Identification of a Signal Peptide and its Application in Genetic Engineering of the Thraustochytrid *Aurantiochytrium* sp.

January 2019

JUNTILA Darryl Joy Caceres

Identification of a Signal Peptide and its Application in Genetic Engineering of the Thraustochytrid *Aurantiochytrium* sp.

A Dissertation Submitted to

the Graduate School of Life and Environmental Sciences,

the University of Tsukuba

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

(Doctoral Program in Integrative Environment and Biomass Sciences)

JUNTILA Darryl Joy Caceres

Abbreviations

DHA	docosahexaenoic acid
EPA	eicosapentaenoic acid
SP1	secreted protein 1
SP2	secreted protein 2
vWF	von Willebrand factor
YE	yeast extract
GTY	glucose, tryptone, yeast extract medium
M4	glucose, low peptone and yeast extract medium
SDS-PAGE	sodium dodecyl sulfate- polyacrylamide gel electrophoresis
qPCR	quantitative polymerase chain reaction
HPLC	high performance liquid chromatography
MS	mass spectrometry
NCBI	National Center for Biotechnology Information
JGI	Joint Genome Institute
LPD	local protein database
AaBgl	*Aspergillus aculeatus* β-glucosidase
pNPG	4-nitrophenyl-β-D-glucopyranoside
pNP	4-nitrophenol
4-MUG	4-methyllumbeferyl-β-D-glucopyranoside
4-MU	4-methylumbeferone
OD_{660}	optical density measured at 660 nm
rpm	revolutions per minute
EF-1α	elongation factor 1 α subunit
Ubi	ubiquitin
Neo^r	neomycin resistance gene
pUbi-Neo^r	plasmid for neomycin resistance gene under ubiquitin promoter
pEF-Neo^r	plasmid for neomycin resistance gene under EF-1α promoter
GY	glucose yeast-extract medium
PY	peptone yeast extract basal medium
TAGs	triacylglycerols

Abstract

Thraustochytrids are osmoheterotrophic protists and are a promising group of oleaginous microorganisms due to their production of valuable oils. Squalene, a triterpene hydrocarbon known for its antioxidant, antitumorigenic and emollient benefits in medicine and cosmetics, is accumulated in some strains of thraustochytrids especially in the genus *Aurantiochytrium*. Thraustochytrids can produce extracellular enzymes such as cellulase, lipase and protease, to break down organic substrates for nutrition. To circumvent production costs in using growth medium, many studies try to grow thraustochytrids in cost-effective substrates such as lignocellulosic wastes. The objective of this study is to cultivate the strain *Aurantiochytrium* sp. 18W-13a, a DHA and squalene producer, in low-cost carbon source such as cellulosic waste. To achieve this, firstly, proteomic analysis of the secretome was done to identify secreted proteins and their signal peptides. After which, application of the signal peptide for targeted secretion of a heterologous cellulase, β-glucosidase from *Aspergillus aculeatus* (AaBgl), was done for cellobiose utilization. Two secreted proteins, SP1 and SP2, were identified with N-terminal signal peptide sequences. SP2 contains the von Willebrand factor (vWF) and PAN/APPLE domains, which are known to be involved in cell to cell interaction. The signal peptide sequence of SP2 was fused to the coding sequence of *AaBgl* and introduced into the cells by combined electroporation and glass bead treatment. The transformant strain, AaBgl^{+}, exhibited increasing growth and β-glucosidase activity with cellobiose as sole carbon source. Native PAGE zymogram showed an active β-glucosidase enzyme in the supernatant of the AaBgl^{+} strain. Both wild-type and AaBgl^{+} strains can produce squalene and the fatty acids DHA and DPA. This work is one of the first reports in targeted secretion of a functional enzyme in thraustochytrids. Moreover, enhanced expression of cellulases has never been done in thraustochytrids. I hope that this study could pave way for establishing lignocellulosic biomass-degrading thraustochytrid strains.

Keywords: secretome analysis, targeted secreted expression, alternative carbon source

General introduction

Thraustochytrids, osmoheterotrophic protists which are decomposers in the marine ecosystem, are emerging as promising sources of valuable lipids such as docosahexaenoic acid (DHA) and squalene. To decompose organic macromolecules in their environment, they are thought to secrete degradation enzymes. However, at present, the secretome analysis of thraustochytrids is still insufficiently studied. Although some strains were reported to exhibit cellulase, protease and lipase activity (Kanchana et al. 2011; Nagano et al. 2011; Liu et al. 2014), in-depth studies of growth and hydrolytic enzyme activity using complex organic substrates are few.

The thraustochytrid strain *Aurantiochytrium* sp. 18W-13a is a promising source for squalene production, as it can produce up to 20% squalene per dry cell weight under optimal conditions. Squalene, a triterpene hydrocarbon, is known as a moisturizing agent because of its natural antioxidant properties (Nakazawa et al. 2012a). However, studies with this strain mainly rely on commercial glucose and peptones as nutrient sources.

The aim of my study is to cultivate this strain in cost-effective substrates such as cellulosic waste, mostly from lignocellulosic biomass which is the most abundant and easily accessible carbon source. However, lignocellulose has a complex and recalcitrant matrix that needs pretreatment and enzymatic hydrolysis for degradation (Isikgor and Becer 2015). In order to achieve this objective, I divided my research into two parts: 1) identification of secreted proteins and their signal peptides in 18W-13a and 2) application of the signal peptide in the genetic engineering of the strain, specifically for targeted secreted expression of a heterologous hydrolytic enzyme.

Because secretome analysis is understudied in thraustochytrids, Part I aims to identify constitutively expressed secreted proteins in 18W-13a under standard laboratory

conditions. By identifying constitutively secreted proteins, I can identify signal peptide sequences that can be used to tag heterologous protein genes for secretion. In expression systems like yeast, the use of homologous or the host's own signal peptide sequence fused with heterologous proteins can enhance protein production, indicating preference of certain signal peptide sequences for different organisms. In my study, 18W-13's own signal peptide is used to ensure secretion of the target protein in this organism.

Part II of the study is the application of the identified signal peptide sequence and its fusion to the coding sequence of the target enzyme, β-glucosidase from *Aspergillus aculeatus* (AaBgl). This enzyme is widely used in engineering yeast strains for lignocellulose conversion into ethanol and is highly active in degrading cellooligosaccharides such as cellobiose (glucose dimer) (Sakamoto et al. 1985b; Kawaguchi et al. 1996; Fujita et al. 2002). Electroporation will be used for the introduction of the gene cassette into cells and initial screening of transformants in selective medium with G418 neomycin antibiotics will be done. Transformant strains with positive insertions of transgenes will be evaluated for their β-glucosidase activity. After which, detailed evaluation of the AaBgl-expressing strain will be done by cultivating the strain under cellobiose as the sole carbon source. Native polyacrylamide gel electrophoresis (PAGE) zymogram analysis will also be done to check for the presence of a functional enzyme in the supernatant of the cultures. Also, biomass and lipid analysis, particularly squalene and DHA contents, will be measured in the wild-type and transformant strains.

Ultimately, the research objective of my study is to generate a transformant that could express and secrete AaBgl, under the influence of 18W-13a's own signal peptide, to utilize cellobiose as its carbon source, as the initial enzyme for cellulose degradation.

Part I

Identification of secreted proteins and signal peptides in *Aurantiochytrium* sp. 18W-13a

Introduction

A wide range of heterotrophic microorganisms are capable of degrading organic matter, contributing a significant role in the recycling of nutrients present in the ecosystem. Decomposers are cosmopolitan in the environment and they play an important role in biomass-rich regions such as mangrove systems (Thompson et al. 2013). Thraustochytrids are osmoheterotrophic protists belonging to the class *Labyrinthulomycetes* (Raghukumar 2008). They are usually found in mangrove, marine and estuarine habitats and were demonstrated to produce a range of extracellular enzymes which can break down and mineralize organic matter in such environments (Liu et al. 2014). The cultures of various species exhibit activities of cellulase (Nagano et al. 2011), lipase (Kanchana et al. 2011), protease, amylase, and glucosidase (Liu et al. 2014; Taoka et al. 2009). However, cultivation of these organisms in media containing high-molecular weight organic substrates as carbon sources is still understudied.

Thraustochytrids are also emerging as a promising group of oleaginous microorganisms due to their high production of valuable oils, especially ω-3 fatty acids docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) (Gupta et al. 2012; Raghukumar, 2008). Previous studies reported DHA and palmitic acid as the two major fatty acids produced by thraustochytrids (Gupta et al. 2016; Gupta et al. 2015; Gupta et al. 2012; Ma et al. 2015). Accumulation of squalene, a hydrocarbon precursor in the sterol synthesis pathway, has also been studied in thraustochytrids. Squalene is mostly used as a moisturizing agent because of its natural antioxidant properties. It has many pharmaceutical and medical applications including tumor suppression, antibacterial and antifungal functions (Nakazawa et al. 2012a). Recently, the thraustochytrid strain *Aurantiochytrium* sp. 18W-13a, was found to produce high amounts of squalene. Under optimal conditions of 25 °C, 25-50 % (v/v)

seawater and 2-6 % (w/v) glucose, squalene content reaches to about 20% of the dry cell weight, which is several hundred-fold higher than those previously reported in literature for other thraustochytrids (Kaya et al. 2011; Nakazawa et al. 2012a). Screening of different labyrinthulomycetes for squalene production showed that *Aurantiochytrium* species accumulate higher amounts of squalene than other genera (Nakazawa et al. 2014).

Among different carbon and nitrogen sources, glucose and peptones are the preferred nutrient sources of many thraustochytrid species (Raghukumar 2008). To circumvent production costs caused by expensive growth medium, many studies attempted to grow several species in cheaper organic substrates. The utilization of organic waste as low-cost carbon source for the cultivation of *Aurantiochytrium* sp. 18W-13a with squalene as its final product is a very promising combinatorial strategy for waste utilization and lipid production. Depending on the chemical composition of the organic waste to be utilized, it is desirable that the strain used can break down the organic material for energy and as source material for its cellular components. Preliminary culture of *Aurantiochytrium* sp. 18W-13a in simple and complex sugars showed its inability to utilize disaccharides and polysaccharides for its growth, which may indicate lack of secreted hydrolytic enzymes for these substrates. Thus, this necessitates secretome analysis to identify secreted proteins and possible hydrolytic enzymes and provide information about adaptation mechanisms of the organism to its environment. Part I of this study aims to identify constitutively expressed secreted proteins of *Aurantiochytrium* sp. 18W-13a under standard laboratory growth conditions.

Materials and Methods

2.1. Strain and culture conditions

The strain *Aurantiochytrium* sp. 18W-13a was kindly provided by Prof. M. M. Watanabe of the Algal Biomass and Energy System Research and Development Center in the University of Tsukuba. In order to identify constitutively expressed secretory proteins from the strain in different growth stages and culture conditions, I grew the cells under two conditions: 1) GTY medium (2% (w/v) glucose, 1% (w/v) Bacto-tryptone (Becton Dickinson, Franklin Lakes, NJ), 0.5% (w/v) Bacto-yeast extract (Becton Dickinson), 50% (v/v) artificial seawater Marine Art SF-1 (Osakayakken, Osaka, Japan)) at 20°C and 2) M4 medium (2% (w/v) glucose, 0.15% (w/v) Bacto-peptone, 0.1% (w/v) Bacto-yeast extract, 0.025% (w/v) KH_2PO_4, 50% (v/v) artificial seawater) at 30°C, both with shaking at 100 rpm. Extracellular proteins and RNA samples were collected at day 0.5, 1.5 and 3, representing lag, exponential and early stationary phases, respectively.

2.2. Extracellular protein collection

Supernatants of the cultures were collected by repeated centrifugation of 100 mL cell cultures at 6000×g, 20 °C for 10 min to remove the cells. The clear supernatant was then concentrated using Amicon Ultra-15 centrifugal filter units (Merck, Germany). Buffer exchange using 50 mM sodium phosphate buffer (pH 7.0) was done six times to remove any residual salts as possible. After concentration, approximately 250 µL of extracellular protein extract was collected and the total protein concentration was measured using Quickstart Bradford Protein Assay (Bio-rad, USA) with bovine serum albumin as a standard.

