Screening and Engineering of *Geobacillus* spp. for Consolidated Bioprocessing of Lignocellulosic Biomass

Ross William Kent, B.Sc.

Submitted by Ross William Kent to the University of Exeter as a dissertation for the degree of Masters by Research in in Biological Sciences, October 2015

This dissertation is available for Library use on the understanding that it is copyright material and that no quotation from the dissertation may be published without proper acknowledgement.

I certify that all material in this dissertation which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

School of Biosciences, University of Exeter

Geoffrey Pope Building, Prince of Wales Road,

Exeter, Devon, EX4 4QD

1. Abstract

Current lignocellulosic fermentation depend on complex and expensive enzymatic hydrolysis of cellulose and hemicellulose. In order to make bioprocessing of lignocellulose economically competitive this study has selected and engineered a host for degradation of cellulose, into simple monomeric sugars. *Geobacillus thermoglucosidans* DSM 2542 was engineered to individually express four cellulase proteins. The proteins chosen were CelA, CelB Cel6B and Cel9A from *Caldicellulosciruptor bescii, Thermotoga neopolitania and Thermobifida fusca respectively*. CelA, CelB and Cel9A were shown to be active against carboxymethylcellulose (CMC). CelA is of particular interest as the protein has been shown to be multifunctional, thermostable at 55 ºC, has a novel mode of action and able to degrade biomass through surface ablation and cavity formation. The cellulase genes were placed under the control of an inducible promoter (P_{araE}) to facilitate regulation of gene expression with arabinose. This study represents considerable progress in the engineering of *G. thermoglucosidans* for use in a consolidated bioprocessing system.

Keywords: Consolidated Bioprocessing *Geobacillus,* Cellulase, CelA.

Contents

1. Abstract 2

2. Table of Figures 6

List of Tables 8

3. Table of Abbreviations 9

Introduction 11

3.1. Biofuels and the energy industry 11

3.2. Production of replica fuel molecules 12

3.3. Cellulose as a carbon source for biofuel production 13

3.4. Geobacillus*:* a thermophilic host for consolidated bioprocessing 20

3.5. Geobacillus*:* a thermophilic host for consolidated bioprocessing 21

3.6. Project aims 22

Materials and Methods 23

3.7. Revival of lyophilized *Geobacillus* spp. 23

3.8. Growth of *Geobacillus* on a simple fermentation medium 23

3.9. Cryo-storage of *Geobacillus* spp. 23

3.10. Assessing the endogenous cellulolytic ability of *Geobacillus* spp. 24

3.10.1. Measuring cellulose degradation as a function of weight loss 24

3.10.2. Screening of Geobacillus spp. for endoglucanase activity 24

3.10.3. Utilisation of short chain polysaccharides by Geobacillus sp. 25

3.10.4. Supernatant protein purification 25

3.10.5. Endoglucanase activity of crude extracellular protein extract 26

3.11. Investigating the growth of *Geobacillus* spp. in Typtic Soy Broth. 28

3.11.1. Preparation of chemically competent Escherichia coli S17-1 28

3.11.2. Heat shock transformation of E. coli S17-1 30

3.11.3. Conjugal transformation of Geobacillus 30

3.11.4. Screening of Geobacillus for successful transformation. 31

3.11.5. Testing competence of selected Geobacillus sp. to transformation conjugal transfer of pS797. 32

3.12. Expression and activity of cellulase proteins in *Geobacillus*..............................32

3.12.3. Confirming heterologous cellulase protein activity and assessing protein localisation in Geobacillus thermoglucosidans..36

3.12.4. Assessing the in vitro activity of cellulase proteins from G. thermoglucosidans with zymography. 37

3.12.5. The effect of cellulase expression on growth 39

3.13. Exchanging P_{ldhA} for P_{araE} to facilitate control of CelA expression...................39

3.7.1. Induction of CelA expression by arabinose 45

4. Results 46

4.1. Chassis selection for consolidated bioprocessing 46

4.1.1. Rationale 46

4.1.2. Growth of Geobacillus on simple medium..46

4.1.3. Revival of Geobacillus following cryo-storage..46

4.1.4. Growth profile of Geobacillus spp. at 55 °C 61

4.1.5. Transformation of Geobacillus spp. by conjugation 61

4.1.6. Summary of Geobacillus species selection..61

5. Improving cellulose degradation in *Geobacillus thermoglucosidans* DSM2542..... 65

5.1. Rationale 65

5.2. Results 69

5.2.1. Transformation of cellulase genes into Geobacillus.......................................69

5.2.2. Cellulase expression and activity in Geobacillus..72

5.2.3. Zymography of cellulase proteins from G. thermoglucosidans 81

5.2.4. Activity of heterologous cellulase proteins in Geobacillus at different temperatures..81

5.2.5. The effect of cellulase expression on cell growth...81

5.2.6. Exchange of P_{ldhA} with the arabinose inducible P_{araE} promoter by Gibson Assembly 86

5.2.7. Part amplification and Gibson assembly 86

5.2.9. Arabinose utilisation by Geobacillus 93

6. Discussion 97

6.1. Selection of a chassis for consolidated bioprocessing 97

6.2. Progress and further work...100

6.3. Improving cellulase activity by rational mutagenesis.......................................101

6.4. The use of solid state fermentation in consolidated bioprocessing 102

6.5. Ancillary enzymes to improve extracellular cellulase activity 103

6.6. Utilisation of hemicellulose by Geobacillus 104

6.7. Geobacillus and synthetic biology..106

7. Acknowledgements 109

8. References 110

9. Appendices.. 126

2. Table of Figures

Figure 1 Metabolic routes for bioprocessing of lignocellulose to synthetic fuel molecules. 14

Figure 2 A simplified representation of lignocellulose, detailing the classical model of lignocellulosic breakdown. 16

Figure 3 Traditional fermentation of biomass compared to consolidated bioprocessing. 17

Figure 4 Mechanisms of cellulose degradation in bacteria. 18

Figure 5 Standard curve of D-glucose quantified using the 3, 5-dinitrosalysylic acid assay. 27

Figure 6 An outline of the conjugal plasmid used to introduce and express genes into *Geobacillus* sp. 29

Figure 7 Overview of Q5 Site Directed Mutagenesis: C-terminal 6x His tag sequence addition to *CelA*, *CelB*, *Cel6B* and *Cel9A*. 38

Figure 8 An outline of Gibson assembly. 44

Figure 9 Weight loss of selected strains of *Geobacillus* cultured in ASM + 1 % Avicel following incubation for 14 days. 54

Figure 10 Summary of CMC overlay assay results 55

Figure 11 The utilisation of short chain polysaccharides by *Geobacillus* spp. 58

Figure 12 The release of reducing sugars from CMC catalysed by putative cellulase proteins. 60

Figure 13 Growth of *Geobacillus* strains at 55 °C over 24 h 62

Figure 14 Transformation of *Geobacillus* strains with pS797 63

Figure 15 Gel electrophoresis of plasmids purified from *Geobacillus thermoglucosidans* DSM2542 showing transformation with pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A*. 70

Figure 16 Successful recovery of pS797 *CelB* following plasmid loss. 71

Figure 17 SDS PAGE showing the production of CelA and CelB by *Geobacillus thermoglucosaidans* DSM2542. 74

Figure 18 Degradation of CMC by CelA, CelB, Cel6B and Cel9A in culture supernatant from *G. thermoglucosidans* DSM2542 and cell pellet after 24 h. 76

Figure 19 PCR amplification of the 5 cellulase genes for His Tag incorporation by Q5 site directed mutagenesis. 79

Figure 20 Western blotting of His tagged cellulase protein in *Geobacillus thermoglucosidans* DSM2542. 80

Figure 21 Zymography demonstrating active cellulase expression in *G. thermoglucosidans.* 83

Figure 22 CMCase activity of cellulase proteins (CelA, CelB, Cel6B and Cel9A) in *Geobacillus thermoglucosidans* culture supernatant at different temperatures. 84

Figure 23 Growth profile of *Geobacillus thermoglucosidans* DSM2542, and the effect of cellulase (CelA, CelB, Cel6B and Cel9A) expression on lag phase length (h). 88

Figure 24 Promoter exchange of P_{ldhA} with P_{araE}, by Gibson assembly based cloning 89

Figure 25 Alignment of pS797 *P*$_{araE}$ *::CelA* expected sequence and the Sanger sequencing data from the araE CelA_Fwd sequence read. 90
Figure 26 Transformation of pS797 *P$_{araE}$::CelA* purified into *Geobacillus thermoglucosidans* by conjugation. 91
Figure 27 Induction of CelA expression by *P$_{araE}$* induction at different concentrations of arabinose. 95
Figure 28 Degradation of arabinose by *Geobacillus thermoglucosidans* DSM2542 wild type and pS797 *CelA* 96

List of Tables

Table 1 Primers used for site directed mutagenesis of cellulase proteins to introduce C-terminal 6x His tag sequence. 35

Table 2 Sequences of primer used for Gibson assembly. **42**

Table 3 Strains selected and screened for cellulolytic activity. **48**

Table 4 The presence or absence of growth of *Geobacillus* spp. on TSB and TSA and the viability of *Geobacillus* spp. following cryo-storage. **51**

Table 5 Additional characteristics of strains predicted to have CMCase and xylanase activity. **57**

Table 6 Summary of host selection for consolidated bioprocessing of lignocellulosic material. **66**

Table 7 Cellulase genes selected for transformation into *G. thermoglucosidans* DSM2542. **68**

Table 8 Predicted signal peptide cleavage sites, secretion pathway and localisation target as predicted by PSORT, SignalP and TatA. **80**

Table 9 A summary of the host selection criteria and final rank of the *Geobacillus* sp. for consolidated bioprocessing. **104**

Table 10 Candidate enzymes for improving cellulose utilisation by pS797 CelA *Geobacillus thermoglucosidans*. **108**

3. Table of Abbreviations

ACP	Acyl Carrier Protein
ADO	Aldehyde Deformylating Oxygenase
ASM	Ammonium Sulphate Medium
bp	Base Pairs
CAPEX	Capital Expenditure
CBP	Consolidated Bioprocessing
CMC	Carboxymethylcellulose
CRISPR	Clustered Randomly Interspersed Short Palindromic Repeats
DNS	3, 5- dinitrosalycylic acid
EPA	Environmental Protection Agency
FAAE	Fatty Acid Acyl Ester
FACS	Fluorescence Associated Cell Sorting
FAME	Fatty Acid Methyl Ester
GB	*Geobacillus*
GTL	Gas to Liquid
HPLC	High Performance Liquid Chromatography
KLD	Kinase Ligase DpnI enzyme mix (NEB, New England, USA)
LB	Luria Broth

NADPH	Nicotinamide adenine dinucleotide phosphate
O.D	Optical Density
OPEX	Operating Expenditure
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
Phusion Polymerase	*Taq* polymerase fused to a Sso7d domain (NEB)
Q5 polymerase	*Taq* polymerase fused to a Sso7d domain (NEB)
rpm	Revolutions Per Minute
RT- Q- PCR	Reverse Transcriptase- Quantitative- PCR
SDM	Site Directed Mutagenesis
SDS- PAGE	Sodium Dodecyl Sulphate- Polyacrylamide Gel Electrophoresis
TBE	Tris Bis EDTA gel electrophoresis buffer
TSA/ K	Tryptic Soy Agar/ Containing 12.5 µg ml^{-1} of Kanamycin
TSB/ K	Tryptic Soy Broth/ Containing 12.5 µg ml^{-1} of Kanamycin
TTE	Tryptone + Trace Elements

Introduction

3.1. Biofuels and the energy industry

Reliance on finite fossil fuel resources is a growing concern. The geopolitical situation surrounding oil and gas is fragile, especially in the Middle-East. Fossil fuels are non-renewable and finite, making the development of sustainable fuel sources essential. Liquid transport fuels, such as petrol, diesel and kerosene make up a large proportion of the fossil fuel demand (Demirbas *et al.* 2007) Gas to Liquid technology (GTL) (Patent number WO20140958 A1, (Hoek *et al*. 2014) provides a way of generating liquid fuel through Fischer-Tropsch synthesis of natural gas to liquid fuel. However GTL technology only strengthens our dependence on fossil fuel reserves. Fossil fuels have a carbon debt related to them as they are non-renewable and release CO_2 when burnt. The carbon debt associated with fossil fuels is an issue for the fuel industry as following the introduction of the Energy Policy Act 2005 (United States Congress 2005), the Clean Air Act 2006 (EPA 2006a), the Renewable Fuels Standards 2006 (EPA 2006b) and the Energy Independence and Security Act 2007 (EPA 2007) companies involved in the production of transport fuels are required to produce renewable liquid fuels. Governmental incentives such as Renewable Identification Number (RIN) credits (EPA 2013) and cellulosic waiver credits (Mcphail *et al*. 2011) encourage carbon offset, making the development of carbon neutral fuels financially viable Given these factors developing replacements to fossil fuels is environmentally important and possibly commercially advantageous.

To date, biofuel generation has been mostly limited to the production of bioethanol and biodiesel. Bioethanol and diesel are classified depending on the whether the feedstock is primarily from food crops (1st generation) or other sources such as lignocellulose, non-food crops or waste (2nd generation). Yeasts are used to ferment plant materials, such as sugarcane bagasse (the residual waste left after soluble sugar extraction) (Rudolf *et al*. 2008)to ethanol. The ethanol produced is then blended with petroleum to produce a useable fuel. Brazilian bioethanol is produced at a price competitive with gasoline (Goldemberg *et al*. 2004).

The main hurdle to replacing petroleum based with bioethanol is the so-called “blend wall”; this is the maximum percentage of ethanol that can be mixed with petroleum fuel before the fuel properties are affected. In the US the maximum percentage of ethanol that can be blended with petrol is 10 % (EPA 2006b).

Another option for liquid fuel replacement is biodiesel. Biodiesel is composed of Fatty Acid Methyl/Acyl Esters (FAME/FAAE), produced by the trans-esterification of glycerol. Biodiesel has numerous advantages over traditional fuels. The lifecycle emissions of biodiesel (NOx, CO_2 and methane) are reduced by 41% if no land change is required for the feedstock growth process, compared to current production levels (NREL 2008). Furthermore, it acts as a lubricant making it compatible with current diesel engines (NREL 2008).

Both bioethanol and biodiesel raises questions about infrastructure, land use change and the feedstock market, particularly if managed incorrectly (Fargione *et al*. 2008). Land use change is of major concern as research has shown that if natural environments such as rainforests, peat lands, savannahs or grasslands are converted for biofuel use, a carbon debt of 17-420 times that of the annual CO_2 reduction provided by the biofuel is created (Fargione *et al*. 2008). Land use change also conflicts with global food production leading to a trade-off between food and fuel.

3.2. Production of replica fuel molecules

Through further refinement or mass market engine modification, biodiesel or bioethanol could replace petroleum based fuels in the road transport market. However changes to the established infrastructure would be required due to the different properties of current biofuels (cold flow, cloud point, volatility, and octane: cetane requirements). This incompatibility means the energy industry is reluctant to convert its infrastructure. The mass market change to biodiesel or ethanol is a goal that would be a gradual and costly to realise.

The biosynthetic production of fuel molecules which, unlike ethanol and biodiesel, are structurally identical to current petroleum based fuels is desirable because of the reduction in downstream processes. A recent study allow a fuel molecule’s chain length to be tailored, (Howard *et al*. 2013) thus removing the need for hydrocracking.

Today’s fuels are made of a complex mix of hydrocarbons, such as alkanes, alkenes and aromatic compounds. These hydrocarbons can be synthesised in nature by a number of different organisms for various purposes, for example the green woodhoopoe (*Phoeniculus purpureus*) secretes hydrocarbons for waterproofing (Burger *et al*. 2004), termites use alkanes as signaling molecules (*Reticulitermes virginicus*) (Howard *et al*. 1982) plants secrete a waxy cuticle which contains large

amounts of hydrocarbons (Rowland *et al*. 2006; Aarts 1995Bernard *et al*. 2012Perera *et al*. 2010von Wettstein-Knowles 2007). Additionally a number of bacterial species are able to synthesis alkenes (Albro & Dittmer 1969; Beller *et al*. 2010; Rude *et al*. 2011).The cyanobacterial pathway for alkane biosynthesis has attracted the most interest specifically the pathway from *Nostoc punctiforme* PCC73102 (Schirmer *et al*. 2010). The cyanobacterial pathway produces C_{15} and C_{17} odd chain alkanes/alkenes through two key steps. Firstly acyl-ACP fatty acids are reduced to aldehyde by an NADPH dependent fatty acyl-ACP reductase.Secondly the aldehyde is then converted to an alkane/alkene by an aldehyde-deformylating oxygenase (ADO). It is possible that a simple enzymatic pathway to alkanes, such as that discussed in Howard *et al*, could be inserted into a industrially relevant host (Figure 1) (Howard *et al*. 2013) that has been engineered to utilise a renewable carbon source. This would allow the production of drop in biofuels from lignocellulosic waste material. An overview of other route to biofuels from lignocellulose is also shown in Figure 1.

3.3. Cellulose as a carbon source for biofuel production

Whilst it has been shown that “drop-in” fuel molecules can be produced in bacterial systems (Akhtar *et al*. 2013; Howard *et al*. 2013) (Figure 1), it is important that an appropriate feed-stock can be utilised by the chosen microbial host. Cellulose is the most abundant biorenewable carbon source on the planet and its utilisation as a carbon source has been a goal of the research community for many years. However, cellulosic material is very complex, recalcitrant to degradation and varies greatly in composition, making it very difficult to degrade (Gomez del Pulgar & Saadeddin 2014; Hemsworth *et al*. 2014). The complex structure of cellulose is shown in Figure 2. Cellulose is currently used as a feedstock for the production of cellulosic ethanol. Typically, lignocellulosic material is broken down using complex treatment processes which use harsh conditions, such as steam explosion, acid pretreatment or mechanical breakdown to break the lignin and partially hydrolyse the cellulose.

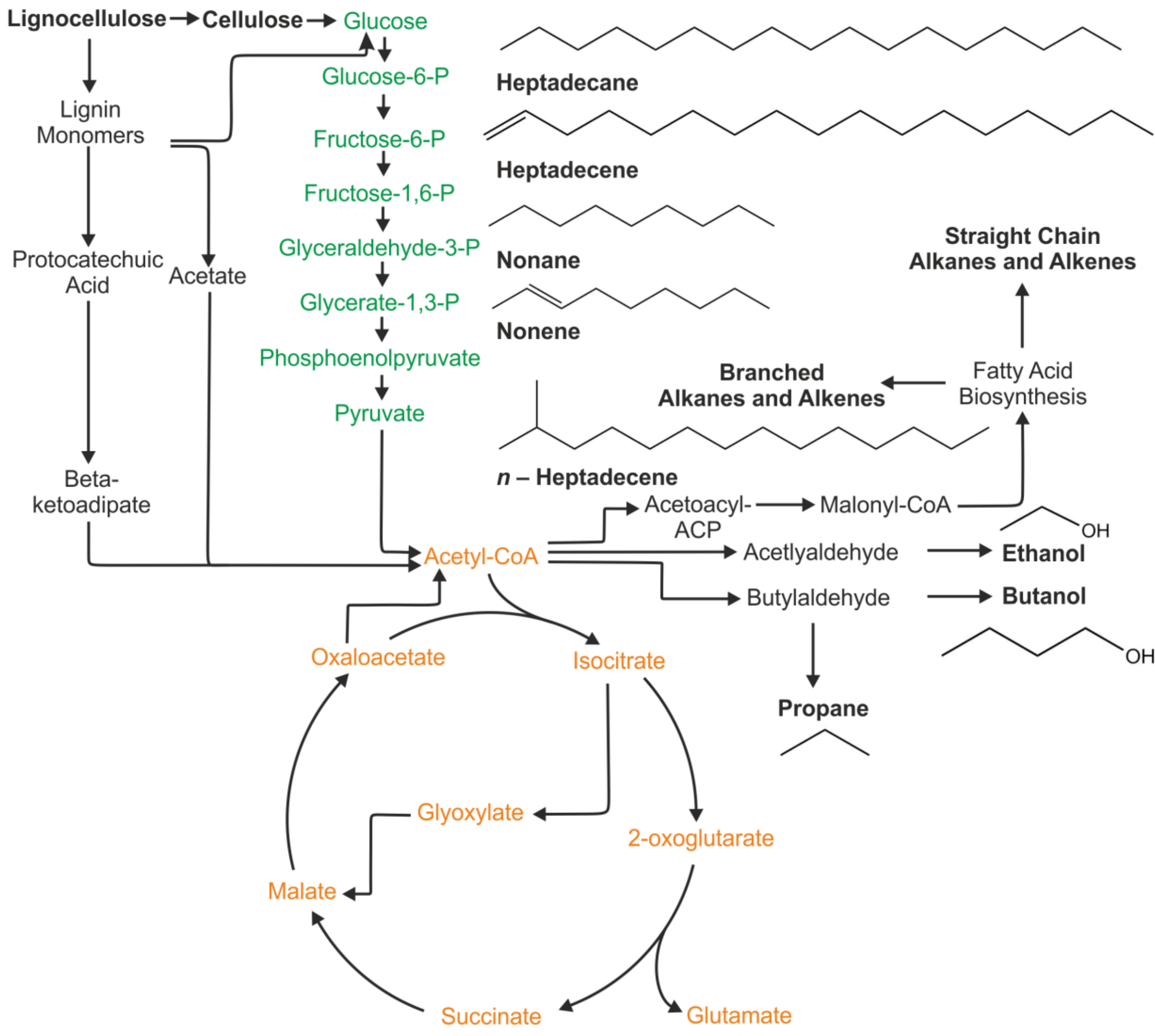


Figure 1 Metabolic routes for bioprocessing of lignocellulose to synthetic fuel molecules.

If lignocellulose can be broken down by cellulase enzymes, to glucose monomers these can be feed into glycolysis (Green). This yields Acetyl-CoA, a precursor to many secondary metabolites, including biofuels. The lignin can be fractionated into monomers, these can then be hydrolysed to protocatechuic acid (PHA) as demonstrated in *Pseudomonas putida* (Linger *et al* 2014). A variety of fuel molecules can be produced, including long, diesel range, straight and branched chain alkanes and alkenes (Howard *et al*. 2013), short, petrol-range straight chain alkanes and alkenes (Akhtar *et al.* 2013), and other fuels such as propane (Kallio *et al.* 2014) and alchohols, such as butanol and ethanol, which can be blended with traditional petroleum fuel (Atsumi *et al*. 2008, Fischer *et al* 2008). A single organism engineered to carry out the catabolic (lignocelluose degradation) and anabolic (fuel molecule synthesis) steps would allow a single fermentation step, for biofuel synthesis from waste material.

This mix of complex sugars is then digested by enzymes to simple sugars such as glucose and cellobiose. The hydrolysate from these treatments can then be used as a feedstock for fermentation, by yeast (such as *Saccharomyces cerevisiae*), to produce ethanol. Multi-step complex processes can be expensive. If these pretreatment, hydrolysis and fermentation steps can be combined into a single fermentation step then this should reduce the CAPEX and OPEX of the conversion of lignocellulose to fuel. Bacteria have been found which produce enzymes able to hydrolyse lignin (Ramachandra *et al.* 1988, Ahmad 2010) and cellulose (Chung *et al.* 2014), and to produce drop in fuel molecules (Howard *et al.* 2013). Consolidated Bioprocessing (Figure 3) is the use of a single fermentation step to convert a feedstock into the desired product; in this case, lignocellulose to replica hydrocarbons.

There are 3 main mechanisms of for degrading cellulose (Figure 4). The first, isfound mainly in *Clostridia* sp. are called cellulosomes. These are large protein complexes of membrane associated, extracellular enzymes for cellulose degradation (Lamed *et al.* 1983a). Another mechanism is the secretion of free cellulases, which are released from the cell (Mingardon *et al.* 2011, Wilson *et al.* 1993). These can either be simple, single domain proteins such as exoglucanases, exoglucanase and β-glucosidases or large, complex, multidomain cellulosome like proteins, such as those produced by *C. bescii* (Brunecky *et al.* 2013), which can contain endo and exo acting glycosyl-hydrolase domains.

Historically fungi have been the main source of cellulolytic enzymes (Mandels & Reese 1957; Lee 1997). *Trichoderma Reesei,* a hyphal strain of fungi has been developed through genetic engineering (Peterson & Nevalainen 2012) and

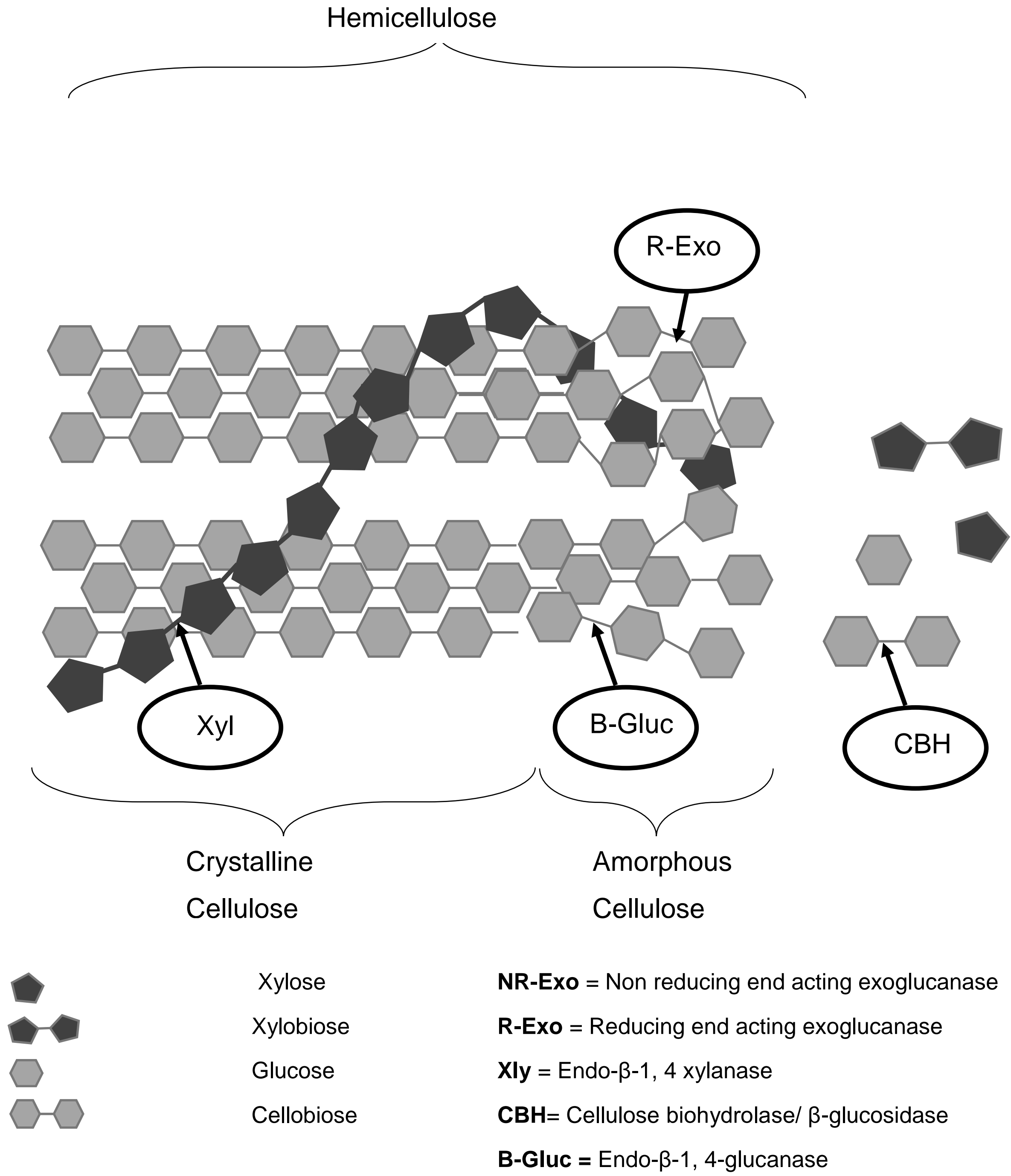


Figure 2 A simplified representation of lignocellulose, detailing the classical model of lignocellulosic breakdown.
Lignocellulose is a highly ordered and complex material. This figure shows the basic structure of cellulose fibrils, cross linked by hemicellulose (xylose), and the main enzymes required for degradation of this complex material. Cellulose is made more recalcitrant to degradation by branched hemicellulose structures and lignin. The most important enzymes in lignocellulose degradation are endo- and exoglucanases and endo- and exoxylanases.

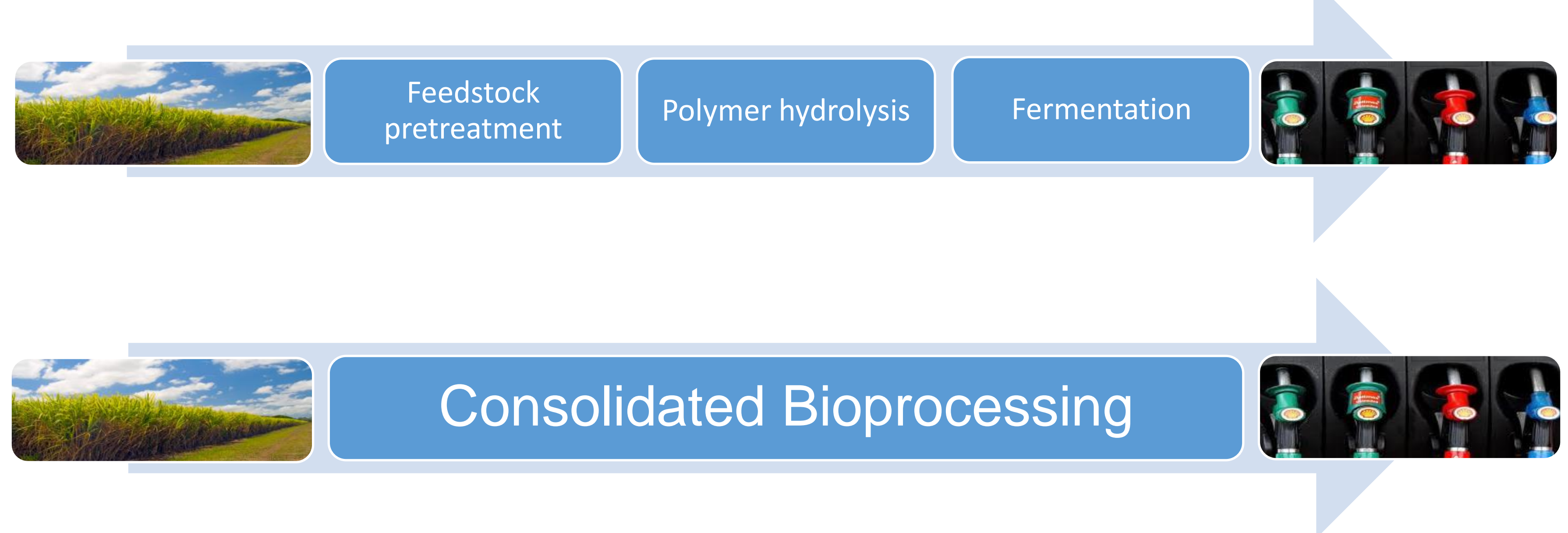


Figure 3 Traditional fermentation of biomass compared to consolidated bioprocessing.

Consolidated bioprocessing aims to remove the complexity from current fermentation workflows. Consolidated bioprocessing reduces capital and operating costs. This research aims to engineer an organism that is able to degrade lignocellulosic biomass and synthesise biofuels.

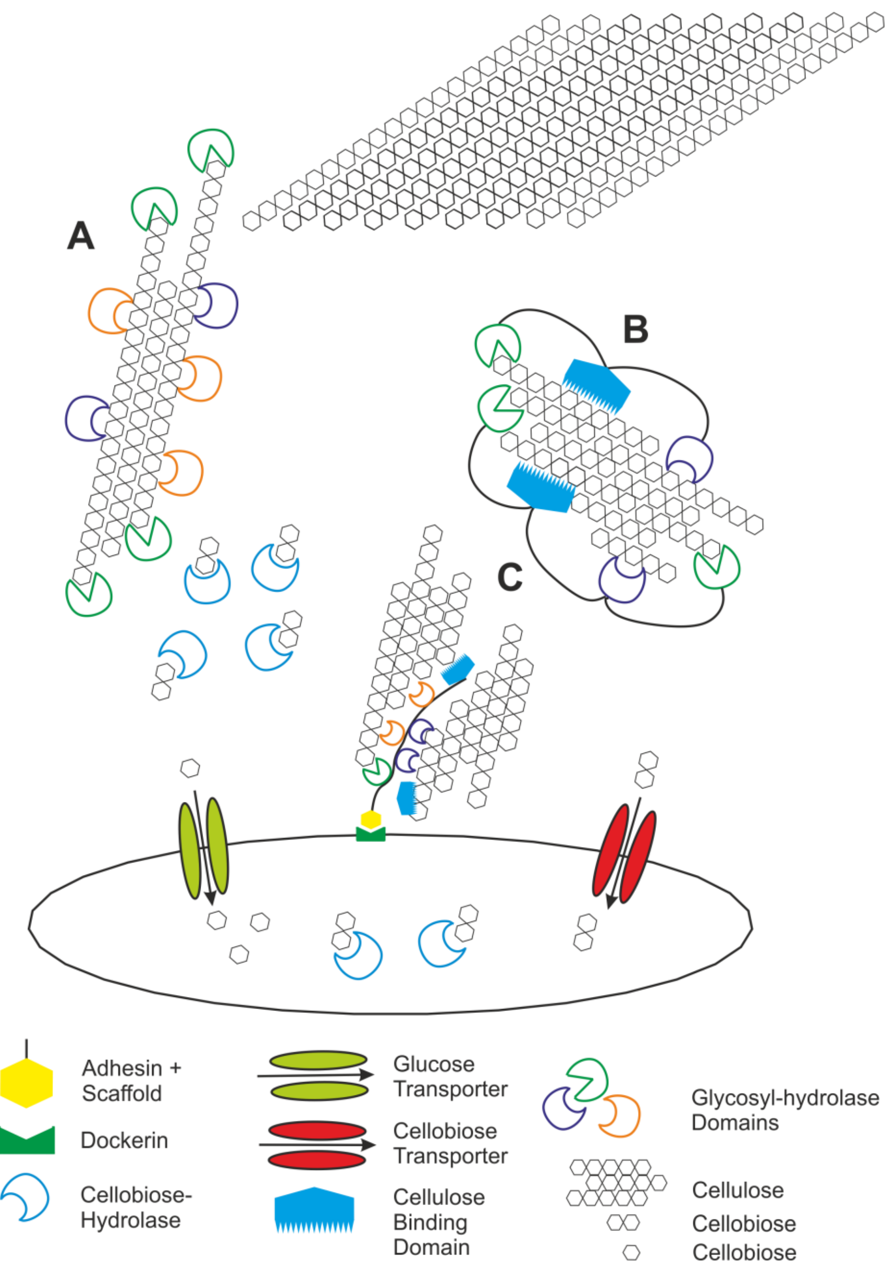


Figure 4 Mechanisms of cellulose degradation in bacteria.

Bacteria are able to degrade cellulose through a number of different mechanisms. Some bacteria secrete free cellulose enzymes (A). Others, such as *Caldicellulosiruptor bescii,* (*C. bescii*), secrete complex multiedomain, multifunctional cellulose enzymes, these are not membrane bound but the domains are linked by long linker arms (B) (Brunecky *et al.* 2013). Some bacterial species, particularly *Clostridia* sp. have highly complex, membrane bound, protein complexes called cellulosomes (C) (Lamed *et al.* 1983b). Both true cellulosomes and free enzyme, cellulose like complexes contain cellulose binding domains. These help to improve cellulose hydrolysis by stabilising the glycosyl-hydrolase interaction with the substrate. Once hydrolysed the cellulose fibres the simple sugars, such as glucose and cellobiose can be transported into the cell, where they are feed into glycolysis (Mingardon *et al.* 2011).

fermentation optimisation (Pakula et al. 2005), but given the hyphal morphology of T. reesei traditional liquid fermentation is difficult (Lee 1997) because the mixing of the medium disrupts the hyphal structure. This makes T. reesei unsuitable for efficient consolidated bioprocessing.Prokaryotic systems that are capable of cellulose degradation are very appealing as they are less complex than eukaryotic systems, making it more feasible to manipulate the endogenous system to produce the desired product, allowing carbon neutral production. Many bacterial species have the ability to degrade cellulose, including *Clostridium thermocellum* (Lamed *et al*. 1983; Dykstra *et al*. 2013)*, Clostridium stercorarium* (Riedel *et al*. 1998)*, Thermobifida fusca* (Irwin *et al*. 1993; Zhang *et al*. 1995; Lykidis *et al*. 2007)*, Thermobifida neopolitania* (Bok *et al*. 1998; Dakhova *et al*. 1993) *Caldicellulosiruptor saccharolyticus* (van de Werken *et al*. 2008; Vanfossen *et al*. 2009; VanFossen *et al*. 2011) and *Caldicellulosiruptor bescii* (Brunecky *et al*. 2013; Kanafusa-Shinkai *et al*. 2013). *C. bescii* is of particular interest as it has high cellulolytic activity; it has been developed to be genetically tractable chassis (Chung *et al*. 2012). Several species of *Geobacillus* have been demonstrated to have cellulolytic capabilities (Zambare *et al*. 2011; Rastogi *et al*. 2010; Assareh *et al*. 2012; Tai *et al*. 2004; Zeigler 2014). Both *Geobacillus thermoglucosidans* DSM 2542 (formerly known as *Geobacillus thermoglucosidasius* (Coorevits *et al*. 2012))and *C. bescii* have successfully been engineered to produce ethanol (Zambare *et al*. 2011; Cripps *et al*. 2009; Chung *et al*. 2014).

Thermophilic bacteria offer numerous advantages for consolidated bioprocessing (Figure 3). Feedstock cooling costs are reduced (Cihan *et al*. 2011). The higher temperatures facilitate the extraction of volatile compounds (Turner *et al*. 2007). Due to the toxicity of the desired products, in this case short chain alkanes, their removal through evaporation is highly desirable as this reduces any inhibitory effects (Wiegel 1980). It is proposed that by using thermophilic bacteria for industrial fermentation, contamination by mesophilic bacteria and bacteriophages will be lower than if a mesophillic culturing environment is used. Although thermophilic bacteriophages do exist (Saunders & Campbell 1966) it has been shown that a greater proportion of thermophilic bacteria possess adaptive viral immune responses (CRISPR based immunity), when compared to mesophillic bacteria (46%) (Weinberger *et al*. 2012) making thermophilic bacteria less susceptible to phage contamination. Whilst the existence of *Geobacillus* spp. specific phages has been demonstrated (Saunders & Campbell 1966; Marks & Hamilton 2014; Liu & Zhang 2008; Wang & Zhang 2010),

current research suggests that the mutation rate is sevenfold lower in thermophilic environments.

3.4. Geobacillus*:* a thermophilic host for consolidated bioprocessing

The economic gains available from commercial scale production of lignocellulosic biofuels could be much greater if operating costs (OPEX) and capital cost (CAPEX) are reduced as much as possible. One way to achieve this is the engineering of an organism that would allow for all pre-treatment, polysaccharide hydrolysis and product synthesis to be performed in a single fermentation step. The generation of this consolidated bioprocessing chassis is the main aim of this body of research.

Geobacillus spp*.* are rod shaped, thermophilic, aerobic or facultative anaerobic endospore-forming bacteria (Zeigler 2014). *Geobacillus* spp. have been isolated from a wide range of environments but their role in composting (Zeigler 2014) makes them of interest to this study. *Geobacillus* spp. have an optimal growth temperature of 45-70 °C (Nazina *et al*., 2001) and remain stable at ambient temperatures.

In recent years, research into *Geobacillus* has increased and important developments have been made. *G. thermoglucosidans* has seen improvements made to its cellulolytic capacity through the expression of the thermophillic cellulase Cel5A from *Thermotoga maritima* (Bartosiak-Jentys *et al*. 2013). There are several reported methods for the genetic transformation of *Geobacillus.* These include electroporation and conjugal transfer (Narumi *et al*. 1992; Cripps *et al*. 2009). Conjugal transfer was selected for the transformation of *Geobacillus.* There are several advantages to using conjugal transfer over electroporation. Because the plasmid is directly transferred to the conjugal recipient via a pilus there is no “free” DNA outside of the cell that can be degraded by extracellular endonucleases, which in turn affects the efficiency of transformation. Furthermore, there are fewer variables that require optimisation with conjugation compared to electroporation.

To date 10 *Geobacillus* genomes have been sequenced giving us a wealth of bioinformatics data from which we can predict genes and operons of interest (Zeigler 2014; Studholme 2014)**.** Approximately 85 % of core genome shared of the *Geobacillus* genus is shared with *Bacillus subtillis* 168 (Zeigler 2014).

3.5. Geobacillus*:* a thermophilic host for consolidated bioprocessing

The economic gains available from commercial scale production of lignocellulosic biofuels could be much greater if operating costs (OPEX) and capital cost (CAPEX) are reduced as much as possible. One way to achieve this is the engineering of an organism that would allow for all pre-treatment, polysaccharide hydrolysis and product synthesis to be performed in a single fermentation step. The generation of this consolidated bioprocessing chassis is the main aim of this body of research.

Geobacillus spp*.* are rod shaped, thermophilic, aerobic or facultative anaerobic endospore-forming bacteria (Zeigler 2014). *Geobacillus* spp. have been isolated from a wide range of environments but their role in composting (Zeigler 2014) makes them of interest to this study. *Geobacillus* spp. have an optimal growth temperature of 45-70 °C (Nazina *et al*., 2001) and remain stable at ambient temperatures.

In recent years, research into *Geobacillus* has increased and important developments have been made. *Geobacillus thermoglucosidans* (*G. thermoglucosidans)*, has seen improvements made to its cellulolytic capacity through the expression of the thermophillic cellulase Cel5A from *Thermotoga maritima* (Bartosiak-Jentys *et al*. 2013). There are several reported methods for the genetic transformation of *Geobacillus.* These include electroporation and conjugal transfer (Narumi *et al*. 1992; Cripps *et al*. 2009). Conjugal transfer was selected for the transformation of *Geobacillus.* There are several advantages to using conjugal transfer over electroporation. Because the plasmid is directly transferred to the conjugal recipient via a pilus there is no “free” DNA outside of the cell that can be degraded by extracellular endonucleases, which in turn affects the efficiency of transformation. Furthermore, there are fewer variables that require optimisation with conjugation compared to electroporation.

To date 10 *Geobacillus* genomes have been sequenced giving us a wealth of bioinformatics data from which we can predict genes and operons of interest (Zeigler 2014; Studholme 2014)**.** Approximately 85 % of core genome shared of the *Geobacillus* genus is shared with *Bacillus subtillis* 168 (Zeigler 2014).

It is hypothesised that this will make the transfer of knowledge easier between the two genera. It is the *pan-genome* (Zeigler 2014; Tettelin *et al*. 2005) that makes *Geobacillus* of even greater interest. *Geobacillus* is a highly diverse genus of bacteria; much of the pan-genomic variation is observed in carbon source utilisation with many species possessing genes for complex polysaccharide utilisation (Zeigler 2014). The intraspecies diversity can be screened for characteristics that would make *Geobacillus* an ideal host for consolidated bioprocessing (CBP) (overview in Figure 3).

Geobacillus spp. have numerous genes and operons that have putative function in the degradation of complex plant polysaccharides (Zeigler 2014), such as a 9.4 kb galactan utilisation cluster (*ganREFGBA*) (Tabachnikov 2013). Genes involved in the uptake of simple monosaccharides and disaccharides, produced from cellulose, arabinose and xylan degradation (Lai & Ingram 1993; Shulami *et al*. 2011), have also been identified.

Geobacillus has been engineered for the production of ethanol from biomass, by a UK based company called TMO renewables. Olson *et al.* (Olson *et al*. 2012) suggested that the modified *Geobacillus thermoglucosidans* DSM2542 of Cripps *et al.* (Cripps *et al*. 2009) was the ethanol production system with the highest titre (g l^{-1}) and rate (g l^{-1} $hour^{-1}$) of currently available thermophilic ethanol platforms, setting an important industrial precedent for the use of *Geobacillus* in an industrial process.

3.6. Project aims

The aim of this research will be to screen all the public available species of *Geobacillus* for relevant characteristics: thermophilic growth, C5 and C6 sugar utilisation, cellulolytic capabilities, and ability to be transformed. Any inherent cellulolytic activity will be assessed and *Geobacillus* will then be engineered to improve any endogenous cellulolytic pathways present to improve growth on cellulosic material, by increasing the available carbon. This will be achieved through the expression of thermostable cellulase genes. Whilst lignocellulose is a complex combination of cellulose and hemicellulose (Figure 2) the main aim of this project is cellulose degradation as cellulose comprises the largest proportion of lignocellulosic biomass. The selected host chassis can later be engineered to produce drop in

biofuels using a pathway similar to those discussed previously (Howard *et al.* 2013; Kallio *et al.* 2014; Akhtar *et al.* 2013).

Materials and Methods

3.7. Revival of lyophilized *Geobacillus* spp.

Strains were obtained from 4 separate culture collections. The cultures obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), American Type Culture Collection (ATCC) and the National Collection of Industrial, Marine and Food Bacteria (NCIMB) were freeze dried cultures, whilst the samples shipped from Bacillus Genetic Stock Centre (BGSC) were shipped on tryptone blood agar base (TBAB) plates. Freeze dried samples were hydrated according to the instructions provided by each respective culture collection. Once fully hydrated, 100 µl of cells was spread plated onto the recommended agar and onto Tryptic Soy Agar (TSA) (15 g l^{-1} Tryptone, 5 g l^{-1} Peptone, 5 g l^{-1} sodium chloride, 15 g l^{-1} agar (plates only), pH adjusted to pH 7.3). A further 100 µl was also inoculated into 5 ml of Tryptic Soy Broth (TSB), in a 50 ml Falcon tube. The strains received from BGSC were on plates and so a single colony was picked from the plate and inoculated into the recommended liquid medium and TSB. All strains were then incubated at 55 °C with shaking at 200 rpm for 16-40 h (depending on the extent of growth after 24 h some strains were left to recover for 48 h). Strains that grew on the TSB were streaked out and incubated for 16 h at 55 °C.

