Analysis of *Euoniticellus intermedius*, larva gut micro-flora: Potential application in the production of biofuels.

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

By

Munamato Mabhegedhe

Supervisors:

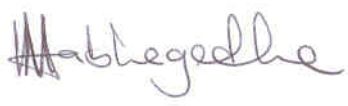
Dr Karl Rumbold and Dr Monde Ntwasa

ABSTRACT

Recent years have seen a dramatic increase in first generation bio-fuel production, mainly driven by concerns of climate change and rising prices of transportation fossil fuels. Due to significant pressure on the few available food sources, second generation bio-fuels have entered the fray, as a sustainable alternative. This research's aim was to search for cellulolytic micro-organisms and enzymes from the gut of the dung beetle, *Euoniticellus intermedius*, (Coleoptera: Scarabaeida) that can be used in the production of second generation bio-fuels. Dung beetle larvae were dissected and the gut micro-flora cultured in cellulose medium. Bacterial growth and cellulase activity was monitored on a daily basis. DNA isolation was then done on the cellulose medium-cultured microbes and the isolated DNA cloned in *E. coli*. The clones were screened for cellulase activity using plate assays. A total of 7 colonies out of 160 screened colonies showed positive CMC (endo-β-1,4-glucanase) and MUC (cellobiohydrolase) activities. Sequencing of these positive colonies yielded mostly bacteria belonging to the *Enterobacteriaceae* family, most of which have not been previously reported to have cellulase activity. This study's findings prove that in addition to this dung beetle's gut being a fruitful source of microbial biodiversity, it is also a potential source of cellulolytic micro-organisms and enzyme activities that will aid the function and design of future bioreactors for the bio-fuel industry.

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.



Munamato Mabhegedhe

11 May 2012

ACKNOWLEDGEMENTS

I start by thanking my supervisors, Dr Karl Rumbold and Dr Monde Ntwasa, for affording me the opportunity to enrol at this great institution and for their continued support and guidance throughout the course of my MSc journey. Secondly, I would like to thank my research advisor, Dr Chamunorwa Alouis Togo, for being the big brother who was always there to pick me up whenever I stumbled. To my fellow research students, thank you very much for being there for me and for showing me the ropes when I needed help. Special mention goes to Maabo Moralo, and Rodney Hull for their invaluable support and contribution towards this MSc research project. I will forever be indebted to you guys!

To the 2011 MCB Postgraduate Committee members, I have learnt a lot during our term of office, thanks to each and every one of you. We have had good times, as well as challenging moments. We might not have achieved everything we had set for ourselves as a team, but I salute you for sacrificing your research time and the selflessness you showed in serving your school and fellow students for no financial gain. I am also grateful to the 2011/2012 Post-Graduate Association committee, for affording me the opportunity to represent the Wits postgraduate community in the Student Representative Council, Senate and Council, to mention a few. Through these leadership experiences, I have matured into a better and more progressive leader than I was when I first walked into Wits on the 23rd of February 2010.

This dream wouldn't have come true without the financial assistance from the Postgraduate Merit Award (PMA) and the Bradlow Scholarship. Special acknowledgements go to the African Union Commission for awarding me the 2011 Mwalimu Nyerere African Union Scholarship (MNAUSS) against stiff competition from other equally exceptional candidates from all over Africa.

A big thank you goes to my big brother, Ennie, for always being there for me in every possible way. More so to my mother, for ushering me into this world and giving me her unconditional love and support, even during trying times. I dedicate this work to the queen of my heart, Esther Vushe, and my unborn child. You give me hope and bring a new meaning to my life. Above all, I give glory to the Almighty Lord for guiding me throughout this endeavour.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

FREQUENTLY USED ABBREVIATIONS

ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
C5 sugars	Sugars containing 5 carbon atoms
C6 sugars	Sugars containing 6 carbon atoms
$CaCl_2$	Calcium chloride
cDNA	Complementary deoxyribonucleic acid
CMC	Carboxymethylcellulose
CO_2	Carbon dioxide
$CoCl_2$	Cobalt chloride
$CuCl_2$	Copper chloride
cm	Centimetre
DI	Deionized water
DNA	Deoxyribonucleic acid
DNS	3,5-Dinitrosalicyclic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agriculture Organisation (of the United Nations)
$FeCl_2$	Ferrous chloride
g	Gram
GHG	Greenhouse gas
H_3BO_3	Boric acid
HCl	Hydrogen chloride
hr	Hour
IPTG	Isopropyl-β-D-thiogalactoside
kb	Kilobase
KCl	Potassium chloride
KH_2PO_4	Potassium hydrogen phosphate
LB	Luria-Bertani Broth
µg	Microgram
µl	Microlitre
M	Molar

M1A	Cellulose media
M1B	Non-cellulose media
mg	Milligram
$MgCl_2$	Magnesium chloride
ml	Millilitre
mm	Millimetre
mM	Millimolar
$MnCl_2$	Manganese chloride
MUC	4-Methylumbellifery-β-D-cellobioside
Mw	Molecular weight
NA	Not applicable
NaCl	Sodium chloride
Na_2MoO_4	Sodium molybdate
NaOH	Sodium hydroxide
Na_2WO_4	Sodium tungstate
$NaSeO_3$	Sodium selenite
$NaSO_4$	Sodium sulphate
NH_4Cl	Ammonium chloride
$NiCl_2$	Nickel chloride
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulphate
sec	Second
TAE	Tris base, acetic acid and EDTA
TE	Tris-EDTA
Tris	Tris (hydroxymethyl)-aminomethane
UV	Ultra Violet
$ZnCl_2$	Zinc chloride

CHAPTER 1

Literature review

1.1 Energy security and climate change

"Energy is the single most valuable commodity of modern society; it is the engine that makes our civilisation go". These are the words of M. Sanjayan, the host of "*Powering the Future*": a science programme on *Discovery* channel.

Global energy demand has been on a constant rise and has increased more than twenty-fold in the last century (Jegannathan *et al.*, 2009), mainly due to the demands of a growing world population and also to power the rapidly growing economies of the developing nations. Most of our energy requirements have been historically met, thanks to the previously abundant, reliable and cheap fossil derived fuels such as coal, petroleum oil and natural gas. Even today, fossil fuels still play a significant role in driving our vehicles and powering our homes and industries (Jegannathan *et al.*, 2009), but due to their finite nature (Adsul *et al.*, 2011; Jegannathan *et al.*, 2009) and the high rate at which we are using them, it is only a matter of time before we run out of them. As the fossil fuels supplies dwindle and our energy demand continue to rise, the price of fuel will subsequently continue to rise astronomically, thereby putting a lot of economic pressure on the oil-importing countries.

The skewed distribution of natural oil resources has also historically led to political vulnerability and economic exploitation not only of the non-oil producing countries by the oil-rich nations, but also of the oil-rich nations by the world super powers in a bid to gain access to this wonderful natural resource. The recent political developments in Iraq and Libya need no further introduction, and are a clear testimony of the political dangers faced by oil-rich nations on a regular basis. This poses a multitude of political threats on the oil-producing countries' national security and economic livelihood.

In addition to the above drawbacks, the continued use of fossil fuels as the major energy source has been widely acknowledged to aggravate the perils of global warming and climate change by releasing the largest proportion of carbon dioxide (CO_2) gas into the atmosphere. Carbon dioxide has been identified as the main greenhouse gas (GHG) in the atmosphere (Adsul *et al.*, 2011; Balat and Balat, 2009; Jegannathan *et al.*, 2009). The effects of global warming and climate change have damaging environmental consequences, and also pose disastrous global risks for life on earth (Jegannathan, *et al.*, 2009).

1.2 Bio-energy

Bio-energy is a type of renewable energy derived from raw materials of biological origin. It comprises of electricity, heat, and a wide range of transportation fuels. In 2010, bio-energy contributed approximately 14% of the total global energy supply (FAO and UNEP, 2010).

Bio-energy systems are identified by two competing paradigms which co-exist within the energy supply chain. The first one is the traditional extractive biomass system, a pattern which has been historically used for many generations to meet domestic and industrial energy demands. The traditional extractive biomass system encompasses traditional firewood burning and charcoal production and is therefore regarded as a low productive and less efficient system (FAO, 2009). According to Cotula *et al.*, (2008), approximately 52% of the developing world and nearly 80% of African countries still rely on the traditional extractive biomass system to cater for most of their energy requirements. The second bio-energy paradigm is the innovative modern system, a system where commercial bioenergy production is done using efficient and environmentally sound technologies. This innovative modern system includes the production of bio-fuels such as bio-ethanol from sugar cane, corn grain or cellulosic biomass and bio-diesel from oil seeds, palm oil or tallow, using more efficient conversion technologies (Cotula *et al.*, 2008).

1.2.1 Bio-fuels

Bio-fuels have been defined by Balat and Balat, (2009) as liquid or gaseous fuels derived from plant matter or residues. Recent years have seen a dramatic increase in bio-fuel production, mainly driven by concerns of climate change and rising prices of transportation fossil fuels, as described above. The prospect of economic development such as employment creation, infrastructure development and additional government revenue from bio-fuel exports has also significantly contributed to the above trend. The need to ensure locally produced energy security has also boosted bio-fuel production since it promises massive cost reduction in the importation of fuel from the oil-rich nations (Cotula *et al.*, 2008). In addition, bio-fuels have the potential to reduce political exploitation of the non-oil producing countries by the oil-rich ones and of the oil-rich countries by the world super powers, thus ensuring political and economic stability.

In terms of modern bio-energy; bio-gas, bio-diesel and bio-ethanol have been cited as the three main bio-fuels being produced in the world today (Yuan *et al.*, 2008).

Biogas is the third modern bio-fuel choice worldwide (Yuan *et al.*, 2008). Traditionally, biogas was obtained by a complex degradation process of organic matter by a set of bacteria under anaerobic conditions. The resulting biogas consisted primarily of methane (CH_4) and CO_2 together with smaller amounts of other gases and vapours (Makaruk *et al.*, 2010). A variety of organic materials ranging from sewage treatment plants, to organic waste utilization in landfill sites, to farm biogas production that utilize pure energy crops, have been successfully employed in bio-methane production. Besides the traditional biogas, methane, bio-hydrogen production by anaerobic processes using waste organic substrates has also been documented (Chen *et al.*, 2009). Lately, production of hydrogen by green algae and microbes has been proposed as a potential source for a third generation biogas (Yuan *et al.*, 2008). Unlike hydrogen production from other biomass sources, algae-based hydrogen production uses a biological water-splitting reaction, in which a hydrogenase enzyme uses the photosynthetic electron transport chain to reduce protons resulting in hydrogen production (Yuan *et al.*, 2008). Production of biogas promises an environmentally less damaging way of obtaining energy and is of major importance for the sustainable use of agrarian biomass as a renewable energy source (Oslaj *et al.*, 2010).

Bio-diesel is produced by the trans-esterification of oils with short-chain alcohols or by the esterification of fatty acids. The trans-esterification reaction consists of the transformation of tri-glycerides into fatty acid alkyl esters, in the presence of an alcohol (such as methanol or ethanol) and a catalyst (an alkali or acid), with glycerol as a by-product (Vasudevan and Briggs, 2008). Oils and fatty acids derived from plant seeds, algae or other biological sources such as animal renderings are trans-esterified for the removal of glycerol as described above and mixed with diesel to form bio-diesel (Yuan *et al.*, 2008). A variety of plant species currently employed in bio-diesel production include soybean, rapeseed, canola, sunflower and palm. Other potential choices for bio-diesel production include using terpenoid products from *Copaifera* species (‘diesel tree’) as biodiesel directly, or engineering the plant terpenoid pathway to produce large amounts of sesquiter-penes and diterpenes. Bio-diesel has attracted attention during the past few years as a renewable and environmentally friendly fuel and it already has a niche in the current transportation fuel system (Vasudevan and Briggs, 2008, Yuan *et al.*, 2008).

Bio-ethanol is an ethyl alcohol (C_2H_5OH) (Balat and Balat, 2009), which is produced from a range of agricultural products such as corn starch (USA), sucrose (Brazil), molasses (India), a variety of sugar rich crops and also lignocellulosic biomass (Adsul *et al.*, 2011). Ethanol

produced from the fermentation of glucose and sucrose from sugarcane (*Saccharum officinarum*), sugar beet (*Beta vulgaris*) or maize (*Zea mays*), is classified under first generation bio-fuels (Balat and Balat, 2009; Yuan *et al.*, 2008). Starch and sugar derived ethanol accounts for more than 10 billion gallons of annual total bio-ethanol output and therefore represent the most conventional and technically feasible option for bio-energy production. However, this method has been heavily criticised as unsustainable, since it creates competition between energy and food supplies. As such, bio-ethanol production from lignocellulosic biomass, regarded as a second generation bio-fuel, has been hailed as the best sustainable choice that would ease the pressure on the food versus fuel debate. Despite offering high hopes, bio-ethanol from lignocellulosic materials is not yet economically feasible at an industrial scale due to the processing cost associated with the recalcitrant nature of lignocellulose matter (Adsul *et al.*, 2011; Margeot *et al.*, 2009). The use of bio-ethanol blended fuel for automobiles can significantly reduce petroleum use and exhaust greenhouse gas emissions (Balat and Balat, 2009). Today, bio-ethanol is the most used non-fossil alternative engine fuel in the world (Demirbas, 2009).

Table 1: Comparison of different bio-ethanol and bio-diesel platforms (adopted fromYuan *et al.*, 2008)

Platforms	**Feedstock**	**NEB**	**NER**	**CO_2 Balance**	**EB**
First generation bio-ethanol	Maize	10-80	1.5-3.0	Positive	+
	Sugarcane	55-80	3.0-4.0	Positive	+
	Sugar beet	40-100	2.5-3.5	Positive	+
	Sweet sorghum	85-300	5-10	Positive	++
Second generation bio-ethanol	Miscanthus	250-550	15-70	Likely negative	+++
	Switchgrass	150-500	10-50	Likely negative	+++
	Poplar	150-250	10-20	Likely negative	+++
Bio-diesel	Soya bean	-20-10	0.2-0.6	Positive	+
	Canola	-5-2	0.7-0.1	Positive	+
	Sunflower	-10-0	0.3-0.9	Positive	+

Abbreviations: NEB – Net energy balance; NER – Net energy ratio; EB – Ecological benefits. Favourable features are indicated by + symbols with +++ being the most favourable.

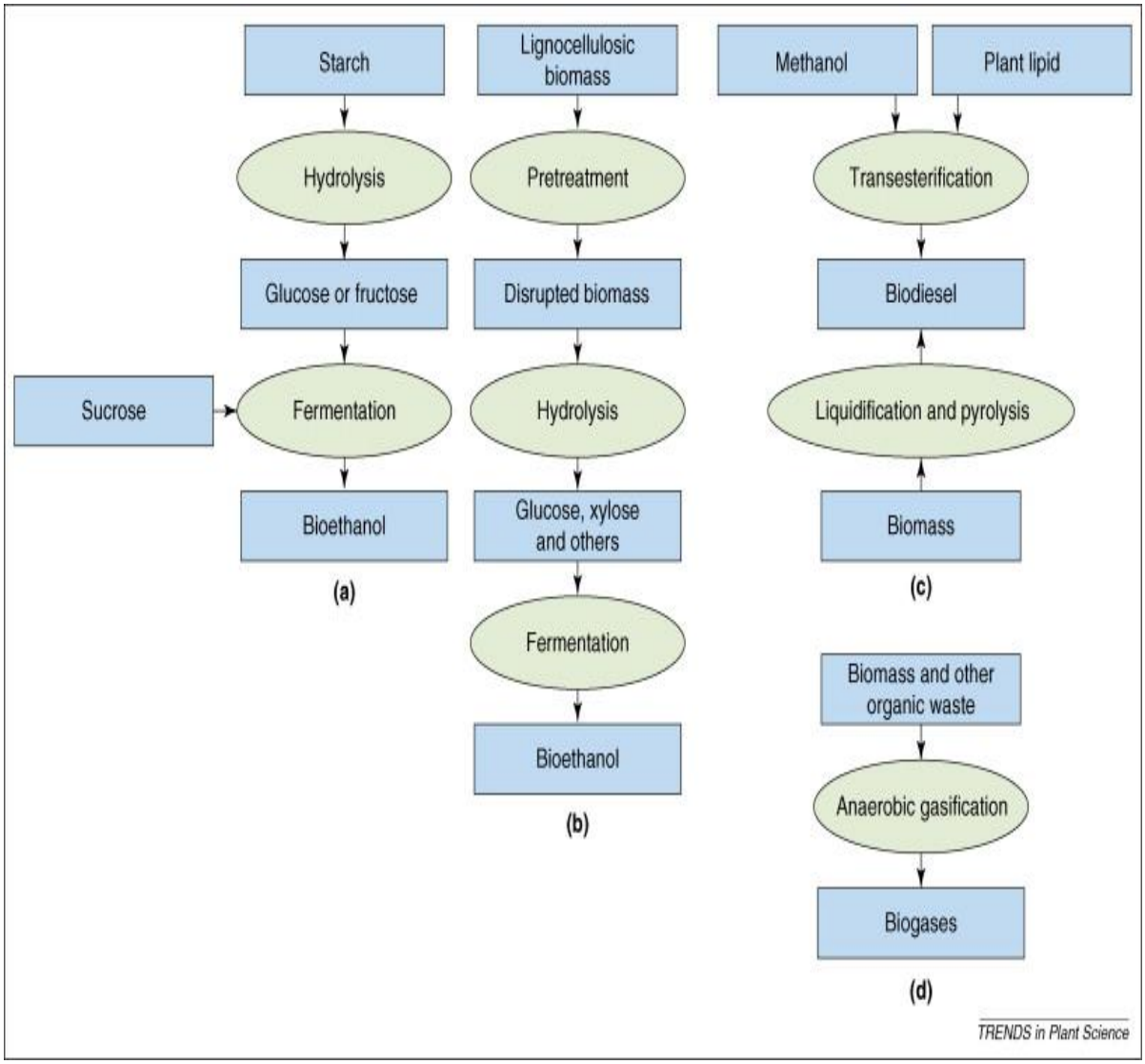


Figure 1: Processing platforms for producing different bio-fuels with various feed stocks as adopted from Yuan *et al.*, (2008). **(a)** Starch and sugar-based first generation bio-ethanol production. In this platform, starch is first hydrolysed to reducing sugars, which are subsequently fermented into bio-ethanol. Sucrose, on the other hand is directly fermented to bio-ethanol. **(b)** Second generation, lignocellulose bio-ethanol production. Lignocellulose ethanol production involves pre-treatment of biomass material to remove lignin and reduce recalcitrant. This is followed by hydrolysis of the disrupted cellulose and hemicelluloses to form reducing sugars, which are undergo fermentation to produce bio-ethanol. **(c)** Bio-diesel is often produced by the trans-esterification of fatty acids. In addition, bio-diesel can be derived from the liquefaction and pyrolysis of biomass. **(d)** Biomass gasification. Biomass can be used to produce methanol, carbon monoxide, hydrogen or other gases through a process of gasification.

1.2.2 Biomass for bio-fuels production

Biomass is regarded as the most common form of carbonaceous materials, and is reportedly widely used through the traditional extractive biomass system in the third world to meet their energy requirements (Demirbas, 2009). The most important biomass energy sources are wood and wood wastes, agricultural crops and their waste by-products. Agricultural residues which have been investigated in biofuel production include cotton stalks, wheat and rice straw, coconut shells, maize and rice husks. They are many different types of biomass crops grown for the sole purpose of energy production. Crops that have been used for bio-fuel production include sugar cane, corn, sugar beets, grains, elephant grass, kelp (seaweed) and many others Second generation bio-fuel production from lignocellulosic biomass is considered the best option to reduce the heavy dependence on fossil fuels. In comparison with their first generation bio-fuel counterparts, biomass derived bio-fuels have the largest potential to meet the current and future energy requirements of planet Earth (Demirbas *et al.*, 2009).

Lignocellulose biomass consists of three types of polymers namely cellulose, hemicellulose, and lignin (Adsul *et al.*, 2011; Lee *et al.*, 2008). Cellulose, the most abundant carbohydrate polymer, is a linear glucose polymer composed of repeating cellobiose units linked by β-1,4-glucosidic bonds (Feng *et al.*, 2007; Zhou *et al.*, 2007). It mainly exists in plant and humic matter and has strong physical and chemical interactions with hemicelluloses and lignin. According to Adsul *et al.*, 2011, native cellulose has about 10,000 glycosyl units in the cellulose chain that form fibrils which are stabilized by strong intermolecular hydrogen bonds between hydroxyl groups of the adjacent molecules. Cellulosic materials have crystalline domains separated by less ordered, amorphous regions. These amorphous regions are the potential points for chemical and enzymatic attacks. The crystalline cellulose is highly resistant to chemical and enzymatic hydrolysis due to its structure in which chains of cellodextrins are precisely arranged (Adsul *et al.*, 2011).

Hemicellulose is a short and highly branched hetero-polymer comprising of mainly xylose, and glucose, mannose, galactose and arabinose and in some cases uronic acids (Adsul *et al.*, 2011). Hemicelluloses are normally categorised as mannans, xylans or galactans according to the predominant type of sugars present in the polymer. The C5 and C6 sugars are joined together by 1,3; 1,6 and 1,4 glycosidic bonds and are often acetylated to form a loose, very hydrophilic structure that glues together cellulose and lignin polymers (Adsul *et al.*, 2011).

Lignin comprises of phenyl-propanoid units derived from the corresponding p-hydroxycinnapyl alcohols. These phenyl-propanoid units are made up of dimethoxylated (syringyl), monomethoxylated (guaiacyl) and non-methoxylated (p-hydroxyphenil) alcohols (Adsul *et al.*, 2011). Lignin is a hydrophobic polymer, which is highly resistant to chemical and biological degradation. It is found in the middle lamella, where it acts as cement that binds together the plant cells. It is also located in the layers of the cell walls where together with hemicelluloses, it forms an amorphous matrix in which cellulose fibrils are embedded and protected against bio-degradation. This association of cellulose, hemicelluloses and lignin imparts the plant cell wall with resistance to both mechanical and biological degradations (Adsul *et al.*, 2011).

1.2.3 Biochemical conversion of lignocellulose to bio-fuels

Lignocellulose biomass can be converted to biofuels by thermo-chemical or biochemical processing (Ruane *et al.*, 2010). The biochemical production of ethanol from forest and agricultural residues, or dedicated lignocellulose crops is still hampered by economic and technical obstacles (Margeot *et al.*, 2009). There are two key parts involved in bio-ethanol production from lignocellulose biomass. Firstly, the cellulose and hemicellulose portions of the biomass must be broken down into simple sugars. This is a major challenge due to the recalcitrant nature of lignocellulose, as such a variety of thermal, chemical and biochemical methods are still being developed to carry out this saccharification step in an efficient and low-cost manner. Secondly, the released sugars must be fermented to make bio-ethanol. The yielded sugars, however, are a complex mixture of 5-carbon and 6-carbon sugars and this provides a greater challenge for complete fermentation into bio-ethanol (Ruane *et al.*, 2010).

The pre-treatment, saccharification, fermentation and distillation steps required for the production of second generation fuels from lignocellulosic biomass renders the overall bio-ethanol production process complex and costly (Margeot *et al.*, 2009; Ruane *et al.*, 2010). In order to make cellulose hydrolysis and fermentation economically feasible, it is important to identify methods to increase enzyme effectiveness and to overcome the barriers of enzymatic hydrolysis and the subsequent fermentation of the released sugars (Lee *et al.*, 2008). There is need to replace the expensive pre-treatment steps with cheaper alternatives and also give more focus on simultaneous saccharification and fermentation in order to effectively decrease the cost of lignocellulosic ethanol production (Yuan *et al.*, 2008).

1.3 Cellulases

Cellulases belong to the glycosyl hydrolase family, which has got approximately more than 100 family units organised into 14 clans (Voget *et al*., 2006). Unlike other glycosyl hydrolases, cellulases have the ability to hydrolyze the β-1,4 glucosidic linkages between glycosyl residues found in cellulose (Voget *et al*., 2006; Zhou *et al.*, 2007). The enzymatic degradation of the cellulose's β-1,4 glucosidic bonds occurs through an acid hydrolysis mechanism through the use of a proton donor and a nucleophile (proton acceptor). The hydrolysis process results in either the inversion or retention (double placement) of the anomeric configuration of carbon-1 at the reducing end (Voget *et al*., 2006).

1.3.1 Categories of cellulases

The degree and extend of cellulose degradation varies considerably in different organisms due to the nature or state of the cellulosic substrate and the type of glycosidases produced by the organism (Baird *et al*., 1990). In general, cellulolytic micro-organisms produce multiple enzyme components to degrade cellulose, known as the cellulase system (Liu *et al.*, 2004). The cellulase system consists of three major categories of enzymes, based on their structural properties (Liu *et al.*, 2004; Saul *et al.*, 1990). These are endoglucanases (EC 3.2.1.4) (also known as β-1,4-D-glucan-4-glucanohydrolase), exoglucanases which include cellodextrinases (EC 3.2.1.74) (also referred to as 1,4-β-D-glucan glucanohydrolases) and cellobiohydrolases (EC 3.2.1.91) (also referred to as β-1,4-D-glucan cellobiohydrolase), and lastly β -glucosidases (EC 3.2.1.21) (also known as β-1, 4-D-glucoside glucohydrolases).

