#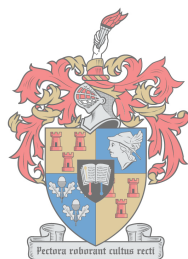 Identification and biochemical characterisation of an aryl-β-glucosidase isolated from a cellulose-acetate rich environment *via* a functional metagenomic approach

Janto Pieters

*Thesis presented in fulfilment of the requirements for the degree of Master of Science in the Faculty of Natural Sciences at Stellenbosch University.*

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918·2018

Supervisor: Prof. Jens Kossmann

Co-supervisors: Dr. Shaun Peters & Dr. Bianke Loedolff

December 2018

DECLARATION: By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof, that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Janto Pieters

December 2018

# Abstract

Functional metagenomics has established itself as an indispensable tool in the search for novel genes by accessing the genetic information of previously unculturable microbes. These novel genes are commonly identified by assessing their function through qualitative observations of their general biochemical activities, prior to nucleic acid sequencing. In this study, a fosmid library was created from a cellulose-acetate rich environment (cigarette waste bin) and functionally screened for cellulose degrading and deacetylation enzymes. Here we report on the identification of a novel aryl-β-glucosidase identified from this library. Following plate based functional screening, one putative β-glucosidase was identified (*BG4*). Next generation sequencing of the 40 Kb fosmid insert identified an open reading frame (ORF) which, contained two distinct glycosyl hydrolase family one domains. The *BG4* coding ORF was isolated, cloned into the pSF-OXB20-NH2-10HIS-EKT bacterial expression vector and heterologously expressed in *E. coli*. While BG4 showed selective hydrolytic activity to β-1,4-glucosidic bonds it displays a natural substrate specificity only to aryl-β-glucosides including arbutin, esculin hydrate, gensitin and salicin. It was most active on esculin hydrate (Km= 5.15 mM, Vmax= 0.28 μmol glucose.min$^{-1}$) and showed a temperature and pH optimum of 40 °C and pH 6. The BG4 activity was stimulated by the presence of manganese chloride and ethyl acetate and inhibited by detergents (SDS, glycerol and Triton X-100) and glucose concentrations exceeding 1.5 mM. The specificity and high activity of BG4 towards aryl-β-glucosides could make it applicable in medical industries to release the biological potent aglycone moiety for glycosylated precursors.

# Opsomming

Funksionele metagenomika is 'n onontbreeklike tegniek in die soektog na onbekende gene omdat dit die gentiese inligting van onkweekbare mikrobes beskikbaar stel. Die gene word geïdentifiseer deur kwalitatiewe obserwasie van funksie voordat DNS volgorde beplaing gedoen word. In die studie is 'n funksionele metagenomiese biblioteek is gemaak vanaf DNS geïsoleer uit 'n sellulose-asetaat ryk omgewing (sigarette stompie asdrom). Dit is daarna geëvalueer vir die teenwoordigheid van sellulose degraderende en deasetilerende gene. Hierdeur is 'n voorlopige β-glukosidase (*BG4*) geïdentifiseer. Volgorde bepaling van *BG4* fosmid het 'n moontlike β-glukosidase oopleesraam geïdentifiseer. Die oopleesraam is daarna gekloneer in pSF-OXB20-NH2-10HIS-EKT bakterïele proteïn uitdrukkings vektor wat daarna in *E. coli* getransformeer is om BG4 te produseer. BG4 het selektiewe hidrolitiese aktiwiteit getoon op β-1,4-glukosiede verbindings en was selektief aktief op ariel-β-glukosied substrate wat arbutin, esculin hidraat, gensitin en salicin insluit. BG4 was mees aktief op esculin hidraat (Km= 5.15 mM, Vmax= 0.28 µmol glucose.min$^{-1}$) en het optimale aktiwiteit getoon by pH 6 en 40 °C. BG4 akiwiteit was gesimuleer deur mangaan chloried en etielasetaat en sterk geïnhibeer deur SDS en glukose bo 1.5 mM. Die hoë aktiwiteit en spesifisiteit van BG4 teenoor ariel-β-glukosiede verleen dit aan moonltike toepassing vir die vrystelling van biologiese aktiewe molekules vanaf hulle onaktiewe, geglukosiseerde voorlopers.

# Acknowledgements

I would like to extend my most sincere gratitude to the following people and institutions

Professor Jens Kossmann, my supervisor and director of the Institute for Plant Biotechnology, for giving me the opportunity to complete my BSc Hons and MSc research at the institute.

The NRF for providing funding.

Drs. James Lloyd, Paul Hills and Christel van der Vyver for their input during lab meetings and all further assistance.

All the staff and students at the Institute for their assistance, friendship and making the last three years truly memorable.

A special thanks to the members of my research group for support during this endeavour.

To my family for never-ending emotional support, financial assistance and motivation during my post-graduate studies.

To Anja Baxter for unconditional love, support and understanding during my post-graduate studies.

Dr. Shaun Peters for the vital assistance given during my whole project but especially the biochemical sections. And for always encouraging my scientific question and critical thinking.

Finally Dr. Bianke Loedolff, my co-supervisor, for all the assistance, patience and support that a post-graduate student can hope for. Also for the countless hours spent on improving not only my work but my mind.

Without all of you none of this would have been possible. I am truly thankful.

# Scientific contributions during Masters Candidature (2017 - 2018)

1. Conference - The South African Academy for Science and Arts symposium, 2-3 November 2017 (University of Pretoria).

    Oral Presentation: Title "Using functional metagenomics for the identification of novel genes".

    Pieters J, Kossmann J, Loedolff B, Peters S

    Award received: Overall 3$^{rd}$ place in group, and winner of Blue Stallion prize.

2. Abstract publication - Pieters J, Loedolff B, Kossmann J (2018). Using functional metagenomics for the identification of novel genes. Die Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie.

3. Pieters J, Kossmann J, Loedolff B, Peters S (2018). Identification and biochemical characterisation of an aryl β-glucosidase isolated from a cigarette filter waste environment. *Manuscript in preparation*. For submission to Frontiers in Microbiology (IF 4)

# Table of content

# Abbreviations

AA – Amino acid

BG – β-Glucosidase

BLAST – Basic Local Alignment Search Tool

BSA – Bovine Serum Albumin

CMC – Carboxymethyl cellulose

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

GH – Glycoside hydrolase

Kb – Kilobase

NCBI – National Center for Biotechnology Information

ORF – Open Reading Frame

PCR – Polymerase chain reaction

pNP – para-Nitrophenyl

SDS – Sodium dodecyl sulfate

U– Enzyme unit

# Background

Cigarette filter pollution is considered one of the most common human-made waste problems globally, owing to an estimated 1.1 billion smokers (2015) who engage in smoking as a recreational activity. Together they manage to dispose an estimated 10 billion cigarettes daily in the environment (World Health Organization, 2017). This pollution was evident subsequent to the International Coastal Clean-up report, reporting over a million cigarette filters collected on United State beaches in 2017 alone  (Belhouari et al., 2017). Consequently, government laws have been instituted to ban the public use of cigarettes (on a global scale) with the hope of decreasing the pollution status and the current health risks associated with smoking (Goodman et al., 2007; Juster et al., 2007; Öberg et al., 2011). From a metagenomic (refer to section 3.1) perspective, pollution such as cigarette filters (ashtrays) creates a potential environment for the discovery of novel enzymes with industrial/medical relevance.

Cigarette filters are produced by the esterification of cellulose (with acetic anhydride), usually obtained from wood pulp (Harris, 2011; Figure 3). The resulting cellulose acetate is spun to produce the fibres from which cigarette filters are produced (Harris, 2011; Novotny et al., 2009; refer to Background, Figure 1). These cellulose acetate filters are photodegradable (can be chemically decomposed by the action of sunlight) and biodegradable (can be decomposed by biological activity of living organisms). However, the degradation period is variable and largely dependent on the environment and the bacterial communities residing within these environments (ranging from < one month to > five years; Bonanomi et al., 2015; Buchanan et al., 1993). The first step to effective degradation is the removal of acetyl moieties which enables the systematic breakdown of the cellulosic backbone of the filter (refer to section 1.1.2, Figure 2).  A triad of enzymes is required for the latter process to occur synergistically, namely endogluconases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21; de Souza, 2013). All of these enzymes belong to the family of glycosyl hydrolases (EC 3.2.1) which are active on the glycosidic bond between two carbohydrates or a carbohydrate and non-carbohydrate moiety (Davies and Henrissat, 1995). β-Glucosidases catalyse the hydrolysis of cellobiose to glucose which is the final and vital step towards complete degradation of cellulose.
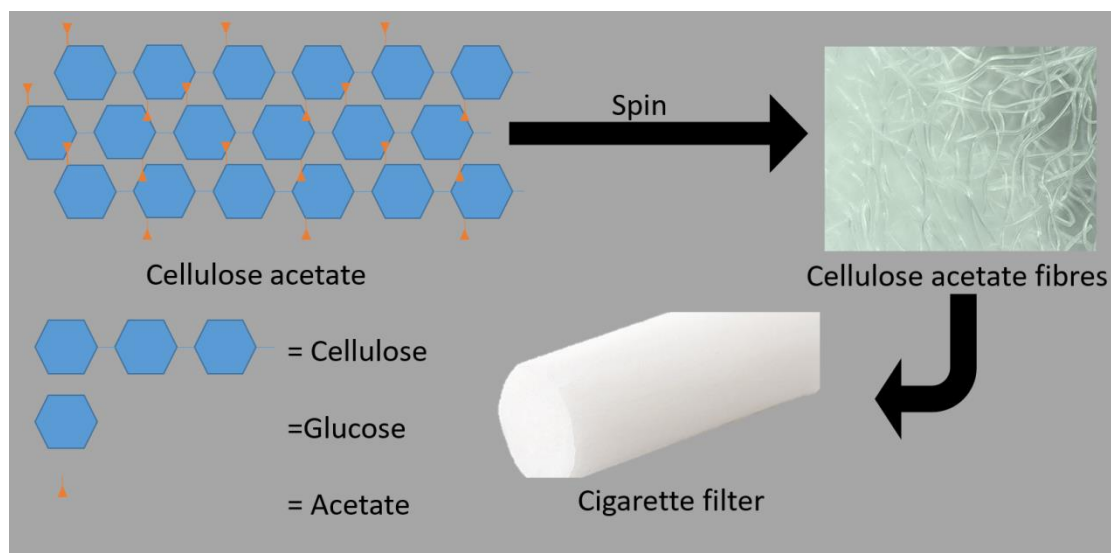
Figure 1. Schematic representation of the components of cellulose acetate cigarette filters. Acetylated cellulose chains are spun into cellulose acetate fibres which are used for cigarette filter production.

The β-glucosidases are extremely versatile enzymes which display a myriad of substrate specificities. Their general function may be categorised by their ability to hydrolyse the β-glycosidic bond present between two carbohydrate molecules or between a carbohydrate and non-carbohydrate molecule (Davies and Henrissat, 1995). This characteristic has caused further classification of β-glucosidases according to specific substrate specificity where they are classified as (i) true β-glucosidases, which hydrolyse cellobiose to produce glucose; (ii) aryl-β-glucosidases, which are only active on aryl-glucosides or (iii) broad specificity β-glucosidases, which show activity on a broad spectrum of substrates (Singhania et al., 2013). Generally, β-glucosidases are classified into group iii with the ability to hydrolyse a wide range of substrates with varying specificity to respective substrates. Aryl-β-glucosidases are the least common microbial β-glucosidases described in literature and have been proposed to contribute to plant decay as aryl-β-glucosides are commonly found in the leaves, bark and flowers of many plants (Marques et al., 2003). Industrially, cellobiose degrading β-glucosidases have enjoyed the most attention due to the critical role performed during bioethanol production (Singhania et al., 2013). Aryl-β-glucosidases also play important industrial roles such as flavonoid- and isoflavonoid glucoside hydrolysis in the food and drink industries and release biologically potent of aglycone moieties in the medical industry (Bhatia et al., 2002). ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,

## 1.1 A complex interplay between enzymes are required for cellulosic material breakdown

### 1.1.1 Classification of glycosyl hydrolyses

Glycosyl hydrolases are ubiquitous enzymes which catalyse the hydrolysis of glycosidic bonds (Davies and Henrissat, 1995; Naumoff, 2011). These bonds exist either between two carbohydrates or a carbohydrate and a non-carbohydrate moiety (Henrissat et al., 1995). The glycosyl hydrolase superfamily consists of numerous enzymes which show high levels of substrate diversity between enzyme activities (Bourne and Henrissat, 2001). This substrate diversity is not surprising when the extensive stereochemical diversity of carbohydrates, their substrates, are considered. The different possible isomers for a reducing hexasaccharide is $>10^{12}$ which is a clear indicator of the potential carbohydrate structural heterogeneity (Laine, 1994). This does not include possible non-carbohydrate moieties which greatly increases the possible complexity. The categorisation of the glycosyl hydrolase superfamily has been attempted numerous times to create a clear classification system. Traditionally, the classification was done exclusively on the substrate specificity but this method is complicated as glycosyl hydrolases (i) have similar substrates but different catalytic mechanisms, (ii) show a broad substrate specificity, (iii) have transglucosylation activity and (iv) could be either exo- or endo-type enzymes (Naumoff, 2011). The current method of classification used relies on similarities between the amino acid sequences of glycosyl hydrolases and have been proposed in the early 1990's by Henrissat (1991). This method was adopted as the amino acid sequence (i) reflect structural features, (ii) reveals the evolutionary relationship and (iii) provides a tool to derive mechanistic information (Singhania et al., 2013). The different families classified according to this methods are continually increasing with 45 families in 1995 (Henrissat et al., 1995), 80 in 2000 (Vasella et al., 2002) and 153 currently (September 2018; Lombard et al., 2014; http://www.cazy.org/Glycoside-Hydrolases.html). The increase is mainly due to the exponential increase in sequencing data availability from known and unknown microbial species. Glycosyl hydrolases are also routinely further classified into clans, a group of families which are suspected to have shared ancestry, of which there is currently 17 in the CAZy database (Lombard et al., 2014; http://www.cazy.org/Glycoside-Hydrolases.html). The classification is done by comparing similarities in (i) tertiary structure, (ii) catalytic residues and (iii) catalytic mechanism (Henrissat and Bairoch, 1996). As more three-dimensional structures are elucidated the clan grouping improve and currently more of the 153 families are classified into clans. The catalytic mechanism of glycosyl

hydrolases have been well studied and there are currently two known mechanisms of action first described by Koshland (1953). The first is inverting glycoside hydrolases and the second retaining glycoside hydrolases which refers to the inversion or retention of the stereochemistry of the substrate (Davies and Henrissat, 1995).

## 1.1.2 Enzymatic degradation of cellulosic materials

Among the glycosyl hydrolase superfamily, there are several families of enzymes which are required for the complete breakdown of cellulose (refer to section 1.1.2, Figure 2).  Cellulose is a linear glucose polysaccharide joined by β-1,4-glycosidic linkages which enables it to be tightly packed to form insoluble crystalline microfibrils (Payne et al., 2015). This occurs due to van der Waals interactions and hydrogen bonds between cellulose chains (Guerriero et al., 2016). The result is a polymer which is recalcitrant to enzymatic degradation especially when found in nature where cellulose forms part of plant cell walls together with hemi-cellulose and lignin (Cragg et al., 2015). Organisms have evolved multiple complementary enzymes, especially glycosyl hydrolases, to degrade this naturally abundant carbon source (Cragg et al., 2015). The complete breakdown of cellulose polymers can be achieved by the synergistic effect of three glycosyl hydrolase family enzymes (refer to section 1.1.2, Figure 2). The endogluconases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) each of which play a vital role in cellulose polymer hydrolysis (Horn et al., 2012).

Firstly, endogluconases cleaves the cellulose polysaccharide chains randomly. This creates reducing and non-reducing ends and releases shorter water soluble carbohydrates (de Souza, 2013; Doerner and White, 1990). Cellobiohydrolases attack reducing and non-reducing ends of the cellulose polymer releasing mainly cellobiose (Guerriero et al., 2016; Horn et al., 2012). In the last step, β-glucosidases hydrolyse the cellobiose produced to glucose (Jeng et al., 2011; Sun and Cheng, 2002; Figure 2).
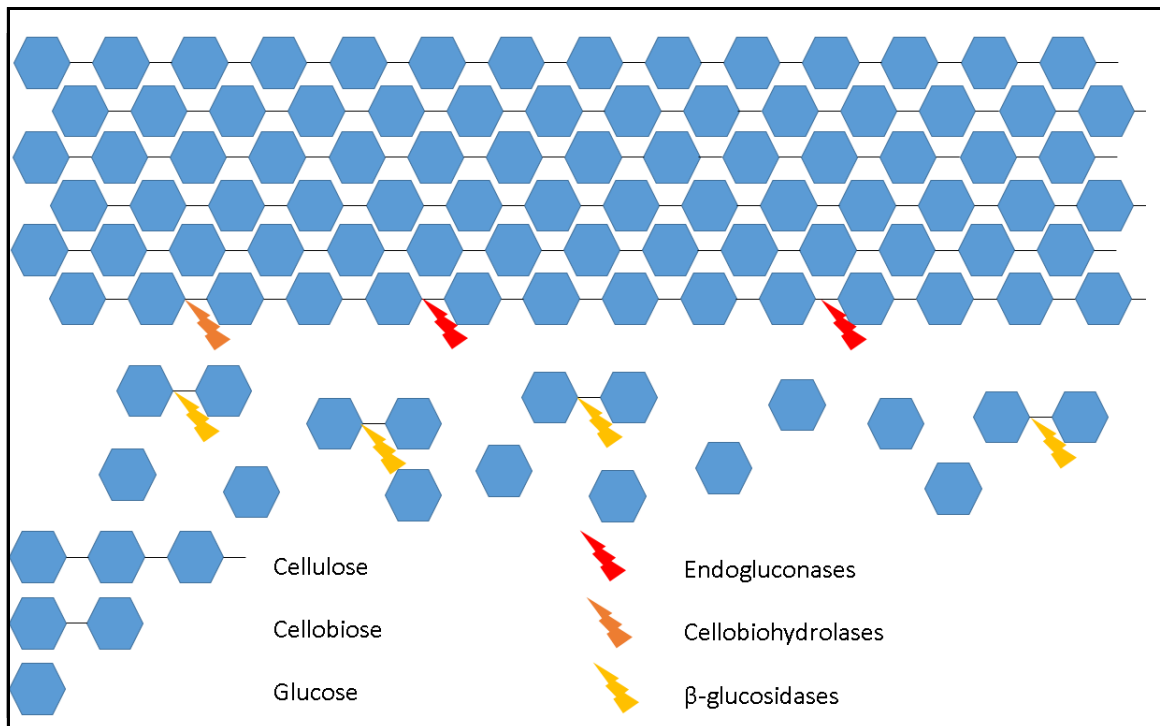
Figure 2. Schematic representation of the synergistic degradation of cellulose by cellulases. Endogluconases (EC 3.2.1.4) hydrolyse internal cellulose bonds, cellobiohydrolases (EC 3.2.1.91) cleaves cellobiose from the cellulose chains and finally cellobiose is hydrolysed by β-glucosidases (EC 3.2.1.21) to liberate glucose monomers.