3.8. Growth of *Geobacillus* on a simple fermentation medium

Geobacillus species were inoculated into 5 ml of TSB and 5 ml of Nutrient Broth (Beef Extract 3 g l^{-1}, Peptone 5 g l^{-1} pH 6.8) medium using a single colony and incubated for 18 h at 55 °C in a shaking incubator, with shaking at 200 rpm. Cultures were then scored for growth by the appearance of colonies after 24 and 48 h.

3.9. Cryo-storage of *Geobacillus* spp.

After growth on simple medium was achieved 5 ml of TSB inoculated with a single colony from a TSA plate were incubated at 55 °C for 16 h. Following this 1 ml of the stationary phase culture was inoculated into 50 ml of TSB medium, which was incubated at the same conditions as described above for 5-6 h. The subsequent culture was then used to make 5x 15 % glycerol cryo-stocks. These were flash frozen in liquid nitrogen and stored at -80 °C. It was important to test the viability of all

stocked strains to recover from the storage process. To test viability cryo-stocked strains were streaked onto TSA plates and incubated at 55 °C for 24 h.

3.10. Assessing the endogenous cellulolytic ability of *Geobacillus* spp.

3.10.1. *Measuring cellulose degradation as a function of weight loss*

All 46 strains of *Geobacillus* were inoculated into 5 ml of TSB medium and grown at 55 °C for 16 h. The overnight culture of *Geobacillus* was then used to inoculate 50 ml of Ammonium Sulphate Medium (ASM) (8 mM citric acid, 5 mM magnesium sulphate, 20 mM sodium di-hydrogen orthophosphate,10 mM potassium sulphate, 25 mM ammonium sulphate, 80 µM calcium chloride, 1.65 µM sodium molybdate, 0.1 % w/v yeast extract, 5 ml l^{-1} Trace Element Solution, 1.44 g l^{-1} zinc sulphate $7H_2O$, 0.56 g l^{-1} cobalt sulphate $6H_2O$, 0.25 g l^{-1} copper sulphate $5H_2O$, 5.56 g l^{-1} ferrous sulphate $6H_2O$, 1.69 g l^{-1} manganese sulphate, boric acid 0.08 g l^{-1}, 12 M sulphuric acid 0.5 % v/v, pH adjusted to pH 7.0, add 40 mM Bis- Tris and 1 % w/v glucose/ cellobiose/ Avicel (microcrystalline cellulose with an average particle size of 50 µm)after autoclaving) Medium supplemented with 1 % Avicel. The 50 ml ASM Glucose/ Cellobiose/ Avicel cultures were grown for 2 weeks and the weight was recorded every 3 – 4 days. The percentage mass loss of the liquid culture was measured compared to a negative control, containing no inoculum. The experiment was performed in triplicate.

3.10.2. *Screening of* Geobacillus spp. *for endoglucanase activity*

Each strain was streak plated from cryo-stock onto TSA. Glycerol stocks were kept in liquid nitrogen during plating. The plates were incubated at 55 °C for 16 h. A single colony was returned and streaked onto TB-ASM agar (ASM, 1% w/v tryptone, 1% w/v cellobiose, 15 % w/v agar). The plates were then incubated for 24 h at 55 °C. Following incubation, 10 ml of 0.8 % agar supplemented with 0.25 % medium viscosity carboxymethylcellulose (CMC) (Teather & Wood 1982) or 1 % beachwood xylan (Wood & Weisz 1987; Teather & Wood 1982) was overlaid onto plates, incubated at 55 °C for 24 h. Approximately 10 ml of 1 % 3, 3’-([1,1´-biphenyl]-4, 4´diyl) bis (4- aminonapthalene -1- sulphonic acid) sodium salt (Congo red) was added to plates following the required incubation and left for 20 min. The stain was then removed and the plates were destained for 20 min with 1 M sodium chloride solution. Following staining, black and white photos of the plate were taken with the myECL imager (ThermoFischer Scientific).

3.10.3. *Utilisation of short chain polysaccharides by* Geobacillus sp.

The 44 selected *Geobacillus* strains were grown on TTE (ASM trace elements supplemented with 1.5 % tryptone and) 1 % w/v cellobiose, 1 % w/v cellotriose, 1 % w/v cellotetraose and 0.43 % w/v cellohexose. Overnight TSB cultures were inoculated from TSA plates with a single colony, grown at 55 °C, with shaking at 200 rpm for 16 h. 50 µl aliquots were transferred to a 96 well plate. Cells were pelleted at 4000 g for 3 min. The resulting pellet was washed with 200 µl of phosphate buffered saline (PBS; sodium chloride 9 g l^{-1}, sodium phosphate dibasic 0.795 g l^{-1}, potassium phosphate monobasic 0.144 g l^{-1}, pH 7.4). The pellet was re-suspended in 100 µl of the TTE + polymer media. The 96 well plate was sealed inside a plastic zip-lock bag and placed at 55 °C shaking at 200 rpm for 48 h. After incubation, the culture from each well was transferred into 1.5ml microfuge tubes. The samples were centrifuged at 15,000 rpm for 10 min to completely pellet the cells. The supernatant was transferred to a 2 ml glass vial with 100 µl inserts for HPLC analysis on an Agilent Hi-Flex Na column with a pore size of 10 µm. A 20 µl sample was injected into the HPLC (Dionex 3000 HPLC with a Shodex RI detector). The mobile phase was deionised water with a flow rate of 0.2 ml min^{-1}. The temperature of the column was 85 °C.

3.10.4. *Supernatant protein purification*

A single colony was inoculated into 5 ml of TSB from TSA and incubated at 55 °C with shaking at 200 rpm, for 16 h. A 1 ml aliquot of this culture was then inoculated into 50 ml of TB-ASM containing 0.25 % CMC (w/v) as a carbon source. After 24 h incubation at the same conditions as before a 1 ml sample was transferred into a 1.5 ml microcentrifuge tube which was centrifuged (15,000 rpm, 5 min, room temperature). A 250 µl sample of supernatant was then transferred to a 2 ml glass vial and this was used for HPLC.

The remaining culture was placed on ice for 10 min, centrifuged at 4000 x g for 10 min and a 10 ml of supernatant sample transferred to a 50 ml centrifuge tube. 40 ml of cold acetone (-20 °C) was added. The lysate and acetone mixture was transferred to -20 °C for 1 h. After incubation the samples were centrifuged (4,000 x g for 10 min). The supernatant was removed and the pellet was placed at room temperature to allow the residual acetone to evaporate. Once dry the remaining pellet was dissolved in 2 ml of PBS and the resulting protein solution stored at 4 °C.

3.10.5. *Endoglucanase activity of crude extracellular protein extract*

Purified protein was incubated with 250 µl of 0.25 % CMC (w/v) in 50 mM citrate buffer (pH 5.8) (Bartosiak-Jentys *et al*. 2013) at a 1:1 volume:volume ratio, vortexed for 5 s and a 9 µl sample was taken and transferred directly into a 96 well PCR plate for quantification of reducing sugars. The remaining solution was incubated horizontally in a 55 °C incubator, shaking at 200 rpm for 6 h. After incubation another 9 µl sample was taken and reducing sugars were quantified.

In order to quantify reducing sugar concentration a colourimetric DNS assay was used (Bartosiak-Jentys *et al*. 2013; Zambare *et al*. 2011; Kanafusa-Shinkai *et al*. 2013; Cava *et al*. 2008; Assareh *et al*. 2012). First, 9 µl of sample was mixed with 171 µl of the 3,5-dinitrosalycylic acid reagent (3,5-dinitrosalysylic acid 10 g l^{-1}, sodium potassium tartrate tetrahydrate 30 g l^{-1}, sodium hydroxide 16 g l^{-1}) in a 96 well PCR plate. The plate was heated in a thermo-cycler at 100 °C for 1 minute and cooled to 20 °C for 2 min. The 180 µl sample was transferred to a black flat bottomed 96 well plate (Corning Costar) and the absorbance was then measure at λ 540 nm and λ 580 nm (TECAN M200 infinite plate reader). Reducing sugar concentration was calculated based on a D-glucose standard curve (Figure 5).

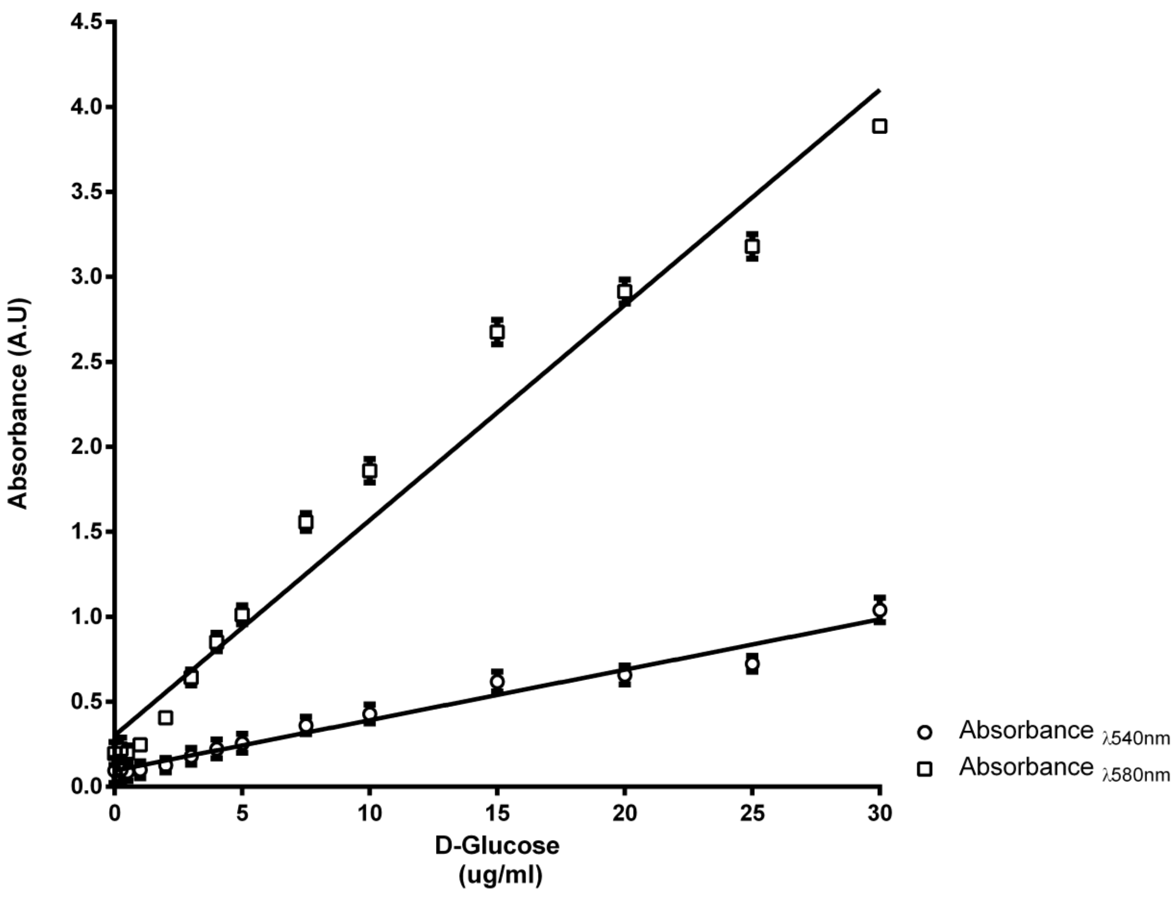


Figure 5 Standard curve of D-glucose quantified using the 3, 5-dinitrosalysylic acid assay.

The graph above shows the absorbance of different D-glucose standards at λ 540 nm and λ 580 nm following a 1 minute incubation of 9 µl of glucose solution with 171 µl of 3, 5-dinitrosalycylic acid. Absorbance was measured using a TECAN M200 plate reader. Readings were taken in triplicate. The mean of these three readings is plotted. The standard curve will be used to calculate the sugar released from cellulose following incubation with cellulase proteins in future experiments. Readings were performed in triplicate. Error bars represent standard deviation. (Line fit using linear regression. λ540 R^2 = 0.9411, λ580 = 0.9785).

3.11. Investigating the growth of *Geobacillus* spp. in Typtic Soy Broth.

Single colonies from TSA plates were used to inoculate 5 ml of TSB medium and incubated for 16 h at 55 °C with shaking at 200 rpm. Following this, 1 ml of 16 h cultutre was inoculated into 50 ml of TSB medium and grown for 24 h under the growth conditions. Two 250 µl samples were aliquoted into a flat black 96 well plate (Corning Costar) and the $O.D_{600}$ was measured using a TECAN M200 infinite plate reader.

3.11.1. *Preparation of chemically competent* Escherichia coli *S17-1*

The conjugal donor strain *Escherichia coli* S17-1 (*genotype TpR SmR recA, thi, pro, hsdR-M+RP4: 2-Tc: Mu: Km Tn7 λpir*) was cultured and used to make chemically competent cells (protocol modified from (Hanahan 1985)) A single colony of *E. coli* S17-1, from a LB plate (provided by Dr. Chloe Singleton at the University of Exeter) was used to inoculate 5 ml of LB medium. *E. coli* S17-1 was grown overnight at 37 °C in a shaking incubator at 225 rpm. The resulting culture was inoculated into 40 ml of LB in a 250 ml flask, and incubated under the same conditions to an optical density ($O.D_{600}$) of ~0.6. The culture was then transferred to a 50 ml centrifuge tube and centrifuged at 4000 g at 4 °C for 5 min. The supernatant was removed and the remaining pellet re-suspended in 8 ml of TF-1 (potassium chloride 7.4 g l^{-1}, calcium chloride $2H_2O$ 5 g l^{-1}, glycerol 150 g l^{-1} and 3 % 1 M potassium acetate pH 7.5 v/v, pH to 6.4 with acetic acid and add 50 ml of 1 M manganese chloride $4H_2O$). The suspended *E. coli* were placed on ice for 15 min and then centrifuged as above. The supernatant was removed and resuspended in 4 ml of TF-2 (potassium chloride 0.74 g l^{-1}, calcium chloride $2H_2O$ 11 g l^{-1}, glycerol 150 g l^{-1}, 20 ml of MOPS pH 6.8 pH adjusted with 5 M potassium hydroxide). These competent cells were stored at -80 °C.

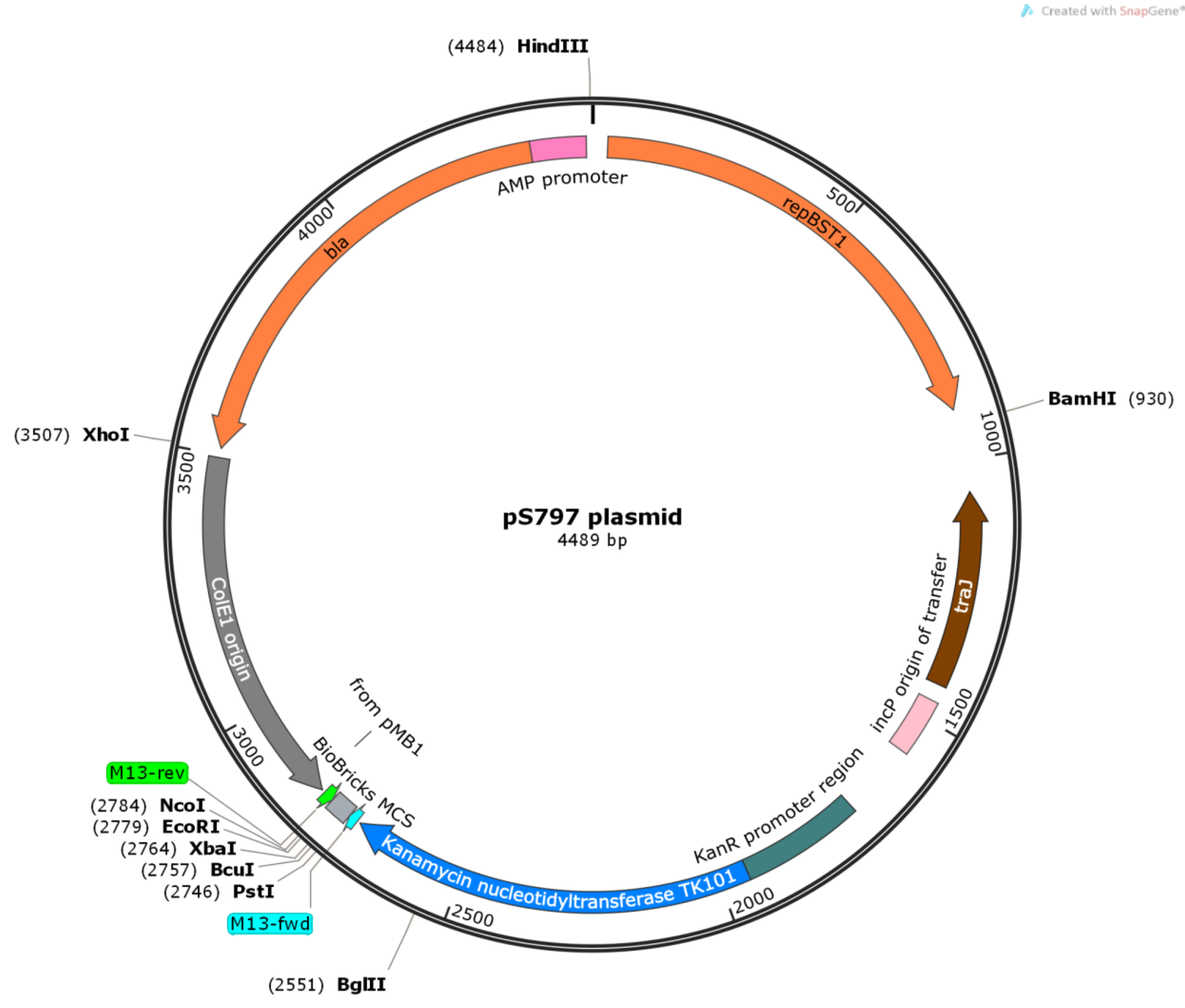


Figure 6 An outline of the conjugal plasmid used to introduce and express genes into *Geobacillus* sp.
The plasmid contain an ColE1 origin of replication and Bla/amp^R for replication and selection of the plasmid in *E.coli* (ampicillin/ carbenicillin 100 µg/ml) and a repBST1 Ori and kan^R cassette for replication and selection in *Geobacillus thermoglucosidans.* An OriT for conjugal transfer is included in the plasmid. pS797 is BioBricks compatible (RFC 10). Full vector sequence is available in appendix 1. Plasmid was obtainined from Dr. Chloe Singleton.

The plasmid vector pS797 (Figure 6) contains two selection marker genes, bla/ Amp^R and Kan^R TK101 for selection of transformed *E. coli* and *Geobacillus* sp. respectively. The plasmid contains two origins of replication; ColE1 Ori, for replication in *E. coli,* and repBST1 for replication in *Geobacillus* sp which has been shown to have homology to a family of theta replicons that have previously only been found in strains of *Bacillus megaterium* (Taylor *et al.* 2008). The plasmid is compatible with BioBricks RCF10 standardised assembly. The required genes and recognition sequences for conjugation are contained within the plasmid making it amenable for transfer by conjugation (Hwang *et al.* 1994; Arutyunov *et al.* 2011; Gubbins *et al.* 2002; Lanka & Wilkins 1995).

3.11.2. *Heat shock transformation of* E. coli *S17-1*

Chemically competent *E. coli* S17-1 were thawed on ice. A 2 µl aliquot of freshly prepared plasmid DNA was added to the competent cells, mixed gently and incubated on ice for 30 min. The cells were transferred to a 42 °C water bath for 30 s, and placed on ice for 2- 3 min, 500 µl of SOC medium was added. The cells were then transferred to a shaking incubator at 37 °C shaking incubator, with shaking at 225 rpm for 1 h. The cells were then pelleted by centrifugation at 15,000 rpm for 1 min. The supernatant was removed and the pellet re-suspended in 50 µl of SOC medium and spread plated out onto LB + ampicillin (100 µg ml $^{-1}$) plates. The inoculated plates were incubated at 37 °C for 16 h.

3.11.3. *Conjugal transformation of* Geobacillus

The conjugal donor strain was transformed with the pS797 vector via heat shock transformation as previously explained. The conjugal recipient strain (*Geobacillus thermoglucosidans* DSM 2542) was plated out from a TSA slope culture onto TSA. The culutre was then incubated for 16 h at 55 °C. A single colony was picked and plated onto TSA, to give a lawn and incubated at 55 °C. The heat shock conjugal donor was streaked out onto LLB + Carbenicillin (100 µg ml $^{-1}$), from the transformation plate and incubated at 37 °C for 16 h. The donor and recipient strains were then mixed thoroughly in 600 µl of LLB liquid medium in an approximate ratio of 1:5 Donor: Recipient, and plated onto LLB agar plates in approximately 20 x 10 µl droplets. These plates were left at room temperature to dry. Once dry the plates were incubated at 37 °C for 7 h. Following the incubation, 1 ml of TSB is pipetted onto the LLB plate and the cells were re-suspended. Once re-suspended the culture was

mixed 1:1 in TSB and 100 µl was plated onto TSA containing 12.5 µg ml $^{-1}$ of kanamycin (TSAK). 100 µl was also inoculated into TSB containing 12.5 µg ml $^{-1}$ of kanamycin (TSBK). These plates were incubated at 55 °C for 48 h (Cripps *et al.* 2009). A WT *E. coli* S17-1 conjugation was performed as a negative control.

3.11.4. *Screening of* Geobacillus *for successful transformation.*

Firstly, single colonies were streak plated onto TSAK plates. If the colonies grew then this confirmed that the selected colony has resistance to kanamycin. A colony from the freshly streaked plate was used to inoculate 5 ml of TSBK and incubated at 55 °C for 16 h. Plasmid DNA was purified from the liquid culture according to the standard protocol (GeneJet Miniprep kit, ThermoFisher Scientific), with some modififcations; plasmid was purified from 10 ml of overnight culture and for the final elution step 30 µl of elution buffer was used. The purified plasmid was digested at 37 °C using *Eco*RI and *Pst*I according to the standard protocol (See below)(Fast digest restriction enzymes, ThermoFisher Scientific) . The resulting digested plasmid was resolved by size using gel electrophoresis (45 mM Tris-Borate/ 1mM EDTA (10.5 g l^{-1} Tris base, 5.5 g l^{-1} boric acid, 0.1 M EDTA) containing 1 % agarose gel). The size of the purified plasmids were calculated based on the separation of a DNA weight marker (1 kb DirectLoad (Sigma Aldrich). Undigested plasmid controls were included in the digest experiment to confirm proper digestion. Wild type *Geobacillus thermoglucosidans* DSM 2542 used as a negative control to confirm that the electrophoretic result was a product of successful transformation.

Digest protocol

	Negative	*Eco*RI	*Pst*I	*EcoR*I / *Pst*I
Fast digest buffer	2 µl	2 µl	2 µl	2 µl
DNA	5 µl	5 µl	5 µl	5 µl
EcoRI	0 µl	1 µl	0 µl	1 µl
PstI	0 µl	0 µl	1 µl	1 µl
H_2O	13 µl	12 µl	12 µl	11 µl

These reactions were incubated at 37 °C for 30 min.

3.11.5. *Testing competence of selected* Geobacillus sp. *to transformation conjugal transfer of pS797.*

From the initial results from chapter 4, the plate overlay assay showed that there were seven species, which were able to degrade both CMC and xylan. The conjugation protocol (modified from Cripps *et al*. 2009) was tested on the seven strains (*G. thermoleovorans* DSM 5366, *G. thermoglucosidans* DSM2542, *G. gargensis* DSM15378, *G.jurassicus* DSM15726, *G. lituanicus* DSM15325, *G. toebii* DSM14590, *Geobacillus* sp. ATCC 15952, *G. stearothermophilus* NCIMB8224), allowing the cross species compatibility of the pS797 vector and conjugation protocol to be assessed.

3.12. Expression and activity of cellulase proteins in *Geobacillus*

Cellulase proteins were synthesised by DNA2.0 (Menlo Park, C.A, USA). Following transformation, the expression and activity of the cellulase proteins were tested by assessing CMC degradation. Expression was assessed using a western blot and CMCase activity was confirmed by a well diffusion assay.

3.12.1. *Incorporation of 6xHistidine tags into the C-terminus of the cellulase proteins*

Site directed mutagenesis was used to add a 6x Histidine tag to the C-terminus of the CelA, CelB, Cel6B, Cel9A and CelZ. Site directed mutagenesis (NEB Q5 site directed mutagenesis standard protocol) was used to incorporate three histidine residues onto each end of the amplicon (Primer sequences are listed in Table 1) and recircularise the resulting PCR product.

<u>PCR reaction mix</u>

Reaction mix:

Q5 master mix	12.5 µl
Template (20 ng/µl)	1 µl
Forward primer 10 µM	1.25 µl
Reverse primer 10 µM	1.25 µl
DMSO (added to *CelA*, *CelZ* and *Cel6B*)	1.5 µl
H_2O	up to 25 µl

Heat cycle:

Denaturation A	98 °C	30 s
Denaturation B	98 °C	10 s
Annealing (*A/B/6B/9A/Z*)	56/67/69/64/67 °C	30 s
Elongation (*A/B/6B/9A/Z*)	72 °C	5/3/2.5/3/3 min
Elongation	72 °C	10 min

Table 1 Primers used for site directed mutagenesis of cellulase proteins to introduce C-terminal 6x His tag sequence.

Primer name	Sequence (5' – 3')
C.besciiCelA_Fwd	caccaccac**TAATAA**AAATGGAAATGATTCACAC
C.besciiCelA_Rev	atgatgatgCTGATTGCCAAACAAGATC
C.saccarolyticus CelB_Fwd	caccaccac**TAATAA**AAATGGAAATGATTCACACCCGATG TG
C.saccarolyticus CelB_Rev	atgatgatgCTTGGATGCCGGGACCGG
C.stercorarium CelZ_fwd	caccaccac**TAATAA**AAATGGAAATGATTCACACCCG
C.stercorarium CelZ_rev	atgatgatgCGGCTCGATACCGCTCAC
T.fuscaCel9A_Fwd	caccaccac**TAATAA**AAATGGAAATGATTCACACCCGATG TG
T.fuscaCel9A_Rev	atgatgatgCGCCAAGGCACACGGCTC
T.fuscaCel6B_Fwd	caccaccac**TAATAA**AAATGGAAATGATTCACACCCGATG
T.fuscaCel6B_Rev	atgatgatgCAGCGGCGGATAGGCATTCG
T.neopolitanaCelB_Fwd	caccaccac**TAATAA**AAATGGAAATGATTCACACCCGATG
T.neopolitanaCelB_Rev	atgatgatgCTCGCCGATCTCCACGCT

Following PCR amplification, a kinase, ligase, DpnI reaction (KLD) was performed according to the standard protocol (see below),(New England Biosciences, USA). A graphic representation of this reaction is shown in Figure 7. The reaction mix was then transformed into *E. coli* TOP10 via heat shock transformation. Following transformation 12 x single colonies were inoculated into liquid LB + ampicillin (100 µg ml $^{-1}$). Plasmid DNA was purified from this culture and a diagnostic digest was carried out using *Eco*RI and *Pst*I to confirm the correct plasmid size, plasmids of the expected size were sent for Sanger sequencing (Intergrated DNA technology (IDT).

<u>Kinase Ligase DpnI Protocol</u>

PCR product	1 µl
2x KLD reaction buffer	5 µl
10x KLD Enzyme mix	1 µl
H_2O	3 µl

Incubated at room temperature for 5 min.

3.12.2. *Detection of cellulase expression by anti-His-HRP conjugated antibody binding*

The following strains were grown from glycerol stocks on TSA (WT) and TSAK (pS797 strains) plates at 55 °C in a static incubator and incubated for 24-48 h.

- Wild type *G. thermoglucosidans*
- pS797 empty vector
- pS797 *CelA*
- pS797 *CelB*
- pS797 *Cel6B*
- pS797 *Cel9A*

After incubation, single colonies were used to inoculate 5 ml of TSB and TSB + Kanamycin (12.5 µg ml $^{-1}$) and was grown overnight and 1 ml was inoculated into 50 ml of TSB. From the 50 ml 2 x 1 ml samples were taken and centrifuged for 1 minute at 15000 rpm. Cell pellets were re-suspended in 32.5 µl of H_20, 12.5 µl of LDS sample buffer and 5 µl of reducing agent. Samples were heated at 95 °C for 10 min and centrifuged at 15000 rpm for 10 min. Following centrifugation, 5 µl of supernatant

was analysed using SDS PAGE on a 4-12% Bis-TRIS SDS PAGEat 165 volts for 35 min. After SDS-PAGE the proteins were transferred onto a nitrocellulose membrane using a Pierce G2 semi dry fast blotter. Proteins were transferred onto the nitrocellulose membrane by applying a potential difference across the membrane stack at 15 volts and 1 A for 20 min. Once the proteins had been transferred the membrane was probed with anti-His HRP conjugated antibody (1:400) using the iBind for 2.5 h (Life Technology, Carlsbad, USA). The antibody probed blot was immersed in 6.4 ml of chemiluminescent substrate 1:1 peroxide solution: chemiluminescent signal enhancer) and incubated at room temperature for 5 min. Chemiluminescent images of the stained western blot were taken with a myECL chemiluminescent imager (Thermo Scientific, Waltham USA).

3.12.3. *Confirming heterologous cellulase protein activity and assessing protein localisation in* Geobacillus *thermoglucosidans*

In order to test the in vivo functionality of the cellulase genes *Geobacillus* pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A* were inoculated into 50 ml of TSB + 12.5 µg ml $^{-1}$ of kanamycin from an overnight culture. Bacteria were cultured at 55 °C with shaking at 200 rpm for 24 h. Two 500 µl aliquots of the 24 h culture were transferred into 1.5 ml Eppendorf tubes. These samples were centrifuged at 15,000 rpm for 2 min. The supernatant was aliquoted into a well in a 0.25 % CMC soft agar plate (0.7% agar). The pellet was resuspened in 500 µl of fresh TSB and was also transferred into the 0.25 % CMC agar plate (0.7 % agar). Wells were made by pouring molten agar around HPLC vials, the vials were then removed leaving an approximately 600 µl cavity in the surface of the agar. Plates were incubated at 55 °C for 24 h. In order to visualise CMC degradation plates were stained with 0.1 % Congo red for 10 min and destained with 1 M sodium chloride for 20 min. Images of the stained wells were taken with a Canon Power shot 5X50 HS camera. Two biological replicates were performed.

The well diffusion assay was used to test the activity of the cellulase proteins at 65 °C and 75 °C. The protocol was carried out on the supernatant of a 24 hour culture of *Geobacillus* and incubated at 65 °C and 75 °C for 24 h and stained as described above.

3.12.4. *Assessing the in vitro activity of cellulase proteins from* G. thermoglucosidans *with zymography.*

Single colonies were inoculated into 5 ml of TSB and TSB containing kanamycin (12.5 µg ml^{-1}) and incubated for 16 h at 55 °C, shaking at 200 rpm. A 1 ml aliquot was inoculated into 50 ml of TSB which was incubated as before. From the 50 ml a 1 ml samples was taken and spun down for 1 minute at 15,000 rpm. The supernatant was removed and combined with 32.5 µl of H_20, 12.5 µl of LDS sample buffer and 5 µl of reducing agent. The cell pellet was re-suspended in 32.5 µl of H_20, 12.5 µl of LDS sample buffer and 5 µl of reducing agent. The samples were heated at 95 °C for 10 min and centrifuged at 15,000 rpm for 10 min. Following this, 5 µl of supernatant was analysed using SDS PAGE on a 4-12 % Bis-TRIS SDS PAGE gel containing 0.1 % CMC and run at 165 volts for 35 min. Following electrophoresis, the gel was rinsed with water, then incubated for 1 hour in 2.5 % (v/v) Triton X-100, sodium phosphate buffer (50 mM pH 6.5). The buffer was then replaced with the same buffer and the gel was incubated for 12 h at 55 °C. The gel was stained with 0.1 % Congo red for 20 min and destained with 1 M NaCl for 20 min (Assareh *et al.* 2012). Supernatant and cell pellet samples were run on individual gels.

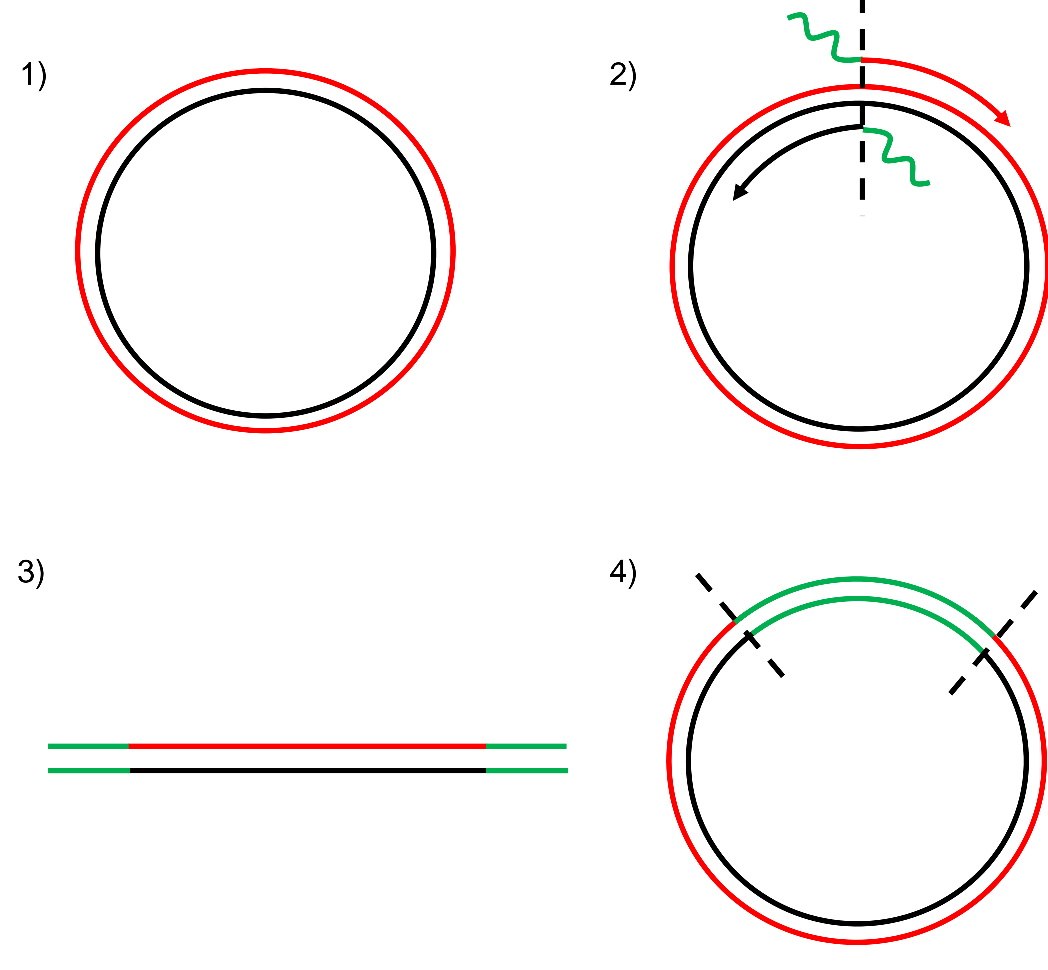


Figure 7 Overview of Q5 Site Directed Mutagenesis: C-terminal 6x His tag sequence addition to *CelA*, *CelB*, *Cel6B* and *Cel9A*.

The figure above outlines Q5 Site Directed Mutagenesis. 1) Plasmid pS797 + *CelA*, *CelB*, *Cel6B* and *Cel9A* genes were purified and used as template DNA for plasmid linearisation by PCR. 2) The primers used to amplify the plasmid have 5´ tails. These incorporate the desired sequence insertion into the original sequence into the plasmid sequence. 3) The linear PCR product has the introduced sequence on each of the DNA strand. 4) The KLD reaction circularises the linear DNA giving the final plasmid with the introduced sequence. This protocol was used in this study to insert a His tag (6xHisTag) onto the C-terminus of the 5 cellulase proteins.

3.12.5. *The effect of cellulase expression on growth*

In order to test the effect of cellulase expression on the growth of *Geobacillus thermoglucosidans* DSM 2542, the growth of wild type *G. thermoglucosidans* DSM2542 and *G. thermoglucosidans* DSM2542 transformed with pS797 vector, pS797 *CelA* pS797 *CelB,* pS797 *Cel6B* or pS797 *Cel9A* were cultured and growth was measured. The six strains were inoculated from a plate. A 5 ml aliquot of TSB +/- kanamycin (12.5 µg/ml) was inoculated with each strain and incubated at 55 °C with shaking at 200 rpm for 16 h. A 1 ml aliquot of this culture was inoculated into 50 ml of TSB +/- kanamycin (12.5 µg ml^{-1}). This was incubated further, under the same conditions for 24 h. The optical density (OD_{600}) of 250 µl of culture was measured using a TECAN M200 Infinite plate reader. Three biological replicates were performed and OD readings of each culture were taken in duplicate.

3.13. *Exchanging P_{ldhA} for P_{araE} to facilitate control of CelA expression*

The cellulase genes were initially synthesised so that gene expression was regulated by a lactate dehydrogenase promoter (P_{ldhA}) taken from *Geobacillus thermoglucosidans* DSM2542. Gene expression under the P_{ldhA} promoter is dependent on the oxygen concentration (Bartosiak-Jentys *et al.* 2012), with gene expression being induced under microaerobic conditions. The requirement of microaerobic conditions limits control of expression and may not be compatible will optimum growth conditions.

An inducible system would allow large amounts of culture to be grown quickly, without the burden of gene expression and then induced by the addition of the required inducer. In order to test this the P_{ldhA} was exchanged for an arabinose inducible promoter (P_{araE}) to yield pS797 P_{araE}::*CelA*. P_{araE} was inserted into switched using Gibson assembly (Gibson *et al.* 2009). An overview of Gibson assembly can be seen in Figure 8.

The two parts, pS797 *CelA* not including the P_{ldhA} promoter (Part A) and P_{araE} (Part B) were amplified using Phusion Polymerase (see below) using the primers listed in Table 2. Following PCR the products were purified using a PCR clean up column (Thermo Scientific). The resulting purified Parts A and B, were quantified (using a Quibit spectrophotometer, Thermo Scientific) and diluted to the appropriate concentration. The two parts were then mixed in a ratio of 1:1 and 1:5 (A:B).

PCR reaction mix

Reaction mix:

Phusion master mix	12.5 µl
Template (20 ng $µl^{-1}$)	1 µl
Forward primer 10 µM	1.25 µl
Reverse primer 10 µM	1.25 µl
DMSO	0.5 µl
H_2O	up to 25 µl

Heat cycle:

Denaturation A	98 °C	30 s
Denaturation B	98 °C	10 s
Annealing (*araE/ CelA*)	60 °C/72 °C	30 s
Elongation (*araE/ CelA*)	72 °C	0.5/5 min
Elongation	72 °C	10 min

Table 2 Sequences of primer used for Gibson assembly.

Primer name	**Sequence**
pS797 *CelA* Forward	ATGAAGCGTTATCGTCGCATCATC
pS797 *CelA* Reverse	CTCTAGAAGCGGCCGCGAATTC
P_{araE} **Forward**	cggccgcttctagagGGAGGTGACCTCAAAAGC
P_{araE} **Reverse**	cgacgataacgcttcatTTTTCCCCCTCCTAAAATTTG
***CelA* + P_{araE} Seq Forward**	CCTACAGCGTGAGCTATGAGAAAGC
***CelA* + P_{araE} Seq Reverse**	GTCAAACCAACCACCCGTCAGG

Gibson assembly reaction mix:

Ratio	1:1	1:5
Gibson assembly master mix	10 µl	10 µl
Part A (0.021 pmol)	2 µl	2 µl
Part B (0.05 pmol)	1 µl	5 µl
H_2O	7 µl	3 µl

The Gibson assembly reaction was incubated in a thermocycler at 50 °C for 15 min. Following this 2 µl of the reaction was transformed into *E. coli* DH5α by heat shock transformation at 42 °C. Colonies returned following transformation were incubated for 16 h in LB containing Ampicillin 100 µg ml $^{-1}$ at 37°C. The plasmid DNA then purified and digested as described below. A total of 12 colonies, 6 from each reaction ratio, were grown for plasmid purification and digest with *Xba*I and *Nde*I to confirm plasmid size.

Plasmid Digest following Gibson assembly

Fast Digest buffer	2 µl
DNA	2 µl
Restriction enzyme (*Xba*I and *Nde*I)	1 µl
H_2O	up to 20 µl

The restriction enzyme digest of plasmids yielded by Gibson assembly was incubated at 37 °C for 30 min. Following digestion the digest reaction mix was separated by gel electrophoresis on a 1 % agarose TBE gel at 120 volts for 90 min to separate the digestion products by size.

The correctly assembled plasmid was transformed into *E. coli* S17-1 as conjugation was carried out as previously described. Once conjugated, ten *Geobacillus* colonies were grown and plasmid DNA was purified using a mini prep kit (Thermo Scientific, Waltham, USA) following slight modifications to the protocol (10 ml of culture used for

miniprep and the purified DNA was eluted into 30 µl of H_2O). The purified DNA was digested as described above.

Once the transformation of the newly constructed pS797 P_{araE}*::CelA* had been verified the required inducer concentration and the effect of induction on cell growth were assessed as described in the following section.

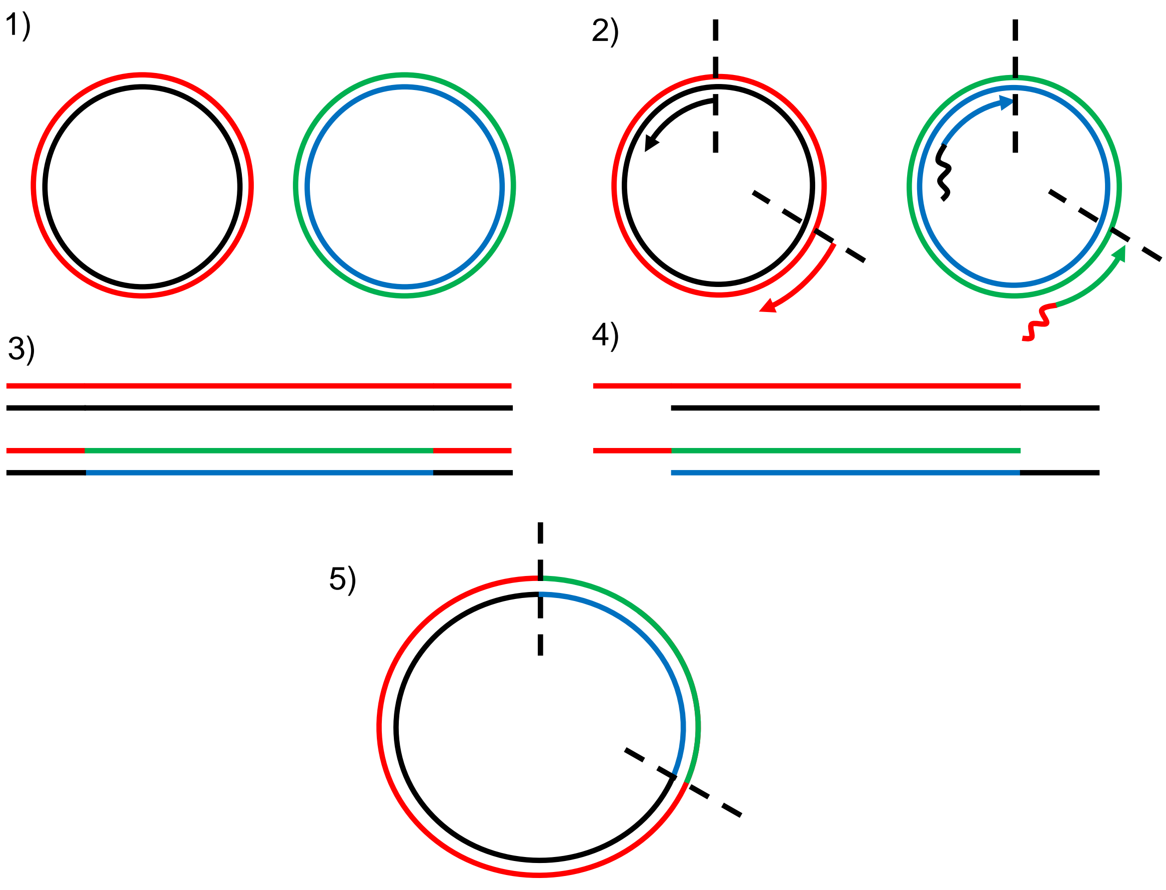


Figure 8 An outline of Gibson assembly.

1) Plasmids containing the respective parts are purified. 2) Primers are designed that will amplify Part A (Black and Red) and Part B (Blue and Green). The primers that amplify Part B each have a 5´ tail sequence that does not anneal to the Part B plasmid, but overlap with the 5´ and 3´ of Part A. 3) PCR is carried out and the resulting products are purified. 4) These PCR products are then added to the Gibson reaction mix at ratios of 1:1 and 1:5 (Part A:Part B) and incubated at 50 °C for 15 min. During this incubation time the 3´ to 5´ exonuclease digests the DNA to give single stranded overhanging sequences. Because of the overlapping sequence added during part amplification these overhangs are complementary to each other. 5) The overhanging sequences align; any gaps in the sequence are then filed by a polymerase and the phosphate backbone of the DNA strands are repaired by a ligase enzyme to give the final plasmid (Gibson *et al.* 2009).