Endoglucanases randomly cleave β-1,4-glycosidic bonds on cellulose chains away from chain ends. Exoglucanases such as cellobiohydrolases attack cellulose from chain ends (Cel7A (cellobiohydrolase I), acts from the reducing ends, and Cel6A (cellobiohydrolase II) acts from the non-reducing ends of the cellulose chains) to produce cellobiose and various oligosaccharides (Bansal *et al*., 2009). On the other hand, β-glucosidase hydrolyses the β-glucosidic linkages of the resultant cellobiose and oligosaccharides to form glucose and a shorter oligosaccharide or aglycone (Bansal *et al*., 2009; Jiang *et al*., 2010; Saul *et al*., 1990). The β-glucosidase is considered the rate limiting enzyme in the cellulose hydrolysing system (Jiang *et al*., 2010). This biologically mediated degradation of cellulose is achieved through the synergistical action of the three above-mentioned cellulase enzyme categories (Duan *et al.*, 2009; Feng *et al.*, 2007; Zhou *et al.*, 2007).

It has been proposed that the enzymatic degradation of cellulose has got more steps than classical enzyme kinetics due to the heterogeneous nature of cellulose. Classical homogeneous enzyme catalysis is modelled by the Michaelis–Menten kinetics, whereas heterogeneous catalysis is modelled by Langmuir–Hinshelwood kinetics, on a catalyst support (Bansal *et al*., 2009). The cellulase degradation of insoluble lignocellulose is a combination of the above two kinetics and the process is affected by additional factors, such as product inhibition, enzyme deactivation, substrate crystallinity, substrate accessibility changes, substrate reactivity changes, fractal nature of the reaction, changes in enzyme synergism and lignin inhibition. These factors consequently contribute to the reduction of cellulose conversion rates (Bansal *et al*., 2009).

The steps involved during this hydrolysis as proposed by Bansal and co-workers (2009) include the adsorption of the cellulase enzyme onto the cellulose substrate via the substrate binding domain. This is followed by the identification of the bond susceptible to hydrolysis on the substrate surface and the formation of the enzyme-substrate complex. The hydrolysis of the β-1,4-glycosidic bond in the cellulose is then achieved by the simultaneous forward movement of the cellulase enzyme along the cellulose substrate. The main hydrolysis products of this reaction are cellobiose and other oligosaccharides. Desorption of the enzyme from the substrate then occurs (a repeat of the identification of the bond susceptible to hydrolysis and the cellulose degradation steps occur if only the catalytic domain detaches from the cellulose chain). Lastly, the produced cellobiose and the other oligosaccharides by-products are hydrolysed to glucose by the action of β-glucosidases.

Highly efficient fungi and bacteria have been demonstrated to produce at least one cellulose degrading enzyme from each of the three cellulase enzyme categories (endoglucanases, exoglucanases or β-glucosidases) (Baird *et al*., 1990). According to Lynd and co-workers (2008), the major improvements attained in the enzymatic hydrolysis of cellulosic biomass to ethanol can now allow lignocellulose derived bio-ethanol to economically compete with first generation bio-ethanol and petroleum derived ethanol. However, challenges such as decreasing rate of hydrolysis due to end-product inhibition, high cellulase enzyme costs, as well as little understanding of the cellulase kinetics on lignocellulosic substrates continue to hinder the successful development of second generation bio-ethanol production (Bansal *et al*., 2009). Bansal and colleagues have come up with various mathematical models to optimize the enzymatic cellulose hydrolysis process and to alleviate the bottlenecks encountered.

1.3.2 Cellulolytic systems

Despite many cellulases having been isolated from fungal sources, there has been a reported increase in the discovery of a wide range of novel cellulases from prokaryotic organisms (Voget *et al*., 2006). Two types of cellulolytic systems have been identified in nature, complexed and non-complexed cellulases. The non-complexed cellulases act discretely and bind directly to the cellulose during enzymatic degradation. These cellulases have a modular structure with non-catalytic carbohydrate binding domains (CBD) that are connected to the catalytic domains by flexible linkers. The carbohydrate binding domains play a major role in the hydrolysis of insoluble cellulose by binding the enzyme to its substrate. The non-complexed cellulases are commonly produced by aerobic fungi and bacteria (Ding *et al*., 2008). In addition, cellulases lacking clearly defined carbohydrate binding domains have also been reportedly identified. These cellulases are non-modular and show poor hydrolytic activity towards insoluble cellulose substrates (Voget *et al*., 2006).

Complexed cellulolytic enzymes, also referred to as cellulosomes, on the other hand, comprises of many cellulases organised on a non-catalytic scaffolding protein that mediates the binding of the enzyme to the cellulose substrate during hydrolysis. The cellulosome is normally produced by anaerobic bacteria (Ding *et al.*, 2008; Voget *et al*., 2006). Cellulases have been discovered through metagenomic and genomic approaches to be produced by a wide range of organisms, including bacteria, yeasts, plants and animals (Feng *et al.*, 2007; Ruane *et al.*, 2010).

1.4 Metagenomics

Soil, water and other environmental samples present a large reservoir of microorganisms and microbial products that could be harnessed to revolutionise the productivity of green biotechnologies. To date, most microbial products have been obtained from microbes that have been isolated and exploited in the laboratory (Singh, 2010). Recent advances in the field of ‘metagenomics’ have dramatically revised our view of microbial biodiversity and its potential for biotechnological applications. It has been widely estimated that up to 99% of microorganisms present in most natural environments are non-culturable, and as a result, very little is known about their genomes, genes and encoded enzymatic activities (Cowan *et al.*, 2005; Ferrer *et al.*, 2005; Singh, 2010; Streit and Schmitz, 2004).

Additionally, many culture-dependent approaches currently used to explore the diversity and potential of microbial communities are biased because of the limitations of cultivation methods (Streit and Schmitz, 2004). Several DNA-based molecular methods based on 16S rRNA gene analysis have been developed to overcome the difficulties and limitations associated with cultivation techniques. However, despite providing extensive information about the taxa and species present in an environment these data usually provide little, if any information at all about the functional role of the different microbes within that environment and the genetic information they contain (Streit and Schmitz, 2004).

The recent development of metagenomic technologies has given us access to a wealth of genetic information through environmental nucleic acid extraction. This has ultimately, provided a means of avoiding the limitations of culture-dependent genetic exploitation (Cowan *et al.*, 2005). The isolation, archiving and analysis of environmental DNA has enabled us to mine microbial diversity, allowing us to access their genomes, identify protein coding sequences and even to reconstruct biochemical pathways, providing insights into the properties and functions of these organisms (Ferrer *et al.*, 2005). This innovative technology is anticipated to accelerate the discovery process of novel useful genes from microorganisms. Currently, various novel enzymes such as nitrile hydratases, cellulases and lipases have already been identified (Singh, 2010; Uchiyama and Miyazaki, 2009).

Metagenomics provides two complimentary approaches in the search for biological products based on either mining for the genetic sequences (sequence-based approach) or the functional screening of clones (Singh, 2010; Uchiyama and Miyazaki, 2009).

In the sequence-based approach, the metagenomic composition is known and the search for a particular function or protein can be performed by mining the sequence data. Once putative homologues are found, the exact sequence information is obtained by either PCR-based or hybridization-based procedures and then expressed in surrogate organisms. Nonetheless, this approach has a major drawback in that the genes obtained through this approach are limited to those having homologies to the probe sequence. Therefore, it would fail to identify novel enzymes that have the same function, but a different structure than known enzymes (Singh, 2010; Uchiyama and Miyazaki, 2009). Despite the limitations of the sequence-based approach, important enzymes such as cellulases, chitinases, carboxypeptidases and lipases have been discovered using it (Singh, 2010).

The functional screening of clones constitutes a function-based assay, in which surrogate organisms are tested for a particular activity, such as reactions catalysed by particular enzymes, or properties attributed to a particular metabolite (Singh, 2010). Function-based screening is a straightforward way to obtain genes that have desired functions, but is often problematic, primarily due to the biased and insufficient expression of anonymous genomic fragments in *Escherichia coli*. They lack genes for homologous recombination (*recA, recBC*) and restriction (*mcrA, mcrBC*), which are useful for cloning variously modified foreign DNA into *E. coli* (Uchiyama and Miyazaki, 2009). The other major problems of this approach are the logistics and the facilities required to screen tens of thousands and up to millions of clones for the desired functions (Singh, 2010).

Despite the limitations of the function-based approach, promising advances have been successfully made in metagenomic screening in the last few years. These include the so called 'substrate-induced gene expression' (SIGEX) technology (Singh, 2010), which selects clones with particular catabolic genes induced by various substrates in concert with fluorescence activated cell sorting (FACS). A further improvement has been the development of a laser-based high throughput screening method, which is claimed to be able to screen one billion clones per day (Singh, 2010).

1.4.1 DNA extraction, library construction, screening

Environmental DNA isolation is usually the first step in metagenomic analysis. A major difficulty associated with this step is the contamination of purified DNA with polyphenolic compounds, which are normally co-purified with the metagenomic DNA. The polyphenolic compounds are thought to interfere with the enzymatic modifications of the isolated DNA and as a result, the construction of environmentally derived DNA libraries with large inserts is often hindered due to the compromised quality of the isolated DNA (Streit and Schmitz, 2004). The above challenges associated with the construction of libraries directly derived from environmental DNA samples have resulted in many laboratories isolating DNA from the metagenomes of laboratory-cultured microbial communities. Despite the fact that laboratory enrichment cultures bear only a limited biodiversity, this technique has been reported to be highly efficient for the rapid isolation of large DNA fragments and cloning of genes with high biotechnological value (Streit and Schmitz, 2004).

DNA isolation and purification is then followed by the construction of DNA libraries in some suitable cloning vectors and host strains. The old approach allowed the construction of small insert libraries, approximately 10 kb or smaller, in a standard sequencing vector, using *Escherichia coli* as a host strain. However, small insert libraries do not allow detection of large gene clusters or operons. To circumvent this limitation, most researchers have been and are still employing large insert libraries, such as cosmid DNA libraries (mostly in the pWE15 vector of Stratagene) with insert sizes ranging from 25-35 kb and/or bacterial artificial chromosome (BAC) libraries with insert sizes up to almost 200 kb (Streit and Schmitz, 2004).

E. coli is still the dominant screening host for functional metagenomics and most commercially available large insert library production systems utilize *E. coli* as a replication host. As such, numerous *E. coli* strains have been engineered that allow for stable replication of various single-copy or multi-copy vectors, minimal recombination and protection from lytic phage. The low copy number status of cosmids or BACs within *E. coli* promotes stable replication and minimizes overexpression of toxic genes. However, the very same attributes that make large insert clones useful for genomic sequencing can become liabilities when searching for function within a metagenomic library. There are a number of genes that cannot be expressed in *E. coli* owing to its differential transcriptional, translational or post-translational controls including, but not limited to, promoter recognition and initiation factors, codon usage, ribosomal entry and protein folding (Taupp *et al.*, 2011).

Functional metagenomic searches for novel genes in metagenomic libraries have been historically performed using highly sophisticated picking and pipetting robots (Streit and Schmitz, 2004). Even today, metagenomic screens for hydrolytic activity still dominate literature and this trend is projected to increase with the renewed interest in the development of improved biomass conversion processes. Indeed, robust and scalable assays for cellulase activity were developed more than three decades ago (Taupp *et al.*, 2011).

Enzyme activities are usually assayed on agar plates supplemented with a specific substrate of interest. By cultivating a metagenomic library on the plates with the substrate, one can identify positive clones through visual screening for the appearance of a clear zone (halo) or colour (Uchiyama and Miyazaki, 2009). The development of novel substrates and detection methods along with increased screening throughput remain the primary activation barriers to more efficient and variable screens. The use of colorimetric or fluorometric substrates in enzyme activity assays allows for optical detection of positive clones (Taupp *et al.*, 2011).

1.4.2 Metagenomes of cellulase producing environments

Research by many scientists has shown that cellulases exist in various natural environments such as water and soil, and are also produced by a wide range of organisms such as microbes, plants, insect and animals (Wang *et al.*, 2009; Willis *et al.*, 2011; Zahura *et al.*, 2011**;** Zhang *et al.*, 2011). The digestive tracts of herbivores and some insects reportedly harbour symbiotic microorganisms that help the hosts to digest the cellulosic feed (Feng *et al.*, 2007).

1.4.2.1 Soil metagenomes

Soil contains diverse microbial species and their metagenomes are a rich source of a wide variety of biocatalysts (Lämmle *et al.*, 2007). Jiang and colleagues successfully identified two novel β-glucosidase genes designated as *bg*l1D and *bg*l1E from a metagenomic library of alkaline-polluted soil samples (Jiang *et al.*, 2011). The *bg*l1D gene reportedly encodes 172 amino acids and the *bg*l1E gene 151 amino acids. Functional analysis of the encoded proteins (Bgl1D and Bgl1E) showed hydrolytic activity towards D-glucosyl-β-(1-4)-D-glucose by degrading it to its glucose monomers. Of the two encoded proteins, Bgl1D showed remarkable activity at low temperatures and across a broad pH range, thus making it more suitable for industrial applications as compared to Bgl1E (Jiang *et al.*, 2011).

A soil metagenome derived-cellulase (endoglucanase), designated as Cel5A, has also been isolated and biochemically characterised by Voget *et al.*, (2006). The enzyme is a single domain cellulase with no separate cellulose binding domain. The gene encoding this enzyme has an open reading frame of 1092 base pairs and is reportedly secreted by *Cellvibrio mixtus* bacterial species. Like most cellulases lacking the catalytic binding domain, the Cel5A enzyme is more active towards soluble cellulose substrates, as compared to their insoluble counterparts. It is considered highly stable since it can retain hydrolytic activity at high temperatures of up to 40°C and a wide pH range (pH 5.5-9). This cellulose degrading enzyme is tolerant to high salt concentrations and remains hydrolytically active even in the presence of diverse divalent cations, detergents or EDTA. What makes Cel5A unique is that it was the first cellulase from a non-extremophile environment to have shown great stability in extreme or harsh conditions. As a result, it has been reported to have great potential of being successfully utilized in industrial applications for bio-fuel and other green chemicals production (Voget *et al.*, 2006).

Endoglucanase and β-glucosidase activity was also identified from a soil-derived metagenome by Wang *et al.*, (2009). Functional screening showed one clone (SC1) showing endoglucanase activity and another one (Sβ1) exhibiting β-glucosidase activity. The endoglucanase enzyme was deduced to belong to glycolsyl hydrolase family 5 (GHF5) and showed 72% amino acid homology to the glycohydrolases derived from buffalo rumen metagenomes. Optimum activity for the endoglucanase enzyme was observed at 55°C and pH 6. The enzyme showed stability over a wide pH range (6-9) and was reported to be tolerant to temperatures as high as 60°C. The β-glucosidase protein on the other hand, was reported to comprise of 761 amino acids and classified to the glycolsyl hydrolase family 3 (GHF3) (Wang *et al.*, 2009).

Table 2: Sequence analyses of soil-derived cellulase genes (adopted from Wang *et al*., 2009)

Active clone	**Accession number**	**Most homologous protein**	**Amino acid identity/ similarity**	**Conserved domains**
SC1	EU282866	Cellulase (ACA61149) from metagenomes of buffalo rumen	72/81	Cellulase (GHF5): aa 25–341
Sβ1	EU292745	β-Glucosidase (ABX76049)	83/91	β-Glucosidase (GHF3): aa 30–252; (GHF3C): aa 317–526

1.4.2.2 Water metagenomes

Cellulolytic enzymes have also been discovered in various water sources such as lakes and hot springs (Pottkämper *et al.*, 2009; Rees *et al.*, 2003; Wang *et al.*, 2008). A thermostable endocellulase (CelDR) was produced by *Bacillus subtilis* (DR) species isolated from a hot spring. The enzyme showed optimal activity at a temperature of 50°C and was proved to retain 70% of its maximum hydrolytic activity (CMCase) at an elevated temperature of 75°C after incubation for 30 minutes. This cellulase strain is a valuable thermostable enzyme because of its extreme heat tolerance (Maki *et al.*, 2009).

Rees *et al.*, (2003) also identified novel bacterial clones closely related to *Bacillus agaradhaerans*, which showed cellulase activity. The bacterium was part of the microbial community isolated from Lake Nakuru, a Kenyan soda lake. The genomic DNA library was made from laboratory cultures which were selectively enriched using carboxymethylcellulose (CMC). Screening for functional activity yielded cellulase-positive clones at a frequency of 1 in 15,000 (Rees *et al.*, 2003).

A Japanese fresh water clam was identified by Sakamoto and Toyohara (2009) to secrete two endogenous β-1,4-glucanase genes, which belong to glycolsyl hydrolase family 45 (GHF45) (*CjCel45A, CjCel45B*). Despite their difference in the 5′ and 3′-untranslated regions and in six nucleotides in the open reading frame, the two genes were both found to encode 208 amino acids from a 627 base pairs open reading frame. The protein encoded by these genes showed considerable β-1,4-glucanase activity (Sakamoto and Toyohara, 2009).

1.4.2.3 Kraft pulping process metagenomes

Ko and colleagues (2007) also identified a novel cellulase producing *Paenibacillus campinasensis* (BL11) species from the black liquor of brown-stock during Kraft pulping. The black liquor environment is highly alkaline and therefore unfavourable for bacterial growth. As such, cellulose degrading enzymes isolated from this kind of environment are thought to tolerate very harsh conditions, including the conditions commonly associated with biomass pre-treatment.

P. campinasensis, is a thermophilic and spore-forming bacterium that can grow over a wide temperature range (25°C-60°C) and a broad pH spectrum (Maki *et al.*, 2009). Its optimal growth has been found to be at neutral pH, at a temperature of 55°C. The physiological property of *P. campinasensis* and the vast number of hydrolytic enzymes it produces makes it very a very important strain in the bio-fuel industry (Ko *et al*., 2007).

The cellulose degrading enzyme produced by *P. campinasensis* was identified through zymographic analysis to be a recombinant protein with a molecular weight of 38 kDa. Hydrolytic activity of this isolate was shown across a variety of saccharides and polysaccharides since it produces a variety of biocatalysts, which include xylanases, cellulases, pectinases and also cyclodextrin glucanotransferases. The purified cellulase showed optimum activity at a temperature of 60°C and pH 7, respectively (Ko *et al.*, 2007).

1.4.2.4 Bio-reactor contents metagenomes

A novel β-glucosidase gene designated as *bgl1T*, was discovered by function based metagenomic screening of bio-reactor contents. The cloned gene has an open reading frame of 1860 base pairs and encodes a 620 amino acid β-glucosidase enzyme, Bgl1T, (65 kDa). Phylogenetic analysis and amino acid sequence of the Bg11T enzyme showed a close relationship with other putative β-glucoside-specific II ABC subunit components. The encoded Bgl1T enzyme hydrolyses D-(+)-cellobiose to glucose, with maximum activity observed at pH 7 and 37℃ when p-nitrophenyl-D-glucoside was employed as the substrate. Bgl1T protein has a Km value of 1.45mM, Vmax value of 20.5U/mg, kcat value of 1370/min and kcat/Km value of 943/mM/min. The Bgl1T glucosidase is more suited for moderate industrial glucose or bio-ethanol production (Jiang *et al*., 2010).

1.4.2.5 Poultry manure metagenomes

A *Paenibacillus* bacterial strain designated B39 was isolated from poultry manure compost in Taichung, Taiwan. This bacterium secretes a high molecular cellulose degrading enzyme that can degrade both carboxymethylcellulose and avicel. The CMCase activity of the secreted cellulase was found to be higher than activity on avicel or filter paper. Optimal CMCase activity was at 60°C, and a pH of 6.5. This enzyme is useful for use in the industrial hydrolysis of soluble cellulose, as well as microcrystalline sources of cellulose because of its thermostability and acidic tolerance (Wang *et al.*, 2008).

1.4.2.6 Swine waste metagenomes

A novel thermophilic, cellulolytic bacterium was isolated from swine waste and identified by Liang and colleagues, (2009), as *Brevibacillus* species, strain JXL. The secreted enzymes were found to use a broad spectrum of substrates such as crystalline cellulose, carboxymethylcellulose, xylan, cellobiose, glucose and xylose. These enzymes successfully retained approximately 50% of their activity after incubation at 100°C for one hour (Liang *et al.*, 2009).

1.4.2.7 Elephant dung metagenomes

Wang and co-workers also reported of two endoglucanase clones (EC1 and EC2) identified from a metagenome library derived from fresh elephant dung (Wang *et al.*, 2009). EC1 is

classified under glycolsyl hydrolase family 9 (GHF9), whereas EC2 belong to glycolsyl hydrolase family 5 (GHF5).

Table 3: Analyses of cellulase genes from elephant dung (adopted from Wang et al., 2009)

Active clone	**Accession number**	**Most homologous protein**	**Amino acid identity/ similarity**	**Conserved domains**
EC1	EU282865	EngO of *Clostridium cellulovorans* (AAT66046)	48/61	Cellulase (GHF9): aa 1–436
EC2	EU282864	Hypothetical protein (zp_02423462) from *Eubacterium siraeum*	39/54	Cellulase (GHF5): aa 101–428

1.4.2.8 Rabbit caecum metagenomes

An activity based metagenomic approach was used by Feng and colleagues (Feng *et al*., 2007) to clone cellulase genes from the rabbit caecum micro-flora. They isolated eleven independent clones showing cellulolytic activity. Endo-β-1,4-glucanase acitivity was observed in four of the clones and seven β-glucosidases were also isolated. To localize the cellulase genes, sub-cloning was performed and the shortest fragments expressing cellulase activity sequenced. Sequencing results yielded eleven cellulase genes, four endo-β-1,4-glucanases belonging to the glycosyl hydrolase family 5 (GHF 5) and seven β-glucosidases which belong to the glycosyl hydrolase family 3 (GHF 3). The sequenced cellulase genes encoded enzymes which shared less than 50% identities and about 70% similarities to cellulases in the databases. Maximum cellulase activities for ten of the eleven cloned cellulases were observed at pH 5.5-7.0 and temperatures of 40°C-55°C, the same conditions as those found in the rabbit caecum (Feng *et al*., 2007).

Table 4: Analyses of rabbit caecum-derived cellulase genes (adopted from Feng *et al*., 2007)

Sub-clone	**Accession number**	**Most homologous protein**	**Amino acid identity/ similarity %**	**Conserved domains/ catalytic residues**
RC1	DQ182491	*Bacillus cellulosilyticus* endoglucanase B (P06565)	46/64	Cellulase (GHF5): aa 94-360 E226 (P), E316 (N)

RC2	DQ182492	*Bacillus subtilis* DLG endo-β-1,4-glucanase (M16185)	46/66	Cellulase (GHF5): aa 116-376 E242 (P), E332 (N)
RC3	DQ916112	*Pectobacterium atrosepticum* endoglucanase N precursor (Q59394)	46/61	Cellulase (GHF5): aa 216-482 E348 (P), E438 (N)
RC5	DQ916113	*Pectobacterium atrosepticum* endoglucanase N precursor (Q59394)	48/64	Cellulase (GHF5): aa 225-491 E357 (P), E447 (N)
RG2	DQ182493	*Clostridium beijerincki* NCIMB 8052 β-glucosidase-related glycosidase (EAP61663)	45/61	GHF 3 C domain: aa 31-305 GHF 3 domain: aa 545∼807 D772 (N)
RG3	DQ182494	*Clostridium beijerincki* NCIMB 8052 β-glucosidase-related glycosidase (EAP61663)	48/64	GHF 3 C domain: aa31-294 GHF 3 domain: aa 545-807 D772 (N)
RG11	DQ916114	*Butyrivibrio fibrisolvens* H17c β-glucosidase A (P16084)	32/48	GHF 3 C domain: aa 35-309 GHF 3 domain: aa 551-811 D776 (N)
RG12	DQ916115	*Butyrivibrio fibrisolvens* H17c β-glucosidase A (P16084)	33/46	GHF 3 C domain: aa 35-307 GHF 3 domain: aa 547-809 D774 (N)
RG14	DQ916116	*Butyrivibrio fibrisolvens* H17c β-glucosidase A (P16084)	33/48	GHF 3 C domain: aa 45-320 GHF 3 domain: aa 563-822 D777 (N)
RG20	DQ916117	*Reinekea* sp. MED297 β-glucosidase-related glycosidase (EAR10786)	33/48	GHF 3 C domain: aa 49-321 GHF 3 domain: aa 561-823 D788 (N)

RG25	DQ916118	*Reinekea* sp. MED297 β-glucosidase-related glycosidase (EAR10786)	32/49	GHF 3 C domain: aa 38-312 GHF 3 domain: aa 555-814 D779 (N)

1.4.2.9 Cow rumen metagenomes

Cow rumen metagenome-derived cellulases, two endoglucanases (CC1 and CC2) and one β-glucosidase (Cβ1) enzymes were identified by Wang *et al.*, (2009). The endoglucanase, CC1 was reported to belong to the glycolsyl hydrolase family 5 (GHF5) and showed 76% amino acid homology to glycohydrolases isolated from buffalo rumen metagenomes. Similarly, the endoglucanase, CC2 also belonged to GHF5 and this enzyme was shown to have 75% homology to the bovine rumen metagenomes. In contrast, the β-glucosidase enzyme was classified under glycolsyl hydrolase family 3 (GHF3). This 759 amino acid cellulose degrading enzyme showed maximum catalytic activity at 35°C and neutral pH (7). It is reportedly stable over a wide range of pH (4.5-8.6) and can retain 80% activity at 40°C (Wang *et al.*, 2009).

