Developing Consolidated Bioprocessing Competent *Saccharomyces cerevisiae* Through the Optimized Expression of Fungal Glycosyl Hydrolases

Lina Mougharbel

A Thesis

In the Department

of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Biology) at

Concordia University

Montreal, Quebec, Canada

February 2018

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Lina Mougharbel**

Entitled: **Developing Consolidated Bioprocessing Competent *Saccharomyces cerevisiae* Through the Optimized Expression of Fungal Glycosyl Hydrolases**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Paul Joyce	Chair
Dr. George B. Szatmari	External Examiner
Dr. Joanne Turnbull	External to Program
Dr. Vincent Martin	Examiner
Dr. Justin Powlowski	Examiner
Dr. Reginald Storms	Thesis Supervisor

Approved by

Dr. Grant Brown, Graduate Program Director

April 5, 2018

Dr. André Roy, Dean
Faculty of Arts and Science

ABSTRACT

Developing Consolidated Bioprocessing Competent *Saccharomyces cerevisiae* Through the Optimized Expression of Fungal Glycosyl Hydrolases

Lina Mougharbel, Ph.D.

Concordia University, 2018

Vascular plant biomass or lignocellulosic biomass is the world's most abundant renewable carbon source. Ethanol production from lignocellulosic biomass has the potential to reduce the use of fossil fuels and the production of greenhouse gasses. Currently, pre-treated lignocellulosic biomass is converted to ethanol in a four-step process: (1) production of glycosylhydrolase enzymes for polysaccharide hydrolysis; (2) hydrolysis of the polysaccharide component of lignocellulosic biomass into fermentable sugars; (3) fermentation of hexose sugars; and (4) fermentation of pentose sugars. A major contributor to the high cost of cellulosic ethanol is the cellulase enzymes produced in step 1 and used in step 2 to hydrolyze cellulose into fermentable sugars. Consolidated bioprocessing (CBP) is a process that uses an organism or organisms that can both produce the enzymes for hydrolysis of the cellulosic component of plant biomass into fermentable glucose, and ferment the glucose into ethanol. CBP-competent *Saccharomyces cerevisiae* should dramatically reduce the cost of cellulosic ethanol production by bypassing the need for separate steps for cellulase enzyme production, enzymatic hydrolysis and fermentation. The goal of my research is to contribute to the development of *S. cerevisiae* strains for application in the production of renewable cellulosic ethanol using the one-step CBP process. This work describes: the construction of recombinant *S. cerevisiae* strains that are capable of efficiently expressing a fungal β-glucosidase and are therefore capable of using cellobiose as their sole carbon source; the identification and cloning of a library of 25 heterologous fungal endoglucanases; screening *S. cerevisiae* transformants expressing the library of cloned endoglucanases for transformants that could produced functional secreted endoglucanase activity; coding region optimization of 5 endoglucanase genes that were able to produce functional endoglucanase in *S. cerevisiae*; the construction of recombinant *S. cerevisiae* strains that

are capable of producing both functional β-glucosidase and endoglucanase at levels sufficient for using the soluble cellulosic polymer carboxymethylcellulose (CMC-4M) as a sole carbon source; and, the identification of several heterologous fungal cellobiohydrolase ORFs that can express functional secreted cellobiohydrolase in *S. cerevisiae*.

Acknowledgement

I wish to thank, first and foremost, my supervisor, Dr. Reginald Storms, for his mentorship and continued support throughout my Ph.D. research. I also wish to thank my committee members, Dr. Vincent Martin and Dr. Justin Powlowski, for their advice and valuable input throughout my research project.

I thank my current and former lab members, Yun, Sophia, Minghui, Susan, Humberto, Greg, Shaghayegh, and Rebecca, for their friendship and moral support.

I wish to express my deepest gratitude to Dr. Edith Munro, her memory will be with me, always.

I thank my friend and mentor, SJB, who helped me get through life challenges.

To my partner, Nadim, who has been there for me from the very beginning and who encouraged and supported me though the challenges of life and graduate school. I thank my loyal companion, Milou, who have kept me company through long hours and very late nights of writing.

Finally, I wish to express my deepest gratitude to my mother, my sisters, and my brother.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1G	First generation
2G	Second generation
A. niger	*Aspergillus niger*
BCA	Bicinchoninic acid
BGL	β-glucosidase
BSA	Bovine serum albumin
CAI	Codon adaptation index
CBH	Cellobiohydrolase
CBM1	Carbohydrate binding module family 1
CBP	Consolidated bioprocessing
CDD	Conserved domain database
CFD	Codon frequency distribution
CMC	Carboxymethyl Cellulose
C_q	Quantification cycle
DCW	Dry cell weight
E. coli	*Escherichia coli*
EG	Endoglucanase
GH	Glycosyl hydrolase
GHG	Green house gasses
MCS	Multiple cloning site
MFα	Mating factor alpha
ORF	Open reading frame
PASC	Phosphoric acid swollen cellulose
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RSCU	Relative synonymous codon usage

S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SHF	Separate hydrolysis and fermentation
SSCF	Simultaneous hydrolysis and co-fermentation
SSF	Simultaneous hydrolysis and fermentation
T. reesei	*Trichoderma reesei*
UPR	Unfolded protein response
VTO	Vector transformant only
WT	Wild type
YNB	Yeast nitrogen base

Introduction

1.1. Cellulosic Ethanol

Conversion of lignocellulosic biomass to ethanol for biofuel has the potential to reduce the use of fossil fuels, reduce the production of the greenhouse gas CO_2 and increase energy security (Lynd, Wyman, Gerngross 1999). The Government of Canada is committed to reduce the emission of green house gases (GHG). One initiative was the Renewable Fuels Regulation SOR/2010-189, which mandated that gasoline produced or imported in Canada must contain an average content of 5% renewable fuel (2016).

Ethanol production from biomass falls into two categories; first generation (1G) ethanol, and second generation (2G) ethanol (Lynd et al. 2002; Sims et al. 2010). First generation ethanol is produced from sugar or starch-containing crops with Brazil and the US being the major producers of 1G ethanol. In Brazil 1G ethanol is mainly produced by fermenting sucrose extracted from sugar cane, whereas in the US ethanol is produced using glucose derived from corn starch (Sims et al. 2010). Even though first generation ethanol technology is well understood, it has several shortcomings, including: 1G ethanol production competes with traditional agriculture crops for land use and water; 1G ethanol is expensive relative to equivalent products derived from fossil fuels; the economic viability of 1G ethanol industries is heavily dependent upon government subsidies; and the green house gas footprint of 1G ethanol is not significantly lower than that of traditional fossil fuels once other factors including land use are taken into consideration (Sims et al. 2010).

Second generation (2G) ethanol is produced from lignocellulosic biomass, the most abundant carbon source on earth. Lignocellulosic feedstocks can be: (i) agricultural residues or by-products (e.g. cereal straw, sugarcane bagasse, and forest residues), (ii)

organic municipal solid wastes, and (iii) dedicated feedstocks (e.g. grasses, short-rotation forests, and energy crops such as sorghum (Sims et al. 2010). Second generation ethanol technology has made significant advances in the past few years. Reflecting this, the first large-scale commercial 2G cellulosic ethanol facility began producing cellulosic ethanol from sugarcane bagasse in late 2014 (http://www.iogen.ca/raizen-project/index.html). This facility was developed by Iogen Corporation of Canada and Brazilian ethanol giant Raízen Energia. Notwithstanding this success, cellulosic ethanol technology is still an immature technology compared to first generation ethanol technology.

1.2. Structure and Composition of Lignocellulosic Biomass

Plant cell walls are mainly composed of cellulose, hemicellulose, and lignin, representing approximately 20-50%, 15-35%, and 10-30%, respectively, of its dry weight (Pauly and Keegstra 2008). The percentage of each of the main components of plant biomass varies significantly across different sources of lignocellulose. For example, the stems of hardwoods are composed of 40-55% cellulose, 24-40% hemicellulose, and 18-25% lignin, whereas wheat straw contains about 50% hemicellulose, 30% cellulose and 15% lignin (Sun and Cheng 2002).

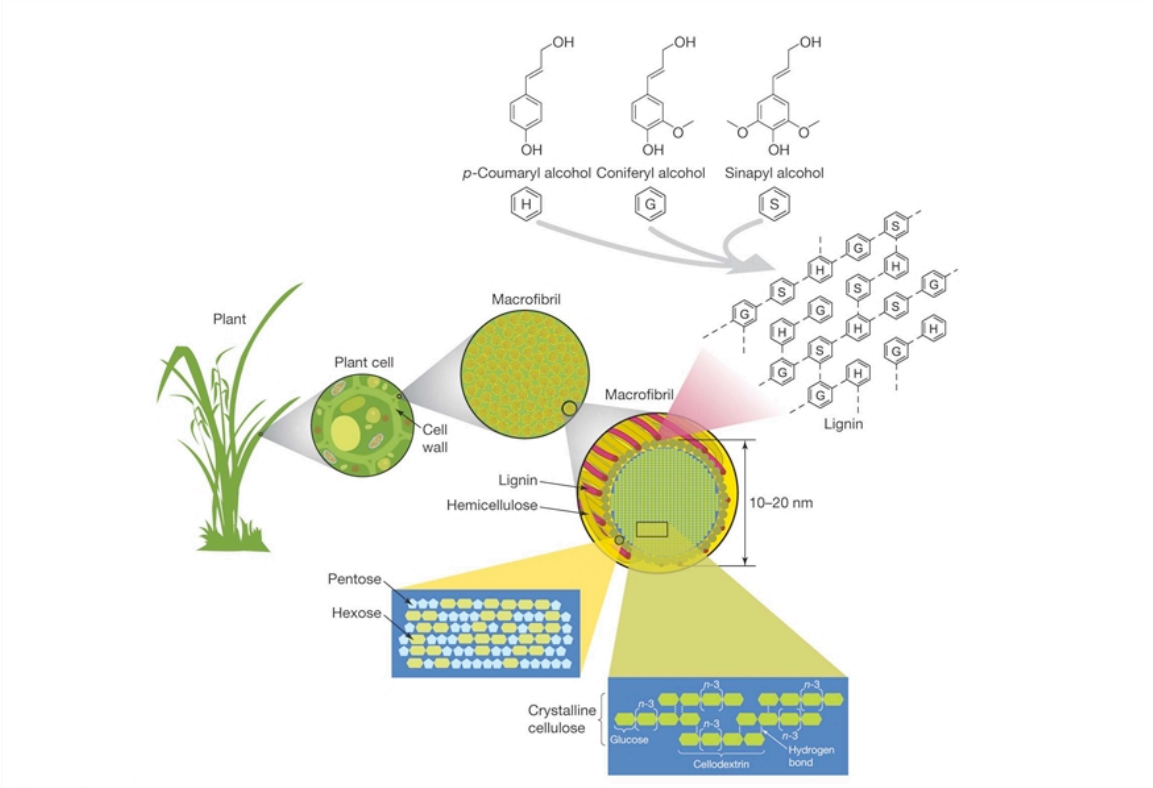


Figure 1.1 – Structure of lignocellulose. Lignocellulose is composed of cellulose, hemicellulose, and lignin. Cellulose, the most abundant component of lignocellulose, is composed of glucose units linked by β-1,4-glycosidic bonds. Hemicellulose, the second most abundant component of lignocellulose, is composed of branched heterogeneous polymers composed of both hexose and pentose sugar monomers and uric acids. Elementary fibrils are packed into larger units called microfibrils that associate to form macrofibrils. Lignin is composed of varying ratios of three major phenolic compounds, *ρ*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. Figure is reprinted from (Rubin 2008) with permission.

1.2.1. Cellulose

Cellulose in plants is present as fibers associated with their cell walls. The cellulose fibers are embedded within a matrix of lignin and hemicellulose. Each cellulose fiber is composed of several hundred thousand cellulose molecules, each molecule consisting of 1,000 to 10,000 β-1,4 linked glucose residues (Chang, Chou, Tsao 1981). Cellulose is composed of D-glucose units linked by β-1,4-glycosidic bonds (Jørgensen, Kristensen, Felby 2007). Each glucose unit in cellulose is rotated 180^{o} with respect to adjacent glucose units making cellobiose the repeating unit in cellulose (Figure 1.2).

Thirty-six cellulose chains are aggregated into elementary fibrils that allow intra- and intermolecular hydrogen bonds due to the linear structure of cellulose (Frey-Wyssling 1954; Jørgensen, Kristensen, Felby 2007; Meier 1962; O'Sullivan 1997). Elementary fibrils are packed into larger units called microfibrils, which are composed of a 100 or more elementary fibrils (Meier 1962; O'Sullivan 1997). Microfibrils are packed together forming macrofibrils (Meier 1962).

Cellulose elementary fibrils are synthesized by cellulose synthase enzyme complexes (CelS), which are rosette structures in the plasma membrane (Doblin et al. 2002; Somerville 2006). Based on the determined cellulose synthase rosette structure of vascular plants it is proposed that the rosette-structured cellulose synthase complexes synthesize and coordinate aggregation of the 36 cellulose molecule elementary fibrils.

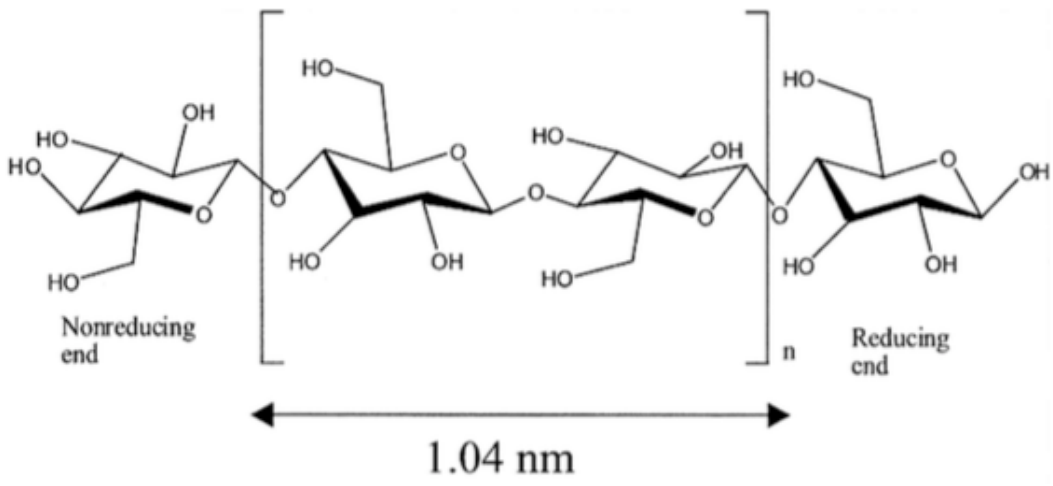


Figure 1.2 – Structure of cellulose. Adjacent glucose units in cellulose are rotated 180^{o} with respect to each other making cellobiose "n" the repeating unit in cellulose. Figure is reprinted from (Zhang and Lynd 2004) with permission.

1.2.2. Hemicellulose

Hemicellulose, the second most abundant component of lignocellulosic biomass, is composed of branched heterogeneous polymers composed of both hexose and pentose sugar monomers and uric acids. The hexose sugars include D-glucose, D-galactose, D-mannose, and L-rhamnose. The pentose sugars include D-xylose and L-arabinose. The uric acids include 4-*O*-methyl- D-glucuronic acid, D-glucuronic acid, and D-galacturonic acid (Schadel et al. 2009). Hemicelluloses can be grouped into four classes based on the type of sugars in the backbone: xylan, xyloglucan, mannan, and mixed linkage β-glucans (Schadel et al. 2009). Xylan, the most abundant hemicellulose in plant cell walls, is a linear polymer of *β*-1,4-linked D-xylose residues and is commonly substituted with α-1,2-linked glucuronosyl and 4-*O*-methyl glucuronosyl residues (Scheller and Ulvskov 2010). Side chain type and distribution varies between different plant species and within different tissues of the same plant (Scheller and Ulvskov 2010). Xyloglucan is a linear polymer of *β*-1,4-linked D-glucose residues and is commonly substituted with D-xylose and D-xylose substituted with D-galactose side-chains (Scheller and Ulvskov 2010). Mannan is a linear polymer of *β*-1,4-linked D-mannose residues that may contain 1,6-linked D-galactose side-chains (Scheller and Ulvskov 2010). Mixed linkage β-glucans are mainly composed of *β*-1,3-linked cellotriose and cellotetraose units (Scheller and Ulvskov 2010).

1.2.3. Lignin

Lignin is composed of varying ratios of three major phenolic compounds, *ρ*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol (Rubin 2008). The ratio of lignin components varies between different plant species and between varying tissues of the same species (Rubin 2008). Lignin provides plants with structural support and pathogen defence, and its hydrophobicity allows for water transport in vascular plant tissues (Del Rio et al. 2007).

1.3. Cellulosic Biomass Hydrolysis

1.3.1. Pretreatment

The lignin and hemicellulose matrix, and the highly crystalline structure of cellulose prevent the penetration of enzymes and water (Lynd et al. 2002) making the cellulose in plant biomass recalcitrant to enzyme degradation. Direct enzymatic hydrolysis yields of native lignocellulose is less than 20% of theoretical, therefore, a pretreatment step is required (Lynd et al. 2002). The pretreatment step disrupts the lignin and hemicellulose matrix surrounding the cellulose fibers (Lynd et al. 2002). The degree of cellulose and hemicellulose depolymerization and lignin solubilization is determined by the pretreatment process (Lynd et al. 2002). The increased porosity of the cell walls coupled with hemicellulose depolymerization and lignin solubilization allows for greater accessibility of the cell wall cellulose component to enzyme hydrolysis (Figure 1.3) (Limayem and Ricke 2012; Lynd et al. 2002; Zheng, Pan, Zhang 2009).

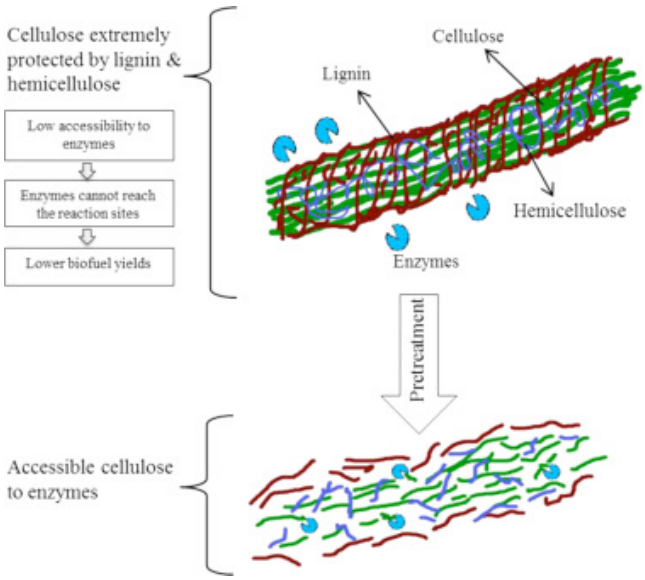


Figure 1.3 – Pretreatment of lignocellulosic biomass. The goal of the pretreatment step is to disrupt the lignin seal and increase the accessibility of cellulose and hemicellulose to hydrolytic enzymes. Figure is reprinted from (Shirkavand et al. 2016) with permission.

In general, pretreatment methods fall into three categories: physical, chemical, and biological (Limayem and Ricke 2012; Zheng, Pan, Zhang 2009). The most common physical pretreatment processes include steam explosion and hot water pretreatment (Limayem and Ricke 2012; Zheng, Pan, Zhang 2009). Physical pretreatment methods do not require the use of chemicals, resulting in the formation of less inhibitors and toxic-end products than are obtained with chemical pretreatment methods (Limayem and Ricke 2012). Physical pretreatment methods are, however, not effective on all biomass materials where chemical pretreatment methods work (Limayem and Ricke 2012). Chemical pretreatment methods include: dilute acid pretreatment; ammonia fiber expansion (AFEX); and, organosolv pretreatment (Limayem and Ricke 2012; Lynd et al. 2002; Sun and Cheng 2002). These processes can be very efficient, but many of them release high levels of inhibitors and toxic products. Finally, in biological pretreatment of biomass materials, brown-rot and white-rot fungi are used to reduce the recalcitrance of lignocellulosic biomass (Limayem and Ricke 2012; Sun and Cheng 2002). Even though biological pretreatment processes are the most environmentally friendly and require very little energy consumption, they are very slow processes (Limayem and Ricke 2012).

When selecting a pretreatment method for lignocellulosic biomass materials, the most important factors to consider are the production of toxic and inhibitory compounds, the cost of the pretreatment process and the efficiency with which the cellulosic component is rendered accessible to enzymatic hydrolysis (Limayem and Ricke 2012).

1.3.2. Hydrolysis of pretreated biomass into fermentable sugars

Due to the recalcitrant nature of lignocellulose, lignocellulose-degrading organisms produce cellulase and hemicellulase systems made up of a mixture of cellulases, hemicellulases and accessory proteins that work synergistically to hydrolyze the cellulose and hemicellulose polymers in lignocellulosic biomass.

Cellulases and hemicellulases belong to a class of enzymes called glycosyl hydrolases. Glycosyl hydrolases hydrolyze the glycosidic bonds between two or more carbohydrates

or between a carbohydrate and a non-carbohydrate (Lynd et al. 2002). Glycosyl hydrolase enzymes have been classified into over 135 families, according to the Carbohydrate-Active Enzymes database (Lombard et al. 2014), based on amino acid sequence similarities within their catalytic domains (Davies and Henrissat 1995; Henrissat 1991; Henrissat and Bairoch 1993). The most widely studied cellulase system is that of *Trichoderma reesei*. The *T. reesei* cellulase system is mainly composed of 6 cellulase degrading glycosylhydrolases, including: three endoglucanases, EGI (Cel7B), EGII (Cel5A), and EGIII (Cel12), belonging to glycosylhydrolase families 7, 5, and 12 respectively (Henrissat 1991; Henrissat and Bairoch 1993; Henrissat and Davies 1997); two cellobiohydrolases, CBHI (Cel7A) and CBHII (Cel6A) belonging to glycosylhydrolase families 7 and 6 respectively; and one β-glucosidase, BGL1 (Cel3A) belonging to glycosylhydrolase family 3 (Henrissat 1991; Henrissat and Bairoch 1993; Henrissat and Davies 1997).

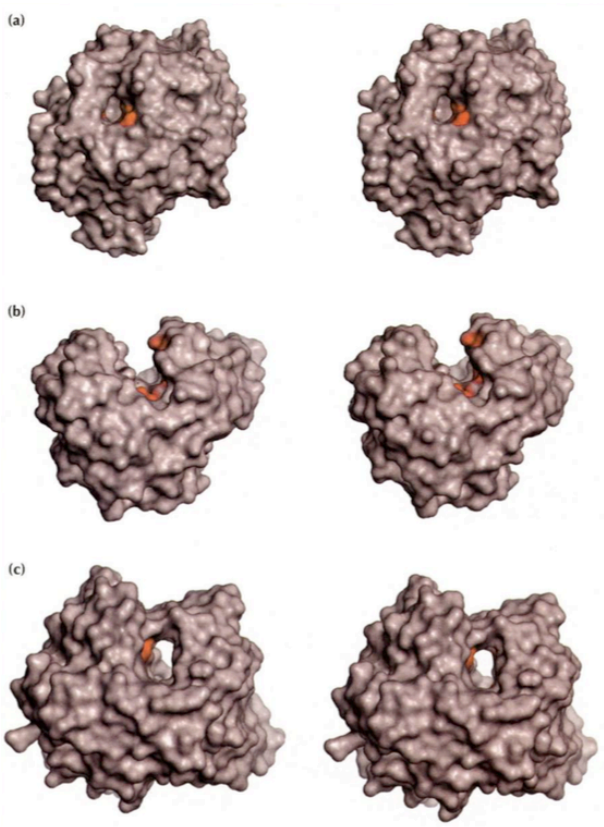


Figure 1.4 – Types of active site topologies found in glycosyl hydrolases. (a) The pocket (e.g. β-glucosidases). (b) The cleft (e.g. Endoglucanases). (c) The tunnel (e.g. Cellobiohydrolases). Figure is reprinted from (Davies and Henrissat 1995) with permission.

Three active site topologies are found in glycosyl hydrolases (Figure 1.4), the pocket, the cleft, and the tunnel (Davies and Henrissat 1995; Henrissat and Davies 1997). The pocket topology (Figure 1.4a) is optimal for the recognition of non-reducing ends of substrates and is common in enzymes whose substrates contain a large number of non-reducing ends (e.g. β-glycosidases, β-galactosidases, and β-amylases) (Davies and Henrissat 1995; Henrissat and Davies 1997). Enzymes with a pocket site topology are not active on highly polymeric substrates like native cellulose, which has very few non-reducing ends (Davies and Henrissat 1995; Henrissat and Davies 1997).

The cleft active site topology (Figure 1.4b) is commonly found in endo-acting polysaccharidases such as endoglucanases and α-amylases, and allows random binding of several sugar units within the chain of its polysaccharide substrate (Davies and Henrissat 1995; Henrissat and Davies 1997).

The tunnel active site topology (Figure 1.4c) apparently arose from the cleft topology after the protein evolved long loops that covered the cleft to form a tunnel with the catalytic subsites enclosed within the tunnel (Davies and Henrissat 1995; Henrissat and Davies 1997). The tunnel active site topology is found in cellobiohydrolases that, due to the position of the catalytic site, are able to release the product, which is usually a short cellodextrin, while remaining firmly bound to the polysaccharide chain, thereby facilitating a processive type action on either the reducing or non-reducing end of the polysaccharide (Davies and Henrissat 1995; Henrissat and Davies 1997).

1.3.3. Enzymatic hydrolysis of cellulose

Although cellulose is a simple polymer of β-1,4 linked glucose residues, its hydrolysis to fermentable glucose units typically requires the cooperative action of at least three distinct types of glycosylhydrolase enzymes (Figure 1.5): endoglucanases (EGs; EC 3.2.1.4), exoglucanases (CBHs; EC 3.2.1.91) and β-glucosidases (BGs; EC 3.2.1.21). The endoglucanases randomly cleave the intramolecular β-1,4-glucosidic bonds of amorphous

cellulose generating oligosaccharides of various lengths, including soluble cellodextrins and cellobiose (Henrissat et al. 1985; Lynd, Wyman, Gerngross 1999). The exoglucanases act on the reducing and non-reducing ends of cellulose generally generating either glucose (glucanohydrolases) or cellobiose (cellobiohydrolases) as their major product (Henrissat et al. 1985; Lynd, Wyman, Gerngross 1999). The β-glucosidases hydrolyze soluble cellodextrins and cellobiose to glucose (Henrissat et al. 1985; Lynd, Wyman, Gerngross 1999).

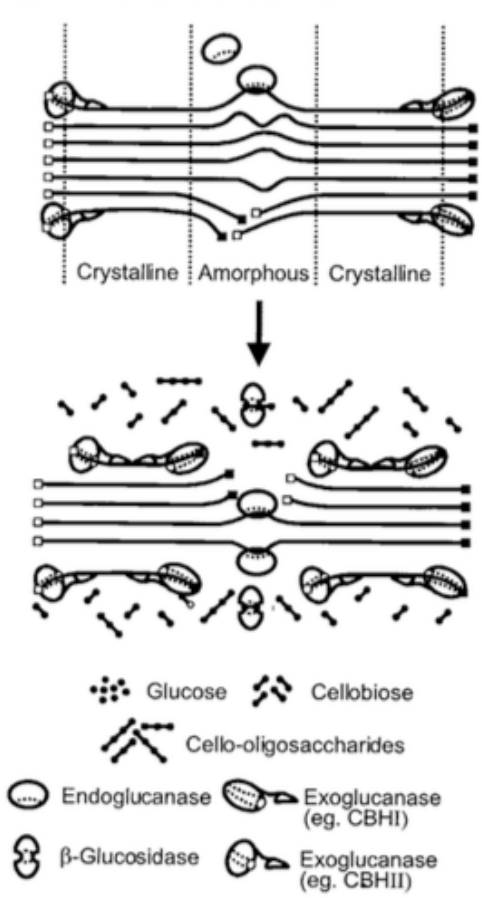


Figure 1.5 – Schematic representation of cellulose hydrolysis. Cellobiohydrolases act on the reducing (CBH1) and non-reducing (CBH2) ends of cellulose, endoglucanases act on the amorphous regions of cellulose, and β-glucosidases hydrolyze soluble cellodextrins and cellobiose to glucose. Figure is reprinted from (Lynd et al. 2002) with permission.

1.3.4. Enzymatic hydrolysis of hemicellulose

The efficient hydrolysis of pretreated lignocellulosic biomass requires the hydrolysis of the hemicellulose component of biomass for an economically viable bioconversion process. Xylan, the second most abundant hemicellulose component in plant cell walls, is hydrolyzed by a group of enzymes called xylanases. Endo-1,4-β-xylanases (EC 3.2.1.8) cleave the glycosidic bonds in the xylan backbone releasing xylooligosaccharides of various lengths (Gírio et al. 2010). β-xylosidases (EC 3.2.1.37) hydrolyse xylooligosaccharides and xylobiose releasing xylose (Gírio et al. 2010). Hydrolysis of mannan-type hemicellulose backbone into simple sugars requires the synergistic action of endo-1,4-β-mannanases (EC 3.2.1.78) and β-mannosidases (EC 3.2.1.25) (Dhawan and Kaur 2007). Side-chain sugars, which may be attached to xylan and mannan backbones, are removed by enzymes such as α-arabinofuranosidases, α-glucuronidase α-galactosidases (EC 3.2.1.22), β-glucosidases (EC 3.2.1.21), and acetyl mannan esterases (Dhawan and Kaur 2007).

1.4. Conversion of Pretreated Biomass to Ethanol

The conversion of pretreated lignocellulose to ethanol requires four biological events: the production of saccharolytic enzymes (cellulases and hemicellulases), the hydrolysis of polysaccharides in pretreated lignocellulosic biomass to fermentable sugars, the fermentation of hexose sugars, and the fermentation of pentose sugars (Lynd 1996).

Four different approaches have been proposed for the production of ethanol from lignocellulosic biomass. These four approaches include; Separate Hydrolysis and Fermentation (SHF), Simultaneous Saccharification and Fermentation (SSF), Simultaneous Saccharification, Co-fermentation (SSCF), and Consolidated Bioprocessing (CBP).

1.4.1. Separate hydrolysis and fermentation (SHF)

In separate hydrolysis and fermentation, exogenous enzymes hydrolyze pre-treated lignocellulosic biomass. Hexose and pentose sugars are then fermented separately. The advantage of using SHF is that the hydrolysis reaction occurs in two different reactors, one reactor at the optimum temperature and pH for the exogenous cellulase and hemicellulase enzymes and a second reactor at the optimum temperature for the fermentation of the hexose and pentose sugars.

Currently, the production of ethanol from lignocellulosic substrates is mainly done using a SHF process that uses three separate stages and therefore three separate reactors; one reactor is needed for the pre-treatment step, one reactor for production of the cellulase enzymes needed for the enzymatic hydrolysis of cellulose polymers into glucose, and a third reactor for the fermentation of glucose into ethanol by *Saccharomyces cerevisiae* (Lynd et al. 2002). Ethanol produced using this three-reactor process is expensive relative to transportation fuels produced from petroleum (Lynd et al. 2002). Major contributions to the high cost of bioethanol production using present day SHF technology are the exogenous enzymes added to hydrolyze pretreated biomass into fermentable sugars, the inability of present day *S. cerevisiae* strains to efficiently ferment pentose sugars (Lynd et al. 2002), and the requirement of three separate reactors for the SHF process.

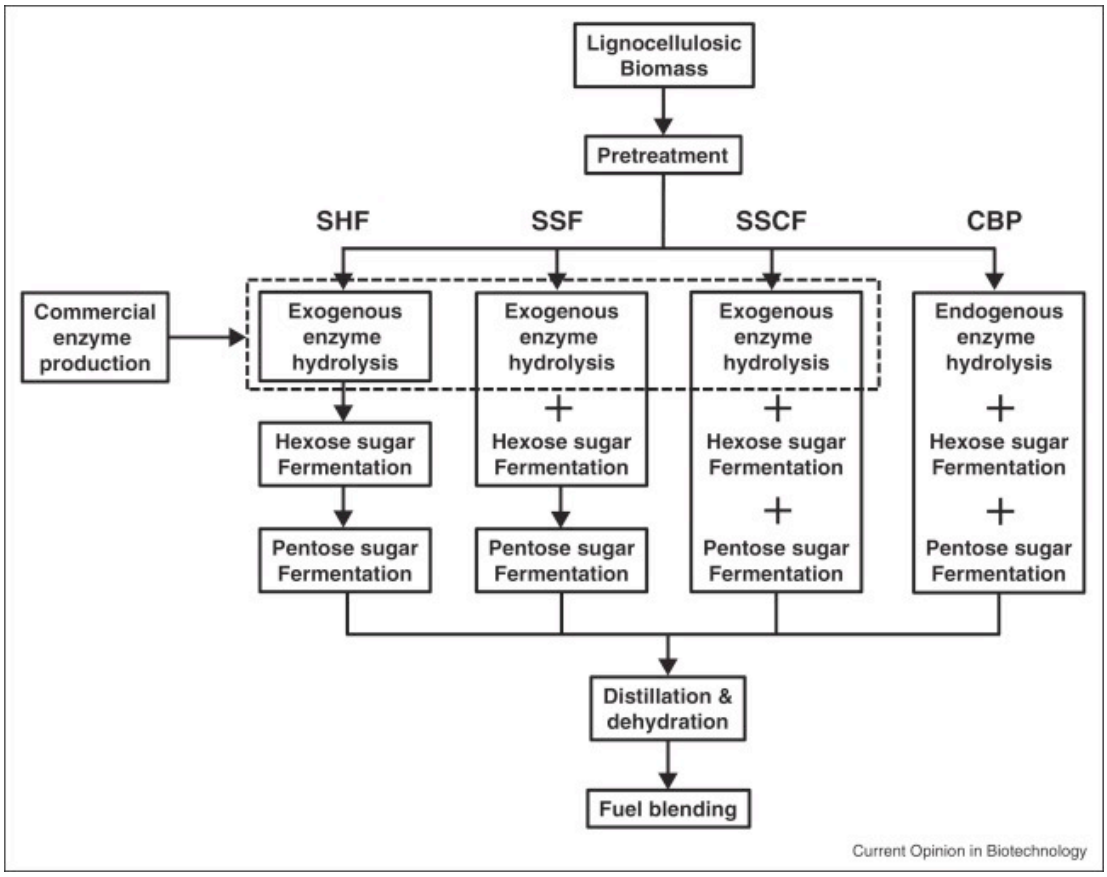


Figure 1.6 – Schematic representation of biomass processing featuring SHF, SSF, SSCF and CBP. Figure is reprinted from (den Haan et al. 2015) with permission.

1.4.2. Simultaneous saccharification and fermentation (SSF)

In simultaneous saccharification and fermentation, exogenous enzymes hydrolyze the pre-treated lignocellulosic biomass in the same reactor and at the same time as the hexose sugars are fermented to ethanol by *S. cerevisiae*. The advantages of SSF over SHF include; SSF limits enzyme feedback inhibition caused by glucose and cellobiose through direct fermentation, and the requirement for two reactors one for pretreatment and one for simultaneous saccharification and fermentation (den Haan et al. 2015).

1.4.3. Simultaneous saccharification and co-fermentation (SSCF)

In SSCF pre-treated biomass is hydrolyzed to glucose and pentose sugars in one reactor and in another reactor either a combination of an *S. cerevisiae* strain to ferment the hexose sugars and a *Zymomonas mobilis* strain to ferment the pentose sugars or an *S. cerevisiae* strain engineered to ferment both hexose and pentose sugars is used to ferment the sugars to ethanol. The advantages of SSCF are that both the hexose and pentose portion of the biomass are used and having enzyme hydrolysis occur in a different reactor than hexose and pentose sugar fermentation, as is the case with SSF, optimizes hydrolysis and fermentation rates when the saccharolytic enzymes and fermentation step are performed at their optimal temperatures (den Haan et al. 2015). A major disadvantage of the SSCF process is the requirement of three separate reactors.

1.4.4. Consolidated bioprocessing (CBP)

In consolidated bioprocessing (CBP), the saccharolytic enzyme production, hydrolysis, and fermentation of hexose sugars, are carried out by a single microbe or a mixed stable culture in a single reactor (Lynd et al. 2002; van Zyl et al. 2007). CBP should dramatically reduce the cost of cellulosic ethanol production by bypassing the need for separate reactors for enzyme production, enzymatic hydrolysis and glucose fermentation

to ethanol, thereby, reducing the capital investment and the equipment associated with a production method requiring three separate reactors (Hasunuma, Ishii, Kondo 2015; van Zyl et al. 2007). The main challenges for CBP technology include the production of sufficient amounts of saccharolytic enzymes without affecting the fermentation capabilities of the CBP organism, co-fermentation of both hexose and pentose sugars, and the tolerance to high ethanol concentration and to residues present due to the pretreatment process (den Haan et al. 2015; Lynd et al. 2005). Several microbes, including the budding yeast *S. cerevisiae,* efficiently ferment glucose and produce ethanol. There are also saprophytic organisms, such as *Trichoderma reesei*, that efficiently hydrolyze cellulose into glucose. Unfortunately, no known organism can both efficiently hydrolyze cellulose and ferment glucose (Lynd et al. 2002).

Theoretically, CBP-competent microorganisms can be developed by three strategies: the native cellulolytic strategy where a naturally occurring cellulolytic microorganism is modified to improve their ethanol yield and titer (Lynd et al. 2005); the recombinant cellulolytic strategy where a microorganism capable of efficiently converting high glucose titers into ethanol is modified so that it expresses the cellulase enzymes needed to degrade cellulose polymers to glucose (Lynd et al. 2005); and the *de novo* strategy where an organism incapable of ethanol production and cellulose hydrolysis is modified into a CBP capable organism.

Native cellulolytic microorganisms that are being developed for CBP include anaerobic bacteria such as *Clostridium thermocellum* (Hasunuma et al. 2013; Jin et al. 2011; Lynd et al. 2005), and fungi such as *Trichoderma reesei* and *Fusarium oxysporum* (van Zyl et al. 2007). Cellulases produced by native cellulolytic organisms exhibit natural synergy (Henrissat et al. 1985) and enzyme levels and ratios are regulated for different substrates, hence, very little, if any, cellulase gene manipulation is required (den Haan et al. 2015). There are also many fungal and bacterial strains that can metabolise hemicellulose that can be chosen for the native cellulolytic strategy; however, these cellulolytic microorganisms produce little to no ethanol and have low tolerance to high ethanol and inhibitor concentrations (den Haan et al. 2015; Lynd et al. 2002). Genetic manipulation of

many desirable cellulolytic organisms to produce high levels of ethanol and improve tolerance to high concentration ethanol and other inhibitors may be difficult because they are not as highly studied as some ethanol producing organisms such as *S. cerevisiae* (den Haan et al. 2015).

1.5. *Saccharomyces cerevisiae* as a CBP Host

The use of yeast in the production of fermented beverages dates as far back as 7000 B.C. (McGovern et al. 2004). Robust *S. cerevisiae* strains are readily available and are extensively used in industry (Demain and Vaishnav 2009). Genetic manipulation of *S. cerevisiae* to express heterologous proteins has been studied for many years. The main challenges in using fermentative strains to develop CBP competent organisms are: 1) secreting high levels of cellulases; 2) expressing appropriate cellulase combinations at ratios suitable for the efficient hydrolysis of different types of pretreated biomass material; 3) expressing hemicellulases for the hydrolysis of the hemicellulose component of pretreated biomass; and 4) the efficient co-fermentation of the resulting pentose sugars (den Haan et al. 2015).

1.5.1. Cellulase expression in *S. cerevisiae*

CBP competent *S. cerevisiae* strains are being developed using three strategies: (i) free enzymes secreted into the media, (ii) enzymes anchored to the cell wall and (iii) enzymes assembled into mini-cellulosomes tethered to the cell wall (Figure 1.7) (den Haan et al. 2015). The free enzyme secretion strategy is limited only by the amount of enzymes being secreted by recombinant yeast and not by the cell surface area (den Haan et al. 2015; Yamada, Hasunuma, Kondo 2013). In a free enzyme secretion system, the enzyme diffusion rate is high, but after the enzymes are secreted, they cannot be recycled (den Haan et al. 2015; Yamada, Hasunuma, Kondo 2013).

Anchoring enzymes to the cell wall directly or assembled into mini-cellulosomes, keeps the enzymes and the released soluble sugars in close proximity to the cells they are

anchored to, providing direct benefit to the cells producing them and allowing the enzymes to be recycled (den Haan et al. 2015). Cell surface display of cellulases improves synergy between the different types of cellulases due to their proximity to each other (Yamada, Hasunuma, Kondo 2013). Potential disadvantages of cell surface display include: the efficiency of processive enzymes are impaired in an immobilized cellulase system thereby impairing crystalline cellulose hydrolysis rates compared to that obtained with a free enzyme system (den Haan et al. 2015); cell surface display of cellulases enzymes can be limited by the cell surface area; and, cellulose hydrolysis rates may also be limited by the low diffusion rates of cellulases due to their reduced mobility when anchored to the cell surface (den Haan et al. 2015; Yamada, Hasunuma, Kondo 2013).

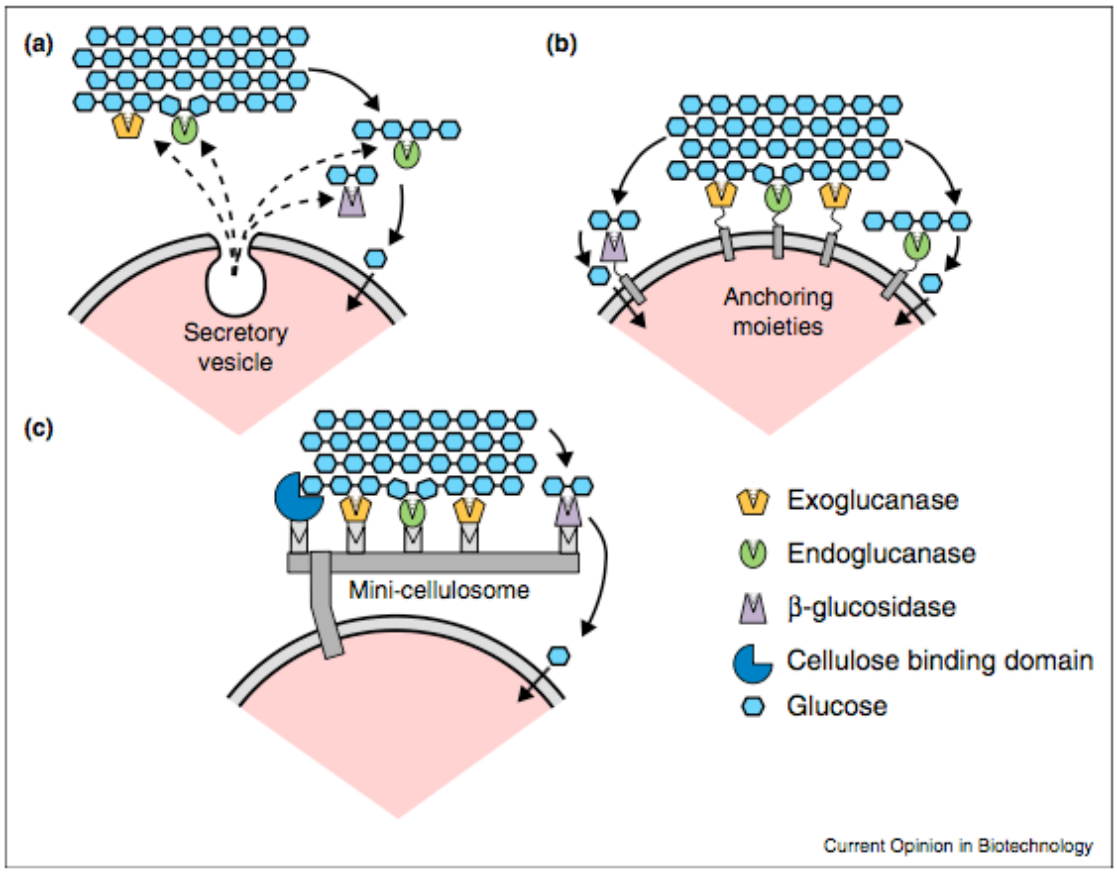


Figure 1.7 – Recombinant cellulolytic strategies. (a) Free enzymes secreted into the media. (b) Enzymes anchored to the cell wall. (c) Enzymes assembled into mini-cellulosomes tethered to the cell wall. Figure is reprinted from (den Haan et al. 2015) with permission.

The complete conversion of insoluble cellulosic substrates by recombinant *S. cerevisiae* strains has not yet been achieved. Recombinant *S. cerevisiae* strains able to grow on and ferment cellobiose at rates comparable to those achieved on glucose have been developed (McBride et al. 2005; van Rooyen et al. 2005; Wilde et al. 2012). To achieve complete hydrolysis of insoluble cellulosic substrates, at least one type of each of the three classes of cellulolytic enzymes have to be functionally expressed as secreted proteins by *S. cerevisiae* (Lynd et al. 2002). Den Haan *et al.* (Den Haan et al. 2007) reported the development of a recombinant *S. cerevisiae* strain co-expressing EG1 from *T. reesei* and BGL1 from *Saccharomycopsis fibuligera* with the ability to grow on and hydrolyze 10 g/l phosphoric acid swollen cellulose (PASC) as the sole carbohydrate source and its one step conversion to ethanol. This study demonstrated that it is possible to use *S. cerevisiae* as a CBP host. Direct ethanol fermentation from PASC was also reported by Yanase *et al.* (Yanase et al. 2010) who demonstrated that the co-expression of CBH and EG improved cellulose degradation by 3- to 5-fold relative to single gene expression.

Other studies reported the expression of cellulases by tethering the enzymes to the cell wall directly (Fujita et al. 2004; Nakatani et al. 2013; Yamada et al. 2010) and by assembly into mini-cellulosomes (Goyal et al. 2011; Tsai, Goyal, Chen 2010; Wen, Sun, Zhao 2010). The heterologous expression of CBHs in high titre has represented a challenge (van Zyl et al. 2007). Ilmen et al. (Ilmen et al. 2011) reported relatively high secretion levels of CBH1 (~0.3 g/L) and CBH2 (~1 g/L) during high cell density fermentations, demonstrating that CBH1 and CBH2 can be expressed at levels where the barrier to CBH sufficiency can eventually be overcome. In general, data reported in the literature on recombinant cellulase expression by *S. cerevisiae* is done under aerobic conditions and high cell densities. These conditions are not representative of industrial CBP conditions, which involve anaerobic cultures and lower cell densities (Olson et al. 2012).

1.5.2. Optimizing cellulase expression in *S. cerevisiae*

When using *S. cerevisiae* as the host for secreted heterologous protein expression the amount of protein produced is generally much lower than that obtained with native proteins (Lambertz et al. 2014). Furthermore, the amount of secreted protein produced can vary dramatically depending upon the heterologous protein (Ilmen et al. 2011). Several approaches have been used to increase the amount of foreign protein produced when using a heterologous host. These approaches include: i) re-synthesizing the gene's coding region so that it is optimized for expression in the heterologous host; ii) using the signal peptide from an efficiently secreted host protein; iii) overexpressing secretory pathway genes that play a role in protein folding, iv) using a strong constitutive promoter to drive the expression of the heterologous cellulase genes; and v) increasing the copy number of the heterologous cellulase encoding genes.

Studies have shown that the expression of a heterologous protein can be enhanced by replacing its native signal peptide with the signal peptide from a yeast protein (Zhu, Yao, Wang 2010). When the native signal peptide of *Trichoderma viride* EG1 was replaced by the *S. cerevisiae* mating factor (*MFα*) prepro-leader sequence the enzyme activity of the heterologous protein increased by 61.5% (Zhu, Yao, Wang 2010). In yeast, there are two known pathways that target proteins to the endoplasmic reticulum, the co-translational pathway and the post-translational pathway (Corsi and Schekman 1996). When proteins are targeted by the co-translational pathway protein translation and translocation across the endoplasmic reticulum membrane occur at the same time (Corsi and Schekman 1996; Osborne, Rapoport, van den Berg 2005). When proteins are targeted by the post-translational translocation pathway protein translocation into the endoplasmic reticulum happens after the protein is completely translated (Corsi and Schekman 1996; Osborne, Rapoport, van den Berg 2005). In yeast, the signal peptides used by the post-translational and the co-translational pathways are different. This distinction may be important because, yeast proteins destined for the extracellular space tend to use the post-translational pathway, which uses signal peptides that are less hydrophobic than the signal peptides used by the co-translational pathway (Corsi and Schekman 1996; Osborne, Rapoport, van den Berg 2005).

Another approach that can be used to improve heterologous protein expression in yeast is re-synthesizing the gene's coding region so that it is optimized for expression in yeast. The degeneracy of the genetic code allows for the synthesis of the same amino acid sequence using many alternative nucleotide sequences. There are 61 codons that encode 20 amino acids and 3 stop codons, with each amino acid being encoded by one to six codons. The use of synonymous codons at different frequencies by different organisms is referred to as codon bias (Davies and Henrissat 1995). Several studies have found that modifying the nucleotide sequence of a foreign gene so that it matches the codon bias of the expression host without changing the protein sequence can increase the expression of the foreign protein (Hillier et al. 2005; Wang et al. 2010). The two most common strategies of modifying the coding sequence for expression in a heterologous host, often referred to as "codon optimization", are the "one amino acid-one codon" strategy where the aim is a high codon adaptation index, and the "codon randomization" strategy using the Monte Carlo algorithm where the codon usage table of an organism is used and codons are randomly assigned based on their frequency distribution in the entire genome or in a set of highly expressed genes of the expression host (Dong et al. 2004). When the codon randomization strategy is used, the codon bias of the optimized sequence will resemble that of the ORF for an average protein or that of the ORF for an average highly expressed protein (Villalobos et al. 2006).

The codon adaptation index (CAI) assesses the degree of bias in the codon usage of a gene (Sharp and Li 1987). The CAI for a gene is calculated by dividing the observed CAI by the maximum possible CAI of a gene with an identical amino acid sequence (Sharp and Li 1987). The relative synonymous codon usage (RSCU) of a codon is calculated by dividing the observed frequency of that codon in a set of highly expressed genes by the frequency expected when synonymous codons are equally used (Sharp and Li 1987). The observed CAI is the geometric mean of the RSCU values of each of the codons used in that gene, and the maximum CAI is the average of the highest possible RSCU values corresponding to each of the codons in that gene (Sharp and Li 1987). Designing a protein-coding region where each codon is substituted by the most frequently used codon for each amino acid in a reference set of highly expressed genes, has a CAI of 1.0.

In general the highly expressed genes of fast growing unicellular organisms tend to have high a CAI, whereas genes that are expressed a lower rates tend to have a low CAI (Welch et al. 2009). Heterologous gDNA and cDNA-derived ORFs are often poorly expressed in common *E. coli* and *S. cerevisiae* laboratory *s*trains. Furthermore, heterologous ORFs generally do not have CAIs similar to that of highly expressed native genes of the expression host. Although resynthesizing the heterologous ORF using only the expression host's high frequently codons, that is maximizing the CAI of an ORF or changing the CAI to that of a typical host gene often significantly improves the gene's expression levels (Gustafsson et al. 2012); however, maximizing the CAI can also have adverse effects on expression.

For example, maximizing the CAI can result in an increased translational error rate (Kerrigan et al. 2008; Sharp and Li 1987). The increase in error rate and reduced mRNA levels are believed to occur because limiting the pool of tRNAs used for translation to just one or two isoacceptors per amino acid causes poor or low levels of tRNA charging (Sharp and Li 1987), and poor growth (Gong, Gong, Yanofsky 2006).

Increased rates of protein translation can also results in increased protein misfolding (Tsai et al. 2008). Considerable evidence from studies mainly with *E. coli* implicate ribosome stalling/pausing as being important for proper protein folding. Mapping studies of codon usage patterns have found that for about 70% of *E. coli* ORFs regions rich in rare synonymous codons are found within short boundary regions separating adjacent regions that encode distinct protein domains. In *E. coli* a rare codon can reduce translation rates as much as 3-fold relative to a frequently used synonymous codon and that the longer the stretch of the boundary region of rare codons is the slower the rate of its translation (Thanaraj and Argos 1996). Experimental evidence also shows that these rare codon boundary regions are important because they facilitate the proper folding of protein domains (reviewed in (Welch et al. 2009). For example, although replacing the rare codons with frequent codons within the boundary regions between the adjacent protein domains in the *E. coli* chloramphenicol acetyltransferase (CAT) gene increased protein expression levels; the CAT specific activity was reduced suggesting the translation rate of the boundary region was important for proper CAT folding (Angov et

al. 2008).

The potential importance of using high frequency codons within protein domain encoding regions and low frequency codons within the boundary regions of a fast growing unicellular organism such as *E. coli* and *S. cerevisiae* is unlikely to be repeated for multicellular eukaryotes since their high frequency codons are translated at the same rate as low frequency codons (Pop et al. 2014; Qian et al. 2012; Sharp and Li 1986).

When the codon randomization strategy is used, the CAI will resemble the CAI of an average gene, which is less than one. In this case, the pool of required tRNAs for translation of the engineered gene will be balanced and not limited to just one or two isoacceptor tRNAs per amino acid. The flexibility of the codon randomization strategy allows for other manipulations of the DNA coding regions such as avoiding or including certain restriction sites, avoiding secondary mRNA structures, and avoiding repetitive elements (Villalobos et al. 2006).

1.6. Thesis objectives

The objectives of my Ph.D. research are to contribute to the development of consolidated bioprocessing *S. cerevisiae* strains by: i) identifying fungal endoglucanases that can be functionally expressed in *S. cerevisiae*; ii) optimizing the expression of fungal endoglucanases in *S. cerevisiae*; iii) co-expressing selected endoglucanases with a functionally expressed β-glucosidase; and iv) identifying fungal cellobiohydrolases that can be functionally expressed in *S. cerevisiae* in order to be potentially expressed with other cellulases in *S. cerevisiae* as part of a cellulase system.

Materials and Methods

2.1. Materials

Carboxymethyl cellulose (CMC-4M) was purchased from Megazyme (Wicklow, Ireland). Cellobiose was purchased from BioShop (Canada). Congo Red, Bovine Serum Albumin (BSA), xylose, and CMC-7M were purchased from Sigma-Aldrich (Oakville, Ontario).

Phosphoric acid swollen cellulose (PASC) was prepared as previously described (Wood 1988) with the following modifications. Twenty grams of microcrystalline cellulose, Avicel PH105 (Sigma-Aldrich), was suspended in 600 ml of 85% phosphoric acid in the fume hood in an ice bath for 4 hours with occasional grinding. The swollen cellulose was washed with ice-cold water until the pH was between 5 and 7. The final wash and suspension was done in 10 mM citrate buffer pH 5.0. The cellulose mixture was blended in a Waring blender to remove any lumps and to homogenize the mixture. The final PASC concentration was estimated by weighing a speed vac dried sample.

2.2. Strains, Plasmids, and Media

The strains used in this study are listed in Table 2.1. *Escherichia coli* strain DH5α was used for the propagation of plasmids and was cultivated in LB media supplemented with 100 µg/mL ampicillin at 37^{o}C. *S. cerevisiae* strain CEN.PK 111-61A (constructed by Dr. P. Koetter, Frankfurt, Germany) was used for expression of the β-glucosidase BGL1 gene ORF, the EG ORFs and the CBH ORFs, Strain CEN.PK 111-61A is referred to herein as the wild type. *E. coli* and *S. cerevisiae* transformations were performed as previously described (Hanahan 1983; Gietz et al. 1995). The plasmids used in this study are listed in Table 2.2. *S. cerevisiae* plasmids p425-TEF and p416-TEF (Mumberg, Muller, Funk 1995) were used to express fungal cellulases genes and derivatives thereof. Plasmid maps were generated using Clone Manager software (Scientific and Educational Software, Denver, CO).

2.2.1. Construction of p425-TEF_M

The 2-micron plasmid p425-TEF_M (Figure 2.1) was derived from p425-TEF by introducing the restriction endonuclease sites *Nhe*I, *Pac*I, *Asc*I, *Mre*I, *Pme*I and *Fse*I. These restriction endonuclease sites were introduced into p425-TEF by directionally cloning an adaptor containing these sites into the backbone of p425-TEF generated by *Spe*I and *Xho*I digestion. Plasmid p425-TEF_M directs the transcription of target gene open reading frames (ORFs) cloned into the multiple cloning site (MCS) of p425-TEF_M using the strong constitutive TEF1 promoter.

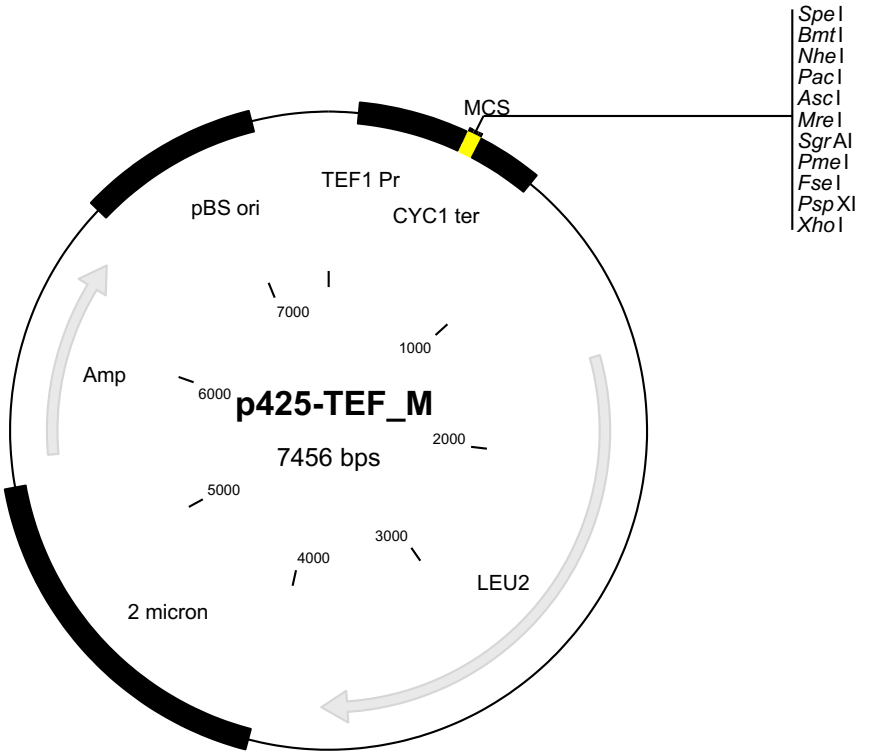


Figure 2.1 – Yeast expression vector p425-TEF_M. p425-TEF_M was derived from p425-TEF (Mumberg, Muller, Funk 1995) by introducing a new multiple cloning site.

2.2.2. Construction of p416TEF-MFα-prepro

Plasmid p416TEF-MFα-prepro (Figure 2.2) was derived from p416-TEF by cloning the MFα-prepro sequence (Kurjan and Herskowitz 1982) into p416-TEF by gap-repair (Orr-Weaver and Szostak 1983). The MFα-prepro sequence was PCR amplified from yeast genomic DNA using primers MFα_F:NheI (AGAATGCTAGCATAATGAGATTTCCTTCAATTTTTACTGC) and MFα_R:FseI (AGACTAGGCCGGCCTCTTTTATCCAAAGATACCCCTTC) then directionally cloned into the *Nhe*I and *Fse*I sites of p425-TEF_M backbone. The MFα-prepro sequence and a portion of the TEF1 promoter sequence at the 5'-end and a portion of the CYC1 transcription termination sequence at the 3'-end were PCR amplified using the p425-TEF_M derivative with the MFα-prepro insert as a template and primers TEF1_F (CTTCAAAACACCCAAGCACAG) and CYC1_R (GCCGCAAATTAAAGCCTTCG). CEN.PK 111-61A was then transformed with the PCR generated MFα-prepro sequence and p416-TEF backbone prepared by digestion with the restriction endonuclease *Xho*I followed by selection for uracil prototrophs.

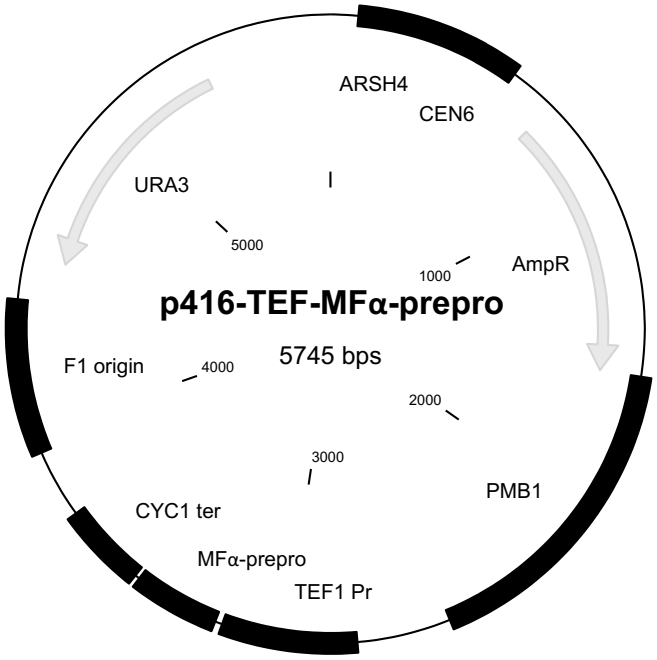


Figure 2.2 – Yeast centromere plasmid p416-TEF-MFα-prepro. p416-TEF-MFα-prepro was derived from p416-TEF (Mumberg, Muller, Funk 1995) by introducing the MFα-prepro sequence downstream of the TEF1 promoter by gap-repair (Orr-Weaver and Szostak 1983).

2.2.3. Construction of plasmids p425-TEF_M_Δ2μ and p425-TEF_M_Δ2μ _δ

The 2-micron sequence was removed from p425-TEF_M by PCR amplifying the rest of the plasmid using 2 micron_NcoI_F and 2micron_NcoI_R primers (Table 2.1). The PCR product was digested with *Nco*I followed by ligation and transformation into *E. coli*. The resulting plasmid, verified by DNA sequencing, was designated p425-TEF_M_Δ2μ.

CEN.PK111-61A genomic DNA was used as the template to PCR amplify the 5'-half (167 nucleotides) and 3'-half (167 nucleotides) of YDRWdelta27 sequence on chromosome IV from coordinates 1206704 to 1207037. 5'δTEF_F1 and 5'δTEF_R1 were used to amplify the 5'-half of YDRWdelta27, and 3'δLEU2_F1 and 3'δLEU2_R1 were used to amplify the 3'-half of YDRWdelta27. The resulting PCR product of the 5'-half of YDRWdelta27 is flanked on each end by 40 nucleotides of homology to p425-TEF_M immediately adjacent to the end of the TEF promoter near the pBS origin of replication. The resulting PCR product of the 3'-half of YDRWdelta27 is flanked by 40 nucleotides of homology on one end and 31 nucleotides of homology on the other end to p425-TEF_M immediately adjacent to the Leu2 marker near the Ampicillin resistance gene. The 5'-half and the 3'-half of YDRWdelta27 were sequentially inserted into the p425-TEF_M_Δ2μ plasmid. The 5'-half of YDRWdelta27 was inserted into p425-TEF_M_Δ2μ by overlap-extension PCR (Bryksin and Matsumura 2010) followed by digestion with 10 units of *Dpn*I (NEB, Cat# R0176S) at 37^{o}C for one hour followed heat inactivation at 80^{o}C for 20 min. The *Dpn*I digested overlap-extension PCR product was then transformed into *E. coli*. The 3'-half of YDRWdelta27 was inserted into p425-TEF_M_Δ2μ derivative with the 5'-delta inserted by overlap-extension PCR followed by digestion with 10 units of *Dpn*1 at 37^{o}C for one hour followed by heat inactivation at 80^{o}C for 20 min. The resulting plasmid, referred to as p425-TEF_M_Δ2μ _δ, was verified by restriction endonuclease mapping and DNA sequencing.