3.7.1. *Induction of* CelA *expression by arabinose*

In order to uncouple cell growth and protein expression the *CelA* gene was placed under control of an arabinose inducible promoter (P_{araE}). *Geobacillus thermoglucosidans* DSM2542 wild type, pS797, and pS797 P_{araE}*::CelA* were grown from single colonies for 24 h at 55 °C with shaking at 200 rpm. A 1 ml inoculum was then added to 50 ml of TSB +/- kan12.5 μg ml $^{-1}$ containing different arabinose concentrations (0, 0.1, 0.2, 0.4, 0.6 % w/v). The bacteria were inoculated for 24 h at 55 °C with shaking at 200 rpm. A 1 ml sample was centrifuged at 15,000 rpm for 1 min. A 250 μl sample of supernatant was taken and used to measure CelA activity using the well diffusion assay (as previously described). Another 250 μl sample was taken for HPLC to assess arabinose degradation. A 20 μl sample was then taken and injected into the HPLC (Dionex 3000 HPLC with a Shodex RI detector). The mobile phase was deionised water with a flow rate of 0.2 ml min^{-1}. The temperature of the column oven was 83 °C.

4. Results

4.1. Chassis selection for consolidated bioprocessing

4.1.1. Rationale

The first aim of this research project was to screen the 44 publicly available species of *Geobacilli* (Table 3), for the ability to degrade cellulose. Other desirable characteristics were assessed such as: ease of storage and transport, growth rate, cellulose degradation, glucose polymer utilisation, utilisation, xylan degradation and competence. The results of the *Geobacillus* spp. trait screening are shown in this chapter.

4.1.2. *Growth of* Geobacillus *on simple medium*

Screening for presence or absence of growth on a simple medium was carried out. Growth on simple medium is a desirable characteristic because it simplifies the optimisation of fermentation. It was found that all the strains tested were able to grow on both TSA and TSB (Table 4).

4.1.3. *Revival of* Geobacillus *following cryo-storage*

All ordered strains were revived from the provided bacterial biomass. Of the strains obtained all except GB17 and GB23 were successfully revived; as such GB17 and GB23 were not used for the remainder of this study. The viability of *Geobacillus* spp. following cryo-storage is highlighted in Table 4.

Table 3 Strains selected and screened for cellulolytic activity**.**

I.D.	Species/Strain	Strain I.D..	Source	Medium	Reference
GB1	*G. thermoleovorans*	DSM 5366	DSMZ	Nutrient agar	(Priest 1989) (Zarilla & Perry 1987)
GB2	*G. thermoleovorans* T4	DSM 14791	DSMZ	TSA	(Tai *et al.* 2004)
GB3	*G. stearothermophilus*	DSM 22	DSMZ	Nutrient agar	-
GB4	*G. stearothermophilus* XL-65-6	9A19	BGSC	TSA	(Lai & Ingram 1993)
GB5	*G. thermoglucosidans* NUB36	-	Could not be sourced	-	-
GB6	*G. thermoglucosidans* NUB3621	9A5	BGSC	TSA	(Chen et al. 1986)
GB7	*G. thermodenitrificans* K1041		University of Exeter	TSA	(Studholme *et al.* 1999)
GB8	*G. stearothermophilus* NRRL B-4419	9A2	BGSC	TSA	-
GB9	*G. stearothermophilus* strain 10	9A21	BGSC	TSA	(Stenesh *et al.* 1968)
GB10	*G. kaustophilus* DSM 7263	DSM 7263	DSMZ	Nutrient agar	(Priest 1989)
GB11	*G. subterraneus*	DSM 13552	DSMZ	Nutrient agar	(Poli *et al.* 2012)
GB12	*G. uzenensis*	DSM 23175	DSMZ	Nutrient agar	(Nazina *et al.* 2001)
GB13	*G. thermocatenulatus*	DSM 730	DSMZ	Nutrient agar	(Golovacheva *et al.*)
GB14	*G. thermodenitrificans*	DSM 465	DSMZ	Nutrient agar	(Manachini *et al.* 2000)
GB15	*G. thermoglucosidans*	DSM 2542	DSMZ	TSA	(Suzuki *et al.* 1983)
GB16	*G. caldoxylosilyticus*	DSM120 41	DSMZ	TSA	(Ahmad *et al.* 2000; Fortina *et al.* 2001)
GB17	*G. debilis*	DSM 16016	DSMZ	TSA	(Banat *et al.* 2004)
GB18	*G. galactosidasius*	DSM 18751	DSMZ	TSA	(Poli *et al.* 2011)
GB19	*G. gargensis*	DSM 15378	DSMZ	Nutrient agar	(Nazina *et al.* 2004)
GB20	*G. jurassicus*	DSM	DSMZ	Plate	(Nazina *et al.* 2005)

		15726		count agar	
GB21	*G. lituanicus*	DSM 15325	DSMZ	Nutrient agar	(Kuisiene *et al.* 2004)
GB22	*G. pallidus*	DSM 3670	DSMZ		(Anon 1988)
GB23	*G. subterraneus* subsp. aromaticivorans	DSM 23066	DSMZ	Nutrient agar	(Poli *et al.* 2012)
GB24	*G. tepidamans*	DSM 16325	DSMZ	TSA	(Schäffer *et al.* 2004)
GB25	*G. thermantarcticus*	DSM 9572	DSMZ	Rich soil extract medium	(Coorevits *et al.* 2012)
GB26	*G. thermodenitrificans* subsp. calidus	DSM 22629	DSMZ	Nutrient agar	(Cihan *et al.* 2011)
GB27	*G. toebii*	DSM 14590	DSMZ	TSA	(Sung *et al.* 2002)
GB28	*G. vulcani*	DSM 13174	DSMZ	Bacto-Marine Broth	(Caccamo *et al.* 2000)
GB29	*Geobacillus* sp.	ATCC BAA-1070	ATCC	TSA	(Fong *et al.* 2006)
GB30	*Geobacillus* sp. NCA 5520	ATCC 15952	ATCC	TSA	(Hammarström & Strominger 1975)
GB31	*Geobacillus* sp.	ATCC BAA 1068	ATCC	TSA	(Fong *et al.* 2006)
GB32	*G. stearothermophilus* N.R.Smith T141	NCIMB 8222	NCIMB	Lysine agar	
GB33	*G. stearothermophilus* N.R.Smith T147	NCIMB 8223	NCIMB	TSA	
GB34	*G. stearothermophilus* N.R.Smith T147	NCIMB 8224	NCIMB	Lysine agar	
GB35	*G. stearothermophilus* N.R.Smith T73	NCIMB8919	NCIMB	TSA	
GB36	*G. stearothermophilus* N.R.Smith T75	NCIMB 8921	NCIMB	Lysine agar	
GB37	*G. stearothermophilus* N.R.Smith T19	NCIMB 8924	NCIMB	Lysine agar	
GB38	*G. stearothermophilus* 186	NCIMB 10278	NCIMB	Lysine agar	
GB39	*G. stearothermophilus* Alpha 1	NCIMB 10279	NCIMB	Lysine agar	
GB40	*G. stearothermophilus*	NCIMB	NCIMB	Lysine	

	Alpha 7	10280		agar
GB41	*G. stearothermophilus*	NCIMB 10814	NCIMB	Lysine agar
GB42	*G. thermodenitrificans* E32-66	NCIMB 11731	NCIMB	Lysine agar
GB43	*G. stearothermophilus*	NCIMB 11780	NCIMB	Nutrient agar
GB44	*G. stearothermophilus* NRS T15	NCIMB 13693	NCIMB	Lysine agar
GB45	*G. thermodenitrificans*	NCIMB 14578	NCIMB	Lysine agar

*

Table 4 The presence or absence of growth of *Geobacillus* spp. on TSB and TSA and the viability of *Geobacillus* spp. following cryo-storage.

I.D	**Growth of *Geobacillus* spp. on TSA**	**Growth of *Geobacillus* spp. on TSB**	**Revival from cryo-storage**
GB1	✓	✓	✓
GB2	✓	✓	✓
GB3	✓	✓	✓
GB4	✓	✓	✓
GB5	-	-	✓
GB6	✓	✓	✓
GB7	✓	✓	✓
GB8	✓	✓	✓
GB9	✓	✓	✓
GB10	✓	✓	✓
GB11	✓	✓	✓
GB12	✓	✓	✓
GB13	✓	✓	✓
GB14	✓	✓	✓
GB15	✓	✓	✓
GB16	✓	✓	✓
GB17	✓	✓	X
GB18	✓	✓	✓
GB19	✓	✓	✓
GB20	✓	✓	✓
GB21	✓	✓	✓
GB22	✓	✓	✓
GB23	✓	✓	X

GB24	✓	✓	✓
GB25	✓	✓	✓
GB26	✓	✓	✓
GB27	✓	✓	✓
GB28	✓	✓	✓
GB29	✓	✓	✓
GB30	✓	✓	✓
GB31	✓	✓	✓
GB32	✓	✓	✓
GB33	✓	✓	✓
GB34	✓	✓	✓
GB35	✓	✓	✓
GB36	✓	✓	✓
GB37	✓	✓	✓
GB38	✓	✓	✓
GB39	✓	✓	✓
GB40	✓	✓	✓
GB41	✓	✓	✓
GB42	✓	✓	✓
GB43	✓	✓	✓
GB44	✓	✓	✓
GB45	✓	✓	✓

4.1.3.1. *Investigating endoglucanase activity in Geobacillus* spp.

Figure 9 shows that over time the mass of all the *Geobacills* spp. cultures decreases, however this decrease is also observed in the controls, suggesting the weight loss is caused by evaporation and not Avicel degradation and CO_2 release. There are two distinct groups within the data set. The strains are plotted along the xaxis in the order in which they were cultured (Control 2 was grown in flask type 2). All strains from 1 – 20 were cultured in flask type 1 and 25 – 46 were cultured in flask type 2. When the mass of the flasks were compared following incubation, (all cultures in flask type 1 were compared to control 1 and cultures in flask type 2 were compared to control 2) there is no significant different in weight loss as all fall within the standard error found within the control. The same is also true when we compare those cultured in flask type 2 and control 2.

It is possible that there was some degradative activity occurring within the culture but that the assay used is not sensitive enough to detect it. A more sensitive assay was required for further analysis of cellulolytic ability. Because of the insensitivity of the assay the data is inconclusive and a better method for testing cellulolytic degradation was investigated. Another assay which has previously been used to identify the cellulolytic bacteria from the rumen of cattle is the Congo red assay (Teather & Wood 1982). This assay is based on the clearing of carboxymethylcellulose from agar plates by bacterial cellulase enzymes. Congo red stains the beta-1,4 glyosidic bonds in polysaccharide polymers, which are degraded by cellulase proteins, thus allowing easy visualisation of cellulolytic activity.

The Congo red method is appropriate for screening of cellulase activity in *Geobacillus* spp. because it is quick and simple allowing for the rapid screening of numerous bacterial species. Images from the plate overlay assay are shown in Figure 10. The strains highlighted in Table 5 were shown to clear xylan and CMC by Congo red staining (Figure 10). Based on this characteristic the strains GB1, GB19, GB20, GB21, GB27, GB30 and GB34 were selected for further characterisation. GB15 was able to degrade CMC but not xylan. The reason that GB15 was included in the further screening is because it has previously been used as a host for industrial fermentation of ethanol and has been engineered to express an active cellulase gene from and exogenous host.

There are issues with the Congo red assay. Because of the washing steps involved following the staining it is possible that the peripheral parts of the colonies can become dislodged. The area under these peripheral areas look like zones of clearing but are caused by the colonies preventing full exposure to the stain. The “washing off” of colonies was something that was observed whilst conducting the assay. For this reason the results are to be treated with caution. The assay is not quantitative, meaning this assay cannot be used to rank the cellulolytic capabilities of the selected strains. An assay that would allow quantitation of cellulolytic activity was also been investigated and will discussed later.

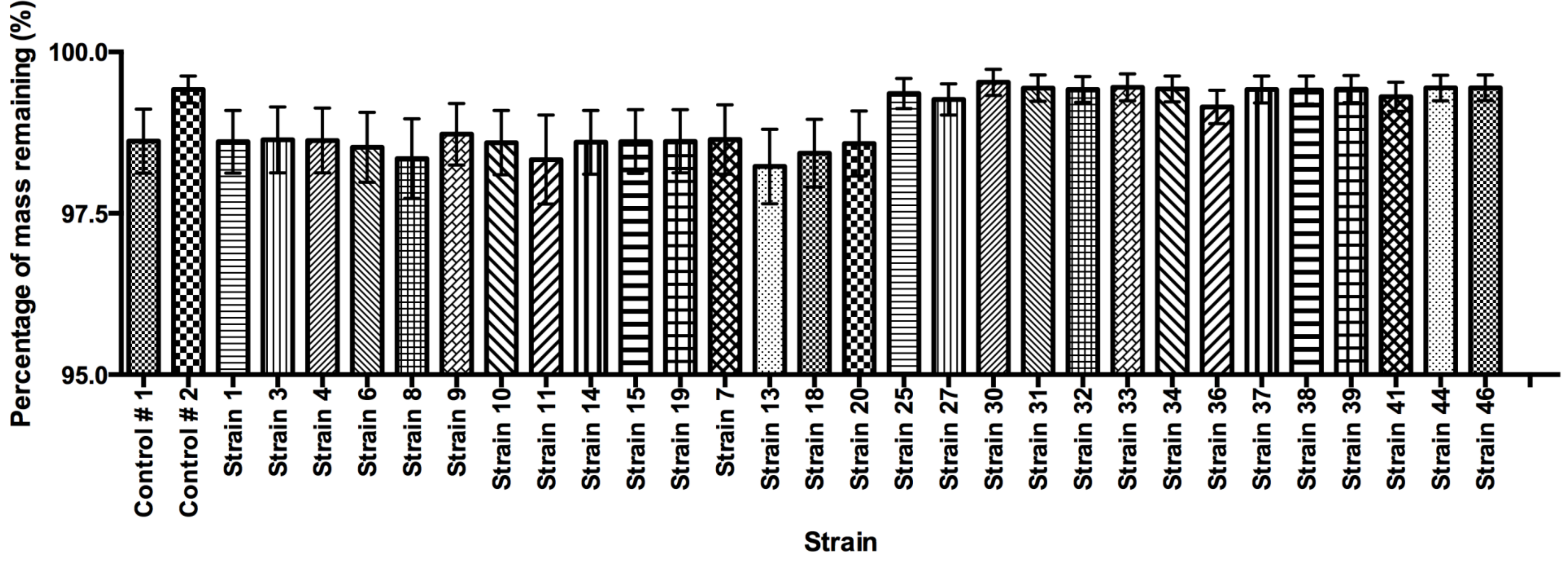


Figure 9 Weight loss of selected strains of *Geobacillus* cultured in ASM + 1 % Avicel following incubation for 14 days.

Geobacillus sp. were grown in ASM + Avicel (1 %) at 55 °C with shaking at 200 rpm for 14 days. The weight of the culture was measured. The change in weight here is represented as % of total weight remaining after 14 days. Control #1 correlates to 1-20 and Control #2 to 25-46. The difference between the controls is because different brands of flask were used, which had different rates of evaporation. This experiment was carried out in triplicate. Error bars represent the standard error.

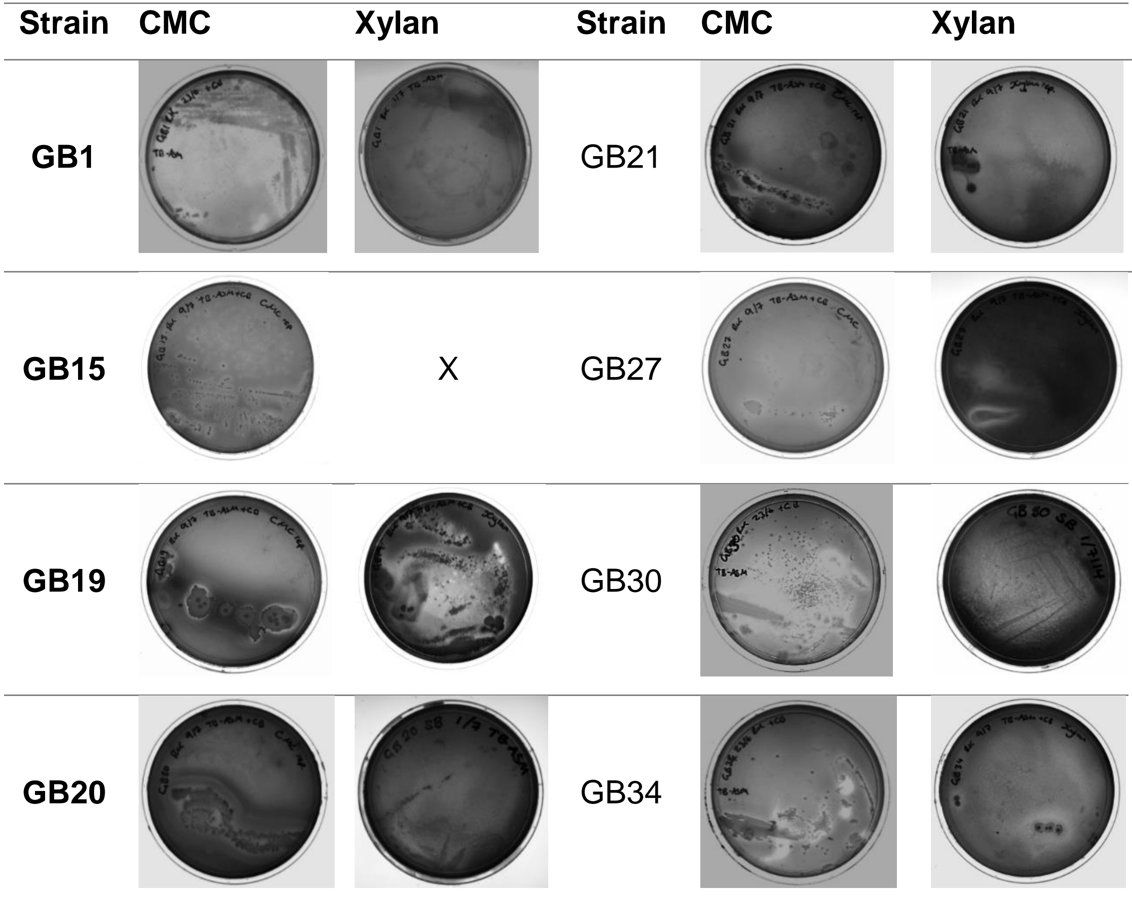


Figure 10 Summary of CMC overlay assay results

*Geobacillus*strains that were able to clear CMC and xylan in 0.7 % agar when overlaid onto streak plated bacteria. The 0.7 % agar was supplemented with 0.25 % CMC or 1 % xylan. This was then stained with 0.1 % Congo red. The images show zones of clearing which indicate CMC and xylan degradation.

Table 5 Additional characteristics of strains predicted to have CMCase and xylanase activity.

I.D #	Strain	Other qualities
GB1	*G. thermoleovorans* DSM 5366	Transformable. HC degrading.
GB15	*G. thermoglucosidans* DSM 2542	Transformable. Similar to C56-YS93. Previous industrial use.
GB19	*G. gargensis* DSM 15378	
GB20	*G. jurassiucus* DSM 15726	
GB21	*G. lituanicus* DSM 15325	
GB27	*G. toebii* DSM 14590	Isolated from hay compost
GB30	*Geobacillus* sp. ATCC 15952	Reported to degrade penicillin.
GB34	*G. stearothermophilus T147 NCIMB8224*	

4.1.3.2. *Utilisation of simple polysaccharides by* Geobacillus spp.

It was important to understand what glucose polymer size could be utilised by *Geobacillus* as glucose polymers are the product of cellulose hydrolysis. By measuring the polysaccharide concentration before and after growth on TTE media containing cellobiose, cellotriose, cellotetraose and cellohexose, the ability of *Geobacillus* spp. to utilise the products of cellulose degradation was assessed.

Analysing glucose polymer degradation will also shed light on the role of *Geobacillus* in a system where cellulolytic activity may be taking place, such as a compost heap. If *Geobacillus* relies on cofermentation then this could be of future interest; for this reason testing the ability of *Geobacillus* spp. to degrade a wide variety of short chain polysaccharides is important.

4.1.3.3. *Short chain polysaccharide utilisation in* Geobacillus spp.

Of the strains tested some show a reduced amount of some of the polysaccharide sizes. GB1, 15, 19, 21, 30 and 34 all show a significant difference in the amount of cellobiose present compared to the control (Figure 11), , and GB1, 19 and 21 show a significant reduction in the amount of cellobiose following fermentation (Multiple T-test, Holm- Sidak method for significance testing α=5.000% (P<0.005) when compared to medium control. This data is presented as the change in HPLC total peak (Refractive Index Units (RIU) area and has not been calculated as a concentration (g l^{-1}) as not enough of the glucose polymers could be obtained to construct a standard curve.

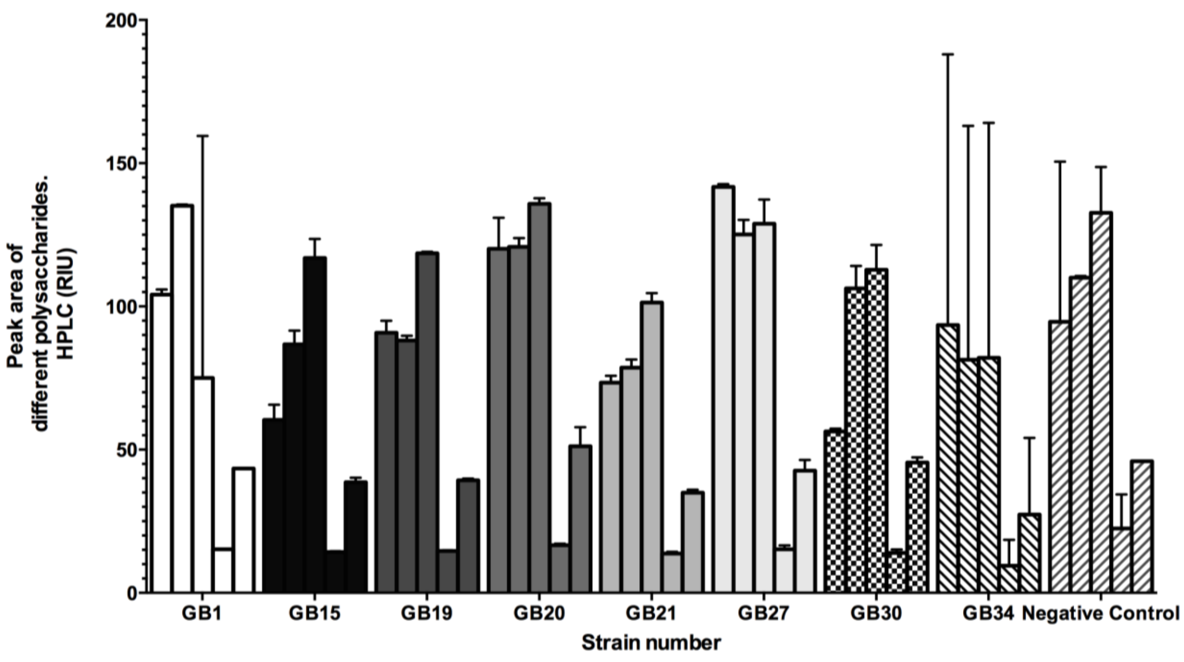


Figure 11 The utilisation of short chain polysaccharides by *Geobacillus* spp.

The change in relative peak area of cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexose following incubation with different strains of *Geobacillus* was measured using HPLC. Strains were incubated for 48 h at 55 °C at 200 rpm. This experiment was carried out in duplicate. Sugar concentration in culture supernatant was measured using HPLC with an Agilent Hi-Flex Na column with a pore size of 10 µm. A 20 µl sample was injected into the HPLC (Dionex 3000 HPLC with a Shodex RI detector). Bars correspond to cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexose respectively from left to right. The negative control represents uninoculated media. Bar are, from left to right, Cellobiose, cellotriose, cellotetraose, cellopentose and cellohexose. One duplicate of medium control was excluded due to large amount of variation; the cellobiose measurement from GB27 which does not appear to have used cellobiose was used in place of this excluded data.

4.1.3.4. *Purification of crude protein extract from* Geobacillus spp. *culture supernatant*
The soluble supernatant protein fraction was purified from *Geobacillus* spp. for 6 h at 55 ºC. Supernatant proteins were precipitated using acetone cooled at -20 ºC. The precipitate was pelleted by centrifugation at 5000 g for 20 min and re-suspended in PBS. The mixture was incubated at a 1:1 (v:v) with 0.25 % CMC in 50 mM citrate buffer. This experiment was carried out in duplicate (technical replicate). Values were normalised against the change in reducing sugar concentration of a negative control.

4.1.3.5. *Endoglucanse activity of crude extracellular protein extract of* Geobacillus spp.
The following experiment was carried out on the proteins purified from culture media of *Geobacillus* spp. grown on CMC in a more quantitative manner than the Congo red plate staining assay. A quantitative method for measuring CMCase activity is required to test CMCase activity. The di-nitrosalycylic acid assay assay was selected as this is has previously been used to assess CMCase activity in a wide variety of systems, including *Geobacillus* (Zambare *et al*. 2011). The supernatant fraction was tested due to the diffuse nature of the zones of clearing observed in the Congo red plate assay and the insolubility of lignocellulose which would require an extracellular system for degradation.

The DNS assay was not conclusive (Figure 12). Whilst initial observations suggested that the supernatant fraction of *Geobacillus* spp. is able to hydrolyse the glucose polymers, statistical tests (Paired two tail t-test: 0 h vs 6 h) showed that this difference is not significant (P>0.05). Initial observations made suggest that there is some cellulolytic activity due to the apparent change in reducing sugar concentration suggests the proteins present in the soluble protein fraction may be able to catalyse the hydrolysis of CMC to reducing sugars, but this activity falls outside of the accuracy and sensitivity of the assay Further work is required here to fully quantify the ability of *Geobacillus* to hydrolyse glucose polymers. Reducing sugar concentration was normalised against a negative control where PBS was added allowing any change in reducing sugar concentration caused by heating at 55 °C to be accounted for.

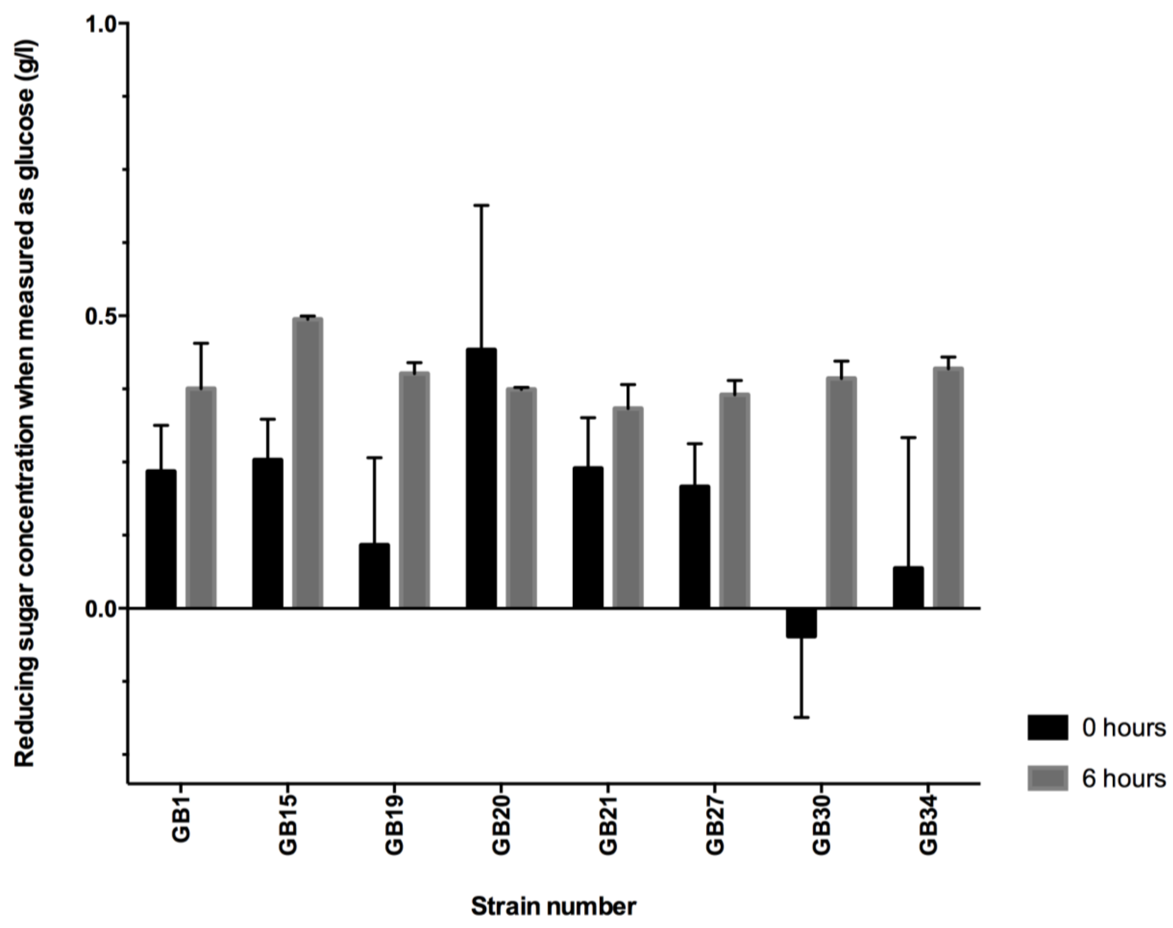


Figure 12 The release of reducing sugars from CMC catalysed by putative cellulase proteins.

The change in reducing sugar concentration following incubation with the soluble supernatant protein fraction of *Geobacillus* spp. for 6 h at 55 °C was measured. Supernatant proteins were precipitated using acetone cooled at -20 °C. This was then pelleted and re-suspended in PBS and incubated at a 1:1 with 0.25 % CMC in 50 mM citrate buffer. Values were normalised against the change in reducing sugar concentration of a negative control. This experiment was carried out in duplicate (technical replicate). These results were shown to be not significantly significant (Paired Two tail T-test).

4.1.4. *Growth profile of* Geobacillus spp. *at 55 °C*

The five strains of *Geobacillus* had similar lag phases when cultivated in TSB medium (Figure 13) (~ 1 hour); GB15 and GB20 had a lag phase of approximately 1.5 h and GB30 which had a lag phase of approximately 45 min. The final optical density of the strains, after 24 h, was between $O.D_{600}$ = 0.63 – 1.22. Most strains except for GB30 were above an $O.D_{600}$ of 0.8. It was also observed that some of the strains show a drop in optical density (GB30, 21, 1). The reason for this has not been confirmed but it is suspected that it is caused by cell flocculation following sporulation or cell lysis. The doubling time of the strains is between 30 and 40 min.

4.1.5. *Transformation of* Geobacillus spp. *by conjugation*

Using the protocol previously described, several strains of *Geobacillus* were transformed with the conjugal plasmid pS797 as shown by the gel electrophoresis image in Figure 14, which shows that a plasmid of the correct size could be purified from *Geobacillus* following conjugation with *E. coli* S17-1, and selection on TSB containing 12.5 µg ml^{-1} kanamycin. The strains that were successfully transformed are GB1, GB15, GB21, GB30 and GB34. Transformed colonies from each conjugation were grown in TSBK, the plasmid was purified and restriction enzymes used to confirm the plasmid size.

4.1.6. *Summary of* Geobacillus *species selection*

Table 56 summarises the *Geobacillus* spp. screening. The first stage in the screening was the plate overlay assay for CMC and xylan degradation as visualised by Congo red staining. Eight strains which showed xylan and CMC degradation were selected for further screening of cellulase activity. This was done by testing CMCase activity of the extracellular protein fraction, glucose polymer utilisation (DP2-DP6) and competence. Following species screening *Geobacillus thermoglucosidans* (GB15). Was selected for further genetic engineering

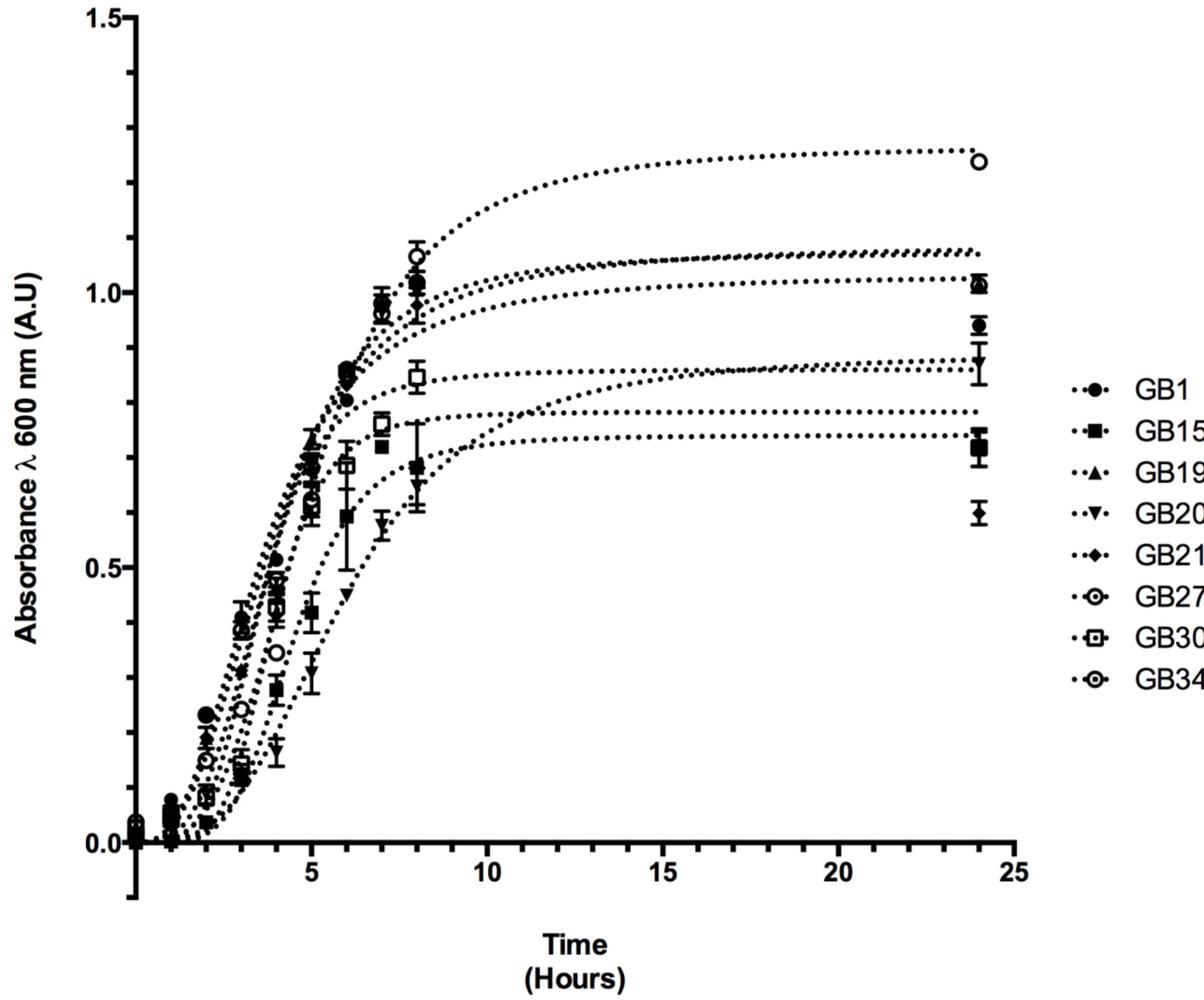


Figure 13 Growth of *Geobacillus* strains at 55 °C over 24 h

The change in optical density of the selected strains of *Geobacillus* when grown in TSB medium at 55 °C in a shaking incubator at 200 rpm. The optical density was recorded every hour for 8 h and then at 24 h. This was recorded using a TECAN M200 plate reader. The data was normalised against a medium blank. This experiment was performed in duplicate. Path length was 0.33 cm. Nonlinear regression curve fitted with Hill Slope. R^2 values were as follows. GB1 = 0.9852, GB15 = 0.9739, GB19 = 0.9942, GB20 = 9891, GB21 = 0.8995, GB27 = 0.9949, GB30 = 0.9548, GB34 = 0.9905.

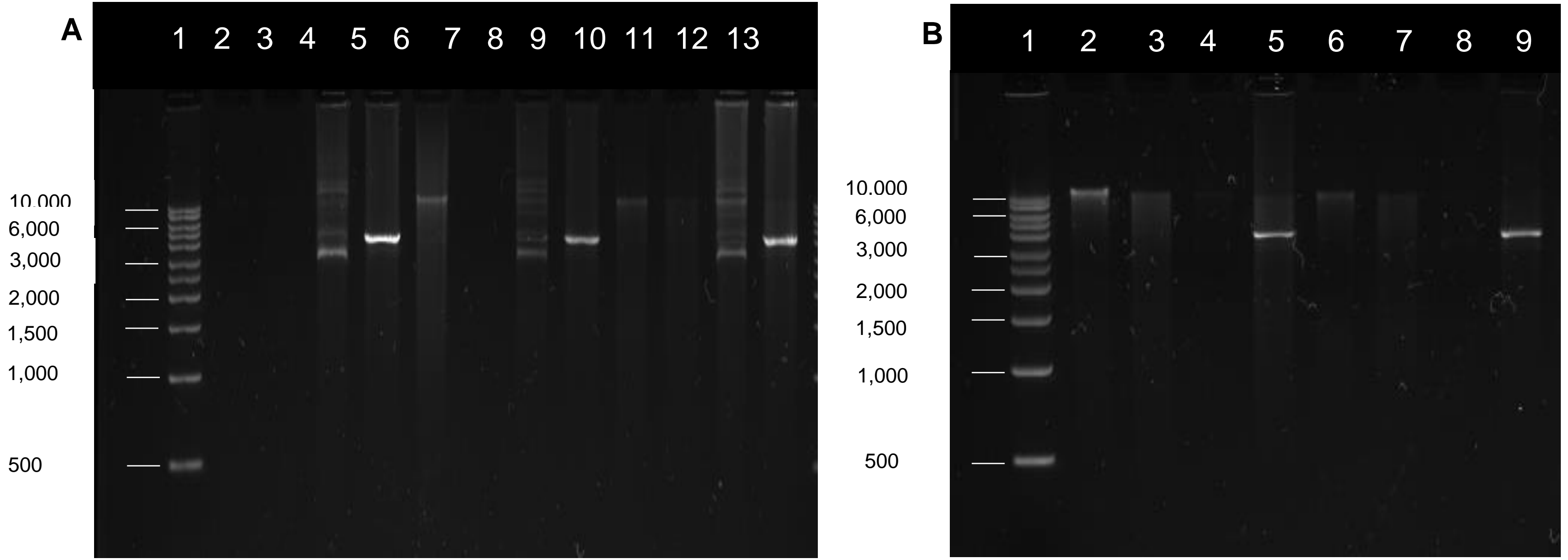


Figure 14 Transformation of *Geobacillus* strains with pS797

DNA purified from *G. thermoleovorans* DSM5366 (GB1), *G. thermoglucosidans* DSM 2542 (GB15), *G. lithuanicus* DSM15325 (GB21), *Geobacillus* sp. ATCC 15952 (GB30) and *G. stearothermophilus* NCIMB 8224 (GB34) was fractionated by gel electrophoresis. The transformed strains are compared to the untransformed wild type. (Lanes are as follows A2-5 GB1, lane 2 is wild type undigested, lane 3 is wild type *Eco*RI digested, lane 4 is pS797 undigested and lane 5 is pS797 EcoRI digest, A6-9 GB15, lanes are as before, A10-13 are GB21, B2-5 are GB30 and B6-9 are GB34. All *Eco*RI digests from the pS797 GB1, GB15, GB21 GB30 and GB34 give a band of the correct size of 4489 bp. DNA weight marker is a Sigma Direct load 1 kb DNA ladder (lanes A1 and B1).

Table 6 Summary of host selection for consolidated bioprocessing of lignocellulosic material.

Strain Name	Strains for further screening	Cryo-storage	Simple medium	Conjugal Competence	Glucose Polymer Utilisation	CMCase-DNS
G. thermoleovorans	**1**	✓	✓	✓	Cellobiose, triose and hexaose	X
G. thermoglucosidans	**15**	✓	✓	✓	Cellobiose	✓
G. gargensis	**19**	✓	✓	X	Cellobiose, triose and hexaose	✓
G. jurassicus	**20**	✓	✓	X	X	X
G. lituanicus	**21**	✓	✓	✓	Cellobiose, triose and hexaose	X
G. toebii	**27**	✓	✓	X	X	✓
***Geobacillus* sp. NCIMB8225**	**30**	✓	✓	✓	Cellobiose	✓
***G. stearothermophilus* N.R.Smith T147**	**34**	✓	✓	✓	X	✓

5. Improving cellulose degradation in *Geobacillus thermoglucosidans* DSM2542.

5.1. Rationale

Geobacillus species were shown to have cellulolytic capabilities (Rastogi *et al*. 2010; Zambare *et al*. 2011). Several of the species that have reported cellulolytic activity are not publically available and have complex intellectual property (I.P) issues and material transfer agreements surrounding their use for commercial purposes. The host highlighted in Chapter 4, *Geobacillus thermoglucosidans* DSM2542, was further engineered to improve cellulose degradation. The genes engineered into *G. thermoglucosidans* DSM2542 were required to operate at the elevated temperatures. At present, control over protein expression in *Geobacillus* is limited by a lack of well characterised, constitutive and inducible promoters and it was important that the required gene network be simple. With these factors in mind the genes shown in Table 7 were selected. The attributes of each of the cellulase proteins are explained below.

CelA from *Caldicellulosiruptor bescii* (previously *Anaerocellum thermophilum* DSM6725 (Yang *et al*. 2010)) was chosen for a number of reasons. It has been shown to have a novel mode of action as it is able to generate cavities in the surface of lignocellulosic material (Brunecky *et al*. 2013) and is essential for growth of *C. bescii* on the insoluble biomass fraction of plant material and untreated lignocellulosic material (Yang *et al*. 2009). *C. bescii* is a Gram positive, thermophilic bacterium and is able to grow at temperatures up to 90 °C (Yang *et al*. 2010), with thermostable cellulase enzymes, published genome, (Kanafusa-Shinkai *et al*. 2013) and has been used as a host for the production of hydrogen from lignocellulosic biomass (Cha *et al*. 2013). The primary cellulase in *C. bescii* is CelA. CelA is highly active up to 80 °C and is multifunctional (Zverlov *et al*. 1998). CelA has endo and exoglucanase functions (Lochner *et al*. 2011) and has even been shown to have xylanase activity (Brunecky *et al*. 2013). CelA has been shown to be the most abundant protein in the *C. bescii* secretome (Yang *et al*. 2009).

Table 7 Cellulase genes selected for transformation into *G. thermoglucosidans* DSM2542.

Species	Encoded Protein	Function	Size (bp / kDa)	Domain Structure
Caldicellulosiruptor bescii	CelA	Endo and exoglucanase activity	5277/195.16	GH9-CBM3-CBM3-CBM3-GH48
Caldicellulosiruptor saccharolyticus	CelB	Endoglucanse		GH10-CBM3-GH5
Clostridium stercorarium	CelZ	Endo 1,4-beta glucanase	2958/106.48	GH9-CBM3-CBM3
Thermobifida fusca	Cel6B	Non-reducing end directed exoglucanase	1788/60.08	GH6-CBM3-CBM3
Thermobifida fusca	Cel9A	Processive endoglucanase	2640/9043	GH9-CBM3-CBM3
Thermotoga neopolitania	CelB	Endo 1,4-betaglucanase	822/29.84	GH12

A second gene was selected for transformation into *Geobacillus* was CelB from *Caldicellulosiruptor saccharolyticus.* This organism was selected as the genome sequence (van de Werken *et al*. 2008) has putative multidomain gylcosyl hydrolases. One of these multidomain glycosyl hydrolases, CelB, was selected from *C. saccharolyticus* because it was shown that when *C. saccharolyticus* was grown on switchgrass CelB was represented in the 10 most abundant transcripts, and was able to degrade xylan, CMC β-glucan and arabinoxylan (VanFossen *et al*. 2011). It is also a putative functional homolog of an ancillary enzyme for which only the partial amino acid sequence could be found in *C. bescii* and so it was hypothesised that the enzymes might work well in combination.

Another candidate is CelZ from *Clostridium stercorarium.* This endo-1,4-beta glucanase has a high level of homology to the endoglucanase domain of *C. bescii* CelA (70.4 %) (Zverlov *et al*. 1998). It is significantly smaller than CelA and has been shown to be able to degrade Avicel and CMC and may have both endo-glucanase and exo-glucanase activity (Bronnenmeier & Staudenbauer 1990). The enzyme is thermostable at 60 °C and no loss of activity was observed after 90 h of incubation (Bronnenmeier & Staudenbauer 1990) and is able to bind to crystalline cellulose, a process which is facilitated by a cellulose binding domain and a thermostablising domain (Riedel *et al*. 1998).

The *CelB* gene from *Thermotoga neoplitana* DSM4359 was chosen for testing in *Geobacillus* because of its similarity to *Cel5A* from *Thermotoga maritima* which has previously been used to improve the cellulolytic capabilities of *Geobacillus* (Bartosiak-Jentys *et al.* 2013).

The final organism from which genes were selected was *Thermobifida fusca* YX. It differs from the other bacteria selected as gene sources because it is aerobic. *T. fusca* has six glycosyl hydrolase genes: *Cel5A, Cel6A, Cel6B, Cel9A, Cel9B and Cel48A*. Of these genes, *Cel9A* was selected because Cel9A showed the highest level of activity against filter paper (Wilson 2004) and its unique structure and processive endoglucanase and exoglucanase function (Sakon *et al*. 1997). When exoglucanase function is observed it has the ability to cleave cellotetraose subunits. This is a substrate

which we expect *Geobacillus* to be able to use, but not necessarily other thermophiles, thus reducing substrate competition with other bacteria. Cel6B was also selected from this organism because it is has processive exoglucanase function (Vuong & Wilson 2009). It is hypothesised that because Cel6B and Cel9A are from the same host, and have complementary function that they would improve cellulose degradation.