Table 5: Sequence analyses of cow rumen-derived cellulase genes (adopted from Wang *et al*., 2009)

Active clone	**Accession number**	**Most homologous protein**	**Amino acid identity/ similarity**	**Conserved domains**
CC1	EU282863	Cellulase (ACA61132) from metagenomes of buffalo rumen	76/86	Cellulase (GHF5): aa 55 - 358
CC2	EU282868	Cellulase (CAJ19135) from metagenomes of bovine rumen	75/84	Cellulase (GHF5): aa 77 - 397
Cβ1	EU282867	β-Glucosidase (ABX76051)	83/91	β-Glucosidase (GHF3) aa: 30 - 252; (GHF3C): aa 317 - 526

1.4.2.10 Buffalo rumen metagenomes

Cellulases were also been discovered in buffalo rumen metagenomes by Duan *et al.*, (2009). A cosmid library was constructed from uncultured buffalo rumen micro-flora and screened for cellulase activities. The library had sixty-one independent clones which expressed cellulase activities; eleven independent clones showed CMCase activity, two clones had MUCace activity and the other forty-eight clones exhibited β-glucosidase activity. Positive CMCase and MUCase clones were sub-cloned with shorter insert sizes and screened for activity once again. Sub-clones that retained the above cellulase activities were then partially or completely sequenced to determine the cellulase genes and their neighbouring open reading frames (ORF). Sequencing results identified twelve endoglucanase genes from the eleven CMCase clones and two MUCase genes from the two MUCase clones. These cellulases showed diverse optimal pH from 4-7, but were most active under acidic conditions (Duan *et al.*, 2009).

Table 6: Analyses of buffalo rumen derived cellulase genes (adopted from Duan *et al.*, 2009)

Sub-clone	**Predicted ORF**	**Protein length (aa)**	**Most homologous protein**	**Amino acid identity/ similarity (%)**	**Conserved domains/ catalytic residues**
DC2 (DC2B)	DC2-2	681	*Prevotella Bryantii* Putative polygalacturonase (AAC97595)	22/41	Putative polygalacturonase
	DC2-3	512	Unidentified microorganism endoglucanase (ABX76048)	50/63	Cellulase/ GHF5 (173-491)/ E323 and E446
	DC2-4	517	Unidentified microorganism cellulase (CAJ19139)	68/82	Cellulase/ GHF5 (152-486)/ E305 and E438
	DC2-5	355	Unidentified microorganism glycosyl hydrolase (CAJ19136)	54/69	Mannase/ GHF26 (22-326)/ E185 and E289
	DC2-6	391	Unidentified microorganism conserved hypothetical protein (CAJ19137)	78/88	Predicted glycosylase/ DUF377 (31-362)

DC3 (DC3b)	DC3-1	520	Unidentified microorganism endoglucanase (ABX76048)	72/81	Cellulase GHF5 (161-502)/ E328 and E459
M8 (M8-5)	M8-2	344	*Clostridium thermocellum* endoglucanase (BAA007930)	44/61	Cellulase GHF5 (15-328)/ E147 and E284
DC9 (DC9BE)	DC9-2	346	Uncultured rumen bacterium β-glucanase (CAP07661)	81/89	Cellulase GHF5 (45-321)/ E177 and E273
M11 (M11-15)	M11-2	335	Uncultured rumen bacterium β-glucanase (CAP07661	80/90	Cellulase GHF5 (34-309)/ E165 and E261
DC20 (DC20HS)	DC20-2	332	Uncultured rumen bacterium β-glucanase (CAP07661	83/91	Cellulase GHF5 (31-307)/ E163 and E259
DC23 (DC23B)	DC23-2	496	Uncultured bacterium endoglycosidase precursor protein (ABB46200)	38/51	Mannase/ GHF26 (150-476)/ E319 and E439
	DC23-3	518	Unidentified microorganism endoglucanase (ABX76048)	51/65	Cellulase GHF5 (178-495)/ E328 and E451
C29 (C29-4X)	C29-2	553	Unidentified microorganism endo-1,4-β-D-glucanase (ABX76045)	76/86	Cellulase GHF5 (39-340)/ E178 and E288
C35 (C35S8)	C35-2	552	Unidentified microorganism endo-1,4-β-D-glucanase (ABX76045)	74/84	Cellulase GHF5 (39-336)/ E174 and E284
C5614 (C5614E7)	C5614-1	537	Unidentified microorganism endo-1,4-β-D-glucanase (ABX76045)	54/66	Cellulase GHF5 (40-334)/ E172 and E282
C67 (C67E4)	C67-1	546	Unidentified microorganism endo-1,4-β-D-glucanase (ABX76045)	67/79	Cellulase GHF5 (40-334)/ E172 and E282

M40 (M40ES)	M40-2	386	*Prevotella ruminicola* cellulase (BAA74515)	83/90	Cellobiosidase GHF5 (46-363)/ E178 and E320
DM1 (DM1P17)	DM1-1	332	*Ruminococcus flavefaciens* cellodextrinase A (P16169)	52/70	Cellodextrinase GHF5 (36-317)/ E157 and E274

1.4.2.11 Insect gut metagenomes

Cellulose digestion has been specifically demonstrated in representatives of widely different taxonomic groups of insects: cockroaches and various wood-eating insects, lower and higher termites (Isoptera), various beetles (Coleoptera and wood wasps (Hymenoptera) (Prins and Kreulin, 1991)). Several groups of insects have the ability to synthesize endoglucanases as these enzymes are present in extracts of their salivary glands and/or mid-gut tissues. However, insects never appear to be capable of forming cellobiohydrolases. This problem seems to have been solved by an overwhelming variety of symbioses that exist between phytophagous insects and micro-organisms (Prins and Kreulen, 1991). Different insects have gut microbial consortia capable of digesting cellulose or hemicelluloses material into many different metabolites, chief amongst them being organic acids, ethanol and methane (Dillion and Dillion, 2004; Lemke *et al*., 2003; Egert *et al*., 2005; Thorsten *et al*., 2003; Tholen *et al*., 1997; Tholen *et al*., 2007).

An endo-β-1,4-glucanase gene was cloned by Inoue and co-workers from the cDNA library of a mixed population of symbiotic protists in the hindgut of the termite, *Coptotermes formosanus* (Inoue *et al*., 2005). Sequencing results showed that the gene nucleotide sequence comprises of 941 nucleotides (including the poly-A tail). The full length cDNA has a 921 bp open reading frame (ORF) and encodes a 33,620 Da protein which is thought to be secreted by a symbiotic protist, *Spirotrichonympha leidyi*. The endoglucanase enzyme showed optimal CMC activity at pH 5.8-6 and 70°C (Inoue *et al*., 2005).

Two endogenous cellulase genes (*CfEG3a and CfEG5*), endo-β-1,4-glucanases, were discovered by Zhang and colleagues from the salivary gland and mid gut of a Formosan subterranean termite, *Coptotermes formosanus* (Zhang *et al*., 2009; Zhang *et al*., 2011). Assigned a gene accession number of EU853671, the gene *CfEGa* has a nucleotide sequence of 1486 bp excluding the poly-A tail. This cellulase gene was cloned and overexpressed in

both the native (nCfEG) and C-terminal His-tagged form (tCfEG). Hydrolytic activity was observed for both nCfEG and tCfEG on cellulose substrates, with nCfEG showing more stability than tCfEG. Optimal enzymatic activity for both forms (nCfEG and tCfEG) was observed at an acidic pH, as compared to neutral or alkaline conditions. Different catalytic activities of the nCfEG and tCfEG forms were observed in the type of hydrolytic products they produced on different substrates. When CMC was used as a substrate, both forms acted as an enduglucanase by randomly hydrolysing the CMC's internal β-1,4-glycosidic bonds to form a smear of polymers with different lengths (cellobiose, cellotriose and cellotetraose were more noticeable). However, the hydrolytic products of tCfEG were one unit sugar less compared to those produced by nCfEG (Zhang *et al*., 2009). The recombinant protein encoded by *CfEG5* showed similar hydrolytic and biochemical properties as those exhibited by nCfEG and tCfEG (Zhang *et al*., 2011).

Willis and co-workers also reported the discovery of a novel endo-β-1,4-glucanase gene from a red flour beetle, *Tribolium castaneum* (Herbst) (Willis *et al*., 2011). The gene was named *TcEG1* (for *T. castaneum* endoglucanase 1), since it was the first endoglucanase gene to be isolated from *T. castaneum*. The full length *TcEG1* cDNA clone had a nucleotide sequence of 1356 base pairs and showed homology to enzymes in the glycosyl hydrolase family 9 (GHF9). This *TcEG1* cDNA reportedly encodes a 49.5 kDa protein, which demonstrated activity on carboxymethyl cellulose (CMC), but not on microcrystalline cellulose (MCC), thus indicating an endoglucanase method of action. The recombinant enzyme, TcEG1 exhibited high relative activity at alkaline pH, and as such can find potential application in the biofuel industry (Willis *et al*., 2011).

1.5 Dung beetles

As described previously, many organisms have the ability to degrade cellulose matter through the use of symbiotic micro-organisms and their own cellulose degrading enzymes. Scarab beetles have been reported as such organisms since they can effectively utilize various cellulose rich plant matter and animal waste as energy sources (Huang *et al*., 2010).

Dung beetles belong to the family Scarabaeidae and order Coleoptera, which is regarded as the largest order in the insect kingdom. Scarab beetles are commonly found in various environmental niches where decaying plant and/or animal wastes form a high proportion of the available biomass (Huang *et al*., 2010). This is because scarabaeids are reportedly

herbivorous or saprophagous and many of their species feed on plant roots, decaying organic matter and animal waste of very low nutritional value (Zhang and Jackson, 2008). As such, scarab beetles play an important role in the ecological control of dung, decaying wood and plant matter (Koyama *et al*., 2003).

1.5.1 Gut morphology and physicochemical properties

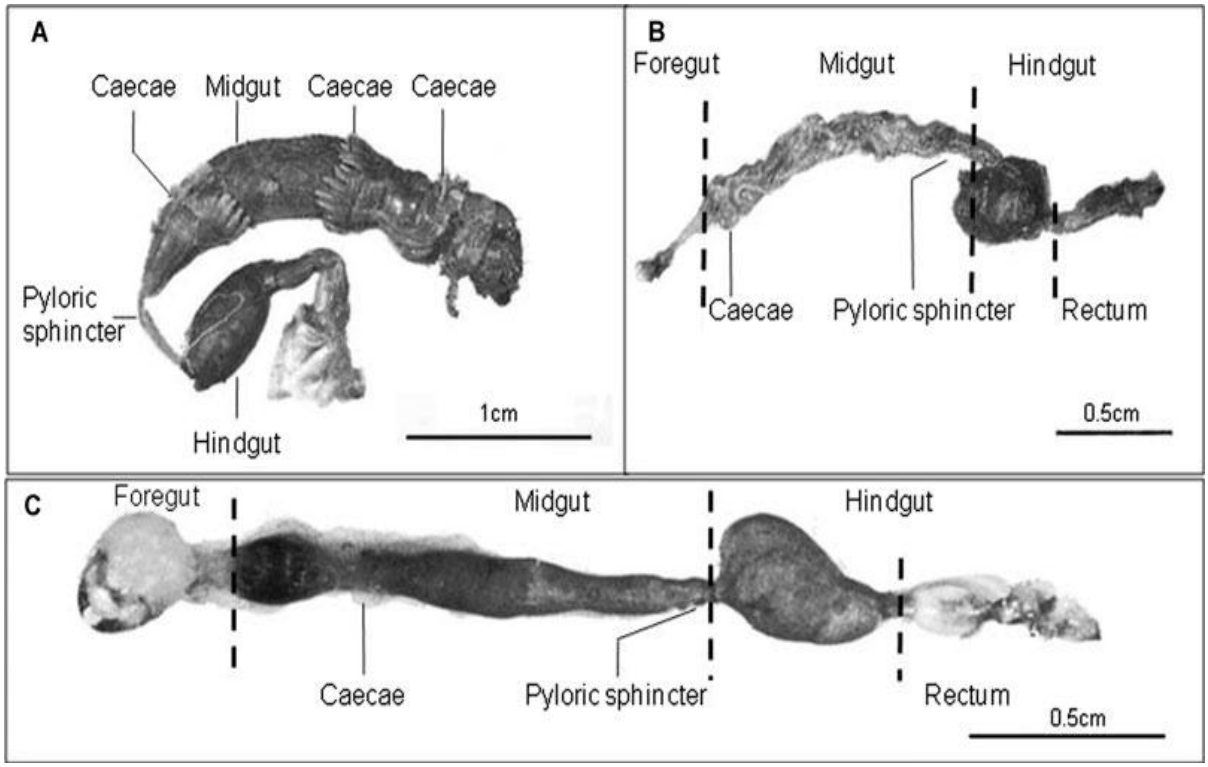


Figure 2: Gut morphology of scarab beetle larvae showing foregut, mid-gut and hindgut locations for (A) *Papuana huebneri*, (B) *Holotrichia parallela*, and (C) *Costelytra zealandica* (adopted from Huang *et al*., 2010).

The dung beetle gut comprises of three sections, a small foregut, an elongated mid-gut and a bulbous hindgut, as shown above (**Figure 2**). The foregut section starts from the oral cavity and extends to the cardiac valve, which marks the beginning of the mid-gut, also known as the mesenteron. The mid-gut consists of an elongated tube of epithelial cells and it extends up to the proctodeum, where the hindgut begins. The mid-gut occupies a large proportion of the dung beetle larvae's body length. The hindgut is a fermentation chamber, lined up with cuticle and lobe-like structures that extend into the lumen (Huang *et al*., 2010).

High alkaline pH values (pH>8) have been observed consistently in the mid-guts of various scarab beetle larvae such as *Melolontha melolontha* and *Pachnoda ephippiata* larvae. The pH values tend to drop to a neutral or slightly alkaline pH as the mid-gut approaches the hindgut section (Ergert *et al*., 2005; Lemke *et al*., 2003). The alkaline pH is thought to assist in the breakdown of organic polymers in humus and dung by extracting hemicelluloses from plant cell walls. Therefore, gut alkalinity has been proposed as a very important adaptation that aids in the essential nutrient uptake from the organic matter consumed by scarab beetle larvae (Huang *et al*., 2010).

Redox potential values tend to shift from oxidizing conditions in the mid-gut toward reducing conditions in the hindgut of most scarab beetle larvae (Ergert *et al*., 2005; Lemke *et al*., 2003). Redox potential values determine whether aerobic oxidation or anaerobic fermentation occurs. The reducing conditions observed in the hindgut accounts for the anaerobic fermentation that prevails in the hindgut section. Both the observed pH and redox potential values observed in the scarab larvae gut are partially determined by the action of symbiotic microbial organisms operating in these sections. These conditions impose physiological constraints on the effectiveness of some classes of the digestive enzymes and as such, play a big role in the digestion, absorption and detoxification of food in the gut of scarab beetle larvae (Huang *et al*., 2010).

1.5.2 Gut micro-flora

The gut of scarab beetle larvae comprises of a rich population of xylan and pectin degrading micro-organisms, which increase in density from the mid-gut to the hindgut (Egert *et al*., 2005; Zhang and Jackson 2008). The scarab hindgut is closely analogous to the rumen of higher mammals, which is the primary site of microbial fermentation of plant organic material. Microbial community composition in the mid-gut is highly varied and is easily affected by the diet or environmental conditions (Huang *et al*., 2010).

Zhang and Jackson, (2008) successfully showed that the number of cultured bacteria in the midgut of *Costelytra zealandica* were significantly less as compared to those of its hindgut. They identified *Clostridium* bacterial species as predominant micro-organisms in the hindgut of *C. zealandica* larvae. The remaining bacteria were aligned to the β-proteobacteria, δ-proteobacteria, and Bacteroidetes.

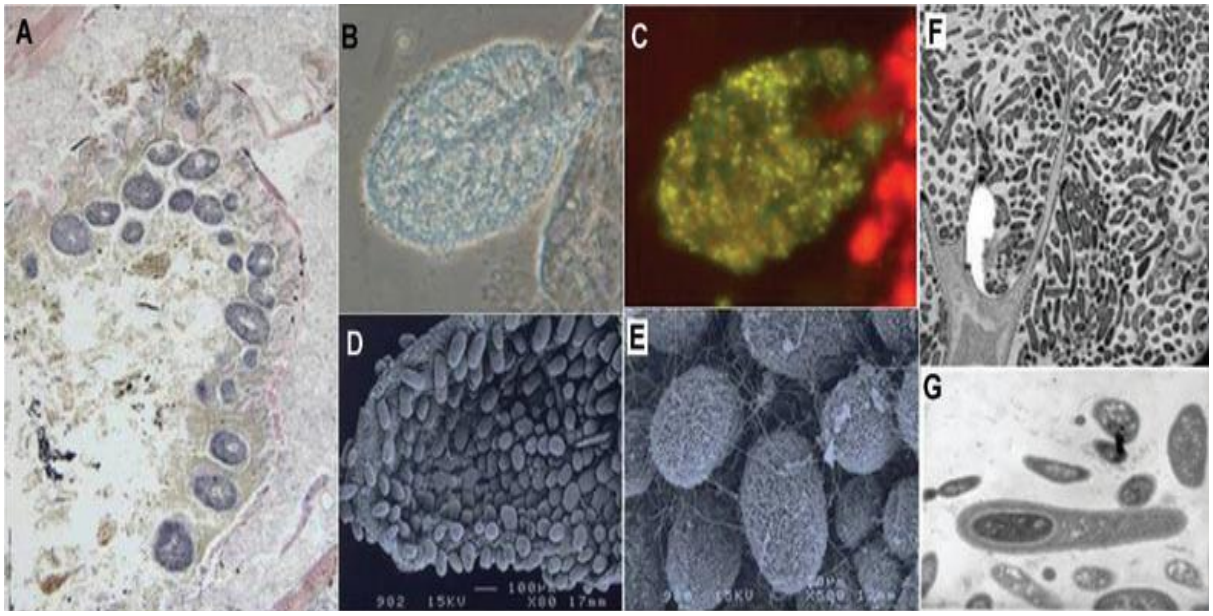


Figure 3: Features of the scarab hindgut and associated bacteria. (A) Thin section of a *C. zealandica* hindgut, showing the characteristic lobe like structures. (B) Wet mount of a lobe from first instar *C. zealandica* larva 5 days post-eclosion observed under phase contrast microscopy. (C) Fluorescence image of the lobe from photo B after staining with LIVE/DEAD Backlight (molecular probes); bacteria in the lobe appear green (SYTO9) and the hindgut cells appear red (propidium iodide). (D) and (E) Scanning electron micrographs of the *C. zealandica* hindgut lobes displaying tree like cuticular growths surrounded by a cuticular meshwork. (F) Transmission electron micrograph of a cross section through the *C. zealandica* hindgut showing a portion of the tree like lobe structure (bottom left) and the diverse bacterial morphotypes. (G) Transmission electron micrograph showing Clostridia-form bacteria from the hindgut lobes of *C. zealandica* (adopted from Zhang and Jackson, 2008).

Egert and co-workers, (2005), demonstrated that *M. melolontha* larvae mid-gut had simple microbial composition compared to the hindgut. Hindgut microbial community species were found to be more diverse and were dominated by clones related to the Clostridiales. The clones related to β- and δ-proteobacteria were found in the hindgut, whereas the lumen was inhabited by a majority of clones related to Actinobacteria, Bacillales, Lactobacillales, as well as γ-proteobacteria. This led to suggestions that different bacterial species were confined to different hindgut compartments (Egert *et al*., 2005). The above observations made in both *C. zealandica* and *M. melolontha* suggests the existence of an intimate symbiotic relationship between Clostridiales and scarab beetle larvae (Huang *et al*., 2010).

1.5.3 Cellulose digestion in scarab beetles

Cellulose and hemicellulose digestion in the scarab beetle larvae is thought to occur in the hindgut section, aided by the symbiotic micro-flora (Zhang and Jackson 2008). The absence of cellulose-degrading microorganisms and the non-reducing conditions in the mid-gut makes cellulose digestion in this gut compartment very unlikely. Thus, the mid-gut only serves to prepare the cellulose or hemicellulose matter for further hindgut cellulose digestion (Huang *et al*., 2010). Digestion of cellulose in the scarab beetle larvae hindgut is more closely related to that of ruminant animals, described previously. The main products of this cellulose and hemicellulose digestion are volatile fatty acids, such as acetic acids, as well as other by-products which are utilized by methanogenic bacteria. The products of cellulose digestion are thought to be quickly absorbed through the cuticular membrane lining the hindgut for subsequent utilization by the insect.

The identification and isolation of bacteria possessing cellulolytic and hemi-cellulolytic activities from scarab beetle larvae digestive tracts, has already been documented (Huang *et al*., 2010). This work seeks to identify cellulolytic organisms and enzymes, such as endoglucanases, exoglucanases or cellobiohydrolases from micro-flora isolated from the guts of the dung beetle (*Euoniticellus intermedius*) larvae for use within the bio-fuel industry.

1.6 Project justification

It is an undisputable fact that the world economy has been largely running on fossil fuel derived energy since the industrial revolution. Previously abundant and cheaply accessible, fossil fuels have undoubtedly played a big role in the industrialisation of most of the developed states we see today, and they are still an integral part of today's energy mix. However, due to their non-renewable nature, fossil fuel reserves have been dwindling at an alarming rate, at a crucial time when the world's energy needs have also been on a constant rise. In response, the price of energy has subsequently sky-rocketed, thereby putting a lot of economic and political pressure, especially on the non-oil producing countries. In addition to rising fuel prices, the damaging environmental effects such as global warming and climate change, associated with rampant fossil fuel use have also compounded the current challenges attributed to the world's heavy reliance on fossil fuels.

To counter the economic and environmental consequences of fossil fuel use, the world has not disappointed, as it has invested a lot in renewable and sustainable energy resources with

bio-fuels being one of the leading pack. Historically, transportation bio-fuels like bio-ethanol have primarily been derived from food sources such as sugar cane and corn, a situation that has sparked a lot of controversy on the food versus fuel debate. This has consequently led to most higher education and energy research institutions becoming more focused towards the production of second generation bio-fuels such as cellulose-derived bio-ethanol. Despite huge financial investment and concerted research efforts in this area, and the abundance of cellulose raw materials, production of second generation bio-ethanol at an industrial scale is still hampered by unreasonably high operating costs due to the recalcitrant nature of cellulose, which ultimately makes its production economically unfeasible.

The rate limiting steps for successful cellulose-derived bio-fuel production has been the cellulose degradation process, due to cellulose structural properties, as well as the difficulty encountered in the conversion of some resultant pentose sugars into ethanol. To tackle these challenges, chemical and biochemical methods for cellulose degradation to form reducing sugars and the subsequent fermentation of these reducing sugars to bio-ethanol have been extensively studied and proposed. Even today, a lot of research is still being undertaken in a bid to find the silver bullet and consequently break the financial barriers which are currently hindering any progress in the industrial production of the second generation bio-fuels.

On the biological front, various cellulose degrading micro-organisms have been discovered through genomic and metagenomic approaches and the potential role of these micro-organisms in advancing the bio-fuel dream can never be over-emphasized. In the same regard, this research seeks to contribute to the growing list of possible biological routes of cellulose degradation by investigating and characterising the cellulose degrading capabilities of the symbiotic micro-flora naturally residing in the gut of the dung beetle, *Euoniticellus intermedius* (Coleoptera: Scarabaeida), larvae, and their potential application and contribution to the sustainable production of cellulose-derived bio-fuels.

1.6.1 Project aim

The aim of this study is to evaluate the cellulose degradation capabilities of the gut micro-flora of *E. intermedius* larvae.

1.6.2 Specific project objectives

- Dung beetle rearing and larvae collection.

- Dissection of *E. intermedius* larvae and culturing of the gut micro-flora on cellulose media.
- Genomic DNA extraction of the cultured micro-flora and construction of a primary cosmid library.
- Screening the cosmid library for cellulase activity.
- Sequencing of the cellulase activity positive clones.