Microbial cellulose degrading enzymes facilitate the degradation of lignocellulosic biomass which is the most abundant polymer on earth (Isikgor and Becer, 2015). The exact chemical composition of lignocellulose in plants differ but generally consist of 50% cellulose, 40% hemicellulose and 10% lignin (Isikgor and Becer, 2015). Microbes have evolved over millions of years to produce enzymes which effectively degrade plant cell walls to utilise this abundant carbon source. The enzymes can be directly translated into other applications such as cigarette filter bioremediation, bioethanol production or the release of potent aglycone from precursors. Cigarette filters' constituents are similar to the cellulosic components in lignocellulose with the exception of added acetyl moieties.
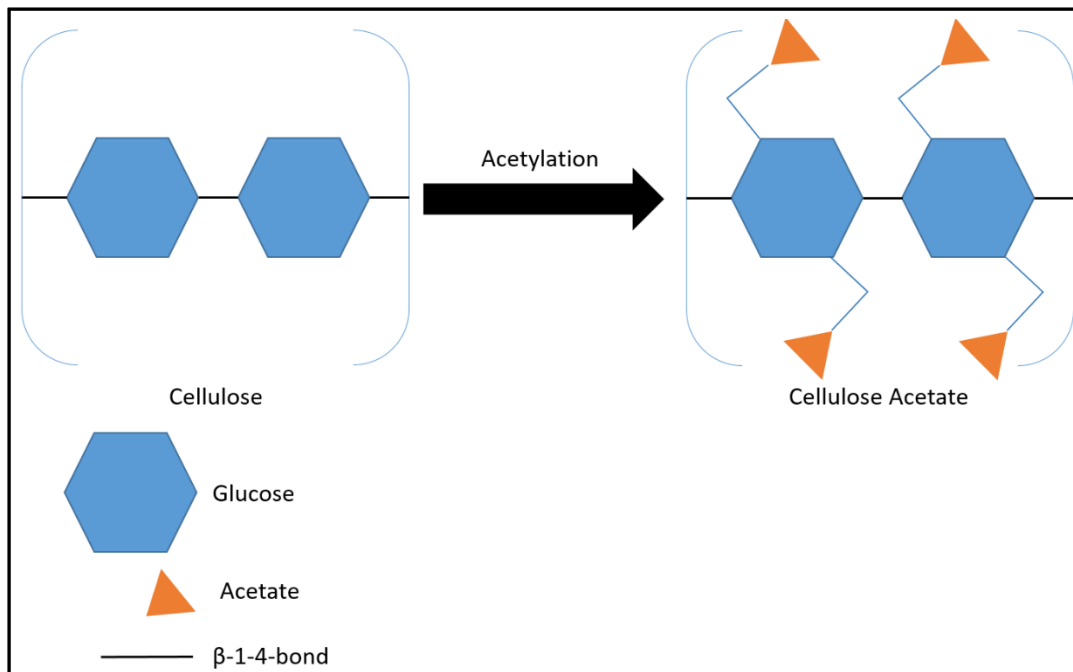
Figure 3: Schematic representation of cellulose and cellulose acetate. Cellulose comprises of glucose units bound by a β-1,4-linkage and cellulose acetate is similar but has acetyl moieties added on the glucose backbone.

### 1.1.3 Classification of β-glucosidases

Glycosyl hydrolases consist of 153 families of specialised carbohydrate hydrolysing enzymes (Lombard et al., 2014; http://www.cazy.org/Glycoside-Hydrolases.html.) One class of glycosyl hydrolases, the β-glucosidases, catalyses the hydrolysis of β-glycosidic bonds occurring between (i) two carbohydrates such as short chain oligosaccharides and disaccharides or (ii) a carbohydrate and a non-carbohydrate moiety such as aryl-, amino-, alkyl- β-D-glucosides (Singhania et al., 2017). Under specific conditions, β-glucosidases can also catalyse the synthesis of a glycosidic bond between different molecules via two modes, reverse hydrolysis and transglycosylation (Singhania et al., 2013). β-Glucosidases are classified according to sequence similarities into glycosyl hydrolase families GH1, GH3, GH5, GH9, GH30 and GH116 with the majority belonging to GH1 and GH3 (Michlmayr and Kneifel, 2014). Glycosyl hydrolases are further alphabetically classified into clans which suggests shared ancestry. GH1, GH5 and GH30 and consist of protein with (β/α) 8-barrel structures and belongs to GH-A, GH116 has (α/α) 6-barrel structure and belongs to GH-O, GH9 also has a (α/α) 6-barrel but is not assigned to a clan and GH3 has two catalytic domains and not assigned to a clan (Lombard et al., 2014).

## 1.2 Industrial application of β-glucosidases

The biotechnological application of β-glucosidases can be classified according either hydrolytic or synthetic applications (Bhatia et al., 2002). The hydrolytic industrial applications of β-glucosidases have been of great interest in industries including biofuel, food and feed and medical. Synthetic application of β-glucosidases, although less studied, has potential in mainly the medicinal industry.

### 1.2.1 Role of β-glucosidases in cellulose degradation: a necessity for biofuels

The current major focus of β-glucosidases in research is their role in second generation biofuel production (Singh et al., 2016). Second generation biofuels require the effective hydrolysis of cellulose present in plant biomass (Horn et al., 2012). This is achieved by treating the biomass with cellulase enzyme mixtures consisting of endogluconases, cellobiohydrolases and β-glucosidases (de Souza, 2013). β-glucosidases are responsible for cellobiose hydrolysis, the final step in efficient cellulose degradation  (de Souza, 2013). β-glucosidases' role is vital as cellobiose degradation (i) relieves the product feedback inhibition of cellobiose on endogluconases and cellobiohydrolases and (ii) produces glucose for fermentation by yeast to produce bioethanol (Aditiya et al., 2016). The lack of adequate β-glucosidase activity has been reported for various commercial enzyme mixtures which are usually produced from fungi (Sun and Cheng, 2002).

β-Glucosidases require specific characteristics to be applicable to the biofuel industry. As β-glucosidases are generally inhibited by glucose ,its product at very low concentrations, glucose tolerance has become one of the major characteristics required for industrial application (Uchiyama et al., 2015). β-Glucosidases have been reported with this unusual trait of exhibiting glucose tolerance and some are even induced by glucose (Yang et al., 2015). The tolerant β-glucosidases mostly form part of GH1 family with several reported in recent years (Fang et al., 2010; Uchiyama et al., 2013). Generally, GH3 β-glucosidases do not show significant glucose tolerance but some have been reported (Li et al., 2014). Even though glucose tolerance has received the most attention the harsh industrial conditions of bioethanol production have also made other characteristics such as temperature stability, pH stability and ethanol tolerance desirable (Singhania et al., 2013). Current commercially available β-glucosidases are not suited for use in these conditions which is why more β-glucosidases are continuously searched for (Elleuche et al., 2014). Various β-glucosidases have

been reported with some of these commercially important traits (Uchiyama et al., 2013; Schröder et al., 2014; Maruthamuthu and Van Elsas, 2017).

## 1.2.2 Food, feed and agricultural applications

The majority of application of β-glucosidases in the food industry is related to the flavour enhancing capabilities of wine, beer and fruit juice. β-glucosidases play a key role in the release of aromatic molecules usually not active due to glycosylation (Ahmed et al., 2017; Bhatia et al., 2002; Kuhad et al., 2011; Singh et al., 2016). Highly efficient microbial β-glucosidases are routinely added as the natural process by plant β-glucosidases can be very time-consuming (Singh et al., 2016) or have limited activity at the industrial production conditions (Ahmed et al., 2017). β-Glucosidases are also used to reduce the bitterness in fruit juices by hydrolysis of naringenin (Fan et al., 2011). β-Glucosidases applied towards flavour enhancement has to be active at a pH of 2.5 to 4, have the ability to hydrolyse glycosylated aromatic precursors and show resilience towards secondary plant metabolites (Fischer and Noble, 1994; Sheehan et al., 2007).  In some food applications high activity on specific β-glucosidic bond is the main attribute required. In the gellan industry β-glucosidases with the ability to hydrolyse glycosyl-rhamnosyl-glucose to produce glucose and rhamnosyl-glucose is required to reduce viscosity (Bhatia et al., 2002). Pyridoxine glucoside hydrolysis by β-glucosidases are required to release vitamin B6 for increased nutritional value (Opassiri et al., 2004). Finally in the soybean cooked syrup industry β-glucosidases that can hydrolyse daidzin and genistin to reduce bitterness is required (Xu and Song, 2014).

In the agricultural sector, β-glucosidases have been suggested to be used in Cassava root detoxification (Gueguen et al., 1997). Cassava root has a cyanogenic glycoside present in the roots which, when consumed, is harmful to human health (Coursey, 1973). Plant protection can also be assisted by β-glucosidases as *Rhizobacteria* produce various biological control agents. Among these are cellulases, which assist in the protection of the plant against pathogenic oomycetes (Menendez et al., 2015). β-Glucosidases are also added to cellulose based feed of single-stomach animals, such as pigs and chickens, where it leads to better degradation and effectively better nutrition (Bhatia et al., 2002).

### 1.2.3 Medical application of β-glucosidases: the release of potent bioactive compounds

In the medical industry, biologically potent molecules can be released from phenolic compounds such as flavonoids, flavanone, flavones and isoflavones through β-glucosidase activity (Ahmed et al., 2017). Soybean isoflavonoids have also been associated with protection from cancer, heart disease and osteoporosis and with the release of aglycone moieties by β-glucosidases the potency was improved (Omoni and Aluko, 2005). The aglycone released from the glycosylated precursor has higher biological activity and is also absorbed faster and in higher amounts (Singh et al., 2016). The release of aglycone moiety from other sources are also of interest such as oleuropein (olives) and amygdalin (peach kernels) which are also used in the medical industry as alternative anticancer agents (Xu and Song, 2014).

### 1.2.4 Other applications

In the paper and pulp industry, the de-inking of paper is done to reduce wood consumption and reduce paper waste. The de-inking can be achieved by chemical and enzymatic methods. The enzymatic method uses β-glucosidases among other enzymes (Ahmed et al., 2017; Menendez et al., 2015). Another proposed application was the release of the biologically active coumarin aglycone from coumarin (Mercer et al., 2013). Coumarins are fragrant molecules present in plants and have medical applications such as antifungal properties (Montagnera et al., 2008). Under defined conditions, some β-glucosidases show transglycosylation activity *in vitro* (Dan et al., 2000). Transglycosylation enables the synthesis of oligosaccharides, aryl-, alkyl- and alcohol-glucosidase with high regio- and stereo-selectivity and without the need for external energy input (Ahmed et al., 2017). Oligosaccharides can be used as therapeutic agents such as heparin or acarbose, the production of carbohydrate based chemicals and as a probiotic agent (Ahmed et al., 2017). Alkyl-glucosides are applied as bio-degradable, non-ionic surfactants. Aryl- and alcohol-glucosides has potential in medicine development, for example, methyl-glucoside serves as a precursor for methyl-laminario oligosaccharides which is applied as AIDS therapeutic agent. Aryl-glucosides have also been shown to have repellent and anti-feed properties or can be applied in the production of natural flavourants in the food industry (Bhatia et al., 2002).

## 1.3 Metagenomics: the search for novel environmental enzymes

Enzymes are natural biocatalysts which significantly increases the rate of biochemical reactions (Beilen and Li, 2002). Biocatalysts are industrially applied toward a plethora of applications and are frequently favoured over chemical reactions as it is a greener alternative without high energy and harsh chemical requirements (Madhavan et al., 2017). Biocatalysts are frequently sourced from microorganisms for specific industrial application as they have evolved to adapt and thrive in every imaginable environment. This immense physiological and functional heterogeneity is now a major source of genetic diversity on earth (Li et al., 2009). Traditionally, when bioprospecting for novel industrially relevant genes microorganisms were cultivated and subsequently screened for the desired phenotype. However, currently more than 99% of all microorganisms cannot be cultured successfully (Xing et al., 2012). This discrepancy in the amount microorganisms present and the amount which is culturable has come to be known as the great plate anomaly  (Madhavan et al., 2017). The great plate anomaly leads to vast amounts of microbial genetic resources not being accessible for bioprospecting. This was addressed by the discovery of metagenomics which made all genetic material available for testing, effectively bypassing the need for any microbial culturing (Madhavan et al., 2017). Metagenomics uses genetic material acquired directly from an environmental source for analysis which can be done in two possible ways: sequence based or functional based analysis (Madhavan et al., 2017; Xing et al., 2012).

### 1.3.1 Sequence based metagenomics *versus* functional based metagenomics

Metagenomics emerged in the early 1990s and established itself as a fundamental technique in the quest to novel enzyme discovery (Escobar-Zepeda et al., 2015). Sequence and functional based metagenomics both start with the isolation of all genomic information within any environment imaginable. The environmental sample source is however vital for bioprospecting as it is used to infer what target genes might be present. The environmental sample can then be sequenced directly using modern next-generation sequencing methods or be used for metagenomic library construction. Direct sequencing methods are typically applied for microbial community analysis and not for biomolecule identification (Madhavan et al., 2017). For bioprospecting purposes, environmental DNA is used to create a metagenomic library, which is vectors housing the isolated DNA. Different vectors systems can be used to create different insert size libraries such as plasmids

(15 KB), cosmids and fosmids (40 KB) or bacterial artificial chromosomes (40 KB). Once the library has been constructed the two systems (sequence- and functional based screening) greatly differ in terms of analysis (Figure 4).
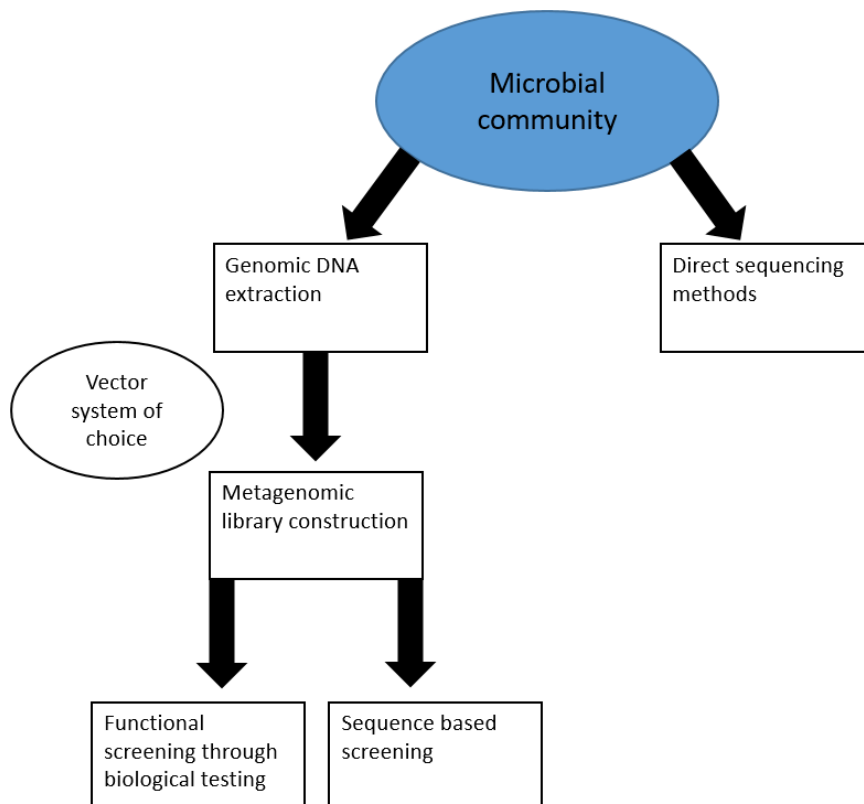


Figure 4: Diagram of strategies employed for metagenomic analysis of environmental community. Adapted from Madhavan et al. (2017). Diagram indicates the steps followed in order to create a functional or sequence based screening library.

Sequence based metagenomic screening involves the direct screening of sequences. There are several direct screening methods which have been applied in gene discovery such as gene specific PCR primers (Henckel et al., 1999), stable isotope probes (Dumont et al., 2006) and integron specific primers (Cowan et al., 2005). These methods are limited as knowledge regarding conserved regions in the gene sequences are required and thus excludes the identification of new families of novel enzymes. The functional based screening of a metagenomic library searches for biological activity prior to any genetic sequence analysis. Once the library is created in the vector system of choice it

11

is screened for biological activity in one of two ways: direct detection of activity or functional complementation. The direct detection employs the use of chemical dyes and insoluble or chromophore-containing derivatives in the growth media. They serve as the substrate for the target enzyme and results in a visual aid for activity (Xing et al., 2012). One screening method used for β-glucosidases identification is when esculin hydrate and ferric citrate is included in growth media which results in a brown halo (Li et al., 2012a). Complementation screening requires a host strain which is deficient of the target gene and thus only clones where the target gene is introduced will be able to grow. The identification of a novel lysine racemase has been achieved through this methods using a lysine deficient host strain (Chen et al., 2009). The major drawback for functional based screening is that in order to identify a gene it has to be effectively expressed and translated into a functional protein by the host system. *Escherichia coli* is generally used for this application and has proved to express 40% of environmental DNA (Gabor et al., 2004). Different expression systems such as *Streptomyces* and *Pseudomonas* can also be applied to increase potential genes expressed (Xing et al., 2012). The ability of functional screening to identify novel genes and gene families resulted in it being the preferred method for novel enzyme bioprospecting.

## 1.3.2 Functional based metagenomics: bioprospecting for industrially relevant β-glucosidases

Bioprospecting for cellulose degrading enzymes is of great industrial importance for industries including biofuel, bioremediation, food, laundry and textiles, medical and paper and pulp (Duan and Feng, 2010). Cellulose consists of a linear β-1,4-linked glucose polymer and requires the synergistic effect of endogluconases, cellobiohydrolases and β-glucosidases for complete hydrolysis (de Souza, 2013). Originally bioprospecting for these classes of enzymes was limited to known natural cellulose degrading organism such as wood degrading fungi (Béguin, 1990). Recent advances in functional metagenomics have contributed to cellulase enzyme identification especially for enzymes with desirable industrial characteristics such as activity under a wide range pH, temperature and ionic conditions (Ilmberger and Streit, 2017). β-Glucosidases play a vital role in cellulose degradation (refer to section 1.2.1) and have been the subject of intense bioprospecting. Recently, β-glucosidases have been identified with industrially important characteristics such as thermo-stability (Fusco et al., 2018; Jabbour et al., 2012; Liu et al., 2012; Park et al., 2005), pH stability (Martin et al., 2014; Maruthamuthu and Van Elsas, 2017), glucose tolerance (Crespim et al., 2016;

Jabbour et al., 2012; Sathe et al., 2017; Uchiyama et al., 2015) an ethanol tolerance (Biver et al., 2014; Karnaouri et al., 2013). β-Glucosidases also have applications in other industries (sections 1.2.3 – 1.2.5) besides cellobiose hydrolysis. Aryl-β-glucosidase specifically have been reported with industrially potential activity such as aromatic improvement of wines/juices (Swangkeaw et al., 2011) and the release of isoflavones from soybean (Mase et al., 2004; Otieno and Shah, 2007).

The use of metagenomics is clearly a powerful tool for the bioprospecting of novel enzymes with industrially relevant activities. The aim of my project was to use a functional metagenomic approach to identify enzymatic activities, specifically focusing on β-glucosidase activity, from a cigarette filter waste (cellulose acetate rich) environment. The objectives were to create and functionally screen a metagenomic library, to sequence and identify putative β-glucosidases, to recombinantly express the identified β-glucosidases and to biochemically characterise the recombinant enzyme.