Table 2.1 – Primers used for the construction of plasmids p425-TEF_M_Δ2μ and p425-TEF_M_Δ2μ_+δ

Primer	5' to 3' primer sequences
2micron_NcoI_F[a]	AGAATCCATGGATAGGAACCCCTATTTGTTTATTTTTCTAAATACATTC
2micron_NcoI_R[a]	AGAATCCATGGATACCTTTGATATTGGATCATACTAAGAAACCAT
5′δTEF_F1[b]	TATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACATGTTGGAATAGAAATCAACT
5′δTEF_R1[c]	GAAACATTTTGAAGCTATGAGCTCCAGCTTTTGTTCCCTTATGTTTATATTCATTGATCCTATTAC
3′δLEU2_F1[d]	CTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAATAAAATGATGATAATAATATTTATAGAATTGTGTAGAA
3′δLEU2_R1[e]	GCTCCCGGAGACGGTCACAGCTTGTCTGTAATGAGAAATGGGTGAATGTTG

[a] The underlined sequences in primers 2micron_NcoI_F and 2micron_NcoI_R are homologous to p425-TEF_M immediately near the ends of the 2-micron sequence. The 5' ends of 2micron_NcoI_F and 2micron_NcoI_R have 5 filler nucleotides followed by the *Nco*I restriction endonuclease site followed by 3 filler nucleotides.

[b] The underlined sequence in the forward primer 5′δTEF_F1 is identical to the first 20 bases of the delta element YDRWdelta27. The 5′ end of the forward primer 5′δTEF_F1 has 40 nucleotides of homology to p425-TEF_M immediately adjacent to the end of the TEF promoter near the pBS origin of replication.

[c] The underlined sequence in the reverse primer 5′δTEF_R1 is homologous to 26 bases of the delta element YDRWdelta27 ending at the 167th nucleotide. The 5' end of the reverse primer 5′δTEF_R1 has 40 nucleotides of homology to p425-TEF_M immediately adjacent to the end of the TEF promoter near the pBS origin of replication.

[d] The underlined sequence in the forward primer 3′δLEU2_F1 is identical to the first 39 bases of the 3' half of the delta element YDRWdelta27. The 5′ end of 3′δLEU2_F1 has 40 nucleotides of homology to p425-TEF_M immediately adjacent to the Leu2 marker near the Ampicillin resistance gene.

[e] The underlined sequence in the reverse primer 3′δLEU2_R1 is homologous to the last 20 bases of the delta element YDRWdelta27. The 5' end of 3′δLEU2_R1 has 31 nucleotides of homology to p425-TEF_M immediately adjacent to the Leu2 marker near the Ampicillin resistance gene.

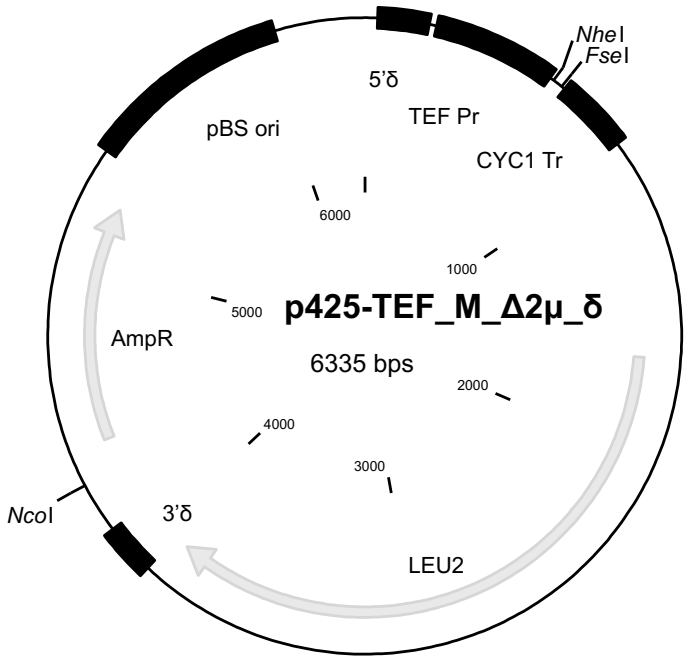


Figure 2.3 Yeast delta-integration plasmid p425-TEF_M_Δ2μ_δ was derived from p425-TEF_M by deleting the 2-micron sequence and introducing the 5'-half of YDRWdelta27 sequence upstream of the TEF1 promoter and the 3'-half of YDRWdelta27 sequence downstream of the LEU2 selection marker.

2.2.4. *S. cerevisiae* strains and *S. cerevisiae* plasmids harbouring cDNA derived fungal glycosylhydrolases

The *S. cerevisiae* strains and plasmids harbouring cDNA derived fungal glycosylhydrolases used in this study are listed in Tables 2.2 and 2.3.

Table 2.2 – Plasmids used in this study and their relevant features

Plasmid	Relevant features	Source
p425-TEF	2 micron, LEU2, AmpR, TEF Pr, CYC1 Tr	Mumberg et al. 1995
p416-TEF	Cen6/ARSH4, URA3, AmpR, TEF Pr, CYC1 Tr	Mumberg et al. 1995
p425-TEF_M	p425-TEF with added MCS *Nhe*I, *Pac*I, *Asc*I, *Mre*I, *Pme*I and *Fse*I	This Study

p425-TEF_M-AfCel7B1	p425-TEF_M with added EG *AfCel7B1*	This Study
p425-TEF_M-AfCel7B2	p425-TEF_M with added EG *AfCel7B2*	This Study
p425-TEF_M-AfCel5A1	p425-TEF_M with added EG *AfCel5A1*	This Study
p425-TEF_M-AfCel5A2	p425-TEF_M with added EG *AfCel5A2*	This Study
p425-TEF_M-AfCel5A3	p425-TEF_M with added EG *AfCel5A3*	This Study
p425-TEF_M-AnCel7B	p425-TEF_M with added EG *AnCel7B*	This Study
p425-TEF_M-AnCel5A1	p425-TEF_M with added EG *AnCel5A1*	This Study
p425-TEF_M-AnCel5A2	p425-TEF_M with added EG *AnCel5A2*	This Study
p425-TEF_M-ApCel5A	p425-TEF_M with added EG *ApCel5A*	This Study
p425-TEF_M-FgCel7B1	p425-TEF_M with added EG *FgCel7B1*	This Study
p425-TEF_M-FgCel7B2	p425-TEF_M with added EG *FgCel7B2*	This Study
p425-TEF_M-FgCel5A1	p425-TEF_M with added EG *FgCel5A1*	This Study
p425-TEF_M-FgCel5A2	p425-TEF_M with added EG *FgCel5A2*	This Study
p425-TEF_M-GtCel12A	p425-TEF_M with added EG *GtCel12A*	This Study
p425-TEF_M-NcCel7B1	p425-TEF_M with added EG *NcCel7B1*	This Study
p425-TEF_M-NcCel7B2	p425-TEF_M with added EG *NcCel7B2*	This Study
p425-TEF_M-NcCel5A	p425-TEF_M with added EG *NcCel5A*	This Study
p425-TEF_M-PsCel12A1	p425-TEF_M with added EG *PsCel12A1*	This Study
p425-TEF_M-PsCel12A2	p425-TEF_M with added EG *PsCel12A2*	This Study
p425-TEF_M-PsCel12A3	p425-TEF_M with added EG *PsCel12A3*	This Study
p425-TEF_M-StCel5A	p425-TEF_M with added EG *StCel5A*	This Study
p425-TEF_M-TrCel7B	p425-TEF_M with added EG *TrCel7B*	This Study
p425-TEF_M-TrCel5A	p425-TEF_M with added EG *TrCel5A*	This Study
p425-TEF_M-TrCel12A	p425-TEF_M with added EG *TrCel12A*	This Study
p425-TEF_M-AfCel7B1opt	p425-TEF_M with added codon-optimized EG *AfCel7B1*	This Study

p425-TEF_M-GtCel12Aopt	p425-TEF_M with added codon-optimized EG *GtCel12A*	This Study
p425-TEF_M-StCel5Aopt	p425-TEF_M with added codon-optimized EG *StCel5A*	This Study
p425-TEF_M-ApCel5Aopt	p425-TEF_M with added codon-optimized EG *ApCel5A*	This Study
p425-TEF_M-AfCel5A1opt	p425-TEF_M with added codon-optimized EG *AfCel5A1*	This Study
p425-TEF_M-preAfCel7B1opt	p425-TEF_M with added codon-optimized EG *AfCel7B1*, Mfα pre signal peptide	This Study
p425-TEF_M-preproAfCel7B1opt	p425-TEF_M with added codon-optimized EG *AfCel7B1*, Mfα prepro signal peptide	This Study
p425-TEF_M-preGtCel12Aopt	p425-TEF_M with added codon-optimized EG *GtCel12A*, Mfα pre signal peptide	This Study
p425-TEF_M-preproGtCel12Aopt	p425-TEF_M with added codon-optimized EG *GtCel12A*, Mfα prepro signal peptide	This Study
p425-TEF_M-preStCel5Aopt	p425-TEF_M with added codon-optimized EG *StCel5A*, Mfα pre signal peptide	This Study
p425-TEF_M-preproStCel5Aopt	p425-TEF_M with added codon-optimized EG *StCel5A*, Mfα prepro signal peptide	This Study
p425-TEF_M-preApCel5Aopt	p425-TEF_M with added codon-optimized EG *ApCel5A*, Mfα pre signal peptide	This Study
p425-TEF_M-preproApCel5Aopt	p425-TEF_M with added codon-optimized EG *ApCel5A*, Mfα prepro signal peptide	This Study
p425-TEF_M-preAfCel5A-1opt	p425-TEF_M with added codon-optimized EG *AfCel5A1*, Mfα pre signal peptide	This Study
p425-TEF_M-preproAfCel5A-1opt	p425-TEF_M with added codon-optimized EG *AfCel5A1*, Mfα prepro signal peptide	This Study
p425-TEF_M-AnBGL1	p425-TEF_M with added BGL *AnBGL1*	This Study
p416-TEF_Matα prepro	Insert: TEF Pr-NheI-Mfα prepro-FseI-CYC1 Tr	This Study

p425-TEF_M_Δ2μ_δ	p425-TEF_M Δ2μ, Insert: 5' δ and 3'δ	This Study
p425-TEF_M_Δ2μ_δ AfCel7B1opt	p425-TEF_M_Δ2μ_δ with added codon-optimized EG *AfCel7B1*	This Study
p425-TEF_M_Δ2μ_δ_ GtCel12Aopt	p425-TEF_M_Δ2μ_δ with added codon-optimized EG *GtCel12A*	This Study
p425-TEF_M_Δ2μ_δ_ StCel5Aopt	p425-TEF_M_Δ2μ_δ with added codon-optimized EG *StCel5A*	This Study
p425-TEF_M_Δ2μ_δ_ preApCel5Aopt	p425-TEF_M_Δ2μ_δ with added codon-optimized EG *ApCel5A*, Mfα prepro signal peptide	This Study
p425-TEF_M_Δ2μ_δ_ preproAfCel5A1opt	p425-TEF_M_Δ2μ_δ with added codon-optimized EG *AfCel5A1*, Mfα prepro signal peptide	This Study
p425-TEF_M-AnCel6A	p425-TEF_M with added CBH *AnCel6A*	This Study
p425-TEF_M-LeCel6A	p425-TEF_M with added CBH *LeCel6A*	This Study
p425-TEF_M-NcCel7A	p425-TEF_M with added CBH *NcCel7A*	This Study
p425-TEF_M-NcCel6A	p425-TEF_M with added CBH *NcCel6A*	This Study
p425-TEF_M-StCel7A	p425-TEF_M with added CBH *StCel7A*	This Study
p425-TEF_M-TrCel7A	p425-TEF_M with added CBH *TrCel7A*	This Study
p425-TEF_M-TrCel6A	p425-TEF_M with added CBH *TrCel6A*	This Study

Table 2.3 – *S. cerevisiae* strains used in this study and their relevant features

S. cerevisiae strains	Relevant features	Source
CEN.PK111-61A	MATα ura3–52 leu2–112 his3-Δ1	Dr. P. Koetter, Frankfurt, Germany
CEN.PK111-61A_p425-TEF_M	ura$^-$ his$^-$	This Study
CEN.PK111-61A_p425-TEF_M-AfCel7B1	ura$^-$ his$^-$, plasmid-borne EG *AfCel7B1*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel7B2	ura$^-$ his$^-$, plasmid-borne EG *AfCel7B2*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel5A1	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A1*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel5A2	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A2*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel5A3	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A3*	This Study
CEN.PK111-61A_p425-TEF_M-AnCel7B	ura$^-$ his$^-$, plasmid-borne EG *AnCel7B*	This Study
CEN.PK111-61A_p425-TEF_M-AnCel5A1	ura$^-$ his$^-$, plasmid-borne EG *AnCel5A1*	This Study
CEN.PK111-61A_p425-TEF_M-AnCel5A2	ura$^-$ his$^-$, plasmid-borne EG *AnCel5A2*	This Study
CEN.PK111-61A_p425-TEF_M-ApCel5A	ura$^-$ his$^-$, plasmid-borne EG *ApCel5A*	This Study
CEN.PK111-61A_p425-TEF_M-FgCel7B1	ura$^-$ his$^-$, plasmid-borne EG *FgCel7B1*	This Study
CEN.PK111-61A_p425-TEF_M-FgCel7B2	ura$^-$ his$^-$, plasmid-borne EG *FgCel7B2*	This Study
CEN.PK111-61A_p425-	ura$^-$ his$^-$, plasmid-borne EG *FgCel5A1*	This Study

TEF_M-FgCel5A1		
CEN.PK111-61A_p425-TEF_M-FgCel5A2	ura$^-$ his$^-$, plasmid-borne EG *FgCel5A2*	This Study
CEN.PK111-61A_p425-TEF_M-GtCel12A	ura$^-$ his$^-$, plasmid-borne EG *GtCel12A*	This Study
CEN.PK111-61A_p425-TEF_M-NcCel7B1	ura$^-$ his$^-$, plasmid-borne EG *NcCel7B1*	This Study
CEN.PK111-61A_p425-TEF_M-NcCel7B2	ura$^-$ his$^-$, plasmid-borne EG *NcCel7B2*	This Study
CEN.PK111-61A_p425-TEF_M-NcCel5A	ura$^-$ his$^-$, plasmid-borne EG *NcCel5A*	This Study
CEN.PK111-61A_p425-TEF_M-PsCel12A1	ura$^-$ his$^-$, plasmid-borne EG *PsCel12A1*	This Study
CEN.PK111-61A_p425-TEF_M-PsCel12A2	ura$^-$ his$^-$, plasmid-borne EG *PsCel12A2*	This Study
CEN.PK111-61A_p425-TEF_M-PsCel12A2	ura$^-$ his$^-$, plasmid-borne EG *PsCel12A3*	This Study
CEN.PK111-61A_p425-TEF_M-StCel5A	ura$^-$ his$^-$, plasmid-borne EG *StCel5A*	This Study
CEN.PK111-61A_p425-TEF_M-TrCel7B	ura$^-$ his$^-$, plasmid-borne EG *TrCel7B*	This Study
CEN.PK111-61A_p425-TEF_M-TrCel5A	ura$^-$ his$^-$, plasmid-borne EG *TrCel5A*	This Study
CEN.PK111-61A_p425-TEF_M-TrCel12A	ura$^-$ his$^-$, plasmid-borne EG *TrCel12A*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel7B1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel7B1opt*	This Study
CEN.PK111-61A_p425-TEF_M-GtCel12Aopt	ura$^-$ his$^-$, plasmid-borne EG *GtCel12Aopt*	This Study
CEN.PK111-61A_p425-TEF_M-StCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *StCel5Aopt*	This Study

CEN.PK111-61A_p425-TEF_M-ApCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *ApCel5Aopt*	This Study
CEN.PK111-61A_p425-TEF_M-AfCel5A1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A1opt*	This Study
CEN.PK111-61A_p425-TEF_M-preAfCel7B1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel7B1opt*, Mfα pre signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preproAfCel7B1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel7B1opt*, Mfα prepro signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preGtCel12Aopt	ura$^-$ his$^-$, plasmid-borne EG *GtCel12Aopt*, Mfα pre signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preproGtCel12Aopt	ura$^-$ his$^-$, plasmid-borne EG *GtCel12Aopt*, Mfα prepro signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preStCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *StCel5Aopt*, Mfα pre signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preproStCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *StCel5Aopt*, Mfα prepro signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preApCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *ApCel5Aopt*, Mfα pre signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preproApCel5Aopt	ura$^-$ his$^-$, plasmid-borne EG *ApCel5Aopt*, Mfα prepro signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preAfCel5A-1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A1opt*, Mfα pre signal peptide	This Study
CEN.PK111-61A_p425-TEF_M-preproAfCel5A1opt	ura$^-$ his$^-$, plasmid-borne EG *AfCel5A1opt*, Mfα prepro signal peptide	This Study
CEN.PK111-61A-δ- AnBGL1a	ura$^-$ leu$^-$ his$^-$, δ-integrated *AnBGL1*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M-AfCel7B1opt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, plasmid-borne EG *AfCel7B1opt*	This Study
CEN.PK111-61A-δ- AnBGL1a	ura$^-$ his$^-$, δ-integrated *AnBGL1*,	This Study

_p425-TEF_M-GtCel12Aopt	plasmid-borne EG *GtCel12Aopt*	
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M-StCel5Aopt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, plasmid-borne EG *StCel5Aopt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M-preApCel5Aopt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, plasmid-borne EG *preApCel5Aopt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M-preproAfCel5A1opt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, plasmid-borne EG *preproAfCel5A1opt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M	ura$^-$ his$^-$, δ-integrated *AnBGL1*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M_Δ2μ_δ-AfCel7B1opt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, δ-integrated EG *AfCel7B1opt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M_Δ2μ_δ-GtCel12Aopt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, δ-integrated EG *GtCel12Aopt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M_Δ2μ_δ-StCel5Aopt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, δ-integrated EG *StCel5Aopt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M_Δ2μ_δ-preApCel5Aopt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, δ-integrated EG *preApCel5Aopt*	This Study
CEN.PK111-61A-δ- AnBGL1a _p425-TEF_M_Δ2μ_δ-preproAfCel5A1opt	ura$^-$ his$^-$, δ-integrated *AnBGL1*, δ-integrated EG *preproAfCel5A1opt*	This Study
CEN.PK111-61A-δ-AnBGL1a _p425-TEF_M_Δ2μ_δ	ura$^-$ his$^-$, δ-integrated *AnBGL1*	This Study
CEN.PK111-61A_p425-TEF_M-AnCel6A	ura$^-$ his$^-$, plasmid-borne CBH *AnCel6A*	This Study
CEN.PK111-61A_p425-TEF_M-LeCel6A	ura$^-$ his$^-$, plasmid-borne CBH *LeCel6A*	This Study
CEN.PK111-61A_p425-TEF_M-NcCel7A	ura$^-$ his$^-$, plasmid-borne CBH *NcCel7A*	This Study

CEN.PK111-61A_p425-TEF_M-NcCel6A	ura$^-$ his$^-$, plasmid-borne CBH *NcCel6A*	This Study
CEN.PK111-61A_p425-TEF_M-StCel7A	ura$^-$ his$^-$, plasmid-borne CBH *StCel7A*	This Study
CEN.PK111-61A_p425-TEF_M-TrCel7A	ura$^-$ his$^-$, plasmid-borne CBH *TrCel7A*	This Study
CEN.PK111-61A_p425-TEF_M-TrCel6A	ura$^-$ his$^-$, plasmid-borne CBH *TrCel6A*	This Study

2.3. Methods

2.3.1. Bioinformatic analysis and cellulase gene annotation

The ORFs of the glycosyl hydrolases cloned in this study were identified by searching the genomes of 12 evolutionary diverse fungi (Table 2.4) for EGs (EG1s, EG2s, and EG3s) using cDNA sequences of *T. reesei* (EG1, EG2, and EG3) and the Basic Local Alignment Search Tool "tblastn" (Altschul et al. 1990). Proteins encoded by the ORFs were annotated using SignalP 4.0 (Petersen et al. 2011) to identify their signal peptides, and their conserved domains, including catalytic domains and carbohydrate binding modules (CBM), using the Conserved Domain Database (CDD) (Marchler-Bauer et al. 2015).

2.3.2. RNA isolation and cDNA synthesis

Fungal spores were streaked out on complete media (CM) plates, prepared as described previously (Kafer 1977), and grown for 2 to 3 days at the temperatures described in Table 2.4. The mycelia was collected from plates using 5 ml of saline tween solution and inoculated into 250 ml Erlenmeyer flasks with 50 ml of liquid media as indicated in Table 2.4 to induce cellulase gene expression. The fungal cultures were grown for 2 to 3 days, shaking at 200 rpm, at the temperatures indicated in Table 2.4. The mycelia in the liquid cultures were recovered by filtration over Miracloth using a Millipore filtration unit, with suction. About 100 mg of the recovered mycelial mass was frozen in liquid nitrogen. The frozen mycelia were then disrupted and homogenized in two steps, first, using a mortar and pestle then using a QIAshredder homogenizer (Qiagen) following the instructions provided by the manufacturer. Total fungal RNA was extracted using the RNeasy Plant Minikit (Qiagen) and the method provided by the supplier. The extracted RNA solution was subjected to on-column DNase digestion using the RNase-Free DNase Set (Qiagen) followed by the RNA clean-up step using the RNeasy Plant Minikit (Qiagen) as recommended by the supplier. All the surfaces were prepared by wiping with RNase away surface decontaminant (Cole-Parmer) prior to RNA extraction. First-strand cDNA was synthesized using Superscript II reverse transcriptase (Invitrogen) and oligo

(dT) 12-18 primer mix. The synthesized cDNAs were used as PCR templates to amplify the EG, CBH, and BGL ORFs.

Table 2.4 – Growth conditions used for the fungal strains

Fungal Species	Source	Growth Conditions
Aspergillus fumigatus	FGSC A1100	V8 medium[a], 37°C, 150 rpm
Aspergillus nidulans	FGSC A4	MM[b] + 1% CMC-7M, 37°C, 150 rpm
Aspergillus niger	FGSC A733	MM[b] + 1% xylose, 37°C, 150 rpm
Aureobasidium pullulans	ATCC 62921	(Semova et al. 2006)
Fusarium graminearum	FGSC 9075	MM[b] + 1% Avicel, 25°C, 150 rpm
Gloeophyllum trabeum	ATCC 11539	(Semova et al. 2006)
Lentinula edodes	ATCC 48564	(Semova et al. 2006)
Nectria haematococca	FGSC 9596	V8 medium[a], 25°C, 150 rpm
Neurospora crassa	FGSC 2489	MM[b] + 1% CMC-7M, 24°C, 150 rpm
Phytophthora sojae	ATCC MYA-4756	V8 medium[a], RT, 150 rpm
Sporotrichum thermophile	ATCC 42464	(Semova et al. 2006)
Trichoderma reesei	ATCC 26921	MM[b] + 1% Avicel, 25°C, 200 rpm

[a] V8 juice medium was prepared as described previously (Miller 1955) except that 2% agar was omitted and calcium carbonate was used at 2 grams per liter.

[b] Minimal media was prepared as described previously (Kafer 1977).

2.3.3. Cellulase cloning

The 33 native ORF DNAs encoding 25 endoglucanases, 7 cellobiohydrolases, and one β-glucosidase, used herein were prepared by PCR from the cDNA templates described above (2.3.2) using Phusion high-fidelity DNA polymerase (NEB) and the primer pairs listed in Tables 2.5 and 2.6. The cDNA derived ORF were then directionally ligated into the backbone of p425-TEF_M (Figure 2.1) following digestion of p425-TEF_M and the ORF DNAs with *Nhe*I and *Fse*I.

The ORFs of several endoglucanases were codon optimized for expression in yeast as described in section 2.3.5. The optimized ORF sequences were synthesized by Genscript® (New Jersey, United States). The optimized ORFs, which were flanked at their 5' and 3' ends by *Nhe*I and *Fse*I sites and p425-TEF_M were prepared by digestion with *Nhe*I and *Fse*I, followed by enzyme inactivation, followed by DNA ligation and transformation into *E. coli*. The resulting recombinant plasmids (Table 2.2) were verified by restriction analysis and sequencing the region immediately upstream and downstream of the *Nhe*I and *Fse*I cloning sites.

Two derivatives of each of the codon-optimized endoglucanases were also constructed (Table 2.2). These derivatives had their signal peptide coding portions replaced with the coding region for either the *S. cerevisiae* MFα-prepro signal sequence or the MFα-pre signal sequence as described in section 2.3.6.

The following system was used to generate unique recombinant gene and protein designations. The two letters at the beginning of each designator correspond the genus and species names of the fungus that encodes the protein (e.g. Af indicates an *Aspergillus fumigatus* protein). The next three letters, for example Cel, indicates that the protein is predicted to be a cellulase. Next is a 1, 2 or 3 digit number (e.g. 7 or 125 would indicate the protein belongs to GH family 7 or 125). Finally, sometimes the gene designator is followed by a capital letter (e.g. an A, B, C etc.) to indicate the protein belongs to a particular GH subfamily A, B, C etc. If the GH subfamily is followed by a number (e.g.

3) this indicates that the designated protein was the third protein of this designation that was identified in a particular fungal species. If the ORF has been codon optimized, “opt” is added to the enzyme designator and if the native signal peptide is replaced by the *S. cerevisiae* MFα-pre or the MFα-prepro signal peptide, the designation “pre” or “prepro” is added.

Table 2.5 – Primers used for endoglucanase ORF amplification

Fungus	EG	CBM	Primer	5' to 3' primer sequences*
Aspergillus fumigatus	*AfCel7B1*	C-terminal	AfCel7B1-F	AGAATGCTAGCATAATGGACTCCAAAAGAGGCGTC
			AfCel7B1-R	AGACTAGGCCGGCCCTACAGACACTGAGAGTACCACG
	AfCel7B2	N/A	AfCel7B2-F	AGAATGCTAGCATAATGGCTCAAACACTGGCAGC
			AfCe l7B2-R	AGACTAGGCCGGCCCTAAACAGAGTAGGTACTGTCAATATCCCC
	AfCel5A1	C-terminal	AfCel5A1-F	AGAATGCTAGCATAATGAAGGCTTCCACTATTATCTGC
			AfCel5A1-R	AGACTAGGCCGGCCCTACAGGCATTGAGAGTAGTAGTC
	AfCel5A2	C-terminal	AfCel5A2-F	GCTCTAGAATGAGAATCAGCAGCTTGATC
			AfCel5A2-R	AGACTAGGCCGGCCTTAAACACACTGGTGGTAGTAAG
	AfCel5A3	N/A	AfCel5A3-F	AGAATGCTAGCATAATGAAATTCGGTAGCATTGTGCTC

			AfCel5A3-R	AGACTAGGCCGGCCTCAACCCAGGTAGGGCTC
Aspergillus nidulans	*AnCel7B*	N/A	AnCel7B-F	AGAATGCTAGCATAATGGCTCTGTTACTATCTCTCAGCCTTC
			AnCel7B-R	AGACTAGGCCGGCCCTAAGACGCCTGGAAAGTACTGCC
	AnCel5A1	N/A	AnCel5A1-F	AGAATGCTAGCATAATGAGGTCTCTCGTCCTTCTGTC
			AnCel5A1-R	AGACTAGGCCGGCCTTATTGACTTCCCACGAAATACGGC
	AnCel5A2	N/A	AnCel5A2-F	AGAATGCTAGCATAATGAAGGTCAACACTCTTTTGGTTGCC
			AnCel5A2-R	AGACTAGGCCGGCCTTAATTCATGTACGCCTCCAACGTACTG
Aureobasidium pullulans	*ApCel5A*	N-terminal	ApCel5A-F	AGAATGCTAGCATAATGAAGTACTCAACTTTCGTCGTCG
			ApCel5A-R	AGACTAGGCCGGCCTTATTTGGAGTTGAAGACACCAGCAATG
Fusarium graminearum	*FgCel7B1*	C-terminal	FgCel7B1-F	AGAATGCTAGCATAATGTATCGCGCCATCGCCAC
			FgCel7B1-R	AGACTAGGCCGGCCTTACTGGCACTGGGAGTAGAAGTCGTTG
	FgCel7B2	N/A	FgCel7B2-F	AGAATGCTAGCATAATGAAGTTCTCTCCTCTTCTCCTCTCAACTCTTC

			FgCel7B2-R	AGACTAGGCCGGCCTTACAGGCGGTGAGCGCCATAC
	FgCel5A1	N/A	FgCel5A1-F	AGAATGCTAGCATAATGCGTTTCACAGATCTTCTTCTC
			FgCel5A1-R	AGACTAGGCCGGCCTTAAGCACCGATGAACTTCTTGATC
	FgCel5A2	N-terminal	FgCel5A2-F	AGAATGCTAGCATAATGAAGTCCCTCCTCGCCCTC
			FgCel5A2-R	AGACTAGGCCGGCCTTAGACATAAGTCTTGAGAAGGGAGTTGTAGTACTGG
Gloeophyllum trabeum	*GtCel12A*	N/A	GtCel12A-F	AGAATGCTAGCATAATGTTCCGCTTCATCTCTGCTTTGC
			GtCel12A-R	AGACTAGGCCGGCCTCACCCGCTCAAGCTGACG
Neurospora crassa	*NcCel7B1*	N/A	NcCel7B1-F	AGAATGCTAGCATAATGGTTCATAAACTCGCCTTCTTAAC
			NcCel7B1-R	AGACTAGGCCGGCCCTAAGCACTCAACCCCTTCG
	NcCel7B2	N/A	NcCel7B2-F	AGAATGCTAGCATAATGTCACGAAGGATTCTCTTGTCG
			NcCel7B2-R	AGACTAGGCCGGCCTCAAGCTTCAAAAGTCGACCCAATC
	NcCel5A	N-terminal	NcCel5A-F	GCTCTAGAATGAAGGCTACGATTCTTGC
			NcCel5A-R	AGACTAGGCCGGCCTTA

				AGGGGTATAGGTCTTGAGAAG
Nectria haematococca	*NhCel12A*	N/A	NhCel12A-F	AGAATGCTAGCATAATGAAGTCCGCCATCGTCG
			NhCel12A-R	AGACTAGGCCGGCCTTAGTATTCAACATCAGCGTGATAGTTGTTG
Phytophthora sojae	*PsCel12A1*	N/A	PsCel12A1-F	AGAATACTAGTATAATGAAGGTTGCGTTCGCTACTG
			PsCel12A1-R	AGACTAGGCCGGCCTTAGACGCGACGAACACTGC
	PsCel12A2	N/A	PsCel12A2-F	AGAATGCTAGCATAATGAAGAGCTCCGCCGTCCTC
			PsCel12A2-R	AGACTAGGCCGGCCTTACTGCTGCTGGACCTGCGAG
	PsCel12A3	N/A	PsCel12A3-F	AGAATGCTAGCATAATGAAGAGCTTTCTCCAACTCGTTG
			PsCel12A3-R	AGACTAGGCCGGCCTTAGTTGACAGCGGCGGAG
Sporotrichum thermophile	*StCel5A*	N-terminal	StCel5A-F	AGAATGCTAGCATAATGAAGTCCTCCATCCTCGCCAG
			StCel5A-R	AGACTAGGCCGGCCTTACGGCAAGTACTTCTTCAAGATCGAGTTGTAG
Trichoderma reesei	*TrCel7B*	C-terminal	TrCel7B-F	AGAATGCTAGCATAATGGCGCCCTCAGTTACACTG
			TrCel7B-R	AGACTAGGCCGGCCCTA

			AAGGCATTGCGAGTAGTAGTCGTTG
TrCel5A	N-terminal	TrCel5A-F	AGAATGCTAGCATAATGAACAAGTCCGTGGCTCCATTG
		TrCel5A-R	AGACTAGGCCGGCCCTACTTTCTTGCGAGACACGAGCTG
TrCel12A	N/A	TrCel12A-F	AGAATGCTAGCATAATGAAGTTCCTTCAAGTCCTCCCTG
		TrCel12A-R	AGACTAGGCCGGCCTTAGTTGATAGATGCGGTCCAGGATG

[a] The underlined sequences in the forward primers (F) and reverse primers (R) represent sequences identical to the ORF non-template strand beginning at the ATG codon and to the ORF template strand beginning at the stop codon, respectively. The 5' end of each forward oligo has 5 filler nucleotides followed by the *Nhe*I restriction endonuclease site followed by 3 filler nucleotides with the following exceptions: AfCel5A2-F and NcCel5A-F 5' ends have two filler nucleotides followed by *Xba*I restriction endonuclease site, and PsCel12A1-F 5' end has 5 filler nucleotides followed by the *Spe*I restriction endonuclease site followed by 3 filler nucleotides. The 5' end of each reverse oligo has 6 filler nucleotides followed by the *Fse*I restriction endonuclease site.

Table 2.6 – Primers used to amplify 7 cellobiohydrolase ORFs

Fungus	CBH	CBM	Primer	5' to 3' primer sequences[a]
Aspergillus niger	*AnCel6A*	N/A	AnCel6A-F	AGAATGCTAGCATAATGCACTCCACCAACATGC
			AnCel6A-R	AGACTAGGCCGGCCTCAAAGGGAAGGATTGGCGTTG
Lentinula edodes	*LeCel6A*	N-terminal	LeCel6A-F	AGAATGCTAGCATAATGAAGATTACTTCCACTGGCTTACTTG
			LeCel6A-R	AGACTAGGCCGGCCTCAAGGATGATCCGGTGTTAGTC
Neurospora crassa	*NcCel7A*	C-terminal	NcCel7A-F	AGAATGCTAGCATAATGCTCGCCAAGTTCG
			NcCel7A-R	AGACTAGGCCGGCCTTACACGCACTGGGAGTAATAGTC
	NcCel6A	N/A	NcCel6A-F	AGAATGCTAGCATAATGCGCGCCCTCTCCCTC
			NcCel6A-R	AGACTAGGCCGGCCTTAACTGTTCTTGACACTCGCATCCGCATTC
Sporotrichum thermophile	*StCel7A*	C-terminal	StCel7A-F	AGAATGCTAGCATAATGTACGCCAAGTTCGCGAC
			StCel7A-R	AGACTAGGCCGGCCTTACAGGCACTGCGAGTACCAG
Trichoderma reesei	*TrCel7A*	C-terminal	TrCel7A-F	AGAATGCTAGCATAATGTATCGGAAGTTGGCCGTCATC
			TrCel7A-R	AGACTAGGCCGGCCTTACAGGCACTGAGAGTAGTAAGGGTTC
	TrCel6A	N-terminal	TrCel6A-F	AGAATGCTAGCATAATGATTGTCGGCATTCTCACCACG

	TrCel6A-R	AGACTAGGCCGGCCTTACAGGAACGATGGGTTTGCGTTTG

[a] The underlined sequences in the forward primers (F) and reverse primers (R) represent sequences identical to the ORF non-template strand beginning at the ATG codon and to the ORF template strand beginning at the stop codon. The 5' end of each forward oligo has 5 filler nucleotides followed by the *Nhe*I restriction endonuclease site followed by 3 filler nucleotides. The 5' end of each reverse oligo has 6 filler nucleotides followed by the *Fse*I restriction endonuclease site.

2.3.4. Screening for functionally expressed endoglucanases

Screening for fungal endoglucanases that could be functionally expressed by *S. cerevisiae* strain CEN.PK 111-61A was performed by spotting 3 µl of culture filtrate, prepared from overnight liquid YNB shake-flask cultures, onto the surface of Congo red indicator plates containing 0.5 % carboxylmethylcellulose (CMC) sodium salt as described previously (Wood 1988).

2.3.5. Coding region optimization

The coding regions of five endoglucanases were optimized using Gene Designer 2.0 (available from https://www.dna20.com/). Gene Designer is a software package that uses the codon randomization strategy in optimizing coding DNA sequences for heterologous protein expression in any organism (Villalobos et al. 2006). When the codon randomization strategy is used, the codon bias of the optimized sequence will resemble that of an average protein coding sequence from the host strain used for expression of the heterologous protein (Villalobos et al. 2006). The flexibility of the codon randomization strategy allows for other manipulations of the DNA coding regions such as avoiding or including certain restriction sites, avoiding secondary mRNA structures, and avoiding repetitive elements (Villalobos et al. 2006). The presence of rare codons or rare codon clusters in regions is not taken into consideration using the codon randomization strategy.

2.3.6. Signal peptide replacement

Protein signal peptides were predicted using SignalP 4.0 (Petersen et al. 2011). Protein ORFs of the 5 EGs with the codon optimized sequences were PCR-amplified without their signal peptides with a pair of primers, where the upstream primer hybridized to about 20 nucleotides of the template strand beginning immediately after the sequence encoding the native signal peptide. The 5’ ends of the MFα1-pre and MFα1-prepro versions of the upstream primer of each EG were identical to the 33 3' bases of the template strand of the MFα1-pre and to the 34 3' bases of the template strand of the MFα1-prepro signal peptide coding region (Table 5). A downstream universal primer CYC1_R (GCCGCAAATTAAAGCCTTCG) hybridized to 20 bases of the coding strand

of the cyc1 transcription terminator. The PCR products were then cloned into plasmid p416-TEF-MFα1 prepro (Figure 2.2) by gap-repair (*in vivo* homologous recombination) (Orr-Weaver and Szostak 1983) in *S. cerevisiae* CEN.PK 111-61A. For this p416-TEF-MFα1 prepro, linearized with *EcoR*1 and *Xho*I, and the codon optimized endoglucanases products were co-transformed into the wild-type yeast strain followed by selection for uracil prototrophs. Congo Red indicator plate screening for functional endoglucanase expression (data not shown) was used to identify p416-TEF plasmids harbouring the desired EGs with MFα1-pre or MFα1-prepro signal peptides. The plasmids were rescued and used as templates to PCR amplify the EGs with the MFα1-pre or the MFα1-prepro signal peptides using two universal primers TEF1_F (CTTCAAAACACCCAAGCACAG) and CYC1_R (GCCGCAAATTAAAGCCTTCG). The upstream primer TEF1_F hybridized to 21 bases of the template strand of the TEF1 promoter, and the downstream universal primer CYC1_R hybridized to 20 bases of the coding strand of the CYC1 Tr region. The gel purified PCR products were then cloned into p425-TEF_M linearized with *Spe*I and *Xho*I by *in vivo* gap-repair (Orr-Weaver and Szostak 1983) using strain CEN.PK 111-61A. The desired recombinant p425-TEF_M plasmids EGs with MFα1-pre or MFα1-prepro signal peptides were identified by selection for leucine prototrophy, followed by Congo Red indicator plate screening for functional endoglucanase expression (data not shown).

Table 2.7 – Forward primers used in signal peptide replacement

EG	Primer	5' to 3' primer sequences[a]
AfCel7B1opt	preAfCel7B1opt-F	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCCAACAACCTGCTGCAAGTAGTG
	preproAfCel7B1opt-F	GCTAAAGAAGAAGGGGTATCTTTGGATAAAAGACAACAACCTGCTGCAAGTAGTG
GtCel12Aopt	preGtCel12Aopt-F	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCGCGACTGTGCTTACCGGTC
	preproGtCel12Aopt-F	GCTAAAGAAGAAGGGGTATCTTTGGATAAAAGAGCGACTGTGCTTACCGGTC
StCel5Aopt	preStCel5Aopt-F	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCCAGTCCGGCCCTTGG
	preproStCel5Aopt-F	GCTAAAGAAGAAGGGGTATCTTTGGATAAAAGACAGTCCGGCCCTTGG
ApCel5Aopt	preApCel5Aopt-F	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCGCATACGGACAATGTGGGG
	preproApCel5Aopt-F	GCTAAAGAAGAAGGGGTATCTTTGGATAAAAGAGCATACGGACAATGTGGGG
AfCel5A1opt	preAfCel5A1opt-F	TTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCGCCCCTAATGCAAAAAGAGC
	preproAfCel5A1opt-F	GCTAAAGAAGAAGGGGTATCTTTGGATAAAAGAGCCCCTAATGCAAAAAGAGC

[a] The underlined sequences in the forward primers (F) represent sequences identical to the ORF non-template strand beginning at the first codon downstream of the signal peptide sequence as identified by using SignalP 4.0 (Petersen et al. 2011). The 5' ends of the MFα1-pre and MFα1-prepro versions of the forward primers of each EG represent sequences identical to the 33 3' bases of the template strand of the MFα1-pre and to the 34 3' bases of the template strand of the MFα1-prepro signal peptide coding region.

2.3.7. Endoglucanase and cellobiohydrolase activity levels

Sugars were removed from culture filtrates using centrifugal filter devices (Ultrafree-0.5, Millipore) and then buffer exchanged into sodium citrate buffer (10 mM, pH 5.0). Enzyme activity was determined by measuring the amount of reducing sugar ends released after cellulose hydrolysis using the bicinchoninic acid reducing sugar assay, which was adapted for 96-well microtiter plates as previously described (Doner and Irwin 1992; Grishutin et al. 2004; Zorov et al. 1997). Hydrolysis reactions were performed in 200 µl reactions with 0.5% w/v CMC-4M (for EGs) or 0.5% w/v PASC (for CBHs), 37.5 mM citrate buffer (pH 5) and 25 µl of desalted culture filtrate at 37^{o}C and shaking at 200 rpm for 24 hours. Activities are presented as U mL^{-1} OD_{600}^{-1}, with one unit defined as the amount of enzyme required for the release 1 µmol of reducing sugar ends per minute under assay conditions.

2.3.8. Chromosomal integration

2.3.8.1 Construction of CEN.PK111-61A-δintegAnBgl1, a strain that grows well on cellobiose

In a previous study, 35 fungal BGLs were screened for their ability to be functionally expressed in *S. cerevisiae* (Wilde et al. 2012). The screen identified *A. niger Bgl1* coding for a family 3 (GH3) glycosylhydrolase that was functionally expressed in *S. cerevisiae* and highly active towards cellobiose. *AnBgl1* was used to engineer *S. cerevisiae*, CEN.PK11-6A derivatives capable of growing with cellobiose as the sole carbon source. *AnBgl1* ORF DNA was prepared by PCR amplification using *AnBgl1* cDNA (Wilde et al. 2012) as the template and the forward primer, AnBgl1-F, and the reverse primer, AnBgl1-R (Table 6). The resulting *AnBgl1* ORF DNA was cloned into plasmid, p425TEF_M, following digestion of the plasmid and *AnBgl1* ORF DNA with *Nhe*I and *Fse*I. The *AnBgl1* cassette was ligated into p425TEF_M. After verification by restriction analysis and sequencing the resulting plasmid was used as the template to PCR amplify the TEF1 Pr_Bgl1 ORF_CYC1 Tr cassette using primers δ_TEF_F and δ_CYC1_R (Table 6). The amplified *AnBgl1* cassette is flanked by transposon δ elements (Figure

2.4). The forward primer, δ_TEF_F, has 46 bases of identity to the non-template strand at the 5'-end of the δ element and is 73 bp long. The reverse primer, δ_CYC1_R, has 51 bases of identity to the template strand at the 3'-end of the δ element and is 71 bases long (Table 2.8). The resulting linear 5'δ _TEF1 Pr_AnBgl1 ORF_CYC1 Tr_3' δ cassette DNA was transformed into the wild type yeast strain CEN.PK111-61A. YNB plates supplemented with histidine, uracil and leucine, with cellobiose as the sole carbon source, were used to select for strains expressing AnBgl1. Two fast growing integrants were selected. Since they both grew similarly on cellobiose plates, one integrant, CEN.PK111-61A-δ-integAnBgl1a, was selected for further studies.

Table 2.8 – Primers used for the construction and δ-integration of the TEF1 Pr_BGL1 ORF_CYC1 Tr cassette

Primer	5' to 3' primer sequences
AnBgl1-F[a]	AGAATGCTAGCATAATGAGGTTCACTTTGATCGAGGC
AnBgl1-R[a]	AGACTAGGCCGGCCTTAGTGAACAGTAGGCAGAGACGC
δ_TEF_F[b]	GAAACGCAAGGATTGATAATGTAATAGGATCAATGAATA TAAACATATAGCTTCAAAATGTTTCTACTCCTTT
δ_CYC1_R[b]	GGGAATCTGCAATTCTACACAATTCTATAAATATTATTAT CATCATTTTATGCAAATTAAAGCCTTCGAGC

[a] The underlined sequences in the forward primer AnBgl1-F and the reverse primer AnBgl1-R represent sequences identical to the ORF non-template strand beginning at the ATG codon and to the ORF template strand beginning at the stop codon. The 5' end of the forward oligo has 5 filler nucleotides followed by the *Nhe*I restriction endonuclease site followed by 3 filler nucleotides. The 5' end of the reverse oligo has 6 filler nucleotides followed by the *Fse*I restriction endonuclease site.

[b] The underlined sequences in the forward primer δ_TEF_F and the reverse primer δ_CYC1_R represent sequences identical to the non-template strand of the TEF1 promoter sequence beginning at ATA and to the template strand of the CYC1 terminator sequence beginning at GCA. The 5' end of the forward oligo δ_TEF_F has 46 nucleotides identical to the non-template strand of the 5' half of the YDRWdelta27 element. The 5' end of the reverse oligo δ_CYC1_R has 51 nucleotides identical to the template strand of the 3' half of the YDRWdelta27 element.

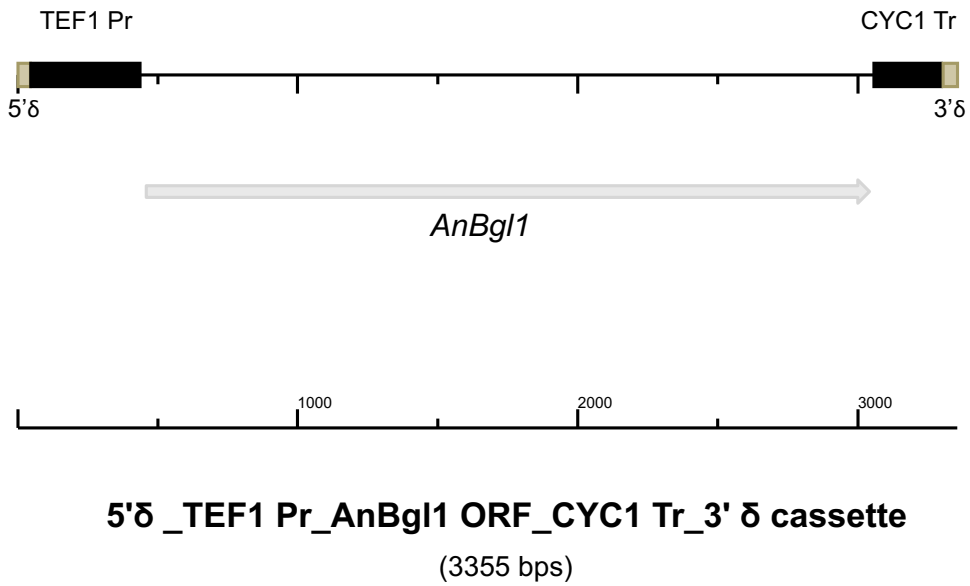


Figure 2.4 – *AnBgl1* expression cassette. This expression cassette was transformed into the yeast strain CEN.PK 111-61A. P425_M-AnBgl1 plasmid was used as the template to PCR amplify *AnBgl1* flanked by TEF1 promoter and CYC1 terminator. The resulting linear DNA was transformed into yeast strain CEN.PK111-61A.

2.3.8.2 Endoglucanase chromosomal integration

The ORFs of five selected endoglucanases were directionally cloned into the *Nhe*I and *Fse*I sites of p425-TEF_M_Δ2μ _δ. The resulting plasmids were transformed into the BGL expressing strain CEN.PK111-61A-δ-AnBgl1a followed by selection for leucine prototrophs.

2.3.8. Growth curves

Yeast overnight cultures were prepared by picking a portion of a fresh colony with a toothpick and inoculating 5 ml of YNB medium containing 2% glucose. The resulting cultures were incubated at 30^{o}C and 200 rpm for 16 h. Cells were then washed twice with distilled water before inoculating into 50 ml of media (YNB with 50 mg/l histidine, 20 mg/l uracil and either 2% glucose, 2% cellobiose, or 2% CMC-4M as the carbon source) in each of three 125 ml shake flasks. The growth was monitored by measuring the increase in OD_{600} over time. Growth rates were determined using at least 7 data points taken during a portion of the exponential growth phase when the culture OD_{600} was between 0.1and 1.0. The R^2 value of the exponential trend line for all the growth rate determinations was greater than 0.98.

2.3.9. SDS-PAGE of secreted proteins

Yeast culture filtrates were desalted and concentrated 10, 20 or 100 times using centrifugal filter devices (Ultrafree-0.5, Millipore) and buffer exchanged into sodium citrate buffer (10 mM, pH 5.0). Concentrated culture filtrates (24 µl) and 6 µl of 5X loading buffer were resolved using 12% SDS-PAGE gels.

2.3.10. Protein deglycosylation

Protein deglycosylation reactions were performed using Protein Deglycosylation Mix P6039S (NEB) according to the manufacturer's instructions. Deglycosylation reactions were performed in 24 µl reactions containing 15.43 µl of concentrated culture filtrates, 8.57 µl deglycosylation reaction mix, and 6 µl of 5X loading buffer. The 30 µl samples were resolved on SDS-PAGE gels.

2.3.11. Estimation of secreted protein levels

Secreted protein levels were estimated using glycosylated and deglycosylated protein samples prepared as described in sections 2.3.9 and 2.3.10 and BSA standards (10 µl) with 0.025, 0.05, 0.075 and 0.1 µg/µl. The protein samples and BSA standards were resolved on SDS-PAGE gels and the resulting gels were stained using Coomassie Brilliant Blue G-250. Gel images were captured using the SynGene G:BOX F^3 gel doc system and analyzed using the GeneTools software from SynGene (http://www.syngene.com). For each gel image the peak areas of the BSA standards included on the gel were used to generate the standard curves of peak area versus BSA protein amounts. The peak areas of experimental Bgl1 and/or EG protein bands, also determined using the GeneTools software, were converted to protein amounts using the standard curve. An example of a protein amount determination is presented in the Appendix section (Figure A4 and A5).

2.3.12. Relationship between *S. cerevisiae* CEN.PK111-61A OD_{600} and dry cell weight

Ten independent cultures of *S. cerevisiae* strain CEN.PK111-61A transformed with p425-TEF_M were grown overnight at 30°C with shaking at 200 rpm in 125 ml Erlenmeyer flasks containing with 25 ml of YNB media supplemented with histidine and uracil. The pellet of 3 ml of each culture was freeze-dried using a ThermoSavant Freeze Dryer Modulyo D and weighed using a Mettler Toledo AX205 analytical scale. The DCW of CEN.PK111-61A was determined to be 0.322 ± 0.019 g $.L^{-1}$. OD_{600}^{-1}.

2.3.13. Determination of gene copy number by quantitative PCR

The number of integrated copies of BGL and EG ORFs was determined by quantitative real-time PCR (qPCR) using the standard curve method with normalization to the housekeeping gene *PGK1* essentially as described previously (Whelan, Russell, Whelan 2003). Genomic DNA was isolated from recombinant yeast strains grown overnight in YPD media using the yeast smash and grab genomic DNA mini-prep method (Hoffman

2001). The primers used for qPCR gene copy number determinations are listed in Table 2.9. PCR primers were designed using the Primer Express™ software to have melting temperature ranging from 58 to 59°C and lengths of 19 to 24 bases. The amplicon length ranged from 81 and 94 bp for the genes of interest and was 114 bp for the housekeeping gene *PGK1*. Quantitative PCR was performed using the Eco Real-Time PCR System (Illumina Inc.) with the MBI *EVOlution* 5X EvaGreen® qPCR mix (Montreal Biotech Inc.). The normalized integrated gene copy number was calculated by the standard curve method using *PGK1* as the native single copy gene. The DNA samples used as templates for making the standard curves used for the calculation of the absolute copy number of *PGK1*, *AnBgl1*, *AfCel7B1opt*, *GtCel12A_opt*, *StCel5Aopt*, *preApCel5Aopt*, and *preproAfCel5A1opt* were gel-purified PGK1 PCR fragment, p425-TEF_M-AnBgl1, p425-TEF_M-AfCel7B1opt, p425-TEF_M-GtCel12Aopt, p425-TEF_M-StCel5Aopt, p425-TEF_M-preApCel5Aopt, and p425-TEF_M-preproAfCel5A1opt, respectively. Standard curves were made using a series of at least five 10-fold serial dilutions of template DNA samples. DNA samples used as template for making the standard curves were quantified using Quant-iT™ PicoGreen® dsDNA reagent (ThermoFisher Scientific) according to the instructions provided by the manufacturer. Normalized integrated gene copy numbers were calculated by dividing the absolute copy number of the gene of interest by the absolute copy number of *PGK1* present in each sample. All qPCR reactions were done in triplicate. Each primer pair in Table 2.9 produced a single peak in the derivative of their melt curve indicating the presence of a single PCR product (Appendix Figure A3). Non-template control reactions were also done in triplicate for all the primer pairs and did not produce any detectable PCR product. None of the genes of interest, except for the housekeeping gene *PGK1*, were detected when the genomic DNA of the wild-type strain CEN.PK111-61A was used as a template (data not shown).

Table 2.9 – Primers used for qPCR analysis

Gene of interest	Primer	5' to 3' primer sequences[a]
PGK1	PGK1_F1	AGCTCACTCTTCTATGGTCGGTTT
	PGK1_R1	AAGAATGGTCTGGTTGGGTTCTC

AnBgl1	AnBGL1_F1	AGGTTGCGGGTGATGAAGTT
	AnBGL1_R1	CGAATTGACGCAGCACGAT
GtCel12Aopt	GtCel12A_F1	TCCTTAAAAACTTCAGGCATCTCA
	GtCel12A_R1	CAAAAGATAGCGTTGTCCAGGTT
StCel5Aopt	StCel5A_F1	TTGCGTTTACCAAAATGATTGG
	StCel5A_R1	GGTTGCGGTTGGTCTAGATGTT
preApCel5Aopt	ApCel5A_F1	AGCTGTAGCGGCCACTTCTG
	ApCel5A_R1	CCGGATGACGCTCCATTAGTAT
AfCel7B1opt	AfCel7B_F1	ACACCACACCCATGTTCTGCTA
	AfCel7B_R1	GGGCCCCAGTAAGACTTTTGA
preproAfCel5A1opt	AfCel5A_F1	CGAATTTGCTGGTGGAGCTAA
	AfCel5A_R1	GCCCATCCACACGTCAGAAT

[a] Quantitative PCR primers were designed using the Primer Express™ software with a melting temperature range of 58 to 59°C and a length of 19 to 24 bases. The amplicon length ranged between 81 and 94 base pairs for the genes of interest and was 114 base pairs for the housekeeping gene *PGK1*.

2.3.14. Cladogram method

Cladograms representing the phylogenetic relationships of the 25 EGs studied herein were generated using MEGA7 (Kumar, Stecher, Tamura 2016). The cladogram was constructed using the maximum likelihood statistical method (Whelan and Goldman 2001) and was tested using 1000 bootstrap replications (Felsenstein 1985). The amino acid sequences of the conserved domains were identified using the NCBI conserved domain database (Marchler-Bauer et al. 2015) and aligned using Muscle (Edgar 2004).

Results

A three-step strategy was used to develop *S. cerevisiae* strains capable of hydrolyzing cellulosic biomass. First, a yeast strain capable of growing on cellobiose as the sole carbon source was developed by transformation of strain CEN.PK111-61A with the 5' δ -TEF1 Pr-*Bgl1*-CYC1 Tr-3; δ cassette (Figure 2.4) followed by selection for growth on cellobiose. In the second step, a library of 25 EGs from 10 evolutionary diverse fungi was screened to find EGs that could be expressed as functional secreted EGs by CEN.PK111-61A. The EGs selected for further analysis were subjected to coding region optimization and signal peptide replacement in order to determine how these modifications affected protein expression. Both plasmid-borne and δ-integrated versions of the selected EGs were expressed in CEN.PK111-61A-δ-AnBgl1a. The resulting transformants were tested to determine their ability to grow on CMC-4M as the sole carbon source. In the third step, 7 CBH ORFs from 5 evolutionary diverse fungi were expressed in the *S. cerevisiae* strain CEN.PK111-61A to identify CBH ORFs that could be expressed as active secreted enzymes by CEN.PK111-61A.

3.1. Strain CEN.PK111-61A Expressing the TEF1 Pr-*AnBgl1*-CYC1 Tr cassette Grows Well using Cellobiose as Its Sole Carbon Source

Developing CBP competent *S. cerevisiae* strains requires efficient and simultaneous expression of at least three different cellulose active enzymes, a BGL, an EG, and a CBH. A three-step approach was taken to produce *S. cerevisiae* strains that can grow on pre-treated lignocellulosic biomass. First, a strain that could grow well with cellobiose as the sole carbon source was generated by transformation of the 5'δ -TEF1 Pr-*AnBgl1*-CYC1 Tr-3'δ cassette (Figure 2.4) into CEN.PK11-61A followed by selection of transformants on YNB plates with cellobiose as the sole carbon source. This identified two transformants that grew fast on cellobiose plates generating large colonies after 48 h. These two strains were designated CEN.PK111-61A-δ-AnBgl1a and CEN.PK111-61A-δ-AnBgl1b.

3.1.1. Growth rate of CEN.PK111-61A-δ-AnBGL1a on glucose and cellobiose

To determine whether levels of secreted AnBgl1 expression by the CEN.PK111-61A-δ-AnBgl1a were sufficient to support the vigorous growth of CEN.PK111-61A-δ-AnBgl1a in liquid cultures with cellobiose as the sole carbon source, we compared the growth rate of CEN.PK111-61A-δ-AnBgl1a and CEN.PK111-61A in liquid YNB medium with glucose and cellobiose as sole carbon sources. The exponential growth phases were used to determine the growth rates and doubling times of CEN.PK111-61A-δ-AnBgl1a and CEN.PK111-61A using glucose and cellobiose as sole carbon sources. The growth of these cultures was followed for 72 h and the growth rates during the exponential growth phase were also determined (Figure 3.1). The R^2 values of the exponential trend lines for all the growth rate determinations were greater than 0.99.

When glucose was used as the carbon source the growth rates and doubling times were 0.33 h^{-1} and 2.1 hours for CEN.PK111-61A-δ-AnBgl1a, and 0.34 h^{-1} and 2.1 hours for CEN.PK111-61A (Table 3.1). These results show that both strains had similar growth rates and, therefore, chromosomal integration of the 5'δ -TEF1 Pr-*AnBgl1*-CYC1 Tr-3'δ cassette into the CEN.PK111-61A strain did not affect the growth rate when glucose was the carbon source. When cellobiose was used as the carbon source the CEN.PK111-61A strain was unable to grow on cellobiose; however, the CEN.PK111-61A-δ-AnBgl1a had a growth rate of 0.33 h^{-1} on glucose and 0.3 h^{-1} on cellobiose (Figure 3.1). These results show that the growth of CEN.PK111-61A-δ-AnBGL1a in liquid YNB medium with cellobiose as the sole carbon source was very similar to the growth rates of either CEN.PK111-61A or CEN.PK111-61A-δ-AnBgl1a when glucose was the carbon source.

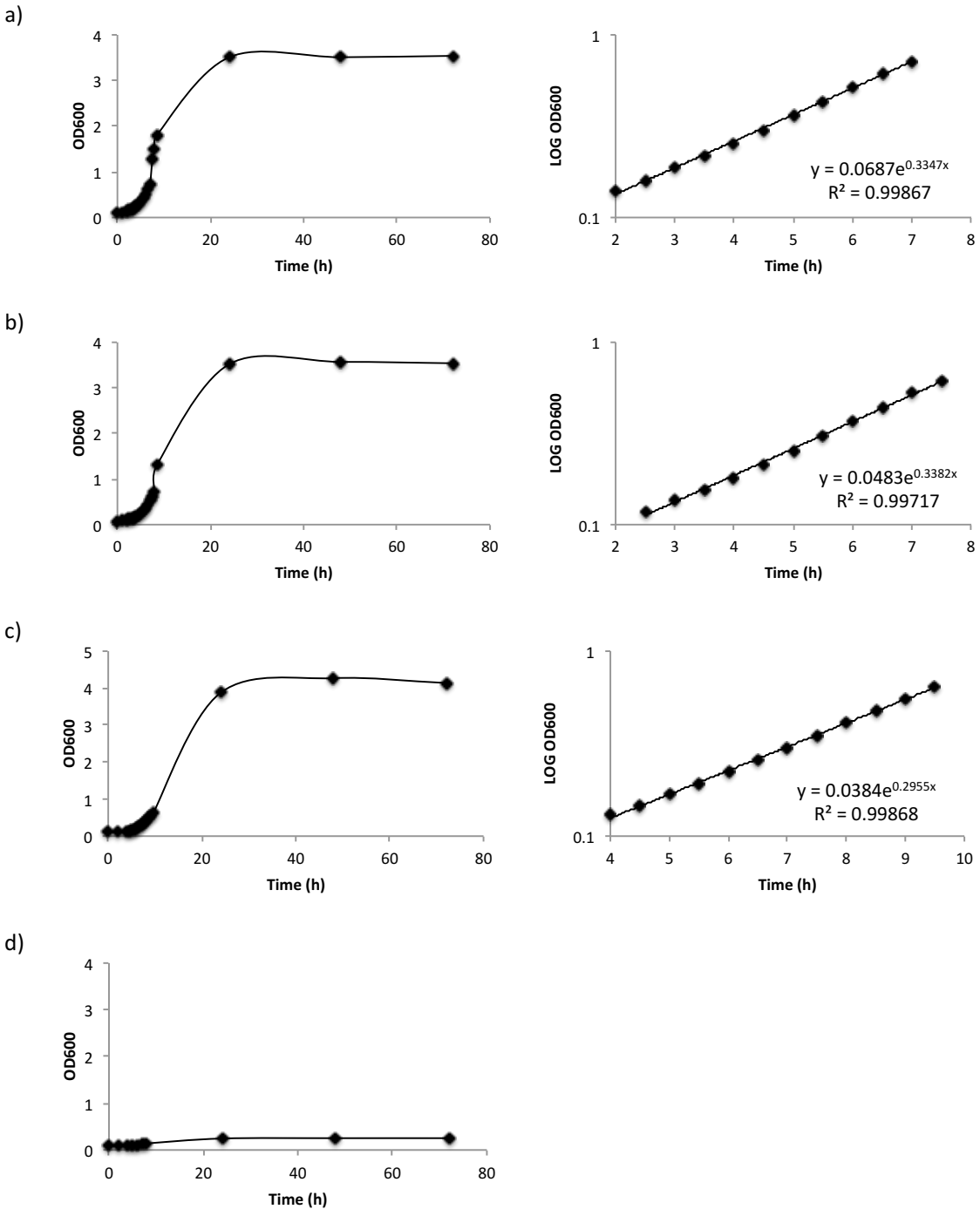


Figure 3.1 – Growth curves of a) CEN.PK111-61A-δ-AnBgl1a and b) CEN.PK111-61A on YNB media with glucose as the carbon source c) CEN.PK111-61A-δ-AnBgl1a and d) the wild-type strain CEN.PK111-61A on YNB media with cellobiose as the carbon source. Panel a (right) shows the exponential trend line for growth rate determination using a portion of the exponential growth phase between OD_{600} 0.1 and 1.0.

Table 3.1 – Growth rates and generation time of the wild-type strain CEN.PK111-61A and CEN.PK111-61A-δ-AnBgl1a with 2% glucose and cellobiose as sole carbon source

	Glucose		Cellobiose	
	growth rate μ (h^{-1})	generation time g (h)	growth rate μ (h^{-1})	generation time g (h)
wild type (wt)	0.34	2.05	N/A*	NG*
CEN.PK111-61A-δ-AnBgl1a	0.33	2.07	0.30	2.35

*NG – No growth

The amount of secreted AnBgl1 produced by CEN.PK111-61A-δ-AnBgl1a was determined using the band intensities in the gel depicted in Figure 3.2. Band intensities in the Coomassie Brilliant Blue G-250 stained SDS-PAGE gel (Figure 3.2) were used to determined protein expression levels using gel images captured using a SynGene G:BOX F^3 gel doc system and image analysis performed using the GeneTools software as described in Materials and Methods section 2.3.11. Levels of secreted AnBgl1 were 0.38 μg/ml when grown on cellobiose and 0.34 μg/ml when grown on glucose (Figure 3.2). These results show that 0.34 μg/ml of secreted AnBgl1 is sufficient to sustain the growth of CEN.PK111-61A-δ-AnBgl1a on cellobiose as the sole carbon source.