CHAPTER 2

Materials and Methods

2.1 Buffers, gels, media and solutions used

Table 7: Buffers and gels laboratory prepared and used in this work

Buffer	Preparation method/Composition
Agarose gels	0.8 g agarose (0.8%), 1 g agarose (1%) or 1 g LMP agarose (1% LMP). Added 20 ml 0.5X TAE buffer and 80 ml distilled water.
Citrate buffer stock (1 M)	210 g citric monohydrate, 750 ml deionised (DI) water, pH adjusted to 4.5 by adding 55 g NaOH.
Citrate working buffer (50 mM)	5 ml 1 M citrate buffer stock plus 95 ml DI water (pH 4.8). Used in cellulase activity analyses (DNS method).
5X TAE buffer (stock)	54 g tris base, 27.5 g acetic acid, 20 ml 0.5 M EDTA pH 8. Added distilled water to 1000 ml, mixed and autoclaved at 121°C for 15 minutes.
1X TAE buffer	50 ml 5X TAE buffer (pH 8) plus 200 ml of distilled water. 25 µl ethidium bromide was added last. Used as a running buffer during gel electrophoresis of DNA.
0.5X TAE running buffer	25 ml 5X TAE buffer (pH 8) plus 225 ml distilled water. 25 µl Ethidium bromide was added last. Used in the preparation of agarose gels and as a running buffer during gel electrophoresis of DNA.
TE buffer	2 ml of 0.5 M EDTA (pH 8) plus 1 ml 1 M Tris-HCl (pH 8) mixed with 97 ml distilled water. Autoclaved at 121°C for 15 minutes Used to re-suspend extracted DNA for storage purposes.
0.05 M Potassium phosphate buffer (pH 7.1)	18.4 ml of 1 M KH_2PO_4, 31.6 ml of 1 M K_2HPO_4, make up to 1000 ml with distilled water.

0.1 M Sodium phosphate buffer (pH 8)	4.66 ml 1M Na_2HPO_4, 0.34 ml 1M NaH_2PO_4, make up to 50 ml with distilled water. Adjust pH with NaOH solution or phosphoric acid.
Phosphate buffered saline solution (pH 7.4)	8.01 g NaCl, 0.2 g KCl, 1.78 g Na_2HPO_4, 0.27 g KH_2PO_4, 1000 ml distilled water.

Table 8: Media and media components used for bacterial culturing

Medium	**Preparation/Composition**
CMC overlay	7 g agarose, 1 g CMC, 1000 ml 0.05 M potassium phosphate buffer.
LB broth	10 g tryptone, 5 g yeast extract, 5 g NaCl, 1000 ml distilled water.
LB agar (1%)	10 g tryptone, 5 g yeast extract, 5 g NaCl, 10 g agar, 1000 ml distilled water.
M1A	1 g NaCl, 0.5 g KCl, 0.1 g $MgCl_2.6H_2O$, 0.015 g $CaCl_2.2H_2O$, 0.3 g NH_4Cl, 0.2 g KH_2PO_4, 0.15 g $NaSO_4$, 0.5 g yeast extract, 40 g filter paper, 1000 ml distilled water.
M1B	1 g NaCl, 0.5 g KCl, 0.1 g $MgCl_2.6H_2O$, 0.015 g $CaCl_2.2H_2O$, 0.3 g NH_4Cl, 0.2 g KH_2PO_4, 0.15 g $NaSO_4$, 0.5 g yeast extract, 0.225 g D-glucose, 0.855 g D-lactose, 1000 ml distilled water.
MUC overlay	Agarose: 7 g; MUC: 500 mg; 0.05 M Potassium phosphate buffer: 1000 ml.
Trace element solution (SL11)	$FeCl_2.4H_2O$: 1.5 g; $CoCl_2.6H_2O$: 100 mg; $MnCl_2.4H_2O$: 100 mg; $ZnCl_2$: 70 mg; $Na_2MoO_4.2H_2O$: 36 mg; $NiCl_2.6H_2O$: 24 mg; H_3BO_3: 6 mg; $CuCl_2.2H_2O$: 2 mg; HCl (25%): 10 mg
Selenium tungstate solution	NaOH: 0.5 g; $Na_2WO_4.2H_2O$: 4 mg; $NaSeO_3.5H_2O$: 3 mg; Distilled water: 1000 ml.

*All media were autoclaved at 121°C for 15 minutes and allowed to cool before use. Media not used immediately were stored at 4°C.

Table 9: Solutions, antibiotics, dyes and other reagents used in this study

Solution	Preparation method/Composition
Ampicillin stock solution (100 mg/ml)	1 g Ampicillin powder plus 10 ml of 70% ethanol. Used to select the transformed bacteria during cloning.
Bradford reagent	100 mg Coomassie Brilliant Blue G-250 dissolved in 50 ml 95% ethanol. To this, 100 ml of 85% (w/v) phosphoric acid was added and mixed. Made up to 1000 ml with distilled water, filtered through a Whitman number 1 filter paper and stored in an amber bottle at 4°C until further use. Used for total protein determination.
Congo red dye solution (1 mg/ml)	1 mg in 1 ml distilled water containing 3 drops of ammonium hydroxide solution. Used during endo-glucanase activity plate assays.
DNS reagent	10.6 g 3.5-dinitro-salicyclic acid, 19.8 g NaOH, 1416 ml distilled water. To this solution, 306 g of Rochelle salts (sodium potassium tartrate), 7.6 ml of phenol and 8.3 g of sodium metabisulfite were added. Used for cellulase activity determination (DNS method).
Sodium acetate (3 M)	408 g sodium acetate in 1000 ml distilled water. Sterilized using syringe micro-filters (0.2 µm). Used during cosmid DNA isolation step.
Glucose stock solution (10.0 mg/ml)	1 g of anhydrous glucose in 100 ml of distilled water. Used for cellulase activity standard curve plotting (DNS method).
Insect Ringer solution	0.9 g NaCl, 0.02 g $CaCl_2$, 0.02g KCl and 0.02g $NaHCO_3$ in 100 ml distilled water. Used during dung beetle larva dissection.
Lactose solution (0.1 M)	36 g in 100 ml distilled water. Sterilized using syringe micro-filters (0.2 µm). Used in CMC/MUC overlay preparation.
Magnesium sulphate solution (10 mM)	2.46 g $MgSO_4$ plus 1000 ml distilled water. Autoclaved at 121°C for 15 min. Used during the DNA cloning experiments.

Protein stock solution (10.0 mg/ml)	10 mg of Bovine γ-globulin in 1 ml of distilled water. Used for standard curve plotting during total protein determination.
Phosphate buffered saline solution	8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 and 0.24 g KH_2PO_4 in 1000 ml distilled water. Sterilized using syringe micro-filters (0.2 µm). Used for homogenizing dung beetle larva guts.
Sodium chloride solution (1M)	5.8 g NaCl in 100 ml distilled water. Sterilized using syringe micro-filters (0.2 µm). Used during endoglucanase activity plate assays.

2.2 Dung beetle collection and maintenance

Dung beetle collection was done at a dairy farm, about 10 km South-West of Johannesburg city, South Africa. The insects were collected between August and September 2010 and bred at the University of the Witwatersrand, in a temperature (28°C) controlled dung beetle rearing room. The live dung beetles were collected directly from the cattle dung and sorted into males and females. Breeding was done in plastic containers measuring 160 mm x 130 mm x 130 mm, with each container carrying not more than three breeding pairs of dung beetles. The containers were half-filled with thoroughly sieved and compacted slightly moist soil.

Fresh cattle dung collected from the same farm as the dung beetles was used as the dung beetle feed and was stored in small 1000 g plastic packets at -20°C. On every third day (72 hours), one packet of dung was added to each container containing the dung beetle breeding pairs.

Once a week, the breeding containers were sieved to remove any brood balls present. These were transferred to plastic containers, measuring 400 x 300 x 200 mm, according to the date of collection. The brood balls were then covered with sieved, slightly moist soil and a wet sponge placed on top of the soil to maintain optimum soil moisture levels. A plastic dish containing some dung was also placed in one corner of the brood ball containers to attract and trap any emerging hatched dung beetles. The number of surviving dung beetle breeding pairs were also determined once per week and transferred to new breeding containers. The newly hatched dung beetles were kept in separate breeding containers from the old breeding pairs and dung beetle numbers were controlled by culling older breeding pairs.

2.3 Dung beetle larvae dissection

A modified dung beetle dissection protocol similar to the one described by Lemke and colleagues (Lemke *et al*., 2003) was used to perform the dissection of 2-3 weeks old, late second or third instar *E. intermedius* larvae. The steel/metal dissection equipment was autoclaved at 121°C for 15 minutes and the preparation dish was left overnight (approximately 12 hours) immersed in 70% ethanol for aseptic technique purposes. Insect ringer solution was prepared according to the protocol described by Hayashi and Kamimura (2001) and autoclaved for 15 minutes at 121ºC.

Larvae dissection was performed in a sterile preparation dish under a dissection microscope, in a laminar flow cabinet. The larvae were first anesthetized by exposing them to a nitrogen, hydrogen and carbon dioxide (71/7/22 vol/vol respectively) (Afrox grade) gas mixture for 15 minutes. They were then fixed onto the preparation dish with steel pins, with the larvae laid on its sides, and sterile insect ringer solution added. The cuticle was cut along the side lines and the ventral integument, circular muscles and trachea were then carefully removed. The larva head was decapitated followed by a circular cut on the anus and the intestinal tract was carefully removed from the body. The beetle larvae guts were stored in 1.5 ml tubes at -70°C until cellulose culturing experiments.

2.4 Culturing of dung beetle larvae gut micro-flora

2.4.1 Media preparation

A basal cellulose medium, M1A, for gut micro-flora cultivation was prepared using Whitman filter paper as the only carbon source, according to a modified method described by Lemke and co-workers (Lemke *et al*., 2003). A non-cellulose basal medium, M1B, was also prepared in the same way as M1A, but in the place of filter paper, glucose and lactose were used as alternative carbon sources. To both M1A and M1B basal media preparations, 1 ml of trace element solution (SL11) and 1 ml of selenium tungstate solution were added and the media pH corrected to pH 8 using a 0.1 M sodium phosphate buffer before autoclaving.

2.4.2 Media inoculation

Dung beetle larvae guts weighing 1 g were homogenized in 9 ml of phosphate buffered saline solution (pH 7.4). The homogenate (1 ml) was then inoculated into 9 ml of M1A (in the place of Whitman filter paper, 0.405 g of cellulose microcrystalline was used in the cellulose media during pre-culturing experiments) and M1B media for pre-culturing. The inoculations were done in triplicate for each respective medium. The pre-cultures were incubated at 30°C for 48 hours in the dark, shaking at a speed of 80 revolutions per minute on a rotary shaker. Bacterial growth was visually observed by the degree of turbidity of cultures after 24 hours and 48 hours of incubation period. Inoculation of pre-culture samples (10 ml) into their respective main cultures (90 ml) was done in triplicate after 48 hours of pre-culturing. The main cultures were incubated at 30°C in the dark, shaking at 80 rpm for 28 days.

Samples for bacterial growth determination and cellulase activity were taken on a daily basis, 3 ml per culture sample and kept at -70°C until analysis.

2.4.3 Bacterial growth determination

Samples (Day 1-Day 28) for bacterial growth determination were taken from both the cellulose media and non-cellulose media cultures. Total protein concentration analyses using the Bradford protein assay protocol (Bradford, 1976) were employed to determine bacterial biomass. This assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. Both hydrophobic and ionic interactions play a crucial role in stabilizing the anionic form of the dye, resulting in a visible colour change (Bradford, 1976; Stoscheck, 1990).

Due to the limited amount of protein (poor bacterial growth) in the M1A and M1B culture samples, the Bradford micro-assay method was used to plot the standard curve. A 10 mg/ml protein stock solution was made and 13 dilutions made from between 1-25 µg/ml. The Bradford stock reagent was diluted 1:5 in distilled water to make the working reagent and 980 µl of the working reagent added to a glass spectrophotometer cuvette. To this, 20 µl of the standard protein dilutions were then added and the mixture left at room temperature for 5 minutes to allow for colour development. Absorbance readings were measured using a spectrophotometer at 595 nm wavelength.

Total protein concentration determination in the test-samples was done by adding test-samples to the Bradford working reagent in a glass spectrophotometer cuvette and incubating the mixture for 5 minutes at room temperature. Volume adjustments were done on the samples to get absorbance values of between 0.1 and 0.9. The absorbance values of the test-samples were then measured using a spectrophotometer at 595 nm wavelength, just like above. The total protein concentrations of the samples were determined by comparing the samples' absorbance values against the plotted Bradford micro-assay standard curve. Any volume adjustments performed on the samples were catered for during the total protein concentration calculations. Measurements were done in quartet for both the cellulose culture and non-cellulose culture samples. Total protein concentration values were regarded as directly proportional to bacterial growth.

2.4.4 Cellulase activity determination

Cellulase activity was determined by measuring the concentration of reducing sugars in the cellulose culture samples using a modified DNS procedure described by Adney and Baker (1996). Considered as a relatively simple, fast and reliable method, the DNS test has found wide applications in the medical field for the determination of sugar levels in blood and urine. It uses the reducing ability of reducing sugars to reduce 3,5-dinitro-salicyclic acid to form a coloured product, 3-amino-5-nitrosalicylate that absorbs light at 540 nm.

To plot the cellulase activity standard curve, a 10 mg/ml glucose stock solution was first made and 6 dilutions made from the stock solution from between 1-7 mg/ml. A constant volume of 0.5 ml of each of the above dilutions was then added to 6 different test tubes containing 1 ml citrate buffer. The control experiment only contained 1.5 ml of citrate buffer. This was followed by the addition of 3 ml of DNS solution to all test tubes (including control) and the mixture boiled for 5-10 minutes in a vigorously boiling water bath (90°C). The tubes were transferred to a cold ice water bath to cool and 0.07 ml of colour-developed mixture added to 0.93 ml of distilled water in a spectrophotometer cuvette. Absorbance values were taken at 540 nm wavelength using a spectrophotometer.

To measure the reducing sugar concentration in the cellulose culture samples, 0.5 ml samples were added to test tubes containing 1 ml citrate buffer and 3 ml of DNS solution added, as described above. The mixture was boiled for 5-10 minutes, allowed to cool, and absorbance values taken without diluting the test-samples at 540 nm. Reducing sugar concentration was

determined by comparing the absorbance values of test-samples against the plotted glucose standard curve. The volume adjustments made for the standard curve were taken into consideration during the test-sample glucose/reducing sugar concentration calculations. Measurements were done in quartet for the cellulose culture samples.

2.5 DNA isolation

After plotting the bacterial growth curve of the original cellulose culture and determining the period where high bacterial growth was observed, a new cellulose culture specifically dedicated for DNA isolation was started. Genomic DNA isolation was done on day 10 samples of the new cellulose culture using the *ZR Fungal/Bacterial DNA MiniPrep kit* (Zymo Research, USA). Due to poor bacterial growth, DNA isolation was performed on a total of 13 replicate samples.

About 1 ml of cellulose culture samples were transferred into 1.5 ml tubes and centrifuged at 13,000 x g for 5 minutes at room temperature. The supernatant was discarded and the pelleted cells from 4 or 5 tubes pooled together (to achieve approximately 20 mg of sample weight) to perform one mini-prep DNA isolation. DNA isolation was performed on the pooled cells according to the procedure described in the *ZR Fungal/Bacterial DNA kit* (Zymo Research, USA).

The DNA isolated from the different replicate tubes was pooled together, concentrated by precipitation in ethanol acetate (24:1) and re-suspended in 100 µl TE buffer. The isolated DNA (5 µl) was mixed with 1 µl of gel loading buffer (Blue) and ran at 100V for 45 minutes on 1% agarose gel in 0.5X TAE buffer. A GeneRuler^{TM} 100 bp DNA ladder (Fermentas, South Africa) (5 µl) was ran alongside the DNA samples and the rate of migration was visualized using UV light. The genomic DNA concentration was determined using a Nano-Drop spectrophotometer (*ND 1000*). The isolated genomic DNA was kept at -20°C until further analyses.

2.6 Cosmid library construction

A complete and unbiased primary cosmid library was constructed from the isolated genomic DNA using a cloning ready, pWEB-TNC™ Cosmid Cloning Kit (EPICENTRE Biotechnologies, USA).

2.6.1 Insert DNA preparation

Previously isolated genomic DNA suspended in TE buffer (41.2077 ug) was first sheared by expelling it twice from a syringe through a small bore needle, as described in the kit instruction manual. Shearing was done to generate random DNA fragments (30-45 kb). The size distribution of the DNA molecules were then examined by running a 5 µl aliquot of genomic DNA on 0.8% agarose gel in 1X TAE buffer using 100 ng of the T7, cosmid control DNA as a size marker.

The DNA was then end-repaired, as described in the kit manual, to generate blunt ends that allow cloning of the DNA into the prepared pWEB-TNC vector. The 80 µl end-repair reaction mixture comprised of 10X End-Repair Buffer (8 µl), 2.5 mM dNTP Mix (8 µl), 10 mM ATP (8 µl), nuclease free water (2 µl), genomic DNA (50 µl, which translates to approximately 20.6 µg) and End-Repair Enzyme Mix (4 µl).

2.6.2 Insert DNA size selection

As described in the kit manual, the end-repaired genomic DNA was size selected by running a 5 µl aliquot on 1% low melting point agarose gel in 1X TAE buffer at 30V overnight, for approximately 12-14 hours. No ethidium bromide solution was added to both the agarose gel preparation and the running buffer. The T7 cosmid control DNA (100 ng) was loaded as a size marker on each of the outer lanes and the end-repaired DNA between these two marker lanes. Following completion of gel electrophoresis, the two outer lanes containing the cosmid control DNA were carefully cut off from the gel, stained with ethidium bromide and viewed under UV light. The position of size markers on the cut outer lanes of the gel were noted using a pipette tip and the gel re-assembled to establish the position of the insert DNA on the unstained gel slice. Using a surgical blade, a 3 mm wide gel slice containing insert DNA migrating between the size markers was cut and stored in 1.5 ml tubes at 4°C until further experiments.

2.6.3 In-gel ligation reaction

The gel slice containing the end-repaired and size-selected DNA was melted and in-gel ligation done according to the *pWEB-TNC™ Cosmid Cloning kit* instruction manual. The reaction was stopped before the *In Vitro* Packaging step, and the reaction mixture stored at -20°C overnight.

2.6.4 *In Vitro* Packaging (using MaxPlax Packaging Extracts)

This was done according to the *pWEB-TNC™ Cosmid Cloning kit* instruction manual. A single packaging extract was thawed and 25 µl of the extract added to a 1.5 ml tube. The ligated cosmid DNA (10 µl) was added to the tube containing the extract, mixed by pipetting and returned to the bottom of the tube by centrifugation for one minute at 10,000 x g. The reaction mixture was incubated at 30°C for 90 minutes, after which another 25 µl of packaging extract was added and the reaction mixture re-incubated at 30°C for another 90 minutes. Phage dilution buffer (500 µl) was added, followed by 25 µl of choloform and the reaction mixed by gentle vortexing after each addition. This was stored at 4°C until further experiments.

The titer of the packaged cosmids was determined by adding 10 µl of the packaged cosmids to 100 µl of previously prepared EPI100-T1^{R} host cells. The reaction mixture was incubated at 37°C for 20 minutes to allow bacterial transformation. The infected bacteria (110 µl) were then spread on LB-ampicillin selection agar plates (90 mm). The inoculated LB-ampicillin plates were incubated in the dark at 37°C overnight (for 12-16 hours). After incubation, colony growth was determined and the titer calculated as follows:

$$\text{Titer} = \frac{\text{(number of colonies) (dilution factor) (1000 µl/ml) colony forming units (cfu)}}{\text{(Volume of phage plated in µl)}}$$

2.6.5 Amplification of cosmid library

The cosmid library was amplified as described by the *pWEB-TNC™ Cosmid Cloning kit* instruction manual. An overnight culture of EPI100-T1^{R} cells (100 µl) were inoculated with 10 µl of the packaged phages. This was incubated in the dark at 37°C, shaking at 100 rpm for 20 minutes. After that, 0.5 ml of LB broth was then added to the infected culture and incubated at 37°C for 45 minutes. The infected culture (600 µl) was added and spread on 4 LB-ampicillin agar plates (90 mm). The plates were incubated upside down in the dark at 37°C overnight, for approximately 12-14 hours (until colonies, 0.2-0.3 mm in diameter, were noticed). The colonies were transferred to a sterile tube by adding 10 ml of LB broth and scrapping them off from the plates using an inoculation loop. The LB broth suspended colonies were then vortex mixed to disrupt any bacterial clumps and sterile glycerol added to a final concentration of 15%. The bacteria were then stored at -70°C in 500 µl aliquots.

2.7 Cellulase activity plate assays

Endo-glucanase and cellobiohydrolase activity plate assays were done on the primary cosmid library as described below.

2.7.1 Endo-glucanase activity assay

Endo-glucanase plate assays were performed according to a modified method described by Teather and Wood (1982). Colonies of the primary cosmid library were replica-plated on LB-ampicillin plates (90 mm) and incubated in the dark at 37°C for approximately 12-14 hours. A carboxymethyl cellulose (CMC) overlay supplemented with ampicillin was prepared. In the place of IPTG, 1 µl of 0.1 M lactose (Merk) was added to 1 ml of the CMC-ampicillin overlay media.

The CMC-ampicillin/lactose overlay mixture (4-8 ml) was added to the previously incubated LB-ampicillin agar plates and these were further incubated at 37°C for 12-16 hours, followed by staining with Congo red dye (1 mg/ml) for 15 minutes. The plates were then flooded with 1 M NaCl and left at room temperature for another 15 minutes. The appearance of a clear yellowish halo on a red background around the colonies was used to determine endo-glucanase activity. After suspected endo-glucanase activity was observed, hydrochloric acid (1 M) was then added to inhibit further enzyme activity. Hydrochloric acid also aids contrast and stabilise the visualized zones, by changing the dye colour from red to blue in the process.

2.7.2 Cellobiohydrolase activity assay

Cellobiohydrolase plate assays were done according to a slightly modified method described by Reinhold-Hurek *et al*., (1993).Colonies that had previously shown positive endo-glucanase activity were replica-plated on LB-ampicillin agar plates (90 mm) and incubated at 37°C for 14 hours. A 4-methylumbelliferyl-β-D-cellobioside (MUC) overlay supplemented with ampicillin and lactose was prepared as described for CMC overlay media above. The MUC overlay (4-8 ml) was then added to the LB-ampicillin agar plates and the plates were further incubated for 12-16 hours. Plates were then exposed to a 302 nm UV Transluminator and active colonies were identified by the appearance of a blue fluorescence.

2.8 Cosmid DNA isolation/purification

Colonies which showed positive CMC and MUC activity were inoculated in 150 ml of LB ampicillin broth. The culture was incubated in the dark at 37°C overnight (for approximately 12-14 hours) on a rotary shaker (100 rpm). Cells were harvested by centrifuging 100 ml of the overnight culture at 4,000 x g for 10 minutes at 20°C.

Cosmid DNA isolation was then performed using a *PureLink™ HiPure Plasmid DNA Purification MidPrep Kit* (Invitrogen, USA) according to a method described in the kit instruction manual until the precipitation step. The supernatant from the precipitation step, 500 µl, was put in 1.5 ml sterile tubes and a phenol: chloroform: isoamylalcohol (25:24:1) mixture (500 µl) added to the tubes. The tubes were centrifuged at 13,000 x g for 2 minutes at 4°C and the aqueous phase (400 µl) transferred to fresh sterile tubes. A 1X volume (400 µl) of chloroform was then added to the tubes containing the aqueous phase and the reaction mixed, followed by centrifuging at 13,000 x g for 2 minutes at 4°C. The top phase (250 µl) was transferred to new sterile 1.5 ml tubes and 1 ml of sodium acetate added to each tube.
The tubes were incubated at -70°C for 20 minutes after which there were then centrifuged at 13,000 x g at 4°C for 10 minutes and the supernatant discarded leaving the cosmid DNA pellet in the tubes. The tubes with the pellet were washed with 500 µl of 70% ethanol and centrifuged at 13,000 x g at 4°C for 5 minutes and again supernatant discarded. The tubes were air-dried at 37°C and the cosmid DNA re-suspended in 50 µl of TE buffer.

2.9 Cosmid DNA restriction digestion

Restriction digestion of the cosmid DNA was performed using *Bam* H I and *Not* I restriction enzymes (Fermentas, South Africa). The cosmid DNA, 20 µl, was mixed with 2.5 µl of 10X buffer, 1 µl of restriction enzyme and 1.5 µl of nuclease free water (to make a total reaction volume of 25 µl). The reaction mixture was thoroughly mixed and incubated at 37°C for 6-12 hours. Agarose gel (1%) electrophoresis at 100V for 60 minutes was done for the restriction digests (5 µl) using GeneRuler™ 1 kb and GeneRuler™ 1 kb Plus (Fermentas, South Africa) as DNA ladders.

2.10 Sequencing

The isolated cosmid DNA samples were sent for sequencing to Inqaba Biotech.

2.11 Sequence analyses

This was done using the National Centre of Biotechnology Information online database (http://www.ncbi.nlm.nih.gov/) and the FASTA, BLASTN and BLASTX tools found on that website.

CHAPTER 3

Results

Positive endo-glucanase and cellobiohydrolase activities were observed from the primary cosmid library containing about $1.2 * 10^3$ colony forming units, constructed with the genomic DNA isolated from the cellulose-cultured *E. intermedius* larvae gut micro-flora. Partial sequencing results attributed the observed cellulase activities to bacteria belonging to the *Enterobacteriacea* family, most of which have no previous cellulase activity record.

To kick start this study, a total of 144 dung beetles were collected from the farm below, **Figure 4**, around August and September 2010 to create the dung beetle rearing culture. According to Edwards (1991), this is regarded as the best time to collect the dung beetles in Johannesburg (South Africa), due to the high rainfall patterns, which results in high reproduction and survival rates. It was reported that in *E. intermedius*, egg production at 25°C ranged from 0-1 per female per week in winter to 1-12 in summer (Edwards, 1991).

