ISTANBUL TECHNICAL UNIVERSITY ★GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DISCOVERY OF NEW CELLULOSE-HYDROLYZING ENZYMES USING METAGENOMIC APPROACH

M.Sc. THESIS

Sevde ŞENCAN

Molecular Biology-Genetics and Biotechnology Department
Molecular Biology-Genetics and Biotechnology Program

DECEMBER 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DISCOVERY OF NEW CELLULOSE-HYDROLYZING ENZYMES USING METAGENOMIC APPROACH

M.Sc. THESIS
Sevde Sencan
(521131118)

Molecular Biology-Genetics and Biotechnology Department
Molecular Biology-Genetics and Biotechnology Program

Thesis Advisor: Prof. Dr. Zeynep Petek ÇAKAR
Co-advisor: Assoc. Prof. Dr. Yavuz ÖZTÜRK

DECEMBER 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

METAGENOMİK YÖNTEM KULLANILARAK SELÜLOZ PARÇALAYAN YENİ ENZİMLERİN KEŞFEDİLMESİ

YÜKSEK LİSANS TEZİ

Sevde ŞENCAN
(521131118)

Moleküler Biyoloji – Genetik ve Biyoteknoloji Anabilim Dalı
Moleküler Biyoloji – Genetik ve Biyoteknoloji Programı

Tez Danışmanı: Prof. Dr. Zeynep Petek ÇAKAR
Eş Danışman: Doç. Dr. Yavuz ÖZTÜRK

ARALIK 2015

Sevde ŞENCAN, a M.Sc. student of ITU Graduate School of Science, Engineering and Technology student ID 521131118, successfully defended the thesis entitled "DISCOVERY OF NEW CELLULOSE-HYDROLYZING ENZYMES USING METAGENOMIC APPROACH", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Zeynep Petek ÇAKAR** İstanbul Technical University	.....
Co-advisor :	**Assoc. Prof.Dr. Yavuz ÖZTÜRK** TÜBİTAK MAM	.....
Jury Members :	**Prof. Dr. Melek ÖZKAN** Gebze Technical University	.....
Jury Members :	**Assoc. Prof. Dr. Ayten KARATAŞ** İstanbul Technical University	.....
Jury Members :	**Assoc. Prof. Dr. Fatma Neşe KÖK** İstanbul Technical University	.....

Date of Submission: 28.11.2015

Date of Defense: 23.12.2015

To my family,

FOREWORD

First of all, I would like to express my sincere gratitude to my supervisors Prof. Dr. Zeynep Petek ÇAKAR and Assoc. Prof. Dr. Yavuz ÖZTÜRK for their guidance and sharing their wisdom with me during the course of research.
I would also like to thank my colleagues Assoc. Prof. Dr. Sevnur Mandacı, Dr. Orhan Özcan, Dr. Dilek Coşkuner Öztürk, İbrahim Sertdemir, and Nurçin Çelik Öztürk for their companionship and help on resolving any technical issues.
I would like to express my deepest appreciation to my father Prof. Dr. Hüner Şencan and my mother Nezahat Şencan for their support, understanding and guidance throughout my life.
I would also like to convey thanks to my jury members Prof. Dr. Melek ÖZKAN, Assoc. Prof. Dr. Ayten KARATAŞ, and Assoc. Prof. Dr. Fatma Neşe KÖK for their valuable comments and help on improvement of this thesis.
This study was financially supported by TÜBİTAK under the project "Development of Enzymes and Microorganisms for Efficient Hydrolysis of Agricultural Wastes and Lignocellulosic Raw Materials" (Project no: 113O123).

December 2015 Sevde ŞENCAN

TABLE OF CONTENTS

ABBREVIATIONS

BAC	**:** Bacterial artificial chromosome
BSA	**:** Bovine serum albumin
CBM	**:** Carbohydrate-binding module
CMC	**:** Carboxymethyl cellulose
DNA	**:** Deoxyribonucleic acid
DNS	**:** 2,4-dinitrosalicyclic acid
dNTP	**:** Deoxynucleotide
EDTA	**:** Ethylenediaminetetraacetic acid
GH	**:** Glycoside hydrolase
IPTG	**:** Isopropyl β-D-1-thiogalactopyranoside
LA	**:** Luria Agar
LB	**:** Luria Broth
mRNA	**:** Messenger ribonucleic acid
ORF	**:** Open reading frame
PCR	**:** Polymerase chain reaction
pNP	**:** p-nitrophenol
pNPG	**:** para-nitrophenyl- beta-D-glucoside
rcf	**:** Relative centrifugal force
rDNA	**:** Ribosomal deoxyribonucleic acid
rpm	**:** Revolutions per minute
SDS	**:** Sodium dodecyl sulfate
SDS-PAGE	**:** Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	**:** Super optimal broth with catabolite repression
TAE	**:** Tris-acetate-EDTA

LIST OF TABLES

LIST OF FIGURES

DISCOVERY OF NEW CELLULOSE-HYDROLYZING ENZYMES USING METAGENOMIC APPROACH

SUMMARY

Fossil fuel is a finite energy source and use of fossil fuels releases harmful chemicals to environment. To fulfill energy requirement, alternative energy sources that can replace fossil fuel must be developed. A high potential environmental friendly alternative is bio-fuel produced from cellulosic biomass. Despite of high number of biofuel researches, production of sustainable and economical biofuel remains a challenge due to recalcitrance of cellulosic materials. The major obstacle for industrial-scale production of biofuel is the inefficient degradation of plant material due to absence of enzymes that catalyze efficient hydrolysis at optimum conditions.

Fully degradation of cellulose requires synergetic action of a complex of enzymes for optimized saccharification. Cellulases, which catalyze cellulolysis, are classified into three groups according to their structure and mechanism of action: endoglucanases, exoglucanases, and β-glucosidases. Unfortunately, commercial biocatalysis is insufficient due to the lack of enzymes with optimal performance for specific applications. A powerful approach, which draws attention lately, for discovery of new enzymes with optimum performance is screening of natural diversity. Microorganisms, which adapted to a wide range of environmental conditions, serve as a good source. Metagenomic approach, which comprises the direct extraction of genomic DNA from environmental samples, can prevent loss of diversity resulting from cultivation, and allows recovery of genomes of all microorganisms present in a given niche.

In this study, the aim was identification of novel cellulolytic genes using metagenomic approach. Environmental samples used in the study comprised water samples from Dalaman Spring, Dalaman Lake, Emet Hot Spring, and Kuzuluk Hot Springs, straw samples that were incubated in water for enrichment of cellulolytic microorganisms, grasshoppers obtained from TUBITAK MAM campus and commercial farm, and grass compost taken from TUBITAK MAM campus. The metagenomic DNA extraction and DNA clean-up (if required) were performed using commercial kits. The isolated DNAs were used for construction of metagenomic library with next-generation sequencing technology. Bioinformatic tools were used for assembly of reads, phylogenetic studies, and gene annotations.

Total of 63 candidate cellulase genes were selected from annotated for further analysis. Candidate genes, which were amplified from metagenomic DNA, were cloned into *E. coli* BL21 (DE3) for production of recombinant protein. 22 of 63 candidate proteins was expressed successfully and screened for biochemical activity against cellulose. Two different substrate, CMC and pNPG, were used for screening of endoglucanases, exoglucanases, and β-glucosidases. Activity assay revealed one active enzyme against CMC, while no active protein that hydrolyzes pNPG was found.

METAGENOMİK YÖNTEM KULLANILARAK SELÜLOZ PARÇALAYAN YENİ ENZİMLERİN KEŞFEDİLMESİ

ÖZET

Fosil yakıtlar tükenebilir kaynaklar olmakla birlikte tüketimleri çevreye zararlı kimyasalların salımına neden olmaktadır. Fosil yakıtların çevreye olan zararlarını azaltmak ve enerjinin sürdürebilirliğini sağlamak için alternatif enerji kaynaklarının geliştirilmesi gerekmektedir. Selülozik biyokütlelerin parçalanması ile elde edilen biyoyakıt enerji gereksiniminin karşılanması için yüksek potansiyelli bir alternatif olarak görülmektedir. Fakat bu alanda yürütülen yoğun araştırmalara rağmen sürdürülebilir ve ekonomik açıdan uygulanabilir biyoyakıt üretimi sağlanamamıştır. Bunun sebebi, selülozik yapının parçalanamaya dirençli yapısıdır. Günümüzde biyoyakıtın endüstriyel ölçüde üretiminin yapılmasının önündeki en büyük engel optimum düzeyde çalışan yüksek kaliteli enzimlerin eksikliğidir.

Biyoyakıt üretiminin yanı sıra selülazlar birçok alanda kendilerine yer bulmaktadır. Selülazların sıklıkla kullanıldığı alanlar gıda endüstrisi, hayvan yemi üretimi, tekstil ve çamaşır sektörü, kimya sektörü, kağıt ve kağıt hamuru endüstrisi, atıkların parçalanması, tıp ve farmasötik endüstrisi, protoplast elde edilmesidir.

Selülozun tamamen parçalanması birden fazla enzimin eş zamanlı çalışmasını gerektirmektedir. Selülozu parçalamakla görevli enzimler, selülazlar, yapılarına ve aksiyon mekanizmalarına bağlı olarak üç grupta sınıflandırılmışlardır: endoglukanaz, ekzoglukanaz ve β-glukozidaz. Endoglukanazlar polisakkarit zincirinin iç bölgelerinde rastgele hidroliz yaparak değişik uzunlukta oligosakkaritler oluştururlar. Ekzoglukanazlar oluşturulan farklı uzunluktaki oligosakkaritlerin indirgenen ve indirgenmeyen ucundan sellobioz unitlerini hidroliz ederler. β-glukozidazlar ise meydana gelen sellobiozun ve sellodekstrinin glikoza parçalanmasından sorumlu enzimlerdir.

Günümüzde amacına uygun optimum şartlarda çalışan enzimlerin yetersizliğinden dolayı selülazların ticari kullanımı sınırlıdır. Yeni enzimlerin keşfi için dikkat çeken yöntemlerden biri doğal çeşitliliğin uygun amaç doğrultusunda taranmasıdır. Her türlü çevresel koşulda yaşamaya adapte olmuş mikroorganizmalar yeni enzimlerin keşfi için önemli kaynak teşkil etmektedir. Metagenomik yaklaşım, çevresel örneklerden direkt DNA izolasyonunu içeren, bu sebeple kültivasyondan kaynaklanan genetik bilgi kayıplarını en aza indirgeyen bir yöntemdir. Bu yöntem ortamda bulunan tüm mikroorganizmanın genetik materyalinin kazanımını sağlamaktadır.

Yapılan çalışmada metagenomik yöntem kullanılarak selüloz parçayan yeni enzimlerin keşfi amaçlanmıştır. Metagenomların eldesi için çeşitli ortamlardan örnekler elde edilerek metagenomik DNA izolasyonları yapılmıştır: Bu örnekler Dalaman Kaynak Suyu, Dalaman Gölü, Emet Kaplıcaları, Kuzuluk Kaplıcaları, Emet

Kaplıcaları ve Dalaman Gölü'nde bekletilen samanlar, çiftlikte yetiştirilen ve TÜBİTAK MAM kampüsünden toplanan çekirgeler ve TÜBİTAK MAM kampüsünden alınan çürümeye başlamış çimendir. Metagenomik DNA'ların örneklerden izolasyonu ve gerekli olduğu durumlarda elde edilen DNA'ların kontaminantlardan arındırılması için ZR Fecal DNA MiniPrep™, Meta-G-Nome™ DNA Isolation Kit ve Genomic DNA Clean & Concentrator kitleri kullanılmıştır. DNA miktarı ve kalitesi agaroz jel elektroforez sitemi ve spektometrik yöntemler kullanılarak tayin edilmiştir.

İzolasyonu yapılan DNA'lar metagenomik kütüphanelerin oluşturulmasında kullanılmıştır. Kütüphane oluşturulurken TruSeq Nano DNA Library Prep Kit (Illumina, USA) üretici firmanın tarifine uygun şekilde kullanılmıştır. HiSeq2500 (Illumina, ABD) dizileme sistemi ile tüm metagenom sekans dizilemesi gerçekleştirilmiştir. HiSeq2500 (Illumina, ABD) dizileme sistemi ile okunan parçaların birleştirilmesinde, filogenetik çalışmalarda ve gen anotasyonunda biyoenformatik araçlardan yararlanılmıştır. Okunan parçalar birleştirilirken Velvet Assembler (v1.2.07) programı kullanılmıştır ve analizle sonucu yaklaşık 68 gigabazlık data oluşturulmuştur. PhyloSift programı kullanılarak örneklerdeki mikrobiyal çeşitlilik belirlenmiştir. MetaGeneMark programı kullanılarak metagenomda bulunan açık okuma pencereleri belirlenmiştir. Belirlenen açık okuma pencereleri selülaz genlerinin belirlenmesinde kullanılmıştır. Selülaz genleri araştırılırken genlerde bulunan selüloz bağlanma bölgeleri ve glikozit hidrolaz dikkate alınmıştır. Analizler sonucu metegenomlarda 1238 selülaz geni olduğu görülmüştür.

Selülaz olarak anotasyonu yapılan genlerden 63 tanesi enzim keşfi için seçilmiştir. Her aday gene uygun özel primer çifti tasarlanmıştır ve uygun reaksiyon koşullarında metagenomik DNA'lar şablon olarak kullanılarak genler çoğaltılmıştır. Seçilen 63 genden 54 tanesi başarıyla çoğaltılmıştır. Amplifikasyonu yapılan genler EZ-Fusion™ Cloning Kit'i üreticinin protokolüne uygun bir şekilde kullanılarak daha önce restriksiyon enzimi ile lineer hale getirilen pET-22b(+) vektörüne klonlanmıştır. Aday geni içeren pET-22b(+) vektörlerinin, proteinlerin ekspresyonu için, *E. coli* BL21 (DE3) suşuna transformasyonu yapılmıştır. Transformasyon ısı şoku yöntemi kullanılarak kimyasal olarak kompetent hale getirilmiş hücreler ile yapılmıştır. Hücrelerin hedef geni içerdiğini göstermek için koloni PZR yapılmıştır. Yapılan doğrulama sonucunda 54 genin başarı ile *E. coli* BL21 hücrelerine klonlandığı gözlemlenmiştir. Rekombinant protein üretimi besi ortamına 1 mM IPTG eklenerek indüklenmiştir. Aynı koloniye ait IPTG ile indüklenen hücreler ve indüklenmeyen hücrelerin protein profilleri SDS-PAGE ile görüntülemiştir. Aday genlerin SDS-PAGE ile analizleri sonucu 22 proteinin başarı ile üretildiği gözlemlenmiştir.

Aday genlerinden başarıyla ekspres edilen 22 tanesi biyokatalitik aktivite için test edilmiştir. Hücrelerin uygun tampon ile sonikasyonundan sonra ham enzim elde edilmiştir. Endoglukanaz ve ekzoglukanaz aktivitesi için CMC substrat olarak kullanılırken, β-glukozidaz için pNPG kullanılmıştır. Aktivite tayini CMC'ye karşı aktivite gösteren bir enzimin varlığı gösterirken, pNPG'ye aktivite gösteren enzim gözlemlenmemiştir. CMC'ye karşı aktivite gösteren gen, 969 bp uzunluğunda olup, protein boyutu 34,40 kDa olarak hesaplanmıştır. BLAST analizi, bulunan genin daha önce bilinen *Rhodanobacter fulvus*'a ait endoglukanaz geni (NCBI Reference Sequence: WP_007079969.1) ile %100 benzer olduğunu göstermiştir.

Genin spesifik aktivitesinin belirlenmesi için HisPur™ Cobalt Purification Kit (Thermo Fisher Scientific) kullanılarak protein saflaştırması yapılmıştır ve işlem için üreticinin tarifi kullanılmıştır. Saflaştırılan proteinin miktar tayini Bradford yöntemi kullanılarak yapılmıştır. Saflaştıran protein enzim aktivite testi için kullanılmış ve spesifik aktivite spektrometrik ölçümlerden yararlanılarak belirlenmiştir. Bulunan endoglukanaz enziminin spesifik aktivitesi bir miligram protein başına 0,44 ünite olarak hesaplanmıştır.

1. INTRODUCTION

1.1 Cellulase, Application Areas and Industrial Importance

1.1.1 Cellulose and lignocellulosic biomass

Cellulose, the most abundant biopolymer produced in the biosphere, is synthesized by photosynthesis and 10^{11}–10^{12} tons of biopolymer is produced annually (Macrae et al., 1993). Cellulose is a homopolysaccharide composed of anhydro-β-1,4-glucose units held together by a β-1,4-glycosidic bond (Lima et al., 2005). The basic chemical formula of cellulose, discovered by Anselme Payen in 1842, is $(C_6H_{10}O_5)_n$ and "n" states for number of monomers (Hon, 1994). Figure 1.1 shows structure of cellulose. The main source of cellulose is the presence of the polymer in the cell wall structure of higher plants due to its firmly linked long linear chains. Beside of plants, other various organisms including green algae, fungi, cyanobacteria and bacteria produce cellulose for different purposes (Kimura & Itoh, 2001; Nobles et al., 2001).

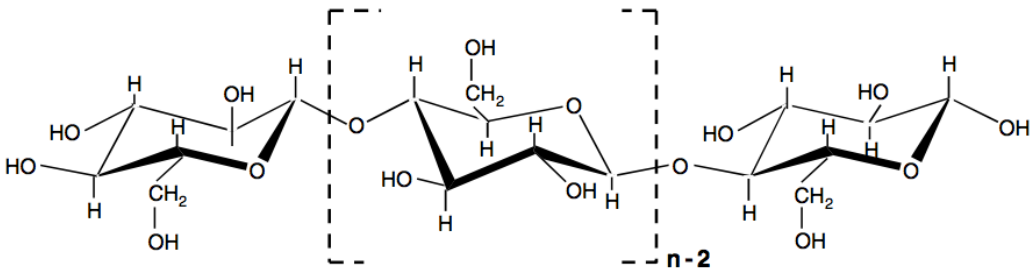

Figure 1.1: Structure of cellulose (Wüstenberg, 2014).

Degradation of cellulose by cellulose-hydrolyzing enzymes is important for flow of carbon from organic carbon sinks to atmospheric CO_2 in treatment of agricultural products and wastes (Berner, 2003). Even though fossil fuel shortage is not a serious problem for near future, cellulosic biomass, complex of cellulose, hemicellulose and lignin (Figure 1.2), is good candidate for production of renewable energy such as biofuel (Mosier et al., 2005). Conversion of biomass to fermentable sugar would bring two advantages; effects of cellulosic waste on environment would be decrease due to reuse of waste and waste material would be converted to a renewable energy

source that can cease our dependence of fossil fuels (Mosier et al., 2005). However, the recalcitrant nature of cellulosic waste due to its crystalline structure is a major limitation for conversion processes. Conversion of cellulosic biomass to fermentable sugars requires three steps: size reduction, fractionation and enzymatic hydrolysis (Wyman, 1994). The enzymatic hydrolysis of cellulose requires synergistic activity of three cellulase enzymes: endoglucanase, exoglucanase or cellobiohydrolase, and β-glucosidase (Bashir et al., 2013).

