

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Constructing Aspartic Acid Synthesis Pathway and Gene Overexpression in *Yarrowia lipolytica*

**Permalink**

<https://escholarship.org/uc/item/7bb2p1nt>

**Author**

Halfon, Alon

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Constructing Aspartic Acid Synthesis

Pathway and Gene Overexpression

in *Yarrowia lipolytica*

A thesis submitted in partial satisfaction  
of the requirements for the degree Master of Science  
in Bioengineering

by

Alon Halfon

2022

© Copyright by

Alon Halfon

2022

## ABSTRACT OF THE THESIS

Constructing Aspterric Acid Synthesis  
Pathway and Gene Overexpression  
in *Yarrowia lipolytica*

by

Alon Halfon

Master of Science in Bioengineering  
University of California, Los Angeles, 2022  
Professor Junyoung O. Park, Co-Chair  
Professor Gerard C. L. Wong, Co-Chair

Modern agriculture relies heavily on a small selection of widespread market-dominant herbicides, many of which share overlapping modes of action such as ALS inhibition from Chlorsulfuron and EPSP inhibition from Round-Up ready. This widespread use has caused a substantial increase in the number and types of herbicide resistant weeds throughout the world which underlies the near future need for herbicides with novel modes of action. Aspterric acid is a sesquiterpenoid herbicide which inhibits the DHAD enzyme of the branched chain amino acid pathway (BCAA); this enzyme is not targeted by any commercial herbicide on the market which makes aspterric acid an ideal candidate for a novel herbicide. *Yarrowia lipolytica*, an oleaginous

yeast strain, stands out as a promising microbial platform for biosynthetic production of aspterric acid due to its high acetyl-CoA flux and a growing body of literature affirming *Y. lipolytica* as particularly suited for the over-production of terpenoids. The four ast genes (A,B,C,D) involved in the aspterric acid synthesis pathway were markerlessly integrated into the *Y. lipolytica* host genome, along with gene copies for HMGCR and FPPS which would overexpress commonly noted rate-limiting enzymes involved in terpenoid synthesis so as to further enhance carbon flux towards terpenoid production. However, the engineered strains of *Y. lipolytica* failed to produce aspterric acid and instead yielded an isomer, demonstrating the continued challenges that follow from attempts at heterologous pathway introduction into this microbe.

The thesis of Alon Halfon is approved.

Yi Tang

Junyoung O. Park, Committee Co-Chair

Gerard C.L. Wong, Committee Co-Chair

University of California, Los Angeles

2022

## Table of Contents

Abstract.....	ii
Committee Page.....	iv
Table of Contents.....	v
List of Figures.....	vi
Acknowledgments.....	viii
Introduction.....	1
Materials and Methods.....	3
Results and Discussion.....	9
Future Directions.....	21
Figures.....	22
Primer and Synthesized DNA List.....	48
Laragen Sequence Reads.....	51
References.....	70

## List of Figures and Tables

Figure 1: The Mevalonate Pathway.....	22
Figure 2: Aspterric Acid Synthesis.....	23
Figure 3: Plasmid Designs.....	24
Figure 4: PCR Screening for Genomic Integration.....	25
Figure 5: Preparation of DNA Fragments.....	26
Figure 6: Lip2t-pFBAin Fragment PCR.....	27
Figure 7: Sequence Verification of Recombinant Plasmids.....	28
Figure 8: ast Gene Introns and Exons.....	29
Figure 9: PCR of <i>A. terreus</i> cDNA.....	30
Figure 10: Clustered Colonies and Normal Colonies.....	31
Figure 11: Intron Retention in astC cDNA.....	32
Figure 12: Method Comparison of Growth for 2A AD and TPS MVL Transformation.....	33
Figure 13: Plasmid Dropout Verification.....	34
Figure 14: Three Primer PCR Contamination.....	35
Figure 15: Aspterric Acid Signal Among Engineered Strains, PO1f, and Standard.....	36
Figure 16: Aspterric Acid Spike-in.....	37
Figure 17: Carbon Labeling of Unknown Metabolite.....	38
Figure 18: Aspterric Acid Signal of EtOAc Extraction of <i>Y. lipolytica</i> Strain and Standard.....	39
Figure 19: Mevalonate Levels of EtOAc and MeOH Extracts of MVL Strains and PO1f.....	40
Figure 20: cDNA PCR of Engineered Strains and PO1f for astA,B,C,D.....	41
Figure 21: Verification of Genomic Integration of AD Genes for AM2 and AD2 Strains.....	42
Figure 22: Genomic Integration of Engineered Strains.....	43



Figure 23: cDNA Synthesis Mix Contamination.....44

Figure 24: Fluorescence Image of Cells in YPD.....45

Figure 25: Fluorescence Imaging of Cells in Minimal Media.....46

Table 1: Engineered Strains' Genotype.....47

## **Acknowledgements**

I would like to acknowledge Glenn Nurwono for his contribution to this work (especially in preparing metabolite extractions), as well as Dr. Yi Tang's lab for providing *Aspergillus terreus* samples for RNA extraction and Dr. Nasim Annabi's lab for lending the use of their fluorescent microscope. I would also like to acknowledge my faculty advisor, Dr. Jun Park for his continual advice and guidance, the past and present members of the Jun Park lab for their feedback, and my thesis committee for theirs.

## Introduction

Terpenoids (or isoprenoids) are an extremely numerous and diverse class of natural products synthesized by the joining of the 5 carbon isoprene units isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) [1]. These isoprenoid precursors are primarily produced by the mevalonate (MVA) pathway (**fig. 1**) and can be chained together multiple times to synthesize increasingly large terpenoids which are themselves often structurally modified by enzymes, lending this class of compounds a diverse set of functions [1]. Given this sizable variety in functions, there is doubtless numerous different terpenoids which could be benefited from by industrial production. The fungal sesquiterpenoid aspterric acid, for instance, is a natural product which functions as a submicromolar competitive inhibitor for dihydroxyacid dehydratase (DHAD) which is a highly conserved enzyme that plays an essential role in the synthesis of branched chain amino acids (BCAA) [2]. The BCAA pathway is not present in animals which makes aspterric acid an ideal candidate for novel herbicide production. The BCAA pathway, however, is present in bacteria, which raises the concern that aspterric acid may affect animal microbiomes if it inhibits bacterial DHAD and will thus need to be further tested before agricultural use. The need for novel herbicides which have different mechanisms of action is beginning to become apparent as the presence of herbicide resistant weeds continue to increase both in number and in type of herbicide resistance [3]. As DHAD is an enzyme that is not targeted by any current herbicide on the market, large scale production of aspterric acid may prove to be commercially useful in the near future [2]. One microbe that seems particularly promising for this task is *Yarrowia lipolytica*.

*Y. lipolytica* is a non-pathogenic, oleaginous yeast strain which is generally regarded as safe by the FDA [4]. It has a fully sequenced genome and a number of genetic tools as well as a

sizable and expanding amount of literature describing it [5]. Furthermore, *Y. lipolytica* possesses a very high acetyl-coA flux due to the enzyme ATP Citrate Lyase (ACL) which catalyzes the conversion of citrate into acetyl-CoA, funneling carbon from the TCA cycle into acetyl-CoA biosynthetic pathways which include lipid production and the mevalonate pathway [4][6]. This seems to make *Y. lipolytica* preferable to more conventional microbes (i.e. *E. coli* and *S. cerevisiae*) when considering a microbial platform for the industrial production of terpenoids such as aspterric acid. This is echoed in literature as a number of papers find enhanced yields of terpenoids when working with *Y. lipolytica* [4][7][8].

In order to leverage *Y. lipolytica*'s metabolism for the production of aspterric acid, the heterologous genes ast A, B, C (**fig. 2**) must be introduced into the new strain along with the ast D self-resistance gene that encodes for a homolog of DHAD which is insensitive to the presence of aspterric acid [2]. Furthermore, terpenoid production within the cell can be enhanced by removing bottlenecks within relevant pathways by overexpression of certain rate-limiting enzymes via gene copy integration; two commonly referenced rate-limiting enzymes in terpenoid synthesis are HMG-CoA Reductase (HMGCR) and FPP Synthase (FPPS) [4][7][9]. Due to existing genetic tools developed for *Y. lipolytica*, these genes can be markerlessly integrated directly into one of five genomic loci while maintaining auxotrophy, allowing for further cycles of genetic engineering [10]. The goal of this research is to construct the aspterric acid synthesis pathway within *Y. lipolytica* and measure its production while developing strains that further enhance terpenoid production.

## Materials and Methods

### Plasmid Design

Homologous Recombination (HR) plasmids for genomic loci: A08, AXP, D17, MFE1, and XPR2 along with their corresponding CRISPR plasmids were obtained from addgene as agar stabs of transformed *E. coli* conferring ampicillin resistance (Catalog: 84608-84617). In order to insert all six genes into the *Y. lipolytica* genome, the HR plasmids were to be modified to contain two genes per plasmid. Two plasmid systems were designed in parallel for this purpose (**fig. 3**) The 2A system (2AS) separates the two genes of interest with the 2A peptide cleavage sequence ERBV-1 in order to produce two proteins from one mRNA transcript; ERBV-1 has been shown to have high cleavage efficiency in *S. cerevisiae*, but has not been characterized in *Y. lipolytica* [11]. The two-promoter system (TPS) was developed as redundancy in case the ERBV-1 sequence proved non-functional in *Y. lipolytica*, it separates the two genes with a native terminator (lip2t) and native, strong constitutive promoter (pFBAin) so that two mRNA transcripts are produced, each translating into one protein [5][12][13].

### Native DNA Fragment Isolation

The PO1f strain of *Y. lipolytica* was purchased from atcc and possesses uracil and leucine auxotrophy. Genomic DNA from *Y. lipolytica* was extracted using the Wizard Genomic DNA Purification Kit from Promega (catalog A1120) and incubating in lyticase for 1 hour from sigmaaldrich (catalog L2524-10KU). Native genes for HMGCR and FPPS were amplified with Gibson overhangs using primers: GA4,51; GA4,56 and GA3,6; GA3,57 (2AS; TPS). The ERBV-1 sequence was synthetically produced from IDT. Native terminator Lip2t and native constitutive promoter pFBAin were amplified with overhangs using primers: P29,30 and P28, GA54.

## **Heterologous DNA Fragment Isolation**

*Aspergillus terreus* fungal culture sample was received from Dr. Yi Tang lab at UCLA. RNA extraction was conducted using PureLink RNA Mini Kit from thermofisher (catalog 12183020). cDNA synthesis was conducted using oligo(dT)<sub>12-18</sub> as a template on poly(A) tails of mRNA and the SuperScript IV First-Strand Synthesis System from thermofisher (catalog 18091050). PCR of cDNA was conducted using primers to amplify genes astA,B,C,D without introns: GA37 GA49; GA37 GA58, GA39 GA40; GA40 GA61, GA34 GA50; GA34 GA60, GA35 GA36; GA35 GA59 (2AS; TPS).

## **Gel Extraction**

All DNA fragments are run on a 1% agarose gel with SYBR safe DNA Gel Stain from thermo fisher (catalog S33102) at 150 V for 20 min alongside a ladder. Gels are imaged on a blue/white light transilluminator from thermofisher (catalog LB0100) and appropriately sized bands for corresponding DNA fragments are excised from the gel and placed in 1.5 ml Eppendorf tubes using a razor. The isolated DNA fragments are extracted from the gel slices using a Monarch DNA Gel Extraction Kit from NEB (catalog T1020L) with an added 10-minute 50° C incubation phase prior to elution to improve DNA yield. A nanodrop is used to confirm yield and purity.

## **Plasmid Construction**

HR plasmids were digested with FastDigest PteI and FastDigest NheI (catalog FD2134 and FD0973) in order to linearize the plasmid and remove the placeholder hrGFP gene. The digest product was run on an agarose gel displaying two bands to validate digestion and the linearized plasmid was then excised and extracted from the gel. Recombinant plasmids were synthesized using linearized plasmids and gene fragments via Gibson Assembly Master Mix from NEB

(catalog E2611S) The recombinant plasmids are as follows: MVL plasmid (containing HMGCR and FPPS for AXP locus), AD plasmid (D17 locus), and CB plasmid (MFE1 locus).

### **Plasmid Harvesting**

Competent DH5 $\alpha$  *E. coli* cells were produced using the Mix and Go *E. coli* Transformation Kit. Gibson products (recombinant plasmids) were transformed into competent *E. coli* via brief incubation on ice then plating on pre-warmed LB ampicillin plates. Transformed colonies were grown in ampicillin LB media and plasmid was extracted using monarch plasmid miniprep kit from NEB (catalog T1010L). Insertion of genes into each plasmid was confirmed via sanger sequencing at Laragen using primers S10,15,16,17 for 2AS plasmids and S10,17,26 and P30 for TPS plasmids. Colonies with the correct recombinant plasmids were grown in 100 ml of ampicillin LB media and harvested for high concentration, high volume plasmid using ZymoPURE II Plasmid Midiprep Kit (catalog D4201).

### ***Y. lipolytica* Transformation**

Fresh transformation media is made by mixing 2.25 ml 50% 3350PEG (sterile filtered), 125  $\mu$ l 2M LiAc, and 125  $\mu$ l 2M DTT. Overnight culture of *Y. lipolytica* in 5 ml YPD is spun down and supernatant decanted leaving approximately 200-300  $\mu$ l of media left which is used to resuspend pellet to form highly concentrated liquid culture. CRISPR and recombinant HR plasmids are added to a 1.5 ml Eppendorf tube in a roughly 1:1 mass ratio to a total of 10  $\mu$ l resulting in at least 2  $\mu$ g per plasmid. 5  $\mu$ l of salmon sperm DNA is added from thermo fisher (catalog 15632011), along with 100  $\mu$ l of the transformation media and approximately 50  $\mu$ l of concentrated liquid culture. The mixture is incubated on a shaking heat block at 28° C for 20 minutes, then the cells undergo heat shock by incubating at 39° C for 45 minutes, following which, the cells are recovered in 1 ml of YPD media at 30° C for 2 hours. The cells are then spun

down and all but 100  $\mu$ l of supernatant is removed in which the cells are resuspended; 50  $\mu$ l of the media is inoculated into 5 ml of -uracil, -leucine media and 50  $\mu$ l of the media is spread on a -uracil, -leucine plate which are both placed in an incubator at 30° C (RPM: 400). Both the plate and media are made using CSM -Leucine, -Uracil powder from Sunrise Science (catalog 1038-010).

### **Colony Selection**

After 2-3 days, colonies are checked for on the auxotrophic plate (only cells that contain both a copy of the CRISPR plasmid which contains the Leucine marker and the HR plasmid which contains the Uracil marker will grow on the plate or in the media). If there are no colonies on the plate and the liquid media has growth, 50  $\mu$ l of the liquid culture will be spread on another auxotrophic plate. If there are colonies on the plate, the cells undergo a three-primer PCR using Phire Plant Direct PCR Master Mix (catalog F160S) in order to verify integration of the genes of interest at a particular genomic locus: AXP- P41,42,53; A08- P39,40,53; D17- P43,44,53; MFE1- P45,46,53; XPR2- P47,48,53. The PCR products are run on a gel, a 1 kb product indicating genomic integration and a 2 kb product indicating lack of integration [10] (**fig. 4**)

Select colonies that produce a 1 kb band are grown in 5-FOA YPD media for 1 day in order to cure/drop out both plasmids. The 5-FOA becomes a toxic product in the presence of the uracil marker so it accelerates plasmid dropout.

### **Final Selection and Verification**

5FOA liquid cultures are streaked on YPD plates and left to grow for 1 day. Colonies from these plates are then selected and transferred over to form YPD, -uracil, and -leucine patch plates and then allowed to grow for 1 day. The presence of growth on either the -uracil plate or the -leucine plate indicates that the plasmids have not dropped out from a specific colony. Colonies on the



YPD patch plate that do not show growth on the auxotrophic plates undergo the same aforementioned PCR step as a final verification step for genomic integration. A colony is selected from the YPD patch plate which shows no growth on auxotrophic plates and shows genomic integration and is inoculated into a 1 ml YPD culture overnight. 500  $\mu$ l of the overnight culture is mixed with 500  $\mu$ l of autoclaved 50% glycerol in a cryovial and stored at  $-80^{\circ}$  C. This new strain is now ready for further rounds of genomic integration.

### **Metabolite Extraction: MeOH Protocol and LCMS**

Four strains of interest are produced: TPS MVL A-D (MVL includes HMGCR and FPPS, while A-D refers to all 4 ast genes) now referred to as AMT, 2AS MVL A-D (AM2), TPS A-D (ADT), and 2AS A-D (AD2). All four strains and PO1f are grown in 5 ml synthetically defined media (YNB, ammonium sulfate, amino acids, 2% glucose) for 3 days. OD of each strain is measured before 2 ml of culture is pipetted onto a nylon filter on a filtering flask connected to a vacuum. The filter paper is then placed face down into a well or petri dish containing 400  $\mu$ l of  $-80^{\circ}$  C extraction solvent (40 methanol: 40 ACN: 20 water) and the plate is tilted until the filter is soaked through. Once all of the cultures have been filtered, the plate containing the wet filters is stored at  $-20^{\circ}$  C for 30 minutes. The plates are then taken out on ice and the filters are flipped over so that the cell side is exposed, the remaining liquid is then pipetted several times to remove the cell debris from the filter; all of the remaining liquid and debris is pipetted into an Eppendorf tube and centrifuged at maximum speed at  $4^{\circ}$  C for 10 minutes. The cell pellet is extracted from a second time by resuspending in an extra 100  $\mu$ l of extraction solvent and then centrifuged again. The supernatant is then harvested and stored at  $-80^{\circ}$  C until LC-MS analysis. All 5 strains as well as an aspartic acid standard from Cayman Chemical (catalog 25919) dissolved in extraction solvent were sampled by the LCMS and analyzed on Maven.

### **Metabolite Extraction: EtOAc Method**

Strains are grown in synthetically defined media (YNB, amino acids, ammonium sulfate, 2% glucose) for 3 days. OD of each culture is determined by spectrophotometer. An equivalent volume of 1:3 acetone:EtOAc is added to a sample of the liquid culture (2 ml) with a scoop of glass beads. The solution is vortexed for 20 minutes to break up *Y. lipolytica* cell walls, then it is centrifuged. The top layer of the solution (the organic phase) is harvested and placed into a speed vac until completely evaporated. The extract is then reconstituted in the extraction solvent (40 methanol: 40 ACN: 20 water) and is stored at -80°C until it is ready to be analyzed on the LCMS.

### **Fluorescent Microscope Sample Preparation**

Strains are grown in minimal media (YNB, ammonium sulfate, uracil, leucine, 1% glucose) overnight. The cultures are then diluted 1:10 in minimal media and 5 µl are pipetted onto a glass microscope slide. The slide is then placed on the microscope to allow for fluorescence imaging.

## Results and Discussion

### Assembling Recombinant Plasmids: Native Genes

Genomic DNA was extracted from PO1f *Y. lipolytica* and used as a template for amplification of native genes and other genetic elements. The HMGCR gene had low yields of amplification, so it was broken down into PCRs of two 1.2 kb fragments which led to more visible bands (**fig 5**). The FPPS gene also suffered from low amplification; this was rectified by using a gradient PCR which determined the ideal melting temperature for this gene's PCR (**fig 5**). The Lip2 terminator and FBAIN promoter fragments were both successfully amplified from *Y. lipolytica* genomic DNA. HR and CRISPR plasmids for AXP, D17, and MFE1 loci were harvested from *E. coli* stock via plasmid miniprep resulting in concentrations ranging from 50 to 200 ng/ $\mu$ l in 50  $\mu$ l. The yields of plasmid were insufficient for *Y. lipolytica* transformation and thus were concentrated using sodium acetate and ethanol. This still did not meet sufficient concentrations of DNA. Plasmid midipreps were then utilized on 100 ml cultures of *E. coli* to attain high yields of highly concentrated HR and CRISPR plasmids ranging from 500 to 2000 ng/ $\mu$ l in 200  $\mu$ l. The HR AXP plasmid was double digested via restriction enzymes resulting in two bands indicating the linearized vector (a strong band at 8 kb) and the excised hrGFP (a faint band at 760 bp) (**fig 5**).

Gel extraction of DNA fragments and vectors was too low for use in Gibson Assembly (ranging from 5-30 ng/ $\mu$ l in 20  $\mu$ l). Use of 50°C water as eluent and addition of a 10-minute, 50°C incubation step on a heat block for the spin column significantly increased DNA yields (40-80 ng/ $\mu$ l in 35  $\mu$ l). Multiple gel slices containing DNA were digested and run through the same column to enhance DNA concentration for linearized vectors to provide necessary yield for Gibson Assembly (>100 ng/ $\mu$ l in 35  $\mu$ l). Competent *E. coli* cells were made and evaluated by

testing with negative and positive controls (growth on Ampicillin plate without HR AXP plasmid and with HR AXP plasmid); the negative control showed no growth and the positive control produced colonies.

Initial Gibson Assembly of the 2A MVL (FPPS and HMGCR) and TPS MVL plasmids and subsequent *E. coli* transformation were unsuccessful and did not lead to colonies on Ampicillin plates. Adjustment of Gibson Assembly mix to contain more linearized vector (250 ng) and a 1:1 ratio of DNA fragments (with the exception of the small 69 bp ERBV-1 fragment which was added at a 5:1 ratio) yielded colonies for the 2A MVL plasmid transformation, but only produced colonies for the TPS MVL plasmid after several attempts. The TPS MVL plasmid is made out of 6 fragments (FPPS, HMGCR1, HMGCR2, AXP, Lip2t, pFB<sub>Ain</sub>) as opposed to the 2A MVL which is made out of 5 fragments (FPPS, HMGCR1, HMGCR2, AXP, ERBV-1) and, as a result, the Gibson assembly for TPS plasmids is less likely to be successful. In order to remedy this issue, Lip2t-pFB<sub>Ain</sub> was amplified as a single fragment using the extracted recombinant plasmid from successful colonies as a template (**fig 6**).

Minipreps of several transformed colonies were performed and the harvested plasmids were sent out for sequencing. Sequencing confirmed correct placement of FPPS, HMGCR1, and HMGCR2 and proper incorporation into AXP vector backbone (between the TEF promoter and CYC terminator) for both plasmids as well as ERBV-1 incorporation for 2A MVL plasmid and Lip2t-pFB<sub>Ain</sub> incorporation for TPS MVL between the two genes (**fig 7**).