# Materials and Methods

## 2.1 *Escherichia coli* genotypes and plasmids

The following *E. coli* strains and plasmids (Table 1) were used in this study:

DH5α (Invitrogen, ThermoFischer Scientific, South Africa): F– Φ80*lac*ZΔM15 Δ (*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1

EPI300™-T1R *E. coli* (Epicentre, Whitehead Scientific, South Africa): F– mcrA Δ (mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ– rpsL nupG trfA tonA dhfr

Table 1: Summarised list of plasmids used, the companies from which acquired, the antibiotic selection and the application of the plasmid.

| Plasmid | Company origin | Selection | Use |
|---------|----------------|-----------|-----|
| CopyControl™ pCC2FOS™ | Epicentre | Chloramphenicol | Vector for production of copy controlled fosmid library |
| pSF-OXB20-NH2-10His-EKT | Oxford Genetic Ltd | Kanamycin | Contains constitutive RecA promoter for high expression levels |

## 2.2 Metagenomic library construction

A metagenomic library was constructed from an environment rich in cellulose acetate (cigarette filter waste). The total environmental genomic DNA was extracted immediately after sampling using the Meta-G-Nome™ DNA isolation kit (Epicentre®, Whitehead Scientific, South Africa) according to the manufacturer's instructions. No further processing was required. The isolated DNA fragments (blunt ended, 40 Kb in length) were ligated into the CopyControl ™ pCC2FOS™ vector (Epicentre®) to construct a fosmid metagenomic library. The ligated DNA was subsequently packaged using the MaxPlax™ lambda packaging extract (Epicentre®) and cloned into the *E. coli* Epi300-T1$^R$ plating strain, according to manufacturer's instructions  (EpiCentre's CopyControl™ HTP Fosmid Library Production Kit with pCC2FOS™ Vector, Epicentre®).The phage titre was determined at $1.2 \times 10^8$ which confirms efficient packaging ($> 10^7$ required according to manufactures specifications). After transduction number of clones was calculated according to manufacturer's instructions, resulting in 840 cones. Fosmid clones were maintained as singly copy and induced to high copy number prior to

subsequent experimentation, using proprietary CopyControl fosmid autoinduction solution (2% v/v, Epicentre®).

## 2.3 Functional plate based screening of library

The fosmid library was screened for β-glucosidase, cellulase and esterase activities using standard plate based screens, as previously described (Li et al., 2012b; Singh et al., 2015; Yusof et al., 1989). Briefly, β-glucosidase activity screens were conducted on Luria Bertani agar plates (LB agar; 1% w/v peptone, 1% w/v sodium chloride, 0.5% w/v yeast extract, 1.5% w/v bacteriological agar) supplemented with ferric citrate (0.05% w/v; Sigma Cat# F3388, Sigma-Aldrich, South Africa), esculin (0.01% w/v; esculin hydrate, Sigma Cat# E8250, Sigma-Aldrich, South Africa). Cellulase activity screens were conducted on CMC-salt plates (0.188% w/v CMC sodium salt, Sigma Cat# C4888, Sigma-Aldrich, South Africa; 0.05% w/v $K_2HPO_4$, 0.025% w/v $MgSO_4.7H_2O$, 0.2% w/v gelatin and 0.5% w/v bacteriological agar) supplemented with Congo red (0.02% w/v Congo Red, Sigma Cat# C6767, Sigma-Aldrich, South Africa ). Finally, esterase activity screens were conducted on tributyrin agar plates (20% w/v tributyrin agar; Sigma Cat# 91015, Sigma-Aldrich, South Africa) supplemented with tributyrin (10% v/v tributyrin, Sigma Cat# 73105, Sigma-Aldrich, South Africa) and victoria blue (0.004% w/v; Victoria Blue B, Sigma Cat# V0753, Sigma-Aldrich, South Africa). All of the fosmid based screening plates included the appropriate antibiotic selection (chloramphenicol 12.5 µg.ml$^{-1}$) and autoinduction solution (2% v/v). Positive clones were identified through a dark brown halo for β-glucosidase-, a yellow halo for cellulase- and a blue, cleared halo for esterase-activity.

## 2.4 Sequencing and metagenomics data assembly

The fosmids identified as positive for the respective functional screens were isolated via FosmidMAX DNA Purification Kit according to manufacturer's specifications (Epicentre®, Whitehead Scientific, South Africa). DNA quality and quantity was determined with a NanoDrop™ spectrophotometer (ND-LITE, ThermoFisher Scientific). Fosmid DNA (500 ng.µL$^{-1}$) was submitted to Inqaba biotec™ (South Africa) and paired end reads were produced using the Illumina Miseq platform. Sequence data was analysed using the CLC Genomics Workbench v7 (Qiagen, Whitehead Scientific, South Africa). The sequence data was assessed for quality and trimmed according to CLC parameters (quality score 0.05, maximum ambiguous nucleotides of two). The trimmed data was used for *de*

*novo* contig assembly according to CLC parameters (word size 45, bubble size 95 and minimum contig length 1000). The contigs were assembled to create consensus sequence from which putative open reading frames were identified using NCBI's ORF finder software.

## 2.5 Sequence analysis

A single fosmid clone displaying β-glucosidase activity in functional screen was chosen to be analysed further for the scope of this project. Following sequencing of the 40 Kb fosmid insert, ORFs were identified using NCBI's ORF finder software (https://www.ncbi.nlm.nih.gov/orffinder/). Putative functional annotation of all the ORFs were determined through nBLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) and a single ORF was identified as a putative β-glucosidase. This nucleotide sequence was translated using the Expasy translate tool (Gasteiger et al., 2003; https://web.expasy.org/translate/; using standard genetic code). The predicted amino acid sequence was then analysed with NCBI domain finder which identified one putative β-glucosidase domain (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), that was then assigned the BG4 nomenclature used to describe the gene in this work. ExPASy PROSITE was used to further identify protein motifs and function was assigned (de Castro et al., 2006; https://prosite.expasy.org/). The protein sequence of the putative β-glucosidase identified was used for pBLAST analysis. The glycosyl hydrolases domain's top six closest related accessions was selected for protein alignment analysis. BioEdit software (BioEdit 7.2.3) was used to align the putative β-glucosidase identified to the top six accessions identified using the default parameters (Hall, 1999).

## 2.6 Restriction digest cloning

The putative β-glucosidase ORF (*BG4*) was selected and used for subsequent cloning and expression. Primers were designed to amplify the ORF, adding a 5' *Kpn*I restriction site in the forward primer sequence (5'- ATA<u>GGTACC</u>GATGACCGATAAACAGAAGAAATTAC- 3') and a 3' *Nhe*I restriction site in the reverse primer sequence (5'- ATA<u>GCTAGC</u>TTATACATCACGTAGATGCTTTAAAAC- 3'). The primer pair was used to amplify the BG4 ORF from the fosmid in a high fidelity PCR reaction (Q5 DNA polymerase, New England Biolabs, Inqaba Biotech, South Africa) under the following conditions (denaturation at 98 °C for 10 sec, primer annealing at 61 °C for 30 sec, elongation at 72 °C for 30

sec; 25 cycles). Reaction products were analysed by agarose gel electrophoresis (1% w/v gel) supplemented with Pronosafe for visualization (0.005 % v/v). The amplicon (1.26 KB) was excised from the gel and purified using a DNA clean-up kit, according to manufacturer's instructions (Wizard® SV Gel and PCR Clean-Up, Promega, Anatech, South Africa). The purified fragment was subjected to a restriction digest using high fidelity versions of *Kpn*I and *Nhe*I (1 µg DNA, 10 U restriction enzymes, 10X CutSmart buffer and ddH$_2$O to a final volume of 50 µl; 37 °C, 1 h). The expression vector pSF (pSF-OXB20-NH2-10His-EKT) was prepared for ligation, using the same restriction conditions. The *BG4* ORF was ligated into the expression vector using T4 DNA ligation (50 ng vector DNA, 75 ng insert DNA, 400 U T4 DNA ligase, 10X ligation buffer and ddH$_2$O to a final volume of 20 µL; 25 °C, 2 h) creating the construct pSF::*BG4*. The ligation reaction was transformed into competent *E. coli* (DH5α, Invitrogen) and the presence of the *BG4* was confirmed via PCR amplification and restriction enzyme digest analysis, as described above.

## 2.7 Heterologous protein expression and quantification

DH5α cells containing either pSF or pSF::*BG4* were inoculated, respectively, into 2 ml of LB supplemented with kanamycin (50 µg.ml$^{-1}$) and allowed to grow overnight (200 rpm, 37 °C). The overnight culture was added to 200 ml of LB with kanamycin (50 µg.ml$^{-1}$) and grown to 0.6 OD$_{595}$ (200 rpm, 37 °C). The cells were centrifuged (3830 g, 4 °C, 15 min) and suspended in 2 ml extraction buffer (pH 7; 50 mM HEPES, 100 mM NaCl, 0.5 mM Phenylmethylsulfonyl fluoride, 0.05% v/v β-mercaptoethanol, 5% v/v glycerol). Lysozyme (1 mg.ml$^{-1}$) was added and incubated on ice (rocking, 30 min). The cells were sonicated (3 times, 5 s, 10 s interval) and centrifuged (10000 g, 4 °C, 15 min). The clarified supernatant was used as crude protein extracts for biochemical analysis. Protein concentrations were determined according to the standard Bradford assay, using BSA as standard (Bradford, 1976).

## 2.8 Para-Nitrophenyl-linked assays

A pNP (Sigma Cat# 1048, Sigma-Aldrich, South Africa) standard curve was created by measuring the absorbance of a concentration range (0.00 µmol, 12.5 µmol, 25µmol, 50µmol and 100µmol) of pNP in assay buffer (HEPES-KOH buffer, PH 7) after the addition of 200 µL sodium bicarbonate stop buffer (0.5 M).

A standard pNP-glycoside assay was adapted for use in the 96 well microtitre plate format (Zhang et al. 2017). The assay was conducted by incubating 0.7 mM pNP-glycoside (pNP-α-D-glucopyranoside, Sigma Cat# N1377; pNP-β-D-glucopyranoside, Sigma Cat# N7006, Sigma-Aldrich, South Africa; pNP-α-D-galactopyranoside, Sigma Cat# N0877, Sigma-Aldrich, South Africa; pNP-β-D-galactopyranoside, Sigma Cat# N1252, Sigma-Aldrich, South Africa) in 45 µL HEPES-KOH buffer (50 mM, pH 7) with 50 µL crude protein extract (30 °C, 10 min). The reaction was terminated by the addition of 200 µL sodium bicarbonate (0.5 M). The colour change was measured at 405 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

## 2.9 Natural substrate assay

A range of natural β-glucoside substrates were chosen which included arbutin (Sigma Cat# A4256, Sigma-Aldrich, South Africa), cellobiose (Sigma Cat# C7252, Sigma-Aldrich, South Africa), cellulose (microcrystalline powder, Sigma Cat # 435236, Sigma-Aldrich, South Africa), cellulose-acetate (Sigma Cat# 180955, Sigma-Aldrich, South Africa), esculin hydrate, lactose, and salicin (Sigma Cat# S0625, Sigma-Aldrich, South Africa). The assay was conducted by incubating 10 mM substrate in 40 µL HEPES-KOH buffer (50 mM, pH 7) with 50 µL crude protein extract (30 °C, 1 h). The reaction was terminated by snap freezing in liquid nitrogen. The liberated glucose was measured with a glucose assay kit (Sucrose/D-Glucose/D-Fructose UV test, R-biopharm, Roche, South Africa) adapted for use in 96 well microtitre plate format. Absorbance was measured at 340 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

## 2.10 Optimum activity parameters

The assay to determine optimum pH for BG4 activity was conducted by incubating 10 mM esculin hydrate in citric acid (50 mM, pH 4 to 5) or McIlvaine buffer (50 mM, pH 5 to 6) or MES-KOH (50 mM, pH 6 to 7) or HEPES-KOH (50 mM, pH 7 to 9) or sodium carbonate (50 mM, pH 9 to 10) buffer with 40 µL crude protein extract (30 °C, 1 h). The reaction was terminated by snap freezing in liquid nitrogen. The liberated glucose was measured with a glucose assay kit (Sucrose/D-Glucose/D-Fructose UV test, R-biopharm, Roche, South Africa) adapted for use in 96 well microtitre plate format. Absorbance was measured at 340 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

The assay to determine optimum temperature for BG4 was conducted by incubating 10 mM esculin hydrate in HEPES-KOH buffer (50 mM, pH 7) with 40 μL crude protein extract over a temperature gradient of 10 °C to 60 °C with intervals of 5 °C (1 h). The reaction was terminated by snap freezing in liquid nitrogen. The liberated glucose was measured with a glucose assay kit (Sucrose/D-Glucose/D-Fructose UV test, R-biopharm, Roche, South Africa) adapted for use in 96 well microtitre plate format. Absorbance was measured at 340 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

## 2.11 Effect of ions, additives, solvents and glucose

A standard pNP-glycoside assay was adapted for use in the 96 well microtitre plate format. The assay to determine the effect of ions, additives, solvents on BG4 activity was set up by incubating 0.7 mM pNP-β-glucopyranoside in 45 μL MES-KOH buffer (50 mM, pH 6) with 50 μL crude protein extract (40 °C, 10 min). Added was one of the following: $Al(SO_4)_3$ (10 mM), $CaCl_2$ (10 mM), $CoCl_2$ (10 mM), KCl (10 mM), $MgCl_2$ (10 mM), $MnCl_2$ (10 mM), NaCl (10 mM), $NiCl_2$ (10 mM), $ZnCl_2$ (10 mM), dithiothreitol (10 mM), β-Mercaptoethanol (10% v/v), ethylenediaminetetraacetic acid (10 mM), sodium dodecyl sulfate (10 mM), glycerol (10% v/v), triton X-100 (10% v/v), dimethyl sulfoxide (10% v/v), ethanol (10% v/v), ethyl acetate (10% v/v) or isopropanol (10% v/v). The reaction was terminated by the addition of 200 μL sodium bicarbonate (0.5 M). The colour change was measured at 405 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

The assay to determine glucose inhibition on BG4 was conducted by incubating 0.7 mM pNP-β-glucopyranoside in 45 μL MES-KOH buffer (50 mM, pH 6) with 50 μL crude protein extract (40 °C, 10 min). Added to this was a concentration range (0 – 1000 mM) of additional free glucose. The reaction was terminated by the addition of 200 μL sodium bicarbonate (0.5 M). The colour change was measured at 405 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software.

The ions, additives and solvents were selected on the basis of previously reported inhibitors of β-glucosidases or carbohydrate active enzymes. Glucose is a well-known and well defined inhibitor of β-glucosidases and was studied more in depth as it is routinely used to define β-glucosidases' biochemical parameters.

19

## 2.12 Determining enzymatic kinetics

The assay to determine enzyme kinetics for natural substrates of BG4 was set up by incubating arbutin (0 mM – 100 mM); esculin (0 mM – 30 mM); salicin (0 mM – 100 Mm) in 50 µL MES-KOH (50 mM, pH 6) buffer with 40 µL crude protein extract (40 °C, 1 h). The reaction was terminated by snap freezing in liquid nitrogen. The liberated glucose was measured with a glucose assay kit (Sucrose/D-Glucose/D-Fructose UV test, R-biopharm, Roche, South Africa) adapted for use in 96 well microtitre plate format. Absorbance was measured at 340 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software. The $K_m$ and $V_{max}$ were subsequently determined by fitting data to a non-linear regression curve of GraphPad Prism 5.

The assay to determine enzyme kinetics of BG4  for artificial substrate pNP-β-glucopyranoside was set up by incubating pNP-β-glucopyranoside (0 mM – 75 mM ) in 50 µL MES-KOH buffer (50 mM, pH 6) with 40 µL crude protein extract (40 °C, 10 min). The reaction was terminated by the addition of 200 µL sodium bicarbonate (0.5 M). The colour change was measured at 405 nm using VersaMax ELISA microtitre plate reader using SoftMax Pro software. The $K_m$ and $V_{max}$ were subsequently determined by fitting data to a non-linear regression curve of GraphPad Prism 5.

All enzymatic assays was done in triplicate and error bars represent standard deviation. One unit of enzyme activity (U) for crude enzyme extracts were defined as the amount of protein needed to liberate 1 µmol of product per minute under standard assay conditions.

## 2.13 Transglycosylation of glucose by BG4

Machereny-Nachel Protino® His-Tag purification kit (Machereny-Nachel, Separations, South Africa) was used to isolated His-tagged BG4 from crude protein extract according to the manufacturer's instructions. All the eluted fractions were assessed for activity using the pNP-β-glucopyranoside assay described earlier (see section 2.8). The transglycosylation ability of purified BG4 was determined by incubating free glucose 10 mM with methanol 5% in 200 µL MES-KOH (50 mM, pH 6) with 250 µL His-tag purified protein in a final volume of 1000 µL (30°C, 24 h).

# Results

## 3.1 Functional metagenomics: plate based screening as an enzyme discovery tool

A fosmid based metagenomic library containing 40 Kb inserts was constructed from the total genomic DNA isolated from an environment identified as cellulose acetate rich (cigarette filter waste bin). Functional plate based screening was used to discern if cellulolytic active enzymes were present in the metagenome of this environment. We found within the metagenomic clones a number of enzyme activities which represented cellulolytic specific activities (Figure 5). These represented (i) esterases which presented blue halo formation on tributyrin agar plates supplemented with victoria blue, (ii) cellulases which presented clear, yellow halo formation on carboxymethylcellulose plates and (iii) β-glucosidase which presented a dark brown halo formation on ferric citrate esculin plates. The clones identified presented distinct activities and showed no overlap when cross-tested on other plate based-screens (Figure 5). From these results, a putative β-glucosidase was identified (BG4) and chosen for further analyses. The fosmid was isolated and fully sequenced using a next-generation approach (Illumina MiSeq platform).
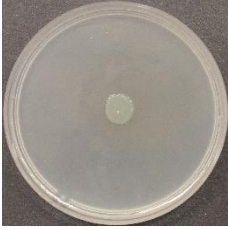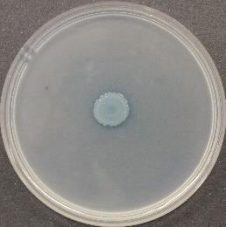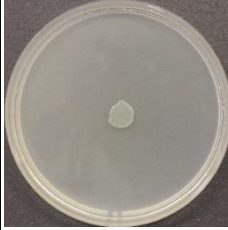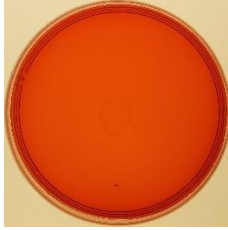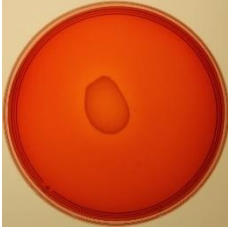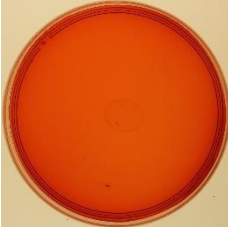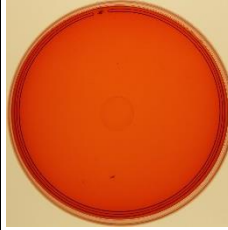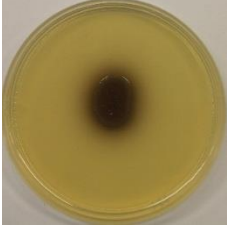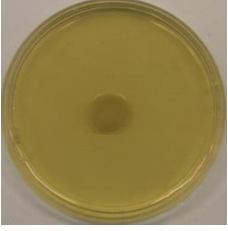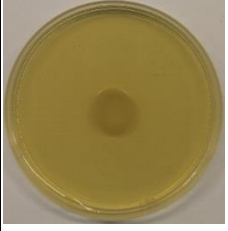
| Plate screen | β-glucosidase (BG4) | Cellulase (C5) | Esterase (E1) | Empty (pCC2Fos) |
|---|---|---|---|---|
| Tributyrin plate |  |  |  |  |
| Carboxmehylcellulose plate |  |  |  |  |
| Ferric citrate esculin plate |  |  |  |  |

Figure 5: Functional plate screens of the metagenomic library for the identification of potential cellulose-acetate degrading enzymes. The metagenomic library was subjected to screens for esterase, cellulase and β-glucosidase enzymes. Tributyrin agar plates indicate lipolytic activity by the formation of a blue halo, carboxymethylcellulose plates indicate cellulase activity the by formation of a yellow halo and ferric citrate esculin indicate β-glucosidase activity by a brown halo formation. Here the phenotypic identification of one esterase (E1), cellulase (C5) and β-glucosidase (BG4) is shown and their activity on other screening methods. *E. coli* (EPI300™-T1R) transformed with empty pCC2FOS™ was included as a control.