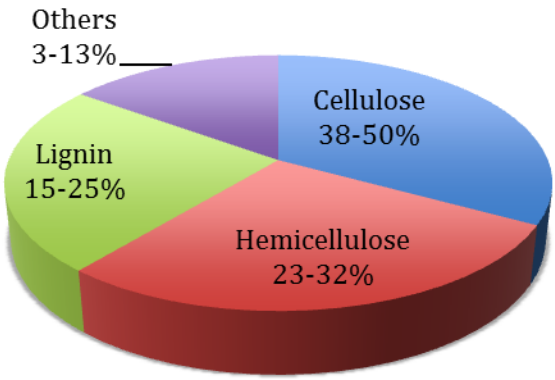

Figure 1.2: Composition of cellulosic biomass. The ratio may differ depending on the plant or tissue type.

1.1.2 Cellulose-hydrolyzing enzymes: endoglucanase, β-glucosidase and exoglucanase

Cellulases, which are responsible for hydrolysis of cellulose, are produced by various organisms including Fungi, Bacteria, Protozoa, plants and animals (Koo, 2001). Cellulases can be classified into three types of enzymes: endoglucanases (EC 3.2.1.4), which hydrolyze internal glycosidic bonds in soluble regions of the cellulose; exoglucanases (EC 3.2.1.91), which hydrolyze reducing and non-reducing ends of cellulose chains and release short-chain oligosaccharides; and β-glucosidases (EC 3.2.1.21), which convert oligosaccharides products to glucose (Golan, 2011). For full degradation of cellulose involves complex interaction and synergetic action of these three enzymes (Figure 1.3).

1.1.2.1 Endoglucanases

Endoglucanase (endo-β-1,4-glucan-4-glucanohydrolases) is the cellulolytic enzyme involved in the first step of cellulose degradation. This enzyme hydrolyzes the inner regions of the cellulose fiber randomly and reduces the degree of polymerization (Golan, 2011). Many bacteria, fungi and plant that produce this enzyme possess different catalytic modules. To exemplify, while fungal endoglucanases possess

single catalytic module and carbohydrate-binding module (CBM), bacterial ones contain multiple catalytic modules and CBMs (Zhou et al., 2008).

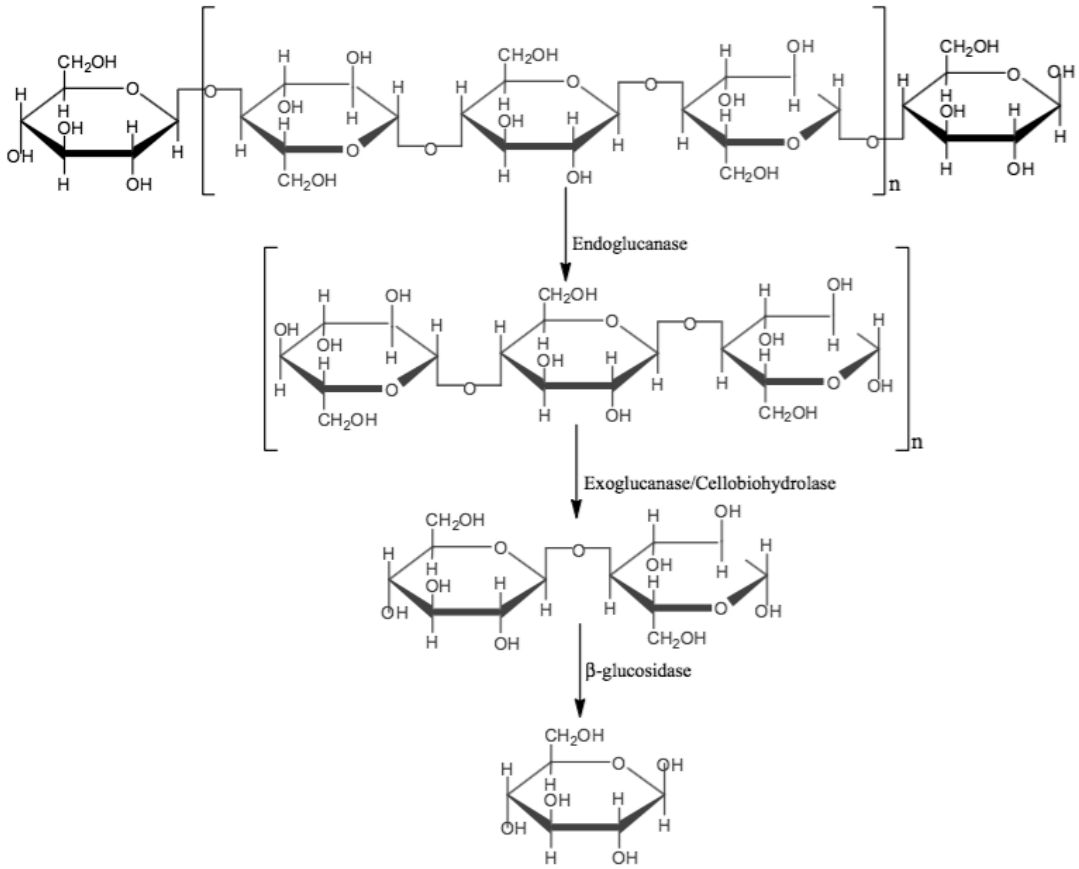

Figure 1.3: Enzymatic hydrolysis of cellulose (Juturu & Wu, 2014).

1.1.2.2 Exoglucanases

Exoglucanases removes cellobiose units from both the reducing and the non-reducing ends of the cellulose (Lynd et al., 2002). Exoglucanase are divided into subgroups: cellobiohydrolase (CBH) and glucanohydrolase (GH) (Golan, 2011). The cellobiohydrolase (β-1,4-D-glucan-cellobiohydrolase) hydrolyzes only the terminal non-reducing end of cellulose oligosaccharides – with some exception of attach on reducing end (Golan, 2011). This enzyme causes physical disruption of cellulose fiber and exposes polymer for further cellulolytic attacks. The CBHs are inhibited by cellobiose, the end product of hydrolysis (Zhou et al., 2008). Glucanohydrolase (β-1,4-D-glucan-glucanohydrolase) is important enzyme since it can directly release glucose polymer (Golan, 2011). Similar to endoglucanases, fungal exoglucanases possess single catalytic module and carbohydrate-binding module (not necessarily), bacterial ones contain multiple catalytic modules and CBMs (Zhou et al., 2008).

1.1.2.3 β-glucosidases

β-glucosidase (β-glucoside glucohydrolase), one of the major group of the glycoside hydrolases (Henrissat & Davies, 1997), catalyze hydrolysis of cellobiose and other cellodextrins into glucose (Golan, 2011). Various organisms including bacteria, fungi, plants, nonvertebrates, and vertebrates produce the enzyme (Matteotti et al., 2011). β-glucosidase is grouped into classes based on substrate specificity: Class I, which includes enzymes with glycosyl beta-glycosidase and aryl beta-glycosidase activity; Class II, that contains enzyme with only glycosyl beta-glucosidase activity; and Class III, that possess enzymes with only aryl/alkyl beta-glucosidase activity (Koffi et al., 2012). β-glucosidase is speed-limiting step of the cellulose degradation, inhibiting the accumulation of cellobiose (Golan, 2011). Similar to cellobiohydrolase, β-glucosidase is inhibited by its own hydrolysis product (Chauve et al., 2010).

Cellulase system involves at least three forms of synergism between these three cellulases: endo-exo synergy, providing reducing and non-reducing end for the action of cellobiohydrolase; exo-exo synergy, acting on reducing and non-reducing ends of cellulose chains; and exoglucanases and β-glucosidases synergy, which removes cellobiose as end products (Lynd et al., 2002).

1.1.3 Sources of cellulases

A variety of organisms including bacteria, fungi, as well as insects are capable of degrade cellulosic materials as an energy source (Maheshwari et al., 2009). However, only a few of these microorganisms are able to secrete a complex of cellulase enzyme that completely hydrolyzes crystalline cellulose (Golan, 2011). Only those organisms that produce required amount of endoglucanase, exoglucanase and β-glucosidase are able to degrade native lignocellulosic biomass. Most of the microorganisms are known to produce either endoglucanase or β-glucosidase (Kumar et al., 2008). β-glucosidase, which hydrolyze cellobiose to glucose, is mostly lacking or produced in small amounts in the cellulase complex and limits the speed of hydrolysis reaction (Bisaria & Ghose, 1981).

Some bacterial cellulases are present as supramolecular machines attached to their outer surface known as cellulosomes (Shoham et al., 1999). Cellulosome structure is given in Figure 1.4. Cellulosome do not necessarily contain CBM, but scaffoldin,

protein attached to cellulases, contains a CBM and binds enzyme complex to cellulose (Wilson, 2011). The cellulosome enables optimal synergism between the three cellulase enzymes and facilitates efficient uptake of hydrolysis product to cell by minimizing the distance (Schwarz, 2001). Bacterial strains producing cellulases includes *Rhodospirillum rubrum*, *Cellulomonas fimi*, *Clostridium stercorarium*, *Bacillus polymyxa*, *Pyrococcus furiosus*, *Acidothermus cellulolyticus*, *Saccharophagus degradans* and *Cytophaga hutchinsonii* (Golan, 2011).

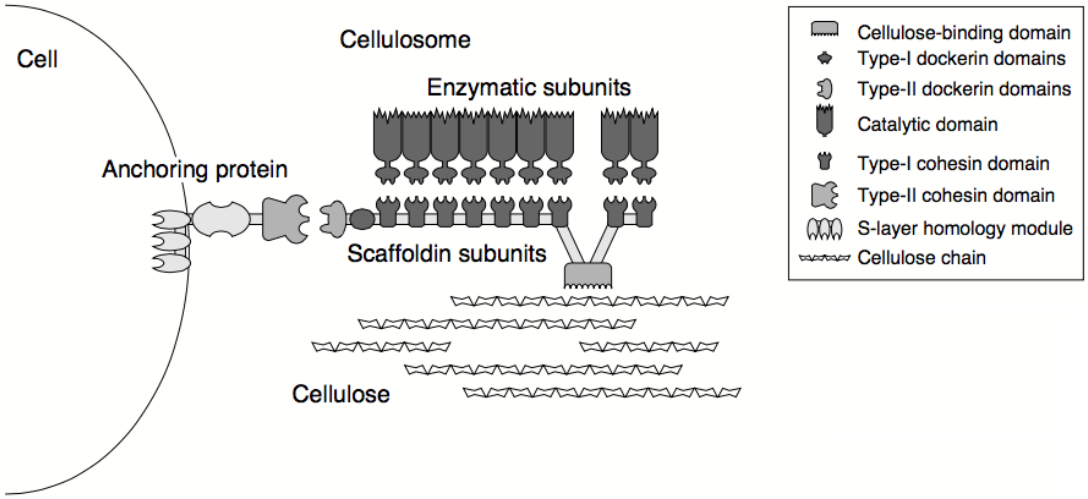

Figure 1.4: Components of cellulosome structure (Shoham et al., 1999).

Several fungi produce significant amounts of extracellular cellulases (Golan, 2011). Similar to bacterial one, fungal cellulases facilitates cellulosic hydrolysis synergistically with endoglucanases, exoglucanases and β-glucosidases (Zhou & Ingram, 2000). The cellulase producing fungi strains include *Agaricus sp.*, *Ganoderma sp.*, *Fusarium oxysporium*, *Piptoporus betulinus*, *Penicillium echinulatum*, *P. purpurogenum*, *Aspergillus niger*, *A. fumigatus*, as well as thermophilic fungi such as *Sporotrichum thermophile*, and *Scytalidium thermophillum*, (Kaur et al., 2004; Martins et al., 2008; Valášková & Baldrian, 2006). Fungal cellulases are not very effective on hydrolysis of crystalline cellulosic substrates (Kaur et al., 2004).

1.1.4 Application of cellulase

Cellulases, which represent 20% of the world‘s sale of industrial enzyme, have potential application in various industries including chemicals, fuel, food, brewing and wine, animal feed, textile and laundry, and pulp and paper sector (Bhat, 2000). Currently, cellulases are the third largest industrial enzyme worldwide, and expected

to increase if fermentation product of sugars becomes a major transportation fuel (Wilson, 2011). Almost all industrial cellulases are produced from *Trichoderma reesei*, an aerobic cellulolytic fungi, due to its ability to produce cellulase enzyme with relatively high enzymatic activity (Miettinen-Oinonen & Suominen, 2002).

Cellulases have a wide range of applications in food industry. Cellulases are used for different purposes including extraction of juices and oil from seeds and fruits, juice clarification, extraction of proteins from soybeans and coconuts, extraction of starch from corn and sweet potato, improving digestibility of food products (Béguin & Aubert, 1994; Coughlan et al., 1985). Also, by removing cell wall, cellulases liberate proteins, enzymes, polysaccharide and flavors. Thus, they increase nutritive quality of food products (Bhat & Bhat, 1997).

In brewery and wine industries, cellulases involve in beer filtration since they hydrolyze β-1,3 glucan and β-1,4 glucan present in barley. Brewery industries have been used recombinant yeast that possesses β-1,3-glucanases and β-1,4-glucanases for production (Bhat, 2000). Usage of β-glucanases with pectinases and hemicellulases contributes clarification and filtration; improvement of color extraction; and improvement of wine quality and stability (Golan, 2011).

Cellulases are capable of modifying cellulosic fibers in a desired manner, and are able to improve the quality of fabrics. Textile and laundry industry exploits cellulases for removal of excess dye from fabrics and jeans (biostoning), removal of cotton microfibrils resulting from washing, fabric softening and brightening (Cortez et al., 2001).

The animal feed industry with 600 million tones production and worth of 50 billion US dollars is important sector (Bhat & Bhat, 1997). The uses cellulases in feed industry incudes: elimination of anti-nutritional factors in feeds, improvement the nutritional value of feed, improvement the digestibility of ruminants (Coughlan et al., 1985).

Pulp and paper industry uses cellulases for different purposes. Biomechanical pulping, which requires cellulases, is the enzymatic treatment of lignocellulosic materials before the mechanical pulping step and it reduces energy consumption during processing (Buchert et al., 1998). Also, cellulases involve in fiber modification and deinking of paper products (Buchert et al., 1998).

Environmental problems due to increased carbon dioxide emission and serious concerns about fossil fuel shortage directed scientist to development of alternative renewable energy sources (Lee et al., 2008). Ethanol is a biodegradable, low toxicity alcohol and does not produces serious environmental problems when it is hydrolyzed to carbon dioxide and water (Bhat & Bhat, 1997). Bioethanol produced from hydrolysis lignocellulosic biomass of has potential to replace fossil fuels (Kumar et al., 2008).

In addition to industrial application, cellulases are used for the preparation of plant and fungal protoplasts, hybrid strains in production and other genetic engineering applications (Bhat & Bhat, 1997).

Cellulases have a wide range of applications, but more research is required to discover new cellulases with better commercial potential. Metagenomic discovery of environment with high cellulolytic activity is one of the most promising methods for determination of novel cellulases. So far, various environments including fresh waters, oceans, guts of insects and ruminants as well as human intestines were used for generation of metagenomic datasets (Sukharnikov et al., 2011).

1.2 Metagenomics and Application Areas

Metagenomics, study of uncultured microorganisms obtained from a certain environment, involves extraction of DNA directly from environment samples (Handelsman et al., 1998). It is estimated that approximately 1% of microbial population is culturable under laboratory conditions (Sharma et al., 2005). Metagenomics offers culture-independent method for investigation of enormous diversity of taxonomically and phylogenetically relevant genes, catabolic genes, and operons (Batraa et al., 2013). All metagenomic approaches mainly involves isolation and examination of DNA extracts that directly isolated from environment. Following DNA extraction, depending on the path to follow metagenomics is divided into two groups: function-based metagenomics, which bases on examination of clones with desired phenotype after preparation of DNA libraries (Batraa et al., 2013); and sequence-based metagenomics, involving the search for genes coding for conserved sequences or motifs (Ferrer et al., 2005). The latter one generates enormous data and enables discovery of millions of novel genes that can be used to enlighten metabolic pathways from uncultured bacteria (Ferrer et al., 2005). High-capacity sequencing

and development of advanced computational analysis resulted in dramatic improvements in gene discovery without performing functional screening (Thomas et al., 2012).

1.2.1 Approaches to metagenomic studies

Both metagenomic approaches requires extraction of DNA directly from naturally occurring microbial populations (Batraa et al., 2013). Quality of isolated DNA should be appropriate for construction of metagenomic library. During DNA extraction, DNA shearing should be minimized and co-extraction of contaminants should be prevented (Cowan et al., 2005). Schematic representation of metagenomic approaches is given in Figure 1.5.

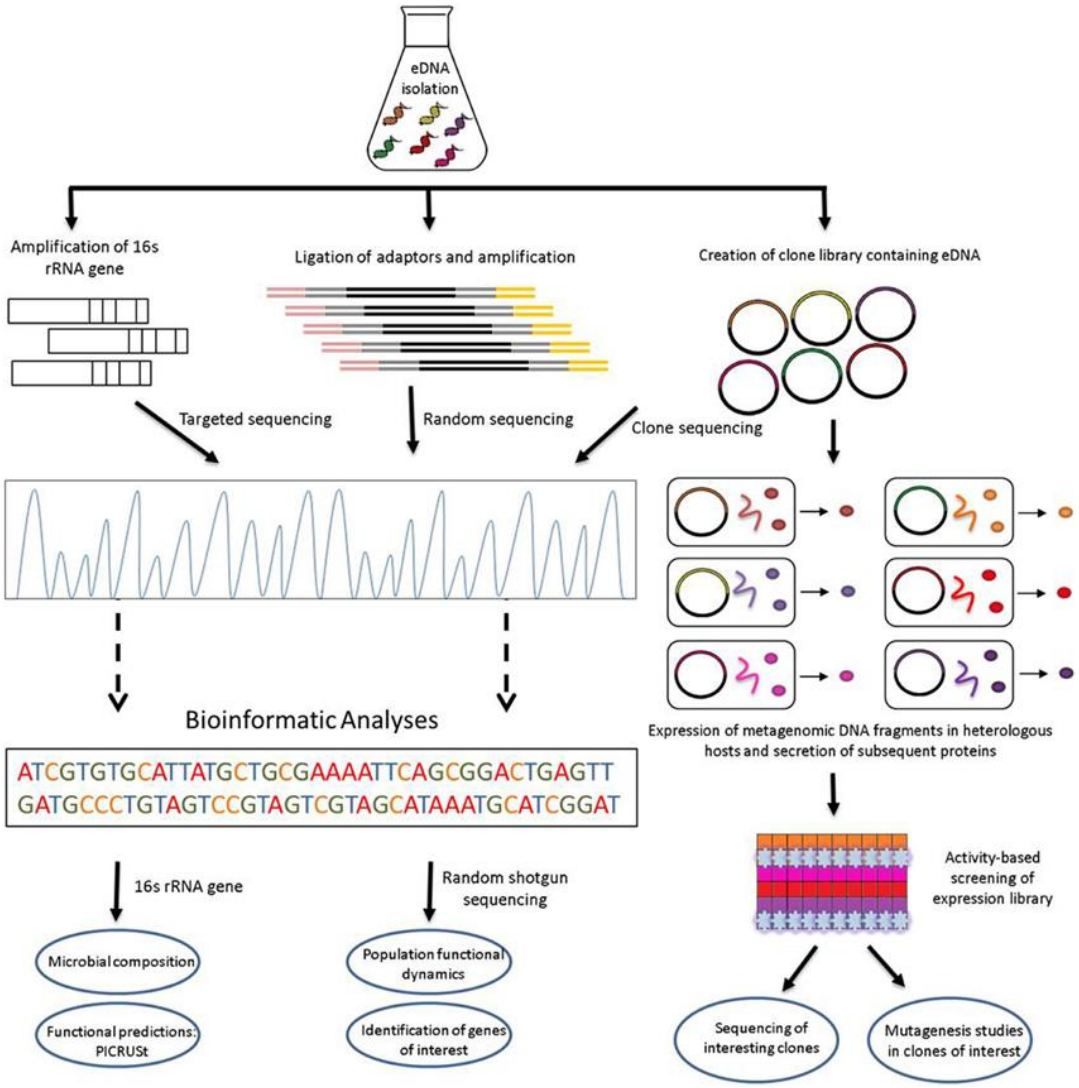

Figure 1.5: Schematic representation of function-based and sequence-based metagenomics (Coughlan et al., 2015)

Sequence-based metagenomics involves DNA extraction from microbial communities, whole-genome sequencing, and assembly of the reads. Following assembly, sequences are analyzed using bioinformatic tools to determine the composition and possible functions of the microbial populations (Ufarte et al., 2015). Schematic representation metagenomics using next-generation sequencing is given in Figure 1.6. Beside of high-throughput sequencing, use of a gene of interest as an anchor to identify clones of interest is possible. After construction of metagenomic library PCR can be used for amplification of the anchor. The clones harboring the anchor are sequenced and used for further analysis (Sabree et al., 2009).

