Investigation and Isolation of Cellulase-Producing microorganisms in the Red Sea

Thesis by

Siham Kamal Fatani

In Partial Fulfillment of the Requirements

For The Degree of

Master of Science

King Abdullah University of Science and Technology,

Thuwal, Saudi Arabia

May, 2016

EXAMINATION COMMITTEE APPROVALS FORM

The thesis of Siham Kamal Fatani is approved by the examination committee.

Committee Chairperson: Prof. Takashi Gojobori
Committee Members: Prof. Stefan Arold, Dr. Katsuhiko Mineta

ABSTRACT

Investigation and Isolation of Cellulase-Producing microorganisms in the Red Sea

Thesis by

Siham Kamal Fatani

Cellulolytic microorganisms are considered to be key players in biorefinery, especially for the utilization of plant biomass. These organisms have been isolated from various environments. The Red Sea is one of the seas with high biodiversity and a unique environment, characterized by high water temperature and high salinity. However, there is little information regarding cellulases in Red Sea environments. The aim of the present study is to evaluate the Red Sea as a gene resource for microbial cellulase. I first surveyed microbial cellulases in the Red Sea using a method called metagenomes, and then investigated their abundance and diversity. My survey revealed that the Red Sea biome has a substantial abundance and a wide range of cellulase enzymes with substantial abundance, when compared with those in other environments. Next, I tried to isolate cellulase-active microorganisms from the Red Sea and I successfully obtained seven strains of four different taxonomic groups. These strains showed a similarity of 99% identity to *Aspergillus ustus*, 99% to *Staphylococcus pasteuri*, 99% to *Bacillus aerius* and 99% to *Bacillus subtilis.*

The enzyme assay I conducted, revealed that these strains actually secreted active cellulases. These results suggest that the Red Sea environment can be, indeed, an excellent gene resource of microbial cellulases.

ACKNOWLEDGEMENTS

The completion of this successful work is the consequence of great help and guidance from many people; faculty, family and friends. I am really happy to have these people by my side whilst undertaking my Master study.

First, I would like to express my profound gratitude and respect to my supervisor, Prof. Takashi Gojobori, Distinguished Professor of Bioscience and Associate Director of Computational Bioscience Research Center for his professional guidance, and regular encouragement and motivation at various stages of this work. I would also like to thank Dr. Katsuhiko Mineta for his support and advice during my research. Moreover, I would like to express my deepest appreciation to Dr. Yoshimoto Saito for his assistance and suggestions throughout my project. I also appreciate Mr. Mohammad Al-arawi and Dr. Hayedeh Behzad for their support and technical advice and without whom this work would not have been possible for me to complete.

In addition, I would like to thank my committee members, Prof. Takashi Gojobori, Prof. Stefan Arold and Dr. Katsuhiko Mineta for giving their time to review my thesis and for offering their insight and suggestions.

I would also give my sincere thanks to my institution King Abdullah University of Science and Technology as well as all lab members, my family and friends' without

whom this project would not be successful.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CMC	Carboxy Methyl Cellulase
NM	Nutrient Media
FPA	Filter Paper Assay
NCBI	National Centre for Biotechnology information
BLAST	Basic Local Alignment Search Tool
Rpm	Revolutions per minute
CAZy	Carbohydrate-Active Enzymes
Pfam	Collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs).
GH	Glycoside hydrolysis
mg	Milligram
µg	Microgram
ml	Milliliter
U	Unit

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction

1.1 Cellulose

Lignocellulosic plant biomass is composed mainly of three types of polymers: cellulose, hemicellulose and lignin. Cellulose is the initial structural component of lignocellulosic biomass, which makes around 40-50% [1]. It is embedded within a matrix of hemicellulose 20-40% and lignin 20-30% that are considered as rich sources of carbohydrates in higher plants [2]. Cellulose is a linear homologous polymeric chain consisting of D-glucose residues that make up to 10,000 glucose residues, linked by β-1,4 glycosidic bonds [3, 4].

Considering that cellulose is the most abundant and renewable feedstock for energy source, this increases its importance as it can be used for the production of alternative liquid fuels by the conversion of lignocellulosic biomass to biofuel, mainly bioethanol and methane. It may be produced through thermochemical processes or biochemical processes [1]. Moreover, Combined with the wide range of applications they are now available to humanity within industries such as food and beverage, pulp and paper, textile, animal feed, detergent and agriculture [5-7].

1.2 Cellulase enzymes and cellulose degradation

Cellulase are a group of enzymes that catalyze the hydrolysis of cellulose into sugars [8]. During hydrolysis of the cellulolytic materials, cellulase enzymes cleave glycosidic bonds by the addition of a water molecule [1, 9]. They differ structurally and mechanistically from each other, and are classified on the catalytic action they undergo (Figure.1) [10]. They are:

- Endoglucanases (EC 3.2.1.4), which cut at random internal amorphous sites in cellulose polysaccharide chain.
- Exoglucanase (EC 3.2.1.91), which cut two to four units of the reducing or non-reducing ends of cellulose polysaccharide chains, resulting in the formation of cellobiose.
- β-Glucosidases (EC 3.2.1.21), which is an enzyme that hydrolase the soluble cellobiose, the major product of endo and exo-glucanase to glucose, from the non-reducing end [11].

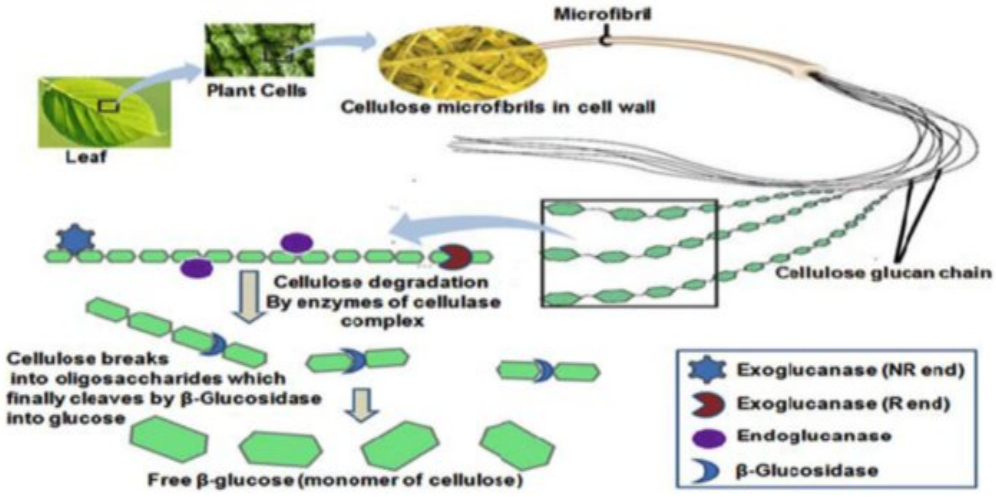


Figure 1. Diagrammatic overview represents the cellulose hydrolysis by cellulase system. The figure was adapted from Singh, G., A. Verma, and V. Kumar.

1.3 The taxonomic diversity of cellulase producers

Many fungal and bacterial strains secrete cellulase enzymes during their growth on cellulosic materials. These enzymes are secreted either in multi-protein complexes or in separate enzymes. A large amount of microbial cellulases (glycosyl hydrolases) have been isolated, classified and characterized from different environments so far [12].

There are many reports of cellulose genes in both bacteria and fungi. For example, *Clostridium*, *Acetivibrio*, *Bacillus*, *Bacteriodes*, *Cellomonas*, *Thermonospora*, *Ruminococcus* and *Erwinia* species strains are known as cellulase-producing bacteria [13], [14] while *Aspergillus*, *Penicillium*, *Trichoderma* and *Fusarium species* are revealed to produce the enzyme in fungi [15]. Cellulase-producing bacteria and fungi were isolated from variant sources such as agricultural wastes, soil, sediments, seawater,animal excreta and so on [9].

Marine microbes possessing cellulase activities play a crucial role in decomposition of organic matters such as pectein pectin?, cellulose, chitin and cycling of nutrients. The distribution of the microbes depends on changes in water temperature, salinity and other parameters [16]. Enzymes that produced from marine microorganisms, might provide several features and advantages over other traditional enzymes due to a wide range of environments [17].

There are also a lot of reports regarding isolation of strains from marine environments. For example, Smira et al (2011) isolated cellulase-active strains of *Streptomyces variabilis, Kocuria rosea* and *Stenotrophomonas maltophilia* from the Persian Gulf. Kim

et al (2009) also reported that *B. subtilis subsp. subtilis A-53* was obtained from a hot spring. To our knowledge, however, there are few reports studying microbial cellulases in the Red Sea environment. In particular, the diversity, abundance and characteristics of cellulase genes or cellulolytic microorganisms are still far from fully explored.

1.4 Microbial diversity in The Red Sea

The Red Sea environment is very unique among various oceans on earth; for example, high salinity, low nutrient concentration, high temperature and so on. Ngugi et al. (2012) revealed that the community structure of bacteria and picoeukarotic plankton in the Red Sea surface water is quite different from those of other marine environments, and that many of these organisms can apparently adapt to harsher conditions and higher temperature. They also stated that numerous species can be found only in the Red Sea environment [18].

