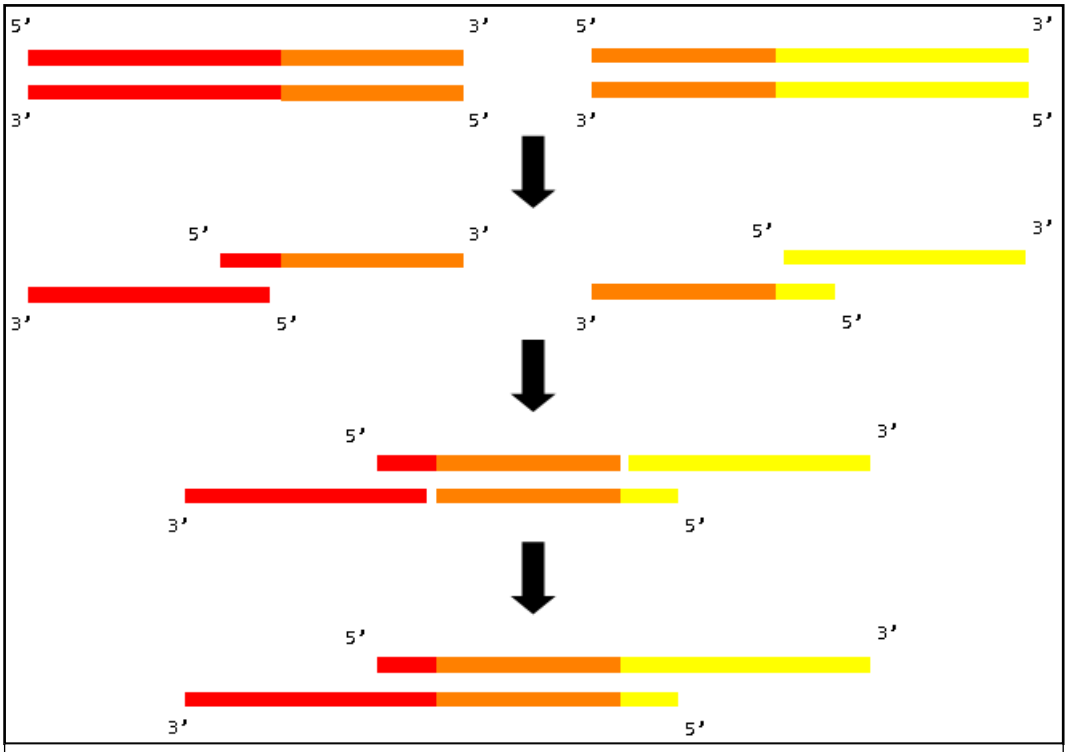


Figure 5: The principle of Gibson assembly. 5' ends of DNA fragments are degraded by T5 exonuclease, leaving long 3' sticky ends. Homologous sticky ends anneal to each other. The gaps are then filled in 5' to 3' by Phusion DNA polymerase and the gaps in the phosphodiester backbone of the DNA are covalently closed by Taq DNA ligase. The 3' sticky ends of the assembled fragment are not repaired in a linear assembly, but will disappear in a circular assembly, where each sticky end is matched with another that can act as a primer for 5' to 3' amplification.

The complete synthesis of artificial genes, which are thousands of bp long, is very expensive compared to the synthesis of shorter nucleotide sequences. It is therefore desirable to design shorter nucleotide sequences that can be scarlessly assembled in the laboratory into longer full gene sequences.

Gibson assembly is a one-pot isothermal DNA assembly method designed by Gibson *et al* (2009) to assemble DNA molecules with homologous ends, with 15 to 40 bp of homology as the standard. The method works by the following principles: The 5' ends of both strands of double stranded DNA (dsDNA) are degraded using a T5-exonuclease until past the homologous regions between DNA fragments to form complementary 3' sticky ends (Gibson *et al*, 2009). The 3' sticky ends anneal to each other, the gaps between the annealed single stranded DNA (ssDNA) strands

and gaps left by the T5-exonuclease are filled in by Phusion DNA polymerase, and the gaps in the new double-stranded DNA backbone are closed by Taq DNA Ligase. Phusion polymerase is used to insert nucleotides into the gaps left by the T5 exonuclease and was chosen by Gibson *et al* (2009) for its proofreading ability. Phusion DNA polymerase may be replaced by Taq DNA polymerase, which carries out the same function in the reaction mix; however, Phusion remains preferable for its far superior accuracy (Gibson *et al*, 2009). This method allows for seamless DNA assembly into a plasmid vector or seamless assembly of a linear DNA fragment. It was chosen as the DNA assembly method of this study to ensure that the artificial genes, which were divided into many separate DNA fragments, were not mutated during the assembly step. The principle of Gibson assembly is shown in Figure 5.

PCR splicing by overlap extension (SOE-PCR)

PCR splicing by overlap extension, also known as SOE-PCR or overlap extension PCR, is a DNA assembly technique that involves at least two DNA fragments using each other as primers for a DNA polymerase in order to create a single fused fragment. The ends of the fragments must have a region of sequence homology with each other to enable priming. The basic principle is shown in Figure 6.

There are various methods of SOE-PCR available which have been optimised to use either ssDNA or dsDNA, and for different DNA polymerases. SOE-PCR may be used to assemble either linear or circular DNA, and the methods may be adjusted accordingly. Stemmer *et al* (1995) were the first to describe the deliberate assembly of non-random oligonucleotides with 20 bp overlapping regions using PCR. Their method involved first assembling the genes through 55 PCR cycles, and then using a small portion of the reaction mix as the DNA template for a new reaction mix. Warrens *et al* (1997) introduced an asymmetric PCR step into SOE-PCR to amplify only the DNA strands that could feasibly function as primers for each other, thereby largely eliminating the fragment that cannot be amplified and preventing the two strands of each template fragment from simply reannealing to each other. The usefulness of this step is highly dependent on the length of the DNA fragments being spliced, the initial DNA concentration for non-exponential amplification of

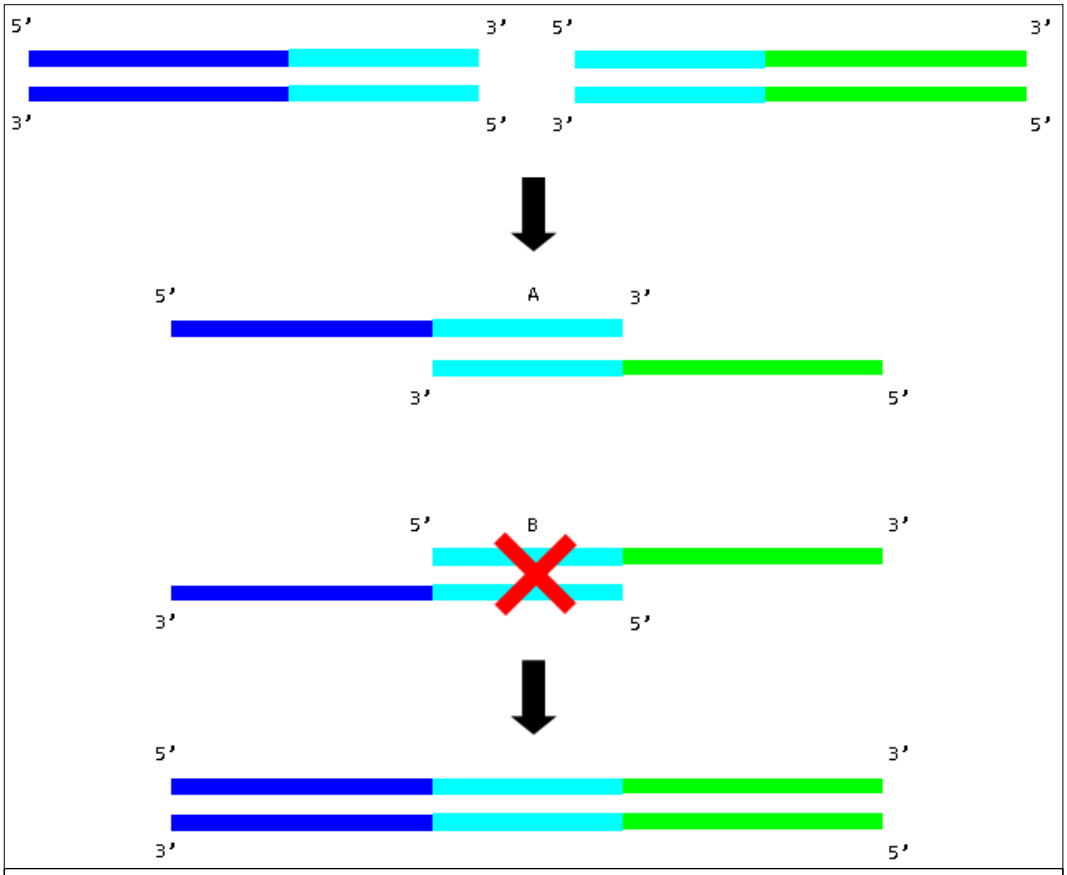


Figure 6: DNA fragments priming each other during SOE-PCR. Fragment A can be amplified into the assembled product. Fragment B cannot be amplified, because there is no way to proceed with amplification in the 5' to 3' direction. Fragment A is amplified into the assembled product.

ssDNA, and which DNA polymerase is used, as certain polymerases are unsuited to synthesising long strands of ssDNA (Veneziano *et al*, 2018). The method developed by Warrens *et al* (1997) allowed them to fuse two challenging pieces of DNA when conventional overlap extension PCR failed despite numerous changes made to the reaction mix and thermal cycling parameters.

More recently, the length, number of fragments assembled, and fidelity of overlap extension PCR has been improved. Kadkhodaei *et al* (2016) employed a novel touchdown cycling method using the Kapa HiFi DNA polymerase to assemble both linear and circular fragments. Their final annealing temperature for all fragments was as high as 70°C; however, instead of using normal touchdown cycling, the annealing temperature slowly reduced from 80°C to 70°C at a steady rate over 100 seconds every cycle before the normal extension phase at 72°C. Hilgarth and Lanigan (2020) found that SOE-PCR could be optimised by utilising standard

touchdown PCR to amplify the template fragments, the overlap PCR, and the amplification of the fused product. They also found that reducing the concentration of primers in the reaction mix also improved the specificity of the reaction (Hilgarth and Lanigan, 2020).

Comparisons between DNA polymerases for SOE-PCR

Kadkhodaei *et al* (2016) found great success assembling very large fragments (up to 20 kbp) out of numerous (up to eight) longer fragments (up to ~2200 bp) using the Kapa HiFi DNA polymerase. Dolgova and Stukolova (2017) compared the ability of Taq, *Pfu*, and Phusion to assemble several oligonucleotides during a single PCR reaction and found that Phusion by far outcompeted both *Pfu* and Taq DNA polymerases, with an error rate of only a third of that of *Pfu*; however, *Pfu* still outcompeted Taq, which failed to produce a band of DNA of the correct size. While Dolgova and Stukolova (2017) provided valuable insight into the mutation rate and efficiency of all three enzymes, and assembled a large number of fragments in a single step, it is limited by the very small size of the template fragments (shorter than 63 bp) and the short length of the final assembled product (581 bp). Hilgarth and Lanigan (2020) compared Phusion and Q5® DNA polymerases and found that Q5® was more consistent than Phusion. This study was limited to a simpler two-fragment assembly, and the largest assembled fragment was <2400 bp.

The choice of DNA polymerase for SOE-PCR also depends on the protocol of SOE-PCR. When employing an asymmetric amplification step like the one introduced by Warrens *et al* (1997) to improve a difficult assembly, certain polymerases are not well-suited to the task. Veneziano *et al* (2018) compared the ability of several commercial DNA polymerases to synthesise long strands of ssDNA. They found that Taq-based enzymes, especially those engineered for higher fidelity, achieved longer ssDNA amplicons than Phusion, and posited that the enhanced 3’ to 5’ exonuclease activity of Phusion might be to blame for this.

Chapter 3:
Materials and Methods

MATERIALS AND METHODS

DNA Design

Choosing the integration locus

The following integration sites were chosen for this study and future studies stemming from it: 1622b, 911b, and 106a at 100% integration efficiency; 208a, 416d, and 1309a at 99% integration efficiency; 308a, HIS3b, and 720a at 98% integration efficiency. For reasons involving the design of the donor DNA explained below, only the integration site 106a remained relevant; however, the other integration sites will be critical for continuing this research. Left and right flanking regions, each with ~1000 bp homology with the yeast chromosome at integration site 106a, are shown in Addendum 1. The flanking regions were determined by downloading the full sequence of each of the 16 *S. cerevisiae* BY4742 chromosomes and searching for the Cas9 target sequence in the appropriate chromosome. Because Cas9 cleaves the DNA between the 17th and 18th bp of the target sequence, the 1000 bp upwards of the cut site, including the first 17 bp of the target sequence, was chosen as the left flanking region. The 1000 bp downwards of the cut site, including the last three bp of the target sequence and the PAM, were chosen as the right flanking region.

Design of the artificial genes

The three proteins chosen for heterologous production are taxadiene synthase (GenBank: ABC25488.1), taxadiene 5α-hydroxylase (UniProtKB/Swiss-Prot: Q6WG30.2) and an NADPH-cytochrome P450 reductase (GenBank: AAT76449.1). The pathway was obtained from the KEGG database (Kanehisa, & Goto, 2000; Kanehisa, 2019; Kanehisa *et al*, 2021). The amino acid sequence of each protein was used to create an artificial gene for each of the enzymes. The protein FASTA sequences of the three genes involved in this study were obtained from the NCBI protein database. The sequences of all three proteins were taken from the *Taxus cuspidata* records. The online IDT Codon Optimization Tool and NovoPro Codon Optimization Tool were both used to generate codon-optimised DNA encoding the target genes based on their amino acid sequences. A blastx

search was conducted comparing the codon-optimised DNA to the *T. cuspidata* protein records. For all three DNA sequences, the DNA obtained from the NovoPro Codon Optimization Tool yielded 100% Query Cover and 100% Percent Identity, whereas the DNA obtained from the IDT Codon Optimization Tool yielded only 99% Query Cover for two of the DNA sequences. For this reason, all DNA sequences used were the ones obtained from the NovoPro Codon Optimization Tool. Similar to the experiments done by Apel *et al* (2017), an MBP tag was added to the C-terminal of each of the proteins to improve solubility. However, the MBP tag sequence used by Apel *et al* (2017) was not used; the protein FASTA sequence of MBP (GenBank: EFK3242586.1) was used to generate codon-optimised DNA. A Gly6 linker, which was also codon-optimised, was used to fuse the MBP tag to the C-terminal of the target protein sequence. Neither codon optimisation tool used adds a stop codon if the input is an amino acid sequence. The codon UAA was used for all sequences.

The promoter sequence was taken from The *PRM9* terminator sequence used for all three artificial genes was taken from the supplementary materials of Curran *et al* (2013). The sequence for the *ICL1* promoter used for all three artificial genes was identified comparing the PCR primers given in the supplementary material of Maury *et al* (2018) to the sequence of *S. cerevisiae* BY4742 chromosome V.

Design of gBlocks for Gibson assembly

To reduce the cost of the order of DNA fragments, an attempt was made to minimise the number of base pairs in the order. This was best achieved by combining the DNA sequences for all three genes into a single sequence and dividing it into gBlocks for Gibson assembly. Each of the three heterologous gene sequences was placed end-to-end, interspersed with the new CRISPR target sites and their PAM sequences (TGG). The order of the assembled DNA sequence was as follows: taxadiene synthase gene, locus 607c CRISPR target sequence and PAM, taxadiene 5α-hydroxylase gene, locus 805a CRISPR target sequence and PAM, locus 1206a CRISPR target sequence and PAM, NADPH-cytochrome P450 reductase gene.

The pCUT method was designed for seamless integration; therefore, from this point forward, no regard was given to the distinction between these three gene sequences.

The flanking regions were too complex and therefore expensive to synthesise, and were therefore excluded from the gBlocks design. The final 40 bp of the left flanking region (Fragment A) and the first 40 bp of the right flanking region (Fragment B) were added to the beginning and the end of the sequence, respectively. This was done to give the flanking regions some overlap with the heterologous DNA to be attached later by Gibson assembly.

The DNA was divided into seven fragments (excluding the flanking regions) with the intention to assemble these fragments into three fragments of equal length for co-transformation with the pCut plasmid into the yeast. Fragments R1 and R2 were each 1875 bp long, and along with Fragment A would make up Insert 1. Fragments R3, R4, and R5 were each 1580 bp long, and would make up Insert 2. Fragments R6 and R7 were each 1875 bp long, and along with Fragment B would make up Insert 3. Each insert would have 80 bp sequence homology with the adjacent insert(s). The length of sequence homology between each fragment was as follows: 40 bp between the left flanking region and R1, and between the R7 and the right flanking region; 30 bp between R1 and R2, between R3 and R4, between R4 and R5, and between R6 and R7; 80 bp between R2 and R3, and R5 and R6. The full DNA sequence is given in Addendum 1.

DNA preparation

DNA synthesised to order

Table 2: Molecular weight of DNA fragments.

Fragment	g/mol
A	586303.03
R1	1158442.17
R2	1158431.39
R3	976168.04
R4	976180.78
R5	976176.86
R6	1158506.85
R7	1158461.77
B	568382.46

The seven DNA fragments were ordered from Gene Universal. Each 500 ng sample of DNA was suspended in 10 µl of nuclease-free water to a concentration of 50 ng/µl. The mols of each fragment of donor DNA were calculated, as shown in Table 2. The resuspended DNA was stored at -20°C.

DNA extraction from yeast

The yeast DNA extraction protocol of Lõoke *et al* (2011) was modified to accommodate for the limits of the Eppendorf Minispin centrifuge. A single colony was picked from a *S. cerevisiae* BY4742 streak plate and suspended in 100 µl of a 1% SDS 0.2 M lithium acetate solution inside a 1.5 ml Eppendorf tube by thorough vortexing at the highest speed setting. The tube was then placed inside a heat block at 70°C for 5 minutes. Afterwards, 300 µl of >99.7% ethanol was added to the tube to precipitate the DNA, and the tube was thoroughly vortexed at the highest setting again. Due to the limitations of the available Eppendorf Minispin, the protocol was modified from the standard set by Lõoke *et al* (2011). The tube was centrifuged at 13 000 rpm, which amounts to 11 337 g, for 6 minutes to force the precipitated DNA into a pellet. Due to concerns about the speed and time modifications for the centrifugation step, special attention was paid to the wash step. The supernatant was discarded and 100 µl was added to the tube. The tube was vortexed again for over a minute at the highest speed setting until as much of the pellet as possible was suspended. The pellet was partially insoluble. The tube was again centrifuged at 13 000 rpm for 6 minutes, and the supernatant was discarded again. The pellet was dissolved in 100 µl of filtered autoclaved distilled water (dH_2O). The tube was centrifuged again at 13 000 rpm for 60 seconds to spin down cell debris.

While the original protocol calls for only 1 µl of the supernatant to be used for PCR, this proved ineffective on multiple occasions, and success was finally found using just under 1 µg of DNA, the maximum amount of DNA allowed by the OneTaq™ protocol. The DNA concentration was determined using a NanoDrop™ spectrophotometer. The DNA was stored at -20°C.

Primer design and PCR amplification of flanking regions

Primers were designed using the NCBI primer blast function to amplify a region as close as possible to 1000 bp on each side of the cut site, with the reverse primer of

the left flanking region and the forward primer of the right flanking region as close to the cut site as possible. The final product expected from the left flanking region was 949 bp and the final product expected from the right flanking region was 920 bp. Each primer sample was suspended in filtered autoclaved dH_2O to a concentration of 100 mM. From each of these samples, a smaller stock with a concentration of 10 mM was made by combining 10 µl of 100 mM stock to 90 µl of filtered autoclaved dH_2O. All primers were stored at -20°C.

Following the protocol of the NEB OneTaq™ 2 X Master Mix with Standard Buffer, 25 µl of Master Mix (MM), 1 µl of forward primer, 1 µl of reverse primer, the maximum allowed amount of DNA (concentration determined by NanoDrop™), and water up to 50 µl was added to a PCR tube to a final volume of 50 µl. The temperatures and times of the thermal cycler are shown in Table A2.3 in Addendum 2. The annealing temperature was chosen using the NEB T_m calculator; however, the annealing temperature given for the right flanking region resulted in no PCR product, therefore an annealing temperature based on the melting temperature of the primers given by the NCBI Primer Blast tool, 52°C, was chosen.

Once PCR was completed, the DNA was purified using gel electrophoresis on a 1% agarose gel and visualised on a UV transilluminator after staining for 30 minutes in 1µg/mL EtBr. Electrophoresis was carried out for one hour at 100 V. The DNA was extracted from the gel using the Zymoclean™ Gel DNA Recovery Kit. The concentration of the recovered DNA was determined using the NanoDrop™. The remaining recovered DNA was then used to carry out another PCR reaction and purification to improve the yield of the target DNA.

Gibson assembly

Gibson assembly of linear fragments

Gibson assembly was attempted using the fragments that were synthesised to order. Each reaction used 100 ng of each fragment. The planned inserts were as follows: Insert 1 would be assembled from fragments A, R1 and R2; Insert 2 would be assembled from fragments R3, R4, and R5; Insert 3 would be assembled from R6, R7, and B. The isothermal buffer mix is shown in Table A2.8 in Addendum 2. The

assembly master mix is shown in Table A2.9 in Addendum 2. The three assemblies were carried out at 50°C for 60 minutes. Gel electrophoresis was performed on a 1% agarose gel at 100 V for one hour and the gel was stained in 1 µg/mL EtBr to visualise the result. In the first attempt, all three assemblies were attempted. In the second attempt, only the assembly of Insert 2 was attempted in an effort to use as little DNA as possible while troubleshooting. Insert 2 was chosen at first because it did not contain either of the PCR-amplified and gel-purified flanking regions, which could introduce unforeseen complications with their relatively lower purity, but assembly of the other inserts was also attempted again to conserve R3, R4, and R5. Assembly of Insert 1 and assembly of Insert 3 were both attempted again to make sure an error had not been made in the protocol. Finally, assembly of Insert 2 was attempted again. After these five attempts this method was abandoned due to the high value of the DNA and fears about it running out.

Primer design for PCR amplification of synthetic gene fragments and pGEM®-T Easy vector

Primers were designed for all nine gBlocks and for three variations of the pGEM®-T Easy vector using the Primer Blast function of the NCBI database. Each primer was analysed using the Multiple Primer Analyzer tool from Thermo Fisher Scientific to check for potential primer dimers.

When trying to create long primers to add homologous regions to the linearised pGEM®-T Easy vector, it became necessary to create them manually. The ends of pGEM®-T Easy differ vastly in melting temperature and did not lend themselves well to primer design by the NCBI Primer Blast algorithm, even with extensive alterations to the Blast parameters. For this reason, the first and last 20 bp of the linearised vector fragment were used to start the design of the vector primers. For the overhang side of the primer, the primers generated using NCBI were used, as they had already been calculated for a low risk of primer dimers. Primer construction is shown in full in Addendum 2. In short, primers A1 and RV2 were used to construct Vector 1 long primers, primers FW3 and RV5 were used to construct Vector 2 long primers, and primers FW6 and B2 were used to construct Vector 3 long primers. As shown in Figure 7, primers could be designed to place

the insert sequence on either the plus or minus strand of the plasmid. For each insert, primer pairs for both conformations were evaluated. For every insert, placing the sequence on the minus strand resulted in better primers with a lower risk of dimerization.