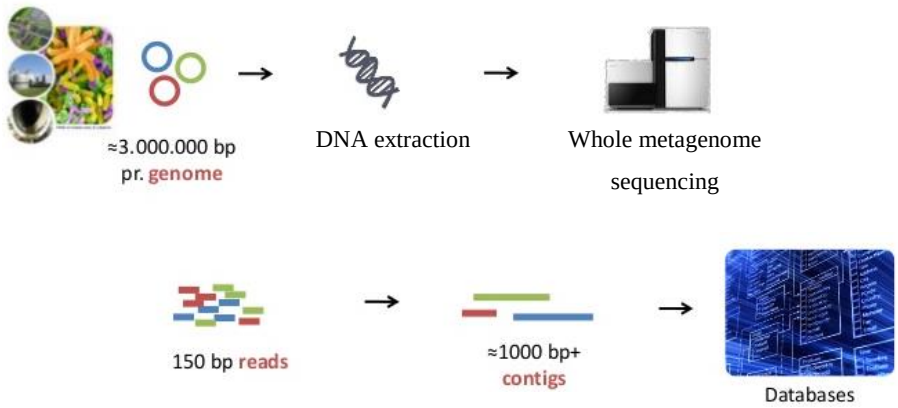

Figure 1.6: Schematic representation of metagenomic approach using next-generation sequencing (Albertsen, 2013).

1.3 Bioprospecting of the Metagenome

Metagenomics has enabled development of biotechnology based on the exploitation of uncultivated microbial species by making exploration and analysis of ecology around us possible (Bashir et al., 2014). Metagenomics finds various applications in industry including identification of new biocatalysts, discovery of new antibiotics, medicine. Also, biosurfactant producing bacteria were exploited for bioremediation of industrial, agricultural, and domestic wastes.

1.3.1 Discovery of bioactives, biosynthetic pathways and antibiotics

Bioactive compounds are nutritional constituents that occur in small quantities in foods (Kris-Etherton et al., 2002). Function-based metagenomics can be used to discover genes that carrying out reactions of interest for the production of bioactives or production of intermediate compounds (Coughlan et al., 2015). Instead of chemical synthesis, the use of microorganisms capable of producing targeted product

offers more environmental friendly alternative for industries. Microbial biosynthetic genes that were identified by functional metagenomic strategies include: biotin (Entcheva et al., 2001), pederin (Piel, 2002), salt tolerance genes (Culligan et al., 2013), and acid resistance genes (Guazzaroni et al., 2013).

Microorganisms produce antibiotic molecules to eliminate competitors in their habitat (Coughlan et al., 2015). Currently, antibiotic therapies are capable of curing most bacterial infections. However, development of antimicrobial resistance is emerging problem for antimicrobial therapy (Alanis, 2005). The improvement in metagenomics allowed screening of microbial populations for novel antimicrobial activity (Jayasuriya et al., 2007). Until now, metagenomic was exploited for mining of novel antimicrobials and studies resulted in discovery of new antibiotics, anti infectives, and antibiotic resistance genes (Coughlan et al., 2015).

1.3.2 Bioremediation

Evaluation of the variation between a contaminated environment and its uncontaminated condition is one of the main subjects of bioremediation (Bella et al., 2012). Metagenomics can give information about microbial and enzymatic diversity of contaminated area and allow the discovery of potentially active genes and organisms (Ufarte et al., 2015). At the same time, metagenomic studies can elucidate how microbial communities respond to changes in environmental conditions such as nutrients, carbon sources, pH, temperature, oxygen, and water content (Bella et al., 2012).

1.3.3 Discovery of novel enzymes

Global industrial enzyme sales were calculated to be $2.3 billion, which were represented by detergents, $789 million; food applications, $634 million; agriculture/feed, $376 million; textile and laundry, $237 million; and others including pulp and paper, leather and bulk chemical, $222 million (Lorenz & Eck, 2005). Moreover, demand for novel enzymes increases each passing day. Production of chemical catalysts that perfectly mimics biological enzymes is not possible and also it burdens serious problems for environment (Lorenz et al., 2002). Currently, metagenomics is accepted to be one of the most potent technologies to acquire the candidate molecules for biocatalytic processes. Some of the main enzymes that were

discovered by metagenomic approach include: cellulases, lipases, xylanases, amylases, proteases, nitrilases, dehydrogenases (Bashir et al., 2014).

1.3.4 Metagenomic approaches to the discovery of cellulases

Due to their high potential applications in biotechnology, cellulases have drawn enormous attention in research and commercial area. Use of environments, in which lignocellulose is being effectively decomposed, for discovery of novel cellulases is the most promising technology (Sukharnikov et al., 2011). Datasets generated from different studies and environments revealed great potential for discovery of novel cellulolytic enzymes. Metagenomics have been used to identification of cellulases from various environmental including soil, gut of higher termite, compost, rabbit cecum, sludge reactors (Duan & Feng, 2010). The rumen with its residing microbiota serves as a natural bioreactor for degradation of cellulosic biomass (Weimer et al., 2009). Thus, high number of metagenomic studies focused on isolation of cellulases from gut symbionts (Duan & Feng, 2010; M. Ferrer et al., 2005; Hess et al., 2011; Shedova et al., 2009). The termite hindgut contains higher number of cellulase domain representatives when compared to microbiome of cow rumen.

Similar to any other function-based metagenomics, insufficient expression of foreign cellulase genes is the major obstacle of enzyme discovery. To overcome these limitations, appropriate hosts for library construction and expression should be selected (Duan & Feng, 2010). Sequence-based approach is gain attention with improvements in sequencing technology. The presence of high-throughput DNA sequencing ceased necessity of preparation of clone libraries for identification of cellulase in metagenome (Rooks et al., 2012). The fragmented metagenomic DNA can be directly used for sequencing and obtained reads are used for assembly. Cellulase encoding genes can be identified from generated library using appropriate databases.

1.4 The Aim of the Study

In this study, sequence-based metagenomic approach was exploited for discovery of novel cellulose-hydrolyzing enzymes. For this purpose, thirteen samples were obtained from various environments that includes: water samples from Dalaman

Spring, Dalaman Lake, Emet hot spring, and Kuzuluk hot springs; straw samples that were incubated in Dalaman lake and Emet Hot Spring for enrichment of cellulolytic microorganisms; and grasshoppers obtained from TUBITAK MAM campus and commercial farm; grass compost from TUBITAK MAM. The DNA isolations were performed for each sample and extracted DNAs were used for generation of metagenomic libraries. Generated libraries were used for seek of cellulase enzymes using bioinformatic tools. Candidate genes were selected and cloned into expression host for enzymatic assays. At the same time, generated libraries were used for evaluation of microbial diversity in samples.

2. MATERIALS AND METHODS

2.1 Culture Media, Buffers, Solutions, Chemicals and Enzymes

2.1.1 Bacterial strains, chemicals and kits

In this study, bacterial strain *E. coli* BL21 (DE3) was used as cloning and expression host. Cells were grown in Luria Broth liquid medium (LB) on a shaker (220 rpm) or on agar plates (LA) at 37^{o}C. Plasmid used for cloning and expression was pET-22b(+) (Figure A.1). Bacterial glycerol stocks were prepared for long-term storage at -80^{o}C. For preparation of stocks, 500 µL overnight grown cultures were added to 500 µL 30% glycerol in a 2 mL screw top cryo tubes and mixed gently. Glycerol stock tubes were froze with liquid nitrogen and stored at -80°C. Kits and chemicals that were used in this study are given in Table 2.1.

Table 2.1: Kits and chemicals.

Kit	Producer
EZ-Fusion™ Cloning Kit	Enzynomics
Genomic DNA Clean & Concentrator	Zymo research
HisPur™ Cobalt Purification Kit	Thermo Fisher Scientific
Meta-G-Nome™ DNA Isolation Kit	Epicentre
Pierce™ Coomassie Plus (Bradford) Assay Kit	Thermo Fisher Scientific
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
ZR Fecal DNA MiniPrep™	Zymo research
Chemicals	**Producer**
2-Mercaptoethanol	Sigma-Aldrich
3,5-dinitrosalicylic acid	Sigma-Aldrich
4-Nitrophenol	Sigma-Aldrich
Acrylamide/Bis-acrylamide, 30% solution	Sigma-Aldrich
Agar	Sigma-Aldrich
Agarose	Sigma-Aldrich
Ampicilline	Sigma-Aldrich
Bromophenol blue	Sigma-Aldrich
Calcium chloride	Merck

Table 2.1 (continued): Kits and chemicals.

Chemicals	Producer
Carboxymethylcellulose sodium salt	Sigma-Aldrich
Citric acid	Sigma-Aldrich
Coomassie blue R-250	Applichem
Dibasic sodium phosphate	Sigma-Aldrich
EDTA	Sigma-Aldrich
Ethanol absolute	Merck
Ethidium bromide	Sigma-Aldrich
Gel Loading Dye, Purple (6X)	New England Biolabs Inc.
Glacial Acetic acid	Riedel-de Haen
Glucose	Sigma-Aldrich
Glycerol	Sigma-Aldrich
Glycine	Sigma-Aldrich
Imidazole	Sigma-Aldrich
IPTG (isopropyl-β-d-thiogalactopyranoside)	Sigma-Aldrich
Luria Broth	Merck
Magnesium chloride	Sigma-Aldrich
Methanol	Carlo Erba
Monobasic sodium phosphate	Sigma-Aldrich
O'GeneRuler DNA Ladder Mix	Thermo Fisher Scientific Inc.
p-nitrophenyl-beta-D-glucoside	Sigma-Aldrich
Potassium chloride	Sigma-Aldrich
SeeBlue® Plus2 Pre-stained Protein Standard	Thermo Fisher Scientific
Sodium carbonate	Sigma-Aldrich
Sodium chloride	Sigma-Aldrich
Sodium citrate dihydrate	Sigma-Aldrich
Sodium dodecyl sulfate	Sigma-Aldrich
Sodium hydroxide	Sigma-Aldrich
Sodium potassium tartrate	Merck
Tris	MP Biomedicals
Tris-Base	Promega
Tris-HCl	MP Biomedicals
Trisodium citrate dihydrate	Merck
Triton-X-100	Calbiochem
Tryptone	Sigma-Aldrich
Yeast extract	Sigma-Aldrich
Enzymes	**Producer**
*Nde*I	New England Biolabs
Q5® High-Fidelity DNA Polymerase	New England Biolabs
Taq DNA polymerase	New England Biolabs
*Xho*I	New England Biolabs

2.1.2 Buffers and solutions

Buffers and solutions that were used in this study including their content are given in Table 2.2.

Table 2.2: Buffers and solutions.

Buffers and Solutions	Content	Amount/Concentration
Ethidium bromide Solution	Ethidium bromide (10 mg/ml)	100 µl
	TAE buffer	Up to 1 l
Grasshopper microbiota collection buffer	EDTA	5 mM
	Glycerol	5%
	Tris base	2 mM
	Triton-X-100	0.2%
SOC medium	Tryptone	2% (w/v)
	Yeast extract	0.5% (w/v)
	NaCl	10 mM
	KCl	2.5 mM
	$MgCl_2$	10 mM
	Glucose	20 mM
Sonication buffer	Glycerol	10%
	Tris-HCl	50 mM
	NaCl	0.58 %
Sodium phosphate buffer	Monobasic sodium phosphate	50 mM, 19 mL
	Dibasic sodium phosphate	50 mM, 81 mL
TAE	Acetic acid	20 mM
	EDTA	2 mM
	Tris base	40 mM
SDS-PAGE BUFFERS		
Coomassie blue R-250 staining	Coomassie blue R-250	0.25 g
	dH_2O	100 mL
	Glacial acetic acid	25 mL
	Methanol	125 mL
Destaining solution	dH_2O	800 mL
	Glacial acetic acid	100 mL
	Methanol	100 mL
Running buffer (5X)	dH_2O	Up to 1 l
	Glycine	72 g
	SDS	5 g
	Tris base	15 g
Sample buffer	Tris HCl (0.5 M)	1 ml
	SDS (10 %)	0.4 mL
	Glycerol	1.6 mL
	dH_2O	3 ml
	Bromophenol blue (0.5 %, w/v)	0.4 mL
	β-mercaptoethanol	0.4 mL

Table 2.2 (continued): Buffers and solutions.

ENZYME ASSAY BUFFERS		
Buffers and Solutions	Content	Amount/Concentration
Substrate solutions	CMC	1 %
	pNPG	10 mM
DNS solution	DNS	35g
	NaOH (2 M)	200 ml
	Sodium potassium tartrate	300 g
	dH_2O	Up to 1 lt
Sodium citrate buffer	Citric acid	50.70, 50 mM
	Trisodium citrate dihydrate	49.30, 50 mM

2.1.3 Equipment

Laboratory equipment that was used in this study was listed in Table 2.3.

Table 2.3: Laboratory equipment.

Laboratory equipment	Producer
Autoclave	CLG, ALP Co., Ltd.
Balance	CPA Analytical Balance, Satorius
Centrifuges	5417 R, Eppendorf
	Megafuge 2.0R, Heraeus
Deep freezers and refrigerators	NEXK, BEKO
	ULT Freezer, Haier
Gel systems	Minicell® Primo EC320, EC
	Mini-PROTEAN® 3 Cell, Bio-Rad
Heat block	Dri-Block®, Techne
Incubators	FN500, Nuve
	Incubator shaker, New Brunswick Scientific Co., Inc.
Laminar flow cabinet	Class II Biological Safety Cabinet, ESCO
Magnetic stirrer	RTCbasic, IKA
Micropipettes	Eppendorf Research
Microwave oven	MD585, Arcelik
pH meter	744 pH meter, Metrohm
Power supplier	Model 200/2.0, Bio-Rad
	PowerPac™ HC, Bio-Rad
Spectrophotometers	ND-1000, Nanodrop®
	SmartSpec™ 3000, Bio-Rad
	Microplate reader, Bio-Rad
Thermocyclers	Mastercycler gradient, Eppendorf
	TC-512, Techne
Gel documentation	Uvi
Ultrapure water system	Purelab, ELGA

Table 2.3 (continued): Laboratory equipment.

Laboratory equipment	Producer
Ultrasonic disintegrator	Soniprep 150, MSE
UV-cabinet	UVT-B-AR, Biosan
Vortex	Benchmixer, Benchmark
Water baths	BTW-U, Biosan

2.2 Collection of Samples

2.2.1 Local grasshopper and farm grasshopper

Total numbers of 203 local grasshoppers were collected from The Scientific and Technological Research Council of Turkey Marmara Research Center (TUBITAK MAM) campus into sterile glass jars using gloves to prevent contamination of grasshoppers.

Total numbers of 432 desert grasshoppers (*Schistocerca gregaria*) were purchased from Mira Canlı Hayvan Bocek Tur. Ins. Tarim Tic. Ltd. Sti. (Antalya, Turkey).

2.2.2 Spring waters and straw-adherent microbiota

Six water samples were collected for metagenomic DNA isolation. Two samples were collected from Kutahya Emet hot spring waters (Kutahya, Turkey, 39.203289, 29.152234; 39.202825, 29.15872) which have temperature of 50°C and 42°C, respectively. 20 liter of water samples were taken into sterile jars and stored at 4°C. Third sample was obtained from Kuzuluk hot spring waters (Sakarya, Turkey, 40.629181, 30.65718) which has temperature of 56 °C. 21 L of water samples were taken into sterile jars and stored at 4°C. Dalaman spring which is also called sulfur spring, located in Dalaman, Mugla, Turkey at 36.726163,28.8147, forms a pool nearby which is called Dalaman Lake and they both release noticeable hydrogen sulfide gas odor (Figure 2.1). Dalaman Spring possess stable temperature, 28-29°C, all around the year. 9.8 liter of lake water and 12.6 liter of spring water were taken into sterile jars and stored at 4 °C.

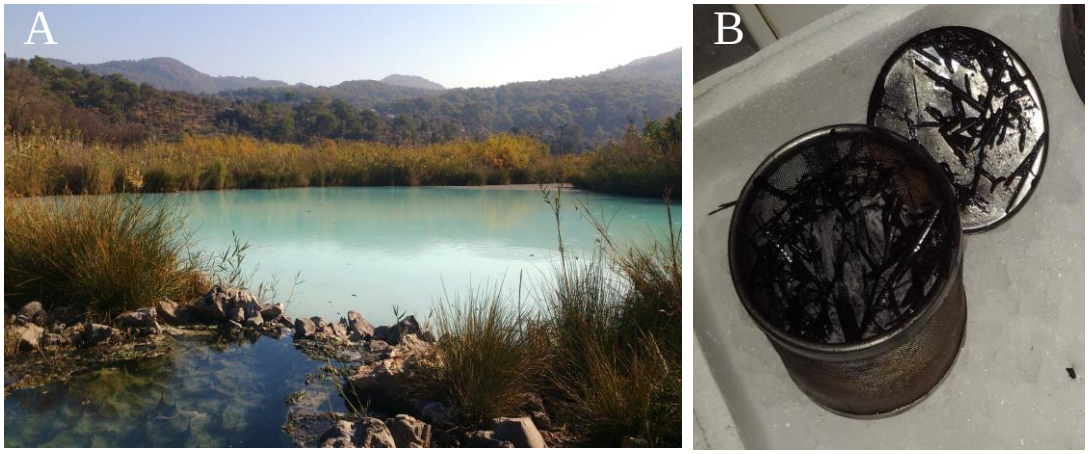

Figure 2.1: Dalaman and straw samples. **A)** Dalaman Spring (bottom left) and Dalaman Lake (upper right). **B)** Straw samples incubated in water samples.