2.3. Electrophoresis of the proteins and proteomic analysis

For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis, 20 µg of the extracellular proteins were loaded unto 7.5% polyacrylamide precast e-PAGEL gel (ATTO, Tokyo, Japan) and was run at 24 mA for 1 h. The gels were stained using Coomassie Brilliant Blue (CBB) R-250 for visualization and Oriole™ Fluorescent Gel Stain for gel excision (Bio-rad). Initially, proteins fractionated on the SDS-PAGE were electrically blotted onto a polyvinylidene difluoride (PVDF) membrane (Immobilon P 0.45 µm, Merck) and stained with CBB R-250. The prominent proteins bands (molecular mass 60-70 kDa) were excised and the amino-terminal amino acid sequences were analyzed by Edman degradation using ABI Procise 494-HT (Applied Biosystems, Foster City, CA) at the Functional Genomics Facility, National Institute for Basic Biology, Aichi, Japan. Another method using mass spectrometry was also performed for the other candidate protein bands (molecular mass > 100 kDa) as previously described (Yoneda et al. 2016). Gel slices were destained and digested using sequence-grade modified trypsin (Promega, Madison, WI). After which, acetonitrile: 5% (v/v) formic acid aqueous solution 1:1 (v/v) was used to extract peptides from the gel pieces. Digested peptides were separated by HPLC with a capillary pump (Agilent 1200, Agilent, Santa Clara, CA) equipped with ZORBAX 300SB-C18 (0.3mm x 150mm, Agilent) column. Mobile phase A (H_2O:acetonitrile 95:5 (v/v) containing 0.1% formic acid) and mobile phase B (H_2O:acetonitrile 1:9 (v/v) containing 0.1% formic acid) were used for elution at 5 $\mu L min^{-1}$ under the following gradient: ratio of mobile phase B started at 5% and increased to 50% for 60 min, then B ratio elevated rapidly to 95% in 1 min and was maintained at 95% for 14 min. The eluted peptides were applied to an electrospray ionization quadrupole time-of-flight (ESI-Q/TOF) system (Agilent 6520 Accurate-Mass QTOF LC/MS, Agilent). The MS scan range was set to *m/z* 105-3,000, and

multicharged ions (+2, +3 and >+3) were preferentially subjected to MS/MS analysis. Obtained data was exported as Mascot generic files. Peptide sequences were queried into a local proteome database using Mascot Server (version 2.2.06, Matrix Science, Boston, MA). The identified proteins were considered as complete sequences of the related peptides. Also, conserved search domain analysis using NCBI (Marchler-Bauer et al. 2015) and BLASTp search analysis (Altschul et al. 2005) using Joint Genome Institute (JGI) *Aurantiochytrium limacinum* ATCC MYA-1381 proteome (http://genome.jgi.doe.gov/Aurli1/Aurli1.home.html) (Nordberg et al. 2014) were done for the identified proteins. Signal peptide prediction using SignalP 4.1 (Petersen et al. 2011) was used to determine the presence of a signal peptide sequence within the protein.

2.4. RNA extraction and qPCR analysis

Combined TRIzol™ Reagent (Ambion, ThermoScientific, Waltham, MA) and RNeasy™ Mini Kit (Qiagen, Hilden, Germany) were used for extraction of RNA samples. The cDNA was synthesized using Primescript™ RT Reagent Kit with gDNA Eraser (Perfect Real Time) (Takara Bio, Ohtsu, Japan). GoTaq™ qPCR Master Mix (Promega) was used for qPCR analysis of transcripts of the extracellular proteins SP1 and SP2, using specific primers *sp1for* (5'-CTATGACAACCACGCCAAAAAG-3'), *sp1rev* (5'-ACACAGAACGCCTGCCAAG-3'), *sp2fo*r (5'-CCTTTTCTGGCTACTTGGC-3'), and *sp2rev* (5'-AGGTTGGTGCGTCTTTGC-3'). Piko 96 Real-Time PCR System (ThermoScientific) and PikoReal Software 2.2 were used to analyze the relative quantification in respect to the reference gene, actin, which was amplified using primers *actinfor* (5-CTCTTCCAGCCGTCCTTCATC-3') and *actinrev* (5'-GGGCGACCATCTCCTTCTC -3').

Results and Discussion

Extracellular proteins ranging from 50-250 kDa were observed in both GTY (20˚C) and M4 (30˚C) media (Fig. 1). Several proteins are constitutively produced under both conditions which may signify essential functions. Firstly, I focused on the prominent protein bands with 60-70 kDa apparent molecular masses on the gel present in both conditions (marked by a bracket in Fig.1) and attempted to determine the N-terminal amino acid sequences using Edman degradation sequencing. Results showed that the 60-70 kDa protein bands, which appear as diffused bands with different mobilities, possessed basically identical 17-amino acid sequence (PEMTSFTAIGNNLAIDA) in their N-terminal regions. However, Blastp search against a local protein database constructed from the draft genome sequence of 18W-13a (unannotated), NCBI's non-redundant protein and JGI's *Aurantiochytrium limacinum* ATCC MYA-1381 databases did not yield any significant hits containing this protein sequence in the N-terminal regions of any database proteins.

Mass spectrometry was then attempted to identify other proteins bands (100-250 kDa). Obtained data were queried against the local protein database. Mascot search showed results for two proteins, designated as SP1 and SP2 (marked with arrows in Fig.1) with IDs of the local protein database, LPD8921 and LPD1644, respectively. The apparent masses of SP1 and SP2 are 250 kDa and 110 kDa, respectively. These values are relatively close to their predicted masses of 251 kDa and 110 kDa which were calculated based on their protein sequences (with lengths of 2,353 and 1,020 amino acids for SP1 and SP2, respectively) (Table 1). Additionally, SignalP 4.1 analysis showed the presence of N-terminal signal peptide tags in these proteins (Fig. 3).

Since the reference database used was the unannotated local protein database, NCBI conserved domain search was done to determine any putative functional domains existing

within the protein sequences. Conserved domain search showed results only for SP2 which contains domains representing the von Willebrand factor (vWF) type A and PAN/APPLE superfamilies (Marchler-Bauer et al. 2015) (Table 1).

The presence of other paralogous and orthologous proteins was then determined through Blastp search of the full-length sequences of SP1 and SP2 against the local database and the annotated protein database of *A. limacinum* ATCC MYA-1381 in JGI (Table 1). For SP1, one paralogous (E value=0) protein, designated as LPD8920, from the local protein database was found. Two proteins, with protein IDs of Aurli1.3202 and Aurli1.3204 from the *A. limacinum* ATCC MYA-1381 database, are orthologous (E-value=0) to SP1. Aurli1.3202 and Aurli1.3204 are annotated as subtilases in the JGI database but the actual portions for their subtilase domains are rather short. Conserved domain search for LPD8920, Aurli1.3202 and Aurli1.3204 did not show any known functional domains (data not shown), similar to the results obtained with SP1. However, interestingly, Constraint-based Multiple Alignment Tool (COBALT) (Papadopoulos and Agarwala 2007) showed conserved regions among SP1 and its orthologous and paralogous proteins. Additionally, this conserved motif appears tandemly within SP1 and LPD8920 protein sequences from 18W-13a while it only exists in a single stretch in Aurli.3202 and Aurli.3204 from *A. limacinum* (Supplementary Fig. 1). These conserved regions could indicate a common function yet to be known. Aurli1.3202 and Aurli1.3204 are shorter in length, 1307 and 1094 amino acids, respectively, in comparison to the 2353 and 2314 amino acid lengths of SP1 and LPD8920, respectively. Another interesting finding is the loci of the coding regions of these proteins within the genomes of their respective organisms. The coding regions for SP1 and its paralogous protein, LPD8920, were positioned flanking each other in the same scaffold in the local genome database of 18W-13a. Similarly, the coding regions for Aurli1.3202 and Aurli1.3204 were also located closely to each other in the genomic database of *A. limacinum*

(JGI). Although functional information about these proteins is lacking, it is interesting to note that these proteins exist in both species of *Aurantiochytrium* and, furthermore, the genes for the proteins may be duplicated in both *Aurantiochytrium* genomes. SP1, LPD8920 and Aurli1.3204 have N-terminal signal peptide sequences based on the SignalP 4.1 analyses (Supplementary Table 1; Fig. 3).

Blastp search showed only one paralogous protein to SP2, designated as LPD9373, from the local protein database. LPD9373 has 824 amino acids while SP2 has 1020 amino acids. Conserved domain search for LPD9373 showed a vWF type A domain within the protein, like SP2. SignalP 4.1 analysis also showed the presence of a N-terminal signal peptide in LPD9373 (Supplementary Table 1).

To confirm the expression of SP1 and SP2 at a transcript level, I conducted a quantitative reverse transcription-PCR analysis using specific primers targeting *sp1* and *sp2* transcripts. Actin gene was used to normalize the expression levels. Results showed constitutive expression of *sp1* and *sp2,* especially in GTY medium. In relation to actin's expression value of 1.0, relative expression reaches 8-20% under GTY (Fig. 2). The low expression seen under M4 could be attributed to lower culture growth, possibly due to less nutrients present, in comparison to GTY.

In previous reports, some species of *Aurantiochytrium* can grow in different forms of cellulose while others cannot. Taoka et al. (2009) reported protease, lipase, urease, phosphatase and α-glucosidase but no amylase, gelatinase, cellulase and chitinase activities for *A. limacinum* ATCC MYA-1381 and *Aurantiochytrium* sp. mh0186. The strain *Aurantiochytrium* sp. KRS101 can grow in carboxymethylcellulose, cellobiose, and pre-treated empty fruit bunch material with better hydrolytic activities observed for cell-associated fractions compared to extracellular fractions (Hong et al. 2012). Another strain of

A. limacinum isolated from a sediment sample can exhibit degradation activity against carboxymethylcellulose (Liu et al. 2014b). It is possible that extracellular enzyme production is strain-specific based on the diverse repertoire of enzymes produced by different species and strains. Another possible explanation is that enzyme production is induced by its substrate and that different species may prefer different substrates for their growth. In the analysis of the extracellular proteins in the study, I could not find any proteins related to the degradation enzymes. This further necessitates research about the secreted proteins and hydrolytic enzymes that 18W-13a can produce.

This study aimed to identify secreted proteins from *Aurantiochytrium* sp. 18W-13a under standard laboratory culture medium GTY and M4. These two media are composed of glucose, tryptone/peptone and yeast extract with GTY medium containing more of the protein sources than M4 medium. The protein banding patterns for both GTY and M4 are similar even at different timepoints, indicating that many proteins are always produced under the influence of standard nutrients such as glucose, tryptone/peptone and yeast extract. Out of several candidate proteins, two proteins, SP1 and SP2, were identified using Mascot search against a local protein database (Fig.1; Table 1). SP1 has no known function based on the absence of known conserved domains in its sequence. Interestingly, this protein has homologous proteins in both 18W-13a and *A. limacinum* ATCC MYA-1381. A conserved region in these proteins was found but its function is yet to be known (Supplementary Fig. 1). SignalP 4.1 prediction confirms the presence of highly probable signal peptide sequence tags in the N-terminal regions of these proteins, which could indicate a role in the secretome of *Aurantiochytrium* species. The orthologous proteins of SP1, Aurli1.3202 and Aurli1.3204, contain a short subtilase region. A previous study reported over 200 transcripts of peptidases including subtilases in a thraustochytrid parasite infecting hard clams, wherein these proteins are suspected to play a role in protein degradation of the host tissue (Rubin et

al. 2014). While *Aurantiochytrium* species are not parasitic, some of its proteins might be related to the degradation of extracellular macromolecules.

SP2 contains von Willebrand factor (vWF) type A and PAN/APPLE-like domains (Table 1). The vWF type A domain mediates adhesion via metal ion-dependent adhesion sites (MIDAS). Proteins containing this domain have several functions including cell adhesion, migration and signal transduction. Most of which are eukaryotic extracellular matrix proteins such as integrin, collagen and the vertebrate von Willebrand factor, the protein that mediates platelet adhesion to collagen. (Whittaker and Hynes 2002; Marchler-Bauer et al. 2015). The PAN/APPLE-like domain have functions related to protein-protein or protein-carbohydrate interactions (Marchler-Bauer et al. 2015). The previously mentioned transcriptome of a thraustochytrid parasite contains genes homologous to ligand-binding proteins such as lectins, integrins and thrombospodin-related proteins (Rubin et al. 2014). With the heterotrophic nature of thraustochytrids, it is possible that secreted proteins may play a role in ligand-binding or protein-binding interactions necessary for cell survival. It is also possible that some extracellular proteins are involved in cell to cell interaction as many thraustochytrids are capable of secreting extracellular polysaccharides (EPS) or extra polysaccharides matrix. This matrix contains sugars, proteins, and lipids and it serves as an energy or moisture reserve to ensure cells' survival under periodic changes in their environment (Jain et al. 2005).

Expression of SP1 and SP2 at the transcript level was quantified to evaluate their expression at different stages of growth. Both are constitutively expressed with relatively higher values (8-20%) under GTY condition in relation to actin (value=1) (Fig. 2). Because of their constitutive expression, these proteins' signal peptide sequence tags may be used in constructing secreted protein recombinant vectors. In yeast, the use of homologous or the host's own signal peptide sequence fused with heterologous proteins can enhance protein

production, indicating preference of certain signal peptide sequences for different organisms. Despite its direct involvement in protein secretion, the exact mechanism of how the signal peptide sequence influences overall expression is yet to be known. Synthetic signal peptide sequences were used to study which amino acids are essential for the functionality of the signal's core region. However, exact interactions of signal peptide tags with their receptor proteins and how these interactions are regulated still remain unsolved (Clérico et al. 2008; Yarimizu et al. 2015). The production of secreted proteins in thraustochytrids has not yet been done. Information about the secretome of these species may pave the way into establishing recombinant expression of extracellular proteins.

Conclusion

In conclusion, two secreted proteins SP1 and SP2, were identified and shown to be constitutively expressed under standard culture conditions. Signal peptide sequences of these two proteins can be used as candidate tags for constructing recombinant extracellular protein expression cassettes.