## 3.2 *BG4* identified as a putative β-glucosidase containing two distinct GH1 family domain signatures.

Fosmid sequence data was analysed using CLC genomic workbench 11 and NCBI's ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) which identified 31 putative ORFs within the 44.7 Kb fosmid insert. Putative functions were assigned using BLASTx (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and only one putative ORF encoding a β-glucosidase (*BG4*)

22

was identified. The *BG4* sequence was further analysed using NCBI's conserved domain analysis (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) which further confirmed that *BG4* assigns to the glycoside hydrolase family one and has a β-glucosidase domain (BglB). The BG4 protein sequence was also analysed using ExPASy PROSITE (https://prosite.expasy.org/) which identified two distinct motifs, a glycosyl hydrolases family one N-terminal signature and a glycosyl hydrolases family one active site (Figure 6B). The predicted amino acid sequence for *BG4* was aligned with six other β-glucosidases which represented significant hits returned from a BLAST search (Figure 7).

A)



B)
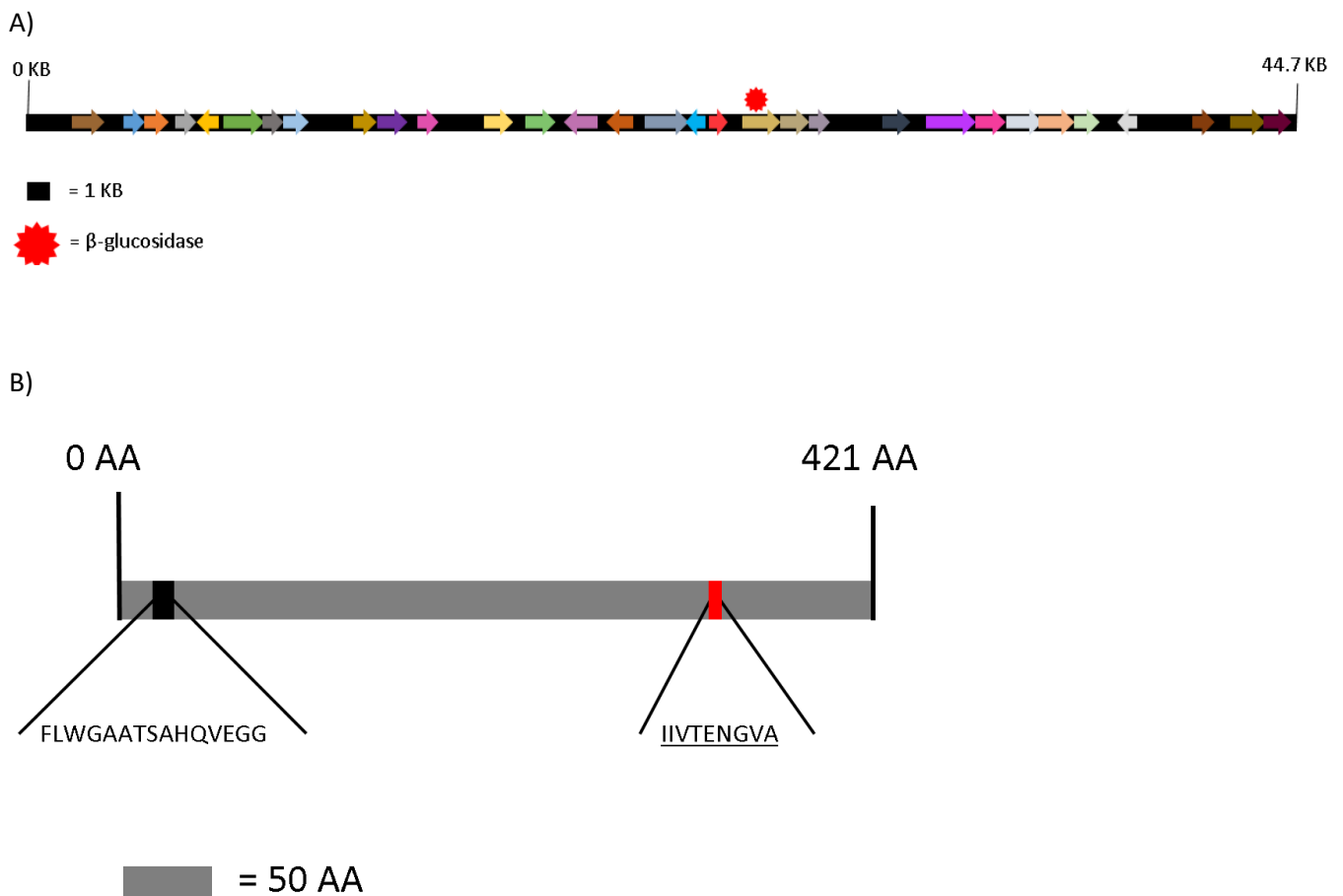


Figure 6: Nucleotide and amino acid sequence analysis of fosmid isolated from β-glucosidase positive clone. Analysis reveals *BG4* presence in fosmid and further sequence analysis. A) Fosmid ORF identification and BLAST to identify putative β-glucosidase. B) ExPASy PROSITE amino acid sequence analysis indicated that BG4 has two distinct GH1 specific domains present. The first is an N-terminal signature domain (FLWGAATSAHQVEGG) and the second an active site domain (IIVTENGVA).
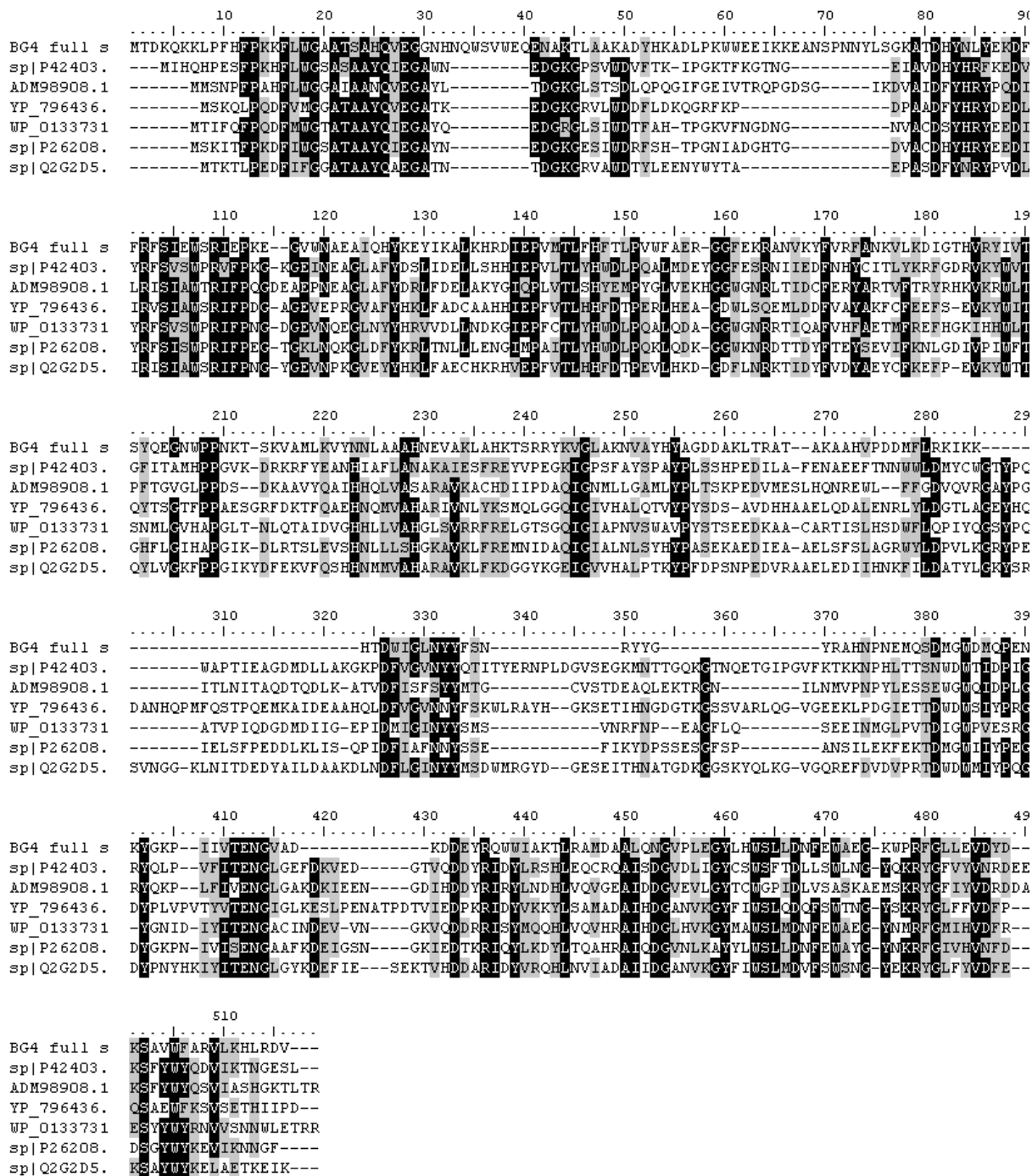
Figure 7: Amino acid sequence alignment of identified putative β-glucosidase (BG4) with six closely related β-glucosidases identified with xBLAST. The similarity is indicated in grey scale with 100% similarity indicated in black. Sequences producing significant alignments are to *Bacillus subtilis* (P42403.1, Fujishima and Yamane, 1995), *Dickeya dadantii* (ADM98908.1), *Lactobacillus paracasei* (YP_796436.1, Makarova et al., 2006), *Paenibacillus polymyxa* (WP_013373108.1), *Ruminiclostridium thermocellum* (P26208.1, Grabnitz et al., 1991) and *Staphylococcus aureus* (Q2G2D5.1, Gillaspy et al., 2006). Alignment was performed using BioEdit version 7.2.3 (Hall, 1999).

## 3.3 Isolation of *BG4* via PCR and construction of a heterologous protein expression construct

Following the identification of the *BG4* ORF, a specific primer pair was designed to incorporate restriction sites that would enable the subcloning of *BG4* into the bacterial expression vector pSF-OXB20-NH2-10His-EKT. This expression construct was subsequently transformed into *E. coli* (DH5α) and the presence of the *BG4* ORF was confirmed by colony PCR analyses (Figure 8).
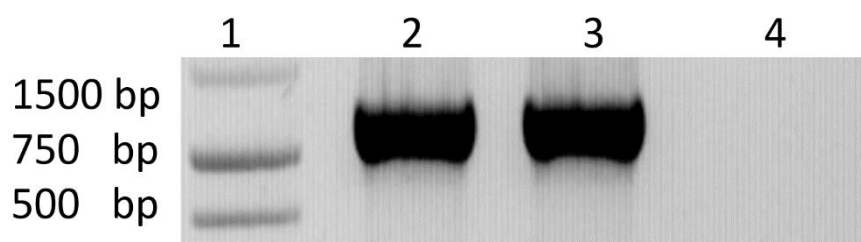


Figure 8: PCR confirmation of pSF-OXB20-NH2-10His-EKT::BG4 expression vector. Lane 1- 1 Kb Promega ladder, lane 2- pCC2FOS::BG4, lane 3- pSF-OXB20-NH2-10His-EKT::BG4, lane 4- pSF-OXB20-NH2-10His-EKT (control). Aliquots (1 ug) of purified plasmid preparations were used in a PCR with primers specific for *BG4* ORF were used and reaction products electrophoresed in a 1.0% w/v agarose gel.

## 3.4 Recombinant BG4 displays β-exohydrolase activity on paranitrophenol substrates

Crude protein extracts from *E. coli* (DH5α) containing the BG4 expression construct were tested for activity on a number of artificial substrates representing galactose and glucose linked to paranitrophenol (pNP) in either an α or β bond configuration. Activity was only observed when crude extracts were presented with pNP-β-glucopyranoside (specific activity 0.73 µmol. min$^{-1}$.mg$^{-1}$, Figure 9). Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activities on any of the pNP-based substrates. These findings confirmed that BG4 identified from the plate-based functional screening of the metagenomic library was a glucosyl β-exohydrolase, an activity typified by the β-Glucosidases (Cairns and Esen, 2010).

Figure 9: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The hydrolytic activity of these extracts was tested on artificial chromogenic substrates containing the two glycosyl moieties in varying bond configurations. Crude protein extracts (50 uL) was incubated with 0.7 mM of each substrate and 45 μL HEPES-KOH buffer (50 mM, pH 7) for 10 min at 30 ° C.  The substrates used were α-Gluc (pNP-α-D-glucopyranoside), β-Gluc (pNP-β-D-glucopyranoside), α-Gal (pNP-α-D-galactopyranoside), β-Gal (pNP-β-D-galactopyranoside). Crude protein-extracts from *E. coli* (DH5α) transformed with the empty pSF-OXB20-NH2-10His-EKT vector were used as a control.

## 3.5 Recombinant BG4 displays a preference toward aryl-β-glucosides and not cellobiose

Crude protein extracts from *E. coli* (DH5α) containing BG4 expression construct were tested for activity on natural β-linked substrates. Activity was observed for all aryl linked glycosides; esculin hydrate (1.432 μmol. min$^{-1}$.  mg$^{-1}$), salicin (0.830 μmol. min$^{-1}$.  mg$^{-1}$), arbutin (0.370 μmol. min$^{-1}$.  mg$^{-1}$) and genistin (0.1265 μmol. min$^{-1}$.  mg$^{-1}$; Figure 10). No activity was observed for cellubiose, lactose, microcrystalline cellulose or cellulose acetate. Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activities on any of the substrates. These findings confirmed that the BG4 ORF identified from the plate-based functional screening of the metagenomic library is an aryl-β-glucosides, classifying BG4 as a group ii β-glucosidase.
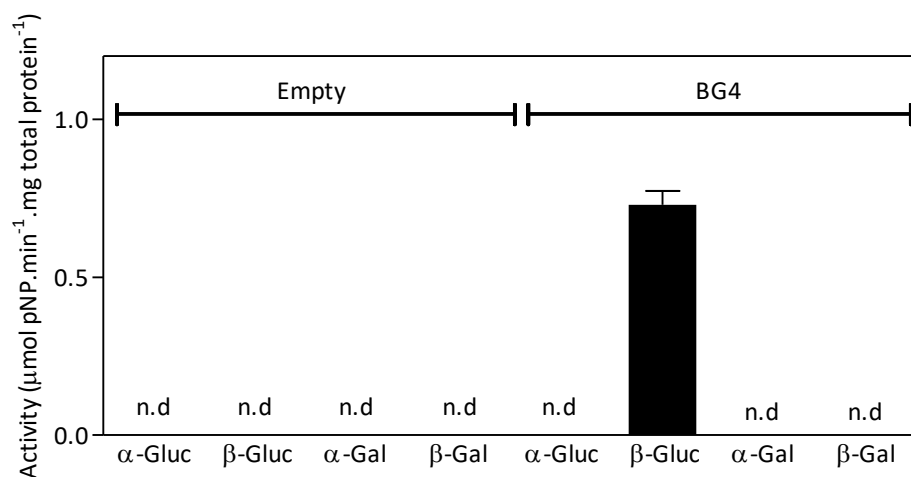
Figure 10: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The preference of BG4 toward various natural substrates was determined using eight natural substrates with β-1,4-bond configuration. The substrates included cellobiose (Cellob), lactose (Lact), microcrystalline cellulose (MCC), cellulose acetate (CelA), genistin (Genis), arbutin (Arb), salicin (Salic), esculin (Escul), and the chromogenic substrate pNP-β-Glucopyranoside (β-Gluc). Crude enzyme extracts (50 µL) were incubated in a 100 µL reaction with 10 mM natural substrate in 40 µL HEPES-KOH buffer (50 mM, pH 7) for 10 min at 30 ° C.

## 3.6 Recombinant BG4 displays optimum activity under slightly acidic conditions and prefers esculin hydrate as substrate

Crude protein extracts from *E. coli* (DH5α) containing BG4 expression construct were tested for activity over a temperature range from 10 °C to 60 °C with intervals of 5 °C. Activity increased incrementally from 10 °C up to 40 °C when it reached its maximum activity of 0.370 µmol.min$^{-1}$.mg$^{-1}$ (Figure 11A). It sharply decreased above 40 °C and activity was abolished at 60 °C. Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activity (data not shown). These findings confirm the optimal temperature for BG4 activity being 40 °C.

Crude protein extracts from *E. coli* (DH5α) containing BG4 expression construct were tested for activity over a pH range of four to ten using various buffers to ensure optimal buffer capacity at each pH. The pH optima for BG4 activity was determined to be pH 6 which was tested for with MES-KOH (0.469 µmol.min$^{-1}$.mg$^{-1}$) and McIlvaine (0.426 µmol.min$^{-1}$.mg$^{-1}$; Figure 11B). No activity was observed below pH 4 or above pH 8. Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activity (data not shown). These findings confirm the optimal pH for BG4 activity being pH 6. All subsequent assays were conducted at 40 °C and pH 6 (MES-KOH).
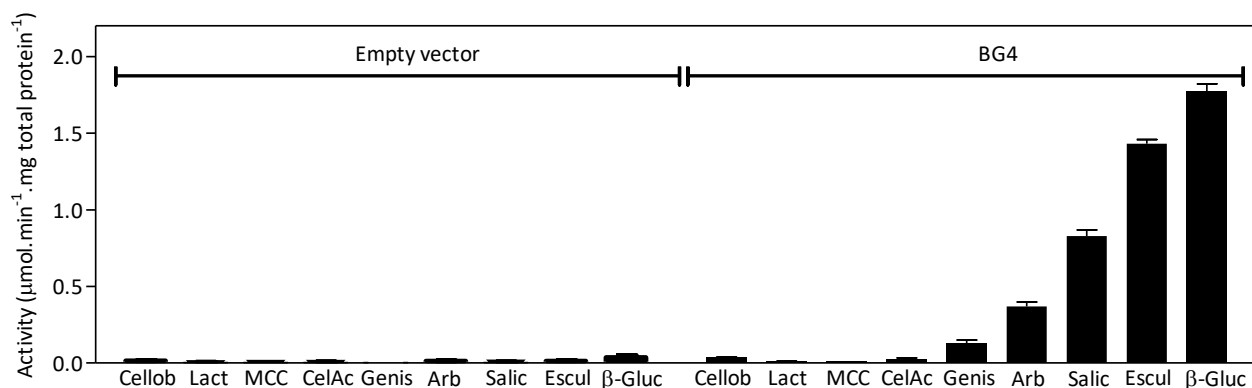
A)



B)



Figure 11: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. Optimum (A) temperature was determined over a temperature range of 10 – 60 °C with intervals of 5 °C (B) pH was determined over pH range4 – 10 with intervals of 1. Crude enzyme extracts (40 µL) were incubated in a 100 µL reaction with 10 mM esculin hydrate in (A) HEPES-KOH (50 mM, pH 7) for 1 h at specified (10 – 60 °C) temperature or (B) specified buffer (pH 4 – 10) for 1 h at 30 ° C.

Crude protein extracts from *E. coli* (DH5α) containing BG4 expression construct were tested to determine the kinetic parameters on four natural substrates with the highest specific activity (Figure 10). The maximum velocity of activity (Vmax) and substrate affinity (Km) were determined for pNP-β-glucopyranoside (0 mM – 75 mM; Figure 12A), esculin hydrate (0 mM – 30 mM; Figure 12B), salicin (0 mM – 100 mM; Figure 12C) and arbutin (0 mM – 100 mM; Figure 12D) at 40 °C and pH 6. Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector

showed no activities on any of the substrates (data not shown). These findings confirmed that the BG4 prefers esculin hydrate showing the highest specificity and activity (table 2).