To increase abundance of cellulolytic microorganisms in samples, autoclave-sterilized straws were put into metal cages and incubated in Dalaman Lake and 42°C Kutahya Emet hot spring waters for 2 months (Figure 2.1). After incubation, samples were immediately fixed with liquid nitrogen and stored at -80°C.

2.2.3 Grass compost pile

Grass compost sample was acquired from pile of cut grass at TUBITAK MAM campus (Figure 2.2). Grass compost was transferred into sterile 50 ml tubes using sterile forceps. Samples were immediately transferred to -20°C and stored there until DNA extraction.

Figure 2.2: Grass pile compost from TUBITAK MAM campus.

2.3 Metagenomic DNA Preparation and Purification

2.3.1 Local grasshopper and farm grasshopper gut microbiota

Prior to any procedure, grasshoppers were swabbed with 70% ethanol and washed with distilled water to remove any surface contaminants. To obtain sufficient high-quality DNA for sequencing, 635 grasshoppers were used for DNA extraction from gut symbionts. The digestive tracts of the grasshoppers were pulled out gently with head after cutting of rare-end of abdomen. Grasshopper guts were separated from head with sterile scissor. Approximately 1 ml of grasshopper microbiota collection buffer was injected through midgut and flow-through was collected in sterile microcentrifuge tubes. Cells were collected by centrifugation at 13000 rpm for 5 minutes and supernatant was discarded. Guts of all grasshoppers and feces of local grasshoppers were also collected into sterile microcentrifuge tubes. All samples were stored at -20°C until further processing.

The metagenomic DNAs were isolated using the ZR Fecal DNA MiniPrep™ (Zymo research, USA) according to the manufacturer's protocol for all grasshopper samples. The isolated metagenomic DNAs were quantified by Nano Drop ND-1000 spectrophotometer and characterized by electrophoresis. Also, for determination of quality of the DNA and quantity of host contamination PCR amplification of conserved 16S rDNA and conserved 18S rDNA were performed.

2.3.2 Spring waters and straw-adherent microbiota

Before DNA extraction, water samples were filter with filter paper to remove particles, which may clog collection filter. Afterwards, waters were filtered through a 0.45μm pore size filter membranes that are provided by Meta-G-Nome™ DNA Isolation Kit (Epicentre, USA). DNA was extracted from biomass-containing filters according to the manufacturer's protocol. Extracted DNAs were purified using Genomic DNA Clean & Concentrator (Zymo research, USA) as determined by its producers.

Genomic DNA of straw-adherent microbiota was isolated using ZR Fecal DNA MiniPrep™ (Zymo research, USA). 200 mg of straw is used for each isolation process.

2.3.3 Grass compost pile

ZR Fecal DNA MiniPrep™ (Zymo research, USA) was used for DNA extraction from grass compost. Approximately 250 mg of grass is used for each isolation process.

2.4 Sequencing of Metagenomic DNA and Bioinformatic Analysis

Sequencing of the metagenomic DNAs derived from samples were performed using HiSeq2500 sequencing system (Illumina, USA) by Advanced Genomics and Bioinformatics Research Group (IGBAM TUBITAK, Turkey). Metagenomic library was constructed with approximately 200 ng of DNA per sample by using TruSeq Nano DNA Library Prep Kit (Illumina, USA) according to the manufacturer's protocol. Briefly, input DNA was fragmented to 400 bp fragments with 3' or 5' overhangs with Covaris shearing. Overhangs resulting from fragmentation were converted into blunt ends. After end repair, the desired library size is selected with appropriate size beads. Blunt fragments were adenylated at 3' ends to inhibit ligation to each other. Also, adenylation provide complementary overhang for ligation to adapter containing "T" at 3' end. After addition of adapters to the ends of the DNA fragments, DNA fragments were enriched using bridge PCR to generate the DNA library. The sequence of DNA fragments were determined by using method called sequencing by synthesis.

The de novo assembly of obtained sequence reads was performed with Velvet Assembler (v1.2.07) (Zerbino & Birney, 2008). Velvet Assembler builds long continuous sequences, which are called contigs, and scaffolds, which are gapped assemblies of contigs, using short-read data obtained from Illumina HiSeq2500. The phylogenetic analysis of microorganisms was carried out with PhyloSift (Darling et al., 2014). PhyloSift searchs sequences for identity to known reference gene families and generates taxonomic summaries. After assembly of reads, the ORFs encoding cellulolytic enzymes that were present in each contig were predicted by MetaGeneMark (Zhu et al., 2010). The annotation was performed by searching the amino acid sequence for glycoside hydrolase (GH) families or for a carbohydrate-binding module that were defined by the CAZy database.

2.5 Cloning of Annotated Genes

2.5.1 PCR amplification of cellulase genes

Isolated metagenomic DNAs were used as template for PCR amplification of candidate genes that were annotated as endoglucanase, exoglucanase and β-glucosidase. PCR reactions were performed using Q5® High-Fidelity DNA Polymerase (NEB, USA) and reaction mixture included 10 µl 5X Q5 reaction buffer, 2 µl 5 mM dNTPs, 1 µl 25µM primers, 0.5 µl DNA template and dH_2O for a total reaction mixture of 50 µl.

Primers that were used in this study were designed for 63 genes in a way that they contained 18 base-pair homologous to one of the two ends of linearized pET-22b(+). Primers used in this study are shown in Table B.1. Numbers stated for gene IDs and F/R stated for forward and reverse, respectively. Underlining indicates restriction sites for *Nde*I (CATATG) and *Xho*I (CTCGAG), and italic type indicates homologous sequence with linear pET-22b(+). Primers 27F-1525R and 515F-1209R are commercial primers that used for amplification of 16S and 18S regions of rDNA. Primers were designed using Primer3Plus (Untergasser et al., 2012) and purchased from Sentegen (Turkey) and Thermo Fisher (USA).

The PCR temperature conditions were 98°C 30 s followed by 30 cycles 98°C 10 s, 50-72°C (depending on T_m of primers) 20 s, 72°C 30-60 s (depending on size of target sequence), and 72°C 2 min for final extension. PCR products were analyzed with 1% agarose gel electrophoresis in 1x TAE at 100V for 35 min and product size was verified. DNA cleanup was performed using QIAquick PCR Purification Kit (Qiagen, USA) and QIAquick Gel Extraction Kit (Qiagen, USA) according to manufacturer's instruction and DNA product was eluted with 30 µl dH_2O. DNA concentration was quantified using NanoDrop ND-1000 Spectrophotometer.

2.5.2 Construction of recombinant plasmid

2.5.2.1 Preparation of cloning vector

The pET-22b(+) plasmid was digested with the specific restriction enzymes *Nde*I and *Xho*I in order to be compatible with the insert DNA that was prepared according to EZ-Fusion Cloning Kit's instructions. Restriction reaction was carried out using 1 µl restriction enzyme for 1 µg DNA and mixture contained 5 µl CutSmart (NEB) buffer

for 50 µl reaction. Digestion was performed at 37 °C overnight. The DNA cleanup was performed with QIAquick PCR Purification Kit and linear pET-22b(+) plasmid was eluted in 30 µL dH_2O. Plasmid DNA concentrations were quantified by using Nanodrop.

2.5.2.2 Cloning of candidate genes into vector

The cloning of candidate cellulolytic genes into pET-22b(+) plasmid was performed by using EZ-Fusion™ Cloning Kit (Enzynomics, Korea). PCR-generated sequences and linearized pET-22b(+), efficiently recognize 18 bp overlapping sequence at their ends.. Reaction mixture contained 1:3 vector:insert molar ratio linear pET-22b(+) plasmid and candidate gene, 1 µL 5X EZ-Fusion™ Cloning PreMIX, and dH_2O for a total reaction mixture of 5 µl. Reaction mixture was incubated at 37°C for 15 minutes and then stored at 4°C until transformation.

2.5.3 Transformation of recombinant plasmid to *E. coli* BL21 (DE3)

2.5.3.1 Competent cell preparation

Competent *E. coli* BL21 (DE3) cells were prepared by using $CaCl_2$ method (Hanahan, 1983) with minor modifications. 500 µl of overnight-grown *E. coli* BL21 (DE3) culture was inoculated in 50 ml LB medium, and incubated at 37° C, 220 rpm until its absorbance at 600 nm reaches range between 0.5 and 0.7. The cells were transferred to sterile 50 ml tube and centrifuged at 4000 rpm for 10 min at 4° C. Supernatant was discarded and pellet was resuspended in ice-cold 25 ml 100 mM $CaCl_2$ solution. After 30 min incubation on ice, centrifugation step is repeated. Following disposal of supernatant, pellet was resuspended in 100 mM $CaCl_2$ solution containing 10% glycerol. The suspension was transferred to microcentrifuge tubes as 100 µl aliquots. Aliquots were immediately frozen in liquid nitrogen and stored at -80°C until transformation process.

2.5.3.2 Heat-shock transformation of recombinant plasmid

Transformation of recombinant plasmid was performed with heat-shock method (Sambrook et al., 1989) with slight changes. Competent *E. coli* cells were thawed on ice. 1.7 µl 1:10 2-mercaptoethanol solution was added, and incubated on ice for 10 minutes. EZ-Fusion reaction mixture (2 µl) containing recombinant plasmid was added into thawed competent cells, and placed on ice for 30 min. Mixture was heat

shocked at exactly 42°C for 45 seconds and then immediately placed on ice for 2 min. SOC medium was added and mixture was incubated at 37°C for 60 minutes and shaken vigorously (250 rpm) for cell regeneration. Cells were harvested by centrifugation at 8000 rpm and 100 µl of culture was spread on LA plate with ampicillin (100 mg/ml). Plates were incubated overnight at 37°C for growth of colonies.

2.5.3.3 Analysis of recombinant pET-22b(+) by colony PCR

For verification of recombinant cells, individual colonies from transformation plate were streaked to new amp-100 plate and incubated overnight at 37°C for growth. Then, individual colonies were selected from these plates with sterile toothpick and dissolved in 40 µl sterile dH_2O and boiled at 100°C for 5 min. After centrifugation at 13000 rpm for 1 min, supernatant was used as template DNA for PCR reaction. Reactions were performed using Taq DNA Polymerase with Standard Taq Buffer (NEB, USA) and mixture for colony PCR included: 5 µl 10X Reaction Buffer, 2 µl 5 mM dNTPs, 1 µl 25µM gene specific primers, 5 µl DNA template and dH_2O for a total reaction mixture of 50 µl. The PCR temperature condition was 95°C 30 s followed by 30 cycles 95°C 30 s, 45-68°C (depending on T_m of primers) 30 s, 68°C 30s-120s (depending on size of target sequence), and 68°C 5 min for final extension. PCR products were analyzed with 1% agarose gel electrophoresis in 1x TAE at 100V for 35 min and product sizes were verified.

2.6 Expression of Candidate Genes

For expression studies, single colony of transformant was picked from freshly grown plate and inoculated into 1 ml LB medium containing ampicillin (100 µg/ml) and incubated in shaking incubator at 37°C overnight. Starter culture was added into fresh LB medium containing ampicillin (100 µg/ml) in ratio of 3:100 and incubated at 37°C with shaking at 250 rpm to an OD_{600} of 0.5-1.0. Prior to induction, culture was split into two equal subcultures.

Protein expression was performed using Isopropyl β-D-1-thiogalactopyranoside (IPTG) induction procedure. IPTG was added one of the subcultures for a final concentration of 1 mM, and the other culture served as an uninduced control. Both of the cultures were incubated at 37°C with shaking for additional 2 hours. After

induction, cells were harvested with centrifugation at 13000 rpm 1 minute. Pellets were washed with 50 mM tris-HCl buffer and centrifugation was repeated. Supernatant was discarded and pellet that contained cells was stored at -20°C until further analysis.

Expression analysis was performed using SDS-PAGE electrophoresis. Cell pellet was dissolved with 100 µl SDS sample buffer and heated at 100°C for 5 min. 5 µl of the sample was loaded on SDS gel for analysis.

2.7 Enzymatic Assays

2.7.1 Preparation of crude extract

The cell pellets, which were contained culture induced with IPTG, were resuspended in sonication buffer. The resuspended pellets were subjected to sonication using a Soniprep 150 (MSE, UK) with 8 short bursts of 15 sec followed by intervals of 15 sec for cooling. Suspensions were kept on ice at all times to prevent denaturation of proteins. Cell debris was removed by centrifugation at 4°C for 20 min at 15000 rcf. Supernatants containing crude enzyme were stored at -4° C until activity assay was performed.

2.7.2 Endoglucanase and exoglucanase assay

Endoglucanase and exoglucanase activities were determined by measuring the hydrolysis of carboxymethyl cellulose (CMC) by method described by Bailey et al. (1992) with minor modifications. 1.8 ml of CMC solution (1% CMC in 50 mM sodium citrate buffer, pH: 4.8) was added into 13 × 100 mm test tube and tempered to 50° C. 200 µl crude enzyme was added into mixture and incubated at 50° C for 30 minutes. The enzymatic reaction was stopped by addition of 3 ml DNS reagent and heating in boiling water for 5-10 min. The absorbance was read at 540 nm against the reagent blank after cooling on ice for 5 min.

2.7.3 β-glucosidase assay

The procedure of Parry et al. (2001) was used to determine β-glucosidase activity with minor modifications. Reactions were prepared in 96-well plates and reaction mixture contained 25 µl of crude enzyme, 25 µl of p-nitrophenyl-β-D-glucopyranoside (pNPG, 10 mM) as substrate in sodium citrate buffer (50 mM, pH

4.8) and 50 µl of sodium citrate buffer (50 mM, pH 4.8). After incubation of reaction mixture at 50ºC for 30 min in thermal cycler (Mastercycler gradient, Eppendorf), the reaction was terminated by adding 100 µl of sodium carbonate (1 M). The developed yellow color due to release of p-nitrophenol (pNP) from pNPG was read at 405 nm using microplate reader.

2.8 Calculation of Specific Activity

2.8.1 Preparation of crude extract

To determine specific activity of biochemically active cellulolytic enzymes, protein isolation is performed. For protein isolation, 800 ml of LB medium was inoculated with 3% overnight grown cell culture. Induction of recombinant protein was carried out by addition of 1 mM IPTG and incubation at 37ºC for 2 hours. Following induction, cells were harvested by centrifugation at 4000g for 15 minutes, 4 ºC. Cells were suspended in 20 ml sodium-phosphate buffer (50 mM, pH 7.4), NaCl (300 mM) and imidazole (10 mM) and sonicated in this buffer to lyse cells. Sonication was performed using a Soniprep 150 (MSE, UK) with 8 short bursts of 30 sec followed by intervals of 30 sec for cooling. Suspension was kept on ice at all times to prevent denaturation of proteins. After removal of cell by centrifugation, crude enzyme was used for protein purification.

2.8.2 Protein purification

Target protein was isolated using HisPur™ Cobalt Purification Kit (Thermo Fisher) according to manufacturer's protocol. Briefly, 2 ml of cobalt resin was added to tube and centrifuged at 700 g for 2 minutes. Resin was washed with two resin-bed volumes of equilibration/wash buffer (sodium-phosphate buffer, 50 mM, pH 7.4; NaCl, 300 mM; and imidazole, 10 mM). Crude enzyme was mixed with resin by gentle shaking for 30 minutes to bind proteins to resin. Following mixture, tube is centrifuged at 700 g for 2 minutes. Resin was washed with two resin-bed volumes of equilibration/wash buffer twice and centrifuged for 2 minutes at 700 g. His-tagged proteins bound to resin was eluted with addition of one resin-bed volume of elution buffer (sodium-phosphate buffer, 50 mM, pH 7.4; NaCl, 300 mM; and imidazole, 300 mM). Purified protein was collected with centrifugation at at 700 g for 2 minutes.

Protein concentration was measured using Pierce™ Coomassie Plus (Bradford) Assay Kit (Thermo Fisher Scientific). Albumin (BSA) standards with different concentrations were prepared as suggested by manufacturer. Purified proteins and BSA standards (10 µL for each) were mixed with 300µL of the Coomassie Plus Reagent and incubated at room temperature for 10 minutes. Absorbance was measured at 595nm with microplate reader. Plotting the absorbance of BSA standards vs. its concentration in µg/mL, standard curve was formed and used for determination of the protein concentration.

2.8.3 Biochemical activity measurement

For calculation of specific activity glucose standards with concentration of 3 mg/ml, 2.5 mg/ml, 2 mg/ml, 1.5 mg/ml, and 1 mg/ml were prepared. Reaction mixture containing of 1% CMC (1.8 ml), glucose standard (200 µl) and DNS reagent (3 ml) were prepared and incubated in boiling water for 10 minutes. The absorbance was read at 540 nm against the reagent blank after cooling on ice for 5 min. The standard line was constructed with absorbance on y-axis and glucose concentration in x-axis.

To calculate amount of released glucose due to catalysis of CMC, 200 µl purified enzyme was added into 1.8 ml of 50°C 1% CMC. Reaction took place at 50°C for 30 minutes. Following reaction, 3 ml DNS solution was added to reaction mixture and incubated in boiling water for 10 minutes. The optical density was measured at 540 nm against the reagent blank after cooling. The standard line was used for calculation of released glucose. Activity and specific activity were calculated using equations 2.1 and 2.2, respectively.

$$\text{Activity} = \frac{\text{mg of glucose released} \times \text{dilution factor}}{\text{Duration of incubation in minutes}} \tag{2.1}$$

$$\text{Specific activity} = \frac{\text{Units}}{\text{mg protein}} \tag{2.2}$$

3. RESULTS AND DISCUSSION

The guts of insects feeding on cellulosic materials provide highly nutrient rich ecological niche for growth of cellulolytic symbionts. It is known that insects lack many of the required metabolites to fully utilize the food and requires commensal bacteria for biomass degradation, nutrient synthesis and other functions (Adlakha et al., 2011; Tartar et al., 2009). Grasshoppers feed on forbs (37.2%), grasses and sedges (58%) and others (Joern, 1983). Deconstruction of these biomasses requires synergistic action of enzymes secreted by the insects and their microbes residing in the guts (Shi et al., 2013). In this research for discovery of cellulolytic enzyme and biodiversity analysis, two grasshoppers that are from different niches and have different diets were selected: farm grasshopper and local grasshopper.

Microorganisms inhabiting in high temperatures mostly produce thermostable enzymes, which are more resistant to chemical denaturation (Schröder et al., 2014). Thermostable enzymes are advantageous for industrial processes requiring elevated thermal conditions. An elevated temperature reduces the chance of contamination and increases substrate solubility by lowering viscosity (Krahe et al., 1996). For instance, water samples from various destinations and temperatures were selected for discovery of cellulase enzyme. To diversify samples, water samples from Dalaman and grass sample from compost in TUBITAK MAM were obtained.

3.1 Extraction and Purification of Metagenomic DNA

In metagenomic researches success of further analysis directly depends on efficient recovery of high-quality microbial DNA from the environmental sample. In this study, to recover DNA with high yield and low fragmentation ZR Fecal DNA MiniPrep™ (Zymo research, USA) and Meta-G-Nome™ DNA Isolation Kit (Epicentre, USA) were used according to manufacturer's protocols. While the latter one used for only DNA isolation from water samples, ZR Fecal DNA MiniPrep™ (Zymo research, USA) was used for DNA extraction from grasshopper samples, grass compost and straws samples. Metagenomic DNA extracted from water samples

had yellowish color because of co-extracted humic acid and contaminants. Since contaminants in DNA solution inhibit enzymatic reactions such as restriction, ligation and amplification (Tebbe & Vahjen, 1993), Genomic DNA Clean & Concentrator (Zymo research, USA) was used for removal of these inhibitory substances. Recovered metagenomic DNA, which was qualified using electrophoretic gel systems (Figure 3.1), was suitable for further steps.