Part II

Application of the signal peptide in genetic engineering of *Aurantiochytrium* sp. 18W-13a

Introduction

Lignocellulosic biomass, the most abundant and sustainable source of organic carbon, is a promising alternative to petroleum and fossil fuels to produce biofuels and biomaterials. Due to its recalcitrant nature, pretreatment methods that depolymerize the lignocellulose matrix consisting of cellulose, hemicellulose, and lignin is essential to expose the substrates for enzymatic or chemical hydrolysis (Isikgor and Becer 2015).

Consolidated bioprocessing is a strategy that aims to simultaneously conduct enzyme production, enzymatic hydrolysis, and fermentation. This is so far the most promising approach in lignocellulosic biomass conversion to bioethanol or butanol by the use of robust cellulolytic microorganisms (Salehi Jouzani and Taherzadeh 2015). Most strains for the consolidated bioprocessing are either naturally cellulolytic or genetically-engineered to produce hydrolytic enzymes, such as *Clostridium* sp., and genetically engineered cellulase-producing strains of *Saccharomyces cerevisiae* (Salehi Jouzani and Taherzadeh 2015; Liu et al. 2016).

As mentioned in Part I, I aimed to cultivate *Aurantiochytrium* sp. 18W-13a in a media containing low-cost carbon sources such as cellulose for biomass and oil production. Preliminary culture experiments showed the strain's inability to degrade certain polysaccharides such as carboxymethylcellulose for its growth, necessitating genetic improvement of the strain's capability to degrade complex carbon substrates. As the first step in developing a strain capable of degrading cellulose, I attempted to express extracellularly a β-glucosidase from *Aspergillus aculeatus* F-50 (AaBgl), a widely used cellulase for cellobiose breakdown into glucose in the consolidated bioprocessing yeast strains (Fujita et al. 2002; Liu et al. 2016). To secrete the enzyme, I applied a native secretion signal sequence, which I previously identified in a secreted protein (SP2) in *Aurantiochytrium* sp. 18W-13a culture (Juntila et al. 2017). This is the first attempt at

developing a thraustochytrid strain with enhanced cellulase expression towards the goal of utilizing cellulosic biomass as its carbon source.

Materials and Methods

2.1. Reagents and chemicals

All reagents and chemicals used were either analytical or HPLC grade. Bacto-peptone, Bacto-yeast extract (YE) and Bacto-agar are purchased from Nippon Becton Dickinson (Tokyo, Japan). Marine Art SF-1 (produced by Tomita Pharmaceutical, Tokushima, Japan and provided by Osaka Yakken, Osaka, Japan) was used to prepare the artificial seawater. Most reagents such as glucose, sodium acetate, trisodium citrate, sodium carbonate, citric acid, and G418 sulfate are from Fujifilm Wako Pure Chemicals (Osaka, Japan). Substrates cellobiose, 4-nitrophenyl-β-D-glucopyranoside (pNPG) and 4-methyllumbeferyl-β-D-glucopyranoside (4-MUG) were purchased from Tokyo Chemical Industry (Tokyo, Japan).

2.2. Strain and culture conditions

Cultivation of *Aurantiochytrium* sp. 18W-13a was done using a basal PY medium (0.15% (w/v) peptone, 0.01% (w/v) YE, 50% (v/v) artificial seawater) supplemented with either 1% glucose or 1% cellobiose as a carbon source at 30°C with shaking at 100 rpm. To measure cell growth, the optical density at 660 nm (OD_{660}) was monitored by a spectrophotometer, UV-1700 PharmaSpec (Shimadzu, Kyoto, Japan).

2.3 Construction of pEF-Neo^r-Ubi-AaBgl Expression vector

Vectors pUbi-Neo^r and pEF-Neo^r, which contain the neomycin resistance gene cassette with the *Thraustochytrium aureum* ubiquitin or the EF-1α promoter and terminator regions, respectively, were kindly provided by Prof. M. Ito from Kyushu University and were constructed as previously reported by Sakaguchi et al*.,* (2012*).* The coding sequence for β-glucosidase from *Aspergillus aculeatus* (*AaBgl*) (Genbank Accession no. D64088; Kawaguchi et al., 1996) was codon optimized based on codon context usage and GC content using the Codon Optimization *On-Line* (CO*OL*) software (Chin et al. 2014). I used calculated values for the codon context usage generated from a local transcriptome database of *Aurantiochytrium* sp. 18W-13a. The original signal peptide of the β-glucosidase was replaced by a native signal peptide (SP) that I previously studied in a constitutively secreted protein SP2 (Juntila et al. 2017) (Fig. 4). Synthetic DNA plasmid with the complete coding sequence was synthesized by Eurofins genomics (Tokyo, Japan). This plasmid, pTAC-2-β-glucosidase, was used as a template for PCR (polymerase chain reaction) amplification of the SP:*AaBgl* coding sequence for infusion ligation (In-Fusion HD Cloning Kit, TakaraBio, Ohtsu, Japan) with the PCR linearized pUbi-Neo^r to replace the Neo^r fragment with the *AaBgl* gene. The resulting plasmid, pUbi-AaBgl, was used as a template for amplification of the Ubi-AaBgl expression cassette for infusion cloning to *Spe*I-linearized pEF-Neo^r to generate the final vector plasmid, pEF-Neo^r-Ubi-AaBgl, which contains the expression cassettes for *Neo^r* and *AaBgl* (Fig. 4). All primers that were used for infusion cloning and PCR are listed in Supplementary Table 2.

2.4. Transformation of Aurantiochytrium sp. 18W-13a

Transformation was done based on the electroporation protocol described by Sakaguchi et al., 2012. Cells were grown in GY medium (0.64% (w/v) glucose, 0.22% (w/v) yeast extract, 50% (v/v) artificial seawater) for 2 d and the cells in the exponential stage (5 x 10^7 cells/500 µL) were harvested. Glass beads (250 mg per 500 µL) were added to the suspension and gently vortexed for 5 sec. After which, the suspension was spun down using a benchtop centrifuge for 15 sec to separate the beads from the cell suspension. Then, the suspension was centrifuged to collect the cells. A single wash with sterile 50% (v/v) artificial seawater was done to remove medium components. After centrifuging and ensuring complete removal of seawater, the cell pellet was resuspended in 80 µL OptiMEM Buffer (Thermo Fisher Scientific, Tokyo, Japan) with 1 µg of pEF-Neo^r-Ubi-Abgl plasmid DNA linearized with *Nde*I. Electroporation was performed by GenePulser Xcell (Bio-Rad, Hercules, CA) under the following conditions: 2 pulses of 750 V, 25 µF, 200 Ω, 1 mm cuvette. All steps until electroporation were done in ice. Immediately after electroporating cells, 1 mL room-temperature GY medium was added to the cuvette for recovery. The cell suspension was transferred into a 1.5 mL microcentrifuge tube and incubated overnight at 20°C with shaking at 100 rpm. All cells were then plated on 1.5% (w/v) agar-solidified GY medium supplemented with 250 µg/mL G418 sulfate. Possible transformants that appeared were streaked again on 1.5% agar-solidified GY medium supplemented with 500 µg/mL G418 sulfate to confirm their G418 resistance. Then, these transformants were cultured in GY medium, and the cells were collected for extraction of the genomic DNA and subsequent PCR of the *AaBgl* and *Neo^r* insertion into the chromosomes. The culture supernatants of transformants were screened for β-glucosidase (BGL) activity using *p*-NPG as a substrate. Reactions were carried out at 200 mM *p*-NPG

in sodium acetate buffer pH 5 at 50 °C for 30 min. After which, 100 mM sodium carbonate was added before measuring colorimetrically the absorbance at 410 nm (UV-1700). Southern blotting was done to determine the number of transgene insertions. Genomic DNA of wild-type and AaBgl+ strains (30 µg each) were digested with *Hind*III and *Pst*I with overnight incubation before electrophoresis separation in 0.8% agarose gel. Gel preparation and capillary blot transfer to Hybond N+ nylon membrane (GE Healthcare Japan, Tokyo, Japan) were done according to the manufacturer's instructions. The DNA was UV crosslinked into the membrane under the optimal settings. After which, a 600 bp DNA fragment corresponding to a part of the coding sequence of AaBgl amplified by PCR was prepared as a probe and labeled for hybridization using the Amersham Gene Images AlkPhos Direct Labelling and Detection system (GE Healthcare Japan). Hybridization and post washes were done according to the kit's instructions. Amersham CDP-Star Detection Reagent (GE Healthcare Japan) was used for chemifluorescence detection and the bands were viewed using AE-9300H Ez-Capture MG (ATTO, Tokyo, Japan). Out of several transformants, one transformant designated as the AaBgl^{+} showing the highest activity with single copy gene insertion was selected for further experiments (Supplementary Fig. 2).

2.5. ApiZYM extracellular enzyme assay

Cells of wild-type and AaBgl^{+} strains were grown overnight in PY medium with 1% glucose before harvesting for ApiZYM strip enzyme assay (bioMérieux, Marcy l'Etoile, France). Cultures were centrifuged at 6,000 x *g* for 10 min and the cell pellet was washed with sterile 50% artificial seawater. Final suspension is at 4.0 x 10^{5} cells per 65 µL sterile 50% artificial seawater which was inoculated into the microcupules of the ApiZYM strip (Gupta et al. 2016). After incubation at 30°C overnight, one drop of ApiZYM A and

B reagents were added into the microcupules and incubated for 15 min. Change and intensity of the developed color were checked for semi-quantitative evaluation of enzyme activities. Preliminary assays were also done using the same method for representative *Aurantiochytrium* species, *Aurantiochytrium* sp. 18W-13a, *Aurantiochytrium mangrovei* NYH1 and *Aurantiochytrium limacinum* SR21, with cells from both exponential and stationary stages screened for extracellular enzyme activity (Supplementary Table 3).

2.6. Growth and p-NPG enzyme assay using cellobiose as a carbon source

From glucose-supplemented PY cultures, the wild-type and AaBgl^{+} strains were precultured with 1% cellobiose PY medium for 2 d before the final cultivation experiment in 1% cellobiose PY medium for 5 d. Each subculture was initiated at a starting OD_{660} of 0.5. To determine β-glucosidase activity, 2 mL culture supernatants were collected by centrifugation of cultures at 6000 x *g*, 20 °C for 10 min, ensuring complete separation of cells. Then, the supernatants were concentrated using Amicon Ultra 4 centrifugal filter units (Merck, Darmstadt, Germany). Buffer exchange using 50 mM sodium phosphate pH 7 was done twice to remove residual salts. Approximately 60 µL of extracellular protein extract was collected and the total protein concentration was measured using DC Protein Assay (Bio-Rad) and Bio-rad Protein Assay (Bio-Rad) according to manufacturer's instructions with bovine serum albumin as a standard. A total of 10 µg protein extract is suspended in 200 mM *p*-NPG in 50 mM sodium acetate buffer pH 5 for incubation at 30°C, 10 min for enzymatic assay. Then, 100 mM sodium carbonate is added before measuring theabsorbance at 410 nm. Normalized enzyme activity is calculated by µmol *p*-nitrophenol (PNP) released per min per µg protein.

2.6. Native PAGE zymogram

For Native PAGE, extracellular protein extracts were analyzed in 10-20% gradient polyacrylamide precast e-PAGEL gel (E-T1020L, ATTO), run at constant 110 V at 4 °C for 2-3 hours. The gel was then rinsed twice with sterile water and 100 mM sodium citrate buffer, pH 5, consecutively. Zymogram method consists of incubating the washed gel with 5 mM 4-MUG in 100 mM sodium citrate buffer, pH 5 at 50 °C for 30 min before observing accumulation of the fluorescent 4-MU under the excitation of UV light. The gel is then rinsed with water for 3 times before Coomassie Blue staining (Bio-Rad) for visualization of the protein bands.

2.7 β-glucosidase activity assay of soluble and insoluble cellular extracts and extracellular extract from the supernatant

Cellular extracts were isolated using sonication with the soluble and insoluble protein fractions separated by centrifugation. To insure complete removal of cell debris that may contribute insoluble proteins, the soluble fraction was filtered using a 0.20 µM filter. Extracellular fraction was obtained by ultracentrifugation using Amicon Ultra-4 centrifugal filter units as previously describe in section 2.6. Total protein concentration was measured using DC Protein Assay (Bio-Rad). . A total of 20 µg protein extract is suspended in 200 mM *p*-NPG in 50 mM sodium acetate buffer pH 5 for incubation at 50°C, 15 min for enzymatic assay. Then, 100 mM sodium carbonate is added before measuring the absorbance at 410 nm. Normalized enzyme activity is calculated by µmol *p*-nitrophenol (PNP) released per min per µg protein.