Table 2: Kinetic analysis of BG4 on four substrates. Enzyme kinetics were determined by measuring BG4 activity over substrate range of: Arbutin (0 mM – 100 mM), Esculin hydrate (0 mM – 30 mM), Salicin (0 mM – 100 mM) and pNP-β-glucopyranoside (0 mM – 75 mM).

| Substrate | Vmax (µmol. min$^{-1}$) | Km (mM) | Substrate range  (mM) |
|---|---|---|---|
| Arbutin | 0.26 | 53.30 | $0 - 100$ |
| Esculin hydrate | 0.25 | 5.15 | $0 - 30$ |
| Salicin | 0.28 | 24.00 | $0 - 100$ |
| pNP-β-glucopyranoside | 0.22 | 13.08 | $0 - 75$ |

C)

| VMAX | 0.2812 |
|------|--------|
| KM | 24.00 |



D)

| VMAX | 0.26 |
|------|------|
| KM | 53.30 |

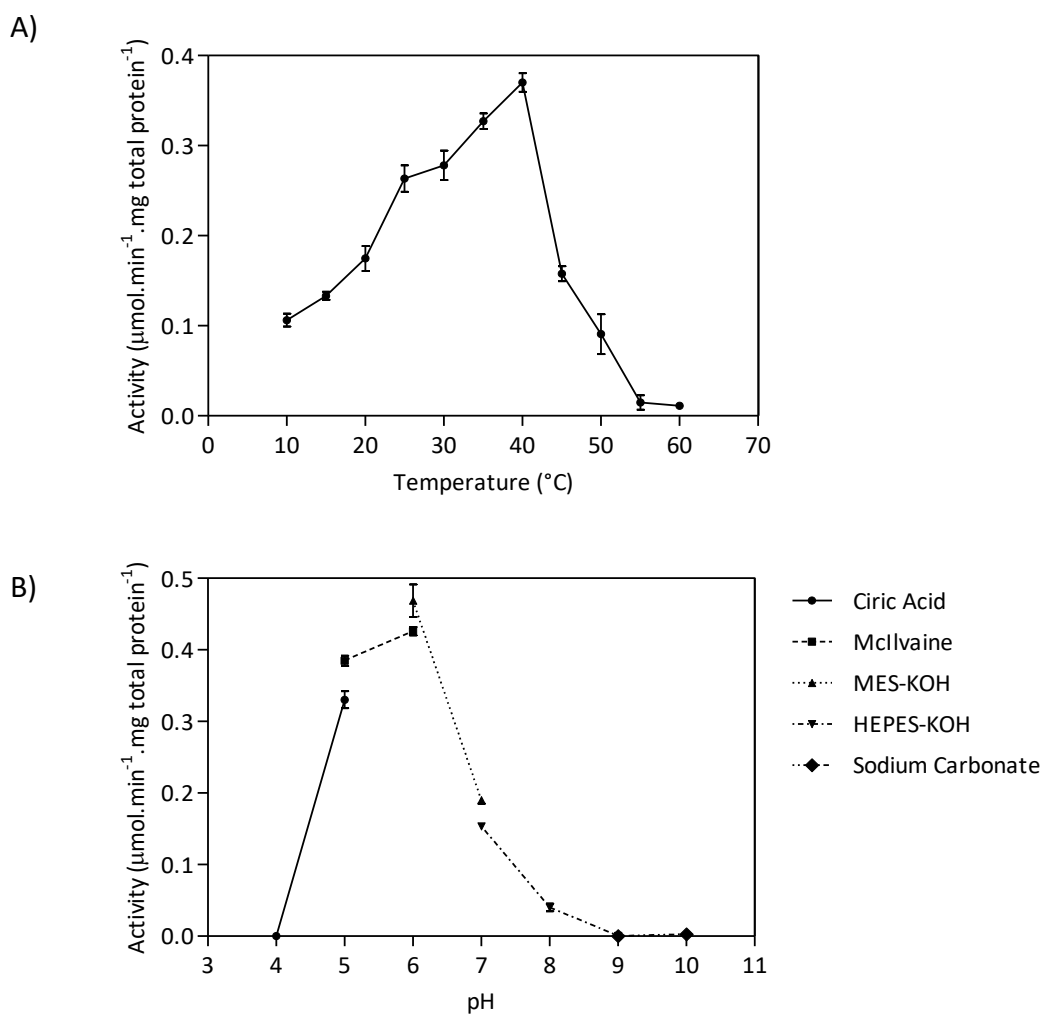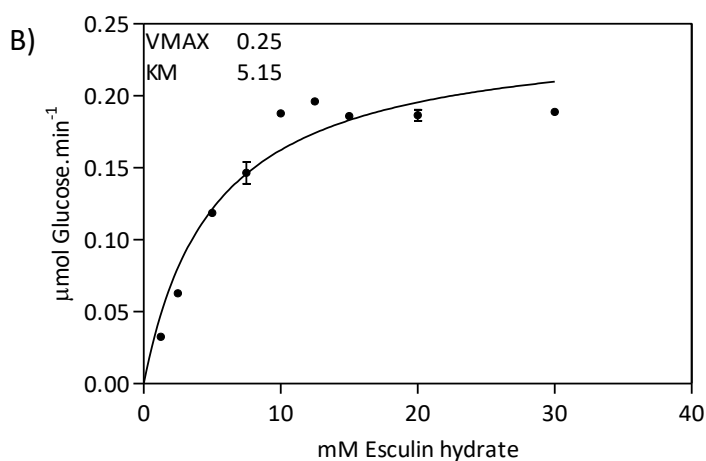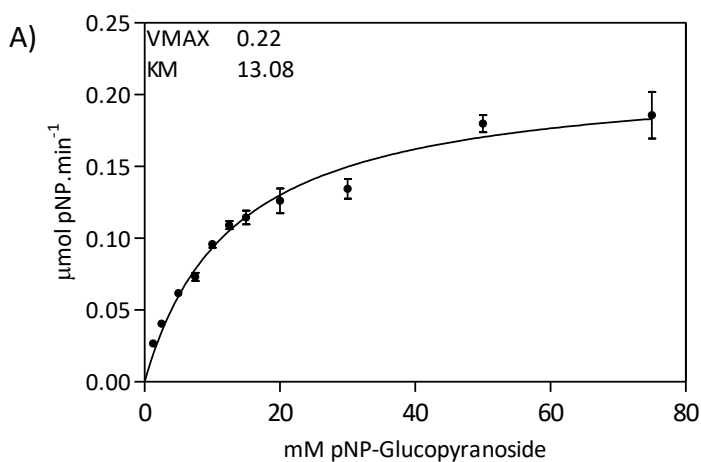Figure 12: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The enzymatic kinetics of BG4 was determined using hydrolysis of substrate gradient and crude extracts obtained from *E. coli* expressing BG4. BG4 activity was assessed for (A) pNP-β-glucopyranoside, (B) Esculin hydrate, (C) Salicin and (D) Arbutin. Crude enzyme extracts (40 μL) was incubated in a 100 μL reaction with a concentration gradient of various substrates in MES-KOH (50 mM, pH 6) for 1 h at 40 °C. Optimum temperature and pH was previously determined using esculin hydrate.

## 3.7 Recombinant BG4 activity is stimulated by MnCl₂ and ethyl acetate and is inhibited by detergents and glucose

Crude protein extract from *E. coli* (DH5α) containing BG4 expression construct was tested for activity in the presence of either 10 mM or 10% v/v ions, additives or solvents (table 3). Activity was measured using the standard pNP-β-glucosidase assay. MnCl₂ showed the highest increase in activity (34%) for the ions. ZnCl₂ and MgCl₂ resulted in a 62% and 33% decrease in activity

30

respectively. The additives did not show any activation but β-mercaptoethanol and SDS resulted in a 60% and 90% inhibition respectively. BG4 was inhibited by all solvents apart from ethyl acetate which showed a 10% increase in activity. DMSO inhibited BG4 by 50%, ethanol by 15% and propan-2-ol by 10%. Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activity (data not shown).

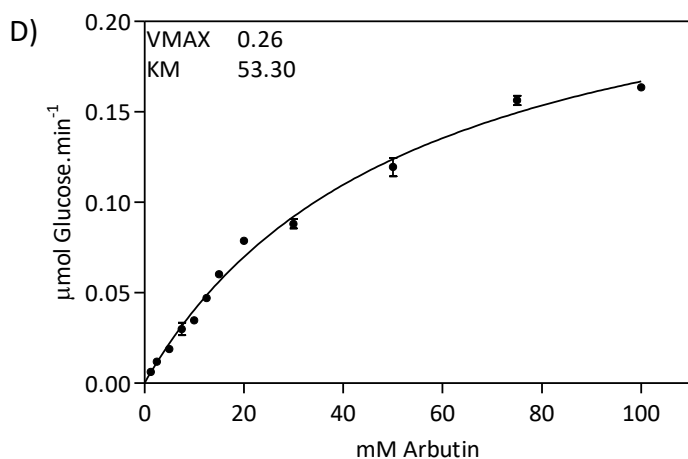Table 3: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The effect of ions, additives and solvents on BG4 activity was determined using pNP-glucopyranoside and crude extracts obtained from *E. coli* expressing BG4. Either 10 mM or 10% (v/v) of ions, additives or solvents were incubated with crude enzyme (50 μL) extracted in 100 μL reaction with 0.7 mM pNP-glucopyranoside in HEPES-KOH buffer (pH 6) for 10 min at 40 °C.

| Treatment | Concentration | Relative activity (%) |
|---|---|---|
| Ions | | |
| MnCl2 | 10 mM | 133.9 |
| CoCl2 | 10 mM | 107.2 |
| NiCl2 | 10 mM | 106.0 |
| CaCl2 | 10 mM | 105.0 |
| KCl | 10 mM | 102.0 |
| NaCl | 10 mM | 93.6 |
| Al(SO4)3 | 10 mM | 83.3 |
| MgCl2 | 10 mM | 77.2 |
| ZnCl2 | 10 mM | 38.2 |
| Additives | | |
| Ethylenediaminetetraacetic acid | 10 mM | 109.0 |
| Dithiothreitol | 10 mM | 104.1 |
| β-Mercaptoethanol | 10% v/v | 41.9 |
| Triton X-100 | 10% v/v | 89.4 |
| Glycerol | 10% v/v | 31.3 |
| Sodium dodecyl sulfate | 10 mM | 7.5 |
| Solvents | | |
| Ethyl acetate | 10% v/v | 110.9 |
| Propan-2-ol | 10% v/v | 91.6 |
| Ethanol | 10% v/v | 85.7 |
| Dimethyl sulfoxide | 10% v/v | 49.2 |
| No treatment | | 100* |

* Specific activity at 100 % was 0.489 μmol pNP. $min^{-1}$. mg total protein $^{-1}$

Crude protein extract from *E. coli* (DH5α) containing BG4 expression construct was tested for glucose tolerance from 0 – 100 mM glucose. Activity was measured using the standard pNP-β-glucosidase assay. At low molarity (1.5 mM) glucose proved to increase activity from 0.029 μmol pNP. min$^{-1}$ to 0.042 μmol pNP. min$^{-1}$. Concentrations higher than 1.5 mM reduced activity and 50% reduction was observed at 200 mM glucose (Figure 13). Crude extracts from *E. coli* (DH5α) containing the empty pSF-OXB20-NH2-10His-EKT vector showed no activity (data not shown).
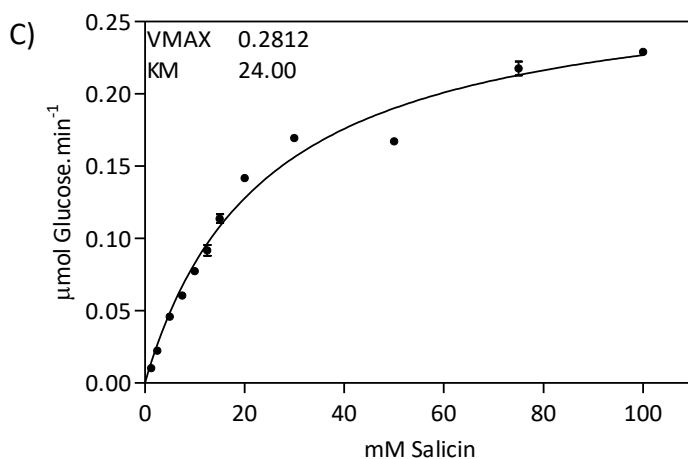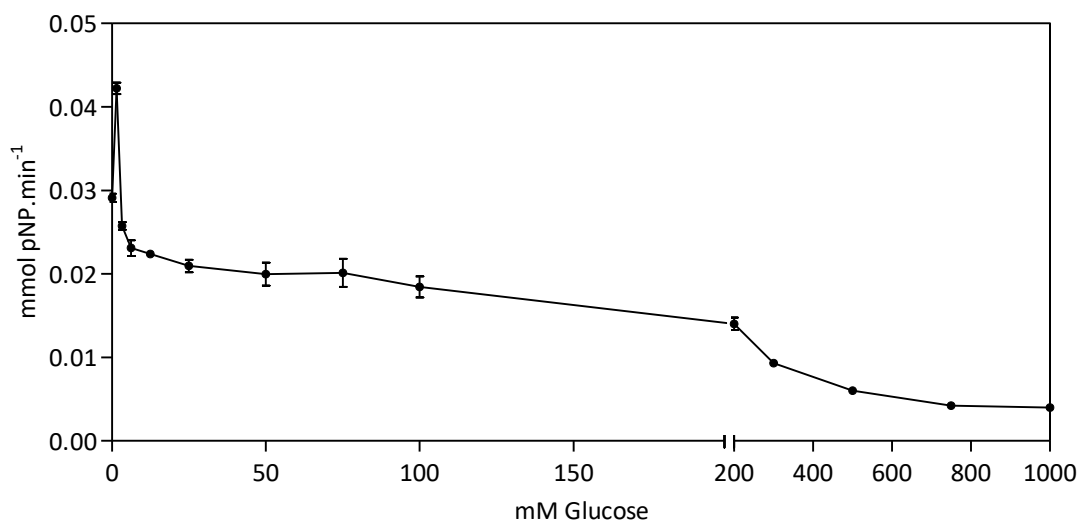


Figure 13: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The inhibition of BG4 by glucose was determined by assessing the reduction in pNP-β-glucopyranoside hydrolysis with increasing amount of glucose present. Crude extracts (40 μL) were incubated incubated with glucose (0 to 1000 mM) in a 100 μL reaction with 0.7 mM pNP-β-D-glucopyranoside in MES-KOH buffer (50 mM, pH 6) for 10 min at 40 ° C.

## 3.8 Purified recombinant BG4 is unable to transglycosylate using methanol as glucose acceptor

Crude protein extract from *E. coli* (DH5α) containing BG4 expression construct was tested for transglycosylation activity. Activity was measured by a reduction of free glucose in the presence of methanol which served as glycosyl acceptor as it is the most frequently used non-carbohydrate acceptor. The crude protein extract showed a clear reduction in free glucose but this was evident in the empty pSF-OXB20-NH2-10His-EKT vector control (data not shown). BG4 was subsequently

purified from the crude protein extract of *E. coli* (DH5α) containing BG4 expression construct using the His-tag purification. To ensure purified BG4 was present all the His-tag purification fractions was assessed for pNP-glucopyranoside hydrolysis and active fractions used subsequently. The active purified BG4 was tested for transglycosylation activity but no reduction in free glucose could be observed (Figure 14). BG4 does not have the ability to transglycosylate glucose using methanol as an acceptor.
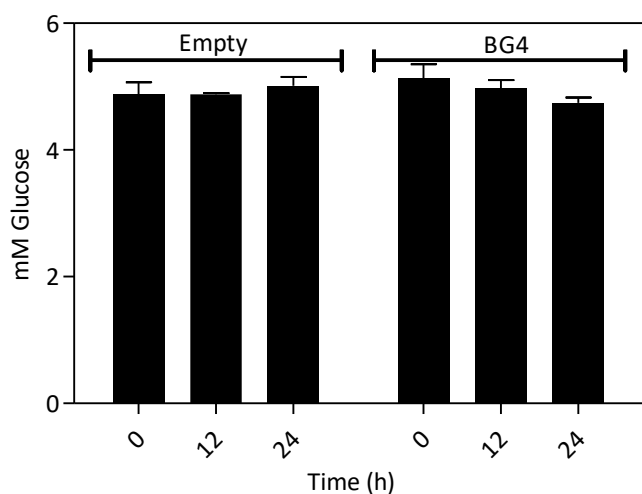


Figure 14: Following 3 h of growth at 37°C, crude protein extracts were isolated from *E. coli* (DH5α) containing the BG4 expression construct. The crude protein extract was subjected to His-tag protein purification (Protino Ni-TED 1000). The purified protein was assessed for transglycosylation ability. Free glucose was used as donor and methanol as acceptor in the reaction.  His-tag purified protein (250 µL) were incubated in a 1000 µL reaction with 10 mM glucose and 5% v/v methanol in HEPES-KOH (50 mM, pH 7) for 0, 12 and 24 h at 30 ° C.

# Discussion

## 4.1 Functional screening and sequence analysis: the search for BG4

Functional metagenomics is a powerful tool for (i) accessing the genetic potential of non-culturable microorganisms and (ii) the identification of novel enzymes since it relies on the functional identification of enzyme activities from any genetic source compared to sequence based identification from culturable  microorganisms (Madhavan et al., 2017). Culturable microorganisms are considered to represent only 1% of all microorganisms thus, metagenomics is a vital tool for accessing and harnessing the genetic potential of microorganisms from any environment (Vartoukian et al., 2010). Although metagenomics has the ability to capture the genetic potential the expression of novel genes in *E. coli* have resulted in low hit rate for gene identification (Uchiyama and Miyazaki, 2009). The average hit rate for identifying an active enzyme (from any environment) is roughly one in 1200 clones screened (Ferrer et al., 2016). Despite this shortcoming, the independence of functional metagenomics from prior sequence knowledge can yield completely novel gene discoveries. In a study of carbohydrate active enzymes, 64 open reading frames were identified from a cow rumen metagenome (Ferrer et al. 2005). Nine open reading frames which were isolated using functional screening methods did not share any sequence similarity to known carbohydrate active enzymes. This represents the portion of genes which would have been overlooked if a sequence based approach was chosen. A study using β-galactosidases complementation to screen a soil metagenome in *E. coli* and *Sinorhizobium meliloti* identified three completely novel β-galactosidases open reading frames. One of the identified genes lacked any conserved domains and shared no homology to known β-galactosidases, ultimately representing a completely novel enzyme family (Cheng et al., 2017). A study searching for β-glucosidase like activity form a soil metagenome also identified a novel β-glucosidase sharing no homology with known β-glucosidases (Jiang et al., 2009). These reports emphasise the necessity of applying functional metagenomics towards novel gene discovery as the enzymes could not have been identified using a sequence based analysis due to low sequence similarity.

In this study, a functional metagenomic approach was applied to identify carbohydrate active enzymes from a cellulose-acetate rich environment. The environment suggested that enzymes involved in cellulosic degradation and deacetylation could be present. Functional screens of the

metagenomic clones were conducted using plate based screening methods. β-Glucosidases (2), cellulases (5) and an esterase (1) were identified through screening of the metagenomic clones using highly specific plate screen methods (refer to section 3.1, table 2). Screening methods can distinguish between different enzymatic activities and allows for preliminary identification of activity. The environment from which the metagenomic library was created is cellulose-acetate rich, which serves as a carbon source for microorganisms present. β-Glucosidases are vital for the complete degradation of cellulose and were therefore expected to be present in this metagenomic library.

The fosmid library construction was efficient as a high fosmid titre ($1.2 \times 10^8$) was obtained indicating high transformation efficiency and high number of clones were obtained (840). The fosmid library yielded only a small number of positive clones when screened for various cellulolytic activities. This however corresponds to previously reported studies where between one (Jiang et al., 2009; Li et al., 2012b) and five (Bergmann et al., 2014) positive clones were identified using similar methodologies from soil metagenomes. The scope of this study was too biochemically characterise one β-glucosidase specifically with regards to characteristics generally required for second generation biofuel production (refer to section 1.2.1). One β-glucosidase indicating black halo formation on the plate base screen was selected, sequenced and analysed to ascertain sequence information of the active fosmid.