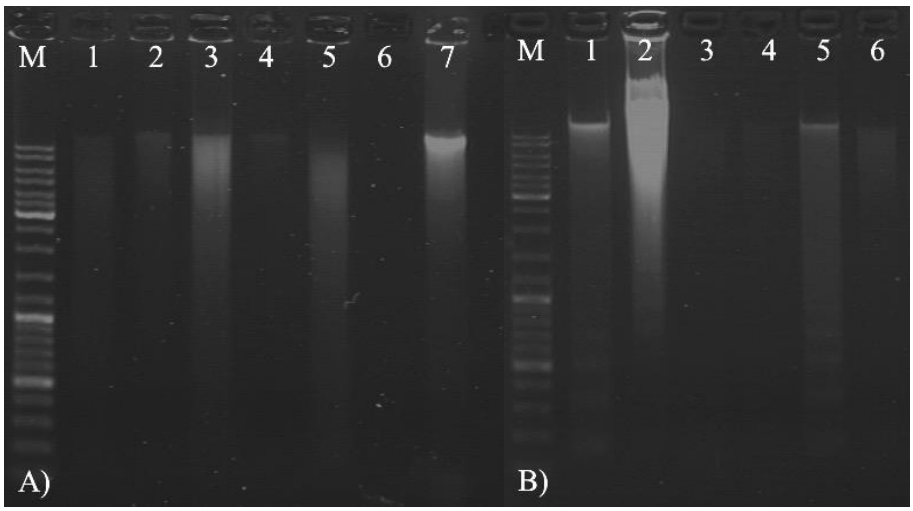

Figure 3.1: Agarose gel electrophoresis of metagenomic DNAs. **A)** M: DNA marker (O'GeneRuler DNA Ladder Mix), 1: Dalaman Lake, 2: Dalaman Spring, 3: Dalaman straw, 4: Emet 42°C, 5: Emet 42°C straw, 6: Emet 50°C, 7: Kuzuluk 56°C; **B)** M: DNA marker, 1: Farm grasshopper flow-through, 2: Farm grasshopper gut, 3: Local grasshopper feces, 4: Local grasshopper flow-through, 5: Local grasshopper gut, 6: Grass compost.

DNA quantification was performed using spectrophotometer (Table 3.1). The ratios of 260/280 and 260/230 were used to assess the purity of DNA. Ratios of 1.8 for 260/280 and 1.8-2.2 for 260/230 are generally accepted as pure for DNA (Scientific, 2008). Concomitantly, successful amplification of 16S and 18S regions of rDNA indicated compatibility of acquired DNAs for sequencing. The ratios lower than these values are indicator of co-purified contaminants such as protein, phenol or other contaminants.

DNA concentrations that were observed with gel electrophoresis were lower than concentrations measured with spectrophotometer. Contaminants, such as humic acid, strongly interfere with spectrophotometric DNA quantification (Steffan et al., 1988) and may be the result of big difference between two DNA concentration. Required amount of metagenomic DNA was given to Advanced Genomics and Bioinformatics Research Group (IGBAM, Turkey) for DNA sequencing.

Table 3.1: Concentration and purity of extracted DNA.

Sample Name	Concentration (ng/µl)	260/280	260/230
Dalaman Lake	37.7	1.22	1.51
Dalaman Spring	38.2	0.97	1.04
Dalaman Straw	58	1.57	1.43
Emet 42°C	20.8	1.23	2.07
Emet 42°C Straw	28.8	1.59	1.18
Emet 50°C	37	1.13	1.02
Kuzuluk 56°C	171.8	1.37	1.19
Farm grasshopper flow-through	82.35	1.66	1.44
Farm grasshopper gut	375.2	1.8	2.18
Local grasshopper feces	26.2	1.71	1.9
Local grasshopper flow-through	17.4	1.15	0.85
Local grasshopper gut	112.3	1.18	0.61
Grass compost	37.1	1.12	0.84

3.2 Whole Metagenome Sequencing and Analysis

3.2.1 Metagenome assembly

In this study, thirteen metagenomic DNA from various environmental samples were sequenced using HiSeq2500 sequencing system (Illumina, USA) by Advanced Genomics and Bioinformatics Research Group (IGBAM TUBITAK, Turkey) to discover new enzymes with cellulolytic activity. Approximately 68 gigabases of sequencing data were generated and all raw Illumina data were assembled with Velvet Assembler (v1.2.07) (Zerbino & Birney, 2008). The metagenome sequencing results were given in Table 3.2. Gene assemblies are composed of contigs which are overlapping sequence reads generated by reassembly of the small DNA fragments of sheared genomic DNA. N50 is a measure of average length of a set of contigs and evaluates quality of assembly. To calculate N50, every contig was ordered from longest to shortest and length of each contig summed starting from the longest one. When sum was equal to half of the total length of all contigs, the length of the shortest contig gives N50 value. The higher the N50 is, the more meaningful the assembly is (Yandell & Ence, 2012). Following de novo assembly, 786012 open frames (ORF) were predicted using MetaGeneMark (Zhu et al., 2010). All predicted ORFs were screened for candidate cellulolytic enzymes.

Table 3.2: Illumina HiSeq sequencing and de novo assembly results.

Sample Name	Number of reads	Mean read length	Number of contig	Mean contig length	N50 contig length	Percentage of assembled bases	Number of identified genes
Dalaman Lake	48941158	100	128561	1174.59	3582	55	55366
Dalaman Spring	107553068	100	1027068	604.55	791	76	201744
Dalaman Straw	57951142	100	94970	649.39	1016	14	53960
Emet 42°C	81491124	100	410540	1133.64	5780	87	149543
Emet 42°C Straw	35624942	100	76155	1589.96	6035	36	85136
Emet 50°C	24837346	100	97901	1001.91	10726	84	23570
Kuzuluk 56°C	48356788	100	131875	1550.25	15720	96	46507
Farm grasshopper flow-through	54885974	100	576958	853.22	480	45	10576
Farm grasshopper gut	30000000	100	308373	460.26	116	44	4610
Local grasshopper feces	49041106	100	170246	528.65	673	72	11732
Local grasshopper flow-through	58690036	100	425478	644.05	595	52	39349
Local grasshopper gut	36680954	100	476078	557.38	232	54	14303
Grass compost	39230526	100	247254	619.14	1115	44	89616

3.2.2 Identification of cellulolytic genes

All predicted genes were subjected to screening for possible enzymes with activity against cellulosic materials. ORFs were screened for presence of catalytic domains characteristic for glycoside hydrolase (GH) families and/or for a carbohydrate-binding module that were defined by the CAZy database (Cantarel et al., 2009). Gene annotation results were given in Table 3.3. GH families related with cellulases are GH1, GH3, GH5, GH6, GH7, GH8, GH9, GH12, GH30, GH45 and GH48 (Hess et al., 2011) and sequences were searched for these domains.

Table 3.3: Gene prediction results and abundance profile of glycoside hydrolases.

Sample name	Gene count	GH family gene count	Cellulase gene count	GH families
Dalaman Lake	55368	182	28	GH3, GH5, GH7, GH8, GH9, GH12
Dalaman Spring	201745	389	59	GH1, GH3, GH5, GH9, GH12
Dalaman Straw	53957	343	65	GH1, GH3, GH5, GH9, GH12
Emet 42°C	149542	586	90	GH1, GH3, GH5, GH9
Emet 42°C Straw	85136	701	189	GH1, GH3, GH5, GH9
Emet 50°C	23583	100	32	GH1, GH3, GH5, GH12
Kuzuluk 56°C	46506	265	42	GH1, GH3, GH5, GH9
Farm grasshopper flow-through	40605	544	225	GH1, GH3, GH5, GH12
Farm grasshopper gut	4994	2	0	
Local grasshopper feces	117331	735	255	GH1, GH3, GH5, GH9
Local grasshopper flow-through	39348	211	77	GH1, GH3, GH5, GH12
Local grasshopper gut	14303	9	0	
Grass compost	89620	541	176	GH1, GH3, GH5, GH9

The GH families found in samples were GH1, GH3, GH5, GH7, GH8, GH9 and GH12. GH families 6, 30 and 48 were not present in our sample. Gene counts were relatively lower for gut samples than flow-through samples due to eukaryotic contamination that resulted in complication in sample processing and sequencing. GH family 7 domain, which was present only in Dalaman Lake, normally found in fungi, the sequence might belong to fungi residing in lake or horizontal gene transfer from fungi to bacteria might be the reason (Hess et al., 2011).

3.2.3 Microbial diversity

Following de novo assembly, PhyloSift (Darling et al., 2014) was used to determine microbial diversity. The community in all samples was dominated by bacteria (82.6%-99.0%), and followed by eukaryota (1.1%-16.9) and archaea (1.1%-3.7%).Figure 3.2 shows distribution of domains for each sample.

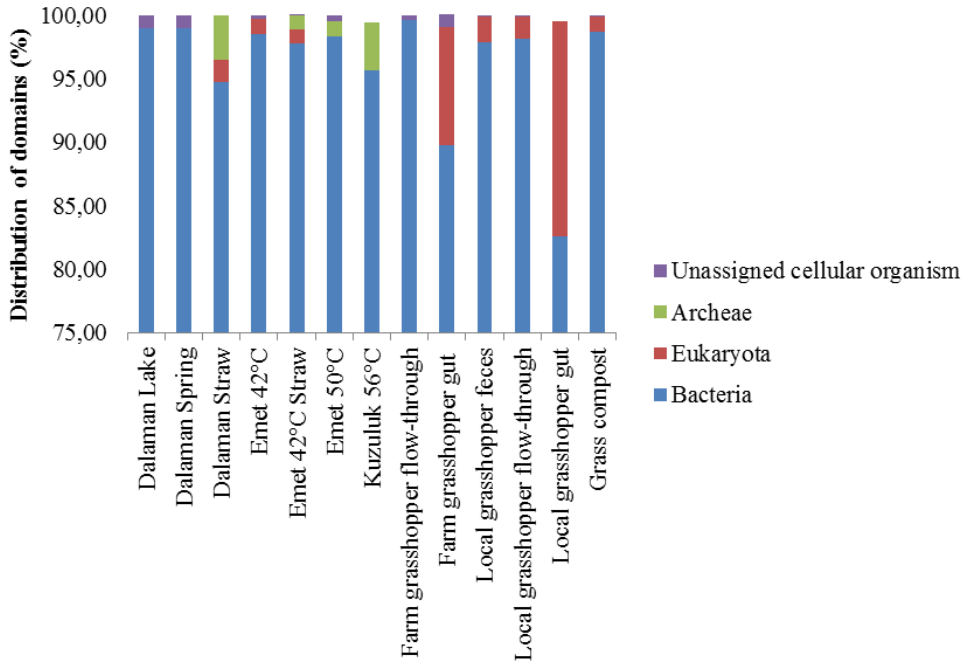

Figure 3.2: Distribution of domains in samples.

The comparison between communities of water samples and straw samples revealed increased abundance of archeae for both Dalaman and Emet 42°C samples. Addition of straw, as an organic matter, might serve as a methanogenic substrate and enhanced population of methanogens (Bengelsdorf et al., 2015; Conrad & Klose, 2006). Frequency of eukaryote was significantly higher in gut samples compared to flow-through samples for both farm and local grasshopper. This result revealed success of used method in elimination of eukaryotic contamination that might complicate sequencing and bioinformatic analysis.

Phyla analysis revealed overall dominance of bacterial phyla Proteobacteria, Firmicutes, Bacteroidetes, Actinobacteria and Aquificae. Community composition showed divergence between samples and dominant phylum was different for almost all samples. Figure 3.3 and Figure 3.4 shows distribution of phyla in water samples. Proteobacteria, which are important microbial groups in hot springs (Lau et al., 2009), were found to be dominant phyla in water samples excluding Kuzuluk 56°C

sample. In Dalaman samples, on average 27% of Proteobacteria were represented by sulfur-reducing bacteria including Campylobacter, Sulfurimonas, Sulfurovum, Desulfobacteraceae and Chlorobaculum as expected. Aquificales (60.80%) was the most dominant phylotype in Kuzuluk 56°C sample consistent with other studies on hot spring waters (Macur et al., 2004; Reysenbach et al., 2005; Vick et al., 2010). Comparison of composition of microbes adherent to straw to microbial population from water samples revealed increased abundance of Bacteroidetes. Bacteroidetes utilize high molecular mass organic matters in aquatic habitats and straw addition might lead enrichment of these organisms (O'Sullivan et al., 2006). Straw, serving as biofilm carrier for microorganisms, might create more stable environment and therefore enhance bacterial diversity compared to water samples.

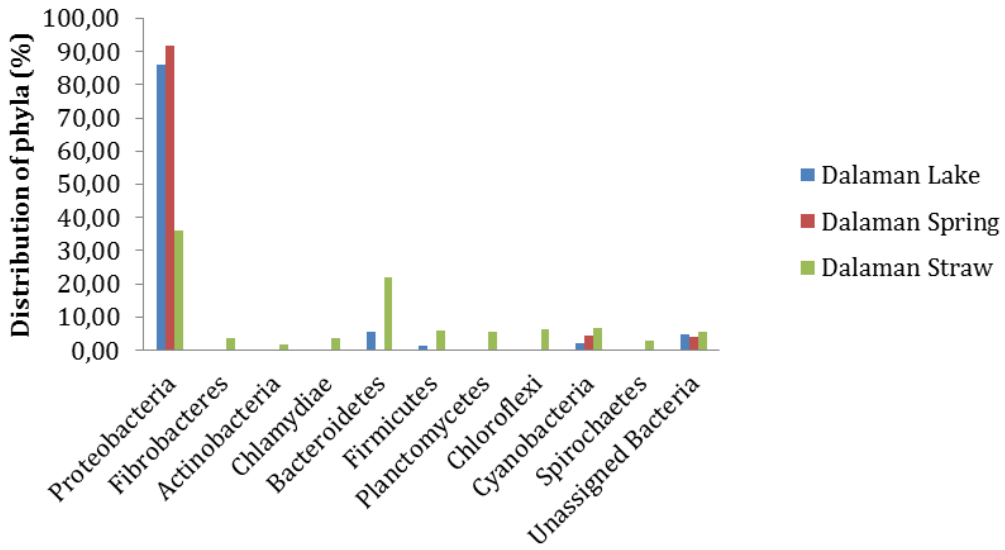

Figure 3.3: Distribution of most abundant phyla in Dalaman samples.

Firmicutes and Proteobacteria dominated microbial community in farm grasshopper and local grasshopper samples, respectively (Figure 3.5). The microbial divergence between two grasshoppers might be due to differences in feeding, or other environmental conditions. It is known that presence of Firmicutes enhance ability to use carbohydrates including cellulose and hemicellulose (Kopke et al., 2010). Thus, high frequency of Firmicutes was expected. However, high frequency of Proteobacteria, especially γ- proteobacteria (87%-97% of Proteobacteria), was unexpected because γ-proteobacteria do not play role in biomass degradation. γ-proteobacteria might involve in degradation of grasses since they are essential for some insects to utilize low nutrient food sources (Shi et al., 2013; Snyder et al.,

2010). The presence of Tenericutes was interesting since they are mostly pathogenic to their host, however, other features of Tenericutes including protecting against parasites were revealed (Garnier et al., 2001; Haselkorn et al., 2013).

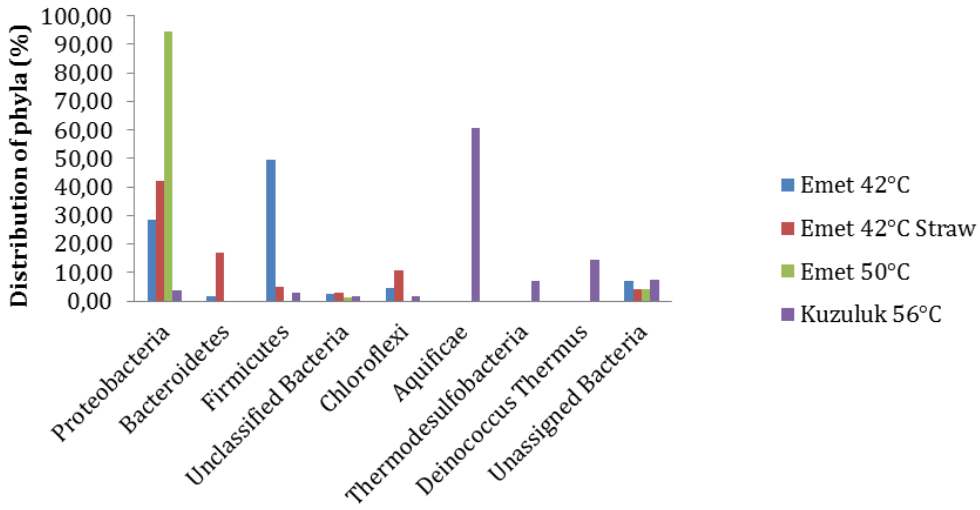

Figure 3.4: Distribution of most abundant phyla in Emet and Kuzuluk samples.

For the grass compost, the phylum Actinobacteria was the dominant group (55.05%), followed by Proteobacteria (25.32%), Bacteroidetes (16.17%), Firmicutes (1.51%) and unassigned bacteria (1.95%). Actinobacteria play an essential role in composting and degradation of recalcitrant compounds such as lignocellulose (Goodfellow & Williams, 1983; Steger et al., 2007).

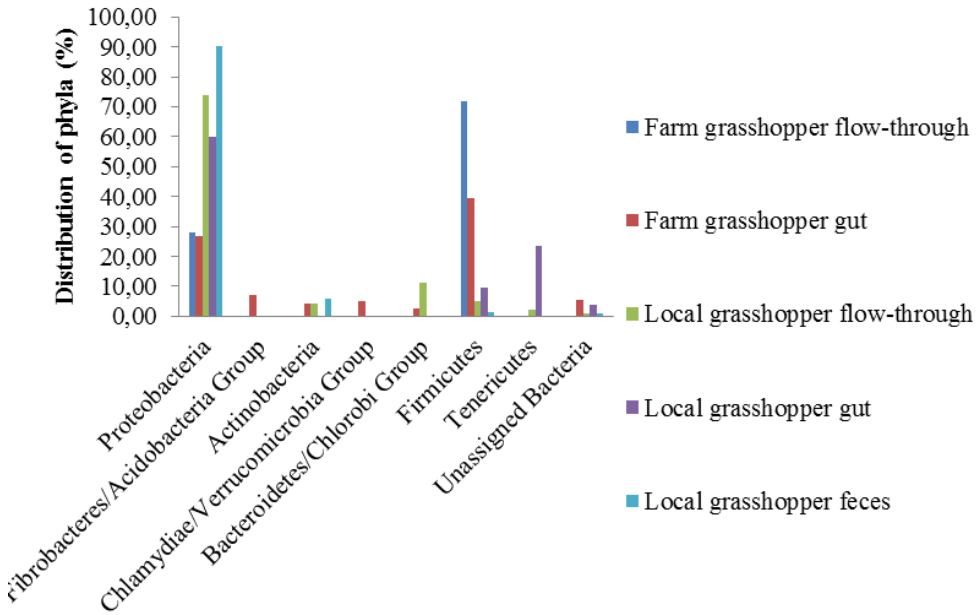

Figure 3.5: Distribution of most abundant phyla in grasshopper samples.