2.8 Biomass and lipid analysis

The cells were grown in the standard medium GTY (2% glucose (w/v), 1% tryptone (w/v) and 0.5% yeast extract (w/v), 50% artificial seawater (v/v)) for 6 days. The AaBgl+ transformant was additionally cultivated in 2% cellobiose instead of glucose and the medium buffered with 50 mM MES, with initial pH 5.8, to aid in enzyme activity. The cells harvested by centrifugation to remove as much medium as possible before drying using a centrifugal vacuum concentrator. Gravimetric measurement of the dry weights was done for calculation of the total biomass yield (g L^{-1}) and for lipid extraction. Dried cells were suspended in 2:1 chloroform:methanol solution with 1 mg squalane and 1 mg tricosanoic acid (C23:0) as the internal standards and incubated overnight. After which, the suspension was filtered to remove cell debris and the resulting filtrate was mixed with 0.9% (w/v) NaCl solution. After agitating the mixture, it was allowed to settle and separate into aqueous and organic phases at room temperature. Lower organic phase was transferred into two pre-weighed new tubes and vacuum evaporated under 370Pa vacuum. After complete evaporation of the solvent, the tubes were weighed again to determine the total lipid weight. One tube is used for squalene analysis wherein hexane was added to dissolve the lipid. The other tube was used for fatty acid analysis and the lipid was dissolved in 0.1 M HCl-methanol solution and heated at 100 °C for esterification for 1 h. After which, addition of hexane-water was done for phase separation. The upper hexane-extract layer was collected, and hexane was evaporated using a centrifugal concentrator. Residual fatty acid methyl esters were dissolved in hexane for GC analysis. Both squalene and fatty acid methyl esters were analyzed in GC-2014 (Shimadzu, Kyoto, Japan) equipped with a CP-Sil5 CB column (Agilent Technologies, Santa Clara, CA) at starting column temperature of 60 °C, then 20 °C/min until 130 °C and a further increase to 270 °C at 4 °C/min.

Results and Discussion

3.1. Generation of the β-glucosidase expressing strain, AaBgl^{+}

Several transformants with positive β-glucosidase activity were isolated after combined glass bead treatment and electroporation protocol (Supplementary Fig. 2). Initial screening of these transformants with confirmed genomic insertions of the *AaBgl* transgene displayed variable β-glucosidase activity. The transformant with the highest activity and single copy insertion was designated as the AaBgl^{+} transformant and was used for further experiments. Transgenes were amplified with the expected bands of 2630 bp and 795 bp for *AaBgl* and *Neo*^{r}, respectively (Fig. 5a, 5b). The AaBgl+ strain can grow with neomycin resistance of up to 1 mgmL^{-1} G418 antibiotic concentration. Very few colonies of the wild-type strain can grow at 100 µg mL^{-1} G418 at high cell concentration (Fig. 5d).

The apiZYM enzyme kit was previously used to semi-quantitatively evaluate the expression of extracellular digestive enzymes in thraustochytrid species. A wide range of enzymatic activities were observed including alkaline phosphatase, protein arylamidases, lipases, napthol-AS-BI-phosphohydrolase and polysaccharide-degrading enzymes such as β-galactosidase and N-acteyl-β-glucosaminidase (Gupta et al. 2016). As for the strain 18W-13a, it exhibited moderate activities of esterase (C4), lipase (C5), leucine and valine arylamidases and high activities of acid phosphatase and napthol-AS-BI-phosphohydrolase, relatively similar to other species *A. mangrovei* NYH1 and *A. limacinum* SR21. However, no prominent carbohydrate-degrading activities were observed (Supplementary Table 3). This supports the previous statement that 18W-13a cannot utilize complex carbohydrates as their C source under standard laboratory conditions (Juntila et al. 2017). The AaBgl^{+} strain

exhibited strong β-glucosidase activity compared to the wild-type strain, as expected due to its constitutive expression under the ubiquitin promoter (Fig. 5c).

Growth assays of the wild-type and AaBgl^{+} strains exhibited similar trends under the basal medium supplemented with 1% glucose as C source. The introduction of the *AaBgl* transgene has no negative effect on the growth of the transformant (Fig. 5e). The pH in the glucose containing media were dropped from 6.8 to around 5.6 in 1 d and then remained at around 5.6 during further cultivation up to 4 d (data not shown). This might be caused by secretion of organic acid wastes from proliferating cells and is observed when cells utilize readily available glucose. Thraustochytrids are reported to tolerate a wide range of pH values for their growth (Raghukumar 2008). At pH 5.6, *Aurantiochytrium* sp. 18W-13a continues to grow, suggesting its tolerance of this pH level. This pH tolerance is beneficial to the aim of degrading cellulose as cellulases are optimal at low pH of 4-5. Also, low pH values may hinder the growth of possible contaminants, especially at large-scale production.

Several methods such as electroporation, particle bombardment and agrobacterium-mediated transformation have been applied in thraustochytrids' transformation, particularly for establishing proof-of-concept reporter gene expression (Cheng et al. 2012; Sakaguchi et al. 2012; Okino et al. 2018). Despite successful transformation results, the applicability of certain methods is highly species-specific and there are cases of reproducibility concerns due to minute changes in the protocols (Adachi et al. 2017). Similarly, I also experienced such difficulties in using standard electroporation method for the 18W-13a strain. The addition of a glass-bead treatment that may partially destroy the cell wall for higher chance of DNA uptake, as suggested by Adachi et al., 2017, led to obtain transformants while I did not get any transformant for electroporation only. The cell wall of thraustochytrids are non-cellulosic and are made up of circular scales of sulfated polysaccharides. The composition

varies under different stages of growth with the zoospores and vegetative cells lacking the sulphated polysaccharides (Fossier Marchan et al. 2017). It is possible that due to thraustochytrids having diverse life cycle stages with changing cell forms and possible changes in cell wall morphologies, the reproducibility and success of transformation are highly variable. As for *Aurantiochytrium* species, cells are thin-walled and are dispersed as single cells. In some cases, an amoeboid stage which is similar to naked protoplasts is observed and these cells may be more susceptible for DNA uptake (Fossier Marchan et al. 2017). Sulphated polysaccharides can possibly hinder DNA interactions due to their negative charge. And the use of glass-bead treatment can disrupt this negatively charged matrix, allowing higher chance of DNA uptake.

3.2. Growth of AaBgl+ in cellobiose as a carbon source

As the first step in cellulose degradation, I tested the growth of the AaBgl+ strain under 1% cellobiose as a sole carbon source. I used a basal-medium with low concentrations of initial peptone and yeast extract to minimize the effect of these components to the growth, with complete dependence on the carbon source as the main nutrient.

Cultures previously grown at 1% cellobiose-supplemented basal medium for 2 days were sub-cultured into the fresh 1% cellobiose medium. A starting optical density of ~0.3-0.4 and initial enzyme activity of 0.134 mmol PNP min^{-1} μg^{-1} protein are enough for the initial growth and effective conversion of cellobiose to glucose, supporting cell growth during cultivation (Fig. 6). Despite the suboptimal conditions of 30°C and initial pH of 6.4-

6.8 for enzyme activity, the AaBgl$^+$ strain continued to grow with more cells expressing the enzyme and the lowering of the medium pH to 5.6 supporting enzymatic activity.

The wild-type strain did not grow and has no BGL activity while the AaBgl$^+$ strain showed a significant increase in growth, reaching the early stationary phase at d 4. A similar trend was observed for β-glucosidase activity (Fig. 6). This is possibly due to oversaturation of the enzyme, as more cellobiose is converted to glucose. Due to the low protein nutrient and nitrogen sources, excess unutilized glucose may have accumulated in the medium which explains the plateauing of both growth and enzyme activity at the later stage of cultivation. These results confirm that the AaBgl$^+$ transformantcan successfully express functional β-glucosidases which support its growth in cellobiose. This is the first report of a fungal enzyme expression, with targeted secretion, in a thraustochytrid species for biotechnological and cultivation purposes.

Aspergillus aculeatus F-50 is a highly cellulolytic fungus whose β-glucosidase is potently active in degrading cellooligosaccharides ranging from cellobiose to cellohexaose and can even degrade insoluble cellooligosaccharides with a degree of polymerization of 20. Optimal activity is achieved under pH 4.0-4.5 and 50°C (Sakamoto et al. 1985a; Kawaguchi et al. 1996). At 30°C, reported relative activity is at 20% of the optimal activity. With the high affinity of the enzyme to cellooligosaccharides, the relative activity does not decrease so much until pH 6. However, there is a sudden decrease to less than 20% activity for pH 7 and inactivity above pH 8 (Sakamoto et al. 1985a). Finding the ideal microorganism for consolidated bioprocessing, which should be capable of high enzyme expression and thermotolerance, has been one of the bottlenecks in lignocellulosic biomass conversion. Even in using yeast, most of the systems are still limited to 30°C temperature (Fitzpatrick et al. 2014). Despite the sub-optimal temperature conditions used in this study, β-glucosidase activity at d 0 was still enough to convert cellobiose to glucose for growth.

With glucose metabolism and possible release of organic acid wastes, the pH lowering to 5.6 further supported β-glucosidase activity which supported latter stages of growth.

Native PAGE analysis of the extracellular protein fraction showed more bands in the AaBgl+ strain than the wild-type strain. This could be due to expression of constitutive extracellular proteins with growth and is not evident in the wild-type cells because of their state of growth inhibition. Notably, the differentially expressed band above the 150 kDa mark showed positive β-glucosidase activity (Fig. 7). Takada et al. 1998 previously reported a gel chromatography purified band of 180 kDa for the AaBgl enzyme expressed in *S. cerevisiae*. However, gel filtration of crude cellulases from *A. aculeatus* F-50 reported the molecular weight of AaBgl as 133 kDa (Sakamoto et al. 1985b). Also, expression in *Aspergillus niger* (Baba et al. 2015) and *Trichoderma reesei* (Nakazawa et al. 2012b) yielded relatively the same mass of 130 kDa. This suggests that the mass of the protein is variable depending on the glycosylation mechanism present in the host system. In my case, the mass of the active enzyme is possibly higher than 150 kDa, with *Aurantiochytrium* having a possibly closer glycosylation mechanism to yeast. However, this is only an estimate as accurate masses cannot be determined through native PAGE analysis. I also analyzed the enzyme activity using different fractions which include cellular fractions of soluble and insoluble proteins against the extracellular protein fraction in the supernatant. Only the extracellular fraction has distinctively high β-glucosidase activity, which indicates that most of the functional enzyme is indeed secreted outside the cell (Fig. 8).

Targeted secretion of the enzyme was done using a native signal peptide sequence tag. No consensus has been found in signal peptides besides similar motifs of N-terminal basic residue and hydrophobic cores. Substituting with the host's own endogenous signal peptides for those of foreign proteins resulted in enhanced heterologous expression especially in yeast. Some studies use synthetic signal peptide sequences to determine

essential amino acids and appropriate lengths for the signal's core regions. However, how the signal sequence affects expression levels still remain unsolved (Yarimizu et al. 2015).

Due to their capability of degrading diverse substrates, the growth of thraustochytrids in complex organic sources has been studied in some native isolates. Cellulase activity against carboxymethylcellulose was previously reported in thraustochytrid genera *Aplanochytrium, Botryochytrium, Oblongichitrium, Parietychytrium, Schizochytrium, Sicyoidochytrium, Thraustochytrium,* and *Ulkenia*, but not in *Aurantiochytrium* (Taoka et al. 2009). *Schizochytrium* DT3 can grow under hemp sugar hydrolysate containing mixed reduced sugars, suggesting that sugar hydrolysate from lignocellulosic biomass can be used for this strain (Gupta et al. 2015). Pretreatment of lignocellulose for hydrolysis into sugars also produces toxic inhibitors such as weak acids and aromatics which necessitates the development of lignocellulosic hydrolysate-tolerant microbial strains such as *Aurantiochytrium* sp. FN21, which was domesticated to grow in sugarcane bagasse hydrolysate (Qi et al. 2017). Most reports use hydrolysates which already contain easily accessible sugars. Only the strain *Aurantiochytrium* sp. KRS101 was reported to exhibit carboxymethylcellulase and cellobiohydrolase activities in both its cell-free lysate and supernatant fractions, and can use cellulosic materials such as cellobiose, carboxymethylcellulose, and pre-treated palm oil empty fruit brunch (Hong et al. 2012). Even with existing native strains that can utilize cellulose, strain improvement through genetic engineering for enhanced enzyme production may be necessary to fully degrade recalcitrant lignocellulosic substrates and utilize the different sugars produced. For example, the commercial strain T18 for DHA production, has the enzymes xylose reductase and xylose isomerase for xylose utilization. However, xylose utilization of the wild-type strain only occurs with additional glucose. Increasing the copy number of xylose isomerase and heterologous xylulose kinase expression increased the xylose usage and

reduced the intermediate product, xylitol (Merkx-Jacques et al. 2018). Identification of cellulase genes and how to induce them are poorly studied in thraustochytrids. In my case, I haven't found a suitable substrate to induce any cellulase activity in 18W-13a and this is the main reason why I tried to overexpress a heterologous cellulase in this strain.

3.3. Biomass and lipid production of AaBgl+ strain

Aurantiochytrium species can accumulate fatty acids such as palmitic acid, DHA, DPA and EPA which are stored in forms of triacylglycerols (TAGs) (Kobayashi et al. 2011; Ishitsuka et al. 2016). They can also produce high amounts of squalene compared to other thraustochytrid species (Nakazawa et al. 2014). I analyzed the biomass and lipid yields, lipid content, squalene content and DHA content between the wild-type and AaBgl+ strains under glucose and cellobiose as a carbon source (Table 2).