### **Assembling Recombinant Plasmids: Heterologous Genes**

The heterologous genes astB, astC, and astD all contain introns (**fig 8**) which must be removed prior to forming the recombined plasmid as it is not known whether or not *Y. lipolytica* will be able to splice them out during mRNA processing. RNA from *A. terreus* fungal culture

was successfully extracted and used as a template for cDNA synthesis. PCR was performed on the cDNA in order to produce ast A, B, C, D gene fragments (**fig 9**) sans introns. The ast A and ast D genes were recombined into the D17 HR plasmid to form 2A AD and TPS AD plasmids; while the ast C and ast B genes were recombined into the MFE1 HR plasmids to form 2A CB and TPS CB plasmids.

A new batch of competent *E. coli* cells were produced and transformed with plasmids: 2A AD, TPS AD, 2A CB, and TPS CB. The resultant colonies for both the transformed cells and the positive Gibson control grew in unusual cluster patterns, clumping about certain areas instead of growing evenly throughout the plate (**fig 10**). Most of these colonies could not grow in ampicillin liquid culture; the few colonies that could grow in ampicillin liquid culture did not produce any recombinant plasmid. The competence kit was determined to have been expired, which suggests that the cells used were not competent and thus resulted in a very low success rate for transformation; it was hypothesized that those few cells that took up foreign DNA containing the ampicillin resistance marker secreted beta-lactamase into their immediate surroundings, degrading the ampicillin on the old plates and allowing non-transformed cells to grow around the transformed colonies in clusters as satellite colonies. To test this hypothesis, new ampicillin plates were made and a fresh batch of *E. coli* competent cells were produced from a new *E. coli* competence kit. Subsequent transformations with the new batch of competent cells and ampicillin plates produced normal colonies (**fig 10**).

All four plasmids were extracted and then sent out for sequencing. The 2A AD and TPS AD plasmids showed proper placement and incorporation of the astA and astD genes as well as the other DNA fragments (ERBV-1/Lip2t-pFB<sub>Ain</sub>) (**fig 7**). The 2A CB and TPS CB plasmids, however, possessed genes with several introns still remaining (**fig 11**). The number of introns

varied depending on the colony, but no colony produced plasmid which did not have at least one intron within the astC gene. This indicates that some population of both the astC and astB mRNA extracted from *A. terreus* contained introns, as it is unlikely to be from contaminating genomic DNA due to the DNase step used on the RNA extraction prior to cDNA synthesis. Furthermore, contaminating genomic DNA would only yield amplified genes with all introns remaining as opposed to partially remaining which would be the product of a spliceosome removing only some of the introns. It seems possible that a degree of intron retention is kept among certain genes when the cell does not need their resultant protein as a form of gene regulation, producing truncated proteins that are quickly recycled and avoiding the cost in energy associated with maintaining high levels of said proteins. In order to resolve this issue, primers were designed to amplify the astC exons using one of the CB plasmids as a template. This CB plasmid was also used as a template to PCR intronless astB. The astC exons were stitched together into a plasmid using Gibson Assembly and transformed into *E. coli*, then the resultant plasmid was used as a template for the amplification of intronless astC. The intronless genes were then recombined into the HR MFE1 plasmid to form 2A CB and TPS CB plasmids which were then transformed into *E. coli* and sequenced after miniprep showing proper placement of the genes and other genetic elements (**fig 7**).

### **Transformation and Verification of Engineered Strains of *Y. lipolytica***

Double transformation of *Y. lipolytica* with either AD or MVL recombinant plasmid and associated CRISPR plasmid on -uracil, -leucine plates yielded very few colonies if any grew at all (**fig 12**). This was remedied by increasing amount of each plasmid in transformation to > 2 µg, adding the two plasmids in approximately equal mass, using an overnight grown culture of *Y. lipolytica* in YPD (instead of plate colonies), and adding a 2 hour recovery step in YPD on a heat

block at 30°C for 2 hours following heat shock (**fig 12**). Half of the transformed cells were plated immediately while the other half were inoculated into -uracil,-leucine liquid culture which would be plated in the case that no colonies grew on the first plating.

The colonies were tested via 3 primer PCR (**fig 4**) to verify that the plasmid conducted homologous recombination with the appropriate genomic locus cut by the CRISPR plasmid. Good colonies were inoculated into rich 5-FOA YPD media to grow for a day and cure both plasmids and were then plated on YPD. Colonies from AD and MVL transformed cells were then patch-plated onto -uracil and -leucine plates to ensure plasmid dropout (**fig 13**). Those colonies which did not show growth on -uracil and -leucine plates were evaluated again with 3 primer PCR to verify genomic integration of the transformed genes. Positive colonies were grown overnight in YPD and then mixed with sterile glycerol in cryovials, resulting in the successful creation of glycerol stocks for the 2A AD, TPS AD, 2A MVL, and TPS MVL strains.

The new strains then underwent another round of transformation; the CB plasmids were transformed into the AD strains and the AD plasmids were transformed into the MVL strains. The process of strain development and verification repeated until the process was halted at the final 3 primer PCR step which consistently showed a 1 kb band on the negative control, indicating contamination of the PCR (**fig 14**). It was hypothesized that the contaminant stems from the highly concentrated plasmids used for transformation which can contain hundreds to thousands of nanograms of plasmid in one microliter. This DNA template contamination may extend to the surface of the workbench, the outside of pipettes, the PCR master mix, the primers, or may be aspirated into the pipette body itself. Working based on this assumption of extensive contamination, the workbench for three-primer PCR was changed, the PCR master mix was replaced, new primers were made from stock, and different pipettes were utilized. This resolved

the contamination issue and the 2A/TPS ABCD and 2A/TPS MVL AD strains were produced (**fig 14**). The MVL AD strains were then subject to another round of transformation with the CB plasmids, resulting in the four main engineered strains of: 2A ABCD (AD2), TPS ABCD (ADT), 2A MVL ABCD (AM2), and TPS MVL ABCD (AMT) (**Table 1**).

### **Metabolic Analysis of Engineered Strains**

The four engineered strains and PO1f were grown in synthetically defined (SD) media (YNB, amino acids, ammonium sulfate, 2% glucose) for 3 days. Cellular metabolites were extracted using the extraction solvent (40 methanol: 40 ACN: 20 water). The cell extracts and the aspterric acid standard were loaded onto the LCMS for analysis. The standard produced a peak at a retention time of 4.12 minutes at aspterric acid's m/z ratio: 265.14365. None of the strains displayed a peak at the appropriate retention time for aspterric acid (**fig 15**); instead, they produced strong peaks at a retention time of around 6.66 minutes representing some unknown metabolite. This peak was not replicated by the wild type strain, the magnitude of its signal at the retention time was small ( $3.7e+04$ ). This indicates that the metabolite is produced as a result of one (or multiple) of the engineered genes (likely the ast genes since the ADT and AD2 strains showed peaks despite not having been engineered to overexpress HMGCR or FPPS). The strains showed significant variability in the strength of their peaks: AMT ( $9.9e+08$ ), AM2 ( $9.6e+07$ ), ADT ( $1.9e+05$ ), and AD2 ( $4.6e+07$ ) with the MVL strains displaying the tallest peaks, suggesting that the metabolite is a terpenoid since the overexpression of HMGCR and FPPS enhances the production of terpenoids.

It was hypothesized that the unknown metabolite could be aspterric acid with a shifted retention time due to matrix effects from the cell's extracted metabolite environment. To test this hypothesis, the aspterric acid standard was spiked into the cells before and after extraction (**fig**



**16)** to observe the cell extract's effects on the retention time of aspterric acid. The retention time of aspterric acid remained unaffected from the presence of the cell extract, demonstrating that the engineered strains' peak represents a different metabolite with the same m/z ratio as aspterric acid.

It was then hypothesized that the unknown metabolite was an isomer of aspterric acid. To investigate this hypothesis, the AMT strain was grown in synthetically defined media with U13 glucose as the sole carbon source. The culture was then extracted and analyzed by the LCMS (**fig 17**) This resulted in a strong peak at an m/z ratio corresponding to 15 labeled carbons (280.19452), the majority (52.31%) of the detected signal from the unknown metabolite also corresponds to 15 labeled carbons. From this it can be concluded that the unknown metabolite has 15 carbons, the same amount as aspterric acid. With both the same number of carbons and m/z ratio as aspterric acid, it is most likely that the metabolite produced by the four engineered strains is an isomer of aspterric acid.

The MeOH extraction method only extracts cellular metabolites, not those present within the supernatant. It was thus possible that the engineered strains may have been secreting aspterric acid into the surrounding media. In order to test this, the engineered strains were grown again for 3 days in SD media and extracted with 1:3 acetone:EtOAc, with glass beads to break the yeast's cell walls. The organic phase was then harvested and evaporated completely and the metabolites were reconstituted in the extraction solvent MeOH:ACN:water. These extracts contained both the metabolites from the cells, as well as those from the supernatant. The extracts were then loaded onto the LCMS for analysis (**fig 18**) The engineered strains all displayed peaks at the retention time of aspterric acid which were orders of magnitude higher than the PO1f signal (AMT 3.3e+06, AM2 2.2e+06, ADT 3.7e+06, AD2 8.6e+06, PO1f 7.7e+04), but they produced

much stronger peaks at the retention time of the unknown metabolite (AMT 2.7e+08, AM2 2.2e+07, ADT 1.4e+07, AD2 2.5e+07). If aspterric acid is being produced by the engineered strains, then it is likely secreted into the supernatant and is heavily outweighed by the production of its isomer.

Since the MVL strains have additional copies of the HMGCR and FPPS genes, it is expected that this will affect the levels of metabolites within the terpenoid synthesis pathway, most significantly the metabolite mevalonate which is catalyzed by the HMGCR enzyme. Further metabolic analysis was performed on the MeOH and EtOAc extracts (**fig 19**). The resultant chromatograms showed strong peaks at around 4 minutes which may be mevalonate. In the MeOH extractions, unexpectedly, PO1f displayed the highest peak (AMT 3.2e+07, AM2 4.0e+07, PO1f 5.0e+07), the AMT strain also displayed a strong peak at a retention time of 9.56 minutes which was replicated in much smaller peaks with the other two strains (AMT 9.2e+08, AM2 3.0e+05, and 6.5e+05). In the EtOAc extractions, PO1f again displayed the highest peak (AMT 2.5e+07, AM2 2.6e+07, PO1f 5.6e+08) and the AM2 and AMT strains displayed a number of additional peaks at 8, 8.7, and 10.8 minutes. These additional peaks may be representative of metabolites present within the supernatant for the engineered strains that was captured by the EtOAc extraction. The low peaks at 4 minutes for the MVL strains may indicate that the gene copies are not being expressed and are thus not producing more mevalonate, but it is difficult to determine without a mevalonate standard to confirm the retention time as either 4 or 9.5 minutes.

### **Troubleshooting: RNA Analysis of Engineered Strains**

In order to investigate the cause of this lack of aspterric acid production among the engineered *Y. lipolytica*, the strains were grown in SD media and extracted of their RNA. This

was done to check for failure at the level of transcription i.e. determine if the incorporated heterologous genes were being transcribed into mRNA. cDNA synthesis was conducted using the RNA extracts as a template and the cDNA was in turn used as a template to amplify 1 kb regions within the astA,B,C,D genes (**fig 20**) All four engineered strains were found to produce astB and astC mRNA, but only the TPS strains produced astA and astD mRNA. Since the AMT and ADT strains were verified to have produced mRNA for the four ast genes, it seems unlikely that the cause of failure to generate aspartic acid is due to transcription.

The lack of mRNA of astA and astD for the 2A strains was hypothesized to be due to lack of genomic integration of the 2A AD plasmid into the D17 locus. To evaluate this hypothesis, new plates were streaked from glycerol stocks of the AM2 and AD2 strains and colonies were selected for three-primer PCR (**fig 21**). The AM2 strain produced 1 kb bands which indicated that the 2A AD plasmid had recombined in the MFE1 locus; however, the AD2 strain did not produce any bands (either 1 kb or 2 kb). This result signifies that PCR is not commencing with the sample and this lack of bands was demonstrated repeatedly with more AD2 colonies suggesting that there is some issue with the template (homologous regions within the MFE1 locus) within the AD2 colonies or some component within the AD2 colonies is inhibiting PCR. The stock of the 2A AD cells (which were used to make the AD2 strain) were also plated and tested with three-primer PCR, producing the same lack of bands. In order to resolve this situation, a new 2A AD strain was produced which was utilized to make a new AD2 (2A astABCD) strain. All four engineered strains were then plated from their stock and underwent three-primer PCR to verify genomic integration (**fig 22**). All four of the strains displayed 1 kb bands in PCR for genomic integration into the AXP, D17, and MFE1 loci.

It seemed strange that the TPS strains would produce mRNA of the ast genes, but the 2A strains would not (for ast A and D) despite using the same genomic loci for the same incorporated genes. It was hypothesized that the increased length of the mRNA in the 2A strains (since both genes would be transcribed onto the same transcript) would be more difficult for the reverse transcriptase to synthesize cDNA from which would in turn result in less template for the PCR. To investigate this, new RNA extracts were made from the AM2 and AD2 strains which were utilized for cDNA synthesis with a lengthened incubation step to account for long strands of mRNA. The resultant cDNA underwent PCR to check for transcription of ast genes; however, the gel controls showed contamination in all the ast PCR samples. After replacing the PCR master mix, remaking primers from stock, changing pipettes, switching workbenches, remaking the RNA extraction and cDNA synthesis, and careful design of negative controls, it was determined that one of the components of cDNA synthesis was contaminated (**fig 23**). This contamination makes it difficult to determine for certain if the AM2 and new AD2 strains produce mRNA for the ast genes.

### **Troubleshooting: Evaluation of 2A System**

The lack of mRNA produced from the 2A strains for certain genes raises the concern that there may be issues with transcription of the 2A gene constructs. This issue, however, can likely be resolved by remaking of the strains. A more significant concern is the possibility that the 2A sequence does not cleave the transcript into two peptides (which cannot be seen from analyzing the mRNA) since the ERBV-1 2A peptide cleavage efficiency has not been evaluated yet within *Y. lipolytica*. In order to quickly assess the 2A system's ability to produce two separate, functional proteins, a series of plasmid constructs were formed to test fluorescence using hrGFP.

The native BGL1 gene was amplified from genomic DNA and the hrGFP gene was amplified from an HR plasmid. These two genes were then isolated and assembled into digested HR A08 plasmids to form three recombinant plasmids via Gibson assembly: a TPS plasmid, a 2A plasmid, and a plasmid with a short linker separating genes which were then transformed into *E. coli*. The hrGFP gene is placed as the second gene within each of the aforementioned plasmid constructs so as to evaluate the effectiveness of the ERBV-1 sequence in separating proteins. It was hypothesized that the TPS strain would produce free hrGFP which would result in fluorescence and that the linker strain would produce fused protein that would not fluoresce; the 2A strain would then either fluoresce if there was a high degree of protein separation or would not fluoresce if there was a significant lack of protein separation, resulting in fused, inactivated proteins.

The TPS BGL-hrGFP (TBG), the 2A BGL-hrGFP (2BG), and the linker BGL-hrGFP (LBG) plasmids were all harvested from transformed *E. coli* with midpreps after being sequence-verified and were then transformed into *Y. lipolytica* with the CRISPR A08 plasmid. The transformed cells were then PCR-checked, grown in 5-FOA YPD, patch-plated onto -uracil and -leucine plates to confirm plasmid dropout, and PCR-checked again to confirm genomic integration. After several attempts, all three strains (**table 1**) were successfully validated and were made into glycerol stocks. The TBG, 2BG, LBG, and PO1f strains were grown in 5 ml of YPD and a few microliters of each were placed on microscope slides to be viewed by a fluorescence microscope which exposed the slides to 276 nm wavelength of light. The resultant images were large green blurs (**fig 24**) from which no useful information could be determined. This was hypothesized to be caused by autofluorescence from components within the YPD media. In order to test this, the strains were grown in SD media and spun down to be

resuspended in 1X PBS buffer in order to provide a “cleaner” media for fluorescence. The images from this batch suffered from significantly less background fluorescence, but it was still present; furthermore, the high density of cells made for crowded and unclear images.

To improve the quality of the images, the TBG, 2BG, LBG, and PO1f strains were grown in minimal media (YNB, ammonium sulfate, uracil, leucine, 1% glucose) overnight and then diluted 1:10 in more minimal media and were imaged by a fluorescent microscope (**fig 25**). The use of minimal media resulted in images that contained almost zero autofluorescence. As hypothesized, the LBG strain showed no fluorescence, indicating that the fusion BGL1-GFP protein is not fluorescent, while the 2BG strain displayed a great deal of fluorescence, demonstrating successful transcription and translation of the 2AS construct. This suggests that the 2BG strain is producing separate hrGFP and BGL1 proteins since the linker strain has shown that the hrGFP-BGL1 fusion protein is non-fluorescent. It is still possible (though unlikely) that the ERBV-1 peptide sequence acts as a linker that allows the BGL1-hrGFP fusion protein to retain fluorescence. In order to have certainty regarding the effectiveness of the ERBV-1 sequence in *Y. lipolytica*, a 2AS construct will need to be designed with His tags appended to the end of the second gene to allow for protein purification and SDS PAGE to evaluate the size of the purified protein allowing for determination of the relative amounts of cleaved and fused protein.

## **Future Directions**

Further investigation into the unknown metabolite's mechanism of synthesis as well as the production of aspterric acid in *Y. lipolytica* may prove to be interesting research opportunities. Even once both these issues are resolved, there will still remain many directions by which to expand this project's scope. These include: evaluating production on growth with differing substrates, overexpression of other relevant bottlenecking enzymes (i.e. isomerase and acetyl coA carboxylase), downregulation of competing biosynthetic pathways via CRISPRi (i.e. fatty acid synthesis), evaluating production scaleup in larger cultures, and further. With our understanding of and genetic tools for *Y. lipolytica* continuing to expand, it seems likely that *Y. lipolytica* will take an increasingly important role in shaping natural product production in the future.

## Figures

**Figure 1: The Mevalonate Pathway**

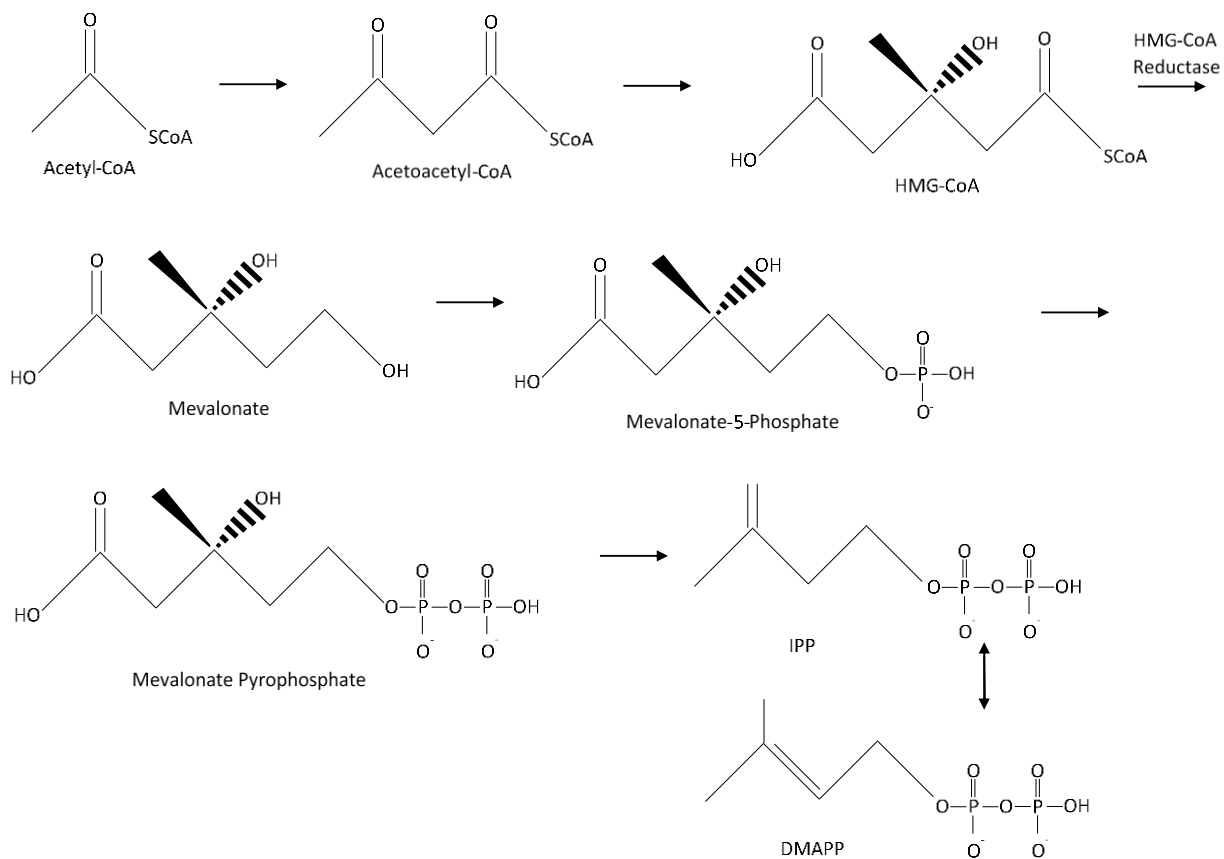


Figure 1: The Mevalonate Pathway catalyzes the conversion of Acetyl-CoA into the isoprenoid precursors: IPP and DMAPP. 3 molecules of Acetyl-CoA, 3 molecules of ATP, and 2 molecules of NADPH are required to produce 1 molecule of IPP or DMAPP. The rate limiting step in this pathway is the conversion of HMG-CoA to mevalonate catalyzed by the enzyme HMG-CoA Reductase which utilizes NADPH.