In spite of its uniqueness of the environment, however, cellulolytic microorganisms and cellulases have remained unstudied in the Red Sea.

1.5 Aim

The aim of this study is to reveal whether the Red Sea is an excellent microbial or genetic resource of cellulase enzymes. For attaining the aim, in this study we conducted the survey of microbial cellulase genes in the Red Sea environment using metagenomes. Moreover, I examined whether we could obtain cellulase active microbial species from the Red Sea by the isolation experiment.

1.6 Objective

The following are the objectives of the present study;

- To characterize and classify the cellulase genes of the Red Sea by the metagenomic approach,
- To reveal characteristics of cellulase genes in the Red Sea by comparing the metagenome with other environments, and
- To isolate cellulase-active microorganisms from the Red Sea.

1.7 Thesis outline

The first chapter of this dissertation includes an introduction covering the topics of cellulose, cellulase enzymes and cellulose degradation, as well as the Red Sea and its marine environment as a source of microbiome. Moreover, previous studies of the isolation of cellulase genes and active cellulase microorganisms were discussed. The aim and motivation for conducting the present thesis has also been described in this chapter.

In the second chapter, the materials and methodology of the identification of cellulase genes from metagenome sequences, were described to reveal the abundance and diversity of cellulases in the Red Sea. The method of isolation and screening of bacterial and fungal strains and the assay of their enzymatic activities were also described in this chapter.

In the third chapter of the dissertation, I included all the results of this study with relevant discussion. In the end, the conclusions of the study are described, along with future perspectives in the relevant field.

Chapter 2. Material and methods

2.1 Survey for the cellulase genes in the Red Sea using metagenome approach

2.1.1 Identification of glycoside hydrolysis families.

About 250,000 cellulase degradation enzymes were classified into 135 Glycoside Hydrolysis families (GH families), according to the CAZy database classification [22, 23]. Among GH families in the CAZy database, I defined the families that contain more than 500 enzymes registered in Uniprot (http://www.uniprot.org/) as major GH families (34 GH families/ 135 in total were included in the major families). I then conducted the functional domain search of sequences in major GH families ((Uniprot-registered sequences only) against the Hmmscan [24] and Pfam database (Pfam-A, http://pfam.xfam.org/). The results obtained clearly showed that each of the major GH families has a unique Pfam domain to distinguish each other.

2.1.2 Metagenome analysis of the Red Sea and other environments

Twenty sequences read archive (SRA)??? files of metagenomes were retrieved from the NCBI-SRA (http://www.ncbi.nlm.nih.gov/sra/?term=SAR) (Table.1). The obtained SRA files were converted to the fastq files using a fastq-dump program in sra tool kit packages [25]. Trimming and quality control has been done using the prinseq-lite.pl [26], which means that the QC score of the resulting reads were more than 20. The obtained reads were assembled using Newbler software version 2.3 (Roche/454 Life Sciences, Brandford, CT) with the default setting. Genes on contigs were predicted using MetaGene software (http://metagene.nig.ac.jp/metagene/metagene.html). The predicted gene sequences were translated into amino acid sequences, and then were annotated using

the Hmmscan and Pfam database Pfam-A (http://pfam.xfam.org/). The sequences annotated by major GH family-specific pfam IDs were extracted as candidate cellulase genes. After that, the Shannon diversity index was calculated in each data of the metagenome following this formula [27].

$$H = -\sum_{i=1}^{S} pi \ln pi$$

Table 1. List of metagenome sequences file obtained from NCBI-SAR archive.

Run ID (SAR)	Sample	NGS	Sites
ERR770995	Red sea, 5m	454 GS FLX Titanium	Seawater
ERR770994	Red sea, 1m	454 GS FLX Titanium	
SRR023397	Sargasso sea, 20m	454 GS FLX Titanium	
SRR023398	Sargasso sea, 50m	454 GS FLX Titanium	
SRR023401	Sargasso sea, 100m	454 GS FLX Titanium	
ERR770970	Greenland, 0m	Illumina MiSeq	
ERR770983	Black sea	Illumina MiSeq	
ERR771001	Adriatic sea	Illumina MiSeq	
ERR771051	Celtic sea	Illumina MiSeq	
SRR091234	Amazon river	Ion Torrent PGM	Freshwater
SRR1146621	USA: Virginia James River	Ion Torrent PGM	
SRR1169884	Mississippi river	Ion Torrent PGM	
ERR358542	Swedish lake Erken	454 GS FLX Titanium	
ERR358543	Swedish lake Ekoln	454 GS FLX Titanium	
ERR358549	American Mendota summer	454 GS FLX Titanium	
ERR358550	American Mendota spring	454 GS FLX Titanium	
DRR014261	Climax forest soil in the island of miyake	454 GS FLX Titanium	Soils
SRR606840	Anguil Rhizospheric Soil Agriculture (Conventional Tillage)	454 GS FLX Titanium	
SRR606844	Anguil Rhizospheric Soil Agriculture (noTillage)	454 GS FLX Titanium	
DRR012573	Bulk soil in the rice paddy field	Illumina HiSeq 2000	

2.2 Isolation of cellulase active microoragnisms from the Red Sea

2.2.1 Collection of samples

All samples were collected in the coastal region of the Red Sea, Saudi Arabia, Thuwal on August 26, 2015 and September 30, 2015. The seawater samples were obtained from the surface of the site at (22° 17.444’N, 39° 03.183 ’E) using a Niskin bottle. Plankton samples were collected by drawing a net of which the mesh size is 0.63 µm, at 1 knot for ten minutes. Seaweed samples were obtained at the KAUST coastal marina (22°18'16.7"N 39°06'12.1"E). Samples were kept in sterile tubes and stored at 4 ˚C until use (Figure.2).

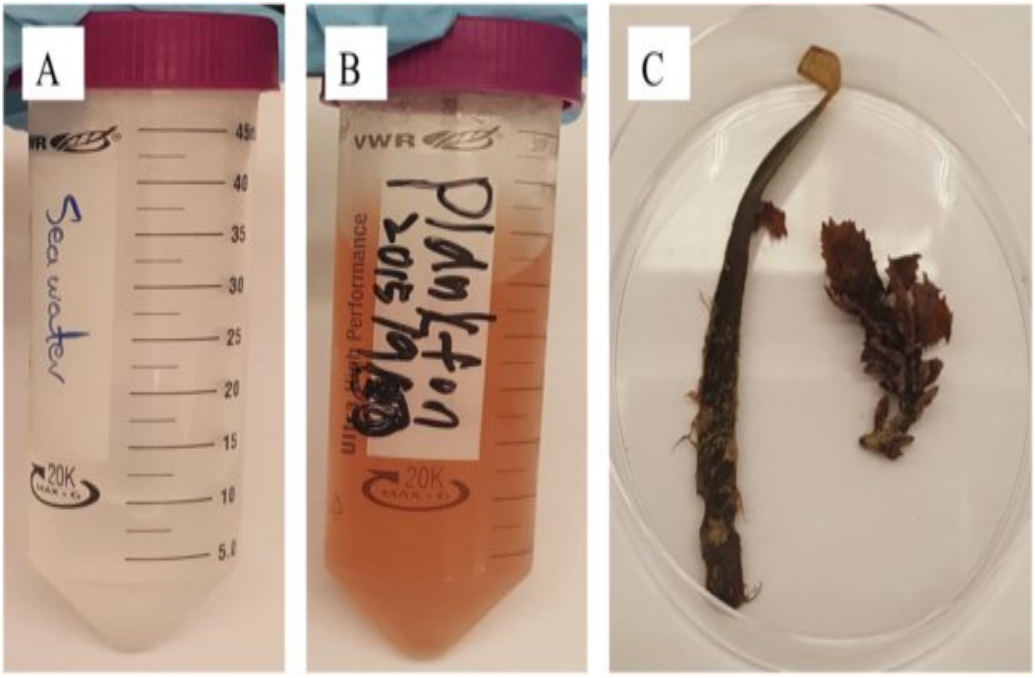


Figure 2. The collected samples from the Red Sea. A: Seawater. B: Plankton. C: Seaweeds.

2.2.2 Preparation of isolation media

Nutrient media (NM media), (0.3% Beef extract, 0.5% Peptone, 0.5% NaCl and 1.7% Agarose) were used for the isolation of microorganisms from the samples [8]. Carboxymethyl cellulase (CMC) media (0.2% $NaNO_3$, 0.1% K_2HPO_4, 0.05% $MgSO_4$, 0.05% KCL, 0.02% peptone, 0.2% CMC and 1.7% Agarose) were also used for the screening of cellulase producing microorganisms [28].