Accumulation of squalene has also been observed in thraustochytrids especially in certain strains of the genus *Aurantiochytrium,* which is also assumed to be directly correlated to its higher carotenoid production (Nakazawa et al. 2014). As stated in part I, the strain *Aurantiochytrium* sp. 18W-13a was extensively studied due to its high production of squalene which can reach 20% dry cell weight under optimal conditions (Kaya et al. 2011; Nakazawa et al. 2012a; Ishitsuka et al. 2016). Interestingly, using coherent anti-stokes Raman scattering spectroscopy (CARS) imaging to visualize and distinguish intracellular lipid forms, there is a clear difference in the detection peaks of squalene and TAGs in 18W-13a. This suggests that squalene and TAGs do not co-localize despite that both are hydrophobic lipids. In particular, squalene is stored in round pore-like vacuole-like organelles, possibly as temporary energy or carbon source (Ishitsuka et al. 2016). This strain can also produce DHA and DPA (docosapentaenoic acid; $C_{22:5}$), at 14-34% and 7-19% of its triacylglycerol (TAG) content, respectively (Matsuura et al. 2012).

Squalene contents findings are lower than previous studies. This is most likely due to the suboptimal conditions used. I analyzed the lipids at d6, early stationary phase, to give an initial comparison between the wild-type and AaBgl^{+} strains. It is highly possible that squalene accumulation comes later during cultivation and is induced by several factors including aeration and available carbon and nitrogen nutrients present in the medium.

Aside from producing squalene, both strains can produce myristic (C14:0), pentadecanoic (C15:0), palmitic (C16:0), heptadecanoic (C17:0), docosapentanoic (DPA/C22:5) and DHA (C22:6) fatty acids, in agreement with previous reports (Matsuura et al. 2012; Tani et al. 2018). Interestingly, DHA contents are at a slightly higher range than squalene contents. Again, both squalene and DHA contents can be improved with modifications in growth conditions.

Conclusion

The transformant strain, AaBgl^{+}, was obtained after transformation using a combined glass bead treatment and electroporation. As expected, transgenes of the selective marker NeoR (neomycin resistance gene) and *AaBgl* gene were present in the transformant, with neomycin resistance of up to 1 mg/mL G418 antibiotic. Only the AaBgl^{+} strain could grow under cellobiose as the C source with increasing growth and enzyme activity during cultivation. Native PAGE analysis of the extracellular proteins showed the presence of a differentially expressed band slightly above 150 kDa which has β-glucosidase activity. With these results, the application of SP2's signal peptide sequence for secretion of AaBgl is successful, in conclusion, to part II. Also, biomass and lipid analysis showed relatively same yields. Both the wild-type and AaBgl^{+} strains can produce both squalene and DHA at relatively the same range, which can be further improved based on culture conditions. This strain is a promising source of both valuable oils.

General Discussion

A promising characteristic of thraustochytrids is their capability to produce extracellular enzymes. It is of great interest to study their mode of nutrition, survival, and competition against other microorganisms in the diverse ecosystems. It is possible that extracellular enzyme production is strain-specific and may be possibly induced by different substrates. This further necessitates research about the secreted proteins and hydrolytic enzymes of thraustochytrids. My work with the secreted proteins of 18W-13 showed the constitutive expression of several proteins under standard laboratory conditions. Two proteins, SP1 and SP2, were further analyzed for their conserved regions and signal peptides. Particularly, SP2 contains vWF and PAN/APPLE-like membranes which may be involved in cell to cell interaction. With this, I decided to use SP2 signal peptide sequence for the targeted expression of the heterologous cellulase, AaBgl, to generate transformants that can utilize cellobiose as the first step in cellulosic biomass utilization.

The enzyme AaBgl, *Aspergillus aculeatus* β-glucosidase, is widely-used for consolidated bioprocessing yeast strains for bioethanol production. Similarly, I aimed to generate a consolidate bioprocessing strain of 18W-13a for conversion of cellulosic waste into biomass and valuable oils.

Thraustochytrids is is a promising expression platform for enzymes and valuable proteins. Targeted expression has also been done in *Aurantiochytrium limacinum* ATCC MYA-1381 for EGFP expression in the mitochondria and endoplasmic reticulum using organelle-specific targeting and/or retaining signals (Okino et al. 2018). Aside from reporter genes used in testing transformation protocols, several studies have been conducted to enhance the fatty acid production of *Aurantiochytrium* species. Suen et al. 2014 reported the

expression of *Vitreoscila stercoraria* hemoglobin (VHb), which can increase dissolved oxygen levels intracellularly by effective uptake of oxygen, in *Aurantiochytrium* sp. MP4 and SK4 after electroporation transformation. Thraustochytrids are aerobic organisms and their growth is greatly affected by dissolved oxygen levels. Both strains show increased biomass with VHb with the MP4 strain producing 44% higher total fatty acids and 9-fold astaxanthin contents than the wild-type strain. Another study increased the amount of EPA production in *A. limacinum* mh0186 by overexpression of the *Thraustochytrium auream* Δ5 desaturase, which could efficiently convert existing or added eicosatetraenoic acid (ETA) to EPA (Kobayashi et al. 2011).

In my work, I was able to a generate a transformant, AaBgl+, that could express and secrete AaBgl, under the influence of its own local or endogenous signal peptide. The secreted AaBgl enzyme was functionally active and can support the growth of the strain under cellobiose as the sole carbon source. This work is one of the first attempts in secreting a functional enzyme in thraustochytrids. Moreover, it is the first step towards the aim of establishing a cellulose-degrading strain as a potential platform for lignocellulosic biomass conversion into valuable products such as squalene and DHA.

Tables and Figures

Table 1. Description of secreted proteins, SP1 and SP2, based on NCBI conserved domain search, Blastp and SignalP 4.1 analyses.

						NCBI Conserved Domain Search	
Name	Length (a.a)	Mass (kDa)	Mascot Score	Seq. Coverage	SignalP 4.1	Accession no.	Name
SP1 (LPD8921)	2353	251.42	546	7%	Secreted	--	---
SP2 (LPD1644)	1020	110.17	316	17%	Secreted	smart00327	von Willebrand factor (vWF) type A domain (61-166/655-786)
						cl00112	PAN/APPLE-like domain (378-426)

(Continued)

Name	Blastp Analysis		
	Protein ID	Score	Annotation
SP1 (LPD8921)	Local protein database		
	LPD8920	7307	---
	A.limacinum ATCC MYA-1381 (JGI)		
	Aurli1.3202	8974	subtilase
	Aurli1.3204	8883	subtilase
SP2 (LPD1644)	Local protein database		
	LPD9373	776	---
	A.limacinum ATCC MYA-1381 (JGI)		
	not detected		

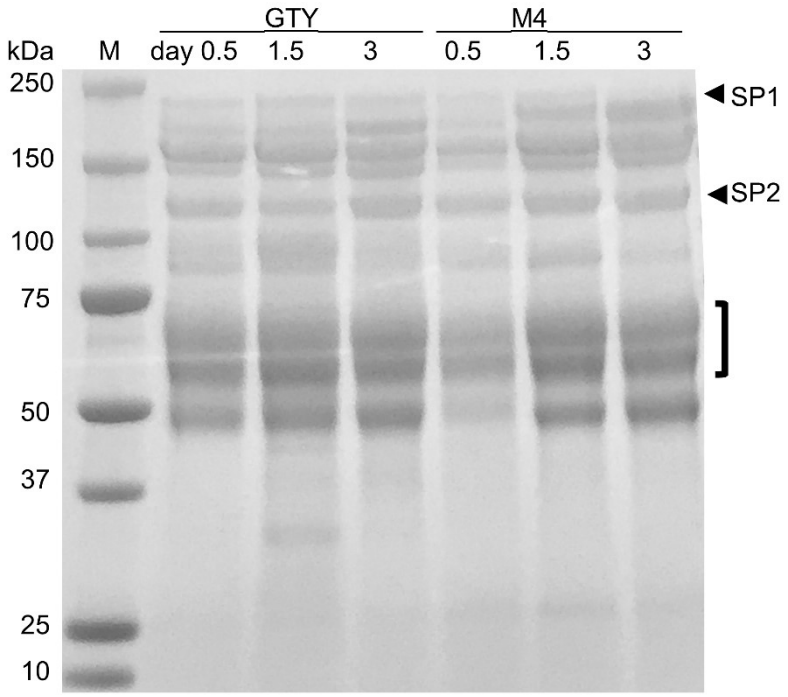


Figure 1. **SDS-PAGE analysis of extracellular proteins expressed under GTY and M4 culture conditions collected at different time points day 0.5, 1.5 and 3.**

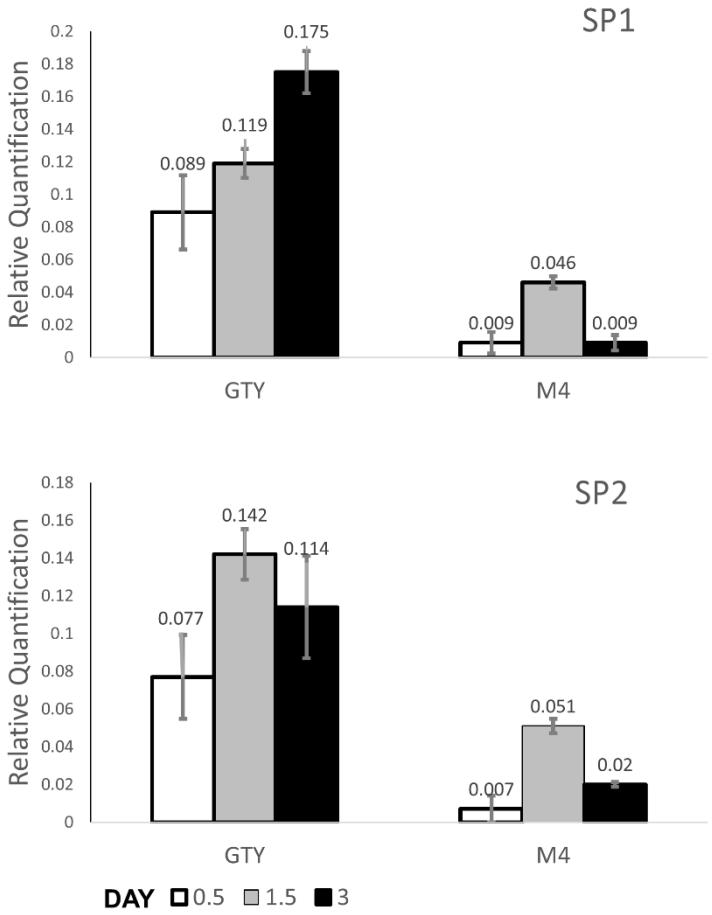


Figure 2. **Relative expression of the *sp1* and *sp2* genes against the gene for actin (value=1.0) under GTY and M4 culture conditions at different time points day 0.5, 1.5 and 3**. Error bars represent standard deviation (n=3).

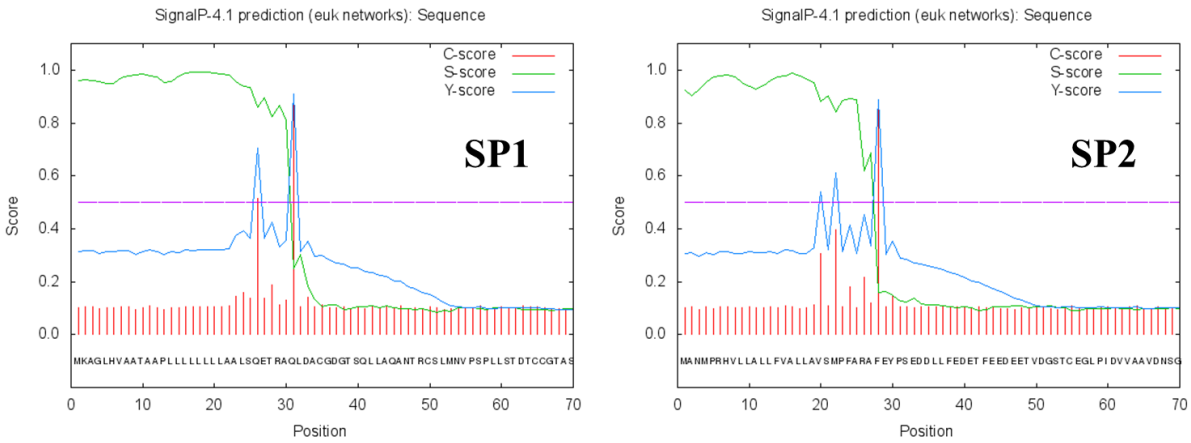


Figure 3. **SignalP 4.1 analysis of the N-terminal sequences (showing first 70 amino acids) of SP1 and SP2.** C-score is the raw cleavage site score used to distinguish the cleavage sites. S-score is the signal peptide score used to distinguish signal peptide regions apart from mature protein regions and proteins without signal peptides. Y-score is a combined geometric average of the C-score and the slope of the S-score which can predict better cleavage site than C-score alone.