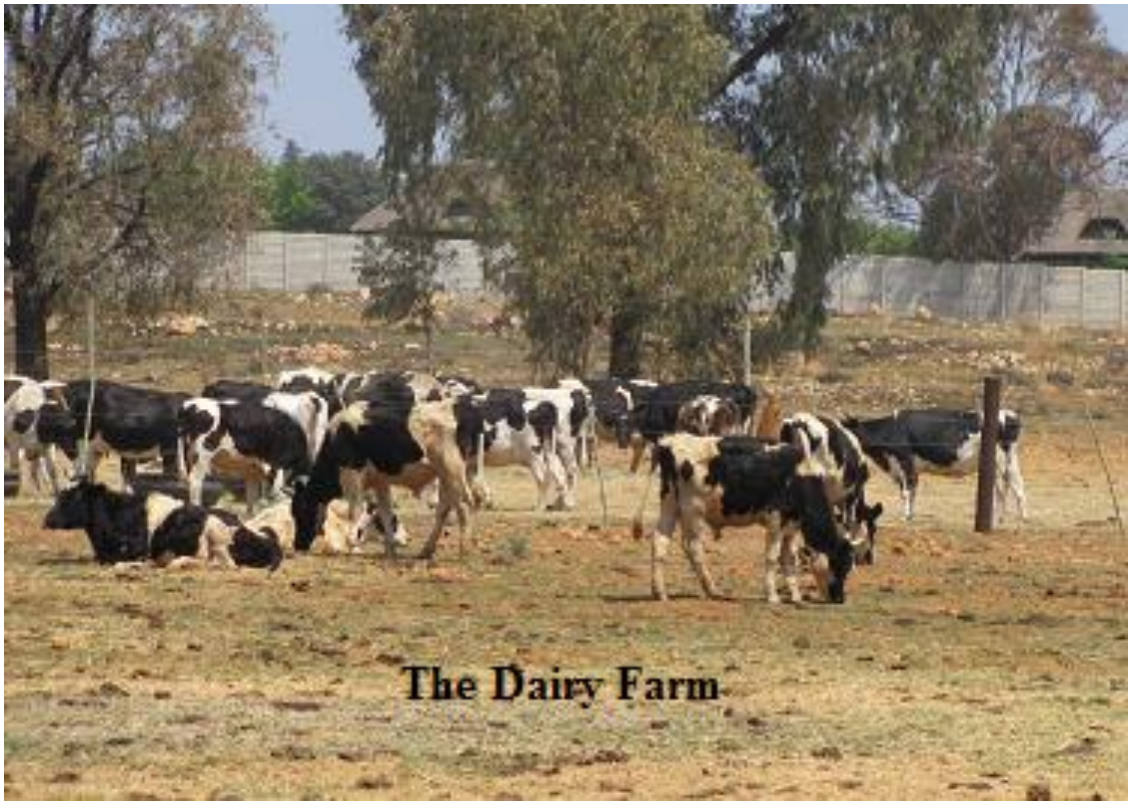


Figure 4: Image of the dairy farm where adult dung beetles and cattle dung were collected.

The adult dung beetles were collected from deep inside the dung or underneath it, as shown in **Figure 5**. The beetles were easier to find in fresh dung as compared to the older dung with low moisture levels. Fresh cattle dung contains high moisture levels and has the largest proportion of the fluid component of dung, which is preferred by adult dung beetles (Edwards, 1991). Residing deep in the dung or underneath it has also been suggested to be a behavioural response by dung beetles to protect themselves from excessive summer temperatures.

Figure 5: Fellow students collecting live dung beetles from cattle dung. The collected adult dung beetles were temporarily put into perforated screw-capped plastic containers.

3.1 Dung beetle rearing

The 144 collected dung beetles were sorted into 17 males and 127 females using the presence or absence of a horn to differentiate the sex. Male dung beetles were identified by the presence of a horn on the head as shown below, **Figure 6**, a feature which is absent in the female dung beetles. Female dung beetle numbers were found to be more than 6 times the male dung beetle numbers. Instead of using 3 pairs of dung beetles in equal male and female proportion, two male dung beetles were paired with four female dung beetles. The rationale behind this pairing was that one male is capable of sexually reproducing with more than one female. Most male dung beetles were also observed to be larger in size as compared to their female counterparts.

Figure 6: Photographic images of female and male dung beetles. Male dung beetles are identified by the presence of a horn on the head, as shown in the picture.

The dung beetle pairs were reared in plastic containers, which were half-filled with compacted soil and cattle dung placed on top. The soil component provided the beetles with tunnel burrowing space to make the brood balls. The brood balls are a protective structure made from a mixture of dung and soil in which female dung beetles lay their eggs, **Figure 7**.

Figure 7: Photographic image of female dung beetle laying eggs inside a brood ball.

An average of 20-30 brood balls per week were harvested from each plastic container with dung beetle breeding pairs. The brood balls from different containers were then pooled together and placed in a separate plastic container, half-filled with moist sand. A wet sponge was placed on top of the sand to maintain the desired moisture levels. The soft dung found inside the brood balls is a future food source for the developing larvae, after the hatching of the egg, **Figure 8**.

Figure 8: Photographic image of dung beetle egg inside a brood ball.

The dung beetle larvae developed inside the brood ball, **Figure 9**, and their nourishment came from the dung contained therein. Unlike the adult dung beetles which reportedly feed on the fluid component of the dung, the larvae feed by ingesting the whole dung particles (Edwards, 1991). To propagate the dung beetle culture, dung beetle larvae were initially allowed to grow for 4-5 weeks and develop into young beetles. Emerging young dung beetles were collected and kept in single-sex plastic containers for 2 weeks to allow them to develop into sexually mature beetles (Lailvaux *et al*., 2005; Pomfret and Knell, 2006).

Figure 9: Photographic image of dung beetle larvae developing inside a brood ball. This is the stage at which the dung beetle larvae were collected for dissection.

3.2 Dung beetle dissection

Once the desired number of adult beetles was achieved, all dung beetle larvae were allowed to develop for approximately 2-3 weeks before harvesting them for dissection. This allowed the larvae to grow into late second instar or third instar, as shown in **Figure 10** below. The gut of dissected larvae showed two distinct regions, an elongated mid-gut occupying most of the larvae's body length and a bulbous hind-gut, commonly referred to as the fermentation chamber (Moyo, 2011). These gut regions are labelled **M** and **H** in **Figure 11** on the next page.

Figure 10: Image of dung beetle larva ready for dissection. Dung beetle larvae were dissected during late second or third instar stages of development.

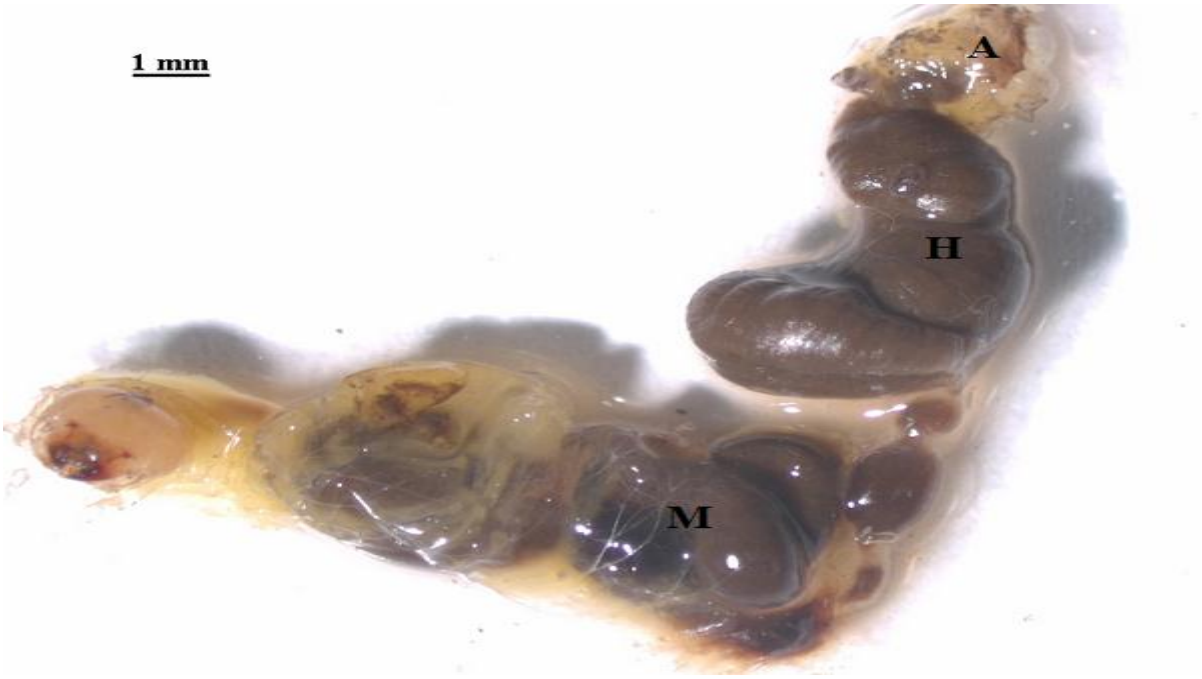


Figure 11: Image of the gut of dissected *E. intermedius* larva. The gut consists of the mid-gut (**M**) and hindgut (**H**) as shown in the picture. The anal opening is represented by **A**.

3.3 Culturing of dung beetle gut micro flora

The homogenised dung beetle larvae were cultured in cellulose medium containing small pieces of filter paper which can be easily seen in **Figure 12,** below. In place of filter paper, non-cellulose medium contained glucose and lactose as carbon sources.

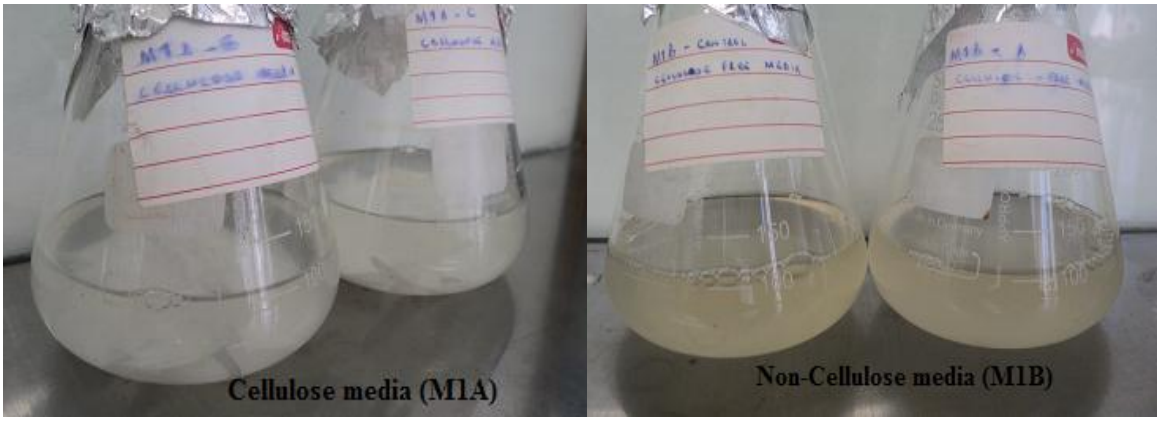


Figure 12: Photographic images of cellulose medium and non-cellulose medium before inoculation with pre-culture samples. Filter paper was used as a sole carbon source in the cellulose media preparation.

After 28 days of incubation, the cellulose media cultures, **B**, contained small particles of filter paper suspended in the liquid medium, as shown in **Figure 13**.

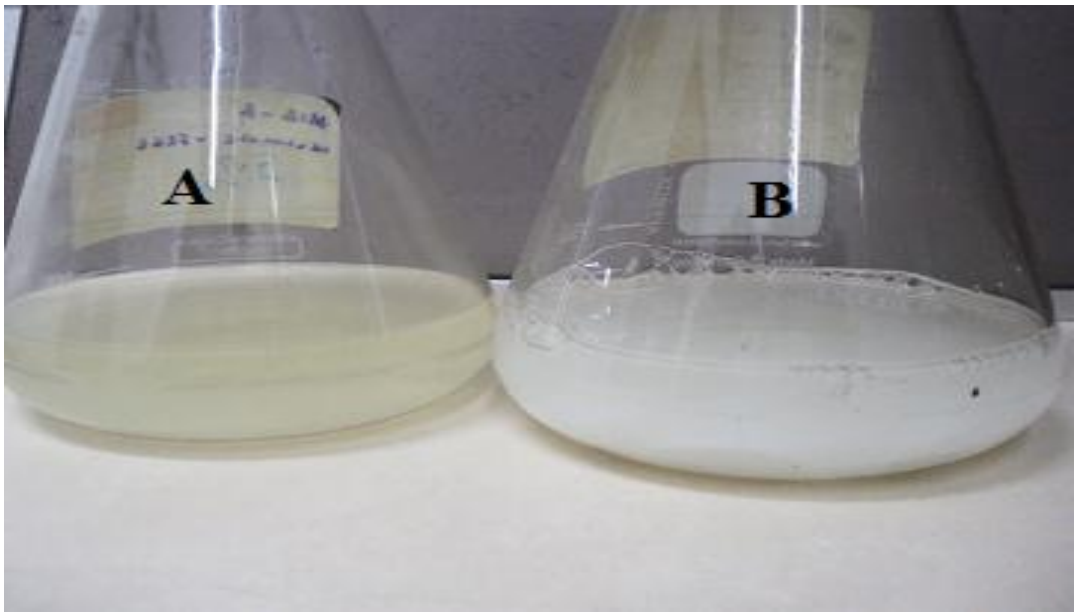


Figure 13: Photographic image of **A**: non-cellulose and **B**: cellulose media cultures after 28 days of culturing.

3.4 Bacterial growth determination

In order to monitor bacterial growth in both the cellulose and non-cellulose media cultures and to determine the best time to perform successful DNA isolation on the cellulose medium-cultures, the time (number of days) required to achieve maximal bacterial growth was established by total protein analysis.

Bacterial growth was determined as a function of total protein concentration, using the Bradford protein assay procedure. Due to limited bacterial growth, a micro-assay standard curve with a straight line function: **y = 0.0294x – 0.35** and **R^2 value of 0.9431** was used to determine total protein concentration in both M1A and M1B media cultures, as shown below in **Figure 14**.

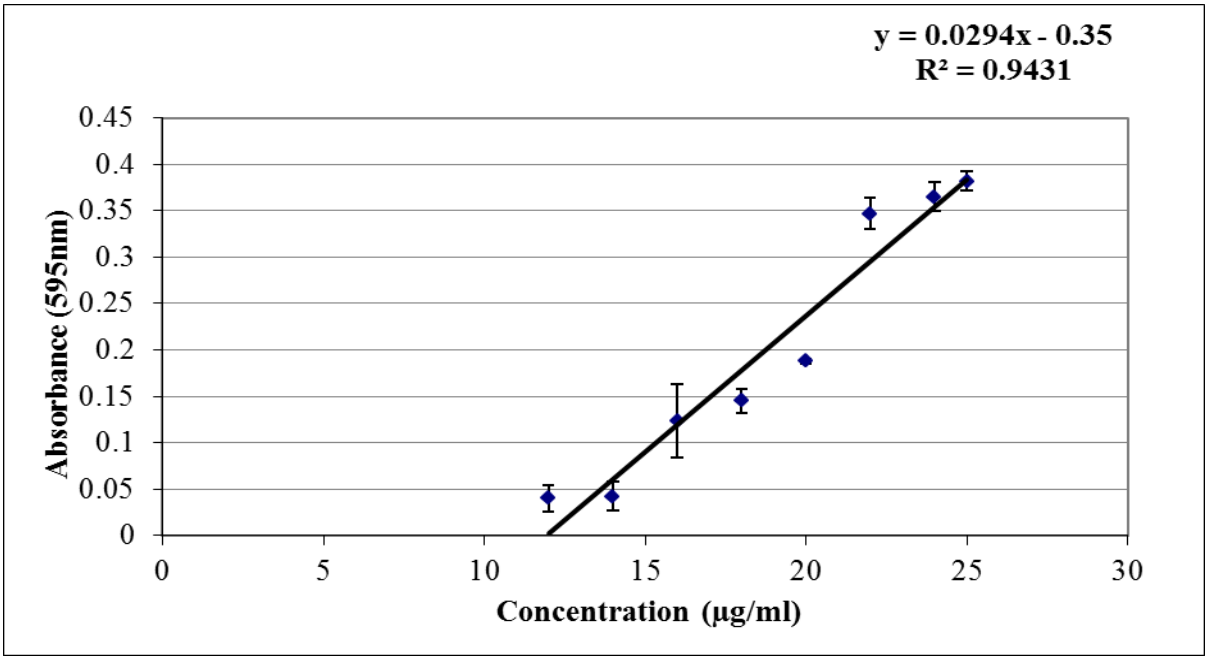


Figure 14: Trend line of the standard curve for total protein analysis. Total protein analysis method was used to determine bacterial growth in the culture samples.

A general increase in total protein concentration or bacterial growth was observed in both the cellulose and non-cellulose media cultures from Day 0 (15.2 µg/ml for cellulose culture and 13.14 µg/ml for non-cellulose culture samples) to Day 10 (24 µg/ml for M1A and 23 µg/ml for M1B media cultures) as shown below, **Figure 15 and Figure 16**. The total protein concentration levels in M1A medium culture then dropped from Day 11 to Day 13, and rose again until it reached a maximum of 25.34 µg/ml on Day 15 (**Figure 15**).

In M1B medium culture, protein concentration dropped from Day 10 to day 11, and rose again until it reached its highest value, 24.5 µg/ml, on Day 15 (**Figure 16**). A steady decline in total protein concentration was then observed in both M1A and M1B media cultures from Day 15 until Day 28.

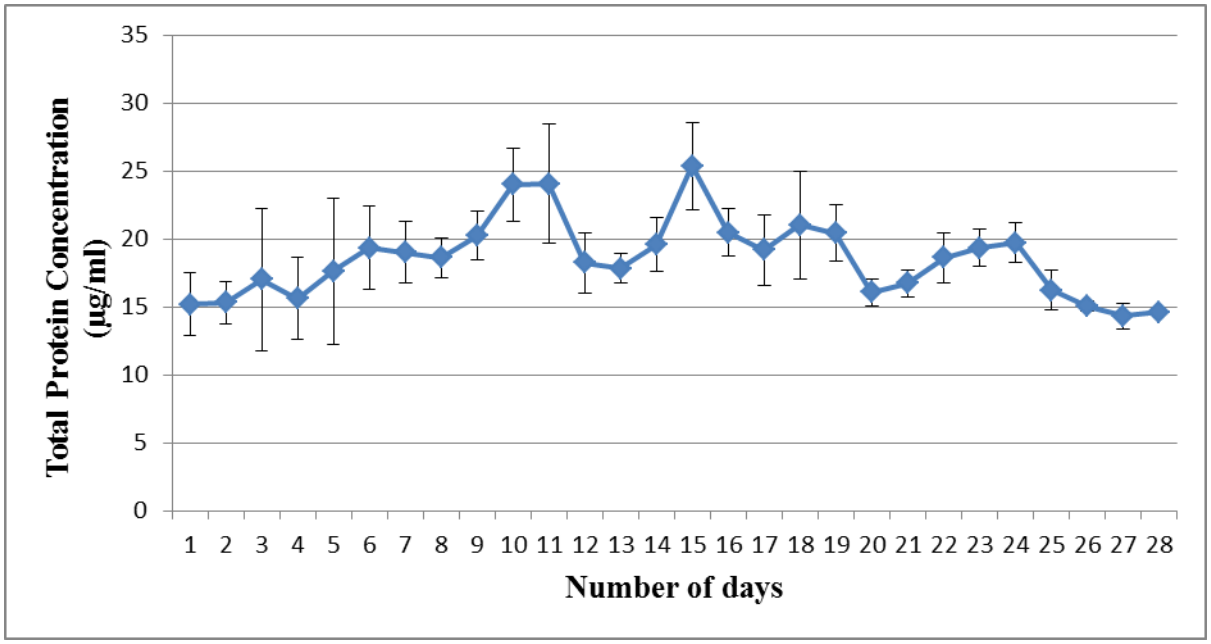


Figure 15: Total protein concentrations of cellulose media culture samples (Day 1-Day 28).

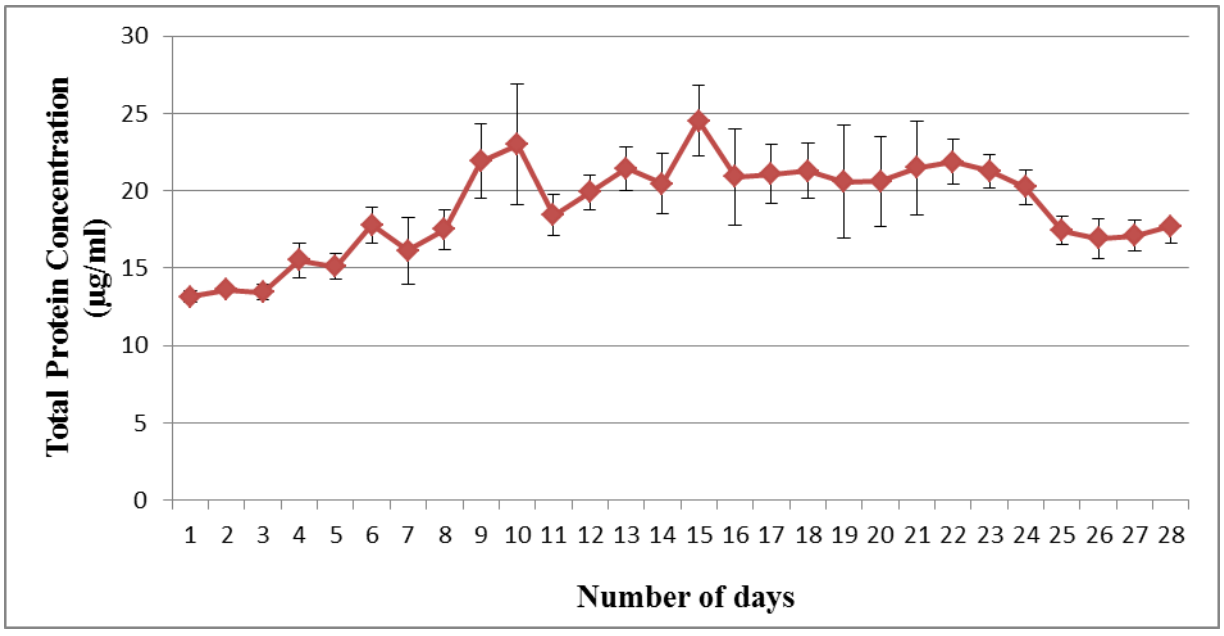


Figure 16: Total protein concentrations of non-cellulose media culture samples (Day 1-Day 28).

The cellulose medium culture showed slightly higher bacterial growth from Day 0 to around Day 8, when compared to the non-cellulose medium culture, as shown in **Figure 17**. However, from Day 15 to Day 28, the non-cellulose medium culture maintained a higher bacterial population than the cellulose medium culture.

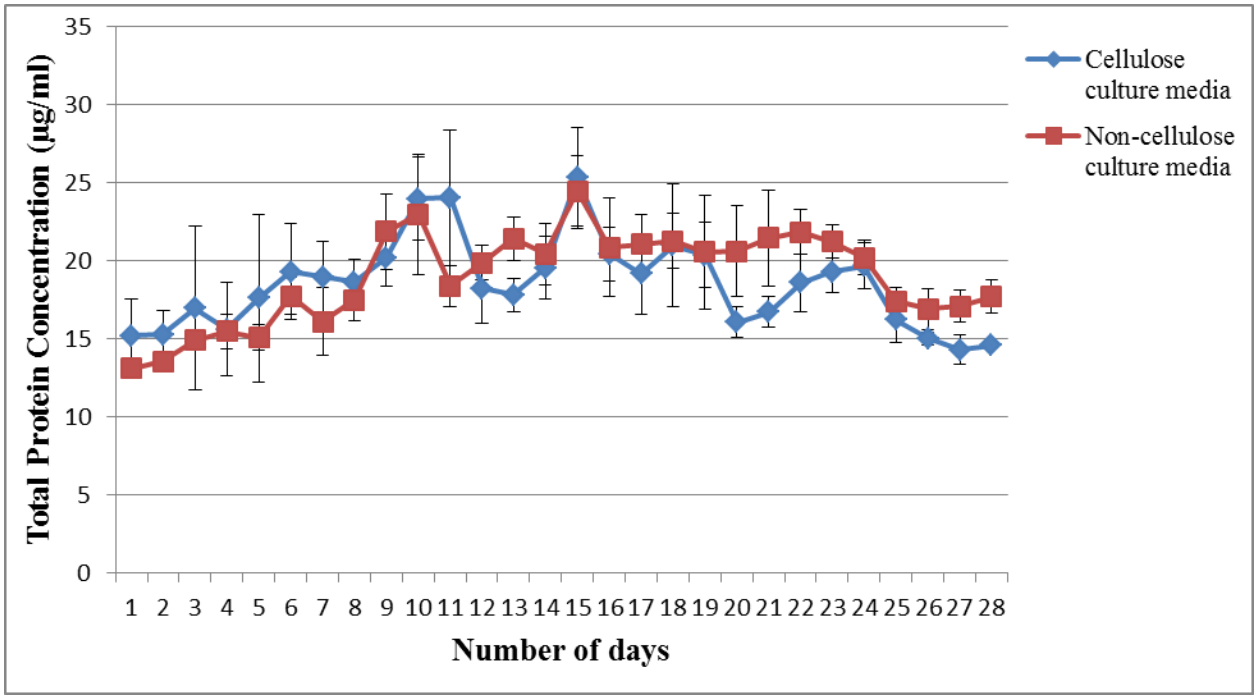


Figure 17: Total protein concentrations of M1A and M1B media culture samples (Day 1-Day 28).