2.2.3 Preparation of collected samples

Seawater and plankton samples tubes were vortexed for 15 min, and then left to settle . The obtained supernatants were collected as an inoculation source. For the seaweeds, 1g of the seaweed was added to 10 ml of sterilized distilled water. 1g of the glass beads (425-600 µm) were dipped in 1 M HCl for 1 hr and then rinsed with distilled water. The resultant acid-washed glass beads were added to the seaweed tubes and vortexed for 10 min, and then left for 5 min to let the solids settle. The supernatant was collected as an inoculation source. The dilution series were constructed from 10^{-1} to 10^{-7} using sterilized water. 100-µl aliquots of each dilution series were spread on the nutrient media and incubated at 30 ^{o}C for 48 hrs. The obtaining colonies were sub-cultured onto the new NM media [29].

2.2.4 Screening of cellulase producing Bacteria and Fungi

The screening of cellulase producing bacteria and fungi was conducted according to the method described by Kasana, R.C., et al. [8], with slight modification. Briefly, the isolates were grown on NM media at 30 ˚C for two days. The growing isolates were transferred on to CMC media and incubated at 30 ˚C for 4 days. For the detection of cellulase enzyme expression, 1ml of 1% Congo red solution was added to the 4-day-old CMC plate after inoculation, and incubated for 30 min. After removing the Congo red solution, the plates were washed with a 1% sodium hydroxide solution for 15 min. After that, strains forming clear halo zones on CMC media were selected as cellulase-producing strains.

2.2.5 Extraction of genomic DNA and amplification of 16S and 18S rRNA gene region.

For the sequence analysis, bacteria and fungal genomic DNA was extracted and purified following the protocol described by Nakada, M., et al [30]. In brief, microbial isolates were grown in an NM broth (0.06 g beef extract, 0.1 g peptone and 0.1 g NaCl) at 30 ˚C for 2 days. The media was centrifuged at 3000 rpm for 5 min at 4 ˚C to collect the cells. For the bacteria strains, the collected pellet was resuspended in extraction buffer (100 mM NaH_2PO_4, pH 8, EDTA 100 mM, 100 mM NaCl, 500 mM Tris [pH 8], 10% [wt/vol] and 1% mercaptoethanol). Fungal mycelia were also collected by the centrifugation at 3000 rpm. The obtained mycelia were removed from the water using filter papers, and then was powdered in liquid nitrogen using a pestle and mortar. The powdered mycelia were then, resuspended in the extraction buffer mentioned above. The

resuspended buffer was then heated at 70 ˚C for 30 min with shaking at 30 rpm and then DNA was extracted using the phenol/chloroform extraction method. Obtained DNA pellets were dissolved into 20 µl TE buffer and stored it at -20^{o}C. The concentration of the DNA was measured using Qubit® dsDNA BR Assay Kit, 500 assay (ThermoFisher scientific®) [31].

The PCR amplification was performed using QIAGEN Multiplex PCR kit. Bacterial 16S rRNA genes were amplified using primers Bac27F (5′-AGAGTTTGGATCMTGGCTCAG-3′) and Univ1492R (5′-CGGTTACCTTGTTACGACTT-3′). Fungal 18S rRNA genes were amplified using primer pairs, Euk1A (5′-CTGGTTGATCCTGCCAG-3′) and Euk516r (5′-ACCAGACTTGCCCTCC-3′). The PCR program for gene amplification consisted of an initial denaturation at 94^{o}C for 15 min, annealing at 55 C^{o} for 30 sec and an extension at 72 ˚C for 30 sec, and a final extension at 72 ˚C for 10 min. The obtained amplicons were purified using the WIZARD cleaning kit (Promega) and cloned into the plasmid pCR™ 2.1 using the TA Cloning® Kit, with pCR™ 2.1.

16S or 18S rRNA sequences of the clones obtained, were determined using the ABI (3730xl-DNA analyzer) sequencing system with primers M13F (5′-GTAAAACGACGGCCAG-3′) and M13R (5′-CAGGAAACAGCTATGAC-3′) [32]. The resulted sequences were used for inference of their taxonomy using NCBI-BLAST search (blastn). Phylogenetic trees were constructed using the MEGA6 software [32]. Alignment construction was performed on the Clustalw software in the MEGA packages.

2.3 Enzyme activity assay

2.3.1 Production of cellulase enzyme using filter paper as substrate and preparation of crude enzyme.

The broth media was prepared using 0.2% $NaNO_3$, 0.1% K_2HPO_4, 0.05% $MgSO_4$, 0.05% KCl, 0.02% peptone and a Whatman no.1 Filter paper as a carbon source. After inoculation, the broth media was incubated for four days at 30 ˚C, shaking at 150 rpm.

2.3.2 Measuring the cellulase activity

Filter paper assay activity was measured by taking the cell-free supernatant and analyzed for cellulase activities. 50 mg of Whatman no.1 filter paper was suspended in a 1.5 ml sodium acetate buffer 0.1 M (PH 5.0), and incubated with 0.5 ml enzyme solution for 2 hrs at 50 ˚C. After incubation, the sample were centrifuged at 10,000 xg for 5 min, and 1 ml of the supernatant was taken for determination of cellulase activity (Filter paper degrading (FPAase) activity) [33]. The reaction was stopped by adding 3 ml of DNS reagent (dinitrosalicylic acid 3.33 g in 200 ml Distilled water, adding 5.33 g NaOH and 100 g Potassium sodium tartrate (Rochelle Salt)) [33, 34] and boiled for 5 min in a water bath, then cooled with cold water for 10 min, and the optical density of the reducing sugar was measured at 540 nm. The generation of monomeric sugar by cellulase was determined using a calibration curve for D-glucose. One unit of FPase enzymatic activity (U) was defined as the amount of enzyme that released 1 μmol of reducing sugars as glucose equivalents min^{-1} [35, 36]. The activities were estimated following the described formula by Thomas, as following [37]:

1 U = 1 μmol/min of glucose equivalent released

= 0.18 mg/min of glucose

The formula used for calculation the enzyme activity:

$$= \frac{amount\ of\ equivalent\ sugar\ generated}{0.18} \times 0.5\ (volume\ of\ enzyme\ solution) \times 120\ \text{min}(reaction\ time)$$

We also prepared filter paper broth media and cultured it without inoculation in the same condition. We used it as a control, calculating the FPase activities of isolates after subtracting the amount of reduced sugar of the control from those of isolates.

Chapter 3. Results and discussion

3.1 Survey for the cellulase genes in the Red Sea using the metagenome approach

3.1.1 Identification of glycoside hydrolysis families.

CAzy is a database of cellulases (Glycoside hydrolyses (GH)), enzymes that hydrolyze the glycosidic bonds in the cellulose, with more than 280,000 cellulase enzymes registered there [23]. These sequences are also classified into 135 families in the database. In this, I selected 34 GH families that contain more than 500 GH sequences that are also registered in the Uniprot database as major GH groups. The amino acid sequences in these GH groups were then annotated by a Pfam functional domain search (see Table.2). I found that all major GH groups can be characterized with specific Pfam domains.

Table 2. The classification of glycoside hydrolase families.

pfam IDS	Families	enzymes	Classification	Domain
PF00232	GH1	β-glucosidase, β-galactosidase, β-mannosidase, exo-β-1,4-glucanase	Cellulase/ Hemicellulase	Glycoside hydrolase family 1
PF02836	GH2C	β-galactosidase, β-mannosidase, exo-β-glucosaminidase	hemicellulase	Glycoside hydrolase family 2
PF00933	GH3N	β-glucosidase, xylan 1,4-β-xylosidase, exo-1,3-1,4-glucanase	Cellulase/ hemicellulase	Glycoside hydrolase family 3
PF01915	GH3C	β-glucosidase, xylan 1,4-β-xylosidase, glucan 1,4-β-glucosidase	Cellulase/ hemicellulase	Glycoside hydrolase family 3
PF02056	GH4	maltose-6-phosphate glucosidase, α-glucosidase, α-galactosidase	Cellulase/ hemicellulase	Glycoside hydrolase family 4

PF00150	GH5	endo-β-1,4-glucanase / cellulase, endo-β-1,4-xylanase, β-glucosidase, β-mannosidase	Cellulase/ hemicellulase	Cellulase (glycosyl hydrolase family 5)
PF01270	GH8	chitosanase; cellulase; licheninase; endo-1,4-β-xylanase; reducing-end-xylose releasing exo-oligoxylanase	Cellulase/ hemicellulase	Glycoside hydrolase family 8
PF00759	GH9	endoglucanase; endo-β-1,3(4)-glucanase, β-glucosidase; exo-β-glucosaminidase	Cellulase/ hemicellulase	Glycoside hydrolase family 9
PF00331	GH10	endo-1,4-β-xylanase; endo-1,3-β-xylanase; tomatinase; xylan endotransglycosylase	Cellulase/ hemicellulose	Glycoside hydrolase family 10
PF00457	GH11	endo-β-1,4-xylanase; endo-β-1,3-xylanase	hemicellulase	Glycoside hydrolase family 11
PF00128	GH13	α-glucosidase; cyclomaltodextrin glucanotransferase; α-amylase; 4-α-glucanotransferase	hemicellulase	Alpha-Amylase
PF00723	GH15	glucoamylase; glucodextranase; α,α-trehalase; dextran dextrinase	hemicellulase	Glycosyl hydrolase family 15
PF00722	GH16	endo-1,3(4)-β-glucanase; β-agarase; endo-β-1,4-galactosidase	Cellulase/ hemicellulase	Glycosyl hydrolase family 16
PF00704	GH18	chitinase; lysozyme; endo-β-N-acetylglucosaminidase; peptidoglycan hydrolase with endo-β-N-acetylglucosaminidase specificity; Nod factor hydrolase; xylanase inhibitor; concanavalin B; narbonin	hemicellulase	Glycosyl hydrolase family 18
PF00182	GH19	chitinase; lysozyme	hemicellulase	Chitinase Class I
PF01464	GH23	lysozyme type G; peptidoglycan lyase; chitinase	hemicellulase	Transglycosylase SLT domain
PF00959	GH24	lysozyme	cellulase/ hemicellulase	Phage_lysozyme/Glycoside hydrolase family 24