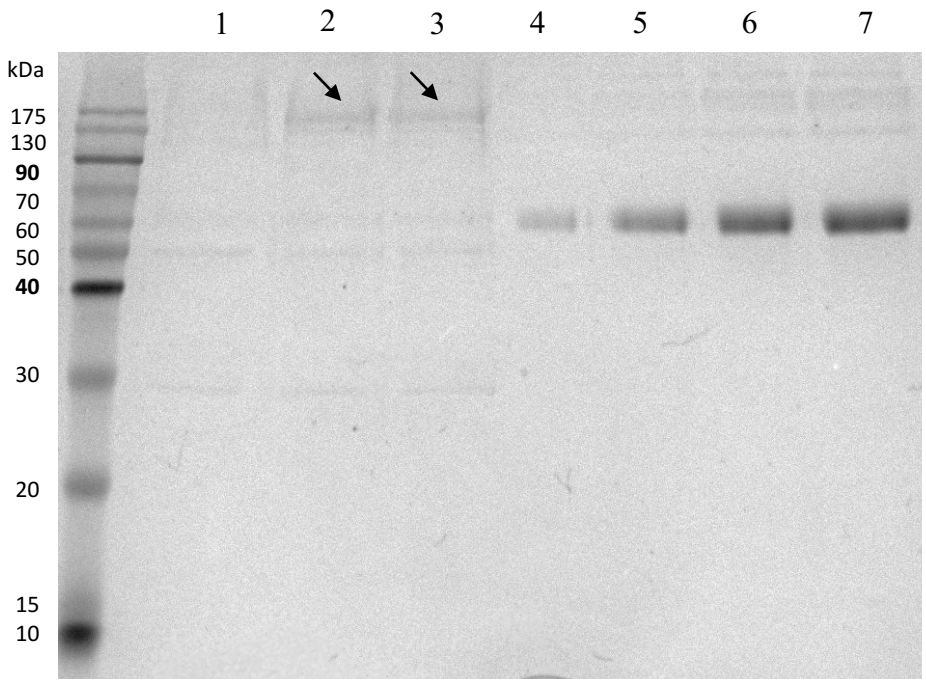


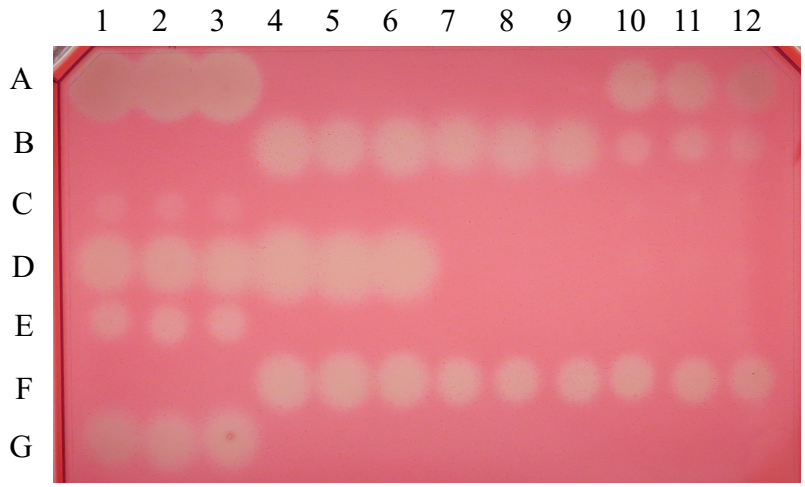
Figure 3.2 – AnBgl1 SDS-PAGE analysis. SDS-PAGE of culture filtrates of the yeast wild-type strain CEN.PK111-61A (lane 1) and the yeast strain with CEN.PK111-61A-δ-Bgl1a grown on cellobiose (lane 2), and glucose (lane 3). The BSA standards 0.25, 0.5, 0.75, and 1 µg are in lanes 4, 5, 6, and 7, respectively. Lanes 1, 2 and 3 were loaded with 30 µl, 24 µl of culture filtrates concentrated 20 times and 6 µl of 5X loading dye were loaded in sample lanes of a 12% SDS-PAGE gel.

3.2. Recombinant Expression of a Library of Fungal Endoglucanases in *S. cerevisiae*

3.2.1. Screening for fungal endoglucanases that can be functionally expressed by *S. cerevisiae*

Once a strain, CEN.PK111-61A-δ-AnBgl1a, capable of very efficient growth using cellobiose as the carbon source was generated, we wanted to identify a fungal EG that could be expressed efficiently by strain CEN.PK111-61A at levels that generated visually detectable halos after 16 h at 37^{o}C (Figure 3.3). For this, the ORFs of 25 fungal endoglucanase ORFs, 11 coding for GH family 5 EGs, 8 coding for GH family 7 EGs and 6 coding for GH family12 EGs, were cloned into the *S. cerevisiae* multi-copy 2-micron yeast expression vector p425_TEF_M. Congo Red indicator plates were used to the determine their relative abilities to be expressed as functional secreted endoglucanases by strain CEN.PK111-61A. This screen revealed that strain CEN.PK111-61A produced functional secreted endoglucanase when transformed with 14 of the 25 endoglucanase genes (Figure 3.3).

Because the Congo Red indicator plate screening showed that strains harbouring plasmids p425_M_AfCel7B1, p425_M_GtCel12A, and p425_M_StCel5A expressed the highest levels of secreted endoglucanase activity, *AfCel7B1*, *GtCel12A* and *StCel5A* were selected for further development of CBP competent *S. cerevisiae*. In addition, *ApCel5A* and *AfCel5A1*, which Congo Red indicator plate screening indicated were expressed a much lower levels, were selected for further analysis to determine whether their expression could be increased by modifying their coding regions, replacing their native signal peptides with a yeast signal peptide, or changing their copy number.



b)

	1	2	3	4	5	6	7	8	9	10	11	12
A	AfCel7B1 **			AfCel7B2			AfCel5A1 **			AfCel5A2 *		
B	AfCel5A3			AnCel7B *			AnCel5A1 *			AnCel5A2 *		
C	ApCel5A **			FgCel7B1			FgCel7B2			FgCel5A1		
D	FgCel5A2 *			GtCel12A **			NcCel7B1			NcCel7B2		
E	NcCel5A *			NhCel12A			PsCel12A1			PsCel12A2		
F	PsCel12A3			StCel5A **			TrCel7B *			TrCel5A *		
G	TrCel12A *			VTO								

* Indicates EGs that generated detectable halos

** Indicates EGs that generated detectable halos and were selected for codon optimization

Figure 3.3 – Congo Red indicator plate screening of recombinant *S. cerevisiae* culture filtrates. Panel A. Congo Red indicator plates seeded with 0.5% CMC-4M were spotted in triplicate with 3 µl of culture filtrates isolated from CEN.PK111-61A transformed with recombinant derivatives of 2 micron expression vector p425-TEF_M harboring the indicated cDNA-derived endoglucanase genes.. Panel B. Map of the EG CEN.PK111-61A transformants used. VTO, vector transformant only, indicates where culture filtrate prepared using the control strain CEN.PK111-61A transformed with the empty expression vector (p425-TEF_M lacking a cloned cDNA-derived endoglucanase gene).

3.2.2. Coding region optimization

In an effort to increase the amount of secreted endoglucanase activity, the coding regions of the five selected endoglucanase ORFs, *ApCel5A*, *GtCel12A*, *StCel5A*, *AfCel7B1* and *AfCel5A1* were subjected to codon optimization. Codon optimization of *AfCel7B1*, *GtCel12A*, and *StCel5A* (Figure 3.4), the three endoglucanses that were selected because Congo Red indicator plate screening identified them as expressing the highest levels of secreted endonuclease activity, resulted in significantly higher levels of secreted endoglucanase activity as determined by two tailed T-test (*AfCel7B1* $p = 0.03$ was significant at $p < 0.05$, *GtCel12A* $p = 0.0079$ was significant at $p < 0.01$ and *StCel5A* $p = 0.070$ was significant at $p < 0.1$). Codon optimization of *ApCel5A*, which was selected because it produced an intermediate level of secreted endoglucanse activity, did not result in significantly increased secreted endoglucanse activity ($p = 1.0$ was not significant at $p < 0.1$). In contrast, codon optimization of *AfCel5A1*, the gene selected because it produced the lowest amount of secreted endoglucanase activity, resulted in the most dramatic increase in production of secreted endoglucanase activity, about 18-fold ($p = 0.001$ was significant at $p < 0.01$).

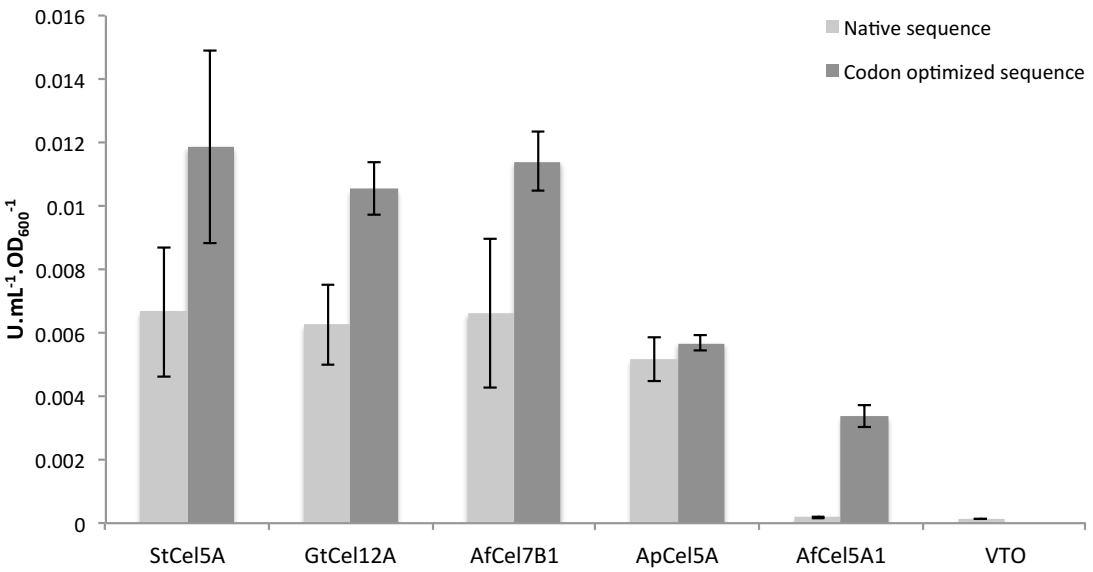


Figure 3.4 – Effect of codon optimization on levels of secreted endonuclease activity. Reducing sugar equivalents produced after 3 h hydrolysis of 0.5% CMC-4M by culture filtrates of yeast CEN.PK111-61A expressing the indicated endoglucanases using gene versions with their native and codon optimized ORFs. The y-axis is the units per ml per OD_{600} where 1 unit is the amount of enzyme required to release 1 µmol reducing sugar ends per minute under assay conditions. Each assay was done in triplicate and the average values with their standard deviations are presented.

3.2.3. Signal peptide replacement

The efficient production of secreted proteins is dependent upon a number of factors including their transport through the secretory pathway. Signal peptides are necessary for targeting proteins to the endoplasmic reticulum. The native signal peptides of the selected endoglucanases may not be as efficient at directing proteins to the *S. cerevisiae* secretory pathway as native *S. cerevisiae* signal peptides. Indeed, the activity levels of secreted *Trichoderma viride* EG1 increased by 61.5% when its signal peptide was replaced by the *S. cerevisiae MFα1* prepro-α-factor signal peptide (Zhu, Yao, Wang 2010). In an attempt to increase the production or secreted endoglucanase activity by the 5 codon-optimized endoglucanases, their native secretion signal peptides were replaced with the *S. cerevisiae* MFα1 pre-α-factor and prepro-α-factor signal peptides. Replacing the signal peptide of the codon optimized EGs with either the *S. cerevisiae MFα1* pre-α-factor or the prepro-α-factor signal peptides significantly reduced production of secreted endoglucanase activity ($p < 0.05$) from p425_M-AfCel7B1, p425_M-StCel5A and p425_M-GtCel12A by at least 50% (Figure 3.5). Replacing the native signal peptide of *ApCel5A* with the pre-α-factor signal peptide significantly increased production of secreted endoglucanase activity about 1.2-fold, while replacing the native signal peptide with the prepro-α-factor signal reduced production of secreted endoglucanase activity by about 30% (Figure 3.5). When the native signal peptide of *AfCel5A1* was replaced with the prepro-α-factor signal peptide production of secreted endoglucanase activity increased about 3.5-fold whereas when the native signal peptide was replaced with the pre-α-factor signal peptide production of secreted endoglucanase activity decreased by about 90%. Remarkably, the combined effect of codon optimization and native signal peptide replacement by the prepro-α-factor signal peptide increased the amount of secreted endoglucanase activity produced by *AfCel5A1* by about 60-fold.

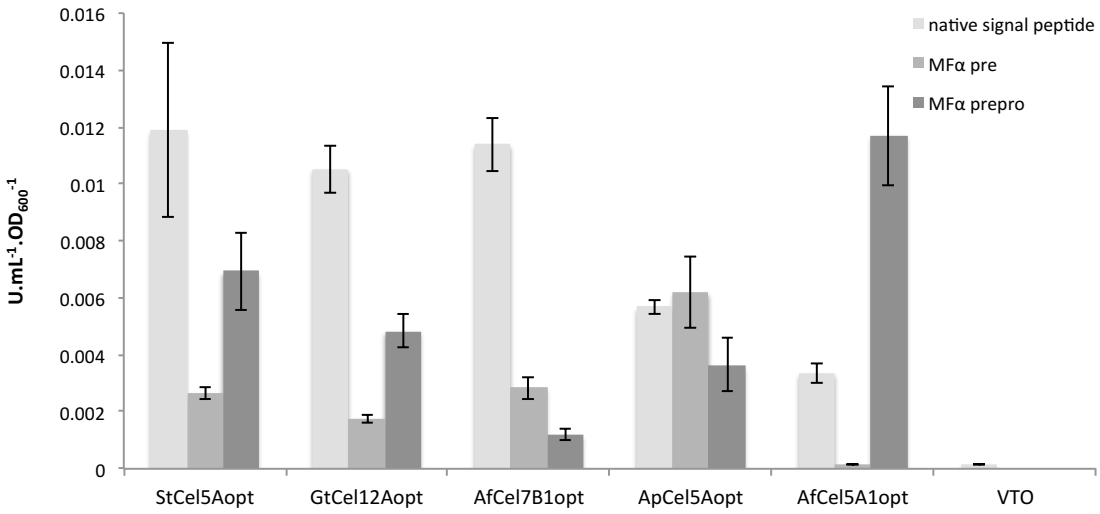


Figure 3.5 – Effect of signal peptide replacement on levels of secreted endonuclease activity. Reducing sugar equivalents produced after a 3 h hydrolysis of 0.5% CMC-4M by culture filtrates produced by yeast expressing the 5 codon-optimized individual endoglucanases with their native, MFα1-pre or the MFα1-prepro signal peptides. The y-axis is the units per ml per OD_{600} where 1 unit is the amount of enzyme required to release 1 μmol of reducing sugar ends per minute under assay conditions. Each assay was done in triplicate and averages with their standard deviations are presented. Pairwise comparisons, performed using a Two tailed T test, between units of activity produced by the native, MFα1-pre and MFα1-prepro signal peptide versions of *StCel5Aopt*, *GtCel12opt*, and *AfCel5Aopt* showed that the native signal peptide supported the production of significantly higher levels of activity than did either the MFα1-pre or MFα1-prepro signal peptides ($p < 0.05$). Pairwise comparisons between units of activity produced by the native, MFα1-pre and MFα1-prepro signal peptide versions of *AfCel5A1opt* showed that the MFα1-prepro signal peptide supported significantly higher levels of activity ($p < 0.01$). Pairwise comparison of the native MFα1-pre and MFα1-prepro signal peptide versions of *ApCel5A* showed that the native and MFα1-pre signal peptides did not produced significantly different levels of endoglucanase activity, although they did support significantly higher levels of secreted protein activity than did the MFα1-prepro signal peptide ($p < 0.1$).

3.2.4. Secreted EG production levels and EG glycosylation

Heterologous proteins expressed by *S. cerevisiae* could be hyper-glycosylated (Ilmen et al. 2011; Jeoh et al. 2008; Skory, Freer, Bothast 1996). Hyper-glycosylation can impact the activity of heterologous proteins (Rasmussen 1992). Buffer-exchanged culture filtrates of the selected endoglucanases were subjected to deglycosylation in order to determine the extent of their glycosylation. SDS-PAGE was performed using culture filtrates prepared from CEN.PK111-61A transformed with p425_TEF_M_StCel5Aopt, p425_M_GtCel12Aopt, p425_M_AfCel7B1opt, p425_M_MFα1_preApCel5Aopt, and p425_M_MFα1_prepreAfCel5A-1_opt (Figure 3.6). Assuming signal peptide cleavage and no glycosylation, the predicted molecular weights as determined the ExPASy ProtParam tool (Gasteiger et al. 2005) for StCel5Aopt, GtCel12Aopt, AfCel7B1opt, preApCel5A-1opt, and preproAfCel5aA1opt were 40.9, 24.1, 46.2, 42.5 and 41.1 kDa, respectively. The observed molecular weights before versus after deglycosylation were about 53 versus 45 for StCel5Aopt, 28 versus 24 for GtCel12Aopt, 47 versus bands of 45 and 46 for AfCel7B1opt, 63 versus 60 for preApCel5A-1opt and 46 versus 46 for AfCel5aA1opt (Figure 3.6). These results show that deglycosylation increases the mobility of all of the proteins except preproAfCel5aA1opt. These results also show that even the deglycosylated version of StCel5Aopt, preApCel5A1opt and preproAfCel5aA1opt, had apparent masses greater than their predicted masses. A common feature associated with the three endoglucanases with SDS-PAGE determined masses greater than their predicted masses were that they were all GH5 endoglucanases with class 1 CBMs.

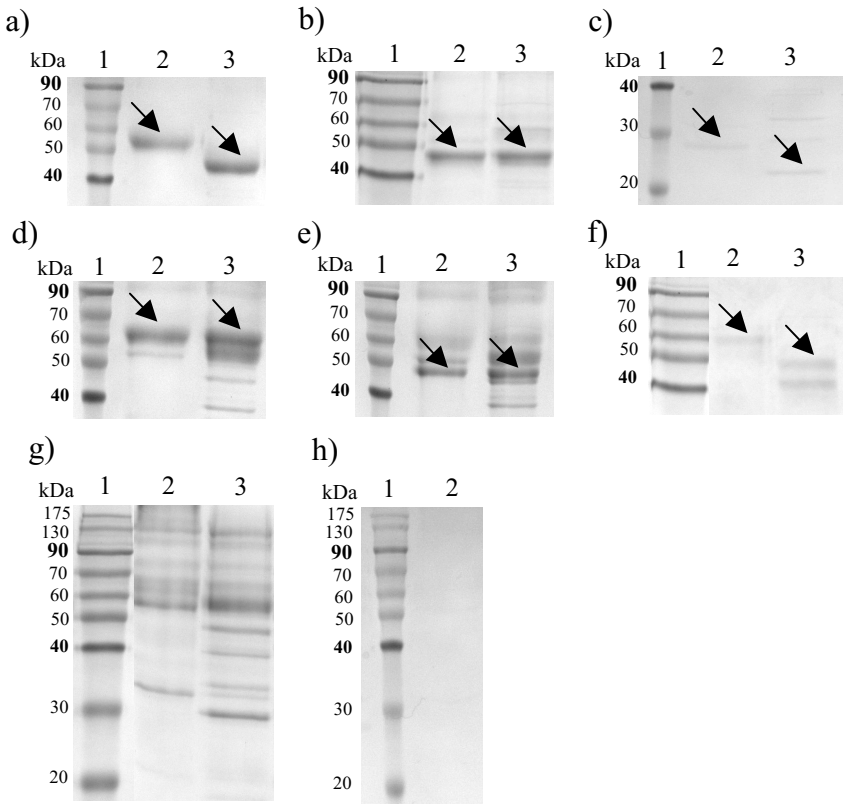


Figure 3.6 – Deglycosylation of culture filtrates produced by recombinant *S. cerevisiae*. *S. cerevisiae* transformed with p425-TEF_M harbouring, *StCel5Aopt* ORF, Panel a; *preproAfCel5A1opt*, Panel b; *GtCel12Aopt* Panel c; *preApCel5Aopt*, Panel d; and *AfCel7B1opt*, Panel e. Filtrates used for Panels a, b, and c, were concentrated 20 times whereas filtrates for panels d and e were concentrated 100 times. Fetuin was used as a positive control, Panel f; g) VTO 100 times concentrated; and h) VTO 20 times concentrated. Molecular weight marker is in lane 1 panels a – h. Lane 2 panels a – h: glycosylated samples, Lane 3 panels a – g: deglycosylated samples. VTO and BSA standards were included on each gel (a – e) but were not included here.

SDS-PAGE images (Figure 3.6) were used to determine the amounts of secreted StCel5Aopt, GtCel12Aopt, AfCel7B1opt, preApCel5Aopt, and prepreAfCel5A1opt present in the culture filtrates (as described in Materials and Methods 2.3.11). The amounts of secreted endoglucanase produced were determined to be 5.7, 0.6, 0.4, 1.1, and 2 μg/ml, respectively. Based on the EG activity levels per ml of culture and the amount of secreted EG protein produced the specific activities of the 5 EGs were determined (Figure 3.7).

Sufficiency analysis (van Zyl et al. 2007) shows that if an attempt to reconstruct the *T. reesei* cellulase system in *S. cerevisiae*, even if EG accounted for all the cellulase system that is not CBH, EG would need to be produced at ~0.3% of total cell protein (van Zyl et al. 2007). The percentage of total cell protein production by the EG was calculated where 1 OD600 = 0.322 g DCW/liter (± 0.019) and total cell protein = 42% DCW (Lange and Heijnen 2001). StCel5Aopt, GtCel12Aopt, AfCel7B1opt, preApCel5Aopt, and preproAfCel5A1opt constituted 0.75, 0.068, 0.054, 0.17, and 0.32 % of total cell protein. These results show that only StCel5Aopt and preproAfCel5A1opt were expressed at levels comparable to that necessary to reconstruct the *T. reesei* cellulase system in *S. cerevisiae*, although production of preApCel5A was close at about 50% of sufficiency levels estimated based on the *T. reesei* system.

The activity levels (Figure 3.5) and protein production levels per ml determined above were used to calculate the specific activities of selected EGs (Figure 3.7). AfCel7B1opt at 130 U mg^{-1} had the highest specific activity, GtCel12Aopt at 85 U mg^{-1} had the second highest specific activity and StCel5Aopt at 9.8 U mg^{-1} had the lowest specific activity. ApCel5Aopt at 20.9 U mg^{-1} and preproAfCel5A1opt at 27.3 U had specific activities than were significantly higher than StCel5Aopt ($P < 0.05$) and significantly lower than AfCel7B1opt and GtCel12Aopt ($p < 0.01$).

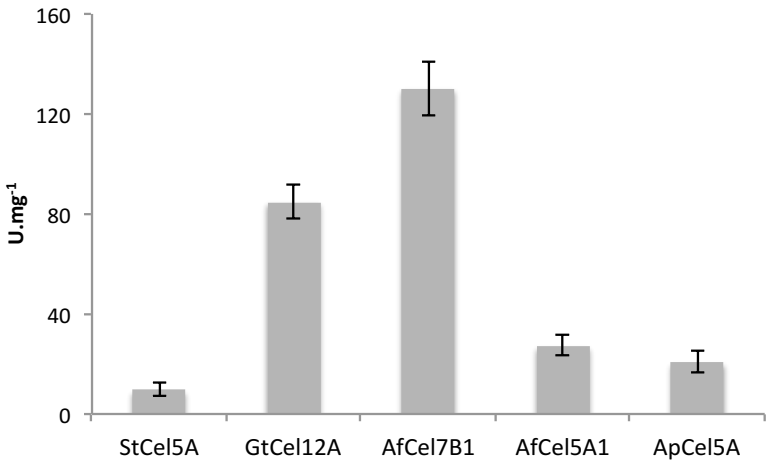


Figure 3.7 – EG specific activity on CMC-4M. One unit is the amount of enzyme required to release 1 μmol of reducing sugar ends per minute under assay conditions.

3.3. *S. cerevisiae* Strains Expressing δ-integrated TEF1 Tr-AnBGL1-CYC1 Tr and Yeast 2 micron Plasmid Borne EGs

The growth rates of *S. cerevisiae* strain CEN.PK111-61A-δ-AnBgl1a on cellobiose and glucose were essentially the same (Figure 3.1 and Table 3.1). The second step in developing a CBP competent *S. cerevisiae* was to develop a derivative of the CEN.PK111-61A-δ-AnBgl1a strain that not only grows efficiently on cellobiose but also grows efficiently on CMC.

Towards developing a strain capable of growth on CMC CEN.PK111-61A-δ-AnBgl1a was transformed with the following plasmids: p425-TEF_M-AfCel7B1opt, p425-TEF_M-GtCel12Aopt, p425-TEF_M-StCel5Aopt, p425-TEF_M-preApCel5Aopt, TEF_M-preproAfCel5A1opt, and p425-TEF_M (VTO). The resulting *S. cerevisiae* strains were tested for their ability to grow using glucose, cellobiose, or CMC-4M as the carbon source.

3.3.1. Expression levels

CEN.PK111-61A-δ-AnBgl1a transformed with p425-TEF_M-AfCel7B1opt, p425-TEF_M-GtCel12Aopt, p425-TEF_M-StCel5Aopt, p425-TEF_M-preApCel5Aopt, p425-TEF_M-preproAfCel5A1opt, and p425-TEF_M (VTO) expressed 1.1, 0.82, 0.32, 0.60, 0.51, and 0.77 μg/ml, of secreted AnBgl1, respectively (Figure 3.8). The amount of secreted AnBgl1 produced by these strains is higher than the 0.38 μg/ml obtained with CEN.PK111-61A-δ-AnBgl1a (Figure 3.2) when it did not harbor a plasmid, with one exception, the p425-TEF_M-StCel5Aopt transformant. The generally higher levels of expression could be because CEN.PK111-61A-δ-AnBgl1a, which required leucine supplementation, plateaued at an OD_{600} of about 4 versus an OD_{600} of about 6 when they did not require leucine.

The amount of secreted EG produced by the p425-TEF_M-StCel5Aopt, p425-TEF_M-preApCel5Aopt, and p425-TEF_M-preproAfCel5A1opt transformants was 0.90, 0.39, and 0.33 µg/ml, respectively, whereas the p425-TEF_M-AfCel7B1opt and p425-TEF_M-GtCel12Aopt transformants did not express detectable amounts. As expected, an EG protein band was not detected in transformants obtained using p425-TEF_M, the expression vector lacking an insert. These results show that levels of secreted EG produced by CEN.PK111-61A-δ-AnBgl1a transformed with p425-TEF_M-StCel5Aopt, p425-TEF_M-preApCel5Aopt, and p425-TEF_M-preproAfCel5A1opt did not attain EG sufficiency levels as determined for a CBP competent *S. cerevisiae* (van Zyl et al. 2007).

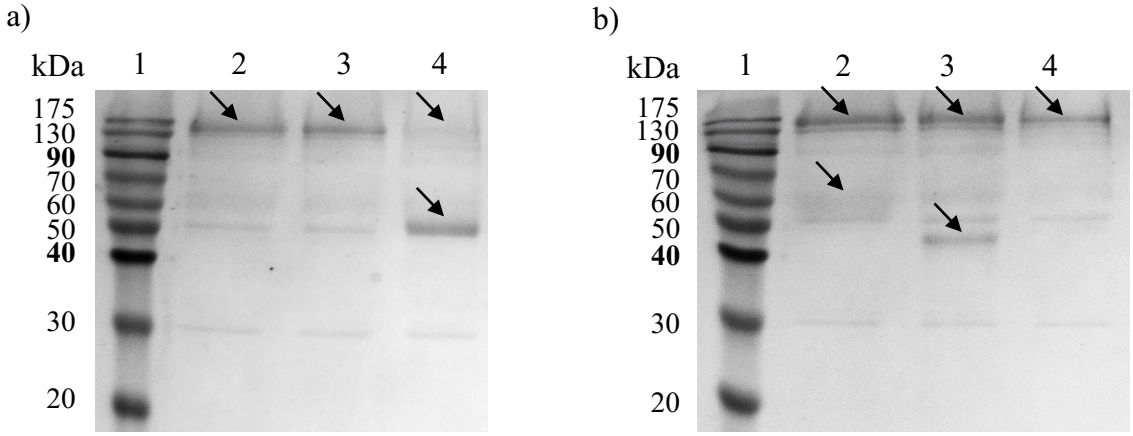
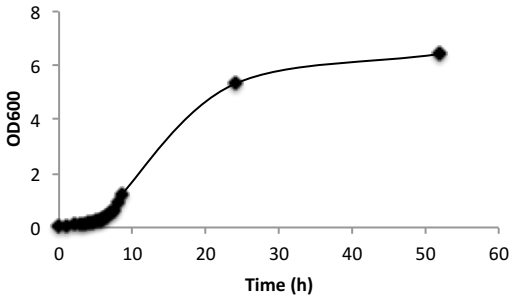


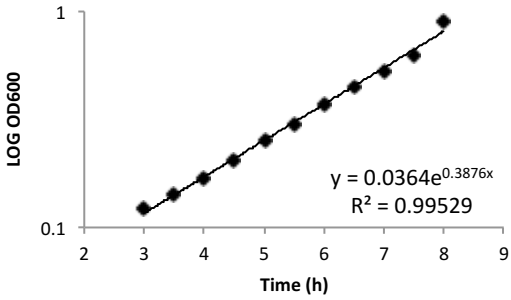
Figure 3.8 – SDS-PAGE analysis of culture filtrates of *S. cerevisiae* expressing δ-integrated *AnBgl1* and plasmid-borne EG. SDS-PAGE of culture filtrates produced by *S. cerevisiae* strain CEN.PK111-61A-δ-AnBgl1a transformed with p425-TEF_M harbouring *AfCel7B1opt* (panel a, lane 2), *GtCel12Aopt* (panel a, lane 3), *StCel5Aopt* (panel a, lane 4), *preApCel5Aopt* (panel b, lane 2), *preproAfCel5A1opt* (panel b, lane 3), and p425-TEF_M alone (panel b, lane 4). AnBgl1 bands in lanes 2, 3 and 4 of panels a and b are indicated by the upper arrows. When present as an identifiable band the EGs are identified by the lower arrow, lane 4 of panel a and lanes 2 and 3 of panel b. These SDS-PAGE gels were used to determine the amount of secreted AnBgl1 and EG protein produced. Protein amounts were determined as described for Figure 3.2.

3.3.2. Growth on glucose, cellobiose, and CMC-4M

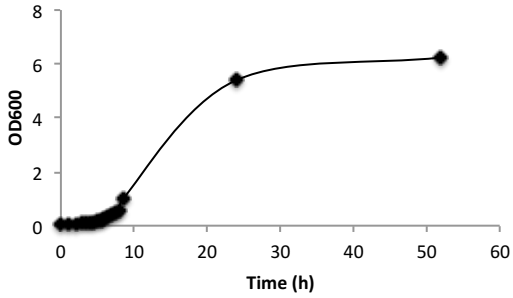
CEN.PK111-61A-δ-AnBgl1a transformed with the p425-TEF_M without and insert and the 5 p425-TEF_M derivatives with the 5 individual EGs as described above (3.2.1) were grown in YNB with glucose as the carbon source. The growth of these strains in YNB liquid cultures was followed for 54 hours (Figure 3.9). The cultures lagged for about 3.5 hours before the initial exponential growth was observed. The growth rates were: 0.39 h^{-1}, 0.37 h^{-1}, 0.35 h^{-1}, 0.35 h^{-1} and 0.28 $h^{-1}$and 0.37 h^{-1} for 425-TEF_M transformants harbouring *AfCel7B1opt*, *GtCel12Aopt*; *StCel5Aopt*; *preApCel5Aopt*; and *preproAfCel5A1opt* and, 0.37 h^{-1} and 1.9 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M (VTO).

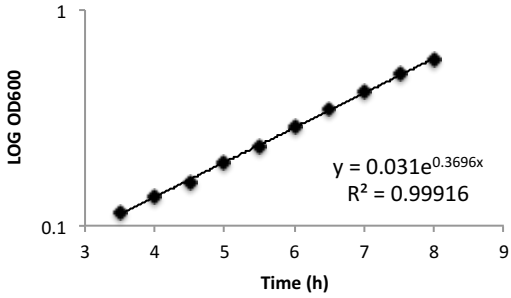
a)


OD600
Time (h)

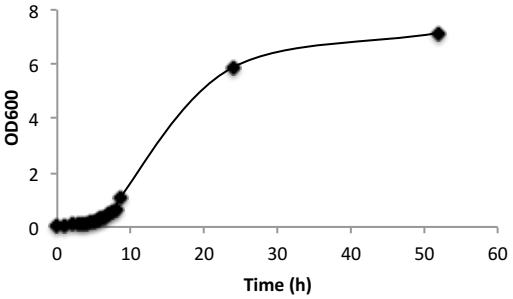

LOG OD600
y = 0.0364e0.3876x
R² = 0.99529
Time (h)

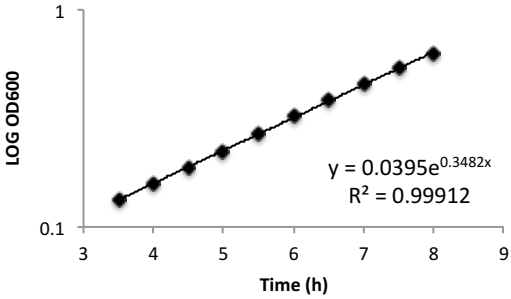
b)


OD600
Time (h)

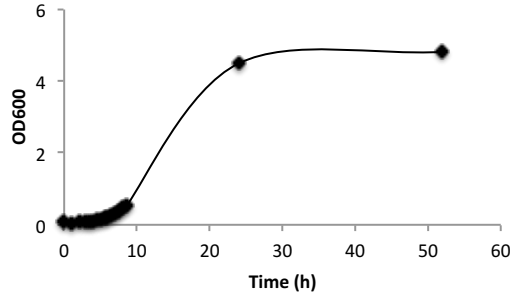

LOG OD600
y = 0.031e0.3696x
R² = 0.99916
Time (h)

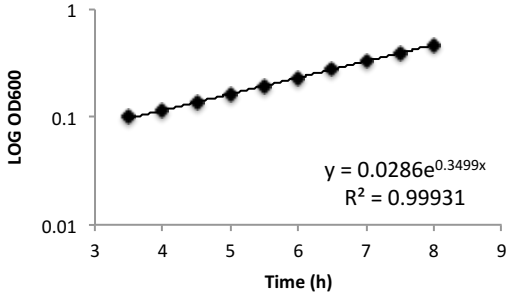
c)


OD600
Time (h)


LOG OD600
y = 0.0395e0.3482x
R² = 0.99912
Time (h)

d)


OD600
Time (h)


LOG OD600
y = 0.0286e0.3499x
R² = 0.99931
Time (h)

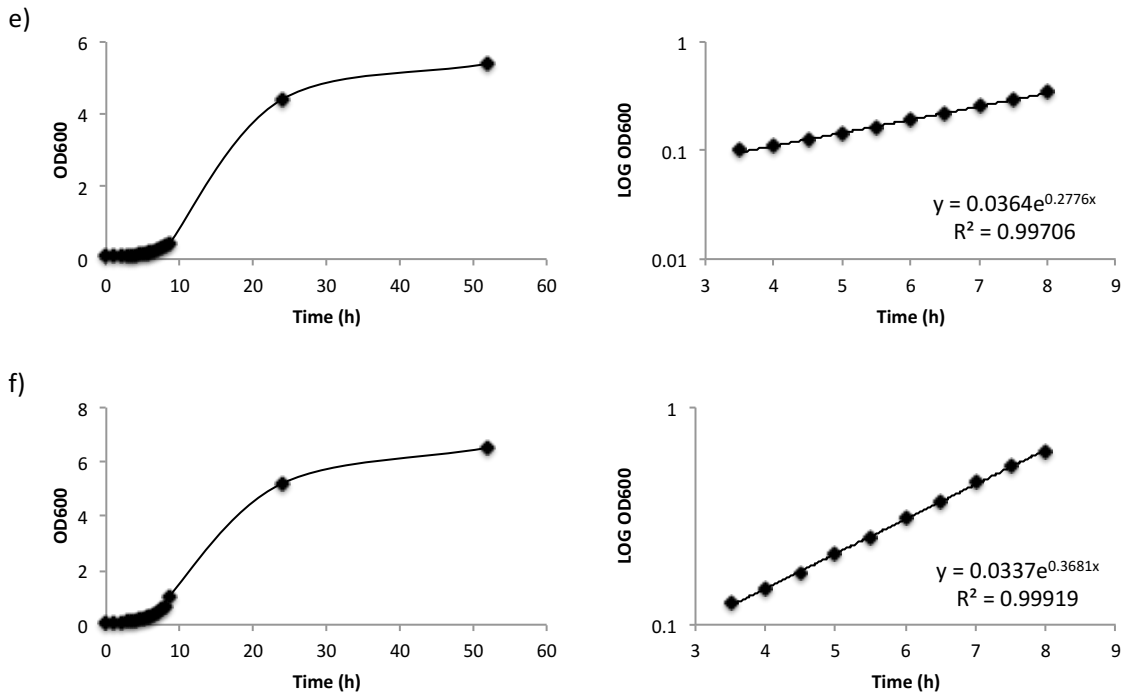
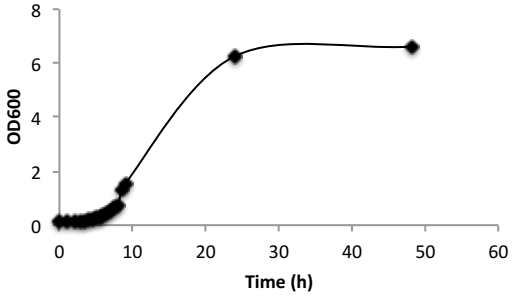
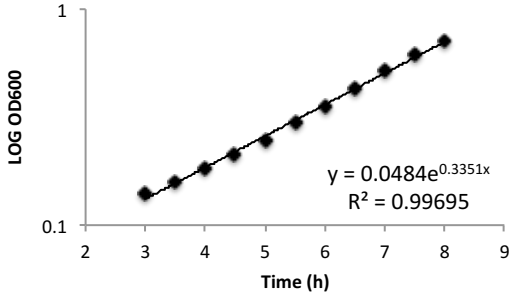


Figure 3.9 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and plasmid borne endoglucanases using glucose. a) *AfCel7B1opt*, b) *GtCel12Aopt*, c) *StCel5Aopt*, d) *preApCel5Aopt*, and e) *preproAfCel5A1opt*. Panel f, growth of CEN.PK111-61A-δ-AnBgl1a transformed with425_M without and insert.

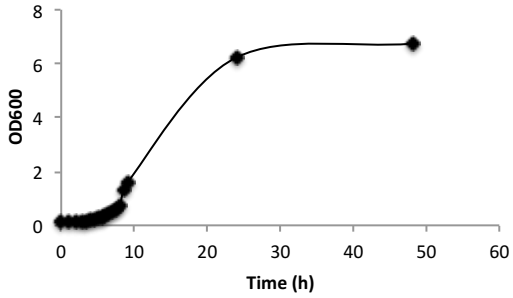
When cellobiose was the carbon source, transformants harbouring p425-TEF_M with *AfCel7B1opt*, *GtCel12Aopt* or *preApCel5Aopt* all had growth rates of 0.33 h^{-1} to 0.34 h^{-1}. These are essentially the same as the growth rate obtained with the CEN.PK111-61A-δ-AnBgl1a parent transformed with p425-TEF_M without an insert (Figure 3.10). In contrast, transformants obtained with p425-TEF_M-StCel5Aopt and p425-TEF_M-preproAfCel5A1opt grew more slowly on cellobiose with growth rates of 0.17 h^{-1} and 0.28 h^{-1}, respectively. The slower growth of these two strains may have resulted from the reduced expression of AnBgl1 that occurred with the p425-TEF_M-StCel5Aopt and p425-TEF_M-preproAfCel5A1opt transformants (see section 3.2.1). These results suggest that AnBgl1 expression levels greater than 0.50 μg ml^{-1} are required to support growth rates on cellobiose that are similar to those observed with glucose as the carbon source.

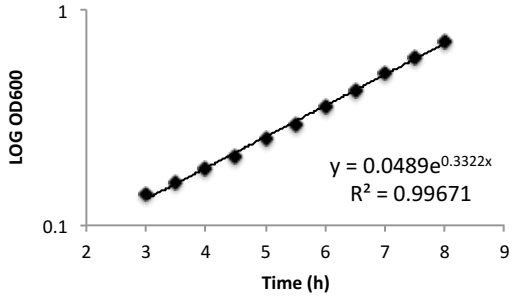
a)


OD600
Time (h)

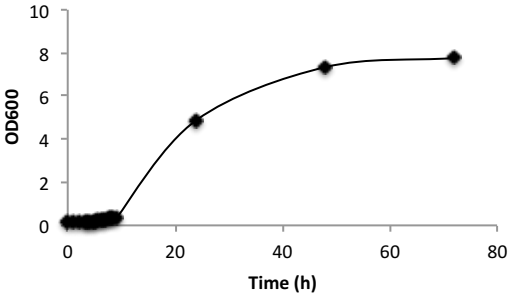

LOG OD600
y = 0.0484e0.3351x
R² = 0.99695
Time (h)

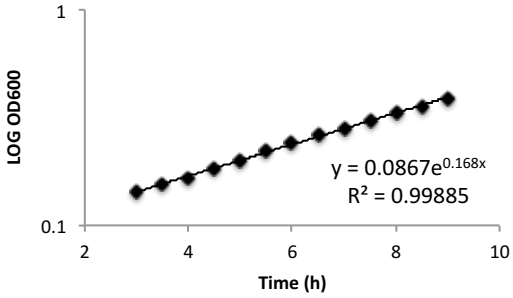
b)


OD600
Time (h)

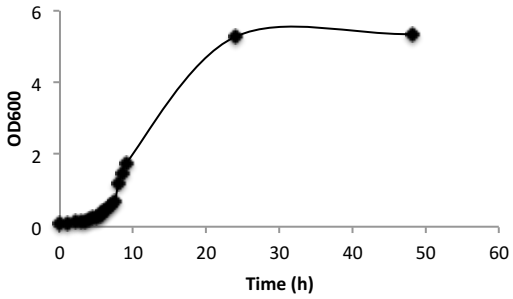

LOG OD600
y = 0.0489e0.3322x
R² = 0.99671
Time (h)

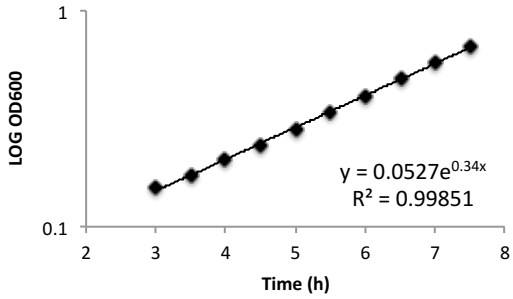
c)


OD600
Time (h)


LOG OD600
y = 0.0867e0.168x
R² = 0.99885
Time (h)

d)


OD600
Time (h)


LOG OD600
y = 0.0527e0.34x
R² = 0.99851
Time (h)

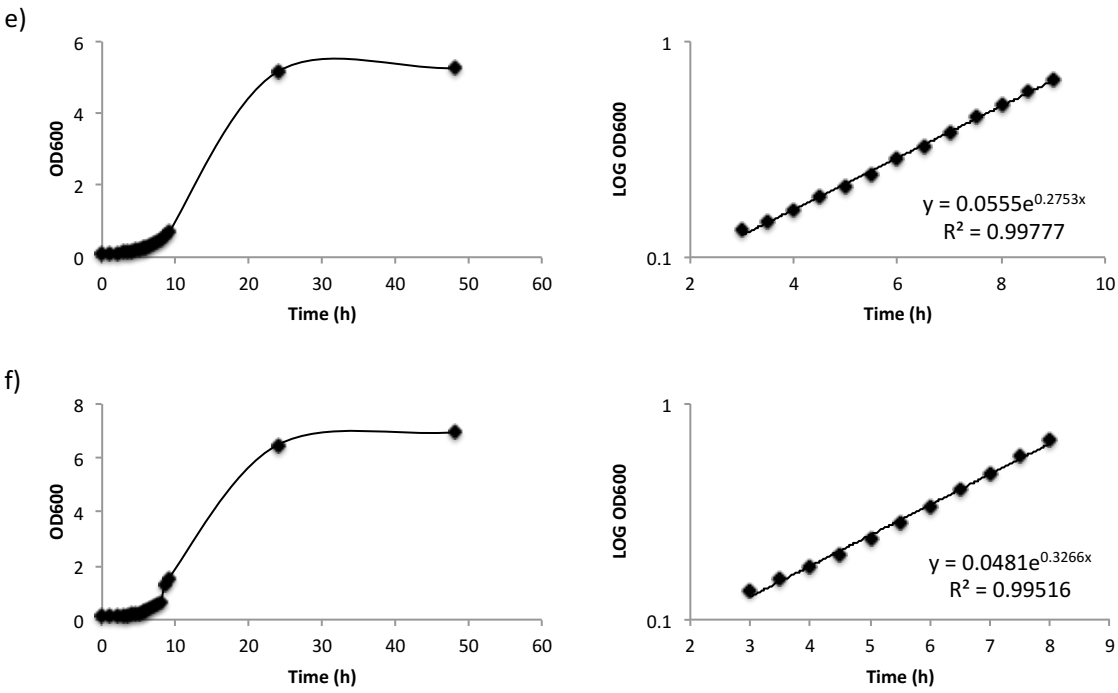
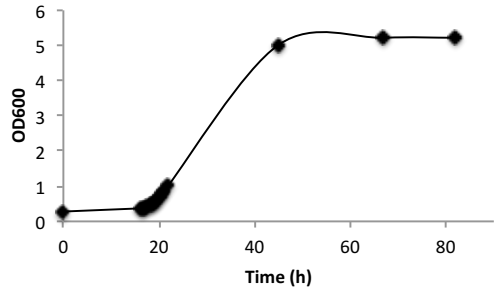


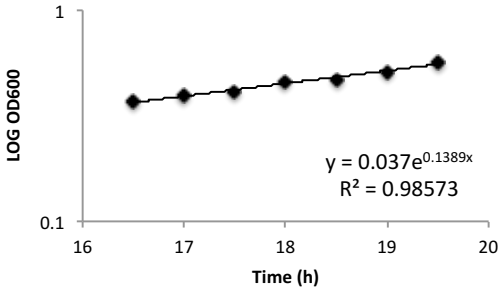
Figure 3.10 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and plasmid borne endoglucanases using cellobiose. a) *AfCel7B1opt*, b) *GtCel12Aopt*, c) *StCel5Aopt*, d) p*reApCel5Aopt*, and e) *preproAfCel5A1opt*.CEN.PK111-61A-δ-AnBgl1a transformed with empty p425_M is used as a control (f).

The growth rates and doubling times of the strains above were also determined using CMC-4M as the carbon source. The growth of these strains using CMC-4M as the carbon source was followed for 80 to 144 h and the growth rates during the exponential growth phase were also determined (Figure 3.11). The R^2 value of the exponential trend line for all the growth rate determinations was greater than 0.98. The calculated growth rates and doubling times for the 6 strains were determined to be: 0.14 h^{-1} and 4.99 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-AfCel7B1opt; 0.2 h^{-1} and 3.46 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_MStCel5Aopt; 0.2 h^{-1} and 3.45 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-preApCel5Aopt; and, 0.33 h^{-1} and 2.12 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-preproAfCel5A1opt. After a 144-hour incubation, CEN.PK111-61A-δ-AnBgl1a _p425-TEF_M-GtCel12Aopt did not grow on CMC-4M. As expected, CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M (VTO), which does not harbour an EG, did not grow with CMC-4M as the carbon source.

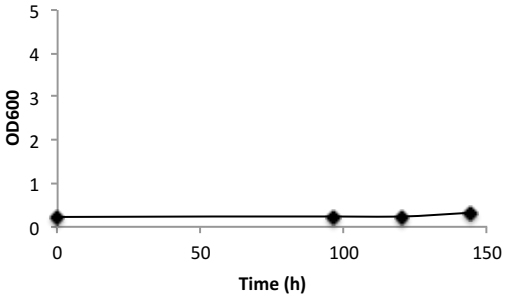
The four strains that were able to grow using CMC-4M exhibited extended lag phases compared to when glucose or cellobiose was the carbon source. The lag phases were about: 16.5 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-AfCel7B1opt; 13.5 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-StCel5Aopt; 28 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-preApCel5Aopt; and, 47 hours for CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M-preproAfCel5A1opt. The extended lag time probably reflect the time required for glucose levels derived from CMC-4M hydrolysis to reach levels sufficient for growth.

a)

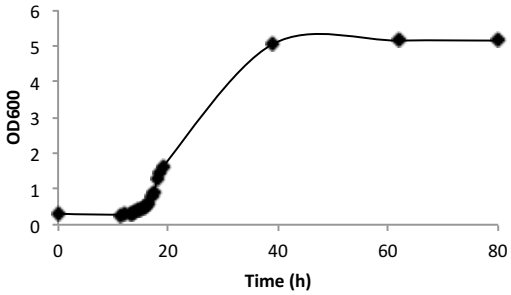

OD600
Time (h)

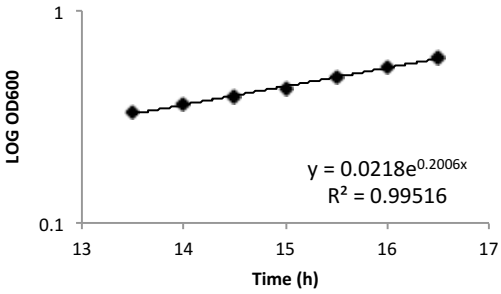

LOG OD600
y = 0.037e0.1389x
R² = 0.98573
Time (h)

b)

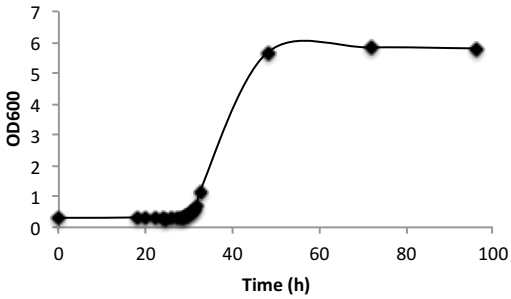

OD600
Time (h)

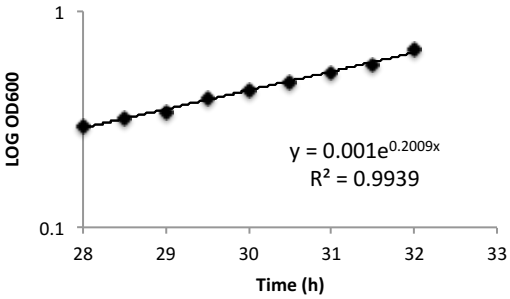
c)


OD600
Time (h)


LOG OD600
y = 0.0218e0.2006x
R² = 0.99516
Time (h)

d)


OD600
Time (h)


LOG OD600
y = 0.001e0.2009x
R² = 0.9939
Time (h)

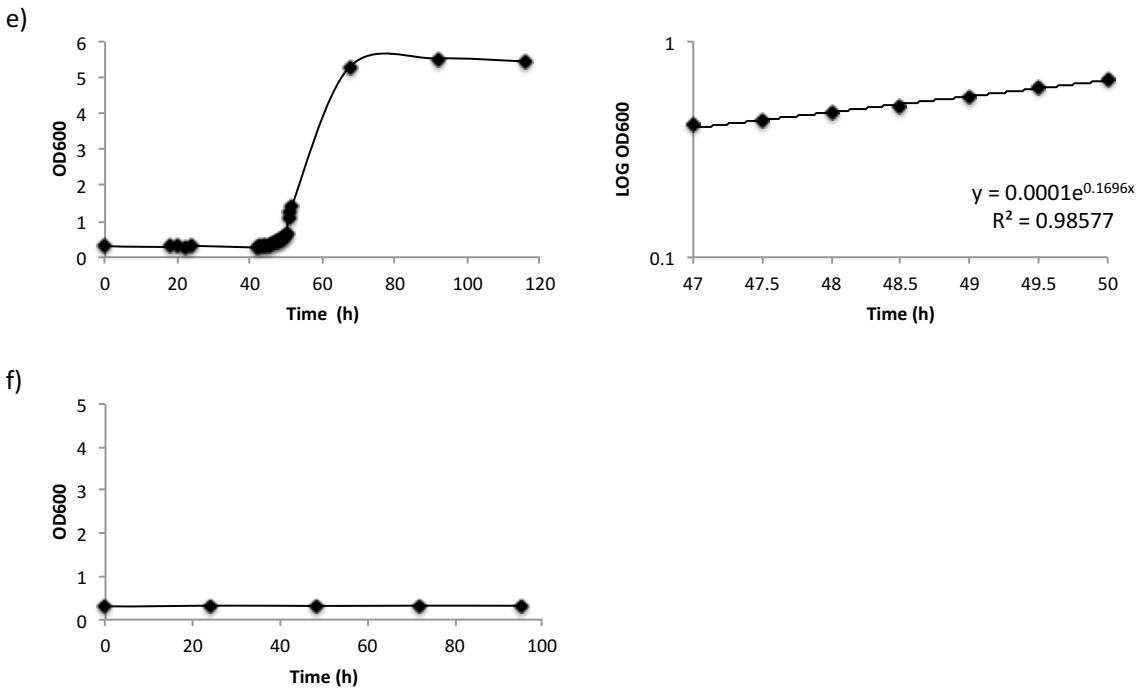


Figure 3.11 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and plasmid borne endoglucanases using CMC-4M. a) AfCel7B1opt, b) GtCel12Aopt, c) StCel5Aopt, d) preApCel5Aopt, and e) preproAfCel5A1opt.CEN.PK111-61A-δ-AnBgl1a transformed with empty p425_M is used as a control (f).

Table 3.2 – Growth rate and generation times of yeast CEN.PK111-61A-δ-AnBgl1a co-expressing δ-*AnBgl1* and either plasmid-borne *AfCel7B1opt*, *GtCel12Aopt*, *StCel5Aopt*, *preApCel5Aopt* or *preproAfCel5A1opt*.

	Glucose		Cellobiose		CMC-4M	
	growth rate μ (h^{-1})	generation time g (h)	growth rate μ (h^{-1})	generation time g (h)	growth rate μ (h^{-1})	generation time g (h)
p425_M-AfCel7B1opt	0.39	1.79	0.34	2.07	0.14	4.99
p425_M-GtCel12Aopt	0.37	1.88	0.33	2.09	NG*	NG*
p425_M-StCel5Aopt	0.35	1.99	0.17	4.13	0.20	3.46
p425_M-preApCel5A	0.35	1.98	0.34	2.04	0.20	3.46
p425_M-preproAfCel5A1opt	0.28	2.50	0.28	2.52	0.17	4.09
p425_M (VTO)	0.37	1.88	0.33	2.12	NG*	NG*

*NG – No growth with CMC-4M as the carbon source.

3.4. *S. cerevisiae* Strains Expressing δ-integrated *AnBgl1* and δ-integrated EGs

There are just over 300 δ-sequences dispersed throughout the haploid S288C genome (Oliveira et al. 2007). To develop strains that do not rely on the maintenance of a plasmid, yeast transposon δ-sequences were used as targeting sequences to integrate the selected endoglucanases, *AfCel7B1opt*, *GtCel12Aopt*, *StCel5Aopt*, *preApCel5Aopt*, and *preproAfCel5A1opt* into the yeast genome. From 50 to 100 integrants were obtained for each of the five endoglucanases. Twenty independent LEU^+ colonies were randomly selected from each transformation and streaked out for single colonies on minimal media plates supplemented with histidine, uracil and CMC-4M as the sole carbon source. One independent integrant from each EG integration experiment representing the transformants that grew the fastest (produced largest colonies after 3 days) with CMC-4M as the carbon source was selected for further analysis.

3.4.1. Expression of AnBgl1 and δ-integrated EGs by strain CEN.PK111-61A-δ-AnBgl1a

The amount of secreted AnBgl1 produced by each of the selected δ-integrated EG transformants, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, CEN.PK111-61A-δ-AnBgl1a _δ-p425-TEF_M-StCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt, and CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M (VTO) during growth with glucose as the carbon source was determined as described for Figure 3.2. The amount of AnBgl1 produced was determined to be 0.75 μg/ml for AfCel7B1opt, 0.38 μg/ml for GtCel12Aopt, 0.94 μg/ml for StCel5Aopt, 0.86 μg/ml for ApCel5Aopt, and 0.77 μg/ml for AfCel5A1opt (Figure 3.12). Thus the expression levels of AnBgl1 by four of the five selected strains were very similar, ranging from 0.75 to 0.94 μg/ml, to the 1.0 μg/ml

AnBgl1 expression levels by CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M, which does not co-express an endoglucanase. AnBgl1 expression by CEN.PK111-61A-δ-AnBgl1a _δ-p425-TEF_M-GtCel12Aopt, however, was reduced about 50% relative to its parent strain.

The amount of secreted EG expressed by the CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt, CEN.PK111-61A-δ-AnBgl1a _δ-p425-TEF_M-preApCel5Aopt, and CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt strains was determined to be 0.75, 0.89, 1.66, 0.38, and 0.57 μg/ml, respectively (Figure 3.12). As expected, detectable EG was not expressed by CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M, the derivative of CEN.PK111-61A-δ-AnBgl1a transformed with the empty δ-p425-TEF_M plasmid.

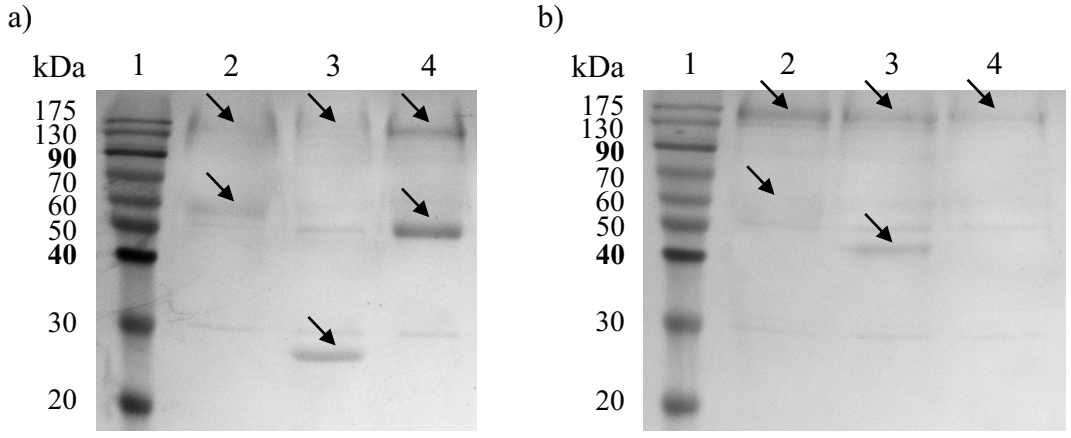


Figure 3.12 – SDS-PAGE analysis of culture filtrates of *S. cerevisiae* expressing δ-integrated *AnBgl1* and δ-integrated EG. SDS-PAGE of culture filtrates produced by *S. cerevisiae* with a δ-integrated *AnBgl1* and transformed with δp425-TEF_M harbouring *AfCel7B1opt* (panel a, lane 2), *GtCel12Aopt* (panel a, lane 3), *StCel5Aopt* (panel a, lane 4), *preApCel5Aopt* (panel b, lane 2), *preproAfCel5A1opt* (panel b, lane 3), and δp425-TEF_M alone (panel b, lane 4). Upper arrows represent AnBgl1 and lower arrows represent EGs.

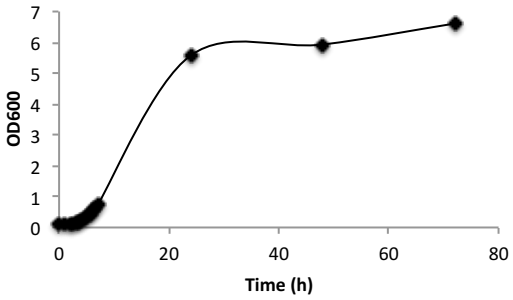
Production levels of AnBgl1 by CEN.PK111-61A-δ-BGL1a with plasmid borne versus integrated versions of the five EGs, was higher when *StCel5Aopt* and *preproAfCel5A1opt* were integrated, lower when *GtCel12Aopt* was integrated, and did not change for *AfCel7B1opt* and *preApCel5Aopt*.

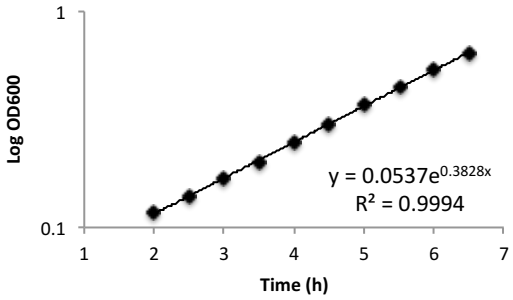
Production of four of the five EGs was at 50% higher when they were integrated. In the case of GtCel12Aopt EG production levels when it was plasmid borne were too low to reveal a detectable band on SDS-PAGE analysis, while 0.89 μg/ml were produced when it was integrated. The fifth EG, ApCel5A was produced at the same level whether integrated or plasmid borne. Interestingly, the combined production of AnBgl1 and the EG by the CEN.PK111-61A-δ-AnBgl1a strain was higher for all five strains when the EG genes were integrated. EG production from 2-micron plasmid borne copies genes versus integrated genes shows that the amount of EG production is not proportional to the gene copy number.

3.4.2. Growth using glucose, cellobiose, and CMC-4M

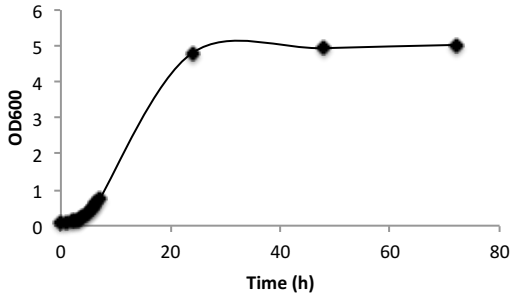
The growth rates and doubling times using glucose as the carbon source were determined for the selected CEN.PK111-61A-δ-AnBgl1a derivative with integrated EGs. A portion of the exponential growth phase between OD_{600} 0.1 and 1.0 was used to determine the growth rate and doubling time of each strain. The growth of these strains using glucose as the carbon source was followed for 72 h and the growth rates during the exponential growth phase were also determined (Figure 3.13). The R^2 value of the exponential trend line used for all the growth rate determinations was greater than 0.99. As observed for the strains with plasmid borne versions of the 5 EGs the exponential growth phase began less than 3 hours after culturing began. The calculated growth rates and doubling times of the five integrated EG strains were determined to be: 0.38 h^{-1}, 1.81 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt; 0.38 h^{-1} and 1.84 for strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt; 0.32 h^{-1} and 2.19 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt; 0.37 h^{-1} and 1.89 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt; and, 0.36 h^{-1} and 1.91 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt. A similar specific growth rate of 0.38 h^{-1} and doubling time of 1.83 h was obtained for the control strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M, a δ-integrant without an EG gene. These results show that all five EG integrant derivatives of CEN.PK111-61A-δ-AnBgl1a and the control strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M grew well with glucose as the carbon source.