It is proposed that by introducing these genes, either individually or in combination, the cellulolytic ability of *Geobacillus* can be improved.

5.2. Results

5.2.1. *Transformation of cellulase genes into* Geobacillus*.*

Geobacillus thermoglucosidans DSM2542 was transformed with pS797 *CelA* (*C. bescii*), pS797 *CelB* (*T. neopolitania*), pS797 *Cel6B* (*T. fusca*), pS797 *Cel9A* (*T. fusca*), using the conjugation protocol as previously described (Cripps *et al*. 2009). Figure 15 shows the purified, digested plasmid, confirming transformation of the *Geobacillus* spp. The selectivity of kanamycin in each of these transformed strains was confirmed by comparison with negative controls, where WT S17-1 *E. coli* were incubated with *Geobacillus*; no growth was observed in the negative controls, indicating that growth on TSA + kanamycin was a result of successful plasmid transfer. Plasmid was purified from 10 ml of overnight liquid culture. DNA was eluted into 30 µl of water, instead of 50 µl in an attempt to increase the plasmid concentration. This plasmid was then digested with *Eco*RI and *Pst*I in single and double digests. Whilst the image is clear for some of the cellulose constructs (Figure 15), most of the expected bands can be observed (pS797 *CelA* 1800 bp, 2000 bp and 4500 bp are present in double digest, pS797 *Cel6B* 500 bp, 1200 bp and 4500 bp are present, pS797 *CelA* 2500 bp and 4500 bp are present. Comparison of the digested DNA from the wild type and transformed strains confirms transformation of *Geobacillus* with *CelA, CelB, Cel6B* and *Cel9A.*

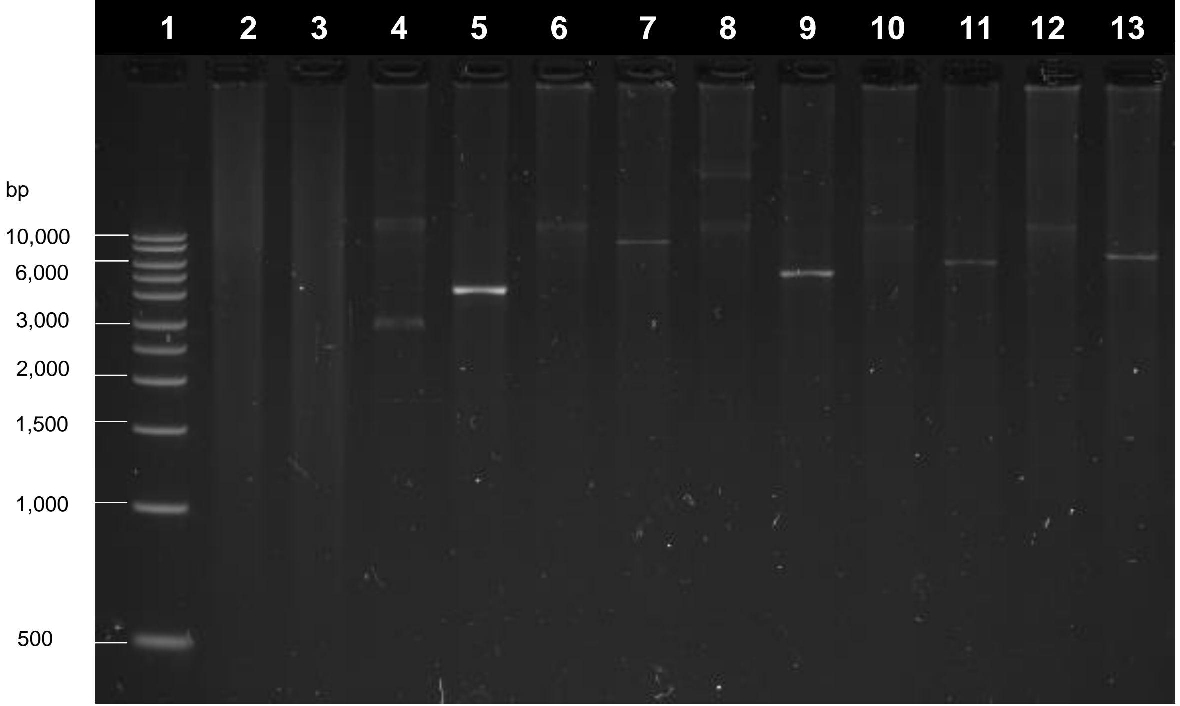


Figure 15 Gel electrophoresis of plasmids purified from *Geobacillus thermoglucosidans* DSM2542 showing transformation with pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A*.

The plasmid DNA was purified from 10 ml of a 16 hour liquid culture of each of the respective strains. The plasmid DNA was then digested with *Eco*RI. The sizes of all the digested bands are of the expected size for each plasmid (pS797 = 4489 bp (lane 5), pS797 *CelA* = 9986 bp (lane 7), pS797 *CelB* = 5531 bp (lane 9), pS797 *Cel6B* = 6497 bp (lane 11) and pS797 *Cel9A* = 7349 bp (lane 13). This gel demonstrates the correct transformation of each of the respective cellulase genes. Wild type control (lane 3) and no digest controls are included (lane2 2, 4 ,6 ,8 ,10 and 12) DNA Weight marker is Sigma Directload 1 kb (lane 1).

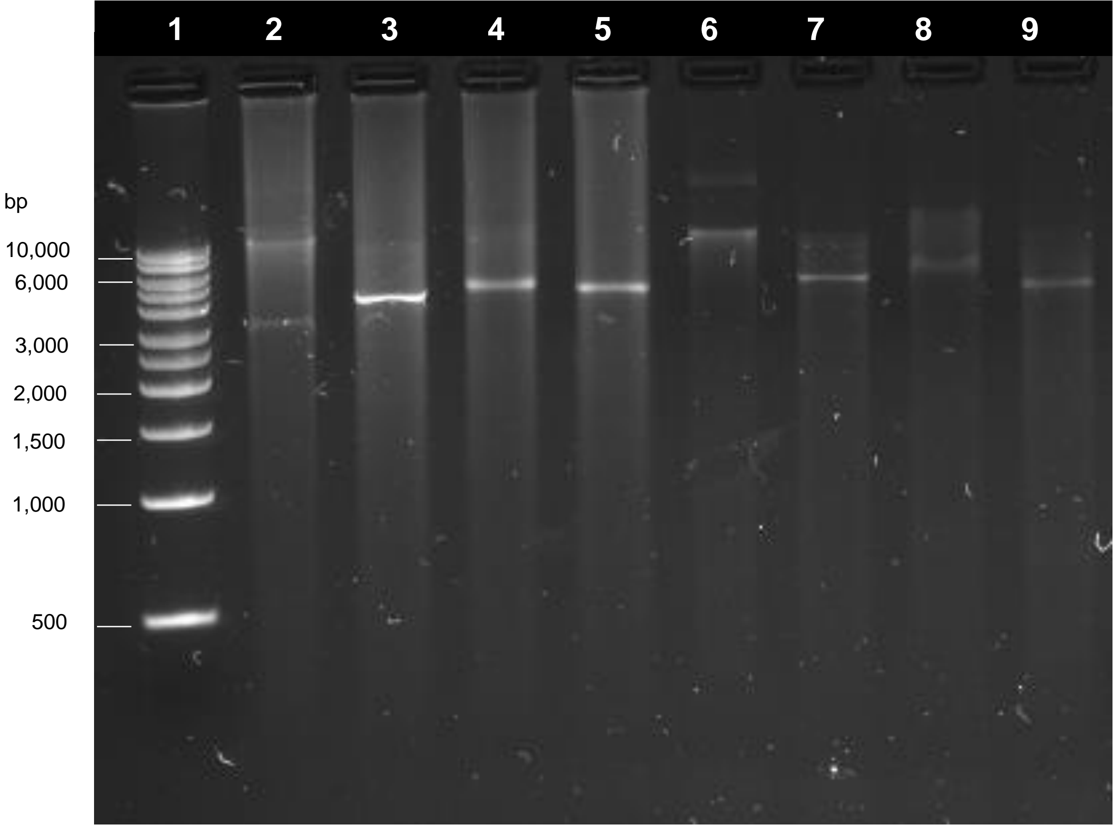


Figure 16 Successful recovery of pS797 *CelB* following plasmid loss.

This figure shows pS797 empty vector and pS797 *CelB* purified from *Geobacillus thermoglucosidans* DSM2542. Lane 2-5 pS797 undigested, *Eco*RI digest, *Pst*I digest and EcoRI + PstI double digest. Lanes 6-9 pS797 *CelB* are undigested, *Eco*RI digest, *Pst*I digest and *Eco*RI + *Pst*I double digest respectively. Size shift shown in lanes 3 and 4 compared to 7 and 8 of corresponds with the size of *CelB* (900 bp). DNA Weight marker is Sigma Directload 1 kb (lane 1).

During experimentation it was observed that *G. thermoglucosidans* pS797 *CelB* had lost its plasmid (data not shown). The reason for the loss of plasmid in *G. thermoglucosidans* DSM2542 pS797 *CelB* (Fig. 16) is unknown but it is suspected that long term storage of transformed strains on agar slopes may have resulted in the loss of plasmid. The analysis was repeated with cultures grown from a glycerol stock of the bacteria. Because the pS797 *CelB* slope cell line was still able to grow on kanamycin, despite plasmid loss, it is suspected that the *Geobacillus* may have incorporated the kanamycin selection marker into its genome.

5.2.2. *Cellulase expression and activity in* Geobacillus*.*

Following successful transformation of cellulase constructs into *Geobacillus* protein expression was confirmed by whole cell lysate SDS-PAGE (figure 17) and western blotting. Activity was then confirmed using an endoglucanase activity, well diffusion assay (Figure 18). SDS-PAGE showed that proteins corresponding to the size of CelA and CelB were present in the pS797 *CelA* and pS797 *CelB* strains of *Geobacillus thermoglucosidans* (Figure 17). Following the addition of His tags (Figure 19 for and Figure 20 for western blot (see appendix 2 and 3 for plasmid digest gel electrophoresis to confirm size and sequencing of His-tag incorporation site for confirmation of correct incorporation))protein expression and size was confirmed by western blotting using an anti- HIS HRP conjugated antibody.

The function of each of the cellulase genes was tested using a well diffusion CMC staining assay (Figure 18. CelA, CelB and Cel9A were all shown to be functional using the assay. Cel6B is reported to be an exoglucanase and so activity against CMC was not expected. It was included in this experiment to test the validity of the literature. The data displayed in Figure 18 is not quantitative and the estimation of the halo size is highly subjective. Halo size (Defined as a change in colour different to that observed in the pS797 empty vector controls, 4 separate measurements taken at 0, 90°, 180° and 270°): CelA supernatant 8.5 mm, cell lysate 2 mm; CelB supernatant 8.5 mm, cell lysate 3.5 mm; Cel6B supernatant 0 mm, cell lysate 0 mm; Cel9A supernatant 5.75 mm, cell lysate 0 mm. Based on this data CelA and CelB have higher rates if clearance than Cel9A.

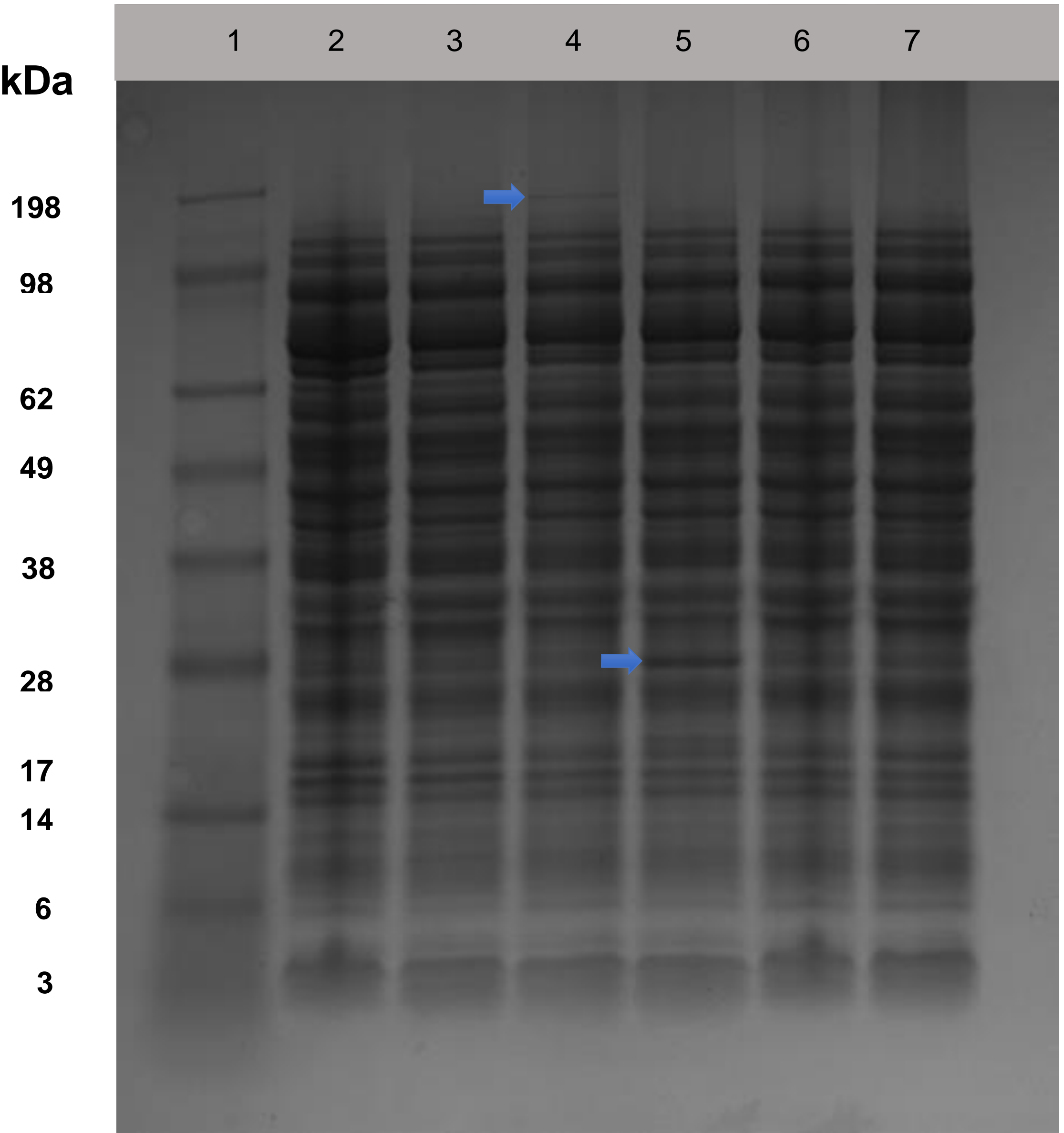


Figure 17 SDS PAGE showing the production of CelA and CelB by *Geobacillus thermoglucosaidans* DSM2542.

SDS PAGE (NuPAGE 4-12 % Bis-TRIS), showing the expression of cellulase proteins in *Geobacillus thermoglucosidans.* Each well contains 5 µl of the soluble fraction of whole cell lysate from 1 ml of a 24 hour culture of WT, pS797 and pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A*. CelA and CelB are indicated by arrows; no obvious band can be seen for Cel6B and Cel9A. Lane 1) SeeBlue Plus2 protein marker (kDa), lane 2) WT, lane 3) pS797 empty vector, lane 4) CelA, lane 5) CelB, lane 6) Cel6B, lane 7) Cel9A. Bands were visualised using Coomassie blue.

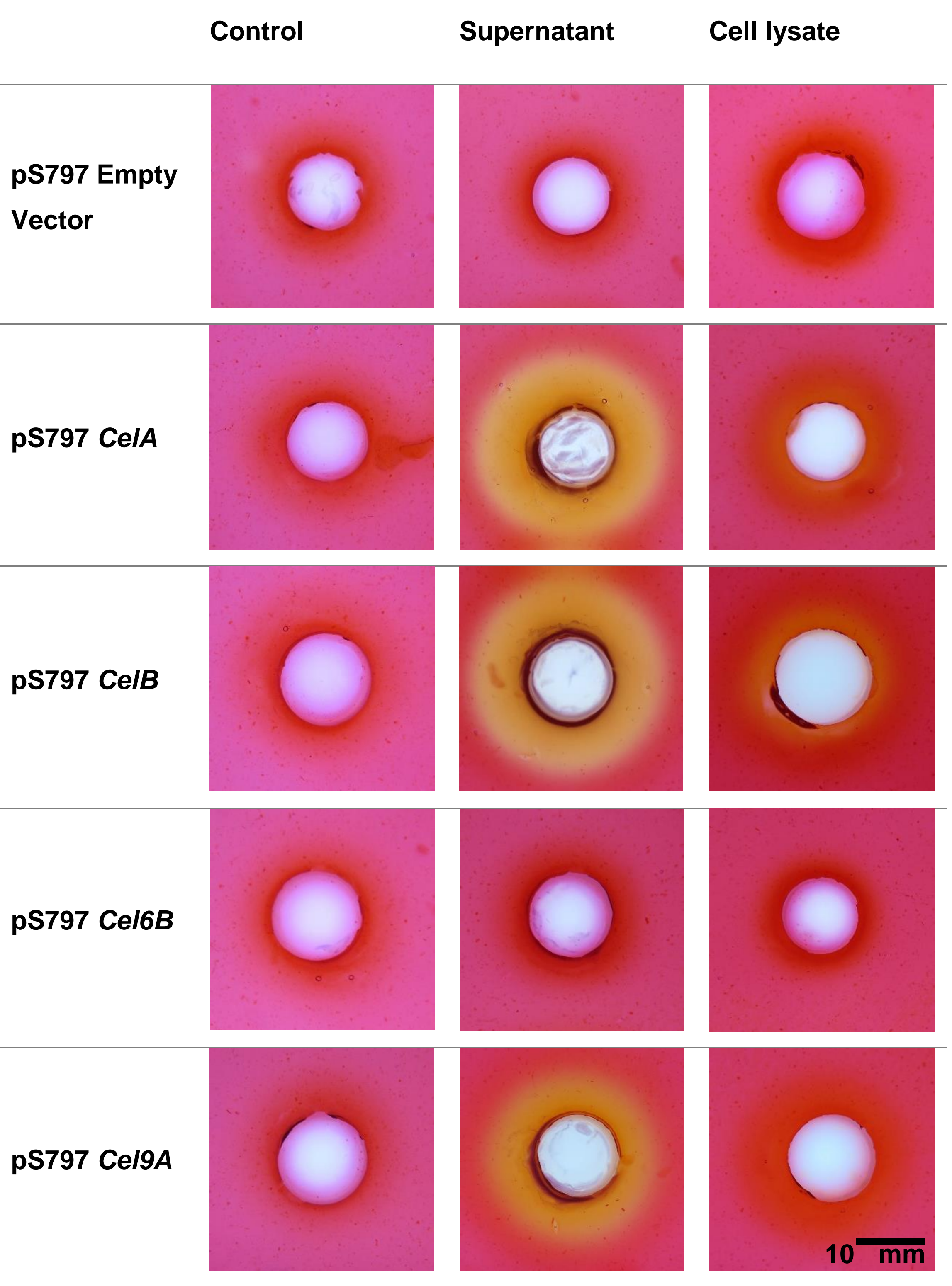

Control
Supernatant
Cell lysate
pS797 Empty Vector
pS797 *CelA*
pS797 *CelB*
pS797 *Cel6B*
pS797 *Cel9A*
10 mm

Figure 18 Degradation of CMC by CelA, CelB, Cel6B and Cel9A in culture supernatant from *G. thermoglucosidans* DSM2542 and cell pellet after 24 h.

The table above shows the ability of culture supernatants and cell pellets of *Geobacillus* to alter the staining of CMC by Congo red (*Geobacillus thermoglucosidans* DSM2542 WT, pS797, pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A*). This alteration of staining indicates CMC degradation. This can be clearly observed as a diffuse halo around the perimeter of the wells. The plate composition is 0.25 % CMC and 0.7 % agar. Following cell harvesting by centrifugation 250 µl of supernatant was transferred to a well in the plate. Wells were made by placing HPLC vials on an agar plate before pouring of molten agar, vials were then removed, leaving a hollow in the agar of a fixed volume (approximately 500 µl total volume). The remaining pellet was resuspened in fresh TSB and this was transferred to the plate. The plate was incubated for 24 h at 55 °C in a static incubator. The plate was washed with Congo red to stain, for 10 min and destained with 1M sodium chloride for 20 min. Halo size (Defined as a change in colour different to that observed in the pS797 empty vector controls, 4 separate measurements taken at 0, 90°, 180° and 270°): CelA supernatant 8.5 mm, cell lysate 2 mm; CelB supernatant 8.5 mm, cell lysate 3.5 mm; Cel6B supernatant 0 mm, cell lysate 0 mm; Cel9A supernatant 5.75 mm, cell lysate 0 mm.

The activity of CelA is of particular interest for a number of reasons. Firstly, it is a very large open reading frame (ORF) (5.2 kb). This is larger than anything that has previously been expressed in *Geobacillus thermoglucosidans* DSM2542 by this group. Secondly, it is a multidomain multifunctional protein that has endo- and exo-glucanase function with the highest cellulolytic activity of all the *Caldicellulosiruptor* cellulase enzymes (Young *et al.* 2014). This dual function of this protein makes it of particular interest as its use may allow for a greatly simplified cellulose degradation system.

Figure 18 also suggests that *G. thermoglucosidans* is able to use putative signal peptides found on the N-terminus of the cellulase proteins (Table 8) to target the protein into the extracellular space. It can be seen that the majority of cellulase activity is observed in the culture supernatant. It is possible that this release of protein is caused by cell lysis, but it is predicted that it is a result of active protein targeting. This is something that requires further investigation, possibly by using LC/MS and tryptic digestion to identify the proteins present in the supernatant.

5.2.2.1. *Confirmation of cellulase expression by western blotting*

The pS797 + *CelA, CelB, Cel6B, Cel9A* and *CelZ* were amplified (Figure 19) using the His-tag Q5 primers (Table 1). Following successful PCR amplification of the pS797 + cellulase gene constructs with the site directed mutagenesis primers the PCR product was purified using a PCR clean up kit and the KLD reaction was carried out as previously described. This was subsequently transformed into *E. coli* DH5α. Single colonies were then grown and plasmid DNA purified. This plasmid was digested and shown to be of the correct expected size (appendix 2). Following size confirmation plasmid DNA was sent for Sanger sequencing (using the M13 fwd sequencing primer site in pS797 (SeqWright, GE)) and the presence of the 6xHis tag between the C terminal end of the respective cellulase coding sequence and the stop codon (TAATAA) was confirmed. Alignments of the original sequence and mutated sequence data and gel images of the digested plasmid are shown in appendix 3.Western blot shows the presence of the proteins CelA ~190 kDa, CelB 30 kDa and Cel9A ~90 kDa. Cel6B ~60 kDa is shown but with another band at ~ 45 kDa (Figure 20). This band is not

Table 8 Predicted signal peptide cleavage sites, secretion pathway and localisation target as predicted by PSORT, SignalP and TatA.

	SignalP	**SignalP D-Score**	**cleavage position**	**TatA**	**cleavage position**	**Psort**
***Thermobifida fusca* YX Cel6B**	Yes	0.745	33-34	potential TAT	33-34	Unknown
***Thermobifida fusca* YX Cel9A**	Yes	0.645	46-47	Yes TAT	46-47	Extracellular
***Caldicellulosiruptor bescii* DSM6725 CelA**	Yes	0.887	24-25	No	-	Extracellular
***Caldicellulosiruptor saccharolyticus* CelB**	Yes	0.63	28-29	No	-	Extracellular
***Clostridium stercorarium* DSM8532 CelZ**	Yes	0.644	27-28	No	-	Extracellular
***Thermotoga neapolitana* DSM4359 CelB**	Yes	0.755	17-18	No	-	Unknown

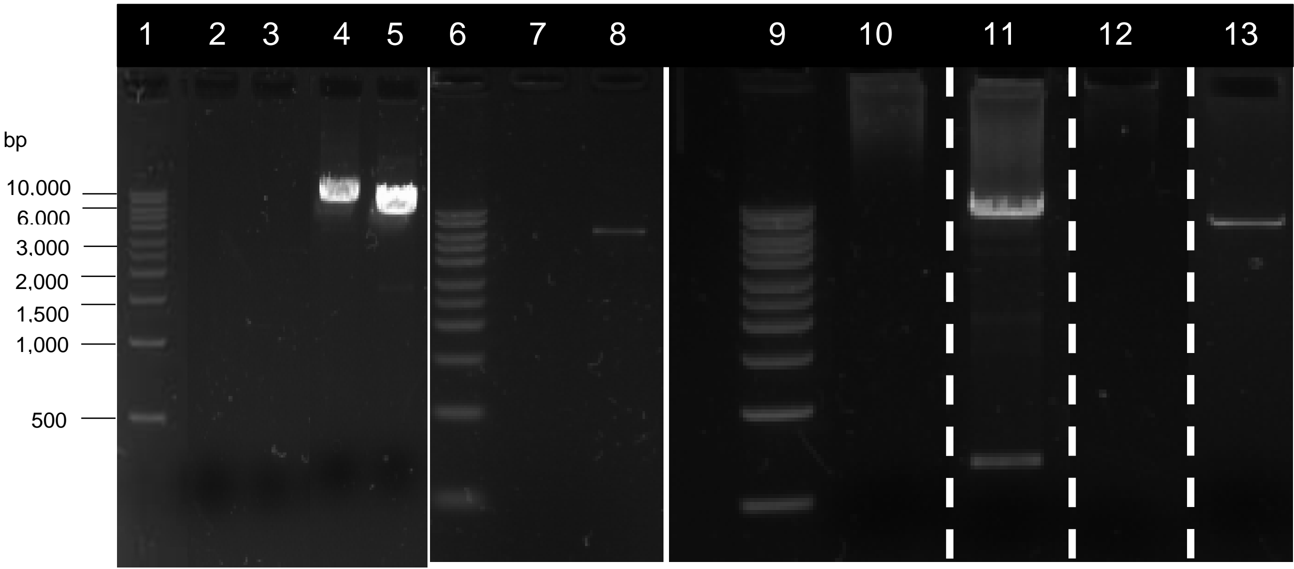


Figure 19 PCR amplification of the 5 cellulase genes for His Tag incorporation by Q5 site directed mutagenesis.

All pS797 cellulase vectors have been successfully amplified, and yield bands of the correct size (pS797 *Cel9A* = 7349 bp, pS797 *CelB* = 5531, pS797 *Cel6B* = 6497, pS797 *CelA* = 9986 bp, pS797 *CelZ* = 7447). Image has been compiled from several gel images. Images have not been modified in any other way. Cropping removed lanes in which the PCR was not successful or where gradient PCR was carried out. If gradient PCR has been carried out the temperature which yielded the brightest bands is shown (pS797 *CelZ* = 64 °C). Gel electrophoresis was at 120 volts for 90 min (pS797 *Cel9A* and *CelB* were run for 60 min).Diagnostic digests of the resulting plasmids are shown in appendix 2.. Lane 1 = Sigma Directload 1 kb ladder, Lane 2 Cel9A negative control, lane 3 = pS797 CelB negative control, lane 4- = pS797 Cel9A + His tag, lane 5 = pS797 CelB + his tag, lane 6, Sigmadirect load 1 kb ladder, lane 7 = pS797 Cel6B negative control, lane 8 pS797 Cel6B + His tag, lane 9 = Sigma directload 1 kb ladder, lane 10 pS797 CelA negative control, lane 11 = pS797 CelA + His tag, lane 12 = pS797 CelZ negative control, lane 13 = pS797 CelZ +His tag.

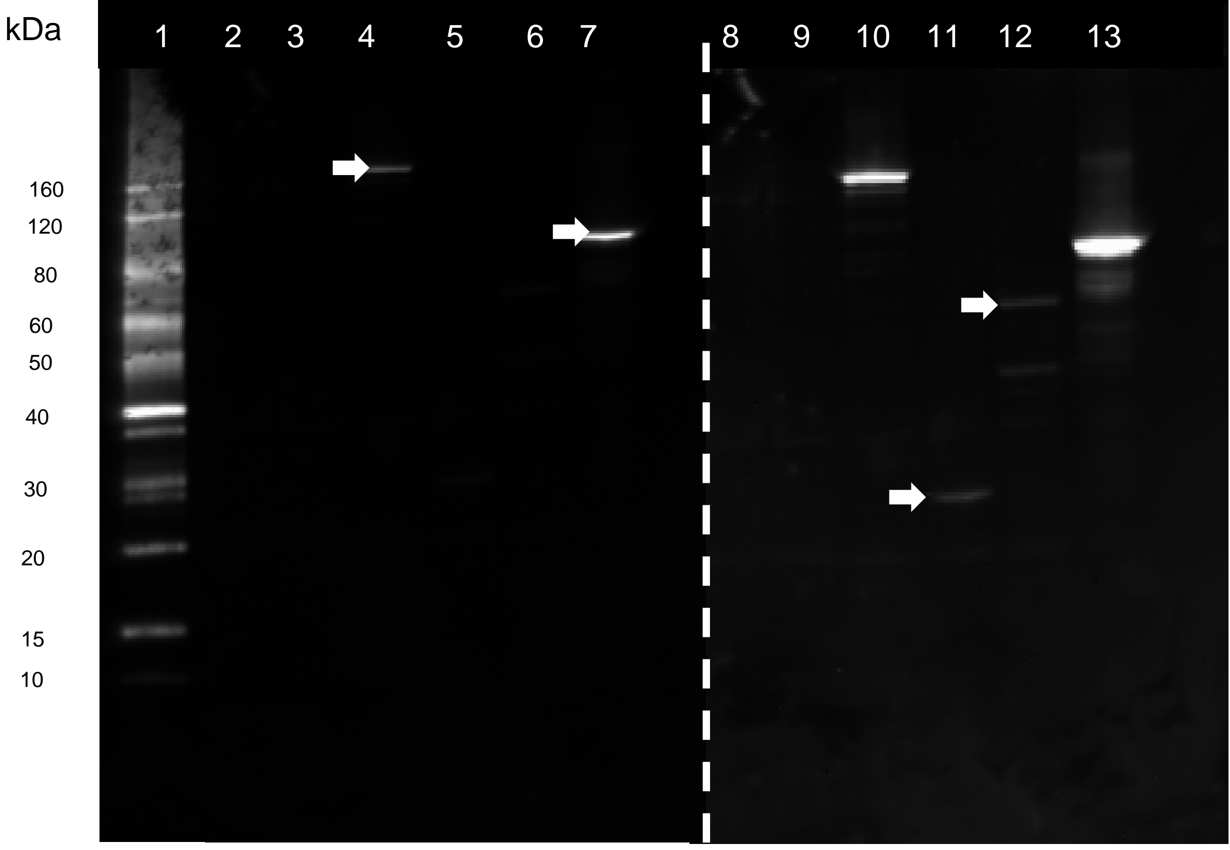


Figure 20 Western blotting of His tagged cellulase protein in *Geobacillus thermoglucosidans* DSM2542.

Blotted proteins were probed with anti-His HRP conjugated antibody at 1:400. Bands can be clearly seen for CelA ~190 kDa, CelB ~30 kDa and Cel9A ~90 kDa. Cel6b has a correct band at ~ 60 kDa and a smaller band at~50 kDa; because of the lack of this band in the wild type control it is suspected that this band is a product of degraded Cel6B. The two images shown are of the same western blot, the left hand image was exposed for 19 s and the right hand image for 2.5 min. The reason for this was the lower abundance of CelB and Cel6B. Protein samples were from whole cell lysate. No further purification was carried out. Lane 1 = Benchmark His tagged protein ladder, lane 2 = *G. thermoglucosidans* Wild Type, lane 3 = pS797 Empty Vector control, lane = 4 CelA, lane 5 = CelB, lane 6 = Cel6B, lane 7 = Cel9A, lane 8 = *G. thermoglucosidans* Wild Type, lane 9 = pS797 Empty Vector, lane = 10 CelA, lane 11 = CelB, lane 12 = Cel6B, lane 13 = Cel9A.

observed in any of the other samples suggesting that either there is a cellular response to the presence of Cel6B that triggers production of a native protein, that is also bound by the anti-His antibody, or that Cel6B is degraded to give a ~45 kDa fragment which is the bound by the antibody.

5.2.3. *Zymography of cellulase proteins from G. thermoglucosidans*

The zymographic analysis of *G. thermoglucosidans* pS797 *CelA,* pS797 *CelB,* pS797 *Cel6B* and pS797 *Cel9A* supernatant and cell lysate shows that CelA, CelB and Cel9A are active against CMC, the wild type and pS797 samples show no zones of clearing suggesting no CMCase activity. As shown in Figure 21 all three proteins show some evidence ofCMC degradation, as shown by the unstained regions in both the cell lysate and supernatant samples. However only CelB shows a clearly defined clearing zone (A, lane 5). The cause of the poor resolution and smearing of the other areas of CMC clearing is unknown. This requires further investigation, to give a clear image of what is happening.

5.2.4. *Activity of heterologous cellulase proteins in* Geobacillus *at different temperatures*

A well diffusion assay was carried out at three different temperatures to test the activity of each of the proteins at elevated temperatures. Plates were incubated at 55 °C, 65 °C and 75 °C. The results of this experiment are shown in Figure 30. It can be clearly seen that both CelA and CelB are active at approximately the same levels at the three incubation temperatures as the zone of clearing is still present at all three temperatures (Figure 22). Halo size (Defined as a change in colour different to that observed in the pS797 empty vector controls, 4 separate measurements taken at 0, 90°, 180° and 270°); are as follows. CelA: 55°C = 7.5 mm, 65°C = 9 mm, 75°C 10 mm, CelB: 55°C = 7.5 mm, 65°C = 6 mm, 75°C = 9 mm, Cel9A 55°C = 4.5 mm, 65°C. Cel9A showed no detectable activity at 75 °C. This was unexpected as Cel9A from *T. fusca* has been previously shown to be active against Avicel at 75 °C, with its optimum temperature being 70 °C (Mingardon *et al*. 2011). This requires further investigation

5.2.5. *The effect of cellulase expression on cell growth*

In order to test the effect of constitutive cellulase expression on the growth of *Geobacillus thermoglucosidans* DSM2542, a growth curve of the WT, empty vector

(pS797) and the four cellulase strains was carried out (Figure 23). It was found that both the expression of cellulase proteins and the burden of harbouring the empty vector had an effect on culture growth. Some cellulases had a different effect on the fitness, with CelB having the greatest effect. The pS797 *CelB* strain had a lag phase of over 8 h. pS797 *CelA* had a lag phase of 6 h, and *Cel6B*, *Cel9A* and empty vector had a lag phase of 5 h. Compared to the wild type *Geobacillus* which had a lag phase of only 2 h it is clear that the burden of plasmid replication, antibiotic selection and cellulase expression is great. Whilst antibiotic selection pressure and plasmid burden cannot be mitigated it is possible to reduce the burden of cellulase expression on growth by switching from a constitutive system to an inducible system. It was for this reason that the arabinose inducible promoter, from the University of Exeter (sequence courtesy of Professor John Love) was investigated. Only expression of CelA and CelB seem to have an effect on the growth rate.

It is interesting that CelA and CelB have a similar effect on growth. CelA is the largest cellulase protein expressed, whilst CelB is the smallest. It may be the burden is similar because the total protein (g/l) being produced is the same but the protein concentration differs between the strain (M). This hypothesis is consistent with the SDS-PAGE data shown in Figure 17 as CelB is more abundant than CelA. This is an interesting observation. Future research should investigate the total protein and protein concentration of the respective strains. It may also be that the different ORFs impact the promoter activity in different ways. This could be investigated using RT-Q PCR.

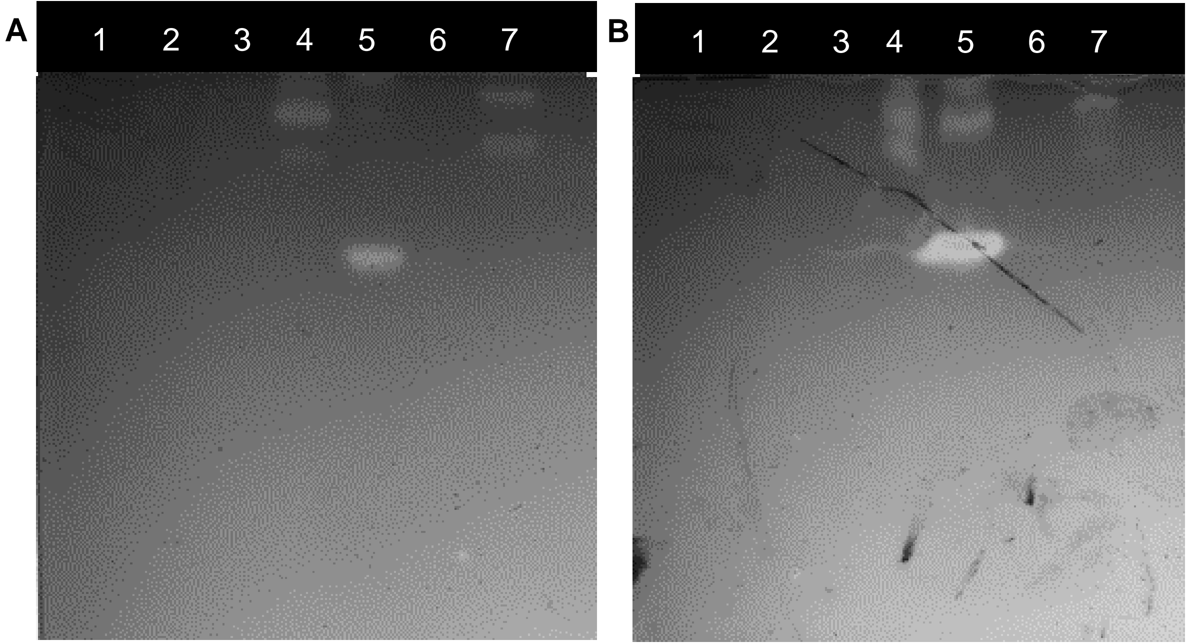


Figure 21 Zymography demonstrating active cellulase expression in *G. thermoglucosidans.*

The activity of CelA, CelB, Cel6B and Cel9A was confirmed by zymography. Supernatant and cell lysate proteins were resolved on 4-10 % stacking Bis-Tris SDS PAGE + 0.1 % CMC for 35 min at 160 volts. Following electrophoresis proteins were renatured in 2.5 % (v/v) Triton X-100, sodium phosphate buffer (50 mM pH 6.5). The buffer was then replaced and the gel was incubated for 12 h at 55 °C. The gels were stained with 0.1 % Congo red for 20 min and destained with 1 M NaCl for 20 min (Assareh *et al.* 2012)**.** Gel A = supernatant samples. Lanes are as follows: 1 = SeeBlue Plus Protein ladder, 2 = *G. thermoglucosidans* Wild type, 3 = pS797 vector control, 4 = pS797 *CelA,* 5 = pS797 *CelB,* 6 = pS797 *Cel6B,* 7 = pS797 *Cel9A.* Gel B = cell lysate samples. Lanes are as follows: 1 = SeeBlue Plus Protein ladder, 2 = *G. thermoglucosidans* Wild type, 3 = pS797 vector control, 4 = pS797 *CelA,* 5 = pS797 *CelB,* 6 = pS797 *Cel6B,* 7 = pS797 *Cel9A.* The light areas on the gel indicate clearing of CMC and therefore cellulase activity.Marker sizes could not be properly assigned to this gel as the ladder diffused from the gel image during the renaturing and incubation process.

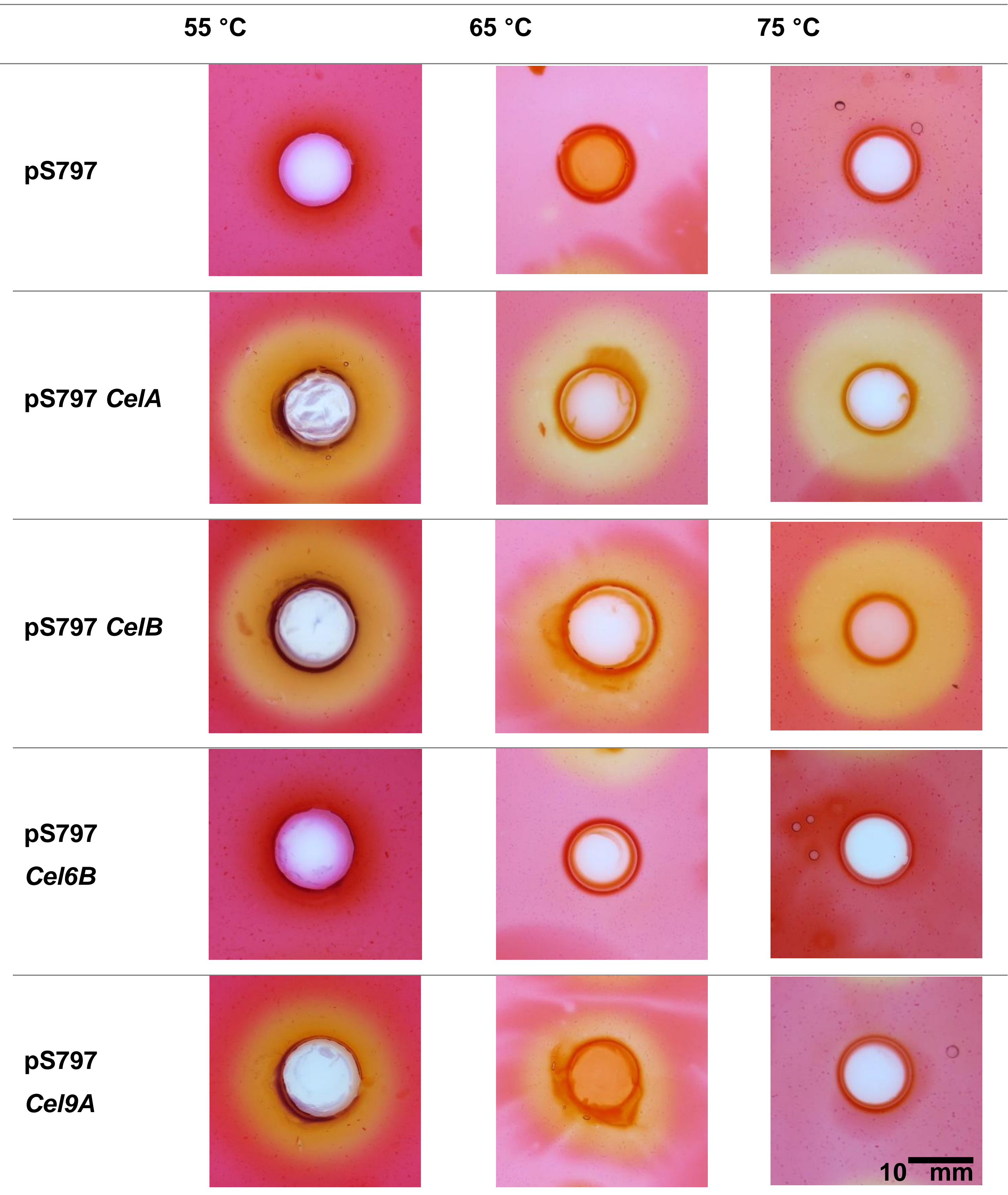


Figure 22 CMCase activity of cellulase proteins (CelA, CelB, Cel6B and Cel9A) in *Geobacillus thermoglucosidans* culture supernatant at different temperatures.

This figure shows the CMCase activity of *Geobacillus thermoglucosidans* culture supernatant after 24 h of growth in TSB +/- kanamycin at 55° c, when transformed with pS797 empty vector, pS797 *CelA*,pS797 *CelB*, pS797 *Cel6B* and pS797 *Cel9A* at 55 °C, 65 °C and 75 °C. Activity was tested using the well diffusion assay as previously described. Halos around the wells show zones of clearing highlighting degradation of CMC by the enzymes. The stain used is 0.1 % Congo red. The data for the wild type control can be found in appendix 4.

5.2.6. *Exchange of P_{ldhA} with the arabinose inducible P_{araE} promoter by Gibson Assembly*

The *ldhA* promoter which controls CelA expression was successfully exchanged, by Gibson assembly (Gibson *et al*. 2009)with the P_{araE} promoter. P_{araE} is an arabinose inducible promoter taken from *Geobacillus thermodenitrificans.* It was shown to switch on gene expression in the prescence of arabinose (Exeter, unpublished data). By placing *CelA* under the control of this inducible promoter, protein expression can be uncoupled from growth. It is hypothesised that by growing *Geobacillus* to a high cell density and then inducing growth a more efficient protocol for CelA expression can be developed.

5.2.7. *Part amplification and Gibson assembly*

Firstly, P_{araE} and pS797 *CelA* backbone were amplified by PCR as shown in Figure 24 part B and C. The resulting parts were assembled using Gibson assembly. The *E. coli* DH5α colonies from the Gibson assembly culture and plasmid DNA was purified, the resulting plasmid was then digest with NdeI and EcoRI

The purified plasmids from 6 colonies of each strain were digested with *Eco*RI and *Nde*I in a single and double digest. *Nde*I cut within the area and nowhere else in the pS797 *CelA* plasmid. *Eco*RI cut once an *E. coli* TOP10 were shown to be positive. Following plasmid digest with *Eco*RI and *Nde*I to confirm the size (appendix 2). Selected samples were then sent for sequencing with the following sequencing primers; araE fwd – CCTACAGCGTGAGCTATGAGAAAGC and P_{araE} Rev - GTCAAACCAACCACCCGTCAGG. Sanger sequencing of the pS797 P_{araE}::*CelA* showed that the P_{araE} had been successfully inserted into the pS797 *CelA* plasmid in the correct location (Figure 25) . There have been no indels or concantamer insertion in the cloning site. It is possible that the PCR of the pS797 *CelA* has introduced errors into the vector. However given the low error rate and proof reading of Phusion polymerase this is unlikely (error rate = 4.4 $\times 10^{-7}$(NEB)). Following sequencing the correct plasmid was transformed into TOP10 for storage and S17-1 for conjugation into *Geobacillus.* Colonies from both TOP10 and S17-1 were all shown to contain correctly assembled plasmid. The *E. coli* S17-1 transformed with pS797 P_{araE}::*CelA* was then used to conjugate *G. thermogulcosidans.* Plasmid DNA was purified from the resulting

Geobacillus and digested with *Nde*I and *Xba*I (Figure 26). This showed that the correct size plasmid was present.