PF01183	GH25	lysozyme	cellulase/ hemicellulase	Glycosyl hydrolase family 18
PF00295	GH28	polygalacturonase; exo-polygalacturonase; exo-polygalacturonosidase; rhamnogalacturonase; endo-xylogalacturonan hydrolase	hemicellulase	Glycoside hydrolase family 28
PF01055	GH31	α-glucosidase; α-galactosidase; α-mannosidase; α-xylosidase; α-glucan lyase; isomaltosyltransferase	cellulase/ hemicellulase	Glycoside hydrolase family 31
PF00251	GH32	β-2,6-fructan 6-levanbiohydrolase; endo-inulinase; fructan β-(2,6)-fructosidase/6-exohydrolase	cellulase/ hemicellulase	Glycosyl hydrolase family 32
PF02012	GH33	sialidase or neuraminidase; trans-sialidase; 2-keto-3-deoxynononic acid hydrolase; anhydrosialidase ; 3-deoxy-D-manno-octulosonic-acid hydrolase	hemicellulase	BNR repeat-like domain
PF00064	GH34	sialidase or neuraminidase	hemicellulase	Neuraminidase/Glycosyl hydrolase family 33
PF01301	GH35	β-galactosidase; exo-β-glucosaminidase; exo-β-1,4-galactanase; β-1,3-galactosidase	hemicellulase	Glycosyl hydrolase family 35
PF16875	GH36N	α-galactosidase; α-N-acetylgalactosaminidase	hemicellulase	Glycosyl hydrolase family 36 N-terminal d
PF01074	GH38N	α-mannosidase; mannosyl-oligosaccharide α-1,2-mannosidase	hemicellulase	Glycosyl hydrolase family 38 N-terminal domain
PF04616	GH43	β-xylosidase; α-L-arabinofuranosidase; xylanase; exo-α-1,5-L-arabinofuranosidase	hemicellulase	Glycoside hydrolase family 43

PF12891	GH44	endoglucanase; xyloglucanase	cellulase/ hemicellulase	Glycoside hydrolase family 44
PF02011	GH48	reducing end-acting cellobiohydrolase; endo-β-1,4-glucanase; chitinase	hemicellulase	Glycoside hydrolase family 48
PF03512	GH52	β-xylosidase	hemicellulase	Glycoside hydrolase family 52
PF03065	GH57	α-amylase; α-galactosidase; 4-α-glucanotransferase	hemicellulase	Glycoside hydrolase family 57
PF01832	GH73	lysozyme; mannosyl-glycoprotein endo-β-N-acetylglucosaminidase;	hemicellulase	Glycoside hydrolase family 73
PF02446	GH77	amylomaltase or 4-α-glucanotransferase	hemicellulase	Glycoside hydrolase family 77
PF13406	GH103	peptidoglycan lytic transglycosylase	hemicellulase	Transglycosylase SLT domain

3.1.2 Metagenome analysis of the Red Sea and other environments

To investigate the abundance and diversity of cellulases, I surveyed cellulase genes in various environments including the Red Sea using metagenomes. To compare the abundance of cellulase genes in various environments, I first calculated the read numbers of cellulase genes per 1,000,000 reads in each metagenomic samples (Figure 3). As a result, I found that among metagenomes of seawater, fresh water and terrestrial soils, fresh water metagenomes had the largest amount of cellulases in general. Seawater environments also show high abundance of cellulases, although the number of cellulases decreased along with the depth. On the other hand, cellulases were much less abundant in soil environments.

To evaluate the diversity of cellulases in these environments, I classified cellulase genes into different GH families. I then calculated the Shannon diversity index using the distribution of reads among the GH families in each metagenome (Figure 4).

Among seawater environments, the Red Sea metagenomes show abundance and diversity similar to the other sea environments (Figures 3 and 4) and much higher than terrestrial soils. Although the Red Sea is known as a severe environment for organisms due to its high temperature, high salinity, etc, these results suggest that the Red Sea biomes have abundant and diverse cellulase genes that are sufficient as a gene or a microbial resource of cellulases. The severe environment may rather allow us to expect the unknown presence of cellulase enzymes in the Red Sea.

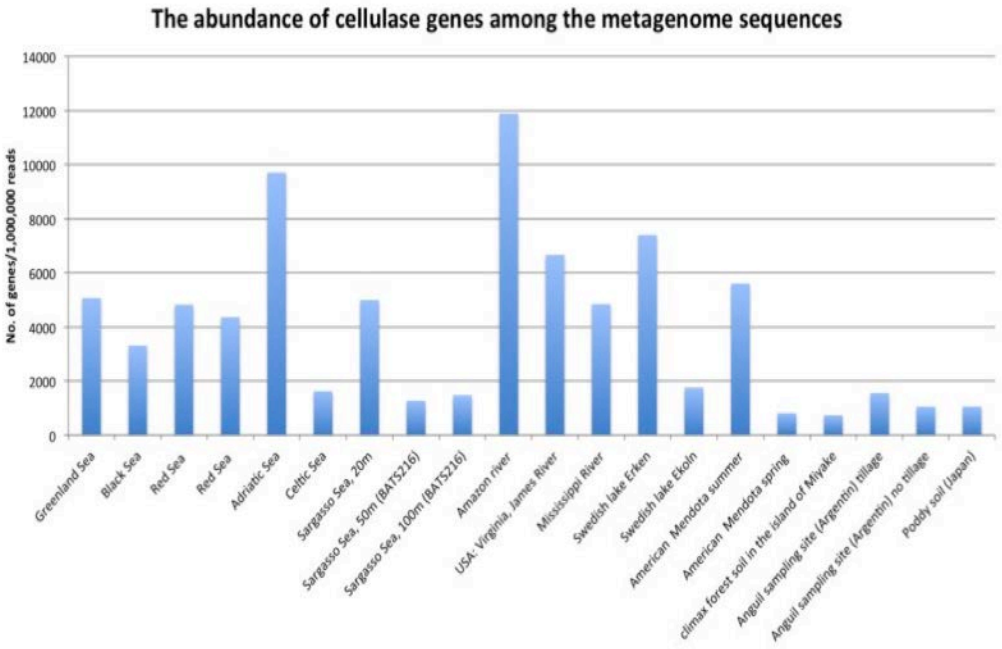


Figure 3. Abundance of cellulase genes in various environments. X-axis shows various locations whereas Y axis shows read number of cellulose genes per 1,000,000 reads.

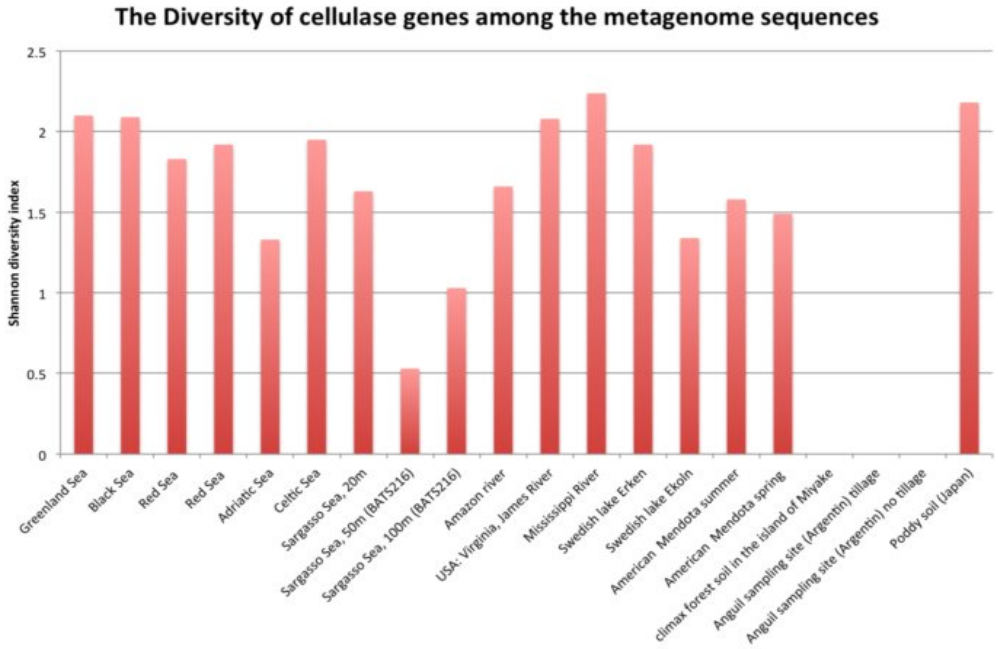


Figure 4. The diversity of cellulase genes in various environments. X-axis shows various locations whereas Y axis shows the Shannon diversity index.