**Figure 2: Aspterric Acid Synthesis**

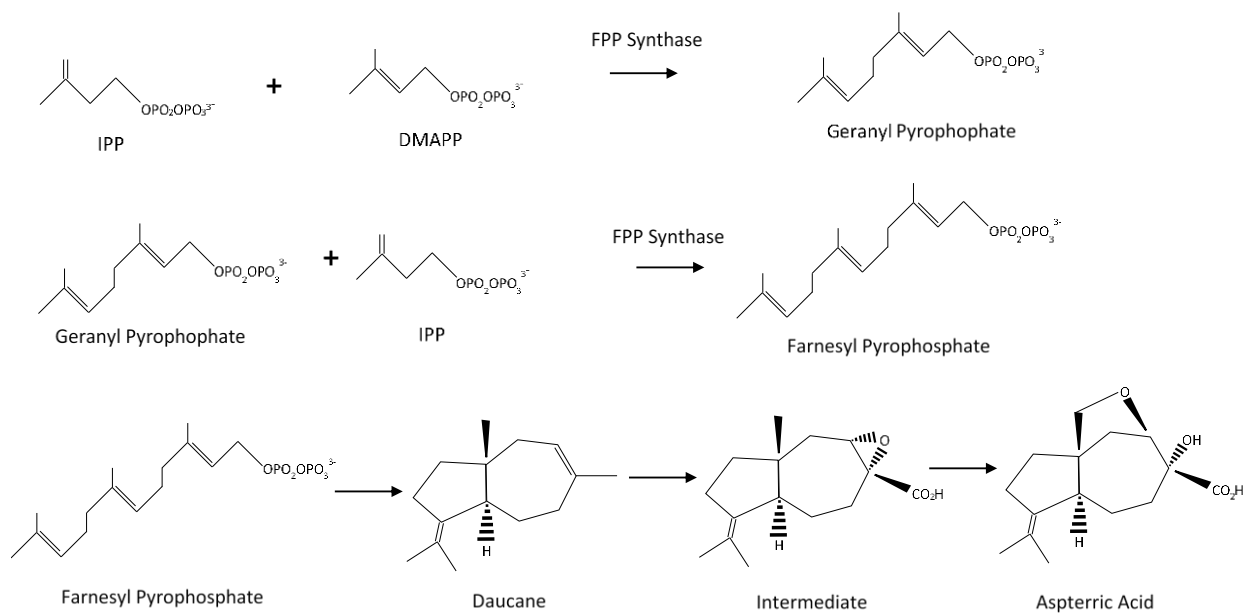


Figure 2: FPP Synthase chains together IPP and DMAPP precursors to form the monoterpene geranyl pyrophosphate and the sesquiterpene farnesyl pyrophosphate. This sesquiterpene is then cyclized by the *astA* enzyme and then sequentially oxygenated by the *astB* and *C* enzymes to form aspterric acid [2]. The *astD* gene codes for a DHAD homologue which does not bind to the competitive inhibitor aspterric acid and thus confers self-resistance to the toxin.

**Figure 3: Plasmid Design**

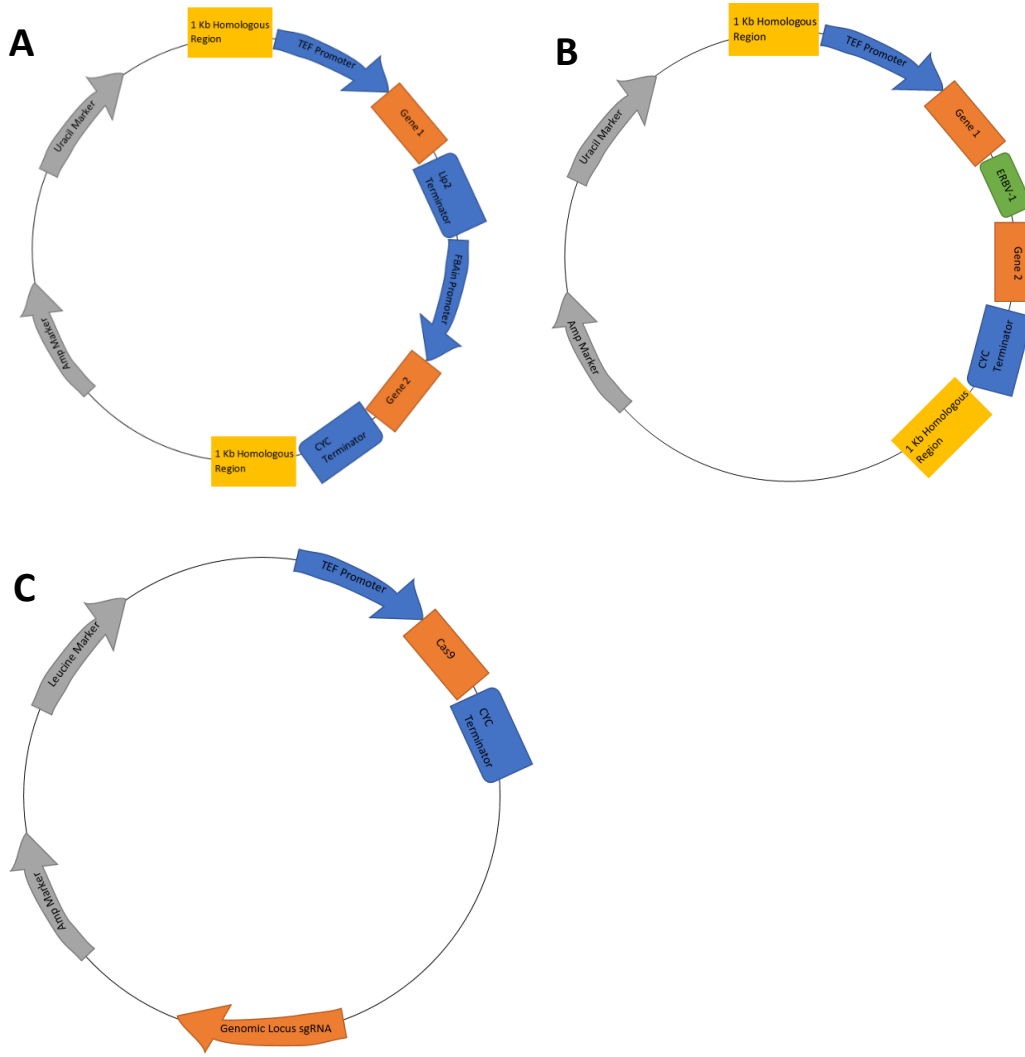


Figure 3: (A) The two-promoter system (TPS) plasmid contain two genes of interest separated by their own respective promoter and terminator; the first gene has a TEF promoter and a natively sourced Lip2 terminator, while the second gene has a natively sourced constitutive FBAin promoter and a CYC terminator. This set of genetic elements is flanked by two 1 kb regions which are homologous to a specific genomic locus (A08, AXP, D17, MFE1, or XPR2) which will allow integration of the contained DNA into this locus via homologous recombination at the site of a double-strand break initiated by CRISPR/Cas9. The plasmid also contains an ampicillin marker to allow for plasmid production in *E. coli* and a uracil marker to allow for selection of transformed PO1f cells. (B) The 2A system (2AS) plasmid also contain an ampicillin and uracil marker as well as 1 kb homologous regions. This plasmid only has a TEF promoter and a CYC terminator which flanks two genes separated by an ERBV-1 2A peptide sequence. This ERBV-1 sequence will initiate ribosomal skipping which will result in the single mRNA transcript producing two peptides. (C) The CRISPR plasmid contains a leucine marker to allow for selection of PO1f colonies along with the corresponding HR plasmid, it also contains an ampicillin marker. The plasmid contains the gene coding for Cas9 which will form a complex with the sgRNA encoded in the plasmid which will allow for targeted cuts at a specific genomic locus.

**Figure 4: PCR Screening for Genomic Integration**

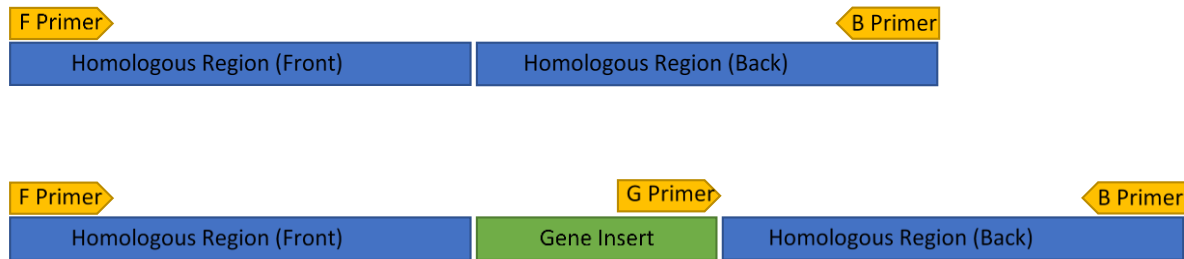
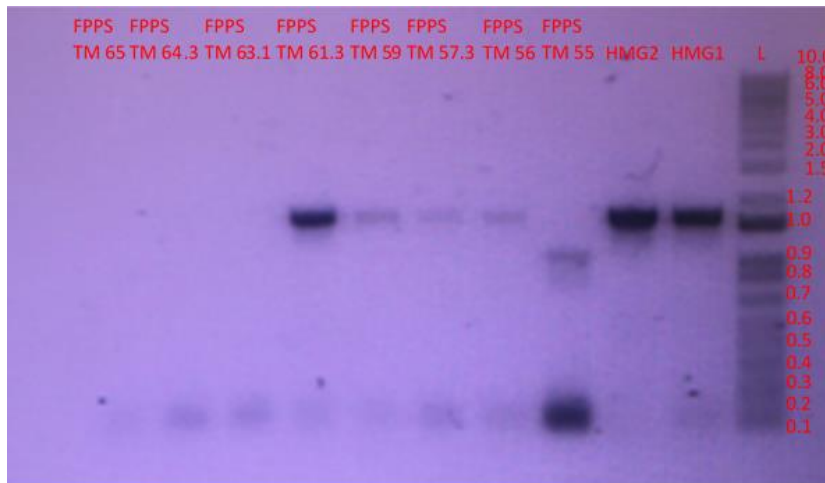


Figure 4: The three primer PCR encounters one of two cases: when there is no integration of genes into a specific genomic locus, the forward primer and reverse primers which bind to the homologous regions produce a 2 kb PCR product which is the case shown at the top. When there is integration of genes into a target genomic locus (as shown in the bottom image), the third primer binds to the CYC terminator region of the genetic insert and forms a 1 kb PCR product with the reverse homologous primer. The forward homologous primer cannot form a PCR product as the extension time during PCR is kept short.

## Figure 5: Preparation of DNA Fragments

**A**



**B**

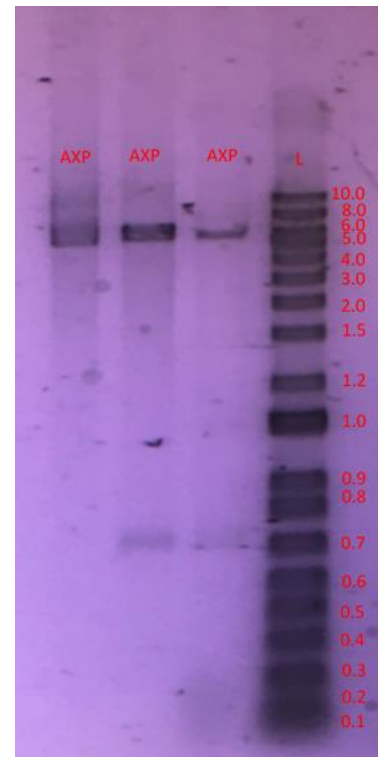


Figure 5: (A) HMGCR gene is broken up into two fragments HMGCR1 (GA2, GA4) and HMGCR2 (GA 1, GA 51) to enhance yield of PCR. Eight samples of the FPPS PCR reaction are run at different melting temperatures so as to determine the optimal  $T_m$  for FPPS amplification. (B) The HR AXP plasmid is double digested by restriction enzymes: *pteI* and *nheI*, producing a linearized vector and a faint hrGFP band.

**Figure 6: Lip2t-pFBain Fragment PCR**

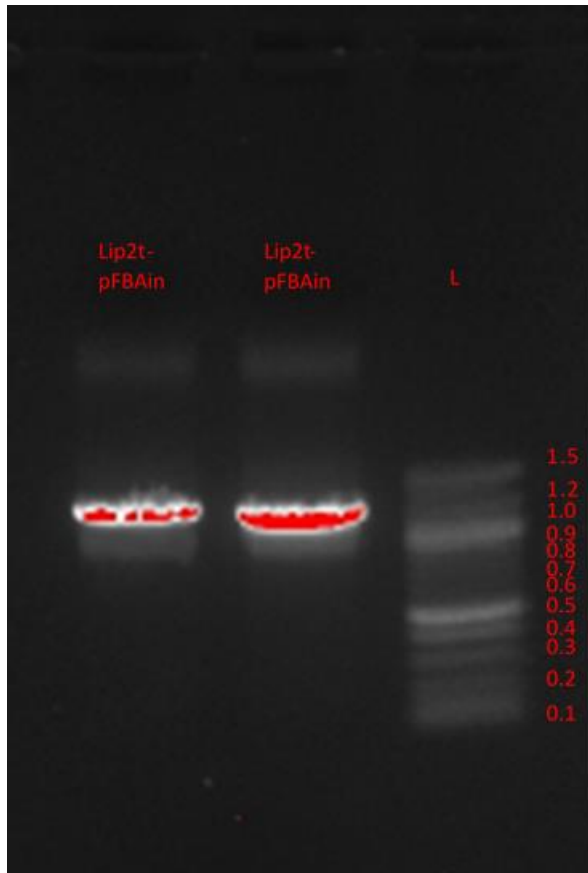


Figure 6: Lip2t-pFBain (P28,29) terminator-promoter DNA fragment (1.2 kb) was amplified from template recombinant TPS MVL plasmid. The fragment was gel extracted for use in Gibson Assembly in order to increase success rate of recombining TPS plasmids.

**Figure 7: Sequence Verification of Recombinant Plasmids**

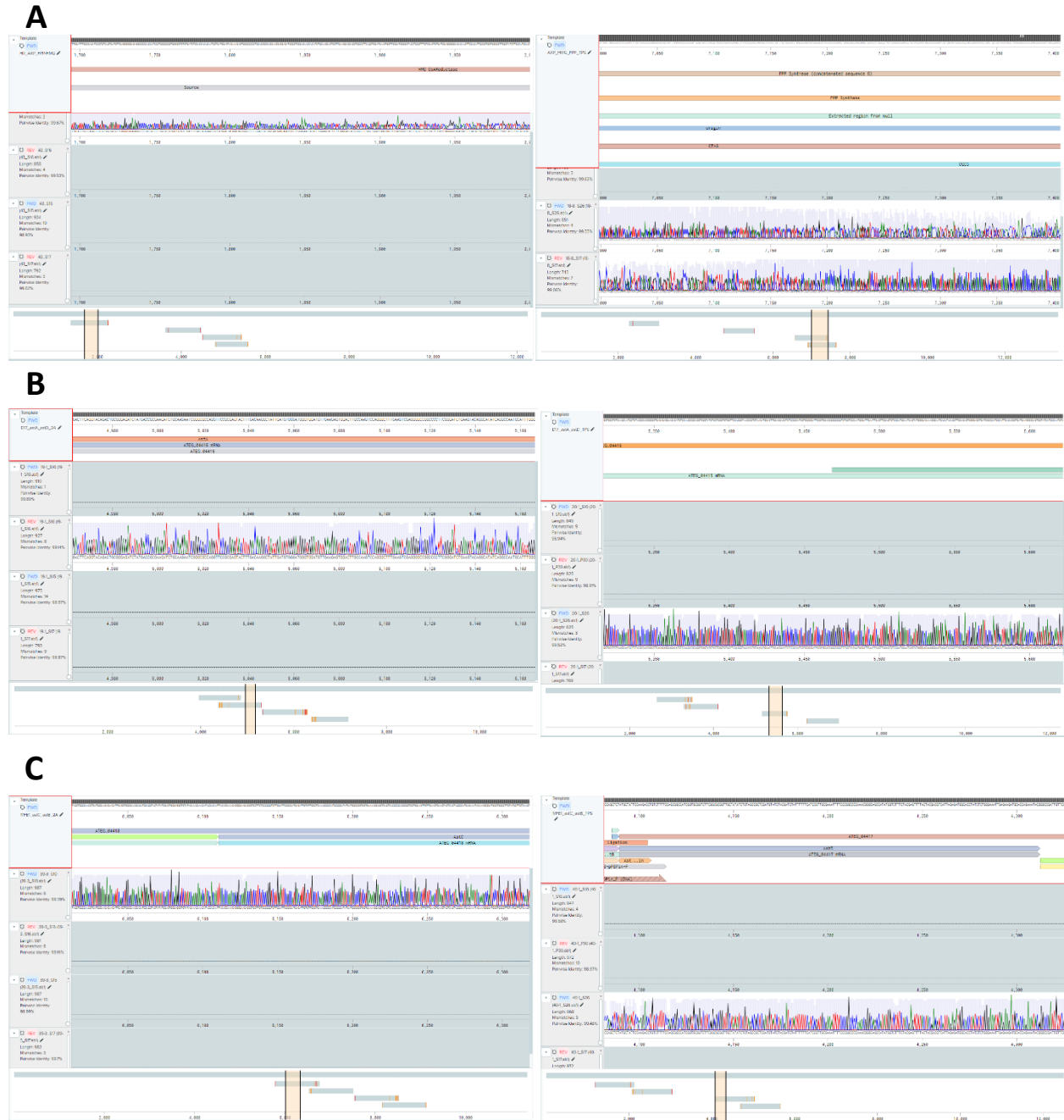


Figure 7: (A) Sequencing is done on short reads from sequencing primers sent to Laragen and aligned to the in silico recombinant plasmid. Since the genes are amplified via high-fidelity polymerase (Q5 and Phusion), the sequencing mainly serves to determine correct placement of the genes and DNA fragments in the recombinant plasmid. The left sequence is for the 2AS and the right is for the TPS. The sequence reads shows that HMGCR and FPPS have been integrated into the HR AXP plasmid between the TEF promoter and the CYC terminator. (B) The sequencing results show that astA and astD were incorporated into the HR D17 plasmid and that astA has been amplified without introns. (C) The sequencing results show that astC and astB were stitched into the HR MFE1 plasmid and that neither gene contain introns.

**Figure 8: ast Gene Introns and Exons**

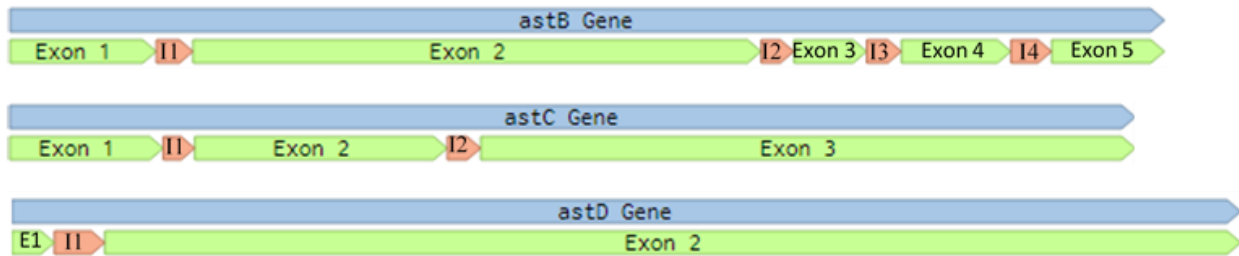
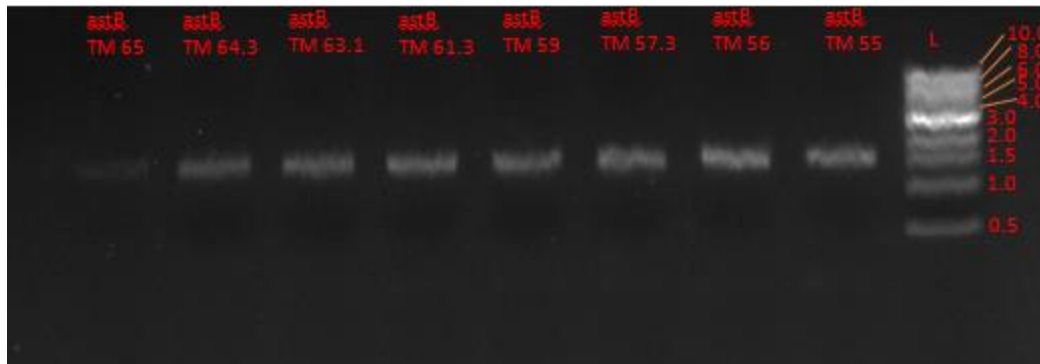


Figure 8: The astA gene contains no introns, however the other ast genes contain them. As depicted in the image above, AstD has one intron, astC has two, and astB has four introns. These introns all contain stop codons in-frame which will result in truncated protein if they are not removed before translation; they are normally spliced out of the RNA transcript in *A. terreus*.

**Figure 9: PCR of *A. Terreus* cDNA**

**A**



**B**



Figure 9: (A) Gradient PCR using cDNA as a template to amplify *astB* (1.5 kb) at a variety of differing melting temperatures. (B) Gradient PCR on cDNA to amplify *astC* (1.6 kb) at a variety of different  $T_m$ . The yield of *ast* genes amplified from cDNA was low, so numerous digested gel slices of the same gene were run through the same spin column in order to produce *ast* genes with sufficient concentration to be useful in Gibson Assembly.