Overall, composition of microorganisms was determined and results indicated that selected environmental samples were good candidates for discovery of biomass degrading enzymes.

3.3 Cloning of Selected Cellulase Genes

Cloning of selected 63 genes involved amplification of genes from metagenomic DNA followed by inserting these genes into pET-22b(+) vector. Constructed plasmids were transformed into *E. coli* BL21 (DE3) and expressions carried out via IPTG induction and expression results were visualized by SDS-PAGE. Table C.1 shows success of each step for each gene. 54 of 63 genes were successfully amplified with PCR. Following transformation of 54 genes containing plasmids into *E. coli* BL21 (DE3), 49 transformants harbored insert gene in plasmid constructs and 22 of these tranformants were able to over-express these genes.

3.3.1 PCR amplification of selected genes

63 candidate genes, which were annotated as β-(1,4) endoglucanases, β-glucosidases, and exo-cellulase, were selected for cloning. The sequence homology–based enzyme discovery has tendency to identify candidate enzyme that are similar to known enzymes and limits discovery of novel enzymes. Also, it was shown that enzymes with high sequence identity are not more likely to be active than low sequence identity enzymes (Hess et al., 2011). For these reasons, candidate genes were selected with different sequence similarity to known enzymes (ranging between 56% and 100%) in order to discover enzymes possessing divergent physicochemical properties. Selected genes were given in Table 3.4, the average gene length was 1246 bp and gene length ranged between 564 bp and 2298 bp. Protein molecular weight prediction was carried out using "The Sequence Manipulation Suite: Protein Molecular Weight" (Stothard, 2000) and average protein size was determined to be 45.76 kilodaltons.

For amplification of selected genes, metagenomic DNAs were used as templates and reactions were performed using Q5® High-Fidelity DNA Polymerase (NEB, USA). Gene specific primers compatible with EZ-Fusion™ Cloning Kit, which contains 18 bp homolog region to pET-22b(+) plasmid, were designed and reaction was prepared

depending on the amplified gene. PCR products were visualized with gel electrophoresis.

Table 3.4: Profile of candidate genes that were selected for cloning.

Gene ID	Enzyme with highest hit	Length (bp)	Predicted protein size (kDa)
DALAMAN SPRING			
176339	Endoglucanase	984	36.33
2484	1,4-beta-D-glucan glucohydrolase	846	30.45
3947	Glycoside hydrolase family 3	1122	40.45
21962	Endoglucanase	1014	37.63
31506	Endoglucanase	564	19.80
38300	Endoglucanase	1215	46.05
50455	Endoglucanase	1188	44.93
EMET 42°C			
46938	Endoglucanase	1419	54.24
53527	Beta-glucosidase	1347	51.82
92869	Beta-glucosidase	1347	52.38
97287	Glycoside hydrolase family 5	696	25.54
100149	Glycoside hydrolase family 5	1029	37.97
111095	Cellulase	714	24.68
EMET 42°C STRAW			
199	Endoglucanase	675	26.11
30829	Glycoside hydrolase family 5	1230	48.17
31013	Endoglucanase	1431	52.86
46252	Beta-glucosidase	1059	38.78
51689	Beta-glucosidase	1293	47.27
72422	Endoglucanase	1362	49.71
KUZULUK 56°C			
11941	Beta-glucosidase	1296	48.59
18880	Beta-glucosidase	1326	51.29
FARM GRASSHOPPER FLOW-THROUGH			
2599	Beta-glucosidase	2226	79.65
2864	Beta-glucosidase	945	36.36
5020	6-phospho-beta-glucosidase	936	36.10
5252	6-phospho-beta-glucosidase	1140	40.24
7430	6-phospho-beta-glucosidase	1431	53.65
7761	6-phospho-beta-glucosidase	1146	42.96
7918	Aryl-phospho-beta-D-glucosidase	1437	55.22
8125	6-phospho-beta-glucosidase	1431	53.65
9683	6-phospho-beta-glucosidase	942	36.44
12130	6-phospho-beta-glucosidase	1053	40.21
12733	Beta-glucosidase	1260	46.73
12836	β-D-glucoside glucohydrolase	2298	83.47
14933	Beta-glucosidase	1998	76.40
16307	6-phospho-beta-glucosidase	1440	55.46

Table 3.4: Profile of candidate genes that were selected for cloning.

Gene ID	Enzyme with highest hit	Length (bp)	Predicted protein size (kDa)
20495	6-phospho-beta-glucosidase	1473	56.34
26764	Beta-glucosidase	1437	54.58
27669	6-phospho-beta-glucosidase	1422	53.69
28464	Beta-glucosidase	1392	53.69
28541	6-phospho-beta-glucosidase	1437	52.83
29016	Glycoside hydrolase family 1	1365	51.31
29062	Beta-glucosidase	1500	56.71
29118	Beta-glucosidase	1437	55.13
29580	6-phospho-beta-glucosidase	1050	40.10
32123	6-phospho-beta-glucosidase	1305	48.96
36243	Aryl-phospho-beta-D-glucosidase	1443	54.21
36838	6-phospho-beta-glucosidase	1410	53.31
37597	Aryl-phospho-beta-D-glucosidase	1410	53.37
38877	6-phospho-beta-glucosidase	1041	39.81
38878	6-phospho-beta-glucosidase	1365	51.24
41502	6-phospho-beta-glucosidase	1401	53.30
42073	6-phospho-beta-glucosidase	1407	52.15
46845	6-phospho-beta-glucosidase	1380	52.43
203416	6-phospho-beta-glucosidase	1368	52.82
LOCAL GRASSHOPPER FECES			
4190	6-phospho-beta-glucosidase	900	34.66
6745	6-phospho-beta-glucosidase	1080	41.59
18167	Cellulase	993	36.16
81997	6-phospho-beta-glucosidase	1410	54.08
102447	6-phospho-beta-glucosidase	1335	49.59
GRASS COMPOST			
960	Cellulase	1083	39.90
66586	Endoglucanase	963	35.27
70910	Endoglucanase	969	34.40
72671	Aryl-phospho-beta-D-glucosidase	954	34.54

Figure 3.6 shows 22 of these genes, which were successfully expressed, as representative results. While 54 genes were successfully amplified with PCR, 9 genes were not replicated. Shearing or breaking down of metagenomic DNA during extraction at selected gene region might prevent binding of primers to DNA.

3.3.2 Analysis of recombinant pET-22b(+) by colony PCR

To determine the presence or absence of insert DNA in plasmid constructs, transformants were analyzed with colony PCR. Selected single colonies were lysed with high temperature and lysates were used as template DNA for PCR. PCR conditions and components of reaction mixture were determined depending on the

target gene and designed primers. In total 49 transformant contained the DNA fragment of interest and Figure 3.7 shows 22 of these genes. Transformation of vectors without insert DNA due to incomplete cleavage of pET-22b (+) with restriction enzymes *Xho*I and *Nde*I or inefficient ligation might result in transformants lacking desired gene. In addition, PCR conditions such as annealing temperature might be problematic for amplification of target gene ("Troubleshooting Guide for Cloning," n.d.).

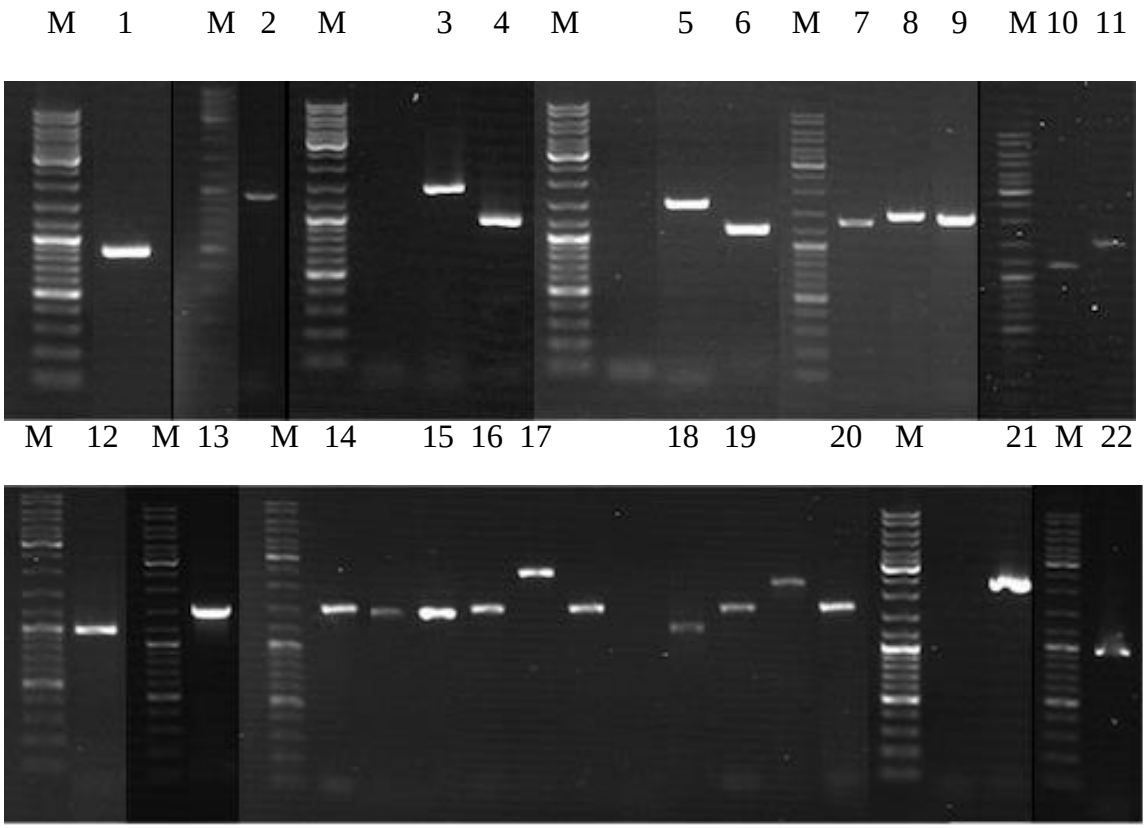

Figure 3.6: Agarose gel electrophoresis of PCR products for selected genes. M: DNA marker (O'GeneRuler DNA Ladder Mix), 1: 2484, 2: 6745, 3: 7430, 4: 9683, 5: 16307, 6: 29580, 7: 28464, 8: 29062, 9: 36838, 10: 46252, 11: 81997, 12: 70910, 13: 92869, 14: 20495, 15: 36243, 16: 8125, 17: 12836, 18: 7761, 19: 28541, 20: 27669, 21:2599, 22: 66586.

3.3.3 Expression of introduced genes

Expression of introduced genes was induced by addition of 1 mM IPTG to the cells at the exponential phase (OD_{600} of 0.5- 1.0). For each sample, cultures with and without IPTG (uninduced control) were incubated for 2 hours at 37°C. Following incubation, 1 ml of the cells aliquot were harvested by centrifugation. Cells were suspended in SDS-PAGE sample loading dye and visualized with SDS-PAGE. Figure 3.8 shows SDS-PAGE results for 22 genes and “i” stands for IPTG-induced.

22 of 49 positive transformant were able to express introduced genes. Toxicity of introduced protein or secondary structure of mRNA that prevent interaction with synthesis machinery might be the reason for unsuccessful expression. Also, problems with host strain due to codon usage might prevent production of target protein (Samuelson, 2011).

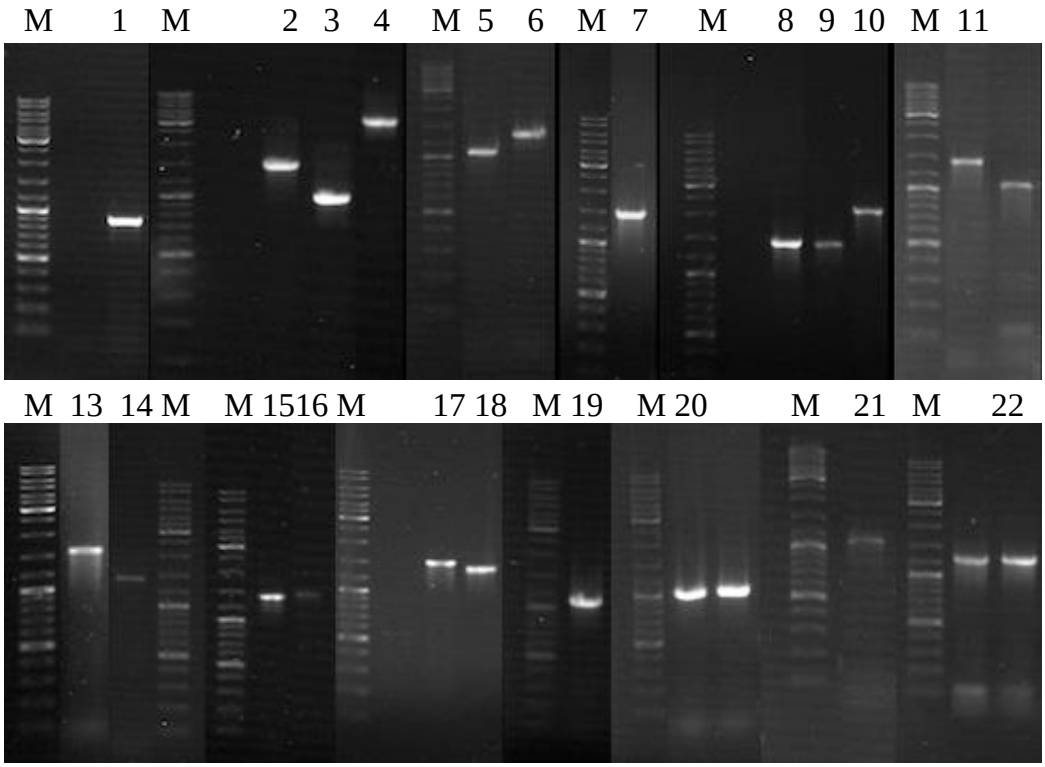

Figure 3.7: Agarose gel electrophoresis of colony PCR for selected genes. M: DNA marker (O'GeneRuler DNA Ladder Mix), 1: 2484, 2: 2599, 3: 7430, 4: 9683, 5:6745, 6: 81997, 7: 7761, 8: 12836, 9: 20495, 10: 28541, 11: 16307, 12: 29580, 13: 27669, 14: 27669, 15: 36838 16: 28464, 17: 29062, 18: 36243, 19: 46252, 20: 66586, 21: 70910, 22: 92869.

3.4 Biochemical Activity of the Candidate Cellulases

To evaluate the biochemical activity of 22 candidate proteins, which were annotated as β-glucosidase, endoglucanase and 1,4-beta-D-glucan glucohydrolase (exoglucanase), methods described by Parry et al. (2001) and Bailey et al. (1992) were used. Following induction of introduced genes with IPTG, transformants were collected with centrifugation. Pellets were suspended in sonication buffer and sonication was performed to lyse cells. After removal of cell debris, crude enzymes were used for screening active cellulolytic enzymes.

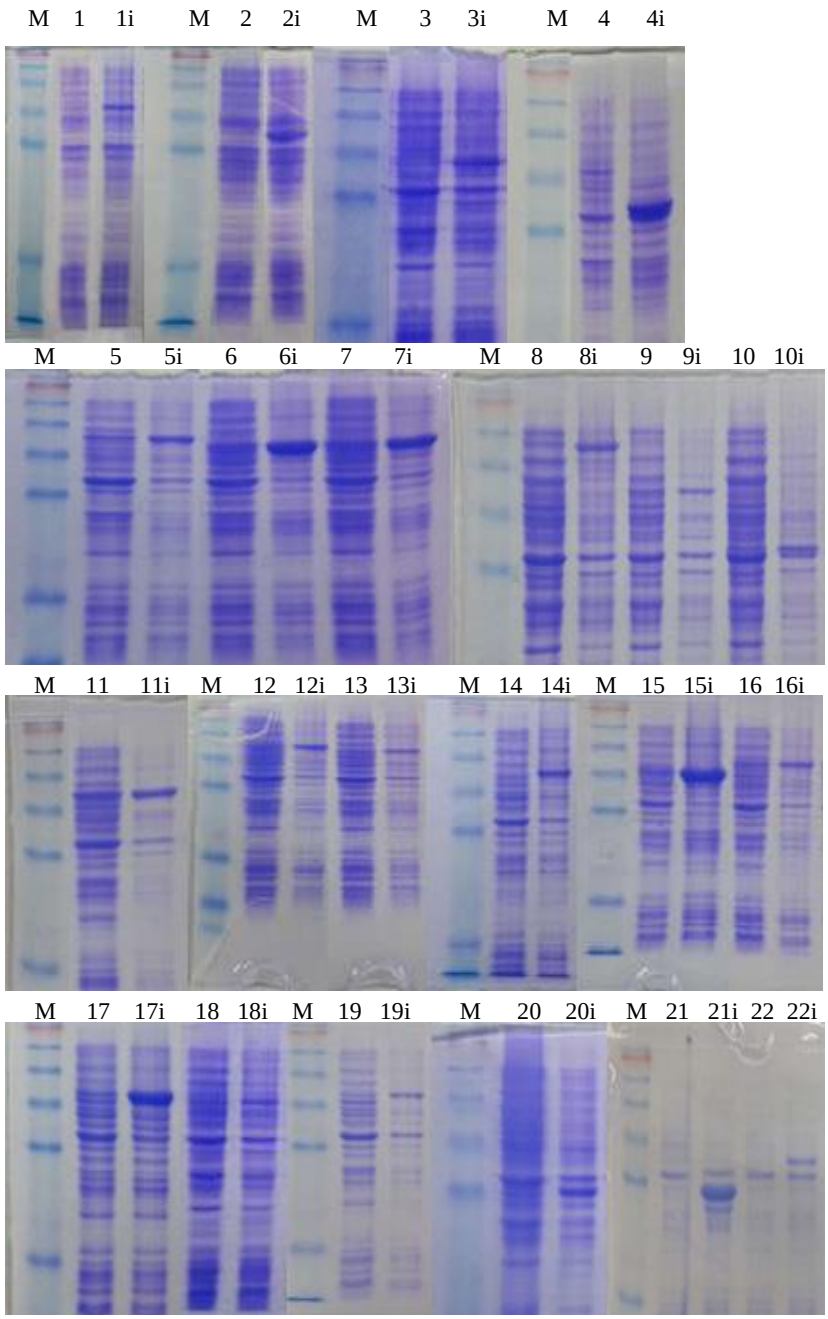

Figure 3.8: SDS-PAGE analysis of expression of introduced genes. 1: 81997, 2: 6745, 3:46252, 4: 66586, 5: 36243, 6: 28541, 7: 27669, 8: 2599, 9: 7430, 10: 9683, 11: 7761, 12: 12836, 13: 20495, 14: 16307, 15: 28464, 16: 29062, 17: 26764, 18: 92869, 19: 36838, 20: 70910, 21: 2484, 22: 29580.