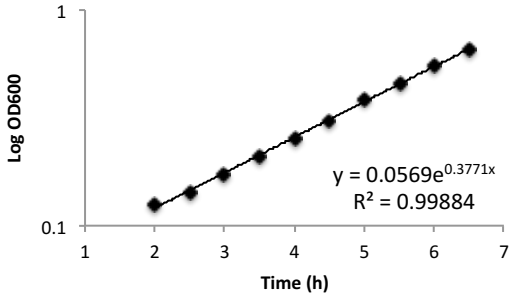
a)


OD600
Time (h)

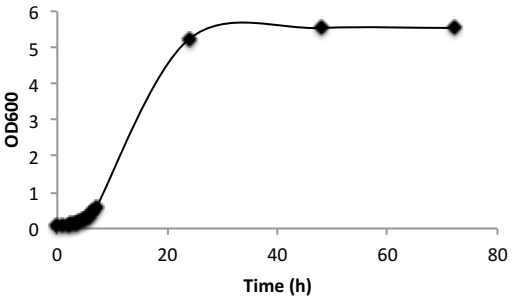

Log OD600
y = 0.0537e0.3828x
R² = 0.9994
Time (h)

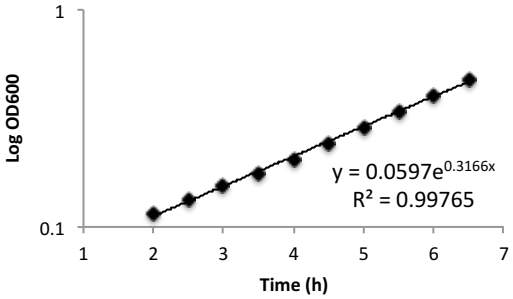
b)


OD600
Time (h)

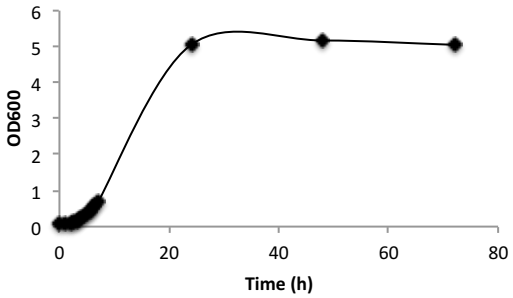

Log OD600
y = 0.0569e0.3771x
R² = 0.99884
Time (h)

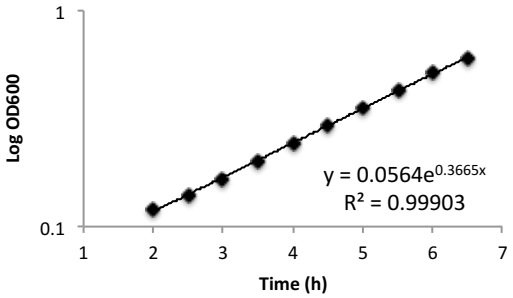
c)


OD600
Time (h)


Log OD600
y = 0.0597e0.3166x
R² = 0.99765
Time (h)

d)


OD600
Time (h)


Log OD600
y = 0.0564e0.3665x
R² = 0.99903
Time (h)

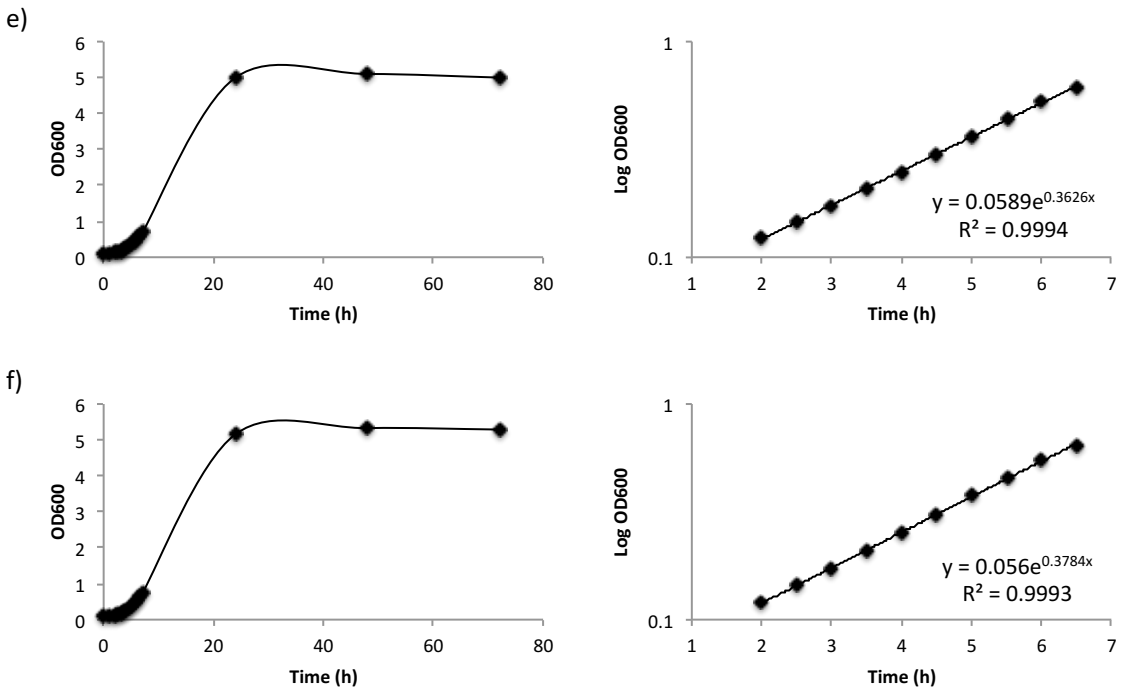
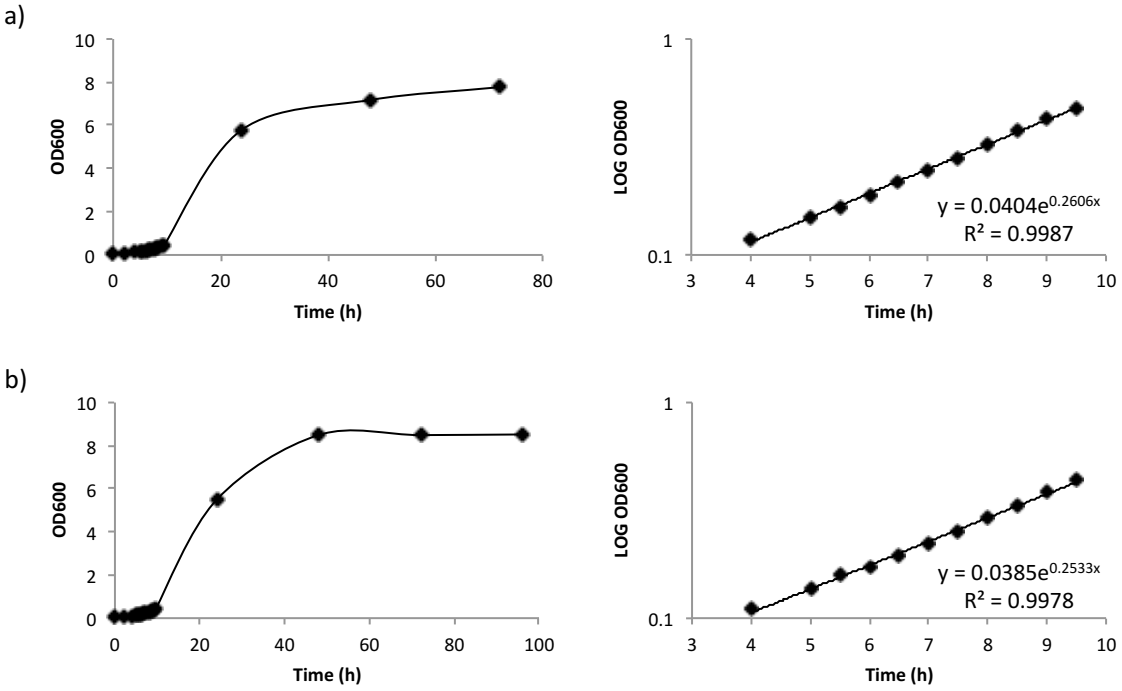


Figure 3.13 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and δ-integrated endoglucanases using glucose. a) *AfCel7B1opt*, b) *GtCel12Aopt*, c) *StCel5Aopt*, d) *preApCel5Aopt*, and e) *preproAfCel5A1opt*. CEN.PK111-61A-δ-AnBgl1 transformed with empty δp425_M is used as a control (f).

The growth rates and doubling times were also determined for the integrated EG strains using cellobiose as the carbon source. A portion of the exponential growth phase between OD_{600} 0.1 and 1.0 was used to determine the growth rate and doubling time. The growth of these strains using cellobiose as the carbon source was followed for 72 to 96 h and the growth rates during the exponential growth phase were also determined (Figure 3.14). The R^2 value of the exponential trend line for all the growth rate determinations was greater than 0.99. The cultures lagged for about 4 hours before growing exponentially. The calculated growth rates and doubling times for 6 strains were determined to be: 0.26 h^{-1}, 2.66 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt; 0.25 h^{-1} and 2.74 for strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt; 0.28 h^{-1} and 2.47 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt; 0.35 h^{-1} and 1.98 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt; 0.34 h^{-1} and

2.07 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt; and, 0.33 h^{-1} and 2.12 h for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M (VTO). These results show that all five EG integrant strains and the control strain CEN.PK111-61A-δ-AnBgl1a _δ-p425-TEF_M, with specific growth rates ranging between 0.26 h^{-1} and 0.35 h^{-1}, grew well having generation times of 2.7 hours or less when using cellobiose as the carbon source. Furthermore, the integrated EGs did not affect the specific growth rates, because, as was observed for the CEN.PK111-61A-δ-AnBgl1a strains with plasmid borne versions of the 5 EGs and the control strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M, cultures of the CEN.PK111-61A-δ-AnBgl1a derivatives with integrated versions of EG genes had similar specific growth rates and entered exponential growth about 3 hours after culturing began.



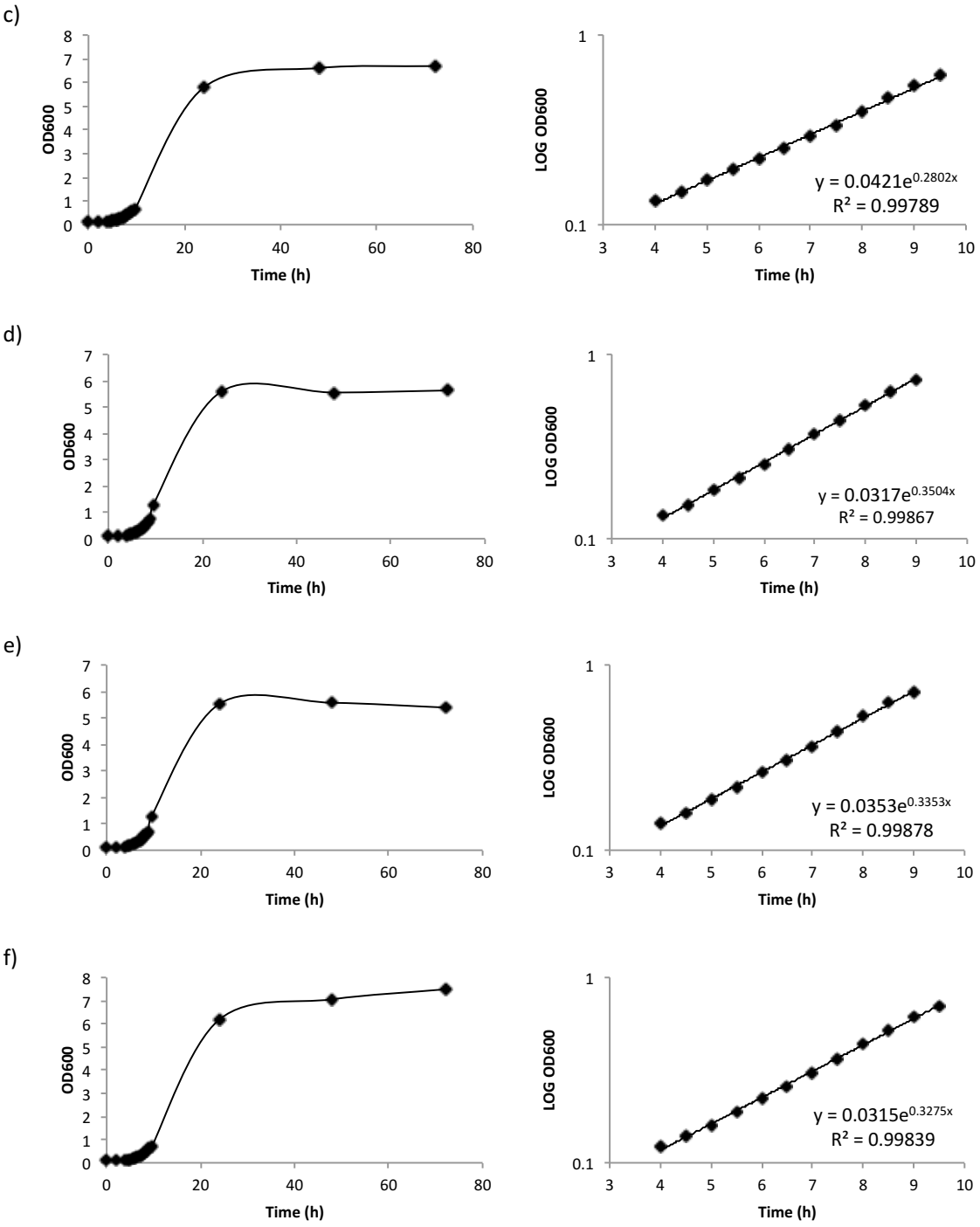


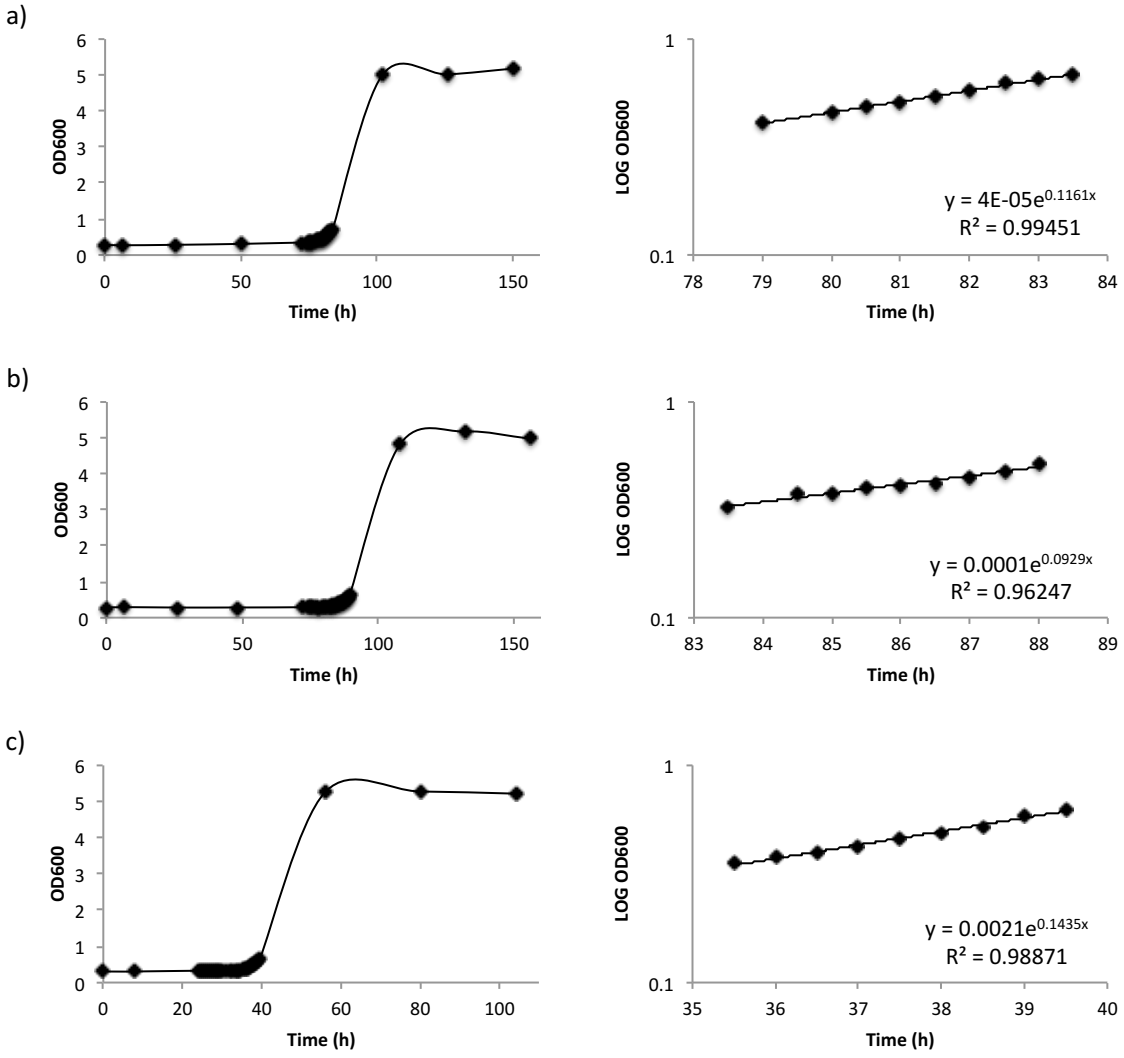
Figure 3.14 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and δ-integrated endoglucanases using cellobiose. a) *AfCel7B1opt*, b) *GtCel12Aopt*, c) *StCel5Aopt*, d) *preApCel5Aopt*, and e) *preproAfCel5A1opt*. CEN.PK111-61A-δ-AnBgl1a transformed with empty δp425_M is used as a control (f).

The growth rates and doubling times were also determined for the above strains using CMC-4M as the carbon source. A portion of the exponential growth phase between OD_{600} 0.1 and 1.0 was used to determine the growth rate and doubling time. The growth of these strains using cellobiose as the carbon source was followed for 96 to 156 h and the growth rates during the initial exponential growth phase were also determined (Figure 3.15). The R^2 value of the exponential trend line for all the growth rate determinations was greater than 0.98 except for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt which was 0.96. The calculated growth rates and doubling times for 6 strains were determined to be: 0.12 h^{-1}, 5.97 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt; 0.09 h^{-1} and 7.46 hours for strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt; 0.14 h^{-1} and 4.83 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt; 0.21 h^{-1} and 3.37 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt; and, 0.11 h^{-1} and 6.06 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt. CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M (VTO) did not grow in CMC-4M (Figure 3.15f).

When CMC-4M was used as the carbon source, these strains exhibited extended lag phases compared to when glucose and cellobiose were used as the carbon source. Exponential growth began at about: 79 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt; 83.5 hours for strain CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt; 35.5 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt; 24.5 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt; and, 50 hours for CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt.

Once exponential growth began the specific growth rates ranged from a low of 0.09 h^{-1} for the AfCel7B1opt integrant to a high of 0.21 h^{-1} for the preApCel5Aopt integrant. These growth rates were similar to those observed for CEN.PK111-61A-δ-AnBgl1a transformed with the plasmid borne versions of the EGs with the exception of

GtCel12Aopt which was unable to support the growth of CEN.PK111-61A-δ-AnBgl1a when it was present on the 2 micron plasmid p425-TEF_M.



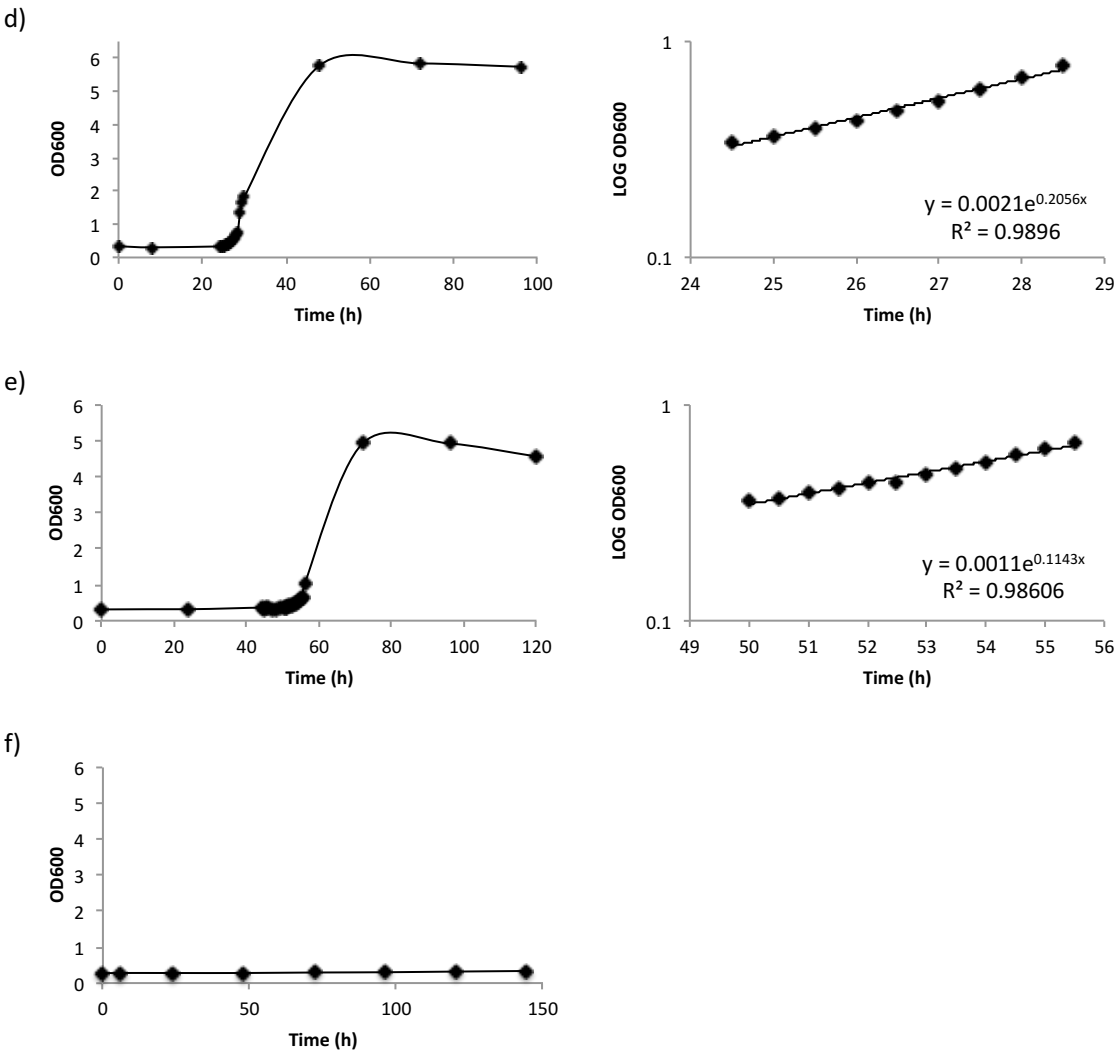


Figure 3.15 – Growth of yeast transformants co-expressing δ-integrated *AnBgl1* and δ-integrated endoglucanases using CMC-4M.a) *AfCel7B1opt*, b) *GtCel12Aopt*, c) *StCel5Aopt*, d) *preApCel5Aopt*, and e) *preproAfCel5A1opt*. CEN.PK111-61A-δ-AnBgl1a transformed with empty δp425_M is used as a control (f).

In summary, the specific growth rates (μ) of CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt, and CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M (VTO) on glucose as the carbon source ranged between 0.32 and 0.38 h^{-1}, while the generation time ranged between 1.81 and 2.19 hours. CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, and CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M (VTO) had the fastest growth rate and the lowest generation time, while CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt had the slowest growth rate and the highest generation time (Table 3.3).

The growth rates on cellobiose as the carbon source ranged from 0.25 to 0.35 h^{-1}, while the generation time ranged from 1.98 to 2.74 hours. CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt had the fastest growth rate and the shortest generation time, while CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt had the slowest growth rate and the longest generation time (Table 3.3).

The growth rates on CMC-4M as the carbon source ranged from 0.09 to 0.21 h^{-1}, while the generation time ranged from 3.37 to 7.46 hours. CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt had the fastest growth rate and the lowest generation time, while CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt had the slowest growth rate and the highest generation time (Table 3.3). As expected, CEN.PK111-61A-δ-AnBgl1a_p425-TEF_M expressing AnBgl1 only, did not grow on CMC-4M. The lag times were consistent when glucose and cellobiose were used as the carbon source. In contrast, a wide range of 24.5 to 83.5 hours of lag time was observed when CMC-4M was used as the carbon source.

Table 3.3 – Growth rate and generation time of yeast CEN.PK111-61A-δ-AnBgl1 co-expressing δ-*AnBgl1* and either δ-integrated EG *AfCel7B1opt*, *GtCel12Aopt*, *StCel5Aopt*, *preApCel5Aopt* or *preproAfCel5A1opt*.

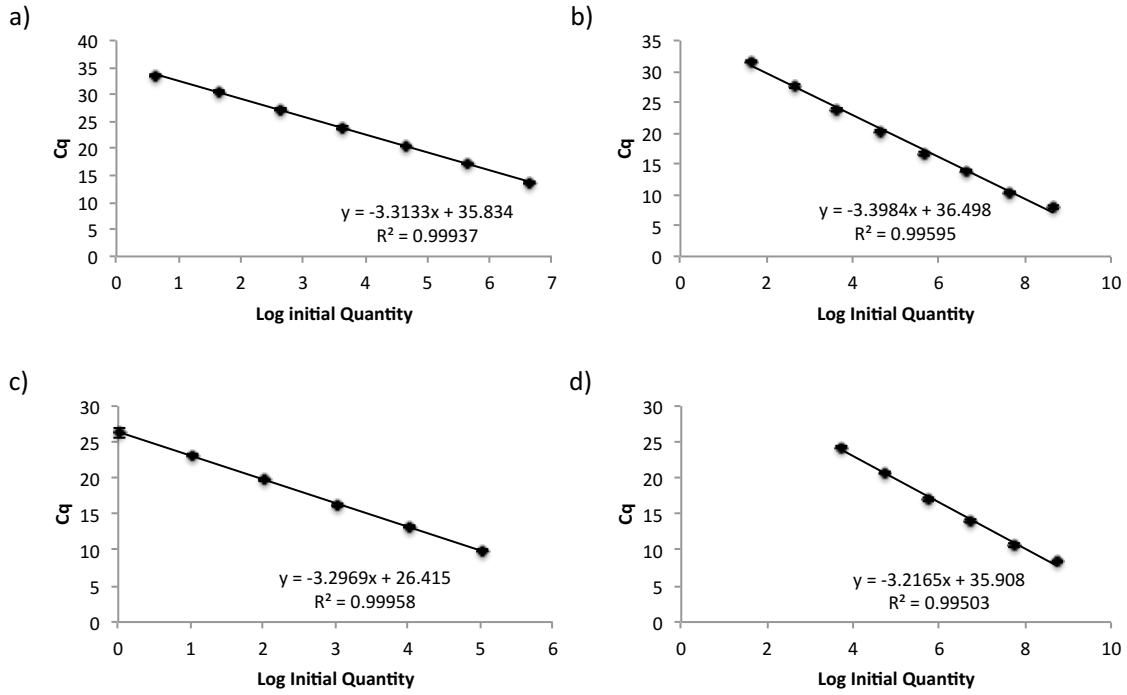
	Glucose		Cellobiose		CMC-4M	
	growth rate μ (h^{-1})	generation time g (h)	growth rate μ (h^{-1})	generation time g (h)	growth rate μ (h^{-1})	generation time g (h)
δ_AfCel7B1opt	0.38	1.81	0.26	2.66	0.12	5.97
δ_GtCel12Aopt	0.38	1.84	0.25	2.74	0.09	7.46
δ_StCel5Aopt	0.32	2.19	0.28	2.47	0.14	4.83
δ_preApCel5A	0.37	1.89	0.35	1.98	0.21	3.37
δ_preproAfCel5A1opt	0.36	1.91	0.34	2.07	0.11	6.06
δ-p425_M (VTO)	0.38	1.83	0.33	2.12	NG*	NG*

*NG – No growth

3.4.3. Copy number quantification using qPCR

The copy number of *AnBgl1* in strain CEN.PK111-61A-δ-AnBgl1a and the number of copies of the EG genes in strains CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt, and CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt were determined by qPCR using the standard curve method and normalization to the housekeeping gene *PGK1*. EvaGreen® was used as the DNA-binding dye for qPCR analysis. Compared to SYBR® Green I, EvaGreen® has much less PCR inhibition, is very stable during PCR and storage conditions, and is shown to be non-mutagenic and non-cytotoxic (Mao, Leung, Xin 2007). The correlation coefficient values (R^2) of the standard curves (Figure 3.16) were ≥ 0.99. The amplification efficiency of the primer pairs used to amplify *AnBgl1*, the 5 selected EG genes and *PGK1* ranged between 96.90 and 105.56% and were all within the efficiency range of 90 to 110% recommended by Illumina (Table 3.4). Absolute gene copy numbers present in each genomic DNA sample were extrapolated from the standard curves (Figure 3.16). The quantification cycle (C_q) value was extrapolated from the amplification plots of qPCR reactions where the standard DNA samples were used as templates (Appendix Figures A1 and A2). The C_q value reflects the fractional cycle at which fluorescence generated with a reaction is high enough to cross the threshold. C_q values were then plotted against the log of the initial quantity of the template DNA used as standard. A linear regression trend line was then added to the plot and its equation was used to extrapolate the initial quantity of the gene of interest and the housekeeping gene in the same genomic DNA sample using the C_q values obtained from the qPCR amplification plots where genomic DNA, with unknown initial amount of gene of interest, were used. Finally, the absolute initial quantity of the gene of interest was divided by the absolute initial quantity of the housekeeping gene to give the copy number for the gene of interest present in the genomes of each strain relative to *PGK1*, the single copy housekeeping gene.

The integrated *AnBgl1* copy number in strain CEN.PK111-61A-δ-AnBgl1a was determined to be 5 copies. The integrated EG copy number in strains CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt, CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B1opt, and CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt were determined to be 1, 2, 1, 1, and 2 copies, respectively (Figure 3.17).



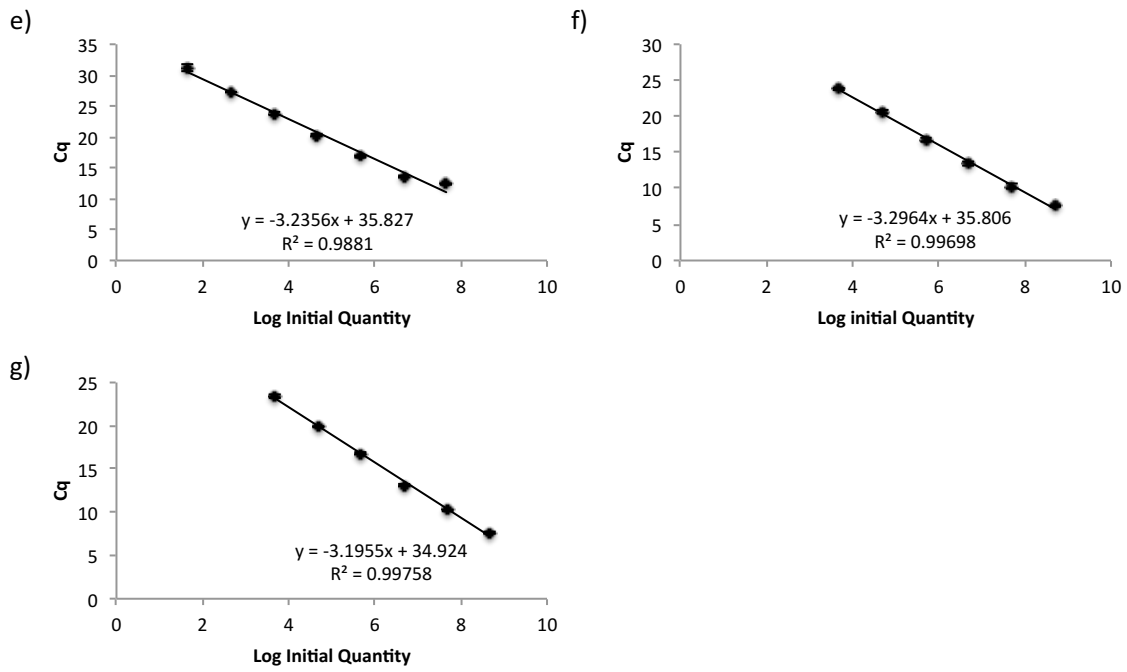


Figure 3.16 – qPCR Standard curves. Standard curves were obtained from at least five 10-fold serial dilutions of DNA samples quantified using PicoGreen. The DNA templates used were a) a gel purified PCR product of *PGK1*, b) p425-TEF_M-AnBgl1 plasmid, c) p425-TEF_M-AfCel7B1opt plasmid, d) p425-TEF_M-GtCel12Aopt plasmid, e) p425-TEF_M-StCel5Aopt plasmid, f) p425-TEF_M-preApCel5Aopt plasmid, and g) p425-TEF_M-preproAfCel5A1opt plasmid.

Table 3.4 – Amplification efficiency of the primer pairs used to detect genes of interest by qPCR

Gene of interest	Primer pair	Slope	Amplification factor	Amplification Efficiency (%)[a]
PGK1	PGK1_F1/ PGK1_R1	-3.2523	2.03	102.99
AnBgl1	AnBgl1_F1/ AnBgl1_R1	-3.3984	1.97	96.90
GtCel12Aopt	GtCel12A_F1/ GtCel12A_R1	-3.2298	2.04	103.99
StCel5Aopt	StCel5A_F1/ StCel5A_R1	-3.2356	2.04	103.73

preApCel5Aopt	ApCel5A_F1/ ApCel5A_R1	-3.2964	2.01	101.08
AfCel7B1opt	AfCel7B_F1/ AfCel7B_R1	-3.2969	2.01	101.06
preproAfCel5A1	AfCel5A_F1/ AfCel5A_R1	-3.1955	2.06	105.56

[a] Amplification efficiencies with the 7 primer pairs ranged from 96.9 to 105.6% and were all well within the efficiency range of 90 to 110% recommended by Illumina.

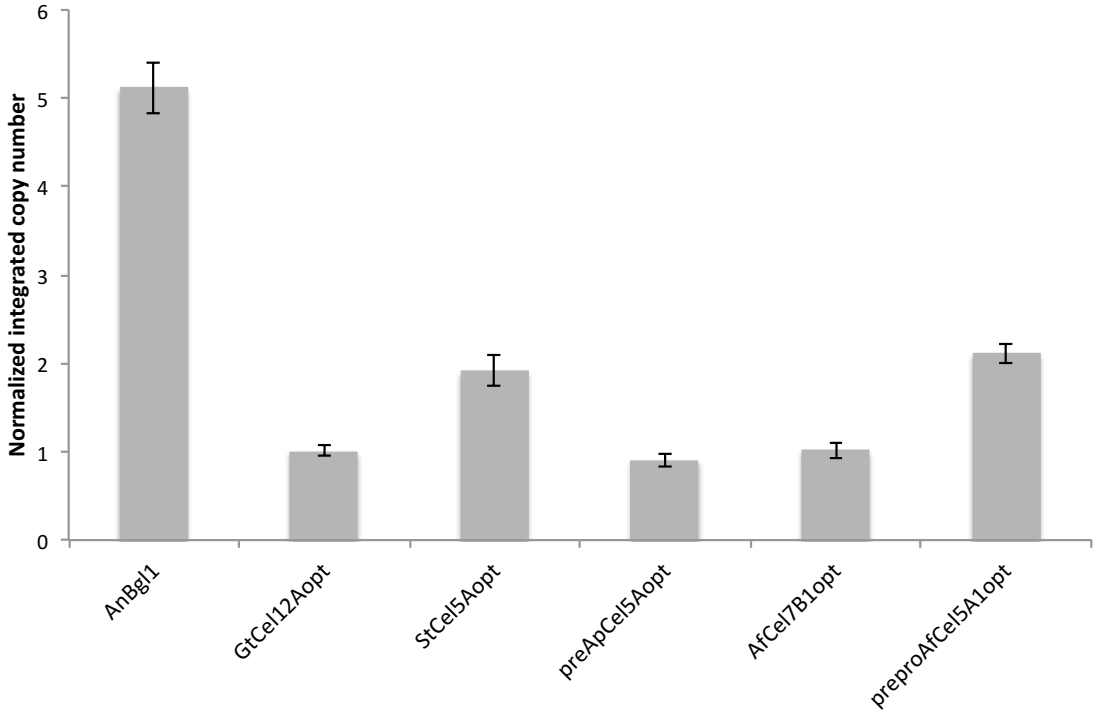


Figure 3.17 – BGL and EG copy number.

3.5. Expression of a Library of Cellobiohydrolases in *S. cerevisiae*

The hydrolysis of the cellulose component of vascular plant biomass requires BGL, EG, and CBH activity (Zhang and Lynd 2004). Having identified a BGL and several EGs that were produced as functional secreted enzymes at sufficient levels to support *S. cerevisiae* growth on CMC-4M, we wanted to identify a CBH enzyme that could be expressed as a secreted and active enzyme at sufficient levels to enable one or more of the 5 strains identified above that could grow on CMC-4M to grow using cellulose as the carbon source. To enable one or more of these strains to produce a cellulase system that can hydrolyze a cellulosic substrate it necessary to also express a cellobiohydrolase (Lynd et al. 2002). Towards this goal 7 fungal cellobiohydrolase ORFs, 4 coding for GH family 6 homologs of *T. reesei* CBH2 and 3 coding for GH family 7 homologs of *T. reesei* CBH1, were cloned into the multi-copy 2-micron expression vector p425_TEF_M. The resulting recombinant plasmids were transformed into strain CEN.PK111-61A and the relative abilities of the 7 cellobiohydrolase ORFs (Table 2.6) to produce functional secreted cellobiohydrolase activity was determined by assaying their ability to hydrolyze phosphoric acid swollen cellulose (PASC).

3.5.1. CBH activity levels on PASC

The ability of CEN.PK111-61A expressing the 7 p425-TEF_M -borne CBH genes to produce functional secreted cellobiohydrolase was determined (Figure 3.18). The results revealed that 4 CBHs, 2 belonging to GH family 7/TrCBH1-like and 2 belonging to GH family 6/TrCBH2-like, produced active cellobiohydrolase. The StCel7A transformant produced the highest level of cellobiohydrolase (CBH) activity ($2.97*10^{-5}$ $U.ml^{-1}.OD_{600}^{-1}$), the *AnCel6A*, *TrCel7A* and *TrCel6A* transformants produced $1.92*10^{-5}$, $1.97*10^{-5}$ and $1.64*10^{-5}$ $U.ml^{-1}.OD_{600}^{-1}$ of cellobiohydrolase activity, *NcCel6A* produced a low level of activity ($0.127*10^{-5}$ $U.ml^{-1}.OD_{600}^{-1}$), and *NcCel7A* and *LeCel6A* both produced less than $0.01*10^{-5}$ $U.ml^{-1}.OD_{600}^{-1}$ of CBH activity.

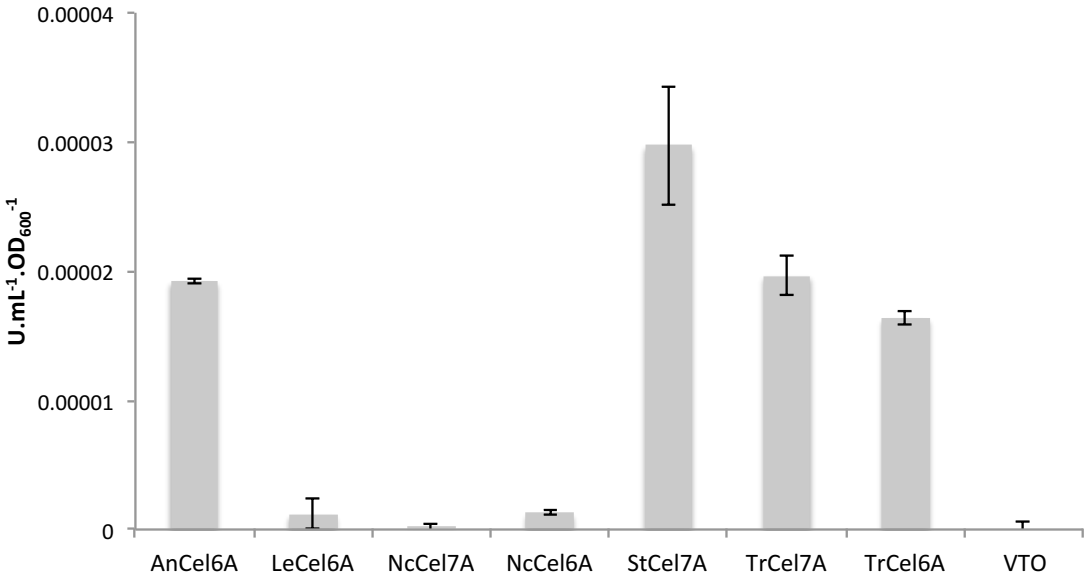


Figure 3.18 – Enzymatic hydrolysis of PASC. Reducing sugar equivalents produced after a 24 h hydrolysis reaction with 0.5% PASC as the substrate and using culture filtrates prepared from CEN.PK111-61A expressing the 7 indicated cellobiohydrolases. The y-axis is the units per ml per OD_{600} where 1 unit is the amount of enzyme required to release 1 µmol of reducing sugar ends per minute under the assay conditions used. Cellobiohydrolases AnCel6A, StCel7A, TrCel7A and TrCel5A produced 13 to 30 times more secreted cellobiohydrolase activity than did the other three other cellobiohydrolases.

3.5.2. CBH expression levels

Assuming signal peptide cleavage and no glycosylation, the predicted molecular masses of StCel7A, AnCel6A, and TrCel6A, were 54, 40.4, and 47.9 kDa, respectively, whereas the observed molecular weights before deglycosylation were 70, 55, and 67 kDa, respectively. After deglycosylation, the observed molecular weights of StCel7A, AnCel6A, and TrCel6A were 67, 51, and 65 kDa, respectively (Figure 3.19). The amount of secreted StCel7A was 0.63 µg/ml, and the amount of secreted AnCel6A and TrCel6A was estimated at 0.3 and 0.44 µg/ml, respectively. TrCel7A expression levels were not high enough to reveal a detectable band (data not show). The proportion of the total cell protein represented by secreted StCel7A, AnCel6A, and TrCel6A was determined to be 0.087%, 0.041%, and 0.058% respectively.

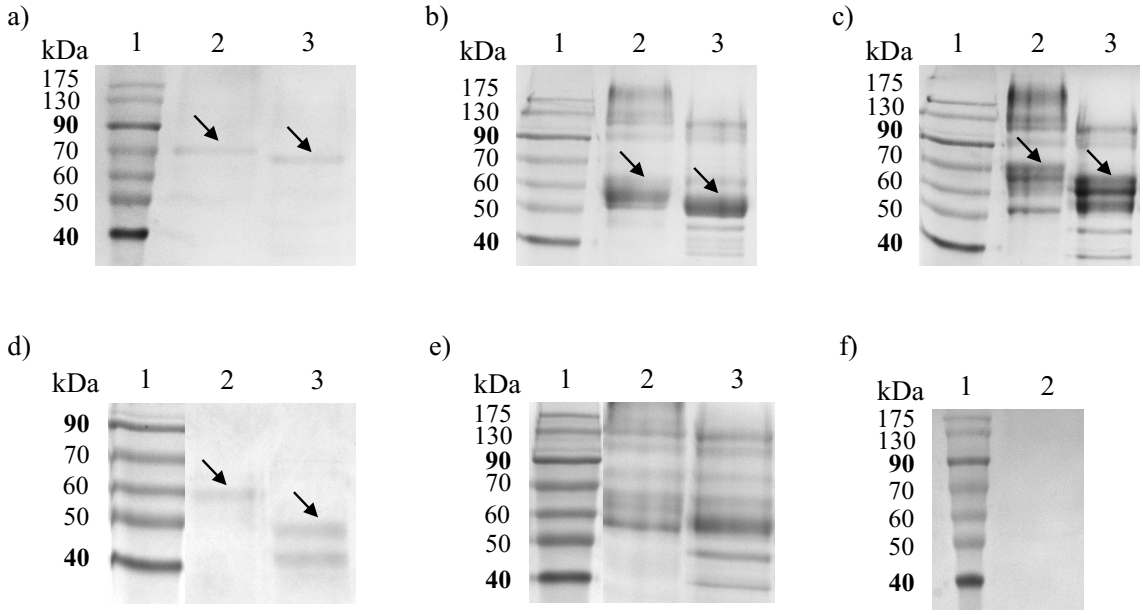


Figure 3.19 – Enzymatic deglycosylation of culture filtrates produced by CBH-expressing CEN.PK111-61A. CEN.PK111-61A transformed with p425-TEF_M harbouring, a) *StCel7A*, b) *AnCel6A*, and c) *TrCel6A*. Panel a, culture filtrates used were concentrated 20 times; Panels b and c, cultures used were concentrated 100 times; Panel d, fetuin was used as a positive control; Panels e and f, culture filtrates of CEN.PK111-61A transformed with p425 TEF_M without an insert concentrated 100 times and 20 times, respectively. Panels a – f lane 1, molecular mass standards; Panels a – f lane 2, glycosylated samples; Panels a – e lane 3, deglycosylated samples (note the fetuin control supplied by NEB has two species of native fetuin).

Discussion

Lignocellulosic biomass is an abundant and renewable commodity that makes an ideal candidate for sustainable bio-ethanol production towards the goal of moving away from the use of finite fossil fuels towards renewable energy. The process of converting lignocellulosic biomass to ethanol starts with its pre-treatment in order to increase the accessibility of glycosylhydrolases to fermentable sugar polymers.

Presently, pre-treated lignocellulosic biomass is converted to ethanol in a four-step process: (1) production of glycosylhydrolase enzymes for polysaccharide hydrolysis; (2) hydrolysis of the polysaccharide component of lignocellulosic biomass into fermentable sugars; (3) fermentation of hexose sugars; and (4) fermentation of pentose sugars (Lynd et al. 2002). Consolidating these processes into a one-step process where pre-treated lignocellulosic biomass is converted to ethanol in a single reactor by a single microbe or a mix of microbes that can both economically produce the enzymes for polysaccharide hydrolysis and ferment hexose and pentose sugars has the potential to dramatically enhance the economics of second generation fuels production.

S. cerevisiae has many characteristics that make it an attractive candidate for the development of a CBP capable host, including; a high ethanol production rate, tolerance to high ethanol concentrations, and US Food and Drug Administration (FDA) assignment as a generally regarded as safe (GRAS) organism. On the other hand, *S. cerevisiae* is unable to directly convert pre-treated lignocellulose into ethanol, because it lacks the enzymes required to break down the polysaccharide polymers in pre-treated lignocellulose into fermentable sugars, and it is not able to efficiently ferment the pentose sugars contained in hemicellulose (Lynd et al. 2002). Developing *S. cerevisiae* strains for CBP of lignocellulose feedstocks, therefore, requires developing strains that can produce the saccharolytic enzymes required to convert the lignocellulose polysaccharide into fermentable sugars and strain improvement to enhance its ability to ferment pentose sugars (Lynd et al. 2002).

This work describes: the construction of a recombinant *S. cerevisiae* strain capable of using cellobiose as the sole carbon source; the identification and cloning of a library of heterologous fungal endoglucanases; the screening of the library of endoglucanases to identify endoglucanases that could be functionally expressed by *S. cerevisiae*; the codon optimization of 5 of the functionally expressed endoglucanase ORFs; the construction of recombinant *S. cerevisiae* strains capable of using CMC-4M as their sole carbon source; and the screening of 7 heterologous fungal cellobiohydrolase ORFs for ones that could produce functional secreted cellobiohydrolase when expressed in *S. cerevisiae*.

4.1. An *S. cerevisiae* Strain That Grows Well on Cellobiose

Wild-type *S. cerevisiae* is unable to grow on cellobiose as a sole carbon (Figure 3.1). BGL expression by *S. cerevisiae* enables its growth on cellobiose (Guo et al. 2011; McBride et al. 2005; Van Rooyen et al. 2005; Wilde et al. 2012). The β-glucosidase *AnBgl1* from *A. niger* was expressed at sufficient levels to sustain the growth of *S. cerevisiae* using cellobiose as the sole carbon source. Although the growth rate of this strain, CEN.PK111-61A-δ-AnBgl1a, on glucose and cellobiose was very similar, the lag time following transfer of glucose grown cells to fresh media was slightly longer with cellobiose than with glucose (Figure 3.1). The longer lag time with cellobiose as the carbon source can be attributed to fact that the CEN.PK111-61A-δ-AnBgl1a must first produce secreted AnBgl1 to then produce sufficient glucose to support logarithmic growth. That the growth rate of the recombinant strain expressing *AnBgl1* on glucose is roughly the same as that on cellobiose shows that levels of secreted AnBgl1 production by CEN.PK111-61A-δ-AnBgl1a were sufficient to support normal logarithmic growth and that the levels of β-glucosidase production by CEN.PK111-61A-δ-AnBgl1a would support growth with cellulose as the carbon source in the presence of a cellulase system depleted of β-glucosidase activity.

Recently, Larue and colleagues (Larue, Melgar, Martin 2016) used directed evolution to develop a version of *AnBgl1* with improved hydrolytic activity on cellobiose and the

synthetic substrate pNPG and decreased sensitivity to both glucose and cellobiose. To improve the efficiency of CEN.PK111-61A-δ-AnBgl1a, the location of the five integrated copies of AnBgl1 could be identified in future studies and the improved version of *AnBgl1* developed by Larue and colleagues (Larue, Melgar, Martin 2016) could perhaps be targeted to these specific δ-sequences.

4.2. Heterologous Endoglucanase Expression by *S. cerevisiae*

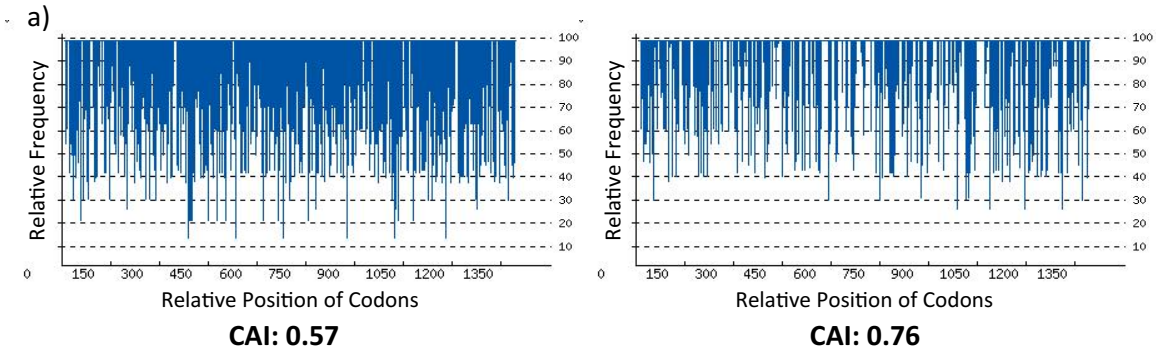
A library of 25 EGs from 10 evolutionary diverse fungal species was expressed in *S. cerevisiae*. The amount of secreted endoglucanase produced varied dramatically across the recombinant EGs even though all the EGs were expressed using the same *S. cerevisiae* strain, the same transcription regulation sequences (the TEF1 promoter region and CYC1 terminator region), and the same episomal plasmid (p425-TEF_M). Other studies have also reported that the ability of *S. cerevisiae* to produce secreted heterologous glycosylhydrolases varies widely. Wilde and colleagues (Wilde et al. 2012) reported a wide range in the amount of secreted protein and activity obtained within a library of 35 genes encoding fungal β-glucosidases belonging to glycosyl hydrolase families 3 and 5. Ilmen et al. (Ilmen et al. 2011) also reported a wide range in the secretion and activity levels of 24 ORFs encoding 14 CBH1 (Cel7A) and 10 CBH2 (Cel6A) cellobiohydrolases of fungal origin. These results and the results herein indicate that some glycosylhydrolases are more compatible with the expression and secretion machinery of *S. cerevisiae* than others.

4.2.1. Coding region optimization

Based on the relative expression levels determined by screening using Congo Red indicator plates, 5 EGs were selected for further study. The selected endoglucanases included the three EGs that were expressed at the highest levels, one that was moderately expressed, and one that was poorly expressed. The effect of coding region optimization on expression levels was highly variable. The highest increase in EG production was observed with codon optimization of *AfCel5A1*, the EG that was expressed at the lowest levels when using its native ORF sequence. Further increases in expression levels could

look at other factors such as the gene copy number, protein folding, rate limiting steps in the secretory pathway, protein degradation due to activation of the unfolded protein response (UPR) pathway, and post-translational modifications.

In an effort to ascertain how coding region optimization improved the expression of *AfCel5A1* but not the other codon optimized EGs, the codon adaptation index (CAI), the GC content, and the codon frequency distribution (CFD) were determined for the ORFs of the 5 selected EGs using the Genscript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis). The use of synonymous codons at different frequencies by different organisms is referred to as codon bias (Bennetzen and Hall 1982). Codon bias variation can vary significantly between different organisms and within organisms between different classes of genes (Gouy and Gautier 1982). The CAI assesses the degree of codon bias within genes (Sharp and Li 1987). The CAI ranged between 0.52 and 0.66 for the native ORFs of the five selected EGs, and between 0.76 and 0.77 for the codon-optimized ORFs (Figure 4.1). Since the CAI of the codon optimized ORFs of all 5 selected EG genes all had CAI values of 0.76 or 0.77 the increased production of AfCel5A1 expression levels compared to the other codon-optimized EGs is unlikely due to its increased CAI for *S. cerevisiae*.



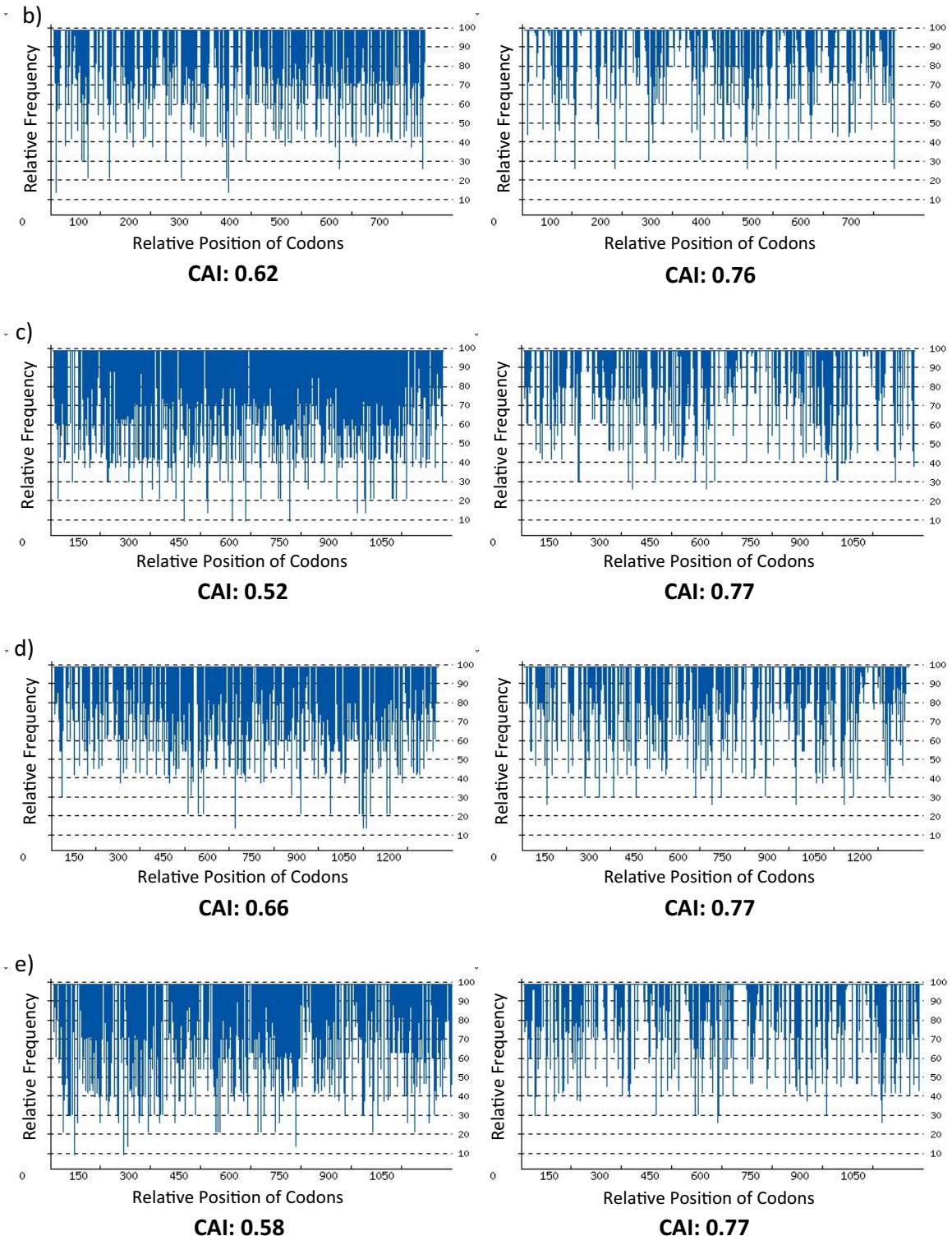
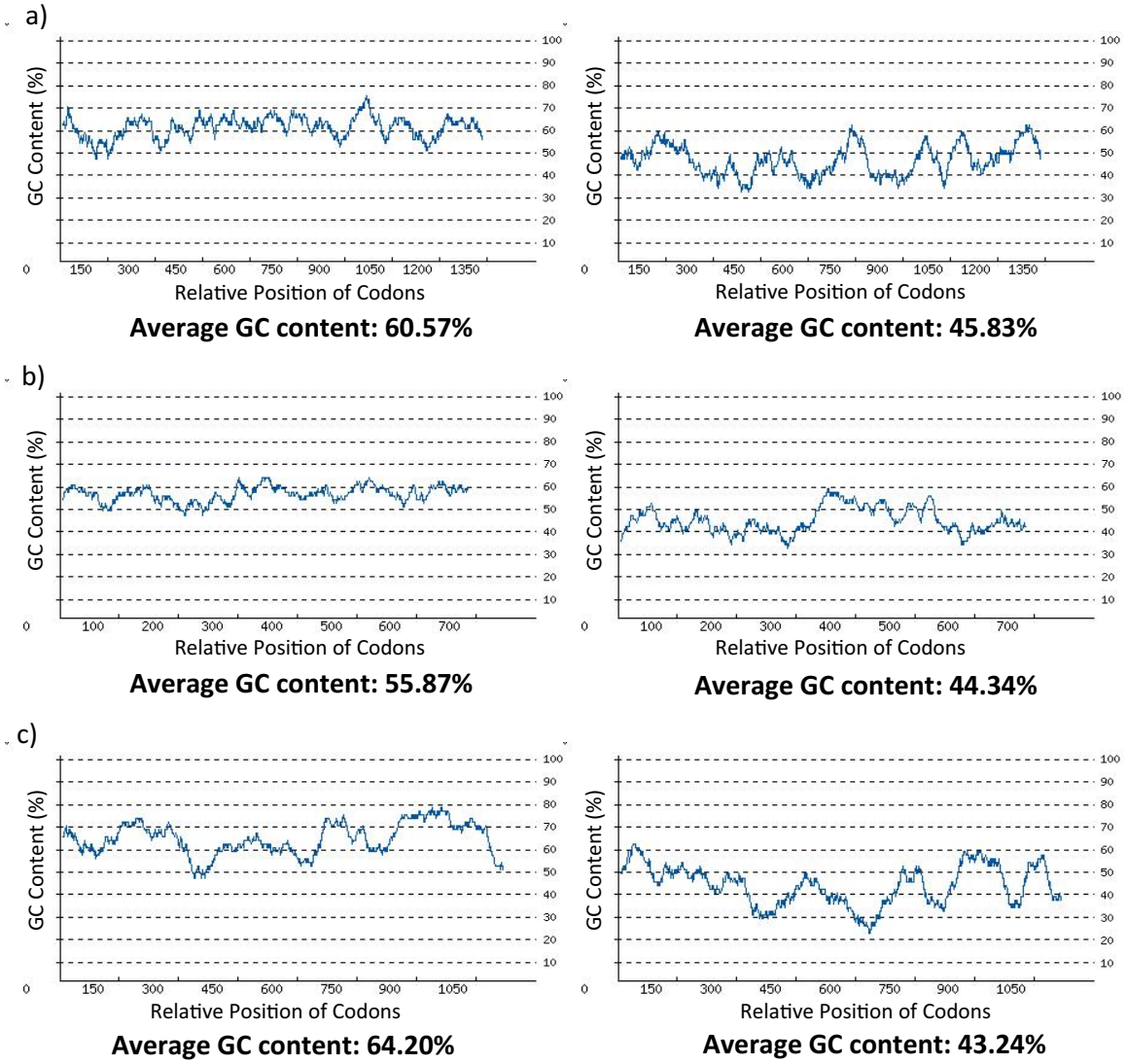


Figure 4.1 – Codon adaptation index analysis. The codon adaptation index of native EG ORF sequences (left) and codon-optimized EG ORF sequences (right) of a) *AfCel7B1*, b) *GtCel12A*, c) *StCel5A*, d) *ApCel5A*, and e) *AfCel5A1*. The CAI values were calculated using the Genscript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis).

The average GC content ranged between 55.9 and 64.2% for the native ORF sequences of the 5 EGs and between 43.2 and 45.9% for the codon-optimized versions of these ORFs (Figure 4.2). The average GC content of the codon-optimized ORFs was much closer to the 42% GC content observed for protein coding regions in *S. cerevisiae* (Kochetov et al. 2002). The variation in the GC content of the native and codon optimized ORFs for *AfCel5A1* were both within the range of the GC contents of the native and codon optimized ORFs of the 5 EGs, thus the GC content adjustment does not seem to be the reason behind the improved production of AfCel5A1.