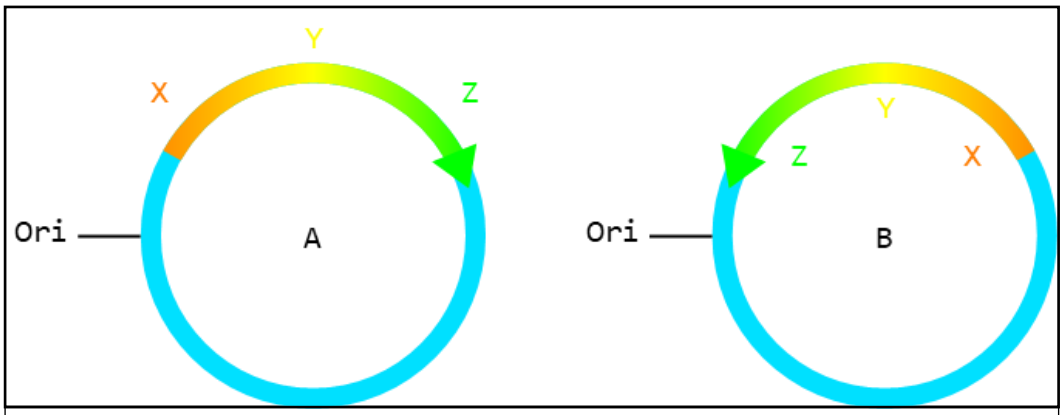


Figure 7: The two possible clones that could be designed when assembling theoretical DNA fragment XYZ into a vector. Vector A shows the fragment on the plus strand of the vector and vector B shows the fragment on the minus strand of the vector. Images not drawn to scale for pGEM®-T Easy vector or any of the inserts used in this study. For each of the three initial vectors of this study, as well as the fourth unplanned vector, the DNA sequence of interest was present on the minus strand.

Complications in the design of certain primers were of note. To form a suitable primer pair, the primers of fragment R4 needed to extend the fragment by 5 bp. The primers for fragment R5 had unavoidable primer dimers, but the melting temperature of these dimers were well below the annealing temperature of R5's PCR reaction. The extremely TA-rich fragment R7 required an unusually long 29 bp reverse primer. The reverse primer of vector fragment V3 only had 16 bp homology with pGem(R)-T Easy to maintain a melting temperature within 5°C of the forward primer. Each primer was resuspended in filtered, autoclaved dH_2O to a concentration of 100 mM. Working primer solutions with a concentration of 10 mM was made from each of the stock primer solutions by mixing 1 µl of stock with 9 µl of nuclease-free water.

Every primer used to amplify one of the fragments, including the flanking regions, is shown in Table 3. Important details of every primer, taken from the NCBI Primer Blast tool, are shown in Table A2.5 in Addendum 2, with the addition of the melting

temperatures given by the NEB Tm calculator, which are specific to the NEB OneTaq® PCR Master Mix. These temperatures can vary from the melting temperature given by Primer Blast and were therefore crucial in the final decision on each primer pair.

Table 3: All primers used in this study

Fragment	Primer	Sequence
A	Forward	TTCCTAAGCCTCCCTCACCA
	Reverse	GCTACCCTGACCGTATCACA
R1	Forward	CATGACCCCGGATCGTCG
	Reverse	CGTCTCTCAGGTACGGTTCAG
R2	Forward	GTAAGTTCGCTGAACCGTACCTG
	Reverse	TAATTAAGTCAACCAAGAACGTCAGGC
R3	Forward	CGGCAAATACGACATAAAAGACG
	Reverse	TCAAGGAAGTTTTGAAGACCAAACC
R4	Forward	AATTTGGTTTGGTCTTCAAAACTTCC
	Reverse	GCTTTGTCTGGCGTAATCTCTG
R5	Forward	AGTGGTTTGCTAGCAGAGATTACG
	Reverse	TTGCTTGCATTTTTCGTTGACTT
R6	Forward	CTTTTCACACCCAAATACCTAACA
	Reverse	ACTGCTGGACCTAATTTTTCTCC
R7	Forward	AGAAGCGGGAGAAAAATTAGG
	Reverse	TCTTCCTTATTATTTTTTTAGTTCTGAAG
B	Forward	CCCTGGTCAAACTTCAGAACTAA
	Reverse	CCAAGGTTTCTGGGTTGTTGG
V1	Forward	TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGG
	Reverse	GCCTGACGTTCTTGGTTGACTTAATTAAATCGAATTCCCGCGGCCGC
V2	Forward	CGTCTTTTATGTCGTATTTGCCGAATCACTAGTGAATTCGCGG
	Reverse	AAGTCAACGAAAAATGCAAGCAAAATCGAATTCCCGCGGCCGC
V3	Forward	TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGG
	Reverse	AAGTCAACGAAAAATGCAAGCAAAATCGAATTCCCGCGGCCGC

PCR amplification of fragments and gel purification

All fragments and vectors were PCR amplified from 0.5 µl template DNA (50 ng of fragments R1-R7 and V1-V3; 34 ng of fragment A; 27 ng of fragment B) using NEB OneTaq DNA polymerase in NEB OneTaq Standard Reaction Buffer in an Applied Biosystems® 2720 thermal cycler. Each PCR was performed in triplicate to maximise the DNA concentration after gel purification. The components of the PCR mix are shown in Table A2.1 in Addendum 2. The thermal cycler times and temperatures are shown in Table A2.3 in Addendum 2. Electrophoresis was performed on a 1% agarose gel using 5 µl of each of the PCR mixes to determine

whether PCR was successful and to ensure that no off-target amplification occurred. Off-target amplification meant that each fragment needed to be purified.

Gel electrophoresis was performed at 100 V for one hour using a 1% TAE-agarose gel and a TAE running buffer. The DNA was extracted from the gel using a "freeze and squeeze" method developed by Rahul Patharkar (2019). For this method, a 1 mL filter-tip pipette was placed inside a 1.5 ml Eppendorf tube. It was necessary to cut off some of the bottom part of the pipette tip to allow it to fit, which is normal for this method. The DNA was cut from the gel and the slice was placed inside the top part of the pipette tip, on top of the filter. The tube was then placed at -80°C for 5 minutes until the gel was completely frozen. The Eppendorf tube was subsequently centrifuged at room temperature at a minimum of 10 000 RCF for 5 minutes. The protocol was adapted to repeat the freezing and centrifuging step to increase the DNA yield. The disadvantages of this technique include the fact that ethidium bromide ends up inside the suspension, which can inhibit PCR (Nath *et al*, 2000), and that the DNA is usually very diluted, often to an extent that makes it impossible to perform Gibson assembly. This was the case for all 12 fragments.

An attempt was made to concentrate all of the fragments using ethanol precipitation with NaCl. This method had mixed rates of success. The concentrations of relatively large vector fragments (>3 kbp) were easily increased to well over 50 ng/µL. The smaller fragments did not concentrate as well, with only three fragments reaching a concentration of greater than 50 ng/µl. No fragment was too diluted to use for Gibson assembly and therefore another round of PCR was not performed. Each fragment or vector was stored at -20°C until Gibson assembly could be performed.

When the concentrated fragments started to run out after subsequent experiments, new PCR was performed using high-fidelity NEB Phusion polymerase. This choice of enzyme was made in the hope of retrieving a higher quality product. Due to technical issues with the previously mentioned thermal cycler, Bio-Rad™ T100 thermal cyclers were used for all of these reactions. Due to the low mutation rate and high product yield of Phusion, only a single tube of PCR reaction was assembled for each fragment. The reaction mixes and thermal cycler temperature

cycles are shown in Tables A2.2 and A2.4 in Addendum 2. All fragments were retrieved using gel electrophoresis with the same parameters as before, followed by the same freeze and squeeze method, but using 200 µL filter tips instead of 1 mL filter tips. The DNA was concentrated using the same ethanol precipitation with NaCl. Although the yields for this protocol were not as high as could be hoped, they were far more consistent than the original protocol, with no fragment yield under 20 ng/µl. The supernatant from each fragment tube was saved in marked Eppendorf tubes in case enough DNA remained in some to be retrieved through further freezing and centrifuging steps.

Gibson assembly of plasmid vectors

The fragments were assembled in groups of 3 (A, 1, and 2; 3, 4, and 5; 6, 7, and B) into their respective PCR-amplified pGEM®-T Easy vectors using Gibson assembly. The reaction was incubated at 50°C for 60 minutes. When this proved unsuccessful or had a very poor product yield, the reaction mix was assembled again and incubated at 50°C for four hours. The yield was noticeably increased. When viewing the gel on the Bio-Rad ChemiDoc™ using its Image Lab™ software, the one-hour assembly reaction yielded a band too faint to be registered by the software, although it was still faintly visible to the human eye. The four-hour Gibson assembly reaction yielded many bands that, although faint, were easily registered by the software. For that reason, subsequent Gibson assembly reactions were incubated for four hours.

After finishing the protocol and confirming the presence of the assembled DNA through gel electrophoresis on a 1% agarose gel, and the product was purified using the abovementioned DNA purification method. The parameters for electrophoresis were the same as for the purification of the fragments, except for the run time, which was increased to one hour and 30 minutes to achieve a clearer separation between larger fragments. Retrieved DNA was stored at -20°C until it could be used for downstream applications. PCR amplification of Insert 2 was attempted, but only a small (~1250 bp) piece of DNA was retrieved. It was discovered that an error had been made in the design of the assembled vectors, and that the forward primer for fragment R3 existed within fragment R5, which led to a PCR reaction that strongly

favoured the amplification of a small section of fragment R5 over the entire assembled Insert 2.

Due to the inability to PCR amplify Insert 2 from the plasmid, the new plan was for the inserts for Vectors 1 and 2 to be combined and assembled into a combination of Vector 1 and 2, henceforth named V1.5. V1.5 was PCR-amplified using the forward primer for V1 and the reverse primer for V2. An unsuccessful attempt was made to assemble fragments A and R1-R5 into V1.5. To reduce the number of fragments involved in the Gibson assembly reaction, a method called nested Gibson assembly, where DNA fragments are assembled in smaller Gibson assembly reactions and these assembled fragments are subsequently assembled into one long fragment, was attempted. In two tubes, Vector 1 was assembled as normal using Gibson assembly, isolated using gel purification and concentrated using the modified ethanol precipitation protocol. PCR was performed to amplify the fragment of interest. The reaction was incubated at 50°C for four hours. One reaction mix was then run on a 1% agarose gel at 100 V for 90 minutes and stained in 1 µg/mL EtBr for 30 minutes. Every assembled band was cut out of the gel and purified using the freeze and squeeze method. For fear of losing the very small yield of assembled plasmid DNA, ethanol precipitation was not performed.

Transformation of competent E. coli with pGEM®-T Easy plasmid

The transformation of the competent *E. coli* was performed as per the instructions on the pGEM®-T Easy user manual. Competent cells were removed from the -80°C freezer and allowed to thaw slowly in an ice bath. The competent cells were very gently mixed with 2 µL of DNA from the Gibson assembly reaction and incubated in an ice bath for 20 minutes. The cells were then heat-shocked for 50 seconds at 42°C and placed back on ice for two minutes before being mixed in with 950 ml of SOC medium and incubated at 37°C for 90 minutes. LB plates with 100 µg/ml ampicillin (LB-Amp), were warmed to 37°C, and 100 µl of cells were spread onto each plate and incubated overnight at 37°C. After initial failure, another plate of LB agar without ampicillin was made to test whether the cells were viable after the transformation protocol. Further attempts at transformation were made with increasing volumes of unpurified Gibson assembly added each time. For every

transformation attempt, 100 µl of cells were plated onto an LB agar plate without ampicillin to ensure that the transformation protocol did not kill them.

This method did not yield any successes after eleven attempts for reasons that were unclear and was shelved in favour of other methods.

PCR amplification of DNA fragments from assembled pGEM®-T Easy plasmid

The gel-purified fragments from each of the two Gibson assembly reactions were used as the template DNA for three respective PCR reactions to amplify each of the three donor fragments. Only 4 µL of isolated DNA was used to prevent PCR inhibition by EtBr. After PCR, gel electrophoresis of each product on a 1% agarose gel was performed to determine whether the correct product had been amplified. PCR was also attempted to isolate smaller fragments to see if parts of the assembly had happened successfully after initial failure to amplify a fully-assembled product. This was done for assemblies of Insert 1 and Insert 3.

After initial failure of PCR on the assembled product, a new method was devised by which fragments might be assembled and PCR amplified. At least four of the important fragments (no vector fragments) were assembled for four hours. Gel electrophoresis was performed at 100V for 90 minutes and the fragment of the correct size was identified and purified from the gel. The assembled fragment served as the base for a new PCR, where primers for the two middle fragments were used in an attempt to isolate a 2-fragment assembled product. After a few unsuccessful attempts, this method was abandoned due to a shortage of reagents and concern over the dwindling amount of the original DNA suspension for each fragment.

SOE-PCR

Initial 3-fragment attempt

The initial attempt at SOE-PCR used 50.0 ng of R1, 50.0 ng of R2, and 25.3 ng of A in a 50 µL reaction. Phusion DNA polymerase and its HF buffer were used for this reaction. PCR was carried out for 30 cycles for the overlap step, and then primers were added to the mix at the normal concentration. PCR was continued for 30 more cycles.

Asymmetric PCR for SOE-PCR

Two attempts were made to use asymmetric amplification of two fragments to achieve overlap extension. The first attempt utilised 25 cycles of asymmetric PCR, followed by combining the two tubes and 25 cycles of allowing the single strands to prime each other and amplify into the combined sequence. Fragments R1 and R2 were chosen at 100 ng each for their longer length and their relative lack of homology to each other apart from the 30 bp overlapping region. The annealing temperatures were determined by NEB Tm calculator. After the tubes were combined, the reaction carried on for another 25 cycles.

After the initial attempt failed, a one-pot method using the same principles was attempted in order to counteract the potential ssDNA-degrading effects of the 3' to 5' exonuclease activity of Phusion. 50 ng of R1 and 50 ng of R2 were added to the PCR reaction mix, along with the forward primer for R1 and the reverse primer for R2. This attempt led to much the same result as the first.

A third attempt was made to splice two pieces of DNA, although a shortage of fragments R1 and R2 led to fragments R6 and R7, both of the same length as R1 and R2, to be used as the new test DNA pieces. To minimise non-specific amplification and encourage the fragments not to simply reanneal to themselves, the primer concentration was reduced by 60% and the template DNA concentration was reduced by 80%. When this also failed, asymmetric PCR was no longer considered a viable option for amplifying these long DNA fragments with a high-fidelity enzyme.

Splicing without primers

The focus was shifted entirely to the splicing step of the reaction. If it could yield as little as 10 ng product that could be isolated from a gel after electrophoresis, even a few diluted microlitres of DNA isolated from the gel would be enough to use as a template for a new PCR.

Fragments R6 and R7 were starting to run low, so focus was shifted to fragments R3, R4, and R5. The gel extraction and ethanol precipitation of these three fragments had been very successful, and very small volumes could be used for high

concentrations of DNA. Fragments R3 and R4 were the first to be used in a splicing attempt.

An initial concentration of 200 ng of each fragment was used for the first splicing attempt. After 35 cycles of PCR with no primers, the results were observed on a gel. This method did not succeed, and attempts were made to optimise it in the hopes of a better result. The first attempt involved reducing the amount of DNA by 50% to discourage reannealing of fragments to their own complementary strands. Increasing the cycle number was tried in the hopes that product was simply not formed at a high enough concentration to be viewed on an electrophoresis gel. In a separate attempt, touchdown annealing temperatures were used. When assembled product could still not be seen on an electrophoresis gel, more extreme action was taken to determine whether the fragments would ever anneal to anything other than their own complementary strand. A gradient of annealing temperatures was set up from 72°C to 63°C and the cycle number was increased to 100. 50 ng of both fragments was used.

After the results of the previous PCR, 3% DMSO was added to the PCR mix to prevent the fragments from self-priming. Shorter DNA fragments R3 and R4 were used simply due to their availability. SOE-PCR with no primers was carried out with the same parameters as the second attempt at primer-free SOE-PCR. When this failed, another attempt was made using 5% DMSO. When a faint band was obtained from the gel, the DNA was extracted using the freeze and squeeze method, but not purified using ethanol precipitation. The band was used for PCR using the forward primer for fragment R3 and the reverse primer for fragment R4. When this produced no results, a simple PCR of an old sample of R1 that had been gel-purified and isolated using the freeze and squeeze method, but not ethanol-precipitated, was performed to determine whether EtBr in the reaction mix was inhibiting PCR. The same PCR mix and thermal cycler parameters were used as with the Phusion amplification of this fragment. The DNA suspension for the old R1 fragment fluoresced more brightly on the UV transilluminator than that of the isolated fused fragment, so for the sake of simplicity it was assumed that the EtBr concentration in the R3 suspension was at least as high as that of the R3+4 assembled fragment.

Testing the integrity of Phusion-amplified DNA fragments after persistent SOE-PCR failure

PCR was performed on the nine important fragments to determine whether the DNA itself was impeding transformation. The reaction was performed first with Phusion DNA polymerase, and then with Kapa HiFi DNA polymerase and a different tube of dNTPs after no results were observed for the first attempt. The third attempt at reamplification used freshly prepared primers. The reaction mixes are shown in Tables A2.2 and A2.10. The thermal cycler times and temperatures are shown in Table A2.4 and A2.11.

pCUT plasmids

Growing the E. coli host bacteria

The pCut Plasmid Toolkit was a gift from Aindrila Mukhopadhyay (Addgene kit #1000000118). The size of the plasmids is ~11 kbp. The *E. coli* host bacteria strains were delivered to the laboratory in a 96-well plate. The host strains were named after each unique plasmid. All plasmids provided ampicillin resistance as a selection marker. For each of the 27 strains, a 5 ml tube of 100 μg/ml ampicillin LB broth was prepared. Each tube was inoculated with 5 μl of a single strain of *E. coli* and incubated overnight at 37°C. The next morning, 600 μl samples of cell suspension were taken from the three strains coding for the most efficient insertion loci for plasmid extraction; the only one that remained relevant was the strain encoding the guide RNA for the 106a cut site. Each tube was then mixed with an equivalent volume of 50% glycerol to a final concentration of 25% glycerol, and 2 ml of each of these glycerol suspensions were frozen at -80°C. When a new liquid stock of any strain becomes necessary, a loop of bacteria is streaked onto an LB-Amp agar plate and incubated at 37°C for about a day, and then a single colony of the plate is transferred into 5 ml of liquid LB-Amp medium and incubated at 37°C for about a day before plasmid extraction.

Plasmid extraction

Plasmids were extracted from the 600 samples of bacterial suspension using the Zyppy™ Plasmid Miniprep Kit from Zymo Research. The plasmids for the

following target sites were chosen: 106a (three extractions), 1206b, and 911b. Plasmid extracts were stored at -80 °C as it was unclear when they would be used. After 18 months, the concentration of the plasmid DNA was determined using the Qubit Broad Range dsDNA assay.

The growth and preparation of the frozen stock of *S. cerevisiae* BY4742

S. cerevisiae was streaked onto YPD agar and incubated at 25°C for two weeks. One colony of yeast from a pre-grown YPD agar plate was streaked onto a fresh YPD agar plate and grown at 30°C until colonies started to show. Two test tubes each containing 5 ml of YPD broth were each inoculated with a single colony from the streak plate. The tubes were incubated at 25°C on a shaker platform at 150 rpm until they reached an OD_{600} of 3. The yeast culture was then diluted with YPD broth to an OD_{600} of around 0.15 and a volume of 100 ml and incubated at 25°C on a shaker at 150 rpm until the OD_{600} reached 0.8-1.2, in accordance with experiments carried out by Apel *et al* (2017) based on the method of preparing and transforming competent *S. cerevisiae* developed by Benatuil *et al* (2010). Unfortunately, due to time constraints, the transformation and selection experiments could not be carried out, so the preparation of competent cells was side-lined in favour of perfecting PCR.

Chapter 4:
Results and Discussion

RESULTS AND DISCUSSION

Yeast DNA extraction and PCR to amplify the flanking regions A and B

The protocol used for DNA extraction at first failed to produce results for PCR. The most likely reason that this failed was that the plates were stored in the incubator and subsequently the refrigerator for too long. Running the extract from old cells on a 1% agarose gel yielded only a single, very faint line of DNA at around 14 kbp, even though the NanoDrop™ measurements showed DNA concentrations of over 1 µg/µL. The necessary alterations made to the protocol to accommodate the limitations of the microcentrifuge are likely not to blame for the failure of the protocol, as this method was performed successfully with fresh yeast; however, these alterations could be responsible for the low DNA yield in successful extractions. It is possible that this method is not suited to old or expired cultures.

After the successful extraction of genomic DNA, the supernatant DNA concentration was measured by the NanoDrop™ to be 121.4 ng/µL. Although the protocol calls for only 1 µl of supernatant to be used, success was only found after increasing the amount of supernatant in the PCR reaction mix to 8 µL. OneTaq can theoretically work with much lower concentrations of DNA. This could mean that either the SDS in the lysis buffer was inhibiting PCR, or that the NanoDrop™ measurements were very inaccurate.

PCR of designer fragments

All PCR products amplified using either OneTaq or Phusion had primer dimers and some smearing around the product after gel electrophoresis. OneTaq had greater smearing around the product and a smaller primer dimer smear, whereas Phusion had a larger primer dimer smear, but sharper product bands. While primer dimers were expected for fragments R5, R7, V1, V2, V3, and V1.5, the primer dimers in the other reaction mixes were surprising. A potential reason for the presence of primer dimers in every reaction, apart from the difference in fidelity of the two enzymes, is the different thermal cyclers that were used during these reactions. The Applied Biosciences® 2720 thermal cycler automatically starts heating the lid when

it is turned on. This ensures that the protocol can begin immediately or almost immediately. In contrast, the Bio-Rad™ T-100 thermal cycler has an option to set the temperature of the heated lid when setting all other PCR parameters. The lid does not start heating until the protocol starts to run, at which point it will start heating from room temperature. This could potentially mean that, in the longer period of time between the lid heating and the initial denaturation step, primer dimers may have formed while the reaction mix gradually heated up before the protocol started. Although the protocols of OneTaq and Phusion were followed strictly, the concentration of the primers might also have contributed to the formation of primer dimers.

Gel purification of fragments

The extraction of DNA from the agarose gels using the Zymoclean™ Gel DNA Recovery Kit yielded a very low concentration of DNA. Gel purification often leads to loss of PCR product, but the loss of PCR product when purifying OneTaq PCR products using this kit was extremely high. For some fragments, the concentration was too low to measure. It is unclear why, but the most likely reason presented in the troubleshooting section of the protocol of the kit is that the gel was not yet fully melted. It is possible the time spent melting the gel was not long enough, but it was not extended past the recommended maximum time due to fears of damaging the DNA. Due to the sheer number of fragments in need of purification and the limited number of uses in a commercial kit, an in-house method became preferable for DNA recovery.

When the freeze and squeeze method was first used, the fragments amplified using OneTaq were amplified in triplicate, which had an overall higher yield, but complicated the extraction and ethanol precipitation step. The target bands on the gel were less sharp than that of Phusion, which led to a greater volume of gel being extracted for each band. The high volume of gel led to a high volume of TAE buffer being extracted with the DNA. The choice of filter-tip pipette was also less than ideal. The 1 mL filter tip was chosen simply for the large space in which six gel slices could fit at once, and was difficult to cut. Its filter would become dislodged if more than two centrifugation steps were used in the freeze and squeeze method.

Where the tip was cut, it often shed plastic shavings into the solution during the high RCF centrifugation steps of the gel extractions. It is possible that these shavings interfered with the formation of the DNA pellet during ethanol precipitation, or resuspended the DNA pellet afterwards by detaching from the Eppendorf tube. If that is the case, most of the DNA would have been lost after the first centrifuging step, when the supernatant was discarded. More DNA could have been lost this way during the wash steps. The yield of DNA showed no significant correlation with the size of the fragment, as shown in Figure 8, but when fragment A, the obvious outlier, was excluded, there was a very weak correlation between size and yield. This could mean that the larger fragments were slightly more resistant to the complications brought on by the plastic shavings, but there is no way to know conclusively, and the outlier fragment A is the second smallest fragment.