**Figure 10: Clustered Colonies and Normal Colonies**

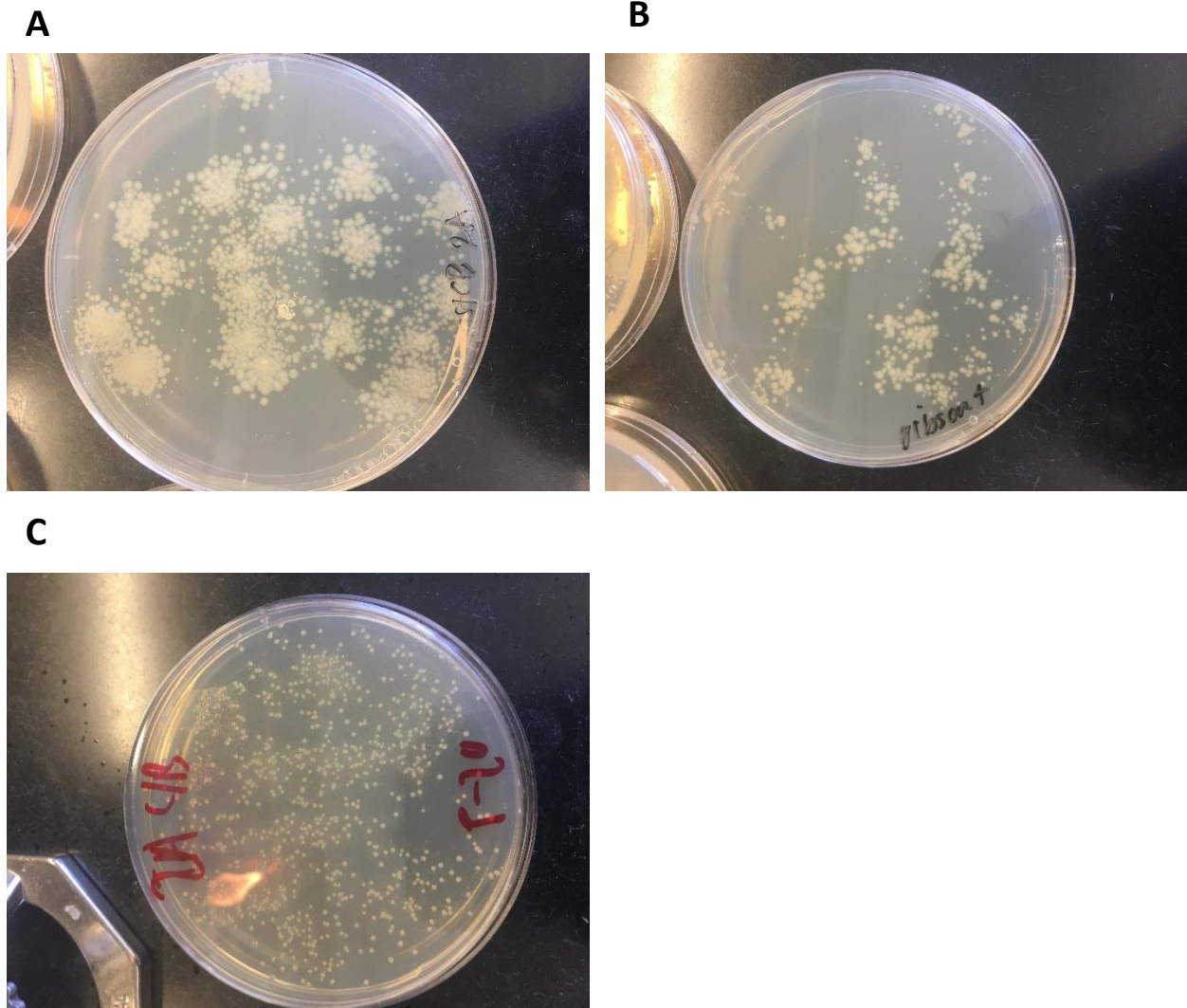


Figure 10: (A) *E. coli* transformed by 2A CB plasmid grows in clusters on ampicillin plate. Most of these colonies cannot grow in ampicillin liquid culture and those that do, fail to produce recombinant plasmid. (B) The cluster-like growth is also seen on the Gibson positive control, indicating that this unusual behavior is not a result of the parts used to form the recombined plasmid. (C) After making new ampicillin plates and competent cells, 2A CB transformed *E. coli* cells form normal-looking colonies which can grow in ampicillin liquid culture.

**Figure 11: Intron Retention in astC cDNA**

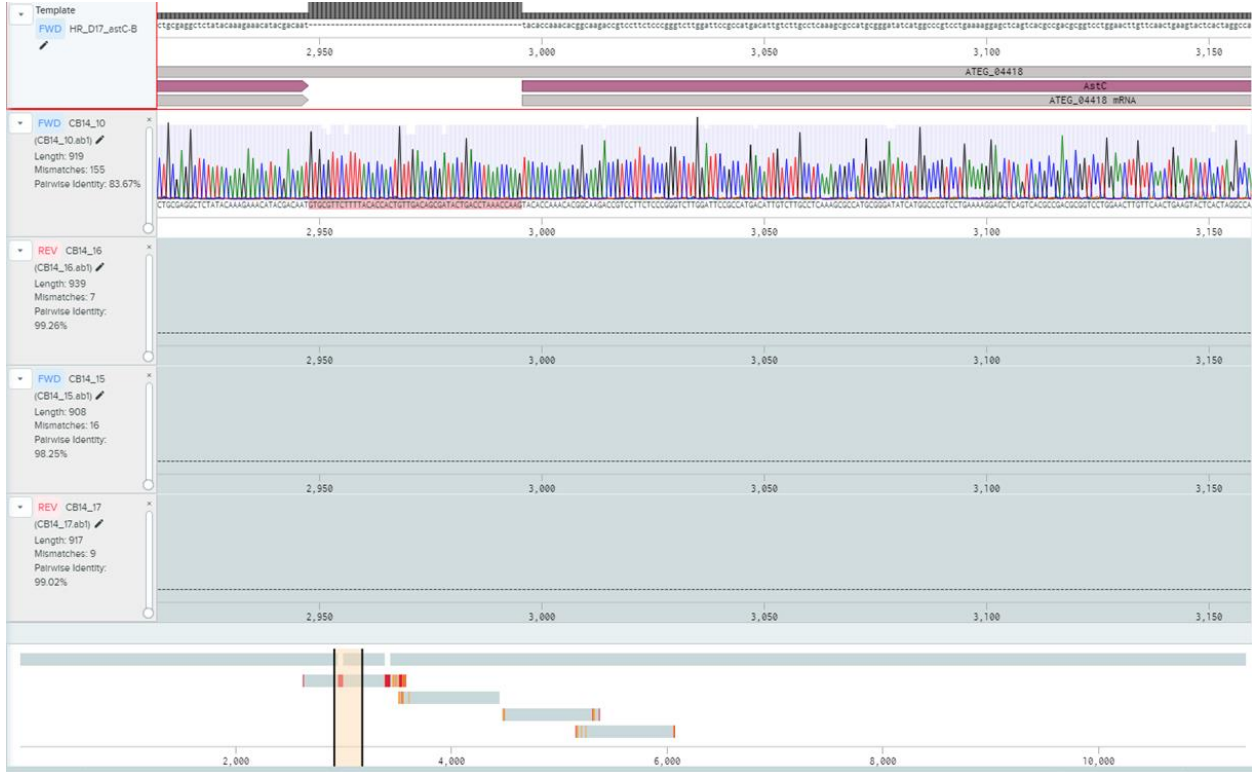
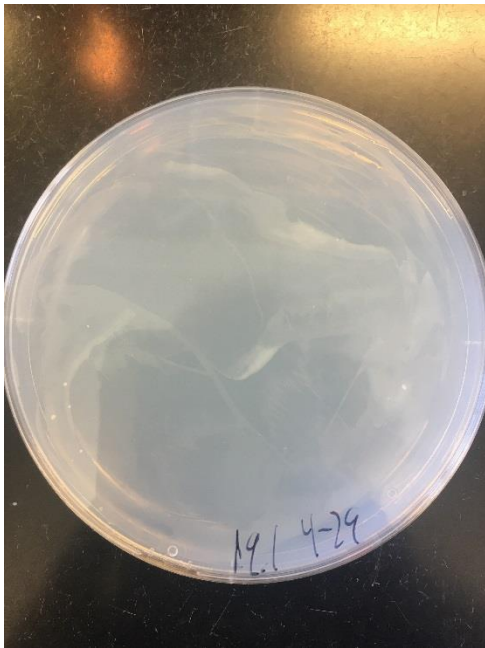


Figure 11: Sequencing of 2A CB plasmid containing *astC* and *astB* amplified from *A. terreus* cDNA using four sequencing primers shows that both introns still remain in the *astC* gene. Some colonies contain varying amounts of introns for *astC* and *astB*, but no colony produced a plasmid with no introns in *astC*

**Figure 12: Method Comparison of Growth for 2A AD and TPS MVL Transformation**

**A**



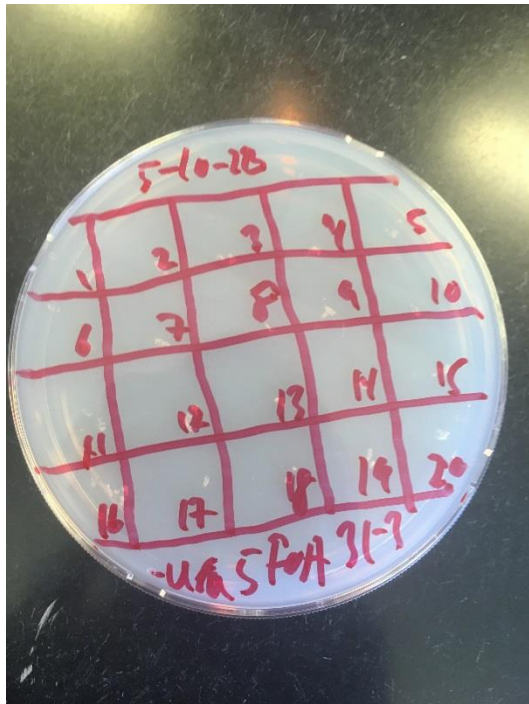
**B**



Figure 12: (A) 2A AD plasmid transformation into *Y. lipolytica* without overnight culture and recovery phase produces few colonies. (B) TPS MVL plasmid transformation into *Y. lipolytica* using overnight culture and 2-hour recovery phase yields many colonies.

Figure 13: Plasmid Dropout Verification

A



B

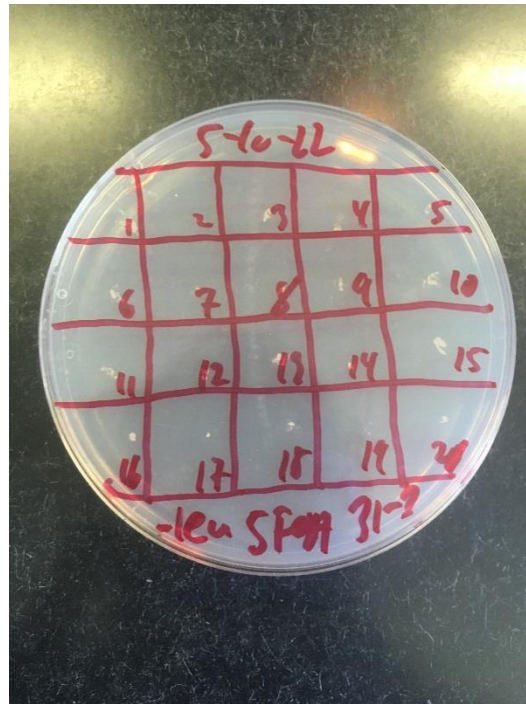
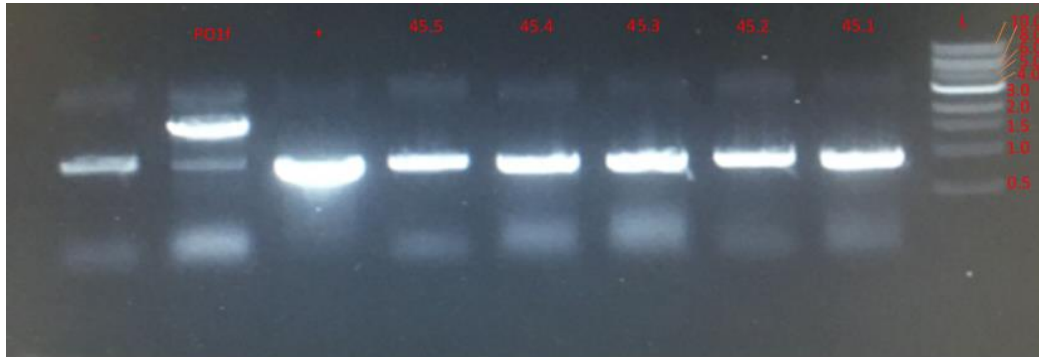


Figure 13: (A) The -uracil plate shows growth on colonies 9 and 19, indicating that these colonies from the TPS AD plate still possess the recombinant plasmid. (B) The -leucine plate shows growth from colonies 10 and 15, indicating that these colonies have not cured the CRISPR plasmid.

**Figure 14: Three Primer PCR Contamination**

**A**



**B**

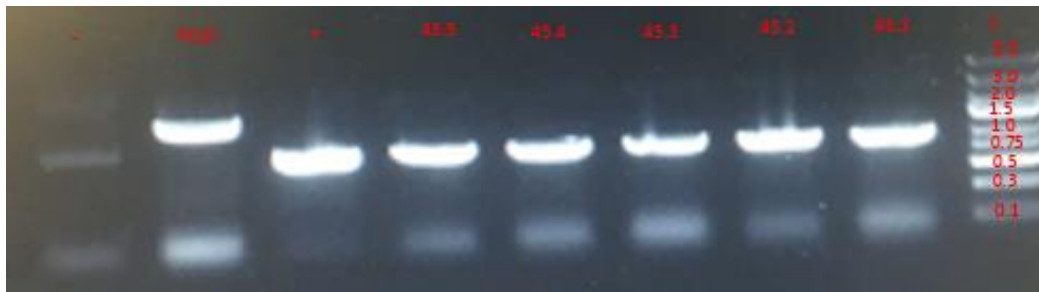


Figure 14: (A) The negative control is the PCR mix without added sample, this control is producing a 1 kb band, indicating template contamination of the PCR mix. As a result of this contamination, the PCR check for genomic integration of the 2A AD into five colonies of the 2A MVL strain (45.1-5) are inconclusive since the 1 kb band could be amplified from the template contaminant, instead of the genomic DNA of the sample, as is the case with the PO1f sample which produces two bands, one at 2 kb (from genomic DNA) and one at 1 kb (from contaminant) (B) The negative control does not show a band, which indicates that the 1 kb bands produced from the five 2A MVL strains are from genomic DNA and thus that the 2A AD construct has been integrated into the genome.

**Figure 15: Aspterric Acid Signal Among Engineered Strains, PO1f, and Standard**

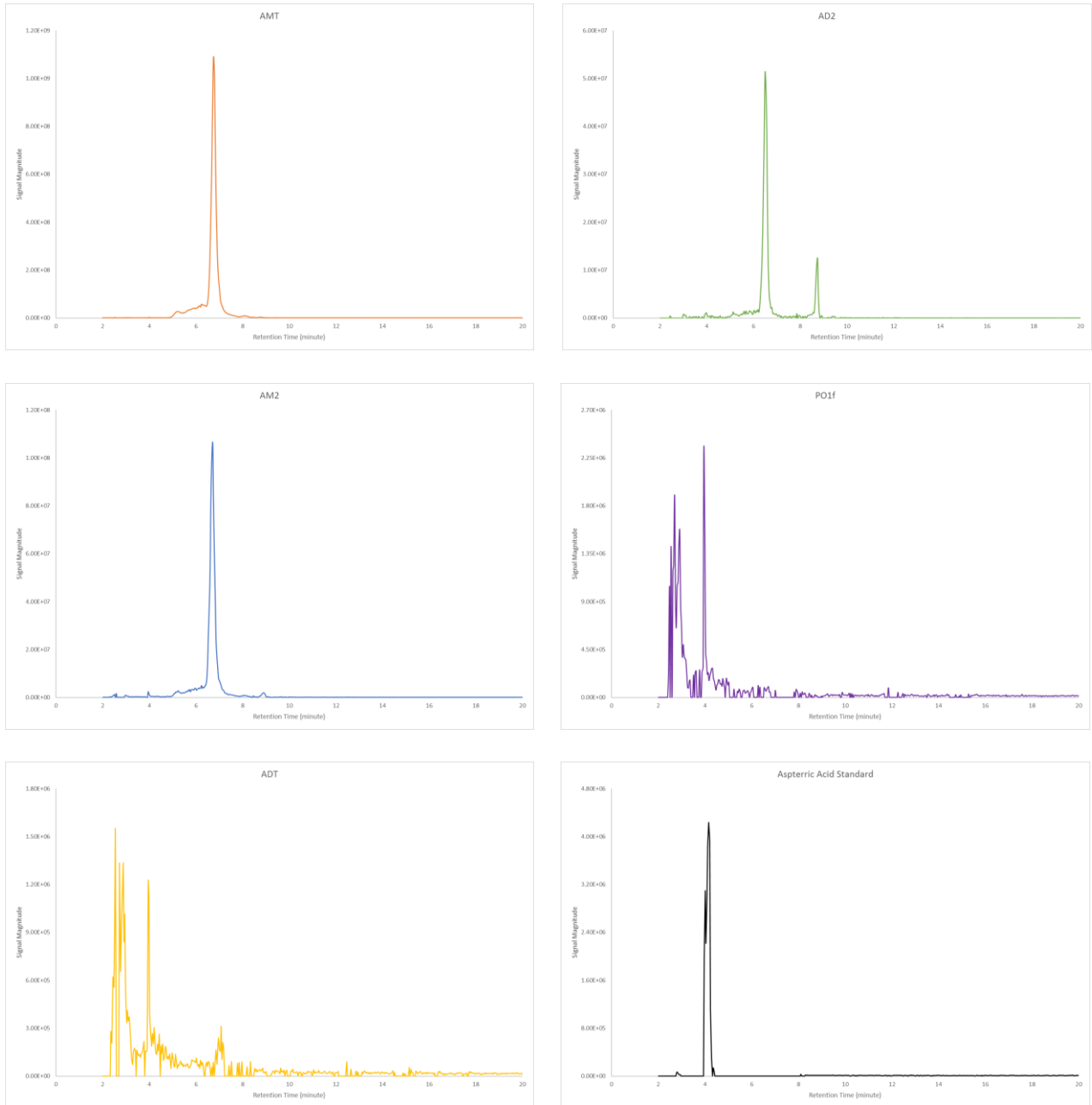


Figure 15: The above chromatograms show the signal intensity over retention time at aspterric acid m/z ratio: 265.14365. This utilizes the MeOH protocol of extraction which harvests cellular metabolites only. The engineered strains AMT, AM2, and AD2 show strong peaks at 6.66 minutes, while the PO1f and ADT strains share a similar signal pattern. The aspterric acid standard shows a strong, single peak at 4.12 minutes which indicates that the metabolite representing the AMT, AM2, and AD2 peaks is not aspterric acid, but is instead likely an isomer.

**Figure 16: Aspterric Acid Spike-in**

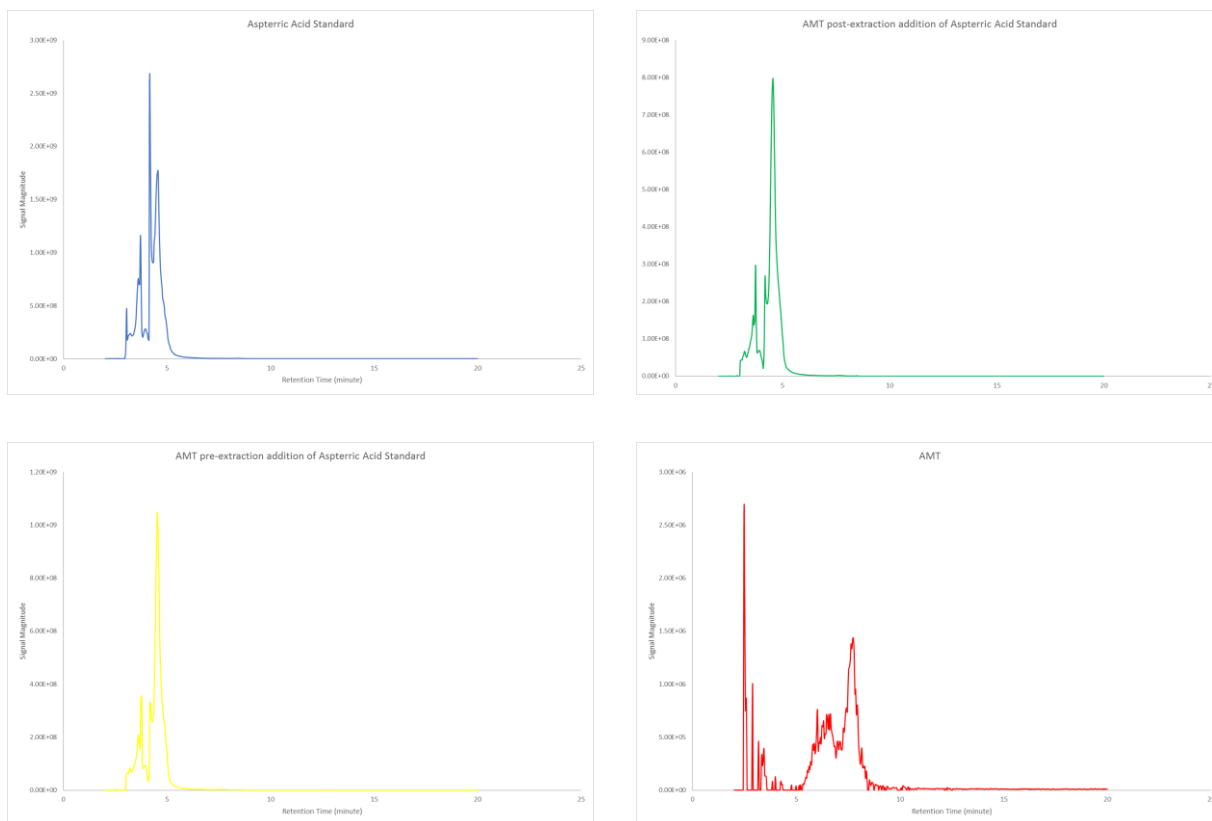


Figure 16: The above chromatograms show signal over retention time for the aspterric acid standard (blue) and AMT strain with the standard added before extraction (yellow), after extraction (green) and without standard added (red). The retention time of aspterric acid remained unchanged both when added before the extraction and after the extraction, indicating that the cell's contents have little effect on this metabolite's retention time (insignificant matrix effects). This confirms that the metabolite produced by the engineered strains is not aspterric acid with a shifted retention time, but is instead a different metabolite altogether.

**Figure 17: Carbon Labeling of Unknown Metabolite**

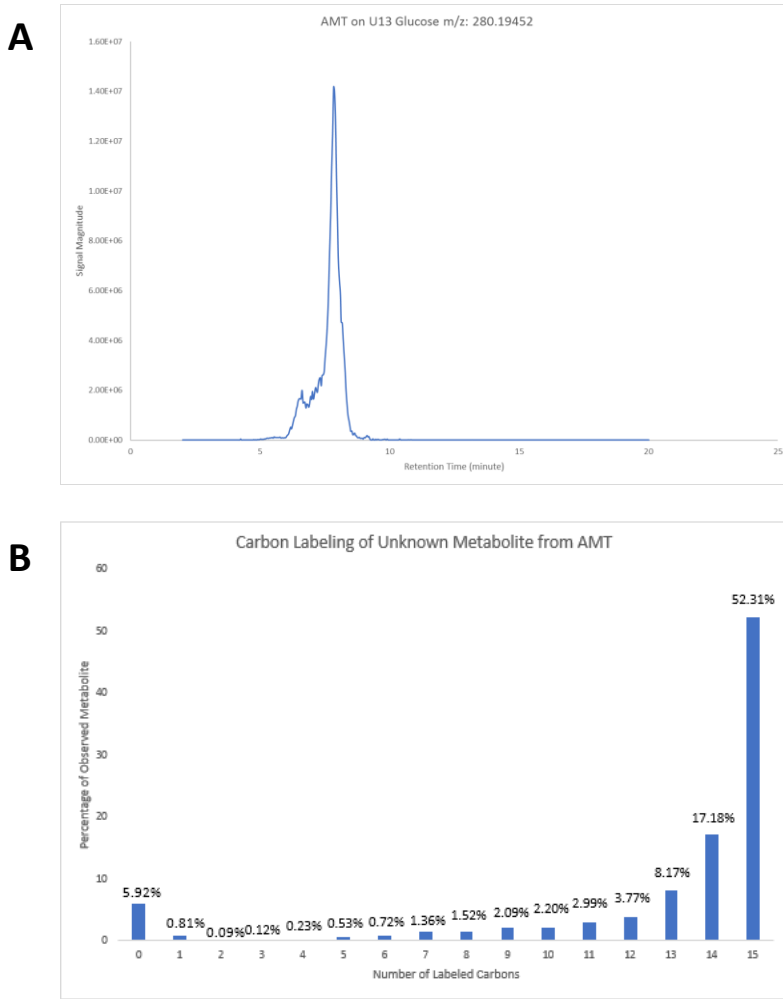


Figure 17: (A) The AMT strain was grown on U13 Glucose as its sole carbon source for three days before extraction. The chromatogram shows a strong signal for the unknown metabolite at  $m/z$ : 280.19452 corresponding to 15 labeled carbons. (B) The graph shows that the majority of the signal (52.4946%) received for the metabolite by the LCMS has 15 carbons labeled which indicates that the metabolite has 15 carbons and is likely an isomer of aspterric acid.