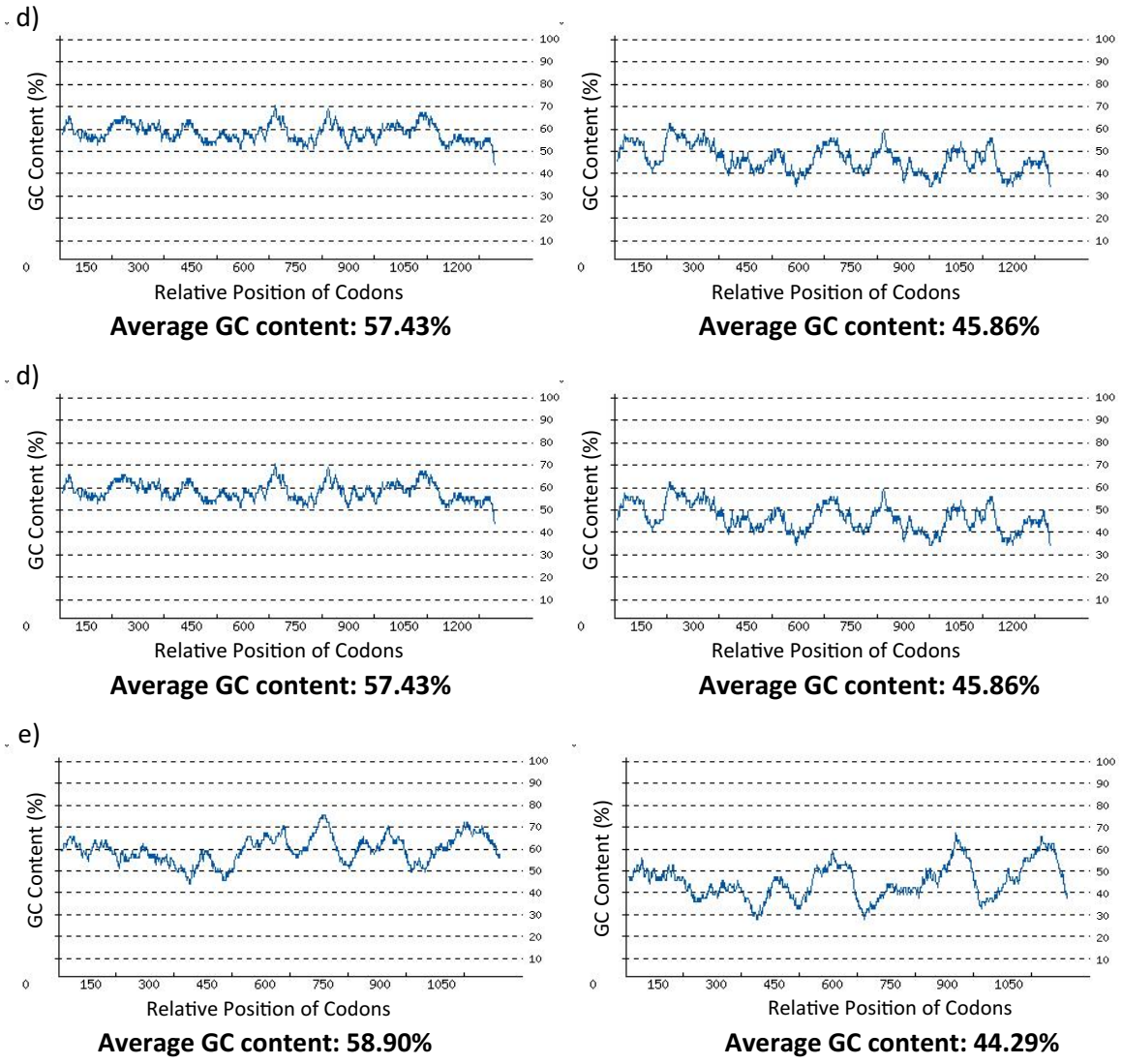
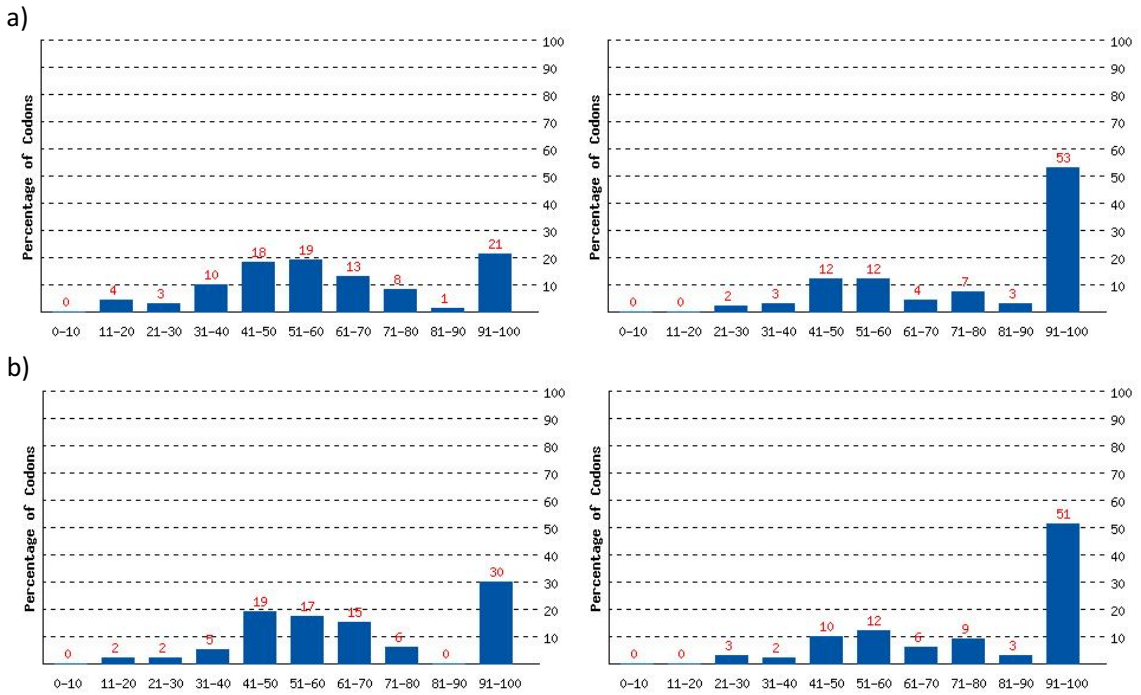


Figure 4.2 – GC content analysis. GC curves of native EG ORF sequences (left) and codon-optimized EG ORF sequences (right) of a) *AfCel7B1*, b) *GtCel12A*, c) *StCel5A*, d) *ApCel5A*, and e) *AfCel5A1*. The GC curves were generated using the Genscript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis).

The codon frequency distribution was determined for the native and codon-optimized ORFs of the selected EGs in order to determine if a high percentage of low frequency codons (synonymous codons that are used less than 30% of the time) were present in the native EG ORF sequences. The percentage of low frequency codons ranged between 3 and 11% for the native ORF sequences, and between 2 and 3% for the codon-optimized

ORF sequences (Figure 4.3). The native ORF sequence of *AfCel5A1* was composed of 11% low frequency codons. This was the highest percentage of low frequency codons among the selected EGs. The codon-optimized ORF sequence of *AfCel5A1* was composed of 2% low frequency codons which corresponds to a 9% decrease in the low frequency codons in the codon-optimized versus native ORF sequence. This decrease in the percentage of low frequency codons in the codon-optimized sequence of *AfCel5A1* was the largest among the 5 EGs subjected to codon optimization. Perhaps the relatively high number of low frequency codons in the *AfCel5A1* ORF, reduced the translation efficiency of its mRNA. Supporting this possibility, low frequency codons can decrease the efficiency of translation by causing ribosomal pausing during elongation in order to wait for the correct and rare corresponding tRNA (Buchan and Stansfield 2007). Pausing during elongation at low frequency codons can also cause the translation machinery to disengage resulting in translation abandonment and low expression levels of the heterologous protein (Buchan and Stansfield 2007).



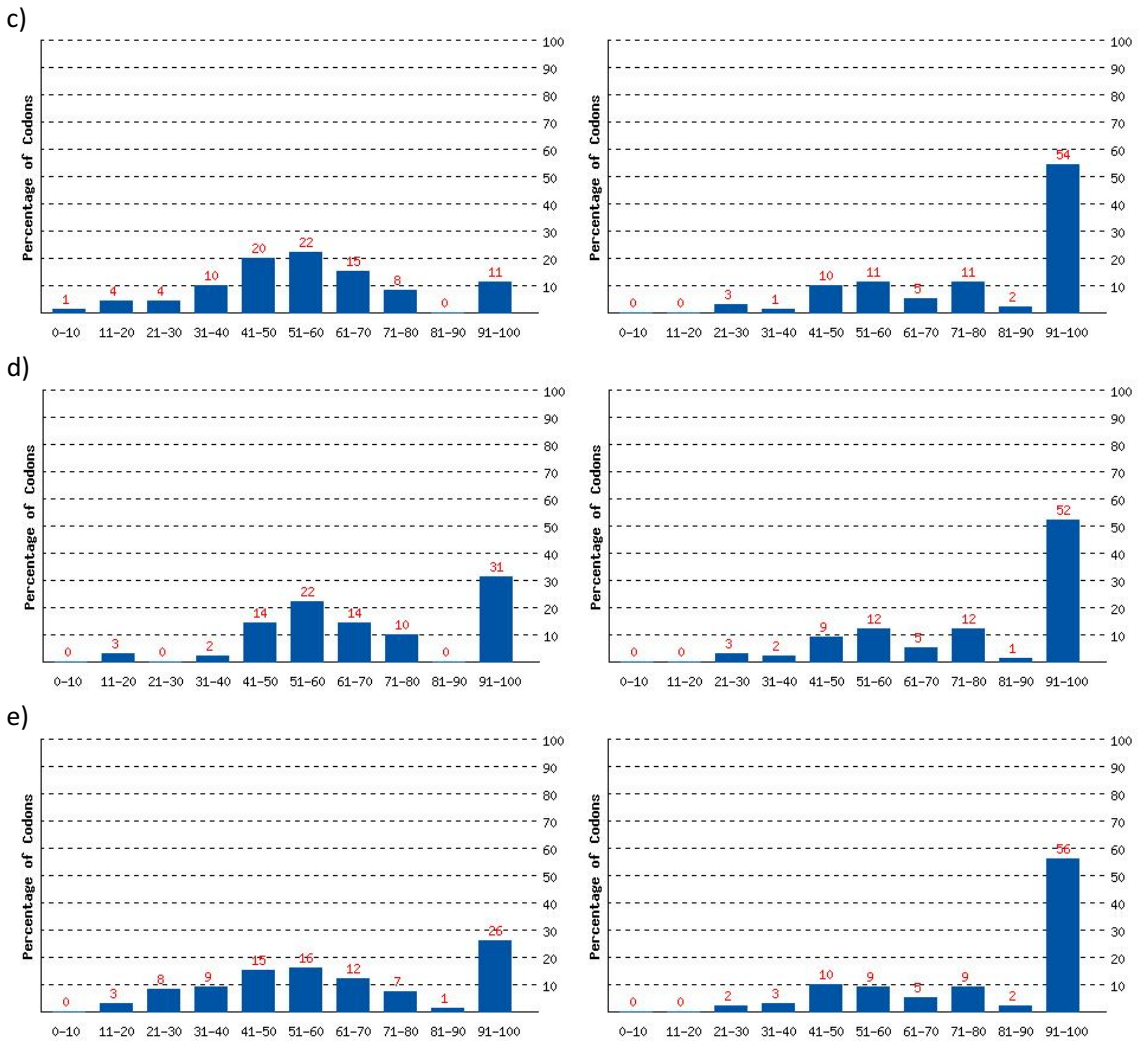


Figure 4.3 – Codon frequency distribution analysis. Codon Frequency Distribution (CFD) of native EG ORF sequences (left) and codon-optimized EG ORF sequences (right) of a) *AfCel7B1*, b) *GtCel12A*, c) *StCel5A*, d) *ApCel5A*, and e) *AfCel5A1*. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in *S. cerevisiae*. The CFD curves were generated using the Genscript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis).

Other research groups also observed variable codon optimization results. The results varied from reduced expression levels (Curran et al. 2013), no change in expression levels (Westfall et al. 2012), and increased expression levels (He et al. 2014; Jia et al. 2012; Zhang et al. 2012). The relative number of low frequency codons may not have been the only factor that resulted in the increased levels of functional secreted AfCel5A1.

This suggests that using codon optimization, as a strategy to improve the expression of heterologous proteins, remains an unreliable strategy.

4.2.2. Signal peptide replacement

Replacing the native signal peptide of target proteins as a strategy to improve heterologous expression levels has been used by many research groups, including (Tang et al. 2013; van Rooyen et al. 2005; van Zyl et al. 2007; Zhu, Yao, Wang 2010). Secretion of a protein requires a targeting signal that directs it to the cell surface. Because the native signal peptides of the selected heterologous EGs may not be efficiently recognized by the *S. cerevisiae* secretory system, the signal peptides of the codon-optimized proteins were replaced by the *S. cerevisiae MFα1* pre-α-factor and prepro-α-factor signal peptides. The results of signal peptide replacement were highly variable. The amount of secreted protein produced decreased or did not change significantly for 9 of the 10 signal peptide replacements that were tested; however, when the native signal peptide of *AfCel5A1opt* was replaced with the *MFα1* prepro-α-factor signal peptide the amount of secreted AfCel5A1 increased 3.5-fold. Impressively, the compounded increase in expression levels of AfCel5A1 activity after coding region optimization and signal peptide replacement was about 60-fold.

4.2.3. EG expression levels

Sufficiency analysis (Lynd et al. 2005) calculated that *T. reesei* would need to produce about 1.5% of its total protein as cellulase in order to sustain an aerobic growth rate of 0.02 h^{-1} on Avicel (Lynd et al. 2005). Assuming the *T. reesei* cellulase system were to be reconstructed where all the non CBH components of the cellulase system were EG proteins, EG would need to make up ~0.3% of total cellular protein (van Zyl et al. 2007). Individually, expression of plasmid borne *AfCel7B1opt*, *GtCel12Aopt*, *StCel5Aopt*, *ApCel5Aopt*, and *preproAfCel5A1opt* by CEN.PK111-61A accounted for about 0.05, 0.07, 0.7, 0.2, and 0.3% of total cell protein, respectively. Thus only two of the five EGs studied herein, *StCel5Aopt* and *preproAfCel5A1opt*, were expressed at the sufficiency levels determined by Zyl and colleagues (van Zyl et al. 2007) for aerobic growth on crystalline cellulose; however, the results presented herein, showed that all five

endoglucanases, event *AfCel7B1opt* that expressed its EG at only 0.05%, were able to support rather robust aerobic growth ($\mu > 0.09$) with CMC-4M as the carbon source (Figure 3.1.5 and section 3.4.2). These results suggest that sufficiency levels of EG production by *S. cerevisiae* may be significantly lower than that calculated for *T. reesei* by Zyl et al. (van Zyl et al. 2007).

4.2.4. Evolutionary clustering of the expressed EGs

As reported previously (Ilmen et al. 2011), the results presented herein show that some EG ORFs are better suited for the yeast expression machinery than others. In an attempt to identify features that correlate with the amount of heterologous EG produced by *S. cerevisiae*, the phylogenetic relationship of the heterologous EGs was determined. The phylogenetic relationship between the conserved domains of the library of EGs studied herein was determined using MEGA7 (Kumar, Stecher, Tamura 2016). As expected the 8 EGs belonging to GH family 7B the 6 EGs belonging to GH family 12A and the 11 EGs belonging to GH family 5A (Figure 4.4) formed independent clusters. With 9 of the 11 GH5 endoglucanases being secreted as active enzymes versus 3 of 8 GH7 enzymes and 2 of 6 GH12 enzymes, it appears that the yeast expression system may be better at producing secreted GH5 enzymes (Figure 4.4). Two of the three active GH7Bs, *AfCel7B1* and *TrCel7B*, were closely associated within the GH7 cluster. The third active GH7B, AnCel7B, was more distant. The two GH12s that were expressed as functional proteins, GtCel12A and TrCel12A, were not close within the GH12 cluster (Figure 4.4). Phylogenetic analysis reveals that EGs belonging to GH family 5A are more compatible with the yeast expression system than EGs belonging to GH family 7B and GH family 12A.

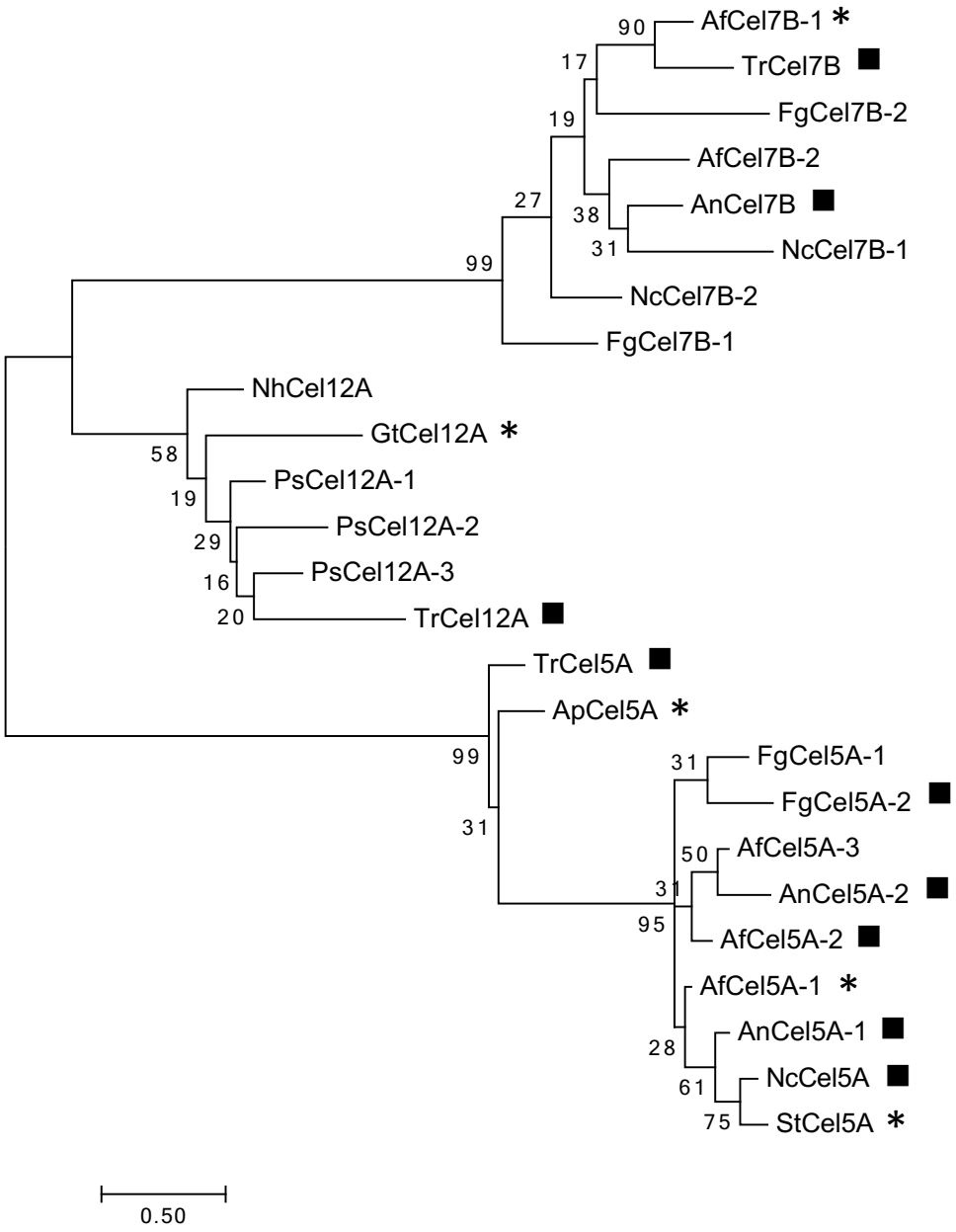


Figure 4.4 – Molecular Phylogenetic analysis. Molecular Phylogenetic analysis by Maximum Likelihood method of the amino acid sequences of the conserved domains of the 25 EGs in this study. The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model(Whelan and Goldman 2001)and 1000 bootstrap replications (Felsenstein 1985). The tree with the highest log likelihood (-3430.7939) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G,

parameter = 3.2977)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 91 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, Stecher, Tamura 2016).

4.3. Co-expression of BGL and EGs

Developing CBH competent *S. cerevisiae* strains that efficiently hydrolyze cellulose requires the co-expression of multiple cellulases within the same strain. Once CEN.PK111-61A-δ-AnBgl1a, a strain capable of efficient growth with cellobiose, was obtained it was used to develop derivative strains that could co-produce both β-glucosidase and endoglucanase at levels that were sufficient to utilize the cellulosic carbon source CMC-4M. Two strategies were used to introduce the above EG genes into CEN.PK111-61A-δ-AnBgl1a; one used episomal 2-micron plasmids, the other used nuclear genome integration.

4.3.1. Plasmid-borne EGs

When the 5 selected EGs were introduced into CEN.PK111-61A p2425 TEF_M, the 2-micron plasmid, *AfCel7B1opt*, *StCel5Aopt*, *preApCel5Aopt*, and, *preproAfCel5A1opt* sustained growth using CMC-4M as the sole carbon source. Only *GtCel12A* was unable to support growth with CMC-4M as the sole carbon source. The inability of *GtCel12A* to support growth with CMC-4M as the carbon source apparently resulted from the very low production of GtCel12A. The very low levels of secreted GtCel12A may have resulted from its inability to compete effectively with AnBgl1 for processing through the secretory pathway.

The amount of AnBgl1 expressed by CEN.PK111-61A-δ-AnBgl1a decreased about four fold when *StCel5Aopt* gene was introduced on the 2-micron plasmid p425TEF_M. Not surprisingly, this resulted in a reduced growth relative to the parent strain CEN.PK111-61A-δ-AnBgl1a when grown with cellobiose as the sole carbon source (Figure 3.10, Table 3.2). It is interesting to note that the growth rate of this strain is very similar on CMC-4M (0.2 h^{-1}) and on cellobiose (0.17 h^{-1}). It therefore appears that StCel5Aopt

might have reduced production of secreted AnBgl1 due to competition for resources such as biosynthetic precursors and translation factors or due to a rate-limiting step in the secretory pathway. It is also interesting that secreted EG production from the p425TEF_M plasmid, when using the CEN.PK111-61A-δ-AnBgl1a strain, resulted in lower amounts of secreted EGs than was obtained when the EGs were expressed using p425TEF_M transformants of strain CEN.PK111-61A. Secreted protein levels of individually expressed *StCel5Aopt*, *preApCel5Aopt*, and *preproAfCel5A1opt* decreased from 5.7, 1.1, and 2 μg/ml to 0.9, 0.39, and 0.3 μg/ml, respectively, after co-expression with *AnBgl1* using the CEN.PK111-61A-δ-AnBgl1a strain. Expression of plasmid borne *AfCel7B1opt* and *GtCel12Aopt* in CEN.PK111-61A produced 0.4 and 0.6 μg/ml of secreted protein, respectively, but no detectable secreted protein when co-expressed from the same plasmids in CEN.PK111-61A-δ-AnBgl1a. Apparently, AnBgl1 competed with the EGs for some cellular resources such as biosynthetic precursors, translation factors or a rate-limiting step in the secretory pathway, and thus limited EG production. This observation is similar to that previously reported for CBH1 and CBH2 competition (Ilmen et al. 2011).

4.3.2. Chromosomal integration of EGs

Unlike the results obtained using CEN.PK111-61A-δ-AnBgl1a with the plasmid-borne EGs, CEN.PK111-61A-δ-AnBgl1a with the 5 EG genes integrated sustained robust growth when using CMC-4M as the sole carbon source. The 5 selected CEN.PK111-61A-δ-AnBgl1a isolates with integrated versions of *AfCel7B1opt*, *StCel5Aopt*, *preApCel5Aopt* and *preproAfCel5A1opt* all produced detectable levels of both AnBgl1 and the 5 individual EGs.

CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt produced the highest levels of secreted EG protein 1.7 μg/ml, corresponding to 0.22% total cell protein while producing AnBgl1 at 0.12% of total cell protein, which was the highest AnBgl1 protein production levels achieved while co-expressing an EG.

In most cases, chromosomal EG integration into the genome of CEN.PK111-61A-δ-AnBgl1a resulted in higher levels of total AnBgl1 and EG protein production. Even though this is the case, the lag times of these strains to reach logarithmic growth on CMC-4M were much higher (24 to 83 hours) than were obtained with the strains having plasmid borne EGs (13 to 47 hours). This could be because the time needed for the secreted EGs to accumulate to levels needed to support logarithmic growth was longer when the EG genes were chromosomally integrated than when they were plasmid-borne EGs because the there were only one or two copies when integrated versus about 40 copies per cell when plasmid borne. The high plasmid copy number of plasmid borne EGs could be causing higher EG expression during the lag phase than is obtained with a much lower copy number for the integrated copies. This higher expression level could enable more EG to be produced during the lag phase thus a shorter lag phase; however, once the cells begin to enter log phase, expression from the strong promoter form the housekeeping *TEF1* gene might be too high when it is plasmid borne and therefore exceed the secretory pathway capacity to process the secreted protein, which in turn would activate the UPR and repress expression or activate EG protein degradation.

Specific activity measurements could explain why the co-expression of the plasmid-borne version or the chromosomally integrated version of *AfCel7B1opt* in the CEN.PK111-61A-δ-AnBgl1a strain sustained growth on CMC-4M, even though levels of secreted AfCel7B1opt protein were too low to be detected in the Coomassie stained SDS-PAGE gels.

4.4. Heterologous Cellobiohydrolase Expression by *S. cerevisiae*

In the *T. reesei* cellulase system, CBHs account for ~70% of the cellulase mass with CBH1 accounting for ~50% of the total cellulase protein (Sandgren et al. 2001). This suggests that CBHs play a very important role in cellulose hydrolysis and that a major portion of further research towards developing a CBP capable *S. cerevisiae* strain will need to focus on enhancing levels of CBH expression. The most highly expressed CBH was *StCel7A*, making up about 0.09% of total cell protein and sufficiency calculations by

Zyl et al. (van Zyl et al. 2007) determined that CBH levels required to support aerobic growth on lignocellulosic feedstocks of *T. reesei* were about 1.2%. Potential strategies for improving CBH production could include; coding region optimization, signal peptide replacement, chromosomal integration and combined mutagenesis and selection. The important finding from the CBH research presented herein is the identification of a CBH that is compatible with the expression machinery of *S. cerevisiae*. As with the findings of Ilmen *et al.* (Ilmen et al. 2011), heterologous CBH expression in *S. cerevisiae* seems to be protein specific and we have yet to identify the features that makes some proteins more compatible with the expression machinery of *S. cerevisiae* than others. It is noteworthy that in this research the best expressed EG and the best expressed CBH originated from *Sporotrichum thermophile*, which is a thermophile. In another study that also looked at the expression levels of fungal cellulases by *S. cerevisiae* (Ilmen et al. 2011) the best-expressed CBH1 was also from a thermophile, *Talaromyces emersonii*. This could indicate that fungal cellulases originating from thermophiles are more compatible for expression in *S. cerevisiae* due to the stability of the protein. Increased thermostability of these proteins could facilitate proper folding and the maintenance of the folded state thus reducing chances of UPR targeted degradation.

4.5. Potential Future Studies

The results presented herein showed that BGL and EGL expression will not be rate-limiting when developing a CBP competent *S. cerevisiae* stain. The efficient hydrolysis of crystalline cellulose requires the cooperative action of β-glucosidase, endoglucanase, and cellobiohydrolase. Having identified several promising EGs and CBHs, the next step would be the CBH expression in BGL and EG expressing strains.

These results indicate that some genes are more compatible with the expression machinery of a host *S. cerevisiae* strain than others. It will be important to identify universal protein features for efficient expression in *S. cerevisiae* by extensive bioinformatics analysis of the best and poorest expressed cellulases.

Optimal ratio of the various classes of cellulase may differ depending on the composition of each substrate. Indeed, the specific requirements of the endoglucanases, cellobiohydrolases and hydrolysis may vary depending on the lignocellulose source and pretreatment method, therefore requiring different cellulase mixtures depending on the specific lignocellulosic substrate and retreatment method. Cellulases with the highest affinity to the expression machinery of *S. cerevisiae* could be selected for expression at different combinations optimized towards various pre-treated lignocellulose biomass.

4.6 Final Conclusion

The results presented herein indicate that some genes are more compatible with the expression machinery of a host *S. cerevisiae* strain than others. Several EGs and CBHs have been identified that are promising candidates for the development of CBP competent *S. cerevisiae*. Co-expression of BGL and EGs sustained the growth of 9 recombinant *S. cerevisiae* strains on YNB media with CMC-4M as the sole carbon source.

References

Environment and Climate Change Canada [Internet]; c2016 [cited 2016 12/22]. Available from: http://www.ec.gc.ca/lcpe-cepa/eng/regulations/detailReg.cfm?intReg=186 .

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3):403-10.

Angov E, Hillier CJ, Kincaid RL, Lyon JA. 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One 3(5):e2189.

Bennetzen JL and Hall BD. 1982. Codon selection in yeast. J Biol Chem 257(6):3026-31.

Bryksin AV and Matsumura I. 2010. Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. BioTechniques 48(6):463-5.

Buchan JR and Stansfield I. 2007. Halting a cellular production line: Responses to ribosomal pausing during translation. Biology of the Cell 99(9):475-87.

Chang MM, Chou TYC, Tsao GT. 1981. Structure, pretreatment and hydrolysis of cellulose. Advances in Biochemical Engineering/Biotechnology 20:15-42.

Corsi AK and Schekman R. 1996. Mechanism of polypeptide translocation into the endoplasmic reticulum. J Biol Chem 271(48):30299-302.

Curran KA, Leavitt JM, Karim AS, Alper HS. 2013. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng 15:55-66.

Davies G and Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure 3(9):853-9.

Del Rio JC, Marques G, Rencoret J, Martinez AT, Gutierrez A. 2007. Occurrence of naturally acetylated lignin units. J Agric Food Chem 55(14):5461-8.

Demain AL and Vaishnav P. 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv 27(3):297-306.

Den Haan R, Rose SH, Lynd LR, van Zyl WH. 2007. Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. Metab Eng 9(1):87-94.

den Haan R, van Rensburg E, Rose SH, Gorgens JF, van Zyl WH. 2015. Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Curr Opin Biotechnol 33:32-8.

Dhawan S and Kaur J. 2007. Microbial mannanases: An overview of production and applications. Crit Rev Biotechnol 27(4):197-216.

Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP. 2002. Cellulose biosynthesis in plants: From genes to rosettes. Plant Cell Physiol 43(12):1407-20.

Doner LW and Irwin PL. 1992. Assay of reducing end-groups in oligosaccharide homologues with 2,2'-bicinchoninate. Anal Biochem 202(1):50-3.

Dong X, Stothard P, Forsythe IJ, Wishart DS. 2004. PlasMapper: A web server for drawing and auto-annotating plasmid maps. Nucleic Acids Res 32:W660-4.

Edgar RC. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics JID - 100965194 .

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39(4):783-91.

Frey-Wyssling A. 1954. The fine structure of cellulose microfibrils. Science 119(3081):80-2.

Fujita Y, Ito J, Ueda M, Fukuda H, Kondo A. 2004. Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. Appl Environ Microbiol 70(2):1207-12.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Walker JM, editor. Humana Press. Inc. 571 p.

Gietz RD, Schiestl RH, Willems AR, Woods RA. 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast 11(4):355-60.

Gírio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Łukasik R. 2010. Hemicelluloses for fuel ethanol: A review. Bioresour Technol 101(13):4775-800.

Gong M, Gong F, Yanofsky C. 2006. Overexpression of tnaC of *Escherichia coli* inhibits growth by depleting tRNA2Pro availability. J Bacteriol 188(5):1892-8.

Gouy M and Gautier C. 1982. Codon usage in bacteria: Correlation with gene expressivity. Nucleic Acids Res 10(22):7055-74.

Goyal G, Tsai S, Madan B, DaSilva NA, Chen W. 2011. Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. Microbial Cell Factories 10(1):89.

Grishutin SG, Gusakov AV, Markov AV, Ustinov BB, Semenova MV, Sinitsyn AP. 2004. Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. Biochimica Et Biophysica Acta (BBA) - General Subjects 1674(3):268-81.

Guo Z, et al. 2011. Development of an industrial ethanol-producing yeast strain for efficient utilization of cellobiose. Enzyme and Microbial Technology 49(1):105-12.

Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M. 2012. Engineering genes for predictable protein expression. Protein Expr Purif 83(1):37-46.

Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166(4):557-80.

Hasunuma T, Ishii J, Kondo A. 2015. Rational design and evolutional fine tuning of *Saccharomyces cerevisiae* for biomass breakdown. Curr Opin Chem Biol 29:1-9.

Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. 2013. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513-22.

He M, Wu D, Wu J, Chen J. 2014. Enhanced expression of endoinulinase from *Aspergillus niger* by codon optimization in *Pichia pastoris* and its application in inulooligosaccharide production. J Ind Microbiol Biotechnol 41(1):105-14.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280 (Pt 2)(Pt 2):309-16.

Henrissat B and Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol 7(5):637-44.

Henrissat B and Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 293 (Pt 3)(Pt 3):781-8.

Henrissat B, Driguez H, Viet C, Schulein M. 1985. Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. Nat Biotech 3(8):722-6.

Hillier CJ, Ware LA, Barbosa A, Angov E, Lyon JA, Heppner DG, Lanar DE. 2005. Process development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for plasmodium falciparum. Infect Immun 73(4):2109-15.

Hoffman CS. 2001. Preparation of yeast DNA. Curr Protoc Mol Biol Chapter 13:Unit13.11.

Ilmen M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-Aho M, la Grange DC, et al. 2011. High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. Biotechnol Biofuels 4:30.

Jeoh T, Michener W, Himmel ME, Decker SR, Adney WS. 2008. Implications of cellobiohydrolase glycosylation for use in biomass conversion. Biotechnol Biofuels 1(1):10,6834-1-10.

Jia H, et al. 2012. High-level expression of a hyperthermostable *Thermotoga maritima* xylanase in *Pichia pastoris* by codon optimization. Journal of Molecular Catalysis B: Enzymatic 78:72-7.

Jin M, Balan V, Gunawan C, Dale BE. 2011. Consolidated bioprocessing (CBP) performance of *Clostridium phytofermentans* on AFEX-treated corn stover for ethanol production. Biotechnol Bioeng 108(6):1290-7.

Jørgensen H, Kristensen JB, Felby C. 2007. Enzymatic conversion of lignocellulose into fermentable sugars: Challenges and opportunities. Biofuels, Bioproducts and Biorefining 1(2):119-34.

Kafer E. 1977. Meiotic and mitotic recombination in *Aspergillus* and its chromosomal aberrations. Adv Genet 19:33-131.

Kerrigan JJ, McNulty DE, Burns M, Allen KE, Tang X, Lu Q, Trulli JM, Johanson KO, Kane JF. 2008. Frameshift events associated with the lysyl-tRNA and the rare arginine codon, AGA, in *Escherichia coli*: A case study involving the human relaxin 2 protein. Protein Expr Purif 60(2):110-6.

Kochetov AV, Sarai A, Vorob'ev DG, Kolchanov NA. 2002. The context organization of functional regions in yeast genes with high-level expression]. Mol Biol (Mosk) 36(6):1026-34.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870-4.

Kurjan J and Herskowitz I. 1982. Structure of a yeast pheromone gene (MF alpha): A putative alpha-factor precursor contains four tandem copies of mature alpha-factor. Cell 30(3):933-43.

Kurland C and Gallant J. 1996. Errors of heterologous protein expression. Curr Opin Biotechnol 7(5):489-93.

Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U. 2014. Challenges and advances in the heterologous expression of cellulolytic enzymes: A review. Biotechnol Biofuels 7(1):135,014-0135-5. eCollection 2014.

Lange HC and Heijnen JJ. 2001. Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. Biotechnol Bioeng 75(3):334-44.

Larue K, Melgar M, Martin VJJ. 2016. Directed evolution of a fungal β-glucosidase in *Saccharomyces cerevisiae*. Biotechnology for Biofuels 9:1-15.

Limayem A and Ricke SC. 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Progress in Energy and Combustion Science 38(4):449-67.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42(Database issue):D490-5.

Lynd LR, Wyman CE, Gerngross TU. 1999. Biocommodity engineering. Biotechnol Prog 15(5):777-93.

Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: An update. Curr Opin Biotechnol 16(5):577-83.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66(3):506,77, table of contents.

Lynd LR. 1996. OVERVIEW AND EVALUATION OF FUEL ETHANOL FROM CELLULOSIC BIOMASS: Technology, economics, the environment, and policy. Annu Rev Energy Environ 21(1):403-65.

Mao F, Leung WY, Xin X. 2007. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. BMC Biotechnol 7:76,6750-7-76.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43(Database issue):D222-6.

McBride J, Zietsman J, Van Zyl W, Lynd L. 2005. Utilization of cellobiose by recombinant β-glucosidase-expressing strains of *Saccharomyces cerevisiae*: Characterization and evaluation of the sufficiency of expression. Enzyme Microb Technol 37(1):93-101.

McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, Moreau RA, Nunez A, Butrym ED, Richards MP, Wang CS, et al. 2004. Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A 101(51):17593-8.

Meier H. 1962. Chemical and morphological aspects of the fine structure of wood. Pure and Applied Chemistry 5(1-2):37-52.

Miller PM. 1955. V-8 juice agar as a general-purpose medium for fungi and bacteria. Phytopathology 45:461-2.

Mumberg D, Muller R, Funk M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156(1):119-22.

Nakatani Y, Yamada R, Ogino C, Kondo A. 2013. Synergetic effect of yeast cell-surface expression of cellulase and expansin-like protein on direct ethanol production from cellulose. Microbial Cell Factories 12(1):66.

Oliveira C, Teixeira JA, Lima N, Da Silva NA, Domingues L. 2007. Development of stable flocculent *Saccharomyces cerevisiae* strain for continuous *Aspergillus niger* beta-galactosidase production. J Biosci Bioeng 103(4):318-24.

Olson DG, McBride JE, Shaw AJ, Lynd LR. 2012. Recent progress in consolidated bioprocessing. Curr Opin Biotechnol 23(3):396-405.

Orr-Weaver TL and Szostak JW. 1983. Yeast recombination: The association between double-strand gap repair and crossing-over. Proc Natl Acad Sci U S A 80(14):4417-21.

Osborne AR, Rapoport TA, van den Berg B. 2005. Protein translocation by the Sec61/SecY channel. Annu Rev Cell Dev Biol 21:529-50.

O'Sullivan AC. 1997. Cellulose: The structure slowly unravels. Cellulose 4(3):173-207.

Pauly M and Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. The Plant Journal 54(4):559-68.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Meth 8(10):785-6.

Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D. 2014. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol Syst Biol 10:770.

Qian W, Yang JR, Pearson NM, Maclean C, Zhang J. 2012. Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet 8(3):e1002603.

Rasmussen JR. 1992. Effect of glycosylation on protein function. Current Opinion in Structural Biology 2(5):682-6.

Rubin EM. 2008. Genomics of cellulosic biofuels. Nature 454(7206):841-5.

Sandgren M, Shaw A, Ropp TH, Wu S, Bott R, Cameron AD, Stahlberg J, Mitchinson C, Jones TA. 2001. The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 A resolution. J Mol Biol 308(2):295-310.

Schadel C, Blochl A, Richter A, Hoch G. 2009. Short-term dynamics of nonstructural carbohydrates and hemicelluloses in young branches of temperate forest trees during bud break. Tree Physiol 29(7):901-11.

Scheller HV and Ulvskov P. 2010. Hemicelluloses. Annu Rev Plant Biol 61:263-89.

Semova N, Storms R, John T, Gaudet P, Ulycznyj P, Min XJ, Sun J, Butler G, Tsang A. 2006. Generation, annotation, and analysis of an extensive *Aspergillus niger* EST collection. BMC Microbiol 6:7.

Sharp PM and Li WH. 1987. The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3):1281-95.

Sharp PM and Li WH. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24(1-2):28-38.

Shirkavand E, Baroutian S, Gapes DJ, Young BR. 2016. Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment – A review. Renewable and Sustainable Energy Reviews 54:217-34.

Sims RE, Mabee W, Saddler JN, Taylor M. 2010. An overview of second generation biofuel technologies. Bioresour Technol 101(6):1570-80.

Skory CD, Freer SN, Bothast RJ. 1996. Expression and secretion of the *Candida wickerhamii* extracellular beta-glucosidase gene, bglB, in *Saccharomyces cerevisiae*. Curr Genet 30(5):417-22.

Somerville C. 2006. Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol 22:53-78.

Sun Y and Cheng J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. Bioresour Technol 83(1):1-11.

Tang H, Hou J, Shen Y, Xu L, Yang H, Fang X, Bao X. 2013. High β-glucosidase secretion in *Saccharomyces cerevisiae* improves the efficiency of cellulase hydrolysis and ethanol production in simultaneous saccharification and fermentation. J Microbiol Biotechnol 23(11):1577-85.

Thanaraj TA and Argos P. 1996. Ribosome-mediated translational pause and protein domain organization. Protein Sci 5(8):1594-612.

Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R. 2008. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J Mol Biol 383(2):281-91.

Tsai S, Goyal G, Chen W. 2010. Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. Appl Environ Microbiol 76(22):7514-20.

Van Rooyen R, Hahn-Hägerdal B, La Grange DC, Van Zyl WH. 2005. Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. J Biotechnol 120(3):284-95.

van Zyl WH, Lynd LR, den Haan R, McBride JE. 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Adv Biochem Eng Biotechnol 108:205-35.

Villalobos A, Ness JE, Gustafsson C, Minshull J, Govindarajan S. 2006. Gene designer: A synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics 7:285.

Wang X, Li X, Zhang Z, Shen X, Zhong F. 2010. Codon optimization enhances secretory expression of *Pseudomonas aeruginosa* exotoxin A in *E. coli*. Protein Expr Purif 72(1):101-6.

Welch M, Villalobos A, Gustafsson C, Minshull J. 2009. You're one in a googol: Optimizing genes for protein expression. J R Soc Interface 6 Suppl 4:S467-76.

Wen F, Sun J, Zhao H. 2010. Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. Appl Environ Microbiol 76(4):1251-60.

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, et al. 2012. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109(3):E111-8.

Whelan JA, Russell NB, Whelan MA. 2003. A method for the absolute quantification of cDNA using real-time PCR. Journal of Immunological Methods 278(1):261-9.

Whelan S and Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691-9.

Wilde C, Gold ND, Bawa N, Tambor JH, Mougharbel L, Storms R, Martin VJ. 2012. Expression of a library of fungal beta-glucosidases in *Saccharomyces cerevisiae* for the development of a biomass fermenting strain. Appl Microbiol Biotechnol 95(3):647-59.

Wood TM. 1988. Preparation of crystalline, amorphous, and dyed cellulase substrates. Meth Enzymol 160:19-25.

Yamada R, Hasunuma T, Kondo A. 2013. Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. Biotechnol Adv 31(6):754-63.

Yamada R, Taniguchi N, Tanaka T, Ogino C, Fukuda H, Kondo A. 2010. Cocktail delta-integration: A novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. Microb Cell Fact 9:32.

Yanase S, Yamada R, Kaneko S, Noda H, Hasunuma T, Tanaka T, Ogino C, Fukuda H, Kondo A. 2010. Ethanol production from cellulosic materials using cellulase-expressing yeast. Biotechnology Journal 5(5):449-55.

Zhang B, Rong C, Chen H, Song Y, Zhang H, Chen W. 2012. *De novo* synthesis of trans-10, cis-12 conjugated linoleic acid in oleaginous yeast *Yarrowia lipolytica*. Microb Cell Fact 11:51,2859-11-51.

Zhang YH and Lynd LR. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol Bioeng 88(7):797-824.

Zheng Y, Pan Z, Zhang R. 2009. Overview of biomass pretreatment for cellulosic ethanol production . International Journal of Agricultural & Biological Engineering 2(3):51-68.

Zhu H, Yao S, Wang S. 2010. MFalpha signal peptide enhances the expression of cellulase *eg1* gene in yeast. Appl Biochem Biotechnol 162(3):617-24.

Zorov NI, et al. 1997. Application of the bicinchoninic method of assay for the reducing sugars to determine carboxymethylcellulase activity of cellulases using a microplate reader. 62(7):704; 6-709.

Appendix

Table of Contents

ORF Nucleotide Sequences

Endoglucanase nucleotide sequences

>AfCel7B1

ATGGACTCCAAAAGAGGCGTCGTGGCCGCGGTGCTGGCCTTGCTTCCTCTCGTTTCTGC
GCAACAACCCGCCGCGAGTTCTGCTGGTAACCCCAAGCTGACGACATACAAATGTACCA
CTGCTGGCGGCTGTGTTGCGCAGGATACATCTGTAGTTCTAGATTGGGGCTACCACTGG
ATCCACACGGTCGATGGGTATACATCATGCACCACATCGTCCGGAGTCGACAGCACCCT
GTGTCCTGATGCGGCCACCTGCGCGAAGAACTGTGTGATCGAGCCGGCCAACTACACCA
GCGCCGGTGTGACCACCTCGGGTGACTCTCTGACGATGTACCAATATGTTCAGAGCAAC
GGCGTCTATACCAACGCCTCGCCTCGCCTCTACCTCCTCGGCCCCGACAAGAACTATGT
CATGCTGAAGCTGCTAGGCCAGGAGCTCACCTTCGACGTGGACCTGTCCACACTCCCCT
GCGGCGAAAACGGCGCCCTGTATCTCTCCGAAATGAGCGCCACCGGTGGTCGCAACGAA
TACAACACCGGCGGCGCCGAGTACGGCTCTGGCTACTGTGACGCCCAATGCCCAGTGAT
CGCCTGGAAGAACGGCACCCTCAACACGAGCGGCGCAAGCTACTGCTGCAACGAGATGG
ACATCCTCGAGGCCAACTCCCGCGCCAACTCGTACACCCCGCACCCCTGCAGTGCAACG
GACTGTGACAAGGGCGGATGCGGCTTCAACCCATACGCTCTCGGCCAAAAGAGCTACTG
GGGGCCCGGCGGCACCGTCGACACGTCTAAGCCCTTCACCATCACCACGCAGTTCATCA
CCAACGACGGCACCACCACCGGCACCCTGTCCGAAATCCGCCGACAGTACATGCAGAAC
GGCAAGGTGATCGCCAATGCCGTTTCCTCCACTGGTGTCAACTCCATTACCGAGGACTG
GTGCACGTCCGTCGACGGCTCGGCCGCCACCTTTGGCGGACTGACCACCATGGGCAAGG
CGCTGGGCCGCGGGATGGTCCTCATCTTCAGCATCTGGAACGATGCCAGCGGCTTTATG
AACTGGCTCGACAGCGGCAACGCAGGCCCTTGCAGCAGCACCGAGGGAAACCCGGACTT
GATCAAGGCGCAGAATCCCACGACCCACGTGGTCTTCTCGAATATCCGCTGGGGAGATA
TCGGATCGACTTTCAAGGGTTCTGATGGCTCGGTGACGACGACTACGTCCACCACGTCG
ACCAAGACCACCACTAGCACCGCGCCTGGGCCAACGCAGACTCACTATGGACAGTGCGG
TGGACAAGGCTGGACTGGACCCACGGCTTGTGCATCGCCCTACACCTGCCAGGTTCTGA
ATCCGTGGTACTCTCAGTGTCTGTAG

>AfCel7B2

ATGGCTCAAACACTGGCAGCCGCCTCGCTGGTTCTCGTCCCCCTTGTGACTGCGCAGCA
GATCGGGTCCATCGCTGAGAACCACCCGGAGCTCAAAACATACAGGTGCGGTTCTCAGG
CTGGCTGCGTTGCACAAAGCACCTCAGTGGTTCTTGACATCAACGCACACTGGATACAC
CAAATGGGAGCCCAAACGTCATGCACCACGAGTAGTGGCCTCGACCCGTCTCTCTGCCC
CGACAAGGTGACCTGCAGCCAGAACTGCGTAGTGGAAGGCATCACCGACTACAGTAGCT
TCGGCGTGCAGAACTCAGGCGACGCCATAACCCTCCGCCAATACCAAGTACAAAACGGC
CAGATCAAGACACTGAGGCCGCGCGTGTACCTCCTCGCCGAGGATGGCATCAACTACAG
CAAGCTGCAGCTCCTCAACCAAGAGTTCACCTTCGACGTTGACGCCTCCAAGCTCCCCT
GCGGCATGAACGGCGCCCTCTACCTCTCCGAAATGGACGCCTCAGGCGGCCGCAGCGCC
CTCAACCCCGCTGGTGCAACCTACGGCACAGGCTACTGCGACGCGCAGTGCTTCAACCC
CGGTCCCTGGATCAACGGCGAGGCCAACACCCTCGGCGCAGGTGCCTGCTGCCAGGAGA
TGGACCTCTGGGAGGCCAACTCGCGCTCCACAATTTTCTCCCCGCACCCGTGCACCACA
GCCGGCCTGTACGCGTGCACCGGCGCCGAATGCTACTCCATCTGCGACGGGTATGGCTG
CACCTACAACCCCTACGAGCTCGGCGCAAAGGATTACTATGGCTACGGTCTCACCGTTG
ACACCGCCAAGCCGATAACCGTCGTCACGCAGTTCGTGACGGCCGATAACACCGCGACA
GGAACTCTGGCCGAGATCCGCAGGCTGTACGTGCAGGAGGGTATGGTGATCGGCAATTC
GGCCGTCGCCATGACGGAGGCTTTCTGCTCGTCGTCGAGGACGTTCGAGGCGCTGGGTG
GGTTGCAGCGTATGGGCGAGGCTCTGGGGAGGGGCATGGTGCCTGTGTTCAGTATCTGG
GATGATCCGAGCCTGTGGATGCATTGGCTTGATAGTGACGGTGCCGGGCCGTGCGGTAG
CACTGAGGGAGATCCTGCTTTCATCCAGGCTAACTATCCCAATACGGCGGTGACTTTTT
CGAAGGTCAGGTGGGGGGATATTGACAGTACCTACTCTGTTTAG

>AfCel5A1

ATGAAGGCTTCCACTATTATCTGCGCACTTCTCCCCCTTGCTGTGGCGGCGCCGAATGC
GAAGCGGGCTTCTGGGTTTGTCTGGTTCGGAAGTAACGAGTCTGGCGCTGAGTTCGGAG
AGACCAAGCTGCCTGGCGTGCTGGGGACGGATTATATCTGGCCCGATGCGTCGACGATC
AAGACTCTGCATGATGCGGGGATGAACATCTTCCGGGTTGCTTTCCGCATGGAGAGGCT
GATCCCGAATCAGATGACGGGGACTCCGGATGCGACGTACCTGAATGATCTCAAGGCGA
CTGTCAATGCGATTACGAGTCTGGGGGCGTACGCGGTTATTGATCCCCATAACTATGGA
AGATACTACGGGAACATCATCTCGTCGACTGACGACTTTGCCGCGTTCTGGAAGACTCT
GGCTGCCCAGTTTGCGTCCAATGACCATGTCATTTTTGACACCAACAACGAGTACCATG

ACATGGACCAGACGCTCGTGCTCAACCTCAACCAGGCTGCCATCAACGCCATCCGTGCT
GCAGGCGCCACGTCGCAGTACATTTTCGTCGAGGGCAACTCGTGGTCCGGCGCGTGGAC
CTGGACCACCGTCAACGACAACCTCAAGGCCCTCACCGACCCGCAGGATAAGATCGTCT
ACGAGATGCACCAGTATCTCGACTCCGACGGGTCCGGCACGTCGGCCACCTGCGTGAGC
TCCACCATCGGCCAGGAGCGCGTGCAGGCCGCCACACAGTGGTTGAAGACCAATGGTAA
GAAAGGTATCATTGGCGAGTTCGCTGGAGGCGCCAACAGCGTGTGCCAGTCCGCTGTCA
CAGGCATGCTTGACTACCTGTCTGCCAACTCGGATGTGTGGATGGGCGCGGCATGGTGG
GCCGCTGGTCCCTGGTGGGCAGATTACATCTTCAACATGGAGCCGCCGTCGGGTACTGC
CTATCAGAACTATCTCTCGTTGTTGAAGCCGTATTTCGTCGGTGGTTCGGGTGGTAACC
CCCCAACGACCACCACAACTACCACCCAGCCTACTACCACCACTACCACGACCACGGCC
GGGAACCCTGGTGGCACCGGACTCGCACAGCACTGGGGCCAATGCGGTGGAATCGGATG
GACCGGGCCAACAGCCTGCGCGAGCCCCTATACCTGCCAGAAGCTGAACGACTACTACT
CTCAATGCCTGTAG

>AfCel5A2

ATGAGAATCAGCAGCTTGATCATGGCCGCCAGCGCTGCTGGTCTTGTCAGCGCCCTGCC
CGTACGCGAGATGACCAAAAAACGCTCTTCTGGATTTACCTGGGTCGGTATCAACGAAT
CTGGTGCGGAATTTGGTAGCAACATCCCCGGAAAGCTGGGTACCGACTACACATGGCCC
GATACCTCCAAGATTCAGACTCTGCGGGATGCGGGTATGAACATCTTCCGGGTCCCATT
CTTGATGGAGCGTCTGGTGCCCAGCTCCATTACCGGTAATCTGGATGCCACTTACCTAA
GTGACCTGAAGAAGACCGTCGAGTTCATTACCGGCAGTGGTGCTTATGCCGTTCTTGAT
CCTCATAACTATGGAAGATACTCTGGAAGTATCATCTCCTCTACTTCTGATTTCCAGGA
ATTCTGGAAGACCGTCGCTAGCGAATTTGCCTCCAATGACAAGGTCATCTTCGATACCA
ACAATGAATACCACGATATGGACCAGACTCTGGTGCTCAATTTGAACCAGGCCGCCATC
AATGGTATCCGTGCTGCCGGTGCCACGACGCAATACATCTTTGTTGAGGGTAACCACTA
CAGTGGTGCCTGGACCTGGGCTGACAACAATGACAACCTTAAGGGTCTGAAGGATCCTG
AGGACAAGATCGTCTTTGAGATGCACCAGTACCTTGATTCGGACGGTTCTGGTACCTCG
GAGTCCTGTGTCAGCTCCACTATTGGCCAGGAGCGTGTGGAGTCTGCCACTCAGTGGCT
GAAGGACAACAGCCTCAAAGGATTCCTCGGAGAGTTCGCCGGCGGTGTCAACTCCCAGT
GCGAGACGGCCGTTGAGGGTCTTCTCTCTTACATGTCCGAGAACAGCGACGTCTGGCTT
GGTGCCGAGTGGTGGTCTGCTGGTCCCATGTGGGGAAATTATATGTATAGTCTGGAGCC
TAGTACTGGCCCTGCCTACTCAAGCTACCTTCCCATTCTGAAGAATTACTTTGTCAGCG

GCACCTCCTCTTCGTCCTCGTCGTCTTCCTCCATCACCAGCTCTACTACTTCTGTGAAC
AACCCCACCACTTCTGCCACTACTCAGGCTCAGGTTGTCAACACTTCTTCCTCTTCCCC
TACTCCTTCTCCAACCTCTGTGCCTGCAGCCGAGGCTCCCGCTCCCACGCCTACTGAGG
CTGCTAAGACTTCCAGTGCTGCTTCGACCACCCCGACGACCGCCGCTGCTTCGACTACA
GCTTCTGCTTGCCGTGCCCCCGAGCCTGCAGACGAAACCACTCTGCCCAGTGCTACTCC
AACTACTTCTGCGTCTGCCTCTGGCATCGCCAAGCACTGGTACCAGTGTGGTGGAATCA
ACTGGACCGGCCCCACAACCTGCGAGAGCGGCTACACCTGTGTTGAGCAGAACCCTTAC
TACCACCAGTGTGTTTAA

>AfCel5A3

ATGAAATTCGGTAGCATTGTGCTCATTGCTGCTGCGGCAGGCTTCGCGGTGGCTGCTCC
TGCAAAGAGAGCTTCGGGGAAAGAATTCATCTTCCCGGACCCTTCTACAATCAGCACAT
TGATCGGGAAGGGCATGAACATCTTCCGGATTCAATTCCTCATGGAGAGACTGGTGCCA
AGCTCTATGACAGGCTCCTATAATGAGGAGTACCTTGCCAATCTGACATCGGTTGTGGA
CGCTGTCACCAAGGCAGGATCTTATGCTATTTTGGACCCACACAACTTTGGCAGATACA
ATGGTCAGATTATCTCCAGCACCGACGACTTCAAGACCTTCTGGCAGAATCTGGCTGGA
AAGTTCAAGTCCAACAATCTCGTCATCTTTGATACTAACAATGAGTATCACGACATGGA
CCAGACACTGGTACTGAACCTCAACCAGGCCGCTATCAACGGTATCCGCGCTGCAGGAG
CCACCTCGCAATACATCTTTGTGGAGGGCAACTCCTGGACCGGCGCCTGGACCTGGGCC
GACGTCAATGACAACCTGAAGGCTCTGACCGACCCCCATGATAAGATCGTCTACGAGAT
GCACCAGTATCTCGACTCGGATGGATCCGGCACCGCGGAGAGCTGCGTGTCTACCACGA
TTGGTAAGGAGCGGGTTTCTGCCGCAACAAAGTGGCTCAAGGATAACGGCAAGGTTGGC
ATCATTGGTGAGTTCGCTGGTGGCGTCAATGATCAGTGCCGGACCGCTATTTCAGGAAT
GCTGGAGTACTTGGCTCAGAACACAGACGTGTGGAAGGGAGCTCTCTGGTGGGCGGCTG
GCCCCTGGTGGGGAAACTATATGTTCAACATGGAGCCTCCGAGCGGTGCAGCTTATGTG
GGCATGTTGGACATCTTGGAGCCCTACCTGGGTTGA

>AnCel7B

ATGGCTCTGTTACTATCTCTCAGCCTTCTTGCCACAACAATCTCAGCCCAACAGATTGG
GACGCCAGAAATCCGGCCACGTCTTACTACCTACCACTGTACTTCCGCCAACGGCTGTA
CAGAGCAGAATACCTCTGTGGTGCTCGATGCCGCCACGCACCCCATCCACGATGCATCC

AACCCCAGTGTCTCCTGCACCACCTCAAATGGGCTAAACCCTGCTCTATGCCCAGACAA
GCAGACTTGCGCAGACAACTGCGTCATCGACGGCATAACTGACTATGCTGCGCACGGAG
TCGAAACCCATGGGTCGCGGTTGACACTCACTCAATACCGAAACGTGAACGGGGCGCTC
TCCTCTGTTTCACCGAGGGTCTATCTCGTTGATGAGTCCGACCCTGATGAGCAGGAGTA
TCGAGCCTTGTCCCTGCTCGCCCAAGAATTTACCTTCACTGTCAACGTCTCCGCGCTCC
CATGCGGGATGAACGGCGCGCTATATCTCTCCGAAATGTCTCCCTCCGGCGGGCGCAGC
GCGCTCAACCCCGCCGGAGCCTCCTATGGCACAGGCTACTGCGATGCCCAATGTTATGT
GAATCCCTGGATCAACGGCGAGGGAAACATCAACGGCTACGGAGCCTGCTGCAACGAGA
TGGACATCTGGGAGGCTAATTCGCGGAGTACGGGGTTCACGCCTCATGCTTGTTTATAT
GAGCCGGAGGAAACTGAGGGAAGAGGGGTATACGAATGCGCCTCCGAAGATGAGTGCGA
TAGCGCGGGCGAGAATGACGGCATCTGCGATAAGTGGGGATGCGGCTTTAACCCGTATG
CTCTGGGAAATACAGAGTACTACGGCCGTGGCCAAGGGTTTGAAGTCGACACTAAAGAG
CCCTTCACGGTCGTGACACAGTTCCTGACGGATGACGGAACAAGTACAGGTGCTTTAAC
CGAGATTAGACGGCTATATATCCAAAACGGGCAGGTCATCGAGAACGCAGTTGTCTCGT
CTGGTGCAGACTCGCTGACCGATTCCCTCTGCGCCTCTACCGCGTCATGGTTCGACTCA
TACGGAGGAATGGAAGGGATGGGAAGGGCGCTTGGCCGTGGGATGGTCCTCGCCATGAG
TATCTGGAATGATGCGGGTGGCTACATGCAGTGGCTCGACGGTGGGGATGCAGGACCCT
GTAATGCCACCGAGGGCGCACCGGAATTTATTGAGGAGCATACGCCGTGGACAAGGGTT
GTCTTTGAAGATTTGAAGTGGGGTGATATTGGCAGTACTTTCCAGGCGTCTTAG

>AnCel5A1

ATGAGGTCTCTCGTCCTTCTGTCGTCCGTCCTGGCCCTGGTGGCACCCTCCAAAGGCGC
CTTCACATGGCTTGGTACCAACGAAGCCGGTGCCGAATTCGGCGAGGGCTCCTACCCCG
GCGAACTGGGCACGGAATACATTTGGCCTGACCTGGGTACGATTGGTACGCTGAGGAAC
GAGGGGATGAACATATTCCGCGTTGCGTTCTCCATGGAGCGCCTGGTGCCTGATTCGTT
GGCTGGACCGGTAGCGGACGAGTATTTTCAGGACTTGGTCGAGACGGTCAACGGCATTA
CGGCCCTGGGTGCGTATGCAGTCCTTGACCCGCATAATTATGGACGGTACTACGGCAAC
ATTATCACTTCAACCGATGACTTTGCAGCCTTCTGGACCATCCTTGCCACCGAGTTCGC
TTCTAATGAGCTTGTCATCTTCGATACCAATAATGAATACCACACCATGGACCAGTCCC
TTGTCCTGAACCTCAATCAGGCGGCCATAGACGCTATTCGCGCCTCCGGTGCCACCAGC
CAATATATCTTCGCGGAAGGAAACTCTTGGACTGGAGCCTGGACCTGGGTGGATGTCAA
CGACAACATGAAAGCGCTCACCGATCCCCAAGATAAGCTCATCTACGAGATGCACCAGT

ATCTTGACTCCGACGGCTCTGGGACGAACACGGCCTGCGTCAGTTCTACTATCGGCAGC
GAACGCGTGACAGCCGCTACAAACTGGCTGCGAGAGAACGGGAAACTGGGGGTGCTCGG
TGAGTTTGCAGGCGCAAACAACCAGGTCTGCAAGGACGCAGTCGCAGATCTGCTGGAGT
ATCTTGAGGAAAACAGTGATGTATGGCTTGGGGCCCTTTGGTGGGCTGCTGGGCCTTGG
TGGGGTGATTATATGTTCAACATGGAGCCTACTTCGGGTATTGCATACCAGGAGTACTC
GGAAATCTTGCAGCCGTATTTCGTGGGAAGTCAATAA

>AnCel5A2

ATGAAGGTCAACACTCTTTTGGTTGCCGTAGCGGCTGGTACCGCGATGGCTGCACCCCA
GCTCAAGAAGCGCGCCGGTTTTACGTTTTTTGGAGTCACCGAGGCTGGCGCTGAGTTCG
GCGAGAAAAGTATTCCTGGTGTTTGGGGCACTGACTACACCTTCCCCGATACTGAATCC
ATCTTGACCCTCATCTCGAAGGGTTTCAACACTTTCCGCATTCCCTTCCTTATGGAGCG
TCTGACGCCTGAGATGACGGGCTCTTTCGACGAAGGATACCTGAAAAATCTGACTTCGG
TCGTCAATGCAGTCACCGATGCAGGAGCATGGGCCATCGTTGACGCCCAAAACTTTGGC
CGATTCAACGGCGAGATCATAAGCAGCGCCTCCGACTTCCAGACTTGGTGGAAGAACGT
AGCGGCCGAGTTCGCGGACAATAAGAACGTCATTTTCGATACCACTAACACGTCTGGGG
CAGACAACGAATTCCACGACATGGACCAGACCCTCGTGCTGGATCTCAACCAAGCCGCC
ATCAACGGCATCCGCGCAGCCGGCGCCACCTCGCAATACATCTTCGTCGAGGGTAACTC
GTACACCGGCGCCTGGACATGGACGGACAACAATGACAACCTGAAATCCCTCACCGACC
CGCAGGACAAGATCGTCTACGAAATGCACCAGTACCTCGATACTGACGGCTCCGGAACA
CACGAGACCTGTGTTTCGGAGACAATCGGCGCCGAGAGGGTTGAATCCGCGACGCAGTG
GCTGAAAGACAACGGCAAACTCGGTGTCATTGGTGAATTTGCGGGCGGCAATAACGAGA
TATGCCGCGCAGCGGTCAAGAGTCTACTGGATGCGCTGAAGGAGAACGATGATGTTTGG
CTCGGTGCGCTTTGGTGGGCTGCTGGGCCTTGGTGGGAAGACTATATGTTCAGCATGGA
GCCGACGGATGGCATTGCGTATACGGGGATGCTCAGTACGTTGGAGGCGTACATGAATT
AA

>ApCel5A

ATGAAGTACTCAACTTTCGTCGTCGCCGCGGCAGCTGGTTCTGCCGCAGCTTCTTCCTC
TGCATACGGCCAATGTGGTGGTAACGGATGGACTGGTGCCACTGACTGTGTCTCCGGCT
ATCACTGTGCTTACCAGAACGACTGGTACTCCCAGTGTGTTCCTGGAGCTGGTTCCGGT

TCCACCTCTGCTGCTGCTGCTGCCACCACCAAGGTTGTGACCTCGGCCAAGGCCAG
CACTACTCAGGCTCCTTCCAAGGCTCCTGTCTCTACTAGCGCCGCCTCTACTAGCAAGG
CTGTTGCCGCTACCTCCGGCAAGGTCAAGTACGCTGGTGTCAACATTGCTGGCTTCGAC
TTCGGTATGGACACCAACGGCGCTTCCTCTGGCTCGTACGTCGACCCTGGAACCACTGG
CCAGAACCAGATGAACCACTTCGTCAAGGACGACAAGCTCAATGCCTTCCGTCTTCCCG
TTGGTTGGCAATACCTCGTCAACAGCCAGCTCGGTGGTACCCTTGACTCTACCTTCTTC
GCCAAGTACGACCAGCAGATGACCTACTGCTTGAACTCCGGCGCCGCCTTGTGTATCCT
TGACCTTCACAACTATGCCCGCTGGAACGGCCAGATTGTTGGCACCAGCGGCGGTCCCA
CCAACGCCCAGTTCGCCAGTGTTTGGAGCCAATTGGCCAAGAAGTACTCGTCTAAGCCC
AAGGTCGCCTTCGCCATCATGAACGAGCCTCATGATCTTCAGGATATCAACGCCTGGGC
CACTACTGTTCAGGCCGCCGTCACCGCCATTCGTCAGGCAGGTGCCACCCAGAACATGA
TTCTGCTCCCTGGTGACGACTGGACACATGCAGCCAATTTTGTGGACAACGGCTCCGCT
GCTGCCCTGAACAAGATCACCAACCTTGACGGCAGCAAGACCAACCTCGTCTTCGACGT
CCACCAATACTCCGACTCTGACGGCAGCGGTACCTCGAGCACCTGCGTTTCCTCTGCCT
CCAACATTGCCGGTTTCACCAAGCTCGCTACTTGGCTCCGCAGCAACAACCGCCAGGCC
ATGCTCACCGAGGCCGGTGGTTCCAACGACCAGTCTTGCCTTACTGCCGTCTGTGACGT
CCTCAACTACCTCATGACCAACTCTGATGTCTACCTTGGCTGGACTGGCTGGTCGGCTG
GTATGTTCGCTACCGACTATGTTCTTTCTGAGGTTCCTACTGGCAGCGCTGGCTCCTAC
AAGGATCAGGCTATTGTCACCAAGTGCATTGCTGGTGTCTTCAACTCCAAATAA