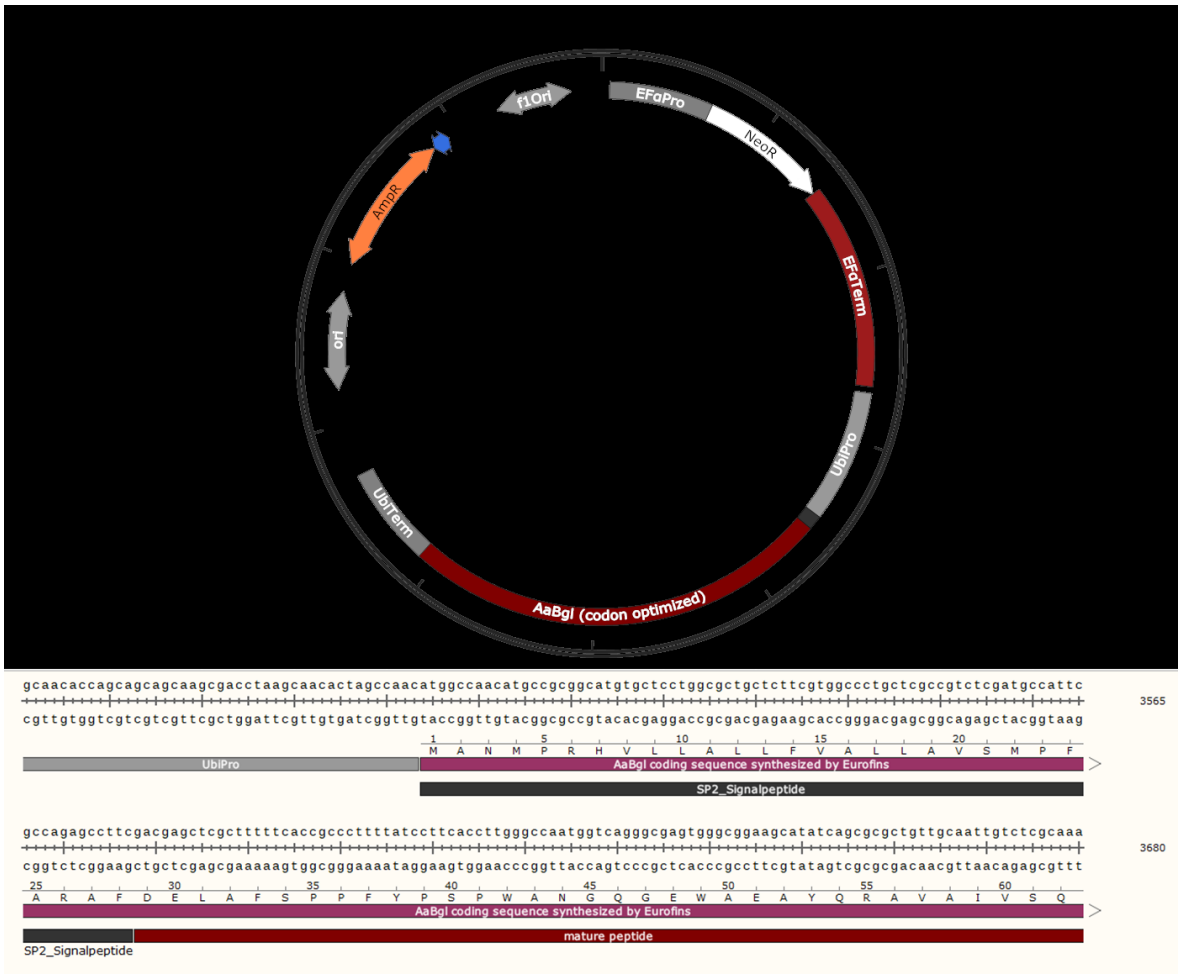


Figure 4. **Vector map of pEF-Neo^r-Ubi-AaBgl showing distinctive features such as promoters (EFαPro and UbiPro), terminators (EFαTerm and UbiTerm) and coding sequences for Neo^r and *AaBgl***. Under the vector map is the N-terminal sequence of the AaBgl coding sequence with designated features for the signal peptide of SP2 and the mature peptide of AaBgl.

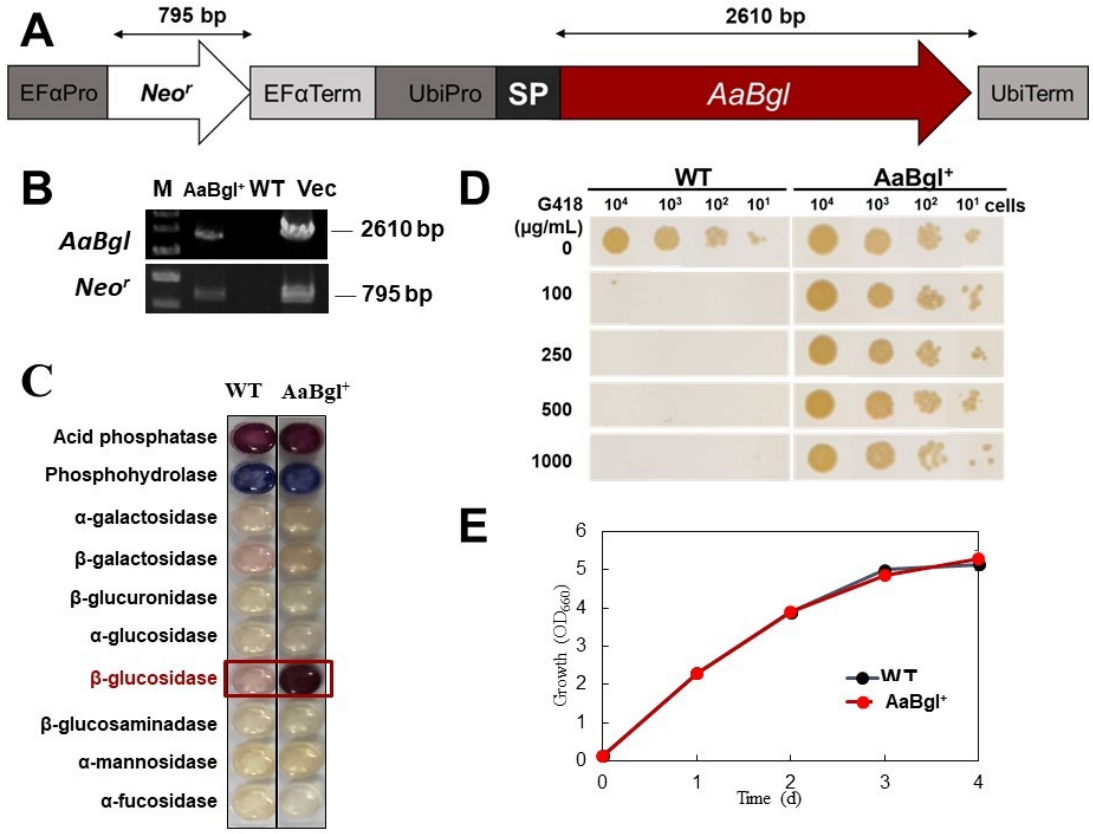


Figure 5. **Generation of AaBgl+ transformant with high expression of *Neo*^{r} and *AaBgl* genes.** A) Expression vector for neomycin resistance (*Neo*^{r}) and *A. aculeatus* beta-glucosidase (*AaBgl*) genes. Strain 18W-13a's own signal peptide (SP) was used as a tag for extracellular expression of *AaBgl*. B) PCR detection of transgenes in wild-type and AaBgl+ strains. C) ApiZYM enzymatic assay kit showing positive acid phosphatase and phosphohydrolase in both strains and high activity of β-glucosidase in AaBgl+ transformant. D) G418 sensitivity assay of WT and AaBgl+ strains. E) Growth (OD_{660}) of WT and AaBgl+ strains under basal PY medium supplemented with 1% glucose as C source.

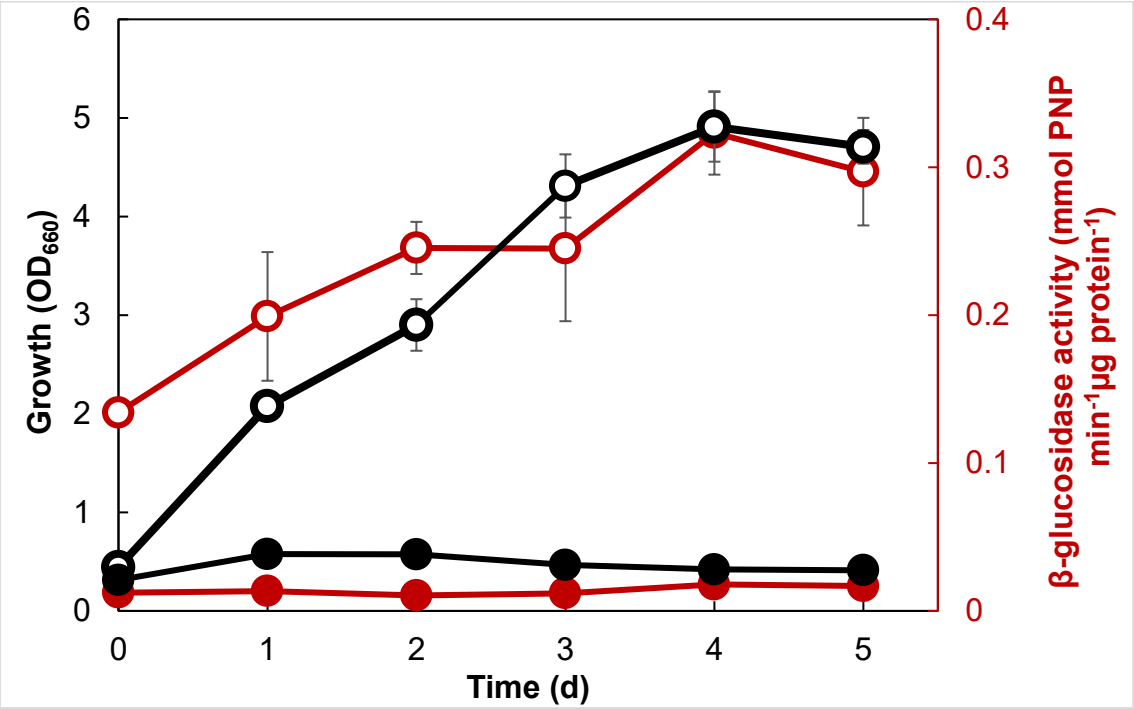


Figure 6. **Growth and β-glucosidase (BGL) activity of wild type (●) and AaBgl+ (○) strains in basal PY medium supplemented with 1% cellobiose as C source.** Black and red colors indicate growth and β-glucosidase activity, respectively.

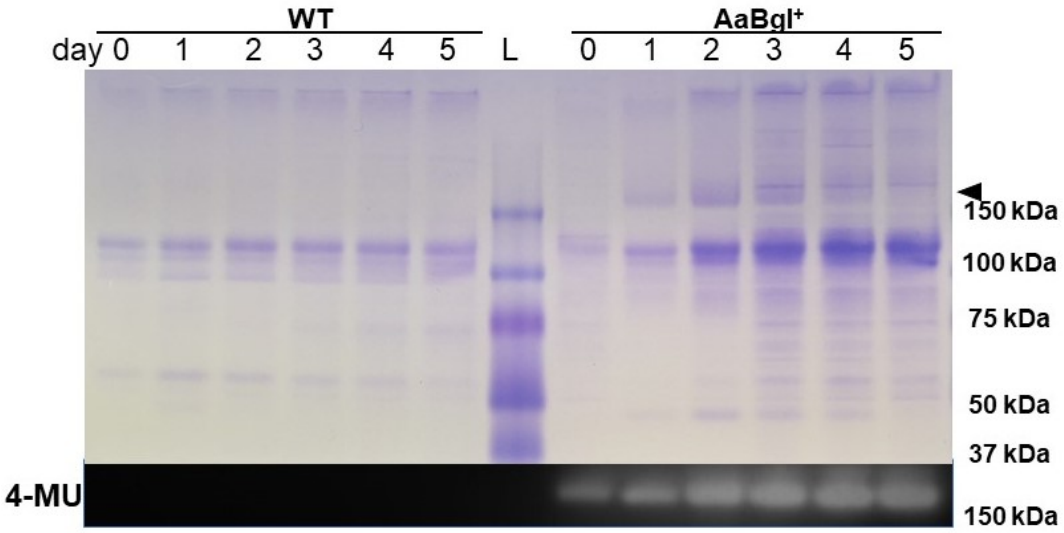


Figure 7**. Native PAGE Zymogram of extracellular proteins in the supernatant of wild-type and AaBgl^{+} strains**. A differentially expressed band (arrow) found above the 150 kDa mark shows positive fluorescence of 4-MU (4-methylumbelliferone) as the product of β-glucosidase activity.