Sequence analysis revealed 31 distinct ORFs in total which were subsequently assessed by BLASTx analysis. One ORF was identified as a putative β-glucosidase and named *BG4*. The top three BLAST hits (66 – 64%) all belonged to candidatus *Saccharibacteria* which is a candidate phylum of bacteria originally identified in soil samples (Rheims et al., 1996). *BG4* was further analysed to ascertain domains present which revealed two distinct GH1 family domains. The presence of multiple domains in glycosyl hydrolases have been described since sequence based classification started (Henrissat and Bairoch, 1993). This two domain architecture, a catalytic C-terminal domain and a non-catalytic N-terminal domain, is routinely observed for GH3 β-glucosidases but is rarely seen in GH1 family β-glucosidases although some reports exists (Kim et al., 2007; L. et al., 2014). The first domain is the N-terminal signature domain (FLWGAATSAHQVEGG) and the second the active site domain (IIVTENGVA) for GH1s. The active site corresponds to the well-defined active site for GH1

([LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN]) with the glutamate (E) being the active site residue (https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS00572). The glutamate active site residue has been proven to be essential for activity of GH1 enzymes (Withers et al., 1990). The original BLASTx results were further assessed and the six β-glucosidases showing the highest similarity towards the GH1 domain were selected for protein sequence alignment (refer to section 3.2, Figure 6). The accessions were all classified as β-glucosidases and phosphor- β-glucosidases except for one, Q2G2D5, which was a β-galactosidase. Both were however GH1 family proteins explaining the similarity. BG4 showed the highest similarity towards P26208.1 which is a characterised β-glucosidase isolated from *Clostridium thermocellum* (Grabnitz et al., 1991). All of the β-glucosidase accessions except P24403.1 were classified as true β-glucosidases with cellobiose being the preferred substrate. P24403.1 was classified as an aryl-β-glucosidase supposedly preferring aryl-glucosides but this was not experimentally proven (Fujishima and Yamane, 1995). Due to the diverse classification of the proteins related to BG4 it could not be deduced what activity was to be expected. One commonality was that all six β-glucosidases were of bacterial origin which suggest BG4 originates from a soil bacterium. This is further corroborated by the similarity candidatus *Saccharibacteria* which is a known soil bacterium.

## 4.2 Characterisation of BG4

According to sequence analyses BG4 it is a β-glucosidase belonging to the GH1 family which is frequently described in literature and assessed for potential for bioethanol production (Biver et al., 2014; Crespim et al., 2016; Jeng et al., 2011). A major advantageous characteristic described for GH1 β-glucosidases is higher glucose tolerance and/or activation by glucose (Pang et al., 2017). Other characteristics also include pH-, temperature- and ethanol tolerance (Biver et al., 2014; Fusco et al., 2018; Martin et al., 2014) and activity on other important substrates (Li et al., 2012b; Swangkeaw et al., 2011). BG4 was recombinantly expressed in order to determine the (i) substrate preference (ii) optimum biochemical parameters and (iii) enzymatic kinetics of β-glucosidase activity.

## 4.2.1 Recombinant BG4 is a group ii β-glucosidase selectively hydrolysing aryl- β-glucosidases

β-Glucosidases hydrolyse the β-glycosidic bonds between two carbohydrates or a carbohydrate and non-carbohydrate moiety. In order to validate BG4's classification the activity towards chromogenic substrates with alternative bond configurations and glycosyl moieties were tested (refer to section 3.4, Figure 8). BG4 showed activity exclusively towards pNP-β-D-glucopyranoside (0.729 µmol. min$^{-1}$. mg$^{-1}$). This confirms the original sequence analysis and validates that classification of BG4 of glucosyl β-exohydrolase. The addition of pNP-α-D-glucopyranoside, pNP-α-D-galactopyranoside and pNP-β-D-galactopyranoside was to elucidate if the BG4 might have activity towards any other natural glycosidic substrates ,which  have been reported for β-glucosidases, such as sophorose (α-1,2-glucose), gentiobiose (β-1,6-glucose), lactose (β-1,4-galactose) and laminarin (α-1,3-glucose) (Uchiyama et al., 2013; Wei et al., 1996).

β-Glucosidase are generally divided into three subgroups with regards to substrates specificity, the (i) true β-glucosidases, which hydrolyse cellobiose/cellooligosaccharides to produce glucose; (ii) aryl-β-glucosidases, which are only active towards aryl-glucosides or (iii) broad specificity β-glucosidases, which show activity on a broad spectrum of substrates (Singhania et al., 2013). Most β-glucosidases are classified in as broad specificity and usually show greater activity towards cellobiose than aryl-glucosides (refer to section 4.2.2, table 4). A broad natural substrate range was used to determine which group BG4 can be classified into (refer to section 3.4, Figure 8). Interestingly, BG4 selectively hydrolysed aryl-β-glycosides which classified it in the least common group of β-glucosidases, the aryl-β-glucosidases (group ii). The specific activities was as follow esculin hydrate (1.43 µmol. min$^{-1}$. mg total protein$^{-1}$), by salicin (0.83 µmol. min$^{-1}$. mg total protein $^{-1}$), arbutin (0.37 µmol. min$^{-1}$. mg total protein $^{-1}$) and genistin (0.13 µmol. min$^{-1}$. mg total protein$^{-1}$).

The specific activities measured for BG4 using pNPG was 1.77 µmol. min$^{-1}$. mg total protein$^{-1}$ which seems quite low when comparing to other β-glucosidases isolated (refer to section 4.2.3, Table 4). Due to the lack of purification of BG4 it cannot be directly compared to listed β-glucosidases as they were all purified to homogeneity. Purification of an enzyme leads to significant increase of enzyme

37

activity (Berg et al., 2002). This is true for β-glucosidases also and would have to be considered in future for full biochemical characterisation (Murray et al., 2004)

## 4.2.2 Recombinant BG4 is optimally active at moderate temperature and slightly acid conditions

The temperature and pH optima were determined using BG4's preferred natural substrate, esculin hydrate. Industrially temperature and pH are important for optimal activity when considering possible applications. BG4 was active over a pH range of 4 to 9 with optimal activity at pH 6. This corresponds to the general pH optimum of β-glucosidases (4 to 7; https://www.brenda-enzymes.org/ enzyme.php?ecno=3.2.1.21). A pH optimum of 6 also corresponds to various enzymes isolated from soil metagenomes (Li et al., 2012a; Nacke et al., 2012; Wang et al., 2012). Aryl-β-glucosidases are routinely applied as flavour enhancers in beverages such as wine, beer and fruit juice, which is achieved by the release of aromatic molecules from their glycosylated inactive form (Bhatia et al., 2002). The pH of these beverages ranges from 2.5 for fruit juices (Birkhed, 1984; Sheehan et al., 2007) to 4.4 for beers (Kaneda et al., 1997) with wine in the middle (2.9 to 3.8; Boulton, 1980; Fischer and Noble, 1994). The acidic nature of these beverages makes BG4 inapplicable for beverage processing described above. Aryl-β-glucosidases are also applied in food processing where it is used to increase bioavailability of vitamin B6 in plant based foods, hydrolyse daidzin and genistin to reduce bitterness and increase nutritional value of soy bean products (soybean cooked syrup, soymilk, tofu) and hydrolysis of cyanogenic glycosides present in cassava root. The pH of these processes allows for BG4 application as they are generally in the neutral (pH 7; Sugimoto and van Buren, 1970) to slightly acidic range (pH 6; Akinrele, 1964). BG4 was active over a temperature range of 10 to 55 °C with optimum activity at 40 °C. This suggests BG4 originated from a mesophilic bacteria having optimal growth between 20 and 45 °C (Ingraham, 1958). Soil isolated enzymes, including β-Glucosidases, routinely show optimum activity under mesophilic conditions (Biver and Vandenbol, 2013; Iqbal et al., 2012; Voget et al., 2006; Yun et al., 2005; refer to section 4.2.2, table 4). Industrially β-glucosidases with high temperature tolerance is preferred for cellulosic biomass degradation but as BG4 is a strict aryl- β-glucosidases this attribute is not vital for industrial applications (Elleuche et al., 2014).

### 4.2.3 Recombinant BG4 is highly active on aryl-β-glucosidases and prefers esculin hydrate as substrate

β-Glucosidases with the ability to hydrolyse aryl-glycosides are not scarce as most show broad specificity towards glycosidic substrates (refer to section 4.2.2, table 4). It is however evident that β-glucosidases with the ability to hydrolyse aryl-glucosides generally do so at a much lower specific activity and it is very rarely the preferred natural substrate. BG4 activity towards substrates with the highest specific activity was further assessed to elucidate the enzymatic kinetics towards each. The preferred substrate of BG4 was esculin hydrate with a Km of 5.15 mM and Vmax of 0.25 μmol glucose. $min^{-1}$. Esculin hydrate was followed by pNP-β-D-glucopyranoside (13.08 mM and 0.22 μmol pNP-β-D-glucopyranoside. $min^{-1}$), salicin (24 mM and 0.28 μmol glucose. $min^{-1}$) and arbutin (53.3 mM and 0.26 μmol glucose. $min^{-1}$). The preference of esculin hydrate over pNP-β-D-glucopyranoside is a unique attribute of BG4 compared to other β-glucosidases which could imply applicability as a coumarin-β-glucosidase in industry.  The general selectivity for aryl-β-glucosides instead of cellobiose makes it applicable in the aryl-β-glucosidases required industries. The higher Km values compared to other characterised β-glucosidases (refer to section 4.2.2, table 4) can be attributed to the substrate ranges applied. In this study the substrate concentrations was high enough for BG4 to reach saturation and activity to plateau. Thus a true representative of Vmax and Km is seen. This is contradictory to most other studies characterising β-glucosidases where a very small substrate range is assessed such as 0.5 – 2.5 μM (Li et al., 2012b),  1 – 10 mM (Jiang et al., 2009) and 0.0625 – 4 mM (Uchiyama et al., 2015). Due to the lack of reaching saturation they tend to show lower Km values.

Table 4: Table indicating temperature optimum, pH optimum, substrate specificity and enzymatic activity of β-Glucosidases.

| Microorganism / source | | Optimum | | Natural substrate peference | Activity (pNPG) | | Reference |
|---|---|---|---|---|---|---|---|
| | | Temperature (°C) | pH | | Specific[1] | Km (mM) | |
| Soil metagenome | | 40 | 6 | Esculin hydrate, Salicin, Arbutin, Genistin | 1.77 | 13.08 | This study |
| Mangrove soil metagenome | | 40 | 6 | Daidzin, Genistin | NR | 0.23 | (Li et al., 2012a) |
| Cattle rumen metagenome | | 38 | 5 | NR | 2500 | 0.31 | (Li et al., 2014) |
| Soil metagenome | | 42 | 8 | Cellobiose, Salicin | 3.36 | 0.19 | (Jiang et al., 2009) |
| Soil metagenome | AS-Esc10 | 60 | 8 | Salicin, Esculin, Arbutin, Cellobiose | 4.08 | 0.2 | (Biver et al., 2014) |
| | AS-Esc6 | 40 | 8 | | 25.77 | 1.14 | |
| Marine microbial metagenome | | 40 | 6.5 | Cellobiose | 50 | 0.39 | (Fang et al., 2010) |
| *Bursaphelenchys xylophilus* metagenome | | 38 | 8 | Cellobiose, Lactose, Salicin, Lichenan, Laminarin, Sopharose | 180.3 | 2.33 | (Zhang et al., 2017) |
| Kusaya gravy metagenome | | 45 | 5 – 6.5 | Cellotetraose, Cellotriose, Laminaribose, Cellobiose, Cellopentaose, Sophorose, Salicin | 53.9 | 0.078 | (Uchiyama et al., 2015) |
| *Thermoascus aurantiacus* | | 80 | 4.5 | Cellobiose, Arbutin, Esculin hydrate, methyl-β-glucoside, salicin | 190 | 0.1137 | (Parry et al., 2001) |
| *Aureobasidium pullulans* | | 75 | 4.5 | Cellobiose, Salicin, CMC, Trehalose, Lactose, Sucrose | 315 | 1.17 | (Saha et al., 1994) |

| *Fomitopsis palustris* | 70 | 4.5 | Cellobiose | 191 | 0.117 | (Yoon et al., 2008) |
|---|---|---|---|---|---|---|
| *Talalaromyces emersonii* | 71.5 | | Salicin, Cellobiose | 512 | 0.13 | (Murray et al., 2004) |
| *Paecilomyces thermophila* | 75 | 6.2 | Cellobiose, Sopharose, Genistin, Gentibiose, Cellotriose, Gentibiose | 97 | 0.26 | (Yang et al., 2008) |
| *Monascus purpureus* | 50 | 5.5 | Maltose, Cellobiose, Salicin | 84 | 0.23 | (Daroit et al., 2008) |

1 Specific activity reported (µmol product produced per minute per milligram of protein); NR Not reported for enzyme

## 4.3 Recombinant BG4's industrially relevant characteristics

## 4.3.1 Recombinant BG4 activity is greatly increased by MnCl$_2$ and shows resistance towards ethyl acetate

The effect of ions, additives and solvents on BG4 activity was determined to assess further possible relevant characteristics (refer to section 3.7 table 3). EDTA showed a slight activation of activity indicating that divalent cations are not required for activity. Manganese chloride did however show result in a great increase in activity (30%) which has been reported for other β-glucosidases (Jeng et al., 2011; Park et al., 2005; Uchiyama et al., 2013). Strong inhibition of 62% was also seen for zinc chloride which corresponds to other metagenomically derived β-glucosidases (Biver et al., 2014; Uchiyama et al., 2015, 2013). This suggest that mangenese chloride should be included in future industrial application analysis. Additives indicated mostly inhibition and most notably a decrease of 92%, 68% and 11% for 10 mM SDS, 10% v/v glycerol and 10% v/v triton respectively, indicating the sensitivity of BG4 towards detergents. This is probably due to detergents disrupting the hydrophobic interactions which contribute to protein stability (Womack et al., 1983). The strong inhibition of SDS have been widely reported for β-glucosidases  (Kitpreechavanich et al., 1986; Painbeni et al., 1992; Riou et al., 1998). BG4 activity was lowered by all solvents except ethyl acetate which increased activity by 10 %. BG4 does however show resistance towards 10 % propan-2-ol and ethanol retaining 90 and 85 % of activity respectively. This resilience is important for solubilisation of water insoluble isoflavones such as daidzin and genistin (Wu et al., 2010). Thus this characteristic enables the utilisation of BG4 to liberate aglycones from aromatic precursors.

## 4.3.2 Recombinant BG4 is inhibited by glucose exceeding 1.5 mM

Glucose is a major and well-defined competitive inhibitor of β-glucosidases (Bhatia et al., 2002; Li et al., 2012a; Singhania et al., 2017). BG4 was activated by low glucose concentration but inhibited by concentrations exceeding 1.5 mM (refer to section 3.7, Figure 13). The activation of BG4 by glucose is similar to activation previously reported for GH1 β-glucosidases (Pang et al., 2017, Yang et al., 2015). BG4 is however still glucose sensitive which does not align with possible industrial applications. Glucose sensitivity is especially important when a batch reaction method is considered because glucose will build up and inhibit β-glucosidase activity. Thus if BG4 is applied in industry glucose liberated will be a major inhibitor and should be continually removed.

### 4.3.3 Recombinant BG4 has no transglycosylation ability with methanol as acceptor

Transglycosylation activity holds potential in the medical industry as various biologically active alryl-, alcohol-glycosides or oligosacchrides can be produced (Ahmed et al., 2017; refer to section 1.2.4). Transglycosylation activity have been reported for some β-glucosidases but not present in all β-glucosidases (Krisch et al., 2010). Among aryl-β-glucosidases none have been reported with transglycosylation ability but numerous group iii (broad specificity β-glucosidases) have been reported that can transglycosylate using aryl-β-glucosides as substrates. Although few reports of transglycosylating aryl- β-glucosidases exist BG4's transglycosylation ability was assessed. During transglycosylation a donor glucose molecule is transferred to either a glycone or aglycone acceptor molecule. Methanol is the most widely reported aglycone acceptor molecule reported in literature and was subsequently use for transglycosylation assessment (peerj, parry2001). BG4 translycosylation was initially assessed using crude total protein but no difference could be seen in the glucose reduction between BG4 and control (data not shown). This is not surprising when considering *E. coli*, thus it's cellular proteins, readily utilises glucose as preffered carbon source (Aidelberg et al., 2014). Thus BG4 was purified using His-tag purification and the fractions was assessed for β-glucosidase activity by using standard a pNP-β-pyranoside hydrolysis assay (data not shown). The active fractions was then used to assess transglycosylation. The purified BG4 showed no translycosylation for methanol as acceptor molecule after 24 h incubation with 5 mM glucose (refer to section 3.8, Figure 14). Interestingly, in recent studies, a possible link between GH1 glucose tolerance and transglycosylation ability have been proposed (De Giuseppe et al., 2014). Based on the protein structure of GH1 β-glucosidases it is proposed that glucose cannot attach to the active domain due to physical exclusion which results in higher glucose tolerance and higher transglycosylation. This could explain BG4's lack of glucose tolerance and transglycosylation ability.

In this study methanol was used as the glycosidic acceptor and although it is widely reported in transglycosylation studies other acceptors also exist such as other alcohols (ethaol and propan-2-ol) or disaccharides (cellobiose, gentiobiose, laminariobiose; Christakopoulos et al., 1994; Park et al., 2005).

## 4.4 Proposed industrial application for recombinant BG4

Medically aryl-β-glucosidases are applied to release biologically active aglycones from glycosylated precursors (refer to section 1.2.3 – 1.2.4). The molecules released are of value as the aglycone shows higher biological activity (Vocadlo and Davies, 2008). BG4 is able to hydrolyse genistin which has industrial value for aglycone release. Daidzin and gensitin are isoflavones which are routinely solubilised in ethyl acetate giving BG4 another advantage as it shows a slight activation by addition of 10 % ethyl acetate. BG4 also hydrolyses salicin and produces salicyl alcohol and glucose (Hudson and Paine, 1909). Salicyl alcohol is a precursor for salicylic acid which is widely applied in the health and personal care sectors (Transparency Market Research, 2013). In the health sector salicylic acid is applied in aspirin production which is used for diverse medical applications such as pain relief, anti-inflammation, anti-platelet and cardiovascular health (Yeomans, 2011). In the personal care sector, which is the fastest growing, it is used for (i) skin ailments such as pimples, acne, black heads and warts and used for (ii) hair products to reduce hair loss and dandruff (Transparency Market Research, 2013). The hydrolysis of esculin hydrate suggests that BG4 can be used to release coumarins from coumarin glycosides, which also show high biological activity and are used as pharmaceuticals to treat lymphedema (Farinola and Piller, 2005),  blood clotting and certain cancers (Rohini and Srikumar, 2014). New coumarin-glycosides reported can also be hydrolysed and biological activity assessed. Arbutin is a hydroquinone glycoside which also indicates that BG4 can be used for the release of hydroquinones. Other industrial important aromatic-glycoside substrates also exist such as

BG4 hydrolyses a range of naturally occurring aromatic glucosides which have vast amounts of known and yet to be discovered biological activities and can that be used as pharmaceuticals. The inhibition of BG4 by glucose, a product of hydrolysis can be overcome by creating a continues flow system where BG4 can be immobilised on a cylindrical reactor and the products continually removed. This method has been proposed previously for β-glucosidases (Fujikawa et al., 1987; Schwarz et al., 2009) and the immobilisation would be beneficial for enzyme reuse and increased stability toward temperature, pH and solvents (Rani et al., 2014).