**Figure 18: Aspterric Acid Signal of EtOAc Extraction of *Y. lipolytica* Strains and Standard**

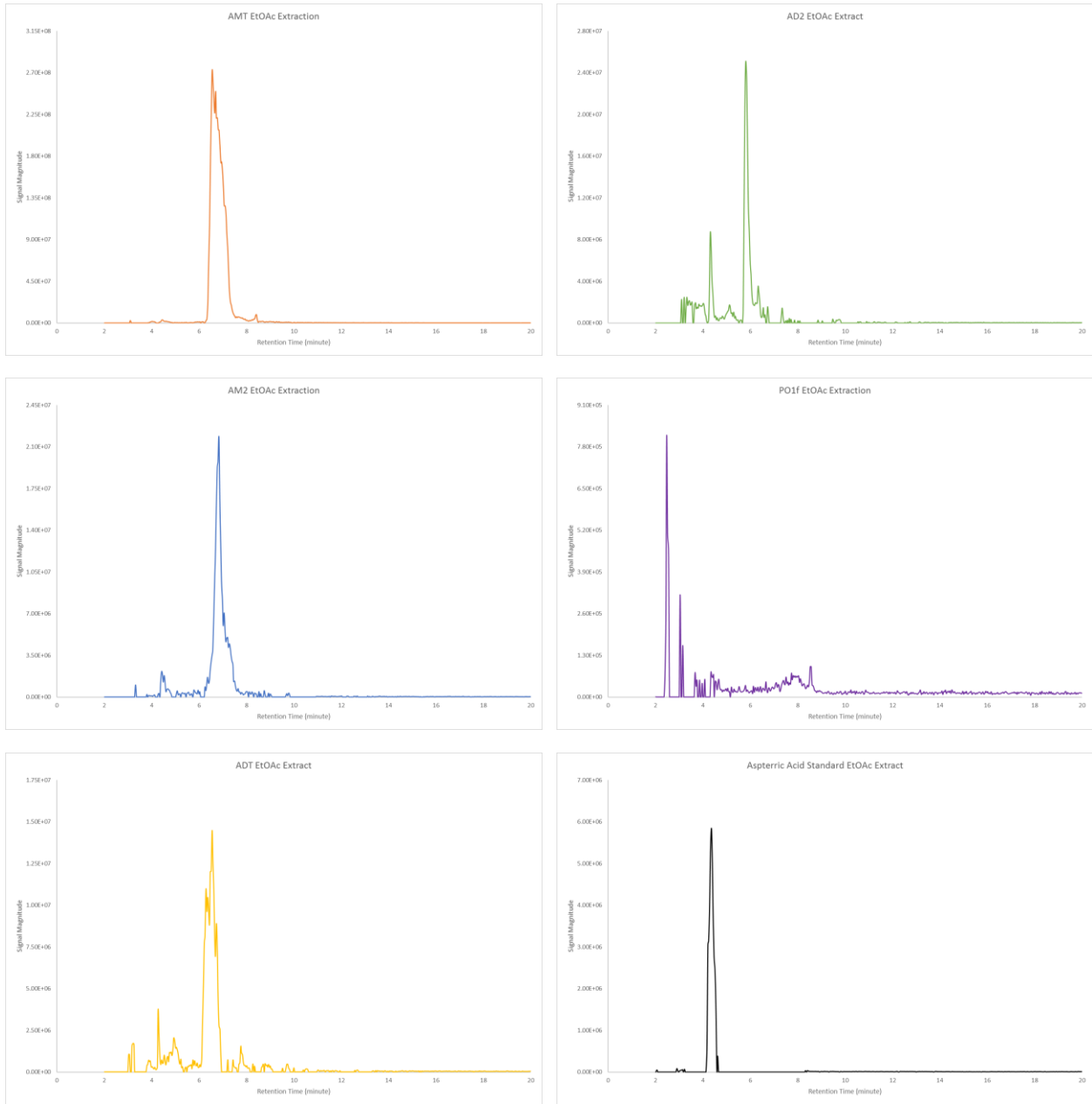


Figure 18: The above chromatograms show signal over retention time for the *Y. lipolytica* strains using the EtOAc extraction protocol which extracts both metabolites from within the cell and from those within the supernatant. The extracts now show peaks at aspterric acid retention time (AMT 3.3e+06, AM2 2.2e+06, ADT 3.7e+06, AD2 8.6e+06) while the wild type peak was negligible (PO1f 7.7e+04). These peaks, however, are dwarfed by the much larger peaks at the unknown metabolite's retention time (AMT 2.7e+08, AM2 2.2e+07, ADT 1.4e+07, AD2 2.5e+07). Even if the strains are producing aspterric acid, it is insignificant in comparison to the production of its isomer.

**Figure 19: Mevalonate Levels of EtOAc and MeOH Extracts of MVL Strains and PO1f**

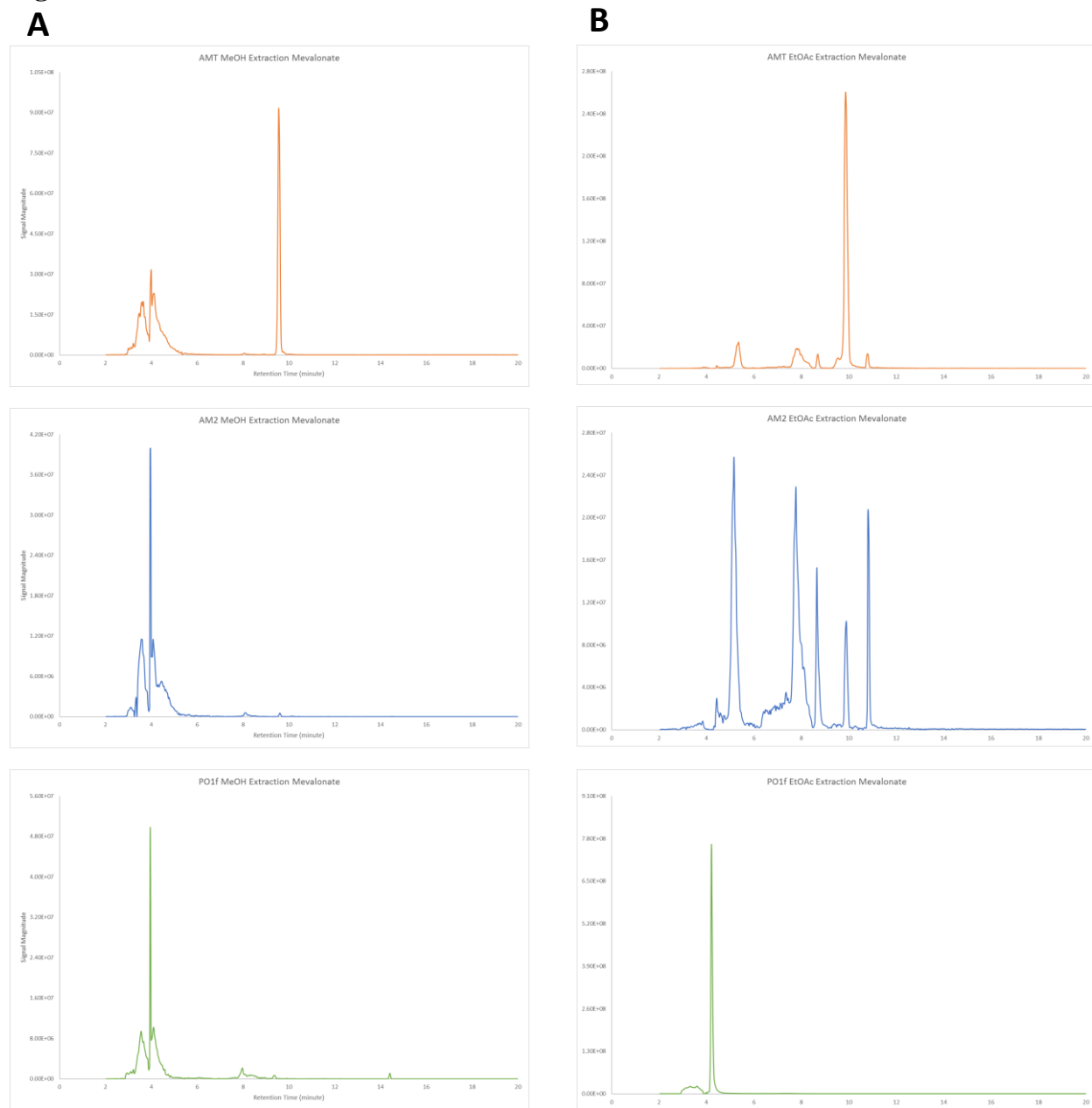


Figure 19: (A) The chromatograms show signal over retention time for the AMT, AM2, and PO1f strains at the m/z ratio for mevalonate (147.06651) utilizing the MeOH extraction protocol. The strains all share a strong peak at around 4.0 minutes making it likely to be the peak representing mevalonate (AMT  $3.2 \times 10^7$ , AM2  $4.0 \times 10^7$ , PO1f  $5.0 \times 10^7$ ) with PO1f displaying the highest peak. The AMT strain also has a large peak at a retention time of 9.56 minutes which is only present in the other two strains as small peaks. (B) These chromatograms utilize the EtOAc extraction protocol and thus contain metabolites from both the cell and supernatant. The PO1f strain has a much higher signal for mevalonate than the other strains (AMT  $2.5 \times 10^7$ , AM2  $2.6 \times 10^7$ , PO1f  $5.6 \times 10^8$ ). The AMT still possesses a strong peak at 9.5 minutes, as well as smaller peaks at 8, 8.7, and 10.8 minutes which are replicated more prominently in the AM2 extract. It seems likely that some of these peaks are from metabolites within the supernatant, captured in this extraction. Due to the lack of a mevalonate standard, it is difficult to determine which peak (those at 4 or 9.5 minutes) represent the metabolite

Figure 20: cDNA PCR of engineered strains and PO1f for astA,B,C,D

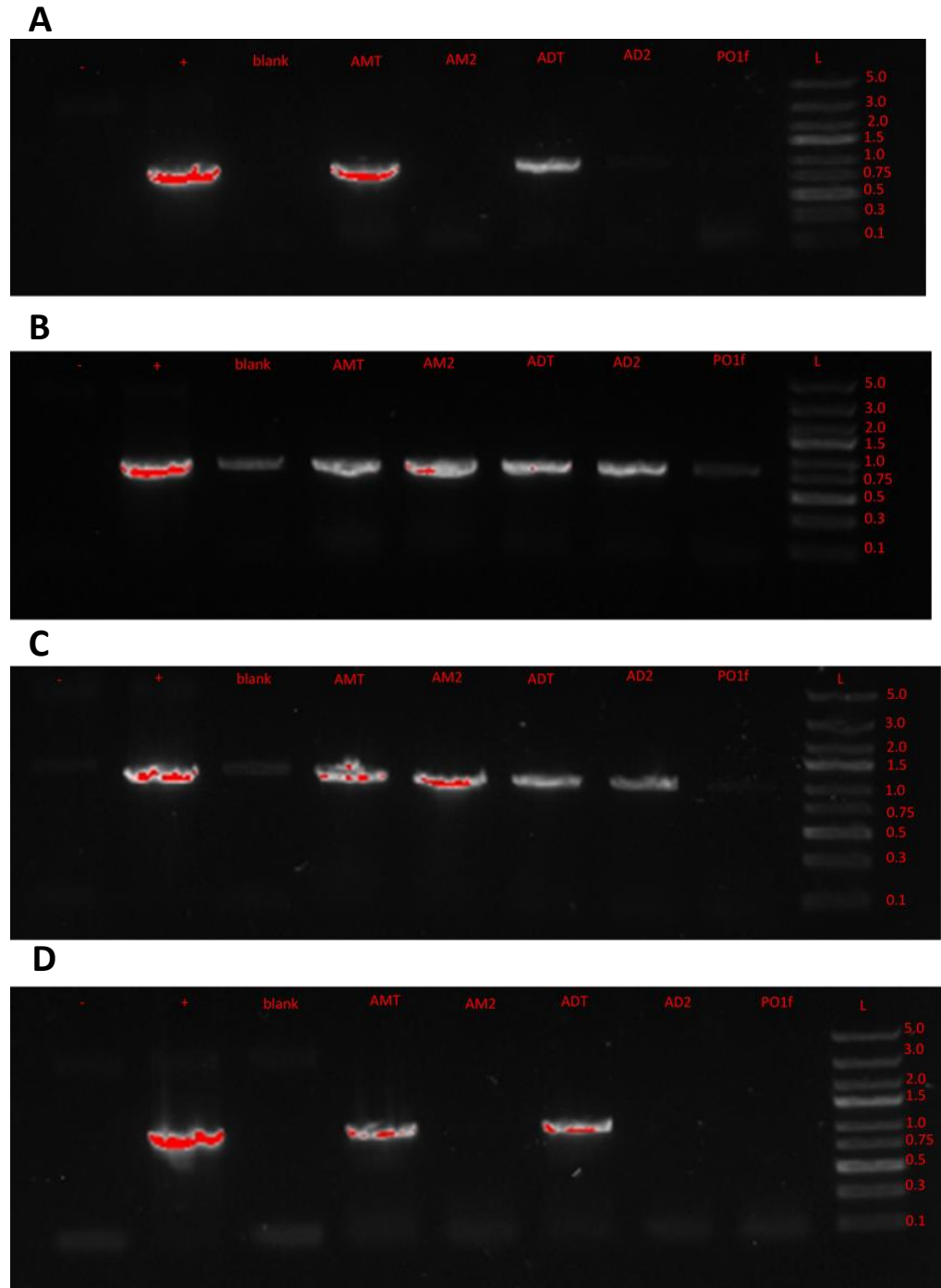


Figure 20: (A) PCR of 1kb astA region (P60,61) for PO1f, AD2, ADT, AM2, and AMT, as well as a blank containing a small sample of the RNA extract to check for gDNA contamination in the PCR, a positive control using one of the recombinant plasmids containing ast AD, and a negative control that does not contain cDNA to check for PCR mix contamination. The AMT and ADT strains showed strong bands for astA, while the AM2 and AD2 strains did not. (B) PCR of 1kb astB region (P58,59) from cDNA showed bands for all four engineered strains. (C) PCR of 1kb astC region (P56,57) from cDNA showed bands for all four engineered strains (D) PCR of 1kb astD region (P54,55) showed bands for AMT and ADT, but not for AM2 and AD2.

**Figure 21: Verification of Genomic Integration of AD Genes for AM2 and AD2 strains**

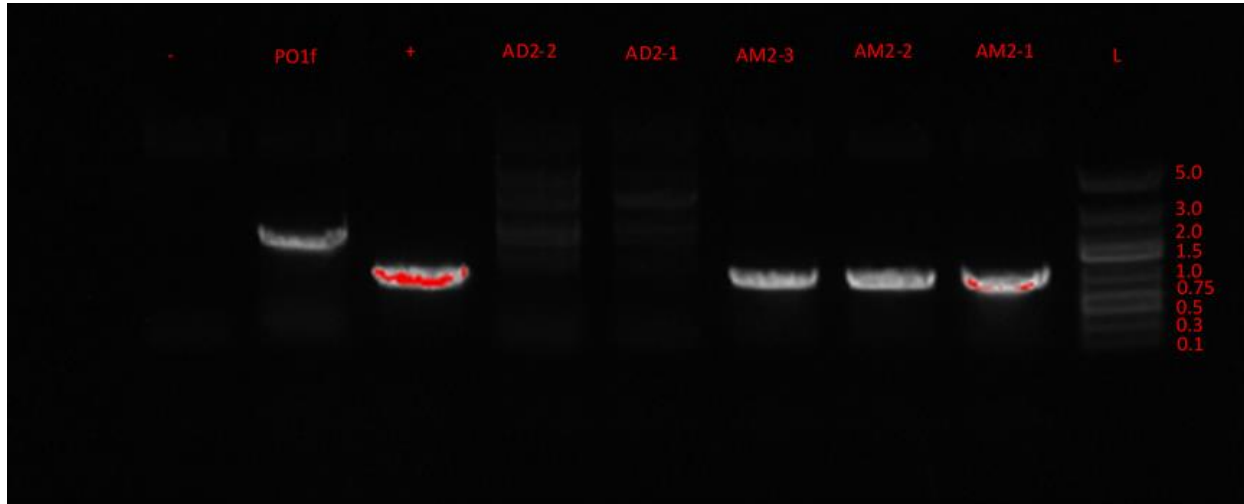


Figure 21: The three selected colonies of the new AM2 plate produced 1 kb bands, indicating genomic integration of the AD genes at the MFE1 locus. Unexpectedly, instead of producing a 2 kb band to indicate no genomic integration (like the PO1f band) or a 1 kb band to indicate genomic integration, the two selected colonies of the new AD2 plate produced no bands at all.

**Figure 22: Genomic Integration of Engineered Strains**

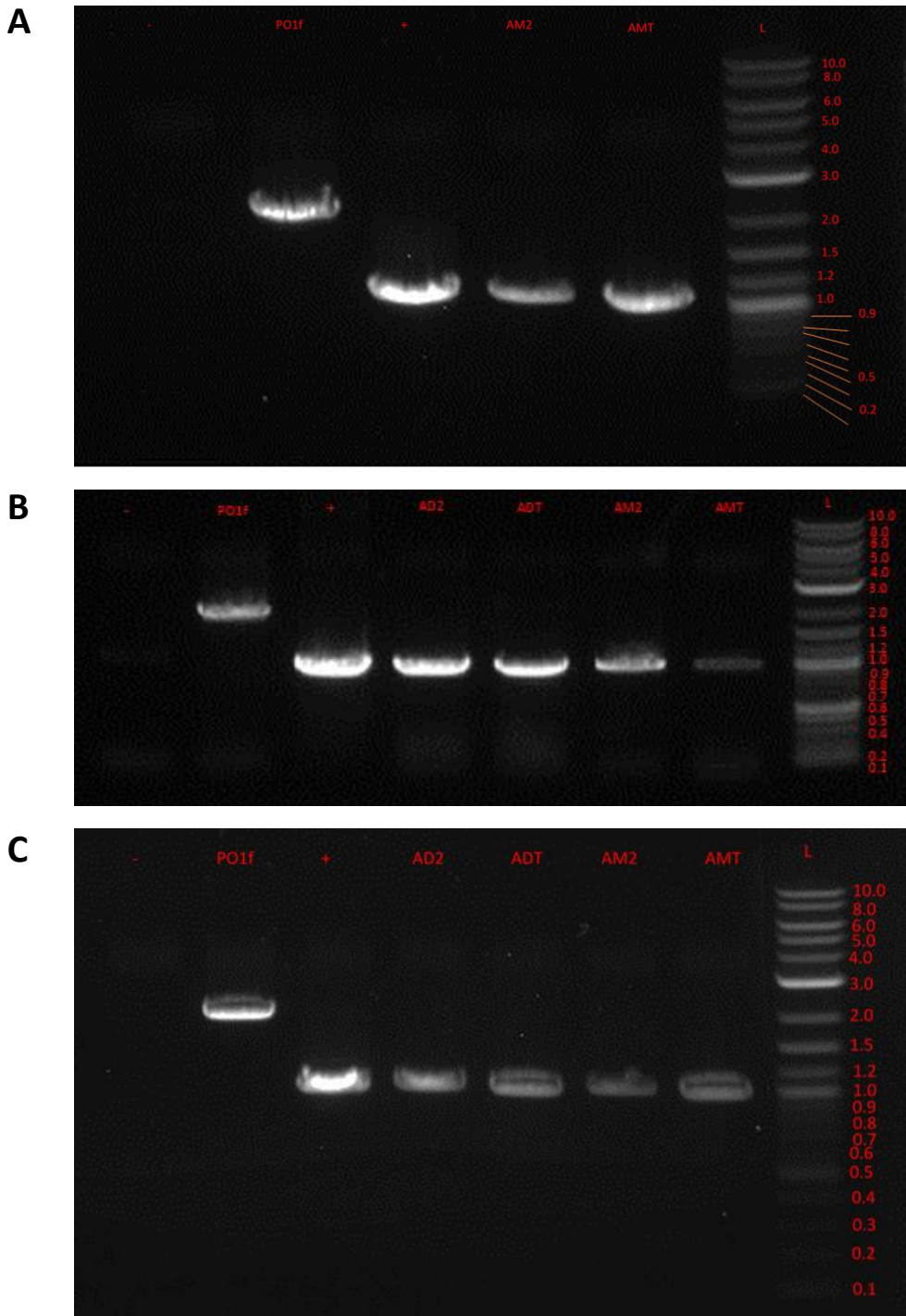


Figure 22: (A) Three-primer PCR verifies integration of FPPS and HMGCR genes into the AXP locus for the AMT and AM2 strains (producing a 1 kb PCR product), while the PO1f strain produces a 2 kb product displaying no integration. A positive control is made by employing PCR on the recombinant plasmid directly. (B) Three-primer PCR verifies integration of *astA* and *astD* genes into the D17 locus for the AMT, AM2, ADT, and AD2 strains. (C) Three-primer PCR verifies integration of *astB* and *astC* genes into the MFE1 locus for the AMT, AM2, ADT, and AD2 strains.