>FgCel7B1

ATGTATCGCGCCATCGCCACCGCCTCGGCGCTCATCGCAGCCGTTCGGGCTCAGCAAGT
TTGCTCTCTCACACAGGAGAGCAAGCCCTCCCTGAACTGGTCCAAGTGTACATCCAGCG
GATGCAGCAACGTCAAGGGATCCGTCACCATCGATGCCAACTGGCGATGGACTCACCAA
GTATCCGGATCTACCAACTGCTACACCGGCAACAAGTGGGACACTTCCGTCTGCACCAG
CGGAAAGGTCTGTGCTGAGAAGTGCTGTCTCGACGGCGCCGACTACGCCAGCACCTACG
GTATCACCTCCAGCGGCGACCAGCTCAGCCTCTCGTTCGTCACCAAGGGACCCTACAGC
ACCAACATCGGTAGCCGTACTTACCTCATGGAGGACGAGAACACCTACCAGATGTTCCA
GCTCCTGGGTAACGAGTTCACCTTTGATGTCGATGTTTCAAACATTGGATGTGGTCTGA
ACGGTGCTCTTTACTTCGTCAGCATGGACGCCGATGGTGGCAAGGCCAAGTACCCCGGT
AACAAGGCCGGAGCCAAGTACGGTACTGGTTACTGTGATGCCCAGTGCCCTCGTGACGT
TAAGTTCATCAACGGCCAGGCCAACTCTGATGGCTGGCAGCCCTCCGACAGCGATGTCA

ACGGTGGTATTGGTAACTTGGGCACATGCTGCCCTGAGATGGACATCTGGGAGGCCAAC
TCCATCTCCACTGCCTACACTCCTCACCCTTGCACTAAGCTCACTCAGCACTCTTGCAC
TGGCGACTCTTGCGGTGGAACCTACTCCAACGACCGTTACGGCGGAACTTGCGATGCTG
ACGGTTGCGATTTCAACTCCTACCGCCAGGGCAACAAGACCTTCTACGGTCCTGGATCT
GGCTTCAACGTTGATACCACTAAGAAGGTCACCGTCGTCACTCAGTTCCACAAGGGCAG
CAACGGACGTCTCTCTGAGATCACCCGTCTCTATGTCCAGAATGGCAAGGTCATTGCCA
ACTCTGAGTCCAAGATTGCTGGAGTTCCCGGAAACTCTCTTACCGCTGACTTCTGCACC
AAGCAGAAGAAGGTCTTTAACGACCCCGATGACTTCACGAAGAAGGGTGCTTGGAGCGG
TATGAGCGATGCTCTTGAGGCTCCCATGGTTCTTGTTATGTCTCTCTGGCACGACCACC
ACTCCAACATGCTCTGGCTCGACTCTACCTACCCCACTGACTCTACCAAGCTCGGCTCT
CAGCGCGGATCTTGCTCTACATCCTCTGGCGTGCCTGCCGACCTTGAGAAGAACGTCCC
CAACTCCAAGGTTGCCTTCTCCAACATCAAGTTCGGTCCCATCGGCAGCACCTACAAGA
GCGACGGCACCACTCCCACCAACCCCACCAACCCTTCTGAGCCCAGCAACACTGCCAAC
CCCAACCCCGGCACCGTTGACCAGTGGGGCCAGTGCGGTGGCAGCAACTACAGCGGTCC
TACCGCCTGCAAGTCTGGCTTCACCTGCAAGAAGATCAACGACTTCTACTCCCAGTGCC
AGTAA

>FgCel7B2

ATGAAGTTCTCTCCTCTTCTCCTCTCAACTCTTCTCGCCAACACGGTCGTGGCCCAGAC
TCCTGACAAGACCCAGGAGAAACACCCCAAGATCGAAACATACCGATGCACAAAGGCAA
AGGGCTGCAAGAAGGCGACAAACTACATCGTCGCCGACGCAGAGCTTCACGGCATCAGC
CAGGCCAACGGCCAGAGCTGCGGTAACTGGGGTGAAGCCGCCAACTCCACTGCTTGCCC
CGACGAGGCAACATGTGCCAAGAACTGCAAGCTCTTTGGCATGAACGAGGCTGCGTACA
AGGCCAAGGGTATCAGCACCTCTGGTAACGCTCTCCGTCTGGAGATGCTGCGCAACGGC
CAGTCTGTTTCTCCCCGTGTTTACCTCCTCGAGGAGAACAAGAACAAGTATGAGATGCT
CAAGCTTACCGGTGCTGAGTTCTCTTTCGATGTTGAGACTCAGAAGCTTCCTTGTGGTA
TGAACGGTGCTCTGTATCTTTCTGAGATGCCTGCCGATGGTGGAAAGAGCACGAGCAAG
TACAGCAAGGTCGGTGCTGCTCAGGGTGGCGGATACTGTGATGCTCAGTGCTATGTCAC
ACCTTTCATCAACGGAGTGGGTAACATCAAGGGCAAGGGTGTCTGCTGTAACGAGATGG
ATATCTGGGAGGCCAACTCTCGCGCCACCCACATTGCTCCTCACCCCTGCAGTGTCCCC
GGCCTTTACGGCTGCACAGGTGCCGAATGCCAGAAGGATGGTATCTGTGACAAGGCTGG
CTGCGGCTGGAACCATAACCGCAACGGCGTTCCTGATTTTTACGGACGCGGCAAGAACT

TCAAGGTCGACACGACCCGCAAGTTCACAGTTGTCTCGTCATTCCCCGCCGACAAGAGT
GGCAAGCTCACAGAGATGCACCGCCACTACATCCAGGACGGCAAGGTGATCAAGAGCGC
CGTCGTCACCCTCCCCGGACCCCCCAAGGTTACCGGTAACATCATCACCGACAACTACT
GCAAGGCCTCGCACGCAGACGACTACATCCGTCTCGGCGGCACAGAGGAGATGGGCGAT
GCCATGACCCGTGGTATGGTTCTCGCCATGAGTGTCTGGTGGAGCGAGGGCGACTCCAT
GGAGTGGCTCGACGGACAGGGTGCTGGCGCTGGACCCTGCACCAAGGAGGAGGGCCTCC
CTAAGAACATTGTCAAGGTCGAGCCCAACCCTGAGGTTACGTTTAGCAACATCCGCATT
GGCGAGATCGGGTCGACACACGCGGTGAAGATGCCTCGCGTGTATGGCGCTCACCGCCT
GTAA

>FgCel5A1

ATGCGTTTCACAGATCTTCTTCTCGCCAGCGCCGGCGCTACACTTGCTCTAGCTGCCCC
TTCCACCGAGAAGCGTGCCGCGGGCAAGTTCCTTTTCACCGGTTCTAATGAATCTGGTG
GTGAGTTTGGCGAGACTCAACTCCCTGGAAAGCTCGGCAAGGACTACATCTGGCCTACC
ACCAAGTCCATCGATACTCTTGCCAGTACCGGCATGAACACCTTCCGTGTTGGTTTCCG
TATGGAGCGCATGACCCCTAGTGGCATCACTGGTGCTCTTGACGAGACCTACTTCAAGG
GTCTTGAGAGTGTTGTCAACCACATCACCAGCAAACACAGGGGCAACTTTGCTGTGATT
GACCCTCACAACTATGGCCGCTACAACAACCAAATCATCCAGAGCACCGCCGACTTTGG
CGCCTGGTGGTCAAAGGTTGCCAAGCGCTTCGCCAACAACAAGAACGTCATCTTTGACA
CCAACAACGAGTACCACGACATGGAAAACTCTCTTGTTGCCGGCCTCAATCAGGCCGCC
ATCGACGCCATCCGCAAGGCCGGCGCCACCTCCCAGTACATCTTTGTTGAGGGTAACTC
TTACACCGGTGCCCACAGCTGGGTCTCCAGCGGCAACGGCGAGGCTCTCAAGAACCTCA
AGGATCCCCAGAACAAGATCATCTACCAGATGCACCAGTACCTCGACTCAGACAACTCG
GGTACCCACGCCGACTGTGTCAGCAGCACCATTGGTGTCGAGCGTGTCAAGGAGGCTAC
CAAGTGGCTCAAGGATAACAAGAAGAGAGGTATCATCGGTGAGACTGCCGCTGGACCTA
ACACTCAGTGCATTGAGGCTCTCAAGGGTGAACTCCAGTACTTGCACGACAATTCTGAC
GTCTGGACTGGTTGGTTGTACTGGGCTGCTGGACCTTGGTGGGGTGACTACATGTACAG
CATGGAGCCTGATACCGGTGCTTCTTACGTCAAGGTTCTCCCTGAGATCAAGAAGTTCA
TCGGTGCTTAA

>FgCel5A2

ATGAAGTCCCTCCTCGCCCTCAGTCTCTTCGCAGGTCTCTCCGTCGCCCAAAGCTCAGC
CTGGGCCCAATGCGGTGGTGAAGGCTTCTCTGGTTCCACATCCTGCGTCTCGGGCTACA
AGTGCACCGTTGTCAACCAGTGGTACAGTCAGTGCCAGCCCGGTACTGCTGAGCCTCCC
TCTACTACCCTCAAGACTACCACTGGCGGTGGCTCTACTCCCACTGGAACACCTGGCGA
TGGAAAGTTCCTCTGGGCTGGTGTCAACGAGGCTGGTGGTGAGTTCGGAGAGAAGAACC
TTCCCGGTACTTGGGGCAAGGACTTTATCTTCCCCGACCCTGCTGCTGTCGACACCCTC
ATTTCTCAGGGCTACAACACCTTCCGTGTGCAGCTCAAGATGGAGCGTGCTAACCCCAG
CGGCTTGACTGGCGCCTACGACCAGGCTTACATCAAGAACCTCACCTCCATTGTCAACC
ACATCACTGGCAAGGGAGCCACTGTTCTTCTTGACCCCCACAACTATGGCCGCTTCTTT
GACAAGATCATCACCTCCACCTCTGACTTCCAGACCTGGTGGAAGAACTTTGCTACTCT
GTTCAAGAGCAACAGCCGCATCATGTTCGATACCAACAACGAGTACCACACCATGGACC
AGACCCTTGTTCTTAACCTCAACCAAGCCGCCATCAACGGTATCCGCGCTGCCGGTGCC
ACTCAGTACATCTTTGTCGAAGGCAACCAGTGGTCCGGCGCCTGGTCGTGGCCCGACGT
CAACGACAACATGAAGGCTCTGACTGACCCCGAGAACAAGCTCATCTACGAGATGCACC
AGTACCTCGACTCTGACAGCTCCGGCACCTCACCTGACTGTGTTTCTACCACCATCGGT
GTCGAGCGTCTCCAGGCTGCTACCAAGTGGCTCCGTGCCAACAAGAAGATCGGCATGAT
CGGAGAGTTTGCTGGTGGTCCCAACGAGACTTGCAAGACTGCTGTCAAGAACATGCTTG
ACTTTATGAAGGCCAACACTGATGTCTGGAAGGGTTTCACATGGTGGTCTGCTGGTCCT
TGGTGGGGTGACTACATGTACAGCTTCGAGCCTCCCAGTGGCTCTGCGTACCAGTACTA
CAACTCCCTTCTCAAGACTTATGTCTAA

>GtCel12A

ATGTTCCGCTTCATCTCTGCTTTGCCCTTTGCCTTGGGCGCATTAAAGCTTGCATCCGC
GGCGACCGTGCTCACTGGTCAATACACCTGTGACACTGCTGGAGATTATACCCTCTGCA
ATGACCAGTGGGGTATTGCCAACGGTGTGGGCTCCCAGACCTCGACTTTGATCAGTGCT
TCTGGCAGCACCATCTCCTGGTCGACCAACTACACTTGGGCTAACAACCCGAATGACGT
CAAGACCTATGCCAATGTTCTCTCCAACACCGCCAAGGGAGTGCAGATCAGTCAAATTA
GCAGCTTCCCCACCACCTGGGATTGGTACTACGAGACTCAGTCGTCTGGCCTCCGCGCT
GATGTTTCGTACGACATGTGGACCGGTACTGCGCAGACCGGCACAGCTGCCAGTTCCTC
TTCATCCTACGAGATCATGATCTGGCTTTCCGGCAAGGGCGGAATCCAACCCGTTGGCT
CTCTGAAGACCTCTGGAATCAGTCTTGCCGGCTATACCTGGAACTTGTGGAGCGGAACC

ACCGAGACCTGGACCACCCTGTCCTTCGTTTCGGCTGACGGGGATATCACTAGCTTCAA
CGCCGAGCTGACTCCCTTCTTCCAGTACTTGGTCGAGAATGAGGGTGTCTCCGCCAGCC
AGTATATCCAGGCCATTCAGACCGGCACGGAAGCCTTCACTGGCACTGCTGAGCTTGTC
ACTACCTCGTTCAGCGTCAGCTTGAGCGGGTGA

>NcCel7B1

ATGGTTCATAAACTCGCCTTCTTAACCGGCCTAACCGCCTCCCTCGTCTCCGCCCAACA
AATCGGTACCATCACCCCCGAATCCCACCCCAAACTGCCCACCAAGCGCTGCACCCTCT
CCGGCGGGTGCCAAACCGTCTCAACCTCCATCGTCCTCGACGCCTTCCAGCGTCCCCTG
CACAAGATCGGCGACCCATCCACCGCTTGCGTCGTAGGCGGTCCTCTCTGCCCCGACGC
CGCCACCTGCGCAGCCAACTGCGCCCTCGAAGGCGTCGATTATGCTTCTTTGGGTGTCA
AGACCGAAGGAGACGCCCTGACGCTGAACCAGTGGGTGTCTGACCCGTCGAACCCAGGA
CATTACAAGACGAGCTCGCCGCGGACTTATCTCGTTGCTGAGGACGGCAAGAACTATGA
AGCTGTCAAGTTGTTGGGCAAGGAGATCTCGTTCGATGTGGATGTTAGTAATTTGCCTT
GTGGCATGAACGGGGCGTTTTATTTGAGTGAGATGTTGATGGATGGTGGACGGGGAGAA
TTCAATGCTGCAGGAGCGGAGTATGGCACGGGGTATTGCGATGCGCAGTGTCCCAAGTT
GGATTTTATCAATGGCGAGGCAAACATCAACAAAACCCACGGCGCGTGCTGCAACGAAA
TGGACATCTTCGAAGCCAACGCCCGCGCAAAGTCCTTCACCCCGCACCCCTGCTCCATC
GAGCGCGTCTACAAGTGCACGGGAGACACCGAGTGCGGCCAGTCCCAGGGCGTCTGCGA
CCAGTGGGGTTGCACTTACAACGAATACCAGAAGGGAGTCCATGATTTCTACGGGCTTG
CTCCTCCCGCCAAAACCATCGACACCACCCAGAAATTCACCGTCACCACACAATTCCTG
ACCGATAATGGCCGTGAAGACGGCGTGTTGGTGGAGATCCGTCGGTTGTGGTCTCAGAA
TGGAAAACTGATCAAGAACGCCAAGATCGCTGTTGATAGTTTATCGACTGATTCAGTGA
GCACGCAGTTCTGCGAAAAGACTTCTTCGTGGACGATGCAGCGCGGCGGACTCAAGACC
ATGGGCGAGGCGATGGGAAGGGGCATGGTGCTGATTTTCAGTATCTGGGCGGATGAGAG
CGGGTTCATGAATTGGTTGGATTCCGGGGACAGTGGACCGTGCAGCGCGACCGAGGGTG
ATCCGAAGCTGATTTTGCAGAAGAAGCCGGATGCGAGGGTGACGTTTTCGAATATCAAG
TGGGGAGAGATGGGTAGTACGTATGCTTCTGCGGGGAAGTATGGTGTTAGACGGGTTGC
GAAGGGGTTGAGTGCTTAG

>NcCel7B2

ATGTCACGAAGGATTCTCTTGTCGGCCTTGCTGGTCGCCGCCGTGTCCGCTCAACAGCC
AGGCAAACTGACACCCGAAGTTCATCCCAAACTTCCAACATGGGCGTGTACCGTTGCCG
ACGGCTGTATCCAGAAGGACACATCCTTGGTCCTCGACTCTGACTACAGATGGGTACAC
ACGGACGACTACACCAACTGCAAGGTCAACGGTCTTAACCCAGCCGTATGTCCCGACGT
TGAGACATGTGCTGCCAACTGTAACCTCGAGGGCGTCGACTACACTGGATCCGGGATTC
ACACCAACGGTAGCGAGCTTACCCTGAATCTCTTTGTCAACCGGACCGACGGTACCACT
AGTCTTGTCTCCCCGCGAGTCTACCTCCTTGCCAACGAGACGACCTACGACATGTTCTC
GCTCCTAGACAAAGAGTTTACCTTTGACGTGGACGTCAGCAAGCTCCCATGCGGAACCA
ATGGAGCTCTTTACTTCAGCGAGATGCTTGCCAATGGAGGCAAGAGCGCGCTGAACCCA
GCCGGCGCAAGCTATGGCACGGGCTACTGTGATGCCCAGTGCCCTACTCCAGCATTCAT
CAACGGCGAGGCCAACCTCGAATCCTACGGCGCCTGCTGCAACGAAATGGACATCTGGG
AAGCCAATTCCCGCGCCACCGCCTTCACTCCTCATCCCTGCAACGTCACCGCGCTCTAC
AAGTGCTCCGGCGATCTCTGCGGTCGTACCGACAAGTACCAATCCGTCTGCGACAAGGA
CGGCTGCGACTACAACCCCTACCGCCTCGGTGACCACCCTTACTACGGTCGCGGCGAGG
GCAACAAAATCGACACCACCCGCCCTTTCACCGTGGTCACTCAGTTCTTCAGCAATACG
ACGAGCGCCGGGGAGAAGGAATTGTCGGCCATTAAGCGGCTGTACCTCCAAGATGGCAA
ACTGATCAGCACCTCTACCATCGCCGTGCCTGGTTTCGATTCAACGAGCGATACCATCA
CCGACGATTATTGCGCCAAGAACAAGCAGATCTTTGGGGGAGTCAACGCCTTTGCCAAC
CAGGGCGGACTTCGACAGATGGGCGAGGCGCTGGATCGGGGCATGGTTCTGGTTTTCAG
CGTGTGGCATGATGCGGGAAGCGCGATGAAGTGGCTGGATGGGACGTTTCCTCCCGGTG
CTGATCCGGAGACACAGCCTGGGACAGAAAGAGGCCCGTGTTTGCCAGGGGAAGGGCAT
GCGGATGATATCCAGAGGGACGCGAGCTGGACGGAAGTCAAGTTTAGTAATGTGAAGAG
TGGGGAGATTGGGTCGACTTTTGAAGCTTGA

>NcCel5A

ATGAAGGCTACGATTCTTGCCAGCACCTTCGCCGCTGGTGCCCTCGCCCAGAGCGGTGC
CTGGGGCCAATGCGGCGGTAACGGTTGGTCCGGCGCTACCAGCTGCATCTCCGGATACG
CCTGCAACTACGTGAACGATTGGTATAGCCAGTGCCAGCCTGGTACTGCCGCTCCTACC
ACCACTGCCGCCGCCACCACCCTCGTCACCTCGACCAAGACCGCCCCTCCTGCTAGCAC
CACCACTGCCACCGCCTCCGGCAAGTTCAAGTGGTTCGGTGTCAACGAGGCCGGCGGTG
AGTTTGGTGATGGTATCTTCCCCGGAAGATGGGGCACTGAGTTCACATTCCCTGACACC

AACACCATCCAGACTCTCCGCAGCCAGGGTTACAACATCTTCCGTGTTGGCTTCGCCAT
GGAGCGCCTTGTCCCCAACACCCTGACGTCGTCTTTCGATAACGGCTATCTCACCAACC
TCACCCAGGTTGTCAACTCTGTTACCAACTCTGGTGCCTATATTGTTCTCGATCCCCAC
AACTATGGCCGTTACTATGGCAAAATCATCACCGATACCGATGCCTTCAAGACTTTCTG
GCAAAATGTTGCTGCCAAGTTTGCTTCCAACTCCAAGGTCATCTTCGATACCAACAACG
AGTACAACACCATGGACCAGACCCTTGTCCTCAACCTCAACCAGGCCGCCATTGACGGT
ATTCGTGCGGCCGGCGCCACTTCGCAGTACATTTTCGTCGAGGGTAACCAATGGACTGG
CGCCTGGTCATGGAATGTGACCAACACCAATTTGGCTGCCCTTACCGATCCCGAGAACA
AGATCGTCTACGAGATGCACCAGTACCTCGACTCCGATAGCTCCGGTACCAGCACTGCC
TGCGTTTCCTCCGAGATCGGTGTTCAGCGCATTGTTGGTGCCACTGCCTGGCTCCGCGC
CAATGGCAAAAAGGGTGTCCTCGGCGAGTTCGCTGGTGGTGCCAACTCAGTCTGCAAGG
CTGCCGTTACTGGCCTTCTTGAGCACCTCAAGGCTAACACTGACGTCTGGGAGGGTGCC
CTTTGGTGGGCTGCCGGTCCCTGGTGGGGTGACTACATGTATAGCTTTGAGCCTCCTTC
GGGCACTGGCTATACCTACTACAACAGCCTTCTCAAGACCTATACCCCTTAA

>NhCel12A

ATGAAGTCCGCCATCGTCGCAGCTCTAGCCGGTCTAGCCGCTGCCTCACCAACCCGCCT
GATTCCCCGCGGCCAGTTCTGCGGCCAGTGGGATTCCGAGACGGCGGGAGCCTACACCA
TCTACAACAACCTCTGGGGCAAGGACAATGCCGAGTCTGGGGAGCAGTGCACCACCAAC
AGCGGCGAGCAGAGCGACGGCAGCATCGCCTGGTCCGTCGAGTGGTCCTGGACCGGTGG
TCAGGGGCAGGTCAAGTCGTATCCCAACGCTGTTGTTGAGATTGAGAAGAAGACTCTCG
GGGAGGTTTCGAGCATTCCTTCGGCTTGGGACTGGACTTATACTGGCGATGGAATCATT
GCCAACGTTGCGTACGACCTCTTTACGAGTTCTACAGAGTCAGGAGACGCCGAGTACGA
GTTCATGATCTGGCTGTCAGCTCTGGGAGGAGCTGGACCGATCAGCAACGACGGATCAC
CCGTCGCAACCGTCGAGCTCGCCGGTACCTCATGGAAGCTGTATCAAGGAAAGAACAAC
CAGATGACTGTCTTCAGCTTCGTAGCCGAGTCAGACGTCAACAACTTCTGCGGCGACCT
GGCAGATTTCACCGACTATCTCGTCGACAACCACGGCGTCAGCAGCTCCCAGATCCTGC
AGAGCGTGGGTGCTGGTACTGAGCCTTTTGAGGGTACCAATGCTGTGTTTACTACCAAC
AACTATCACGCTGATGTTGAATACTAA

>PsCel12A1

ATGAAGGTTGCGTTCGCTACTGCCATGGCCGCTGCAGCGTTGGCGGCTGCCTATGCCGA
CGACTTCTGCGACCAGTGGGGCACCACAACCACGGACAACTACATCATCTACAACAACC
TGTGGGGCGAGAGCTACGCCACGTCGGGCAGTCAGTGCACTGGCCTGGACAGCAGCAGC
GGCTCCACGGTTGCGTGGCACACCAACTGGACGTGGACCGGCGCATCGTCCAATGTCAA
GTCGTTCGCCAACGCCGCCCTCCAGTTTGACGCTGTGCAGCTCTCCAGCGTCTCGTCCA
TCCCGACGACCATGGAGTACTCGCTCGAGTACAGCGGGAACATCGCCGCCGATGTTTCG
TACGATTTGTTCACGGCCTCGACATCCACCGGCGACAATGAGTTCGAGATTATGATCTG
GCTGGCTGCTCTTGGTGGCGCAGGCCCCATTTCGAGCACGGGCTCCGCTGTAGCCACGA
CGACCATCGCGGACACCTCGTTCAGTCTGTACACCGGAGCTAACGGTGACACGACAGTG
TACTCATTTGTGGCTTCCGACACAGTCAAGAGCTTTTCAGGCGACCTCATGGACTTCTT
CACGTACCTCATTGATAACGAGGGCTTCTCTTCGAGCCAGTACCTGAACACTGTTCAGG
CCGGCACGGAGCTTTTCACCGGCACCGACGTAACTCTGACGGTGTCGTCTTACTCGGCC
GCAGTCAACATGGGAGCTTCCAGTGGAACTACTGCGACAACTTCGACGGCCAGCTCGTC
CACAGGAACCTCCGTCTCGACGAAGGAGCTCATCGCGAGTAGCTCGTCGACTGAGTCGG
ACTCGCAAACCGCCGCGCTGACTGCTACGACGGCCACCCCGTCCACGGACTCGTCGTAC
AGCCAGCAGGAAACAACTGCGCCGTCGACGTCCAGCACGTCAGTGAGCGGTGAAGCCAC
AACTTCAAGCACGAGCTCGACCTGGAACGAACAGGAATCGACTGCGGCCTCTGCATCTG
AGACCACGGCTCCGTCAACTACTTCGAGTACGTCCGAGGAGACCACGTCGTCTAGCGCC
ACGGAAACTACGACTACCAGCGTTACCGCGGCCCCCAGCACCTCGACGACGACTGGAAA
GAAGTGCGCTAGCCGCAGTGTTCGTCGCGTCTAA

>PsCel12A2

ATGAAGAGCTCCGCCGTCCTCGCTGCTTCCGCTCTCGCCATCGCGGCCTCGCCTGCCTT
CGCCCAAGAAGAGTTCTGTGGCCAGTGGAACTTGACCAAGACGGACGACTACATCCTCT
ACAACAACCTCTGGGGCGCCTTCGACGACCCCACCGGCACCCAGTGCACGGGTCTCGAC
AGCGTCAAAGGCTCCACCGTCGCGTGGCACACGAGCTTCGGCTGGTCGGGCTCCAAGAC
GCAGGTCAAGTCGTTCGCCAACGTCGCGCTCCAGTTCGACCACCTGCCGATCTCCGAGG
TGACATCCATCCCCTCGACCATCCACTTCAAGTACGACTACGAAGAGAGCCTCATCGCC
AACGTGGCGTTCGACCTGTTCACGTCCTCCACGCCTGACGGCGACGCCGAGTACGAGAT
CATGGTGTGGCTGGCGGCTATCGGCGGCGCGGGTCCCATCTCGAGCACGGGCTCCGCTG
TGGACCAGGTGACGGTGGGCGGCGTCGACTTCAGCCTGTACGCCGGCAAGAACGGCAAC

ATGACGGTCTACTCGTTCGTCGCGTCCACGATGGTCAACCGGTACGCGACCGACTTTAA
GCAGTTCTTCGACGTGCTCCCCAGGAATCTCACCATCGACCCGAGCCAGTACCTGATCA
ACGTCCAGGCCGGCACTGAGCCGTTCGTCGGCAACGGCACGCTCACCGTGTCCAAGTAC
TCGGCCGCGGTCAACCCTGCGGAGTACTCGCAGGTCCAGCAGCAGTAA

>PsCel12A3

ATGAAGAGCTTTCTCCAACTCGTTGTCGTCGTCGCCGCCTTGCTGTCCGTGTCCACCGC
CGACTTCTGCAGTCAGTGGCGTCTGTCCAAGGCCGGCAAGTACGTGATCTACAACAACC
TCTGGAACAAGAACGCCGCCGCATCGGGCAGCCAATGCACGGGCGTCGACAAGATCAGC
GGCTCCACCATCGCCTGGCACACGTCGTACACCTGGACGGGAGGCGCGGCCACGGAGGT
CAAGTCGTACTCGAACGCCGCGCTGGTCTTCTCCAAGAAGCAGATCAAGAACATCAAGT
CGATCCCCACCAAGATGAAGTACTCGTACTCGCACTCCTCGGGCACGTTCGTCGCTGAC
GTGTCGTACGACCTGTTCACGAGCTCCACCGCCAGCGGCAGCAACGAGTACGAGATCAT
GATCTGGCTGGCCGCGTACGGCGGCGCCGGCCCGATCTCCAGCACGGGCAAGGCCATCG
CCACCGTCACCATCGGCAGCAACAGCTTCAAGCTGCCTTCCACTCAGTACCTGACCACG
CTCGAAGCCGGCACGGAGCCCTTCACGGGCTCGAACGCCAAGATGACCGTATCCTCGTT
CTCCGCCGCTGTCAACTAA

>StCel5A

ATGAAGTCCTCCATCCTCGCCAGCGTCTTCGCCACGGGCGCCGTGGCTCAAAGTGGTCC
GTGGCAGCAATGTGGTGGCATCGGATGGCAAGGATCGACCGACTGTGTGTCGGGCTACC
ACTGCGTCTACCAGAACGATTGGTACAGCCAGTGCGTGCCTGGCGCGGCGTCGACAACG
CTGCAGACATCGACCACGTCCAGGCCCACCGCCACCAGCACCGCCCCTCCGTCGTCCAC
CACCTCGCCTAGCAAGGGCAAGCTGAAGTGGCTCGGCAGCAACGAGTCGGGCGCCGAGT
TCGGGGAGGGCAATTACCCCGGCCTCTGGGGCAAGCACTTCATCTTCCCGTCGACTTCG
GCGATTCAGACGCTCATCAATGATGGATACAACATCTTCCGGATCGACTTCTCGATGGA
GCGTCTGGTGCCCAACCAGTTGACGTCGTCCTTCGACCAGGGTTACCTCCGCAACCTGA
CCGAGGTGGTCAACTTCGTGACGAACGCGGGCAAGTACGCCGTCCTGGACCCGCACAAC
TACGGCCGGTACTACGGCAACATCATCACGGACACGAACGCGTTCCGGACCTTCTGGAC
CAACCTGGCCAAGCAGTTCGCCTCCAACTCGCTCGTCATCTTCGACACCAACAACGAGT
ACAACACGATGGACCAGACCCTGGTGCTCAACCTCAACCAGGCCGCCATCGACGGCATC

CGGGCCGCCGGCGCGACCTCGCAGTACATCTTCGTCGAGGGCAACGCGTGGAGCGGGGC
CTGGAGCTGGAACACGACCAACACCAACATGGCCGCCCTGACGGACCCGCAGAACAAGA
TCGTGTACGAGATGCACCAGTACCTCGACTCGGACAGCTCGGGCACCCACGCCGAGTGC
GTCAGCAGCACCATCGGCGCCCAGCGCGTCGTCGGAGCCACCCAGTGGCTCCGCGCCAA
CGGCAAGCTCGGCGTCCTCGGCGAGTTCGCCGGCGGCGCCAACGCCGTCTGCCAGCAGG
CCGTCACCGGCCTCCTCGACCACCTCCAGGACAACAGCGACGTCTGGCTGGGTGCCCTC
TGGTGGGCCGCCGGTCCCTGGTGGGGCGACTACATGTACTCGTTCGAGCCTCCTTCGGG
CACCGGCTATGTCAACTACAACTCGATCTTGAAGAAGTACTTGCCGTAA

>TrCel7B

ATGGCGCCCTCAGTTACACTGCCGTTGACCACGGCCATCCTGGCCATTGCCCGGCTCGT
CGCCGCCCAGCAACCGGGTACCAGCACCCCCGAGGTCCATCCCAAGTTGACAACCTACA
AGTGTACAAAGTCCGGGGGGGTGCGTGGCCCAGGACACCTCGGTGGTCCTTGACTGGAAC
TACCGCTGGATGCACGACGCAAACTACAACTCGTGCACCGTCAACGGCGGCGTCAACAC
CACGCTCTGCCCTGACGAGGCGACCTGTGGCAAGAACTGCTTCATCGAGGGCGTCGACT
ACGCCGCCTCGGGCGTCACGACCTCGGGCAGCAGCCTCACCATGAACCAGTACATGCCC
AGCAGCTCTGGCGGCTACAGCAGCGTCTCTCCTCGGCTGTATCTCCTGGACTCTGACGG
TGAGTACGTGATGCTGAAGCTCAACGGCCAGGAGCTGAGCTTCGACGTCGACCTCTCTG
CTCTGCCGTGTGGAGAGAACGGCTCGCTCTACCTGTCTCAGATGGACGAGAACGGGGGC
GCCAACCAGTATAACACGGCCGGTGCCAACTACGGGAGCGGCTACTGCGATGCTCAGTG
CCCCGTCCAGACATGGAGGAACGGCACCCTCAACACTAGCCACCAGGGCTTCTGCTGCA
ACGAGATGGATATCCTGGAGGGCAACTCGAGGGCGAATGCCTTGACCCCTCACTCTTGC
ACGGCCACGGCCTGCGACTCTGCCGGTTGCGGCTTCAACCCCTATGGCAGCGGCTACAA
AAGCTACTACGGCCCCGGAGATACCGTTGACACCTCCAAGACCTTCACCATCATCACCC
AGTTCAACACGGACAACGGCTCGCCCTCGGGCAACCTTGTGAGCATCACCCGCAAGTAC
CAGCAAAACGGCGTCGACATCCCCAGCGCCCAGCCCGGCGGCGACACCATCTCGTCCTG
CCCGTCCGCCTCAGCCTACGGCGGCCTCGCCACCATGGGCAAGGCCCTGAGCAGCGGCA
TGGTGCTCGTGTTCAGCATTTGGAACGACAACAGCCAGTACATGAACTGGCTCGACAGC
GGCAACGCCGGCCCCTGCAGCAGCACCGAGGGCAACCCATCCAACATCCTGGCCAACAA
CCCCAACACGCACGTCGTCTTCTCCAACATCCGCTGGGGAGACATTGGGTCTACTACGA
ACTCGACTGCGCCCCCGCCCCCGCCTGCGTCCAGCACGACGTTTTCGACTACACGGAGG
AGCTCGACGACTTCGAGCAGCCCGAGCTGCACGCAGACTCACTGGGGGCAGTGCGGTGG

CATTGGGTACAGCGGGTGCAAGACGTGCACGTCGGGCACTACGTGCCAGTATAGCAACG
ACTACTACTCGCAATGCCTTTAG

>TrCel5A

ATGAACAAGTCCGTGGCTCCATTGCTGCTTGCAGCGTCCATACTATATGGCGGCGCCGC
TGCACAGCAGACTGTCTGGGGCCAGTGTGGAGGTATTGGTTGGAGCGGACCTACGAATT
GTGCTCCTGGCTCAGCTTGTTCGACCCTCAATCCTTATTATGCGCAATGTATTCCGGGA
GCCACTACTATCACCACTTCGACCCGGCCACCATCCGGTCCAACCACCACCACCAGGGC
TACCTCAACAAGCTCATCAACTCCACCCACGAGCTCTGGGGTCCGATTTGCCGGCGTTA
ACATCGCGGGTTTTGACTTTGGCTGTACCACAGATGGCACTTGCGTTACCTCGAAGGTT
TATCCTCCGTTGAAGAACTTCACCGGCTCAAACAACTACCCCGATGGCATCGGCCAGAT
GCAGCACTTCGTCAACGACGACGGGATGACTATTTTCCGCTTACCTGTCGGATGGCAGT
ACCTCGTCAACAACAATTTGGGCGGCAATCTTGATTCCACGAGCATTTCCAAGTATGAT
CAGCTTGTTCAGGGGTGCCTGTCTCTGGGCGCATACTGCATCGTCGACATCCACAATTA
TGCTCGATGGAACGGTGGGATCATTGGTCAGGGCGGCCCTACTAATGCTCAATTCACGA
GCCTTTGGTCGCAGTTGGCATCAAAGTACGCATCTCAGTCGAGGGTGTGGTTCGGCATC
ATGAATGAGCCCCACGACGTGAACATCAACACCTGGGCTGCCACGGTCCAAGAGGTTGT
AACCGCAATCCGCAACGCTGGTGCTACGTCGCAATTCATCTCTTTGCCTGGAAATGATT
GGCAATCTGCTGGGGCTTTCATATCCGATGGCAGTGCAGCCGCCCTGTCTCAAGTCACG
AACCCGGATGGGTCAACAACGAATCTGATTTTTGACGTGCACAAATACTTGGACTCAGA
CAACTCCGGTACTCACGCCGAATGTACTACAAATAACATTGACGGCGCCTTTTCTCCGC
TTGCCACTTGGCTCCGACAGAACAATCGCCAGGCTATCCTGACAGAAACCGGTGGTGGC
AACGTTCAGTCCTGCATACAAGACATGTGCCAGCAAATCCAATATCTCAACCAGAACTC
AGATGTCTATCTTGGCTATGTTGGTTGGGGTGCCGGATCATTTGATAGCACGTATGTCC
TGACGGAAACACCGACTGGCAGTGGTAACTCATGGACGGACACATCCTTGGTCAGCTCG
TGTCTCGCAAGAAAGTAG

>TrCel12A

ATGAAGTTCCTTCAAGTCCTCCCTGCCCTCATACCGGCCGCCCTGGCCCAAACCAGCTG
TGACCAGTGGGCAACCTTCACTGGCAACGGCTACACAGTCAGCAACAACCTTTGGGGAG
CATCAGCCGGCTCTGGATTTGGCTGCGTGACGGCGGTATCGCTCAGCGGCGGGGCCTCC

TGGCACGCAGACTGGCAGTGGTCCGGCGGCCAGAACAACGTCAAGTCGTACCAGAACTC
TCAGATTGCCATTCCCCAGAAGAGGACCGTCAACAGCATCAGCAGCATGCCCACCACTG
CCAGCTGGAGCTACAGCGGGAGCAACATCCGCGCTAATGTTGCGTATGACTTGTTCACC
GCAGCCAACCCGAATCATGTCACGTACTCGGGAGACTACGAACTCATGATCTGGCTTGG
CAAATACGGCGATATTGGGCCGATTGGGTCCTCACAGGGAACAGTCAACGTCGGTGGCC
AGAGCTGGACGCTCTACTATGGCTACAACGGAGCCATGCAAGTCTATTCCTTTGTGGCC
CAGACCAACACTACCAACTACAGCGGAGATGTCAAGAACTTCTTCAATTATCTCCGAGA
CAATAAAGGATACAACGCTGCAGGCCAATATGTTCTTAGCTACCAATTTGGTACCGAGC
CCTTCACGGGCAGTGGAACTCTGAACGTCGCATCCTGGACCGCATCTATCAACTAA

Codon optimized endoglucanase nucleotide sequences

>AfCel7B1opt

ATGGATTCCAAGAGAGGAGTAGTAGCTGCCGTTCTTGCTCTATTGCCGTTAGTTAGTGC
TCAACAACCTGCTGCAAGTAGTGCCGGTAATCCCAAATTAACTACGTATAAGTGTACGA
CCGCCGGTGGTTGTGTTGCCCAAGACACATCCGTCGTGTTGGATTGGGGTTACCACTGG
ATCCACACGGTTGACGGCTATACGAGTTGCACAACGTCATCTGGTGTAGATTCCACTTT
GTGCCCAGACGCTGCTACCTGCGCTAAAAATTGTGTTATAGAACCAGCAAATTATACAA
GTGCCGGCGTTACTACATCTGGAGACTCATTAACAATGTATCAGTATGTGCAATCTAAT
GGCGTTTATACCAACGCCTCTCCTAGATTGTACCTGTTGGGACCAGACAAAAATTACGT
TATGTTAAAGTTATTAGGTCAAGAATTGACGTTCGATGTCGACCTATCTACATTGCCAT
GCGGTGAGAATGGAGCACTATATTTGTCTGAGATGAGCGCTACTGGTGGTAGGAATGAA
TATAACACTGGTGGAGCTGAGTATGGCTCCGGTTATTGTGATGCTCAATGTCCGGTAAT
AGCTTGGAAAAATGGAACACTAAATACTTCTGGTGCATCTTATTGTTGCAACGAAATGG
ATATCCTAGAGGCAAACTCTAGAGCTAATTCTTATACACCACACCCATGTTCTGCTACT
GATTGCGATAAAGGTGGTTGCGGCTTCAATCCGTATGCACTGGGTCAAAAGTCTTACTG
GGGCCCAGGCGGCACGGTGGACACATCAAAGCCATTTACTATCACCACCCAATTTATTA
CGAACGATGGAACAACTACTGGTACTCTTTCTGAAATTCGTAGGCAGTATATGCAAAAT
GGCAAGGTCATAGCAAATGCTGTTTCTTCTACAGGAGTTAATTCAATTACTGAGGATTG
GTGTACTTCCGTTGATGGTTCAGCAGCTACTTTCGGGGGTTTGACAACTATGGGTAAGG
CGCTTGGCCGTGGCATGGTCTTAATCTTTAGCATTTGGAATGATGCTTCTGGTTTTATG

AACTGGTTAGATTCTGGGAACGCAGGCCCATGCAGCAGTACCGAGGGTAACCCCGATTT
GATCAAGGCACAGAATCCCACTACACATGTGGTCTTTTCTAATATCAGGTGGGGTGATA
TAGGGTCTACATTTAAAGGCAGCGACGGATCTGTGACAACTACTACGTCTACTACGAGC
ACTAAGACTACGACTAGCACTGCTCCTGGTCCAACACAAACACATTATGGCCAATGCGG
CGGGCAAGGTTGGACTGGCCCTACCGCATGCGCATCCCCATATACTTGTCAGGTTTTGA
ACCCGTGGTACTCACAATGTCTGTAA

>AfCel5A1opt

ATGAAGGCATCAACGATTATATGCGCTTTATTACCGCTGGCAGTAGCAGCCCCTAATGC
AAAAAGAGCGTCCGGTTTTGTCTGGTTTGGTTCGAATGAAAGTGGCGCTGAATTTGGCG
AGACTAAGTTACCTGGAGTTTTGGGTACCGACTATATCTGGCCCGATGCATCGACAATA
AAGACTCTACATGATGCAGGTATGAACATTTTCAGAGTAGCATTCAGAATGGAAAGATT
GATTCCTAACCAGATGACTGGTACTCCAGATGCTACCTATTTAAACGACTTAAAGGCTA
CGGTGAATGCTATTACTAGCCTGGGAGCTTATGCTGTAATTGATCCTCATAATTATGGT
AGATATTACGGTAATATTATCAGTTCTACAGATGATTTTGCAGCGTTTTGGAAGACTCT
GGCAGCTCAATTCGCTAGCAACGATCATGTGATTTTTGACACCAATAATGAGTACCATG
ATATGGATCAAACATTGGTTTTGAACCTAAACCAAGCTGCGATAAACGCGATTAGAGCC
GCTGGCGCCACTTCACAATACATTTTTGTCGAAGGCAACTCCTGGTCAGGGGCGTGGAC
ATGGACGACTGTCAATGACAACTTAAAAGCCTTAACGGATCCACAAGATAAAATTGTTT
ATGAAATGCATCAATATTTAGACTCAGATGGATCCGGAACAAGTGCTACATGTGTTTCA
TCAACAATAGGTCAAGAAAGGGTACAAGCTGCAACTCAATGGCTAAAGACTAACGGTAA
AAAGGGTATTATTGGCGAATTTGCTGGTGGAGCTAATTCTGTTTGCCAGAGCGCTGTAA
CGGGTATGTTGGATTATCTTTCGGCTAATTCTGACGTGTGGATGGGCGCAGCATGGTGG
GCTGCCGGACCATGGTGGGCGGACTACATTTTTAACATGGAACCACCATCTGGAACAGC
ATATCAGAATTACCTATCTTTATTGAAACCATATTTTGTGGGTGGTTCTGGAGGAAATC
CTCCAACTACCACTACTACTACAACTCAACCAACCACGACTACTACTACCACAACTGCA
GGAAATCCCGGGGGAACCGGCTTAGCTCAACATTGGGGCCAATGTGGAGGCATCGGTTG
GACCGGCCCAACCGCTTGTGCTTCCCCATATACGTGCCAAAAACTTAATGATTACTATA
GCCAATGTTTGTAA

>Apul15654opt

ATGAAATACTCAACTTTTGTCGTAGCTGCTGCTGCTGGATCTGCAGCTGCATCAAGTTC
AGCATACGGACAATGTGGGGGTAACGGTTGGACAGGCGCAACAGATTGCGTCTCTGGTT
ATCATTGCGCTTATCAAAATGATTGGTATAGCCAATGCGTTCCTGGAGCTGGTTCTGGT
TCTACTTCCGCAGCCGCTGCTGCTGCGACGACTAAAGTCGTGACATCTGCAAAGGCATC
AACGACTCAAGCCCCGTCAAAGGCTCCAGTTTCAACAAGCGCTGCTTCCACCAGTAAAG
CTGTAGCGGCCACTTCTGGCAAGGTTAAATACGCTGGTGTAAATATTGCCGGCTTTGAC
TTTGGAATGGATACTAATGGAGCGTCATCCGGTAGTTATGTTGATCCAGGTACTACCGG
TCAGAATCAGATGAATCACTTTGTAAAGGACGATAAGCTAAACGCTTTCAGGCTACCAG
TAGGCTGGCAGTACTTGGTTAATTCCCAATTGGGTGGTACCTTGGATAGTACCTTTTTC
GCCAAATATGATCAACAGATGACCTATTGTTTAAATAGCGGTGCAGCGCTATGTATCTT
AGACCTGCATAACTACGCAAGGTGGAACGGACAAATCGTCGGGACATCTGGTGGTCCTA
CTAACGCACAATTCGCGAGTGTCTGGAGTCAACTAGCTAAGAAGTACTCTAGCAAACCA
AAGGTCGCCTTTGCTATTATGAATGAGCCTCATGATCTGCAAGATATTAATGCTTGGGC
TACTACGGTGCAGGCTGCAGTTACTGCCATCAGACAAGCTGGTGCGACGCAAAATATGA
TATTGTTGCCTGGAGATGATTGGACTCATGCTGCAAACTTTGTTGATAATGGTTCAGCA
GCTGCCTTGAACAAAATTACTAACCTTGACGGGTCAAAAACCAATTTAGTATTTGACGT
CCATCAATACTCAGATTCAGATGGTTCTGGTACTAGCTCGACCTGTGTTAGCTCCGCAT
CGAATATCGCTGGCTTTACTAAGTTGGCTACCTGGCTTAGGTCAAATAATAGACAAGCT
ATGCTTACGGAAGCCGGGGGTTCTAACGATCAAAGCTGCTTGACTGCGGTTTGCGATGT
GCTGAACTATTTAATGACAAATTCAGATGTTTATTTAGGTTGGACTGGTTGGTCAGCAG
GTATGTTTGCTACAGATTACGTTTTAAGTGAAGTTCCCACTGGATCAGCGGGCTCATAC
AAGGATCAAGCTATAGTTACAAAGTGTATCGCTGGAGTTTTCAACTCTAAATAA

>GtCel12Aopt

ATGTTTAGGTTTATTTCTGCTTTACCTTTTGCTTTGGGAGCTTTGAAATTAGCTAGTGC
TGCGACTGTGCTTACCGGTCAATATACCTGCGATACTGCTGGGGACTATACTTTATGTA
ATGATCAATGGGGTATCGCTAATGGTGTGGGCTCACAAACTTCAACCTTGATTAGTGCA
TCTGGGTCTACTATCTCATGGTCTACGAATTATACTTGGGCAAATAACCCTAACGATGT
TAAAACATACGCGAATGTGCTGAGTAATACCGCCAAGGGTGTGCAAATATCACAAATTT
CTTCATTTCCAACTACATGGGATTGGTACTATGAAACACAGTCATCTGGATTACGTGCT

GATGTCTCATACGATATGTGGACAGGCACTGCCCAAACCGGCACGGCTGCCAGTTCTAG
CTCATCCTATGAAATCATGATATGGCTGAGCGGGAAGGGTGGAATCCAACCCGTCGGTT
CCTTAAAAACTTCAGGCATCTCATTGGCTGGGTATACATGGAATTTATGGTCCGGCACC
ACCGAAACCTGGACAACGCTATCTTTTGTTTCAGCCGACGGCGACATAACTAGTTTTAA
TGCAGAATTGACTCCATTCTTCCAGTACTTAGTTGAGAATGAAGGTGTTTCTGCAAGCC
AGTATATACAGGCTATTCAAACCGGTACTGAAGCATTCACTGGAACAGCCGAATTGGTT
ACAACAAGTTTCAGTGTTAGTCTATCTGGGTAA

>StCel5Aopt

ATGAAGTCAAGTATATTGGCATCTGTTTTTGCTACCGGAGCTGTTGCTCAGTCCGGCCC
TTGGCAACAATGCGGTGGTATCGGCTGGCAAGGATCCACAGACTGTGTTAGTGGCTACC
ATTGCGTTTACCAAAATGATTGGTACTCCCAATGTGTCCCTGGTGCGGCGTCCACAACT
TTGCAAACTTCTACAACATCTAGACCAACCGCAACCTCTACTGCTCCTCCTTCCAGTAC
TACATCACCTTCCAAAGGAAAGCTTAAGTGGCTTGGTTCAAATGAATCTGGAGCTGAAT
TTGGAGAAGGTAACTACCCGGGTCTATGGGGGAAGCATTTCATATTTCCTTCCACTTCA
GCAATACAGACTCTGATTAATGATGGTTACAATATATTTCGTATCGATTTTAGTATGGA
AAGACTTGTCCCAAATCAACTTACTTCAAGTTTTGACCAAGGCTATTTAAGGAACTTAA
CGGAGGTGGTAAACTTCGTTACCAACGCTGGTAAATACGCGGTCCTTGATCCACATAAT
TACGGTAGGTATTACGGGAACATCATAACTGATACTAACGCGTTTAGAACTTTCTGGAC
TAATTTGGCTAAGCAATTCGCATCAAATTCATTGGTAATTTTCGATACAAATAATGAAT
ACAATACTATGGATCAAACGTTAGTTTTGAATTTAAATCAAGCAGCAATTGATGGTATA
AGAGCTGCTGGTGCTACATCTCAGTATATTTTTGTCGAAGGTAACGCCTGGTCTGGCGC
ATGGTCTTGGAATACTACCAACACTAATATGGCAGCTTTGACGGATCCTCAAAATAAGA
TTGTATATGAGATGCATCAGTATTTAGATTCCGATTCAAGTGGAACTCACGCCGAATGC
GTAAGCTCTACAATAGGAGCGCAGAGGGTGGTTGGCGCCACTCAATGGTTACGTGCGAA
CGGTAAACTGGGTGTTCTGGGCGAATTTGCTGGCGGAGCTAATGCTGTTTGCCAGCAAG
CTGTTACTGGTTTATTAGATCATTTACAAGATAATTCAGATGTTTGGTTGGGAGCACTT
TGGTGGGCAGCCGGTCCTTGGTGGGGTGATTATATGTATTCTTTCGAGCCGCCATCCGG
TACAGGATATGTCAATTACAATTCTATCTTGAAAAAATATCTTCCCTAA

Cellobiohydrolase nucleotide sequences

>AnCel6A

ATGCACTCCACCAACATGCGAGCCATCTGGCCCCTCGTTTCTCTTTTCTCTGCCGTCAA
GGCCCTCCCCGCCGCAAGCGCAACTGCTTCAGCGTCCGTTGCTGCATCGAGCTCTCCGG
CGCCAACGGCCTCTGCTACGGGCAATCCCTTTGAGGGATACCAGCTCTATGCGAACCCC
TACTATAAGTCACAAGTGGAGAGTTCGGCCATTCCATCATTGTCTGCTAGTTCACTGGT
CGCGCAGGCGAGTGCTGCTGCAGATGTGCCTTCGTTCTACTGGCTGGACACGGCCGACA
AGGTACCTACAATGGGTGAATATCTGGAAGACATCCAGACACAAAATGCTGCTGGAGCG
AGCCCCCCAATTGCTGGTATCTTCGTTGTCTATGACCTACCAGATCGGGATTGTTCTGC
CTTGGCTAGTAATGGGGAATACTCGATCAGTGATGGTGGTGTGGAGAAGTACAAGGCGT
ACATTGATTCTATTCGGGAGCAGGTCGAGACGTACTCGGATGTTCAGACTATTCTGATT
ATTGAACCCGATAGCTTGGCTAACCTGGTGACGAATCTCGATGTGGCTAAATGCGCTAA
TGCTGAATCCGCTTACCTGGAATGCACCAACTATGCCCTTGAGCAACTGAATCTGCCGA
ACGTGGCTATGTATCTTGATGCTGGCCATGCGGGATGGCTGGGATGGCCTGCCAACATC
GGTCCCGCGGCGCAACTCTACGCATCGGTGTATAAGAATGCGTCGTCCCCAGCTGCTGT
TCGCGGCCTCGCCACCAATGTAGCTAACTTCAATGCCTGGAGCATCGACTCTTGCCCCT
CTTATACATCGGGCAACGATGTCTGTGATGAGAAAAGCTACATCAATGCCATTGCGCCG
GAGCTGTCTAGTGCTGGGTTTGATGCCCACTTCATTACCGATACGGGTCGCAATGGGAA
GCAACCCACCGGTCAGAGCGCGTGGGGTGACTGGTGCAATGTCAAGGATACCGGCTTCG
GTGCTCAACCGACGACCGATACTGGAGACGAGCTGGCTGATGCCTTTGTCTGGGTCAAG
CCGGGCGGAGAGAGCGACGGGACATCGGATACCAGCTCTTCTCGCTACGACGCGCATTG
CGGATATAGCGATGCCTTGCAGCCTGCTCCGGAGGCCGGAACCTGGTTCCAGGCATACT
TTGAGCAACTTTTGACCAACGCCAATCCTTCCCTTTGA

>LeCel6A

ATGAAGATTACTTCCACTGGCTTACTTGCTCTCTCATCTTTACTGCCGTTTGCGCTCGG
TCAGTCTCAGTTATATGGCCAATGCGGCGGTATAGGCTGGAGTGGCGCGACTACCTGTG

TTTCCGGCGCGACCTGCACGGTCGTAAACGCTTACTACTCACAGTGCCTTCCTGGTTCG
GCCTCGGCACCTCCAACTTCGACAAGTAGCATCGGAACAGGAACTACTACTTCATCCGC
GCCCGGAAGCACTGGAACCACGACACCTGCTGCCGGCAACCCCTTCACTGGTTACGAGA
TCTATCTCAGTCCTTACTACGCCAACGAAATAGCTGCTGCAGTCACGCAGATCAGTGAT
CCTACCACTGCCGCTGCCGCAGCTAAAGTCGCAAACATCCCCACGTTCATTTGGCTGGA
TCAAGTTGCCAAAGTACCAGATTTAGGAACGTACCTGGCGGATGCTAGTGCCAAGCAAA
AGTCCGAAGGAAAGAATTACCTCGTACAGATTGTAGTATACGACCTCCCAGACCGCGAC
TGTGCTGCTTTGGCGTCCAACGGAGAATTCACCATTGCTGATAATGGCGAAGCAAACTA
CCACGACTATATTGATCAAATCGTGGCACAAATCAAACAATACCCCGATGTCCACGTTG
TTGCTGTTATTGAACCTGACTCTTTGGCCAACCTTGTCACAAACCTCAGCGTCGCCAAG
TGCGCGAACGCCCAGACCACTTACTTGGAATGTGTCACATATGCCATGCAACAACTTTC
TGCCGTCGGCGTTACCATGTATCTCGATGCTGGTCATGCCGGCTGGTTGGGCTGGCCGG
CGAATTTTTCCCGGCTGATCAGCTGTTCACATCCTTATACTCCAATGCTGGGTCACCCT
CTGGCGTTCGAGGACTTGCTACCAACGTTGCCAACTATAATGCCCTCGTTGCTAATACA
CCAGATCCTATTACCCAAGGTGATCCCAACTACGACGAGATGCTTTACATCGAGGCTCT
GGCTCCTTTGCTTGGCTCCTTCCCTGCTCATTTTATCGTTGACCAAGGTCGTTCAGGTG
TCCAGGACATCAGACAGCAATGGGGTGACTGGTGTAACGTTCTTGGTGCTGGATTCGGA
ACGCAGCCTACGACTAACACCGGATCATCCTTGA

>NcCel7A

ATGCTCGCCAAGTTCGCTGCCCTTGCGGCCCTTGTGGCCTCTGCCAACGCCCAGGCTGT
TTGCTCTCTCACTGCCGAGACCCATCCCTCCCTCAACTGGTCCAAGTGCACTTCCTCCG
GATGCACGAACGTCGCCGGATCCATCACTGTTGATGCCAACTGGCGCTGGACTCACATT
ACTTCTGGTAGCACCAACTGCTACAGCGGCAACGAGTGGGACACCTCCCTCTGCAGCAC
CAACACCGATTGCGCGACCAAGTGCTGCGTTGATGGTGCTGAGTACTCTTCTACCTATG
GTATTCAGACCAGCGGCAACTCGCTCAGCCTTCAGTTCGTCACCAAGGGCTCGTACTCG
ACCAACATTGGTTCCCGTACTTACCTTATGAACGGTGCCGATGCCTACCAGGGCTTCGA
GCTCCTTGGCAACGAGTTCACCTTCGATGTCGATGTGTCCGGCACCGGCTGCGGTCTTA
ACGGCGCCCTCTACTTCGTCTCCATGGATCTTGATGGTGGCAAGGCCAAGTACACCAAC
AACAAGGCTGGTGCCAAGTACGGTACCGGTTACTGCGACGCTCAGTGCCCCCGTGATCT
CAAGTACATCAACGGTATCGCCAACGTTGAGGGCTGGACCCCCTCCACCAACGATGCTA
ACGCTGGTATTGGTGACCACGGTACTTGCTGCTCTGAGATGGATATCTGGGAAGCGAAC

AAAGTCTCTACAGCGTTCACCCCGCACCCCTGCACCACCATCGAACAGCACATGTGCGA
GGGTGACTCCTGCGGTGGTACCTATTCCGACGACCGCTATGGCGGTACTTGCGATGCCG
ATGGTTGTGACTTCAACAGCTACCGCATGGGCAACACCACCTTCTACGGTGAGGGCAAG
ACTGTCGATACCAGCTCCAAGTTCACCGTTGTCACCCAGTTCATCAAGGACTCCGCTGG
CGATCTTGCTGAGATCAAGCGCTTCTACGTCCAGAACGGAAAAGTCATTGAGAACTCTC
AGTCCAACGTTGATGGAGTTTCTGGCAACTCCATCACCCAGTCTTTCTGCAATGCTCAG
AAGACTGCTTTCGGCGATATCGATGACTTCAACAAGAAGGGTGGCCTGAAGCAAATGGG
CAAGGCCCTTGCCAAGCCCATGGTCCTCGTCATGTCCATCTGGGACGACCATGCCGCCA
ACATGCTCTGGCTCGACTCCACCTACCCTGTCGAGGGTGGCCCCGGTGCTTACCGTGGC
GAGTGCCCTACCACCTCGGGTGTCCCCGCTGAGGTCGAGGCCAATGCTCCCAACTCCAA
GGTCATCTTCTCCAACATCAAGTTCGGCCCCATCGGCTCTACCTTTAGCGGCGGCTCTT
CCGGCACCCCTCCTTCCAACCCTTCGAGCTCCGTCAAGCCCGTCACCTCCACTGCTAAG
CCTTCTTCCACCTCTACTGCCTCCAACCCCAGCGGTACCGGTGCTGCTCACTGGGCTCA
GTGCGGTGGTATTGGCTTCTCTGGCCCCACCACTTGCCAGAGCCCTTACACTTGCCAAA
AGATCAACGACTATTACTCCCAGTGCGTGTAA

>NcCel6A

ATGCGCGCCCTCTCCCTCCTCGCCGCCCTCTCGGCCGCCGGCGCCACCGTCACCGCTTC
CCCCGTTGGCCCTGCTCATGCCCCACGGGCCCTCTCCGACTCCAACCCCTTTGTCGGCA
AGAAGCTCTTCGCCAACCCCAAATGGGCCTCTAAGCTCGAAGACACATACAAGGCCTTT
ACTTCCAGGGGCGACTCGGAAAACGCAGCCGTAGTCCGCAAGGTCCAAAAGATCGGTTC
CTTCGTCTGGGTGTCCTCGCGCTCGGGGCTCTCCGAGATCGACGAAGCCATCCAAGCCG
CTCGTGCGGAGCAGAGCCGGACGGGACAGAAGCAGATTGTCGGGTTGGTGATGTACAAC
ATACCTGATCGTGACTGCTCTGCGGGCGAGTCGGCGGGCGAGCTGAGCACCAAGAACGA
TGGACTGAGAATCTACAAGGAGCAGTTTGTCAAGCCGTATGCGGCCAAGCTGGCGGCGG
CCAAGGACTTGCAGTTTGCTGTGGTGGTGGAGCCGGACTCGTTGGGCAATTCGGTCACC
AACTCTGCCGTCGAGTTTTGCAAGCAGGCCATTCCTACCTATCGGGAAGGTATCGCTTT
TGCCATCAAGAGCTTGCAGTTGGATAACGTGGCGGTGTATTTGGATGCTGCGAATGGTG
GATGGTTGGGCTGGGGCGATAGTTTGCCCAAGGCCGCCGACGAAATCGCCACCATCCTC
TCCATGGCCTCCCCCGCCAAAGTCCGCGGTTTCTCCACCAACGTCTCCAACTACAACCC
CTACCTCGCCACCAAGCGCGACTCCTTCACCTCGGGCTCCCCGTCCTACGACGAATCGC
ACTACGCCTCCTCCCTCGCGCCCTACCTTCAACAACACGGGCTCCCTGCGCACTTCATC

ATCGACCAGGGCCGTGTGGCTTATCCGGGCGCCAGGAAGGACTGGGGTGACTGGTGCAA
CGTCAACCCGGCTGGGTTTGGCCATATCCCCACGACGGACCAGAGCGTGCTGCAGAACT
CGAATGTCGATGCGATTGTGTGGATCAAGCCCGGTGGTGAGAGTGATGGACAGTGTGGA
TTGAAGGGGGCGCCGATTGCGGGAGCGTGGTTTGGTGGGTATGCGGAGATGTTGACGGG
GAATGCGGATGCGAGTGTCAAGAACAGTTAA