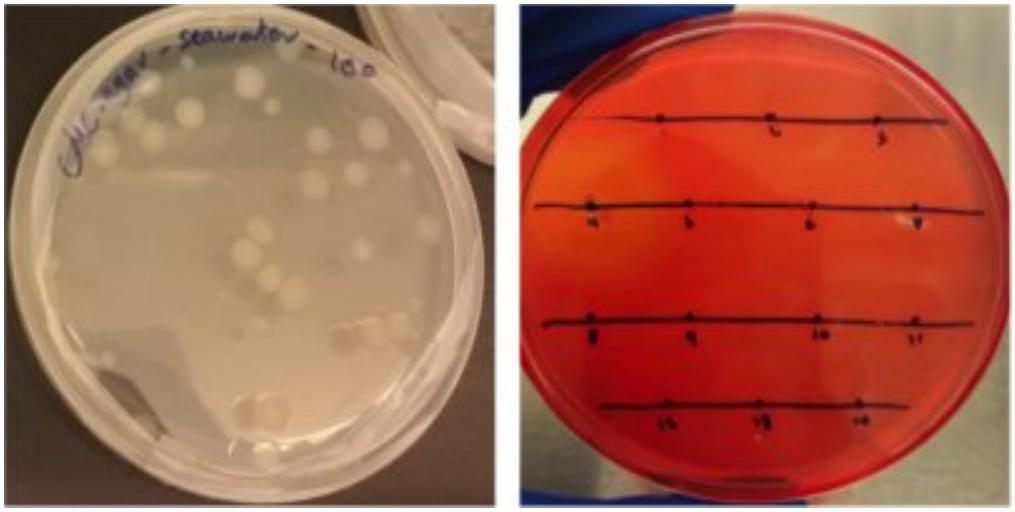
3.2 Isolation of cellulase active microorgnisms from the Red Sea

3.2.1 Isolation of fungal and bacterial cultures

The survey in this study suggests that Red Sea has cellulase-active microorganisms. Next, I tried to isolate microbial strains that possess cellulase activities from the Red Sea samples.

To get cellulase-active strains, I tried to isolate microorganisms from seawater samples using a nutrient rich (NM) media, I have then successfully obtained 28 bacterial strains (SWB1-SWB28). I then, conducted screening of cellulase-active strains using CMC media. However, I was unable to gather any positive strains from them (Figure.5). It is

easily suggested that although there are a lot of microbial species reported to have cellulase genes, actual cellulase-active microbiomes may be present on the surface of cellulose containing substances. I have, therefore, tried to isolate cellulase-active



microorganisms from the surface of seaweeds and planktons that were collected from the Red Sea.

Figure 5. Isolation (A) and screening (B) of cellulase active microorganisms from the seawater of the Red Sea.

I obtained 64 bacterial colonies (SB1-SB64) on an NM media from the seaweed while ten microbial isolates are obtained on an NM media from Plankton samples (one is a fungal strain (PF1) and the others were bacterial strains (PB1-9) (Figure 6). The obtained strains were then used for screening of cellulase-active strains on CMC media. After growing the strains on CMC media, I stained them with Congo red. As Congo red interacts with (1→4)-β-d-glucans, (1→3)-β-d-glucans and (1→4)-β-d-xylans [23], I was able to find a cellulase activity on CMC media by the formation of a halo zone (not stained by Congo

Red due to the CMC hydrolysis). In fact, I detected four strains showing a clear halo zone around the colonies on the media. Using these strains, I conducted the second screening with the same condition of the first one. All of four strains (two strains (SB4 and SB5) from the seaweed samples and the other two (PF1 and PB4) from plankton samples) formed significant zone of clearance (Figure.7), suggesting that these strains actually have cellulase activities. The four strains obtained were used for further analyses.

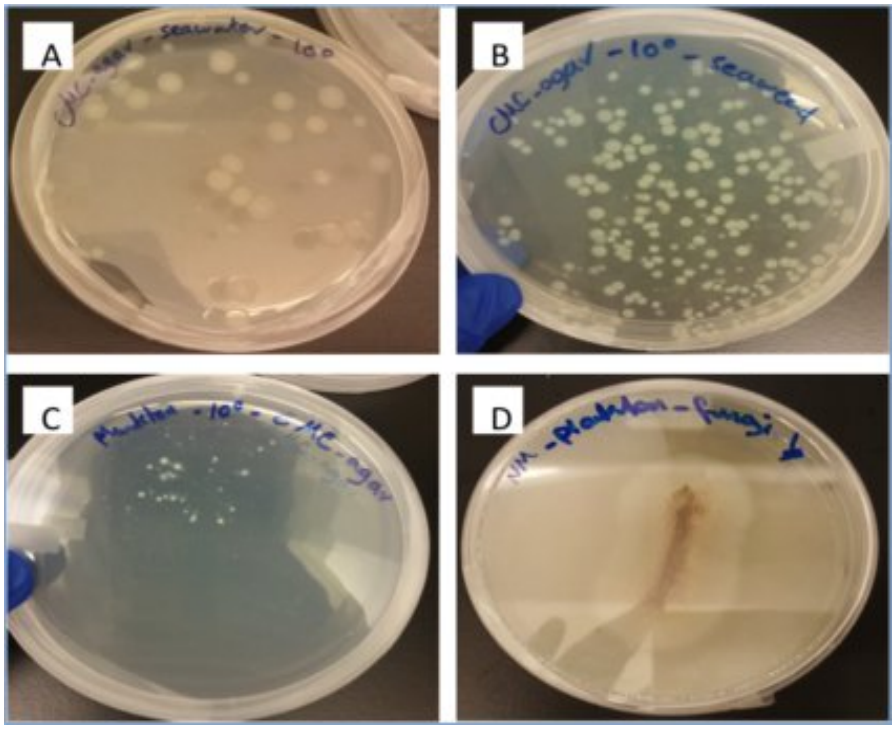


Figure 6. Isolation of microorganisms from different samples. a: Bacterial colonies from seawater. b: Isolates from seaweeds. c: Bacterial isolates From planktons. d: Fungi from plankton.

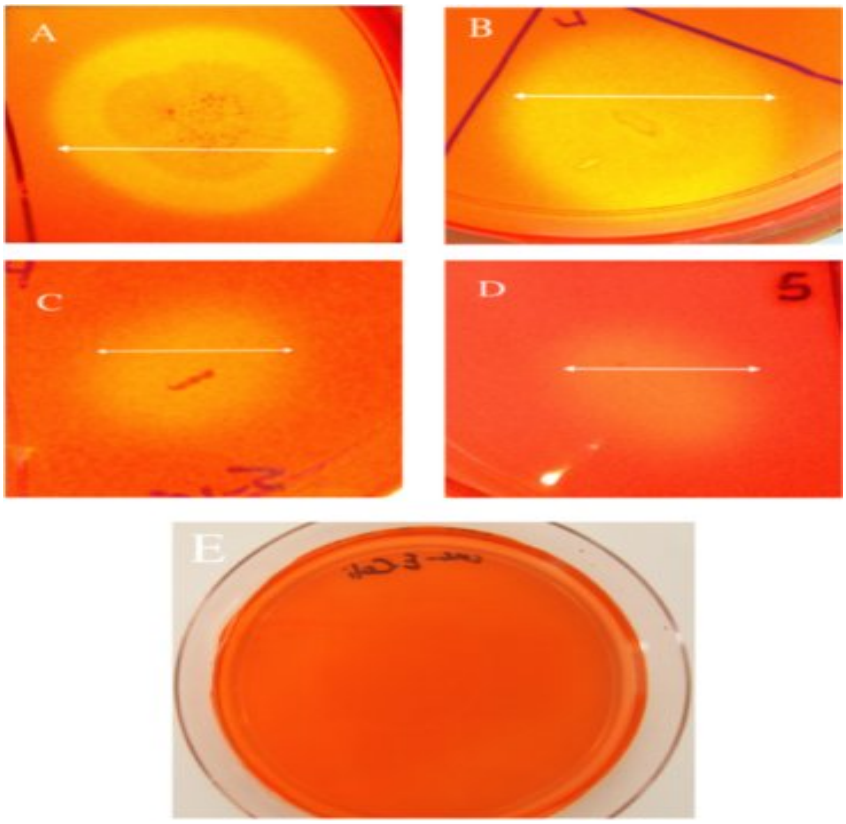


Figure 7. Screening by Congo red of the cellulase producing isolates on CMC agar plate. A: Fungal isolate from plankton (PF1). B: Bacterial isolate from plankton sample (PB4). C and D: Bacterial isolate from the seaweeds (SB4 and SB5). E: *E.coli* negative control.