In contrast, fragments amplified by Phusion had consistently adequate yields from the gel. The single PCR reaction, split into two bands on the agarose gel, fit easily inside the well provided by a 200 µL filter pipette tip. The 200 µL tip was much easier to cut to fit the Eppendorf tube, and left no visible plastic shavings even after several rounds of centrifugation. The problem of the filter detaching and falling into the DNA suspension was also not observed even when the centrifugation step of the gel extraction was performed three or four times. The small fluid volume allowed for much easier ethanol precipitation, and each fragment left a visible small clear DNA pellet. These fragments showed no correlation between their length and their final yield, which means such a correlation likely does not exist at least for the size difference between these fragments. Excluding the wash steps had no discernible impact on downstream applications. Final concentrations for each purification method measured by Qubit 2.0 are shown in Table 4.

Table 4: DNA yield by method.			
Fragment	bp	Yield per purification method	
		Freeze and squeeze (OneTaq)	Truncated freeze and squeeze (Phusion)
A	949	228.00	20.4
1	1875	27.08	40.0
2	1875	135.00	60.0
3	1580	13.27	60.2
4	1580	13.96	199.8
5	1580	90.20	60.6
6	1875	7.82	43.8
7	1875	16.16	52.4
B	920	21.40	62.4
V1	3064	99.40	53.0
V2	3063	126.00	-
V3	3062	74.40	47.6
V1.5	3060	-	23.2

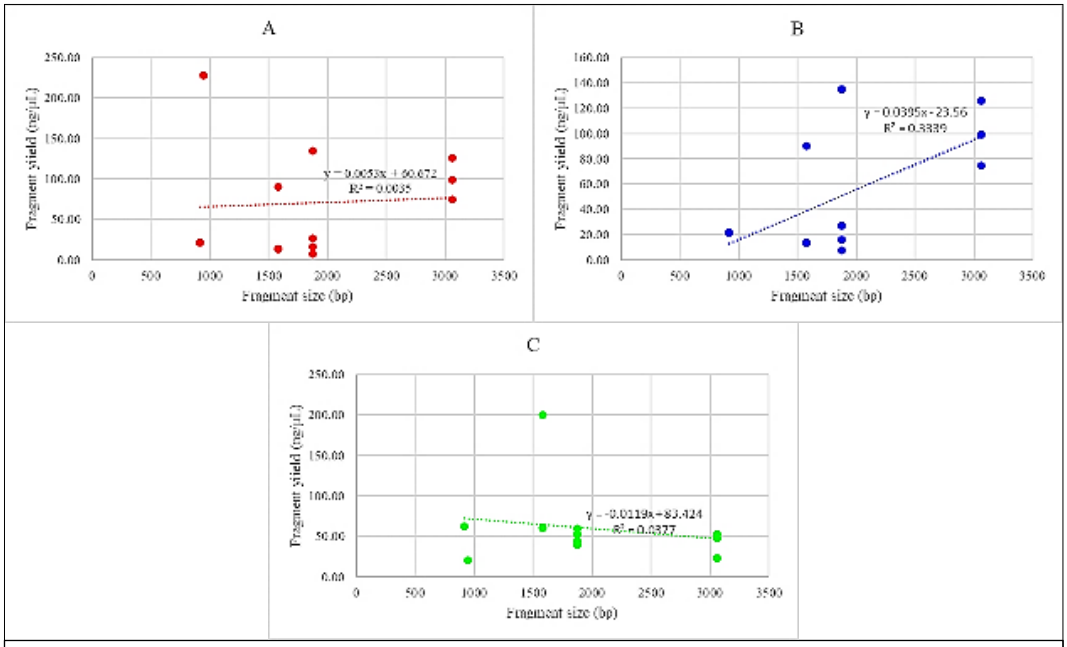


Figure 8: Graphs comparing the size of the purified fragment with the final yield of the fragment. A) When all fragments of the first purification were included in the calculation, there was no correlation between fragment size and the final yield. B) When outlier fragment A of the first purification was removed, there was only a very weak correlation between the fragment size and the final yield. C) Fragments amplified with Phusion and purified in the second purification attempt showed no correlation between fragment size and yield, with the outlier fragment R4 as neither the longest nor the shortest.

Gibson assembly

Linear Gibson assembly

The first attempts at Gibson assembly used only the reconstituted DNA fragments that had been ordered for this purpose, and used an isothermal buffer donated by a colleague. The reasons why this failed are not entirely clear, as Gibson assembly can in theory be used to assemble a linear fragment. It is possible that the isothermal buffer was expired. It was stored at -20°C, which is the preferred short-term storage temperature for NAD^+; however, it is not certain how long it had been stored at this temperature. As Gibson assembly is used primarily to assemble plasmids, it is also possible, although less likely, that assembly of a linear fragment was simply less efficient than assembly of a circular fragment. Due to limited DNA and time, this was not tested.

Gibson assembly into a plasmid vector

It was difficult to tell whether any circular product was formed because circular DNA does not migrate at the same pace as linear DNA (Schmidt *et al*, 2001). Gibson assembly yielded very little assembled product, ranging from no product to a faint glow where a band of linear product would be expected when assembling OneTaq PCR product, to very faint bands when Phusion PCR product was assembled. Figure 9 shows three new bands for every reaction, which are faint for the assembly of Vector 1 and very faint for the assembly of Vector 2. The bands at ~4000 bp are likely assemblies of fragments V1 and A, and fragments V3 and B, respectively. These bands do not appear to have travelled far enough to be assemblies of R1 and R2 / R6 and R7, which would be 3750 bp in length; however, it is not impossible considering the anomaly of DNA bands moving slower near the centre of the gel. The fragments at ~5000 bp are likely assemblies of fragments V1 and R1/R2, and fragments V3 and R6/R7, respectively. This band may also be assemblies of the three non-vector fragments: A, R1, and R2, or R6, R7, and B, respectively. The fragments at ~6000 bp are likely assemblies of three fragments: V1, R1/R2, and A, or V3, R6/R7, and B. There is no band ~7700, which would indicate a linear assembly of all four fragments for either Vector 1 or Vector 2.

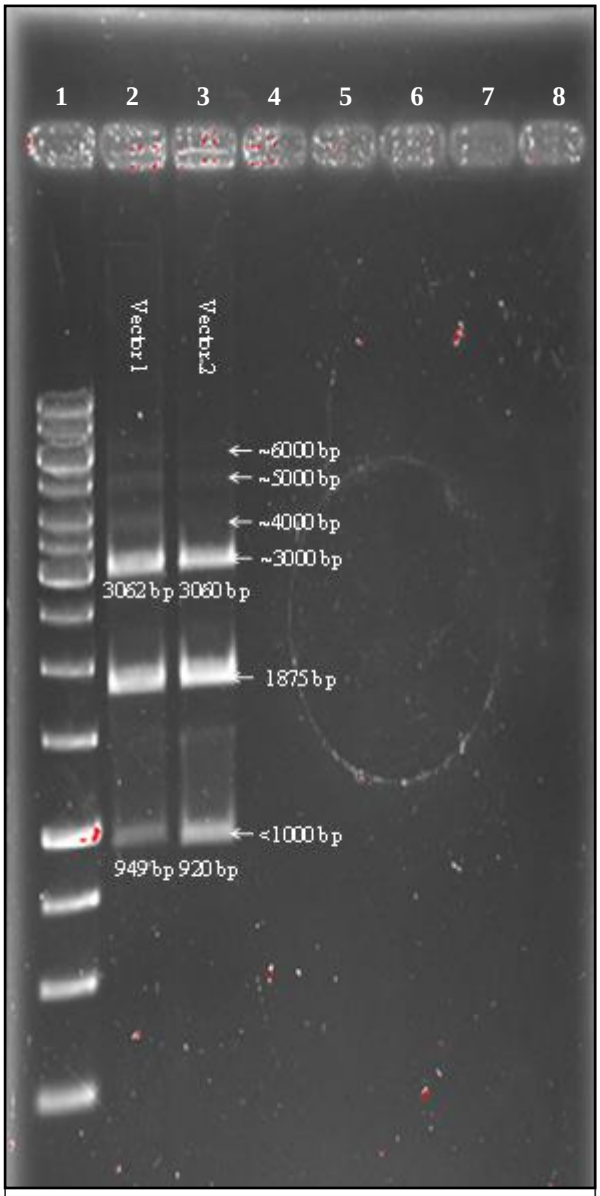


Figure 9: Partially successful Gibson assembly of Phusion-amplified fragments after 1 hour of incubation at 50°C. Lane 1 contains GeneRuler 1kb DNA ladder. Lane 2 contains the products for the Gibson assembly of Vector 1. Lane 3 contains the products for the Gibson assembly of Vector 3. DNA migration was slower nearer to the centre of the gel than at the edges, producing a slanted appearance.

While the poor assembly when using fragments amplified by OneTaq is likely due to mismatches on the 3' ends of the DNA generated by OneTaq (Clark, 1988), Gibson assembly also did not progress smoothly when using fragments amplified with Phusion, which means other factors likely also contributed. A potential cause is that PCR inhibitors lingered in the DNA suspensions even after ethanol precipitation. Ethidium bromide is the most likely culprit, but the final suspensions were briefly checked for fluorescence over a UV light, and no such fluorescence was apparent. Another potential cause is DNA damage after repeated exposure to EtBr and UV radiation, which could have left single-strand breaks in the DNA that would not be obvious even after subsequent electrophoresis, but would interfere with the assembly. If this is the case, it also explains the difficulty in amplifying assembled product from both diluted, unpurified Gibson assembly product and from gel-purified Gibson assembly product. Gibson assembly improved when a new assembly master mix was created, which indicates that NAD^+ degradation may have been at least partially to blame.

Gibson assembly results improved dramatically when the reaction mix was incubated for four hours instead of only one. The results are shown in Figure 10. It is possible that the DNA had unforeseen complexities that made assembly more difficult than anticipated, like a tendency to form secondary structures or damage that could not be seen on an electrophoresis gel. Although smearing is still present above the original fragments, it is not present to the same extent as after the one hour assembly.

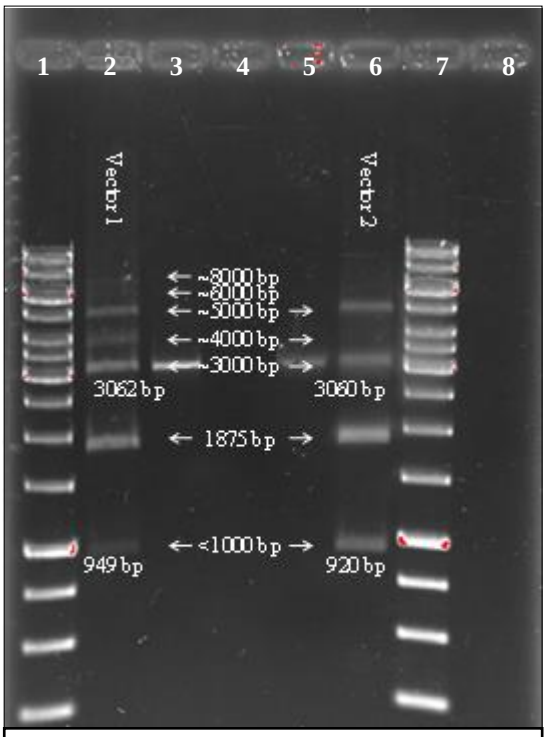


Figure 10: Partially successful Gibson assembly after 4 hours of incubation at 50°C. Lanes 1 and 7 contain GeneRuler 1 kb DNA ladder. Lane 2 contains the Gibson assembly products for the assembly of Vector 1. Lane 3 contains the equivalent amount of vector fragment V1 to compare the fluorescence and travel distance of the unassembled vector fragment. Lane 5 contains unassembled vector fragment V3 at an equal amount to the assembly reaction mix of Vector 3. Lane 6 contains the Gibson assembly products for the assembly of Vector 3.

For the four-hour assemblies, bands were observed for Vector 1 and Vector 3 at ~4000 bp and ~5000 bp, and for Vector 1 between 6000 and 8000 bp. The band at ~4000 bp could be an assembly between the V1 and A / V3 and B, or between the two synthesised DNA fragments, fragments R1 and R2 / fragments R6 and R7. The ~5000 bp fragment could be an assembly between the three insert fragments, fragments A, R1, and R2 / fragments R6, R7, and B, or between fragments V1 and R1 or R2 / fragments V3 and R6 or R7. The band between 6000 and 8000 bp could be an assembly between all four fragments of Vector 1 that failed to circularise. The fact that the assembly was supposed to produce circular DNA complicates matters. Because of the poor assembly rate and the multiple fragments and multiple products, it is unclear to where CCC-DNA or nicked circular DNA would migrate. The bands at ~4000 bp or ~5000 bp could

contain CCC-DNA, and the band between 6000 bp and 8000 bp could contain nicked circular DNA.

Gibson assembly of Inserts 1 and 2 into the same vector

The initial attempt at Gibson assembly of the combination of Inserts 1 and 2 failed for unknown reasons. It is possible that the seven different fragments in the reaction mix were simply too numerous to assemble correctly, especially for a master mix that consistently yielded poor assembly results. The enzymes were bought from NEB, who discourage using more than five fragments per assembly (excluding the vector).

PCR amplification of Gibson assembly product

PCR amplification of the insert in Vector 2 was proven impossible due to the forward primer for fragment R3 appearing in its entirety inside fragment R5. This was unfortunately unavoidable due to the composition of the three artificial genes, which all used the same promoter, terminator, and solubility tag, and how they were divided evenly between fragments. PCR confirmed that the formation of the short ~1.3 kb fragment far outcompeted the formation of the much larger 4.86 kb fragment. For this reason, the decision was made to assemble fragments A and R1-R5 into the combination Vector 1.5. While it would not be possible to change the primers to successfully amplify Insert 2, failure to plan for this was an avoidable oversight.

PCR amplification of DNA bands isolated after gel electrophoresis of the other two Gibson assembly products yielded only a long smear throughout the entire gel lane, or occasionally an empty lane, as did PCR amplification of diluted Gibson assembly product that had not been purified. It is possible that the Taq DNA ligase was not functioning, which would result in single strand nicks in the DNA making normal amplification during PCR impossible. This would mean that even in a reaction where both priming sites are present on one of the fragments, as is the case where either the unassembled vector or assembled vector is present, the amplification of shorter single strands would proceed exponentially and overwhelm the reaction, leading to a smear. The effect may be compounded by the 3’ to 5’ exonuclease

activity of Phusion polymerase (Veneziano *et al*, 2018). Degradation of ssDNA of various lengths could have contributed to the size of the smear.

If it is true that the Taq DNA ligase stopped functioning, it may be due simply to the degradation of the NAD^+ in the Gibson assembly master mixes. While the NAD^+ stock, Gibson assembly isothermal buffer, and Gibson assembly master mixes were treated very carefully and always thawed on ice in the absence of bright light, NAD^+ degradation is still possible.

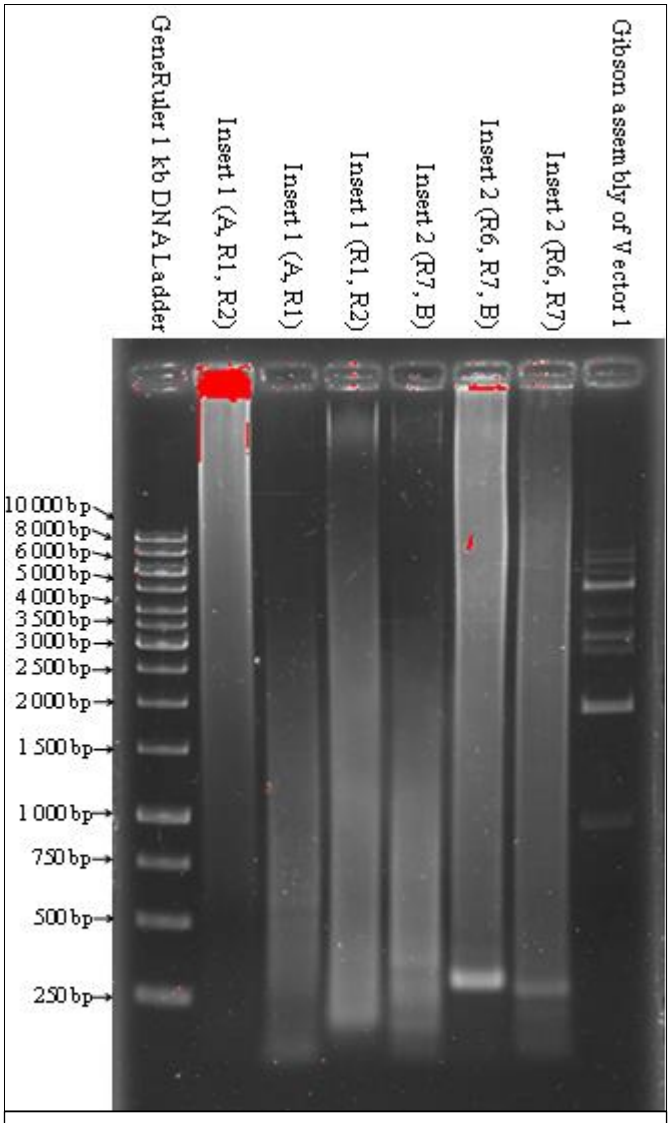


Figure 11: Gel electrophoresis of PCR products amplified directly from diluted unpurified Gibson assembly product. Lane 1 contains DNA ladder. Lanes 2-4 contain PCR results from Insert 1. Lanes 5-7 contain PCR results from Insert 3. Lane 8 contains a Gibson assembly of Vector 1.

The attempt to PCR-amplify smaller fragments from after Gibson assembly also failed, leaving the same long smear of nonspecific amplified products. The results of this attempt are shown in Figure 11. This provides further evidence for the possibility that the ligation step of the assembly did not work or that the DNA fragments were damaged; however, due to the nature of linear Gibson assembly products, this is not conclusive proof. Linear Gibson assembly products may lack the 5' ends of the template strand, which would prevent amplification of the full fragment. The result would be non-exponential amplification of ssDNA that lacks the second priming site. It should be noted that these

strands could anneal to each other and prime each other, following the basic principle of SOE-PCR with an asymmetric amplification step, but with the difficulties of using Phusion to amplify ssDNA (Veneziano *et al*, 2018), and the fixed amount of template for each of the fragments (as opposed to exponentially increasing template in normal PCR), it might not be feasible without extensive and time-consuming optimisation.

Transformation of competent E. coli with assembled plasmid vectors

Transformation of *E. coli* JM109 competent cells with cloned pGEM®-T Easy plasmid vector led to consistent failure regardless of the amount of assembly mix used. It is unclear why this was the case. The cells, when taken straight from the freezer and not transformed, grew reasonably well on plain LB agar without ampicillin; therefore, the frozen stocks were still viable. The cells showed no obvious decline in growth on plain LB agar after undergoing the transformation procedure, occasionally producing a lawn on the agar if left for two days, which means it is unlikely that the transformation protocol is causing great enough harm to kill a significant proportion of the cells. LB-Amp agar was remixed to ensure that the ampicillin was present at the correct concentration, but no change was observed. It is possible that the sheer size of the insert, which is greater than the size of the vector itself, was impeding transformation. It is likewise possible that the transformation failed due to some component of the Gibson assembly reaction inhibiting it. It is also possible that the cells have lost their competency, which would have been obvious with a normal pGEM®-T Easy ligation reaction, where some plasmids would ligate without the insert and allow some cells to survive; however, Gibson assembly would only yield intact plasmids if seamless cloning had successfully occurred.

SOE-PCR

Initial attempt

The first attempt at overlap extension PCR was made using fragments amplified with OneTaq, and the overlap phase was performed by Phusion. This reaction left only a smear on the gel around where the original bands were expected. One

potential reason for this is that the reaction simply required optimisation. The annealing temperature was also chosen based on faulty input in the online NEB Tm Calculator tool: while the sequence was correct, the concentration of the fragments was not adjusted based on the molar concentration of DNA in the solution, leading to a higher annealing temperature being used than would have been recommended had the correct concentration been used for the calculation.

SOE-PCR using asymmetric amplification

The initial attempt at SOE-PCR using asymmetric amplification, where two separate asymmetric PCR reactions were carried out for 30 rounds before the reaction mixes were combined so the ssDNA strands could prime each other, produced nothing but a long smear on the gel, with a slightly denser smear around the area where fragments R1 and R2 would be expected (1875 bp). The choice of Phusion as the enzyme may have contributed to the failure of this method. Fragments R1 and R2 are much longer than what Phusion is typically able to reproduce as ssDNA (Veneziano *et al*, 2018).

The second attempt at SOE-PCR using asymmetric amplification in a one-pot mix also produced a smear. It is possible that the primer concentration was simply too high. If that were the case, single strands may have been synthesised preferentially to the exclusion of longer fused DNA fragments. The high template concentration may also have contributed to the synthesis of ssDNA being favoured over the fused fragments.

The third attempt at asymmetric SOE-PCR, which used a low concentration of DNA and a low concentration of primers, produced mostly empty lanes on the gel, with a faint smear from around 1500 bp growing more pronounced towards the bottom of the gel. It is possible that the template DNA concentration and cycle number was simply too low for sufficient non-exponential amplification, and that optimisation of this protocol would eventually lead to the successful generation of an assembled product; however, Phusion was likely not the correct enzyme for this method and OneTaq was undesirable for its significantly higher error rate. For this reason, this method was shelved in favour of SOE-PCR with no primers.

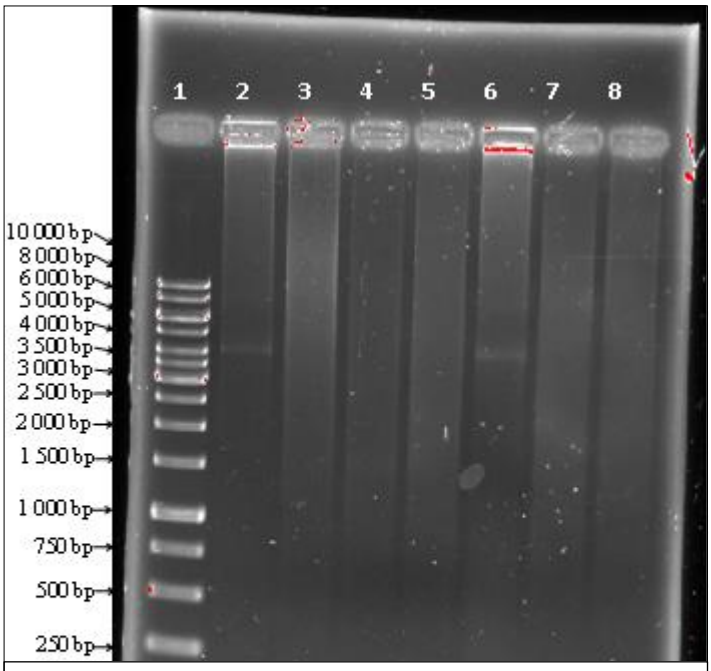


Figure 12: Gel electrophoresis results of the 100-cycle primer-free SOE-PCR. Lane 1 contains GeneRuler 1 kb DNA ladder. Lane 2 and lane 6 have a faint band just over 4000 bp. The other lanes have smears from the top. No smear was brightest around where the expected assembled fragment would be.

SOE-PCR with no primers

The initial attempt at SOE-PCR with no primers left a smear around where the bands for R1 and R2 would be expected, but no band or even smear around where the assembled fragment (3720 bp) would be expected. The smearing around where a band for the unassembled two fragments would be expected was also present when the concentration of DNA was reduced by 50%, albeit less intense. Overlap extension PCR with touchdown annealing yielded a similar smear.