3.5 Cellulase activity determination

The ability of the cultured dung beetle gut micro-flora to utilize cellulose (filter paper) as the only carbon source was demonstrated by the ability of the gut micro-flora to grow in cellulose medium, as shown in **Figure 15** and **Figure 17** above. It was hypothesized that the gut micro-flora were able to survive in the M1A medium culture because of its ability to degrade cellulose and then utilizing the by-products of the degradation process to meet its energy and growth requirements. To prove this hypothesis, the amount of reducing sugars released into the cellulose medium culture was determined as a measure of the cellulose degradation activities of the cultured dung beetle gut micro-flora. The cellulase activity of the gut micro-flora was regarded to be directly proportional to the reducing sugar concentration of the M1A medium culture.

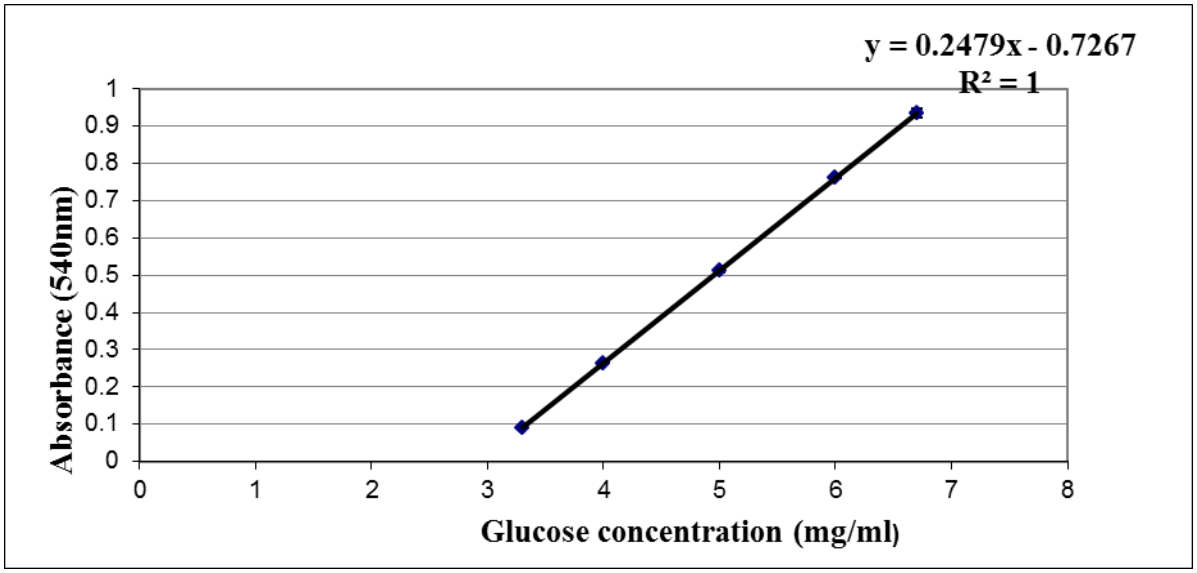


Figure 18: Trend line of the standard curve for cellulase activity analysis. Cellulase activity of cultured dung beetle larvae gut micro-flora was determined against a glucose standard.

Cellulase activity of the cultured micro-flora was determined by measuring the concentration of reducing sugars in using the DNS method. A standard curve with a straight line equation of **y = 0.2479x – 0.7267** and **R^2 value of 1** was used to determine the concentration of reducing sugars released through the degradation of cellulose, in M1A culture, **Figure 18** above.

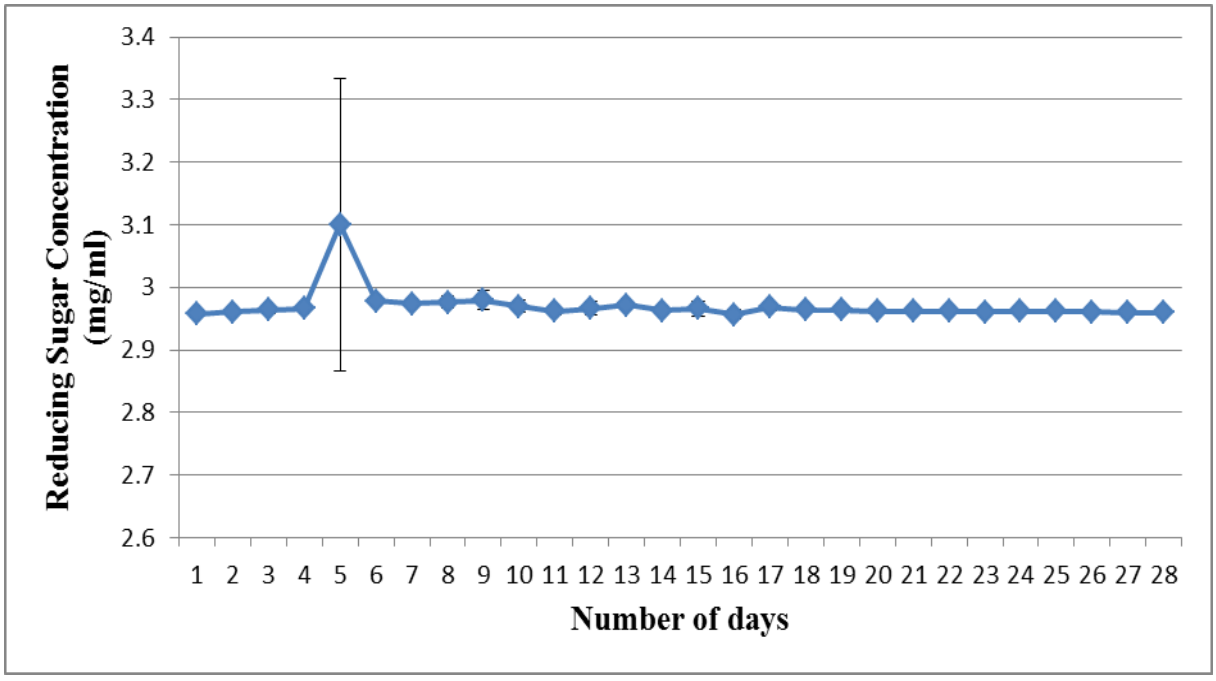


Figure 19: Reducing sugar concentrations of cellulose media culture samples (Day 1-Day 28).

Information from **Figure 19** above shows that reducing sugar concentrations in M1A medium culture remained relatively constant at 2.9 mg/ml from Day 0 to Day 28, with an exception of Day 5 where a concentration of 3.1 mg/ml was observed. Taking a closer look at the values, however, reveals that a slight increase was actually observed from Day 0 (2.958 mg/ml) to Day 9 (2.980 mg/ml). This accounts for an increment in reducing sugar concentration of 0.022 mg/ml, which is quite a significant difference. A steady decline was then observed from Day 10 to Day 28 (2.96 mg/ml), due to the declining bacterial growth, **Figure 15**.

One may also argue that the reducing sugar concentrations shown in **Figure 19** above do not truly represent the cellulose activity of the gut micro-flora. The ability of the non-cellulose medium cultured gut micro-flora to utilize glucose and lactose (M1B medium) can also mean that probably large amounts of reducing sugars might have been produced, but these were immediately consumed by the gut micro-flora before being accounted for.

3.6 Genomic DNA isolation

DNA isolation was done on Day 10 samples of the cellulose media culture after realizing that bacterial growth was at its maximum during that period (**Figure 15**). Due to poor bacterial growth levels, DNA isolation was done on many samples to yield a total amount of DNA of 41.2077 µg, **Figure 21**.

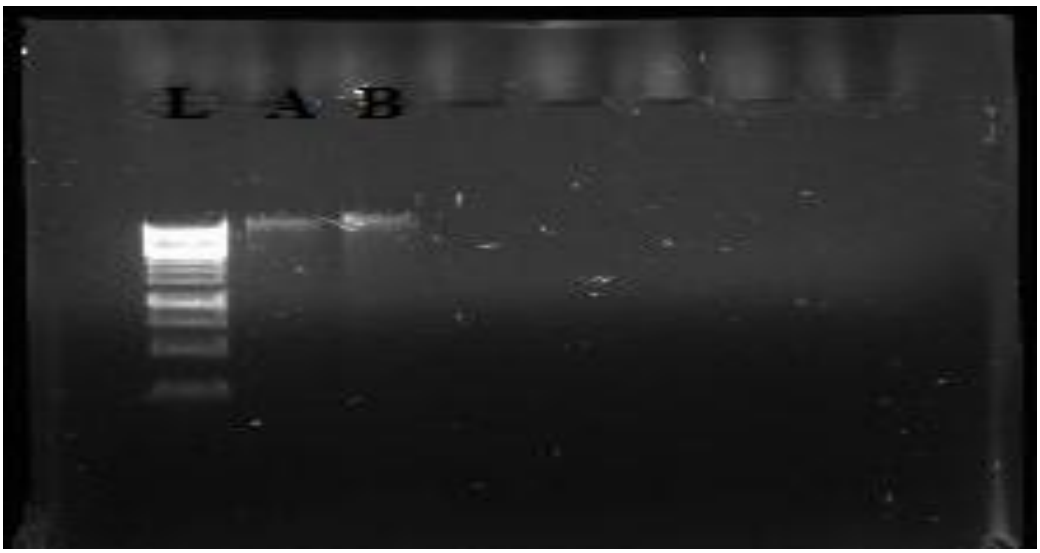


Figure 20: Gel electrophoresis picture of isolated genomic DNA replicate samples (**A** and **B**). A 100 bp DNA ladder was used (**L**).

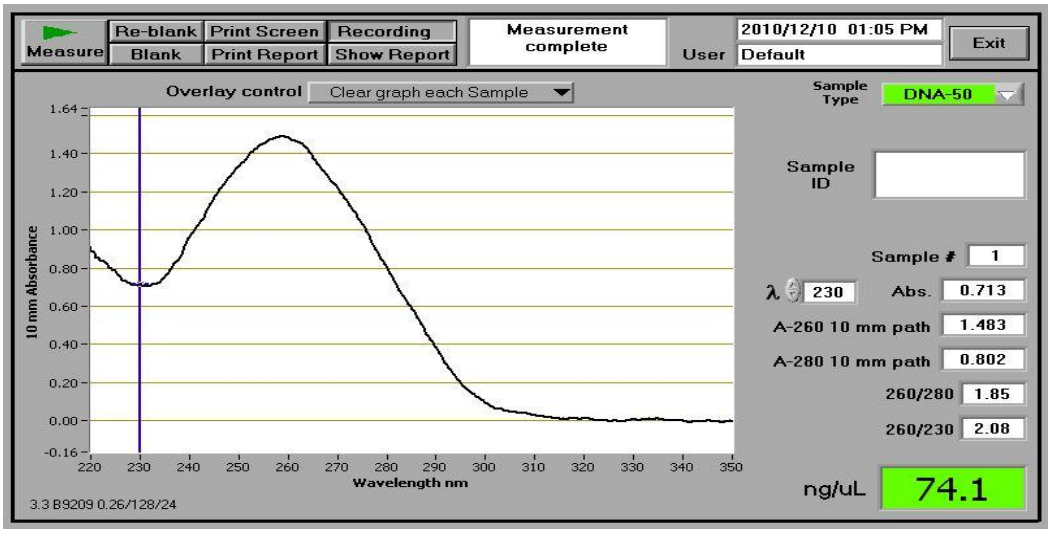


Figure 21: Concentration of one the isolated DNA samples. A total genomic DNA amount of 41.2077 µg was pooled together from 13 DNA extracted samples.

3.7 Cosmid DNA library construction

A cosmid DNA library is a collection of cloned fragments of genomic DNA inserted into a cosmid cloning vector. These clones are stored and propagated in a population of micro-organisms (host) through the process of molecular cloning. The purpose of creating a cosmid library is to investigate entire structures of genes and to study regulation of gene expression. This is also essential for genetic transformation. In this project a primary cosmid library was generated so as to study gene expression with regards to cellulase activity of *E. intermedius* larvae gut micro-flora.

To construct a complete and unbiased primary cosmid library, genomic DNA was first sheared using a syringe needle to generate highly random DNA fragments (approximately 20-40 kb). This is in contrast to more biased libraries that result from fragmenting the DNA by partial restriction endonuclease digestion.

3.7.1 Insert DNA size selection

The isolated and sheared genomic DNA was run on a gel to determine if it was the right size for successful insertion into the cosmid vector, **Figure 22** below.

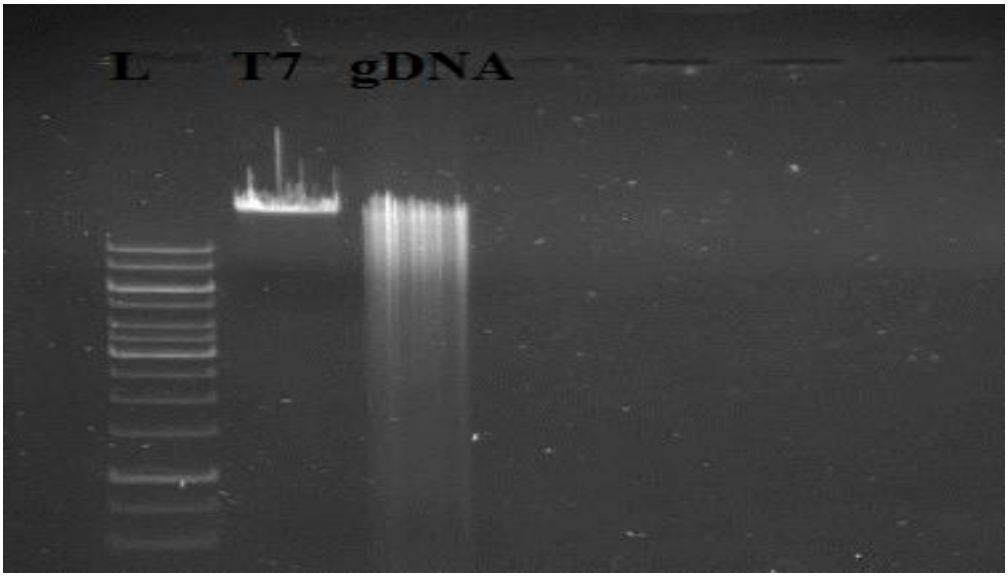


Figure 22: Gel electrophoresis image of genomic DNA (**gDNA**) as compared against the **T7** cosmid control DNA. The position of genomic DNA compared to that of the cosmid control DNA, indicates that it is the right size (20-45 kb) for the End-Repair reaction. A 1 kb DNA ladder (**L**) was used.

3.7.2 Titer determination

Determination of the titer of the packaged cosmids is important to determine the efficiency of the packaging reaction. Titer was calculated using the average number of colonies obtained per 10 µl of packaged phage inoculated in an LB-ampicillin agar plate (90 mm), since the experiment was done in quartet.

The average number of colonies obtained in the four LB-ampicillin agar plates used were 12. Therefore:

$$\text{Titer} = \frac{(\text{number of colonies})\,(\text{dilution factor})\,(1000\ \mu l/ml)}{(\text{volume of phage plated in } \mu l)}\ \text{colony forming units (cfu)}$$

$$= \frac{(12)\,(1)\,(1000\ \mu l/ml)}{10\ \mu l}\ \text{cfu}$$

$$\text{Titer} = 1.2 * 10^3\ \text{cfu/ml}$$

Since the total volume of packaged phage was 1000 µl or 1.0 ml, the total number of colonies contained within the library was calculated as follows:

Total number of colonies = Titer * Total volume of packaged phage

$= 1.2 * 10^3$ cfu/ml * 1.0 ml

$= \underline{1.2 * 10^3 \text{ cfu}}$

3.7.3 Amplification of cosmid library

Cosmid libraries are generally unstable at 4°C for long term storage when stored as packaged phage suspended in phage dilution buffer. In order to preserve the cosmid library for cellulose activity plate assays, the cosmid library was amplified. During amplification, the total numbers of colonies observed in all the four LB-ampicillin agar plates used were 378, **Figure 23**. These were scrapped off the plates using LB broth and preserved in 15% glycerol at -70°C.

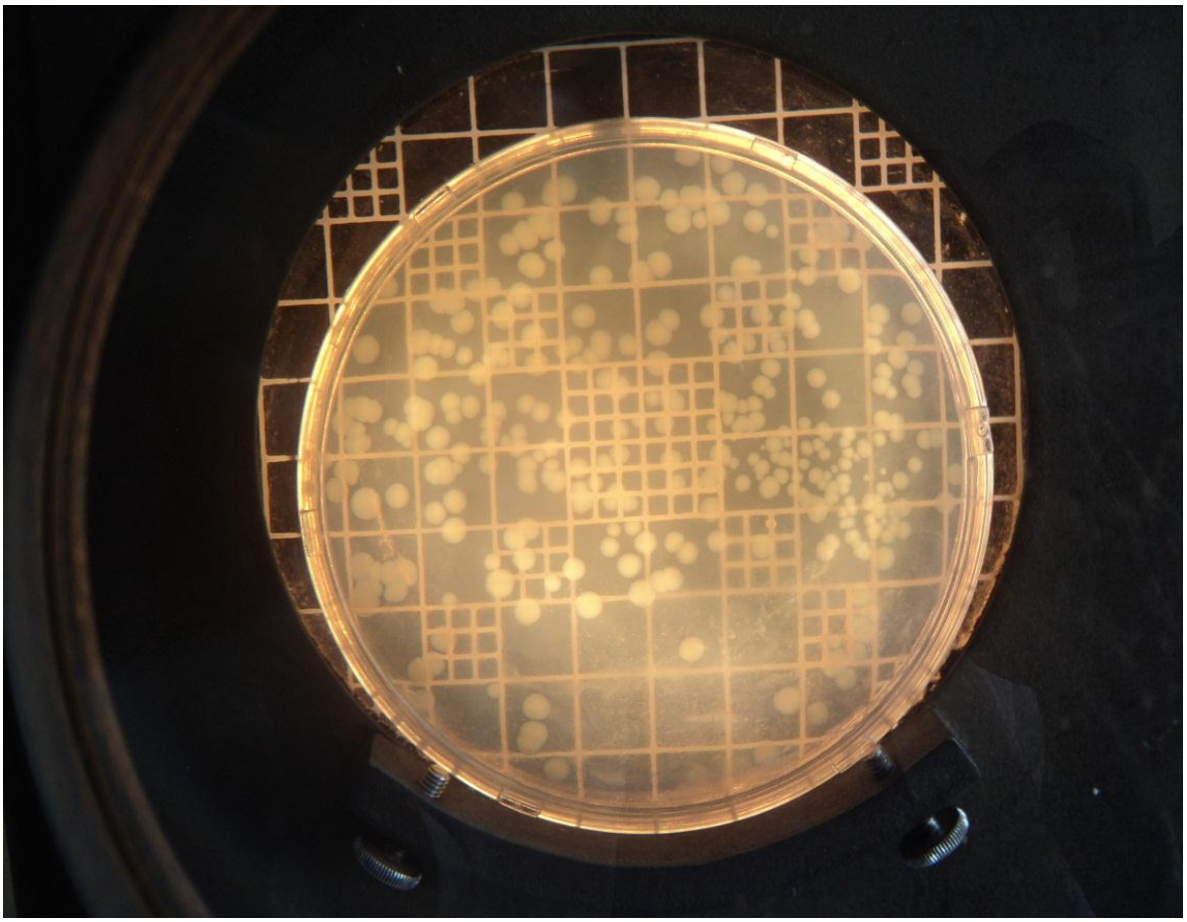


Figure 23: A photographic image of one of the LB-ampicillin plates used for cosmid library amplification. The total number of colonies observed in the four LB-ampicillin plates used for amplification was 378.

3.8 Cellulase activity screening using plate assays

Plate assays were used to screen for cellulase activities of the primary cosmid library. The cellulose activities that were investigated are endo-β-1,4-glucanase activity and cellobiohyrolase activity. No β-glucosidase activity assay was done.

Endo-glucanases (EC 3.2.1.4), degrade cellulose by randomly cleaving the β-1,4-glycosidic bonds on cellulose chains away from the chain ends. Cellobiohydrolases (EC 3.2.1.91), on the other hand, produce cellobiose by attacking cellulose from the chain ends from the reducing ends, cellobiohydrolase I, and also from the non-reducing ends of the cellulose chain, cellobiohydrolase II. Lastly, β-glucosidases (EC 3.2.1.21) convert the produced cellobiose to glucose (Bansal *et al*., 2009).

3.8.1 Endo-β-1,4-glucanase (CMC) activity

Positive endo-β-1,4-glucanase activity was observed in 3 colonies out of a total of 77 screened colonies from 10 different LB-ampicillin agar plates. Positive activity was observed by the appearance of a clear halo around the colony on a blue background as shown in **Figure 24** and **Figure 25**, below.

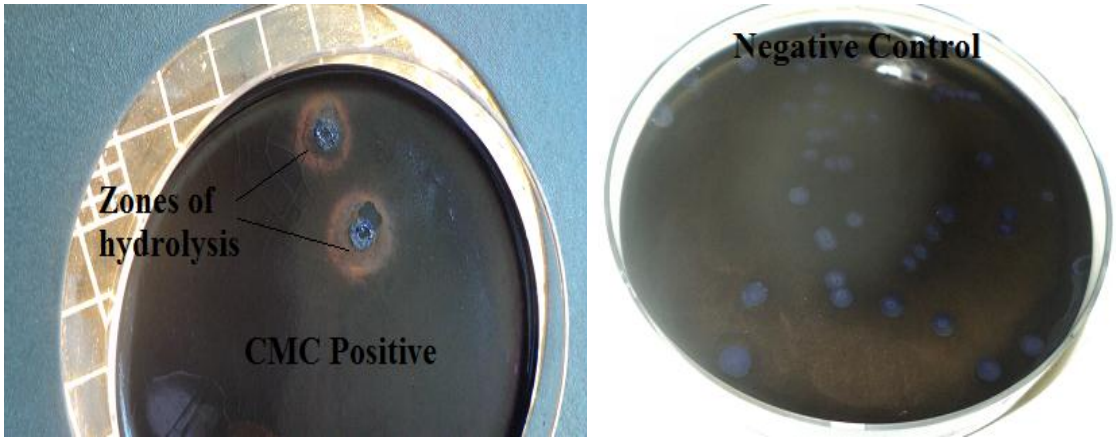


Figure 24: Photographic images of colonies exhibiting positive and negative endo-β-1,4-glucanase activity, respectively.

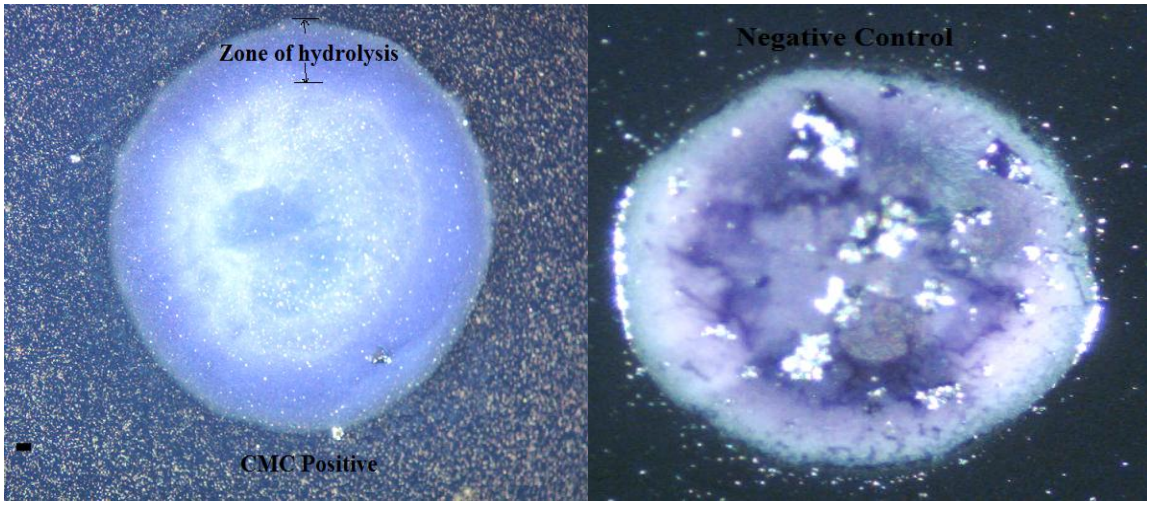


Figure 25: Microscopic images of colonies exhibiting positive and negative endo-β-1,4-glucanase activity, respectively. There is no zone of hydrolysis around the negative control colony.