**Figure 23: cDNA Synthesis Mix Contamination**

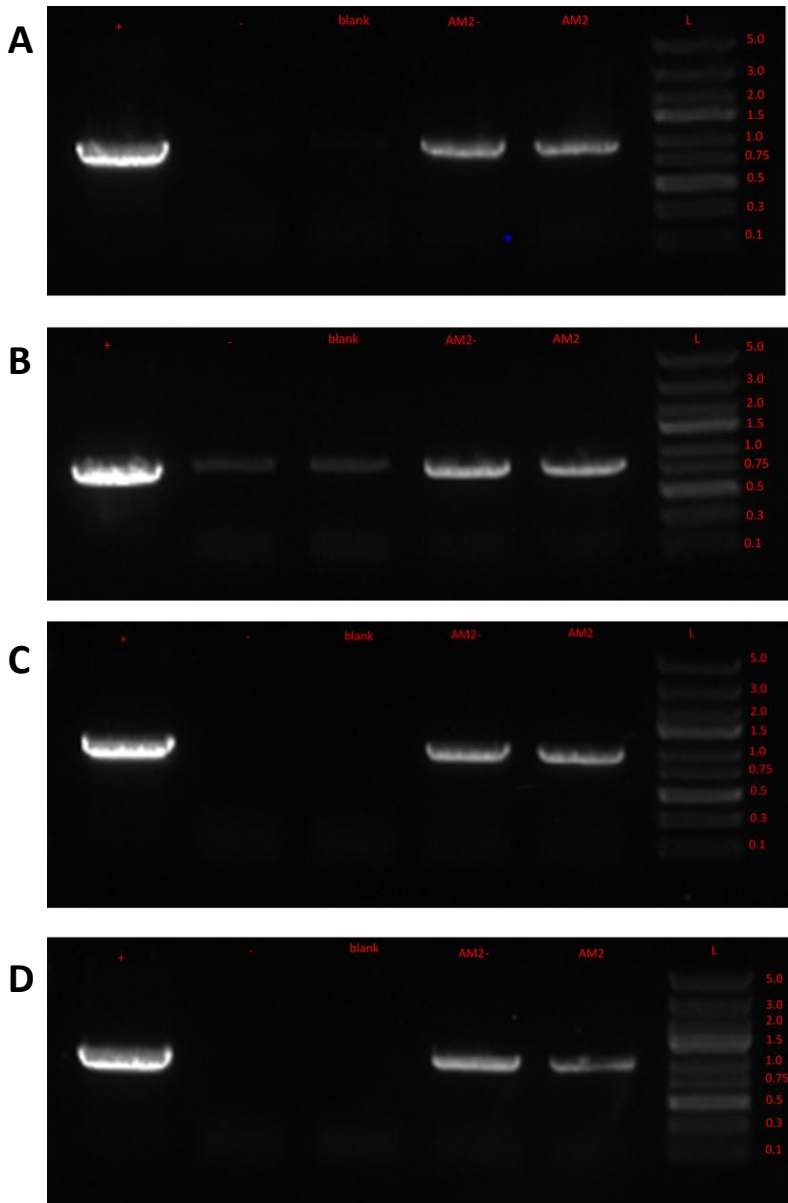


Figure 23: (A) PCR of *astA* from AM2 cDNA shows a band, the negative control AM2- which represents cDNA synthesis with no RNA extract added also shows a band indicating that one of the components for cDNA is contaminated with DNA template. The blank sample is a small amount of RNA extract to check for genomic DNA contamination, the negative control checks for PCR mix contamination as it has no cDNA added, and the positive control uses a plasmid template for amplification of *ast* genes. (B) PCR of *astB* shows the same result of bands in both the AM2 and AM2- samples. (C) PCR of *astC* shows bands for both AM2 and AM2- samples. (D) PCR of *astD* shows bands for both AM2 and AM2- samples.

**Figure 24: Fluorescence Image of Cells in YPD**

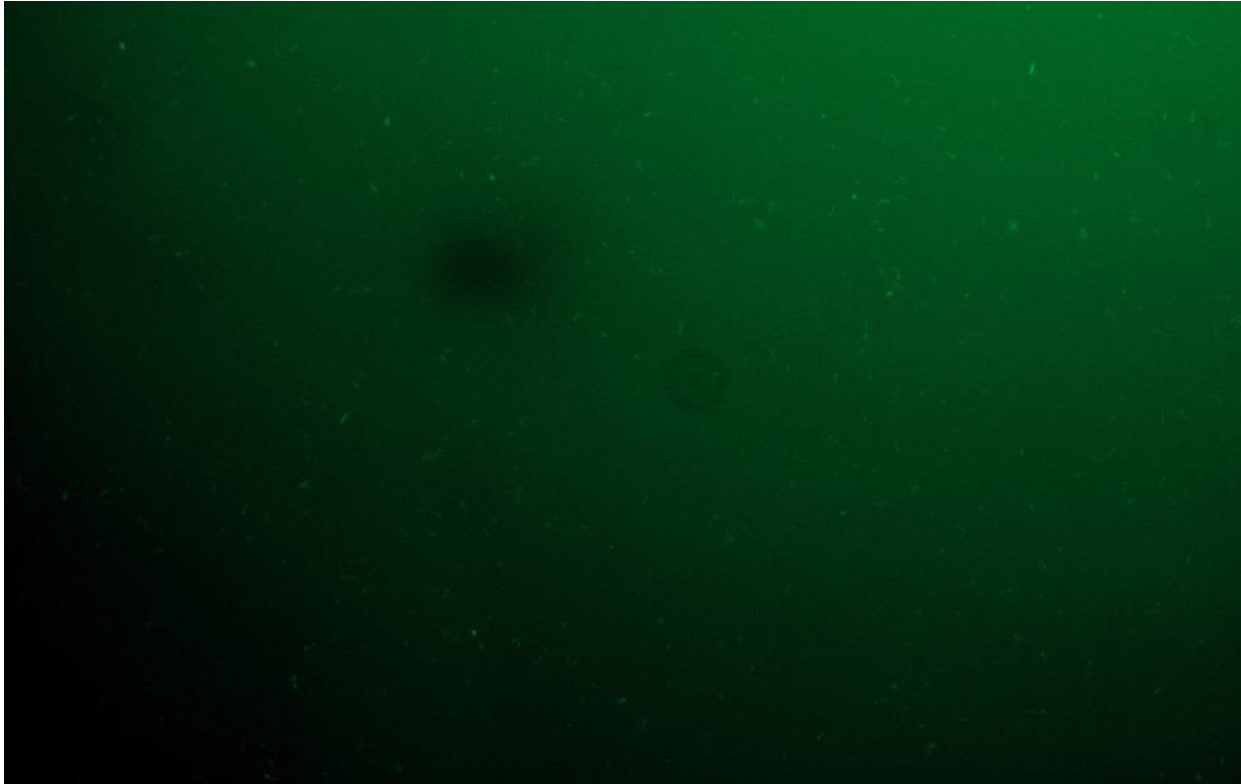


Figure 24: There is a great deal of background fluorescence within the image depicting PO1f cells. A number of components within YPD are capable of autofluorescence and will make fluorescent signals from cells difficult to detect or interpret. A high cell density, as is shown in the image, also makes fluorescence imaging more difficult.

**Figure 25: Fluorescence Imaging of Cells in Minimal Media**

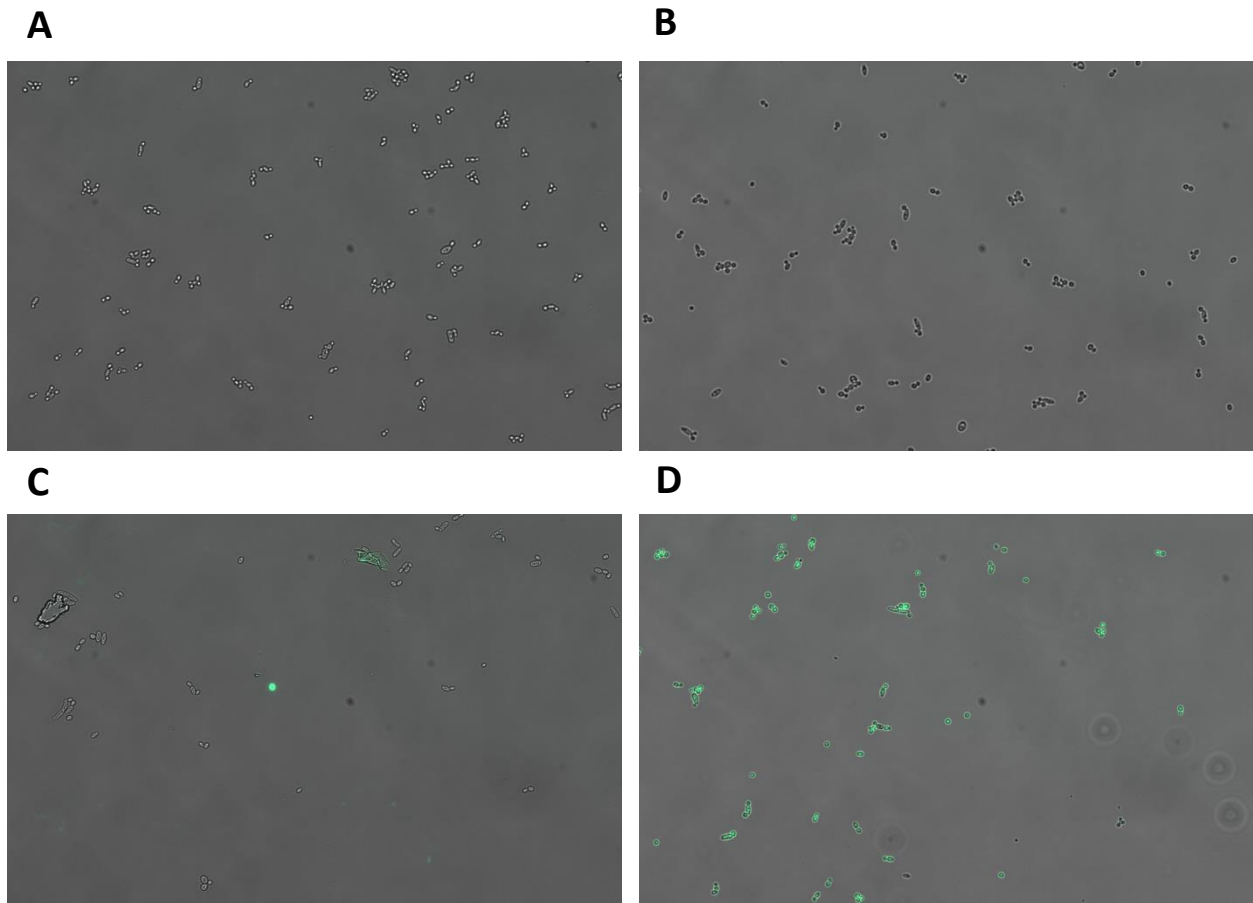


Figure 25: (A) Overlaid fluorescence and microscope image of PO1f strain, displaying no fluorescence as expected. (B) LBG strain overlaid image shows no fluorescence, indicating that the fused BGL1-hrGFP protein is inactive. (C) TBG strain overlaid image shows some degree of fluorescence within the cells, signifying the production of functional hrGFP (D) 2BG strain overlaid image shows a high degree of fluorescence within cells, suggesting the cleavage of the transcript into functional hrGFP and BGL1 proteins.



**Table 1: Engineered Strains' Genotype**

Genomic Loci and Genotype									
	AXP Locus		D17 Locus		MFE1 Locus		A08 Locus		
Strains	TPS HMGCR + FPPS	2A HMGCR + FPPS	TPS astA + astD	2A astA + astD	TPS astC + astB	2A astC + astB	2A BGL1 + hrGFP	TPS BGL1 + hrGFP	Link BGL1 + hrGFP
AMT	+		+		+				
AM2		+		+		+			
ADT			+		+				
AD2				+		+			
2AD				+					
TAD			+						
2MVL		+							
TMVL	+								
2AM		+		+					
TAM	+		+						
LBG									+
TBG								+	
2BG							+		

Table 1: Each engineered strain has two genes inserted per genomic integration. HMGCR and FPPS genes are inserted into the AXP locus, the astA and astD genes are inserted into the D17 locus, the astC and astB genes are inserted into the MFE1 locus, and the BGL1 and hrGFP genes are inserted into the A08 locus. Each strain uses only either 2A gene inserts (each gene is separated by the ERBV-1 2A sequencer) or two-promoter gene inserts (each gene is separated by lip2t-pFB<sub>Ain</sub> terminator-promoter fragment). The intermediary strains (2AD, TAD, 2MVL, TMVL, 2AM, and TAM) were utilized to make the strains used for metabolic analysis. These four main strains (highlighted in grey) contain all four ast genes needed for aspartic acid synthesis and resistance. The LBG strain separates the BGL1 and hrGFP genes with a short linker sequence, resulting in fused protein.

## Primer and Synthesized DNA List (5' to 3')

ERBV-1 sequence:

GGTTCTGGTGGTGCTACTAATTTTTCTTTGTTGAAATTGGCTGGTGATGTTGA  
ATTGAATCCAGGTCCA

GA 1: CCTGAACGTCTACCTGTTTGGAGC

GA 2: GCTGGGCTTCTCAATGACGATAGGAAC

GA 3:

GCGTGACATAACTAATTACATGAGGCTACTTCTGTTCGCTTGTAATCTTGG

GA 4:

TTCCTTCTGAGTATAAGAATCATTCAAAGGCGCGCATGCTACAAGCAGCTATT  
GGAAAGATTGTG

GA 6:

TGGTGATGTTGAATTGAATCCAGGTCCAATGTCCAAGGCGAAATTCGAAAGC

GA 34:

TTCCTTCTGAGTATAAGAATCATTCAAAGGATGGGAGCTTCTACTTTCTC

GA 35:

TGTAAGCGTGACATAACTAATTACATGAGGCTAGATCGGTCCGTCCGTG

GA 36: GTGATGTTGAATTGAATCCAGGTCCAATGTTTCGCGTCGAGGATCC

GA 37:

TTCCTTCTGAGTATAAGAATCATTCAAAGGATGGACATGAATACCTTCCCC

GA 39: TGTTGAATTGAATCCAGGTCCAATGCTATTCCAAGACCTGTCTTTTC

GA 40:

TGTAAGCGTGACATAACTAATTACATGAGGCTACTTGCAGAGACCCATAACT  
C

GA 49:

CAAAGAAAAATTAGTAGCACCACCAGAACCTGCGTTGCCTAGCGGGAG

GA 50:

CAAAGAAAAATTAGTAGCACCACCAGAACCGCTTCGTCTCTTTATCCTGAT

GA 51:

CAAAGAAAAATTAGTAGCACCACCAGAACCTGACCGTATGCAAATATTCGAA  
CC

GA 54:

GGCAGCTTGAGAAGATTGACACAGGTGGAGAACAGTGTACGCAGTACTATAG  
AG

GA 56:  
AGGTAGAAGTTGTAAAGAGTGATAAATAGCCTATGACCGTATGCAAATATTC  
GAAC

GA 57:  
CACACTCTCTACACAACTAACCCAGCTCTATGTCCAAGGCGAAATTCGAAA  
G

GA 58:  
AGGTAGAAGTTGTAAAGAGTGATAAATAGCTTATGCGTTGCCTAGCGGGAG

GA 59:  
CACACTCTCTACACAACTAACCCAGCTCTATGTTTCGCGTCGAGGATCC

GA 60:  
AGGTAGAAGTTGTAAAGAGTGATAAATAGCCTAGCTTCGTCTCTTTATCCTG

GA 61:  
CACACTCTCTACACAACTAACCCAGCTCTATGCTATTCCAAGACCTGTCTTT  
TC

P 28: AGAGCTGGGTTAGTTTGTGTAGAG

P 29: GCTATTTATCACTCTTTACAAC

P 30: CTCCACCTGTGTCAATCTTC

P 39: AACCCGCATACACGTGCATAATACTTGCATAACGATAGC

P 40: GGAACGCACTTCCGTA CTAGTATCTTCTC

P 41: CACCAGCTGGTTAATGTGGGGTTGAGAGTGC

P 42: GCATCCAACAATGTTGAACCTGCAATTGAATCACTGC

P 43: GTCTCTTCAACTCGCTCAAAAAACAACCGATTGCTC

P 44: TGGCAGATGGCCTTGGACGCTGGAC

P 45:  
CTTTGAAATGATGACATTTTCGAGAAAATGAGAACAACCGAATATTTGTAC

P 46: GAGCTCGAGCATCTCCTCAGTGAAGACAG

P 47: GAGCCTCGACCTGGGCAGGGGTG

P 48: CAAGCTTGATTGGAAATTAACATATGAGCTGCGTGCTTTTTGC

P 53: GAGAAGGTTTTGGGACGCTCGAAGG

P 54: ACGCCCAGTGGTATGATGCG

P 55: CCCTTGGGGCCTTCATACCG

P 56: AGCGAGCCGATTTACCGTCG  
P57: TCAGCCATCCACCGATTTCGC  
P 58: ACACTGCCGATGTCGTCACC  
P 59: ACATCCTCTGCTGCCCTTGC  
P 60: ATCCATCGTGCGCTTCCTGG  
P 61: AAGTACTCGCGGAACCTGGC  
S 10: TTTGTGGTTGGGACTTTAGCCAAG  
S 15: GGTTCCTGGTGGTGCTACTA  
S 16: TGGACCTGGATTCAATTCAA  
S 17: CCTTCCTTTTCGGTTAGAGC  
S 26: TGTAATGACACAATCCGAAAGTC

## Laragen Sequence Reads

### 2A AXP MVL Sequence Reads

S10:

TACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTTAAGCAT  
TTCCTTCTGAGTATAAGAATCATTCAAAGGCGCGCATGCTACAAGCAGCTATTGGAA  
AGATTGTGGGATTTGCGGTCAACCGACCCATCCACACAGTTGTCCTGACGTCCATCG  
TGGCGTCAACCGCATAACCTCGCCATCCTCGACATTGCCATCCCGGGTTTCGAGGGCA  
CACAACCCATCTCATACTACCACCCTGCAGCAAAATCTTACGACAACCCTGCTGATT  
GGACCCACATTGCAGAGGCCGACATCCCTTCAGACGCCTACCGACTTGCATTTGCC  
AGATCCGTGTCAGTGATGTTTCAGGGCGGAGAGGCCCCACCATCCCTGGCGCCGTG  
GCCGTGTCTGATCTCGACCACAGAATCGTCATGGACTACAAACAGTGGGCCCCCTGG  
ACCGCCAGCAACGAGCAGATCGCCTCGGAGAACCACATCTGGAAGCACTCCTTCAA  
GGACCACGTGGCCTTCAGCTGGATCAAGTGGTTCCGATGGGCCTACCTGCGTTTGTC  
CACTCTCATCCAGGGGGCAGACAACCTTCGACATTGCCGTGGTCGCCCTTGGCTATCT  
TGCCATGCACTACACCTTCTTCAGTCTCTTCCGATCCATGCGAAAGGTTGGCTCGCA  
CTTTTGGCTTGCCTCCATGGCTCTGGTCTCTTCCACCTTCGCTTTCCTGCTTGCGGTGG  
TGGCTTCCTCTAGCCTGGGTACCGACCTAGCATGATCACCATGTCCGAGGGCCTGC  
CCTTCCTCGTGGTCGCCATTGGCTTTGACCGAAAGGTCAACCTGGCTAGCGAGGTGC  
TCACATCCAAGAGCAGCCAGCTCGCT

S16:

AAAGGCCTTCAACTCCACCTCTCGATTTGCTCGTCTCCAGTCTCTTCACTCTACCCTT  
GCTGGTAACGTGCTGTTTATTCGATTCCGAACCACCACTGGTGATGCCATGGGCATG  
AACATGATCTCCAAGGGCGTCGAACACTCTCTGGCCGTCATGGTCAAGGAGTACGG

CTTCCCTGATATGGACATTGTGTCTGTCTCGGGTAACTACTGCACTGACAAGAAGCC  
CGCAGCGATCAACTGGATCGAAGGCCGAGGCAAGAGTGTTGTTGCCGAAGCCACCA  
TCCCTGCTCACATTGTCAAGTCTGTTCTCAAAAGTGAGGTTGACGCTCTTGTTGAGCT  
CAACATCAGCAAGAATCTGATCGGTAGTGCCATGGCTGGCTCTGTGGGAGGTTTCAA  
TGCACACGCCGCAAACCTGGTGACCGCCATCTACCTTGCCACTGGCCAGGATCCTGC  
TCAGAATGTGAGTCTTCCAAGTGCATCACGCTGATGAGCAACGTCGACGGTAACCT  
GCTCATCTCCGTTTCCATGCCTTCTATCGAGGTCGGTACCATTGGTGGAGGTAAT  
TTGGAGCCCCAGGGGGCTATGCTGGAGATGCTTGGCGTGCGAGGTCCTCACATCGA  
GACCCCGGTGCCAACGCCAACAGCTTGCTCGCATCATTGCTTCTGGAGTTCTTGC  
AGCGGAGCTTTCGCTGTGTTCTGCTCTTGCTGCCGGCCATCTTGTGCAAAGTCATATG  
ACCCACAACCGGTCCCAGGCTCCTACTCCGGCCAAGCAGTCTCAGGCCGATCTGCAG  
CGTCTACAAAACGGTTCGAATATTTGCATACGGTCAGGTTCTGGTGGTGCTACTAA

S15:

TTGNNCCAGGTCCAATGTCCAAGGCGAAATTCGAAAGCGTGTTCCCCGAATCTCCG  
AGGAGCTGGTGCAGCTGCTGCGAGACGAGGGTCTGCCCCAGGATGCCGTGCAGTGG  
TTTTCCGACTCACTTCAGTACAAGTGTGTGGGTGGAAAGCTCAACCGAGGCCTGTCT  
GTGGTGCACACCTACCAGCTACTGACCGGCAAGAAGGAGCTCGATGACGAGGAGTA  
CTACCGACTCGCGCTGCTCGGCTGGCTGATTGAGCTGCTGCAGGCGTTTTTCTCGTG  
TCGGACGACATTATGGATGAGTCCAAGACCCGACGAGGCCAGCCCTGCTGGTACCT  
CAAGCCCAAGGTCCGCATGATTGCCATCAACGATGCTTTCATGCTAGAGAGTGGCAT  
CTACATTCTGCTTAAGAAGCATTTCGACAGGAGAAGTACTACATTGACCTTGTCGA  
GCTGTTCCACGACATTTTCGTTCAAGACCGAGCTGGGCCAGCTGGTGGATCTTCTGAC  
TGCCCCGAGGATGAGGTTGATCTCAACCGGTTCTCTCTGGACAAGCACTCCTTTAT