# Conclusion and future work

In conclusion, BG4 is a unique β-glucosidase isolated form a cellulose-acetate rich environment. BG4 was highly specific towards aromatic glucosides and had no activity towards cellobiose. The identification of a strict aryl-β-glucosidase in this environment is surprising as the high prevalence of cellulose-acetate might have suggested that a cellobiose active β-glucosidase would rather be present. It is however vital to consider that the acetate moieties present on cellulose-acetate has to be removed prior to cellulose hydrolysis and here I suggest that BG4 could have this primary activity in its natural environment.  The ether bond that exists between the aromatic glycosides which BG4 is active on is the same as the bond between glucose and acetate. This suggests an accessory role of BG4 during cellulose acetate hydrolysis. The deacetylation of glucose-acetate and cellulose-acetate by BG4 has to be assessed and BG4 could thus play an important role in future cellulose-acetate bioremediation efforts.

BG4 crude protein specific activity was rather low when compared to other β-glucosidases. This is however prior to purification which is expected to increase the activity as previously seen when purifying. BG4 had very high selectivity towards naturally occurring aromatic glycosides tested and can be used to release biologically potent aglycones. BG4 shows optimal activity at moderate conditions (temperature and pH) which makes it inapplicable for the beverage industries as they require activity at lower pH but BG4 could still be applied in the food and medical industries. BG4 has the highest specificity for esculin hydrate suggesting it can be used for coumarin glycoside hydrolysis to release hydroquinones. It also shows high activity towards salicin and can be used for salicyl alcohol production, which in turn is used for production of cosmetics. Finally, BG4 has the ability to hydrolyse daidzin and genstin which are isoflavonoids with potent aglycones. The robustness of BG4 towards solvents is essential for this application as aromatic-glycosides have low solubility in water.

The future experimentation is required to assess the full industrial application. These include testing a broader range of aromatic glycosidase with industrial potentials such as flavonoid glucosides, flavonol glucosides and coumarin glucosides. The lack of transglycosylation ability can be explained by the low glucose tolerance of BG4 but could also be due to the wrong acceptor molecule applied and will be reassessed by applying other acceptors such as a disaccharide or aryl-glucoside which

will require HPLC analysis. To produce an industrially relevant production method a continuous flow reactor will have to be assessed. This will be done by optimising the method of immobilisation, reassessing the optimal biochemical conditions, determining the enzyme kinetics for selected substrate and testing the continuous flow reactor.

This study applied functional metagenomics towards the identification of a novel β-glucosidase (BG4) indicating the power for novel gene identification. BG4 is a group ii β-glucosidase with the unique ability to selectively hydrolyse aryl-β-glucosides and have no activity towards cellobiose. This is the least reported group of β-glucosidases and BG4 can thus can provide vital information regarding structural characteristics which contribute to the substrate specificity. Furthermore, BG4 can be applied towards cosmetic and medical applications where aryl-β-glucosides hydrolysis is required for the release of biologically potent aglycone moieties.

# Reference list

Aditiya, H.B., Mahlia, T.M.I., Chong, W.T., Nur, H., Sebayang, A.H., 2016. Second generation bioethanol production: A critical review. Renew. Sustain. Energy Rev. 66, 631–653. https://doi.org/10.1016/j.rser.2016.07.015

Ahmed, A., Nasim, F. ul-H., Batool, K., Bibi, A., 2017. Microbial β-Glucosidase: Sources, Production and Applications. J. Appl. Environ. Microbiol. Vol. 5, Pages 31-46. https://doi.org/10.12691/JAEM-5-1-4

Aidelber, G., Towbin, B.D., Rothschild, D., Dekel, E., Bren, E., Alon, U., 2014. Hierarchy of non-glucose saugars in Eschrichia coli. BMC Syst. Biol. 8. https://bmcsystbiol.biomedcentral.com/track/pdf/10.1186/s12918-014-0133-z

Akinrele, I.A., 1964. Fermentation of cassava. J. Sci. Food Agric. 15, 589–594. https://doi.org/10.1002/jsfa.2740150901

Béguin, P., 1990. Molecular biology of cellulose degradation of. Annu. Rev. Microbiol. 44, 1317–1542. https://doi.org/10.1152/physrev.00004.2012

Beilen, J.B. va., Li, Z., 2002. Enzyme technology: An overview. Curr. Opin. Biotechnol. 13, 338–344. https://doi.org/10.1016/S0958-1669(02)00334-8

Belhouari, Y., Farnum, B., Jenkins, C., Kieser, J., Román, A.L. de, McCauley, D., Chelsea Rochman, P.D., Schreiber, R., Schwartz, E., Taylor, H., Trott, S., 2017. Together for our Ocean. International Coastal Clean Up 2017 Report.

Berg, J., Tymoczko, J., Stryer, L., 2002. Biochemistry. 5th ed. New York, W.H. Freeman

Bergmann, J.C., Costa, O.Y.A., Gladden, J.M., Singer, S., Heins, R., D'haeseleer, P., Simmons, B.A., Quirino, B.F., 2014. Discovery of two novel β-glucosidases from an Amazon soil metagenomic library. FEMS Microbiol. Lett. 351, 147–155. https://doi.org/10.1111/1574-6968.12332

Bhatia, Y., Mishra, S., Bisaria, V.S., 2002. Microbial β-glucosidases: Cloning, properties, and applications. Crit. Rev. Biotechnol. 22, 375–407. https://doi.org/10.1080/07388550290789568

Birkhed, D., 1984. Sugar content, acidity and effect on plaque pH of fruit juices, fruit drinks, carbonated beverages and sport drinks. Caries Res. 18, 120–127.

Biver, S., Stroobants, A., Portetelle, D., Vandenbol, M., 2014. Two promising alkaline β-glucosidases isolated by functional metagenomics from agricultural soil, including one showing high tolerance towards harsh detergents, oxidants and glucose. J. Ind. Microbiol. Biotechnol. https://doi.org/10.1007/s10295-014-1400-0

Biver, S., Vandenbol, M., 2013. Characterization of three new carboxylic ester hydrolases isolated by functional screening of a forest soil metagenomic library. J. Ind. Microbiol. Biotechnol. 40, 191–200. https://doi.org/10.1007/s10295-012-1217-7

Bonanomi, G., Incerti, G., Cesarano, G., Gaglione, S.A., Lanzotti, V., 2015. Cigarette butt decomposition and associated chemical changes assessed by $^{13}$C CPMAS NMR. PLoS One 10, 1–16. https://doi.org/10.1371/journal.pone.0117393

Boulton, R., 1980. The General Relationship Between Potassium, Sodium and pH in Grape Juice and Wine. Am J Enol Vitic 31, 182–186.

Bourne, Y., Henrissat, B., 2001. Glycoside hydrolases and glycosyltransferases: Families and functional modules. Curr. Opin. Struct. Biol. 11, 593-600. https://doi.org/10.1016/S0959-440X(00)00253-0

Buchanan, C.M., Gardner, R.M., Komarek, R.J., Gardver, R.M., Komarek, R.J., 1993. Aerobic Biodegradation of Cellulose Acetate. J. Appl. Polym. Sceince 47, 1709–1719. https://doi.org/10.1007/s10924-010-0258-0

Cairns, J.R.K., Esen, A., 2010. β-Glucosidases. Cell. Mol. Life Sci. 67, 3389–3405. https://doi.org/10.1007/s00018-010-0399-2

Chen, I.C., Lin, W. De, Hsu, S.K., Thiruvengadam, V., Hsu, W.H., 2009. Isolation and characterization of a novel lysine racemase from a soil metagenomic library. Appl. Environ. Microbiol. 75, 5161–5166. https://doi.org/10.1128/AEM.00074-09

Cheng, J., Romantsov, T., Engel, K., Doxey, A.C., Rose, D.R., Neufeld, J.D., Charles, T.C., 2017. Functional metagenomics reveals novel β-galactosidases not predictable from gene sequences. PLoS One 12, 1–20. https://doi.org/10.1371/journal.pone.0172545

Christakopoulos, P., Goodenough, P.W., Kekos, D., Macris, B.J., Claeyssens, M., Bhat, M.K., 1994. Purification and characterization of an extracellular beta-glucosidase with transglycosylation

and exo-glucosidase activities from Fusarium oxysporum. Eur. J. Biochem. 224, 379–385.

Coursey, D.G., 1973. Cassava as Food: Toxicity and Technology 27–36.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., Wittwer, P., 2005. Metagenomic gene discovery: Past, present and future. Trends Biotechnol. 23, 321–329. https://doi.org/10.1016/j.tibtech.2005.04.001

Cragg, S.M., Beckham, G.T., Bruce, N.C., Bugg, T.D.H., Distel, D.L., Dupree, P., Etxabe, A.G., Goodell, B.S., Jellison, J., McGeehan, J.E., McQueen-Mason, S.J., Schnorr, K., Walton, P.H., Watts, J.E.M., Zimmer, M., 2015. Lignocellulose degradation mechanisms across the Tree of Life. Curr. Opin. Chem. Biol. 29, 108–119. https://doi.org/10.1016/j.cbpa.2015.10.018

Crespim, E., Zanphorlin, L.M., de Souza, F.H.M., Diogo, J.A., Gazolla, A.C., Machado, C.B., Figueiredo, F., Sousa, A.S., Nóbrega, F., Pellizari, V.H., Murakami, M.T., Ruller, R., 2016. A novel cold-adapted and glucose-tolerant GH1 β-glucosidase from Exiguobacterium antarcticum B7. Int. J. Biol. Macromol. https://doi.org/10.1016/j.ijbiomac.2015.09.018

Dan, S., Marton, I., Dekel, M., Bravdo, B.A., He, S., Withers, S.G., Shoseyov, O., 2000. Cloning, expression, characterization, and nucleophile identification of family 3, Aspergillus niger β-glucosidase. J. Biol. Chem. 275, 4973–4980. https://doi.org/10.1074/jbc.275.7.4973

Daroit, D.J., Simonetti, A., Hertz, P.F., Brandelli, A., 2008. Purification and characterization of an extracellular β-glucosidase from Monascus purpureus. J. Microbiol. Biotechnol. 18, 933–41.

Davies, G., Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases. Structure. 3, 853-859. https://doi.org/10.1016/S0969-2126(01)00220-9

de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, 362–365. https://doi.org/10.1093/nar/gkl124

De Giuseppe, P.O., Souza, T.D.A.C.B., Souza, F.H.M., Zanphorlin, L.M., Machado, C.B., Ward, R.J., Jorge, J.A., Furriel, R.D.P.M., Murakami, M.T., 2014. Structural basis for glucose tolerance in GH1 β-glucosidases. Acta Crystallogr. Sect. D Biol. Crystallogr. 70, 1631–1639. https://doi.org/10.1107/S1399004714006920

De Souza, W.R., 2013. Microbial Degradation of Lignocellulosic Biomass. Sustain. Degrad. Lignocellul. Biomass - Tech. Appl. Commer. https://doi.org/10.5772/54325

Doerner, K.C., White, B.A., 1990. Assessment of the endo-1,4-β-glucanase components of Ruminococcus flavefaciens FD-1. Appl. Environ. Microbiol. 56, 1844–1850.

Duan, C.J., Feng, J.X., 2010. Mining metagenomes for novel cellulase genes. Biotechnol. Lett. 32, 1765–1775. https://doi.org/10.1007/s10529-010-0356-z

Dumont, M.G., Radajewski, S.M., Miguez, C.B., McDonald, I.R., Murrell, J.C., 2006. Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. Environ. Microbiol. 8, 1240–1250. https://doi.org/10.1111/j.1462-2920.2006.01018.x

Elleuche, S., Schröder, C., Sahm, K., Antranikian, G., 2014. Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. Curr. Opin. Biotechnol. 29, 116–123. https://doi.org/10.1016/j.copbio.2014.04.003

Escobar-Zepeda, A., De León, A.V.P., Sanchez-Flores, A., 2015. The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. Front. Genet. 6, 1–15. https://doi.org/10.3389/fgene.2015.00348

Fan, G., Xu, Y., Zhang, X., Lei, S., Yang, S., Pan, S., 2011. Characteristics of immobilised β-glucosidase and its effect on bound volatile compounds in orange juice. Int. J. Food Sci. Technol. 46, 2312–2320. https://doi.org/10.1111/j.1365-2621.2011.02751.x

Fang, Z., Fang, W., Liu, J., Hong, Y., Peng, H., Zhang, X., Sun, B., Xiao, Y., 2010. Cloning and characterization of a β-glucosidase from marine microbial metagenome with excellent glucose tolerance. J. Microbiol. Biotechnol. 20, 1351–1358. https://doi.org/10.4014/jmb.1003.03011

Farinola, N., Piller, N., 2005. Pharmacogenomics: Its Role in Re-establishing Coumarin as Treatment for Lymphedema. Lymphat. Res. Biol. 3, 81–86. https://doi.org/10.1089/lrb.2005.3.81

Ferrer, M., Golyshina, O. V., Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Martins Dos Santos, V.A.P., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N., Golyshin, P.N., 2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ. Microbiol. 7, 1996–2010. https://doi.org/10.1111/j.1462-

2920.2005.00920.x

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O. V., Golyshin, P.N., 2016. Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. Microb. Biotechnol. 9, 22–34. https://doi.org/10.1111/1751-7915.12309

Fischer, U., Noble,  A.C., 1994. The Effect of Ethanol, Catechin Concentration, and pH on Sourness and Bitterness of Wine. Am. J. Enol. Vitic. 45, 6–10.

Fujikawa, S., Yokota, T., Koga, K., Kumada, J., 1987. The Continuous Hydrolysis of Geniposide To Genipin Using Immobilized Beta-Glucosidase on Calcium Alginate Gel. Biotechnol. Lett. 9, 697–702. https://doi.org/10.1007/bf01024600

Fujishima, Y., Yamane, K., 1995. A 10 kb nucleotide sequence at the 5' flanking region (32") of srfAA of the Bacillus subtilis chromosome. Microbiology 141, 277–279.

Fusco, F.A., Fiorentino, G., Pedone, E., Contursi, P., Bartolucci, S., Limauro, D., 2018. Biochemical characterization of a novel thermostable β-glucosidase from Dictyoglomus turgidum. Int. J. Biol. Macromol. https://doi.org/10.1016/j.ijbiomac.2018.03.018

Gabor, E.M., Alkema, W.B.L., Janssen, D.B., 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ. Microbiol. 6, 879–886. https://doi.org/10.1111/j.1462-2920.2004.00640.x

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788. https://doi.org/10.1093/nar/gkg563

Gillaspy, A.F., Worrell, V., Roe, B., Dyer, D., Orvis, J., Iandolo, J., 2006. The Staphylococcus aureus NCTC 8325 Genome. 381–412. In Fischetti V., Novick R., Ferretti J., Portnoy D., Rood J., Gram-Positive Pathogens, Second Edition. ASM Press, Washington, DC. https://doi.org/10.1128/9781555816513.ch32

Goodman, P., Agnew, M., McCaffrey, M., Paul, G., Clancy, L., 2007. Effects of the Irish smoking ban on respiratory health of bar workers and air quality in Dublin pubs. Am. J. Respir. Crit. Care Med. 175, 840–845. https://doi.org/10.1164/rccm.200608-1085OC

Grabnitz, F., Seiss, M., Rucknagel, K.P., Staudenbauer, W.L., Munchen, T.U., 1991. Structure of the

β-glucosidase gene bglA of Clostridium thermocellum. Reactions 309, 301–309.

Gueguen, Y., Chemardin, P., Labrot, P., Arnaud,  a, Galzy, P., 1997. Purification and characterization of an intracellular β-glucosidase from a new strain of Leuconostoc mesenteroides isolated from cassava. J. Appl. Microbiol. 82, 469–476. https://doi.org/DOI 10.1046/j.1365-2672.1997.00136.x

Guerriero, G., Hausman, J.F., Strauss, J., Ertan, H., Siddiqui, K.S., 2016. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. Eng. Life Sci. 16, 1–16. https://doi.org/10.1002/elsc.201400196

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98. https://doi.org/citeulike-article-id:691774

Harris, B., 2011. The intractable cigarette "filter problem." Tob. Control 20, 10–16. https://doi.org/10.1136/tc.2010.040113

Henckel, T., Friedrich, M., Conrad, R., 1999. Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. Appl. Environ. Microbiol. 65, 1980–1990.

Henrissat, B., 1991. A classification of glycosyl hydrolases based sequence similarities amino acid. Biochem. J. 280, 309–316. https://doi.org/10.1007/s007920050009

Henrissat, B., Bairoch, A., 1996. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316, 695–696. https://doi.org/10.1042/bj3160695

Henrissat, B., Bairoch, A., 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 293, 781–788. https://doi.org/10.1042/bj2930781

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G., 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proc. Natl. Acad. Sci. 92, 7090-7094. https://doi.org/10.1073/pnas.92.15.7090

Horn, S.J., Vaaje-kolstad, G., Westereng, B., Eijsink, V.G.H., 2012. Novel enzymes for the degradation of cellulose Novel enzymes for the degradation of cellulose. Biotechnol. Biofuels 5, 1–13.

https://doi.org/10.1186/1754-6834-5-45

Hudson, C.S., Paine, H.S., 1909. The hydrolysis of salicin by the enzyme emulsin. J. Am. Chem. Soc. 31, 1242–1249. https://doi.org/10.1021/ja01941a010

Ilmberger, N., Streit, W.R. Screening for Cellulase Encoding Clones in Metagenomic Libraries. In Streit, W.R., Rolf, D., 2017. Metagenomics. Mehods in Molecular Biology 1539, 205–217. https://doi.org/10.1007/978-1-4939-6691-2

Ingraham, J.L., 1958. GROWTH OF PSYCHROPHILIC BACTERIA. J. Bacteriol. 76, 75–80.

Iqbal, H.A., Feng, Z., Brady, S.F., 2012. Biocatalysts and small molecule products from metagenomic studies. Curr. Opin. Chem. Biol. 16, 109–116. https://doi.org/10.1016/j.cbpa.2012.02.015

Isikgor, F.H., Becer, C.R., 2015. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym. Chem. 6, 4497–4559. https://doi.org/10.1039/C5PY00263J

Jabbour, D., Klippel, B., Antranikian, G., 2012. A novel thermostable and glucose-tolerant β-glucosidase from Fervidobacterium islandicum. Appl. Microbiol. Biotechnol. 93, 1947-1956. https://doi.org/10.1007/s00253-011-3406-0

Jeng, W.Y., Wang, N.C., Lin, M.H., Lin, C.T., Liaw, Y.C., Chang, W.J., Liu, C.I., Liang, P.H., Wang, A.H.J., 2011. Structural and functional analysis of three β-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis. J. Struct. Biol. 173, 46-56. https://doi.org/10.1016/j.jsb.2010.07.008

Jiang, C., Ma, G., Li, S., Hu, T., Che, Z., Shen, P., Yan, B., Wu, W., 2009. Charcterization of a novel β-glucosidase-like activity from a soil metagenome. J. Microbiol. 47, 542-548 https://doi.org/10.1007/s12275-009-0024-y

Juster, H.R., Loomis, B.R., Hinman, T.M., Farrelly, M.C., Hyland, A., Bauer, U.E., Birkhead, G.S., 2007. Declines in hospital admissions for acute myocardial infarction in new york state after implementation of a comprehensive smoking ban. Am. J. Public Health 97, 2035–2039. https://doi.org/10.2105/AJPH.2006.099994

Kaneda, H., Takashio, M., Tamaki, T., Osawa, T., 1997. Influence of pH on flavour staling during beer storage. J. Inst. Brew. 103, 21–23. https://doi.org/10.1002/j.2050-0416.1997.tb00932.x

Karnaouri, A., Topakas, E., Paschos, T., Taouki, I., Christakopoulos, P., 2013. Cloning, expression and characterization of an ethanol tolerant GH3 β-glucosidase from *Myceliophthora thermophila*. PeerJ. 1 https://doi.org/10.7717/peerj.46

Kim, S.-J., Lee, C.-M., Kim, M.-Y., Yeo, Y.-S., Yoon, S.-H., Kang, H.-C., Koo, B.-S., 2007. Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA. J. Microbiol. Biotechnol. 17, 905–912.