Analysis of candidate two endoglucanase and one 1,4-beta-D-glucan glucohydrolase (exoglucanase) with dinitrosalicylic colorimetric method revealed presence of one active endoglucanase enzyme (Figure 3.9). The only active cellulolytic enzyme, 70910, was 969 bp long and its molecular weight was estimated 34.40 kDa. The BLAST search revealed 100% sequence identity with non-redundant protein

endoglucanase of *Rhodanobacter fulvus* (NCBI Reference Sequence: WP_007079969.1). The source of biochemically active enzyme, 70910, was grass compost.

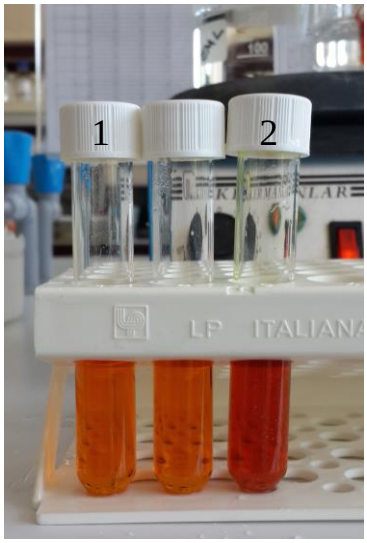

Figure 3.9: Result of DNS colorimetric method for 70910. 1: Negative control (without enzyme), 2: crude enzyme 70910.

To determine specific activity of enzyme 70910, protein purification was performed using HisPur™ Cobalt Purification Kit (Thermo Fisher). Calculated protein concentration was 0.52 mg/ml. After purification, SDS-PAGE revealed presence of more than one protein band in elution solution. The activity was determined to be 0.23 U/ml/min and specific activity was calculated to be 0.44 units per mg of protein. However, this calculation might not reflect the true values since calculated protein amount includes proteins other than target protein. Further experiments should be done for isolation of more purified enzyme for activity calculation.

To test activity of 19 β-glucosidases, pNPG was used as substrate and yellow color released after hydrolysis was measured spectrometrically. Analysis revealed no active β-glucosidase enzyme.

Overall, ratio of active protein was lower than expected. False-positive prediction of carbohydrate active genes, misfolding of recombinant protein or inappropriate reaction conditions might be responsible for high number of inactive proteins (Hess et al., 2011).

4. CONCLUSIONS

The aim of the study was discovery of novel cellulase genes. For this purpose, gut microbiome of grasshoppers, hot spring waters, and grass compost were exploited. One biochemically active cellulose-hydrolyzing enzyme was discovered from grass compost sample. Number of active enzymes was lower than expected. The problems due to misfolding, insufficient expression of recombinant protein or wrong annotation of genes might be resulted in inactive enzymes. Use of other expression mechanisms that involves different host or different vector can resolve these problems. Also, improvement of sequencing process for better library generation might result in more reliable data for seek of cellulase candidate. In discovery of microbial diversity, presence of unassigned organisms revealed out limited information in phylogenetic and it has shown there is more to discover. Further studies are required including characterization of found cellulolytic enzyme and exploitation of this enzyme for bioreactor. Also, enormous data generated from samples can be used for discovery of other enzymes such as lipases, proteinases,and etc.

REFERENCES

Adlakha, N., Rajagopal, R., Kumar, S., Reddy, V. S., & Yazdani, S. S. (2011). Synthesis and characterization of chimeric proteins based on cellulase and xylanase from an insect gut bacterium. *Applied Environmental Microbiology, 77*(14), 4859-4866.

Alanis, A. J. (2005). Resistance to antibiotics: are we in the post-antibiotic era? *Archives of medical research, 36*(6), 697-705.

Albertsen, M. (2013, November 18). Genomics and metagenomics. [PowerPoint slides] Retrieved from http://www.slideshare.net/MadsAlbertsen/20131029-albertsen-genomics-metagenomics

Bailey, M. J., Biely, P., & Poutanen, K. (1992). Interlaboratory testing of methods for assay of xylanase activity. *Journal of biotechnology, 23*(3), 257-270.

Bashir, Y., Pradeep Singh, S., & Kumar Konwar, B. (2014). Metagenomics: An Application Based Perspective. *Chinese Journal of Biology, 2014*.

Bashir, Z., Kondapalli, V. K., Adlakha, N., Sharma, A., Bhatnagar, R. K., Chandel, G., & Yazdani, S. S. (2013). Diversity and functional significance of cellulolytic microbes living in termite, pill-bug and stem-borer guts. *International Journal of Scientific Reports, 3*, 2558.

Batraa, N., Bhatiaa, S., Behala, A., Singhb, J., & Joshic, A. (2013). Metagenomic Research: Methods and Ecological Applications.

Béguin, P., & Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews, 13*(1), 25-58.

Bella, T. H., Greerb, C. W., & Yergeaub, E. (2012). Metagenomics Potential for Bioremediation.

Bengelsdorf, F. R., Gabris, C., Michel, L., Zak, M., & Kazda, M. (2015). Syntrophic microbial communities on straw as biofilm carrier increase the methane yield of a biowaste-digesting biogas reactor.

Berner, R. A. (2003). The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature, 426*(6964), 323-326.

Bhat, M. K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnology Advances, 18*(5), 355-383.

Bhat, M. K., & Bhat, S. (1997). Cellulose degrading enzymes and their potential industrial applications. *Biotechnology Advances, 15*(3), 583-620.

Bisaria, V. S., & Ghose, T. K. (1981). Biodegradation of cellulosic materials: substrates, microorganisms, enzymes and products. *Enzyme and microbial technology, 3*(2), 90-104.

Buchert, J., Oksanen, T., Pere, J., Siika-Aho, M., Suurnäkki, A., & Viikari, L. (1998). Applications of Trichoderma reesei enzymes in the pulp and paper industry. *Trichoderma and Gliocladium, 2*, 343-363.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research, 37*(Database issue), D233-238.

Chauve, M., Mathis, H., Huc, D., Casanave, D., Monot, F., & Ferreira, N. L. (2010). Comparative kinetic analysis of two fungal β-glucosidases. *Biotechnology and Biofuels, 3*(1), 3.

Conrad, R., & Klose, M. (2006). Dynamics of the methanogenic archaeal community in anoxic rice soil upon addition of straw. *European journal of soil science, 57*(4), 476-484.

Cortez, J. M., Ellis, J., & Bishop, D. P. (2001). Cellulase finishing of woven, cotton fabrics in jet and winch machines. *Journal of biotechnology, 89*(2-3), 239-245.

Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordonez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in microbiology, 6*, 672.

Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in microbiology, 6*.

Coughlan, M. P., Wood, T. M., Montenecourt, B. S., & Mendels, M. (1985). Cellulases: production, properties and applications. *Biochemical Society Transactions, 13*, 405-406.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R., & Wittwer, P. (2005). Metagenomic gene discovery: past, present and future. *Trends in Biotechnology, 23*(6), 321-329.

Culligan, E. P., Sleator, R. D., Marchesi, J. R., & Hill, C. (2013). Functional environmental screening of a metagenomic library identifies stlA; a unique salt tolerance locus from the human gut microbiome. *PLoS One, 8*(12), e82985.

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A. t., Bik, H. M., & Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ, 2*, e243.

Duan, C. J., & Feng, J. X. (2010). Mining metagenomes for novel cellulase genes. *Biotechnology Letters, 32*(12), 1765-1775.

Entcheva, P., Liebl, W., Johann, A., Hartsch, T., & Streit, W. R. (2001). Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied Environmental Microbiology, 67*(1), 89-99.

Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., Golyshin, P. N. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology, 7*(12), 1996-2010.

Ferrer, M., Martínez-Abarca, F., & Golyshin, P. N. (2005). Mining genomes and 'metagenomes' for novel catalysts. *Current Opinion in Biotechnology, 16*(6), 588-593.

Garnier, M., Foissac, X., Gaurivaud, P., Laigret, F., Renaudin, J., Saillard, C., & Bové, J. M. (2001). Mycoplasmas, plants, insect vectors: a matrimonial triangle. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie, 324*(10), 923-928.

Golan, A. E. (2011). *Cellulase: Types and Action, Mechanism, and Uses*: Nova Science Publishers.

Goodfellow, M., & Williams, S. T. (1983). Ecology of actinomycetes. *Annual Reviews in Microbiology, 37*(1), 189-216.

Guazzaroni, M. E., Morgante, V., Mirete, S., & Gonzalez-Pastor, J. E. (2013). Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environmental Microbiology, 15*(4), 1088-1102.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology, 166*(4), 557-580.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews, 68*(4), 669-685.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology, 5*(10), R245-249.

Haselkorn, T. S., Cockburn, S. N., Hamilton, P. T., Perlman, S. J., & Jaenike, J. (2013). Infectious adaptation: potential host range of a defensive endosymbiont in Drosophila. *Evolution, 67*(4), 934-945.

Henrissat, B., & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology, 7*(5), 637-644.

Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., & Rubin, E. M. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science, 331*(6016), 463-467.

Hon, D. S. (1994). Cellulose: a random walk along its historical path. *Cellulose, 1*(1), 1-25.

Jayasuriya, H., Herath, K. B., Zhang, C., Zink, D. L., Basilio, A., Genilloud, O., & Singh, S. B. (2007). Isolation and structure of platencin: a FabH and FabF dual inhibitor with potent broad-spectrum antibiotic activity. *Angewandte Chemie International Edition, 46*(25), 4684-4688.

Joern, A. (1983). Host plant utilization by grasshoppers (Orthoptera: Acrididae) from a sandhills prairie. *Journal of Range Management*, 793-797.

Juturu, V., & Wu, J. C. (2014). Microbial cellulases: Engineering, production and applications. *Renewable and Sustainable Energy Reviews, 33*, 188-203.

Kaur, G., Kumar, S., & Satyanarayana, T. (2004). Production, characterization and application of a thermostable polygalacturonase of a thermophilic mould Sporotrichum thermophile Apinis. *Bioresource Technology, 94*(3), 239-243.

Kimura, S., & Itoh, T. (2001). Occurence of High Crystalline Cellulose in the Most Primitive Tunicate, Appendicularian. *Progress in Biotechnology, 18*, 121-125.

Koffi, Y. G., Konan, K. H., Kouadio, E. J. P., Dabonné, S., & Kouamé, L. P. (2012). Purification and biochemical characterization of beta-glucosidase from cockroach Periplaneta americana. *Journal of Animal and Plant Sciences, 13*, 1747-1757.

Koo, Y.-M. (2001). Pilot-scale production of cellulase using Trichoderma reesei Rut C-30 in fed-batch mode. *Journal of Microbiology and Biotechnology, 11*(2), 229-233.

Kopke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., & Durre, P. (2010). Clostridium ljungdahlii represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences, 107*(29), 13087-13092.

Krahe, M., Antranikian, G., & Märkl, H. (1996). Fermentation of extremophilic microorganisms. *FEMS Microbiology Reviews, 18*(2- 3), 271-285.

Kris-Etherton, P. M., Hecker, K. D., Bonanome, A., Coval, S. M., Binkoski, A. E., Hilpert, K. F., & Etherton, T. D. (2002). Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. *American Journal of Medicine, 113*(9), 71-88.

Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of Industrial Microbiology & Biotechnology, 35*(5), 377-391.

Lau, M. C. Y., Aitchison, J. C., & Pointing, S. B. (2009). Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles, 13*(1), 139-149.

Lee, Y.-J., Kim, B.-K., Lee, B.-H., Jo, K.-I., Lee, N.-K., Chung, C.-H., . . . Lee, J.-W. (2008). Purification and characterization of cellulase produced by Bacillus amyoliquefaciens DL-3 utilizing rice hull. *Bioresource Technology, 99*(2), 378-386.

Lima, A. O. S., Quecine, M. C., Fungaro, M. H. P., Andreote, F. D., Maccheroni Jr, W., Araújo, W. L., . . . Azevedo, J. L. (2005). Molecular characterization of a β-1, 4-endoglucanase from an endophytic Bacillus pumilus strain. *Applied Microbiology and Biotechnology, 68*(1), 57-65.

Lorenz, P., & Eck, J. (2005). Metagenomics and industrial applications. *Nature Reviews Microbiology, 3*(6), 510-516.

Lorenz, P., Liebeton, K., Niehaus, F., & Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Current Opinion in Biotechnology, 13*(6), 572-577.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and molecular biology reviews, 66*(3), 506-577, table of contents.

Macrae, R., Robinson, R. K., & Sadler, M. J. (1993). Encyclopaedia of food science, food technology, and nutrition.

Macur, R. E., Langner, H. W., Kocar, B. D., & Inskeep, W. P. (2004). Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic- rich, acid- sulphate- chloride geothermal spring. *Geobiology, 2*(3), 163-177.

Maheshwari, P., Mishra, B. K., Kumar, S., & Doshi, A. (2009). Cellulolytic enzyme profile of different mushrooms species on maize residue. *Environment and Ecology, 27*(2), 689-691.

Martins, L. F., Kolling, D., Camassola, M., Dillon, A. J. P., & Ramos, L. P. (2008). Comparison of Penicillium echinulatum and Trichoderma reesei cellulases in relation to their activity against various cellulosic substrates. *Bioresource Technology, 99*(5), 1417-1424.

Matteotti, C., Thonart, P., Francis, F., Haubruge, E., Destain, J., Brasseur, C., . . . Vandenbol, M. (2011). New glucosidase activities identified by functional screening of a genomic DNA library from the gut microbiota of the termite Reticulitermes santonensis. *Microbiological Research, 166*(8), 629-642.

Miettinen-Oinonen, A., & Suominen, P. (2002). Enhanced production of Trichoderma reesei endoglucanases and use of the new cellulase preparations in producing the stonewashed effect on denim fabric. *Applied Environmental Microbiology, 68*(8), 3956-3964.

Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., & Ladisch, M. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology, 96*(6), 673-686.

Nobles, D. R., Romanovicz, D. K., & Brown, R. M. (2001). Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? *Plant physiology, 127*(2), 529-542.

O'Sullivan, L. A., Rinna, J., Humphreys, G., Weightman, A. J., & Fry, J. C. (2006). Culturable phylogenetic diversity of the phylum 'Bacteroidetes' from river epilithon and coastal water and description of novel members of the family Flavobacteriaceae: Epilithonimonas tenax gen. nov., sp. nov. and Persicivirga xylanidelens gen. nov., sp. nov. *International journal of systematic and evolutionary microbiology, 56*(1), 169-180.

Parry, N. J., Beever, D. E., Owen, E., Vandenberghe, I., Van Beeumen, J., & Bhat, M. K. (2001). Biochemical characterization and mechanism of action of a thermostable beta-glucosidase purified from Thermoascus aurantiacus. *Biochemical Journal, 353*(Pt 1), 117-127.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proceedings of the National Academy of Sciences, 99*(22), 14002-14007.

Reysenbach, A.-L., Banta, A., Civello, S., Daly, J., Mitchel, K., Lalonde, S., . . . Takacs-Vesbach, C. (2005). Aquificales in Yellowstone National Park. *Geothermal biology and geochemistry in Yellowstone National Park, 1*, 129-142.

Rooks, D. J., McDonald, J. E., & McCarthy, A. J. (2012). Metagenomic approaches to the discovery of cellulases. *Methods in Enzymology, 510*, 375-394.

Sabree, Z. L., Rondon, M. R., & Handelsman, J. (2009). Metagenomics. *Encyclopedia of Microbiology*, 622-632.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning* (Vol. 2): Cold spring harbor laboratory press New York.

Samuelson, J. C. (2011). Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations *Heterologous Gene Expression in E. coli* (pp. 195-209): Springer.

Schröder, C., Elleuche, S., Blank, S., & Antranikian, G. (2014). Characterization of a heat-active archaeal β-glucosidase from a hydrothermal spring metagenome. *Enzyme and microbial technology, 57*, 48-54.

Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiol Biotechnology, 56*(5-6), 634-649.

Scientific, T. F. (2008). NanoDrop 1000 Spectrophotometer V3. 7 User's Manual. *Thermo Fisher Scientific*, 1-105.

Sharma, R., Ranjan, R., Kapardar, R. K., & Grover, A. (2005). Unculturable'bacterial diversity: an untapped resource. *Current Science, 89*(1), 72-77.

Shedova, E. N., Berezina, O. V., Lunina, N. A., Zverlov, V. V., Schwartz, W. H., & Velikodvorskaia, G. A. (2009). [Cloning and characterization of a large metagenomic DNA fragment containing glycosyl-hydrolase genes]. *Molekuliarnaia genetika, mikrobiologiia i virusologiia.*(1), 11-15.

Shi, W., Xie, S., Chen, X., Sun, S., Zhou, X., Liu, L., . . . Yuan, J. S. (2013). Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genetics, 9*(1), e1003131.

Shoham, Y., Lamed, R., & Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in microbiology, 7*(7), 275-281.

Snyder, A. K., Deberry, J. W., Runyen-Janecky, L., & Rio, R. V. M. (2010). Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proceedings of the Royal Society of London B: Biological Sciences, 277*(1692), 2389-2397.

Steffan, R. J., Goksoyr, J., Bej, A. K., & Atlas, R. M. (1988). Recovery of DNA from soils and sediments. *Applied Environmental Microbiology, 54*(12), 2908-2915.

Steger, K., Sjögren, Å., Jarvis, Å., Jansson, J. K., & Sundh, I. (2007). Development of compost maturity and Actinobacteria populations during full- scale composting of organic household waste. *Journal of applied microbiology, 103*(2), 487-498.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques, 28*(6), 1102, 1104.

Suenaga, H., Ohnuki, T., & Miyazaki, K. (2007). Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology, 9*(9), 2289-2297.

Sukharnikov, L. O., Cantwell, B. J., Podar, M., & Zhulin, I. B. (2011). Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends in biotechnology, 29*(10), 473-479.

Tartar, A., Wheeler, M. M., Zhou, X., Coy, M. R., Boucias, D. G., & Scharf, M. E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes. *Biotechnology Biofuels, 2*, 25.

Tebbe, C. C., & Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied Environmental Microbiology, 59*(8), 2657-2665.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation, 2*(1), 3.

Troubleshooting Guide for Cloning. (n.d.). Retrieved October 15, 2015, from https://http://www.neb.com/tools-and-resources/troubleshooting-guides/troubleshooting-guide-for-cloning

Uchiyama, T., & Miyazaki, K. (2009). Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology, 20*(6), 616-622.

Ufarte, L., Laville, E., Duquesne, S., & Potocki-Veronese, G. (2015). Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnology Advances*.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research, 40*(15), e115.

Valášková, V., & Baldrian, P. (2006). Degradation of cellulose and hemicelluloses by the brown rot fungus Piptoporus betulinus–production of extracellular enzymes and characterization of the major cellulases. *Microbiology, 152*(12), 3613-3622.

Vick, T. J., Dodsworth, J. A., Costa, K. C., Shock, E. L., & Hedlund, B. P. (2010). Microbiology and geochemistry of Little Hot Creek, a hot spring environment in the Long Valley Caldera. *Geobiology, 8*(2), 140-154.

Weimer, P. J., Russell, J. B., & Muck, R. E. (2009). Lessons from the cow: what the ruminant animal can teach us about consolidated bioprocessing of cellulosic biomass. *Bioresource Technology, 100*(21), 5323-5331.