The entire gradient of annealing temperatures when paired with 100 PCR cycles yielded no product, only a smear near the top of the gel with their brightest areas nowhere near the expected length of the assembled product. Two of the lanes yielded a band at just over 4000 bp. The product would be around 3720 bp, which indicates that there was likely a self-priming problem with at least one of the two fragments. Figure 11 shows the electrophoresis results of the gradient of annealing temperatures.

A potential contributing factor for the smears in linear two-fragment SOE-PCR with no primers is the initial high concentration of template DNA fragments. These reactions, which utilised NEB Phusion DNA polymerase, used a DNA concentration roughly 10-20 times the maximum concentration suggested for normal PCR by NEB per individual fragment. Roughly half the weight of the template fragments is made up of strands that cannot prime each other. These single

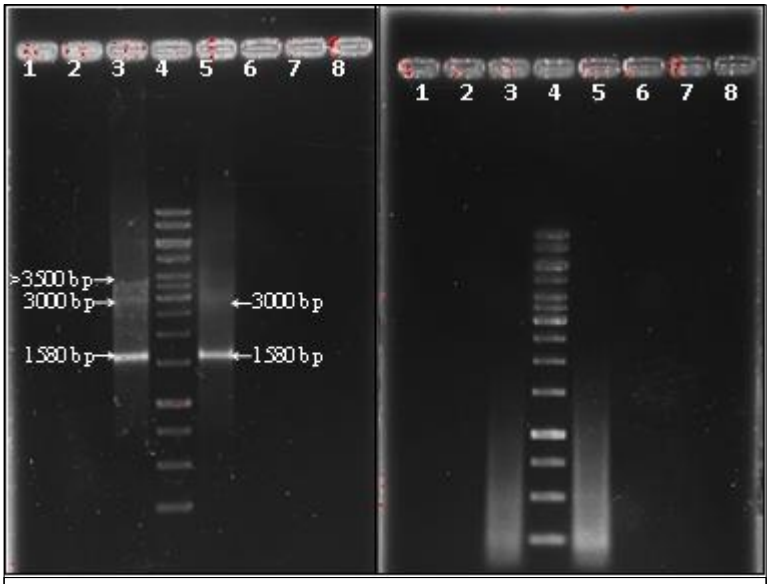


Figure 13: Results of SOE-PCR without primers and the subsequent attempt at reamplification. The contrast of both images was increased by 40% and the brightness reduced by 20% to show the very faint bands more clearly. On both gels, lane 4 contains GeneRuler 1 kb DNA ladder and lanes 3 and 5 contain the PCR results, split into two volumes of 30 µL each including loading buffer. The gel on the right shows the initial results for overlap extension PCR of fragment R3+4.

strands are long and may bind to other DNA non-specifically, resulting in short or long mutant amplicons. Even the priming strands may bind nonspecifically to their complementary strand or a strand on the other DNA fragment.

The addition of 3% DMSO in the final attempt at SOE-PCR without primers helped prevent excessive nonspecific binding, as shown in Figure 13. The bands of unassembled 1580 bp fragments R3 and R4 are much sharper than in previous PCR attempts. Two faint bands were observed after the primer-free first step. The top band was around the expected size for an assembled fragment, but was less visible than the other band at around 3000 bp. Although the correct band was successfully isolated and extracted from the gel, subsequent PCR left only smears, with the brightest part of the smear around where primer dimers are usually observed. This could mean that the band isolated around where the assembled product was expected was not the assembled product, but a nonspecific amplicon that happened to be around the correct size. The presence of the second, lower band indicates that nonspecific priming and perhaps self-priming occurred in this reaction, so it is possible that it occurred in two distinct ways. However, because the DNA was not purified by ethanol precipitation prior to PCR, it is possible that PCR was simply inhibited in this case. Prior experimental evidence against this theory includes the very successful amplification of the short part of Insert 2 between the forward

primer for R3 inside R5 and the reverse primer of R5. The template DNA for this reaction was also not ethanol precipitated. While this fragment is only about a third as large as the desired amplicon in the SOE-PCR reaction, it resulted in minimal smearing on the gel and a bright band of amplified product. The subsequent PCR of an old suspension of fragment R1, which had also been gel purified and extracted but never purified through ethanol precipitation, yielded more smearing, but with a clear band of the target amplicon.

It is possible that the DNA obtained from the first step in SOE-PCR was damaged by a combination of EtBr and UV light. Cutting the fragment out of the gel took longer than normal because of difficulty viewing the band on the UV transilluminator. If the DNA was damaged, it is likely that only short single strands of varying length were produced during the second PCR.

Reamplification of fragments to determine DNA integrity

Reamplification of the nine important fragments formerly amplified and purified by gel electrophoresis consistently failed regardless of choice of enzyme or primer freshness. Considering that these fragments still travelled normally on a gel when observed after Gibson assembly or SOE-PCR with no primers, this provides further evidence that the DNA had been damaged by a combination of EtBr and UV light. EtBr has long been known as an intercalating molecule that could cause mutagenesis (Waring, 1965). UV light primarily causes damage to DNA by causing the formation of pyrimidine dimers between adjacent pyrimidine bases (Durbeej & Eriksson, 2002). These factors were the only consistent experiences each of the fragments had. Gibson Assembly and SOE-PCR did not use all fragments equally, which makes contamination of all of the fragments unlikely. The working solutions for all the primers were four months old and stored in autoclaved dH_2O, even though the stock solutions for the primer for fragments A and B were much older. Replacing the primers did not lead to successful PCR either, which rules their age out as the problem. With nine reaction mixes, it is possible that a reagent could have been left out of one or two tubes, but highly improbable that it was left out of every single one after being marked off on a checklist, and especially improbable that it was done twice. Cutting the bands out of the gel involved prolonged exposure to

UV light, with at least six bands to excise per gel. More care was put into cutting off excess gel than into ensuring the UV exposure was minimal. In the future, the latter will be prioritised, as even small DNA bands lose enough fluid during freeze and squeeze extraction to dilute the DNA beyond workable concentrations.

Growth of yeast

In the 30°C incubator, *S. cerevisiae* BY4742 grew very rapidly, showing significant growth by day 3. The yeast liquid culture was not grown at 30°C due to a mechanical error of the incubator preventing the use of the shaker. At 25°C, it took roughly twice as long to reach the ideal optical density both in the 5 mL culture and the 100 mL culture. There were no problems with growth otherwise.

Chapter 5:
Conclusion and Future Research

CONCLUSION AND FUTURE RESEARCH

The attempt to assemble three linear DNA inserts for simultaneous integration into the same locus of *S. cerevisiae* BY4742 was ultimately unsuccessful. The reason why the yeast could not be successfully transformed is that the three inserts could not be assembled from their smaller fragments using common molecular methods. The reason why this was the case and why the plan for assembly and integration had very room for change can be attributed to poor decisions made during the design of the DNA fragments.

Although care was put into the design of the artificial genes with regards to how well they would be expressed and how functional they would be in a foreign environment, not much care was put into limiting complications like large areas of homology between fragments. While budgetary constraints remained a limiting factor in how the DNA could be designed, features of the donor DNA like the reused promoter, solubility tag, and terminator were easily avoidable problems that had far-reaching consequences for this study.

If the choice had not been made to insert all three homologous genes into the same locus, the homology between the three genes would not have been a problem unless the yeast was transformed with all three genes simultaneously. However, one of the primary advantages of the pCut method is the potential for multiple simultaneous integrations (Apel *et al*, 2017), and therefore it would have been advantageous to use different promoters and terminators.

Regardless of whether or not the unintentional regions of sequence homology were eliminated, the decision to integrate all three genes into a single locus severely limited the direction in which the study was able to proceed. The intention was to limit the total number of base pairs used while attempting to construct three inserts of equal length for the co-transformation into competent *S. cerevisiae*. Assembling the DNA into a single long sequence and dividing the sequence with no regard for where genes began or ended made it impossible both for each gene to be expressed separately, or for the method of transformation to be altered. While it would be relatively easy to switch between integration methods for single genes, the targeted scarless integration of nearly 14 000 bp poses a significant challenge. Likewise, if

one of the three genes posed a significant challenge with regards to assembly or integration, the gene could not simply be removed from the study or put aside until a solution was found. Instead of this all-or-nothing approach to integrating these genes, the number of bp could simply have been limited by removing the NADPH-cytochrome P450 reductase gene from the study completely. Since the aim of this study was to provide a strain of *S. cerevisiae* to act as a scaffold for future integrations of the taxol biosynthesis pathway, removing an enzyme that is not even strictly necessary, and which can be added in a future study, would drastically improve the possibilities for the design of the DNA fragments. Each gene could then have its own set of gBlocks completely separate from the other. The pCut method was designed using three inserts with 30-80 bp sequence homology between them (Apel *et al*, 2017). If assembly of gBlocks proved problematic, as was the case in this study, co-transforming the yeast with the separate fragments would have likely still led to a successful integration thanks to the HDR mechanism of the cell and the regions of sequence homology designed into the gBlocks for Gibson assembly.

If the genes were on separate gBlocks, yeast integrating plasmids (YIps) could have been used to integrate the genes into the yeast genome, however, these plasmids require a selection marker, may have a low integration efficiency, and may integrate a given gene more than once (Sikorski & Hieter, 1989; Zhang *et al*, 1996). Yeast centromere plasmids (YEps), plasmids with a partial centromere developed by Clarke and Carbon (1980) to act and replicate like a miniature yeast chromosome, could also be used for separate genes; however, they integrate more than once and could place a high metabolic burden on the cell. The single controlled insertion of each gene of interest into the yeast genome using CRISPR-Cas9-targeted integration remains the preferred method for the reconstruction of the complex taxol biosynthesis pathway; therefore, close attention must be paid to the failures of the DNA assembly.

The reason for the failure of Gibson assembly is uncertain. The most likely explanation is that the NAD^+ in the reaction mix was degraded, and therefore Taq DNA ligase could not function properly; however, even if the DNA backbone was not repaired, the assembly of the DNA fragments should still have produced a

stronger band. The subsequent success of PCR amplifications using Phusion rules it out as the culprit, which leaves the T5 exonuclease as the potential cause of the poor assembly efficiency. However, as mentioned before, the DNA contained many unintentional regions of homology, which could have interfered with the assembly. It is also possible that the DNA formed unforeseen secondary structures that inhibited Gibson assembly. If the situation arises again, the addition of DMSO to the reaction mix might be helpful. The potential for multi-fragment scarless assembly continues to make Gibson assembly the preferable method for DNA assembly in future studies. The purchase of Gibson assembly positive control fragments may be helpful in the future to determine whether the reaction mix is still working.

While SOE-PCR cannot rival Gibson assembly in terms of simplicity, it provides a reasonable backup plan for if Gibson assembly does not work. Such a wide variety of SOE-PCR protocols exist that it is almost difficult to choose which one to use; however, the wide variety of successful methods complicates troubleshooting, and the only unambiguously positive effect caused by a change to the reaction mix or protocol was the addition of DMSO to the reaction mix. One benefit SOE-PCR has over Gibson assembly is that the product can be PCR-amplified to increase the yield, whereas linear Gibson assembly product lacks the 5' ends of each DNA strand. During PCR, primers can bind to the complementary strand and synthesise a new strand, but the reverse primer cannot bind to the new strand. Likewise, the forward primer cannot bind to a strand synthesised by the reverse primer.

It is unfortunate that the DNA samples were lost before the SOE-PCR protocol by Kadkhodaei *et al* (2016) could be attempted. Kadkhodaei *et al* (2016) employed a novel thermal cycling method which involved a 100-second-long annealing step where the temperature was lowered slowly (0.1°C/s) from 80°C to 70°C before extension at 72°C. The sustained high temperatures could have had a positive impact on the specificity of the mutual priming of DNA fragments.

For future yeast genomic integration studies, the pCut method remains an excellent choice for high-accuracy, multiple simultaneous DNA integrations into the yeast

genome. Future studies will continue the aim of building towards a *S. cerevisiae* cell factory that expresses every enzyme in the taxol biosynthesis pathway.

References

REFERENCES

Apel, A.R., d'Espaux, L., Wehrs, M., Sachs, D., Li, R.A., Tong, G.J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N.J., Keasling, J.D., & Mukhopadhyay, A. (2017). A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 45(1): 496-508. [Available at https://doi.org/10.1093/nar/gkw1023]

Arbuck, S.G., Christian, M.C., Fisherman, J.S., Cazenave, L.A., Sarosy, G., Suffness, M., Adams, J., Canetta, R., Cole, K.E., & Friedman, M.A. (1993). Clinical Development of Taxol. *Journal of the National Cancer Institute Monographs* No. 15: 11-24.

Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A.A., Hoseinpoor, R., Rahbarnia, L., & Aria, M. (2019) Yeast Expression Systems: Overview and Recent Advances. *Molecular Biotechnology*, 61: 365-384. [Available at https://doi.org/10.1007/s12033-019-00164-8]

Barnes, W.M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *PNAS*, 91(6): 2216-2220. [Available at https://doi.org/10.1073/pnas.91.6.2216]

Benatuil, L., Perez, J.M., Belk, J., & Hsieh, C.-M. (2010). An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design & Selection*, 23(4): 155–159. [Available at https://doi.org/10.1093/protein/gzq002]

Boge, T.C., Hepperle, M., Vander Velde, D.G., Gunn, C.W., Grunewald, G.L., & Georg, G.I. (1999). The oxetane conformational lock of paclitaxel: Structural analysis of D-secopaclitaxel. *Bioorganic & Medicinal Chemistry Letters*, 9(20): 3041-3046. [Available at https://doi.org/10.1016/S0960-894X(99)00521-1]

Bonetti, B., Fu, L., Moon, J., & Bedwell, D.M. (1995). The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences in *Saccharomyces cerevisiae*. *Journal of Molecular Biology*, 251(3): 334-345. [Available at https://doi.org/10.1006/jmbi.1995.0438]

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D. (1998). Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, 14(2): 115-132. [Available at https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2]

Chew, S.Y., Chee, W.J.Y. & Than, L.T.L. (2019). The glyoxylate cycle and alternative carbon metabolism as metabolic adaptation strategies of *Candida glabrata*: perspectives from *Candida albicans* and *Saccharomyces cerevisiae*. *Journal of Biomedical Science*, 26**:**52. [Available at https://doi.org/10.1186/s12929-019-0546-5]

Clark, J.M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research*, 16(20): 9677-9686. [Available at https://doi.org/10.1093/nar/16.20.9677]

Clarke, L., & Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 287: 504–509. [Available at https://doi.org/10.1038/287504a0]

Croteau, R., Ketchum, R.E.B., Long, R.M., Kaspera, R., Wildung, M.R. (2006). Taxol Biosynthesis and Molecular Genetics. *Phytochemistry Reviews*, 5: 75-97. [Available at https://doi.org/10.1007/s11101-005-3748-2]

Curran, K.A., Karim, A.S., Gupta, A., & Alper, H.S. (2013). Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering*, 19: 88-97. [Available at https://doi.org/10.1016/j.ymben.2013.07.001]

DeJong, JH.M., Liu, Y., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D., & Croteau, R.D. (2005). Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, 93(2): 212-224. [Available at https://doi.org/10.1002/bit.20694]

Dolgova, A.S., & Stukolova, O.A. (2017). High-fidelity PCR enzyme with DNA-binding domain facilitates de novo gene synthesis. *3 Biotech*, 7(2):128. [Available at https://doi.org/10.1007/s13205-017-0745-2]

Durbeej, B., & Eriksson, L.A. (2002). Reaction mechanism of thymine dimer formation in DNA induced by UV light. *Journal of Photochemistry and Photobiology A: Chemistry*, 152(1-3): 95-101. [Available at https://doi.org/10.1016/S1010-6030(02)00180-6]

Egger, E., Tauer, C., Cserjan-Puschmann, M., Grabherr, R., & Striedner, G. (2020). Fast and antibiotic free genome integration into *Escherichia coli* chromosome. *Scientific Reports*, 10:16510. [Available at https://doi.org/10.1038/s41598-020-73348-x]

El-Sayed, A.S.A., Fathalla, M., Yassin, M.A., Zein, N., Morsy, S., Sitohy, M, & Sitohy, B. (2020). Conjugation of *Aspergillus flavipes* Taxol with Porphyrin Increases the Anticancer Activity of Taxol and Ameliorates Its Cytotoxic Effects. *Molecules*, 25:263. [Available at https://doi.org/10.3390/molecules25020263]

Fernández, E., Moreno, F., & Rodicio, R. (1992). The *ICL1* gene from *Saccharomyces cerevisiae*. *The FEBS Journal*, 204(3): 983-990. [Available at https://doi.org/10.1111/j.1432-1033.1992.tb16720.x]

Foland, T.B., Dentler, W.L., Suprenant, K.A., Gupta, M.L., & Himes, R.H. (2005). Paclitaxel-induced microtubule stabilization causes mitotic block and apoptotic-like cell death in a

paclitaxel-sensitive strain of *Saccharomyces cerevisiae*. *Yeast*, 22(12): 971-978. [Available at https://doi.org/10.1002/yea.1284]

Gallego-Jara, J., Lozano-Terol, G., Sola-Martínez, R.A., Cánovas-Díaz, M., De Diego Puente, T. (2020). A Compressive Review about Taxol®: History and Future Challenges. *Molecules*, 25(24):5986. [Available at https://doi.org/10.3390/molecules25245986]

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., & Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5): 343-345. [Available at https://doi.org/10.1038/nmeth.1318]

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1): 31-46. [Available at https://doi.org/10.1158/2159-8290.CD-21-1059]

Heiman, M.G., & Walter, P. (2000). Prm1p, a Pheromone-Regulated Multispanning Membrane Protein, Facilitates Plasma Membrane Fusion during Yeast Mating. *Journal of Cell Biology*, 151 (3): 719-730. [Available at https://doi.org/10.1083/jcb.151.3.719]

Herskowitz, I. (1988). Life Cycle of the Budding Yeast *Saccharomyces cerevisiae*. *Microbiological Reviews*, 52(4): 536-553. [Available at https://doi.org/10.1128/mr.52.4.536-553.1988]

Hilgarth, R.S., & Lanigan, T.M. (2020). Optimization of overlap extension PCR for efficient transgene construction. *MethodsX*, 7:100759. [Available at https://doi.org/10.1016/j.mex.2019.12.001]

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6): 1262-1278. [Available at https://doi.org/10.1016/j.cell.2014.05.010]

Hughes, R.A., & Ellington, A.D. (2017) Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harbor Perspectives in Biology*. 9(1):a023812. [Available at https://doi.org/10.1101/cshperspect.a023812]

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, 169(12): 5429-5433. [Available at https://doi.org/10.1128/jb.169.12.5429-5433.1987]

Jennewein, S., Park, H., DeJong, JH.M., Long, R.M., Bollon, A.P., & Croteau, R.B. (2005). Coexpression in Yeast of *Taxus* Cytochrome P450 Reductase With Cytochrome P450 Oxygenases Involved in Taxol Biosynthesis. *Biotechnology and Bioengineering*, 89(5): 588-598. [Available at https://doi.org/10.1002/bit.20390]

Jiang, Y., Proteau, P., Poulter, D., & Ferro-Novick, S. (1995). BTS1 encodes a geranylgeranyl diphosphate synthase in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 270(37): 21793-21799. [Available at https://doi.org/10.1074/jbc.270.37.21793]

Kadkhodaei, S., Memari, H.R., Abbasiliasi, S., Rezaei, M.A. Movahedi, A., Shung, T.J., & Ariff, A.B. (2016). Multiple overlap extension PCR (MOE-PCR): an effective technical shortcut to high throughput synthetic biology. *RSC Advances*, 71: 66682-66694. [Available at https://doi.org/10.1039/C6RA13172G]

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1): 27-30. [Available at https://doi.org/10.1093/nar/28.1.27]

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Science*, 28: 1947-1951. [Available at https://doi.org/10.1002/pro.3715]

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1): D545-D551. [Available at https://doi.org/10.1093/nar/gkaa970]

Kennedy, E.M., Cullen, B.R. 2015. Bacterial CRISPR/Cas DNA endonucleases: A revolutionary technology that could dramatically impact viral research and treatment. *Virology*, 478-480: 213-220. [Available at https://doi.org/10.1016/j.virol.2015.02.024]

Kuang, X., Sun, S., Wei, J., Li, Y., & Sun, C. (2019). Iso-Seq analysis of the *Taxus cuspidata* transcriptome reveals the complexity of Taxol biosynthesis. *BMC Plant Biology*, 19:210. [Available at https://doi.org/10.1186/s12870-019-1809-8]

Kudla, G., Murray, A.W., Tollervey, D., & Plotkin, J.B. (2009). Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science*, 324(5924): 255-258. [Available at https://doi.org/10.1126/science.1170160]

Lahue, C., Madden, A.A., Dunn, R.R., & Heil, C.S. (2020). History and Domestication of *Saccharomyces cerevisiae* in Bread Baking. *Frontiers in Genetics*, 11:584718. [Available at https://doi.org/10.3389/fgene.2020.584718]

Lazarow, P.B. (2016). "Peroxisomes." in Bradshaw, R.A, & Stahl, P.D. (eds.) *Encyclopedia of Cell Biology Volume 2*, pp. 248-272. Academic Press, Waltham (MA). ISBN 9780123947963. [Available at https://doi.org/10.1016/B978-0-12-394447-4.20022-9]

Lim, P.T., Goh, B.H., & Lee, W.-L. (2022). "Taxol: Mechanisms of action against cancer, an update with current research" in Swamy, M.K., Pullaiah, T., & Chen, Z.-S. (eds.) *Paclitaxol*, pp. 47-71. Academic Press. ISBN 9780323909518. [Available at https://doi.org/10.1016/B978-0-323-90951-8.00007-2]

Liu, M.-S., Gong, S., Yu, H.-H., Taylor, D.W., Johnson, K.A. (2019). Chapter Thirteen - Kinetic characterization of Cas9 enzymes. *Methods in Enzymology*, 616: 289-311. [Available at https://doi.org/10.1016/bs.mie.2018.10.022]

Liu, W.C., Gong, T., & Zhu, P. (2016). Advances in exploring alternative taxol sources. *RSC Advances*, 6: 48800-48809. [Available at https://doi.org/10.1039/C6RA06640B]

Lõoke, M., Kristjuhan, K., & Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques*, 50(5): 325–328. [Available at https://doi.org/10.2144/000113672]

Mamadalieva, N.Z., & Mamedov, N.A. (2020). "Taxus brevifolia a High-Value Medicinal Plant, as a Source of Taxol" in Mathé, A. (ed.) *Medicinal and Aromatic Plants of North America Volume 6*, 201-218. Springer. ISBN: 978-3-030-44928-5. [Available at https://doi.org/10.1007/978-3-030-44930-8]

Martin, V. (1993). Overview of Paclitaxel (TAXOL®). *Seminars in Oncology Nursing*, 9(4(2)): 2-5. [Available at https://doi.org/10.1016/S0749-2081(16)30035-3]

Maury, J., Kannan, S., Jenson, N.B., Öberg, F.K., Kildegaard, K.R., Forster, J., Nielsen, J., Workman, C.T., & Borodina, I. (2018). Glucose-Dependent Promoters for Dynamic Regulation of Metabolic Pathways. *Frontiers in Bioengineering and Biotechnology*, 6:63. [Available at https://doi.org/10.3389/fbioe.2018.00063]

McElroy, C., Jennewein S. (2018). Taxol® Biosynthesis and Production: From Forests to Fermenters. In: Schwab W., Lange B., Wüst M. (eds) *Biotechnology of Natural Products*, 145-185. Springer [Available at https://doi.org/10.1007/978-3-319-67903-7_7]

Mojica, F.J.M., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloforax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology* 17(1), 85-93. [Available at https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x]

Mojica, F.J.M., Juez, G., & Rodríguez-Valera, F. (1993). Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified *Pst*I sites. *Molecular Microbiology*, 9(3): 613-621. [Available at https://doi.org/10.1111/j.1365-2958.1993.tb01721.x]

Moriyama, E.N., & Powell, J.R. (1998). Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Research*, 26(13): 3188-3193. [Available at https://doi.org/10.1093/nar/26.13.3188]

Nakata, A, Amemura, M., & Makino, K. (1989). Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *Journal of Bacteriology*, 171(6): 3553-3556. [Available at https://doi.org/10.1128/jb.171.6.3553-3556.1989]

National Center for Biotechnology Information. PubChem Compound Summary for CID 443489, 10-Deacetyl-2-debenzoylbaccatin III. [Available at https://pubchem.ncbi.nlm.nih.gov/compound/443489] Accessed June 25, 2022.