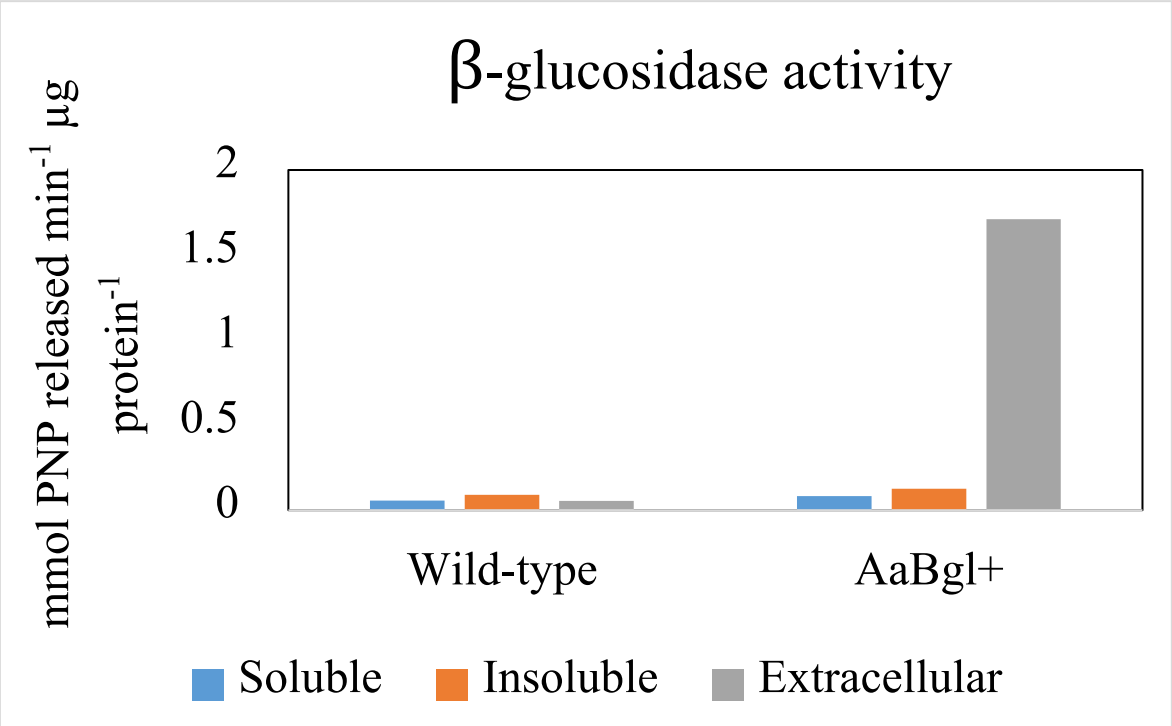
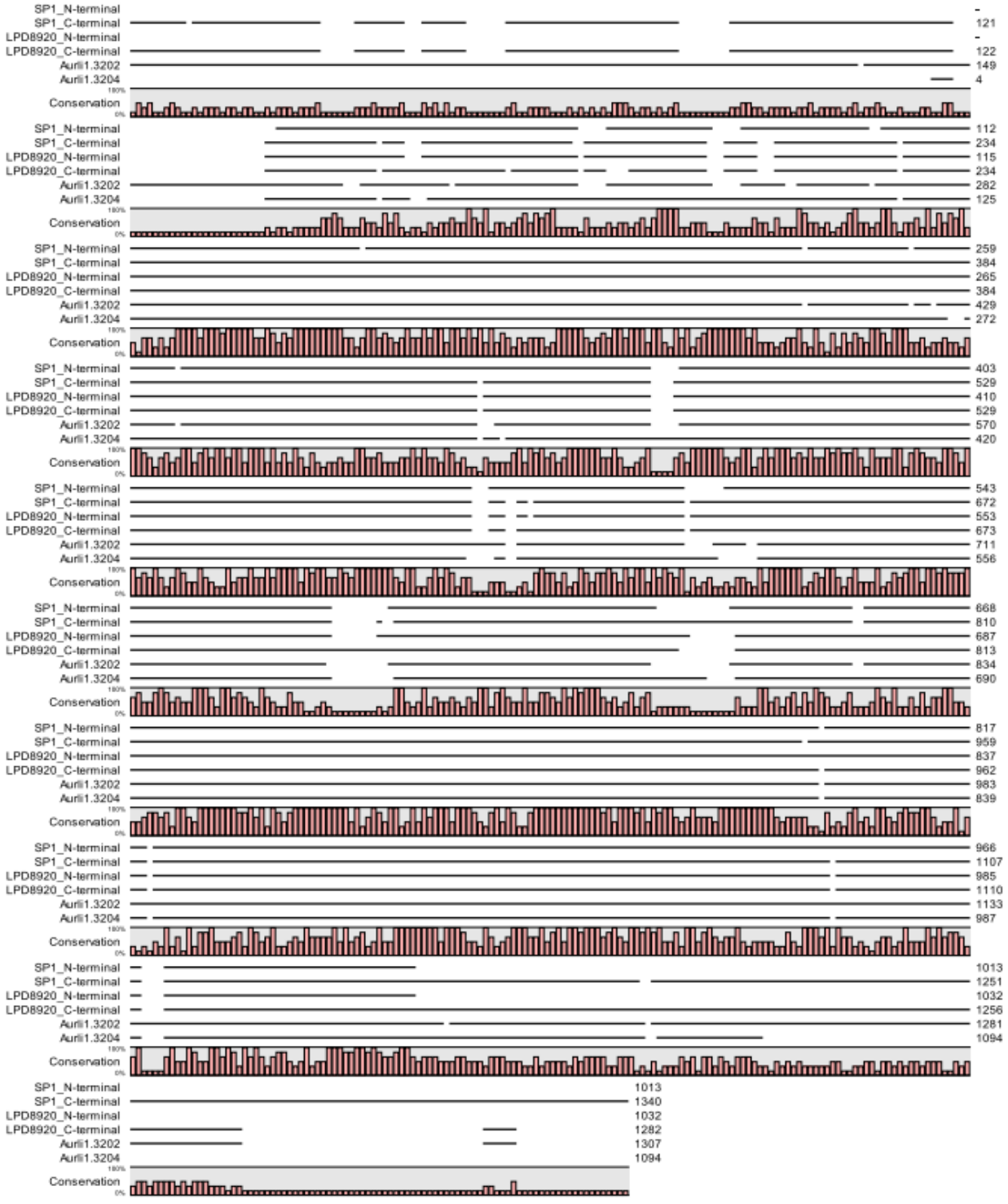


Figure 8. β-glucosidase activities of the cellular extracts of soluble and insoluble proteins and the extracellular extract from the supernatant. Enzyme activity is normalized to the protein content tested using the p-NPG assay.

Table 2. Biomass, lipid and squalene analysis of wild-type and AaBgl+ strains under glucose and cellobiose as C sources after 6 days cultivation.

Strain & C source	Biomass yield (gL^{-1})	Lipid yield (gL^{-1})	Lipid Content (% dw)	DHA Content (%dw)	Squalene Content (% dw)
Wild-type					
Glucose	5.67 ± 0.37	2.15 ± 0.07	38 ± 2.00	0.41 ± 0.03	0.27 ± 0.02
Cellobiose	-	-	-	-	-
AaBgl+					
Glucose	4.28 ± 0.35	1.26 ± 0.21	29.33 ± 2.31	1.32 ± .09	0.60 ± 0.21
Cellobiose	4.72 ± 0.23	1.28 ± 0.25	27.33 ± 6.43	0.71 ± .07	0.66 ± 0.05

Supplementary Material



Supplementary Figure 1. Alignment of the homologous regions of SP1, LPD8920, Aurli1.3204 and Aurli1.3202 using COBALT (graphics produced through CLC Sequence Viewer 7.7). Guide numbers for amino acid position are seen on the right side of the sequence. Bars show conservation percentage of a certain amino acid at each position.

Supplementary Table 1. List of paralogous and orthologous proteins to SP1 and SP2.

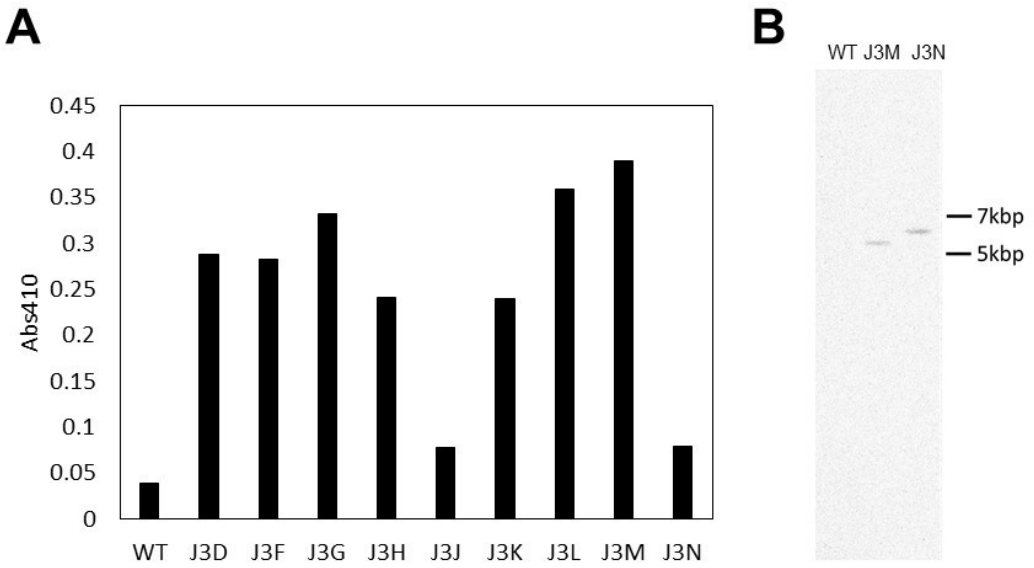
Paralogous Proteins (local protein database)			
Name	Amino acid length	NCBI conserved search domain	SignalP 4.1
LPD8920	2314	-	Secreted
LPD9373	824	vWFA domain (552-760) PAN/APPLE domain (216-291) PT repeat (298-320)	Secreted
Orthologous Proteins (JGI *A. limacinum* ATCC MYA-1381)			
Name	Amino acid length	Annotation	SignalP 4.1
Aurli1.3202	1307	Subtilase	Not secreted
Aurli1.3204	1094	Subtilase	Secreted

Supplementary Table 2. List of primers used for infusion and PCR.

Primer Name	Sequence	Purpose
INF_ubi_AABgl_F1	GCAACACTAGCCAACATG GCCAACATGCCGCGG	infusion to pUbi-Neo^r, check ABgl insertion to genome
INF_ubi_AABgl_R1	CATACTACAGATAGCTTA CTGCACTTTGGGAAGTGC TGC	infusion to pUbi-Neo^r check ABgl insertion to genome
ubiProR	GTTGGCTAGTGTTGCTTA GGTCGCT	linearization of pUbi-Neo^r for infusion
ubitermF	GCTATCTGTAGTATGTGC TATTCTC	linearization of pUbi-Neo^r for infusion
pEFubi1611F2	CCGTCTTTCGACTAGCCTT ATCGTTTAGGGAAGG	Infusion of Ubi-Abgl to pEF-Neo ^{r}
pEFubi1611R2	GCTTGCATGCACTAGGA ATTGGGCCCGACGTCG	Infusion of Ubi-Abgl to pEF-Neo ^{r}
NeoRFor	ATGATTGAACAGGACGG CCTTC	Check Neo^r insertion to genome
NeoRRev	TCAAAAGAACTCGTCCA GGAGG	Check Neo^r insertion to genome

Supplementary Table 3. Semi-quantitative results for ApiZYM extracellular enzyme activity assays for strains *Aurantiochytrium* sp. 18W-13a (18W-13a), *A. mangrovei* NYH1 (NYH1) and *A. limacinum* SR21 (SR21) at exponential (E) and stationary stages (S). Values range from 0 (no activity) to 5 (highest activity).

	18W-13a		**NYH1**		**SR21**	
Cultures	E	S	E	S	E	S
Control	0	0	0	0	0	0
Alkaline phosphatase	0	0	0	0	0	0
Esterase (C14)	1	1	1	1	1	1
Esterase (C8)	1	1	1	1	2	2
Lipase (C14)	0	0	0	0	1	0
Leucine arylamidase	1	1	2	1	5	5
Valine arylamidase	1	1	1	1	2	1
Cysteine arylamidase	0	0	0	0	1	0
Trypsin	0	0	0	0	0	0
α-chymotrypsin	0	0	0	0	0	0
Acid phosphatase	3	5	5	5	5	5
Phosphohydrolase	3	4	4	5	5	5
α-galactosidase	0	0	0	0	0	0
β-galactosidase	0	0	0	0	0	0
β-glucuronidase	0	0	0	0	0	0
α-glucosidase	0	0	0	0	0	0
β-glucosidase	0	0	1	1	0	0
β-glucosaminadase	0	0	0	0	0	0
α-mannosidase	0	0	0	0	0	0
α-fucosidase	0	0	0	0	0	0



Supplementary Figure 2. A) Screening of BGL activity in the supernatants of transformants with *AaBgl* gene insertions. B) Southern blot hybridization of *AaBgl* probe (600 bp) using 30 µg genomic DNA digested with *Hind*III and *Pst*I from the WT, J3M and J3N strains. J3M, designated as AaBgl+ transformant strain, was selected for its high BGL activity and single copy insertion.