TGTGCGATAACAAGACTGCTTACTACTCCTTCTACCTGCCCGTTGTTCTAGCCATGTAC  
GTGGCCGGCATTACCAACCCCAAGGACCTGCAGCAGGCCATGGATGTGCTGATCCC  
TCTCGGAGAGTACTTCCAGGTCCAGGACGACTACCTTGACAACCTTTGGAGACCCCGA  
GTTTATTGGTAAGATCGGCACCGACATCCAGGACAACAAGTGCTCCTGGCTCGTTAA  
CAAAGCCCTTCAGAANGCCACCCCGAGCAGCGACAGATCCTCGAGGACAACACTACG  
GCGTCAAGGACAAGTCCAAGGAGCTCGTCATCAAGAACTGTATGATGACATGNAG  
ATTGAGC

S17:

ACGACATTATGGATGAGTCCAAGACCCNGACGAGGCCAGCCCTGCTGGTACCTCAA  
GCCCAAGGTCGGCATGATTGCCATCAACGATGCTTTCATGCTAGAGAGTGGCATCTA  
CATTCTGCTTAAGAAGCATTTCGACAGGAGAAGTACTACATTGACCTTGTCGAGCT  
GTTCCACGACATTTTCGTTCAAGACCGAGCTGGGCCAGCTGGTGGATCTTCTGACTGC  
CCCCGAGGATGAGGTTGATCTCAACCGGTTCTCTCTGGACAAGCACTCCTTTATTGT  
GCGATAACAAGACTGCTTACTACTCCTTCTACCTGCCCGTTGTTCTAGCCATGTACGTG  
GCCGGCATTACCAACCCCAAGGACCTGCAGCAGGCCATGGATGTGCTGATCCCTCTC  
GGAGAGTACTTCCAGGTCCAGGACGACTACCTTGACAACCTTTGGAGACCCCGAGTTC  
ATTGGTAAGATCGGCACCGACATCCAGGACAACAAGTGCTCCTGGCTCGTTAACA  
AGCCCTTCAGAAGGCCACCCCGAGCAGCGACAGATCCTCGAGGACAACACTACGGCG  
TCAAGGACAAGTCCAAGGAGCTCGTCATCAAGAACTGTATGATGACATGAAGATT  
GAGCAGGACTACCTTGACTACGAGGAGGAGGTTGTTGGCGACATCAAGAAGAAGAT  
CGAGCAGGTTGACGAGAGCCGAGGCTTCAAGAAGGAGGTGCTCAACGCTTTCCTCG  
CCAAGATTTACAAGCGACAGAAGTAGCCTCATGTAATTAGTTANGTCACGCNTACA

TPS AXP MVL Reads

S10:

AATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTTAAG  
CATTTCTTCTGAGTATAAGAATCATTCAAAGGCGCGCATGCTACAAGCAGCTATTG  
GAAAGATTGTGGGATTTGCGGTCAACCGACCCATCCACACAGTTGTCCTGACGTCCA  
TCGTGGCGTCAACCGCATACTCGCCATCCTCGACATTGCCATCCCGGGTTTCGAGG  
GCACACAACCCATCTCATACTACCACCCTGCAGCAAAATCTTACGACAACCCTGCTG  
ATTGGACCCACATTGCAGAGGCCGACATCCCTTCAGACGCCTACCGACTTGCATTTG  
CCCAGATCCGTGTCAGTGATGTTTCAGGGCGGAGAGGCCCCACCATCCCTGGCGCC  
GTGGCCGTGTCTGATCTCGACCACAGAATCGTCATGGACTACAAACAGTGGGCCCCC  
TGGACCGCCAGCAACGAGCAGATCGCCTCGGAGAACCACATCTGGAAGCACTCCTT  
CAAGGACCACGTGGCCTTCAGCTGGATCAAGTGGTTCCGATGGGCCTACCTGCGTTT  
GTCCACTCTCATCCAGGGGGCAGACAACCTTCGACATTGCCGTGGTCGCCCTTGGCTA  
TCTTGCCATGCACTACACCTTCTTCAGTCTCTTCCGATCCATGCGAAAGGTTGGCTCG  
CACTTTTGGCTTGCCTCCATGGCTCTGGTCTCTTCCACCTTCGCTTTCCTGCTTGCGGT  
GGTGGCTTCTCTAGCCTGGGTTACCGACCTAGCATG

P30:

TGTGTCTGTCTCGGGTAACTACTGCACTGACAAGAAGCCCGCAGCGATCAACTGGAT  
CGAAGGCCGAGGCAAGAGTGTTGTTGCCGAAGCCACCATCCCTGCTCACATTGTCA  
AGTCTGTTCTCAAAGTGAGGTTGACGCTCTTGTGAGCTCAACATCAGCAAGAATC  
TGATCGGTAGTGCCATGGCTGGCTCTGTGGGAGGTTTCAATGCACACGCCGCAAACC  
TGGTGACCGCCATCTACCTTGCCACTGGCCAGGATCCTGCTCAGAATGTCGAGTCTT  
CCAAGTGCATCACGCTGATGAGCAACGTCGACGGTAACTGCTCATCTCCGTTTCCA



TGCCTTCTATCGAGGTCGGTACCATTGGTGGAGGTA CTATTTTGGAGCCCCAGGGGG  
CTATGCTGGAGATGCTTGGCGTGCGAGGTCCTCACATCGAGACCCCCGGTGCCAACG  
CCCAACAGCTTGCTCGCATCATTGCTTCTGGAGTTCTTGCAGCGGAGCTTTCGCTGTG  
TTCTGCTCTTGCTGCCGGCCATCTTGTGCAAAGTCATATGACCCACAACCGGTCCCA  
GGCTCCTACTCCGGCCAAGCAGTCTCAGGCCGATCTGCAGCGTCTACAAAACGGTTC  
GAATATTTGCATACGGTCATAGGCTATTTATCACTCTTTACA ACTTCTACCTCAACTA  
TCTACTTTAATAAATGAATATCGTTTATTCTCTATGATTACTGTATATGCGTTCCTCT  
AAGACAAATCGAAACCAGCATGCGATCGAATGGCATA CAAAAGTTCTTCCGAAG

S26:

CTCTATGTCCNAGGCGAAATTCGAAAGCGTGTTCCCCCGAATCTCCGAGGAGCTGGT  
GCAGCTGCTGCGAGACGAGGGTCTGCCCCAGGATGCCGTGCAGTGGTTTTCCGACTC  
ACTTCAGTACA ACTGTGTGGGTGGAAAGCTCAACCGAGGCCTGTCTGTGGTCGACAC  
CTACCAGCTACTGACCGGCAAGAAGGAGCTCGATGACGAGGAGTACTACCGACTCG  
CGCTGCTCGGCTGGCTGATTGAGCTGCTGCAGGCGTTTTTCCTCGTGTCGGACGACA  
TTATGGATGAGTCCAAGACCCGACGAGGCCAGCCCTGCTGGTACCTCAAGCCCAAG  
GTCGGCATGATTGCCATCAACGATGCTTTCATGCTAGAGAGTGGCATCTACATTCTG  
CTTAAGAAGCATTTCGACAGGAGAAGTACTACATTGACCTTGTCGAGCTGTTCCAC  
GACATTTTCGTTCAAGACCGAGCTGGGCCAGCTGGTGGATCTTCTGACTGCCCCCGAG  
GATGAGGTTGATCTCAACCGGTTCTCTCTGGACAAGCACTCCTTTATTGTGCGATAC  
AAGACTGCTTACTACTCCTTCTACCTGCCCGTTGTTCTAGCCATGTACGTGGCCGGCA  
TTACCAACCCCAAGGACCTGCAGCAGGCCATGGATGTGCTGATCCCTCTCGGAGAGT  
ACTTCCAGGTCCAGGACGACTACCTTGACA ACTTTGGAGACCCCCGAGTTCATTGGTA

AGATCGGCACCGACATCCAGGACAACAAGTGCTCCTGGCTCGTTAACAAGCCCTT  
CAGAA-GNCACCCCCGAGCAGCGACAGATCCTCGAGGACAACACTACNGCGTCAAGGA

S17:

TACNTCAAGCCCAAGGTCGGCATGATNGCCATCAACGATGCTTTCATGCTAGAGAGT  
GGCATCTACATTCTGCTTAAGAAGCATTNCGGACAGGAGAAGTACTACATTGACCT  
TGTCGAGCTGTTCCACGACATTTTCGTTCAAGACCGAGCTGGGCCAGCTGGTGGATCT  
TCTGACTGCCCCGAGGATGAGGTTGATCTCAACCGGTTCTCTCTGGACAAGCACTC  
CTTTATTGTGCGATAACAAGACTGCTTACTACTCCTTCTACCTGCCCCGTTGTTCTAGCC  
ATGTACGTGGCCGGCATTACCAACCCCAAGGACCTGCAGCAGGCCATGGATGTGCT  
GATCCCTCTCGGAGAGTACTTCCAGGTCCAGGACGACTACCTTGACAACCTTGGAGA  
CCCCGAGTTCATTGGTAAGATCGGCACCGACATCCAGGACAACAAGTGCTCCTGGCT  
CGTTAACAAGCCCTTCAGAAGGCCACCCCCGAGCAGCGACAGATCCTCGAGGACA  
ACTACGGCGTCAAGGACAAGTCCAAGGAGCTCGTCATCAAGAACTGTATGATGAC  
ATGAAGATTGAGCAGGACTACCTTGACTACGAGGAGGAGGTTGTTGGCGACATCAA  
GAAGAAGATCGAGCAGGTTGACGAGAGCCGAGGCTTCAAGAAGGAGGTGCTCAAC  
GCTTTCCTCGCCAAGATTTACAAGCGACAGAAGTAGCCTCATGTAATTAGTTANGTC  
ACGCTAC

2A D17 AD Reads

S10:

ATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTTAAGC  
ATTTCTTCTGAGTATAAGAATCATTCAAAGGATGGACATGAATACCTTCCCCGCCA  
GCACATACTGCGAATCCATCGTGCGCTTCTGGATGCCATTGAGTACCATGATGACA  
ACCTGACACACGAGGAGCGGGTTCGAAGGCCTCCGTCATGTTTCATTCCAAGACTGCA

CAATACTTTACCGAGCCTCTTCCGAGAAGTATCCTCAAAGGGGTGGCTCCCCGCCGG  
ATTGCTGCTGTCACTCGGACCATTTCCCATTTTCATCGTCTATTGTTGGAGCAAGCTAC  
CCCGGGAAGCTCAAGTAGACGTCTCCATTTACCTGTCCATCATCAACGTGCTGGACG  
ACGAGATCAGCAGCGAACCAGCACCCAGATGACCAGCTTCTGGTCAGACCTGATC  
CAGGGCAAGCAGCCGAAGCACCCCTTCTGGGTGCTGTTCAATTCACATCTGCCCCGT  
CTCCTCCGACACTACGGCAGCTTCTGCGCCTTCAACATCATGCGCTGCACATTCGAT  
TACTTCGAAGGCTGTTGGATCGAGCAGCACAATTTCCAGGGCTACCCTGGCGCCGAC  
TGCTACCCGTCGTTCCCTCCGACGACTGAACTGCCTCGGCGGTGCTGTTCGCAGGGACC  
ATCTTCCCCGCCCAAGTTCGATGAGCAGAAGCTCTTTGCGCAGATGTCTTGCGTG  
ATGGCGCAAATCGACGGACCCGTGCCCCTGATGAATGATCTGTTCTCTTTCTACAAA  
GAATATGATCAGGACGAGGCCAATCTGGTGAGCAACTGGTGCACTGTGGANGGAAT  
CACGATGGACCAGGCCCTCACACGTCTCACCGACGATAACAATCCATGCTTGCGTA

S16:

AACCCAGCACCCAGATGACCAGNTTCTGNTCAGACNTGATCCAGGGCAAGCAGCCG  
NAAGCACCCNTTCTGGGTGCTGTTCAANTTCACATCTGCCCCGTCTCCTCCGACACTA  
CGGCAGCTTCTGCGCCTTCAACATCATGCGCTGCACATTCGATTACTTCGAAGGCTG  
TTGGATCGAGCAGCACAATTTCCAGGGCTACCCTGGCGCCGACTGCTACCCGTCGTT  
CNTCCGACGACTGAACTGCCTCGGCGGTGCTGTTCGCAGGGACCATCTTCCCCGCCGC  
CAAGTTCGATGAGCAGAAGCTCTTTGCGCAGATGTCTTGCGTGATGGCGCAAATCGA  
CGGACCCGTCGCCCTGATGAATGATCTGTTCTCTTTCTACAAAGAATATGATCAGGA  
CGAGGCCAATCTGGTGAGCAACTGGTGCACTGTGGACGGAATCACGATGGACCAGG  
CCCTCACACGTCTCACCGACGATAACAATCCATGCTTGCGTACGGATTCTGGATATCT  
TGAAGGACAAGGACCCGGACATGCTGGCCACCATCCGCGGGTTTATCCACGGGTAT

GCGACATGGCACATCTGCGACTTCAGGTACAGACTGCGGGAGATCTATGACCGCGA  
AGATCTGCAAGAATCGGGCGCCAGGTTCCGCGAGTACTTTGACAAGGCTATTGATGT  
GGGATGGGTGGATGTTGAAGAGTGGACTTGCCAAGTCCAGGGCTTTGAAGTCGAGG  
GGCCGGCCCCCTTCGGGATCTGAAGTACAGGCATATCAGGCCAATGCATTTGGGTTCT  
CGGTCGATACACAGCGGCACCACAACATAGATTACGTGGGGAGTTCAATACTTGGT  
CTGTTTGAATGGGCGACGCGTTATCTGAGGGGGAAACTCCCGCTAGGCAACGCAGG  
TTCTGGTGGTGCTACTAATTTC

S15:

GCTGGTGATGTTGATTGAATCCAGGTCCAATGTTGCGGTCGAGGATCCGATCGAGAG  
CTCTGGGGTTGCACCCACGAGCCCGGTTTGAGAATACTCGTCTTCCGGCTTCCACTA  
CCGGGCGCCGTTACAAGTCCGACGAGACCCTCAATCGGGTCTCCTCGAAAATCACA  
CAACCCAAATCTCAGGGTGCCTCTCAAGCAATGCTCTACGCCACCGGCCTCACAGAG  
GAAGACATGTCCAAGCCGCAGGTCGGTATCTCCTCGGTGTGGTTCGAGGGTAACCC  
TGTAACATGCACTTGCACGACCTCTCCGCCATCGTGCGCGATTCCGTGCACCGGGCG  
GGTCTGGTCCCAATGCGATTCAATTCAGTTGGAGTGTGCGGATGGAATCAGTATGGGC  
ACAAAGGGAATGAGATACAGCTTGCAGAGTCGGGAGCTGATCGCCGATGGCATTGA  
GACGGTGATGAACGCCCAGTGGTATGATGCGAATGTGTCGCTCCCGGGTTGCGATA  
AGAACATGCCGGGTGTGTTGATGGCGATGGGACGCACGAATCGACCTAGTATCATG  
GTCTATGGCGGCAGTATCAAGCCCGGATGCAGTGCAAAGGGCCAAAAGCTGGACCT  
GGTTAGCGCGTTCCAGTCGTATGGACAGTTCATCACCGGCCAGATCGACGAGAAGG  
AGCGGTTTCGATATTATTCGCAATGCATGCCCCGGCAGANGTGCCTGCGGTGGCATGT  
ACACGGCCAATACCCTGGCCACGGCCATTGAGACTATGGGTATGACCGTTCCCGGTA  
GCAGCAGTTGCCCGGCAGACGATCCCAAGAAGCTGGTTCGAGTGCGAAAACATCGGC

GAGGTGGTTAAGACTATGCTCNGGGAANATATCAAGCCCAGGGATGTCTTGACGCG  
TCAAGCTTTCNAGATGCATGATTGTGGTGATATCCTGGNTGGCAGCACCACGCCNNC  
TGCNTCTGATTGCT

S17:

AGCCATNTGGAAAATNCNTTATGANCGACCTATAACAACATCGGNGGCACACCAGCC  
CTCNTCAAATATCTTNTGAAGGAGGGCNTGATCGACGGNTCAGGAATTACTGTCACT  
GGCAAACAATGAAGGAGAACGTGGCCTCATGGCCCGATTTCCCTGCCGACCAGGA  
CATTATCCGCCCCCTCAGCAACCCTATCAAACCATCTGGCCATCTCCAGATTCTTCGC  
GGGTCGCTGGCACCGGGCGGTTCCGTGGGTAAGATTACTGGCAAGGAGGGTCTGCG  
GTTTCGAGGGTACGGCCAAGTGCTACGACTACGAAGATGCATTTATTGAGTCCCTCGA  
GCGGGGCGAAATCAAGAAGGGCGAGAAGACGGTCGTGATTATCCGGTATGAAGGCC  
CCAAGGGTGGCCAGGAATGCCTGAGATGCTCAAGCCCAGCGCGGCCATTATGGGT  
GCCGGTCTGGGCCAGGACGTTGCGCTTCTCACGGACGGAAGATTCTCGGGTGGCAGT  
CACGGATTCTTGATCGGACATATCGTGCCGGAAGCCATGGAGGGCGGCCCGATCGC  
CTTGGCCCGGGACGGTGACCGGATCGTGATTGATGCTGAGGAAAGAGTGGTTGATC  
TGGATATCCCGACTGAAGAGCTCGAGAAGAGGAGGAAGGAGTGGAAAGCACCCCC  
GCTCCGATAACCAGAAGGGAACCTTTGAAGAAGTACTGCACGCTGGTCAGCGATGCCA  
GCCATGGATGTGTCACGGACGGACCGATCTAGCCTCATGTAATTAGTTATGTCACGC  
TTAC

TPS D17 AD Reads

S10:

TACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTTAAGCAT  
TTCCTTCTGAGTATAAGAATCATTCAAAGGATGGACATGAATACCTTCCCCGCCAGC

ACATACTGCGAATCCATCGTGCGCTTCCTGGATGCCATTGAGTACCATGATGACAAC  
CTGACACACGAGGAGCGGGTCGAAGGCCTCCGTCATGTTTCATTCCAAGACTGCACA  
ATACTTTACCGAGCCTCTTCCGAGAAGTATCCTCAAAGGGGTGGCTCCCCGCCGGAT  
TGCTGCTGTCACTCGGACCATTTCCCATTTTCATCGTCTATTGTTGGAGCAAGCTACCC  
CGGGAAGCTCAAGTAGACGTCTCCATTTACCTGTCCATCATCAACGTGCTGGACGAC  
GAGATCAGCAGCGAACCCAGCACCCAGATGACCAGCTTCTGGTCAGACCTGATCCA  
GGGCAAGCAGCCGAAGCACCCCTTCTGGGTGCTGTTCAATTCACATCTGCCCCGTCT  
CCTCCGACACTACGGCAGCTTCTGCGCCTTCAACATCATGCGCTGCACATTCGATTA  
CTTCGAAGGCTGTTGGATCGAGCAGCACAATTTCCAGGGCTACCCTGGCGCCGACTG  
CTACCCGTCGTTCTCCGACGACTGAACTGCCTCGGCGGTGCTGTCCANGGACCAT  
CTTCCCCGCCGCAAGTTCGATGAGCAGAAGCTCTTTGCGCANATGTCTTGCGTGAT  
GGCGCAAATCGANNACCCGTCGCCCTGATGAATGATCTGTTCTCTTTCTACAAGAAT  
ATGATCAGGACGAGGNCAATCTGGTGAGCAACTGGTGNACTGNGGACGGA

P30:

GACTGANCTGCCTCGGCGGTGCTGTCGCAGGGACCATCTTCCCCGCCGCCAANNTC  
GATGAGCAGAAGCTCTTTGCGCAGATGTCTTGCGTGATGGCGCAAATCGACGGNCC  
CGTCGCCCTGATGAATGATCTGTTNTCTTTNTTACAAAGAATATGATCAGGACGAGG  
CCAATCTGGTGAGCAACTGGTGCCTGTGGACGGAATCACGATGGACCAGGCCCTC  
ACACGTCTCACCGACGATAACAATCCATGCTTGCGTACGGATTCTGGATATCTTGAAG  
GACAAGGACCCGGACATGCTGGCCACCATCCGCGGGTTTATCCACGGGTATGCGAC  
ATGGCACATCTGCGACTTCAGGTACAGACTGCGGGAGATCTATGACCGCGAAGATC  
TGCAAGAATCGGGCGCCAGGTTCCGCGAGTACTTTGACAAGGCTATTGATGTGGGAT  
GGGTGGATGTTGAAGAGTGGACTTGCCAAGTCCAGGGCTTTGAAGTCGAGGGGCCG

GCCCCTTCGGGATCTGAAGTACAGGCATATCAGGCCAATGCATTTGGGTCTCGGTC  
GATACACAGCGGCACCACAACATAGATTACGTGGGGAGTTCAATACTTGGTCTGTTT  
GAATGGGCGACGCGTTATCTGAGGGGGAAACTCCCGCTAGGCAACGCATAAGCTAT  
TTATCACTCTTTACAACCTTCTACCTCAACTATCTACTTTAATAAATGAATATCGTTTA  
TTCTCTATGATTACTGTATATGCGTTCCTCTAAGACAAATCGAAACCAGCATGCGAT  
CGAATGGCATACAAAAAGTTCTTCCGAAGTTGA

S26:

ACTAACCAGCTCTATGTTTCGCGTCGAGGATCCGATCGAGAGCTCTGGGGTTGCACC  
CACGAGCCCGGTTTGAGAATACTCGTCTTCCGGCTTCCACTACCGGGCGCCGTTACA  
AGTCCGACGAGACCCTCAATCGGGTCTCCTCGAAAATCACACAACCCAAATCTCAG  
GGTGCCTCTCAAGCAATGCTCTACGCCACCGGCCTCACAGAGGAAGACATGTCCAA  
GCCGCAGGTCGGTATCTCCTCGGTGTGGTTCGAGGGTAACCCCTGTAACATGCACTT  
GCACGACCTCTCCGCCATCGTGC GCGATTCCGTGCACCGGGCGGGTCTGGTCCCAAT  
GCGATTCAATTCAGTTGGAGTGTCGGATGGAATCAGTATGGGCACAAAGGGAATGA  
GATACAGCTTGCAGAGTCGGGAGCTGATCGCCGATGGCATTGAGACGGTGATGAAC  
GCCCAGTGGTATGATGCGAATGTGTCGCTCCCGGGTTGCGATAAGAACATGCCGGGT  
GTGTTGATGGCGATGGGACGCACGAATCGACCTAGTATCATGGTCTATGGCGGCAGT  
ATCAAGCCCGGATGCAGTGCAAAGGGCNAAAAGCTGGNNCTGGTTAGCGCGTTCCA  
GTC