Nath, K., Sarosy, J.W., Hahn, J., & Di Como, C.J. (2000). Effects of ethidium bromide and SYBR Green I on different polymerase chain reaction systems. *Journal of Biochemical and Biophysical Methods*, 42(1–2): 15-29. [Available at https://doi.org/10.1016/S0165-022X(99)00033-0]

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Doemae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014) Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, 156(5): 935-949. [Available at https://doi.org/10.1016/j.cell.2014.02.001]

Peng, R., Lin, G. and Li, J. (2016). Potential pitfalls of CRISPR/Cas9-mediated genome editing. *The FEBS Journal*, 283(7): 1218–1231. [Available at http://doi.wiley.com/10.1111/febs.13586]

Plotkin, J.B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, 12: 32-42. [Available at https://doi.org/10.1038/nrg2899]

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba S., Zhang, Y., & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154: 1380-1389. [Available at https://doi.org/10.1016/j.cell.2013.08.021]

Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*, 117: 119-128. [Available at https://doi.org/10.1016/j.biochi.2015.03.025]

Samaranayake, G., Magri, N.F., Jitrangsri, C., & Kingston, D.G.I. (1991). Modified taxols. 5. Reaction of taxol with electrophilic reagents and preparation of a rearranged taxol derivative with tubulin assembly activity. *The Journal of Organic Chemistry*, 56: 17, 5114-5119. [Available at https://doi.org/10.1021/jo00017a024]

Sanchez-Muñoz, R., Perez-Mata, E., Almagro, L., Cusido, R. M., Bonfill, M., Palazon, J., *et al*. (2020). A novel hydroxylation step in the Taxane biosynthetic pathway: a new approach to Paclitaxel production by synthetic biology. *Frontiers in Bioengineering and Biotechnology*, 8:410. [Available at https://doi.org/10.3389/fbioe.2020.00410]

Schepartz, S.A. (1993). Summary—Session I: Supply—Harvest of *Taxus brevifolia*. *Journal of the National Cancer Institute Monographs* No. 15: 5-6.

Schmidt, T., Friehs, K. and Flaschel, E. (2001). 'Structures of Plasmid DNA.' Schleef, M. (Ed.). *Plasmids for Therapy and Vaccination*. Weinheim, Germany: Wiley-VCH, pp. 29-43. [Available at https://doi.org/10.1002/9783527612833.ch02]

Sharp, P.M., & Li, W.-H. (1987). The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*. 15(3): 1281–1295. [Available at https://doi.org/10.1093/nar/15.3.1281]

Sikorski, R.S., & Hieter, P. (1989). A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces Cerevisiae*. *Genetics*, 122(1): 19-27. [Available at https://doi.org/10.1093/genetics/122.1.19]

Snyder, J.P., Nettles, J.H., Cornett, B., Downing, K.H., & Nogales, E. (2001). The binding conformation of Taxol in β-tubulin: A model based on electron crystallographic density. *PNAS*, 98(9): 5312-5316. [Available at https://doi.org/10.1073/pnas.051309398]

Soliman, S.S.M., Greenwood, J.S., Bombarely, A., Mueller, L.A., Tsao, R., Mosser, D.D., & Raizada, M.N. An Endophyte Constructs Fungicide-Containing Extracellular Barriers for Its Host Plant. *Current Biology*, 25(19): 2570-2576. [Available at https://doi.org/10.1016/j.cub.2015.08.027]

Song, T.-Q., Ding, M.-Z., Zhai, F., Liu, D., Xiao, W.-H., & Yuan, Y.J. (2017). Engineering *Saccharomyces cerevisiae* for geranylgeraniol overproduction by combinatorial design. *Scientific Reports*, 7:14991. [Available at https://doi.org/10.1038/s41598-017-15005-4]

Stajich, J.E., Dietrich, F.S. & Roy, S.W. (2007). Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biology*, 8(10): R223. [Available at https://doi.org/10.1186/gb-2007-8-10-r223]

Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M., & Heyneker, H.L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 164(1): 49-53. [Available at https://doi.org/10.1016/0378-1119(95)00511-4]

Subban, K., Subramani, R., Srinivasan, V.P.M., Johnpaul, M., & Chelliah, J. (2019). Salicylic acid as an effective elicitor for improved taxol production in endophytic fungus *Pestalotiopsis microspora*. *PLoS ONE*, 14(2): e0212736. [Available at https://doi.org/10.1371/journal.pone.0212736]

Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J., & Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *PNAS*, 99(11): 7536-7541. [Available at https://doi.org/10.1073/pnas.112047299]

Veneziano, R., Shepherd, T.R., Ratanalert, S., Bellou, L., Tao, C., & Bathe, M. (2018). *In vitro* synthesis of gene-length single-stranded DNA. *Scientific Reports*, 8:6548. [Available at https://doi.org/10.1038/s41598-018-24677-5]

Walker, K., & Croteau, R. (2000). Taxol biosynthesis: Molecular cloning of a benzoyl-CoA:taxane 2α-O-benzoyltransferase cDNA from *Taxus* and functional expression in

Escherichia coli. *PNAS*, 97(25): 13591-13596. [Available at https://doi.org/10.1073/pnas.250491997]

Walker, K.D., Klettke, K., Akiyma, T., Croteau, R. (2004). Cloning, Heterologous Expression, and Characterization of a Phenylalanine Aminomutase Involved in Taxol Biosynthesis. *Journal of Biological Chemistry*, 297(52): 53947-53954. [Available at https://doi.org/10.1074/jbc.M411215200]

Wang, M., Cornett, B., Nettles, J., Liotta, D.C., & Snyder, J.P. (2000). The Oxetane Ring in Taxol. *The Journal of Organic Chemistry*, 65(4): 1059-1068. [Available at https://doi.org/10.1021/jo9916075]

Wang, T., Li, L., Zhuang, W., Zhang, F., Shu, X., Wang, N., & Wang, Z. (2021). Recent Research Progress in Taxol Biosynthetic Pathway and Acylation Reactions Mediated by *Taxus* Acyltransferases. *Molecules*, 26(10):2855. [Available at https://doi.org/10.3390/molecules26102855]

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P., & McPhail, A.T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93(9): 2325-2327. [Available at https://pubs.acs.org/doi/pdf/10.1021/ja00738a045 (PDF only)]

Waring, M.J. (1965) Complex formation between ethidium bromide and nucleic acids. *Journal of Molecular Biology*, 13(1): 269-282. [Available at https://doi.org/10.1016/S0022-2836(65)80096-1]

Warrens, A.N., Jones, M.D., & Lechler, R.I. (1997). Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene*, 186(1): 29-35. [Available at https://doi.org/10.1016/S0378-1119(96)00674-9]

Zhang, Z., Moo-Young, M., & Chisti, Y. (1996). Plasmid stability in recombinant *Saccharomyces cerevisiae*. *Biotechnology Advances*, 14(4): 401-435. [Available at https://doi.org/10.1016/s0734-9750(96)00033-x]

Addendum 1:

DNA Construction

ADDENDUM 1: DNA CONSTRUCTION

Components of DNA construction

Protein-encoding sequence

>Taxadiene synthase (2586 bp)

ATGGCTCAGTTGAGTTTCAATGCTGCCCTTAAAATGAATGCTCTGGGAAATAGAGCCATTCACAACCCGACCAACTGCAGGGCTAAGTCT
GAGGGACAAATGATGTGGGTGTGTTCCAAAAGTGGTAGGACAAGAGTTAAAATGTCTAGAGGGTCAGGAGGTCCAGGTCCTGTTGTTATG
ATGAGCTCTAGCACAGGTACCTCTAAAGTTGTAAGTGAAACATCATCTACAATAGTTGACGACATTCCTAGGCTTAGTGCTAATTATCAT
GGGGATTTATGGCATCATAATGTTATTCAAACATTGGAAACCCCATTCAGAGAAAGCAGTACCTATCAGGAAAGAGCGGACGAGCTGGTC
GTTAAAATCAAAGATATGTTTAATGCTTTAGGTGATGGAGACATCAGTCCTAGTGCCTATGATACTGCTTGGGTCGCAAGAGCTGCCACA
ATATCCAGCGATGGCTCCGAGAAGCCTAGGTTTCCGCAAGCACTTAACTGGGTGTTCAATAACCAGCTACAAGATGGTTCCTGGGGCATA
GAGTCCCACTTTAGTTTATGTGATAGACTTCTAAACACTTCAAATTCTGTGATAGCATTGTCCGTTTGGAAGACCGGACATAGTCAAGTT
GAACAGGGCACAGAATTCATTGCCGAAAATTTAAGGTTATTGAACGAAGAAGATGAGCTTTCCCCTGACTTCGAAATTATATTCCCAGCG
CTGCTTCAGAAAGCAAAAGCCCTGGGTATCAATTTACCTTATGATTTACCGTTCATTAAGTCACTGTCTACAACAAGAGAGGCCCGTTTG
ACGGACGTTTCAGCTGCGGCTGACAATATACCTGCTAACATGTTAAACGCTTTGGAAGGATTAGAAGAAGTCATTGACTGGAATAAAATA
ATGAGATTTCAATCCAAGGATGGAAGCTTCTTGTCCTCACCAGCCTCTACAGCATGCGTCCTTATGAATACTGGAGACGAAAAATGCTTC
ACTTTTCTAAACAATTTATTAGACAAGTTCGGTGGTTGCGTCCCCTGTATGTATTCCATTGATCTGTTAGAACGTTTGAGTCTTGTAGAT
AATATTGAACACTTGGGAATTGGAAGACATTTTAAGCAGGAAATAAAAGGTGCTTTGGACTATGTCTACAGACATTGGTCTGAACGTGGT
ATAGGTTGGGGACGTGATAGCTTAGTACCTGATCTAAACACAACGGCTTTAGGTTTGCGTACGTTGCGTACACACGGATATGATGTGTCT
TCAGATGTTCTTAATAACTTCAAAGATGAGAATGGAAGATTTTTTAGCAGCGCAGGGCAAACACATGTCGAATTAAGATCAGTTGTGAAT
CTATTTAGGGCTTCCGATTTAGCTTTCCCTGATGAAGGAGCCATGGATGACGCTCGTAAGTTCGCTGAACCGTACCTGAGAGACGCTTTG
GCAACCAAAATCTCCACTAATACTAAGCTATTTAAAGAGATAGAATACGTGGTGGAATACCCTTGGCATATGTCCATTCCAAGATTGGAA
GCTAGATCATATATAGATTCCTATGACGACGACTACGTGTGGCAACGTAAAACACTATATAGAATGCCATCCTTGTCCAACTCTAAGTGC
TTGGAGCTGGCAAAACTGGATTTTAACATAGTACAGAGCTTACACCAGGAGGAACTGAAATTATTGACTAGATGGTGGAAAGAAAGTGGC
ATGGCTGACATAAACTTCACAAGACACAGAGTTGCTGAAGTCTACTTCAGTTCCGCAACTTTCGAACCAGAGTATAGCGCGACCAGAATA
GCCTTTACTAAAATTGGTTGTTTGCAGGTTCTTTTCGATGATATGGCAGACATATTTGCTACCCTGGATGAATTAAAATCTTTTACAGAA
GGCGTCAAACGTTGGGACACTTCTCTATTACACGAAATTCCCGAGTGCATGCAAACGTGTTTCAAGGTGTGGTTCAAACTTATGGAAGAA
GTTAATAACGACGTCGTAAAAGTTCAAGGCAGAGATATGCTGGCGCACATTAGAAAACCTTGGGAACTTTATTTCAATTGCTACGTTCAA
GAAAGAGAATGGCTGGAAGCAGGTTACATTCCTACTTTCGAAGAATATTTGAAAACGTATGCAATAAGCGTGGGCTTGGGACCCTGTACA
TTACAGCCGATATTGTTGATGGGTGAACTTGTAAAGGATGATGTAGTGGAAAAGGTCCACTATCCTTCCAATATGTTCGAATTAGTTTCA
TTGTCATGGAGACTTACAAACGATACGAAGACTTACCAGGCAGAGAAAGCTAGAGGTCAACAGGCCTCCGGTATAGCTTGTTATATGAAA
GACAATCCTGGAGCCACTGAGGAAGATGCTATTAAACACATTTGTAGAGTAGTAGATAGAGCTTTAAAAGAGGCCAGTTTTGAATACTTC
AAACCCAGTAACGATATTCCGATGGGTTGTAAAAGCTTTATTTTCAATCTTAGGTTGTGCGTCCAGATTTTCTATAAATTTATTGATGGC
TACGGTATAGCCAACGAAGAGATCAAAGATTATATCAGAAAAGTGTACATTGATCCAATCCAAGTC

>Taxadiene 5α-hydroxylase (1497 bp)

ATGGATGCCCTATACAAATCTACAGTGGCCAAATTTAATGAGGTCACACAGTTAGATTGCAGTACTGAATCTTTCTCTATCGCTCTATCT
GCTATAGCAGGTATACTATTATTGCTACTTCTTTTCAGATCCAAAAGACACTCTTCTTTGAAGTTGCCACCTGGGAAGTTGGGGATCCCC
TTTATTGGCGAATCATTCATTTTCTTAAGGGCTTTAAGATCCAATTCTCTGGAACAATTTTTTGACGAAAGAGTCAAAAAATTTGGTTTG
GTCTTCAAAACTTCCTTGATAGGTCATCCAACAGTAGTTCTTTGTGGCCCTGCGGGCAATAGATTGATACTGTCTAACGAAGAAAAATTA
GTCCAAATGAGCTGGCCAGCACAATTTATGAAACTTATGGGTGAAAATTCTGTCGCAACAAGAAGGGGTGAAGATCATATCGTAATGCGT
TCTGCCCTTGCCGGTTTTTTTGGTCCTGGTGCTTTGCAATCATATATCGGTAAGATGAATACTGAAATTCAGTCCCACATAAACGAAAAG
TGGAAAGGAAAAGACGAAGTAAACGTATTGCCATTGGTTAGAGAACTAGTTTTCAACATATCCGCAATCCTATTTTTTAATATTTACGAT
AAACAGGAACAGGATAGATTACACAAATTACTAGAGACGATCCTGGTGGGTTCATTTGCCTTGCCTATCGATTTGCCTGGTTTCGGTTTC
CATAGAGCATTGCAAGGAAGAGCCAAGTTGAATAAAATCATGTTATCCTTAATTAAAAAGCGTAAGGAGGATTTACAGTCAGGTTCCGCA

ACAGCAACTCAAGACCTGTTATCAGTCCTTTTGACCTTTAGAGACGATAAAGGAACACCCTTGACTAACGACGAAATCCTTGATAACTTT
TCAAGTTTACTACACGCATCCTATGATACGACGACAAGTCCGATGGCGCTTATATTCAAACTGCTGTCTTCCAACCCTGAATGCTATCAG
AAAGTGGTCCAGGAACAACTGGAAATTCTGTCCAATAAAGAAGAGGGAGAGGAGATCACATGGAAAGATTTGAAAGCTATGAAGTACACC
TGGCAAGTGGCTCAGGAAACTTTAAGAATGTTCCCTCCTGTTTTTGGTACTTTTAGAAAAGCTATTACAGATATTCAGTACGATGGTTAT
ACTATACCAAAGGGATGGAAATTATTATGGACTACTTATTCTACACATCCCAAAGATTTGTATTTCAACGAACCTGAAAAGTTTATGCCT
AGCAGATTCGACCAGGAGGGTAAACATGTTGCACCCTACACGTTTCTACCGTTTGGAGGCGGACAAAGGTCCTGTGTCGGATGGGAATTC
TCTAAAATGGAAATCTTGTTGTTCGTACATCATTTTGTTAAAACATTTAGCAGTTATACACCTGTGGATCCCGATGAGAAGATTAGTGGG
GACCCGTTGCCCCCCTTACCTTCCAAGGGATTTTCCATAAAACTATTTCCACGTCCA

>NADPH-cytochrome P450 reductase (2151 bp)

ATGCAAGCAAACAGTAATACAGTCGAGGGCGCATCTCAAGGTAAATCCCTGCTAGATATTAGTAGACTTGATCATATATTTGCTTTGTTG
CTAAACGGAAAAGGTGGTGACTTGGGCGCTATGACTGGTAGTGCTTTGATCCTGACCGAGAATTCACAAAATCTAATGATACTGACTACA
GCATTAGCGGTTCTGGTGGCTTGCGTATTTTTTTTCGTGTGGAGAAGAGGTGGATCAGATACGCAGAAGCCTGCTGTAAGACCCACTCCG
TTGGTGAAAGAAGAAGACGAGGAGGAAGAAGATGATAGTGCCAAAAAAAAAGTAACAATTTTTTTCGGTACTCAAACCGGCACGGCTGAG
GGCTTTGCCAAGGCATTGGCGGAAGAAGCCAAAGCTCGTTACGAGAAAGCAGTATTTAAGGTCGTGGATCTGGATAACTATGCTGCTGAT
GACGAACAATATGAAGAAAAACTTAAAAAAGAAAAGCTAGCTTTTTTTATGCTAGCCACCTACGGTGATGGCGAGCCTACCGACAATGCA
GCCCGTTTTTACAAATGGTTCTTGGAAGGGAAAGAGAGGGAACCTTGGTTGTCCGACCTTACATACGGTGTTTTTGGTTTAGGCAACAGA
CAGTACGAACATTTTAATAAAGTGGCGAAAGCCGTCGATGAGGTTTTGATAGAACAAGGCGCGAAGAGACTTGTACCCGTAGGCTTAGGT
GATGACGATCAGTGCATTGAAGATGATTTTACGGCATGGAGAGAACAAGTGTGGCCTGAGCTGGATCAGTTGTTGAGAGACGAAGATGAC
GAACCTACCTCTGCCACGCCTTATACTGCAGCTATTCCCGAATACAGAGTGGAAATCTACGATTCCGTCGTATCAGTTTATGAAGAAACA
CATGCTCTAAAGCAGAACGGGCAAGCCGTATATGACATTCATCATCCGTGTAGATCAAACGTCGCTGTAAGGAGAGAACTGCACACACCT
CTTTCAGATCGTTCCTGTATTCATCTGGAATTTGATATCAGTGACACCGGTCTGATATATGAGACTGGTGACCATGTTGGCGTTCATACA
GAAAATTCTATAGAAACCGTGGAAGAGGCCGCAAAGTTATTAGGCTATCAATTGGATACCATATTTTCCGTTCATGGTGATAAAGAAGAC
GGCACTCCATTAGGCGGCTCTTCTTTGCCACCACCCTTTCCAGGGCCCTGTACATTAAGGACCGCCTTAGCACGTTATGCCGACCTGTTA
AATCCGCCGAGGAAAGCGGCTTTCTTGGCGTTGGCCGCGCATGCTTCTGATCCTGCCGAAGCTGAAAGACTAAAATTTTTATCCTCCCCA
GCCGGCAAAGATGAATACTCACAGTGGGTGACTGCATCACAAAGAAGTCTTTTAGAAATCATGGCTGAATTTCCGAGCGCAAAGCCTCCG
CTAGGTGTCTTCTTCGCAGCCATTGCACCAAGACTTCAACCAAGATATTATTCAATTTCATCAAGTCCTAGATTTGCCCCTTCCAGAATT
CACGTTACATGCGCATTGGTCTACGGCCCATCTCCGACTGGCAGAATTCATAAGGGTGTTTGTTCTAACTGGATGAAGAACAGTCTTCCT
TCTGAGGAAACACATGATTGCTCTTGGGCGCCTGTTTTTGTTAGACAAAGTAACTTCAAGCTTCCTGCGGACTCAACCACACCGATTGTG
ATGGTGGGACCAGGAACTGGCTTTGCACCTTTCAGAGGCTTCCTACAGGAAAGGGCAAAATTGCAAGAAGCGGGAGAAAAATTAGGTCCA
GCAGTTTTATTTTTCGGTTGTAGGAACAGGCAAATGGATTACATATACGAAGATGAGTTAAAGGGATACGTGGAAAAGGGTATTTTGACC
AATCTGATTGTTGCCTTCAGCAGGGAAGGCGCAACGAAGGAATATGTGCAACACAAAATGCTAGAAAAGGCATCTGACACTTGGTCACTA
ATAGCCCAAGGTGGCTATTTATATGTGTGTGGCGATGCTAAGGGGATGGCCAGAGATGTCCACAGAACTTTACACACTATCGTACAAGAA
CAGGAATCAGTCGACTCCTCAAAGGCTGAATTCTTAGTTAAGAAGCTTCAGATGGACGGTAGATATTTGAGGGATATTTGG

Linker between protein and tag (18 bp)

>Gly6

GGAGGTGGCGGTGGTGGA

Solubility tag

>MBP protein tag (1191 bp)

ATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAA
GGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAA
GTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCAC
GACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGG

GATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAAT
CCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCG
TATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGAT
AATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAA
GCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTA
ACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAG
CTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTG
AAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCC
CAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCC
CAAACTAGGATCACAAAGTAA

Promoter

>*ICL1* promoter (400 bp)

TTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATT
AACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCA
AAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGC
TTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAG
AGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAA

Terminator

>*PRM9* terminator (250 bp)

ACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTT
GTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTG
ACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAAT

Artificial gene sequences

Taxadiene synthase (4445 bp)

TTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATT
AACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCA
AAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGC
TTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAG
AGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGCTCAGTTGAGTTTCAATGCTGCCCTTAAAATGAATGCTCTGGGAAA
TAGAGCCATTCACAACCCGACCAACTGCAGGGCTAAGTCTGAGGGACAAATGATGTGGGTGTGTTCCAAAAGTGGTAGGACAAGAGTTAA
AATGTCTAGAGGGTCAGGAGGTCCAGGTCCTGTTGTTATGATGAGCTCTAGCACAGGTACCTCTAAAGTTGTAAGTGAAACATCATCTAC
AATAGTTGACGACATTCCTAGGCTTAGTGCTAATTATCATGGGGATTTATGGCATCATAATGTTATTCAAACATTGGAAACCCCATTCAG
AGAAAGCAGTACCTATCAGGAAAGAGCGGACGAGCTGGTCGTTAAAATCAAAGATATGTTTAATGCTTTAGGTGATGGAGACATCAGTCC
TAGTGCCTATGATACTGCTTGGGTCGCAAGAGCTGCCACAATATCCAGCGATGGCTCCGAGAAGCCTAGGTTTCCGCAAGCACTTAACTG
GGTGTTCAATAACCAGCTACAAGATGGTTCCTGGGGCATAGAGTCCCACTTTAGTTTATGTGATAGACTTCTAAACACTTCAAATTCTGT
GATAGCATTGTCCGTTTGGAAGACCGGACATAGTCAAGTTGAACAGGGCACAGAATTCATTGCCGAAAATTTAAGGTTATTGAACGAAGA
AGATGAGCTTTCCCCTGACTTCGAAATTATATTCCCAGCGCTGCTTCAGAAAGCAAAAGCCCTGGGTATCAATTTACCTTATGATTTACC
GTTCATTAAGTCACTGTCTACAACAAGAGAGGCCCGTTTGACGGACGTTTCAGCTGCGGCTGACAATATACCTGCTAACATGTTAAACGC
TTTGGAAGGATTAGAAGAAGTCATTGACTGGAATAAAATAATGAGATTTCAATCCAAGGATGGAAGCTTCTTGTCCTCACCAGCCTCTAC
AGCATGCGTCCTTATGAATACTGGAGACGAAAAATGCTTCACTTTTCTAAACAATTTATTAGACAAGTTCGGTGGTTGCGTCCCCTGTAT
GTATTCCATTGATCTGTTAGAACGTTTGAGTCTTGTAGATAATATTGAACACTTGGGAATTGGAAGACATTTTAAGCAGGAAATAAAAGG
TGCTTTGGACTATGTCTACAGACATTGGTCTGAACGTGGTATAGGTTGGGGACGTGATAGCTTAGTACCTGATCTAAACACAACGGCTTT