References

Adachi T, Sahara T, Okuyama H, Morita N (2017) Glass Bead-based Genetic Transformation:An Efficient Method for Transformation of Thraustochytrid Microorganisms. J Oleo Sci 66:791–795. doi: 10.5650/jos.ess17084

Altschul SF, Wootton JC, Gertz EM, et al (2005) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272:5101–5109. doi: 10.1111/j.1742-4658.2005.04945.x

Baba Y, Sumitani J-I, Tani S, Kawaguchi T (2015) Characterization of *Aspergillus aculeatus* β-glucosidase 1 accelerating cellulose hydrolysis with T*richoderma cellulase* system. AMB Express 5:3. doi: 10.1186/s13568-014-0090-3

Cheng R, Ma R, Li K, et al (2012) *Agrobacterium tumefaciens* mediated transformation of marine microalgae Schizochytrium. Microbiol Res 167:179–186. doi: 10.1016/j.micres.2011.05.003

Chin JX, Chung BKS, Lee DY (2014) Codon Optimization OnLine (COOL): A web-based multi-objective optimization platform for synthetic gene design. Bioinformatics 30:2210–2212. doi: 10.1093/bioinformatics/btu192

Clérico EM, Maki JL, Gierasch LM (2008) Use of synthetic signal sequences to explore the protein export machinery. Biopolym - Pept Sci Sect 90:307–319. doi: 10.1002/bip.20856

Fitzpatrick J, Kricka W, James TC, Bond U (2014) Expression of three *Trichoderma reesei* cellulase genes in *Saccharomyces pastorianus* for the development of a two-step process of hydrolysis and fermentation of cellulose. J Appl Microbiol 117:96–108. doi: 10.1111/jam.12494

Fossier Marchan L, Lee Chang KJ, Nichols PD, et al (2017) Taxonomy, ecology and biotechnological applications of thraustochytrids: A review. Biotechnol Adv 1–21.

doi: 10.1016/j.biotechadv.2017.09.003

Fujita Y, Takahashi S, Ueda M, et al (2002) Direct and Efficient Production of Ethanol from Cellulosic Material with a Yeast Strain Displaying Cellulolytic Enzymes Direct and Efficient Production of Ethanol from Cellulosic Material with a Yeast Strain Displaying Cellulolytic Enzymes. Appl Environ Microbiol 26:668–673. doi: 10.1128/AEM.68.10.5136

Gupta A, Abraham RE, Barrow CJ, Puri M (2015) Omega-3 fatty acid production from enzyme saccharified hemp hydrolysate using a novel marine thraustochytrid strain. Bioresour Technol 184:373–378. doi: 10.1016/j.biortech.2014.11.031

Gupta A, Barrow CJ, Puri M (2012) Omega-3 biotechnology: Thraustochytrids as a novel source of omega-3 oils. Biotechnol Adv 30:1733–1745. doi: 10.1016/j.biotechadv.2012.02.014

Gupta A, Singh D, Byreddy AR, et al (2016) Exploring omega-3 fatty acids, enzymes and biodiesel producing thraustochytrids from Australian and Indian marine biodiversity. Biotechnol J 11:345–355. doi: 10.1002/biot.201500279

Hong WK, Kim CH, Rairakhwada D, et al (2012) Growth of the oleaginous microalga *Aurantiochytrium* sp. KRS101 on cellulosic biomass and the production of lipids containing high levels of docosahexaenoic acid. Bioprocess Biosyst Eng 35:129–133. doi: 10.1007/s00449-011-0605-0

Ishitsuka K, Koide M, Yoshida M, et al (2016) Identification of intracellular squalene in living algae, *Aurantiochytrium mangrovei* with hyper-spectral coherent anti-Stokes Raman microscopy using a sub-nanosecond supercontinuum laser source. J Raman Spectrosc. doi: 10.1002/jrs.4979

Isikgor FH, Becer CR (2015) Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym Chem 6:4497–4559. doi:

10.1039/C5PY00263J

Jain R, Raghukumar S, Tharanathan R, Bhosle NB (2005) Extracellular polysaccharide production by thraustochytrid protists. Mar Biotechnol 7:184–192. doi: 10.1007/s10126-004-4025-x

Juntila DJ, Yoneda K, Suzuki I (2017) Identification of extracellular proteins from *Aurantiochytrium* sp. 18W-13a. J Appl Phycol 30:63–69. doi: 10.1007/s10811-017-1171-x

Kanchana R, Muraleedharan UD, Raghukumar S (2011) Alkaline lipase activity from the marine protists, thraustochytrids. World J Microbiol Biotechnol 27:2125–2131. doi: 10.1007/s11274-011-0676-8

Kawaguchi T, Enoki T, Tsurumaki S, et al (1996) Cloning and sequencing of the cDNA encoding beta-glucosidase 1 from A*spergillus aculeatus*. Gene 173:287–288. doi: 0378111996001795 [pii]

Kaya K, Nakazawa A, Matsuura H, et al (2011) Thraustochytrid *Aurantiochytrium* sp. 18W-13a Accummulates High Amounts of Squalene. Biosci Biotechnol Biochem 75:2246–2248. doi: 10.1271/bbb.110430

Kobayashi T, Sakaguchi K, Matsuda T, et al (2011) Increase of eicosapentaenoic acid in thraustochytrids through thraustochytrid ubiquitin promoter-driven expression of a fatty acid Δ5 desaturase gene. Appl Environ Microbiol 77:3870–3876. doi: 10.1128/AEM.02664-10

Liu Y, Singh P, Sun Y, et al (2014a) Culturable diversity and biochemical features of thraustochytrids from coastal waters of Southern China. Appl Microbiol Biotechnol 98:3241–3255. doi: 10.1007/s00253-013-5391-y

Liu Y, Singh P, Sun Y, et al (2014b) Culturable diversity and biochemical features of thraustochytrids from coastal waters of Southern China. Appl Microbiol Biotechnol

98:3241–3255. doi: 10.1007/s00253-013-5391-y

Liu Z, Ho S-H, Sasaki K, et al (2016) Engineering of a novel cellulose-adherent cellulolytic *Saccharomyces cerevisiae* for cellulosic biofuel production. Sci Rep 6:24550. doi: 10.1038/srep24550

Ma Z, Tan Y, Cui G, et al (2015) Transcriptome and gene expression analysis of DHA producer *Aurantiochytrium* under low temperature conditions. Sci Rep 5:14446. doi: 10.1038/srep14446

Marchler-Bauer A, Derbyshire MK, Gonzales NR, et al (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222–D226. doi: 10.1093/nar/gku1221

Matsuura H, Nakazawa A, Ueda M, et al (2012) On the Bio-Rearrangement into Fully Saturated Fatty Acids-Containing Triglyceride in *Aurantiochytrium* sp. Procedia Environ Sci 15:66–72. doi: 10.1016/j.proenv.2012.05.011

Merkx-Jacques A, Rasmussen H, Muise DM, et al (2018) Engineering xylose metabolism in thraustochytrid T18. Biotechnol Biofuels 11:248. doi: 10.1186/s13068-018-1246-1

Nagano N, Matsui S, Kuramura T, et al (2011) The Distribution of Extracellular Cellulase Activity in Marine Eukaryotes, Thraustochytrids. Mar Biotechnol 13:133–136. doi: 10.1007/s10126-010-9297-8

Nakazawa A, Kokubun Y, Matsuura H, et al (2014) TLC screening of thraustochytrid strains for squalene production. J Appl Phycol 26:29–41. doi: 10.1007/s10811-013-0080-x

Nakazawa A, Matsuura H, Kose R, et al (2012a) Optimization of culture conditions of the thraustochytrid *Aurantiochytrium* sp. strain 18W-13a for squalene production. Bioresour Technol 109:287–291. doi: 10.1016/j.biortech.2011.09.127

Nakazawa H, Kawai T, Ida N, et al (2012b) Construction of a recombinant T*richoderma reesei* strain expressing *Aspergillus aculeatus* β-glucosidase 1 for efficient biomass

conversion. Biotechnol Bioeng 109:92–99. doi: 10.1002/bit.23296

Nordberg H, Cantor M, Dusheyko S, et al (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42:26–31. doi: 10.1093/nar/gkt1069

Okino N, Wakisaka H, Ishibashi Y, Ito M (2018) Visualization of Endoplasmic Reticulum and Mitochondria in *Aurantiochytrium limacinum* by the Expression of EGFP with Cell Organelle-Specific Targeting/Retaining Signals. Mar Biotechnol 20:182–192. doi: 10.1007/s10126-018-9795-7

Papadopoulos JS, Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073–1079. doi: 10.1093/bioinformatics/btm076

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–6. doi: 10.1038/nmeth.1701

Qi F, Zhang M, Chen Y, et al (2017) A lignocellulosic hydrolysate-tolerant *Aurantiochytrium* sp. mutant strain for docosahexaenoic acid production. Bioresour Technol 227:221–226. doi: 10.1016/j.biortech.2016.12.011

Raghukumar S (2008) Thraustochytrid marine protists: Production of PUFAs and other emerging technologies. Mar Biotechnol 10:631–640. doi: 10.1007/s10126-008-9135-4

Sakaguchi K, Matsuda T, Kobayashi T, et al (2012) Versatile transformation system that is applicable to both multiple transgene expression and gene targeting for thraustochytrids. Appl Environ Microbiol 78:3193–3202. doi: 10.1128/AEM.07129-11

Sakamoto R, Arai M, Murao S (1985a) Enzymic properties of three β-glucosidases from

Aspergillus aculeatus no. F-50. Agric Biol Chem 49:1283–1290. doi: 10.1080/00021369.1985.10866915

Sakamoto R, Kanamoto J, Arai M, Murao S (1985b) Purification and Physicochemical Properties of Three β-Glucosidases from *Aspergillus aculeatus* No. F-50. Agric Biol Chem 49:1275–1281. doi: 10.1080/00021369.1985.10866914

Salehi Jouzani G, Taherzadeh MJ (2015) Advances in consolidated bioprocessing systems for bioethanol and butanol production from biomass: a comprehensive review. Biofuel Res J 5:152–195. doi: 10.18331/BRJ2015.2.1.4

Tani N, Yoneda K, Suzuki I (2018) The effect of thiamine on the growth and fatty acid content of *Aurantiochytrium* sp. Algal Res 36:57–66. doi: 10.1016/j.algal.2018.10.012

Taoka Y, Nagano N, Okita Y, et al (2009) Extracellular Enzymes Produced by Marine Eukaryotes, Thraustochytrids. Biosci Biotechnol Biochem 73:180–182. doi: 10.1271/bbb.80416

Thompson CE, Beys-da-Silva WO, Santi L, et al (2013) A potential source for cellulolytic enzyme discovery and environmental aspects revealed through metagenomics of Brazilian mangroves. AMB Express 3:65. doi: 10.1186/2191-0855-3-65

Whittaker CA, Hynes RO (2002) Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. Mol Biol Cell 13:3369–3387. doi: 10.1091/mbc.E02-05-0259

Yarimizu T, Nakamura M, Hoshida H, Akada R (2015) Synthetic signal sequences that enable efficient secretory protein production in the yeast *Kluyveromyces marxianus*. Microb Cell Fact 14:20. doi: 10.1186/s12934-015-0203-y

Yoneda K, Yoshida M, Suzuki I, Watanabe MM (2016) Identification of Major Lipid Droplet Protein in a Marine Diatom P*haeodactylum tricornutum*. Plant Cell Physiol 57:397–406. doi: 10.1093/pcp/pcv204

Acknowledgments

I would like to express my gratitude to Prof. Makoto M. Watanabe of the Algal Biomass and Energy System Research and Development Center, University of Tsukuba for providing the *Aurantiochytrium* sp. 18W-13a. Also, to Dr. Masaki Yoshida from the University of Tsukuba and Dr. Ryo Koyanagi from Okinawa Institute of Science and Technology for the local protein, transcriptome, and genome databases used in my study. The service for the Edman protein sequencing was provided by Yumiko Makino in the Functional Genomics Facility, National Institute for Basic Biology and Dr. Shoji Mano in National Institute for Basic Biology. Most importantly, to Dr. Makoto Ito for his contribution of the backbone plasmids used in this study and electroporation protocol and insights.

Thank you to Dr. Iwane Suzuki for his guidance and support during the course of my work. To Dr. Kohei Yoneda for his contribution in proteomic analysis. To my fellow laboratory Ph.D. students which are already doctors now, Bakku, Asada-san, Machida-san, for their help and friendship. Thank you to all members of the PLMET laboratory.

Thank you to my family, Mama Juliet, Papa Danilo, Inday Dawn, Daphne and Manoy Dandy, who stood by my side in the toughest times. Thank you for my friends and loved ones, especially fellow students Grace and Lhannie, for always hearing me out. To my sisters in faith, Ate Joy, Ate Fe, Ate Noemi and Cha, for always empowering me. To my closest friends Jireh, Nikki and Yev, for the long-distance calls. To my MBB friends and professors (too many to mention, special thanks to Dr. Bautista and Dr. Prieto) who continue to believe in me. To the doctors who helped me during my sickness, Dr. Amosco and Dr. Yamaguchi, thank you so much for your patience.

Thank you to the Lord for His grace, strength and provision. For giving me purpose and hope in this life. Jeremiah 29:11