3.8.2 Cellobiohydrolase (MUC) activity

MUC activity was observed in at least 4 colonies out of a total of 83 colonies screened for MUC activity. Positive colonies exhibited a blue fluorescence when exposed to a 302 nm UV radiation on a UV Transilluminator, as shown below (**Figure 26**).

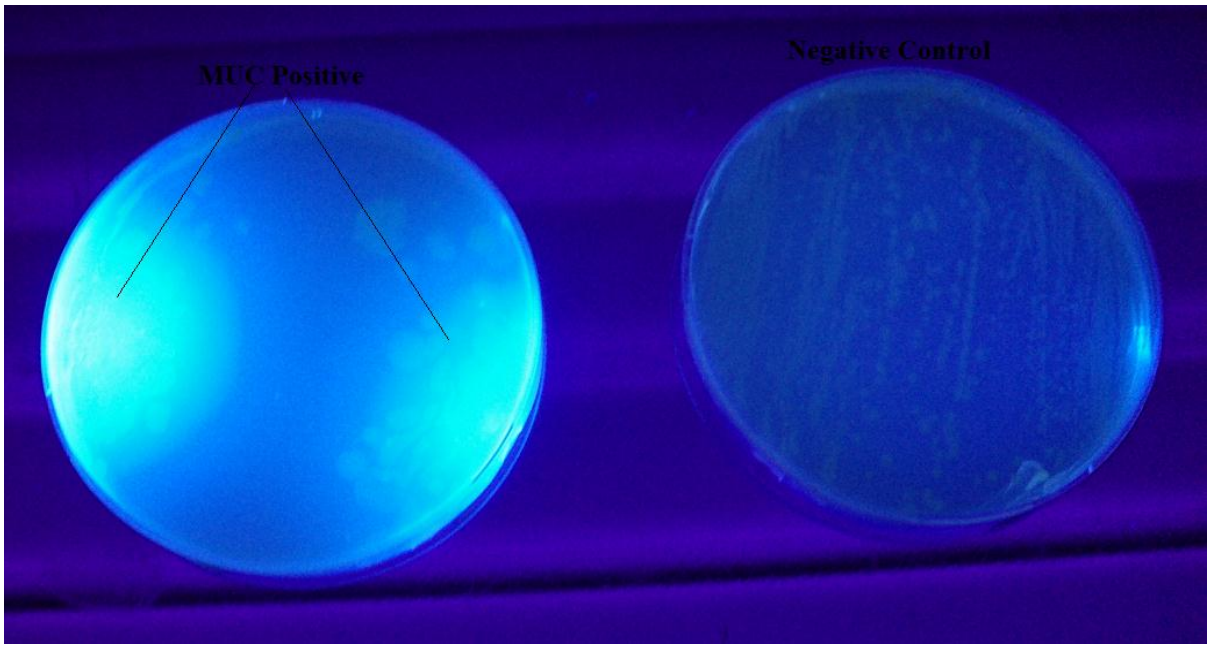


Figure 26: Photographic image of positive MUC activity, as indicated by blue fluorescing colonies. The negative control showed no blue fluorescence.

3.9 Cosmid DNA isolation from colonies with positive MUC/CMC activity

Isolation of cosmid DNA was done on the colonies which had shown both positive CMC and MUC activity. The isolated cosmid DNA, **Figure 27** and **Figure 28**, was the sent for sequencing to Inqaba Biotech.

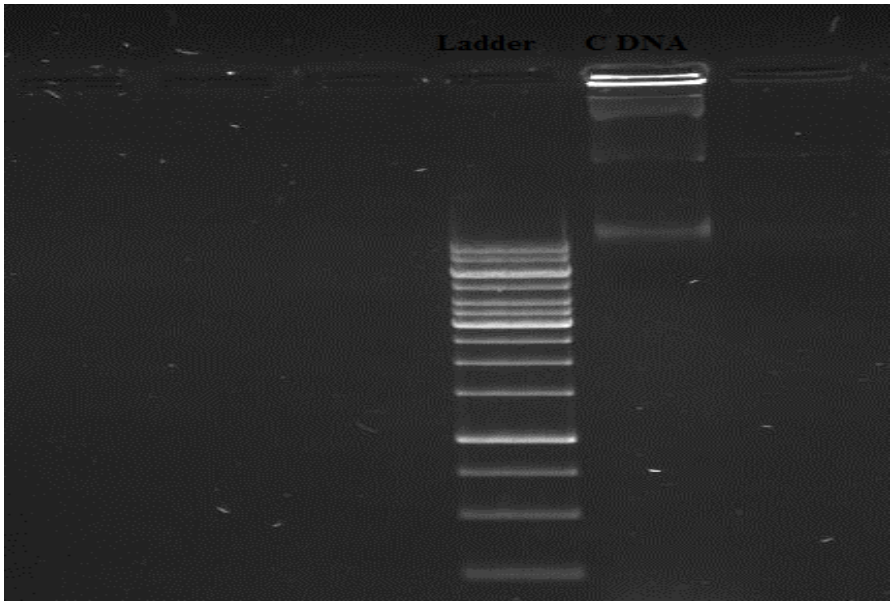


Figure 27: Gel electrophoresis image of extracted cosmid DNA, **C DNA**. **Ladder**: 1 kb DNA ladder was used. Due to the large size of the cosmid DNA, most of the DNA remained trapped in the well during electrophoresis. A total cosmid DNA amount of 32.29 µg was extracted.

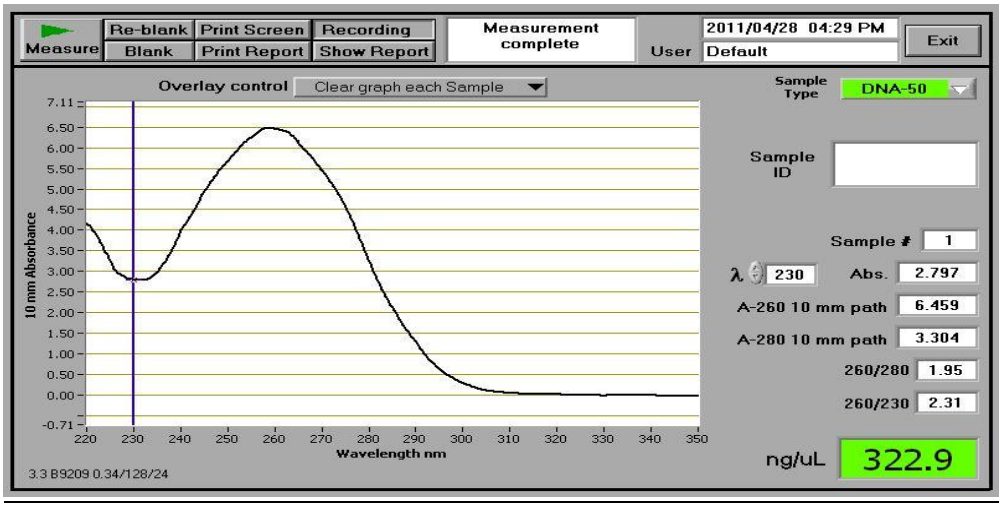


Figure 28: Cosmid DNA concentration of colonies with positive MUC/CMC activity.

3.9.1 Restriction digestion of cosmid DNA

Restriction digestion of cosmid DNA isolated from the colonies with positive endo-glucanase and cellobiohydrolase activity was done to determine the insert size of the DNA.

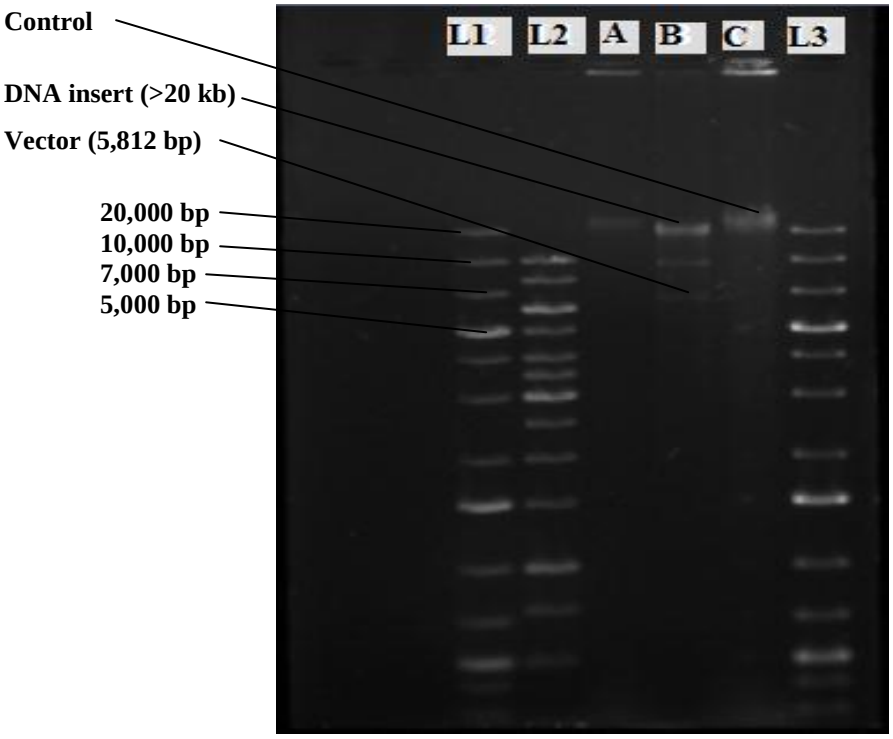


Figure 29: Image of restriction digested cosmid DNA. **L1**: DNA ladder, GeneRuler^{TM} 1 kb Plus; **L2**: DNA ladder, GeneRuler^{TM} 1 kb; **L3**: DNA ladder GeneRuler^{TM} 1 kb Plus; **A**: *Not* I restriction digested cosmid DNA; **B**: *Bam* H I restriction digested cosmid DNA; and **C**: Undigested cosmid DNA control.

Restriction digestion was performed using *Bam* H I and *Not* I restriction enzymes and gel electrophoresis of the restriction products done as shown in **Figure 29**. Only *Bam* H I catalysed restriction yielded multiple bands (>20,000 base pairs, approximately 10,000 base pairs and slightly less than 7,000 base pairs), **Figure 29: B**, after gel electrophoresis. Using the GeneRuler^{TM} 1 kb Plus DNA ladder (**L1** and **L3**) as a reference, the DNA insert size was estimated to be >20,000 base pairs or >20 kb, corresponding to the largest band and the known pWEB-TNC^{TM} vector size (5,812 base pairs) corresponded to the band slightly less than 7,000 base pairs. It was also observed that the undigested cosmid DNA (control) band was slightly larger than that of the insert DNA

3.10 Cosmid DNA sequencing

To identify the bacterial species responsible for the observed cellulase activities, the cosmid DNA isolated from the colonies expressing positive MUC/CMC activity was sent for sequencing to Inqaba Biotech (South Africa). Sequencing was done by the standard Sanger method from both ends of the cosmid DNA insert (approximately 1kb), using M13F and T7 as primers. However, due to the large size of the cosmid DNA insert (>20 kb), and the difficulties encountered in designing the required subsequent primers, sequencing of the entire insert DNA was not complete at the time of submission of this study. As such, only partial sequencing results (approximately 1 kb sequences from both ends of the cosmid DNA insert: M13F and T7 sequences) were used in the compilation of this thesis.

The T7 derived sequence did not produce any significant BLAST (BLASTN and BLASTX) results, most likely due to the short length of the produced nucleotide sequence. On the other hand, the M13F derived sequence BLAST results (BLASTN) produced bacterial genes (dominated by *E. coli* genes) most of which belonged to the *Enterobacteriaceae* family and a few unidentified/uncultured clones, as well as one or two cloning vectors as illustrated below, **Table 12**. There was no significant nucleotide sequence homology to known cellulase genes in the databases (BLASTN). BLASTX results produced mainly peptides or proteins closely related to bacterial species (dominated by *E.coli* species) from the *Enterobacteriaceae* family and a few proteins of unknown function. These results are shown in **Table 11**, below.

Table 10: Selected M13F sequence BLASTN results (NCBI)

Accession number (s)	Description	Query Coverage	E (expect) value	Max Identity
AM946981.2 CP001509.3	*Escherichia coli* BL21(DE3), complete genome	98%	0.0	99%
EU078592.1	*Enterobacteria* phage DE3, complete genome	98%	0.0	99%
U39286.1	Cloning vector TLF97-3, phage lambda lacZ translational fusion vector, complete sequence	98%	0.0	99%
AC198536.1	*Monosiga brevicollis* clone JGIACYI-5B3, complete sequence	98%	0.0	99%

AC150248.3	Uncultured bacterium clone zdt-9n2, complete sequence	98%	0.0	99%
CP001063.1	*Shigella boydii* CDC 3083-94, complete genome	62%	6e-176	94%
AE005674.2	*Shigella flexneri* 2a str. 301, complete genome	66%	2e-130	86%
FN543502.1	*Citrobacter rodentium* ICC168, complete genome	11%	9e-40	94%
AJ556162.1	Phage BP-4795 complete genome	13%	6e-22	83%

Table 11: Selected M13F sequence BLASTX results (NCBI)

Accession number (s)	Description	Query Coverage	E (expect) value	Max Identity
EFX10836.1	putative host specificity protein [*Escherichia coli* O157:H7 str. G5101]	85%	1e-128	95%
EFW49240.1	Phage tail fiber protein [*Shigella dysenteriae* CDC 74-1112]	87%	1e-128	96%
EFU32434.1	fibronectin type III domain protein [*Escherichia coli* MS 85-1]	87%	4e-128	95%
EGT68149.1	hypothetical protein C22711_2179 [*Escherichia coli* O104:H4 str. C227-11]	87%	2e-123	92%
ZP_07590551.1	Protein of unknown function DUF1983 [Escherichia coli W] >gb|EFN39579.1| Protein of unknown function DUF1983 [*Escherichia coli* W] >gb|ADT75156.1| hypothetical protein ECW_m1685 [*Escherichia coli* W] >gb|ADX50870.1| protein of unknown function DUF1983 [*Escherichia coli* KO11FL]	87%	5e-121	95%

YP_860048.1	unnamed protein product [*Escherichia coli* APEC O1] >gb|ABJ03924.1| putative tail component of prophage [*Escherichia coli* APEC O1]	87%	9e-120	93%

CHAPTER 4

Discussion

The impetus to capture new biotechnologies for lignocellulose matter degradation for use in the production of second generation bio-fuels and other green chemicals has now reached new heights. This is highlighted by the extensive domination of current biotechnological research by the genomic and metagenomic mining of novel enzymes or bio-catalysts from animal and insect gut micro-biomes (Wang *et a*l., 2008).

In this study, a cosmid DNA library was constructed from the genomic DNA of *E. intermedius* larvae gut micro-flora, which were cultured in a cellulose-enriched medium. The library exhibited both endo-glucanase and cellobiohydrolase activities. This shows that the gut micro-flora of *E. intermedius* larvae contain have cellulose degrading capabilities. These observed results were in agreement and confirmed previous research findings of many scientists who postulated that scarab beetle larvae had cellulose degradation capabilities due to the existence of a symbiotic relationship with their cellulolytic gut micro-flora (Bayon and Mathelin, 1980; Egert *et al*., 2003; Li and Brune, 2005; Zhang and Jackson, 2008).

Whilst a number of studies had been done on scarab beetle larvae such as *C. zealandica*, *P. ephippiata* and *Melolontha melolontha* (Huang *et al*., 2010), little or nothing has been done to investigate *E. intermedius*'s cellulose degradation capabilities. Therefore, this is so far the first study on the cloning of cellulase genes from the gut microbes found in *E. intermedius* larvae.

4.1 Cellulose degradation by gut micro-flora

The cellulose degradation capabilities of *E. intermedius* gut micro-flora were evaluated by culturing the guts on a basal medium containing cellulose (filter paper) as the sole carbon source. There was observed structural disintegration of filter paper strips in the cellulose medium culture as culturing progressed from Day 1 to Day 28, probably due to the action of the cellulolytic gut micro-flora. It can also be argued that the progressive disintegration of filter paper strips may also be attributed to the shaking motion of the rotary shaker used during culturing experiments. However, what is most important is the mere ability of the gut micro-flora to grow and survive on a cellulose-enriched growth medium. This alone confirms that a sub-set of gut micro-flora from *E. intermedius* can in fact degrade cellulose matter.

Poor bacterial growth rates were observed in the cellulose medium cultures during the course of the 28 day culturing period. The poor growth patterns observed might be attributed to the disruptive effect of cellulose medium culturing (by only selecting cellulose degrading micro-flora and excluding non-cellulose degrading organisms) to the inherent community structure and relationships which naturally exists in the dung beetle larva gut. In the dung beetle larva digestive tract, different types of micro-organisms work synergistically, by performing different roles which collectively contribute to the survival and well-being of the whole microbial community. Any disruption to one or more of these different microbial processes negatively impacts on the growth and survival of the entire microbial community (Huang *et al*., 2010).

The observed concentration of reducing sugars in the cellulose medium culture showed negligible increments over the 28 day culturing period. Low reducing sugar concentration may be attributed to the ability of the larva gut micro-flora to utilize the produced reducing sugars as a carbon source (non-cellulose medium contained glucose and lactose as carbon sources and larva gut micro-flora were also able to grow in this medium). This can mean that any reducing sugars produced during the cellulose degradation process were immediately utilized most probably by another sub-section of the cultured gut micro-flora before they could be accounted for using the DNS assay.

Secondly, complete cellulose digestion to produce the reducing sugar glucose requires the synergistic action of three cellulase enzymes, endo-glucanases, cellobiohydrolases and finally β-glucosidases (Karmakar and Ray, 2011), as shown in **Figure 30**. In this study, only endo-glucanase and cellobiohydrolase activity was evaluated and positively identified. Since no β-glucosidase activity was evaluated, one can argue that the cellulose digestion process may not have been done to full completion, probably due to the absence of β-glucosidases which hydrolyse cellobiose to glucose and other shorter oligosaccharides. The β-glucosidase catalysed step is considered the rate limiting step of the whole cellulose digestion process (Jiang *et al*., 2010). In the dung beetle gut for instance, cellulose digestion has been reported to yield mainly acetic acid due to incomplete cellulose digestion (Huang *et al*., 2010). To improve reducing sugar or glucose yields during cellulose digestion, in the cultures, there is need to select only cellulose digesting micro-flora which did not consume the produced reducing sugars in the culture, and to make sure that all the three types of cellulase enzymes are present to carry out cellulose degradation process to full completion.

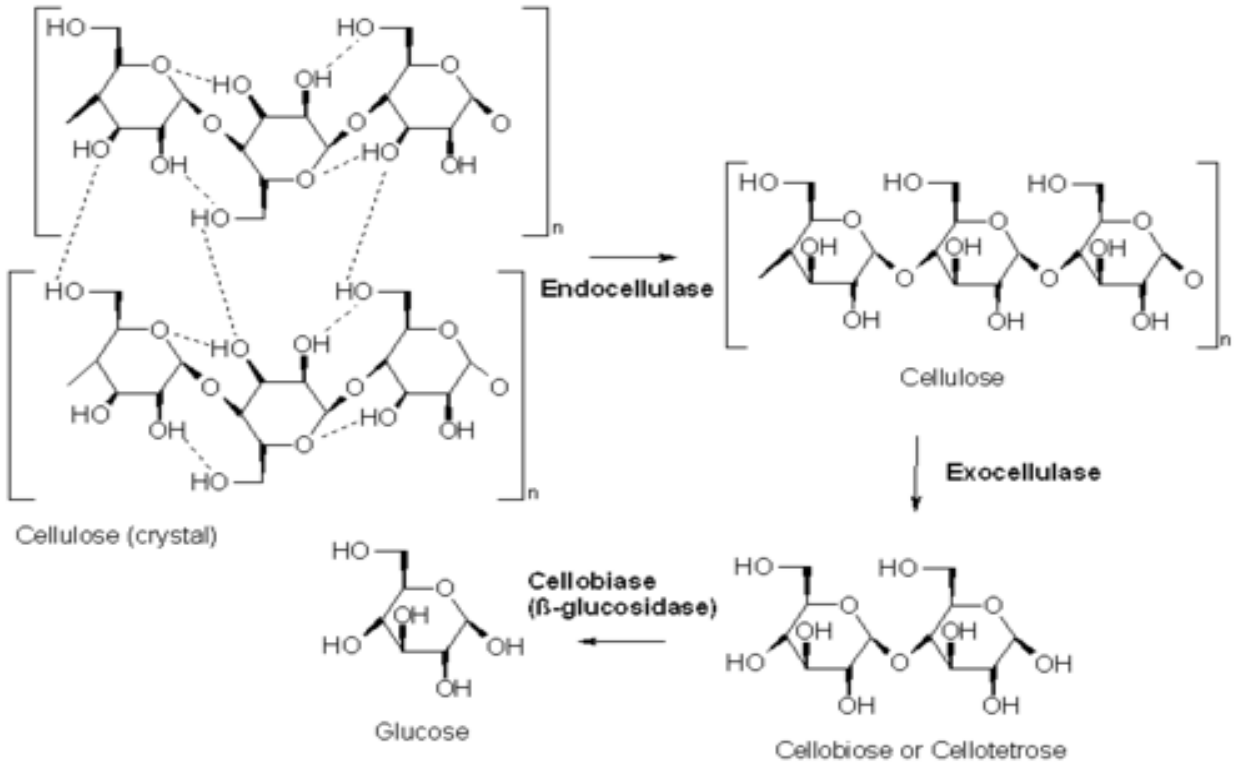


Figure 30: Mode of action of various components of cellulase (adopted from Karmakar and Ray, 2011)

4.2 Cellulase activities

Endo-glucanase (CMC) and cellobiohydrolase (MUC) cellulase activities of the cosmid library were assayed on LB-ampicillin agar plates containing appropriate cellulose substrates. By cultivating a metagenomic library on the plates with the substrate, one can identify positive clones through visual screening for the appearance of a clear zone (halo) or colour (Uchiyama and Miyazaki, 2009).

Endo-glucanase activity was evaluated using CMC as the cellulose substrate and the plates stained with Congo red dye to check for colony clearing zones. Endo-glucanases have been reported to be very active on the amorphous regions of cellulose and as such their activity are easily assayed using soluble cellulose substrates like CMC (Karmakar and Ray, 2011). According to Teather and Wood, (1982), Congo red is a pH dependent dye that binds the β-1,4-glucosidic bonds of cellulose. Endo-glucanases, (**Figure 30**) randomly cleave these β-1,4-glucosidic bonds along the cellulose chain and in the process releases the Congo red dye. This leads to the formation of a clearing zone or halo around the endo-glucanase positive clones.

Cellobiohydrolase activity, on the other hand, was evaluated using MUC as the cellulose substrate source. This type of cellulase is exo-acting, as it cleaves β-1,4-glucosidic bonds from the cellulose chain ends (Karmakar and Ray, 2011). Positive MUC activity was evaluated by the appearance of blue fluorescing colonies when the incubated plates were exposed to 302 nm UV light. The use of fluorometric MUC cellulose substrate in this assay allowed the detection of cellobiohydrolase positive clones (Taupp *et al.*, 2011).

It should be however noted that the clones which had previously tested positive for CMC activity were replica-plated on LB-ampicillin agar containing MUC substrate and tested for MUC activity. Therefore, clones which tested positive for MUC activity possessed both CMC and MUC activities. According to Karmakar and Ray, (2011), there is evidence that some cellulases possess both endo- and exo- cellulase activity.

The observed high number of positive colonies can be attributed to the selective cellulose medium-culturing step which was performed prior to DNA isolation. Despite the fact that laboratory enrichment cultures bear only a limited biodiversity, this technique has been reported to be highly efficient for the rapid isolation of large DNA fragments and cloning of genes with high biotechnological value (Streit and Schmitz, 2004). This strategy is commonly used in metagenomics to increase the likelihood of detecting clones displaying the desired trait (Voget *et al*., 2006).

4.3 Cellulolytic bacterial genes

Partial sequence analysis of a positive MUC/CMC colony using BLASTN and BLASTX tools yielded mostly bacterial genes and species belonging to the *Enterobacteriaceae* family. This was according to expectations since all the cellulose-cultured microbes were extracted from the gut of the dung beetle larvae, which is populated by Proteobacteria representatives, to which *Enterobacteriaceae* species belong (Huang *et al*., 2010). Cellulolytic activity of most of the obtained genes and bacterial species has never been previously documented suggesting that *E. intermedius* larvae gut micro-flora might be potential sources of new cellulolytic enzymes. In addition, a few uncultured bacterial genes and proteins of unknown function were also obtained, thus indicating that our culture might contain previously unidentified bacterial species which might be of biotechnology significance. Cellulolytic bacterial genes could not be conclusively verified, due to the partial sequencing of the clone.

4.4 Conclusion

With the successful culturing and cellulose screening of micro-flora possessing cellulolytic activities from *E. intermedius* larvae digestive tracts, a new source of cellulolytic bacteria and enzymes has been established. This study's findings prove that this dung beetle's gut is a new source of cellulolytic micro-organisms and enzyme activities that will aid the function and design of future bioreactors for the bio-fuel industry. I have no doubt that this research will significantly contribute to the potential of scarab-associated cellulolytic activities for use within the bio-fuel industry.