Wilson, D. B. (2011). Microbial diversity of cellulose hydrolysis. *Current Opinion in Microbiology*, *14*(3), 259-263. doi: 10.1016/j.mib.2011.04.004

Wüstenberg, T. (2014). *Cellulose and Cellulose Derivatives in the Food Industry: Fundamentals and Applications*. Weinheim, Germany: John Wiley & Sons.

Wyman, C. E. (1994). Ethanol from lignocellulosic biomass: technology, economics, and opportunities. *Bioresource Technology, 50*(1), 3-15.

Yandell, M., & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature Review Genetics*, *13*(5), 329-342.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research, 18*(5), 821-829.

Zhou, J., Wang, Y. H., Chu, J., Zhuang, Y. P., Zhang, S. L., & Yin, P. (2008). Identification and purification of the main components of cellulases from a mutant strain of Trichoderma viride T 100-14. *Bioresource Technology, 99*(15), 6826-6833.

Zhou, S., & Ingram, L. O. (2000). Synergistic Hydrolysis of Carboxymethyl Cellulose and Acid-Swollen Cellulose by Two Endoglucanases (CelZ and CelY) fromErwinia chrysanthemi. *Journal of Bacteriology, 182*(20), 5676-5682.

Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, *38*(12), e132.

APPENDICES

Appendix A: Map of pET-22b(+) vector

Appendix B: Primers that were used in the study

Appendix C: Success of each step for each candidate gene

Appendix A

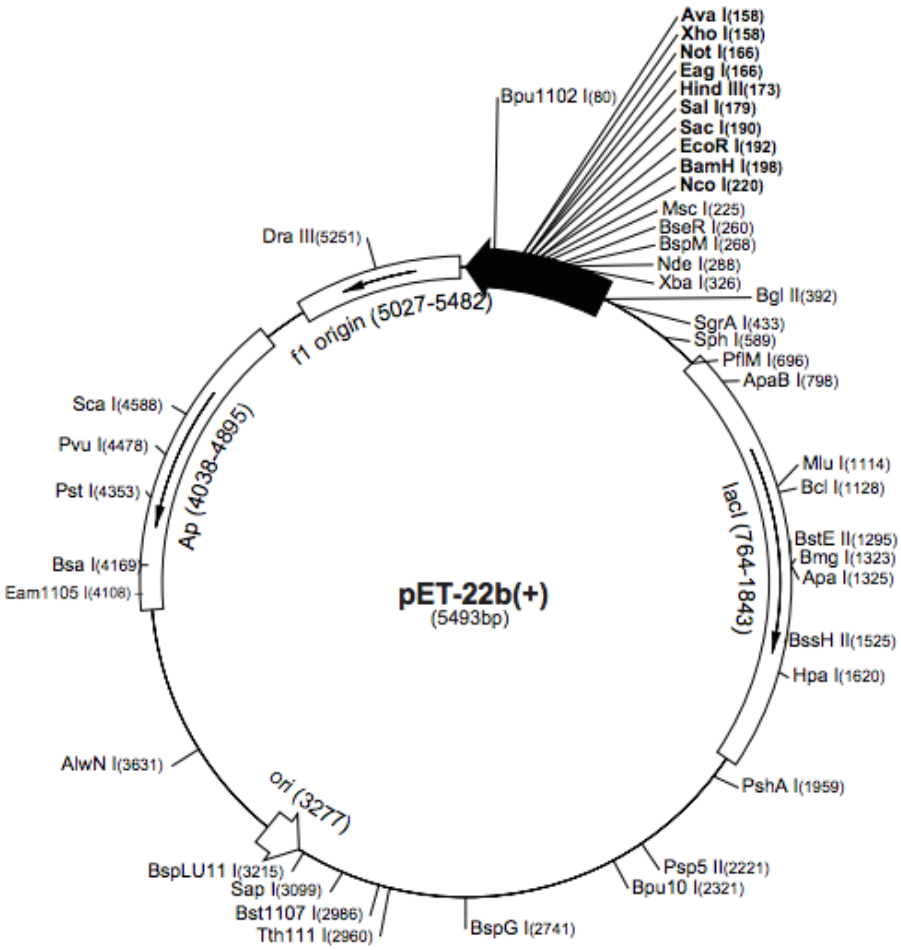

Figure A.1: Map of pET-22b(+) vector.

Appendix B

Table B.1: Primers that were used in the study.

Primers	Sequence (5'-3')
50455F	*GAAGGAGATATA*CATATGCGAGAAGAAATCAGCAGGA
50455R	*GTGGTGGTGGTG*CTCGAGGCCATTTTTTGATTTGAAAATAATCCA
2484F	*GAAGGAGATATA*CATATGGCCAAGGAACTGCGC
2484R	*GTGGTGGTGGTG*CTCGAGGATGCTCTCGGCGGCGTC
2599F	*GAAGGAGATATA*CATATGACCCAACAACTCACGCA
2599R	*GTGGTGGTGGTG*CTCGAGATTGTAACGGTCTCCACCAAGA
7430/5252F	*GAAGGAGATATA*CATATGACTAAGAATGTATTTCCAGA
7430R	*GTGGTGGTGGTG*CTCGAGAAGATTTACACCGTTAGATGCGA
9683F	*GAAGGAGATATA*CATATGAAGTACTGGTTGACGTT
9683R	*GTGGTGGTGGTG*CTCGAGCCCCAAATCCTCTCCATTTGAG
5252R	*GTGGTGGTGGTG*CTCGAGACAACGTCTTCCGCACCCNN
5020F	*GAAGGAGATATA*CATATGACTTTAAGAAAAGATTTCCTATGG
5020R	*GTGGTGGTGGTG*CTCGAGTATAATAACTAAAGCCCCGA
46845F	*GAAGGAGATATA*CATATGGGGCGGAGCGCAACAG
46845R	*GTGGTGGTGGTG*CTCGAGCTTGTTTGTACCATCAAATGA
2864F	*GAAGGAGATATA*CATATGTTTTATGGGGCGGAGCAG
2864R	*GTGGTGGTGGTG*CTCGAGTGAATAATAGTAAGAAATCGCT
12130F	*GAAGGAGATATA*CATATGCGGACTTATACATTTCCCCA
12130R	*GTGGTGGTGGTG*CTCGAGTCAGAATATCGTAGATGCCTTTTTCG
16307F	*GAAGGAGATATA*CATATGAAGAAAAATAAGAAGCAATTTCCT
16307R	*GTGGTGGTGGTG*CTCGAGATCGAGGACTTCTCCGTTAGA
29580F	*GAAGGAGATATA*CATATGAAGGTTCAATTTCCAAATGA
29580R	*GTGGTGGTGGTG*CTCGAGATATATCATAAATACCTCTTTCA
32123F	*GAAGGAGATATA*CATATGGGACTTAAGTTGAATGATGA
32123R	*GTGGTGGTGGTG*CTCGAGGTCAATTGGTCCCCATGACA
38877F	*GAAGGAGATATA*CATATGAAAAGTTATTTTCCAAAAGA
38877R	*GTGGTGGTGGTG*CTCGAGTAGAGTTGATTCAATGCTGTTCGC
38878F	*GAAGGAGATATA*CATATGTTGGAAGGAGGAAAAGGATT
38878R	*GTGGTGGTGGTG*CTCGAGCTTGCCGCCTTTCAAAAATGA
41502F	*GAAGGAGATATA*CATATGAAACAGACATTTCCAACT
41502R	*GTGGTGGTGGTG*CTCGAGCTCCTCCAACGAAGGGATCG
42073F	*GAAGGAGATATA*CATATGAGTCGAGGAGGAATGAA
42073R	*GTGGTGGTGGTG*CTCGAGAAACGTCAAGGTTCCTTTGCC
12733F	*GAAGGAGATATA*CATATGAATGAAAAAAAACTGAAGCAACT
12733R	*GTGGTGGTGGTG*CTCGAGAAGTAATGCTGTCCTGCGGT
176339F	*GAAGGAGATATA*CATATGGTAGATCAAAACAATAATCA
176339R	*GTGGTGGTGGTG*CTCGAGATTGGAACTTATTAAATCAATAAGTGC
4190F	*GAAGGAGATATA*CATATGTTTTATCATCGCTATCCGGA
4190R	*GTGGTGGTGGTG*CTCGAGCACGATAAACAGCGGCTTTTGA
6745F	*GAAGGAGATATA*CATATGAAATACCGCTTTCCTGA
6745R	*GTGGTGGTGGTG*CTCGAGCCACGGAATATTGCCGTAGC
18167F	*GAAGGAGATATA*CATATGACTAAAACCGTGATCGCC
18167R	*GTGGTGGTGGTG*CTCGAGGGCTGCCGCTGGCGTCGC
81997F	*GAAGGAGATATA*CATATGACCACATATGCATTCCCG
81997R	*GTGGTGGTGGTG*CTCGAGCGCGGAACGTTCCTCACA
102447F	*GAAGGAGATATA*CATATGATGAAGAGTTTCCCGAA
102447R	*GTGGTGGTGGTG*CTCGAGCTGTTTTAGCGACGCACCG
66586F	*GAAGGAGATATA*CATATGGATGTAACCAAAGCCATACG
66586R	*GTGGTGGTGGTG*CTCGAGCAACCCCAATGCCTTAAGCG
72671F	*GAAGGAGATATA*CATATGAGTGCAGAATTGCAGTTACC
72671R	*GTGGTGGTGGTG*CTCGAGGTTGCTCATGTAGTAGCTGAAACC
21962F	*GAAGGAGATATA*CATATGCTGGGTGTCGACAACAT

Table B.2 (continued): Primers that were used in the study.

Primers	Sequence (5'-3')
21962R	*GTGGTGGTGGTG*CTCGAGCGATGAGCTGGAACTGCTTG
38300F	*GAAGGAGATATA*CATATGGAACTCTATGTAGGAAACATGT
38300R	*GTGGTGGTGGTG*CTCGAGCTTGGCTTCTTCGAAGCCTG
31013F	*GAAGGAGATATA*CATATGCTAATCAATTTTAAAAAACTCA
31013R	*GTGGTGGTGGTG*CTCGAGATCAGCCGCTGAAACAATACC
46252F	*GAAGGAGATATA*CATATGACCACGTACAAATTCCC
46252R	*GTGGTGGTGGTG*CTCGAGTGGATTTTGCCGTCGGCG
51689F	*GAAGGAGATATA*CATATGACATTCCCGAGCGACTT
51689R	*GTGGTGGTGGTG*CTCGAGGGCTGTCCTTGGGAATCCTC
72422F	*GAAGGAGATATA*CATATGCGGAGGGTCTTGGCC
72422R	*GTGGTGGTGGTG*CTCGAGTCTCGTCGAAGTATGCCTTCTC
46938F	*GAAGGAGATATA*CATATGCTTCAAGTTCGTGGTAATCG
46938R	*GTGGTGGTGGTG*CTCGAGATGCAAATGCTTGCAGAGAA
53527F	*GAAGGAGATATA*CATATGAACACACTGAACTTTCCAGC
53527R	*GTGGTGGTGGTG*CTCGAGTACAGAATTAATTTGTGCTAAACCGA
92869F	*GAAGGAGATATA*CATATGGAGGCATTCATGATTGA
92869R	*GTGGTGGTGGTG*CTCGAGGCCCTCAGCGGGGAAAATAT
11941F	*GAAGGAGATATA*CATATGGCCGAGAACGCCGAAAA
11941R	*GTGGTGGTGGTG*CTCGAGGAGCTGGGCCCGCGCGATC
18880F	*GAAGGAGATATA*CATATGGTGAAATTCACTTTCCATTTCC
18880R	*GTGGTGGTGGTG*CTCGAGGACCCCGAACATTTCCTGGA
960F	*GAAGGAGATATA*CATATGTTGGCGCGTCTTCGCACG
960R	*GTGGTGGTGGTG*CTCGAGGCCGGAGAGGCCGCTGGTC
70910F	*GAAGGAGATATA*CATATGACTTGCGAAAAACCCTGC
70910R	*GTGGTGGTGGTG*CTCGAGCCAGCTGCGCACGATCGTC
31506F	*GAAGGAGATATA*CATATGGAAGCCTACGGCCAC
31506R	*GTGGTGGTGGTG*CTCGAGGTCGGCGTTCTTACCGCC
30829F	*GAAGGAGATATA*CATATGAAAACAAGAAATAACTTTCT
30829R	*GTGGTGGTGGTG*CTCGAGTTTCAGCCCTAGAGCTTTTA
100149F	*GAAGGAGATATA*CATATGACTCTCCCCGGAGGATC
100149R	*GTGGTGGTGGTG*CTCGAGAAGGAATTGTTGGGCACAAGG
111095F	*GAAGGAGATATA*CATATGCTGGGGCTCTCCGTCGTG
111095R	*GTGGTGGTGGTG*CTCGAGACTGCGCCTGGCGAGGCCAG
97287F	*GAAGGAGATATA*CATATGGTGGGAGCCGACGTTACC
97287R	*GTGGTGGTGGTG*CTCGAGTAATGCTTGGCTTGATCGATCTC
199F	*GAAGGAGATATA*CATATGATCTGGAGAAAAATTGCTGAGC
199R	*GTGGTGGTGGTG*CTCGAGATAAATAGTTCCAGGAATTTGATTGTC
26764F	*GAAGGAGATATA*CATATGTTGGATCATAAACAACTAAAAGA
26764R	*GTGGTGGTGGTG*CTCGAGGGTATTGGTTTTAGGAGGCTGG
7918F	*GAAGGAGATATA*CATATGGAACATCAAAAACGAACTTCA
7918R	*GTGGTGGTGGTG*CTCGAGTAGATTTTCGCCGTTTTCTTCGA
14933F	*GAAGGAGATATA*CATATGAAAACGCAACGTCAACT
14933R	*GTGGTGGTGGTG*CTCGAGATCAACATACTTAAGAACTGCGACT
29016F	*GAAGGAGATATA*CATATGAGTAAATTTCCAGAAGCA
29016R	*GTGGTGGTGGTG*CTCGAGCTCTCCTAACCCATTTTTATCGATGAC
29118F	*GAAGGAGATATA*CATATGAAACAATTTCCAAAAAA
29118R	*GTGGTGGTGGTG*CTCGAGCAAATTTTCTCCATTAGACTGGA
203416F	*GAAGGAGATATA*CATATGAAATTTCCAACTGATTTTCTATGG
203416R	*GTGGTGGTGGTG*CTCGAGCTCCCCTTCTCTCGTCCCG
36838F	*GAAGGAGATATA*CATATGAAACAGACATTTCCAACT
36838R	*GTGGTGGTGGTG*CTCGAGCTCCTCCAACGAAGGGATCG
27669F	*GAAGGAGATATA*CATATGCGGTTCCCTGAGGGG
27669R	*GTGGTGGTGGTG*CTCGAGGCCCACATCCTCGCCGTT
28464F	*GAAGGAGATATA*CATATGACTAACCTGCATTTTCCT
28464R	*GTGGTGGTGGTG*CTCGAGGTCGACAAAGCCGTTTTGTAGA

Table B.3 (continued): Primers that were used in the study.

Primers	Sequence (5'-3')
29062F	*GAAGGAGATATA*CATATGCAACAAAAACAATGTAAAAAGGGG
29062R	*GTGGTGGTGGTG*CTCGAGTTCTGCCAAACTTGCGCCAT
37597F	*GAAGGAGATATA*CATATGGCTTTTCCAACTGGTTTCC
37597R	*GTGGTGGTGGTG*CTCGAGTTCTTCCGTCAAACTACGACCA
36243F	*GAAGGAGATATA*CATATGAGTGTATTTCCAAAAAATTTCCT
36243R	*GTGGTGGTGGTG*CTCGAGCTTGCCGCCTTTCAAAAATGA
28541F	*GAAGGAGATATA*CATATGGGATTTCCAGAAAACTT
28541R	*GTGGTGGTGGTG*CTCGAGAGCCAAATCTTCCCCATTTGAG
20495F	*GAAGGAGATATA*CATATGAGTTTACCAGAAGAATTTTTATGG
20495R	*GTGGTGGTGGTG*CTCGAGAGGCTTCGCTTGGATCAAGG
12836F	*GAAGGAGATATA*CATATGAAATGGCTATGTTCTGTAGGTG
12836R	*GTGGTGGTGGTG*CTCGAGTTGCAGTTCGAATTCGGCTT
8125F	*GAAGGAGATATA*CATATGACTAAGAATGTATTTCCAGA
8125R	*GTGGTGGTGGTG*CTCGAGAAGATTTACACCGTTAGATGCGA
7761F	*GAAGGAGATATA*CATATGGGACTAAAGTTGAATGATGA
7761R	*GTGGTGGTGGTG*CTCGAGCAAGTCCACACCATGTGAGG
3947F	*GAAGGAGATATA*CATATGAAAAAGAGCCTCCTCGC
3947R	*GTGGTGGTGGTG*CTCGAGTCATGGTGCTGGGGTGATG
27F	AGAGTTTGATCCTGGCTCAG
1525R	AAGGAGGTGATCCAGCC
515F	GTGCCAAGCAGCCGCGGTAA
1209R	GGGCATCACAGACCTG

Appendix C

Table C.1: Success of each step for each candidate gene.

Gene ID	Gene amplification	Transformation verification	Expression	Enzyme activity
199	√	√	×	
960	√	√	×	
2484	√	√	√	×
2599	√	√	√	×
2864	√	×		
3947	√	×		
4190	√	√	×	
5020	×			
5252	√	√	×	
6745	√	√	√	×
7430	√	√	√	×
7761	√	√	√	×
7918	√	×		
8125	√	√	×	
9683	√	√	√	√
11941	√	√	×	
12130	√	√	×	
12733	√	√	×	
12836	√	√	√	×
14933	√	√	×	
16307	√	√	√	×
18167	×			
18880	×			
20495	√	√	√	×
21962	×			
26764	√	√	√	×
27669	√	√	√	×
28464	√	√	√	×
28541	√	√	√	×
29016	√	√	×	
29062	√	√	√	
29118	√	×		
29580	√	√	√	×
30829	×			
31013	√	√	×	
31506	√	√	×	
32123	√	√	×	
36243	√	√	√	×
36838	√	√	√	×
37597	√	√	√	×
38300	×			
38877	√	√	×	
38878	√	√	×	
41502	√	√	×	
42073	√	√	×	
46252	√	√	√	×

Table C.2: Success of each step for each candidate gene.

Gene ID	Gene amplification	Transformation verification	Expression	Enzyme activity
46845	×			
46938	√	√	×	
50455	√	√	×	
51689	√	×		
53527	×			
66586	√	√	√	×
70910	√	√	√	√
72422	√	√	×	
72671	√	√	×	
81997	√	√	√	×
92869	√	√	√	×
97287	√	√	√	×
100149	√	√	×	
102447	√	√	×	
111095	×			
176339	√	√	×	
203416	√	√	×	

CURRICULUM VITAE

Name Surname: Sevde Şencan

Place and date of birth: İstanbul, 03.09.1990

Permanent Address: Sulukule TOKİ Evleri Şüheda sk. Ada118B D:7 Fatih, ISTANBUL

E-mail: sevdesencan@gmail.com

EDUCATION:

2008-2013 Biology (B.Sc.), Middle East Technical University
Ankara, Turkey