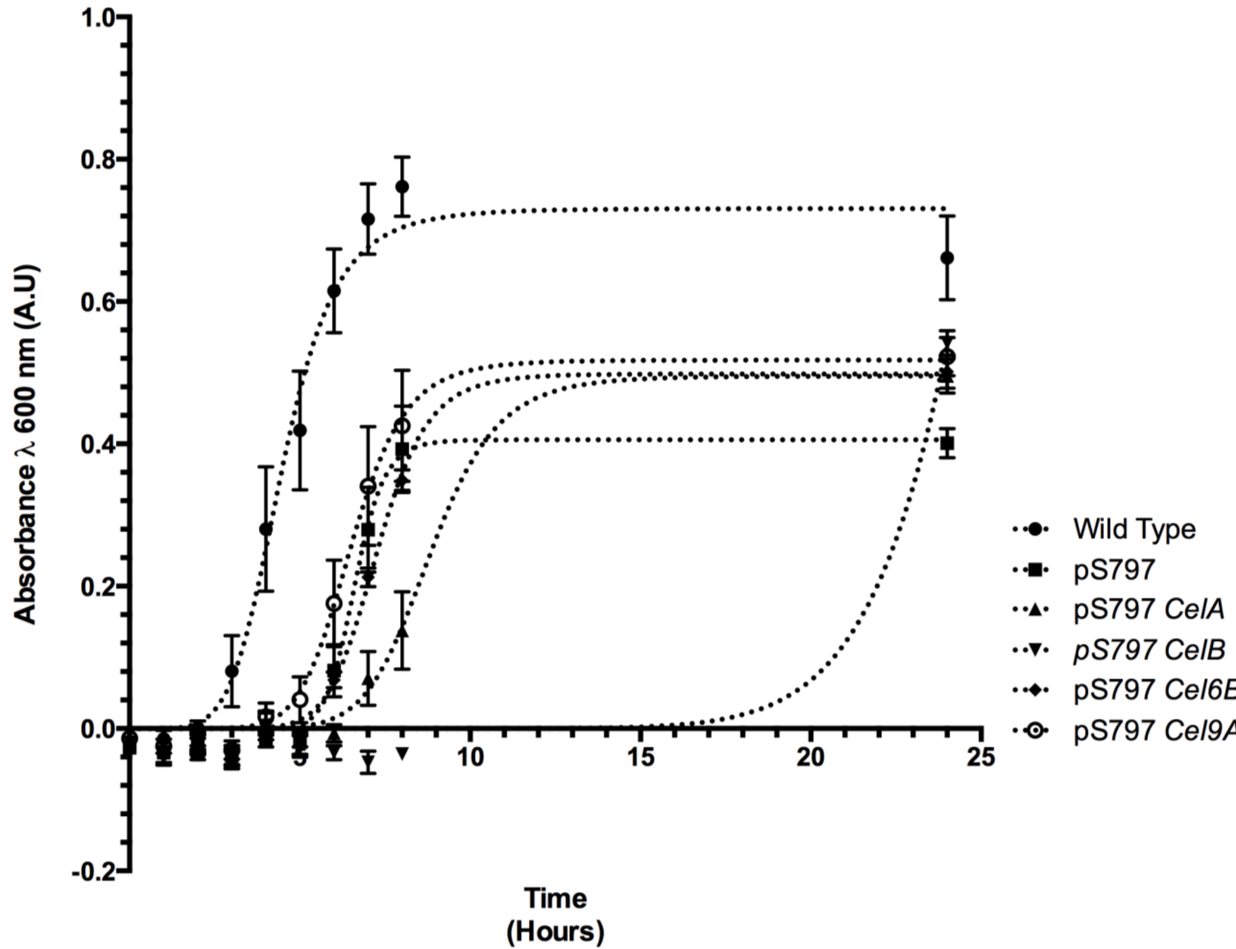
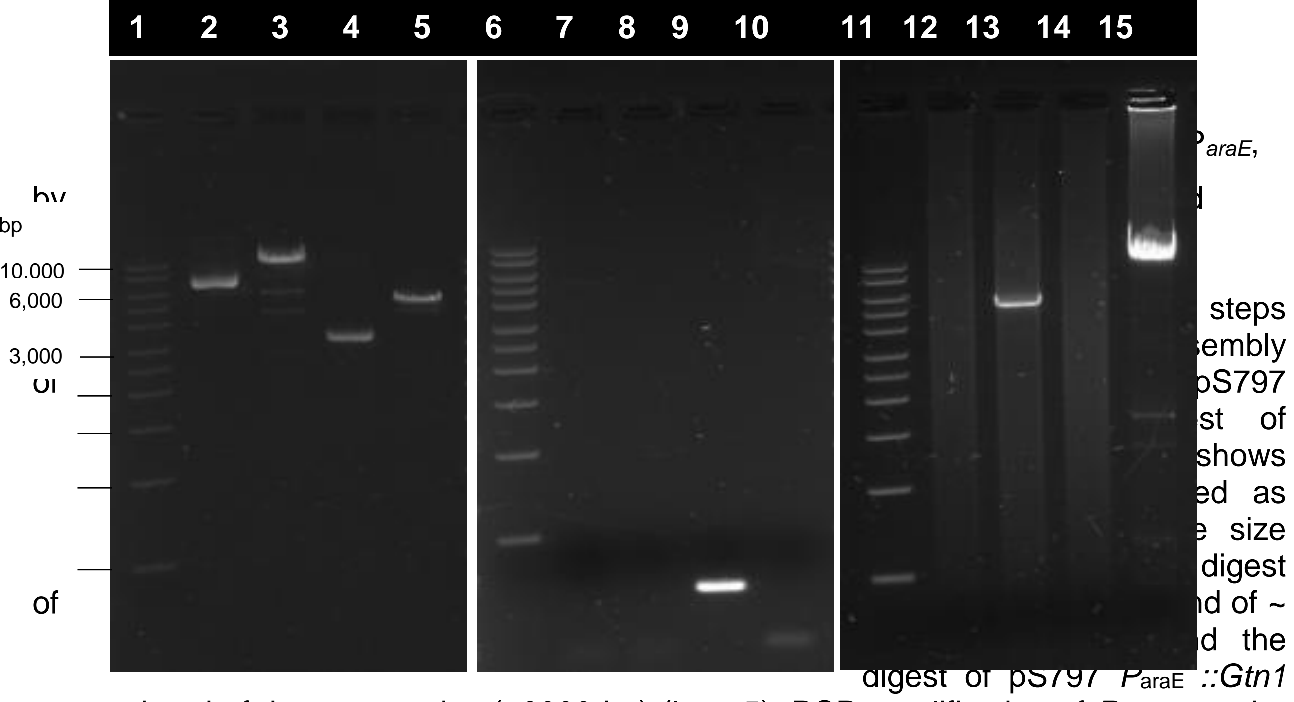


Figure 23 Growth profile of *Geobacillus thermoglucosidans* DSM2542, and the effect of cellulase (CelA, CelB, Cel6B and Cel9A) expression on lag phase length (h).
The impact of constitutive cellullase expression was investigated by observing the change in optical density following growth at 55 °C for 24 h in 50 ml of TSB medium. The optical density of the cultures was measured using a TECAN M200 infinite plate reader and normalised against a medium blank. 250 µl samples were taken for each reading. This experiment was carried out in triplicate and each reading was taken in duplicate giving 6 readings per strain at each time point. Data points represent the mean of these replicates. Error bars represent standard error. Data is fitted using a Hill Slope. R^2 = wild type = 0.8706, pS797 = 0.8901, pS797 *CelA* = 0.900, pS797 *CelB* = 0.9780, pS797 *Cel6B* = 0.9767, pS797 *Cel9A* = 0.8388. Blank corrected against TSB medium. The cause of the negative reading early in the growth curve is unknown as these were blanked against the same media used for growth.



digest of pS797 P_{araE} ::*Gtn1* gave a band of the correct size (~6000 bp) (lane 5). PCR amplification of P_{araE}gave the expected PCR band forP_{araE} (250 bp) and pS797 *CelA* (9500 bp) are present (lane 9 and lane 15) indicating correct part amplification. Problems were encountered when amplifying pS797 *CelA* from plasmid stocks (lanes 7, 8, 12 and 13), plasmid was freshly prepared and this led to full amplification (Lane 15) All gels shown were a 1 % agarose gel run at 120 volts. Lane 1 = Sigma Directload 1 kb ladder, lane 2 = undigested pS797 *CelA*,lane 3 = *Eco*RI digested pS797 CelA, lane 4 = undigested pS797 P_{araE} *:: Gtn1, lane 5* = *Eco*RI digested pS797 P_{araE}*::Gtn1,* lane 6 = Sigma directload 1 kb ladder, lane 7 pS797 *CelA* negative control (failed PCR), lane 8 = pS797 *CelA* (failed PCR), lane 9 = amplified P_{araE},lane 10 = P_{araE} negative control, lane 11 = Sigma directload 1 kb ladder, lane 12 = pS797 *CelA* negative control (failed PCR- old template), lane 13 = pS797 *CelA* (failed PCR, gave truncated band size with old template), lane 14 = pS797 *CelA* negative control, lane 15 = amplified pS797 *CelA*

```
CelA     GGANAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG     80
pS797    GGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG     1560

CelA     GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG     140
pS797    GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG     1620

CelA     ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTT     200
pS797    ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTT     1680

CelA     TTTACAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGGGAGCTGACCTCA     260
pS797    TTTACAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGGGAGCTGACCTCA     1740

CelA     AAAGCCGCCTGCCTTCCACAGACATACTAGATATCGATTCAAGGAAGATTCATTGTTTAC     320
pS797    AAAGCCGCCTGCCTTCCACAGACATACTAGATATCGATTCAAGGAAGATTCATTGTTTAC     1800

CelA     ATATGGTAATCAAAAAGGGTGTCCCAATTTGTAGGGACATCTTTCTTTTTTTGAAAACGC     380
pS797    ATATGGTAATCAAAAAGGGTGTCCCAATTTGTAGGGACATCTTTCTTTTTTTGAAAACGC     1860

CelA     TTCAATAAAGTTAAATTTTTATAAAAATCGTTAATTGATTAAAAAGACGTACAATTTAAT     440
pS797    TTCAATAAAGTTAAATTTTTATAAAAATCGTTAATTGATTAAAAAGACGTACAATTTAAT     1920

CelA     CAATGATATAGTAAAAGCACGAAGAAAATATACGTACAAATTTTAGGAGGGGGAAAAATG     500
pS797    CAATGATATAGTAAAAGCACGAAGAAAATATACGTACAAATTTTAGGAGGGGGAAAAATG     1980

CelA     AAGCGTTATCGTCGCATCATCGCTATGGTCGTGACGTTCATCTTCATTCTGGGTGTGGTG     560
pS797    AAGCGTTATCGTCGCATCATCGCTATGGTCGTGACGTTCATCTTCATTCTGGGTGTGGTG     2040

CelA     TACGGCGTTAAACCGTGGCAAGAGGTTCGTGCTGGCAGCTTCAATTACGGTGAAGCGCTC     620
pS797    TACGGCGTTAAACCGTGGCAAGAGGTTCGTGCTGGCAGCTTCAATTACGGTGAAGCGCTC     2100

CelA     CAGAAAGCGATCATGTTCTACGAGTTTCAGATGAGCGGTAAGCTGCCGAACTGGGTCCGC     680
pS797    CAGAAAGCGATCATGTTCTACGAGTTTCAGATGAGCGGTAAGCTGCCGAACTGGGTCCGC     2160
```

Figure 25 Alignment of pS797 P_{araE} *::CelA* expected sequence and the Sanger sequencing data from the araE CelA_Fwd sequence read.

Sequence alignment (Clustal) shows that the plasmid has been correctly assembled (Red = BioBrick Non coding sequence prefix, Purple = P_{araE}, 5´ end of the *CelA* coding sequence).

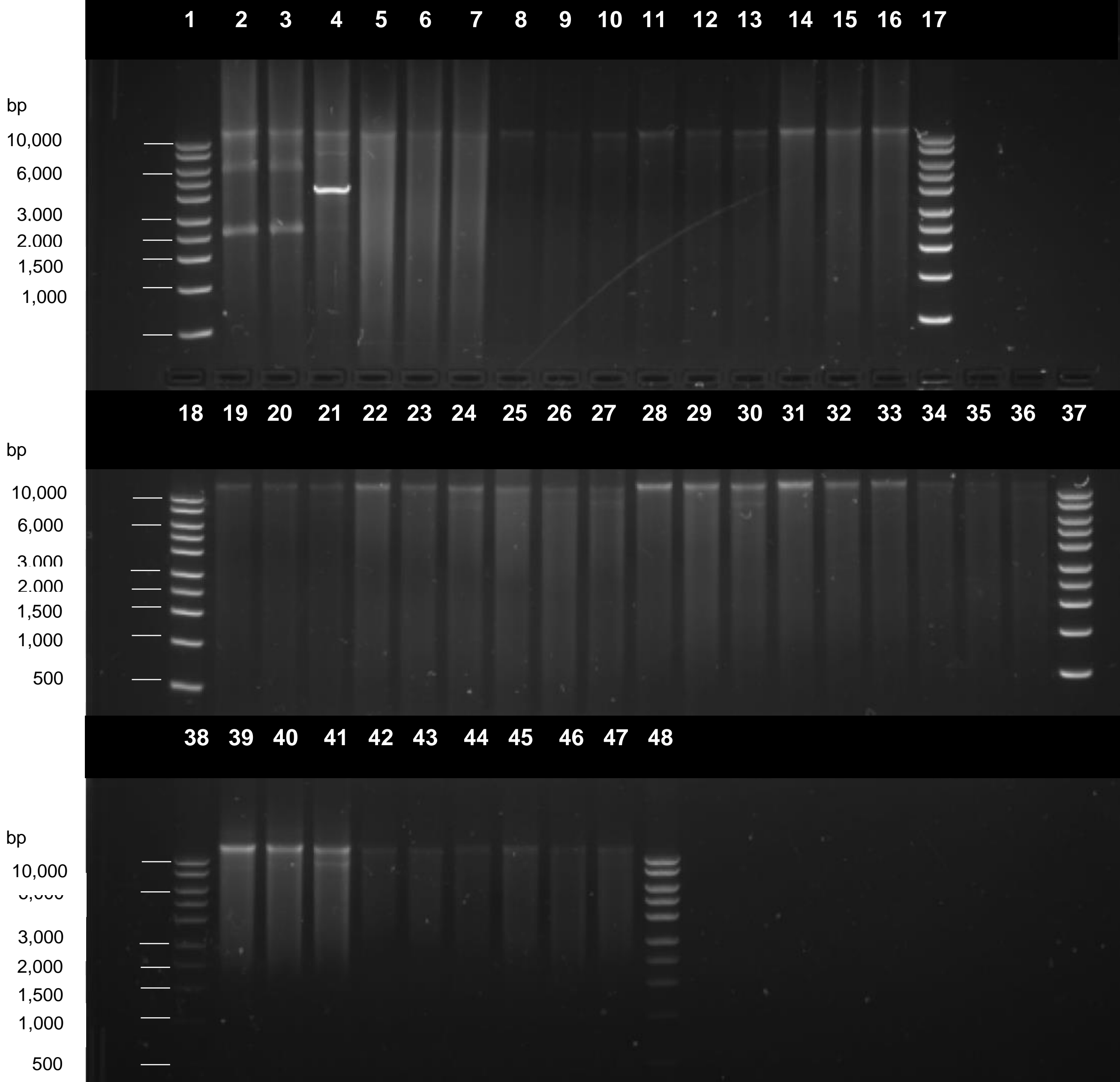


Figure 26 Transformation of pS797 P_{araE}::*CelA* purified into *Geobacillus thermoglucosidans* by conjugation.

Gel electrophoresis image of purified DNA from *Geobacillus thermoglucosidans* showing transformation with the pS797 P_{araE}::*CelA.* Plasmid purified from GB15 was digested with *Eco*RI and *Nde*I (Gel E). Colony #2 and Colony #6 number have bands of the expected size following the *Xba*I and *Nde*I digest. Lanes are grouped in threes according to the treatment of the purified DNA as follows: undigested, *Nde*I digest and *Xba*I digest as follows lanes2-4 = pS797, lanes 5-7 Wt, lanes 8-10 = pS797 P_{araE}::*CelA* #1, lanes 11-13 = P_{araE}::*CelA* #2, lanes 14-16 = P_{araE}::*CelA* #3, lanes 19-21 = P_{araE}::*CelA* #4, lanes 22-24 = P_{araE}::*CelA* #5, lanes 25-27 = P_{araE}::*CelA* #6, lanes 28-30 = P_{araE}::*CelA*

#7, lanes 31-33 = P_{araE}*::CelA* #8, lanes 34-36 = P_{araE}*::CelA* #9, lanes 39-41 = P_{araE}*::CelA* #10, lanes 42-44 = P_{araE}*::CelA* #11, lanes 45-47 = P_{araE}*::CelA* #12. Lanes 1, lanes 17, 18, 37, 38 and 48 contain DNA ladder (Sigma Aldrich Directload 1 kb).

5.2.8. *Investigating induction of pS797 P_{araE} :: CelA by arabinose*

Following transformation of GB15 with pS797 P_{araE}::*CelA* cultures were grown overnight in TSA medium and then inoculatated into TSA +/- Arabinose at different concentrations (0, 0.1, 0.2, 0.4 and 0.6 % w/v). Cultures were assayed using the well diffusion assay previously described. The results of this are shown in Figure27. The figure shows that CelA was expressed at all concentrations of arabinose, including 0%. This is demonstrated in Figure 27 by zones which do not stain with Congo red. The diameter is these halos was measured. The halo diameters were as follows, 0 % arabinose = 6.9 mm, 0.1 % = 6.25 mm, 0.2 % 6.25 mm, 0.3 % = 5 mm, 0.4 % 6.25 mm. The cause of this is unknown but it may be caused by the presence of arabinose in the tryptone or the peptone used in the medium preparation. It has been previously shown that tryptone is able to induce a commonly used arabinose inducible promoter (P_{bad}). Another possible explaination for CelA expression in the absence of Arabinose is that the promoter itself is "leaky" so in the absence of an inducer there is a basal level of transcription.

5.2.9. *Arabinose utilisation by* Geobacillus

One of the aims using the *araE* promoter was to uncouple growth with protein expression by using arabinose as an inducer of protein expression. However, HPLC analysis of *G. thermoglucosidans* supernatant after growth for 24 h at 55 °C on TSB +/- arabinose (0.1, 0.2, 0.4, 0.6 % w/v) showed that *G. thermoglucosidans* is able to degrade arabinose. After 24 h incubation all of the arabinose in the TSB + 0.1 % arabinose had been utilised by *G. thermoglucosidans* wild type, pS797 and pS797 CelA. The *G. thermoglucosidans* strains in TSB + 0.2 % arabinose had used 92 %, 99 % and 96 % of the available arabinose. In TSB + 0.4 % arabinose cultures 54 %, 58 % and 59% of the arabinose had been utilised and in the TSB + 0.6 % arabinose culture 44 %, 40 % and 43 % of the arabinose was degraded (Figure 28). This data suggests that arabinose is not a suitable inducer for reliable expression of protein.

In the Lactose inducible system a synthetic analogue of lactose is used (IPTG). If an arabinose analog can be found that is able to switch on CelA expression but that cannot be broken down then this would overcome the issue of arabinose degradation by *Geobacillus.*

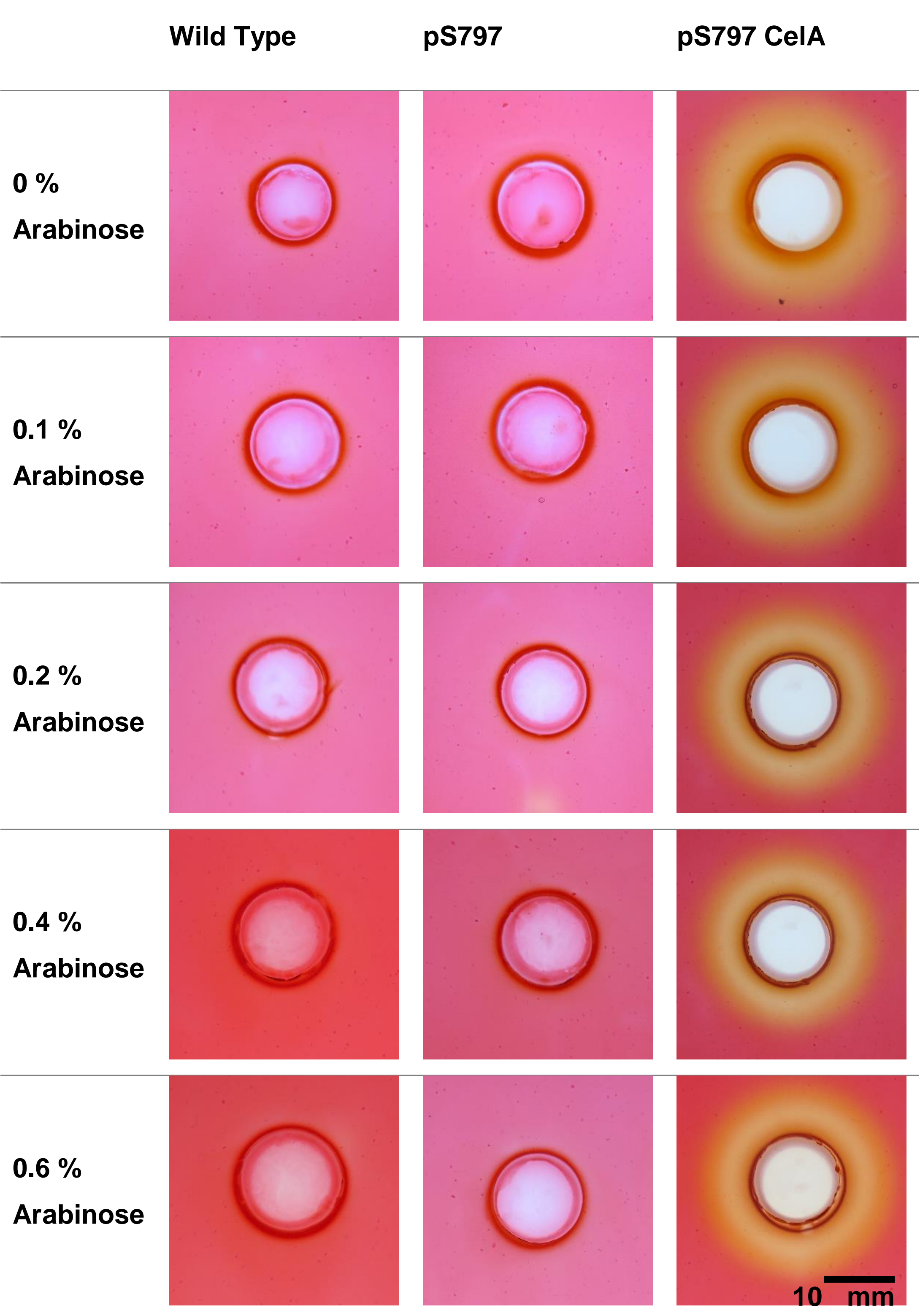

Wild Type
pS797
pS797 CelA
0 %
Arabinose
0.1 %
Arabinose
0.2 %
Arabinose
0.4 %
Arabinose
0.6 %
Arabinose
10 mm

Figure 27 Induction of CelA expression by P_{araE} induction at different concentrations of arabinose.

The figure above shows Congo Red staining of CMC agar following incubation with 250 µl of culture supernatant from *Geobacillus thermoglucosidans* WT, pS797 and pS797 P_{araE}*::CelA* grown in TSB containing different concentration of arabinose (0. 0.1, 0.2, 0.4 and 0.6 % w/v). The agar was stained with 1 M Congo red for 10 min and then destained with 1 M NaCl for 20 min. Halos, as observed in all pS797 P_{araE}*::CelA* wells indicate degradation of the CMC.

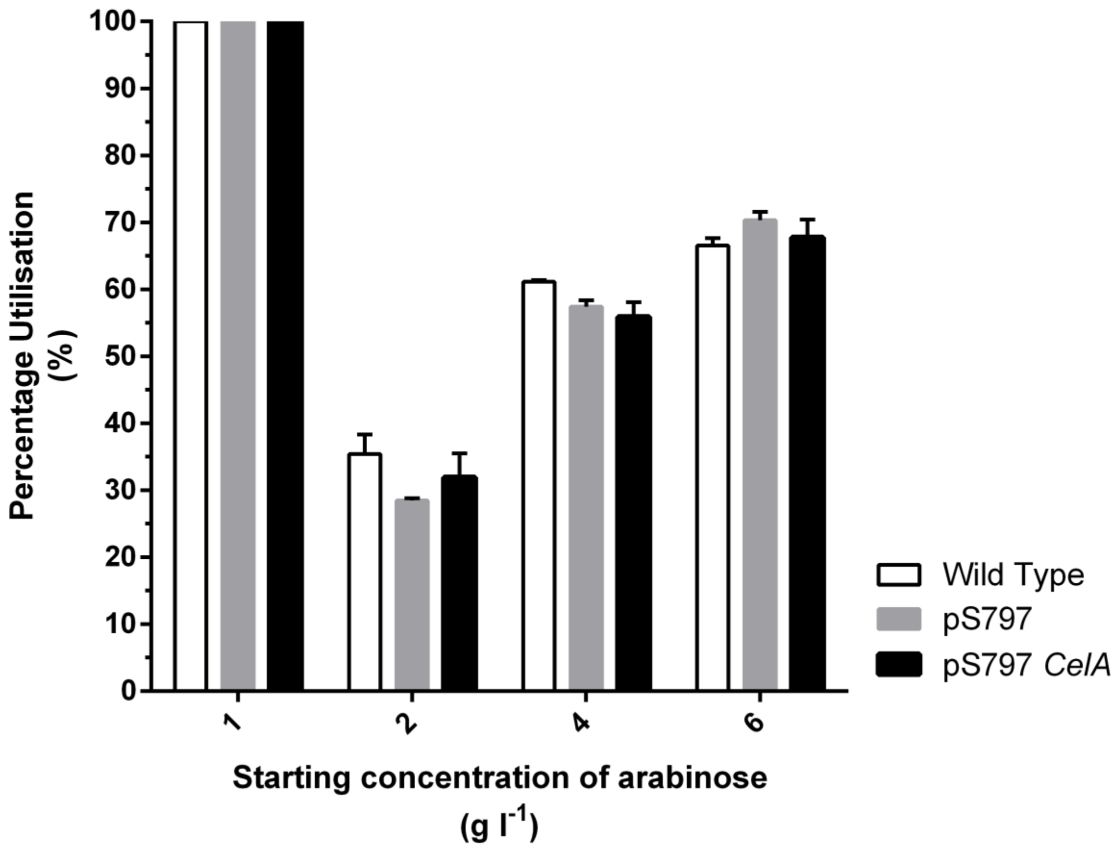


Figure 28 Degradation of arabinose by *Geobacillus thermoglucosidans* DSM2542 wild type and pS797 *CelA*

The graph above shows the amount of arabinose used during growth of *G. thermoglucosidans* for 24 h at 55 °C with shaking at 200 rpm. Arabinose usage by 3 strains of *G. thermoglucosidans* was investigated, Wild Type, pS797 and pS797 P_{araE}*::CelA.* These 3 strains were grown on TSB containing 0, 1, 2, 4 and 6 g l $^{-1}$ of arabinose. The arabinose concentration was measured using HPLC. No arabinose was detected in the media of the 1 g l^{-1} arabinose sample, following fermentation and so the amount utilised was considered to be 100 %.The experiment was carried out in duplicate.

6. Discussion

6.1. *Selection of a chassis for consolidated bioprocessing*

Geobacillus spp. are a good choice of host for the Consolidated Bioprocessing (CBP) of cellulosic biomass into useful products. The host selection process yielded seven strains able to degrade CMC and xylan. *Geobacillus* spp. are very fast growing, with a doubling time between 30-40 min when grown on TSB at 55 °C in 250 ml non-baffled shake flasks. Genomic data and experimental data (George Taylor, MRes 2014) show that *Geobacillus* sp. are able to utilise a wide range of carbon sources including C5 and C6 sugars. The claim is supported by the data presented in this report which shows that numerous strains (GB1, GB19, GB20, GB21 GB27, GB30 and GB34) of *Geobacillus* spp. are able to degrade both CMC and xylan. Of the listed strains, four were successfully transformed by conjugal mating with the pS797 plasmid. Due to the demonstrated CMCase activity, cellobiose utilisation and the ability to transform the strain, GB15 seemed to be a promising strain for further genetic engineering.

Further to this, a large amount of work has gone into developing *G. thermoglucosidans* DSM2542 by collaborating groups (University of Exeter) and so the the final decision was made that *G. thermoglucosidans* DSM2542 would be the host of choice for this project. It has the most developed molecular toolkit and a genome sequence is available (unpublished, Exeter University).

This study has progressed the use of molecular tools in five species of *Geobacillus* (*G. thermoleovorans* DSM 5366, *G. thermoglucosidans* DSM2542, *G. lituanicus* DSM15325, *and Geobacillus* sp. ATCC15952, *G. stearothermophilus* N.R.Smith T147 MCIMB 8224). All used the same plasmid, indicating that the origin of replication and kanamycin cassette are functional in a range of *Geobacillus* species. Some of these strains can be transformed and have xylanase activity (GB21, GB30 and GB34). However, as the molecular toolkit has so far been constructed with *G. thermoglucosidans* in mind, further strain development will be continued in *G. thermoglucosidans* DSM2542 for the time being.A summary of the species screening process can be found in Table 9..

Table 9 A summary of the host selection criteria and final rank of the *Geobacillus* spp. for consolidated bioprocessing.

Strain I.D.	Strain Name	Transfor-mation	CMC utilisation	Xylan utilisation	CMCase activity DNS	Heterologous cellulase activity	Additional comments	Strain Rank
1	*G. thermoleovorans* DSM 5366	✓	✓	✓	X	-	HC degrading	**-**
15	*G. thermoglucosidans* DSM 2542	✓	✓	X	✓	✓	Previous use[(1)]	**1**
19	*G. gargensis* DSM 15378	X	✓	✓	✓	-	-	**-**
20	*G. jurassiucus* DSM 15726	X	✓	✓	-	-	-	**-**
21	*G. lituanicus* DSM 15325	✓	✓	✓	X	-	-	**4**
27	*G. toebii* DSM 14590	X	✓	✓	✓	-	Hay compost isolate	**-**
30	*Geobacillus* sp. ATCC 15952	✓	✓	✓	✓	-	Degrades penicillin	**2**
34	*G. stearothermophilus* 38224	✓	✓	✓	X	-	-	**3**

- (1) **TMO Renewables**

Three functional cellulases (CelA from *C. bescii,* CelB from *T. neopolitana* and Cel9A from *T. fusca)* have been expressed in *G. thermoglucosidans.* The cellulases have been shown to be active at degrading CMC. CelA and CelB are stable at 75 °C, making them well suited to consolidated bioprocessing of cellulose into simple sugar*.* The simple sugars (glucose, cellobiose and xylose) released by the enzymatic degradation of cellulose and hemicellulose can be utilised by *G. thermoglucosidans,* to produce fatty acids. These fatty acids, can in turn, be converted into alkanes and alkenes*.*

The expressed proteins (CelA*,* CelB and Cel9A*)* have been shown to have activity against CMC, a modified, soluble form of cellulose. This work further develops *G. thermoglucosidans* as a host for consolidated bioprocessing of lignocellulosic material. The protein of greatest interest for lignocellulose bioprocessing is CelA.

CelA is of interest as it has exoglucanase, endoglucanase and xylanase activity (60 % conversion of xylan in native switch grass (Brunecky *et al*. 2013). CelA has high activity against avicel and is active at much higher temperatures than other cellulase enzymes (up to 85°C) (Brunecky *et al*. 2013). The high activity of CelA is attributed to the proximity of the endoglucanase and cellobiohydrolase (exoglucanase) domains, afforded by the multi-domain structure of CelA (Brunecky *et al*. 2013). The use of CelA may remove the need for pre-treatment entirely. It has been shown that the activity of CelA against switch grass was actually reduced by pre-treatment of switch grass (Brunecky *et al*. 2013). CelA is able to act on cellulose in two ways: through surface ablation and cavity formation (Brunecky *et al*. 2013). However it is a much larger protein than CelB (~170 kDa and 30 kDa respectively) (Zverlov *et al*. 1998; Bok *et al*. 1998). The characteristics discussed make CelA a very promising enzyme for the consolidated processing of lignocellulosic biomass. However, without the ability to tailor and regulate the gene expression levels, the burden of expressing a protein of this size (195 kDa) could be deleterious. The effect of this burden might be further exacerbated by the propensity of *Geobacillus* to sporulate under stress conditions. These are challenges to the progress of this work.

6.2. *Progress and further work*

Further work on this project should include development of a quantitative assay for cellulase activity against CMC, microcrystalline cellulose and complex biomass. An *in vitro* method for quantitative testing of the cellullase proteins would be to purify the proteins, using the same His tag motif used in this study for western blotting. The purified cellulase protein can then be incubated with CMC or Avicel and activity can be assessed by HPLC. A quantitative assay would allow the study of the enzyme kinetics, cofactor requirements and the optimisation of environmental factors, such as pH, temperature, O_2 concentration and expression level, to improve CelA activity in *Geobacillus*.

During preparation of this manuscript, further work was carried out to develop such an assay, using the tools developed here. This work has successfully quantified the activity of CelA (purified from *G. thermoglucosidans* culture supernatant) against Avicel, However, it is important to remember that improved *in vitro* activity does not translate directly to *in vivo* activity. *In vivo* assays for cellulase activity have been developed (Personal communication, David Parker: 10th July 2015) and it is recommended that future work investigate *in vivo* activity using an assay which is high throughput, highly sensitive and highly specific assay.

A method that would fulfil these requirements is a FACS-based vanadium bromoperoxidase reducing sugar detection assay (Prodanovic *et al*. 2012; Ostafe *et al*. 2013). Single bacterial cells are isolatedin a double emulsion containing a glucose oxidase enzyme (GOx) and vanadium bromoperoxidase and sodium bromide. Cellulase activity is detected when the hydrogen peroxide generated by GOx is converted to hypobromide by vanadium bromoperoxide. As a result of hypobromide reacting with the probe, 3-carboxy-7-(4'-aminophenoxy)-coumarine, fluorescent 3-carboxy-coumarine is released. This fluorescence can be detected using flow cytometry, allowing measurement of cellulase activity. By generating a standard curve based on reducing sugar concentration this can be developed into a quantitative *in vivo* assay for cellulase activity in *Geobacillus*.

6.3. *Improving cellulase activity by rational mutagenesis*

Now that a set of functional cellulase genes has been expressed in *Geobacillus* these proteins can be altered to improve their compatibility with the host. The optimal temperature of the cellulase can be optimised to better match with that of *Geobacillus*. The cellulase genes are functional between 70 °C-85 °C, whilst *Geobacillus thermoglucosidasius* DSM2542 has a growth optimum of 55 °C (Nazina *et al*. 2001) and *Geobacillus stearothermophilus* N.R.Smith T147 has an optimum of 55 °C (NCIMB). Through modification of the protein structure it is possible for the optimal temperature of the proteins to be reduced whilst still maintaining optimal activity.

Protein structures are reengineered using numerous approaches. Historically, these approaches tend to be based on mutagenesis of the DNA, through either chemical mutagenesis or error prone PCR (Liang *et al*. 2011). However, recently developments in high throughput DNA synthesis have seen a reduction in the cost of DNA synthesis (Carlson 2009). The ability to write DNA as well as read it has allows high throughput protein engineering without needing to perform mutagenesis on your gene of interest (Liao *et al*. 2007). This technology has been successful in altering both substrate specificity (Govindarajan *et al*. 2014) and activity (Liao *et al*. 2007). It is also possible to use protein engineering to improve the temperature range and optimal temperature for an enzyme through computational modelling of protein folding (Trudeau *et al*. 2014). By redesigning the cellulase proteins expressed here, it may be possible to tailor their kinetics so that they best suit the host and the fermentation system selected.

Once protein variants have been generated it is important that they can be screened using a fast and simple activity assay. The FACS GOx assay discussed previously, is a good candidate for this and it is strongly suggested that this assay is developed for use in *Geobacillus.* This assay has been previously used to screen cellulase mutant libraries for increased activity (Liang *et al*. 2011). Future studies should adapt this assay for use with *Geobacillus.*

Whilst progress has been made toward generation of cellulolytic *Geobacillus,* it seems that engineering of cellulolytic pathways into a host is a difficult approach to take. However, given the industrial development of alkane synthesis pathways, the possible routes to success are limited. The biggest restriction to progress is the oxygen dependence of the ADO gene (Schirmer *et al.* 2010). Recent progress in the development of thermophiles, like *C. bescii,* for biofuel production are very attractive for CBP due to its ability to utilise lignocellulose as a carbon source (Brunecky *et al.* 2013) for the production of bioethanol (Chung *et al.* 2014). However, *C. bescii* is an anaerobic organism, making it incompatible with known alkane synthesis pathways and the generation of economically viable biofuels. Although there has been debate over the enzyme's oxygen dependency for the alkane production mechanism (Li *et al.* 2012; Warui *et al.* 2011; Eser *et al.* 2011), the isotopic tracer analysis carried out recently (Li *et al.* 2012) shows that the enzyme requires O_2 for its activity. Whilst it may be possible to adjust the process of fuel production to enable the use of an anaerobic organism, this would increase operating capital and therefore reduce the economic feasibility of the entire process. Alkane synthesis pathways that do not require oxygen may be compatible with alkane production in *C. bescii,* but currently characterised pathways are not.

6.4. *The use of solid state fermentation in consolidated bioprocessing*

The oxygen requirements of current alkane pathways and the high solid content of lignocellulosic material, makes bioprocessing of biomass difficult in a liquid based fermentation system. The solid fraction presents a particular challenge because of material settling and blockage of liquid handling system, i.e. addition or removal of substrate and product. There is also a very high water usage associated with a liquid fermentation system, which adds cost to the process. Liquid fermentation requires large, high cost fermentation tanks, liquid handling systems, pressure regulation and heating and cooling. All of these factors add capital expenditure to an industrial scale project. This increased capital cost means that the bioprocessing must be more efficient for the process to remain viable, which is a particular issue given that liquid transport fuels are a low value, high volume product.

A solid state, consolidated process, such as a silage based system, would negate the issues of a liquid system. It would be very cheap to set up and maintain. Like a garden compost heap, it is expected that a silage pile would maintain the optimal temperature of the composting microbes, in this case *Geobacillus.* There will of course be a temperature gradient from the inside to the outside of the pile, but the use of *Geobacillus,* with its broad temperature tolerance should mean the impact of any temperature fluctations is reduced.

However, as with liquid fermentation, there are barriers to solid state fermentation. Monitoring of variables such as cell growth, pH and temperature becomes difficult. There are still a lot of factors that need to be considered before *Geobacillus* is ready to be deployed in a solid state system.

6.5. *Ancillary enzymes to improve extracellular cellulase activity*

Whilst the proper expression of CelA is promising it is important to highlight the complexity of natural lignocellulosic systems. The degradation mechanism found in nature often consist of numerous genes; for example *C. bescii* DSM6725 has 32 genes whose expression is up-regulated in response to growth on cellulose, many of which are multifunctional (Dam *et al*. 2011). The expression of a single cellulase enzyme may not prove to be sufficient to achieve optimum biomass degradation. The role of other ancillary enzymes needs to be investigated as their use could lead to significant increases in yield. An example of this is the use of a thermostable beta glucosidase enzyme from *T. maritima* (bglA) (Brunecky 2013). By supplementing CelA with this enzyme at a ratio of 14 mg g^{-1} glucan: 1 mg g^{-1} glucan, the total glucan conversion over 160 h, with Avicel as a substrate was doubled at 75 °C (Brunecky *et al*. 2013). This increase in total conversion also held true on dilute acid pretreated corn stover, where supplementation gave an increase in in conversion from 30 to ~75 %. It is hypothesised that this large increase in rate of conversion was caused by removing product inhibition.

The effect of sondary enzymes in a *Geobacillus* lignocellulose fermentation system is an important area of this research and requires further investigation. To aid this a number of candidate enzymes have been selected and are listed in the table below (Table 10).

6.6. *Utilisation of hemicellulose by* Geobacillus

A constituent part of lignocellulose that has not been focussed on is this study is hemicellulose, made up of mostly xylan and arabinan. A recent paper (De Maayer *et al.* 2014) highlighted the variation between loci in the currently sequenced strains of *Geobacillus* and the highly variable hemicellulose utilisation loci, which encodes the hemicellulose utilisation system genes (HUS), which encode xlyanases, xylanosidases, arabinofuranosidases and glucuronidases. The authors highlight this as a target for re-engineering (De Maayer *et al.* 2014). It was also observed that there was some diversity in the Arabinose Utilisation System with several strains (Y412MC52,Y412MC61, T-6 and WSUCF1) having a AUS incorporated into the HUS locus (De Maayer et al. 2014). This research experimentally validates some of the findings of De Maayer *et al.* 2014 as it has been shown that the 46 strains tested vary in their ability to utilise hemicellulose (xylan) and CMC. Whilst the focus of this research was cellulose utilisation it may be important that hemicellulases are utilised as well. Previous work on GB15 has shown xylose utilisation (George Taylor, Exeter University MRes, unpublished data), and this study demonstrated arabinose utilisation. The use of hemicellulose as a carbon source should be investigated further. Further assessment of the HUS and AUS variability of *Geobacillus* will provide further insight into chassis development.

By developing the toolkit available the diverse pan genome of *Geobacillus* spp. can be combined into one host. This would require complex engineering and optimisation, but with proper strain development is feasible. The HUS of *G. stearothermophilus* T-6 was recently sequenced (NCBI Acc. # DQ868502) and shown to contain 13 gene clusters for hemicellulose degradation making T-6 an additional candidate for host screening (Not currently available from a tractable source).

Table 10 Candidate enzymes for improving cellulose utilisation by pS797 *CelA Geobacillus thermoglucosidans.*

Gene I.D	Organism	Protein Size	Ascension #	Localisation	Function
Cel9b	*C. bescii*	151.06	ACM60953	Extracellular	Beta glucanase and mannase activity
CelE	*C. bescii*	142.19	ACM60947	Extracellular	Beta glucosidase and mannase activity
CelF	*C. bescii*	164.8	ACM60945	Extracellular	Xylanase and cellobiohydrolase
PecB	*C. bescii*	49.16		Extracellular	Pectate lyase
BglA	*T. maritima*	51.54	CAA52276.1	Intracellular	Beta glucosidase
BglB	*Geobacillus thermoglucosidasius C56-YS93*	55.58	AEH46291.1	Intracellular	Beta glucosidase
Geoth_2227	*Geobacillus thermoglucosidasius C56-YS93*	54.89	AEH48161.1	Extracellular	Beta glucosidase
BglA	*Geobacillus NBRC 107763*	55.51	GAJ43011.1	Intracellular	Beta glucosidase
BglB	*Geobacillus C56-T3*	54.54	ADI26683.1	Extracellular	Beta glucosidase
BglB	*Geobacillus C56-T3*	52.74	ADI26193.1	Extracellular	Beta glucosidase
BglA	*C. saccharolyticus*	53.5	CAA31087.1	Extracellular	Beta glucosidase

Long term population stability is required in a bioprocessing system, therefore factors such as mutagenic drift and contamination must be considered. In a nutrient limited environment it becomes necessary for the host to express the heterologous gene in order to survive. If there is excess free sugar around then it is no longer necessary for the host to express our cellulases. This lack of selection pressure would allow gene deletion or mutation even in a system where genes of interest have been integrated into the genome. A number of closely related species of *Geobacillus* are able to transport cellobiose using a cellobiose phosphotransferase operon (Lai & Ingram 1993; Zeigler 2014). This presents another interesting way of removing substrate inhibition. This approach is a “pull through” instead of a “push through” approach; a method which works with the metabolic flux by removing the effect of feedback and therefore hoping to reduce metabolic bottlenecks and the overproduction of product. In essence, it prevents supply outweighing demand and so by balancing “push and pull” carbon flux can be tailored to maximise production of the desired molecule i.e. transport fuels.

6.7. Geobacillus *and synthetic biology*

Recent developments in genetic tools and resources has been recognised as the catalyst for the recent advances in the use of thermophiles in biotechnology (Taylor *et al*. 2011). This increase in tools and resources is observed in *Geobacillus*. Gene KO systems have been developed (Cripps *et al*. 2009; Suzuki *et al*. 2012) and the number of promoter and reporter systems available is being expanded (Bartosiak-Jentys *et al*. 2013; Bartosiak-Jentys *et al*. 2012). This development opens a number of possibilities which will be essential for the development of *Geobacillus* as a chassis for CBP; pathway optimisation will be essential and a reporter system, such as green fluorescent protein (GFP) can be used to test predictable expression systems in *G. thermoglucosidans* as done in *E. coli* (Kosuri *et al*. 2013), allowing the construction of more finely tuned pathways, resulting in more efficient protein production, carbon utilisation and therefore higher productivity. Given that the host being developed in this project will be used for alkane production, this will require genetic modification at the upstream and downstream ends of carbon metabolism. This will no doubt require a large suite of genes that will require expression optimisation.

Now that gene function has been demonstrated it becomes possible to begin optimisation of the protein itself. There are several companies that offer this kind of service (DNA2.0, Synthace). By using Design of Experiment (DOE) and machine learning, it is possible to predict and test the effect of DNA sequence modifications and to alter the expression level (Gustafsson *et al*. 2012) and functional parameters (Liao *et al*. 2007) of your protein of choice. Importantly, this design is carried out in a similar fashion irrespective of protein function or mechanism meaning that it can be used to optimise any protein. Another recent paper showed a 14 fold increase in protein expression by applying the N-terminal bias for rare codon usage across an entire gene sequence (Goodman *et al*. 2013).