S17:

TTATGAACGANNATACAACATCGGTGGCACACCAGCCCTCCTCAAATATCTTCTGAA  
GGAGGGCCTGATCGACGGCTCAGGAATTA CTGTC ACTGGCAAAACAATGAAGGAGA  
ACGTGGCCTCATGGCCCGATTTCCCTGCCGACCAGGACATTATCCGCCCCCTCAGCA

ACCCTATCAAACCATCTGGCCATCTCCAGATTCTTCGCGGGTCGCTGGCACCCGGGCG  
GTTCCGTGGGTAAGATTACTGGCAAGGAGGGTCTGCGGTTTCGAGGGTACGGCCAAG  
TGCTACGACTACGAAGATGCATTTATTGAGTCCCTCGAGCGGGGCGAAATCAAGAA  
GGGCGAGAAGACGGTCGTGATTATCCGGTATGAAGGCCCAAGGGTGGCCCAGGAA  
TGCCTGAGATGCTCAAGCCCAGCGCGGCCATTATGGGTGCCGGTCTGGGCCAGGAC  
GTTGCGCTTCTCACGGACGGAAGATTCTCGGGTGGCAGTCACGGATTCTTGATCGGA  
CATATCGTGCCGGAAGCCATGGAGGGCGGCCCGATCGCCTTGGCCCGGGACGGTGA  
CCGGATCGTGATTGATGCTGAGGAAAGAGTGGTTGATCTGGATATCCCGACTGAAG  
AGCTCGAGAAGAGGAGGAAGGAGTGGAAAGCACCCCCGCTCCGATACCAGAAGGG  
AACTTTGAAGAAGTACTGCACGCTGGTCAGCGATGCCAGCCATGGATGTGTCACGG  
ACGGACCGATCTAGCCTCATGTAATTAGTTATGTCACG

2A MFE1 CB Reads

S10:

CCCGATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTT  
AAGCATTTCCTTCTGAGTATAAGAATCATTCAAAGGATGGGAGCTTCTACTTTCTCC  
CAGTCTTTTGCCGAAGGATATGCGGCATGGGCTCTGATGCTTCCTGCTTTAGTAGGA  
TGCGCGTTGCTTATCTATCGGGCATTCTTTGCCATCCGATATCCCGCCAACCTCCCAC  
TAGCTGGCGAACCTGATGGGAAGAGAACATTCAGCTGGCGAACGAGATGGAGGTAC  
TATATCGACTGCGAGGCTCTATACAAAGAAACATACGACAATTACACCAAACACGG  
CAAGACCGTCCTTCTCCCGGGTCTTGGATTCCGCCATGACATTGTCTTGCTCAAAG  
CGCCATGCGGGATATCATGGCCCGTCCTGAAAAGGAGCTCAGTCACGCCGACGCGG  
TCCTGGAACCTTGTTCAACTGAAGTACTCACTAGGCCATGAGAAATACAAAGCCGATC  
CCTGGCCTTGTATGCTTGTCAAGTCGGACATCAACTCCAAGCTGGAGGCAGTCTGCG



ATGGCATGAACGAGGAGCTGAAGTATGCATTTGATAAATATGTTGGGTGTGATAACC  
GAATCGTGGAAAGAAGTCGATCTGCTGGAGACAATCCGAATGATCATCATGGCCGC  
AGCGAGCCGATTTACCGTCGGATTTCCGCTCTGCCGCAGCGAAGCGTACCTTCGAGC  
CTGTTGGAAAGTCAATGATGGGATAATGATGAACGGCGGTCTGACCGGCGCGACTC  
CACGTCTTCTACGCCAATTGTTGGCCCGTTAGTTACGATGAAACTCCGCCAGAGCA  
TCGAGCANGTCAAGAAACATGTAGAGCCTATCTACCGCCAGCGGGGTGCAGGCATT  
GAGCCAGCAAANAGCGCGGAAAAGCCCGCCAGCGACGAAACGCAGGATCTCTTCC  
AGCAAATGCTCCGGTACGCCCAGA

S16:

TCAATGATNGGATAATGATGAACGGCGNTCTGACCGGCGCGANNCNNCGTNNTTCT  
ACGCCCAATTGTTGGCCCGTTAGTTACGATGAAACTCCGCCAGAGCATCGAGCAGGT  
CAAGAAACATGTAGAGCCTATCTACCGCCAGCGGGTGCAGGCATTGAGCCAGCAAA  
ACAGCGCGGAAAAGCCCGCCAGCGACGAAACGCAGGATCTCTTCCAGCAAATGCTC  
CGGTACGCCCAGAGAGAGCGACCAGGCGAACTGCACGACCTCCCCAGCATGTGCAG  
ACGGCTGTGTTTTGCCAACTTCGCCGCGTGCACCAGACAACCCTGCTTGTGACAAA  
CATGGTCCTCAACATCGTCAGCTCAGACCCGCAACACAACACCATCTCCGTCCTCCG  
AGACGAAGTGAAGGATGTCATCGGCCCTGACAGCAACGCCAAATGGACCAAGTACA  
AAGTCGCGCAGATGATCAAATCCGACAGCGTGGCGCGTGAAACCATGCGGTTATAT  
TCAAACACCAACCGGGGCGTCTTCCGCAAGGTCCTTGTCGAGGGCATCAAAACAGA  
AGACGGAATCGAGCTGCCCAAAGGCGCATACTCTCCTTCCCTGGGCGTCCACTGCA  
ATGCGACCCCGAAACTTTCGAGGATCCTTTCGAGTACAATCCCTTCCGGTTCTCTCG  
GATCCGCGAACAGGCGCCCAGAGACACGAAGGGCCGGTCTAGCGCGAGCCATCTGA  
GCTTTGTGTCGACTTCGCCGGAGCATCTGCCCTTTGGACATGGGGGCCATTTCGTGTC

CGGGGAGGTTCTGGTGGACTTTGAGGTCAAGATGATTGTTGCTTATCTTCTGATGA  
ATTATGATGTCGAGTTTCCGGCCGAGTATAAGGGACAGAGACCTGCGAATCGGTGG  
ATGGCTGAGGCTCTGATGCCGCCTTCTGGTGC GCGGATCAGGATAAAGAGACGAAG  
CGTTCTGGTGGTGCTACTAAT

S15:

GGTGATGTTGATTGAATCCAGGTCCAATGCTATTCCAAGACCTGTCTTTTCCAGCGG  
CCATCGGTGCGGTCTTCGGCGCGGTTGCTATATCTGTAGCCGCTCGATGTATCTACG  
ATCTATTTTTTCATCCGTTGCGAAATTTTCCCGGCCCAAACGGGCAGCCATATGGTC  
TTTCTACGAGTTTTACTACGACGTCATTAGAGATGGCACCTATCTCTGGGAAATTGA  
GAAGATGCACCAGAAATACGGGCCGATTGTTTCGCATCAACTCGAGATCTTTGCACAT  
TCACGATCCGGAATACTTTAACACCATCTACGCTGGAAGTGGCCGCAAGGTAAACAA  
AGAACTGTCCGCCGTGTCTGGCTACACATTCCCGCATTCCACCATCTCGACCCTCGA  
CCATGACCTCCACCGCAAACGCCGCGCTATTGTGAGCCCGTACTTTTCTAAAAGGGC  
CATCGCGGAAATCGAGCCGGTCATCCACGAGCGTCTGAATGTATTGATCTCGCGTCT  
TGCAGAAGCAAAGGGTAGCATCGTGGACCTGACGTCTGCCTTCTCTGCCTACACTGC  
CGATGTCGTCACCTACCACTTCTATGGGACTCATGCAAACCTACATTGGCAGCAAGGA  
CTTCAAGTACGGACTCAAGGATGCACTGACTGTGCTGTTGAATCTGTACAACCTCAC  
GCGCTTCCTGCCAGTCCCCGCAAACACACTCAAAAACCTCCCATTGCCAATCCTCGG  
GTTGATCAACCCGAACTTCCCTCTGGTCGTGTCGGCGCGGGANGCTAACAAAAAGAT  
GGTTTTGGGCTACCTCAACAAACCTGACGAAGACAAAAAGGCCATGAAAGACGCGA  
GATCCAAGTCTGTCATTGTGAGCGCATTGACAGATCCCAATGTCCCCGACGCGGA-  
NAAANACTGGACCGTCTGNTGGATGAGGGCGAAACCATCATCTTTTGCNNGGATTG  
ACNNCACCGCTAGGACACTCG

S17:

ACATNNGCAGCAAGGACTTCAAGTACGGACTCAAGGATGCACTGACTGTGCTGTTG  
AATCTGTACAACCTCACGCGCTTCCCTGCCAGTCCCCGCAAACACACTCAAAAACCTC  
CCATTGCCAATCCTCGGGTTGATCAACCCGAACTTCCCTCTGGTCGTGTCGGCGCGG  
GAGGCTAACAAAAGATGGTTTTGGGCTACCTCAACAAACCTGACGAAGACAAAA  
GGCCATGAAAGACGCGAGATCCAAGTCTGTCATTGTGAGCGCATTGACAGATCCCA  
ATGTCCCCGACGCGGAGAAAACACTGGACCGTCTGTTGGATGAGGGCGAAACCATC  
ATCTTTGCGGGGATTGACACCACCGCTAGGACACTCGGCGTTGCATTGTTCCACCTG  
CTCAACAATAAGGATGTGCTGATGAAGCTGCGGAAAGAATTGCAGGCTGTTGCGAA  
GCCGGATGGTCAGCAATGGACCACGACAGAGCTCGAGGCTGTGCCTTATATGAGAG  
GCGTCGTCCAGGAAGCAATCCGCCTTGCGTACGGTCTGGTCGTGCGCATTCCCAGAA  
TCTCCCCACACGAAGCTCTGCGGTACAACGGTTTCGTAATTCCACCGGGAACACCGG  
TCAGCCAGTCAACCTACCTGGTCAACAACGACCCTTCCGTATTCCCAAATCCTCAAG  
TCTTCGATCCCGAGCGTTGGGTCAAGGCTGCACAGGACGGCGTCAGCCTCGACAAA  
TACATGGTCAGCTTCAGCAAGGGCAGCAGAGGATGTCTCGGTATCAACCTGGCATAT  
GCAAACCTCTACCTCGGTATCGCAAGAGTTGCAACATCCTTGGACATGGAGCTGTTT  
GAAACGACTGCGAAAGCTATTTCTGTATACCATACACGGGGCTTCGCCTTTCCAAG  
GAAGGCGACGGTGCAGTCAAGGCGCGAGTTATGGGTCTCTGCAAGTAGCCTCATGT  
AATTAGTTANGTCACGCTACA

TPS MFE1 CB Reads

S10:

ATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAATCGTTAAGC  
ATTTCTTCTGAGTATAAGAATCATTCAAAGGATGGGAGCTTCTACTTTCTCCCAGTC

TTTTGCCGAAGGATATGCGGCATGGGCTCTGATGCTTCCTGCTTTAGTAGGATGCGC  
GTTGCTTATCTATCGGGCATTCTTTGCCATCCGATATCCCGCCAACCTCCCCTAGCT  
GGCGAACCTGATGGGAAGAGAACATTCAGCTGGCGAACGAGATGGAGGTAATATAT  
CGACTGCGAGGCTCTATACAAAGAAACATACGACAATTACACCAAACACGGCAAGA  
CCGTCCTTCTCCCGGGTCTTGGATTCCGCCATGACATTGTCTTGCCTCAAAGCGCCAT  
GCGGGATATCATGGCCCGTCCTGAAAAGGAGCTCAGTCACGCCGACGCGGTCTGG  
AACTTGTTCAACTGAAGTACTCACTAGGCCATGAGAAATACAAAGCCGATCCCTGGC  
CTTGTATGCTTGTCAAGTCGGACATCAACTCCAAGCTGGAGGCAGTCTGCGATGGCA  
TGAACGAGGAGCTGAAGTATGCATTTGATAAATATGTTGGGTGTGATACCGAATCGT  
GGAAAGAAGTCGATCTGCTGGAGACAATCCGAATGATCATCATGGCCGCAGCGAGC  
CGATTTACCGTCGGATTTCCGCTCTGCCGCAGCGAAGCGTACCTTCGAGCCTGTTGG  
AAAGTCAATGATGGGATAATGATGAACGGCGGTCTGACCGGCGCGACTCCACGTCT  
TCTACGCCCAATTGTTGGCCCGTTAGTTACGATGAAACTCCGCCAGAGCATCGAGC-  
NNTCAAGAAACATGTAGAGCCTATCTACCGCCAGCGGGTGCAGGCATTGAGCCAGC  
AAACAGCGCGGAAAAGCCCGCCAGCGACGAAACGC

P30:

GGGTGCAGNCATNGAGCCAGCAAAAANAGCGNGGAAAAGCCCGCCAGCGACGAAAC  
GCAGGATCTCTTCCAGCAAATGNTCCGGTACGCCAGAGAGAGCGACCAGGCGAAC  
TGCACGACCTCCCCAGCATGTGCAGACGGCTGTGTTTTGCCAACTTCGCCGCCGTGC  
ACCAGACAACCCTGCTTGTGACAAACATGGTCCTCAACATCGTCAGCTCAGACCCGC  
AACACAACACCATCTCCGTCNTCCGAGACGAAGTGAAGGATGTCATCGGCCCTGAC  
AGCAACGCCAAATGGACCAAGTACAAAGTCGCGCAGATGATCAAATCCGACAGCGT  
GGCGCGTGAAACCATGCGGTTATATTCAAACACCAACCGGGGCGTCTTCCGCAAGG

TCCTTGTCGAGGGCATCAAAACAGAAGACGGAATCGAGCTGCCCAAAGGCGCATAAC  
GTCTCCTTCCTGGGCCGTCCACTGCAATGCGACCCCGAAACTTTCGAGGATCCTTTC  
GAGTACAATCCCTTCCGGTTCTCTCGGATCCGCGAACAGGGCGCCCAGAGACACGAA  
GGGCCGGTCTAGCGCGAGCCATCTGAGCTTTGTGTCGACTTCGCCGGAGCATCTGCC  
CTTTGGACATGGGGGCCATTCGTGTCCGGGGAGGTTCCCTGGTGGACTTTGAGGTCAA  
GATGATTGTTGCTTATCTTCTGATGAATTATGATGTCGAGTTTCCGGCCGAGTATAAG  
GGACAGAGACCTGCGAATCGGTGGATGGCTGAGGCTCTGATGCCGCCTTCTGGTGC  
GCGGATCAGGATAAAGAGACGAAGCTAGGCTATTTATCACTCTTTACAACCTTCTACC  
TCAACTATCTACTTTAATAAATGAATATCGTTTATTCTCTATGATTACTGTATATGCG  
TTCCTCTAAGACAAATCGAAACCAGCATGCGATCGAATGGCATAACAAAAGTTCTCCG  
AAGTGATCAA

S26:

ACTAACCAGCTCTATGCTATTCCAAGACCTGTCTTTTCCAGCGGCCATCGGTGCGG  
TCTTCGGCGCGGTTGCTATATCTGTAGCCGCTCGATGTATCTACGATCTATTTTTTCA  
TCCGTTGCGAAATTTCCCGGCCCAAACGGGCAGCCATATGGTCTTTCTACGAGTT  
TACTACGACGTCATTAGAGATGGCACCTATCTCTGGGAAATTGAGAAGATGCACCA  
GAAATACGGGCCGATTGTTTCGCATCAACTCGAGATCTTTGCACATTCACGATCCGGA  
ATACTTTAACACCATCTACGCTGGAAGTGGCCGCAAGGTTAACAAAGAACTGTCCGC  
CGTGTCTGGCTACACATTCCCGCATTCACCATCTCGACCCTCGACCATGACCTCCA  
CCGCAAACGCCGCGCTATTGTGAGCCCGTACTTTTCTAAAAGGGCCATCGCGGAAAT  
CGAGCCGGTCATCCACGAGCGTCTGAATGTATTGATCTCGCGTCTTGCAGAAGCAAA  
GGGTAGCATCGTGGACCTGACGTCTGCCTTCTCTGCCTACACTGCCGATGTCGTAC  
CTACCACTTCTATGGGACTCATGCAAACACTACATTGGCAGCAAGGACTTCAAGTACGG

ACTCAAGGATGCACTGACTGTGCTGTTGAATCTGTACAACCTCACGCGCTTCCTGCC  
AGTCCCCGCAAACACACTCAAAAACCTCCCATTGCCAATCCTCGGGTTGATCAACCC  
GAACTTCCCTCTGGTCGTGTCGGCGCGGGANGCTAACAAAAAGATGGTTTTGGGCTA  
CCTCAACAAACCTGACGAAGACAAAAAGGCCATGAAAGACGCGAGATCCAAGTCTG  
TCATTGTGAGCGCATTGACAGATCCCAATGTCCCCGACGCGGANAAAACACTGGAC  
CGTCTGTTGGATGAGGGCGAAACCATCATCTTTGNNGGATTGACACCACCGCTAGGA

S17:

GGCAGCAAGGACTTCAAGTACGGACTCANGGATGCACTGACTGTGCTGTTGAATCT  
GTACAACCTCACGCGCTTCCTGCCAGTCCCCGCAAACACACTCAAAAACCTCCCATT  
GCCAATCCTCGGGTTGATCAACCCGAACTTCCCTCTGGTCGTGTCGGCGCGGGAGGC  
TAACAAAAAGATGGTTTTGGGCTACCTCAACAAACCTGACGAAGACAAAAAGGCCA  
TGAAAGACGCGAGATCCAAGTCTGTCATTGTGAGCGCATTGACAGATCCCAATGTCC  
CCGACGCGGAGAAAACACTGGACCGTCTGTTGGATGAGGGCGAAACCATCATCTTT  
GCGGGGATTGACACCACCGCTAGGACACTCGGCGTTGCATTGTTCCACCTGCTCAAC  
ATAAGGATGTGCTGATGAAGCTGCGGAAAGAATTGCAGGCTGTTGCGAAGCCGGA  
TGGTCAGCAATGGACCACGACAGAGCTCGAGGCTGTGCCTTATATGAGAGGCGTCG  
TCCAGGAAGCAATCCGCCTTGCGTACGGTCTGGTCGTGCGCATTCCCAGAATCTCCC  
CACACGAAGCTCTGCGGTACAACGGTTTCGTAATTCCACCGGGAACACCGGTCAGC  
CAGTCAACCTACCTGGTCAACAACGACCCTTCCGTATTCCCAAATCCTCAAGTCTTC  
GATCCCGAGCGTTGGGTCAAGGCTGCACAGGACGGCGTCAGCCTCGACAAATACAT  
GGTCAGCTTCAGCAAGGGCAGCAGAGGATGTCTCGGTATCAACCTGGCATATGCAA  
AACTCTACCTCGGTATCGCAAGAGTTGCAACATCCTTGGACATGGAGCTGTTTGAAA  
CGACTGCGAAAGCTATTTCTGTATAACCATAACACGGGGCTTCGCCTTTCCCAAGGAAG

GCGACGGTGCAGTCAAGGCGCGAGTTATGGGTCTCTGCAAGTAGCCTCATGTAATTA  
GTTANGTCACG

## References

1. A. Ludwiczuk, K. Skalicka-Wozniak, M.I. Georgiev. Terpenoids. *Pharmacognosy Academic Press* (2017), pp. 233-266
2. Y. Yan, Q. Liu, X. Zang, S. Yuan, U. Bat-Erdene, C. Nguyen, J. Gan, J. Zhou, S. E. Jacobsen, Y. Tang. Resistance-Gene Directed Discovery of a Natural Product Herbicide with a New Mode of Action. *Nature* 2018 Jul 11; 559(7714): 415-418
3. I. Heap. Global Perspective of Herbicide-Resistant Weeds. *Pest Management Science* 2014 Sep; 70(9): 1306-1315
4. W. Tang, D. Wang, Y. Tian, X. Fan, C. Wang, X. Lu, P. Li, X. Ji, H. Liu. Metabolic Engineering of *Yarrowia lipolytica* for Improving Squalene Production. *Bioresource Technology* 2021 Mar; 323: 124652
5. J. Nicaud. *Yarrowia lipolytica*. *Yeast* 2012 Sep 07; 29(10): 409-418
6. S. Rodriguez, C.M. Denby, T.V. Vu, E.E.K. Baidoo, G. Wang, J.D. Keasling. ATP Citrate Lyase Mediated Cytosolic Acetyl-CoA Biosynthesis Increases Mevalonate Production in *Saccharomyces Cerevisiae*. *Microbial Cell Factories* 2016 Mar 03; 15(48)
7. M. Marsafari, P. Xu. Debottlenecking Mevalonate Pathway for Antimalarial Drug Precursor Amorphadiene Biosynthesis in *Yarrowia lipolytica*. *Metabolic Engineering Communications* 2020 Jun; 10: e00121
8. S. Moser, H. Pichler. Identifying and Engineering the Ideal Microbial Terpenoid Production Host. *Applied Microbiology and Biotechnology* 2019 May 25; 103: 5501-5516



9. S. Rubat, I. Varas, R. Sepulveda, D. Almonacid, F. Gonzalez-Nilo, E. Agosin. Increasing the Intracellular Isoprenoid Pool in *Saccharomyces Cerevisiae* by Structural Fine-Tuning of a Bifunctional Farnesyl Diphosphate Synthase. *FEMS Yeast Research* 2017 Jun 17; 17(4): fox032
10. C. Schwartz, M. Shabbir-Hussain, K. Frogue, M. Blenner, I. Wheeldon. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. *ACS Synthetic Biology* 2017; 6(3): 402-409
11. T.M. Souza-Moreira, C. Navarrete, X. Chen, C.F. Zanelli, S.R. Valentini, M. Furlan, J. Nielsen, A. Krivoruchko. Screening of 2A Peptides for Polycistronic Gene Expression in Yeast. *FEMS Yeast Research* 2018 Aug; 18(5): foy036
12. X. Xiong, S. Chen. Expanding Toolbox for Genes Expression of *Yarrowia lipolytica* to Include Novel Inducible, Repressible, and Hybrid Promoters. *ACS Synthetic Biology* 2020 Jun 25; 9(8): 2208-2213
13. M. Larroude, T. Rossignol, J.M. Nicaud, R. Ledesma-Amaro. Synthetic Biology Tools for Engineering *Yarrowia lipolytica*. *Biotechnology Advances* 2018 Dec; 36(8): 2150-2164