4.5 Recommendations

Future work should focus on the sub-cloning of the obtained MUC/CMC positive clones to localize the cellulase genes and shorten the inserts for effective sequencing and biochemical characterization (Feng *et al*., 2007). Other cellulase activities like β-glucosidase should also be investigated in the future since this study only characterized endo-glucanase and cellobiohydrolase activities.

CHAPTER 5

Appendices

5.1 **Appendix A**: Total protein analysis results

Table 12: Total protein analysis standard curve

Concentration (µg/ml)	Absorbance 1 (595 nm)	Absorbance 2 (595 nm)	Average	Standard deviation
12	0.05	0.03	0.04	0.014142136
14	0.053	0.031	0.042	0.015556349
16	0.151	0.095	0.123	0.03959798
18	0.154	0.136	0.145	0.012727922
20	0.19	0.186	0.188	0.002828427
22	0.359	0.335	0.347	0.016970563
24	0.376	0.354	0.365	0.015556349
25	0.389	0.375	0.382	0.009899495

Table 13: Total protein concentration of cellulose medium culture (µg/ml)

Sample / Day	1	2	3	4	Average concentration	Standard deviation
1	18.571	14.49	13.129	14.592	15.1955	2.347080811
2	17.483	14.49	14.082	15.136	15.29775	1.520086483
3	14.32	14.32	14.49	24.83	16.99	5.227281001
4	19.218	12.823	13.469	16.973	15.62075	3.012541604
5	17.925	13.367	14.116	25.102	17.6275	5.367778839
6	19.49	21.565	14.932	21.293	19.32	3.066822351

7	19.592	16.531	17.993	21.837	18.98825	2.27365277
8	18.98	20.544	17.789	17.245	18.6395	1.46185282
9	18.367	19.456	20.476	22.585	20.221	1.795927059
10	27.313	21.599	22.041	25	23.98825	2.681908447
11	21.667	30.34	20.714	23.537	24.0645	4.344882545
12	18.163	15.306	18.844	20.612	18.23125	2.2064435
13	16.293	17.993	18.231	18.776	17.82325	1.071521154
14	18.129	22.313	19.796	18.061	19.57475	1.994040851
15	22.876	23.401	25.102	30	25.34475	3.245672542
16	22.041	21.769	18.503	19.49	20.45075	1.73045743
17	17.103	22.172	20.483	16.966	19.181	2.573298402
18	19.517	17.897	26.793	19.828	21.00875	3.947996317
19	17.379	21.241	22.138	20.828	20.3965	2.08466632
20	14.621	16.759	16.379	16.552	16.07775	0.983511862
21	15.897	17	16.034	18	16.73275	0.977083202
22	18.517	17.897	21.138	16.862	18.6035	1.822373635
23	19.069	17.966	19	21.241	19.319	1.377068626
24	18.828	18.517	19.69	21.759	19.6985	1.460530155
25	15.517	18.379	15.379	15.655	16.2325	1.435429204
26	14.931	14.793	15.655	14.828	15.05175	0.406409789
27	14.276	13.448	15.621	13.931	14.319	0.932065448
28	14.724	14.724	14.586	14.31	14.586	0.195161472

Table 14: Total protein concentration of non-cellulose medium culture (µg/ml)

Sample / Day	1	2	3	4	Average concentration	Standard deviation
1	13.707	12.925	12.925	12.993	13.1375	0.381017497
2	13.435	13.639	13.639	13.639	13.588	0.102
3	13.469	13.469	14.048	12.823	13.45225	0.500478021
4	15.442	15.068	17.007	14.456	15.49325	1.087932404
5	15.136	14.252	16.224	14.796	15.102	0.831900635
6	16.973	18.027	19.32	16.769	17.77225	1.169863062
7	15.952	13.129	17.211	18.163	16.11375	2.186189893
8	18.639	17.279	15.748	18.299	17.49125	1.297916375
9	19.116	22.993	24.592	20.85	21.88775	2.400976798
10	25.986	26.701	19.592	19.694	22.99325	3.879755781
11	17.313	17.381	20.102	18.81	18.4015	1.327255941
12	18.367	19.728	20.68	20.782	19.88925	1.120347379
13	21.701	19.558	22.993	21.497	21.43725	1.417165334
14	20.612	19.524	18.537	23.095	20.442	1.96121204
15	23.878	26.667	21.565	25.884	24.4985	2.281260251
16	19.014	20.51	18.571	25.442	20.88425	3.149714736
17	21.897	18.276	21.483	22.621	21.06925	1.920636418
18	23.103	21.379	18.897	21.793	21.293	1.758248371
19	24.621	15.793	21.414	20.448	20.569	3.649545451
20	20.897	16.517	21.759	23.31	20.62075	2.912316189
21	20.414	17.759	24.793	22.931	21.47425	3.058509155
22	19.966	21.586	23.345	22.586	21.87075	1.459925883
23	20.31	22.034	20.414	22.276	21.2585	1.040759178
24	19.345	20.069	19.69	21.828	20.233	1.103677791
25	18.241	16.172	17.655	17.655	17.43075	0.883465289
26	16.034	15.931	16.966	18.724	16.91375	1.293508504
27	18.517	17.172	16.345	16.379	17.10325	1.017004875
28	18.655	17.31	16.379	18.448	17.698	1.059662525

Table 15: Total protein concentration comparison of cellulose and non-cellulose media cultures.

Day	Cellulose culture (M1A) average concentration (µg/ml)	Non-cellulose culture (M1B) average concentration (µg/ml)
1	15.1955	13.1375
2	15.29775	13.588
3	16.99	14.9065
4	15.62075	15.49325
5	17.6275	15.102
6	19.32	17.77225
7	18.98825	16.11375
8	18.6395	17.49125
9	20.221	21.88775
10	23.98825	22.99325
11	24.0645	18.4015
12	18.23125	19.88925
13	17.82325	21.43725
14	19.57475	20.442
15	25.34	24.4985
16	20.45075	20.88425
17	19.181	21.06925
18	21.00875	21.293
19	20.3965	20.569
20	16.07775	20.62075
21	16.73275	21.47425
22	18.6035	21.87075
23	19.319	21.2585
24	19.6985	20.233
25	16.2325	17.43075
26	15.05175	16.91375
27	14.319	17.10325
28	14.586	17.698

5.2 **Appendix B**: Cellulase activity measurement results

Table 16: Cellulase activity standard curve

Concentration (mg/ml)	Absorbance 1 (54 0nm)	Absorbance 2 (540 nm)	Average	Standard deviation
3.3	0.094	0.09	0.092	0.002828427
4	0.265	0.2648	0.2649	0.000141421
5	0.508	0.516	0.512	0.005656854
6	0.7611	0.7603	0.7607	0.000565685
6.7	0.927	0.943	0.935	0.011313708

Table 17: Reducing sugar concentration of cellulose medium culture (mg/ml)

Sample / Day	1	2	3	4	Average concentration	Standard deviation
1	2.957	2.958	2.959	2.957	2.95775	0.000957427
2	2.959	2.959	2.966	2.957	2.96025	0.003947573
3	2.966	2.959	2.972	2.958	2.96375	0.006551081
4	2.966	2.969	2.972	2.959	2.9665	0.005567764
5	2.987	3.451	2.982	2.983	3.10075	0.233509993
6	2.981	2.973	2.978	2.979	2.97775	0.00340343
7	2.975	2.971	2.977	2.973	2.974	0.002581989
8	2.973	2.97	2.99	2.972	2.97625	0.009251126
9	2.972	2.999	2.984	2.964	2.97975	0.015239751
10	2.963	2.981	2.974	2.959	2.96925	0.010078856
11	2.959	2.966	2.968	2.956	2.96225	0.005678908

12	2.957	2.978	2.972	2.959	2.9665	0.010148892
13	2.971	2.979	2.968	2.967	2.97125	0.005439056
14	2.967	2.961	2.965	2.958	2.96275	0.004031129
15	2.956	2.975	2.978	2.955	2.966	0.012192894
16	2.962	2.949	2.963	2.951	2.95625	0.007274384
17	2.964	2.973	2.973	2.963	2.96825	0.0055
18	2.963	2.969	2.966	2.959	2.96425	0.004272002
19	2.961	2.965	2.966	2.962	2.9635	0.002380476
20	2.959	2.964	2.966	2.959	2.962	0.003559026
21	2.96	2.964	2.965	2.96	2.96225	0.002629956
22	2.959	2.964	2.964	2.958	2.96125	0.003201562
23	2.959	2.964	2.963	2.958	2.961	0.00294392
24	2.959	2.964	2.963	2.959	2.96125	0.002629956
25	2.961	2.963	2.964	2.96	2.962	0.001825742
26	2.959	2.964	2.962	2.959	2.961	0.00244949
27	2.959	2.963	2.959	2.959	2.96	0.002
28	2.958	2.961	2.96	2.958	2.95925	0.0015

5.3 **Appendix C**: Sequencing results

5.3.1 M13F nucleotide sequence

>M13F

ATTCGGGGGGGGGGAAGGCGCGCTCGCCGCCTTTACAATGTCCCCGACGATTTTT
TCCGCCCTCAGCGTACCGTTTAGAAGCAGTTTTCAGCTATCGTCACATTACTGAG
CGTCCCGGAGTTCGCATTCACACTGCCACTGATATCCGCATTTTTAGCGGTCAGC
TTTCCGTCCGGTGTCAGGGAAAAGGCCGGAGGATTGCCGCCGCTGGTAATGGTG
GGGGCCGTCAGGCGCTTCAGGAACACGTCGTTCATGAATATCTGGTTGCCCTGCG
CCACAAACATCGGCGTTTCATTCCCGTTTGCCGGGTCAATAAATGCGATACGATT
GGCGGCAACCAGAAACTGGCTCAGTTTGCCTTCCTCCGTGTCCTCCATGCTGAGG
CCAATACCCGCGACATAATGTTTGCCGTCTTTGGTCTGCTCAATTTTGACAGCCC
ACATGGCATTCCACTTATCACTGGCATCCTTCCACTCTTTCGAAAACTCCTCCAGT
CTGCTGGCGTTATCCTCCGTCAGCTCGACTTTTTCCAGCAGCTCCTTGCCGAGATG
GGATTCGGTTATCTTGCCTTTGAAAAAATCCAGGTAACCTTCCGCATCATCGCTC
GCCCGACCGACGGCCTCCACGAATGCCGATTTGCCAACGGTGTTCACACTGCGG
ATATAAAAGTAATAATCATGGCCCGGTTTGATATTGATACTGGCGGCTATCCAGT
ACAGCGCCGTACCAAGATAACGCGTGCTGGTTTCAACCTGTCTGATATCCGCAAT
CTGCTTTTCCGAGAACCAGAACTCAAACTGTACCGTCGGGTCATAAAACGGCAA
GATGCGGCGTGGCGGTTATCTGAAAATAGCCCGGCGTCAGCTCAATCCTCGACG
GTGCTGCCGGTGCGGCAATCCAGAACGATACCGACGCCGGGATCGGCTTGCTGC
CCCACGCATTTACCGCCCGGACTGTCAGCCTGTAGTTCCCAGC

5.3.2 T7 nucleotide sequence

>T7

GTGAGAGCCAAATCAGGAGACAACATCGCGcCGGCAACACTCTGGAAGCAAGCA
TGTTTAGTGTGCTAGGGGGGCCAACGCCGCTGCTTGGTTAGGCTAGGCCGTGCTC
CAAAAAAACGT

References

Adney, B and Baker, J. 1996. Measurement of cellulase activities. Laboratory Analytical Procedures (LAP). ***Technical Report*** **NREL/TP-510-42628.** January 2008.

Adsul, M. G., Singhvi, M. S., Gaikaiwari, S. A. and Gokhale, D. V. 2011. Development of biocatalysts for production of commodity chemicals from lignocellulosic biomass. *Bioresource Technology*. **102**: 43042-4312.

Baird, S. D., Johnson, D. A. and Seligy, V. L. 1990. Molecular cloning, expression, and characterization of endo-3-1,4-glucanase genes from *Bacillus polymyxa* and *Bacillus circulans*. *Journal of Bacteriology*. **172**: 1576-1586.

Balat, M., and Balat, H. 2009. Recent trends in global production and utilization of bio-ethanol fuel. *Applied Energy*. **86**: 2273-2282.

Bansal, P., Hall, M., Realff, M. J., Lee, J. H. and Bommarius, A. S. 2009. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances*. **27**: 833-848.

Bayon, C. and Mathelin, J. 1980. Carbohydrate fermentation and by-product absorption studied with labelled cellulose in *Oryctes Nascornis* larvae (Coleoptera: Scarabaeidae). *J. Insect Physiol*. **26**: 833-840.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* **72***:* 248-254.

Chen, S., Lo, Y., Lee, K., Huang, T. and Chang, J. 2009. Sequencing batch reactor enhances bacterial hydrolysis of starch promoting continuous bio-hydrogen production from starch feedstock. *International Journal of Hydrogen Energy*. **34**: 8549-8557.

Cotula, L., Dyer, N. and Vermeulen, S. 2008. Fuelling exclusion? The biofuels boom and poor people's access to land. *FAO and IIED*. London. **ISBN: 978-1-84369-702-2.**

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., and Wittwer, P. 2005. Metagenomic gene discovery: past, present and future. *Trends in Biotechnology*. **23**: 321-329.

Demirbas, A. 2009. Biofuels securing the planet's future energy needs. *Energy Conversion and Management*. **50**: 2239-2249.

Demirbas, M. F., Balat, M., and Balat, H. 2009. Potential contribution of biomass to the sustainable energy development. *Energy Conversion and Management*. **50**: 1746-1760.

Ding, S., Xu, Q., Crowley, M., Zeng, Y., Nimlos, M., Lamed, R., Bayer, E. A. and Himmel, M. E. 2008. A biophysical perspective on the cellulosome: new opportunities for biomass conversion. *Current Opinion in Biotechnology*. **19**: 218-227.

Dillion, R. J. and Dillion, V. M. 2004. The gut bacteria of insects: Non-pathogenic interactions. *Annual Review Entomology*: **49**: 71-92.

Duan, C., Xian, L., Zhao, G., Feng, Y., Pang, H., Bai, X., Tang, J., Ma, Q. and Feng, J. 2009. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal of Applied Microbiology*. **107**: 245-256.

Edwards, P. B. 1991. Seasonal variation in the dung of African grazing mammals, and its consequences for coprophagous insects. *Functional Ecology*. **5**: 617-628.

Egert, M., Stingl, U., Bruun, L. D., Pommerenke, B., Brune, A. and Friedrich, M. W. 2005. Structure and topology of microbial communities in the major gut compartments of *Melolontha melolontha* larvae (Coleoptera: Scarabaeidae). *Applied and Environmental Microbiology*: **71**: 4556-4566.

FAO and UNEP. 2010. A Decision Support Tool for Sustainable Bioenergy. *UN Energy*. **ISBN number: 978-92-5-106638-6**

FAO. 2009**.** Small scale bioenergy initiatives: brief description and preliminary lessons on livelihood impacts from case studies in Asia, Latin America and Africa. *Environment and Natural Resources Management*.

Feng, Y., Duan, C., Pang, H., Mo, X., Wu, C., Yu, Y., Hu, Y., Wei, J., Tang, J. and Feng, J. 2007. Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Applied Microbiology and Biotechnology*. **75**: 319-328.

Ferrer, M., Martı´nez-Abarca, F. and Golyshin, P. N. 2005. Mining genomes and 'metagenomes' for novel catalysts. *Current Opinion in Biotechnology*. **16**: 588-593.

Hayashi, F and Kamimura, Y. 2001. The potential for incorporation of male derived proteins into developing eggs in the leafhopper *Bothrogonia ferruginea*. *Journal of Insect Physiology*. **48**: 153-159.

Huang, S., Zhang, H., Marshall, S. and Jackson T. A. 2010. The scarab gut: A potential bioreactor for bio-fuel production. *Insect Science*. **17**: 175-183.

Inoue, T., Moriya, S., Ohkuma, M. and Kudo, T. 2005. Molecular cloning and characterization of a cellulase gene from a symbiotic protist of the lower termite, *Coptotermes formosanus*. *Gene*. **349**: 67-75.

Jegannathan, K. R., Chan, E. and Ravindra, P. 2009. Harnessing biofuels: A global Renaissance in energy production? *Renewable and Sustainable Energy Reviews*. **13**: 2163-2168.

Jiang, C., Hao, Z., Jin, K., Li, S., Che, Z., Ma, G. and Wu, B. 2010. Identification of a metagenome-derived β-glucosidase from bioreactor contents. *Journal of Molecular Catalysis B: Enzymatic*. **63**: 11-16.

Karmakar, M. and Ray, R. R. 2011. Current trends in research and application of microbial cellulases. *Research Journal of Microbiology*. **6**: 45-53.

Ko, C. H., Chen, W. L., Tsai, C. H., Jane, W. N., Liu, C. C. and Tu, J. 2007. *Paenibacillus campinasensis* BL11: a wood material-utilizing bacterial strain isolated from black liquor. *Bioresource Technology*. **14**:2727-2733.

Koyama, M., Iwata, R., Yamane, A., Katase, T. and Ueda, S. 2003. Nutrient intake in the third instar larvae of *Anomala cuprea* and *Protaetia orientalis submarmorea* (Coleoptera: Scarabaeidae) from a mixture of cow dung and wood chips: Results from stable isotope analyses of nitrogen and carbon. *Applied Entomology and Zoology*. **38**: 305-311.

Lailvaux, S. P., Hathway, J., Pomfret, J. and Knell, R. J. 2005. Horn size predicts physical performance in the beetle *Euoniticellus intermedius* (Coleoptera: Scarabaeidae). *Functional Ecology*. **19**: 632-639.

Lämmle, K ., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H. and Rupp, S. 2007. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology*. **127**: 575-592.

Lee, J. S., Parameswaran, B., Lee, J. P. and Park, S. C. 2008. Recent developments of key technologies on cellulosic ethanol production. *Industrial Research*. **67**: 865-873.

Lemke, T., Stingl, U., Egert, M., Friedrich, M. W. and Brune, A. 2003. Physicochemical conditions and microbial activities in the highly alkaline gut of the humus feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Applied and Environmental Microbiology*. **69**: 6650-6658

Li, X., and Brune, A. 2005. Digestion of microbial biomass, structural polysaccharides, and protein by the humivorous larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Soil Biol Biochem*. **37**:107-116

Liang, Y., Yesuf, J., Schmitt, S., Bender, K., Bozzola, J. 2009. Study of cellulases from a newly isolated thermophilic and cellulolytic *Brevibacillus* sp. strain JXL. *Journal of Industrial Microbiology and Biotechnology*.

Liu, Y., Zhang, J., Liu, Q., Zhang, C. and Ma, Q. 2004. Molecular cloning of novel cellulase genes cel9A and cel12A from *Bacillus licheniformis* GXN151 and synergism of their encoded polypeptides. *Current Microbiology*. **49**: 234-238.

Makaruk, A., Miltner, M. and Harasek, M. 2010. Membrane biogas upgrading processes for the production of natural gas substitute. *Separation and Purification Technology*. **74**: 83-94.

Maki, M., Leung, K. T. and Qin, W. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*. **5**: 500-516.

Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R. and Monot, F. 2009. New improvements for lignocellulosic ethanol. *Current Opinion in Biotechnology*. **20**: 372-380.

Moyo, P. 2011. The use of dung beetle larvae gut microbial consortia in the production of biogas and bio-ethanol. *Unpublished MSc thesis*. 13-14.

National Center for Biotechnology Information: www.ncbi.nlm.nih.gov/Blast.cgi [accessed 10 January 2012].

Oslaj, M., Mursec, B. and Vindis, P. 2010. Biogas production from maize hybrids. *Biomass and Bioenergy*. **34**: 1538-1545.

Pomfret, J. C. and Knell, R. J. 2006. Sexual selection and horn allometry in the dung beetle *Euoniticellus intermedius*. *Animal Behaviour*. **71**: 567-576.

Pottkämper, J.,Barthen, P., IImberger, N., Schwaneberg, U., Schenk, A., Schulte, M., Ignatiev, N. and Streit, W. R. 2009. Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. *Green Chemistry*. **11**: 957-965.

Prins, R. A. and Kreulen, D. A. 1991. Comparative aspects of plant cell wall digestion in insects. *Animal Feed Science and Technology*. **32**: 101-118.

Rees, H. C., Grant, S., Jones, B., Grant, W. D. and Heaphy, S. 2003. Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles: Life Under Extreme Conditions*. **7**: 415-421.

Reinhold-Hurek B., Hurek, T., Claeyssens, M. and Montagu, M 1993. Cloning, expression in *Escherichia coli*, and characterization of cellulolytic enzymes of *Azoarcus* sp., a root-invading diazotroph. *Journal of Bacteriology*. **175**:7056-7065

Ruane, J., Sonnino, A. and Agostini, A. 2010. Bioenergy and the potential contribution of agricultural biotechnologies in developing countries. *Biomass and Bioenergy*. **34**: 1427-1439.

Sakamoto, A. and Toyohara, H. 2009. Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam *Corbicula japonica*. *Comparative Biochemistry and Physiology, Part B*. **152**: 390-396.

Saul, D. J., Williams, L. C., Grayling, R. A., Chamley, L. W., Love, D. R. and Bergquist, P. L. 1990. celB, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*". *Applied and Environmental Microbiology*. **56**. 3117-3124.

Singh, B. 2010. Exploring microbial diversity for biotechnology: the way forward. *Trends in Biotechnology*. **28**. 111-116.

Stoscheck, C. M. 1990. Quantitation of Protein. *Methods in Enzymology*. **182***:* 50-69.

Streit, W. R. and Schmitz, R. A. 2004. Metagenomics – the key to the uncultured microbes. *Current Opinion in Microbiology*. **7**: 492-498.

Taupp, M., Mewis, K. and Hallam, S. J. 2011. The art and design of functional metagenomic screens. *Current Opinion in Biotechnology*. **22**: 1-8.

Teather, R. M. and Wood, P. J. 1982 Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied Environmental Microbiology*. **43**:777-780.

Thorsten, L., Ulrich, S., Markus, E., Michael, W. F. and Brune, A. 2003. Microbial community structure in the midgut and hindgut of the humus feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Applied and Environmental Microbiology*. **69**: 6659-6668.

Tholen, A., Schink, B. and Brune, A. 1997. The gut microflora of *Reticulitermes flavipes*, its relation to oxygen, and evidence for oxygen dependent acetogenesis by the most abundant *Enterecoccus* species. *FEMS Microbiology Ecology*. **24**: 137-149.

Tholen, A., Pester, M. and Brune, A. 2007. Simultaneous methanogenesis and the oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiology Ecology*. **62**: 303-312.

Uchiyama, T. and Miyazaki, K. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*. **20**: 616-622.

Vasudevan, P. T. and Briggs, M. 2008. Biodiesel production-current state of the art and challenges. *Journal of Industrial Microbiology and Biotechnology*. **35**: 421-430.

Voget, S., Steele, H. L. and Streit W. R. 2006. Characterization of a metagenome-derived halotolerant cellulose. *Journal of Biotechnology*. **126**: 26-36.

Wang, F., Li, F., Chen, G. and Liu, W. 2009. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiological Research*. **164**: 650-657.

Willis, J. D., Oppert, B., Oppert, C., Klingeman, W. E. and Jurat-Fuentes, J. L. 2011. Identification, cloning, and expression of a GHF9 cellulase from *Tribolium castaneum* (Coleoptera: Tenebrionidae). *Journal of Insect Physiology*. **57**: 300-306.

Yuan, J. S., Tiller, K. H., Al-Ahmad, H., Stewart, N. R. and Stewart Jr., C. N. 2008. Plants to power: bioenergy to fuel the future. *Trends in Plant Science*. **13**. 421-429.

Zahura, U. A., Rahman, M. M., Inoue, A., Tanaka, H. and Ojima, T. 2011. cDNA cloning and bacterial expression of an endo-β-1,4-mannanase, AkMan, from *Aplysia kurodai*. *Comparative Biochemistry and Physiology, Part B*. **159**: 227-235.

Zhang, D., Lax, A. R., Raina, A. K. and Bland, J. M. 2009. Differential cellulolytic activity of native-form and C-terminal tagged-form cellulase derived from *Coptotermes*

formosanus and expressed in E. coli. *Insect Biochemistry and Molecular Biology*. **39**. 516-522.

Zhang, D., Lax, A. R., Bland, J. M. and Allen, A. B. 2011. Characterization of a new endogenous endo-b-1,4-glucanase of Formosan subterranean termite (*Coptotermes formosanus*). *Insect Biochemistry and Molecular Biology*. **41**: 211-218.

Zhang, H.Y. and Jackson, T.A. 2008. Autochthonous bacterial flora indicated by PCR-DGGE of 16S rRNA gene fragments from the alimentary tract of *Costelytra zealandica* (Coleoptera: Scarabaeidae*). Journal of Applied Microbiology*. **105**: 1277–1285

Zhou, X., Smith, J. A., Oi, F. M., Koehler, P. G., Bennett, G. W. and Scharf, M. E. 2007. Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene*. **395**: 29-39.