This design-model-build-test cycle has become the defining feature of a newly emerged field of biology known as synthetic biology. This principle can be applied to chassis construction. This rational design of the pathways required for CBP will allow for the generation of an efficient and productive system. Synthetic control system of gene expression has been hugely advanced in recent years allowing the syncronised “switching on” (Elowitz & Leibler 2000) of genes through the introduction of synthetic external stimuli (Dixon *et al*. 2010), logic based computation of inputs for programmable gene expression (Wang *et al*. 2011; Bonnet *et al*. 2013), and tighter control over the expression of genes at the level of transcription (Larson *et al*. 2013) and translation. Recent work of translational regulation mediated by riboregulators has allowed for the development of a highly effective bacterial kill switch (Callura *et al*. 2010), an achievement that received great attention from the media. If a system like this is approved for use then this could drastically change the use of Genetically Modified (GM) organisms, not just for CBP of biofuels but for the whole biotechnology industry.

Another recent paper from the Collins lab tested riboregulated networks that can be used to redirect carbon flux of cells (Callura *et al*. 2012). It may also be possible to use CRISPRi based systems to achieve this (Larson *et al*. 2013). Often when bacteria are used in biotechnology it is necessary to redirect the carbon flux through gene knockouts (Young *et al*. 2014; Cripps *et al*. 2009). It can be time consuming construct KO libraries and so these technologies are of great interest to industry.

The development of synthetic biology techniques is of great importance to industrial scale bioprocessing. In an industrial setting knowledge transfer can often be a limiting factor. Synthetic biology aims to increase the reproducibility of research. This will make knowledge and technology transfer less risky and therefore a more attractive choice for businesses.

7. Acknowledgements

First of all I would like to thank my supervisors Professor John Love and Professor David Parker. Without their effort, guidance and insight this work would not have been possible. Secondly, to Professor Rob Lee, for hosting me in the Biodomain Lab at Shell Technology Centre Houston and making it all possible; it was a fantastic year, and everybody there taught me a lot. Finally to Shell U.S. Hosting Company and the University of Exeter for this opportunity.

My deepest gratitude also to the members of the Biodomain Lab and Mezzanine Lab, you have all helped me and put up with my mess. I thank you all but in particular those who have helped tackle *Geobacillus*: Dr. Thomas Howard, Dr. Chloe Singleton, Dr. Thomas Lux, Dr. Paul James Dr. Trevor Zurroff, Dr. Yi Liang, Dr. Nadim Ashraf, Dr. Kun Zhang Ashley Baugh, Lindsay Crowley, Christopher Dealing, Kellie Hull, Crystal Warner, George Taylor, Jamie Gilman and Dagmara Kolak. Thanks also to Steve Brown, Alex Johns, Peggy Dousseaud and Sam Buddington who offered help criticism, guidance and morale. and an extra set of eyes when it came to proofreading. I owe you all a beer(s).

Thank you.

8. References

Aarts, M., 1995. Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *The Plant cell: Online*, 7(12), pp.2115–2127. Available at: http://www.plantcell.org/content/7/12/2115.abstract [Accessed September 26, 2014].

Ahmad, S., Scopes, R, K., Rees, G.N., Patel, B.K.,, 2000. *Saccharococcus caldoxylosilyticus* sp. nov., an obligately thermophilic, xylose-utilizing, endospore-forming bacterium. *International Journal of Systematic and Evolutionary Microbiology*, 50 (2), pp.517–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10758855 [Accessed March 7, 2014].

Ahmad, M., Taylor, C., Pink, D., Burton, Kerry., Eastwood, Daniel., Bending, G., Bugg, T.D.H., 2010. Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Molecular Biosystems,* 6 (5) pp.815-21. Available at: http://pubs.rsc.org/en/content/articlehtml/2010/mb/b908966g Accessed {February 2, 2016]

Akhtar, M.K., Turner, N.J. & Jones, P.R., 2013. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), pp.87–92. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3538209&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Albro, P.W. & Dittmer, J.C., 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. 3. The metabolic relationship of long-chain fatty acids and hydrocarbons and other aspects of hydrocarbon metabolism in Sarcina lutea. *Biochemistry*, 8(5), pp.1913–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5785213 [Accessed September 26, 2014].

Anon, 1988. Validation of the publication of new names and new combinations previously effectively published outside the IJSB: List No. 24. *International Journal of Systematic Bacteriology*, 38(1), pp.136–137. Available at: http://ijs.sgmjournals.org/content/38/1/136.abstract [Accessed October 27, 2014].

Arutyunov, D., Rodriguez-Maillard, J.M. & Frost, L.S., 2011. A PAS domain within F plasmid TraJ is critical for its function as a transcriptional activator. *Biochemistry and Cell Ciology = Biochimie et Biologie Cellulaire*, 89(4), pp.396–404. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21774634 [Accessed October 13, 2014].

Assareh, R., Shahbani Zahiri, H., Akbari Noghabi, K., 2012. Characterization of the newly isolated *Geobacillus* sp. T1, the efficient cellulase-producer on untreated barley and wheat straws. *Bioresource technology*, 120, pp.99–105. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22784959 [Accessed March 10, 2014].

Atsumi, S., Cann, A.F., Connor, M.R., Shen, C.R., Smith, K.M., Brynildsen, M.P., Chou, K.J.Y., Hanai, T., Liao, J.C., 2008. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic engineering*, 10(6), pp.305–11. Available at: http://www.sciencedirect.com/science/article/pii/S109671760700047X [Accessed

June 25, 2015].

Banat, I.M., Marchant, R. & Rahman, T.J., 2004. *Geobacillus debilis* sp. nov., a novel obligately thermophilic bacterium isolated from a cool soil environment, and reassignment of *Bacillus pallidus* to *Geobacillus pallidus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 54(Pt 6), pp.2197–201. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15545458 [Accessed March 7, 2014].

Bartosiak-Jentys, J., Hussein, A., Lewis, C.J., Leak, D.J., 2013. Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*. *Microbiology (Reading, England)*, 159(Pt 7), pp.1267–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23704786 [Accessed August 19, 2013].

Bartosiak-Jentys, J., Eley, K. & Leak, D.J., 2012. Application of pheB as a reporter gene for *Geobacillus* spp., enabling qualitative colony screening and quantitative analysis of promoter strength. *Applied and Environmental Microbiology*, 78(16), pp.5945–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3406106&tool=pmcentrez&rendertype=abstract [Accessed November 19, 2013].

Beller, H.R., Goh, E.-B. & Keasling, J.D., 2010. Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Applied and Environmental Microbiology*, 76(4), pp.1212–23. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2820947&tool=pmcentrez&rendertype=abstract [Accessed August 6, 2014].

Bernard, A., Domergue, S., Pascal, S., Jetter, R., Renne, C., Faure, J., Haslam, R.P., Nappier, J.A., Lessire, R., Joubes, J., 2012. Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *The Plant Cell*, 24(7), pp.3106–18. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3426135&tool=pmcentrez&rendertype=abstract [Accessed September 17, 2014].

Bok, J.-D., Yernool, D.A. & Eveleigh, D.E., 1998. Purification, characterization, and molecular analysis of thermostable cellulases CelA and CelB from *Thermotoga neapolitana*. *Appl. Envir. Microbiol.*, 64(12), pp.4774–4781. Available at: http://aem.asm.org/content/64/12/4774.short [Accessed March 18, 2014].

Bonnet, J. Yin, P., Ortiz, M., Substoontorn, P., Endy, D., 2013. Amplifying genetic logic gates. *Science (New York, N.Y.)*, 340(6132), pp.599–603. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23539178 [Accessed July 11, 2014].

Bronnenmeier, K. & Staudenbauer, W.L., 1990. Celluloase hydrolysis by a highly thermostable eno-1,4-beta-glucanase (Avicelase I) from *Clostridium stercorarium*. *Enzyme and Microbial Technology*, pp.431–436. Available at: http://ac.els-cdn.com/014102299090053S/1-s2.0-014102299090053S-main.pdf?_tid=143fe974-5f90-11e4-a65d-00000aab0f6c&acdnat=1414603494_e5e5b34dc3c10c172e9b3a54923ff8be [Accessed October 29, 2014].

Brunecky, R. et al., Alahuhta, M., Xu, Q., Donohoe, B., Crowley, M., Kataeva, I., Yang,

S.J., Resch, M.G., Adams, M.W.W., Lunin, V., Himmel, M.E., Bonble, Y., 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science (New York, N.Y.)*, 342(6165), pp.1513–6. Available at: http://www.sciencemag.org/content/342/6165/1513 [Accessed July 14, 2014].

Burger, B. V. Reiter, B., Borzyk, O., du Plessis, M.A., 2004. Avian exocrine secretions. I. Chemical characterization of the volatile fraction of the uropygial secretion of the green woodhoopoe, *Phoeniculus purpureus*. *Journal of Chemical Ecology*, 30(8), pp.1603–1611. Available at: http://link.springer.com/10.1023/B:JOEC.0000042071.65335.f3 [Accessed September 26, 2014].

Caccamo, D. Gugliandolo, C., Stackebrandt, E., Maugeri, T.L., 2000. *Bacillus vulcani* sp. nov., a novel thermophilic species isolated from a shallow marine hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology*, 50 Pt 6, pp.2009–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11155974 [Accessed March 7, 2014].

Callura, J.M. Dwyer, D. Isaacs, F., Cantor, C.R., Collins, J., 2010. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), pp.15898–903. Available at: http://www.pnas.org/content/early/2010/08/09/1009747107.abstract [Accessed March 19, 2014].

Callura, J.M., Cantor, C.R. & Collins, J.J., 2012. Genetic switchboard for synthetic biology applications. *Proceedings of the National Academy of Sciences of the United States of America*, 109(15), pp.5850–5. Available at: http://www.pnas.org/content/109/15/5850.full [Accessed October 31, 2014].

Carlson, R., 2009. The changing economics of DNA synthesis. *Nature biotechnology*, 27(12), pp.1091–4. Available at: http://dx.doi.org/10.1038/nbt1209-1091 [Accessed March 23, 2015].

Cava, F.de Pedro, M., Blas-Galindo, E., Waldo, G.S., Westblade, L.F., Berenguer, J., 2008. Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environmental Microbiology*, 10(3), pp.605–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18190515 [Accessed August 8, 2013].

Cha, M., Chung, D., Elkins, J., Guss, A.M., Westphelig, J., 2013. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. *Biotechnology for Biofuels*, 6(1), p.85. Available at: http://www.biotechnologyforbiofuels.com/content/6/1/85 [Accessed October 8, 2014].

Chen, Z.F., Wojcik, S.F. & Welker, N.E., 1986. Genetic analysis of *Bacillus stearothermophilus* by protoplast fusion. *Journal of Bacteriology*, 165(3), pp.994–1001. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=214527&tool=pmcentrez&rendertype=abstract [Accessed October 27, 2014].

Chung, D., Cha, M., Guss, A.M., Westphelig, J.,et al., 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), pp.8931–6. Available at: http://www.pnas.org/content/111/24/8931.abstract [Accessed September 24, 2014].

Chung, D., Farkas, J., Huddleston, J.R., Olivar, E., Westphelig, J., et al., 2012. Methylation by a unique α-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. *PloS one*, 7(8), p.e43844. Available at: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0043844#pone-0043844-g004 [Accessed February 20, 2014].

Cihan, A.C., Ozcan, B., Tekin, N., Cokmus, C., et al., 2011. *Geobacillus thermodenitrificans* subsp. *calidus*, subsp. nov., a thermophilic and α-glucosidase producing bacterium isolated from Kizilcahamam, Turkey. *The Journal of General and Applied Microbiology*, 57(2), pp.83–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21606609 [Accessed March 7, 2014].

Coorevits, A. Dinsdale, A., Halket, G., Lebbe, L., De Vos, P., van Landschoot, A., Logan, N.A., et al., 2012. Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus, G. stearothermophilus, G. jurassicus, G. toebii, G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly “thermoglucosidasius”); transfer of *Bacillus thermantarcticus* . *International Journal of Systematic and Evolutionary Microbiology*, 62(Pt 7), pp.1470–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21856988 [Accessed March 7, 2014].

Cripps, R.E., Eley, K., Leak, D., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., Atkinson, T., 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering*, 11(6), pp.398–408. Available at: http://www.sciencedirect.com/science/article/pii/S1096717609000688 [Accessed December 3, 2013].

Dakhova, O.N. Kurepina, N., Zverlov, V., Svetlichnyi, V.A., Velikodvorskaya, G.A., 1993. Cloning and expression in *Escherichia coli* of *Thermotoga neapolitana* genes coding for enzymes of carbohydrate substrate degradation. *Biochemical and Biophysical Research Communications*, 194(3), pp.1359–64. Available at: http://www.sciencedirect.com/science/article/pii/S0006291X83719741 [Accessed March 21, 2014].

Dam, P., Kataeva, I., Yang, S., Zhou, F., Yin, Y., Chou, W., Poole, F.L., Westphelig, J., Hettich, R., et al., 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Research*, 39(8), pp.3240–54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3082886&tool=pmcentrez&rendertype=abstract [Accessed January 16, 2015].

Demirbas, A., 2007. Importance of biodiesel as transportation fuel. *Energy Policy,* 35 (9) pp.4661-4670. Available at: http://www.sciencedirect.com/science/article/pii/S0301421507001516 [Accessed February 02, 2016].

Dixon, N.Duncan, J., Geerlings, T., Dunstan, M., McCarthy, J.E.G., Leys, D., Micklefield, J., ., 2010. Reengineering orthogonally selective riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.2830–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2840279&tool=pmcentrez&rendertype=abstract [Accessed July 14, 2014].

Dykstra, A.B. Rodriguez, M., Raman, B., Cook, K., Hettich, R.L., 2013. Characterizing the range of extracellular protein post-translational modifications in a cellulose-degrading bacteria using a multiple proteolyic digestion/peptide fragmentation approach. *Analytical chemistry*, 85(6), pp.3144–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23406086 [Accessed March 27, 2014].

Elowitz, M.B. & Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), pp.335–8. Available at: http://dx.doi.org/10.1038/35002125 [Accessed March 20, 2014].

EPA, U., 2006a. Air Pollution and the Clean Air Act | US Environmental Protection Agency. Available at: http://www.epa.gov/air/caa/ [Accessed June 3, 2015].

EPA, U., 2007. *Energy Independence and Security Act*, United States of America. Available at: http://www.gpo.gov/fdsys/pkg/PLAW-110publ140/pdf/PLAW-110publ140.pdf [Accessed June 3, 2015].

EPA, U., 2013. Regulation of Fuels and Fuel additives: 2013 Renewable Fuel Standards. United States of America. Available at: http://www.epa.gov/otaq/fuels/renewablefuels/documents/rfs-2013-standards-nrpm.pdf [Accessed October 9, 2014].

EPA, U., 2006b. *Renewable Fuel Standard*, United States of America. Available at: http://epa.gov/oms/renewablefuels/420d06008.pdf [Accessed June 3, 2015].

Eser, B.E. Das, D., Han, H., Jones, P., Marsh, N.E.G., 2011. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. *Biochemistry*, 50(49), pp.10743–50. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3235001&tool=pmcentrez&rendertype=abstract.

Fargione, J. Hill, J., Tilman,D., Polasky, S., Hawthorne, P., 2008. Land clearing and the biofuel carbon debt. *Science (New York, N.Y.)*, 319(5867), pp.1235–8. Available at: http://www.sciencemag.org/content/319/5867/1235.short [Accessed July 9, 2014].

Fong, J.C.N., Svenson, C., Nakasugi, K., Leong, C.T.C., Bowman, J.P., 2006. Isolation and characterization of two novel ethanol-tolerant facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles : Life Under Extreme Conditions*, 10(5), pp.363–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16532362 [Accessed October 27, 2014].

Fortina, M.G., Mora, D., Schumann, P., Parini, C., Manachini, P.L., Stackebrandt, E., 2001. Reclassification of *Saccharococcus caldoxylosilyticus* as *Geobacillus caldoxylosilyticus* (Ahmad et al. 2000) comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 51(Pt 6), pp.2063–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11760948 [Accessed March 7, 2014].

Gibson, D.G. Young, L., Chuang, R., Venter, J.C., Hutchison, C., Smith, H., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), pp.343–5. Available at: http://dx.doi.org/10.1038/nmeth.1318 [Accessed July 14, 2014].

Goldemberg, J. Coelho, S., Nastari, P., Lucon, O., 2004. Ethanol learning curve—the Brazilian experience. *Biomass and Bioenergy*, 26(3), pp.301–304. Available at: http://www.sciencedirect.com/science/article/pii/S0961953403001259 [Accessed September 24, 2014].

Golovacheva, R.S. Loginova, L., Salikhov, T., Noleskinov, A.A., Zaitseva, G.N., [New species of thermophilic bacilli--*Bacillus thermocatenulatus* nov. sp]. *Mikrobiologiia*, 44(2), pp.265–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1226140 [Accessed October 27, 2014].

Gomez del Pulgar, E.M. & Saadeddin, A., 2014. The cellulolytic system of *Thermobifida fusca*. *Critical Reviews in Microbiology*, 40(3), pp.236–47. Available at: http://informahealthcare.com/doi/abs/10.3109/1040841X.2013.776512 [Accessed March 25, 2014].

Goodman, D.B., Church, G.M. & Kosuri, S., 2013. Causes and effects of N-terminal codon bias in bacterial genes. *Science (New York, N.Y.)*, 342(6157), pp.475–9. Available at: http://www.sciencemag.org/content/342/6157/475.short [Accessed July 14, 2014].

Govindarajan, S. Mannervik, B., Silverman, J.A., Wright, K., Regitsky, D., Hegazy, U., Purcell, T.J., Welch, M., Minshull, J., Gustafsson, C., 2014. Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase. *ACS Synthetic Biology*. Available at: http://dx.doi.org/10.1021/sb500242x [Accessed February 20, 2015].

Gubbins, M.J. Lau, I., Will, W., Manchak, J., Raivio, T., Frost, L., 2002. The positive regulator, TraJ, of the *Escherichia coli* F plasmid is unstable in a cpxA* background. *Journal of Bacteriology*, 184(20), pp.5781–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139618&tool=pmcentrez&rendertype=abstract [Accessed October 13, 2014].

Gustafsson, C. Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., Welch, M., 2012. Engineering genes for predictable protein expression. *Protein Expression and Purification*, 83(1), pp.37–46. Available at: http://www.sciencedirect.com/science/article/pii/S1046592812000629 [Accessed October 7, 2014].

Hammarström, S. & Strominger, J.L., 1975. Degradation of penicillin G to phenylacetylglycine by D-alanine carboxypeptidase from Bacillus stearothermophilus. *Proceedings of the National Academy of Sciences of the United States of America*, 72(9), pp.3463–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=433014&tool=pmcentrez&rendertype=abstract [Accessed October 27, 2014].

Hanahan, D., 1985. *DNA Cloning Techniques: A Pratical Approach* D.M.Glover, ed., Oxford: IRL Press.

Hemsworth, G.R. Henrissat, B., Davies, G., Walton, P., 2014. Discovery and

characterization of a new family of lytic polysaccharide monooxygenases. *Nature Chemical Biology*, 10(2), pp.122–6. Available at: http://dx.doi.org/10.1038/nchembio.1417 [Accessed May 27, 2014].

Hoek, A., Scholten, W.J.F. & Sharshani, A.M.A.S. Al, 2014. Integrated gas-to-liquid condensate process. Available at: https://www.google.com/patents/WO2014095815A1?cl=en&dq=gas+to+liquid+qatar&hl=en&sa=X&ei=UGwlVKHfC4SWyQS4hoKADg&ved=0CDQQ6AEwAw [Accessed September 26, 2014].

Howard, R.W. McDaniel,C., Nelson, D., Blomquist, G.L., Gelbaum, L.T., Zalkow, L.H., 1982. Cuticular hydrocarbons of *Reticulitermes virginicus* (Banks) and their role as potential species- and caste-recognition cues. *Journal of Chemical Ecology*, 8(9), pp.1227–39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24413965 [Accessed September 26, 2014].

Howard, T.P. Middelhaufe, S., Moore, K., Edner, C., Kolak, D., Taylor, G., Parker, D., Lee, R., Smirnoff, N., Aves, S., Love, J., 2013. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), pp.7636–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3651483&tool=pmcentrez&rendertype=abstract [Accessed August 15, 2014].

Hwang, I. Li, p., Zhang, L., Piper, K.R., Cook, D.M., Tate, M.E., Farrand, S.K., 1994. TraI, a LuxI homologue, is responsible for production of conjugation factor, the Ti plasmid N-acylhomoserine lactone autoinducer. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11), pp.4639–43. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=43843&tool=pmcentrez&rendertype=abstract [Accessed October 13, 2014].

Irwin, D.C. Spezio, M., Walker, L., Wilson, D.B., 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnology and Bioengineering*, 42(8), pp.1002–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18613149 [Accessed March 27, 2014].

Kallio, P. Pasztor, A., Thiel, K., Ahktar, M.K., Jones, P.R., 2014. An engineered pathway for the biosynthesis of renewable propane. *Nature Communications*, 5, p.4731. Available at: http://www.nature.com/ncomms/2014/140902/ncomms5731/full/ncomms5731.html [Accessed September 2, 2014].

Kanafusa-Shinkai, S. Wakayama, J., Tsukamoto, K., Hayashi, N., Miyasaki, Y., Ohmori, H., Tajima, K., Yokoyama, H., 2013. Degradation of microcrystalline cellulose and non-pretreated plant biomass by a cell-free extracellular cellulase/hemicellulase system from the extreme thermophilic bacterium *Caldicellulosiruptor bescii*. *Journal of Bioscience and Bioengineering*, 115(1), pp.64–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22921519 [Accessed September 25, 2014].

Kosuri, S. Goodman. D.B., Cambray, G., Mutalik, V., Gao, Y., Endy, D., Church, G., 2013. Composability of regulatory sequences controlling transcription and

translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23924614 [Accessed August 9, 2013].

Kuisiene, N., Raugalas, J. & Chitavichius, D., 2004. *Geobacillus lituanicus* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 54(Pt 6), pp.1991–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15545423 [Accessed March 7, 2014].

Lai, X. & Ingram, L.O., 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *Journal of Bacteriology*, 175(20), pp.6441–50. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=206752&tool=pmcentrez&rendertype=abstract [Accessed March 7, 2014].

Lamed, R., Setter, E. & Bayer, E.A., 1983a. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J. Bacteriol.*, 156(2), pp.828–836. Available at: http://jb.asm.org/content/156/2/828.short [Accessed March 18, 2014].

Lamed, R., Setter, E., Kenig, R., Bayer, E.A., 1983b Cellulosome: a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic cellulose binding and various cellulolytic activities. *Biotechnol. Bioeng. Symp,* 13, pp.163-181. Available at: http://agris.fao.org/agris-search/search.do?recordID=US201302651685 Accessed[February 3, 2016]

Lanka, E. & Wilkins, B.M., 1995. DNA processing reactions in bacterial conjugation. *Annual Review of Biochemistry*, 64, pp.141–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7574478 [Accessed September 25, 2014].

Larson, M.H.Glibert, L.,Wang, X., Lim, W., Weissmann, J., Qi, L., 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, 8(11), pp.2180–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24136345 [Accessed March 19, 2014].

Lee, J., 1997. Biological conversion of lignocellulosic biomass to ethanol. *Journal of Biotechnology*, 56(1), pp.1–24. Available at: http://www.sciencedirect.com/science/article/pii/S0168165697000734 [Accessed October 7, 2014].

Li, N.Chang, W., Warui, D., Booker, S.J., Krebs, C., Bollinger, J.M., , 2012. Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry*, 51(40), pp.7908–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22947199 [Accessed September 8, 2014].

Liang, C. Fioroni, M., Rodriquez-Ropero, F., Xue, F., Schwaneberg, U., Ma, Y., 2011. Directed evolution of a thermophilic endoglucanase (Cel5A) into highly active Cel5A variants with an expanded temperature profile. *Journal of Biotechnology*, 154(1), pp.46–53. Available at: http://www.sciencedirect.com/science/article/pii/S0168165611001763 [Accessed January 28, 2015].

Liao, J. Warmuth, M., Govindarajan, S., Ness, J.E., Wang, R., Gustrafsson, C., Minshull, J., 2007. Engineering proteinase K using machine learning and synthetic genes. *BMC Biotechnology*, 7(1), p.16. Available at: http://www.biomedcentral.com/1472-6750/7/16 [Accessed October 29, 2014].

Linger, J.G. Vardon, D., Guarnieri, M.T., Karp, E.M., Hunsinger, G.B., Franden, M.A., Johnsaon, C.W., Chupka, G., Strathmann, T.J., Plenkos, P.T., 2014. Lignin valorization through integrated biological funneling and chemical catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(33), pp.12013–8. Available at: http://www.pnas.org/content/111/33/12013.short [Accessed January 26, 2016].

Liu, B. & Zhang, X., 2008. Deep-sea thermophilic *Geobacillus* bacteriophage GVE2 transcriptional profile and proteomic characterization of virions. *Applied Microbiology and Biotechnology*, 80(4), pp.697–707. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18636255 [Accessed August 14, 2013].

Lochner, A. Giannone, R., Rodriguez, M., Shah, M., Mielenz, J., Keller, M., Antranakian, G., Graham, D., Hettich, R., 2011. Use of label-free quantitative proteomics to distinguish the secreted cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis*. *Applied and Environmental Microbiology*, 77(12), pp.4042–54. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3131623&tool=pmcentrez&rendertype=abstract [Accessed October 14, 2014].

Lykidis, A. Mavromatis, K., Ivanova, N., Anderson I., Land, M., DiBartolo, G., Martinez, M., Lapidus, A., Lucas, S., 2007. Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *Journal of Bacteriology*, 189(6), pp.2477–86. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1899369&tool=pmcentrez&rendertype=abstract [Accessed March 8, 2014].

De Maayer, P. Brumm, P., Mead, D., Cowan, D., 2014. Comparative analysis of the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis. *BMC Genomics*, 15(1), p.836. Available at: http://www.biomedcentral.com/1471-2164/15/836 [Accessed February 25, 2015].

Manachini, P.L. Mora, D., Nicastro, G., Parini, C, Stackebrandt, E., Pukall, R., Fortina, M.G., 2000. *Bacillus thermodenitrificans* sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology*, 50 Pt 3, pp.1331–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10843079 [Accessed October 27, 2014].

Mandels, M. & Reese, E.T., 1957. Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. *J. Bacteriol.*, 73(2), pp.269–278. Available at: http://jb.asm.org/content/73/2/269.citation [Accessed October 7, 2014].

Marks, T.J. & Hamilton, P.T., 2014. Characterization of a thermophilic bacteriophage of *Geobacillus kaustophilus*. *Archives of Virology*, 159(10), pp.2771–5. Available at: http://www.researchgate.net/publication/262073346_Characterization_of_a_thermophilic_bacteriophage_of_Geobacillus_kaustophilus [Accessed December 4, 2014].

Mcphail, L., Westcortt, P. & Lutman, H., 2011. The renawable identification number system and U.S biofuel mandates. *United state Department of Agriculture,*

Economic Research Service. Available at: http://www.ers.usda.gov/media/138383/bio03.pdf [Accessed October 9, 2014].

Mingardon, F. Bagert, J., Maisonnier, C., Trudeau, D., Arnold, F., 2011. Comparison of family 9 cellulases from mesophilic and thermophilic bacteria. *Applied and Environmental Microbiology*, 77(4), pp.1436–42. Available at: http://aem.asm.org/content/77/4/1436.full [Accessed January 29, 2015].

Narumi, I. Sawakami, K., Makamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N., Kihara, H., 1992. A newly isolated *Bacillus stearotheromophilus* K1041 and its transformation by electroporation. *Biotechnology Techniques*, 6(1), pp.83–86. Available at: http://link.springer.com/10.1007/BF02438695 [Accessed October 10, 2013].

Nazina, T.N. Lebedeva, E., Poltaraus, A., Tourova, T., Grigoryan, A., Sokolova, D., Lysenko, A., Osipov, G., 2004. *Geobacillus gargensis* sp. nov., a novel thermophile from a hot spring, and the reclassification of *Bacillus vulcani* as *Geobacillus vulcani* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 54(Pt 6), pp.2019–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15545427 [Accessed March 7, 2014].

Nazina, T.N. Sokolova, D., Grigoryan, A., Shestokova, N., Mikahailova, E., Poltaraus, A., Tourova, T., Lysenko, A., Osipov, G., Belyaev, S., 2005. *Geobacillus jurassicus* sp. nov., a new thermophilic bacterium isolated from a high-temperature petroleum reservoir, and the validation of the *Geobacillus* species. *Systematic and Applied Microbiology*, 28(1), pp.43–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15709364 [Accessed March 7, 2014].

Nazina, T.N. Tourova, T., Poltaraus, A., Novikova, E,V., Grigoryan, A., Ivanova, A.E., Lysenko, A., Petrunyaka, V., Osipov, G., Belyaev, S., Ivanov, M.V., 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus, Bacillus th*. *International Journal of Systematic and Evolutionary Microbiology*, 51(Pt 2), pp.433–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11321089 [Accessed March 7, 2014].

NREL, 2008. Biodiesel handling and use guide, fourth edition. Available at: http://www.nrel.gov/vehiclesandfuels/pdfs/43672.pdf [Accessed September 26, 2014].

Olson, D.G. McBride, J., Shaw, A., Lynd, L., 2012. Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology*, 23(3), pp.396–405. Available at: http://www.sciencedirect.com/science/article/pii/S0958166911007373 [Accessed July 9, 2014].

Ostafe, R.Prandanovic, R., Commandeur, U., Fischer, R., 2013. Flow cytometry-based ultra-high-throughput screening assay for cellulase activity. *Analytical Biochemistry*, 435(1), pp.93–8. Available at: http://www.sciencedirect.com/science/article/pii/S0003269712005623 [Accessed May 15, 2014].

Pakula, T.M. Salonen, K., Uusitalo, J., Penttila, M., 2005. The effect of specific growth

rate on protein synthesis and secretion in the filamentous fungus Trichoderma reesei. *Microbiology (Reading, England)*, 151(Pt 1), pp.135–43. Available at: http://mic.sgmjournals.org/content/151/1/135.full [Accessed October 7, 2014].

Perera, M.A.D.N., Qin., Yandeau-Nelson, M., Fan, L., Dixon, P., Nikolau, B., 2010. Biological origins of normal-chain hydrocarbons: a pathway model based on cuticular wax analyses of maize silks. *The Plant Journal : for Cell and Molecular Biology*, 64(4), pp.618–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21070415 [Accessed September 26, 2014].

Peterson, R. & Nevalainen, H., 2012. *Trichoderma reesei* RUT-C30--thirty years of strain improvement. *Microbiology (Reading, England)*, 158(Pt 1), pp.58–68. Available at: http://mic.sgmjournals.org/content/158/1/58.full#ref-49 [Accessed October 7, 2014].

Poli, A. Laezza, G., Gul-Guven, R., Orlando, P., Nicolaus, B., 2011. *Geobacillus galactosidasius* sp. nov., a new thermophilic galactosidase-producing bacterium isolated from compost. *Systematic and Applied Microbiology*, 34(6), pp.419–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21645983 [Accessed October 27, 2014].

Poli, A. Guven, K., Romano, I., Pirincciglu, H., Guven, R.G., Euzeby, J.P.M., Matpan, F., Acer, O., Orlando, P., Nicolaus, B., 2012. *Geobacillus subterraneus* subsp. *aromaticivorans* subsp. nov., a novel thermophilic and alkaliphilic bacterium isolated from a hot spring in Sırnak, Turkey. *The Journal of General and Applied Microbiology*, 58(6), pp.437–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23337579 [Accessed March 7, 2014].

Priest, F.., 1989. Validation of the publication of new names and new combinations previoulsy effectively published outside the IJSB. *Int.J.Syst.Bacteriol*, pp.93–94. Available at: http://ijs.sgmjournals.org/content/38/2/220.full.pdf [Accessed October 27, 2014].

Prodanovic, R. Ostafe, R., Blanusa, M., Schwaneburg, U., 2012. Vanadium bromoperoxidase-coupled fluorescent assay for flow cytometry sorting of glucose oxidase gene libraries in double emulsions. *Analytical and Bioanalytical Chemistry*, 404(5), pp.1439–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22811062 [Accessed May 15, 2014].

Rastogi, G. Bhalla, A., Adhikari, A., Bischoff, K., Hughes, S., Christopher, L.P., Sani, R., 2010. Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresource Technology*, 101(22), pp.8798–806. Available at: http://www.sciencedirect.com/science/article/pii/S0960852410009594 [Accessed February 21, 2014].

Ramachandra, M., Crawford, D., Hertel, G., 1988. Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete Streptomyces viridosporus. *Applied and Environmental Micrbiology,* 54(12), pp3057-3063. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0024256283&partnerID=tZOtx3y1 Accessed [February 2, 2016].

Riedel, K. Ritter, J., Bauer, S., Bronnemeirer, K., 1998. The modular cellulase CelZ of the thermophilic bacterium *Clostridium stercorarium* contains a thermostabilizing

domain. *FEMS Microbiology Letters*, 164(2), pp.261–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9682475 [Accessed May 12, 2014].

Rowland, O. Zheng, H., Hepworth, S., Lam, P., Jetter, R., Kunst, L., 2006. CER4 encodes an alcohol-forming fatty acyl-coenzyme A reductase involved in cuticular wax production in *Arabidopsis*. *Plant Physiology*, 142(3), pp.866–77. Available at: http://www.plantphysiol.org/content/142/3/866.short [Accessed September 26, 2014].

Rude, M.A. Baron, T., Brubaker, S., Alibhai, M., Del Cadayre, S., Schirmer, A., 2011. Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from Jeotgalicoccus species. *Applied and Environmental Microbiology*, 77(5), pp.1718–27. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3067255&tool=pmcentrez&rendertype=abstract [Accessed September 8, 2014].

Rudolf, A., Baudel, H., Zacchi, G., Hahn-Hagerdahl, B., Linden, G., Simultaneous saccharification and fermentation of steam-pretreated bagasse using Saccharomyces cerevisiae TMB3400 and Pichia stipitis CBS6054. *Biotechnolgy and Bioengineering,* 99(4) pp.783-90 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17787015 [Accessed February 2, 2016]

Sakon, J. Irwin, D., Wilson, D., Karplus, A., 1997. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology*, 4(10), pp.810–818. Available at: http://dx.doi.org/10.1038/nsb1097-810 [Accessed October 29, 2014].

Saunders, G.F. & Campbell, L.L., 1966. Characterization of a Thermophilic Bacteriophage for *Bacillus stearothermophilus*. *J. Bacteriol.*, 91(1), pp.340–348. Available at: http://jb.asm.org/content/91/1/340.short [Accessed December 4, 2014].

Schäffer, C. Franck, W., Scheberl, A., Kosma, P., McDermott, T., Messner, P., 2004. Classification of isolates from locations in Austria and Yellowstone National Park as *Geobacillus tepidamans* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 54(Pt 6), pp.2361–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15545484 [Accessed March 7, 2014].

Schirmer, A.Rude, M.A., Li, X., Popova, E., Del Cadayre, S., 2010. Microbial biosynthesis of alkanes. *Science (New York, N.Y.)*, 329(5991), pp.559–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20671186 [Accessed January 28, 2013].

Shulami, S. Raz-Pasteur, A., Tabachnikov, O., Gilead-Grooper, S., Shner, I., Shoham, Y., 2011. The L-Arabinan utilization system of *Geobacillus stearothermophilus*. *Journal of Bacteriology*, 193(11), pp.2838–50. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3133107&tool=pmcentrez&rendertype=abstract [Accessed November 26, 2013].

Stenesh, J., Roe, B.A. & Snyder, T.L., 1968. Studies of the deoxyribonucleic acid from mesophilic and thermophilic bacteria. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 161(2), pp.442–454. Available at: http://www.sciencedirect.com/science/article/pii/0005278768901214 [Accessed October 27, 2014].

Studholme, D.J., 2014. Some (bacilli) like it hot: genomics of *Geobacillus* species. *Microbial Biotechnology*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25195706 [Accessed October 30, 2014].

Studholme, D.J., Jackson, R.A. & Leak, D.J., 1999. Phylogenetic analysis of transformable strains of thermophilic *Bacillus* species. *FEMS Microbiology Letters*, 172(1), pp.85–90. Available at: http://doi.wiley.com/10.1111/j.1574-6968.1999.tb13454.x [Accessed October 27, 2014].

Sung, M.H. Kim, H., Bae, J., Rhee, S.K., Jeon, C.O., Kim, K., Kim, J.J., Hong, S.P., Lee, G.H., Yoon, S.J., Park, Y.H., Baek, D.H., ., 2002. Geobacillus toebii sp. nov., a novel thermophilic bacterium isolated from hay compost. *International journal of systematic and evolutionary microbiology*, 52(Pt 6), pp.2251–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12508894 [Accessed March 7, 2014].

Suzuki, H., Murakami, A. & Yoshida, K., 2012. Counterselection system for *Geobacillus kaustophilus* HTA426 through disruption of pyrF and pyrR. *Applied and Environmental Microbiology*, 78(20), pp.7376–83. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3457123&tool=pmcentrez&rendertype=abstract [Accessed November 5, 2013].

Suzuki, Y. Kishigami, T., Inoue, K., Mizoguchi, Y., Eto, N., Takagi, M., Abe, S., 1983. *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic bacilli. *Systematic and Applied Microbiology*, 4(4), pp.487–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23194806 [Accessed October 27, 2014].

Tai, S.-K. Lin, H., Kuo, J., Liu, J.K., Leak, D.J., Cowan, D.A., 2004. Isolation and characterization of a cellulolytic *Geobacillus thermoleovorans* T4 strain from sugar refinery wastewater. *Extremophiles : Life Under Extreme Conditions*, 8(5), pp.345–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15480864 [Accessed March 5, 2014].

Taylor, M.P. van Zyl, L., Tuffin, I., 2011. Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts. *Microbial Biotechnology*, 4(4), pp.438–48. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3815256&tool=pmcentrez&rendertype=abstract [Accessed March 11, 2014].

Taylor, M.P., Esteban, C.D. & Leak, D.J., 2008. Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. *Plasmid*, 60(1), pp.45–52. Available at: http://www.sciencedirect.com/science/article/pii/S0147619X08000395 [Accessed October 8, 2013].

Teather, R.M. & Wood, P.J., 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology*, 43(4), pp.777–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=241917&tool=pmcentrez&rendertype=abstract [Accessed June 15, 2014].

Tettelin, H. Masignani, V., Cieslewicz, M., Donati, C., Medini, D., Ward, N.L., Anguioli, S., Crabtree, J., Jones, A.L., Fraiser, C *et al.*2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of*

America, 102(39), pp.13950–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1216834&tool=pmcentrez&rendertype=abstract [Accessed October 5, 2014].

Trudeau, D.L., Lee, T.M. & Arnold, F.H., 2014. Engineered thermostable fungal cellulases exhibit efficient synergistic cellulose hydrolysis at elevated temperatures. *Biotechnology and Bioengineering*, 111(12), pp.2390–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24916885 [Accessed January 28, 2015].

Turner, P., Mamo, G. & Karlsson, E.N., 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microbial Cell Factories*, 6(1), p.9. Available at: http://www.microbialcellfactories.com/content/6/1/9 [Accessed October 5, 2014].

United States Congress, 2005. *Energy Policy Act 2005*, United State. Available at: http://energy.gov/sites/prod/files/2013/10/f3/epact_2005.pdf [Accessed May 8, 2015].

Vanfossen, A.L. Verhaart, M., Kengen, S., Kelly, R., 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. *Applied and Environmental Microbiology*, 75(24), pp.7718–24. Available at: http://aem.asm.org/content/75/24/7718.short [Accessed March 17, 2014].

VanFossen, A.L. et al., 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Biotechnology and Bioengineering*, 108(7), pp.1559–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21337327 [Accessed February 23, 2014].

Vuong, T. V & Wilson, D.B., 2009. Processivity, synergism, and substrate specificity of Thermobifida fusca Cel6B. *Applied and Environmental Microbiology*, 75(21), pp.6655–61. Available at: http://aem.asm.org/content/75/21/6655.full [Accessed October 30, 2014].

Wang, B. Kitney, R., Joly, N., Buck, M., 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications*, 2, p.508. Available at: http://www.nature.com/ncomms/journal/v2/n10/abs/ncomms1516.html?WT [Accessed November 10, 2013].

Wang, Y. & Zhang, X., 2010. Genome analysis of deep-sea thermophilic phage D6E. *Applied and Environmental Microbiology*, 76(23), pp.7861–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2988599&tool=pmcentrez&rendertype=abstract [Accessed December 4, 2014].

Warui, D.M. Li, N., Norgaard, H., Krebs, C., Bollinger, J.M., Booker, SJ., 2011. Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society*, 133(10), pp.3316–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3069495&tool=pmcentrez&rendertype=abstract [Accessed September 8, 2014].

Weinberger, A.D. Wolf, Y.I., Lobkovsky, A.E., Gilmore, M.S., Koonin, E.V., 2012. Viral

diversity threshold for adaptive immunity in prokaryotes. *MBio*, 3(6), pp.e00456–12. Available at: http://mbio.asm.org/content/3/6/e00456-12 [Accessed May 1, 2015].

van de Werken, H.J.G. Verhaart, M., VanFossen, A., Wilquist, K., Lewis, D.L., Goorissen, H., Mongodin, M.F., Nelson, K., van Niel, E.W.J., Kengen, S. *et al.*2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Applied and Environmental Microbiology*, 74(21), pp.6720–9. Available at: http://aem.asm.org/content/74/21/6720.short [Accessed March 13, 2014].

von Wettstein-Knowles, P., 2007. Analyses of barley spike mutant waxes identify alkenes, cyclopropanes and internally branched alkanes with dominating isomers at carbon 9. *The Plant Journal : for Cell and Molecular Biology*, 49(2), pp.250–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17241448 [Accessed September 26, 2014].

Wiegel, J., 1980. Formation of ethanol by bacteria. A pledge for the use of extreme thermophilic anaerobic bacteria in industrial ethanol fermentation processes. *Experientia*, 36(12), pp.1434–1446. Available at: http://link.springer.com/10.1007/BF01960144 [Accessed November 4, 2014].

Wilson, D.B., 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chemical Record (New York, N.Y.)*, 4(2), pp.72–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15073875 [Accessed March 6, 2014].

Wood, P.J. & Weisz, J., 1987. Detection and assay of (1-4)-beta-D-glucanse, (1-3)-beta-D-glucanse, (1-3)(1-4)-beta-D-glucanase and xylanase based on complex formation of substrate with Congo red. *Cereral Chemistry*. Available at: http://www.aaccnet.org/publications/cc/backissues/1987/Documents/64_8.pdf [Accessed June 10, 2014].

Yang, S.-J. Kataeve, I., Wiegel, J., Yin, Y., Dam, P., Xu, Y., Westphelig, J., Adams, M.W.W., 2010. Classification of "*Anaerocellum thermophilum*" strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 60(Pt 9), pp.2011–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19801388 [Accessed October 29, 2014].

Yang, S.-J. Kataeve, I., Hamilton-Brehm, S., Engle, N., Tschalpinski, T.J., Doeppke, C., Davis, M., Westphelig, J., 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. *Applied and Environmental Microbiology*, 75(14), pp.4762–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2708433&tool=pmcentrez&rendertype=abstract [Accessed October 29, 2014].

Young, J. Chung, D., Bomble, Y., Himmel, M., Westphelig, J., 2014. Deletion of C*aldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. *Biotechnology for Biofuels*, 7(1), p.142. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4195899&tool=pmcentrez&rendertype=abstract [Accessed October 28, 2014].

Zambare, V.P. Bhalla, A., Muthukumarappan, K., Sani, R.K., Christopher, L.P., 2011. Bioprocessing of agricultural residues to ethanol utilizing a cellulolytic extremophile.

Extremophiles : Life Under Extreme Conditions, 15(5), pp.611–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21837419 [Accessed February 16, 2014].

Zarilla, K.A. & Perry, J.J., 1987. *Bacillus thermokovorans*, sp. nov., a species of obligately thermophilic hydrocarbon utilizing endospore-forming bacteria. *Systematic and Applied Microbiology*, 9(3), pp.258–264. Available at: http://www.sciencedirect.com/science/article/pii/S0723202087800310 [Accessed October 27, 2014].

Zeigler, D.R., 2014. The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? *Microbiology (Reading, England)*, 160(Pt 1), pp.1–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24085838 [Accessed September 9, 2014].

Zhang, S., Lao, G. & Wilson, D.B., 1995. Characterization of a *Thermomonospora fusca* exocellulase. *Biochemistry*, 34(10), pp.3386–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7880834 [Accessed March 25, 2014].

Zverlov, V. Mahr, S., Riedel, K., Bronnenmeier, K.,1998. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile "*Anaerocellum thermophilum*" with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology (Reading, England)*, 144 (Pt 2, pp.457–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9493383 [Accessed March 14, 2014].