3.2.2 Taxonomic prediction of cellulase-active strains

To predict taxonomies, I determined the sequences of a small subunit rRNA gene from each isolate. PCR reactions were performed using the primer pair that encompassed about 1500bp of 16S rRNA gene region for bacterial strains, while using the pair targeting around 420 bp of 18S rRNA gene region for a fungal strain. Amplicons with expected sizes were obtained from each strain. After sequencing each of the amplicons, subsequent phylogenetic analyses revealed that all the bacterial strains examined here were identified as Firmicutes strains.

SB5 was suggested to be related to *B. sonorensis, B. licheniformis* and *B. aerius,* whereas PB4 was included in a large group of *B. amyloliquefaciens, B. mojavensis, B. siamensis, B. methylotrophicus* and *B. subtilis*. On the other hand, SB4 was included in the *Staphylococcus* cluster and identified as relatives of *S. waneri* and *S. pasteuri* (Figure.5). The fungal strain PF1 was identified to be an ascomycota, *Aspergillus* species and is closely-related to *A. ustus*. *The* Blastn search showed that the sequence of the strain SB4 showed 99% identity to *Staphylococcus pasteuri,* and SB5 showed 99% to *Bacillus aerius.* Plankton fungal strains PF1, showed identity to the *Aspergillus ustus* isolate by 99%, while PB4 strains showed 99% identity to *Bacillus subtilis* (Figure.8 and 9).

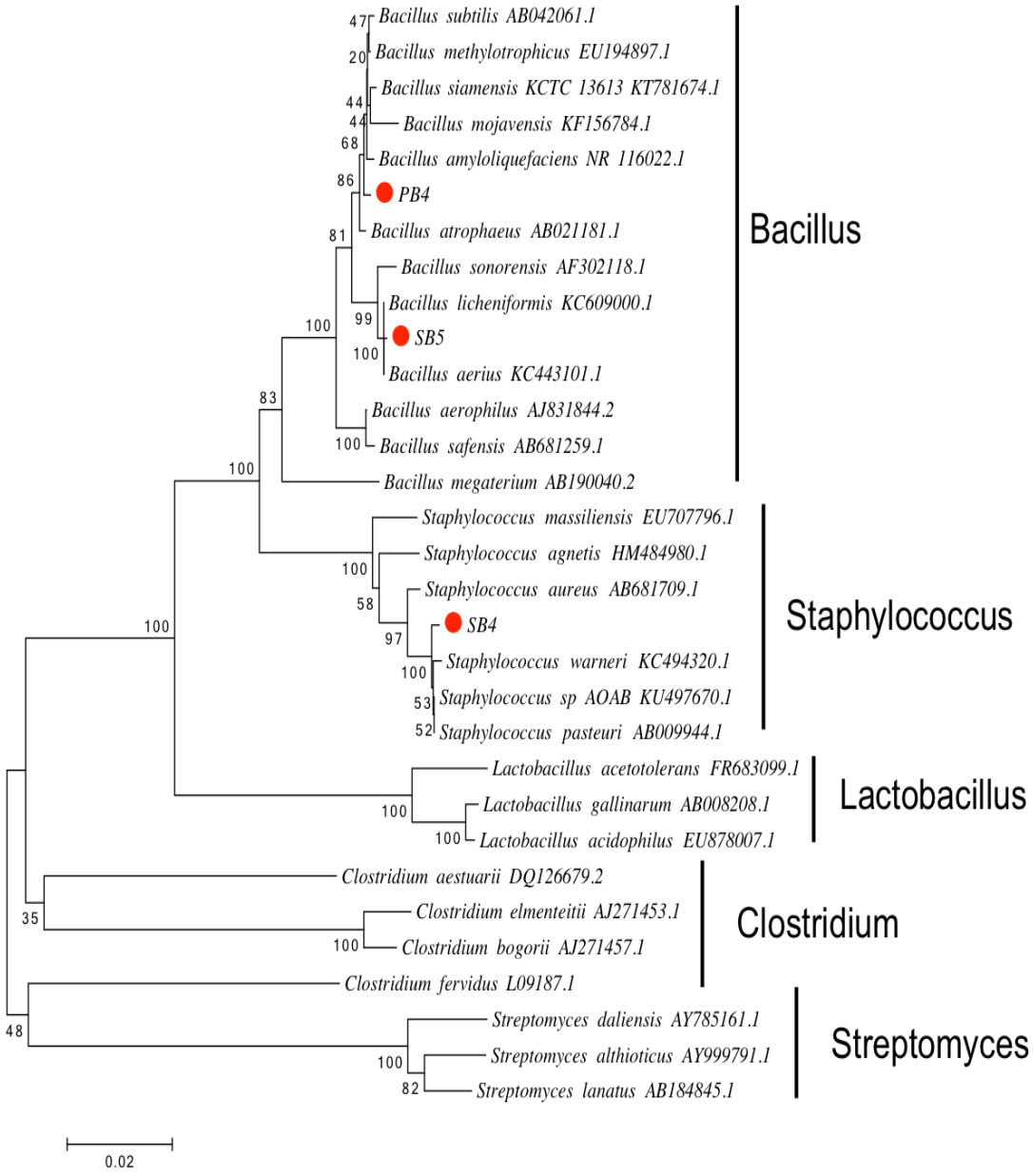


Figure 8. Neighbor-joining phylogenetic tree based on the 16s rRNA of the strains PB4, SB4 and SB5. The Taxa clustered together in the bootstap test (100 replicates). The evolutionary distance bar show the unit of the number of base substitutions per site.

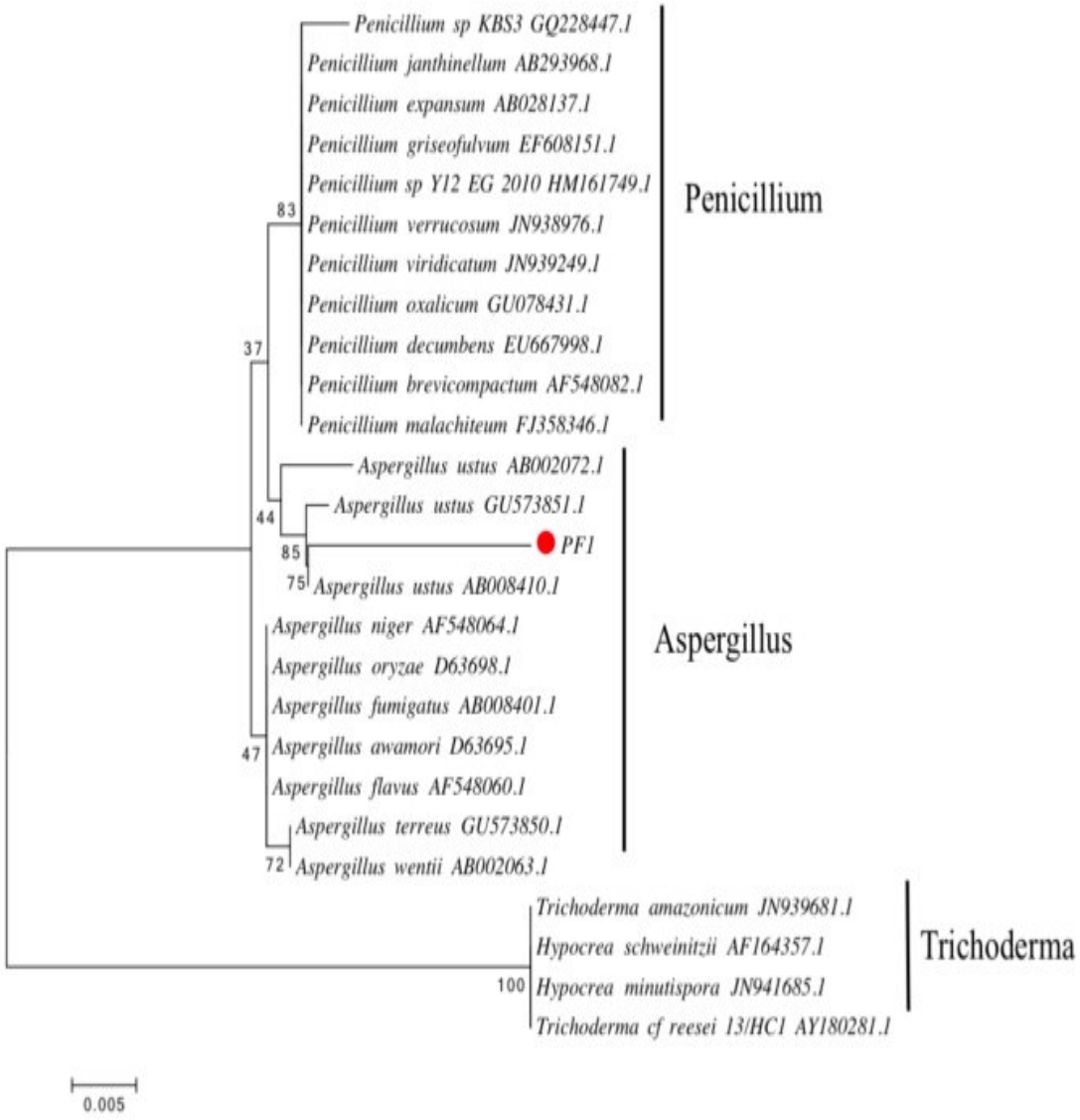


Figure 9. Neighbour-joining phylogenetic tree based on the 18s rRNA of the strains PF1. The Taxa clustered together in the bootstap test (100 replicates). The evolutionary distance bar shows the unit of the number of base substitutions per site.