Kitpreechavanich, V., Hayashi, M., Nagai, S., 1986. Purification and Characterisation of Extracellular Beta-Xylosidase and Beta-Glucosidase from Aspergillus fumigatus. Agric. Biol. Chem. 50, 1703–1711.

Koshland, D.E., 1953. Stereochemistry and The Mechanism of Enzymatic Reactions. Biol. Rev. 28, 416–436. https://doi.org/10.1111/j.1469-185X.1953.tb01386.x

Krisch, J., Takó, M., Papp, T., Vágvölgyi, C., 2010. Characteristics and potential use of β -glucosidases from Zygomycetes. In Méndez-Vilas, A. Current research, technology and education topics in applied microbiology and microbial biotechnology, vol 2, Formatex Research Center, 891–896.

Kuhad, R.C., Gupta, R., Singh, A., 2011. Microbial Cellulases and Their Industrial Applications. Enzyme Res. 2011, 1–10. https://doi.org/10.4061/2011/280696

Laine, R.A., 1994. Invited commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 × 10 structures for a reducing hexasaccharide: The Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology 4, 759–767. https://doi.org/10.1093/glycob/4.6.759

Li, G., Jiang, Y., Fan, X.J., Liu, Y.H., 2012a. Molecular cloning and characterization of a novel β-glucosidase with high hydrolyzing ability for soybean isoflavone glycosides and glucose-tolerance from soil metagenomic library. Bioresour. Technol. 123, 15-22. https://doi.org/10.1016/j.biortech.2012.07.083

Li, G., Jiang, Y., Fan, X.J., Liu, Y.H., 2012b. Molecular cloning and characterization of a novel β-glucosidase with high hydrolyzing ability for soybean isoflavone glycosides and glucose-tolerance from soil metagenomic library. Bioresour. Technol. 123, 15–22. https://doi.org/10.1016/j.biortech.2012.07.083

Li, L.L., McCorkle, S.R., Monchy, S., Taghavi, S., van der Lelie, D., 2009. Bioprospecting metagenomes: Glycosyl hydrolases for converting biomass. Biotechnol. Biofuels 2, 1–11. https://doi.org/10.1186/1754-6834-2-10

Li, Y., Liu, N., Yang, H., Zhao, F., Yu, Y., Tian, Y., Lu, X., 2014. Cloning and characterization of a new β-glucosidase from a metagenomic library of rumen of cattle feeding with Miscanthus sinensis. BMC Biotechnol. 14, 85. https://doi.org/10.1186/1472-6750-14-85

Liu, D., Zhang, R., Yang, X., Zhang, Z., Song, S., Miao, Y., Shen, Q., 2012. Characterization of a thermostable β-glucosidase from Aspergillus fumigatus Z5, and its functional expression in Pichia pastoris X33. Microb. Cell Fact. 11. https://doi.org/10.1186/1475-2859-11-25

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, 490–495. https://doi.org/10.1093/nar/gkt1178

Madhavan, A., Sindhu, R., Parameswaran, B., Sukumaran, R.K., Pandey, A., 2017. Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting. Appl. Biochem. Biotechnol. 183, 636–651. https://doi.org/10.1007/s12010-017-2568-3

Makarova, K.S., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E. V., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D.M., Hawkins, T., Plengvidhya, V., Welker, D.L., Hughes, J.E., Goh, Y., Benson, A., Baldwin, K.A., Lee, J.-H., Díaz-Muñiz, I., Dosti, B., Smeianov, V. V., Wechter, W., Barabote, R., Lorca, G.L., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J.R., Hutkins, R., O'Sullivan, D., Steele, J.L., Unlu, G., Saier, M.H., Klaenhammer, T.R., Richardson, P., Kozyavkin, S., Weimer, B.C., Mills, D.A., 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. U. S. A. 103, 15611–15616. https://doi.org/10.1073/pnas.0607117103

Marques, A.R., Coutinho, P.M., Videira, P., Fialho, A.M., Sá-Correia, I., 2003. Sphingomonas paucimobilis beta-glucosidase Bgl1: a member of a new bacterial subfamily in glycoside hydrolase family 1. Biochem. J. https://doi.org/10.1042/BJ20021249

Martin, M., Biver, S., Steels, S., Barbeyron, T., Jam, M., Portetelle, D., Michel, G., Vandenbol, M., 2014. Identification and characterization of a halotolerant, cold-active marine endo-β-1,4-

glucanase by using functional metagenomics of seaweed-associated microbiota. Appl. Environ. Microbiol. 80, 4958–4967. https://doi.org/10.1128/AEM.01194-14

Maruthamuthu, M., Van Elsas, J.D., 2017. Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library. Biotechnol. Biofuels 10, 1–17. https://doi.org/10.1186/s13068-017-0808-y

Mase, T., Yokoe, M., Mori, S., 2004. Purification , characterization , and a potential application of β -glucosidase from Aspergillus pulverulentus YM-80. J Apply Glycosci. 51, 211-216.

Menendez, E., Garcia-Fraile, P., Rivas, R., 2015. Biotechnological applications of bacterial cellulases. AIMS Bioeng. 2, 163–182. https://doi.org/10.3934/bioeng.2015.3.163

Mercer, D.K., Robertson, J., Wright, K., Miller, L., Smith, S., Stewart, C.S., O'Neil, D.A., 2013. A prodrug approach to the use of coumarins as potential therapeutics for superficial mycoses. PLoS One 8. https://doi.org/10.1371/journal.pone.0080760

Michlmayr, H., Kneifel, W., 2014. β-Glucosidase activities of lactic acid bacteria: Mechanisms, impact on fermented food and human health. FEMS Microbiol. Lett. 352, -10. https://doi.org/10.1111/1574-6968.12348

Montagnera, C., de Souzaa, S.M., Groposoa, C., Monacheb, F.D., Smania, E.F.A., Smania, A., 2008. Antifungal activity of Coumarins. Zeitschrift für Naturforschung C: A journal of Biosciences. 63, 21–28. https://doi.org/10.1515/znc-2008-1-205

Murray, P., Aro, N., Collins, C., Grassick, A., Penttilä, M., Saloheimo, M., Tuohy, M., 2004. Expression in Trichoderma reesei and characterisation of a thermostable family 3 β-glucosidase from the moderately thermophilic fungus Talaromyces emersonii. Protein Expr. Purif. 38, 248–257. https://doi.org/10.1016/j.pep.2004.08.006

Nacke, H., Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J., Daniel, R., 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol. Lett. 34, 663–675. https://doi.org/10.1007/s10529-011-0830-2

Naumoff, D.G., 2011. Hierarchical Classification of Glycoside Hydrolases. Biochem. 76, 622–635.

Novotny, T.E., Lum, K., Smith, E., Wang, V., Barnes, R., 2009. Cigarettes butts and the case for an

environmental policy on hazardous cigarette waste. Int. J. Environ. Res. Public Health 6, 1691–1705. https://doi.org/10.3390/ijerph6051691

Öberg, M., Jaakkola, M.S., Woodward, A., Peruga, A., Prüss-Ustün, A., 2011. Worldwide burden of disease from exposure to second-hand smoke: A retrospective analysis of data from 192 countries. The Lancet 377, 139–146. https://doi.org/10.1016/S0140-6736(10)61388-8

Omoni, A.O., Aluko, R.E., 2005. Soybean foods and their benefits: Potential mechanisms of action. Nutr. Rev. 63, 272–283. https://doi.org/10.1301/nr.2005.aug.272-283

Opassiri, R., Hua, Y., Wara-Aswapati, O., Akiyama, T., Svasti, J., Esen, A., Ketudat Cairns, J., 2004. β-Glucosidase, exo-β-glucanase and pyridoxine transglucosylase.Biochem J. 379, 125-131.

Otieno, D.O., Shah, N.P., 2007. Endogenous β-glucosidase and β-galactosidase activities from selected probiotic micro-organisms and their role in isoflavone biotransformation in soymilk. J. Appl. Microbiol. 103, 910-917. https://doi.org/10.1111/j.1365-2672.2007.03438.x

Painbeni, E., Vallés, S., Polaina, J., Flors, A., 1992. Purification and characterization of a Bacillus polymyxa β-glucosidase expressed in Escherichia coli. J. Bacteriol. 174, 3087–3091.

Pang, P., Cao, L. chuang, Liu, Y. huan, Xie, W., Wang, Z., 2017. Structures of a glucose-tolerant β-glucosidase provide insights into its mechanism. J. Struct. Biol. 198, 154–162. https://doi.org/10.1016/j.jsb.2017.02.001

Park, T.H., Choi, K.W., Park, C.S., Lee, S.B., Kang, H.Y., Shon, K.J., Park, J.S., Cha, J., 2005. Substrate specificity and transglycosylation catalyzed by a thermostable β-glucosidase from marine hyperthermophile Thermotoga neapolitana. Appl. Microbiol. Biotechnol. 69, 411-422. https://doi.org/10.1007/s00253-005-0055-1

Parry, N.J., Beever, D.E., Owen, E., Vandenberghe, I., Van Beeumen, J., Bhat, M.K., 2001. Biochemical characterization and mechanism of action of a thermostable β-glucosidase purified from Thermoascus aurantiacus. Biochem. J. 353, 117–127. https://doi.org/10.1042/0264-6021:3530117

Payne, C.M., Knott, B.C., Mayes, H.B., Hansson, H., Himmel, M.E., Sandgren, M., Ståhlberg, J., Beckham, G.T., 2015. Fungal cellulases. Chem. Rev. 115, 1308–1448. https://doi.org/10.1021/cr500351c

Rani, V., Mohanram, S., Tiwaru, R., Nian, L., Arora, A., 2014. β-Glucosidase: Key Enzyme in Determining Efficiency of Cellulase and Biomass Hydrolysis. J. Bioprocess. Biotech. 05. https://doi.org/10.4172/2155-9821.1000197

Rheims, H., Rainey, F.A., Stackebrandt, E., 1996. A molecular approach to search for diversity among bacteria in the environment. J. Ind. Microbiol. Biotechnol. 17, 159–169. https://doi.org/10.1007/BF01574689

Riou, C., Salmon, J.M., Vallier, M.J., Günata, Z., Barre, P., 1998. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β-glucosidase from Aspergillus oryzae. Appl. Environ. Microbiol. 64, 3607–3614. https://doi.org/10.1002/adsc.200700082

Rohini, K., Srikumar, P.S., 2014. Therapeutic Role of Coumarins and Coumarin-Related Compounds. J. thermodynacis Catal. 05, 10–12. https://doi.org/10.4172/2157-7544.1000130

Saha, B.C., Freer, S.N., Bothast, R.J., 1994. Production, Purification, and Properties of a Thermostable β-Glucosidase from a Color Variant Strain of Aureobasidium pullulans. Appl. Environ. Microbiol. 60, 3774–3780.

Sathe, S.S., Soni, S., Ranvir, V.P., Choudhari, V.G., Odaneth, A.A., Lali, A.M., Chandrayan, S.K., 2017. Heterologous expression and biochemical studies of a thermostable glucose tolerant β-glucosidase from Methylococcus capsulatus (bath strain). Int. J. Biol. Macromol. 102, 805-812. https://doi.org/10.1016/j.ijbiomac.2017.04.078

Schröder, C., Elleuche, S., Blank, S., Antranikian, G., 2014. Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. Enzyme Microb. Technol. 57, 48–54. https://doi.org/10.1016/j.enzmictec.2014.01.010

Schwarz, A., Thomsen, M.S., Nidetzky, B., 2009. Enzymatic synthesis of Beta-glucosylglycerol using a continuous-flow microreactor containing thermostable β-glycoside hydrolase celB immobilized on coated microchannel walls. Biotechnol. Bioeng. 103, 865–872. https://doi.org/10.1002/bit.22317

Sheehan, V.M., Ross, P., Fitzgerald, G.F., 2007. Assessing the acid tolerance and the technological robustness of probiotic cultures for fortification in fruit juices. Innov. Food Sci. Emerg. Technol. 8, 279–284. https://doi.org/10.1016/j.ifset.2007.01.007

Singh, G., Verma, A.K., Kumar, V., 2016. Catalytic properties, functional attributes and industrial applications of β-glucosidases. 3 Biotech 6, 1–14. https://doi.org/10.1007/s13205-015-0328-z

Singh, S., Thavamani, P., Megharaj, M., Naidu, R., 2015. Multifarious activities of cellulose degrading bacteria from Koala (Phascolarctos cinereus) faeces. J. Anim. Sci. Technol. 57. https://doi.org/10.1186/s40781-015-0056-2

Singhania, R.R., Patel, A.K., Pandey, A., Ganansounou, E., 2017. Genetic modification: A tool for enhancing beta-glucosidase production for biofuel application. Bioresour. Technol. 245, 1352–1361. https://doi.org/10.1016/j.biortech.2017.05.126

Singhania, R.R., Patel, A.K., Sukumaran, R.K., Larroche, C., Pandey, A., 2013. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. Bioresour. Technol. 127, 500–507. https://doi.org/10.1016/j.biortech.2012.09.012

Sugimoto, H., van Buren, J.P., 1970. REMOVAL FROM SOY MILK BY AN ENZYME. J. Food Sci. 35, 655–660.

Sun, Y., Cheng, J., 2002. Hydrolysis of lignocellulosic materials for ethanol production : a review. Bioresour. Technol. 83, 1–11. https://doi.org/10.1016/S0960-8524(01)00212-7

Swangkeaw, J., Vichitphan, S., Butzke, C.E., Vichitphan, K., 2011. Characterization of β-glucosidases from Hanseniaspora sp. and Pichia anomala with potentially aroma-enhancing capabilities in juice and wine. World J. Microbiol. Biotechnol. 27, 423–430. https://doi.org/10.1007/s11274-010-0474-8

Transparency Market Research, 2013. Salicylic Acid Market for Pharmaceutical , Skin care , Hair care and Other Applications : Global Industry Analysis , Size , Share , Growth , Trends and Forecast 2013 - 2019.

Uchiyama, T., Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr. Opin. Biotechnol. 20, 616–622. https://doi.org/10.1016/j.copbio.2009.09.010

Uchiyama, T., Miyazaki, K., Yaoi, K., 2013. Characterization of a novel β-Glucosidase from a compost microbial metagenome with strong transglycosylation activity. J. Biol. Chem. 288, 18325–18334. https://doi.org/10.1074/jbc.M113.471342

Uchiyama, T., Yaoi, K., Miyazaki, K., 2015. Glucose-tolerant β-glucosidase retrieved from a Kusaya gravy metagenome. Front. Microbiol. 6, 1–9. https://doi.org/10.3389/fmicb.2015.00548

Vartoukian, S.R., Palmer, R.M., Wade, W.G., 2010. Strategies for culture of "unculturable" bacteria. FEMS Microbiol. Lett. 309, 1–7. https://doi.org/10.1111/j.1574-6968.2010.02000.x

Vasella, A., Davies, G.J., Böhm, M., 2002. Glycosidase mechanisms. Curr. Opin. Chem. Biol. 6, 619-629. https://doi.org/10.1016/S1367-5931(02)00380-0

Vocadlo, D.J., Davies, G.J., 2008. Mechanistic insights into glycosidase chemistry. Curr. Opin. Chem. Biol. 12, 539-555. https://doi.org/10.1016/j.cbpa.2008.05.010

Voget, S., Steele, H.L., Streit, W.R., 2006. Characterization of a metagenome-derived halotolerant cellulase. J. Biotechnol. 126, 26–36. https://doi.org/10.1016/j.jbiotec.2006.02.011

Wang, Q., Qian, C., Zhang, X.Z., Liu, N., Yan, X., Zhou, Z., 2012. Characterization of a novel thermostable β-glucosidase from a metagenomic library of termite gut. Enzyme Microb. Technol. 51, 319–324. https://doi.org/10.1016/j.enzmictec.2012.07.015

Wei, D.L., Kirimura, K., Usami, S., Lin, T.H., 1996. Purification and characterization of an extracellular β-glucosidase from the wood-grown fungus Xylaria regalis. Curr. Microbiol. 33, 297–301. https://doi.org/10.1007/s002849900117

Withers, S.G., Street, I.P., Rupitz, K., Kempton, J.B., Warren, A.J., Aebersold, R., 1990. Unequivocal Demonstration of the Involvement of a Glutamate Residue as a Nucleophile in the Mechanism of a "Retaining" Glycosidase. J. Am. Chem. Soc. 112, 5887–5889. https://doi.org/10.1021/ja00171a043

Womack, M.D., Kendall, D. a, MacDonald, R.C., 1983. Detergent effects on enzyme activity and solubilization of lipid bilayer membranes. Biochim. Biophys. Acta. 733, 210–215. https://doi.org/10.1016/0005-2736(83)90524-2

World Health Organization, 2017. WHO Report on the Global Tobacco Epidemic,2017, World Health Organization. https://doi.org/Licence: CC BY-NC-SA 3.0 IGO.

Wu, J., Ge, J., Zhang, Y., Yu, Y., Zhang, X., 2010. Solubility of Genistein in Water, Methanol, Ethanol, Propan-2-ol, 1-Butanol, and Ethyl Acetate from ( 280 to 333 ) J Chem Eng Data. 55, 5286–5288.

Xing, M.N., Zhang, X.Z., Huang, H., 2012. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnol. Adv. 30, 920–929. https://doi.org/10.1016/j.biotechadv.2012.01.021

Xu, X., Song, Z., 2014. Advanced research on anti-tumor effects of amygdalin. J. Cancer Res. Ther. 10, 3-7. https://doi.org/10.4103/0973-1482.139743

Yang, S., Jiang, Z., Yan, Q., Zhu, H., 2008. Characterization of a thermostable extracellular β-glucosidase with activities of exoglucanase and transglycosylation from Paecilomyces thermophila. J. Agric. Food Chem. 56, 602–608. https://doi.org/10.1021/jf072279+

Yang, Y., Zhang, X., Yin, Q., Fang, W., Fang, Z., Wang, X., Zhang, X., Xiao, T., 2015. A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases. Scientific Reports. 5.

Yeo S.L., Shazilah K., Suhaila, S., Abu Bakar, F.D., Murad, A.M.A., 2014. In-silico analysis of Aspergillus niger beta-glucosidases. AIP Confrence Proceedings 1614, 537–543. https://doi.org/10.1063/1.4895257

Yeomans, N.D., 2011. Aspirin: Old drug, new uses and challenges. J. Gastroenterol. Hepatol. 26, 426–431. https://doi.org/10.1111/j.1440-1746.2010.06569.x

Yoon, J.J., Kim, K.Y., Cha, C.J., 2008. Purification and characterization of thermostable β-glucosidase from the brown-rot basidiomycete Fomitopsis palustris grown on microcrystalline cellulose. J. Microbiol. 46, 51–55. https://doi.org/10.1007/s12275-007-0230-4

Yun, J., Kang, S., Park, S., Yoon, H., Kim, M., Heu, S., Ryu, S., 2005. Characterization of a novel amylolytic enzyme encoded by a gene from a soil derived metagenomic library. Appl. Environ. Microbiol. 70, 7229–7235. https://doi.org/10.1128/AEM.70.12.7229

Samad, M.Y.A., Razak, C.N.A., Salleh, A.B., Yunus, W., Ampon, K., Basri, M., 1989. A plate assay for primary screening of lipase activity. J. Microbiol. Methods 9, 51–56. https://doi.org/10.1016/0167-7012(89)90030-4

Zhang, L., Fu, Q., Li, W., Wang, B., Yin, X., Liu, S., Xu, Z., Niu, Q., 2017. Identification and characterization of a novel β-glucosidase via metagenomic analysis of Bursaphelenchus xylophilus and its microbial flora. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-14073-w