9. Appendices

Appendix 1 Sequence of pS797 empty vector

>ps797 conjugal vector

aattcgggagaaacgggaggagaacggaaatgtacataatcactagcgaacaattcacgcaactg
aaagaaatcgagagagtcacagatttcgtttataaacgccgtaagattaaaaatgagctgatgga
gcgtatcgccgacgcattaggccgcaagttccgcaaagcgcaaaaaacaacgcgcgtcatcgacg
aaattgtttggctcgcaacagaccgcggtttttcgttcccagggcgcgaaacgcttgccaaaaag
ttaggcgtgtccttgccgactgttgatcgtgctactcgcatgctcaaagactcgggcgaagtcgt
cgtttgctatcgtgagaatccgaatagtaacggacctaaaacgcctgtattcattttccggagcc
atgcaaatttcgagcgtattgcagccgttttaaacttgcgtgacaacgaagctgataaagtagaa
aacgccgaaaagcctacggaatcaagcgatttagcgcgtaaaacggacgctaccattagttcacc
agttgaaaaacatgacgatattaataaattaatagacaatacacatccgttggataaaattgtaa
aatacgtcactattaaagtgaacgaagtgcagaagcgtcatccttacggcattcgatatctatcc
gcatacataaacaaaacactcgacgacctaatgcaacgtgcaaaagcgaaggtcaaagccgaaaa
agctcgcgcccaacaacgcaaacaagccgactatcaaccgaccgcgaccggcttaatttggtacg
acttcacggcggacacaagcgccagccagccgacaagccagcgcaagccagcgtttgacagcgaa
gacatgtttgcgccgttattgcaggctgttgcacgaattaagggcgttgaggtgtgacacactcg
ttttaaggcgtctaacaggggatcccggtcttgccttgctcgtcggtgatgtacttcaccagctc
cgcgaagtcgctcttcttgatggagcgcatggggacgtgcttggcaatcacgcgcacccccggc
cgttttagcggctaaaaaagtcatggctctgccctcgggcggaccacgcccatcatgaccttgcc
aagctcgtcctgcttctcttcgatcttcgccagcagggcgaggatcgtggcatcaccgaaccgcg
ccgtgcgcgggtcgtcggtgagccagagtttcagcaggccgcccaggcggcccaggtcgccattg
atgcgggccagctcgcggacgtgctcatagtccacgacgcccgtgattttgtagccctggccgac
ggccagcaggtaggccgacaggctcatgccggccgccgccttttcctcaatcgctcttcgtt
cgtctggaaggcagtacaccttgataggtgggctgcccttcctggttggcttggtttcatcagcc
atccgcttgccctcatctgttacgccggcggtagccggccagcctcgcagagcaggattcccgtt
gagcaccgccaggtgcgaataagggacagtgaagaaggaacacccgctcgcgggtgggcctactt
cacctatcctgcccggctgacgccgttggatacaccaaggaaagtctacacgaaccctttggcaa
aatcctgtatatcgtgcgaaaaaggatggatataccgaaaaaatcgctataatgaccccgaagca
gggttatgcagcggaaaaggatctgccgatgaagatggattttctattattgcaatgtggaattg
ggaacggaaaaattattttattaaagagtagttcaacaaacgggccagtttgttgaagattagat

gctataattgttattaaaaggattgaaggatgcttaggaagacgagttattaatagctgaataag
aacggtgctctccaaatattcttatttagaaaagcaaatctaaaattatctgaaaagggaatgag
aatagtgaatggaccaataataatgactagagaagaaagaatgaagattgttcatgaaattaagg
aacgaatattggataaatatggggatgatgttaaggctattggtgtttatggctctcttggtcgt
cagactgatgggccctattcggatattgagatgatgtgtgtcatgtcaacagaggaagcagagtt
cagccatgaatggacaaccggtgagtggaaggtggaagtgaatttttatagcgaagagattctac
tagattatgcatctcaggtggaatcagattggccgcttacacatggtcaatttttctctattttg
ccgatttatgattcaggtggatacttagagaaagtgtatcaaactgctaaatcggtagaagccca
aaagttccacgatgcgatttgtgcccttatcgtagaagagctgtttgaatatgcaggcaaatggc
gtaatattcgtgtgcaaggaccgacaacatttctaccatccttgactgtacaggtagcaatggca
ggtgccatgttgattggtctgcatcatcgcatctgttatacgacgagcgcttcggtcttaactga
agcagttaagcaatcagatcttccttcaggttatgaccatctgtgccagttcgtaatgtctggtc
aactttccgactctgagaaacttctggaatcgctagagaatttctggaatgggattcaggagtgg
acagaacgacacggatatatagtggatgtgtcaaaacgcataccattttgaggatcttgtaaaac
gacggccagtcctgcaggcggccgctactagtatctagaagcggccgcgaattccatggtcatag
ctgtttcctgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcat
cacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtt
tccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccg
cctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtg
taggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcctt
atccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca
ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcct
aactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcgg
aaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgttt
gcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacgggg
tctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattattgacagctcg
agttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagt
tgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctg
caatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccgga
agggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccg
ggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggca
tcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcga

gttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcag
aagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtca
tgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgt
atgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaac
tttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgt
tgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcacc
agcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacg
gaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtc
tcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgctcgaaa
gctt

>C bescii CelA

MKRYRRIIAMVVTFIFILGVVYGVKPWQEVRAGSFNYGEALQKAIMFYEFQMSGKLPNWVRNNWR
GDSALKDGQDNGLDLTGGWFDAGDHVKFNLPMSYTGTMLSWAVYEYKDAFVKSGQLEHILNQIEW
VNDYFVKCHPSKYVYYYQVGDGSKDHAWWGPAEVMQMERPSFKVTQSSPGSTVVAETAASLAAAS
IVLKDRNPTKAATYLQHAKELYEFAEVTKSDAGYTAANGYYNSWSGFYDELSWAAVWLYLATNDS
TYLTKAESYVQNWPKISGSNTIDYKWAHCWDDVHNGAALLLAKITGKDIYKQIIESHLDYWTTGY
NGERIKYTPKGLAWLDQWGSLRYATTTAFLAFVYSDWVGCPSTKKEIYRKFGESQIDYALGSAGR
SFVVGFGTNPPKRPHHRTAHSSWADSQSIPSYHRHTLYGALVGGPGSDDSYTDDISNYVNNEVAC
DYNAGFVGALAKMYQLYGGNPIPDFKAIETPTNDEFFVEAGINASGTNFIEIKAIVNNQSGWPAR
ATDKLKFRYFVDLSELIKAGYSPNQLTLSTNYNQGAKVSGPYVWDASKNIYYILVDFTGTLIYPG
GQDKYKKEVQFRIAAPQNVQWDNSNDYSFQDIKGVSSGSVVKTKYIPLYDGDVKVWGEEPGTSGA
TPTPTATATPTPTPTVTPTPTPTPTSTATPTPTPTPTVTPTPTPTPTATPTATPTPTSTPSSTPV
AGGQIKVLYANKETNSTTNTIRPWLKVVNTGSSSIDLSRVTIRYWYTVDGDKAQSAISDWAQIGA
SNVTFKFVKLSSSVSGADYYLEIGFKSGAGQLQAGKDTGEIQIRFNKSDWSNYNQGNDWSWMQSM
TNYGENVKVTAYIDGVLVWGQEPSGATPTPTATPAPTVTPTPTPTPTSTPTATPTATPTPTPTPS
STPVAGGQIKVLYANKETNSTTNTIRPWLKVVNTGSSSIDLSRVTIRYWYTVDGDKAQSAISDWA
QIGASNVTFKFVKLSSSVSGADYYLEIGFKSGAGQLQAGKDTGEIQIRFNKSDWSNYNQGNDWSW
MQSMTNYGENVKVTAYIDGVLVWGQEPSGATPTPTATPAPTVTPTPTPAPTPTPTPTPTATPTPT
PTPTPTATPTVTATPTPTPSSTPSVLGEYGQRFMWLWNKIHDPANGYFNQDGIPYHSVETLICEA
PDYGHLTTSEAFSYYVWLEAVYGKLTGDWSKFKTAWDTLEKYMIPSAEDQPMRSYDPNKPATYAG

EWETPDKYPSPLEFNVPVGKDPLHNELVSTYGSTLMYGMHWLMDVDNWYGYGKRGDGVSRASFIN
TFQRGPEESVWETVPHPSWEEFKWGGPNGFLDLFIKDQNYSKQWRYTDAPDADARAIQATYWAKV
WAKEQGKFNEISSYVAKAAKMGDYLRYAMFDKYFKPLGCQDKNAAGGTGYDSAHYLLSWYYAWGG
ALDGAWSWKIGSSHVHFGYQNPMAAWALANDSDMKPKSPNGASDWAKSLKRQIEFYRWLQSAEGA
IAGGATNSWNGRYEKYPAGTATFYGMAYEPNPVYHDPGSNTWFGFQAWSMQRVAEYYYVTGDKDA
GALLEKWVSWVKSVVKLNSDGTFAIPSTLDWSGQPDTWNGAYTGNSNLHVKVVDYGTDLGITASL
ANALLYYSAGTKKYGVFDEGAKNLAKELLDRMWKLYRDEKGLSAPEKRADYKRFFEQEVYIPAGW
IGKMPNGDVIKSGVKFIDIRSKYKQDPDWPKLEAAYKSGQAPEFRYHRFWAQCDIAIANATYEIL
FGNQ

>C. saccharolyticus CelB

MKRNLFRIVSRVVLIAFIASISLVGAMSYFPVETQAAPDWSIPSLCESYKDDFMIGVAIPARCLS
NDTDKRMVLKHFNSITAENEMKPESLLAGQTSTGLSYRFSTADAFVDFASTNKIGIRGHTLVWHN
QTPDWFFKDSNGQRLSKDALLARLKQYIYDVVGRYKGKVYAWDVVNEAIDENQPDSYRRSTWYEI
CGPEYIEKAFIWAHEADPNAKLFYNDYNTEISKKRDFIYNMVKNLKSKGIPIHGIGMQCHINVNW
PSVSEIENSIKLFSSIPGIEIHITELDMSLYNYGSSENYSTPPQDLLQKQSQKYKEIFTMLKKYK
NVVKSVTFWGLKDDYSWLRSFYGKNDWPLLFFEDYSAKPAYWAVIEASGVTTSSPTPTPTPTVTV
TPTPTPTPTPTVTATPTPTPTPVSTPATGGQIKVLYANKETNSTTNTIRPWLKVVNSGSSSIDLS
RVTIRYWYTVDGERAQSAVSDWAQIGASNVTFKFVKLSSSVSGADYYLEIGFKSGAGQLQPGKDT
GEIQIRFNKSDWSNYNQGNDWSWLQSMTSYGENEKVTAYIDGVLVWGQEPSGATPAPTMTVAPTA
TPTPTLSPTVTPTPAPTQTAIPTPTLTPNPTPTSSIPDDTNDDWLYVSGNKIVDKDGRPVWLTGI
NWFGYNTGTNVFDGVWSCNLKDTLAEIANRGFNLLRVPISAELILNWSQGIYPKPNINYYVNPEL
EGKNSLEVFDIVVQTCKEVGLKIMLDIHSIKTDAMGHIYPVWYDEKFTPEDFYKACEWITNRYKN
DDTIIAFDLKNEPHGKPWQDTTFAKWDNSTDINNWKYAAETCAKRILNINPNLLIVIEGIEAYPK
DDVTWTSKSSSDYYSTWWGGNLRGVRKYPINLGKYQNKVVYSPHDYGPSVYQQPWFYPGFTKESL
LQDCWRPNWAYIMEENIAPLLIGEWGGHLDGADNEKWMKYLRDYIIENHIHHTFWCFNANSGDTG
GLVGYDFTTWDEKKYSFLKPALWQDSQGRFVGLDHKRPLGTNGKNINITTYYNNNEPEPVPASK

>T. neopolitania CelB

MRLVVSFLLVVSAFLFSAEVVLTDIGATDITFKGFPVTMELNFWNVKSYEGETWLKFDGEKVQFY
ADIYNIVLQNPDSWVHGYPEIYYGYKPWAAHNSGTEILPVKVKDLPDFYVTLDYSIWYENDLPIN
LAMETWITRKPDQTSVSSGDVEIMVWFYNNILMPGGQKVDEFTTTIEINGSPVETKWDVYFAPWG

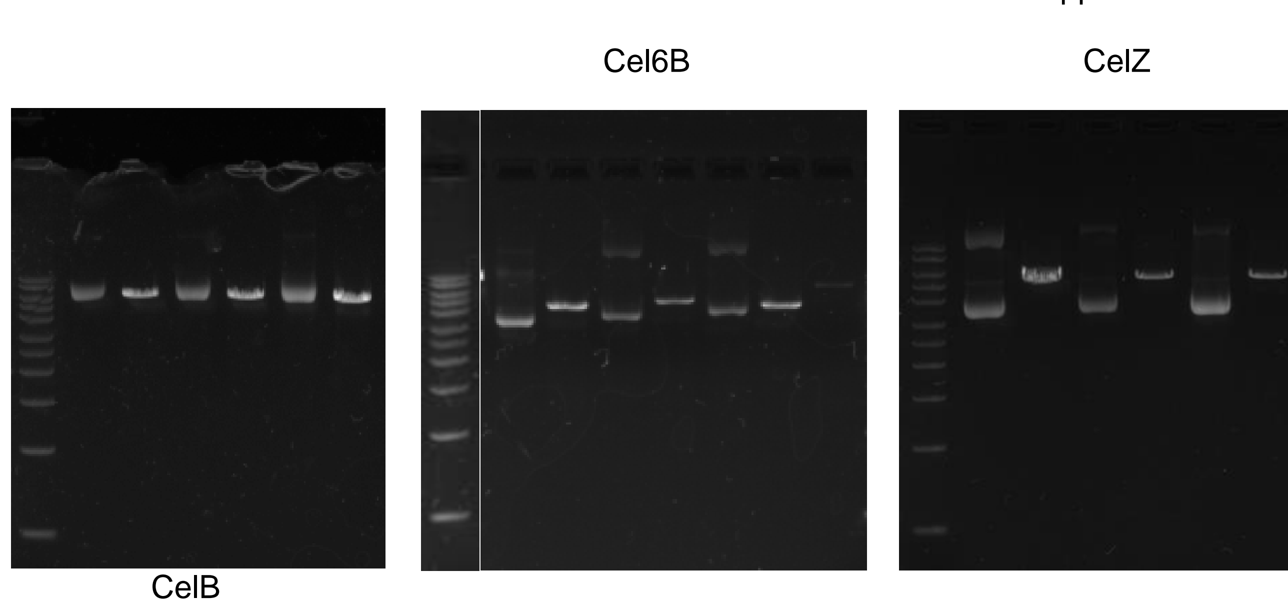
WDYLAFRLTTPMKDGRVKFNVKDFVEKAAEVIKKHSTRVENFDEMYFCVWEIGTEFGDPNTTAAK
FGWTFKDFSVEIGE

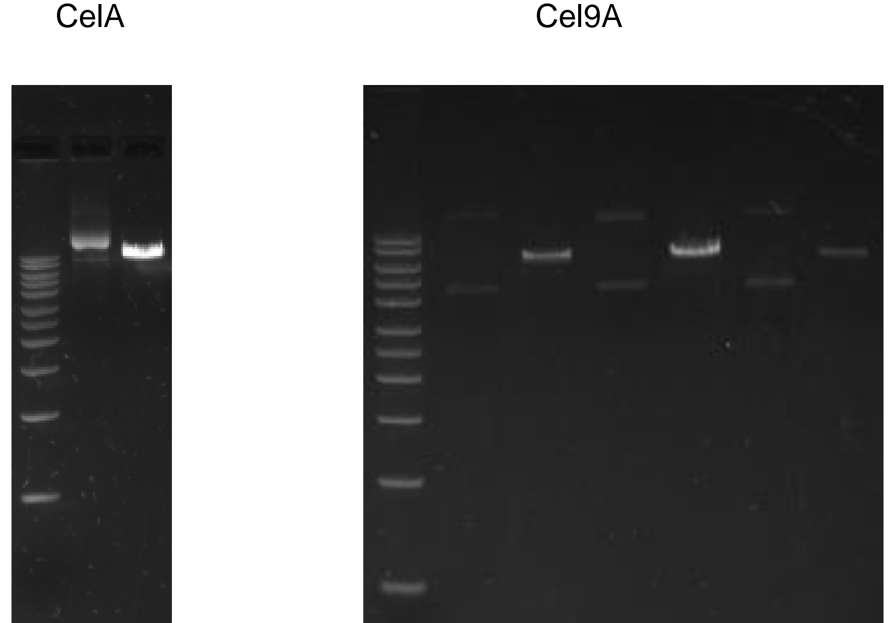
>T. fusca Cel6B

MSKVRATNRRSWMRRGLAAASGLALGASMVAFAAPANAAGCSVDYTVNSWGTGFTANVTITNLGS
AINGWTLEWDFPGNQQVTNLWNGTYTQSGQHVSVSNAPYNASIPANGTVEFGFNGSYSGSNDIPS
SFKLNGVTCDGSDDPDPEPSPSPSPSPSPTDPDEPGGPTNPPTNPGEKVDNPFEGAKLYVNPVWS
AKAAAEPGGSAVANESTAVWLDRIGAIEGNDSPTTGSMGLRDHLEEAVRQSGGDPLTIQVVIYNL
PGRDCAALASNGELGPDELDRYKSEYIDPIADIMWDFADYENLRIVAIIEIDSLPNLVTNVGGNG
GTELCAYMKQNGGYVNGVGYALRKLGEIPNVYNYIDAAHHGWIGWDSNFGPSVDIFYEAANASGS
TVDYVHGFISNTANYSATVEPYLDVNGTVNGQLIRQSKWVDWNQYVDELSFVQDLRQALIAKGFR
SDIGMLIDTSRNGWGGPNRPTGPSSSTDLNTYVDESRIDRRIHPGNWCNQAGAGLGERPTVNPAP
GVDAYVWVKPPGESDGASEEIPNDEGKGFDRMCDPTYQGNARNGNNPSGALPNAPISGHWFSAQF
RELLANAYPPL

>T. fusca Cel9A

MSVTEPPPRRRGRHSRARRFLTSLGATAALTAGMLGVPLATGTAHAEPAFNYAEALQKSMFFYEA
QRSGKLPENNRVSWRGDSGLNDGADVGLDLTGGWYDAGDHVKFGFPMAFTATMLAWGAIESPEGY
IRSGQMPYLKDNLRWVNDYFIKAHPSPNVLYVQVGDGDADHKWWGPAEVMPMERPSFKVDPSCPG
SDVAAETAAAMAASSIVFADDDPAYAATLVQHAKQLYTFADTYRGVYSDCVPAGAFYNSWSGYQD
ELVWGAYWLYKATGDDSYLAKAEYEYDFLSTEQQTDLRSYRWTIAWDDKSYGTYVLLAKETGKQK
YIDDANRWLDYWTVGVNGQRVPYSPGGMAVLDTWGALRYAANTAFVALVYAKVIDDPVRKQRYHD
FAVRQINYALGDNPRNSSYVVGFGNNPPRNPHHRTAHGSWTDSIASPAENRHVLYGALVGGPGSP
NDAYTDDRQDYVANEVATDYNAGFSSALAMLVEEYGGTPLADFPPTEEPDGPEIFVEAQINTPGT
TFTEIKAMIRNQSGWPARMLDKGTFRYWFTLDEGVDPADITVSSAYNQCATPEDVHHVSGDLYYV
EIDCTGEKIFPGGQSEHRREVQFRIAGGPGWDPSNDWSFQGIGNELAPAPYIVLYDDGVPVWGTA
PEEGEEPGGGEGPGGGEEPGEDVTPPSAPGSPAVRDVTSTSAVLTWSASSDTGGSGVAGYDVFLR
AGTGQEQKVGSTTRTSFTLTGLEPDTTYIAAVVARDNAGNVSQRSTVSFTTLAENGGGPDASCTV
GYSTNDWDSGFTASIRITYHGTAPLSSWELSFTFPAGQQVTHGWNATWRQDGAAVTATPMSWNSS
LAPGATVEVGFNGSWSGSNTPPTDFTLNGEPCALA





Appendix 2 His Tag diagnostic digest data

CelZ gel is laid out differently, one undigested control and 6 digested plasmids, all others are undigested then digested etc.

Appendix 3 Sequencing of pS797 *CelA::6xHis* showing introduction of 6x His Tag by Q5 Site Directed Mutagenesis.

```
> pS797 C. bescii CelA

AAGCTTTCGAGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA
ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCT
GCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATC
ATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTT
GCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
TCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGC
CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTCGAGCTGTCAATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT
GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACA
GCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGCTGCCTCGTCCATTTTTTT
GCTTAATGGAGGTTGTCATGAAAATGACAAACAACGTCCAAACAATTGCCATAATCGTTTACGCATAGTTTCGATTTCATCGCGTAAAATAATTTGT
GAATGTATTCACAATAATAAGAAGGGAGAATAGTGATGAAGCGTTATCGTCGCATCATCGCTATGGTCGTGACGTTCATCTTCATTCTGGGTGTGGT
GTACGGCGTTAAACCGTGGCAAGAGGTTCGTGCTGGCAGCTTCAATTACGGTGAAGCGCTCCAGAAAGCGATCATGTTCTACGAGTTTCAGATGAGC
GGTAAGCTGCCGAACTGGGTCCGCAATAACTGGCGTGGTGACTCTGCGCTGAAAGACGGTCAAGACAACGGTCTGGACCTGACGGGTGGTTGGTTTG
ACGCTGGCGATCACGTGAAGTTCAATCTGCCGATGTCTTATACCGGCACCATGCTGTCTTGGGCGGTATACGAGTATAAAGACGCGTTTGTTAAGAG
CGGCCAACTGGAACACATTCTGAATCAAATCGAGTGGGTTAATGACTATTTCGTCAAGTGTCACCCGAGCAAATACGTGTACTATTATCAGGTCGGT
GATGGTAGCAAAGATCATGCGTGGTGGGGTCCGGCCGAGGTCATGCAGATGGAACGTCCGTCGTTTAAGGTCACGCAGTCTAGCCCGGGTAGCACGG
TTGTGGCGGAGACTGCCGCAAGCCTGGCTGCGGCCTCCATTGTTCTGAAAGATCGTAATCCAACCAAAGCCGCGACCTATCTGCAGCACGCGAAAGA
ATTGTACGAGTTTGCAGAGGTGACCAAGTCCGACGCAGGCTACACCGCGGCAAATGGCTATTACAATAGCTGGAGCGGTTTCTATGACGAACTGAGC
TGGGCGGCAGTGTGGCTGTACCTGGCAACGAATGACAGCACCTACCTGACCAAAGCGGAGAGCTACGTCCAAAATTGGCCGAAGATCTCCGGTAGCA
ACACCATTGATTACAAATGGGCACACTGCTGGGACGATGTCCATAATGGCGCAGCGCTTCTGTTGGCAAAAATCACCGGTAAAGACATCTACAAACA
GATTATTGAGAGCCACCTGGACTACTGGACCACCGGTTACAATGGTGAACGCATCAAGTATACCCCGAAAGGCCTGGCGTGGCTGGATCAATGGGGC
AGCTTACGTTATGCAACCACGACCGCTTTTCTGGCCTTCGTCTACAGCGACTGGGTTGGTTGTCCGAGCACCAAGAAAGAGATTTATCGCAAATTCG
GTGAAAGCCAGATCGACTACGCGCTGGGTTCAGCCGGTCGTAGCTTTGTTGTCGGTTTTGGCACGAACCCGCCGAAACGTCCGCACCACCGTACTGC
GCACAGCAGCTGGGCCGATTCCCAATCTATTCCGAGCTATCACCGTCATACCCTGTACGGTGCTCTGGTGGGTGGTCCGGGTAGCGACGACAGCTAT
ACCGATGACATTAGCAATTATGTTAACAATGAAGTGGCGTGCGACTACAACGCTGGTTTCGTTGGTGCCCTGGCGAAGATGTACCAGCTGTATGGCG
GCAACCCGATTCCGGACTTTAAAGCAATCGAAACCCCGACGAACGATGAGTTCTTCGTCGAGGCAGGCATCAACGCCAGCGGTACCAACTTTATCGA
GATCAAGGCCATTGTCAATAATCAGAGCGGCTGGCCAGCGCGTGCGACCGATAAACTGAAATTTCGTTATTTCGTGGACTTGAGCGAGCTGATTAAA
GCTGGCTACAGCCCGAATCAACTGACCCTGTCTACCAATTACAATCAGGGCGCAAAAGTTTCTGGCCCGTACGTTTGGGACGCCAGCAAGAATATCT
ACTATATCCTGGTTGACTTTACGGGTACCTTGATCTACCCGGGTGGCCAGGATAAGTACAAGAAAGAGGTCCAGTTCCGTATTGCAGCACCGCAAAA
CGTTCAGTGGGACAATAGCAACGACTATAGCTTTCAAGACATCAAAGGCGTGAGCAGCGGTAGCGTCGTTAAGACCAAGTACATTCCATTGTATGAT
GGTGATGTCAAGGTGTGGGGTGAGGAACCGGGTACGAGCGGTGCAACCCCGACCCCGACGGCAACTGCGACCCCTACGCCGACCCCTACCGTTACTC
CGACCCCGACTCCGACGCCGACGTCCACGGCCACGCCAACGCCTACCCCAACTCCGACCGTAACGCCGACCCCGACCCCGACTCCGACCGCTACCCC
GACGGCAACGCCGACGCCGACCTCTACCCCGAGCAGCACCCCTGTTGCGGGTGGTCAGATTAAGGTACTGTACGCAAACAAAGAAACCAACAGCACC
```

ACCAACACGATTCGCCCGTGGTTGAAAGTCGTTAACACCGGTAGCTCCTCGATTGATCTGAGCCGCGTTACCATCCGCTACTGGTACACCGTCGATG
GCGATAAGGCACAGAGCGCGATCAGCGATTGGGCCCAGATTGGTGCAAGCAATGTCACCTTTAAGTTCGTGAAGCTGTCATCCAGCGTTTCTGGCGC
GGACTATTATCTGGAAATTGGTTTCAAATCCGGTGCGGGCCAACTGCAGGCTGGTAAGGATACCGGTGAAATCCAGATTCGCTTTAACAAAAGCGAT
TGGAGCAACTATAACCAAGGTAACGATTGGTCCTGGATGCAAAGCATGACCAACTACGGCGAGAATGTCAAGGTGACCGCCTACATTGATGGTGTCT
TAGTCTGGGGTCAAGAGCCGTCCGGTGCGACGCCTACCCCGACGGCGACGCCGGCACCGACCGTTACCCCTACGCCGACTCCGACTCCGACGAGCAC
CCCGACCGCCACGCCTACCGCAACCCCGACCCCGACCCCAACGCCGAGCTCCACCCCGGTTGCCGGTGGTCAGATTAAAGTGCTGTACGCGAACAAG
GAAACCAACAGCACCACCAATACGATTCGTCCGTGGCTGAAAGTTGTGAACACGGGCAGCAGCAGCATTGACCTGTCTCGTGTGACCATTCGTTATT
GGTATACCGTGGATGGCGATAAAGCGCAGTCTGCAATCAGCGATTGGGCTCAGATCGGCGCTTCAAACGTTACTTTTAAGTTCGTTAAACTGAGCAG
CAGCGTTTCCGGTGCAGACTACTATCTGGAAATCGGTTTCAAGAGCGGTGCCGGCCAATTGCAAGCGGGCAAGGATACGGGCGAGATCCAGATCCGC
TTCAATAAGAGCGACTGGAGCAACTATAATCAGGGCAACGACTGGAGCTGGATGCAATCCATGACGAATTACGGCGAGAACGTGAAGGTCACGGCGT
ATATCGATGGTGTTCTGGTGTGGGGCCAAGAGCCGAGCGGTGCCACCCCGACCCCAACGGCGACCCCTGCGCCGACGGTCACCCCGACCCCAACGCC
TGCACCAACCCCAACCCCGACGCCGACTCCGACTGCGACTCCTACCCCTACTCCGACGCCAACCCCGACGGCCACGCCGACCGTGACGGCAACCCCG
ACCCCGACGCCGTCCAGCACGCCGAGCGTTTTGGGTGAATACGGTCAGCGCTTCATGTGGCTGTGGAATAAGATCCACGACCCGGCGAATGGTTATT
TCAACCAAGATGGCATTCCGTACCACTCGGTTGAAACCCTGATCTGCGAGGCTCCGGATTACGGCCATCTGACGACCTCCGAAGCATTCAGCTATTA
CGTGTGGCTGGAAGCGGTTTACGGCAAGCTGACGGGCGATTGGAGCAAGTTTAAGACCGCTTGGGACACTCTGGAGAAGTACATGATTCCGAGCGCT
GAAGATCAACCGATGCGTTCCTACGACCCGAACAAGCCGGCTACCTACGCGGGTGAATGGGAAACGCCGGACAAATATCCGAGCCCGCTGGAGTTCA
ATGTGCCGGTTGGCAAAGATCCGCTGCACAATGAACTGGTGAGCACGTATGGCAGCACCCTGATGTATGGTATGCACTGGCTGATGGATGTCGATAA
TTGGTATGGTTACGGTAAGCGCGGTGACGGTGTGAGCCGCGCATCTTTCATTAACACCTTTCAGCGTGGTCCGGAAGAGAGCGTGTGGGAAACCGTC
CCTCACCCGAGCTGGGAAGAGTTTAAGTGGGGTGGCCCGAACGGTTTCTTGGATCTGTTTATCAAAGATCAGAACTATTCGAAGCAATGGCGTTACA
CCGATGCGCCGGACGCGGACGCTCGTGCCATCCAGGCGACCTATTGGGCCAAAGTGTGGGCGAAAGAGCAAGGTAAGTTTAATGAAATCAGCAGCTA
TGTGGCGAAAGCGGCGAAGATGGGCGACTATCTGCGTTATGCGATGTTTGACAAATACTTTAAGCCATTGGGCTGTCAAGATAAGAACGCTGCTGGC
GGCACCGGCTATGATAGCGCGCACTATCTGCTGAGCTGGTATTACGCGTGGGGTGGTGCGCTGGACGGTGCGTGGTCGTGGAAGATTGGTAGCTCTC
ATGTGCATTTTGGTTACCAAAACCCGATGGCGGCCTGGGCCCTGGCTAATGATAGCGACATGAAGCCGAAGTCGCCGAATGGTGCGTCGGATTGGGC
GAAGTCCTTGAAACGTCAAATTGAGTTCTACAGATGGCTGCAGAGCGCAGAGGGTGCGATCGCGGGTGGTGCCACCAACTCTTGGAACGGTCGTTAT
GAGAAATATCCGGCAGGCACGGCAACCTTTTACGGTATGGCGTACGAGCCGAATCCGGTGTACCATGATCCGGGCAGCAACACCTGGTTCGGTTTCC
AAGCGTGGAGCATGCAACGCGTTGCAGAGTACTACTACGTCACGGGTGACAAAGACGCAGGCGCGCTGCTGGAGAAATGGGTGAGCTGGGTTAAGAG
CGTGGTGAAGCTGAACAGCGACGGTACGTTCGCAATCCCGAGCACCCTGGACTGGTCGGGCCAGCCGGATACGTGGAACGGTGCGTATACCGGTAAC
AGCAATCTGCATGTCAAAGTCGTCGATTATGGTACTGACCTGGGTATTACCGCGAGCCTGGCCAATGCACTGTTGTACTACAGCGCGGGTACCAAAA
AGTATGGCGTTTTTGATGAAGGCGCAAAGAACCTGGCGAAAGAACTGCTGGACCGTATGTGGAAACTGTACCGTGACGAGAAAGGCCTGTCTGCACC
AGAGAAACGTGCGGACTACAAGCGTTTCTTCGAGCAAGAAGTCTACATCCCGGCTGGTTGGATTGGCAAGATGCCGAACGGTGATGTGATCAAATCT
GGCGTCAAATTCATCGACATTCGCAGCAAATACAAACAGGACCCTGACTGGCCGAAACTGGAAGCGGCCTACAAAAGCGGCCAAGCGCCGGAGTTTC
GCTACCATCGTTTCTGGGCGCAGTGCGATATTGCCATTGCGAATGCGACTTATGAGATCTTGTTTGGCAATCAGTAATAAAAATGGAAATGATTCAC
ACCCGATGTGGACTTGCCTCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAA
ATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTTCTC
AGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGAAGCG
CTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTGTCGGTCCTT
GCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACTTTTGGGCTTCTACCGATTT
AGCAGTTTGATACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCGGCCAATCTGATTCCACC
TGAGATGCATAATCTAGTAGAATCTCTTCGCTATAAAAATTCACTTCCACCTTCCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCCTCTG
TTGACATGACACACATCATCTCAATATCCGAATAGGGCCCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCATA
TTTATCCAATATTCGTTCCTTAATTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGGTCCATTCACTATTCTCATTCCCTTTTCA
GATAATTTTAGATTTGCTTTTCTAAATAAGAATATTTGGAGAGCACCGTTCTTATTCAGCTATTAATAACTCGTCTTCCTAAGCATCCTTCAATCCT
TTTAATAACAATTATAGCATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAAATAATTTTTCCGTTCCCAATTCCACATTGC
AATAATAGAAAATCCATCTTCATCGGCAGATCCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTT
TCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACC
CGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTA
ACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGA
TTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCA

CGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGT

GATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCAT

GACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGA

AGTACATCACCGACGAGCAAGGCAAGACCGCCTAGGCCTGTTAGACGCCTTAAAACGAGTGTGTCACACCTCAACGCCCTTAATTCGTGCAACAGCC

TGCAATAACGGCGCAAACATGTCTTCGCTGTCAAACGCTGGCTTGCGCTGGCTTGTCGGCTGGCTGGCGCTTGTGTCCGCCGTGAAGTCGTACCAAA

TTAAGCCGGTCGCGGTCGGTTGATAGTCGGCTTGTTTGCGTTGTTGGGCGCGAGCTTTTTCGGCTTTGACCTTCGCTTTTGCACGTTGCATTAGGTC

GTCGAGTGTTTTGTTTATGTATGCGGATAGATATCGAATGCCGTAAGGATGACGCTTCTGCACTTCGTTCACTTTAATAGTGACGTATTTTACAATT

TTATCCAACGGATGTGTATTGTCTATTAATTTATTAATATCGTCATGTTTTTCAACTGGTGAACTAATGGTAGCGTCCGTTTTACGCGCTAAATCGC

TTGATTCCGTAGGCTTTTCGGCGTTTTCTACTTTATCAGCTTCGTTGTCACGCAAGTTTAAAACGGCTGCAATACGCTCGAAATTTGCATGGCTCCG

GAAAATGAATACAGGCGTTTTAGGTCCGTTACTATTCGGATTCTCACGATAGCAAACGACGACTTCGCCCGAGTCTTTGAGCATGCGAGTAGCACGA

TCAACAGTCGGCAAGGACACGCCTAACTTTTTGGCAAGCGTTTCGCGCCCTGGGAACGAAAAACCGCGGTCTGTTGCGAGCCAAACAATTTCGTCGA

TGACGCGCGTTGTTTTTTGCGCTTTGCGGAACTTGCGGCCTAATGCGTCGGCGATACGCTCCATCAGCTCATTTTTAATCTTACGGCGTTTATAAAC

GAAATCTGTGACTCTCTCGATTTCTTTCAGTTGCGTGAATTGTTCGCTAGTGATTATGTACATTTCCGTTCTCCTCCCGTTTCTCCCGAATT

CelA his tag addition 1 Fwd

NNNNNNNNNNNNNNNNNGANAGCAAAAAACACCCGATGAGGCAAGTCCACATCGGGTGTGAATCATTTCCATTTTTATTAG

TGGTGGTGATGATGATGCTGATTGCCAAACAAGATCTCATAAGTCGCATTCGCAATGGCAATATCGCACTGCGCCCAGAA

ACGATGGTAGCGAAACTCCGGCGCTTGGCCGCTTTTGTAGGCCGCTTCCAGTTTCGGCCAGTCAGGGTCCTGTTTGTATT

TGCTGCNAATGTCNATGAATTTGACGCCNGANTTGATCACATCACCGTTCGGCATCTTGCCAATCCAACCAGCCGGGATG

TAGACTTCTTGCTCGAANAAACGCTTGTANTCCGCACGTTTCTCTGGTGCANACAGGCCTTTCTCGNCACGGTACAGTTT

CCACATACGGTCCANCAGNTCNTTCGCCAGGTTCTTTGCGCCTTCATCAAAAACGCCNTANTTTTTGGTACCCGCGCTGT

ANNACANCANTGNNTTGGCCANNNTCGCGGTAATACCCANGTCAGTACCATANTCNACGACTTTGACATGCNNATTGCTGNCANCGNTCNNCGTATC

CGGCTGGCCNGACCANNCNAGGGTGCTCGGGATTGCGAACGTACCGTCGCT

GTTCANCTTCACCACGCTCTTAACCCANCTCACNNATTTCTCCAGCANCGCGCCTGCNTCTTTGTCACNNGTGACGTANT

ANTACTCTGCAACGCGTTGCATGCTCCACGCTTGNAAACCGAACCANGTGTTGCTGCCCGGATCATGGNACACCGNANTC

NGCTCGNANGCCATACCGTAAAANGNTGCCGNGCCTGCCGGANATNTCTCATAACGACNNNNNNNNNAGTTGGTGGCACC

ANNNNCGATCGCACCCTCTGCGCTCTGCANCCNTCTGNAGAACTNNNNNTGNNNNTTNANGACTTCGNCNNNTNNGANNN

NNNNTCGGCGANTTCGGNNTNNNGNNGNTATCNNNANCNNGNNNNNNCCNNNNTNNGGTTNNNNANNCNNNNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNCNNNNANCCNNNNNNNNANNANNNNNNNNNNNGCANNNNNNNNN

NNNNNNNNNNNNNNNNNGGNN

Red= STOP

Green= His Tag

Purple= CelA

pS797 CelA + His M13 fwd RC & pS797 CelA

CelA CAGGACCCTGACTGGCCGAAACTGGAAGCGGCCTACAAAAGCGGCCAAGCGCCGGAGTTT 968

pS797 CAGGACCCTGACTGGCCGAAACTGGAAGCGGCCTACAAAAGCGGCCAAGCGCCGGAGTTT 7080

****

CelA CGCTACCATCGTTTCTGGGCGCAGTGCGATATTGCCATTGCGAATGCGACTTATGAGATC 1028

```
pS797           CGCTACCATCGTTTCTGGGCGCAGTGCGATATTGCCATTGCGAATGCGACTTATGAGATC 7140
                ************************************************************


CelA            TTGTTTGGCAATCAGCATCATCATCACCACCACTAATAAAAATGGAAATGATTCACACCC 1088
pS797           TTGTTTGGCAATCAG------------------TAATAAAAATGGAAATGATTCACACCC 7182
                ***************                  ***************************


CelA            GATGTGGACTTGCCTCATCGGGTG-TTTTTTGCTNTCNNNNNNNNNNNNNNNN------- 1140
pS797           GATGTGGACTTGCCTCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCT 7242
                ************************
```

>CelZ M13 fwd RC

NNNNNNNNNNNGNNNNNANNNNNACNANNNNNNNNNTCANCNNTNNCNANNNNNNNGNCNTTNNNNNNNNNANNNNCNGGNNNANNNNNNNNNNNNAAN
NATTCAANNNGGTCNTNNNCNNNNNNGNTAACNNNNTNNANNGNCATTNAGTANGNTNANCNNNTNTNTGCNNNAGGGNACCGGNNTNNNCGGT
NAGNNGNGACACNGNGACGATNNNNAANTNTTTNNNAANCANNNNTGATNCCAGNNCCGTTCANNTGATTTTNGATTTCAGCGNGGGTCGTGAN
CCGNTTNTGACCGTTAACATTANNGNCANCNNCACGAGCNNCTCCNTCGTCCCGANTACTGCGGATTTCGATAAGAATNNGATGCAAGCCGCGA
TGTCAAGGTGAAACTGGTTCCGAATGGCAACACGCTGCTGGCAGTGAAGAAAGACGGCGAGGCGCTGGTGTTGGGCCGTGATTACAGCATCGAT
GGCGATGAAAGTTACCATTTTCCGTGAGTATTTGGCGGACCAGCCAGTGGGTCGTGTTACCCTGACCTTCGACTTCGACCGTGGCACGGATCCG
GTCCTGACCATCAACATCACCGACAGCCGCCAGGTTGAAACGGGTGTGATTCAAATTCAGATGTTTAATGGTAATACGAGCGATAAAACCAATG
GTATCATGCCGCGCTACCGTCTGACGAATACCGGTACCACCCCGATTCGCCTGTCCGATGTGAAGATCCGCTATTACTACACTATTGACGGTGA
AAAGGACCAGAATTTCTGGTGCGACTGGTCCTCAGTCGGCTCGAACAACATTACGGGCACCTTCGTAAAGATGGCGGAGCCGAAAGAGGGCGCA
GACTATTACCTGGAGACTGGTTTTACGGACGGTGCGGGCTATCTGCAACCGAATCAAAGCATTGAAGTCCAGAATCGCTTCAGCAAGGCCGATT
GGACGGACTACATCCAAACGAATGACTACAGCTTCAGCACCAATACGAGCTATGGCAGCAACGACCGTATTACCGTTTACATTAGCGGTGTGTT
GGTGAGCGGTATCGAGCCGCATCATCATCNCCNNCNNNNNNNNAAAANGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGNNTNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

> pS797 C. stercorarium CelZ

AAGCTTTCGAGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCA
ATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTG
AGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGA
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTT
TGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCC
TGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGAT
AAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGAT
CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTCGAGCTGTCAATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC

AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACAGGAAAC
AGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGCTGCCTCGTCCATTTTTTTGCTTAATGGAGGTTGTCATGAAAATGACAAACAACGTCCA
AACAATTGCCATAATCGTTTACGCATAGTTTCGATTTCATCGCGTAAAATAATTTGTGAATGTATTCACAATAATAAGAAGGGAGAATAGTGAT
GCGCAAGTTTTGGAGCTTTGCGATTATCATTAGCCTGCTGGTGACGGGTCTGTTCATTCATACCCCGAAGGCCGAGGCAGCCGGCTACAATTAC
GGTGAGGCTCTGCAAAAAGCCATTATGTTCTACGAGTTCCAGCGTTCTGGCAAATTGCCTGAGAACAAGCGTGACAACTGGCGCGGTGATTCTG
GTTTGAATGACGGTGCAGACGTCGGCCTTGATCTGACCGGTGGCTGGTATGATGCGGGCGATCACGTTAAGTTTAATCTGCCGATGGCGTACAG
CCAAACCATGTTGGCTTGGGCGGCATACGAAGCAGAAGAAGCGCTGGAAAGATCCGGCCAGATGGGTTATCTGCTGGACGCAATCAAATGGGTT
AGCGATTACCTGATTAAGTGTCACCCAAGCCCGAATGTGTTCTACTATCAGGTCGGCGACGGTCACCTGGATCACTCCTGGTGGGGTCCGGCTG
AGGTTATGCAGATGGATCGTCCGGCGTACAAAGTTGATCTGGCGAACCCGGGTAGCACCGTGGTCGCGGAAGCTGCTGCAGCTCTGGCAAGCGC
AGCCGTCGTGTTTGCGGATCGTGATCCGGCCTATGCGGCGACGTGTATTCAGCATGCGAAAGAACTGTACAACTTTGCGGAGATTACCAAGAGC
GACAGCGGTTACACCGCAGCCAGCGGCTTCTATGACTCTCATAGCGGTTTCTACGATGAGCTGAGCTGGGCTGGCGTGTGGCTGTACCTGGCGA
CCGGTGACGAAACCTACTTAAACAAAGCAGAGCAGTATGTTGCGTACTGGGGCACCGAACCGCAGACGAACATCATCAGCTACAAATGGGCGCA
CTGCTGGGATGACGTACACTATGGCGCCTGCTTGCTGCTGGCGAAGATCACCGGTAAACAAATCTATAAAGAAGCGATTGAGCGTCACTTGGAC
TACTGGAGCGTTGGTTACAATGGCGAGCGTGTGCACTATACGCCGAAAGGTCTGGCGTGGCTGGACAGCTGGGGTTCGCTGCGTTATGCCACGA
CCACTGCATTTCTGGCCTCCGTTTACGCGGACTGGGAAGGTTGCTCGCGTGAGAAAGCGGCGATTTACAACGATTTCGCGAAGCAGCAAATCGA
CTACGCTCTGGGTTCAAGCGGCCGTTCTTACGTTGTTGGTTTCGGCGTCAACCCGCCGAAGCGTCCGCATCACCGCACCGCGCATTCCAGCTGG
GCGGATTCTATGAGCGTGCCGGACTACCATCGCCATGTTCTGATCGGTGCTCTGGTCGGTGGTCCTGGCAAAGATGATAGCTATACCGACGACA
TCAATAACTATATCAATAACGAAGTCGCGTGTGATTACAACGCAGGCTTTGTCGGTGCACTGGCCAAAATGTATGAAGATTATGGTGGCAGCCC
GATTCCAGACCTGAACGCATTTGAAGAAATCACTAATGACGAGTTCTTTGTTATGGCGGGTATTAACGCTAGCGGTCAGAATTTCATCGAAATC
AAGGCCCTGCTGCACAATCAGAGCGGTTGGCCGGCACGTGTTGCAGACAAACTGAGCTTTCGTTACTTTGTTGATCTCACCGAGCTGATTGAGG
CAGGCTACAGCGCAAGCGATGTGACCATTACCACTAACTACAACGCGGGTGCGAAGGTTACGGGTCTGCACCCTTGGAATGAAGCCGAGAACAT
CTACTATGTCAACGTGGATTTCACGGGTACCAAGATTTACCCGGGTGGTCAATCTGCCTATCGCAAAGAAGTCCAGTTCCGCATCGCCGCACCG
CAAAACACGAACTTTTGGAATAACGACAATGACTACTCCTTCCGTGACATCAAAGGTGTGACGAGCGGCAATACGGTCAAAACCGTCTATATCC
CGGTGTATGACGATGGCGTTCTGGTCTTTGGTGTGGAGCCAGAGGGTGGCAGCGGTGAGAATAACAGCTCCATCAGCATTACCAATGCGACCTT
TGACAAGAATCCGGCGAAACAAGAGAATATTCAAGTGGTCATGAACCTGAACGGTAACACGTTGAACGGCATTAAGTATGGTAACACCTATCTG
CGTGAGGGTACCGACTATACGGTGAGCGGCGACACCGTGACGATCTTGAAGTCTTTTCTGAACAGCTTTGATACCAGCACCGTTCAACTGATTT
TTGATTTCAGCGCGGGTCGTGACCCGGTTCTGACCGTTAACATTATTGACACCACCACGAGCGCCTCCATCGTCCCGACTACTGCGGATTTCGA
TAAGAATCCTGATGCAAGCCGCGATGTCAAGGTGAAACTGGTTCCGAATGGCAACACGCTGCTGGCAGTGAAGAAAGACGGCGAGGCGCTGGTG
TTGGGCCGTGATTACAGCATCGATGGCGATGAAGTTACCATTTTCCGTGAGTATTTGGCGGACCAGCCAGTGGGTCGTGTTACCCTGACCTTCG
ACTTCGACCGTGGCACGGATCCGGTCCTGACCATCAACATCACCGACAGCCGCCAGGTTGAAACGGGTGTGATTCAAATTCAGATGTTTAATGG
TAATACGAGCGATAAAACCAATGGTATCATGCCGCGCTACCGTCTGACGAATACCGGTACCACCCCGATTCGCCTGTCCGATGTGAAGATCCGC
TATTACTACACTATTGACGGTGAAAAGGACCAGAATTTCTGGTGCGACTGGTCCTCAGTCGGCTCGAACAACATTACGGGCACCTTCGTAAAGA
TGGCGGAGCCGAAAGAGGGCGCAGACTATTACCTGGAGACTGGTTTTACGGACGGTGCGGGCTATCTGCAACCGAATCAAAGCATTGAAGTCCA
GAATCGCTTCAGCAAGGCCGATTGGACGGACTACATCCAAACGAATGACTACAGCTTCAGCACCAATACGAGCTATGGCAGCAACGACCGTATT
ACCGTTTACATTAGCGGTGTGTTGGTGAGCGGTATCGAGCCGTAATAAAAATGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGTGT
TTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAAATGGTATGCGTTTTGACACATCCACT
ATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACCAGACA
TTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGAAGCGCTCGTCGTATAACAGATGCG
ATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTGTCGGTCCTTGCACACGAATATTACGC
CATTTGCCTGCATATTCAAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACTTTTGGGCTTCTACCGATTTAGCAGTTTGATACA
CTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCGGCCAATCTGATTCCACCTGAGATGCATA
ATCTAGTAGAATCTCTTCGCTATAAAAATTCACTTCCACCTTCCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCCTCTGTTGACATG
ACACACATCATCTCAATATCCGAATAGGGCCCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCATATTTAT
CCAATATTCGTTCCTTAATTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGGTCCATTCACTATTCTCATTCCCTTTTCAGA
TAATTTTAGATTTGCTTTTCTAAATAAGAATATTTGGAGAGCACCGTTCTTATTCAGCTATTAATAACTCGTCTTCCTAAGCATCCTTCAATCC
TTTTAATAACAATTATAGCATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAAATAATTTTTCCGTTCCCAATTCCACA
TTGCAATAATAGAAAATCCATCTTCATCGGCAGATCCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCA

```
TCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAG
TAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGC
TACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCA
GACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGG
GCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGA
CCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGC
GTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATC
AAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGCAAGGCAAGACCGCCTAGGCCTGTTAGACGCCTTAAAACGAGTGTGTCA
CACCTCAACGCCCTTAATTCGTGCAACAGCCTGCAATAACGGCGCAAACATGTCTTCGCTGTCAAACGCTGGCTTGCGCTGGCTTGTCGGCTGG
CTGGCGCTTGTGTCCGCCGTGAAGTCGTACCAAATTAAGCCGGTCGCGGTCGGTTGATAGTCGGCTTGTTTGCGTTGTTGGGCGCGAGCTTTTT
CGGCTTTGACCTTCGCTTTTGCACGTTGCATTAGGTCGTCGAGTGTTTTGTTTATGTATGCGGATAGATATCGAATGCCGTAAGGATGACGCTT
CTGCACTTCGTTCACTTTAATAGTGACGTATTTTACAATTTTATCCAACGGATGTGTATTGTCTATTAATTTATTAATATCGTCATGTTTTTCA
ACTGGTGAACTAATGGTAGCGTCCGTTTTACGCGCTAAATCGCTTGATTCCGTAGGCTTTTCGGCGTTTTCTACTTTATCAGCTTCGTTGTCAC
GCAAGTTTAAAACGGCTGCAATACGCTCGAAATTTGCATGGCTCCGGAAAATGAATACAGGCGTTTTAGGTCCGTTACTATTCGGATTCTCACG
ATAGCAAACGACGACTTCGCCCGAGTCTTTGAGCATGCGAGTAGCACGATCAACAGTCGGCAAGGACACGCCTAACTTTTTGGCAAGCGTTTCG
CGCCCTGGGAACGAAAAACCGCGGTCTGTTGCGAGCCAAACAATTTCGTCGATGACGCGCGTTGTTTTTTGCGCTTTGCGGAACTTGCGGCCTA
ATGCGTCGGCGATACGCTCCATCAGCTCATTTTTAATCTTACGGCGTTTATAAACGAAATCTGTGACTCTCTCGATTTCTTTCAGTTGCGTGAA
TTGTTCGCTAGTGATTATGTACATTTCCGTTCTCCTCCCGTTTCTCCCGAATT
```