3.2.3 Enzyme activity assay

The four strains PF1, PB4, SB4 and SB5 were examined for the FPase activity after cultivation in a 100 ml liquid medium that was provided with the Whatman no.1 Filter paper. All media have shown degradation of the filter paper, whereas the control, i.e. the same media cultured without bacterial inoculums showed only a amount of small degradation(Figure.10). It suggested that the degradation of the FP has taken place by the activity of the isolates. The enzymatic analysis using the supernatant of both media showed that cellulase activities reached its peak after 72 hrs in all strains (Figure 11). All the strains have shown similar activities during the four day incubation period [38].

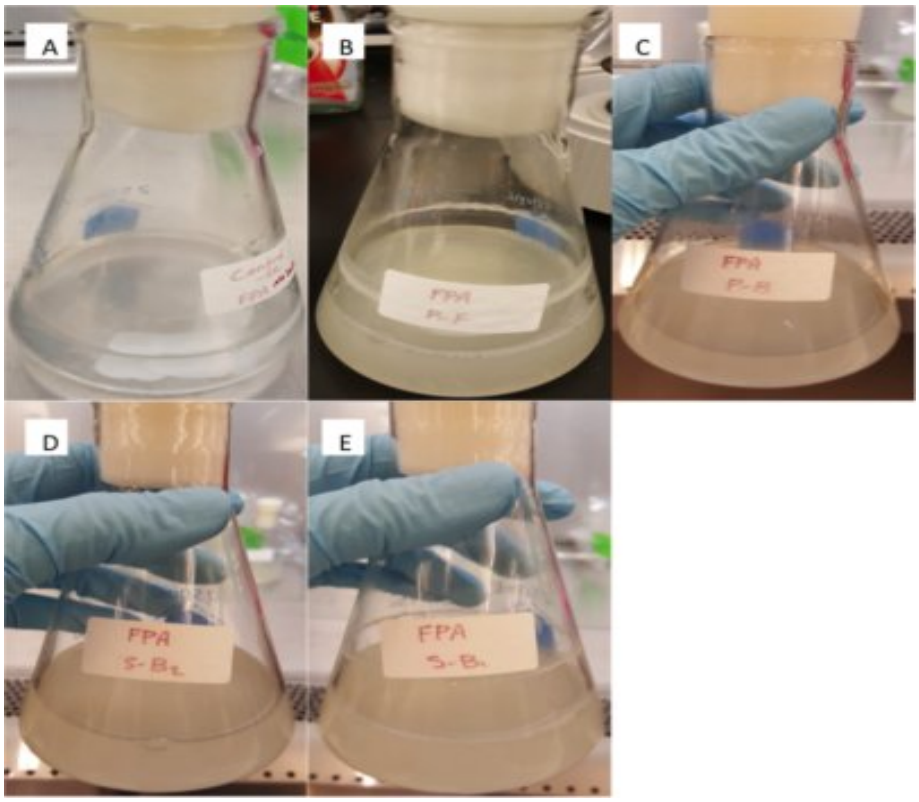


Figure 10. Filter paper assay. The degradation of the Whatman. no.1 filter paper by the four strains PF1(b), PB4(c), SB4(d), B5(e) and negative control (a).

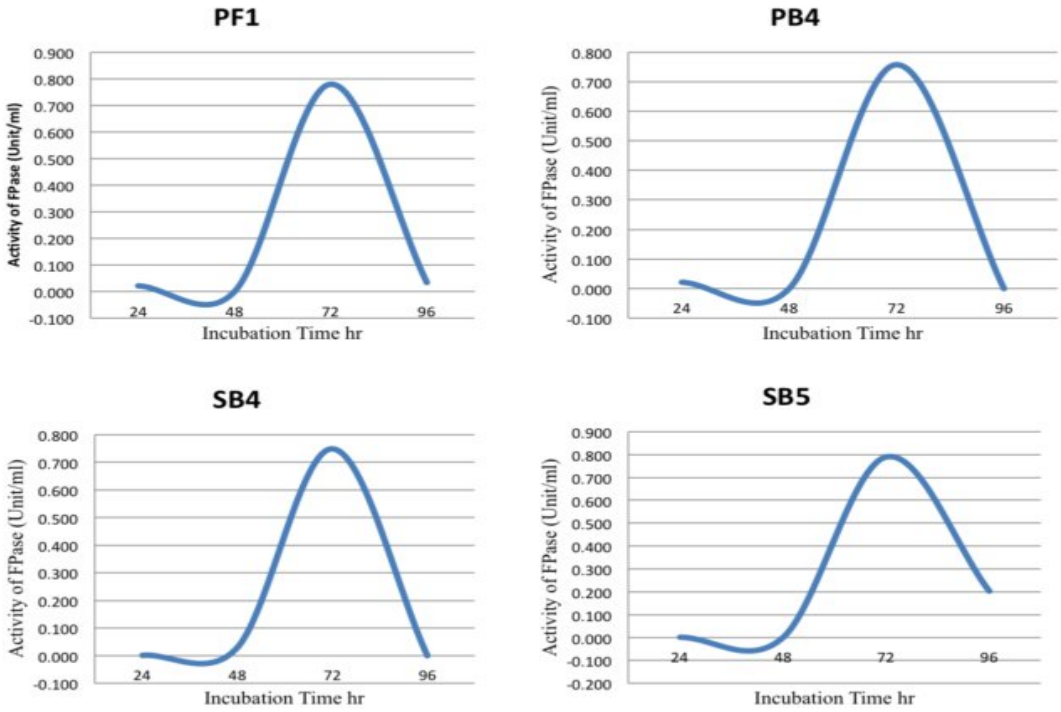


Figure 11. The Filter paper assay activity measurement of the four strains.

3.2.4 Prediction of cellulase genes in isolates

To predict what kind of enzymes the isolates use for cellulolysis, I investigated cellulase genes from the genomes of their relative species. The genomic survey of the *Bacillus licheniformis* WX_02, DSM13 and 5NAP23 strains that are closely related to SB5, revealed that they have 41,43 and 42 cellulase genes which are classified into 11 groups for each genome. The *B. subtillus* strain 168, strain B-1 and strain SPIZIZENII TU-B-10 that are closely related to PB4, have also 22, 24 and 28 cellulase genes of 9 groups. On the other hand, the *S. pasteuri* strain BAB3, SP1 and the *S. warneri* strain SG1 that are relative species of SB4, have only 13, 10 and 14 genes of 6 groups. In the case of fungi, I

found about 131, 136 and 114 cellulase genes of 12 groups in each genome of *A. terreus* NIH2624, *A. oryzae* 3.042 and *A. niger* strains (Table.3-6). These results indicated two possibilities regarding cellulase activities in the obtained isolates. 1) Cellulase activity is the sum of many cellulases expressed in each isolate, 2) cellulase activity is composed of one or a few cellulases that have strong activities. Further investigations are needed to reveal what type of cellulases are actually playing a key role in their glycoside hydrolysis reactions in the Red Sea.

Table 3. Cellulase genes from genomes of relative species of SB4.

Pfam	*Staphylococcus pasteuri strain BAB3 (Firmicutes)*	*Staphylococcus pasteuri SP1 (firmicutes)*	*Staphylococcus warneri SG1 (firmicutes)*
PF00150	6	4	-
PF00232	2	2	5
PF00933	1	-	3
PF01183	2	2	2
PF04616	2	2	2
PF01915	-	-	2
Total number of cellulase genes	13	10	14

Table 4. Cellulase genes from genomes of relative species of SB5.

Pfam	*Bacillus licheniformis WX_02*	*Bacillus licheniformis DSM13*	*Bacillus licheniformis 5NAP23*
PF00150	10	10	10
PF00232	7	7	7
PF00295	2	2	2
PF00759	1	1	1
PF01183	1	1	-
PF01270	1	1	1
PF01915	2	3	3
PF02011	3	2	2
PF04616	12	12	12
PF12891	2	2	2
PF00933	-	2	2
Total number of cellulase genes	41	43	42

Table 5. Cellulase genes from genomes of relative species of PB4.