AGGTTTGCGTACGTTGCGTACACACGGATATGATGTGTCTTCAGATGTTCTTAATAACTTCAAAGATGAGAATGGAAGATTTTTTAGCAG
CGCAGGGCAAACACATGTCGAATTAAGATCAGTTGTGAATCTATTTAGGGCTTCCGATTTAGCTTTCCCTGATGAAGGAGCCATGGATGA
CGCTCGTAAGTTCGCTGAACCGTACCTGAGAGACGCTTTGGCAACCAAAATCTCCACTAATACTAAGCTATTTAAAGAGATAGAATACGT
GGTGGAATACCCTTGGCATATGTCCATTCCAAGATTGGAAGCTAGATCATATATAGATTCCTATGACGACGACTACGTGTGGCAACGTAA
AACACTATATAGAATGCCATCCTTGTCCAACTCTAAGTGCTTGGAGCTGGCAAAACTGGATTTTAACATAGTACAGAGCTTACACCAGGA
GGAACTGAAATTATTGACTAGATGGTGGAAAGAAAGTGGCATGGCTGACATAAACTTCACAAGACACAGAGTTGCTGAAGTCTACTTCAG
TTCCGCAACTTTCGAACCAGAGTATAGCGCGACCAGAATAGCCTTTACTAAAATTGGTTGTTTGCAGGTTCTTTTCGATGATATGGCAGA
CATATTTGCTACCCTGGATGAATTAAAATCTTTTACAGAAGGCGTCAAACGTTGGGACACTTCTCTATTACACGAAATTCCCGAGTGCAT
GCAAACGTGTTTCAAGGTGTGGTTCAAACTTATGGAAGAAGTTAATAACGACGTCGTAAAAGTTCAAGGCAGAGATATGCTGGCGCACAT
TAGAAAACCTTGGGAACTTTATTTCAATTGCTACGTTCAAGAAAGAGAATGGCTGGAAGCAGGTTACATTCCTACTTTCGAAGAATATTT
GAAAACGTATGCAATAAGCGTGGGCTTGGGACCCTGTACATTACAGCCGATATTGTTGATGGGTGAACTTGTAAAGGATGATGTAGTGGA
AAAGGTCCACTATCCTTCCAATATGTTCGAATTAGTTTCATTGTCATGGAGACTTACAAACGATACGAAGACTTACCAGGCAGAGAAAGC
TAGAGGTCAACAGGCCTCCGGTATAGCTTGTTATATGAAAGACAATCCTGGAGCCACTGAGGAAGATGCTATTAAACACATTTGTAGAGT
AGTAGATAGAGCTTTAAAAGAGGCCAGTTTTGAATACTTCAAACCCAGTAACGATATTCCGATGGGTTGTAAAAGCTTTATTTTCAATCT
TAGGTTGTGCGTCCAGATTTTCTATAAATTTATTGATGGCTACGGTATAGCCAACGAAGAGATCAAAGATTATATCAGAAAAGTGTACAT
TGATCCAATCCAAGTCGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGAT
GTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGT
CGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTAC
CGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAA
AGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCT
TTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGG
TAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAA
CGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCA
CATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTC
CAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTC
CGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGT
GAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAA
TGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGG
GAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACC
AGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATA
GTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCA
TTATGCAACGCTTCGGAAAATACGATGTTGAAAAT

Taxadiene 5α-hydroxylase (3356 bp)

TTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATT
AACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCA
AAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGC
TTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAG
AGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGATGCCCTATACAAATCTACAGTGGCCAAATTTAATGAGGTCACACA
GTTAGATTGCAGTACTGAATCTTTCTCTATCGCTCTATCTGCTATAGCAGGTATACTATTATTGCTACTTCTTTTCAGATCCAAAAGACA
CTCTTCTTTGAAGTTGCCACCTGGGAAGTTGGGGATCCCCTTTATTGGCGAATCATTCATTTTCTTAAGGGCTTTAAGATCCAATTCTCT
GGAACAATTTTTTGACGAAAGAGTCAAAAAATTTGGTTTGGTCTTCAAAACTTCCTTGATAGGTCATCCAACAGTAGTTCTTTGTGGCCC
TGCGGGCAATAGATTGATACTGTCTAACGAAGAAAAATTAGTCCAAATGAGCTGGCCAGCACAATTTATGAAACTTATGGGTGAAAATTC
TGTCGCAACAAGAAGGGGTGAAGATCATATCGTAATGCGTTCTGCCCTTGCCGGTTTTTTTGGTCCTGGTGCTTTGCAATCATATATCGG
TAAGATGAATACTGAAATTCAGTCCCACATAAACGAAAAGTGGAAAGGAAAAGACGAAGTAAACGTATTGCCATTGGTTAGAGAACTAGT
TTTCAACATATCCGCAATCCTATTTTTTAATATTTACGATAAACAGGAACAGGATAGATTACACAAATTACTAGAGACGATCCTGGTGGG
TTCATTTGCCTTGCCTATCGATTTGCCTGGTTTCGGTTTCCATAGAGCATTGCAAGGAAGAGCCAAGTTGAATAAAATCATGTTATCCTT

AATTAAAAAGCGTAAGGAGGATTTACAGTCAGGTTCCGCAACAGCAACTCAAGACCTGTTATCAGTCCTTTTGACCTTTAGAGACGATAA
AGGAACACCCTTGACTAACGACGAAATCCTTGATAACTTTTCAAGTTTACTACACGCATCCTATGATACGACGACAAGTCCGATGGCGCT
TATATTCAAACTGCTGTCTTCCAACCCTGAATGCTATCAGAAAGTGGTCCAGGAACAACTGGAAATTCTGTCCAATAAAGAAGAGGGAGA
GGAGATCACATGGAAAGATTTGAAAGCTATGAAGTACACCTGGCAAGTGGCTCAGGAAACTTTAAGAATGTTCCCTCCTGTTTTTGGTAC
TTTTAGAAAAGCTATTACAGATATTCAGTACGATGGTTATACTATACCAAAGGGATGGAAATTATTATGGACTACTTATTCTACACATCC
CAAAGATTTGTATTTCAACGAACCTGAAAAGTTTATGCCTAGCAGATTCGACCAGGAGGGTAAACATGTTGCACCCTACACGTTTCTACC
GTTTGGAGGCGGACAAAGGTCCTGTGTCGGATGGGAATTCTCTAAAATGGAAATCTTGTTGTTCGTACATCATTTTGTTAAAACATTTAG
CAGTTATACACCTGTGGATCCCGATGAGAAGATTAGTGGGGACCCGTTGCCCCCCTTACCTTCCAAGGGATTTTCCATAAAACTATTTCC
ACGTCCAGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGATGTTTTCTGC
ATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTAAGAA
ATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCGACGG
ACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTTTCCA
AGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTTTAAT
TTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGC
GTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATA
CGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGC
TGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGA
CACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTAT
TAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGA
TAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAA
AGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAAC
TGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGT
ATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTG
CCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAAC
GCTTCGGAAAATACGATGTTGAAAAT

NADPH-cytochrome P450 reductase (4010 bp)

TTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATT
AACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCA
AAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGC
TTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAG
AGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGCAAGCAAACAGTAATACAGTCGAGGGCGCATCTCAAGGTAAATCCCT
GCTAGATATTAGTAGACTTGATCATATATTTGCTTTGTTGCTAAACGGAAAAGGTGGTGACTTGGGCGCTATGACTGGTAGTGCTTTGAT
CCTGACCGAGAATTCACAAAATCTAATGATACTGACTACAGCATTAGCGGTTCTGGTGGCTTGCGTATTTTTTTTCGTGTGGAGAAGAGG
TGGATCAGATACGCAGAAGCCTGCTGTAAGACCCACTCCGTTGGTGAAAGAAGAAGACGAGGAGGAAGAAGATGATAGTGCCAAAAAAAA
AGTAACAATTTTTTTCGGTACTCAAACCGGCACGGCTGAGGGCTTTGCCAAGGCATTGGCGGAAGAAGCCAAAGCTCGTTACGAGAAAGC
AGTATTTAAGGTCGTGGATCTGGATAACTATGCTGCTGATGACGAACAATATGAAGAAAAACTTAAAAAAGAAAAGCTAGCTTTTTTTAT
GCTAGCCACCTACGGTGATGGCGAGCCTACCGACAATGCAGCCCGTTTTTACAAATGGTTCTTGGAAGGGAAAGAGAGGGAACCTTGGTT
GTCCGACCTTACATACGGTGTTTTTGGTTTAGGCAACAGACAGTACGAACATTTTAATAAAGTGGCGAAAGCCGTCGATGAGGTTTTGAT
AGAACAAGGCGCGAAGAGACTTGTACCCGTAGGCTTAGGTGATGACGATCAGTGCATTGAAGATGATTTTACGGCATGGAGAGAACAAGT
GTGGCCTGAGCTGGATCAGTTGTTGAGAGACGAAGATGACGAACCTACCTCTGCCACGCCTTATACTGCAGCTATTCCCGAATACAGAGT
GGAAATCTACGATTCCGTCGTATCAGTTTATGAAGAAACACATGCTCTAAAGCAGAACGGGCAAGCCGTATATGACATTCATCATCCGTG
TAGATCAAACGTCGCTGTAAGGAGAGAACTGCACACACCTCTTTCAGATCGTTCCTGTATTCATCTGGAATTTGATATCAGTGACACCGG
TCTGATATATGAGACTGGTGACCATGTTGGCGTTCATACAGAAAATTCTATAGAAACCGTGGAAGAGGCCGCAAAGTTATTAGGCTATCA
ATTGGATACCATATTTTCCGTTCATGGTGATAAAGAAGACGGCACTCCATTAGGCGGCTCTTCTTTGCCACCACCCTTTCCAGGGCCCTG
TACATTAAGGACCGCCTTAGCACGTTATGCCGACCTGTTAAATCCGCCGAGGAAAGCGGCTTTCTTGGCGTTGGCCGCGCATGCTTCTGA
TCCTGCCGAAGCTGAAAGACTAAAATTTTTATCCTCCCCAGCCGGCAAAGATGAATACTCACAGTGGGTGACTGCATCACAAAGAAGTCT

TTTAGAAATCATGGCTGAATTTCCGAGCGCAAAGCCTCCGCTAGGTGTCTTCTTCGCAGCCATTGCACCAAGACTTCAACCAAGATATTA
TTCAATTTCATCAAGTCCTAGATTTGCCCCTTCCAGAATTCACGTTACATGCGCATTGGTCTACGGCCCATCTCCGACTGGCAGAATTCA
TAAGGGTGTTTGTTCTAACTGGATGAAGAACAGTCTTCCTTCTGAGGAAACACATGATTGCTCTTGGGCGCCTGTTTTTGTTAGACAAAG
TAACTTCAAGCTTCCTGCGGACTCAACCACACCGATTGTGATGGTGGGACCAGGAACTGGCTTTGCACCTTTCAGAGGCTTCCTACAGGA
AAGGGCAAAATTGCAAGAAGCGGGAGAAAAATTAGGTCCAGCAGTTTTATTTTTCGGTTGTAGGAACAGGCAAATGGATTACATATACGA
AGATGAGTTAAAGGGATACGTGGAAAAGGGTATTTTGACCAATCTGATTGTTGCCTTCAGCAGGGAAGGCGCAACGAAGGAATATGTGCA
ACACAAAATGCTAGAAAAGGCATCTGACACTTGGTCACTAATAGCCCAAGGTGGCTATTTATATGTGTGTGGCGATGCTAAGGGGATGGC
CAGAGATGTCCACAGAACTTTACACACTATCGTACAAGAACAGGAATCAGTCGACTCCTCAAAGGCTGAATTCTTAGTTAAGAAGCTTCA
GATGGACGGTAGATATTTGAGGGATATTTGGGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGC
CTTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAA
CGGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCC
TCAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGA
GATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGAT
TGCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGA
ACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGC
CTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTT
AATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGG
TCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATT
TGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGA
AGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGC
AGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAAT
TAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTA
GCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATAT
ACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTT
ATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAAT

DNA between cassettes (Cas9 target loci and attached PAM sequence)

>805a target sequence

TTATTTGAATGATATTTAGTTGG

>1206a target sequence

CGAACATTTTTCCATGCGCTTGG

>607c target sequence

CTATTTTTGCTTTCTGCACATGG

DNA fragments for Gibson assembly

A (amplified from genomic DNA of S. cerevisiae BY4742) (949 bp)

TTCCTAAGCCTCCCTCACCATAAATTACCTTTATTACTTGCATGACTATTATTAGCAGAGCATGTAGTATGGGACTCAAGACCGATATGA
TACACACCAAAGACGTAGGCACCGGCGATTAAATCAAAGGCTCCGATAGCCGAAAAGTGAGAAGAAAAAAAAAAGGAAAAAAAGGAATTGT
CCTAATGAGCGGTGTGGCCGACTTGCCATAATATCAGTTAGGGCTACTATCAATGTTTTATCTACGTTGGAGTAAGATCGTTTATCACTT
CCATATTTGGACCAAATGAAAAGTTCAATCGGCCAAGTATTTCATGGATGGAATGACGTTTGGTAAGGAAGTGCTTTTTCTTTTTCCACA
TATTTTCCCTTTCTCTCGGGGAAATTTTGTTTCTAAACATAAAAAATAAAGCAACAGCAAAAAAGAGGGTCTGTCCAGCGAATAAGAAGA

AAACCTCCTTTTCGGCTTTTGAAGATAGGTTGCAGTTGTCTGCGGGCACAAAATGGGCAATTTTTTTAATACTTTTTACGTATGAGACAA
GATTTTTTTCGCCAATTATATCGCATGAAGAATAACCAGAGTTTTTCTCCGAACGTTAAGGGAGTTGAAGTAAAAATAAAGAAAGGACCA
AATGAGAATGGGTATGCTTGGTCTTAGTCTTCGAATCAAATTCTGCTTCCCTGTTCATGGCAACGTCACCTCAATTATTTGGAAAGGGGG
GGTTTTCCGACTTTATTTGAGATGACTTGAGATGTGTGTCAATGCTAGTATTTTGGAGATTAATCTCAGTACAAAACAATATTAAAAAGA
GGTGAATTATTTTTCCCCCCTTATTTTTTTTTTGTTAGAATTGATCCAAATGTAAATAAACAATCACAAGGAAAAAAAAAAAAAAAAAA
AATAGCCGCCATGACCCCGGATCGTCGGTTGTGATACGGTCAGGGTAGC

R1 (1875 bp)

CATGACCCCGGATCGTCGGTTGTGATACGGTCAGGGTAGCTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTC
ATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAG
TCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATC
GGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTA
TCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGCTCAGT
TGAGTTTCAATGCTGCCCTTAAAATGAATGCTCTGGGAAATAGAGCCATTCACAACCCGACCAACTGCAGGGCTAAGTCTGAGGGACAAA
TGATGTGGGTGTGTTCCAAAAGTGGTAGGACAAGAGTTAAAATGTCTAGAGGGTCAGGAGGTCCAGGTCCTGTTGTTATGATGAGCTCTA
GCACAGGTACCTCTAAAGTTGTAAGTGAAACATCATCTACAATAGTTGACGACATTCCTAGGCTTAGTGCTAATTATCATGGGGATTTAT
GGCATCATAATGTTATTCAAACATTGGAAACCCCATTCAGAGAAAGCAGTACCTATCAGGAAAGAGCGGACGAGCTGGTCGTTAAAATCA
AAGATATGTTTAATGCTTTAGGTGATGGAGACATCAGTCCTAGTGCCTATGATACTGCTTGGGTCGCAAGAGCTGCCACAATATCCAGCG
ATGGCTCCGAGAAGCCTAGGTTTCCGCAAGCACTTAACTGGGTGTTCAATAACCAGCTACAAGATGGTTCCTGGGGCATAGAGTCCCACT
TTAGTTTATGTGATAGACTTCTAAACACTTCAAATTCTGTGATAGCATTGTCCGTTTGGAAGACCGGACATAGTCAAGTTGAACAGGGCA
CAGAATTCATTGCCGAAAATTTAAGGTTATTGAACGAAGAAGATGAGCTTTCCCCTGACTTCGAAATTATATTCCCAGCGCTGCTTCAGA
AAGCAAAAGCCCTGGGTATCAATTTACCTTATGATTTACCGTTCATTAAGTCACTGTCTACAACAAGAGAGGCCCGTTTGACGGACGTTT
CAGCTGCGGCTGACAATATACCTGCTAACATGTTAAACGCTTTGGAAGGATTAGAAGAAGTCATTGACTGGAATAAAATAATGAGATTTC
AATCCAAGGATGGAAGCTTCTTGTCCTCACCAGCCTCTACAGCATGCGTCCTTATGAATACTGGAGACGAAAAATGCTTCACTTTTCTAA
ACAATTTATTAGACAAGTTCGGTGGTTGCGTCCCCTGTATGTATTCCATTGATCTGTTAGAACGTTTGAGTCTTGTAGATAATATTGAAC
ACTTGGGAATTGGAAGACATTTTAAGCAGGAAATAAAAGGTGCTTTGGACTATGTCTACAGACATTGGTCTGAACGTGGTATAGGTTGGG
GACGTGATAGCTTAGTACCTGATCTAAACACAACGGCTTTAGGTTTGCGTACGTTGCGTACACACGGATATGATGTGTCTTCAGATGTTC
TTAATAACTTCAAAGATGAGAATGGAAGATTTTTTAGCAGCGCAGGGCAAACACATGTCGAATTAAGATCAGTTGTGAATCTATTTAGGG
CTTCCGATTTAGCTTTCCCTGATGAAGGAGCCATGGATGACGCTCGTAAGTTCGCTGAACCGTACCTGAGAGACG

R2 (1875 bp)

GTAAGTTCGCTGAACCGTACCTGAGAGACGCTTTGGCAACCAAAATCTCCACTAATACTAAGCTATTTAAAGAGATAGAATACGTGGTGG
AATACCCTTGGCATATGTCCATTCCAAGATTGGAAGCTAGATCATATATAGATTCCTATGACGACGACTACGTGTGGCAACGTAAAACAC
TATATAGAATGCCATCCTTGTCCAACTCTAAGTGCTTGGAGCTGGCAAAACTGGATTTTAACATAGTACAGAGCTTACACCAGGAGGAAC
TGAAATTATTGACTAGATGGTGGAAAGAAAGTGGCATGGCTGACATAAACTTCACAAGACACAGAGTTGCTGAAGTCTACTTCAGTTCCG
CAACTTTCGAACCAGAGTATAGCGCGACCAGAATAGCCTTTACTAAAATTGGTTGTTTGCAGGTTCTTTTCGATGATATGGCAGACATAT
TTGCTACCCTGGATGAATTAAAATCTTTTACAGAAGGCGTCAAACGTTGGGACACTTCTCTATTACACGAAATTCCCGAGTGCATGCAAA
CGTGTTTCAAGGTGTGGTTCAAACTTATGGAAGAAGTTAATAACGACGTCGTAAAAGTTCAAGGCAGAGATATGCTGGCGCACATTAGAA
AACCTTGGGAACTTTATTTCAATTGCTACGTTCAAGAAAGAGAATGGCTGGAAGCAGGTTACATTCCTACTTTCGAAGAATATTTGAAAA
CGTATGCAATAAGCGTGGGCTTGGGACCCTGTACATTACAGCCGATATTGTTGATGGGTGAACTTGTAAAGGATGATGTAGTGGAAAAGG
TCCACTATCCTTCCAATATGTTCGAATTAGTTTCATTGTCATGGAGACTTACAAACGATACGAAGACTTACCAGGCAGAGAAAGCTAGAG
GTCAACAGGCCTCCGGTATAGCTTGTTATATGAAAGACAATCCTGGAGCCACTGAGGAAGATGCTATTAAACACATTTGTAGAGTAGTAG
ATAGAGCTTTAAAAGAGGCCAGTTTTGAATACTTCAAACCCAGTAACGATATTCCGATGGGTTGTAAAAGCTTTATTTTCAATCTTAGGT
TGTGCGTCCAGATTTTCTATAAATTTATTGATGGCTACGGTATAGCCAACGAAGAGATCAAAGATTATATCAGAAAAGTGTACATTGATC
CAATCCAAGTCGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGATGTTTT
CTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTA

AGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCG
ACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTT
TCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTT
TAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGT
CTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCA
AATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTA

R3 (1580 bp)

CGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCA
CATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTC
CAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTC
CGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGT
GAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAA
TGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGG
GAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACC
AGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATA
GTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCA
TTATGCAACGCTTCGGAAAATACGATGTTGAAAATTTATTTGAATGATATTTAGTTGGCGAACATTTTTCCATGCGCTTGGTTCCATTCA
TCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATTAACACTCAA
CGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCAAAAGAGAAA
AGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGCTTCACTGGC
AAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAGAGAAAACTC
TTAGCATAACATAACAAAAAGTCAACGAAAAATGGATGCCCTATACAAATCTACAGTGGCCAAATTTAATGAGGTCACACAGTTAGATTG
CAGTACTGAATCTTTCTCTATCGCTCTATCTGCTATAGCAGGTATACTATTATTGCTACTTCTTTTCAGATCCAAAAGACACTCTTCTTT
GAAGTTGCCACCTGGGAAGTTGGGGATCCCCTTTATTGGCGAATCATTCATTTTCTTAAGGGCTTTAAGATCCAATTCTCTGGAACAATT
TTTTGACGAAAGAGTCAAAAAATTTGGTTTGGTCTTCAAAACTTCCTTGA

R4 (1580 bp)

AATTTGGTTTGGTCTTCAAAACTTCCTTGATAGGTCATCCAACAGTAGTTCTTTGTGGCCCTGCGGGCAATAGATTGATACTGTCTAACG
AAGAAAAATTAGTCCAAATGAGCTGGCCAGCACAATTTATGAAACTTATGGGTGAAAATTCTGTCGCAACAAGAAGGGGTGAAGATCATA
TCGTAATGCGTTCTGCCCTTGCCGGTTTTTTTGGTCCTGGTGCTTTGCAATCATATATCGGTAAGATGAATACTGAAATTCAGTCCCACA
TAAACGAAAAAGTGGAAAGGAAAAGACGAAGTAAACGTATTGCCATTGGTTAGAGAACTAGTTTTCAACATATCCGCAATCCTATTTTTTA
ATATTTACGATAAACAGGAACAGGATAGATTACACAAATTACTAGAGACGATCCTGGTGGGTTCATTTGCCTTGCCTATCGATTTGCCTG
GTTTCGGTTTCCATAGAGCATTGCAAGGAAGAGCCAAGTTGAATAAAATCATGTTATCCTTAATTAAAAAGCGTAAGGAGGATTTACAGT
CAGGTTCCGCAACAGCAACTCAAGACCTGTTATCAGTCCTTTTGACCTTTAGAGACGATAAAGGAACACCCTTGACTAACGACGAAATCC
TTGATAACTTTTCAAGTTTACTACACGCATCCTATGATACGACGACAAGTCCGATGGCGCTTATATTCAAACTGCTGTCTTCCAACCCTG
AATGCTATCAGAAAGTGGTCCAGGAACAACTGGAAATTCTGTCCAATAAAGAAGAGGGAGAGGAGATCACATGGAAAGATTTGAAAGCTA
TGAAGTACACCTGGCAAGTGGCTCAGGAAACTTTAAGAATGTTCCCTCCTGTTTTTGGTACTTTTAGAAAAGCTATTACAGATATTCAGT
ACGATGGTTATACTATACCAAAGGGATGGAAATTATTATGGACTACTTATTCTACACATCCCAAAGATTTGTATTTCAACGAACCTGAAA
AGTTTATGCCTAGCAGATTCGACCAGGAGGGTAAACATGTTGCACCCTACACGTTTCTACCGTTTGGAGGCGGACAAAGGTCCTGTGTCG
GATGGGAATTCTCTAAAATGGAAATCTTGTTGTTCGTACATCATTTTGTTAAAACATTTAGCAGTTATACACCTGTGGATCCCGATGAGA
AGATTAGTGGGGACCCGTTGCCCCCCTTACCTTCCAAGGGATTTTCCATAAAACTATTTCCACGTCCAGGAGGTGGCGGTGGTGGAATGA
AAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTA
AACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCA
CCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCACGACA
GGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGAC

R5 (1580 bp)

AGTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAA
CTTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATT
CCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCT
GCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTG
ACGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACT
GCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGC
CAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAAC
TACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCC
AAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCC
GTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACA
GAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTA
GTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACA
AAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATCTATTTTTGCTTTCTGCACATGG
TTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATT
AACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCA
AAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGC
TTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAG
AGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGCAAGCAA

R6 (1875 bp)

CTTTTCACACCCAAATACCTAACAATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGCAAGCAAACAGTAATAC
AGTCGAGGGCGCATCTCAAGGTAAATCCCTGCTAGATATTAGTAGACTTGATCATATATTTGCTTTGTTGCTAAACGGAAAAGGTGGTGA
CTTGGGCGCTATGACTGGTAGTGCTTTGATCCTGACCGAGAATTCACAAAATCTAATGATACTGACTACAGCATTAGCGGTTCTGGTGGC
TTGCGTATTTTTTTTCGTGTGGAGAAGAGGTGGATCAGATACGCAGAAGCCTGCTGTAAGACCCACTCCGTTGGTGAAAGAAGAAGACGA
GGAGGAAGAAGATGATAGTGCCAAAAAAAAAAGTAACAATTTTTTTCGGTACTCAAACCGGCACGGCTGAGGGCTTTGCCAAGGCATTGGC
GGAAGAAGCCAAAGCTCGTTACGAGAAAGCAGTATTTAAGGTCGTGGATCTGGATAACTATGCTGCTGATGACGAACAATATGAAGAAAA
ACTTAAAAAAGAAAAGCTAGCTTTTTTTATGCTAGCCACCTACGGTGATGGCGAGCCTACCGACAATGCAGCCCGTTTTTACAAATGGTT
CTTGGAAGGGAAAGAGAGGGAACCTTGGTTGTCCGACCTTACATACGGTGTTTTTGGTTTAGGCAACAGACAGTACGAACATTTTAATAA
AGTGGCGAAAGCCGTCGATGAGGTTTTGATAGAACAAGGCGCGAAGAGACTTGTACCCGTAGGCTTAGGTGATGACGATCAGTGCATTGA
AGATGATTTTACGGCATGGAGAGAACAAGTGTGGCCTGAGCTGGATCAGTTGTTGAGAGACGAAGATGACGAACCTACCTCTGCCACGCC
TTATACTGCAGCTATTCCCGAATACAGAGTGGAAATCTACGATTCCGTCGTATCAGTTTATGAAGAAACACATGCTCTAAAGCAGAACGG
GCAAGCCGTATATGACATTCATCATCCGTGTAGATCAAACGTCGCTGTAAGGAGAGAACTGCACACACCTCTTTCAGATCGTTCCTGTAT
TCATCTGGAATTTGATATCAGTGACACCGGTCTGATATATGAGACTGGTGACCATGTTGGCGTTCATACAGAAAATTCTATAGAAACCGT
GGAAGAGGCCGCAAAGTTATTAGGCTATCAATTGGATACCATATTTTCCGTTCATGGTGATAAAGAAGACGGCACTCCATTAGGCGGCTC
TTCTTTGCCACCACCCTTTCCAGGGCCCTGTACATTAAGGACCGCCTTAGCACGTTATGCCGACCTGTTAAATCCGCCGAGGAAAGCGGC
TTTCTTGGCGTTGGCCGCGCATGCTTCTGATCCTGCCGAAGCTGAAAGACTAAAATTTTTATCCTCCCCAGCCGGCAAAGATGAATACTC
ACAGTGGGTGACTGCATCACAAAGAAGTCTTTTAGAAATCATGGCTGAATTTCCGAGCGCAAAGCCTCCGCTAGGTGTCTTCTTCGCAGC
CATTGCACCAAGACTTCAACCAAGATATTATTCAATTTCATCAAGTCCTAGATTTGCCCCTTCCAGAATTCACGTTACATGCGCATTGGT
CTACGGCCCATCTCCGACTGGCAGAATTCATAAGGGTGTTTGTTCTAACTGGATGAAGAACAGTCTTCCTTCTGAGGAAACACATGATTG
CTCTTGGGCGCCTGTTTTTGTTAGACAAAGTAACTTCAAGCTTCCTGCGGACTCAACCACACCGATTGTGATGGTGGGACCAGGAACTGG
CTTTGCACCTTTCAGAGGCTTCCTACAGGAAAGGGCAAAATTGCAAGAAGCGGGAGAAAAATTAGGTCCAGCAGT

R7 (1875 bp)

AGAAGCGGGAGAAAAATTAGGTCCAGCAGTTTTATTTTTCGGTTGTAGGAACAGGCAAATGGATTACATATACGAAGATGAGTTAAAGGG
ATACGTGGAAAAGGGTATTTTGACCAATCTGATTGTTGCCTTCAGCAGGGAAGGCGCAACGAAGGAATATGTGCAACACAAAATGCTAGA
AAAGGCATCTGACACTTGGTCACTAATAGCCCAAGGTGGCTATTTATATGTGTGTGGCGATGCTAAGGGGATGGCCAGAGATGTCCACAG
AACTTTACACACTATCGTACAAGAACAGGAATCAGTCGACTCCTCAAAGGCTGAATTCTTAGTTAAGAAGCTTCAGATGGACGGTAGATA
TTTGAGGGATATTTGGGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGAT
GTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGT
CGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTAC
CGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAA
AGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCT
TTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGG
TAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAA
CGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCA
CATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTC
CAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTC
CGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGT
GAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAA
TGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGG
GAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACC
AGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATA
GTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCA
TTATGCAACGCTTCGGAAAATACGATGTTGAAAATCCCTGGTCAAACTTCAGAACTAAAAAAATAATAAGGAAGA

B (amplified from genomic DNA of S. cerevisiae BY4742) (920 bp)

CCCTGGTCAAACTTCAGAACTAAAAAAATAATAAGGAAGAAAAAAATAGCTAATTTTTCCGGCAGAAAGATTTTCGCTACCCGAAAGTTT
TTCCGGCAAGCTAAATGGAAAAAGGAAAGATTATTGAAAGAGAAAGAAAGAAAAAAAAAAAATGTACACCCAGACATCGGGCTTCCACAAT
TTCGGCTCTATTGTTTTCCATCTCTCGCAACGGCGGGATTCCTCTATGGCGTGTGATGTCTGTATCTGTTACTTAATCCAGAAACTGGCA
CTTGACCCAACTCTGCCACGTGGGTCGTTTTGCCATCGACAGATTGGGAGATTTTCATAGTAGAATTCAGCATGATAGCTACGTAAATGT
GTTCCGCACCGTCACAAAGTGTTTTCTACTGTTCTTTCTTCTTTCGTTCATTCAGTTGAGTTGAGTGAGTGCTTTGTTCAATGGATCTTA
GCTAAAATGCATATTTTTTCTCTTGGTAAATGAATGCTTGTGATGTCTTCCAAGTGATTTCCTTTCCTTCCCATATGATGCTAGGTACCT
TTAGTGTCTTCCTAAAAAAAAAAAAAGGCTCGCCATCAAAACGATATTCGTTGGCTTTTTTTTCTGAATTATAAATACTCTTTGGTAACT
TTTCATTTCCAAGAACCTCTTTTTTCCAGTTATATCATGGTCCCCTTTCAAAGTTATTCTCTACTCTTTTTCATATTCATTCTTTTTCAT
CCTTTGGTTTTTTATTCTTAACTTGTTTATTATTCTCTCTTGTTTCTATTTACAAGACACCAATCAAAACAAATAAAACATCATCACAAT
GTCTAGATTAGAAAGATTGACCTCATTAAACGTTGTTGCTGGTTCTGACTTGAGAAGAACCTCCATCATTGGTACCATCGGTCCAAAGAC
CAACAACCCAGAAACCTTGG

V1 (Amplified from pGEM®-T Easy plasmid vector) (3064 bp)

TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATA
GCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC
CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCA
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTC

AGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAAT
CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTT
CATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATA
ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGG
GGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
AAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAA
TCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCA
ACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTA
AAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAG
CGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGG
GCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTC
ACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTAATTAAGTCAACCAAGAACGTC
AGGC

V2 (Amplified from pGEM®-T Easy plasmid vector) (3063 bp) [Not used for final method]

CGTCTTTTATGTCGTATTTGCCGAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGC
ATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAC
TGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCT
CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT
TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC
AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAA
GCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTAT
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTG
GTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTA
AATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCG
CCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAG
GCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGT

GCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGC
GACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT
CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA
TACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATC
AGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACT
CCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCC
GTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGA
AAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTAC
AGGGCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGA
CTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTTGCTTGCATTTTTCGTTGA
CTT

V3 (Amplified from pGEM®-T Easy plasmid vector) (3062 bp)

TGTTAGGTATTTGGGTGTGAAAAGAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATG
CATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA
ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCA
CTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT
CGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCC
GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTG
TGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCAT
CAGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGC
AAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGC

CGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG
AAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTA
CAGGGCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAG
GGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACG
ACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCCAAGGTTTCTGGGTTGTT
GG

V1.5 (Amplified from pGEM®-T Easy plasmid vector) (3060 bp)

TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATA
GCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTC
CACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCA
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTC
AGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
ACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAAT
CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTT
CATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCT
CCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATA
ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGG
GGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAC
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAA
TAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
AAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAA
TCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCA
ACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTA
AAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAG
CGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGG
GCGCGTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGG
ATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTC
ACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTTTGCTTGCATTTTTCGTTGACTT

1000 bp homologous arms for the three inserted target sequences

Inserted target sequence 805a

>Left homologous region

AACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTAC
ATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGG
TGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATT
TAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTT
ACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAA
AGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTA
TGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTC

AGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAG
GATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCA
TTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTC
GTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATTTATTTG
AATGATATTT

>Right homologous region (not applicable if inserted 1206a target sequence is used first)

AGTTGGCGAACATTTTTCCATGCGCTTGGTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAA
TCAAAACTGAAGCCAATCACCACAAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGC
ATCCCAGTATGTGACTCAATATTGGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAA
AACGTTTCTTTATTATATAAAGGAGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTC
TTTTCACACCCAAATACCTAACAATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGATGCCCTATACAAATCT
ACAGTGGCCAAATTTAATGAGGTCACACAGTTAGATTGCAGTACTGAATCTTTCTCTATCGCTCTATCTGCTATAGCAGGTATACTATTA
TTGCTACTTCTTTTCAGATCCAAAAGACACTCTTCTTTGAAGTTGCCACCTGGGAAGTTGGGGATCCCCTTTATTGGCGAATCATTCATT
TTCTTAAGGGCTTTAAGATCCAATTCTCTGGAACAATTTTTTGACGAAAGAGTCAAAAAATTTGGTTTGGTCTTCAAAACTTCCTTGATA
GGTCATCCAACAGTAGTTCTTTGTGGCCCTGCGGGCAATAGATTGATACTGTCTAACGAAGAAAAATTAGTCCAAATGAGCTGGCCAGCA
CAATTTATGAAACTTATGGGTGAAAATTCTGTCGCAACAAGAAGGGGTGAAGATCATATCGTAATGCGTTCTGCCCTTGCCGGTTTTTTT
GGTCCTGGTGCTTTGCAATCATATATCGGTAAGATGAATACTGAAATTCAGTCCCACATAAACGAAAAGTGGAAAGGAAAAGACGAAGTA
AACGTATTGC

Inserted target sequence 1206a

>Left homologous region (not applicable if inserted 805a target sequence is used first)

TAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATG
GAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCT
TGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGA
CTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCT
CCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTAT
TGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATC
CTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAA
CTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACG
GGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGAT
ATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGC
TCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATTTATTTGAATGATATTTAGTTGGCGAACAT
TTTTCCATGC

>Right homologous region

GCTTGGTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCAC
AAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATT
GGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGG
AGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACA
ATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGATGCCCTATACAAATCTACAGTGGCCAAATTTAATGAGGT
CACACAGTTAGATTGCAGTACTGAATCTTTCTCTATCGCTCTATCTGCTATAGCAGGTATACTATTATTGCTACTTCTTTTCAGATCCAA
AAGACACTCTTCTTTGAAGTTGCCACCTGGGAAGTTGGGGATCCCCTTTATTGGCGAATCATTCATTTTCTTAAGGGCTTTAAGATCCAA
TTCTCTGGAACAATTTTTTGACGAAAGAGTCAAAAAATTTGGTTTGGTCTTCAAAACTTCCTTGATAGGTCATCCAACAGTAGTTCTTTG
TGGCCCTGCGGGCAATAGATTGATACTGTCTAACGAAGAAAAATTAGTCCAAATGAGCTGGCCAGCACAATTTATGAAACTTATGGGTGA
AAATTCTGTCGCAACAAGAAGGGGTGAAGATCATATCGTAATGCGTTCTGCCCTTGCCGGTTTTTTTGGTCCTGGTGCTTTGCAATCATA

TATCGGTAAGATGAATACTGAAATTCAGTCCCACATAAACGAAAAGTGGAAAGGAAAAGACGAAGTAAACGTATTGCCATTGGTTAGAGA
ACTAGTTTTC

Inserted target sequence 607c

>Left homologous region

AACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTAC
ATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGG
TGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATT
TAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTT
ACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAA
AGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTA
TGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTC
AGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAG
GATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCA
TTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTC
GTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATCTATTTT
TGCTTTCTGC

>Right homologous region

ACATGGTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCAC
AAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATT
GGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGG
AGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACA
ATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGCAAGCAAACAGTAATACAGTCGAGGGCGCATCTCAAGGTAA
ATCCCTGCTAGATATTAGTAGACTTGATCATATATTTGCTTTGTTGCTAAACGGAAAAGGTGGTGACTTGGGCGCTATGACTGGTAGTGC
TTTGATCCTGACCGAGAATTCACAAAATCTAATGATACTGACTACAGCATTAGCGGTTCTGGTGGCTTGCGTATTTTTTTTCGTGTGGAG
AAGAGGTGGATCAGATACGCAGAAGCCTGCTGTAAGACCCACTCCGTTGGTGAAAGAAGAAGACGAGGAGGAAGAAGATGATAGTGCCAA
AAAAAAAGTAACAATTTTTTTCGGTACTCAAACCGGCACGGCTGAGGGCTTTGCCAAGGCATTGGCGGAAGAAGCCAAAGCTCGTTACGA
GAAAGCAGTATTTAAGGTCGTGGATCTGGATAACTATGCTGCTGATGACGAACAATATGAAGAAAAACTTAAAAAAGAAAAGCTAGCTTT
TTTTATGCTAGCCACCTACGGTGATGGCGAGCCTACCGACAATGCAGCCCGTTTTTACAAATGGTTCTTGGAAGGGAAAGAGAGGGAACC
TTGGTTGTCC

Final DNA sequence

The sequence of DNA if assembled correctly and without mutations. Presented in full for future applications using the three inserted target sequences.

Legend

Flanking regions

30 bp overlap (highlighted over flanking regions highlight)

80 bp overlap

FW3 and RV5 on fragment 5 (highlighted over 80 bp overlap highlight)

805a target sequence

1206a target sequence

607c target sequence

DNA sequence

TTCCTAAGCCTCCCTCACCATAAATTACCTTTATTACTTGCATGACTATTATTAGCAGAGCATGTAGTATGGGACTCAAGACCGATATGA
TACACACCAAAGACGTAGGCACCGGCGATTAAATCAAAGGCTCCGATAGCCGAAAAGTGAGAAGAAAAAAAAAAGGAAAAAAAGGAATTGT
CCTAATGAGCGGTGTGGCCGACTTGCCATAATATCAGTTAGGGCTACTATCAATGTTTTATCTACGTTGGAGTAAGATCGTTTATCACTT
CCATATTTGGACCAAATGAAAAGTTCAATCGGCCAAGTATTTCATGGATGGAATGACGTTTGGTAAGGAAGTGCTTTTTCTTTTTCCACA
TATTTTCCCTTTCTCTCGGGGAAATTTTGTTTCTAAACATAAAAAAATAAAGCAACAGCAAAAAAGAGGGTCTGTCCAGCGAATAAGAAGA
AAACCTCCTTTTCGGCTTTTGAAGATAGGTTGCAGTTGTCTGCGGGCACAAAATGGGCAATTTTTTTAATACTTTTTACGTATGAGACAA
GATTTTTTTCGCCAATTATATCGCATGAAGAATAACCAGAGTTTTTCTCCGAACGTTAAGGGAGTTGAAGTAAAAATAAAGAAAGGACCA
AATGAGAATGGGTATGCTTGGTCTTAGTCTTCGAATCAAATTCTGCTTCCCTGTTCATGGCAACGTCACCTCAATTATTTGGAAAGGGGG
GGTTTTCCGACTTTATTTGAGATGACTTGAGATGTGTGTCAATGCTAGTATTTTGGAGATTAATCTCAGTACAAAACAATATTAAAAAGA
GGTGAATTATTTTTCCCCCCTTATTTTTTTTTTGTTAGAATTGATCCAAATGTAAATAAACAATCACAAGGAAAAAAAAAAAAAAAAAAA
AATAGCCGCCATGACCCCGGATCGTCGGTTGTGATACGGTCAGGGTAGCTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTT
TTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAA
TATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACC
AGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAAT
TTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAA
TGGCTCAGTTGAGTTTCAATGCTGCCCTTAAAATGAATGCTCTGGGAAATAGAGCCATTCACAACCCGACCAACTGCAGGGCTAAGTCTG
AGGGACAAATGATGTGGGTGTGTTCCAAAAGTGGTAGGACAAGAGTTAAAATGTCTAGAGGGTCAGGAGGTCCAGGTCCTGTTGTTATGA
TGAGCTCTAGCACAGGTACCTCTAAAGTTGTAAGTGAAACATCATCTACAATAGTTGACGACATTCCTAGGCTTAGTGCTAATTATCATG
GGGATTTATGGCATCATAATGTTATTCAAACATTGGAAACCCCATTCAGAGAAAGCAGTACCTATCAGGAAAGAGCGGACGAGCTGGTCG
TTAAAATCAAAGATATGTTTAATGCTTTAGGTGATGGAGACATCAGTCCTAGTGCCTATGATACTGCTTGGGTCGCAAGAGCTGCCACAA
TATCCAGCGATGGCTCCGAGAAGCCTAGGTTTCCGCAAGCACTTAACTGGGTGTTCAATAACCAGCTACAAGATGGTTCCTGGGGCATAG
AGTCCCACTTTAGTTTATGTGATAGACTTCTAAACACTTCAAATTCTGTGATAGCATTGTCCGTTTGGAAGACCGGACATAGTCAAGTTG
AACAGGGCACAGAATTCATTGCCGAAAATTTAAGGTTATTGAACGAAGAAGATGAGCTTTCCCCTGACTTCGAAATTATATTCCCAGCGC
TGCTTCAGAAAGCAAAAGCCCTGGGTATCAATTTACCTTATGATTTACCGTTCATTAAGTCACTGTCTACAACAAGAGAGGCCCGTTTGA
CGGACGTTTCAGCTGCGGCTGACAATATACCTGCTAACATGTTAAACGCTTTGGAAGGATTAGAAGAAGTCATTGACTGGAATAAAATAA
TGAGATTTCAATCCAAGGATGGAAGCTTCTTGTCCTCACCAGCCTCTACAGCATGCGTCCTTATGAATACTGGAGACGAAAAATGCTTCA
CTTTTCTAAACAATTTATTAGACAAGTTCGGTGGTTGCGTCCCCTGTATGTATTCCATTGATCTGTTAGAACGTTTGAGTCTTGTAGATA
ATATTGAACACTTGGGAATTGGAAGACATTTTAAGCAGGAAATAAAAGGTGCTTTGGACTATGTCTACAGACATTGGTCTGAACGTGGTA
TAGGTTGGGGACGTGATAGCTTAGTACCTGATCTAAACACAACGGCTTTAGGTTTGCGTACGTTGCGTACACACGGATATGATGTGTCTT
CAGATGTTCTTAATAACTTCAAAGATGAGAATGGAAGATTTTTTAGCAGCGCAGGGCAAACACATGTCGAATTAAGATCAGTTGTGAATC
TATTTAGGGCTTCCGATTTAGCTTTCCCTGATGAAGGAGCCATGGATGACGCTCGTAAGTTCGCTGAACCGTACCTGAGAGACGCTTTGG