CelZ + His M13 fwd RC & pS797 CelZ

```
CelZ            AACCGAATCAAAGCATTGAAGTCCAGAATCGCTTCAGCAAGGCCGATTGGACGGACTACA 952
pS797           AACCGAATCAAAGCATTGAAGTCCAGAATCGCTTCAGCAAGGCCGATTGGACGGACTACA 4735
                ************************************************************

CelZ            TCCAAACGAATGACTACAGCTTCAGCACCAATACGAGCTATGGCAGCAACGACCGTATTA 1012
pS797           TCCAAACGAATGACTACAGCTTCAGCACCAATACGAGCTATGGCAGCAACGACCGTATTA 4795
                ************************************************************

CelZ            CCGTTTACATTAGCGGTGTGTTGGTGAGCGGTATCGAGCCGCATCATCATCNCCNNCNNN 1072
pS797           CCGTTTACATTAGCGGTGTGTTGGTGAGCGGTATCGAGCCGTAATA-------------- 4841
                ***************************************** *: *

CelZ            NNNNAAAANGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGNNTNNNNNNNNN 1132
pS797           ----AAAATGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGTGTTTTTTTGCT 4897
                    ****.***************************************..*....... .

CelZ            NNNNNNNNNNNNNNNNNNNNNNNNNNNNN------------------------------- 1161
pS797           ATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAAA 4957
                 .      .  . .. . ..  .  ..
```

>Cel6b M13 fwd RC

```
NNNNNNNNANNCNANGNNNNNNNNNNNNNNNGGNNNNNTNNANNNNTNNNNAANNNNNNNNTANNNNNNNNCNNTNNNNNNNTCANNNGGGNNTTNNNG
GANTACGAAACTNNGNATNGTCGCGNTNTNGAAATNGACAGCCNNCNGNACNNGNTTACCNANGTCGNNGGTAATGGCGGCNCCGAGCTGTGCGCAT
ATATGAAGCAGNNNGNNGGCTACGTTAATGGTGTGGGNTACGCTTTGCGCAAGCTGGGTGAAATTCCTAATGTTTACAACTATATTGACGCGGCACN
CCATGGTTGGATCGGTTGGGACAGCAATTTCGGCCCGAGCGTGGACATCTTTTATGAAGCGGCGAACGCNTCCGGCAGCACGGTTGACTATGTCCAC
GGTTTTATCAGCAACACCGCTAACTATTCAGCGACCGTTGAGCCGTACCTGGATGTTAACGGTACCGTCAATGGCCAACTGATCCGTCAAAGCAAAT
GGGTCGATTGGAATCAGTACGTGGACGAACTGTCCTTTGTCCAAGACTTGCGTCAGGCCCTGATTGCGAAAGGCTTCCGTAGCGATATCGGTATGCT
```

eGATTGATACCAGCCGCAACGGTTGGGGCGGTCCGAACCGTCCGACGGGTCCAAGCTCCTCCACGGACCTGAATACCTACGTTGACGAGAGCCGCAT
CGACCGTCGTATTCACCCGGGCAATTGGTGTAATCAGGCCGGTGCGGGTCTGGGTGAGCGTCCGACCGTGAACCCTGCCCCAGGCGTGGACGCGTAC
GTGTGGGTTAAACCACCGGGCGAGAGCGACGGTGCGTCTGAAGAGATTCCGAATGATGAAGGCAAGGGTTTTGACCGCATGTGTGACCCGACCTATC
AAGGCAACGCGCGCAACGGCAATAATCCGAGCGGTGCACTGCCGAATGCACCGATCAGCGGTCACTGGTTTAGCGCGCAGTTCCGCGAGCTGCTGGC
GAATGCCTATCCGCCGCTGCATCATCATCACCACCACTAATANNNNTGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGTGTTNNTGCTA
TNNNNNNNNNNNNNNNNNNN

>pS797 T. fusca Cel6B

AAGCTTTCGAGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA
ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCT
GCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATC
ATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTT
GCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
TCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGC
CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTCGAGCTGTCAATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT
GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACA
GCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGCTGCCTCGTCCATTTTTTT
GCTTAATGGAGGTTGTCATGAAAATGACAAACAACGTCCAAACAATTGCCATAATCGTTTACGCATAGTTTCGATTTCATCGCGTAAAATAATTTGT
GAATGTATTCACAATAATAAGAAGGGAGAATAGTGATGTCTAAGGTCCGTGCTACGAACCGTCGTTCCTGGATGCGTCGTGGTCTGGCTGCGGCTAG
CGGTCTGGCGCTGGGCGCCTCGATGGTAGCGTTCGCGGCACCGGCCAATGCAGCGGGCTGCAGCGTCGATTACACCGTTAATAGCTGGGGTACCGGT
TTCACCGCCAACGTGACCATCACCAACCTGGGTTCTGCAATCAATGGCTGGACGCTGGAGTGGGACTTCCCGGGTAATCAACAAGTTACCAACCTGT
GGAATGGCACCTATACCCAGAGCGGTCAACATGTTTCCGTTTCTAACGCTCCGTACAACGCCAGCATTCCGGCAAACGGTACGGTTGAATTTGGTTT
CAATGGTAGCTATAGCGGTAGCAATGATATTCCGAGCAGCTTCAAGTTGAACGGTGTGACTTGTGATGGCTCCGACGATCCAGACCCGGAGCCGAGC
CCGAGCCCGTCGCCGAGCCCAAGCCCGACGGACCCGGACGAGCCGGGTGGCCCGACGAACCCTCCGACTAACCCGGGTGAGAAAGTGGATAACCCGT
TTGAGGGTGCGAAGCTGTACGTCAATCCGGTGTGGAGCGCCAAAGCTGCAGCCGAGCCTGGCGGCTCGGCGGTCGCAAACGAATCTACCGCGGTGTG
GCTGGATCGCATTGGTGCGATTGAGGGCAACGATAGCCCGACGACGGGTAGCATGGGTCTGAGAGATCATCTGGAAGAAGCAGTGCGTCAGAGCGGT
GGCGACCCGCTGACTATTCAGGTTGTGATTTACAATTTACCGGGTCGTGATTGCGCGGCACTCGCAAGCAATGGCGAGCTGGGTCCGGATGAACTGG
ATCGCTACAAAAGCGAGTATATCGATCCGATTGCGGATATCATGTGGGATTTCGCGGACTACGAAAACTTGCGTATCGTCGCGATTATCGAAATCGA
CAGCCTGCCGAACTTGGTTACCAACGTCGGTGGTAATGGCGGCACCGAGCTGTGCGCATATATGAAGCAGAACGGTGGCTACGTTAATGGTGTGGGC
TACGCTTTGCGCAAGCTGGGTGAAATTCCTAATGTTTACAACTATATTGACGCGGCACACCATGGTTGGATCGGTTGGGACAGCAATTTCGGCCCGA
GCGTGGACATCTTTTATGAAGCGGCGAACGCCTCCGGCAGCACGGTTGACTATGTCCACGGTTTTATCAGCAACACCGCTAACTATTCAGCGACCGT
TGAGCCGTACCTGGATGTTAACGGTACCGTCAATGGCCAACTGATCCGTCAAAGCAAATGGGTCGATTGGAATCAGTACGTGGACGAACTGTCCTTT
GTCCAAGACTTGCGTCAGGCCCTGATTGCGAAAGGCTTCCGTAGCGATATCGGTATGCTGATTGATACCAGCCGCAACGGTTGGGGCGGTCCGAACC
GTCCGACGGGTCCAAGCTCCTCCACGGACCTGAATACCTACGTTGACGAGAGCCGCATCGACCGTCGTATTCACCCGGGCAATTGGTGTAATCAGGC
CGGTGCGGGTCTGGGTGAGCGTCCGACCGTGAACCCTGCCCCAGGCGTGGACGCGTACGTGTGGGTTAAACCACCGGGCGAGAGCGACGGTGCGTCT
GAAGAGATTCCGAATGATGAAGGCAAGGGTTTTGACCGCATGTGTGACCCGACCTATCAAGGCAACGCGCGCAACGGCAATAATCCGAGCGGTGCAC
TGCCGAATGCACCGATCAGCGGTCACTGGTTTAGCGCGCAGTTCCGCGAGCTGCTGGCGAATGCCTATCCGCCGCTGTAATAAAAATGGAAATGATT
CACACCCGATGTGGACTTGCCTCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTC
AAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTT

```
CTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGAA
GCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTGTCGGTC
CTTGCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACTTTTGGGCTTCTACCGA
TTTAGCAGTTTGATACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCGGCCAATCTGATTCC
ACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATAAAAATTCACTTCCACCTTCCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCCT
CTGTTGACATGACACACATCATCTCAATATCCGAATAGGGCCCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCC
ATATTTATCCAATATTCGTTCCTTAATTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGGTCCATTCACTATTCTCATTCCCTTT
TCAGATAATTTTAGATTTGCTTTTCTAAATAAGAATATTTGGAGAGCACCGTTCTTATTCAGCTATTAATAACTCGTCTTCCTAAGCATCCTTCAAT
CCTTTTAATAACAATTATAGCATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAAATAATTTTTCCGTTCCCAATTCCACAT
TGCAATAATAGAAAATCCATCTTCATCGGCAGATCCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCT
TTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCC
ACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGC
GTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAG
CGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGA
GCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTC
GGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGC
CATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCTGG
TGAAGTACATCACCGACGAGCAAGGCAAGACCGCCTAGGCCTGTTAGACGCCTTAAAACGAGTGTGTCACACCTCAACGCCCTTAATTCGTGCAACA
GCCTGCAATAACGGCGCAAACATGTCTTCGCTGTCAAACGCTGGCTTGCGCTGGCTTGTCGGCTGGCTGGCGCTTGTGTCCGCCGTGAAGTCGTACC
AAATTAAGCCGGTCGCGGTCGGTTGATAGTCGGCTTGTTTGCGTTGTTGGGCGCGAGCTTTTTCGGCTTTGACCTTCGCTTTTGCACGTTGCATTAG
GTCGTCGAGTGTTTTGTTTATGTATGCGGATAGATATCGAATGCCGTAAGGATGACGCTTCTGCACTTCGTTCACTTTAATAGTGACGTATTTTACA
ATTTTATCCAACGGATGTGTATTGTCTATTAATTTATTAATATCGTCATGTTTTTCAACTGGTGAACTAATGGTAGCGTCCGTTTTACGCGCTAAAT
CGCTTGATTCCGTAGGCTTTTCGGCGTTTTCTACTTTATCAGCTTCGTTGTCACGCAAGTTTAAAACGGCTGCAATACGCTCGAAATTTGCATGGCT
CCGGAAAATGAATACAGGCGTTTTAGGTCCGTTACTATTCGGATTCTCACGATAGCAAACGACGACTTCGCCCGAGTCTTTGAGCATGCGAGTAGCA
CGATCAACAGTCGGCAAGGACACGCCTAACTTTTTGGCAAGCGTTTCGCGCCCTGGGAACGAAAAACCGCGGTCTGTTGCGAGCCAAACAATTTCGT
CGATGACGCGCGTTGTTTTTTGCGCTTTGCGGAACTTGCGGCCTAATGCGTCGGCGATACGCTCCATCAGCTCATTTTTAATCTTACGGCGTTTATA
AACGAAATCTGTGACTCTCTCGATTTCTTTCAGTTGCGTGAATTGTTCGCTAGTGATTATGTACATTTCCGTTCTCCTCCCGTTTCTCCCGAATT
```

```
pS797           CGACGGTGCGTCTGAAGAGATTCCGAATGATGAAGGCAAGGGTTTTGACCGCATGTGTGA 3539
Cel6b           CGACGGTGCGTCTGAAGAGATTCCGAATGATGAAGGCAAGGGTTTTGACCGCATGTGTGA 861
                ************************************************************

pS797           CCCGACCTATCAAGGCAACGCGCGCAACGGCAATAATCCGAGCGGTGCACTGCCGAATGC 3599
Cel6b           CCCGACCTATCAAGGCAACGCGCGCAACGGCAATAATCCGAGCGGTGCACTGCCGAATGC 921
                ************************************************************

pS797           ACCGATCAGCGGTCACTGGTTTAGCGCGCAGTTCCGCGAGCTGCTGGCGAATGCCTATCC 3659
Cel6b           ACCGATCAGCGGTCACTGGTTTAGCGCGCAGTTCCGCGAGCTGCTGGCGAATGCCTATCC 981
                ************************************************************

pS797           GCCGCTGTAATAAAAATGGAAATGATTCACACCCGATGTGGACTTGCCTCATCGGGTGTT 3719
Cel6b           GCCGCTGCA---------------------------------------TCATC------- 995
                ******* *                                       *****

pS797           TTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGAT 3779
```

```
Cel6b            --------ATCACCACCACTAATA------------------------------------ 1011
                         ****  ** **** **

pS797            CCTCAAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCGTTCTGTCCACTCC 3839
Cel6b            ------------------------------------------------------------
```

Poor alignment but the His tag sequence can be clearly seen on the sequence data (Cel6b) and not on the original template (pS797).

```
> pS797 T. fusca Cel9A

AAGCTTTCGAGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA
ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTC
GCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCT
GCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATC
ATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTT
GCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
TCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGC
CCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTCGAGCTGTCAATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCT
GTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACA
GCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAG
CTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGCTGCCTCGTCCATTTTTTT
GCTTAATGGAGGTTGTCATGAAAATGACAAACAACGTCCAAACAATTGCCATAATCGTTTACGCATAGTTTCGATTTCATCGCGTAAAATAATTTGT
GAATGTATTCACAATAATAAGAAGGGAGAATAGTGATGAGCGTTACCGAGCCGCCTCCTAGACGCCGTGGTCGTCATAGCCGTGCGCGTCGCTTTCT
TACCAGCCTGGGTGCGACTGCGGCGCTCACGGCGGGTATGCTGGGCGTCCCGCTGGCGACCGGCACCGCCCACGCAGAGCCGGCATTTAACTATGCT
GAGGCTCTGCAGAAATCTATGTTTTTCTACGAAGCGCAGCGTTCCGGCAAGCTGCCGGAGAATAACCGTGTCAGCTGGCGTGGTGACAGCGGCTTGA
ATGACGGTGCCGATGTGGGTCTGGATCTGACGGGTGGTTGGTACGATGCGGGCGACCACGTGAAATTTGGTTTTCCGATGGCGTTCACGGCCACGAT
GCTGGCCTGGGGTGCAATTGAGAGCCCGGAAGGTTATATCCGCTCGGGTCAGATGCCGTATCTGAAGGATAACTTGCGTTGGGTGAATGACTACTTC
ATTAAAGCGCATCCGTCCCCAAACGTCCTGTACGTCCAGGTTGGTGATGGCGACGCGGATCACAAATGGTGGGGTCCGGCGGAAGTTATGCCGATGG
AACGTCCGTCTTTCAAGGTAGACCCGAGCTGTCCGGGCAGCGACGTTGCAGCCGAAACGGCGGCAGCCATGGCCGCTAGCAGCATCGTTTTCGCAGA
CGATGACCCGGCGTATGCGGCGACCCTGGTCCAACACGCCAAGCAACTGTACACCTTTGCAGATACCTACCGTGGCGTGTACAGCGACTGCGTGCCG
GCTGGTGCCTTCTATAACTCCTGGAGCGGCTATCAAGACGAGCTGGTGTGGGGCGCTTACTGGCTGTACAAAGCTACGGGCGACGACAGCTATCTGG
CAAAAGCGGAGTACGAGTACGATTTCCTGTCCACGGAGCAACAGACCGATCTGCGTTCTTACCGTTGGACCATCGCGTGGGACGACAAGAGCTACGG
TACGTACGTGCTGCTGGCGAAAGAAACCGGCAAGCAGAAGTACATCGACGACGCGAATCGCTGGCTGGATTACTGGACCGTTGGTGTTAATGGTCAG
CGCGTCCCGTACAGCCCGGGTGGCATGGCCGTTCTGGATACTTGGGGCGCATTGCGTTATGCAGCGAATACCGCTTTTGTGGCGTTGGTGTATGCAA
AAGTGATTGATGATCCGGTGCGCAAGCAGCGTTACCATGACTTCGCGGTCCGTCAAATCAACTACGCGCTGGGCGACAATCCGCGTAACTCTTCCTA
TGTCGTTGGTTTCGGTAACAATCCGCCACGTAATCCGCACCACCGCACGGCACACGGTAGCTGGACCGATAGCATCGCGTCACCGGCGGAGAATCGT
CACGTGTTGTATGGTGCACTGGTCGGTGGCCCAGGTAGCCCGAACGACGCCTATACTGATGACCGCCAAGACTACGTTGCAAATGAAGTCGCCACCG
ACTACAATGCGGGCTTCTCTAGCGCGCTGGCCATGCTGGTCGAGGAATATGGTGGCACGCCACTGGCAGATTTTCCGCCGACCGAGGAACCGGACGG
TCCGGAGATCTTTGTTGAGGCTCAGATTAACACGCCGGGTACGACGTTTACTGAGATTAAGGCAATGATCCGCAACCAGAGCGGTTGGCCTGCGCGC
ATGCTGGACAAAGGTACTTTCCGTTACTGGTTTACCCTGGACGAAGGCGTCGATCCGGCGGATATTACCGTGAGCAGCGCGTACAACCAATGCGCTA
```

CCCCGGAAGATGTTCATCATGTCAGCGGCGATCTGTACTATGTGGAAATTGACTGCACCGGCGAGAAAATCTTCCCGGGTGGTCAGAGCGAGCATCG
TCGCGAGGTCCAATTTCGTATCGCGGGTGGCCCTGGCTGGGATCCGAGCAATGACTGGAGCTTCCAGGGCATCGGTAACGAACTGGCACCGGCTCCG
TATATTGTGCTGTATGACGACGGTGTGCCGGTTTGGGGTACCGCGCCAGAGGAAGGCGAGGAACCTGGTGGCGGTGAGGGTCCGGGTGGTGGTGAAG
AACCGGGTGAAGATGTTACGCCGCCTAGCGCACCGGGCAGCCCGGCAGTGCGCGATGTGACCAGCACGAGCGCAGTACTGACGTGGAGCGCGAGCAG
CGATACCGGTGGCAGCGGCGTTGCAGGCTATGATGTTTTCTTGCGTGCGGGCACCGGCCAAGAGCAAAAGGTCGGTAGCACCACCCGTACTTCCTTC
ACCCTGACCGGTCTGGAGCCGGATACCACCTATATTGCCGCAGTTGTCGCTCGTGATAATGCGGGCAACGTCAGCCAGCGCTCTACGGTGTCGTTCA
CCACCCTGGCAGAGAACGGTGGCGGTCCGGATGCGAGCTGCACGGTCGGCTATAGCACCAATGACTGGGACAGCGGTTTTACCGCCAGCATCCGCAT
TACCTACCACGGTACGGCACCGCTGTCGTCCTGGGAACTGAGCTTTACCTTTCCAGCCGGTCAACAAGTTACGCACGGCTGGAATGCTACGTGGCGT
CAGGACGGTGCGGCCGTGACCGCGACGCCGATGAGCTGGAACAGCAGCCTGGCTCCGGGTGCGACCGTCGAGGTTGGCTTCAATGGTTCATGGAGCG
GTTCCAACACTCCGCCAACCGATTTTACCCTGAACGGTGAGCCGTGTGCCTTGGCGTAATAAAAATGGAAATGATTCACACCCGATGTGGACTTGCC
TCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAAATGGTATGCGTTTTGACA
CATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTTCTCAGAGTCGGAAAGTTGACC
AGACATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACCGAAGCGCTCGTCGTATAACAGATG
CGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTGTCGGTCCTTGCACACGAATATTACGCC
ATTTGCCTGCATATTCAAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACTTTTGGGCTTCTACCGATTTAGCAGTTTGATACACTTT
CTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCGGCCAATCTGATTCCACCTGAGATGCATAATCTAGT
AGAATCTCTTCGCTATAAAAATTCACTTCCACCTTCCACTCACCGGTTGTCCATTCATGGCTGAACTCTGCTTCCTCTGTTGACATGACACACATCA
TCTCAATATCCGAATAGGGCCCATCAGTCTGACGACCAAGAGAGCCATAAACACCAATAGCCTTAACATCATCCCCATATTTATCCAATATTCGTTC
CTTAATTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGGTCCATTCACTATTCTCATTCCCTTTTCAGATAATTTTAGATTTGCT
TTTCTAAATAAGAATATTTGGAGAGCACCGTTCTTATTCAGCTATTAATAACTCGTCTTCCTAAGCATCCTTCAATCCTTTTAATAACAATTATAGC
ATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAAATAATTTTTCCGTTCCCAATTCCACATTGCAATAATAGAAAATCCATC
TTCATCGGCAGATCCTTTTCCGCTGCATAACCCTGCTTCGGGGTCATTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGAT
TTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTT
CTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGG
ATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGG
CGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCG
CATCAATGGCGACCTGGGCCGCCTGGGCGGCCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCC
CTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAA
AACGGCCGGGGGGTGCGCGTGATTGCCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGC
AAGGCAAGACCGCCTAGGCCTGTTAGACGCCTTAAAACGAGTGTGTCACACCTCAACGCCCTTAATTCGTGCAACAGCCTGCAATAACGGCGCAAAC
ATGTCTTCGCTGTCAAACGCTGGCTTGCGCTGGCTTGTCGGCTGGCTGGCGCTTGTGTCCGCCGTGAAGTCGTACCAAATTAAGCCGGTCGCGGTCG
GTTGATAGTCGGCTTGTTTGCGTTGTTGGGCGCGAGCTTTTTCGGCTTTGACCTTCGCTTTTGCACGTTGCATTAGGTCGTCGAGTGTTTTGTTTAT
GTATGCGGATAGATATCGAATGCCGTAAGGATGACGCTTCTGCACTTCGTTCACTTTAATAGTGACGTATTTTACAATTTTATCCAACGGATGTGTA
TTGTCTATTAATTTATTAATATCGTCATGTTTTTCAACTGGTGAACTAATGGTAGCGTCCGTTTTACGCGCTAAATCGCTTGATTCCGTAGGCTTTT
CGGCGTTTTCTACTTTATCAGCTTCGTTGTCACGCAAGTTTAAAACGGCTGCAATACGCTCGAAATTTGCATGGCTCCGGAAAATGAATACAGGCGT
TTTAGGTCCGTTACTATTCGGATTCTCACGATAGCAAACGACGACTTCGCCCGAGTCTTTGAGCATGCGAGTAGCACGATCAACAGTCGGCAAGGAC
ACGCCTAACTTTTTGGCAAGCGTTTCGCGCCCTGGGAACGAAAAACCGCGGTCTGTTGCGAGCCAAACAATTTCGTCGATGACGCGCGTTGTTTTTT
GCGCTTTGCGGAACTTGCGGCCTAATGCGTCGGCGATACGCTCCATCAGCTCATTTTTAATCTTACGGCGTTTATAAACGAAATCTGTGACTCTCTC
GATTTCTTTCAGTTGCGTGAATTGTTCGCTAGTGATTATGTACATTTCCGTTCTCCTCCCGTTTCTCCCGAATT

```
pS797        GGCCGTGACCGCGACGCCGATGAGCTGGAACAGCAGCCTGGCTCCGGGTGCGACCGTCGA 4436
Cel9A        GGCCGTGACCGCGACGCCGATGAGCTGGAACAGCAGCCTGGCTCCGGGTGCGACCGTCGA 950
             ************************************************************

pS797        GGTTGGCTTCAATGGTTCATGGAGCGGTTCCAACACTCCGCCAACCGATTTTACCCTGAA 4496
```

```
Cel9A           GGTTGGCTTCAATGGTTCATGGAGCGGTTCCAACACTCCGCCAACCGATTTTACCCTGAA 1010
                ************************************************************

pS797           CGGTGAGCCGTGTGCCTTGGCG------------------TAATAAAAATGGAAATGATT 4538
Cel9A           CGGTGAGCCGTGTGCCTTGGCGCATCATCATCACCACCACTAANNNAAATGGAAATGATT 1070
                **********************                  ***   **************

pS797           CACACCCGATGTGGACTTGCCTCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCG 4598
Cel9A           CACACCCGATGTGGACTTGCCTCATCGGGTGTTNNNNNN-TNTCAAAANNNNNNNNNNNN 1129
                *********************************       * ******

pS797           GCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAAATGGTATGCGTTTTGACACA 4658
Cel9A           NNNN-------------------------------------------------------- 1133


pS797           TCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGC 4718
Cel9A           ------------------------------------------------------------
```

```
>pS797 T. neopolitana CelB
AAGCTTTCGAGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCA
ATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTG
AGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGA
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTT
TGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCC
TGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGAT
AAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGAT
CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTCGAGCTGTCAATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTT
CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC
GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACAGGAAAC
AGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGCTGCCTCGTCCATTTTTTTGCTTAATGGAGGTTGTCATGAAAATGACAAACAACGTCCA
AACAATTGCCATAATCGTTTACGCATAGTTTCGATTTCATCGCGTAAAATAATTTGTGAATGTATTCACAATAATAAGAAGGGAGAATAGTGAT
GCGTCTGGTTGTCTCCTTCCTGCTGGTCGTTAGCGCGTTTTTGTTTAGCGCAGAGGTCGTGCTGACCGACATTGGTGCGACGGACATCACGTTC
AAGGGCTTCCCGGTCACCATGGAACTGAATTTCTGGAATGTGAAATCCTATGAAGGCGAAACCTGGCTGAAGTTCGATGGTGAGAAAGTTCAGT
TTTACGCCGACATCTACAATATCGTCCTGCAGAACCCAGACTCTTGGGTTCACGGCTATCCGGAGATTTACTACGGCTATAAGCCGTGGGCTGC
GCATAACAGCGGTACTGAGATTCTGCCGGTCAAGGTTAAAGATCTGCCGGACTTTTATGTTACGCTGGACTATAGCATTTGGTACGAAAATGAT
CTGCCTATTAACCTGGCAATGGAAACCTGGATTACCCGCAAGCCGGATCAAACTTCGGTGAGCAGCGGTGATGTTGAGATCATGGTGTGGTTTT
ACAATAACATTCTGATGCCGGGTGGTCAAAAAGTGGATGAGTTCACCACCACCATCGAGATCAACGGTAGCCCGGTGGAAACCAAGTGGGACGT
GTACTTCGCCCCGTGGGGTTGGGACTATTTGGCGTTCCGTCTGACGACCCCGATGAAAGACGGCCGTGTCAAGTTCAACGTAAAAGACTTTGTT
GAAAAAGCTGCGGAAGTCATCAAAAAGCACAGCACGCGCGTTGAGAATTTCGATGAAATGTACTTTTGCGTGTGGGAGATTGGTACCGAGTTCG
GCGATCCGAACACCACGGCAGCGAAGTTCGGTTGGACGTTTAAAGATTTTAGCGTGGAGATCGGCGAGTAATAAAAATGGAAATGATTCACACC
```

CGATGTGGACTTGCCTCATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCGTTTTACAAGATCCTCAAA

ATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCGTTCTGTCCACTCCTGAATCCCATTCCAGAAATTCTCTAGCGATTCCAGAAGTTT

CTCAGAGTCGGAAAGTTGACCAGACATTACGAACTGGCACAGATGGTCATAACCTGAAGGAAGATCTGATTGCTTAACTGCTTCAGTTAAGACC

GAAGCGCTCGTCGTATAACAGATGCGATGATGCAGACCAATCAACATGGCACCTGCCATTGCTACCTGTACAGTCAAGGATGGTAGAAATGTTG

TCGGTCCTTGCACACGAATATTACGCCATTTGCCTGCATATTCAAACAGCTCTTCTACGATAAGGGCACAAATCGCATCGTGGAACTTTTGGGC

TTCTACCGATTTAGCAGTTTGATACACTTTCTCTAAGTATCCACCTGAATCATAAATCGGCAAAATAGAGAAAAATTGACCATGTGTAAGCGGC

CAATCTGATTCCACCTGAGATGCATAATCTAGTAGAATCTCTTCGCTATAAAAATTCACTTCCACCTTCCACTCACCGGTTGTCCATTCATGGC

TGAACTCTGCTTCCTCTGTTGACATGACACACATCATCTCAATATCCGAATAGGGCCCATCAGTCTGACGACCAAGAGAGCCATAAACACCAAT

AGCCTTAACATCATCCCCATATTTATCCAATATTCGTTCCTTAATTTCATGAACAATCTTCATTCTTTCTTCTCTAGTCATTATTATTGGTCCA

TTCACTATTCTCATTCCCTTTTCAGATAATTTTAGATTTGCTTTTCTAAATAAGAATATTTGGAGAGCACCGTTCTTATTCAGCTATTAATAAC

TCGTCTTCCTAAGCATCCTTCAATCCTTTTAATAACAATTATAGCATCTAATCTTCAACAAACTGGCCCGTTTGTTGAACTACTCTTTAATAAA

ATAATTTTTCCGTTCCCAATTCCACATTGCAATAATAGAAAATCCATCTTCATCGGCAGATCCTTTTCCGCTGCATAACCCTGCTTCGGGGTCA

TTATAGCGATTTTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTATCCAACG

GCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGG

AATCCTGCTCTGCGAGGCTGGCCGGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGGGCAGCC

CACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCGGCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGC

CGTCGGCCAGGGCTACAAAATCACGGGCGTCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGGC

CTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCGCCCTGCTGGCGAAGATCGAAGAGAAGCAGG

ACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCGCCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTG

CCAAGCACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGAGCAAGGCAAGACCGCCTAGGCCTG

TTAGACGCCTTAAAACGAGTGTGTCACACCTCAACGCCCTTAATTCGTGCAACAGCCTGCAATAACGGCGCAAACATGTCTTCGCTGTCAAACG

CTGGCTTGCGCTGGCTTGTCGGCTGGCTGGCGCTTGTGTCCGCCGTGAAGTCGTACCAAATTAAGCCGGTCGCGGTCGGTTGATAGTCGGCTTG

TTTGCGTTGTTGGGCGCGAGCTTTTTCGGCTTTGACCTTCGCTTTTGCACGTTGCATTAGGTCGTCGAGTGTTTTGTTTATGTATGCGGATAGA

TATCGAATGCCGTAAGGATGACGCTTCTGCACTTCGTTCACTTTAATAGTGACGTATTTTACAATTTTATCCAACGGATGTGTATTGTCTATTA

ATTTATTAATATCGTCATGTTTTTCAACTGGTGAACTAATGGTAGCGTCCGTTTTACGCGCTAAATCGCTTGATTCCGTAGGCTTTTCGGCGTT

TTCTACTTTATCAGCTTCGTTGTCACGCAAGTTTAAAACGGCTGCAATACGCTCGAAATTTGCATGGCTCCGGAAAATGAATACAGGCGTTTTA

GGTCCGTTACTATTCGGATTCTCACGATAGCAAACGACGACTTCGCCCGAGTCTTTGAGCATGCGAGTAGCACGATCAACAGTCGGCAAGGACA

CGCCTAACTTTTTGGCAAGCGTTTCGCGCCCTGGGAACGAAAAACCGCGGTCTGTTGCGAGCCAAACAATTTCGTCGATGACGCGCGTTGTTTT

TTGCGCTTTGCGGAACTTGCGGCCTAATGCGTCGGCGATACGCTCCATCAGCTCATTTTTAATCTTACGGCGTTTATAAACGAAATCTGTGACT

CTCTCGATTTCTTTCAGTTGCGTGAATTGTTCGCTAGTGATTATGTACATTTCCGTTCTCCTCCCGTTTCTCCCGAATT

>CelB

NNNTNNNNNCNATNNNANNTNNNNNCNNNTNNNNNNNNTGCNTNNNCNTTTTTTTNNNANGNNTNTCANNAAANGNCNANNNANNNCNAANCNNN

NCCATAATNGTTTACGCATAGTTTNGATTTCATCGCGNAAANTAATTTGNGAANGTATTCNCAATANTANGANGGGAGANTAGTGATGCGTCTG

NNNTCTCCTTCCTGCTGGTCGTTAGCGCGTTTTTGTTTAGCGCAGAGGTCGTGCTGACCGACATTGGTGCGACGGACATCACGTTCAAGGGCTT

CCCGGTCACCATGGAACTGAATTTCTGGAATGTGAAATCCTATGAAGGCGAAACCTGGCTGAAGTTCGATGGTGAGAAAGTTCAGTTTTACGCC

GACATCTACAATATCGTCCTGCAGAACCCAGACTCTTGGGTTCACGGCTATCCGGAGATTTACTACGGCTATAAGCCGTGGGCTGCGCATAACA

GCGGTACTGAGATTCTGCCGGTCAAGGTTAAAGATCTGCCGGACTTTTATGTTACGCTGGACTATAGCATTTGGTACGAAAATGATCTGCCTAT

TAACCTGGCAATGGAAACCTGGATTACCCGCAAGCCGGATCAAACTTCGGTGAGCAGCGGTGATGTTGAGATCATGGTGTGGTTTTACAATAAC

ATTCTGATGCCGGGTGGTCAAAAAGTGGATGAGTTCACCACCACCATCGAGATCAACGGTAGCCCGGTGGAAACCAAGTGGGACGTGTACTTCG

CCCCGTGGGGTTGGGACTATTTGGCGTTCCGTCTGACGACCCCGATGAAAGACGGCCGTGTCAAGTTCAACGTAAAAGACTTTGTTGAAAAAGC

TGCGGAAGTCATCAAAAAGCACAGCACGCGCGTTGAGAATTTCGATGAAATGTACTTTTGCGTGTGGGAGATTGGTACCGAGTTCGGCGATCCG

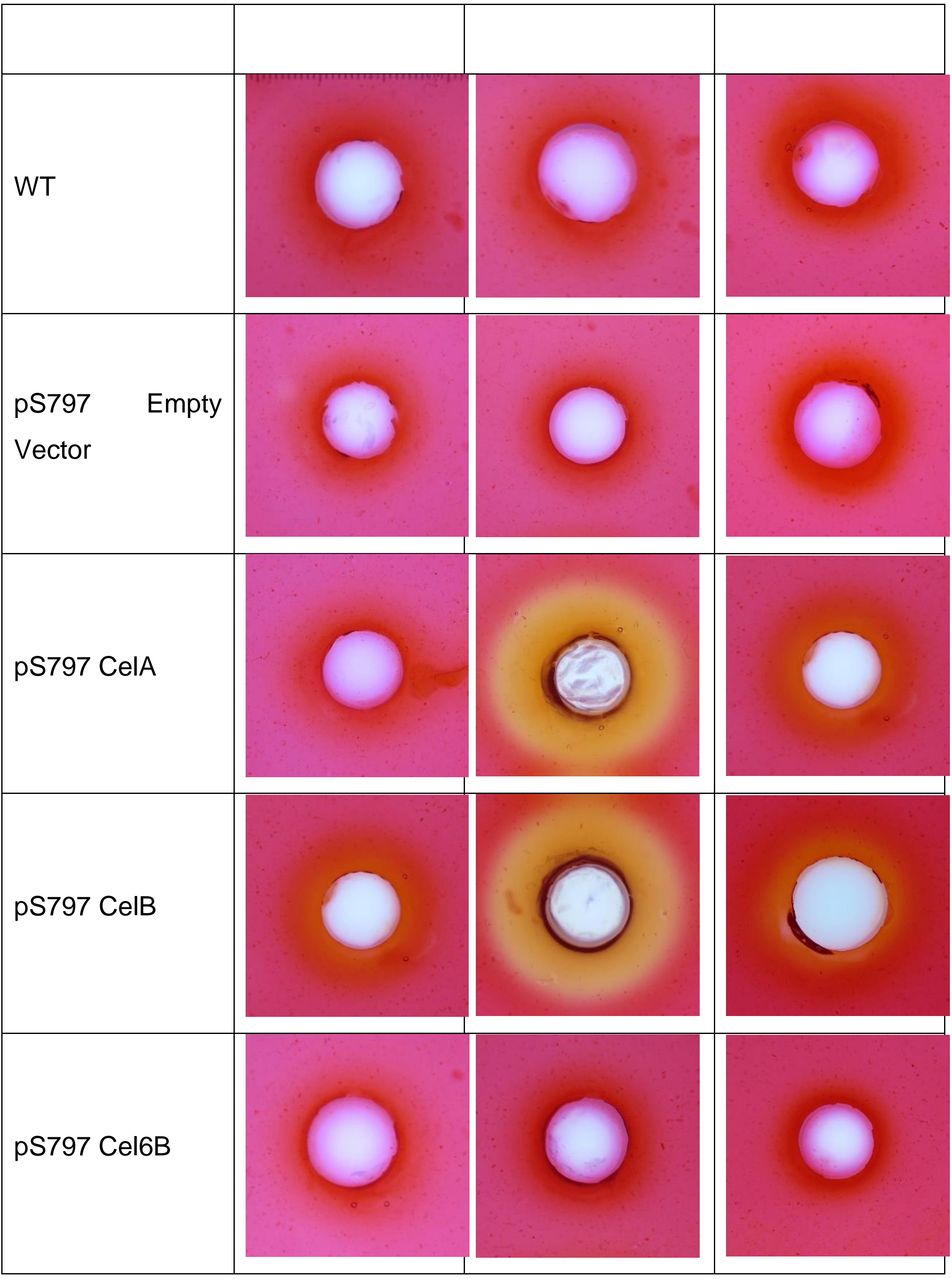
AACACCACGGCAGCGAAGTTCGGTTGGACGTTTAAAGATTTTAGCGTGGAGATCGGCGAGCATCATCATCACCACCACTAATAAAAATGGAAAT

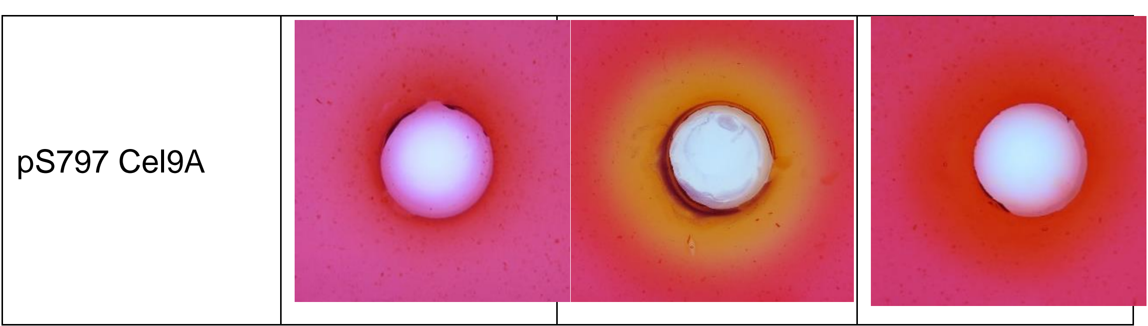
GATTCACACCCGATGTGGACTTGCCTCATCGGGTGTTNNTGCTATCAANNNTNNNNNNNNNNNNN

```
pS797          AACGTAAAAGACTTTGTTGAAAAAGCTGCGGAAGTCATCAAAAAGCACAGCACGCGCGTT 2580
CelB           AACGTAAAAGACTTTGTTGAAAAAGCTGCGGAAGTCATCAAAAAGCACAGCACGCGCGTT 880
               ************************************************************
```

	Control	Supernatant	Cell lysate

```
pS797
GAGAATT
TCGATGA
AATGTAC
TTTTGCG
TGTGGGA
GATTGGT
ACCGAGT
TCGGCGA
TCCG
2640
CelB
GAGAATT
TCGATGA
AATGTAC
TTTTGCG
TGTGGGA
GATTGGT
ACCGAGT
TCGGCGA
TCCG
940

*******
*******
*******
*******
*******
*************************

pS797       AACACCACGGCAGCGAAGTTCGGTTGGACGTTTAAAGATTTTAGCGTGGAGATCGGCGAG 2700
CelB        AACACCACGGCAGCGAAGTTCGGTTGGACGTTTAAAGATTTTAGCGTGGAGATCGGCGAG 1000
            ************************************************************

pS797       ------------------TAATAAAAATGGAAATGATTCACACCCGATGTGGACTTGCCT 2742
CelB        CATCATCATCACCACCACTAATAAAAATGGAAATGATTCACACCCGATGTGGACTTGCCT 1060
                              ******************************************

pS797       CATCGGGTGTTTTTTTGCTATCAAAACTACTAGTAGCGGCCGCCTGCAGACTGGCCGTCG 2802
CelB        CATCGGGTGTTNNT--GCTATCAANNNTNNNNNNNNNNNNNN------------------ 1099
            ***********  *  ********   *

pS797       TTTTACAAGATCCTCAAAATGGTATGCGTTTTGACACATCCACTATATATCCGTGTCGTT 2862
CelB        ------------------------------------------------------------
```


WT
pS797 Empty Vector
pS797 CelA
pS797 CelB
pS797 Cel6B

pS797 Cel9A			

Appendix 4

Table showing Congo red staining of CMC agar following incubation with *Geobacillus thermoglucosidans* DSM2541, pS797, pS797 *CelA*, pS797 *CelB*, pS797 *Cel6B*, pS797 *Cel9A* supernatant for 24 hours at 55°C*.*