Characterization and whole genome sequencing of *Klebsiella variicola* HSTU-AAM51 strain isolated from Goat rumen contents: prospects of lignocellulase enzyme production

A THESIS

BY

MD. ABDULLAH-AL-MAMUN

Registration No.: 1805350

Session: 2018-2019

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY

DINAJPUR-5200

DECEMBER, 2019

Characterization and whole genome sequencing of *Klebsiella variicola* HSTU-AAM51 strain isolated from Goat rumen contents: prospects of lignocellulase enzyme production

A THESIS

BY

MD. ABDULLAH-AL-MAMUN

Registration No: 1805350

Session: 2018-2019

Submitted to the Department of Biochemistry and Molecular Biology, Hajee Mohammad Danesh Science and Technology University, Dinajpur-5200In partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE

IN

BIOCHEMISTRY AND MOLECULAR BIOLOGY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY

DINAJPUR-5200

DECEMBER, 2019

Characterization and whole genome sequencing of *Klebsiella variicola* HSTU-AAM51 strain isolated from Goat rumen contents: prospects of lignocellulase enzyme production

A THESIS

BY

MD. ABDULLAH-AL-MAMUN

Registration No.: 1805350

Session: 2018-2019

Approved as to Style and Content By

Supervisor **Dr. Md. Azizul Haque** Assistant Professor Department of Biochemistry and Molecular Biology	Co-Supervisor **Dr. Md. Yeasin Prodhan** Assistant Professor Department of Biochemistry and Molecular Biology

Dr. Md. Abu Sayed
Chairman
Examination Committee and Chairman
Department of Biochemistry and Molecular Biology

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY

DINAJPUR-5200

DECEMBER, 2019

CERTIFICATE

This is to certify that the research work entitled "**Characterization and whole genome sequencing of *Klebsiella variicola* HSTU-AAM51 strain isolated from Goat rumen contents: prospects of lignocellulase enzyme production"** is an original research work and submitted in partial fulfillment of the requirements for the degree of MS in Biochemistry and Molecular Biology. This research work done by **MD. ABDULLAH-AL-MAMUN** bearing student ID No.**1805350** since 2018-2019, under my supervision, as a fellow for the degree of Master of sciences.

Supervisor

Dr. Md. Azizul Haque

Assistant professor

Department of Biochemistry and Molecular Biology

Hajee Mohammad Danesh Science and Technology University

Dinajpur-5200, Bangladesh


Dedicated
To
My Beloved Parents

ACKNOWLEDGEMENTS

All praises are for the almighty Allah for whom the author was able to complete this research work and prepare this manuscript in partial fulfillment of the requirements for the degree of Master of Science in Biochemistry and Molecular Biology from Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur within the appointed time.

The author wishes to convey his cordial respect and pleasant appreciation to his respected supervisor ***Dr. Md. Azizul Haque****, Assistant professor, Department of Biochemistry and Molecular Biology, Hajee Muhammad Danesh Science and Technology University (HSTU), Dinajpur, Bangladesh, for his scholastic supervision, precious suggestion, and continuous encouragement during the entire period of this research work.*

The author is extremely delightful to express his heartfelt respect, and profound appreciation to his benevolent research co-supervisor, ***Dr. Md. Yeasin Prodhan****, Assistant professor, Department of Biochemistry and Molecular Biology, HSTU. for his valuable assistance and indispensable guidance throughout the period of this research work.*

I would like to express my grateful appreciation to ***Dr. Md. Abu Sayed****, Associate Professor and chairman, Department of Biochemistry and Molecular Biology, Hajee Mohammad Danesh Science and Technology University, Dinajpur for his for the help extended to me.*

I would like to state my special thanks to ***Dr. Mst. Nur-E-Nazmun Nahar****, Associate professor, Department of Biochemistry and Molecular Biology, HSTU.*

I would like to convey my respect and thanks to Md. Abdul Wadud, lab instructor in Agri-chemistry department and Md. Tajul Ahmed, lab instructor in Pathology department, Hajee Mohammad Danesh Science and Technology University, Dinajpur for their instrumental support and kind behavior and I would like to state my special thanks to Sharmin Sultana, Syngdha rani Das, Md. Ashikujjaman Ashik, Md. Bokhtiar Rahman, Simo, Shamima yesmin, Goutom Chandra Devnath, Md. Al Amin Hossain, Shahriar Aziz, Md. Shohorab Hossain, Zoherul Hasan for their support in my research.

The Author

ABSTRACT

Due to shortage of fossil fuels in Bangladesh, we need to source alternative energy one of choices may be the lignocellulosic bioethanol production. However the commercial enzymes for bioethanol production are not available in Bangladesh. Therefore lignocellulase producing strains is the target of our research.