CAACCAAAATCTCCACTAATACTAAGCTATTTAAAGAGATAGAATACGTGGTGGAATACCCTTGGCATATGTCCATTCCAAGATTGGAAG
CTAGATCATATATAGATTCCTATGACGACGACTACGTGTGGCAACGTAAAACACTATATAGAATGCCATCCTTGTCCAACTCTAAGTGCT
TGGAGCTGGCAAAACTGGATTTTAACATAGTACAGAGCTTACACCAGGAGGAACTGAAATTATTGACTAGATGGTGGAAAGAAAGTGGCA
TGGCTGACATAAACTTCACAAGACACAGAGTTGCTGAAGTCTACTTCAGTTCCGCAACTTTCGAACCAGAGTATAGCGCGACCAGAATAG
CCTTTACTAAAATTGGTTGTTTGCAGGTTCTTTTCGATGATATGGCAGACATATTTGCTACCCTGGATGAATTAAAATCTTTTACAGAAG
GCGTCAAACGTTGGGACACTTCTCTATTACACGAAATTCCCGAGTGCATGCAAACGTGTTTCAAGGTGTGGTTCAAACTTATGGAAGAAG
TTAATAACGACGTCGTAAAAGTTCAAGGCAGAGATATGCTGGCGCACATTAGAAAACCTTGGGAACTTTATTTCAATTGCTACGTTCAAG
AAAGAGAATGGCTGGAAGCAGGTTACATTCCTACTTTCGAAGAATATTTGAAAACGTATGCAATAAGCGTGGGCTTGGGACCCTGTACAT
TACAGCCGATATTGTTGATGGGTGAACTTGTAAAGGATGATGTAGTGGAAAAGGTCCACTATCCTTCCAATATGTTCGAATTAGTTTCAT
TGTCATGGAGACTTACAAACGATACGAAGACTTACCAGGCAGAGAAAGCTAGAGGTCAACAGGCCTCCGGTATAGCTTGTTATATGAAAG
ACAATCCTGGAGCCACTGAGGAAGATGCTATTAAACACATTTGTAGAGTAGTAGATAGAGCTTTAAAAGAGGCCAGTTTTGAATACTTCA
AACCCAGTAACGATATTCCGATGGGTTGTAAAAGCTTTATTTTCAATCTTAGGTTGTGCGTCCAGATTTTCTATAAATTTATTGATGGCT
ACGGTATAGCCAACGAAGAGATCAAAGATTATATCAGAAAAGTGTACATTGATCCAATCCAAGTCGGAGGTGGCGGTGGTGGAATGAAAA
TCAAGACAGGCGCTAGAATATTGGCTTTATCTGCCTTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAAC
TGGTAATCTGGATCAACGGAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCG
TCGAACATCCAGATAAGTTAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGT
TTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAG
TACGTTATAATGGAAAACTTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTA
AAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTA
CATGGCCACTTATTGCTGCTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCG
GTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTAATTAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCAT
TTAACAAAGGTGAAACTGCGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCT
TACCGACGTTCAAAGGCCAACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGA
AAGAATTTCTAGAGAACTACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCT
ATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGT
CAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTA
GGATCACAAAGTAAACAGAAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTC
ATTTATGATTTTTTGTAGTAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCT
CGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATTTATTT
GAATGATATTTAGTTGGCGAACATTTTTCCATGCGCTTGGTTCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTC
ATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATTAACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAG
TCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCAAAAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATC
GGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGCTTCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTA
TCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAGAGAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGGATGCCC
TATACAAATCTACAGTGGCCAAATTTAATGAGGTCACACAGTTAGATTGCAGTACTGAATCTTTCTCTATCGCTCTATCTGCTATAGCAG
GTATACTATTATTGCTACTTCTTTTCAGATCCAAAAGACACTCTTCTTTGAAGTTGCCACCTGGGAAGTTGGGGATCCCCTTTATTGGCG
AATCATTCATTTTCTTAAGGGCTTTAAGATCCAATTCTCTGGAACAATTTTTTGACGAAAGAGTCAAAAAATTTGGTTTGGTCTTCAAAA
CTTCCTTGATAGGTCATCCAACAGTAGTTCTTTGTGGCCCTGCGGGCAATAGATTGATACTGTCTAACGAAGAAAAATTAGTCCAAATGA
GCTGGCCAGCACAATTTATGAAACTTATGGGTGAAAATTCTGTCGCAACAAGAAGGGGTGAAGATCATATCGTAATGCGTTCTGCCCTTG
CCGGTTTTTTTGGTCCTGGTGCTTTGCAATCATATATCGGTAAGATGAATACTGAAATTCAGTCCCACATAAACGAAAAGTGGAAAGGAA
AAGACGAAGTAAACGTATTGCCATTGGTTAGAGAACTAGTTTTCAACATATCCGCAATCCTATTTTTTAATATTTACGATAAACAGGAAC
AGGATAGATTACACAAATTACTAGAGACGATCCTGGTGGGTTCATTTGCCTTGCCTATCGATTTGCCTGGTTTCGGTTTCCATAGAGCAT
TGCAAGGAAGAGCCAAGTTGAATAAAATCATGTTATCCTTAATTAAAAAGCGTAAGGAGGATTTACAGTCAGGTTCCGCAACAGCAACTC
AAGACCTGTTATCAGTCCTTTTGACCTTTAGAGACGATAAAGGAACACCCTTGACTAACGACGAAATCCTTGATAACTTTTCAAGTTTAC
TACACGCATCCTATGATACGACGACAAGTCCGATGGCGCTTATATTCAAACTGCTGTCTTCCAACCCTGAATGCTATCAGAAAGTGGTCC
AGGAACAACTGGAAATTCTGTCCAATAAAGAAGAGGGAGAGGAGATCACATGGAAAGATTTGAAAGCTATGAAGTACACCTGGCAAGTGG
CTCAGGAAACTTTAAGAATGTTCCCTCCTGTTTTTGGTACTTTTAGAAAAGCTATTACAGATATTCAGTACGATGGTTATACTATACCAA
AGGGATGGAAATTATTATGGACTACTTATTCTACACATCCCAAAGATTTGTATTTCAACGAACCTGAAAAGTTTATGCCTAGCAGATTCG

ACCAGGAGGGTAAACATGTTGCACCCTACACGTTTCTACCGTTTGGAGGCGGACAAAGGTCCTGTGTCGGATGGGAATTCTCTAAAATGG
AAATCTTGTTGTTCGTACATCATTTTGTTAAAACATTTAGCAGTTATACACCTGTGGATCCCGATGAGAAGATTAGTGGGGACCCGTTGC
CCCCCTTACCTTCCAAGGGATTTTCCATAAAACTATTTCCACGTCCAGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAA
TATTGGCTTTATCTGCCTTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACG
GAGATAAAGGATATAACGGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGT
TAGAGGAAAAATTTCCTCAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAA
GTGGTTTGCTAGCAGAGATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAAC
TTATCGCATACCCGATTGCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTC
CAGCTTTAGACAAAGAACTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTG
CTGATGGAGGCTACGCCTTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGA
CGTTCTTGGTTGACTTAATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTG
CGATGACTATTAACGGTCCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCC
AACCCTCCAAGCCATTTGTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACT
ACTTATTGACGGATGAAGGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCA
AAGATCCTAGGATTGCAGCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCG
TTAGAACTGCGGTAATTAATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAG
AAGACGGGAGACACTAGCACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAG
TAGGATATAAATATATACAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAA
AGTAGCTCACTGCCTTATTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATCTATTTTTGCTTTCTGCACATGGT
TCCATTCATCCGAGCGATCACTTATCTGACTTCGTCACTTTTTCATTTCATCCGAAACAATCAAAACTGAAGCCAATCACCACAAAATTA
ACACTCAACGTCATCTTTCACTACCCTTTACAGAAGAAAATATCCATAGTCCGGACTAGCATCCCAGTATGTGACTCAATATTGGTGCAA
AAGAGAAAAGCATAAGTCAGTCCAAAGTCCGCCCTTAACCAGGCACATCGGAATTCACAAAACGTTTCTTTATTATATAAAGGAGCTGCT
TCACTGGCAAAATTCTTATTATTTGTCTTGGCTTGCTAATTTCATCTTATCCTTTTTTTCTTTTCACACCCAAATACCTAACAATTGAGA
GAAAACTCTTAGCATAACATAACAAAAAGTCAACGAAAAATGCAAGCAAACAGTAATACAGTCGAGGGCGCATCTCAAGGTAAATCCCTG
CTAGATATTAGTAGACTTGATCATATATTTGCTTTGTTGCTAAACGGAAAAGGTGGTGACTTGGGCGCTATGACTGGTAGTGCTTTGATC
CTGACCGAGAATTCACAAAATCTAATGATACTGACTACAGCATTAGCGGTTCTGGTGGCTTGCGTATTTTTTTTCGTGTGGAGAAGAGGT
GGATCAGATACGCAGAAGCCTGCTGTAAGACCCACTCCGTTGGTGAAAGAAGAAGACGAGGAGGAAGAAGATGATAGTGCCAAAAAAAAA
GTAACAATTTTTTTCGGTACTCAAACCGGCACGGCTGAGGGCTTTGCCAAGGCATTGGCGGAAGAAGCCAAAGCTCGTTACGAGAAAGCA
GTATTTAAGGTCGTGGATCTGGATAACTATGCTGCTGATGACGAACAATATGAAGAAAAACTTAAAAAAGAAAAGCTAGCTTTTTTTATG
CTAGCCACCTACGGTGATGGCGAGCCTACCGACAATGCAGCCCGTTTTTACAAATGGTTCTTGGAAGGGAAAGAGAGGGAACCTTGGTTG
TCCGACCTTACATACGGTGTTTTTGGTTTAGGCAACAGACAGTACGAACATTTTAATAAAGTGGCGAAAGCCGTCGATGAGGTTTTGATA
GAACAAGGCGCGAAGAGACTTGTACCCGTAGGCTTAGGTGATGACGATCAGTGCATTGAAGATGATTTTACGGCATGGAGAGAACAAGTG
TGGCCTGAGCTGGATCAGTTGTTGAGAGACGAAGATGACGAACCTACCTCTGCCACGCCTTATACTGCAGCTATTCCCGAATACAGAGTG
GAAATCTACGATTCCGTCGTATCAGTTTATGAAGAAACACATGCTCTAAAGCAGAACGGGCAAGCCGTATATGACATTCATCATCCGTGT
AGATCAAACGTCGCTGTAAGGAGAGAACTGCACACACCTCTTTCAGATCGTTCCTGTATTCATCTGGAATTTGATATCAGTGACACCGGT
CTGATATATGAGACTGGTGACCATGTTGGCGTTCATACAGAAAATTCTATAGAAACCGTGGAAGAGGCCGCAAAGTTATTAGGCTATCAA
TTGGATACCATATTTTCCGTTCATGGTGATAAAGAAGACGGCACTCCATTAGGCGGCTCTTCTTTGCCACCACCCTTTCCAGGGCCCTGT
ACATTAAGGACCGCCTTAGCACGTTATGCCGACCTGTTAAATCCGCCGAGGAAAGCGGCTTTCTTGGCGTTGGCCGCGCATGCTTCTGAT
CCTGCCGAAGCTGAAAGACTAAAATTTTTATCCTCCCCAGCCGGCAAAGATGAATACTCACAGTGGGTGACTGCATCACAAAGAAGTCTT
TTAGAAATCATGGCTGAATTTCCGAGCGCAAAGCCTCCGCTAGGTGTCTTCTTCGCAGCCATTGCACCAAGACTTCAACCAAGATATTAT
TCAATTTCATCAAGTCCTAGATTTGCCCCTTCCAGAATTCACGTTACATGCGCATTGGTCTACGGCCCATCTCCGACTGGCAGAATTCAT
AAGGGTGTTTGTTCTAACTGGATGAAGAACAGTCTTCCTTCTGAGGAAACACATGATTGCTCTTGGGCGCCTGTTTTTGTTAGACAAAGT
AACTTCAAGCTTCCTGCGGACTCAACCACACCGATTGTGATGGTGGGACCAGGAACTGGCTTTGCACCTTTCAGAGGCTTCCTACAGGAA
AGGGCAAAATTGCAAGAAGCGGGAGAAAAATTAGGTCCAGCAGTTTTATTTTTCGGTTGTAGGAACAGGCAAATGGATTACATATACGAA
GATGAGTTAAAGGGATACGTGGAAAAGGGTATTTTGACCAATCTGATTGTTGCCTTCAGCAGGGAAGGCGCAACGAAGGAATATGTGCAA
CACAAAATGCTAGAAAAGGCATCTGACACTTGGTCACTAATAGCCCAAGGTGGCTATTTATATGTGTGTGGCGATGCTAAGGGGATGGCC
AGAGATGTCCACAGAACTTTACACACTATCGTACAAGAACAGGAATCAGTCGACTCCTCAAAGGCTGAATTCTTAGTTAAGAAGCTTCAG
ATGGACGGTAGATATTTGAGGGATATTTGGGGAGGTGGCGGTGGTGGAATGAAAATCAAGACAGGCGCTAGAATATTGGCTTTATCTGCC

TTAACGACTATGATGTTTTCTGCATCTGCTTTGGCTAAGATCGAAGAAGGTAAACTGGTAATCTGGATCAACGGAGATAAAGGATATAAC
GGTTTAGCCGAGGTCGGTAAGAAATTCGAAAAAGATACTGGAATAAAAGTCACCGTCGAACATCCAGATAAGTTAGAGGAAAAATTTCCT
CAAGTTGCGGCTACCGGCGACGGACCCGATATTATTTTCTGGGCTCACGACAGGTTTGGTGGATACGCACAAAGTGGTTTGCTAGCAGAG
ATTACGCCAGACAAAGCTTTCCAAGACAAATTGTATCCATTTACTTGGGATGCAGTACGTTATAATGGAAAACTTATCGCATACCCGATT
GCTGTTGAGGCTCTTTCTTTAATTTATAATAAGGATCTGCTACCGAATCCACCTAAAACCTGGGAAGAGATTCCAGCTTTAGACAAAGAA
CTGAAGGCTAAAGGTAAGTCTGCGTTGATGTTTAATCTTCAAGAACCGTATTTTACATGGCCACTTATTGCTGCTGATGGAGGCTACGCC
TTCAAATACGAAAACGGCAAATACGACATAAAAGACGTCGGTGTTGATAATGCCGGTGCAAAAGCAGGCCTGACGTTCTTGGTTGACTTA
ATTAAAAATAAGCACATGAATGCTGACACAGATTATTCAATTGCTGAAGCCGCATTTAACAAAGGTGAAACTGCGATGACTATTAACGGT
CCATGGGCTTGGTCCAACATCGACACCAGCAAGGTGAATTACGGTGTAACTGTCTTACCGACGTTCAAAGGCCAACCCTCCAAGCCATTT
GTGGGAGTATTGTCCGCTGGTATTAACGCCGCTAGTCCTAATAAAGAGCTGGCGAAAGAATTTCTAGAGAACTACTTATTGACGGATGAA
GGTTTGGAAGCTGTGAACAAGGATAAGCCTCTTGGTGCGGTAGCATTGAAGTCCTATGAGGAAGAATTGGCCAAAGATCCTAGGATTGCA
GCCACAATGGAAAATGCGCAGAAAGGTGAAATAATGCCCAATATTCCCCAGATGTCAGCGTTTTGGTATGCCGTTAGAACTGCGGTAATT
AATGCTGCATCTGGGAGGCAAACTGTCGATGAGGCTCTTAAGGACGCCCAAACTAGGATCACAAAGTAAACAGAAGACGGGAGACACTAG
CACACAACTTTACCAGGCAAGGTATTTGACGCTAGCATGTGTCCAATTCAGTGTCATTTATGATTTTTTGTAGTAGGATATAAATATATA
CAGCGCTCCAAATAGTGCGGTTGCCCCAAAAACACCACGGAACCTCATCTGTTCTCGTACTTTGTTGTGACAAAGTAGCTCACTGCCTTA
TTATCACATTTTCATTATGCAACGCTTCGGAAAATACGATGTTGAAAATCCCTGGTCAAACTTCAGAACTAAAAAAATAATAAGGAAGAA
AAAAATAGCTAATTTTTCCGGCAGAAAGATTTTCGCTACCCGAAAGTTTTTCCGGCAAGCTAAATGGAAAAAGGAAAGATTATTGAAAGA
GAAAGAAAGAAAAAAAAAAAATGTACACCCAGACATCGGGCTTCCACAATTTCGGCTCTATTGTTTTCCATCTCTCGCAACGGCGGGATTC
CTCTATGGCGTGTGATGTCTGTATCTGTTACTTAATCCAGAAACTGGCACTTGACCCAACTCTGCCACGTGGGTCGTTTTGCCATCGACA
GATTGGGAGATTTTCATAGTAGAATTCAGCATGATAGCTACGTAAATGTGTTCCGCACCGTCACAAAGTGTTTTCTACTGTTCTTTCTTC
TTTCGTTCATTCAGTTGAGTTGAGTGAGTGCTTTGTTCAATGGATCTTAGCTAAAATGCATATTTTTTCTCTTGGTAAATGAATGCTTGT
GATGTCTTCCAAGTGATTTCCTTTCCTTCCCATATGATGCTAGGTACCTTTAGTGTCTTCCTAAAAAAAAAAAAAGGCTCGCCATCAAAA
CGATATTCGTTGGCTTTTTTTTCTGAATTATAAATACTCTTTGGTAACTTTTCATTTCCAAGAACCTCTTTTTTCCAGTTATATCATGGT
CCCCTTTCAAAGTTATTCTCTACTCTTTTTCATATTCATTCTTTTTCATCCTTTGGTTTTTTATTCTTAACTTGTTTATTATTCTCTCTT
GTTTCTATTTACAAGACACCAATCAAAACAAATAAAACATCATCACAATGTCTAGATTAGAAAGATTGACCTCATTAAACGTTGTTGCTG
GTTCTGACTTGAGAAGAACCTCCATCATTGGTACCATCGGTCCAAAGACCAACAACCCAGAAACCTTGG

Addendum 2:
PCR Tables

ADDENDUM 2: PCR TABLES

Table A2.1: PCR mix for all fragments (OneTaq)			
Component	**Initial concentration**	**Final concentration**	**Volume (µL)**
OneTaq® Standard reaction buffer	5x	1x	10
OneTaq® DNA polymerase	5 U/µL	0.025 U/µL	0.25
dNTPs	10 mM	0.2 mM	1
Forward Primer	10 µM	0.2 µM	1
Reverse Primer	10 µM	0.2 µM	1
DNA	25-50 ng/µL	12.5-25 ng/µL	0.5
dH_2O	-	-	36.25

Table A2.2: PCR mix for all fragments (Phusion)			
Component	**Initial concentration**	**Final concentration**	**Volume (µL)**
Phusion® HF buffer	5x	1x	10
Phusion® DNA polymerase	5 U/µL	0.025 U/µL	0.5
dNTPs	10 mM	0.5 mM	1
Forward Primer	10 µM	0.5 µM	2.5
Reverse Primer	10 µM	0.2 µM	2.5
DNA	25-50 ng/µL	12.5-25 ng/µL	0.5
dH_2O	-	-	33

Table A2.3: Thermal cycler parameters for PCR amplification of gBlocks (OneTaq)				
Step	**Fragment**	**Temperature (°C)**	**Time (s)**	
Initial denaturing	All	94	30	
Denaturing	All	94	30	x30
Annealing	A	54	60	
	1	55		
	2	54		
	3	52		
	4	51		
	5	51		
	6	49		
	7	46		
	B	52		
	V1	66		
	V2	63		
	V3	62		
Extension	A	68	60	
	1		120	
	2		120	
	3		105	
	4		105	
	5		105	
	6		120	
	7		120	
	B		60	
	V1		195	
	V2		195	
	V3		195	
Final Extension	All	68	300	
Hold	All	4	∞	

Table A2.4: Thermal cycler parameters for PCR amplification of gBlocks (Phusion)				
Step	**Fragment**	**Temperature (°C)**	**Time (s)**	
Initial denaturing	All	98	30	
Denaturing	All	98	10	x40
Annealing	A	64	30	
	1	64		
	2	66		
	3	63		
	4	64		
	5	63		
	6	62		
	7	60		
	B	63		
	V1	–	No annealing step	
	V2	–		
	V3	–		
Extension	A	72	30	
	1		60	
	2		60	
	3		50	
	4		50	
	5		50	
	6		60	
	7		60	
	B		30	
	V1		105	
	V2		105	
	V3		105	
Final Extension	All	72	600	
Hold	All	12	∞	

Table A2.5: PCR primers for gBlocks						
		Sequence (5'->3')	Length (bp)	GC%	NEB OneTaq Tm (°C)	NEB Phusion Tm (°C)
R1	Forward primer	CATGACCCCGGATCGTCG	18	66.7	62	62
	Reverse primer	CGTCTCTCAGGTACGGTTCAG	21	57.1	60	61
	Product length	1875				
R2	Forward primer	GTAAGTTCGCTGAACCGTACCTG	23	52.2	61	63
	Reverse primer	TAATTAAGTCAACCAAGAACGTCAGGC	27	40.7	59	63
	Product length	1875				
R3	Forward primer	CGGCAAATACGACATAAAAGACG	23	43.5	57	60
	Reverse primer	TCAAGGAAGTTTTGAAGACCAAACC	25	40.0	58	62
	Product length	1580				
R4	Forward primer	AATTTGGTTTGGTCTTCAAAACTTCC	26	34.6	56	61
	Reverse primer	GCTTTGTCTGGCGTAATCTCTG	22	50	59	61
	Product length	1585				
R5	Forward primer	AGTGGTTTGCTAGCAGAGATTACG	24	45.8	59	62
	Reverse primer	TTGCTTGCATTTTTCGTTGACTT	23	34.8	56	60
	Product length	1580				
R6	Forward primer	CTTTTCACACCCAAATACCTAACA	24	37.5	54	59
	Reverse primer	ACTGCTGGACCTAATTTTTCTCC	23	43.5	57	60
	Product length	1875				
R7	Forward primer	AGAAGCGGGAGAAAAATTAGG	21	42.9	55	58
	Reverse primer	TCTTCCTTATTATTTTTTTAGTTCTGAAG	29	24.1	51	56
	Product length	1875				

Table A2.6: PCR primers for flanking regions						
		Sequence	Length (bp)	GC%	NEB OneTaq Tm (°C)	NEB Phusion Tm (°C)
A	Forward Primer	TTCCTAAGCCTCCCTCACCA	20	55.00	61	62
	Reverse Primer	GCTACCCTGACCGTATCACA	20	55.00	59	61
	Product length	949				
B	Forward Primer	CCCTGGTCAAACTTCAGAACTAA	23	43.48	57	60
	Reverse Primer	CCAAGGTTTCTGGGTTGTTGG	21	52.38	59	61
	Product length	920				

Table A2.7: PCR primers for vectors								
		Sequence	**Length (bp)**	**GC%**	**Homology with vector**	**Homology with insert**	**NEB OneTaq Tm (°C)**	**NEB Phusion Tm (°C)**
V1	**Forward primer**	TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGG	40	50	20	20	71	74
	Reverse primer	GCCTGACGTTCTTGGTTGACTTAATTAAATCGAATTCCCGCGGCCGC	47	51	20	27	74	77
	Product length	3062						
V2	**Forward primer**	CGTCTTTTATGTCGTATTTGCCGAATCACTAGTGAATTCGCGG	43	44	20	23	68	72
	Reverse primer	AAGTCAACGAAAAATGCAAGCAAAATCGAATTCCCGCGGCCGC	43	49	20	23	73	77
	Product length	3061						
V3	**Forward primer**	TGTTAGGTATTTGGGTGTGAAAAGAATCACTAGTGAATTCGCGG	44	41	20	24	67	72
	Reverse primer	CCAACAACCCAGAAACCTTGGAATCGAATTCCCGCGG	37	54	16**	21	72	75
	Product length	3060						
V1.5 *	**Forward primer**	TGGTGAGGGAGGCTTAGGAAAATCACTAGTGAATTCGCGG	40	50	20	20	71	74
	Reverse primer	AAGTCAACGAAAAATGCAAGCAAAATCGAATTCCCGCGGCCGC	43	49	20	23	73	77
	Product length	3058						

* V1.5 is the combination of V1 and V2

** Lower homology with vector to reduce Tm

Table A2.8: Gibson isothermal buffer mix			
Component	Concentration	6 ml	1 ml
Tris-HCl	1 M	3000	500
MgCl2	2 M	150	25
dNTPs	10 mM each	600	100
DTT	1 M	300	50
PEG-8000	g	1.5	0.25
NAD^+	50 mM	600	100

Table A2.9: Gibson 2X master mix			
Component	Concentration	1200 µL	180 µL
Iso buffer	5X	320	48
T5 exo	1 U/µL	6.4	0.96
Phusion	2 U/µL	20	3
Taq ligase	40 U/µL	160	24
water	N/A	293.6	104.04

Table A2.10: PCR mix for all fragments (Kapa)			
Component	Initial concentration	Final concentration	Volume (µL)
Buffer	5X	1X	5.00
dNTPs	10 mM	0.3 mM	0.75
Kapa	1 U/µL	0.4 U/µL	0.50
DNA	Variable	Variable	0.50
Fw primer	10 µM	0.3 µM	0.75
Rv primer	10 µM	0.3 µM	0.75
Water	-	-	16.75

Table A2.11: Thermal cycler parameters for PCR amplification of gBlocks (Kapa)				
Step	**Fragment**	**Temperature (°C)**	**Time (s)**	
Initial denaturing	All	95	180	
Denaturing	All	98	20	x40
Annealing	A	61	15	
	1	61		
	2	63		
	3	60		
	4	61		
	5	60		
	6	59		
	7	57		
	B	60		
Extension	A	72	30	
	1		60	
	2		60	
	3		50	
	4		50	
	5		50	
	6		60	
	7		60	
	B		30	
Final Extension	All	72	600	
Hold	All	12	∞	