>StCel7A

ATGTACGCCAAGTTCGCGACCCTCGCCGCCCTTGTGGCTGGCGCCGCTGCTCAGAACGC
CTGCACTCTGACCGCTGAGAACCACCCCTCGCTGACGTGGTCCAAGTGCACGTCTGGCG
GCAGCTGCACCAGCGTCCAGGGTTCCATCACCATCGACGCCAACTGGCGGTGGACTCAC
CGGACCGATAGCGCCACCAACTGCTACGAGGGCAACAAGTGGGATACTTCGTACTGCAG
CGATGGTCCTTCTTGCGCCTCCAAGTGCTGCATCGACGGCGCTGACTACTCGAGCACCT
ATGGCATCACCACGAGCGGTAACTCCCTGAACCTCAAGTTCGTCACCAAGGGCCAGTAC
TCGACCAACATCGGCTCGCGTACCTACCTGATGGAGAGCGACACCAAGTACCAGATGTT
CCAGCTCCTCGGCAACGAGTTCACCTTCGATGTCGACGTCTCCAACCTCGGCTGCGGCC
TCAATGGCGCCCTCTACTTCGTGTCCATGGATGCCGATGGTGGCATGTCCAAGTACTCG
GGCAACAAGGCAGGTGCCAAGTACGGTACCGGCTACTGTGATTCTCAGTGCCCCCGCGA
CCTCAAGTTCATCAACGGCGAGGCCAACGTAGAGAACTGGCAGAGCTCGACCAACGATG
CCAACGCCGGCACGGGCAAGTACGGCAGCTGCTGCTCCGAGATGGACGTCTGGGAGGCC
AACAACATGGCCGCCGCCTTCACTCCCCACCCTTGCACCGTGATCGGCCAGTCGCGCTG
CGAGGGCGACTCGTGCGGCGGTACCTACAGCACCGACCGCTATGCCGGCATCTGCGACC
CCGACGGATGCGACTTCAACTCGTACCGCCAGGGCAACAAGACCTTCTACGGCAAGGGC
ATGACGGTCGACACGACCAAGAAGATCACGGTCGTCACCCAGTTCCTCAAGAACTCGGC
CGGCGAGCTCTCCGAGATCAAGCGGTTCTACGTCCAGAACGGCAAGGTCATCCCCAACT
CCGAGTCCACCATCCCGGGCGTCGAGGGCAACTCCATCACCCAGGACTGGTGCGACCGC
CAGAAGGCCGCCTTCGGCGACGTGACCGACTTCCAGGACAAGGGCGGCATGGTCCAGAT
GGGCAAGGCCCTCGCGGGGCCCATGGTCCTCGTCATGTCCATCTGGGACGACCACGCCG
TCAACATGCTCTGGCTCGACTCCACCTGGCCCATCGACGGCGCCGGCAAGCCGGGCGCC
GAGCGCGGTGCCTGCCCCACCACCTCGGGCGTCCCCGCTGAGGTCGAGGCCGAGGCCCC
CAACTCCAACGTCATCTTCTCCAACATCCGCTTCGGCCCCATCGGCTCCACCGTCTCCG
GCCTGCCCGACGGCGGCAGCGGCAACCCCAACCCGCCCGTCAGCTCGTCCACCCCGGTC
CCCTCCTCGTCCACCACATCCTCCGGTTCCTCCGGCCCGACTGGCGGCACGGGTGTCGC

TAAGCACTATGAGCAATGCGGAGGAATCGGGTTCACTGGCCCTACCCAGTGCGAGAGCC
CCTACACTTGCACCAAGCTGAATGACTGGTACTCGCAGTGCCTGTAA

>TrCel7A

ATGTATCGGAAGTTGGCCGTCATCTCGGCCTTCTTGGCCACAGCTCGTGCTCAGTCGGC
CTGCACTCTCCAATCGGAGACTCACCCGCCTCTGACATGGCAGAAATGCTCGTCTGGTG
GCACGTGCACTCAACAGACAGGCTCCGTGGTCATCGACGCCAACTGGCGCTGGACTCAC
GCTACGAACAGCAGCACGAACTGCTACGATGGCAACACTTGGAGCTCGACCCTATGTCC
TGACAACGAGACCTGCGCGAAGAACTGCTGTCTGGACGGTGCCGCCTACGCGTCCACGT
ACGGAGTTACCACGAGCGGTAACAGCCTCTCCATTGGCTTTGTCACCCAGTCTGCGCAG
AAGAACGTTGGCGCTCGCCTTTACCTTATGGCGAGCGACACGACCTACCAGGAATTCAC
CCTGCTTGGCAACGAGTTCTCTTTCGATGTTGATGTTTCGCAGCTGCCGTGCGGCTTGA
ACGGAGCTCTCTACTTCGTGTCCATGGACGCGGATGGTGGCGTGAGCAAGTATCCCACC
AACACCGCTGGCGCCAAGTACGGCACGGGGTACTGTGACAGCCAGTGTCCCCGCGATCT
GAAGTTCATCAATGGCCAGGCCAACGTTGAGGGCTGGGAGCCGTCATCCAACAACGCGA
ACACGGGCATTGGAGGACACGGAAGCTGCTGCTCTGAGATGGATATCTGGGAGGCCAAC
TCCATCTCCGAGGCTCTTACCCCCCACCCTTGCACGACTGTCGGCCAGGAGATCTGCGA
GGGTGATGGGTGCGGCGGAACTTACTCCGATAACAGATATGGCGGCACTTGCGATCCCG
ATGGCTGCGACTGGAACCCATACCGCCTGGGCAACACCAGCTTCTACGGCCCTGGCTCA
AGCTTTACCCTCGATACCACCAAGAAATTGACCGTTGTCACCCAGTTCGAGACGTCGGG
TGCCATCAACCGATACTATGTCCAGAATGGCGTCACTTTCCAGCAGCCCAACGCCGAGC
TTGGTAGTTACTCTGGCAACGAGCTCAACGATGATTACTGCACAGCTGAGGAGGCAGAA
TTCGGCGGATCCTCTTTCTCAGACAAGGGCGGCCTGACTCAGTTCAAGAAGGCTACCTC
TGGCGGCATGGTTCTGGTCATGAGTCTGTGGGATGATTACTACGCCAACATGCTGTGGC
TGGACTCCACCTACCCGACAAACGAGACCTCCTCCACACCCGGTGCCGTGCGCGGAAGC
TGCTCCACCAGCTCCGGTGTCCCTGCTCAGGTCGAATCTCAGTCTCCCAACGCCAAGGT
CACCTTCTCCAACATCAAGTTCGGACCCATTGGCAGCACCGGCAACCCTAGCGGCGGCA
ACCCTCCCGGCGGAAACCCGCCTGGCACCACCACCACCCGCCGCCCAGCCACTACCACT
GGAAGCTCTCCCGGACCTACCCAGTCTCACTACGGCCAGTGCGGCGGTATTGGCTACAG
CGGCCCCACGGTCTGCGCCAGCGGCACAACTTGCCAGGTCCTGAACCCTTACTACTCTC
AGTGCCTGTAA

>TrCel6A

ATGATTGTCGGCATTCTCACCACGCTGGCTACGCTGGCCACACTCGCAGCTAGTGTGCC
TCTAGAGGAGCGGCAAGCTTGCTCAAGCGTCTGGGGCCAATGTGGTGGCCAGAATTGGT
CGGGTCCGACTTGCTGTGCTTCCGGAAGCACATGCGTCTACTCCAACGACTATTACTCC
CAGTGTCTTCCCGGCGCTGCAAGCTCAAGCTCGTCCACGCGCGCCGCGTCGACGACTTC
TCGAGTATCCCCCACAACATCCCGGTCGAGCTCCGCGACGCCTCCACCTGGTTCTACTA
CTACCAGAGTACCTCCAGTCGGATCGGGAACCGCTACGTATTCAGGCAACCCTTTTGTT
GGGGTCACTCCTTGGGCCAATGCATATTACGCCTCTGAAGTTAGCAGCCTCGCTATTCC
TAGCTTGACTGGAGCCATGGCCACTGCTGCAGCAGCTGTCGCAAAGGTTCCCTCTTTTA
TGTGGCTGGATACTCTTGACAAGACCCCTCTCATGGAGCAAACCTTGGCCGACATCCGC
ACCGCCAACAAGAATGGCGGTAACTATGCCGGACAGTTTGTGGTGTATGACTTGCCGGA
TCGCGATTGCGCTGCCCTTGCCTCGAATGGCGAATACTCTATTGCCGATGGTGGCGTCG
CCAAATATAAGAACTATATCGACACCATTCGTCAAATTGTCGTGGAATATTCCGATATC
CGGACCCTCCTGGTTATTGAGCCTGACTCTCTTGCCAACCTGGTGACCAACCTCGGTAC
TCCAAAGTGTGCCAATGCTCAGTCAGCCTACCTTGAGTGCATCAACTACGCCGTCACAC
AGCTGAACCTTCCAAATGTTGCGATGTATTTGGACGCTGGCCATGCAGGATGGCTTGGC
TGGCCGGCAAACCAAGACCCGGCCGCTCAGCTATTTGCAAATGTTTACAAGAATGCATC
GTCTCCGAGAGCTCTTCGCGGATTGGCAACCAATGTCGCCAACTACAACGGGTGGAACA
TTACCAGCCCCCCATCGTACACGCAAGGCAACGCTGTCTACAACGAGAAGCTGTACATC
CACGCTATTGGACCTCTTCTTGCCAATCACGGCTGGTCCAACGCCTTCTTCATCACTGA
TCAAGGTCGATCGGGAAAGCAGCCTACCGGACAGCAACAGTGGGGAGACTGGTGCAATG
TGATCGGCACCGGATTTGGTATTCGCCCATCCGCAAACACTGGGGACTCGTTGCTGGAT
TCGTTTGTCTGGGTCAAGCCAGGCGGCGAGTGTGACGGCACCAGCGACAGCAGTGCGCC
ACGATTTGACTCCCACTGTGCGCTCCCAGATGCCTTGCAACCGGCGCCTCAAGCTGGTG
CTTGGTTCCAAGCCTACTTTGTGCAGCTTCTCACAAACGCAAACCCATCGTTCCTGTAA

BGL nucleotide sequence

>AnBgl1

ATGAGGTTCACTTTGATCGAGGCGGTGGCTCTGACTGCCGTCTCGCTGGCCAGCGCTGA
TGAATTGGCCTACTCCCCGCCGTATTACCCCTCCCCTTGGGCCAATGGCCAGGGTGACT

GGGCGGAAGCATACCAGCGCGCTGTTGATATCGTCTCGCAGATGACATTGGCTGAGAAG
GTCAATTTGACTACGGGAACTGGATGGGAATTGGAATTATGTGTTGGTCAGACTGGAGG
TGTTCCCCGATTGGGAATTCCGGGAATGTGTGCACAGGATAGCCCTCTGGGTGTTCGTG
ACTCCGACTACAACTCTGCGTTCCCTGCCGGTGTCAACGTGGCCGCAACCTGGGACAAG
AATCTGGCTTACCTTCGTGGCCAGGCTATGGGTCAGGAGTTTAGTGACAAGGGTGCTGA
TATCCAATTGGGTCCAGCTGCCGGCCCTCTCGGTAGAAGTCCCGACGGCGGTCGTAACT
GGGAGGGCTTCTCCCCCGACCCGGCCCTCAGTGGTGTGCTCTTTGCAGAGACAATCAAG
GGTATTCAGGATGCTGGTGTGGTTGCAACGGCTAAGCACTACATCGCCTACGAGCAGGA
GCATTTCCGTCAGGCGCCTGAAGCTCAAGGCTACGGATTCAATATTACCGAGAGTGGAA
GCGCGAACCTCGACGATAAGACTATGCATGAGCTGTACCTCTGGCCCTTCGCGGATGCC
ATCCGTGCAGGTGCCGGTGCTGTGATGTGCTCGTACAACCAGATCAACAACAGCTATGG
CTGCCAAAACAGCTACACTCTGAACAAGCTGCTCAAGGCTGAGCTGGGTTTCCAGGGCT
TTGTCATGAGTGATTGGGCGGCTCACCATGCCGGTGTGAGTGGTGCTTTGGCGGGATTG
GACATGTCTATGCCGGGAGACGTCGATTACGACAGTGGCACGTCTTACTGGGGTACCAA
CTTGACCATCAGTGTGCTCAACGGGACGGTGCCCCAATGGCGTGTTGATGACATGGCTG
TCCGCATCATGGCCGCCTACTACAAGGTCGGCCGTGACCGTCTGTGGACTCCTCCCAAC
TTCAGCTCATGGACCAGAGATGAATACGGCTTCAAGTACTACTATGTCTCGGAGGGACC
GTATGAGAAGGTCAACCAGTTCGTGAACGTGCAACGCAACCATAGCGAGTTGATCCGCC
GTATTGGAGCAGACAGCACGGTGCTCCTCAAGAACGATGGCGCTCTTCCCTTGACTGGA
AAGGAGCGCTTGGTCGCCCTTATCGGAGAAGATGCGGGTTCCAATCCTTATGGTGCCAA
CGGCTGCAGTGACCGTGGGTGCGACAATGGAACATTGGCGATGGGCTGGGGAAGTGGCA
CTGCCAACTTTCCCTACTTGGTGACCCCCGAGCAGGCCATCTCGAACGAGGTGCTCAAG
AACAAGAATGGCGTATTCACTGCGACCGATAACTGGGCTATTGATCAGATTGAGGCGCT
TGCTAAGACCGCCAGTGTCTCTCTTGTCTTTGTCAACGCCGACTCTGGTGAGGGTTATA
TCAATGTCGACGGAAACCTGGGTGACCGCAGGAACCTGACCCTGTGGAGGAACGGCGAC
AATGTGATCAAGGCTGCTGCTAGCAACTGCAACAACACGATCGTTATTATTCACTCTGT
CGGCCCAGTCTTGGTTAACGAGTGGTACGACAACCCCAATGTTACCGCTATTCTCTGGG
GTGGTCTTCCCGGTCAGGAGTCTGGCAACTCCCTCGCCGACGTGCTCTACGGCCGTGTC
AACCCCGGTGCCAAGTCGCCCTTCACCTGGGGCAAGACTCGTGAGGCCTACCAAGATTA
CTTGTACACCGAGCCCAACAACGGCAACGGAGCGCCCCAGGAAGACTTCGTCGAGGGCG
TCTTCATTGACTACCGCGGATTTGACAAGCGCAACGAGACTCCTATCTATGAGTTCGGC
TATGGTCTGAGCTACACCACCTTCAACTACTCGAACCTTCAGGTGGAGGTTCTGAGCGC

CCCTGCGTACGAGCCTGCTTCGGGCGAGACTGAGGCAGCGCCGACTTTCGGAGAGGTCG
GAAATGCGTCGGATTACCTCTACCCCGATGGACTGCAGAGAATCACCAAGTTCATCTAC
CCCTGGCTCAACAGTACCGATCTTGAGGCGTCTTCTGGGGATGCTAGCTATGGGCAGGA
TGCCTCAGACTATCTTCCCGAGGGAGCCACCGATGGCTCTGCGCAACCGATCCTGCCTG
CCGGTGGTGGTGCTGGCGGCAACCCTCGCCTGTACGACGAGCTCATCCGCGTGTCGGTG
ACTATCAAGAACACCGGCAAGGTTGCGGGTGATGAAGTTCCTCAACTGTATGTTTCTCT
TGGCGGCCCTAACGAACCCAAGATCGTGCTGCGTCAATTCGAGCGTATCACGCTGCAGC
CGTCGGAAGAGACGCAGTGGAGCACAACTCTGACGCGCCGTGACCTTGCGAACTGGAAT
GTTGAGACGCAGGACTGGGAGATTACGTCGTATCCCAAGATGGTGTTTGTCGGAAGCTC
CTCGCGGAAGCTGCCGCTCCGGGCGTCTCTGCCTACTGTTCACTAA

Amino Acid Sequences

Sequences in **bold** indicate the conserved domain of the designated protein.

Endoglucanase amino acid sequences

>AfCel7B1

MDSKRGVVAAVLALLPLVSAQQPAASSAGN**PKLTTYKCTTAGGCVAQDTSVVLDWGYHW**
IHTVDGYTSCTTSSGVDSTLCPDAATCAKNCVIEPANYTSAGVTTSGDSLTMYQYVQSN
GVYTNASPRLYLLGPDKNYVMLKLLGQELTFDVDLSTLPCGENGALYLSEMSATGGRNE
YNTGGAEYGSGYCDAQCPVIAWKNGTLNTSGASYCCNEMDILEANSRANSYTPHPCSAT
DCDKGGCGFNPYALGQKSYWGPGGTVDTSKPFTITTQFITNDGTTTGTLSEIRRQYMQN
GKVIANAVSSTGVNSITEDWCTSVDGSAATFGGLTTMGKALGRGMVLIFSIWNDASGFM
NWLDSGNAGPCSSTEGNPDLIKAQNPTTHVVFSNIRWGDIGSTFKGSDGSVTTTTSTTS
TKTTTSTAPGPTQTHYGQCGGQGWTGPTACASPYTCQVLNPWYSQCL

>AfCel7B2

MAQTLAAASLVLVPLVTAQ**QIGSIAENHPELKTYRCGSQAGCVAQSTSVVLDINAHWIH**
QMGAQTSCTTSSGLDPSLCPDKVTCSQNCVVEGITDYSSFGVQNSGDAITLRQYQVQNG
QIKTLRPRVYLLAEDGINYSKLQLLNQEFTFDVDASKLPCGMNGALYLSEMDASGGRSA

LNPAGATYGTGYCDAQCFNPGPWINGEANTLGAGACCQEMDLWEANSRSTIFSPHPCTT
AGLYACTGAECYSICDGYGCTYNPYELGAKDYYGYGLTVDTAKPITVVTQFVTADNTAT
GTLAEIRRLYVQEGMVIGNSAVAMTEAFCSSSRTFEALGGLQRMGEALGRGMVPVFSIW
DDPSLWMHWLDSDGAGPCGSTEGDPAFIQANYPNTAVTFSKVRWGDIDSTYSV

>AfCel5A1

MKASTIICALLPLAVAAPNAKRASGFVWFGSNESGAEFG**ETKLPGVLGTDYIWPDASTI**
KTLHDAGMNIFRVAFRMERLIPNQMTGTPDATYLNDLKATVNAITSLGAYAVIDPHNYG
RYYGNIISSTDDFAAFWKTLAAQFASNDHVIFDTNNEYHDMDQTLVLNLNQAAINAIRA
AGATSQYIFVEGNSWSGAWTWTTVNDNLKALTDPQDKIVYEMHQYLDSDGSGTSATCVS
STIGQERVQAATQWLKTNGKKGIIGEFAGGANSVCQSAVTGMLDYLSANSDVWMGAAWW
AAGPWWADYIFNMEPPSGTAYQNYLSLLKPYFVGGSGGNPPTTTTTTTQPTTTTTTTTA
GNPGGTGLAQHWGQCGGIGWTGPTACASPYTCQKLNDYYSQCL

>AfCel5A2

MRISSLIMAASAAGLVSALPVREMTKKRSSGFTWVGI**NESGAEFGSNIPGKLGTDYTWP**
DTSKIQTLRDAGMNIFRVPFLMERLVPSSITGNLDATYLSDLKKTVEFITGSGAYAVLD
PHNYGRYSGSIISSTSDFQEFWKTVASEFASNDKVIFDTNNEYHDMDQTLVLNLNQAAI
NGIRAAGATTQYIFVEGNHYSGAWTWADNNDNLKGLKDPEDKIVFEMHQYLDSDGSGTS
ESCVSSTIGQERVESATQWLKDNSLKGFLGEFAGGVNSQCETAVEGLLSYMSENSDVWL
GAEWWSAGPMWGNYMYSLEPSTGPAYSSYLPILKNYFVSGTSSSSSSSSITSSTTSVN
NPTTSATTQAQVVNTSSSSPTPSPTSVPAAEAPAPTPTEAAKTSSAASTTPTTAAASTT
ASACRAPEPADETTLPSATPTTSASASGIAKHWYQCGGINWTGPTTCESGYTCVEQNPY
YHQCV

>AfCel5A3

MKFGSIVLIAAAAGFAVAAPAKRA**SGKEFIFPDPSTISTLIGKGMNIFRIQFLMERLVP**
SSMTGSYNEEYLANLTSVVDAVTKAGSYAILDPHNFGRYNGQIISSTDDFKTFWQNLAG
KFKSNNLVIFDTNNEYHDMDQTLVLNLNQAAINGIRAAGATSQYIFVEGNSWTGAWTWA
DVNDNLKALTDPHDKIVYEMHQYLDSDGSGTAESCVSTTIGKERVSAATKWLKDNGKVG
IIGEFAGGVNDQCRTAISGMLEYLAQNTDVWKGALWWAAGPWWGNYMFNMEPPSGAAYV

GMLDILEPYLG

>AnCel7B

MALLLSLSLLATTISAQQIG**TPEIRPRLTTYHCTSANGCTEQNTSVVLDAATHPIHDAS**
NPSVSCTTSNGLNPALCPDKQTCADNCVIDGITDYAAHGVETHGSRLTLTQYRNVNGAL
SSVSPRVYLVDESDPDEQEYRALSLLAQEFTFTVNVSALPCGMNGALYLSEMSPSGGRS
ALNPAGASYGTGYCDAQCYVNPWINGEGNINGYGACCNEMDIWEANSRSTGFTPHACLY
EPEETEGRGVYECASEDECDSAGENDGICDKWGCGFNPYALGNTEYYGRGQGFEVDTKE
PFTVVTQFLTDDGTSTGALTEIRRLYIQNGQVIENAVVSSGADSLTDSLCASTASWFDS
YGGMEGMGRALGRGMVLAMSIWNDAGGYMQWLDGGDAGPCNATEGAPEFIEEHTPWTRV
VFEDLKWGDIGSTFQAS

>AnCel5A1

MRSLVLLSSVLALVAPSKGAFTWLGTNEAGAEFGEGSYPGEL**GTEYIWPDLGTIGTLRN**
EGMNIFRVAFSMERLVPDSLAGPVADEYFQDLVETVNGITALGAYAVLDPHNYGRYYGN
IITSTDDFAAFWTILATEFASNELVIFDTNNEYHTMDQSLVLNLNQAAIDAIRASGATS
QYIFAEGNSWTGAWTWVDVNDNMKALTDPQDKLIYEMHQYLDSDGSGTNTACVSSTIGS
ERVTAATNWLRENGKLGVLGEFAGANNQVCKDAVADLLEYLEENSDVWLGALWWAAGPW
WGDYMFNMEPTSGIAYQEYSEILQPYFVGSQ

>AnCel5A2

MKVNTLLVAVAAGTAMAAPQLKKRAGFTFFG**VTEAGAEFGEKSIPGVWGTDYTFPDTES**
ILTLISKGFNTFRIPFLMERLTPEMTGSFDEGYLKNLTSVVNAVTDAGAWAIVDAQNFG
RFNGEIISSASDFQTWWKNVAAEFADNKNVIFDTTNTSGADNEFHDMDQTLVLDLNQAA
INGIRAAGATSQYIFVEGNSYTGAWTWTDNNDNLKSLTDPQDKIVYEMHQYLDTDGSGT
HETCVSETIGAERVESATQWLKDNGKLGVIGEFAGGNNEICRAAVKSLLDALKENDDVW
LGALWWAAGPWWEDYMFSMEPTDGIAYTGMLSTLEAYMN

>ApCel5A

MKYSTFVVAAAAGSAAASSSAYGQCGGNGWTGATDCVSGYHCAYQNDWYSQCVPGAGSG

STSAAAAAATTKVVTSAKASTTQAPSKAPVSTSAASTSKAVAATSGKVKYAG**VNIAGFD**
FGMDTNGASSGSYVDPGTTGQNQMNHFVKDDKLNAFRLPVGWQYLVNSQLGGTLDSTFF
AKYDQQMTYCLNSGAALCILDLHNYARWNGQIVGTSGGPTNAQFASVWSQLAKKYSSKP
KVAFAIMNEPHDLQDINAWATTVQAAVTAIRQAGATQNMILLPGDDWTHAANFVDNGSA
AALNKITNLDGSKTNLVFDVHQYSDSDGSGTSSTCVSSASNIAGFTKLATWLRSNNRQA
MLTEAGGSNDQSCLTAVCDVLNYLMTNSDVYLGWTGWSAGMFATDYVLSEVPTGSAGSY
KDQAIVTKCIAGVFNSK

>FgCel7B1

MYRAIATASALIAAVRAQ**QVCSLTQESKPSLNWSKCTSSGCSNVKGSVTIDANWRWTHQ**
VSGSTNCYTGNKWDTSVCTSGKVCAEKCCLDGADYASTYGITSSGDQLSLSFVTKGPYS
TNIGSRTYLMEDENTYQMFQLLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGKAKYPG
NKAGAKYGTGYCDAQCPRDVKFINGQANSDGWQPSDSDVNGGIGNLGTCCPEMDIWEAN
SISTAYTPHPCTKLTQHSCTGDSCGGTYSNDRYGGTCDADGCDFNSYRQGNKTFYGPGS
GFNVDTTKKVTVVTQFHKGSNGRLSEITRLYVQNGKVIANSESKIAGVPGNSLTADFCT
KQKKVFNDPDDFTKKGAWSGMSDALEAPMVLVMSLWHDHHSNMLWLDSTYPTDSTKLGS
QRGSCSTSSGVPADLEKNVPNSKVAFSNIKFGPIGSTYKSDGTTPTNPTNPSEPSNTAN
PNPGTVDQWGQCGGSNYSGPTACKSGFTCKKINDFYSQCQ

>FgCel7B2

MKFSPLLLSTLLANTVVAQTPDK**TQEKHPKIETYRCTKAKGCKKATNYIVADAELHGIS**
QANGQSCGNWGEAANSTACPDEATCAKNCKLFGMNEAAYKAKGISTSGNALRLEMLRNG
QSVSPRVYLLEENKNKYEMLKLTGAEFSFDVETQKLPCGMNGALYLSEMPADGGKSTSK
YSKVGAAQGGGYCDAQCYVTPFINGVGNIKGKGVCCNEMDIWEANSRATHIAPHPCSVP
GLYGCTGAECQKDGICDKAGCGWNHNRNGVPDFYGRGKNFKVDTTRKFTVVSSFPADKS
GKLTEMHRHYIQDGKVIKSAVVTLPGPPKVTGNIITDNYCKASHADDYIRLGGTEEMGD
AMTRGMVLAMSVWWSEGDSMEWLDGQGAGAGPCTKEEGLPKNIVKVEPNPEVTFSNIRI
GEIGSTHAVKMPRVYGAHRL

>FgCel5A1

MRFTDLLLASAGATLALAAPSTEKRAAGKFLFTGSNESGGEFG**ETQLPGKLGKDYIWPT**

TKSIDTLASTGMNTFRVGFRMERMTPSGITGALDETYFKGLESVVNHITSKHRGNFAVI
DPHNYGRYNNQIIQSTADFGAWWSKVAKRFANNKNVIFDTNNEYHDMENSLVAGLNQAA
IDAIRKAGATSQYIFVEGNSYTGAHSWVSSGNGEALKNLKDPQNKIIYQMHQYLDSDNS
GTHADCVSSTIGVERVKEATKWLKDNKKRGIIGETAAGPNTQCIEALKGELQYLHDNSD
VWTGWLYWAAGPWWGDYMYSMEPDTGASYVKVLPEIKKFIGA

>FgCel5A2

MKSLLALSLFAGLSVAQSSAWAQCGGEGFSGSTSCVSGYKCTVVNQWYSQCQPGTAEPP
STTLKTTTGGGSTPTGTPGDGKFLWAG**VNEAGGEFGEKNLPGTWGKDFIFPDPAAVDTL**
ISQGYNTFRVQLKMERANPSGLTGAYDQAYIKNLTSIVNHITGKGATVLLDPHNYGRFF
DKIITSTSDFQTWWKNFATLFKSNSRIMFDTNNEYHTMDQTLVLNLNQAAINGIRAAGA
TQYIFVEGNQWSGAWSWPDVNDNMKALTDPENKLIYEMHQYLDSDSSGTSPDCVSTTIG
VERLQAATKWLRANKKIGMIGEFAGGPNETCKTAVKNMLDFMKANTDVWKGFTWWSAGP
WWGDYMYSFEPPSGSAYQYYNSLLKTYV

>GtCel12A

MFRFISALPFALGALKLASAATVLTGQYTCDTAGD**YTLCNDQWGIANGVGSQTSTLISA**
SGSTISWSTNYTWANNPNDVKTYANVLSNTAKGVQISQISSFPTTWDWYYETQSSGLRA
DVSYDMWTGTAQTGTAASSSSSYEIMIWLSGKGGIQPVGSLKTSGISLAGYTWNLWSGT
TETWTTLSFVSADGDITSFNAELTPFFQYLVENEGVSASQYIQAIQTGTEAFTGTAELV
TTSFSVSLSG

>NcCel7B1

MVHKLAFLTGLTASLVSAQ**QIGTITPESHPKLPTKRCTLSGGCQTVSTSIVLDAFQRPL**
HKIGDPSTACVVGGPLCPDAATCAANCALEGVDYASLGVKTEGDALTLNQWVSDPSNPG
HYKTSSPRTYLVAEDGKNYEAVKLLGKEISFDVDVSNLPCGMNGAFYLSEMLMDGGRGE
FNAAGAEYGTGYCDAQCPKLDFINGEANINKTHGACCNEMDIFEANARAKSFTPHPCSI
ERVYKCTGDTECGQSQGVCDQWGCTYNEYQKGVHDFYGLAPPAKTIDTTQKFTVTTQFL
TDNGREDGVLVEIRRLWSQNGKLIKNAKIAVDSLSTDSVSTQFCEKTSSWTMQRGGLKT
MGEAMGRGMVLIFSIWADESGFMNWLDSGDSGPCSATEGDPKLILQKKPDARVTFSNIK
WGEMGSTYASAGKYGVRRVAKGLSA

>NcCel7B2

MSRRILLSALLVAAVSAQ**QPGKLTPEVHPKLPTWACTVADGCIQKDTSLVLDSDYRWVH**
TDDYTNCKVNGLNPAVCPDVETCAANCNLEGVDYTGSGIHTNGSELTLNLFVNRTDGTT
SLVSPRVYLLANETTYDMFSLLDKEFTFDVDVSKLPCGTNGALYFSEMLANGGKSALNP
AGASYGTGYCDAQCPTPAFINGEANLESYGACCNEMDIWEANSRATAFTPHPCNVTALY
KCSGDLCGRTDKYQSVCDKDGCDYNPYRLGDHPYYGRGEGNKIDTTRPFTVVTQFFSNT
TSAGEKELSAIKRLYLQDGKLISTSTIAVPGFDSTSDTITDDYCAKNKQIFGGVNAFAN
QGGLRQMGEALDRGMVLVFSVWHDAGSAMKWLDGTFPPGADPETQPGTERGPCLPGEGH
ADDIQRDASWTEVKFSNVKSGEIGSTFEA

>NcCel5A

MKATILASTFAAGALAQSGAWGQCGGNGWSGATSCISGYACNYVNDWYSQCQPGTAAPT
TTAAATTLVTSTKTAPPASTTTATASGKFKWFGVNEAGGEFGDGIFPGR**WGTEFTFPDT**
NTIQTLRSQGYNIFRVGFAMERLVPNTLTSSFDNGYLTNLTQVVNSVTNSGAYIVLDPH
NYGRYYGKIITDTDAFKTFWQNVAAKFASNSKVIFDTNNEYNTMDQTLVLNLNQAAIDG
IRAAGATSQYIFVEGNQWTGAWSWNVTNTNLAALTDPENKIVYEMHQYLDSDSSGTSTA
CVSSEIGVQRIVGATAWLRANGKKGVLGEFAGGANSVCKAAVTGLLEHLKANTDVWEGA
LWWAAGPWWGDYMYSFEPPSGTGYTYYNSLLKTYTP

>NhCel12A

MKSAIVAALAGLAAASPTRLIPRGQFCGQWDSETAGAYTIYNNLWGKDNAESGEQCTTN
SGEQSDGSIAWSVEWSWTGGQGQVKSYPNAVVEIE**KKTLGEVSSIPSAWDWTYTGDGII**
ANVAYDLFTSSTESGDAEYEFMIWLSALGGAGPISNDGSPVATVELAGTSWKLYQGKNN
QMTVFSFVAESDVNNFCGDLADFTDYLVDNHGVSSSQILQSVGAGTEPFEGTNAVFTTN
NYHADVEY

>PsCel12A1

MKVAFATAMAAAALAAAYADDFCDQWGTTTTDNYIIYNNLWGESYATSGSQCTGLDSSS
GSTVAWHTNWTWTGASSNVKSFANAALQFDA**VQLSSVSSIPTTMEYSLEYSGNIAADVS**

YDLFTASTSTGDNEFEIMIWLAALGGAGPISSTGSAVATTTIADTSFSLYTGANGDTTV
YSFVASDTVKSFSGDLMDFFTYLIDNEGFSSSQYLNTVQAGTELFTGTDVTLTVSSYSA
AVNMGASSGTTATTSTASSSTGTSVSTKELIASSSSTESDSQTAALTATTATPSTDSSY
SQQETTAPSTSSTSVSGEATTSSTSSTWNEQESTAASASETTAPSTTSSTSEETTSSSA
TETTTTSVTAAPSTSTTTGKKCASRSVRRV

>PsCel12A2
MKSSAVLAASALAIAASPAFAQEEFCGQWNLTKTDDYILYNNLWGAFDDPTGTQCTGLD
SVKGSTVAWHTSFGWSGSKTQVKSFANVALQFDH**LPISEVTSIPSTIHFKYDYEESLIA**
NVAFDLFTSSTPDGDAEYEIMVWLAAIGGAGPISSTGSAVDQVTVGGVDFSLYAGKNGN
MTVYSFVASTMVNRYATDFKQFFDVLPRNLTIDPSQYLINVQAGTEPFVGNGTLTVSKY
SAAVNPAEYSQVQQQ

>PsCel12A3
MKSFLQLVVVVAALLSVSTADFCSQWRLSKAGKYVIYNNLWNKNAAASGSQCTGVDKIS
GSTIAWHTSYTWTGGAATEVKSYSNAALVFS**KKQIKNIKSIPTKMKYSYSHSSGTFVAD**
VSYDLFTSSTASGSNEYEIMIWLAAYGGAGPISSTGKAIATVTIGSNSFKLPSTQYLTT
LEAGTEPFTGSNAKMTVSSFSAAVN

>StCel5A
MKSSILASVFATGAVAQSGPWQQCGGIGWQGSTDCVSGYHCVYQNDWYSQCVPGAASTT
LQTSTTSRPTATSTAPPSSTTSPSKGKLKWLGS**NESGAEFGEGNYPGLWGKHFIFPSTS**
AIQTLINDGYNIFRIDFSMERLVPNQLTSSFDQGYLRNLTEVVNFVTNAGKYAVLDPHN
YGRYYGNIITDTNAFRTFWTNLAKQFASNSLVIFDTNNEYNTMDQTLVLNLNQAAIDGI
RAAGATSQYIFVEGNAWSGAWSWNTTNTNMAALTDPQNKIVYEMHQYLDSDSSGTHAEC
VSSTIGAQRVVGATQWLRANGKLGVLGEFAGGANAVCQQAVTGLLDHLQDNSDVWLGAL
WWAAGPWWGDYMYSFEPPSGTGYVNYNSILKKYLP

>TrCel7B
MAPSVTLPLTTAILAIARLVAAQ**QPGTSTPEVHPKLTTYKCTKSGGCVAQDTSVVLDWN**

YRWMHDANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAASGVTTSGSSLTMNQYMP
SSSGGYSSVSPRLYLLDSDGEYVMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGG
ANQYNTAGANYGSGYCDAQCPVQTWRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSC
TATACDSAGCGFNPYGSGYKSYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKY
QQNGVDIPSAQPGGDTISSCPSASAYGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDS
GNAGPCSSTEGNPSNILANNPNTHVVFSNIRWGDIGSTTNSTAPPPPPASSTTFSTTRR
SSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL

>TrCel5A

MNKSVAPLLLAASILYGGAAAQQTVWGQCGGIGWSGPTNCAPGSACSTLNPYYAQCIPG
ATTITTSTRPPSGPTTTTRATSTSSSTPPTSSGVRFAGVNIAGFDFGCTTDGTCVTSKV
YPPLKNFT**GSNNYPDGIGQMQHFVNDDGMTIFRLPVGWQYLVNNNLGGNLDSTSISKYD**
QLVQGCLSLGAYCIVDIHNYARWNGGIIGQGGPTNAQFTSLWSQLASKYASQSRVWFGI
MNEPHDVNINTWAATVQEVVTAIRNAGATSQFISLPGNDWQSAGAFISDGSAAALSQVT
NPDGSTTNLIFDVHKYLDSDNSGTHAECTTNNIDGAFSPLATWLRQNNRQAILTETGGG
NVQSCIQDMCQQIQYLNQNSDVYLGYVGWGAGSFDSTYVLTETPTGSGNSWTDTSLVSS
CLARK

>TrCel12A

MKFLQVLPALIPAALAQTSCDQWATFTGNGYTVSNNLWGASAGSGFGCVTAVSLSGGAS
WHADWQWSGGQNNVKSYQNSQ**IAIPQKRTVNSISSMPTTASWSYSGSNIRANVAYDLFT**
AANPNHVTYSGDYELMIWLGKYGDIGPIGSSQGTVNVGGQSWTLYYGYNGAMQVYSFVA
QTNTTNYSGDVKNFFNYLRDNKGYNAAGQYVLSYQFGTEPFTGSGTLNVASWTASIN

Cellobiohydrolase amino acid sequences

>AnCel6A

MHSTNMRAIWPLVSLFSAVKALPAASATASASVAASSSPAPTASATGNPFEGYQL
YANPYYKSQVESSAIPSLS**ASSLVAQASAAADVPSFYWLDTADKVPTMGEYLEDI**

QTQNAAGASPPIAGIFVVYDLPDRDCSALASNGEYSISDGGVEKYKAYIDSIREQ
VETYSDVQTILIIEPDSLANLVTNLDVAKCANAESAYLECTNYALEQLNLPNVAM
YLDAGHAGWLGWPANIGPAAQLYASVYKNASSPAAVRGLATNVANFNAWSIDSCP
SYTSGNDVCDEKSYINAIAPELSSAGFDAHFITDTGRNGKQPTGQSAWGDWCNVK
DTGFGAQPTTDTGDELADAFVWVKPGGESDGTSDTSSSRYDAHCGYSDALQPAPE
AGTWFQAYFEQLLTNANPSL

>LeCel6A

MKITSTGLLALSSLLPFALGQSQLYGQCGGIGWSGATTCVSGATCTVVNAYYSQCLPGS
ASAPPTSTSSIGTGTTTSSAPGSTGTTTPAAGNPFTGYEIYLSPYYANEIAAAVTQISD
PTTAAAAA**KVANIPTFIWLDQVAKVPDLGTYLADASAKQKSEGKNYLVQIVVYDLPDRD**
CAALASNGEFTIADNGEANYHDYIDQIVAQIKQYPDVHVVAVIEPDSLANLVTNLSVAK
CANAQTTYLECVTYAMQQLSAVGVTMYLDAGHAGWLGWPANFSRLISCSHPYTPMLGHP
LAFEDLLPTLPTIMPSLLIHQILLPKVIPTTTRCFTSRLWLLCLAPSLLILSLTKVVQV
SRTSDSNGVTGVTFLVLDSERSLRLTPDHP

>NcCel7A

MLAKFAALAALVASANAQA**VCSLTAETHPSLNWSKCTSSGCTNVAGSITVDANWRWTHI**
TSGSTNCYSGNEWDTSLCSTNTDCATKCCVDGAEYSSTYGIQTSGNSLSLQFVTKGSYS
TNIGSRTYLMNGADAYQGFELLGNEFTFDVDVSGTGCGLNGALYFVSMDLDGGKAKYTN
NKAGAKYGTGYCDAQCPRDLKYINGIANVEGWTPSTNDANAGIGDHGTCCSEMDIWEAN
KVSTAFTPHPCTTIEQHMCEGDSCGGTYSDDRYGGTCDADGCDFNSYRMGNTTFYGEGK
TVDTSSKFTVVTQFIKDSAGDLAEIKRFYVQNGKVIENSQSNVDGVSGNSITQSFCNAQ
KTAFGDIDDFNKKGGLKQMGKALAKPMVLVMSIWDDHAANMLWLDSTYPVEGGPGAYRG
ECPTTSGVPAEVEANAPNSKVIFSNIKFGPIGSTFSGGSSGTPPSNPSSSVKPVTSTAK
PSSTSTASNPSGTGAAHWAQCGGIGFSGPTTCQSPYTCQKINDYYSQCV

>NcCel6A

MRALSLLAALSAAGATVTASPVGPAHAPRALSDSNPFVGKKLFANPKWASKLEDTYKAF
TSRG**DSENAAVVRKVQKIGSFVWVSSRSGLSEIDEAIQAARAEQSRTGQKQIVGLVMYN**

IPDRDCSAGESAGELSTKNDGLRIYKEQFVKPYAAKLAAAKDLQFAVVVEPDSLGNSVT
NSAVEFCKQAIPTYREGIAFAIKSLQLDNVAVYLDAANGGWLGWGDSLPKAADEIATIL
SMASPAKVRGFSTNVSNYNPYLATKRDSFTSGSPSYDESHYASSLAPYLQQHGLPAHFI
IDQGRVAYPGARKDWGDWCNVNPAGFGHIPTTDQSVLQNSNVDAIVWIKPGGESDGQCG
LKGAPIAGAWFGGYAEMLTGNADASVKNS

>StCel7A

MYAKFATLAALVAGAAAQN**ACTLTAENHPSLTWSKCTSGGSCTSVQGSITIDANW**
RWTHRTDSATNCYEGNKWDTSYCSDGPSCASKCCIDGADYSSTYGITTSGNSLNL
KFVTKGQYSTNIGSRTYLMESDTKYQMFQLLGNEFTFDVDVSNLGCGLNGALYFV
SMDADGGMSKYSGNKAGAKYGTGYCDSQCPRDLKFINGEANVENWQSSTNDANAG
TGKYGSCCSEMDVWEANNMAAAFTPHPCTVIGQSRCEGDSCGGTYSTDRYAGICD
PDGCDFNSYRQGNKTFYGKGMTVDTTKKITVVTQFLKNSAGELSEIKRFYVQNGK
VIPNSESTIPGVEGNSITQDWCDRQKAAFGDVTDFQDKGGMVQMGKALAGPMVLV
MSIWDDHAVNMLWLDSTWPIDGAGKPGAERGACPTTSGVPAEVEAEAPNSNVIFS
NIRFGPIGSTVSGLPDGGSGNPNPPVSSSTPVPSSSTTSSGSSGPTGGTGVAKHY
EQCGGIGFTGPTQCESPYTCTKLNDWYSQCL

>TrCel7A

MYRKLAVISAFLATARAQS**ACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANW**
RWTHATNSSTNCYDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSI
GFVTQSAQKNVGARLYLMASDTTYQEFTLLGNEFSFDVDVSQLPCGLNGALYFVS
MDADGGVSKYPTNTAGAKYGTGYCDSQCPRDLKFINGQANVEGWEPSSNNANTGI
GGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGTYSDNRYGGTCDP
DGCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQ
PNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSGGMVLVMSLWD
DYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKF
GPIGSTGNPSGGNPPGGNPPGTTTTRRPATTTGSSPGPTQSHYGQCGGIGYSGPT
VCASGTTCQVLNPYYSQCL

>TrCel6A

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSN
DYYSQCLPGAASSSSSTRAASTTSRVSPTTSRSSSATPPPGSTTTRVPPVGSGTA
TYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAAAAVA**KVPSFMWLDTLDK**
TPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAKYK
NYIDTIRQIVVEYSDIRTLLVIEPDSLANLVTNLGTPKCANAQSAYLECINYAVT
QLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRALRGLATNVANY
NGWNITSPPSYTQGNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQ
QQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFDSHCA
LPDALQPAPQAGAWFQAYFVQLLTNANPSFL

BGL amino acid sequence

>AnBgl1

MRFTLIEAVA**LTAVSLASADELAYSPPYYPSPWANGQGDWAEAYQRAVDIVSQMTLAEK**
VNLTTGTGWELELCVGQTGGVPRLGIPGMCAQDSPLGVRDSDYNSAFPAGVNVAATWDK
NLAYLRGQAMGQEFSDKGADIQLGPAAGPLGRSPDGGRNWEGFSPDPALSGVLFAETIK
GIQDAGVVATAKHYIAYEQEHFRQAPEAQGYGFNITESGSANLDDKTMHELYLWPFADA
IRAGAGAVMCSYNQINNSYGCQNSYTLNKLLKAELGFQGFVMSDWAAHHAGVSGALAGL
DMSMPGDVDYDSGTSYWGTNLTISVLNGTVPQWRVDDMAVRIMAAYYKVGRDRLWTPPN
FSSWTRDEYGFKYYYVSEGPYEKVNQFVNVQRNHSELIRRIGADSTVLLKNDGALPLTG
KERLVALIGEDAGSNPYGANGCSDRGCDNGTLAMGWGSGTANFPYLVTPEQAISNEVLK
NKNGVFTATDNWAIDQIEALAKTASVSLVFVNADSGEGYINVDGNLGDRRNLTLWRNGD
NVIKAAASNCNNTIVIIHSVGPVLVNEWYDNPNVTAILWGGLPGQESGNSLADVLYGRV
NPGAKSPFTWGKTREAYQDYLYTEPNNGNGAPQEDFVEGVFIDYRGFDKRNETPIYEFG
YGLSYTTFNYSNLQVEVLSAPAYEPASGETEAAPTFGEVGNASDYLYPDGLQRITKFIY
PWLNSTDLEASSGDASYGQDASDYLPEGATDGSAQPILPAGGGAGGNPRLYDELIRVSV
TIKNTGKVAGDEVPQLYVSLGGPNEPKIVLRQFERITLQPSEETQWSTTLTRRDLANWN
VETQDWEITSYPKMVFVGSSSRKLPLRASLPTVH

Plasmid Sequences

p425-TEF_M

Table 1: p425-TEF_M plasmid features

Feature	Location
TEF1 Pr	113..513
MCS	521..567
CYC1 Tr	521..567
LEU2	1543..3760
2 micron	4032..5379
AmpR	5490..6350
pBS ori	6498..7165

>p425-TEF_M

TATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATG
ATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCATAGCT
TCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCC
GTACCACTTCAAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGG
GTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCT
TTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAAATTT
TTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGT
CTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTC
TTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTGCTAGC
TTAATTAAGGCGCGCCGGCGTTTAAACGGCCGGCCCTCGAGTCATGTAATTAGTTATGT
CACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGA
CAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTT
ATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTA
TACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCCGG
TACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAA

CGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCC
TTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT
GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTT
CGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATC
GGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTT
GATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCA
ACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGG
TTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTT
TACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAT
ATCGACGGTCGAGGAGAACTTCTAGTATATCCACATACCTAATATTATTGCCTTATTAA
AAATGGAATCCCAACAATTACATCAAAATCCACATTCTCTTCAAAATCAATTGTCCTGT
ACTTCCTTGTTCATGTGTGTTCAAAAACGTTATATTTATAGGATAATTATACTCTATTT
CTCAACAAGTAATTGGTTGTTTGGCCGAGCGGTCTAAGGCGCCTGATTCAAGAAATATC
TTGACCGCAGTTAACTGTGGGAATACTCAGGTATCGTAAGATGCAAGAGTTCGAATCTC
TTAGCAACCATTATTTTTTTCCTCAACATAACGAGAACACACAGGGGCGCTATCGCACA
GAATCAAATTCGATGACTGGAAATTTTTTGTTAATTTCAGAGGTCGCCTGACGCATATA
CCTTTTTCAACTGAAAAATTGGGAGAAAAAGGAAAGGTGAGAGCGCCGGAACCGGCTTT
TCATATAGAATAGAGAAGCGTTCATGACTAAATGCTTGCATCACAATACTTGAAGTTGA
CAATATTATTTAAGGACCTATTGTTTTTTCCAATAGGTGGTTAGCAATCGTCTTACTTT
CTAACTTTTCTTACCTTTTACATTTCAGCAATATATATATATATATTTCAAGGATATAC
CATTCTAATGTCTGCCCCTAAGAAGATCGTCGTTTTGCCAGGTGACCACGTTGGTCAAG
AAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTCGTTCCAATGTC
AAGTTCGATTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTTCC
ACTTCCAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTG
TGGGTGGTCCTAAATGGGGTACCGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATC
CGTAAAGAACTTCAATTGTACGCCAACTTAAGACCATGTAACTTTGCATCCGACTCTCT
TTTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTGACTTCGTTGTTGTCA
GAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGAAGACGATGGTGATGGTGTC
GCTTGGGATAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGAATGGCCGC
TTTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCTTGGATAAAGCTAATC

TTTTGGCCTCTTCAAGATTATGGAGAAAAACTGTGGAGGAAACCATCAAGAACGAATTC
CCTACATTGAAGGTTCAACATCAATTGATTGATTCTGCCGCCATGATCCTAGTTAAGAA
CCCAACCCACCTAAATGGTATTATAATCACCAGCAACATGTTTGGTGATATCATCTCCG
ATGAAGCCTCCGTTATCCCAGGTTCCTTGGGTTTGTTGCCATCTGCGTCCTTGGCCTCT
TTGCCAGACAAGAACACCGCATTTGGTTTGTACGAACCATGCCACGGTTCTGCTCCAGA
TTTGCCAAAGAATAAGGTTGACCCTATCGCCACTATCTTGTCTGCTGCAATGATGTTGA
AATTGTCATTGAACTTGCCTGAAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTT
TTGGATGCAGGTATCAGAACTGGTGATTTAGGTGGTTCCAACAGTACCACCGAAGTCGG
TGATGCTGTCGCCGAAGAAGTTAAGAAAATCCTTGCTTAAAAAGATTCTCTTTTTTTAT
GATATTTGTACATAAACTTTATAAATGAAATTCATAATAGAAACGACACGAAATTACAA
AATGGAATATGTTCATAGGGTAGACGAAACTATATACGCAATCTACATACATTTATCAA
GAAGGAGAAAAAGGAGGATAGTAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTC
AACGTGATAAGGAAAAAGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACC
ACACAAAAAGTTAGGTGTAACAGAAAATCATGAAACTACGATTCCTAATTTGATATTGG
AGGATTTTCTCTAAAAAAAAAAAAATACAACAAATAAAAAACACTCAATGACCTGACCA
TTTGATGGAGTTTAAGTCAATACCTTCTTGAAGCATTTCCCATAATGGTGAAAGTTCCC
TCAAGAATTTTACTCTGTCAGAAACGGCCTTACGACGTAGTCGATATGGTGCACTCTCA
GTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCT
GACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGT
CTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAA
AGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAG
TATGATCCAATATCAAAGGAAATGATAGCATTGAAGGATGAGACTAATCCAATTGAGGA
GTGGCAGCATATAGAACAGCTAAAGGGTAGTGCTGAAGGAAGCATACGATACCCCGCAT
GGAATGGGATAATATCACAGGAGGTACTAGACTACCTTTCATCCTACATAAATAGACGC
ATATAAGTACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATAT
ATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTC
GCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGAAGTTCCTATTCCGAAGT
TCCTATTCTCTAGAAAGTATAGGAACTTCAGAGCGCTTTTGAAAACCAAAAGCGCTCTG
AAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAGCTACTAAAATATTG
CGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCT
ATATCCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGA
CCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATTGTAGTAA

GAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAACATTTCCTATACGTAG
TATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTGACCAATGAAGAATCATCA
ACGCTATCACTTTCTGTTCACAAAGTATGCGCAATCCACATCGGTATAGAATATAATCG
GGGATGCCTTTATCTTGAAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGG
GAAGTGGAGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTTCT
AACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTATAGAGCGCACAA
AGGAGAAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATGCATT
TTTGTAGAACAAAAAAGAAGTATAGATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGC
ATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCG
TTGCATTTTTGTTTTACAAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTC
GCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGC
TCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTCAGGTGGCACT
TTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT
GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
GTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTT
CCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGG
TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA
TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA
GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTG
ACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGT
AACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG
ACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTA
CTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG
ACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCG
GTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGT
ATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGAT
CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCAT
ATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATC
CTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT

GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGT
CCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT
ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT
ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG
GGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTAC
AGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG
GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC
CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTC
CCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGC
GGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTT
TACACTTTATGCTTCCGGCTCC

δp425TEF_M

Table 2: δp425-TEF_M plasmid features

Feature	Location
5' δ	41..207
TEF1 Pr	230..630
CYC1 Tr	687..930
LEU2	1660..3877
3' δ	3928..4094
AmpR	4369..5229
pBS ori	5377..6044

> δp425TEF_M

TATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACATGTTGGAATAGAAATCAAC
TATCATCTACTAACTAGTATTTACATTACTAGTATATTATCATATACGGTGTTAGAAGA
TGACGCAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGA
TAATGTAATAGGATCAATGAATATAAACATAAGGGAACAAAAGCTGGAGCTCATAGCTT
CAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCG
TACCACTTCAAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGGG
TGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTT
TTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAAATTTT
TTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGTC
TTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCT
TGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTGCTAGCT
TAATTAAGGCGCGCCGGCGTTTAAACGGCCGGCCCTCGAGTCATGTAATTAGTTATGTC
ACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGAC
AACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTA
TTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTAT
ACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCGGCCGGT
ACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAAC
GTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCT
TTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCG
CAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTG
TGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTC
GCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTG
ACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGT
TAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTT
ACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATA
TCGACGGTCGAGGAGAACTTCTAGTATATCCACATACCTAATATTATTGCCTTATTAAA
AATGGAATCCCAACAATTACATCAAAATCCACATTCTCTTCAAAATCAATTGTCCTGTA
CTTCCTTGTTCATGTGTGTTCAAAAACGTTATATTTATAGGATAATTATACTCTATTTC

TCAACAAGTAATTGGTTGTTTGGCCGAGCGGTCTAAGGCGCCTGATTCAAGAAATATCT
TGACCGCAGTTAACTGTGGGAATACTCAGGTATCGTAAGATGCAAGAGTTCGAATCTCT
TAGCAACCATTATTTTTTTCCTCAACATAACGAGAACACACAGGGGCGCTATCGCACAG
AATCAAATTCGATGACTGGAAATTTTTTGTTAATTTCAGAGGTCGCCTGACGCATATAC
CTTTTTCAACTGAAAAATTGGGAGAAAAAGGAAAGGTGAGAGCGCCGGAACCGGCTTTT
CATATAGAATAGAGAAGCGTTCATGACTAAATGCTTGCATCACAATACTTGAAGTTGAC
AATATTATTTAAGGACCTATTGTTTTTTCCAATAGGTGGTTAGCAATCGTCTTACTTTC
TAACTTTTCTTACCTTTTACATTTCAGCAATATATATATATATTTCAAGGATATACC
ATTCTAATGTCTGCCCCTAAGAAGATCGTCGTTTTGCCAGGTGACCACGTTGGTCAAGA
AATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTCGTTCCAATGTCA
AGTTCGATTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTTCCA
CTTCCAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGT
GGGTGGTCCTAAATGGGGTACCGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATCC
GTAAAGAACTTCAATTGTACGCCAACTTAAGACCATGTAACTTTGCATCCGACTCTCTT
TTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTGACTTCGTTGTTGTCAG
AGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGAAGACGATGGTGATGGTGTCG
CTTGGGATAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGAATGGCCGCT
TTCATGGCCCTACAACATGAGCCACCATTGCCTATTTGGTCCTTGGATAAAGCTAATCT
TTTGGCCTCTTCAAGATTATGGAGAAAAACTGTGGAGGAAACCATCAAGAACGAATTCC
CTACATTGAAGGTTCAACATCAATTGATTGATTCTGCCGCCATGATCCTAGTTAAGAAC
CCAACCCACCTAAATGGTATTATAATCACCAGCAACATGTTTGGTGATATCATCTCCGA
TGAAGCCTCCGTTATCCCAGGTTCCTTGGGTTTGTTGCCATCTGCGTCCTTGGCCTCTT
TGCCAGACAAGAACACCGCATTTGGTTTGTACGAACCATGCCACGGTTCTGCTCCAGAT
TTGCCAAAGAATAAGGTTGACCCTATCGCCACTATCTTGTCTGCTGCAATGATGTTGAA
ATTGTCATTGAACTTGCCTGAAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTTT
TGGATGCAGGTATCAGAACTGGTGATTTAGGTGGTTCCAACAGTACCACCGAAGTCGGT
GATGCTGTCGCCGAAGAAGTTAAGAAAATCCTTGCTTAAAAAGATTCTCTTTTTTTATG
ATATTTGTACATAAACTTTATAAATGAAATTCATAATAGAAACGACACGAAATTACAAA
ATGGAATATGTTCATAGGGTAGACGAAACTATATACGCAATCTACATACATTTATCAAG
AAGGAGAAAAAGGAGGATAGTAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTCA
ACGTGATAAGGAAAAAGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCA
CACAAAAAGTTAGGTGTAACAGAAAATCATGAAACTACGATTCCTAATTTGATATTGGA

GGATTTTCTCTAAAAAAAAAAAAATACAACAAATAAAAAACACTCAATGACCTGACCAT
TTGATGGAGTTTAAGTCAATACCTTCTTGAAGCATTTCCCATAATGGTGAAAGTTCCCT
CAAGAATTTTACTCTGTCAGAAACGGCCTTACGACGTAGTCGATATGGTGCACTCTCAG
TACAATCTGCTCTGATGCCGCATAGTTAAGCCAATAAAATGATGATAATAATATTTATA
GAATTGTGTAGAATTGCAGATTCCCTTTTATGGATTCCTAAATCCTTGAGGAGAACTTC
TAGTATATTCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTA
TCTCAACATTCACCCATTTCTCATTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCAT
GTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATAC
GCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGTATGATCCAATATCA
AAGGAGAATCCATGGATAGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT
GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
GTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTT
CCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGG
TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTA
TTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAA
TGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA
GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTG
ACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGT
AACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTG
ACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTA
CTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG
ACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCG
GTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGT
ATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGAT
CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCAT
ATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATC
CTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCT
GCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAG
CTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGT
CCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT

ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT
ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG
GGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTAC
AGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG
GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGAT
GCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC
CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTC
CCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGC
GGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTT
TACACTTTATGCTTCCGGCTCC

p416TEF-MFα-prepro

Table 3: p416TEF-MFα-prepro plasmid features

Feature	Location
ARSH4	82..456
CEN6	457..573
AmpR	715..1575
PMB1	1576..2516
TEF1 Pr	2794..3194
MFα prepro	3216..3471
CYC1 Tr	3485..3736
URA3	4526..5329

> p416TEF-MFα-prepro

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTT
TCTTAGGACGGATCGCTTGCCTGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAA

TAATTTGGGAATTTACTCTGTGTTTATTTATTTTTATGTTTTGTATTTGGATTTTAGAA
AGTAAATAAAGAAGGTAGAAGAGTTACGGAATGAAGAAAAAAAAATAAACAAAGGTTTA
AAAAATTTCAACAAAAAGCGTACTTTACATATATATTTATTAGACAAGAAAAGCAGATT
AAATAGATATACATTCGATTAACGATAAGTAAAATGTAAAATCACAGGATTTTCGTGTG
TGGTCTTCTACACAGACAAGATGAAACAATTCGGCATTAATACCTGAGAGCAGGAAGAG
CAAGATAAAAGGTAGTATTTGTTGGCGATCCCCCTAGAGTCTTTTACATCTTCGGAAAA
CAAAAACTATTTTTTCTTTAATTTCTTTTTTTACTTTCTATTTTTAATTTATATATTTA
TATTAAAAAATTTAAATTATAATTATTTTTATAGCACGTGATGAAAAGGACCCAGGTGG
CACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAA
ATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGG
AAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTG
CCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGT
TGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGT
TTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC
GGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTC
AGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACA
GTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACT
TCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATC
ATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGA
ACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG
CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA
GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTC
CCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGAC
AGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTAC
TCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAA
GATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTA
ATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCA
AGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATA
CTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTG

TCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAA
CGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC
CTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTG
TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACG
GTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT
CTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACG
ACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCC
TCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGA
AAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAG
GCTTTACACTTTATGCTTCCGGCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATT
TCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAA
GGGAACAAAAGCTGGAGCTCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCA
GATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAAGCACAGCATACT
AAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAA
AAGAAAAAAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTT
ATCACGTTTCTTTTTCTTGAAAATTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCC
CATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCT
TGTTCTATTACAACTTTTTTTACTTCTTGCTCATTAGAAAGAAAGCATAGCAATCTAAT
CTAAGTTTTCTAGAACTAGTGCTAGCATAATGAGATTTCCTTCAATTTTTACTGCAGTT
TTATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAAC
GGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTTAGATTTAGAAGGGGATTTCGATG
TTGCTGTTTTGCCATTTTCCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACT
ATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTTTGGATAAAAGAGGCCGGCCCTC
GAGTCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTA
ACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGT
TATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGAC
GCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCG
AAGGCTTTAATTTGCGGCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGC
TCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG

ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGC
GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
CGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCT
TTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTG
ATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATT
TTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAA
TTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCT
GTGCGGTATTTCACACCGCATAGGGTAATAACTGATATAATTAAATTGAAGCTCTAATT
TGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTTTAGTTTTGCTGGCCGCA
TCTTCTCAAATATGCTTCCCAGCCTGCTTTTCTGTAACGTTCACCCTCTACCTTAGCAT
CCCTTCCCTTTGCAAATAGTCCTCTTCCAACAATAATAATGTCAGATCCTGTAGAGACC
ACATCATCCACGGTTCTATACTGTTGACCCAATGCGTCTCCCTTGTCATCTAAACCCAC
ACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTTTGAG
CAATAAAGCCGATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTA
TATTCTCCAGTAGATAGGGAGCCCTTGCATGACAATTCTGCTAACATCAAAAGGCCTCT
AGGTTCCTTTGTTACTTCTTCTGCCGCCTGCTTCAAACCGCTAACAATACCTGGGCCCA
CCACACCGTGTGCATTCGTAATGTCTGCCCATTCTGCTATTCTGTATACACCCGCAGAG
TACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTTCGAAGAGTAAAAA
ATTGTACTTGGCGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATGGAAAAATCAG
TCAAGATATCCACATGTGTTTTTAGTAAACAAATTTTGGGACCTAATGCTTCAACTAAC
TCCAGTAATTCCTTGGTGGTACGAACATCCAATGAAGCACACAAGTTTGTTTGCTTTTC
GTGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCT
TATATGTAGCTTTCGACATGATTTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAGT
TGGGTTAAGAATACTGGGCAATTTCATGTTTCTTCAACACTACATATGCGTATATATAC
CAATCTAAGTCTGTGCTCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATTACCGAATCAA
AAAAATTTCAAAGAAACCGAAATCAAAAAAAAGAATAAAAAAAAAATGATGAATTGAAT
TGAAAAGCTGTGGTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAG
CCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGG
CATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCA
CCGTCATCACCGAAACGCGCGA

δ-TEFpr-BGL-cyc1tt-δ

Table 3: δ-TEFpr-AnBgl1-cyc1tt-δ plasmid features

Feature	Location
5'δ	1..46
TEF1 Pr	47..437
AnBgl1	460..3042
CYC1 Tr	3058..3297
3'δ	3305..3355

> δ-TEFpr-AnBgl1-cyc1tt-δ

GAAACGCAAGGATTGATAATGTAATAGGATCAATGAATATAAACATATAGCTTCAAAAT
GTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCAC
TTCAAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGT
TAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTTTCTT
CGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAAATTTTTTTTTT
GATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGTCTTCAAT
TTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGCTCA
TTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGAACTAGTATGAGGTTCACTT
TGATCGAGGCGGTGGCTCTGACTGCCGTCTCGCTGGCCAGCGCTGATGAATTGGCCTAC
TCCCCGCCGTATTACCCCTCCCCTTGGGCCAATGGCCAGGGTGACTGGGCGGAAGCATA
CCAGCGCGCTGTTGATATCGTCTCGCAGATGACATTGGCTGAGAAGGTCAATTTGACTA
CGGGAACTGGATGGGAATTGGAATTATGTGTTGGTCAGACTGGAGGTGTTCCCCGATTG
GGAATTCCGGGAATGTGTGCACAGGATAGCCCTCTGGGTGTTCGTGACTCCGACTACAA
CTCTGCGTTCCCTGCCGGTGTCAACGTGGCCGCAACCTGGGACAAGAATCTGGCTTACC
TTCGTGGCCAGGCTATGGGTCAGGAGTTTAGTGACAAGGGTGCTGATATCCAATTGGGT
CCAGCTGCCGGCCCTCTCGGTAGAAGTCCCGACGGCGGTCGTAACTGGGAGGGCTTCTC