Pfam	*Bacillus Subtillus 168*	*Bacillus Subtillus strain B-1*	*Bacillus Subtillus SPIZIZENII TU-B-10*
PF00150	5	7	6
PF00232	5	4	7
PF00457	1	1	3
PF00933	1	2	1
PF01915	1	2	1
PF04616	6	2	5
PF12891	4	2	4
PF00759	-	2	-
PF01270	-	2	1
Total number of cellulase genes	22	24	28

Table 6. Cellulase genes from genomes of relative species of PF1.

Pfam	*Aspergillus Terreus NIH2624 Ascomycetes*	*Aspergillus Oryzae 3.042 Ascomycetes*	*Aspergillus Niger Ascomycetes*
PF00150	35	22	19
PF00232	3	1	5
PF00295	9	20	22
PF00331	7	7	2
PF00457	2	4	4
PF00759	1	-	1
PF00933	21	21	19
PF01183	2	1	-
PF01915	22	26	21
PF04616	26	21	13
PF07691	3	5	3
PF12123	-	7	5
Total number of cellulase genes	131	136	114

Chapter 4. Conclusion

In this study, I surveyed cellulase genes using metagenomes of the Red Sea. Since cellulase is a large family containing enzymes of different orthologous groups, classified cellulases based on the results of my current research. I then investigated cellulase genes from the metagenome. This study is the first attempt to evaluate both the abundance and diversity of cellulase genes in various environments.

My cellulase survey revealed that the Red Sea biome has a large amount and wide range of cellulase enzymes. The Red Sea differs from most other oceans by its high salinity, low nutrient concentrations, and high average??? temperatures year-around, allowing us to expect the presence of novel cellulase or cellulase-active organism that may have unique enzymatic or ecological characteristics.

I also isolated cellulase-active microorganisms from the Red Sea. I isolated these strains from plankton and seaweed, that are, cellulose substrates, suggesting that these strains have cellulases functioning in Red Sea environments. By the FPase activity measurement, I actually detected cellulase enzymatic activity from their culturing media. Although further analyses are needed for the identification and enzymatic characterization of cellulase genes in these strains, the present results allow me to expect that the Red Sea is readily available as a microbial or genetic resource of cellulase enzymes.

On the other hand, no cellulase-active microorganisms were isolated from the seawater in this study, though a lot of cellulase genes were found from seawater metagenomes of the

Red Sea. Further examination of the seawater samples are required. In particular, I may change the type of media for isolation. In addition, the use of another method such as a metagenome library based on screening method combined with microfluid system, for example, may also reveal more numbers of novel cellulase genes in the Red Sea.

REFERENCES

1. Horn, S.J., et al., *Novel enzymes for the degradation of cellulose.* Biotechnology for Biofuels, 2012. 5(1): p. 1-13.

2. Koeck, D.E., et al., *Genomics of cellulolytic bacteria.* Current opinion in biotechnology, 2014. 29: p. 171-183.

3. Fukuoka, A. and P.L. Dhepe, *Catalytic conversion of cellulose into sugar alcohols.* Angewandte Chemie International Edition, 2006. 45(31): p. 5161-5163.

4. Teeri, T.T., *Crystalline cellulose degradation: new insight into the function of cellobiohydrolases.* Trends in Biotechnology, 1997. 15(5): p. 160-167.

5. Trivedi, N., et al., *Solid state fermentation (SSF)-derived cellulase for saccharification of the green seaweed Ulva for bioethanol production.* Algal Research, 2015. 9: p. 48-54.

6. Kuhad, R.C., R. Gupta, and A. Singh, *Microbial cellulases and their industrial applications.* Enzyme research, 2011. 2011.

7. Karmakar, M. and R. Ray, *Current trends in research and application of microbial cellulases.* Research Journal of Microbiology, 2011. 6(1): p. 41.

8. Kasana, R.C., et al., *A Rapid and Easy Method for the Detection of Microbial Cellulases on Agar Plates Using Gram's Iodine.* Current Microbiology, 2008. 57(5): p. 503-507.

9. Gunavathy, P. and M. Boominathan, *Isolation and Molecular characterization of Cellulase Producing Bacteria from Soil of Sacred Grove, Puducherry, India.* Int. J. Curr. Microbiol. App. Sci, 2015. 4(12): p. 584-590.

10. Singh, G., A. Verma, and V. Kumar, *Catalytic properties, functional attributes and industrial applications of β-glucosidases.* 3 Biotech, 2016. 6(1): p. 1-14.

11. Sadhu, S. and T.K. Maiti, *Cellulase production by bacteria: a review.* British Microbiology Research Journal, 2013. 3(3): p. 235.

12. Wang, F., et al., *Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches.* Microbiological research, 2009. 164(6): p. 650-657.

13. Jo, K.-I., et al., *Pilot-scale production of carboxymethylcellulase from rice hull by Bacillus amyloliquefaciens DL-3.* Biotechnology and Bioprocess Engineering, 2008. 13(2): p. 182-188.

14. Robson, L.M. and G.H. Chambliss, *Cellulases of bacterial origin.* Enzyme and Microbial Technology, 1989. 11(10): p. 626-644.

15. Galante, Y., A. De Conti, and R. Monteverdi, *Application of Trichoderma enzymes in the food and feed industries.* Trichoderma and Gliocladium, 1998. 2: p. 327-342.

16. Samira, M., R. Mohammad, and G. Gholamreza, *Carboxymethyl-cellulase and filter-paperase activity of new strains isolated from Persian Gulf.* Microbiology Journal, 2011. 1(1): p. 8-16.

17. Das, S., P. Lyla, and S.A. Khan, *Marine microbial diversity and ecology: importance and future perspectives.* Current Science, 2006. 90(10): p. 1325-1335.

18. Ngugi, D.K., et al., *Biogeography of pelagic bacterioplankton across an antagonistic temperature–salinity gradient in the Red Sea.* Molecular ecology, 2012. 21(2): p. 388-405.

19. Healy, F., et al., *Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose.* Applied microbiology and biotechnology, 1995. 43(4): p. 667-674.

20. Ferrer, M., et al., *Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora.* Environmental Microbiology, 2005. 7(12): p. 1996-2010.

21. Kim, S.-J., et al., *Characterization of a gene encoding cellulase from uncultured soil bacteria.* FEMS microbiology letters, 2008. 282(1): p. 44-51.

22. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic acids research, 2014. 42(D1): p. D490-D495.

23. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics.* Nucleic acids research, 2009. 37(suppl 1): p. D233-D238.

24. Eddy, S.R., *Accelerated profile HMM searches.* PLoS Comput Biol, 2011. 7(10): p. e1002195.

25. Oki, S., et al., *SraTailor: Graphical user interface software for processing and visualizing ChIP - seq data.* Genes to Cells, 2014. 19(12): p. 919-926.

26. Schmieder, R. and R. Edwards, *Quality control and preprocessing of metagenomic datasets.* Bioinformatics, 2011. 27(6): p. 863-864.

27. Rosenzweig, M.L., *Species diversity in space and time*. 1995: Cambridge University Press.

28. Chantarasiri, A., *Novel halotolerant cellulolytic Bacillus methylotrophicus RYC01101 isolated from ruminant feces in Thailand and its application for bioethanol production.* KMUTNB: International Journal of Applied Science and Technology, 2014. 7(3): p. 63-68.

29. Patagundi, B.I., C. Shivasaran, and B. Kaliwal, *Isolation and characterization of cellulase producing bacteria from soil.* International Journal of Current Microbiology and Applied Sciences, 2014. 3(5): p. 59-69.

30. Nakada, M., et al., *RFLP analysis for species separation in the generaBipolaris andCurvularia.* Mycoscience, 1994. 35(3): p. 271-278.

31. Saitoh, Y., K. Izumitsu, and C. Tanaka, *Phylogenetic analysis of heavy-metal ATPases in fungi and characterization of the copper-transporting ATPase of Cochliobolus heterostrophus.* Mycological research, 2009. 113(6): p. 737-745.

32. TOPO, T., *Cloning Kit for Sequencing.* Version K. Invitrogen, 35p, 2003.

33. Shareef, I. and S.R. Ashalatha, *Isolation, Screening and Purification of Cellulase from Cellulase Producing Klebsiella variicola RBEB3 (KF036184. 1).*

34. Shareef, I., M. Satheesh, and S.X. Christopher, *Isolation and Identification of Cellulose Degrading Microbes.* International Journal of Innovative Research in Science, Engineering and Technology, 2015. 8(4): p. 6788-6793.

35. Gupta, P., K. Samant, and A. Sahu, *Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential.* International Journal of Microbiology, 2012. 2012.

36. Kim, Y.-K., et al., *Isolation of cellulolytic Bacillus subtilis strains from agricultural environments.* ISRN microbiology, 2012. 2012.

37. Wood, T.M. and K. Bha, *Methods for measuring cellulase activities.* 1988.

38. Thayer, D. and C.A. David, *Growth of "seeded" cellulolytic enrichment cultures on mesquite wood.* Applied and environmental microbiology, 1978. 36(2): p. 291-296.