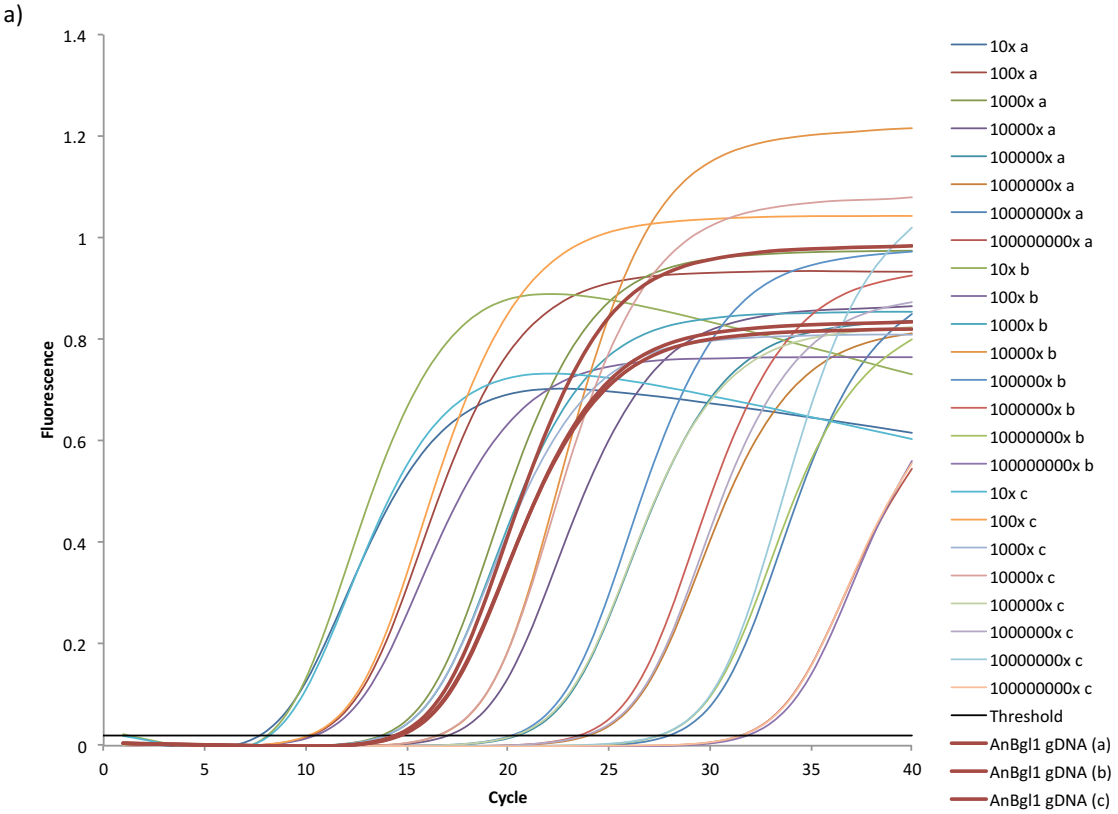
CCCCGACCCGGCCCTCAGTGGTGTGCTCTTTGCAGAGACAATCAAGGGTATTCAGGATG
CTGGTGTGGTTGCAACGGCTAAGCACTACATCGCCTACGAGCAGGAGCATTTCCGTCAG
GCGCCTGAAGCTCAAGGCTACGGATTCAATATTACCGAGAGTGGAAGCGCGAACCTCGA
CGATAAGACTATGCATGAGCTGTACCTCTGGCCCTTCGCGGATGCCATCCGTGCAGGTG
CCGGTGCTGTGATGTGCTCGTACAACCAGATCAACAACAGCTATGGCTGCCAAAACAGC
TACACTCTGAACAAGCTGCTCAAGGCTGAGCTGGGTTTCCAGGGCTTTGTCATGAGTGA
TTGGGCGGCTCACCATGCCGGTGTGAGTGGTGCTTTGGCGGGATTGGACATGTCTATGC
CGGGAGACGTCGATTACGACAGTGGCACGTCTTACTGGGGTACCAACTTGACCATCAGT
GTGCTCAACGGGACGGTGCCCCAATGGCGTGTTGATGACATGGCTGTCCGCATCATGGC
CGCCTACTACAAGGTCGGCCGTGACCGTCTGTGGACTCCTCCCAACTTCAGCTCATGGA
CCAGAGATGAATACGGCTTCAAGTACTACTATGTCTCGGAGGGACCGTATGAGAAGGTC
AACCAGTTCGTGAACGTGCAACGCAACCATAGCGAGTTGATCCGCCGTATTGGAGCAGA
CAGCACGGTGCTCCTCAAGAACGATGGCGCTCTTCCCTTGACTGGAAAGGAGCGCTTGG
TCGCCCTTATCGGAGAAGATGCGGGTTCCAATCCTTATGGTGCCAACGGCTGCAGTGAC
CGTGGGTGCGACAATGGAACATTGGCGATGGGCTGGGGAAGTGGCACTGCCAACTTTCC
CTACTTGGTGACCCCCGAGCAGGCCATCTCGAACGAGGTGCTCAAGAACAAGAATGGCG
TATTCACTGCGACCGATAACTGGGCTATTGATCAGATTGAGGCGCTTGCTAAGACCGCC
AGTGTCTCTCTTGTCTTTGTCAACGCCGACTCTGGTGAGGGTTATATCAATGTCGACGG
AAACCTGGGTGACCGCAGGAACCTGACCCTGTGGAGGAACGGCGACAATGTGATCAAGG
CTGCTGCTAGCAACTGCAACAACACGATCGTTATTATTCACTCTGTCGGCCCAGTCTTG
GTTAACGAGTGGTACGACAACCCCAATGTTACCGCTATTCTCTGGGGTGGTCTTCCCGG
TCAGGAGTCTGGCAACTCCCTCGCCGACGTGCTCTACGGCCGTGTCAACCCCGGTGCCA
AGTCGCCCTTCACCTGGGGCAAGACTCGTGAGGCCTACCAAGATTACTTGTACACCGAG
CCCAACAACGGCAACGGAGCGCCCCAGGAAGACTTCGTCGAGGGCGTCTTCATTGACTA
CCGCGGATTTGACAAGCGCAACGAGACTCCTATCTATGAGTTCGGCTATGGTCTGAGCT
ACACCACCTTCAACTACTCGAACCTTCAGGTGGAGGTTCTGAGCGCCCCTGCGTACGAG
CCTGCTTCGGGCGAGACTGAGGCAGCGCCGACTTTCGGAGAGGTCGGAAATGCGTCGGA
TTACCTCTACCCCGATGGACTGCAGAGAATCACCAAGTTCATCTACCCCTGGCTCAACA
GTACCGATCTTGAGGCGTCTTCTGGGGATGCTAGCTATGGGCAGGATGCCTCAGACTAT
CTTCCCGAGGGAGCCACCGATGGCTCTGCGCAACCGATCCTGCCTGCCGGTGGTGGTGC
TGGCGGCAACCCTCGCCTGTACGACGAGCTCATCCGCGTGTCGGTGACTATCAAGAACA
CCGGCAAGGTTGCGGGTGATGAAGTTCCTCAACTGTATGTTTCTCTTGGCGGCCCTAAC

GAACCCAAGATCGTGCTGCGTCAATTCGAGCGTATCACGCTGCAGCCGTCGGAAGAGAC
GCAGTGGAGCACAACTCTGACGCGCCGTGACCTTGCGAACTGGAATGTTGAGACGCAGG
ACTGGGAGATTACGTCGTATCCCAAGATGGTGTTTGTCGGAAGCTCCTCGCGGAAGCTG
CCGCTCCGGGCGTCTCTGCCTACTGTTCACTAAGGCCGGCCCTCGAGTCATGTAATTAG
TTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGA
GTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAG
AACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTA
ACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGC
ATAAAATGATGATAATAATATTTATAGAATTGTGTAGAATTGCAGATTCCC

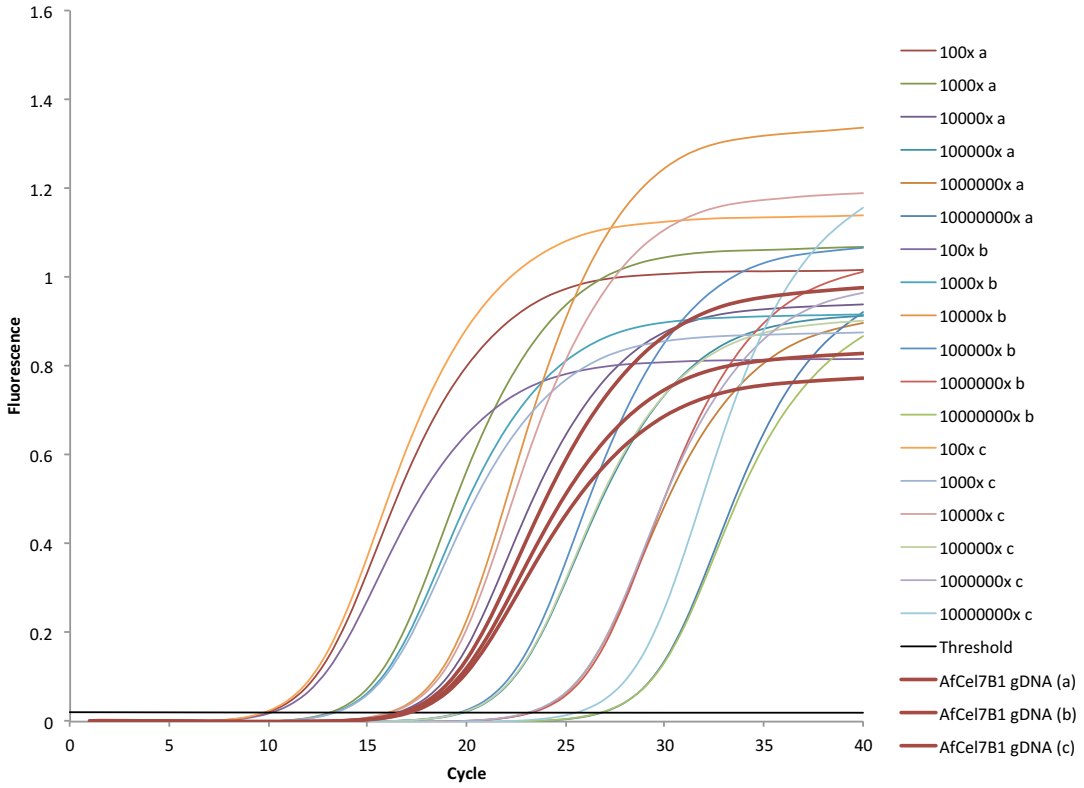
qPCR Supplementary Figures

qPCR amplification plots

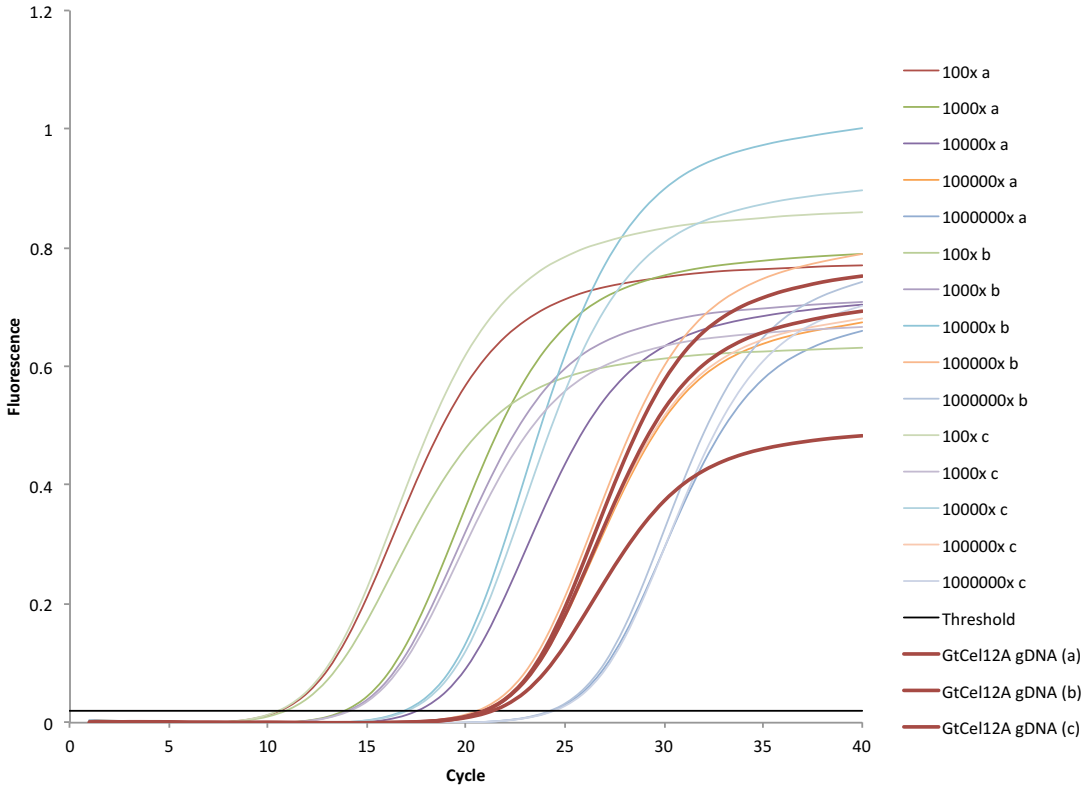
a)



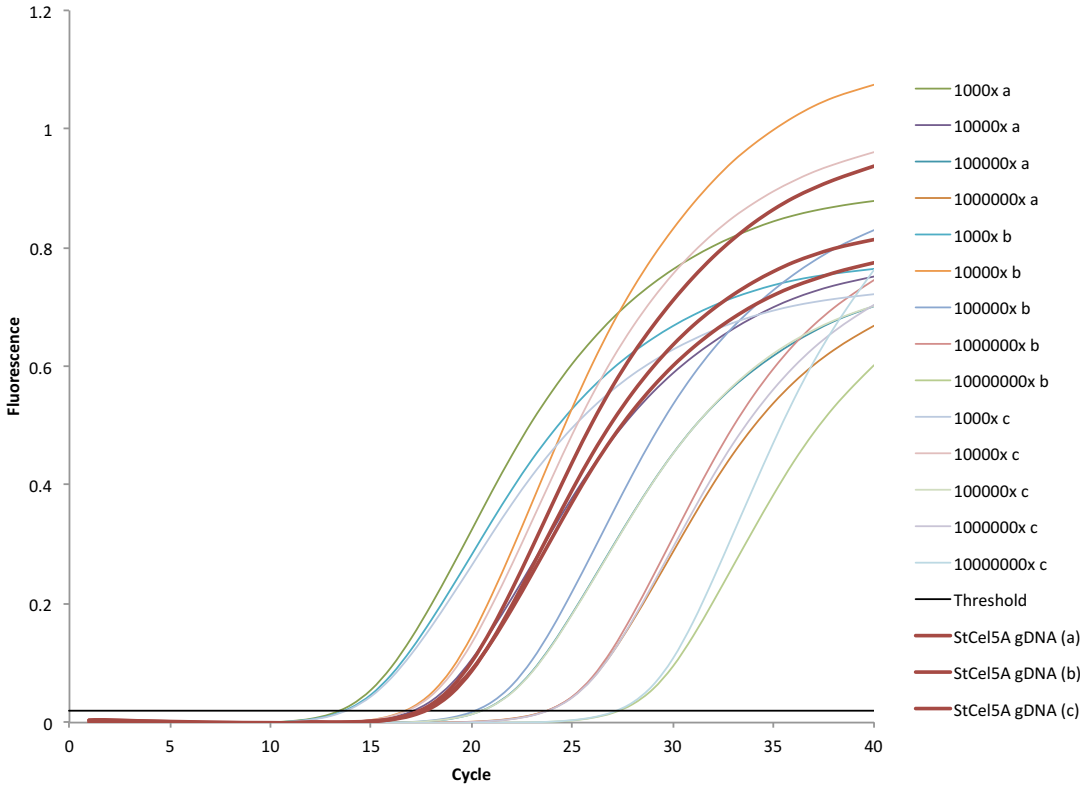
b)


Fluorescence
Cycle
100x a
1000x a
10000x a
100000x a
1000000x a
10000000x a
100x b
1000x b
10000x b
100000x b
1000000x b
10000000x b
100x c
1000x c
10000x c
100000x c
1000000x c
10000000x c
Threshold
AfCel7B1 gDNA (a)
AfCel7B1 gDNA (b)
AfCel7B1 gDNA (c)

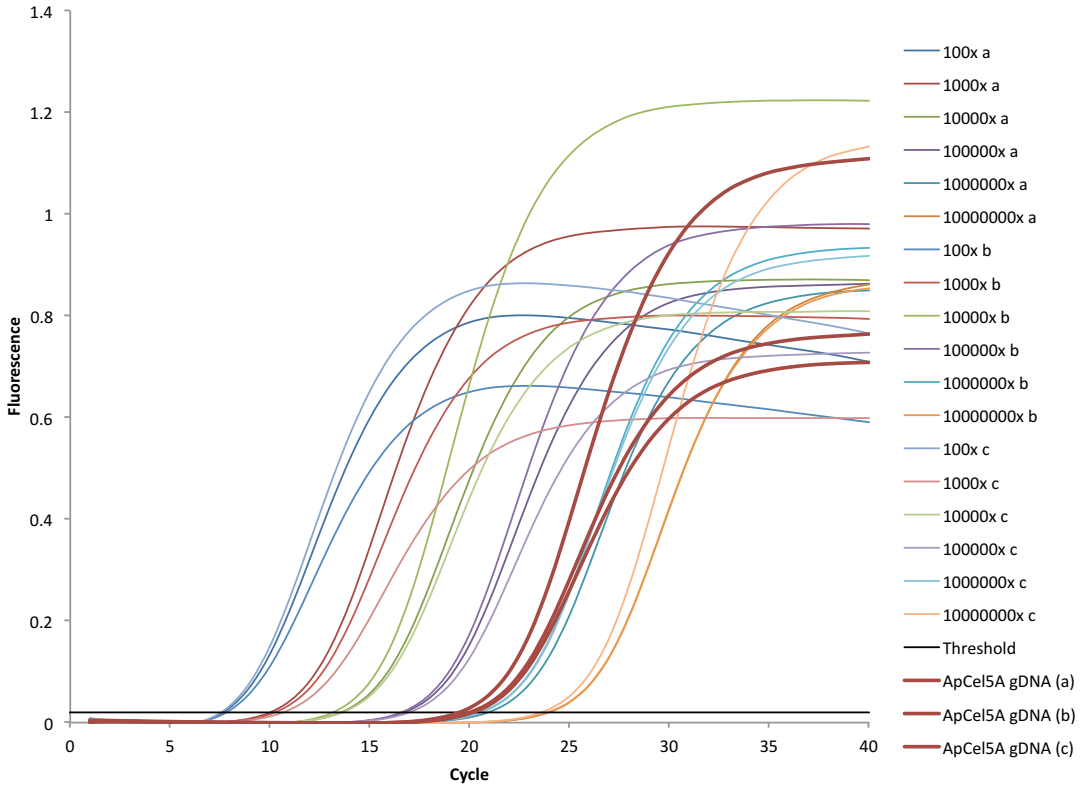
c)


Fluorescence
Cycle
100x a
1000x a
10000x a
100000x a
1000000x a
100x b
1000x b
10000x b
100000x b
1000000x b
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
GtCel12A gDNA (a)
GtCel12A gDNA (b)
GtCel12A gDNA (c)

d)


Fluorescence
Cycle
1000x a
10000x a
100000x a
1000000x a
1000x b
10000x b
100000x b
1000000x b
10000000x b
1000x c
10000x c
100000x c
1000000x c
10000000x c
Threshold
StCel5A gDNA (a)
StCel5A gDNA (b)
StCel5A gDNA (c)

e)


Fluorescence
Cycle
100x a
1000x a
10000x a
100000x a
1000000x a
10000000x a
100x b
1000x b
10000x b
100000x b
1000000x b
10000000x b
100x c
1000x c
10000x c
100000x c
1000000x c
10000000x c
Threshold
ApCel5A gDNA (a)
ApCel5A gDNA (b)
ApCel5A gDNA (c)

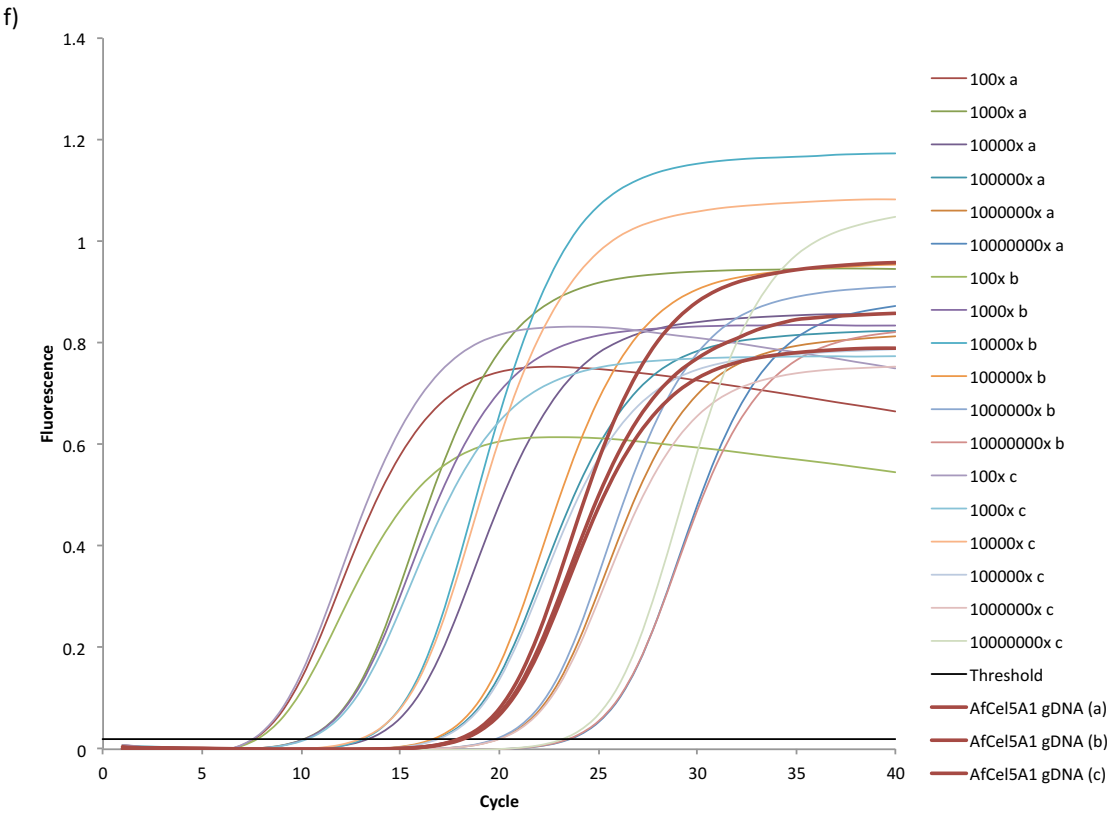
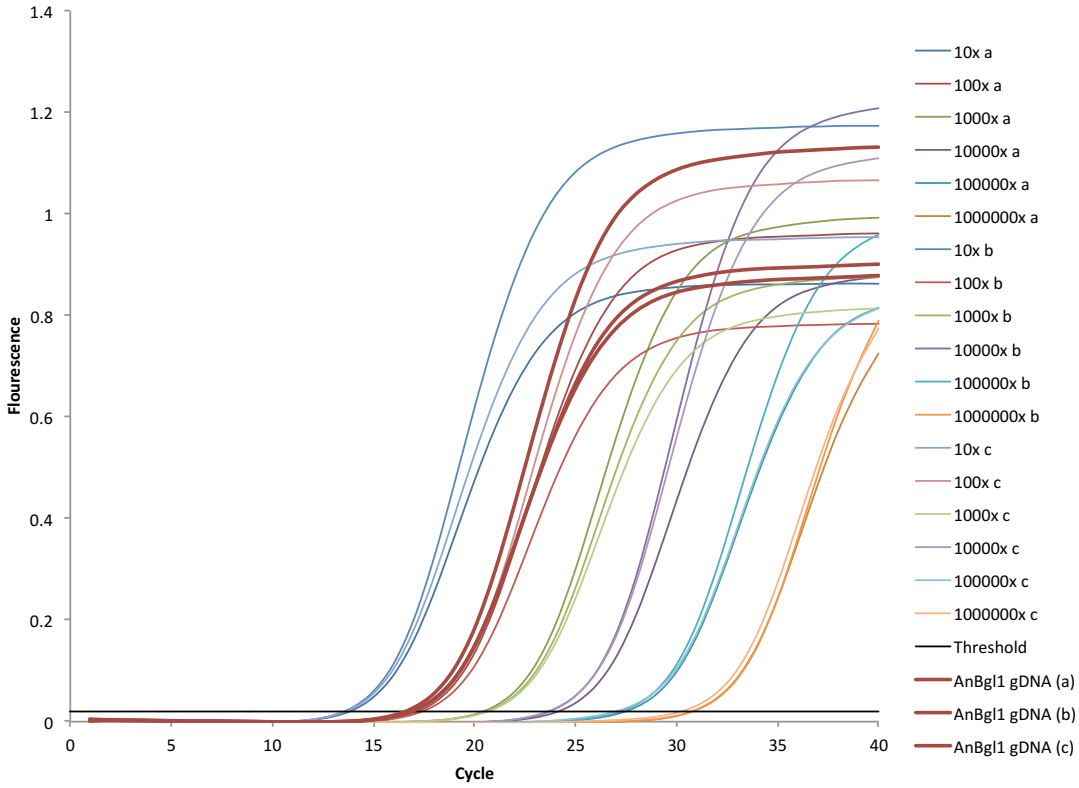
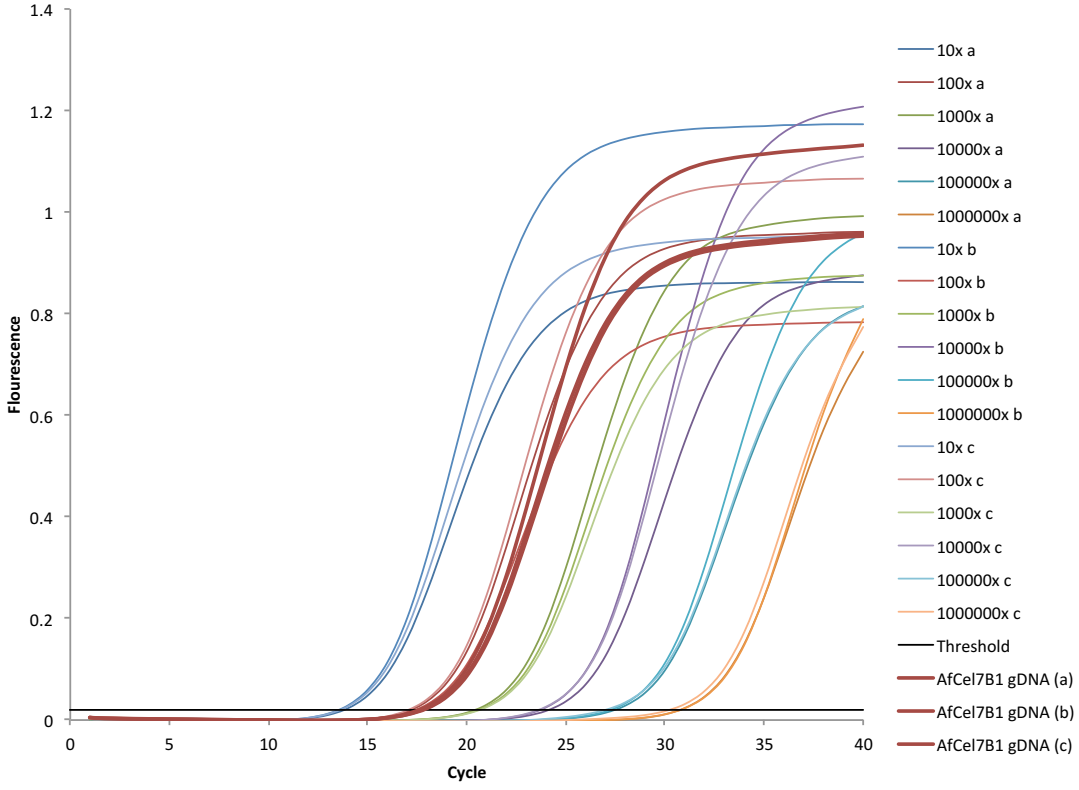


Figure A1 – Amplification plots obtained from serial dilutions of DNA samples quantified using PicoGreen. Each curve represents a single dilution. Amplification plots were generated in triplicate (a, b, and c) for each dilution. The dilution factor of each DNA sample is indicated adjacent each panel. Amplification plots obtained from the serial dilution of a) p425-TEF_M-AnBgl1 plasmid, b) p425-TEF_M-AfCel7B1opt plasmid, c) p425-TEF_M-GtCel12Aopt plasmid, d) p425-TEF_M-StCel5Aopt plasmid, e) p425-TEF_M-preApCel5Aopt plasmid, and f) p425-TEF_M-preproAfCel5A1opt plasmid. The qPCR plots for the genomic DNA samples are in bold maroon and were also generated in triplicate (a, b, and c).

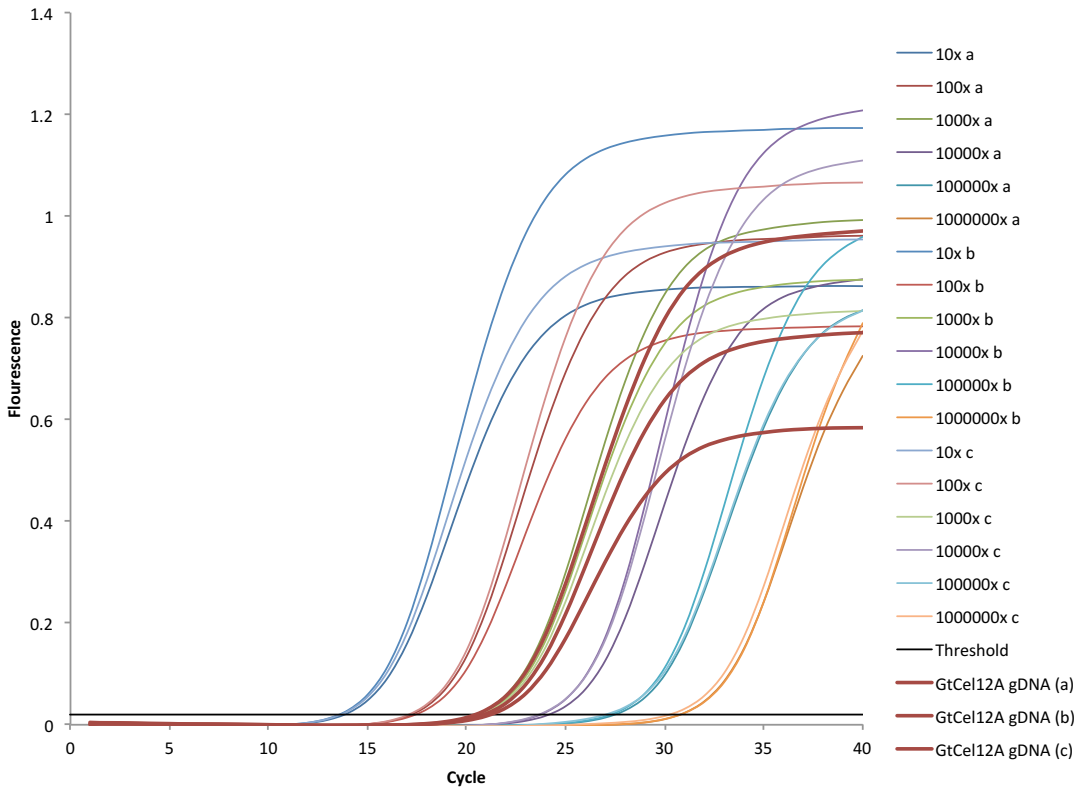
a)


Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
AnBgl1 gDNA (a)
AnBgl1 gDNA (b)
AnBgl1 gDNA (c)

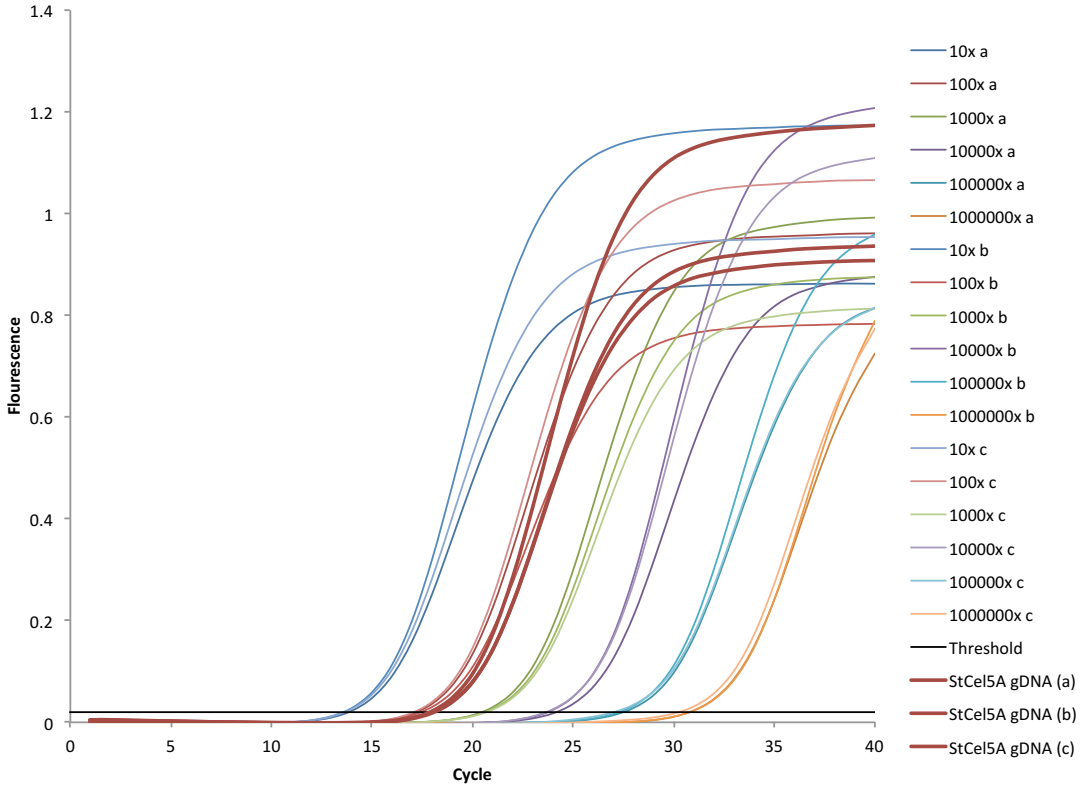
b)


Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
AfCel7B1 gDNA (a)
AfCel7B1 gDNA (b)
AfCel7B1 gDNA (c)

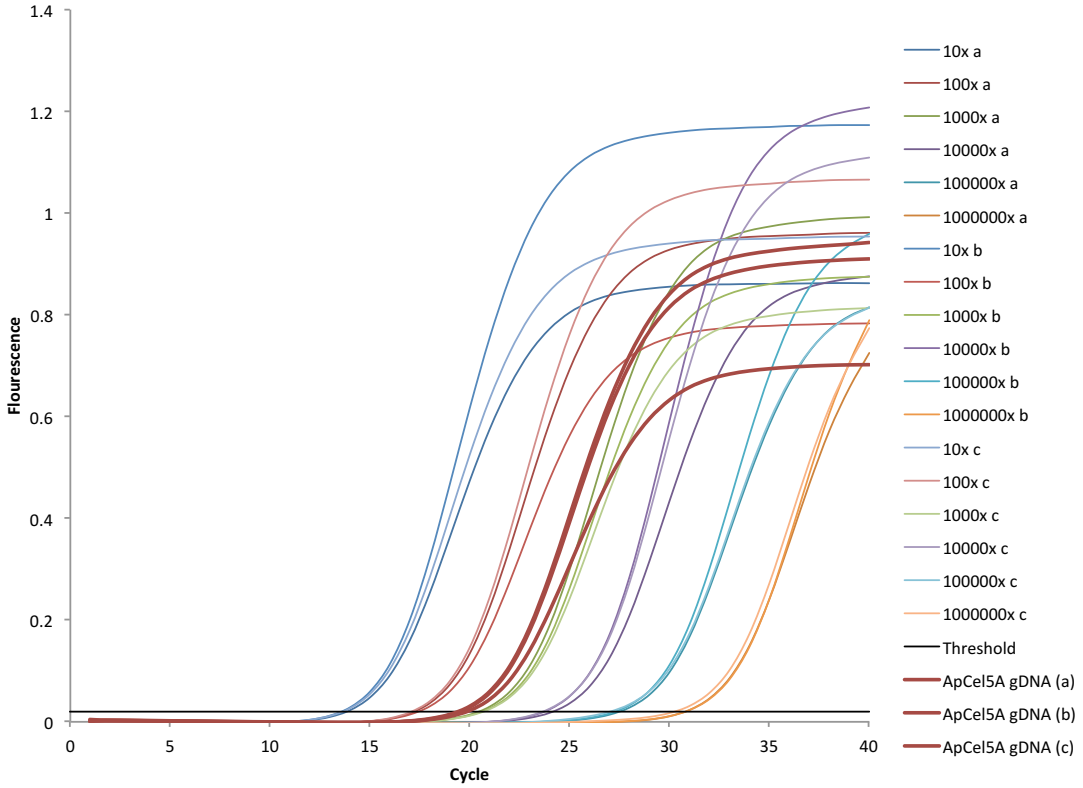
c)


Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
GtCel12A gDNA (a)
GtCel12A gDNA (b)
GtCel12A gDNA (c)

d)


Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
StCel5A gDNA (a)
StCel5A gDNA (b)
StCel5A gDNA (c)

e)


Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
ApCel5A gDNA (a)
ApCel5A gDNA (b)
ApCel5A gDNA (c)

f)

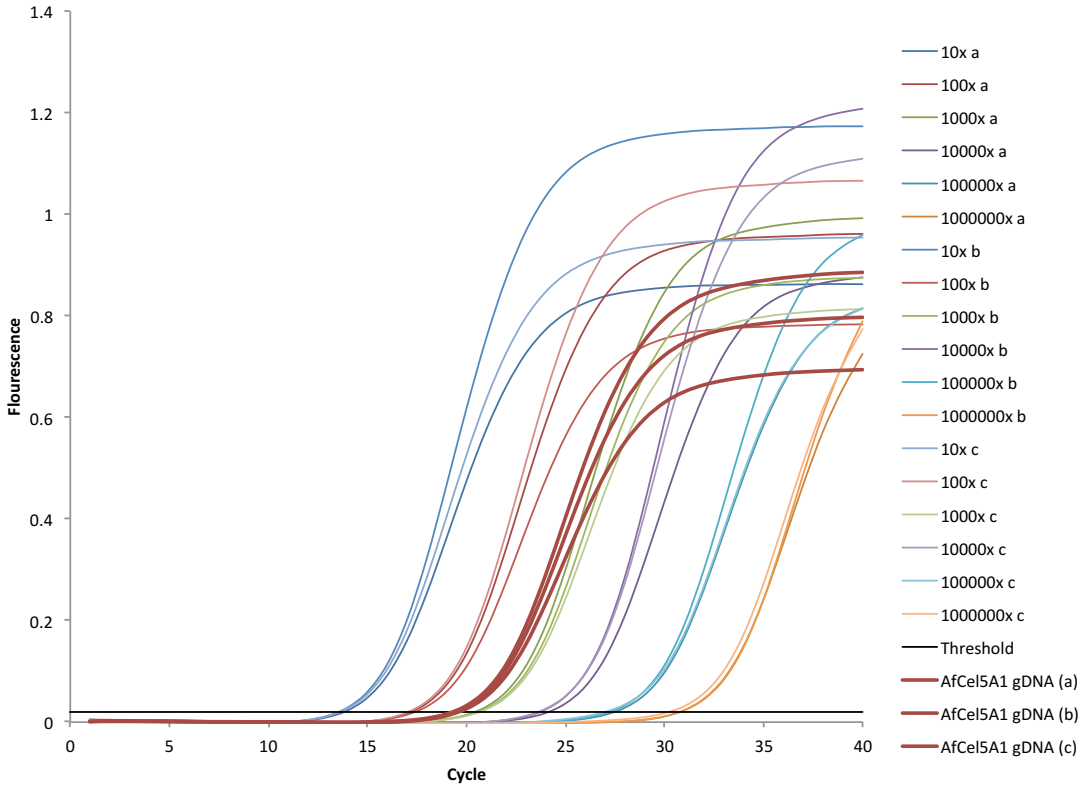
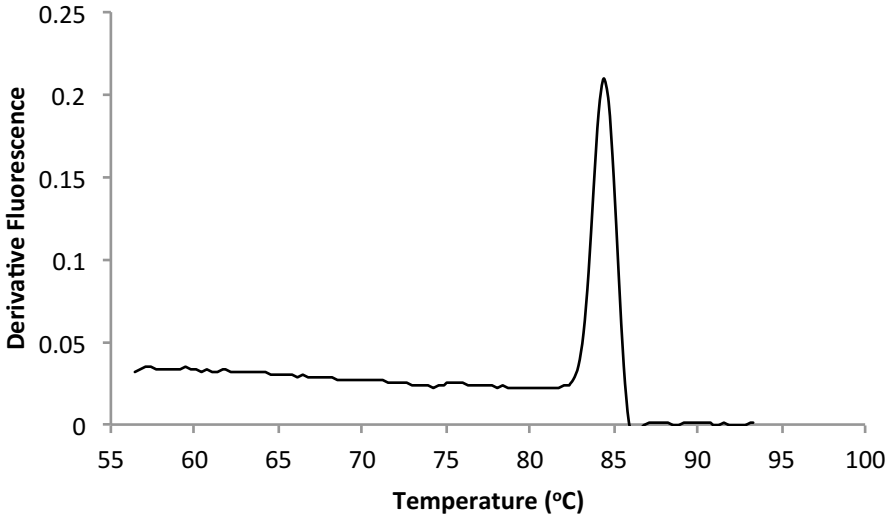

Flourescence
Cycle
10x a
100x a
1000x a
10000x a
100000x a
1000000x a
10x b
100x b
1000x b
10000x b
100000x b
1000000x b
10x c
100x c
1000x c
10000x c
100000x c
1000000x c
Threshold
AfCel5A1 gDNA (a)
AfCel5A1 gDNA (b)
AfCel5A1 gDNA (c)

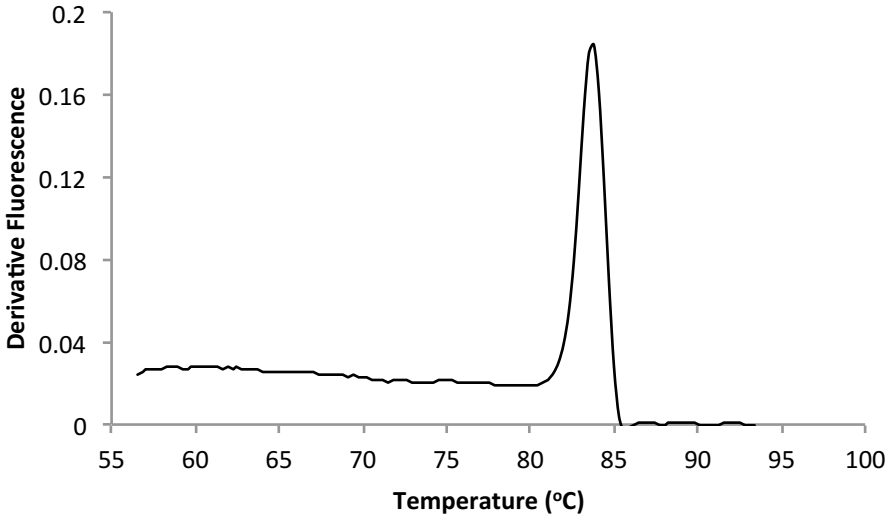
Figure A2 – Amplification plots obtained from six 10-fold serial dilutions of a gel purified PCR product of PGK1 quantified using PicoGreen. Each curve represents a single dilution. Amplification plots were generated in triplicate (a, b, and c) for each dilution. The dilution factor is indicated adjacent each panel. Amplification plots of *PGK1* obtained from the genomic DNA of a) CEN.PK111-61A-δ-AnBgl1a, b) CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-AfCel7B-1opt, c) CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-GtCel12Aopt, d) CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-StCel5Aopt, e) CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preApCel5Aopt, and f) CEN.PK111-61A-δ-AnBgl1a_δ-p425-TEF_M-preproAfCel5A1opt as template are in bold maroon and were also generated in triplicate (a, b, and c).

qPCR derivative melt plots

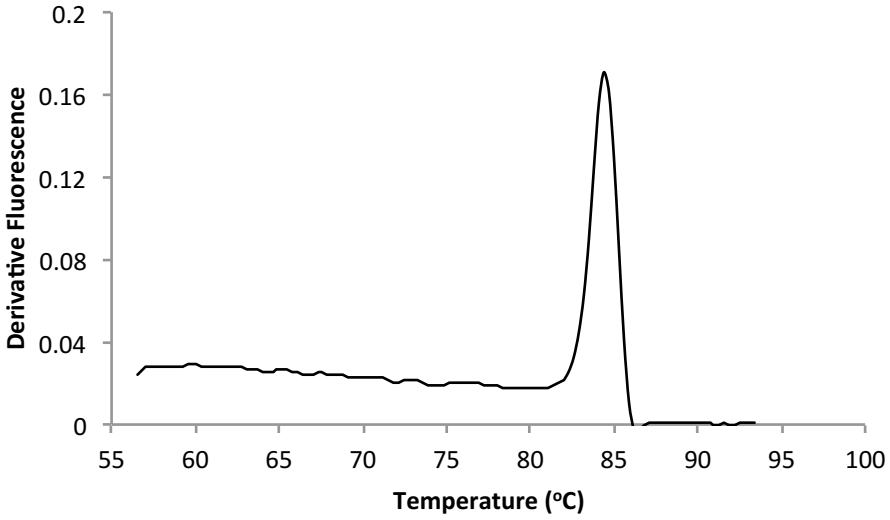
a)



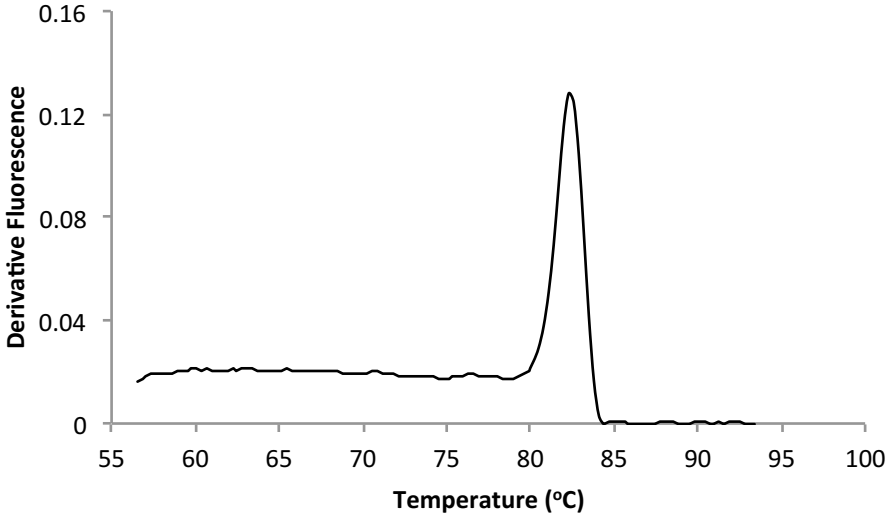
b)



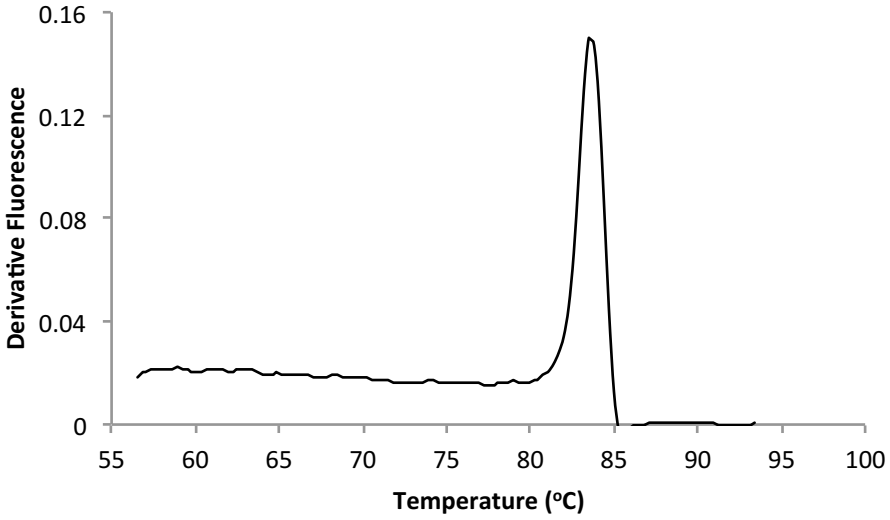
c)


0.2
0.16
0.12
0.08
0.04
0
Derivative Fluorescence
55
60
65
70
75
80
85
90
95
100
Temperature (°C)

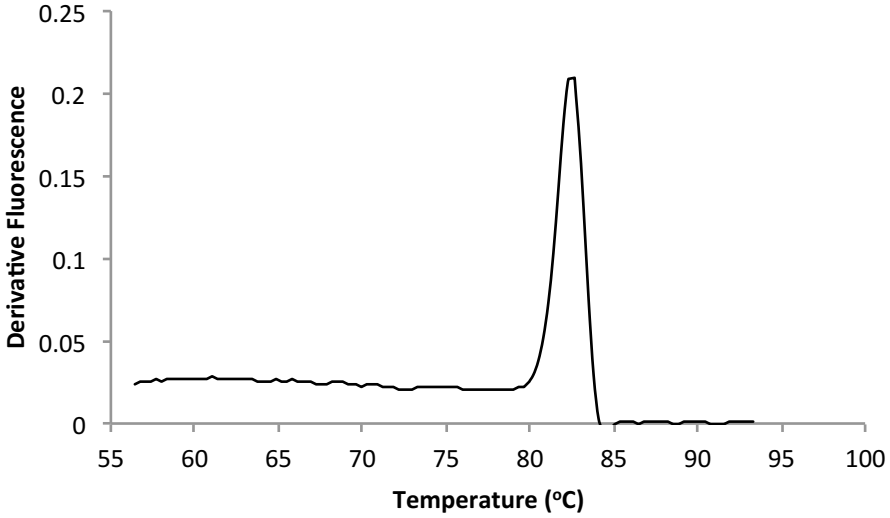
d)


0.16
0.12
0.08
0.04
0
Derivative Fluorescence
55
60
65
70
75
80
85
90
95
100
Temperature (°C)

e)


0.16
0.12
0.08
0.04
0
Derivative Fluorescence
55
60
65
70
75
80
85
90
95
100
Temperature (°C)

f)


0.25
0.2
0.15
0.1
0.05
0
Derivative Fluorescence
55
60
65
70
75
80
85
90
95
100
Temperature (°C)

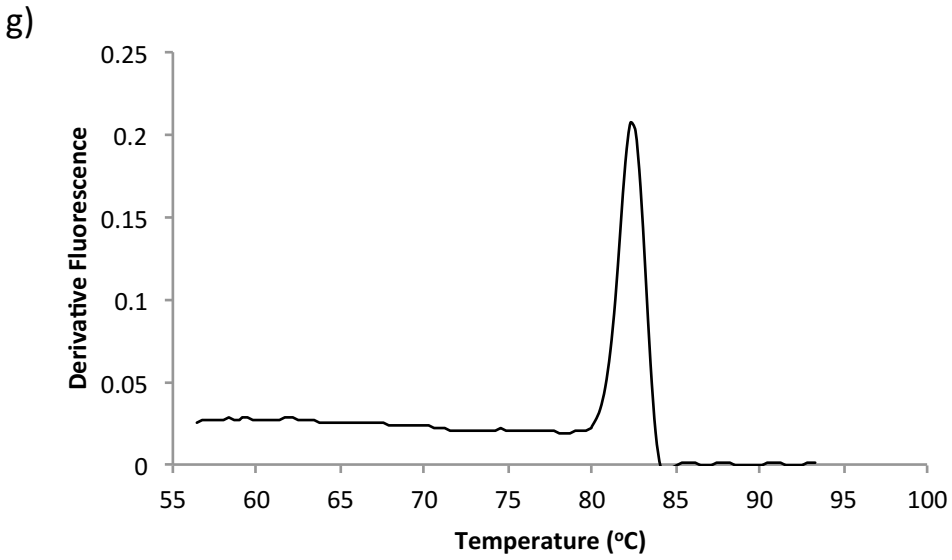


Figure A3 – Melt curves displayed as negative derivative of fluorescence versus temperature. Melt curves were produced using prime pairs a) PGK1_F1/ PGK1_R1, b) AnBgl1_F1/ AnBgl1_R1, c) AfCel7B_F1/ AfCel7B_R1, d) GtCel12A_F1/ GtCel12A_R1, e) StCel5A_F1/ StCel5A_R1, f) ApCel5A_F1/ ApCel5A_R1, and g) AfCel5A_F1/ AfCel5A_R1. Each of these primer pairs produced a single peak in the derivative of their melt curve indicating the presence of a single PCR product.

SDS-PAGE Analysis

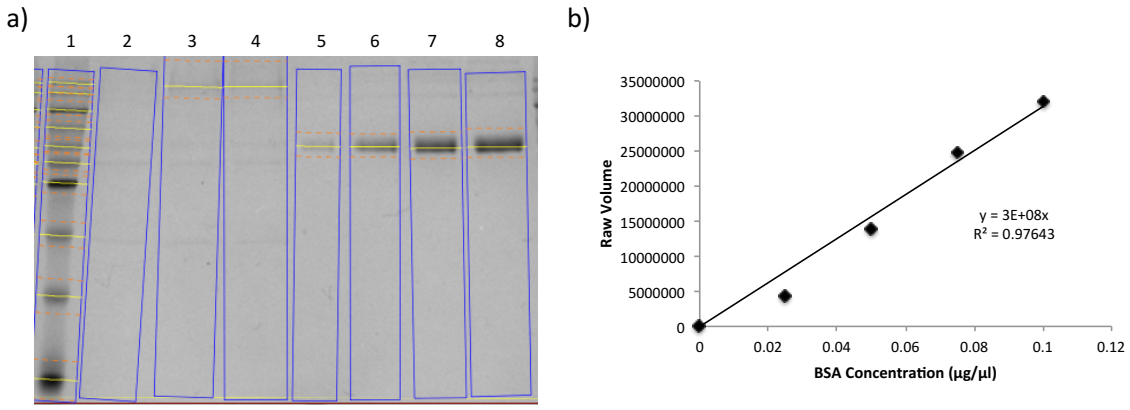
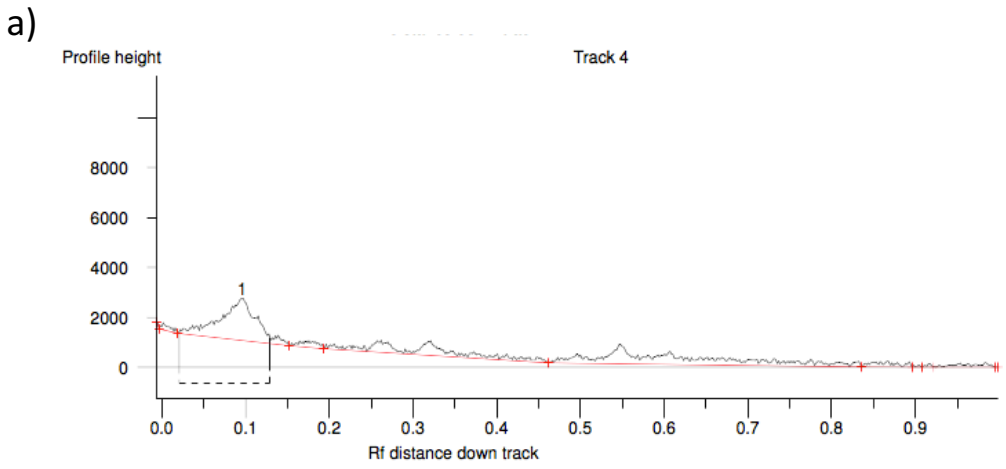
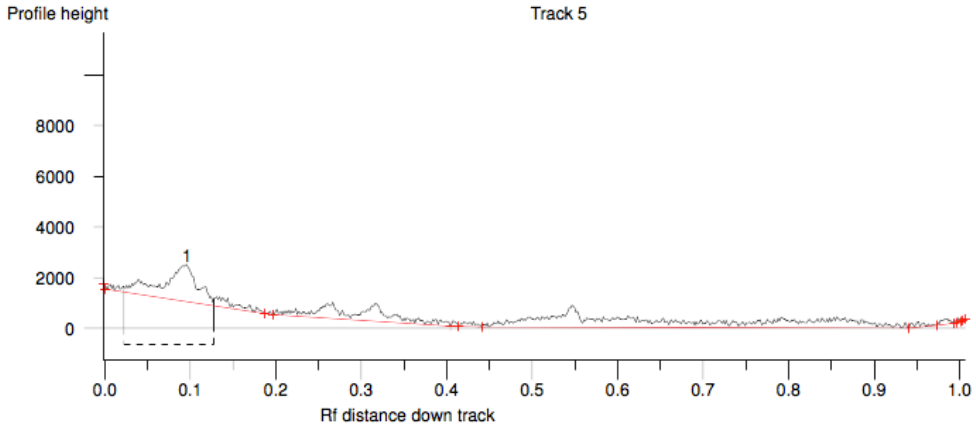


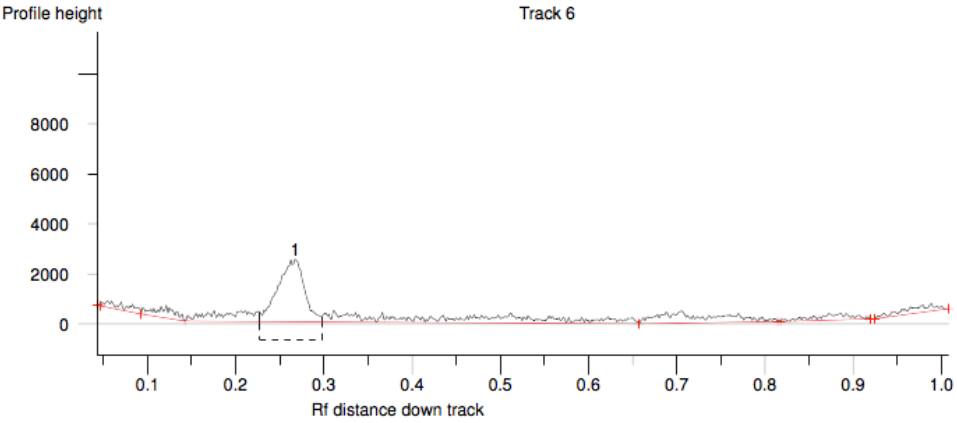
Figure A4 – SDS-PAGE sample analysis. A) Lane 1: MW marker; lane 2: wt (control); lane 3: AnBgl1 from the culture supernatant of CEN.PK111-61A-δ-AnBgl1a grown on cellobiose media; lane 4: AnBgl1 from the culture supernatant of CEN.PK111-61A-δ-AnBgl1a grown on glucose media; lane 5: 0.25 µg BSA; lane 6: 0.5 µg BSA; lane 7: 0.75 µg BSA; and, lane 8: 1 µg BSA. B) BSA standard curve.



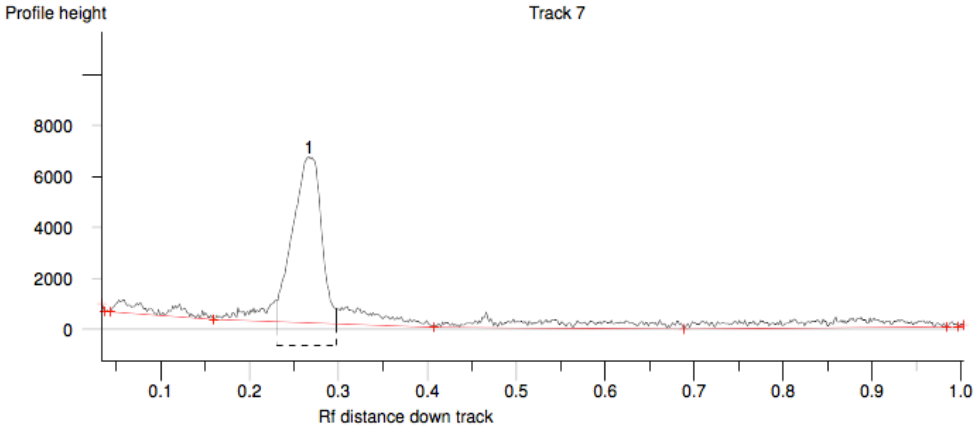
b)


Profile height
Track 5
8000
6000
4000
2000
0
1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Rf distance down track

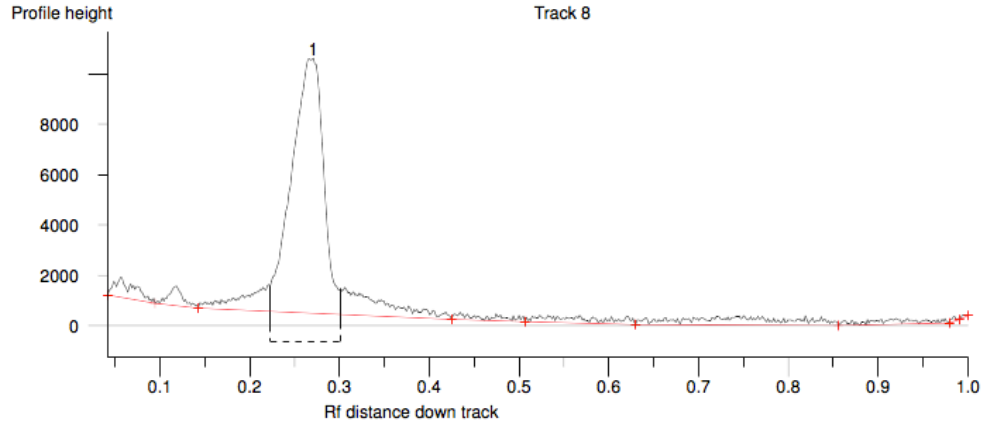
c)


Profile height
Track 6
8000
6000
4000
2000
0
1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Rf distance down track

d)


Profile height
Track 7
8000
6000
4000
2000
0
1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Rf distance down track

e)



f)

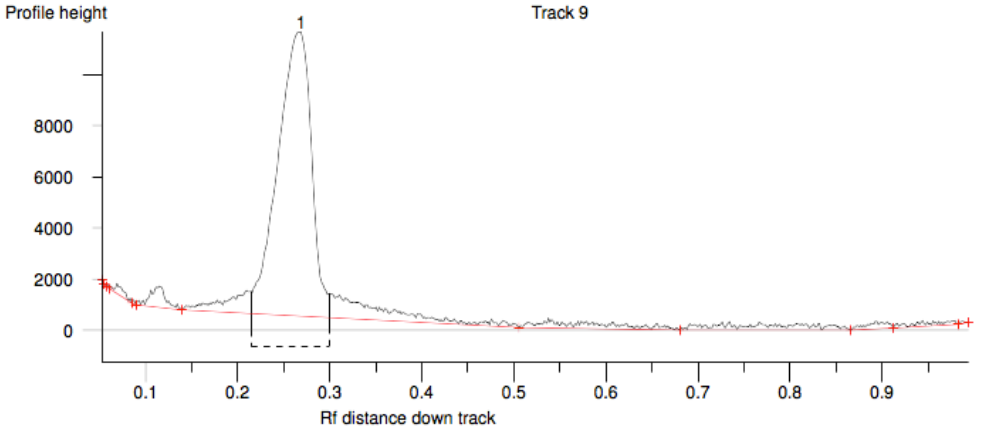


Figure A5 – Densitometry profiles. The selected peaks correspond to the proteins of interest in lanes 3 – 8 in the SDS-PAGE gel in figure A4. The proteins of interest are a) AnBgl1 from the culture supernatant of CEN.PK111-61A-δ-AnBgl1a grown on cellobiose media; b) AnBgl1 from the culture supernatant of CEN.PK111-61A-δ-AnBgl1a grown on glucose media; c) 0.25 µg BSA; d) 0.5 µg BSA; e) 0.75 µg BSA; and, f) 1 µg BSA.