The study is aimed to characterize a lignocellose degrading strain and to revealing their mechanism through genome wide investigation. The lignocellulose degrading bacteria was isolated from Goat rumen contents by screening on the CMC, pectin, xylanand starch containing agar media. The isolate was named as *Klebsiella variicola* HSTU-AAM51. The HSTU-AAM51 showed MR negative, VP negative, catalase positive, oxidase positive, MIU test negative, citrate negative, TSI, , sucrose, dextrose, lactose, urease and maltose positive in vitro test. *Klebsiella variicola*. HSTU- AAM51 was isolated from goat rumen contents and confirmed to possess xylanase, cellulose and pectinase activity by whole genome sequence analysis. The complete genome of*Klebsiella vaiicola* HSTU- AAM51 strain comprises of 5,564,045 bp in a circular chromosome with a G + C content of 57.2%. Among the predicted 5187 protein-coding genes(CDS), 45 genes are involved in the degradation of lignocellulose and other polysaccharides, including6cellulose degrading genes (GH3, GH5, GH13, GH94) , 4xylanose degrading genes (GH10,CE4,GH43), 1 mannose degrading gene (GH5), 4 hemicellulose degrading genes (GH27, GH2, GH13, GH53) , 1 pectin degrading gene (PE1) and 29 lignin degrading genes. Lignocellulose degrading genes are detected by computational analysis, where as some CAZy (GH5, GH94, GH10, GH5, CE4, GH53 and PE1) genes showed signal peptide prediction positive, which proves that they have extracellular enzymes activity. Finally lignocelluloses degrading enzyme activity was ensured by molecular docking analysis where ascellulase enzymes (GH3, GH94 and GH5) showed ability to break down ligand respectively **β** -D-glucopyranose, alpha & beta –D-Glucose and cellopentaose which release energy respectively ΔG=-10.51K cal/mol, ΔG=-6.8 K cal/mol and ΔG=-10.3 K cal/mol. In addition toxylanase enzymes (GH10 and CE4) showed ability to break down ligand respectively**β**-D-xylopyranose(XYP),4-O-**β** -D-xylopyranosyl-**β** -D-xylopyranose(BXP) and 2-(N-morpholino)-ethanesulfonic acid(MES), 2-(acetylamino)-2-deoxy-A-D-glucopyranose (NDG) which release energy respectively ΔG=-7.3, -6.3 K cal/mol and ΔG=-7.3, -6.46 K cal/mol. Moreover hemicellulase enzymes (GH53 and GH13) showed ability to break down ligand respectively imidazole(IMD),N-acetyl-D-glucoseamine(NAG), **β**-D-glactose(GAL) and 2-[bis-(2-hydroxy-ethyl)-amino]-2-hydroxymethyl-propane-1,3-diol(BTB)

which release energy respectively ΔG=-9.4, -6.7, -7.0 K cal/mol, ΔG=-7.74 K cal/mol and pectinase enzymes (PE1) has capability to break down ligand **β**-D-galactopyranuronic acid (Pectin) which release energy respectively ΔG=-6.23 K cal/mol. This genome-based analysis facilitates our understanding of the mechanism underlying the biodegradation of lignocelluloses. This genome information including CAZyme repertoire will be useful to understand lingocellulolytic machinery. This study first time reports the whole genome analysis of *Klebsiella variicola* HSTU-AAM51 which are enriched with carbohydrate and lignin degrading useful enzymes for bioethanol production.

CONTENTS

CONTENTS (Contd.)

CONTENTS (Contd.)

CONTENTS (Contd.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Cond.)

LIST OF ABBREVIATIONS

g	:	Gram
ml	:	Millilitre
µl	:	Microlitre
g/L	:	Gram per litre
mg/ml	:	Milligram per millilitre
ng/µl	:	Nanogram per microlitre
pH	:	Negative logarithm of Hydrogen ion concentration
rpm	:	Revolutions per minute
V	:	Volt
U/ml	:	Units per millilitre
U/mg	:	Units per milligram
RNA	:	Ribonucleic acid
DNA	:	Deoxyribonucleic acid
bp	:	Base pair
kb	:	Kilobase pair
NCBI	:	National Centre for Biotechnology Information
BLAST	:	Basic Local Alignment Search Tool
°C	:	Degree Celsius
$(NH_4)2SO_4$	:	Ammonium sulphate
$MgSO_4 \cdot 7H_2O$	:	Magnesium sulphateheptahydrate
KH_2PO_4	:	Monopotassium phosphate
$CaCl_2 .2H_2O$	:	Calcium chloride dihydrate
NaH_2PO_4		Monosodium phosphate
Na_2HPO_4	:	Disodium phosphate
NaOH	:	Sodium hydroxide
HCl	:	Hydrochloric acid
NaCl	:	Sodium chloride
DNS	:	3,5- Dinitrosalicylic acid
nm	:	Nanometre
ANI	:	Average nucleotide identity
RAST	:	Rapid Annotation using Subsystem Technology

KAAS	:	KEGG Automatic Annotation Server
BBH	:	bi-directional best hit
GR	:	Goat Rumen

Alk. $CuSO_4$	:	Alkaline copper sulphate
M	:	Molar
mM	:	Millimolar
N	:	Normal
EDTA	:	Ethylenediaminetetraacetic acid
UV	:	Ultra violet
PCR	:	Polymerase Chain Reaction
RFP	:	Restriction Fragment Length Polymorphism
SDSPAGE SP	:	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis Species
TAE	:	Tris Acetate EDTA
TBE	:	Tris Borate EDTA
KDa	:	KilloDaltone
NA	:	Nutrient Agar media
cm^{-1}		Per centimeter

CHAPTER-I
INTRODUCTION

1.1 Introduction

Lignocellulosic biomass which comprises of cellulose, hemicellulose, lignin and pectin, is the most abundant form of fixed carbon on Earth (109 tons/annum) and its breakdown is a critical component of the global carbon cycle [1]. It is of major interest due to its renewable and sustainable nature as a feedstock to replace fossil fuels for the production of biofuels and chemicals [2]. The use of agricultural residues including rice straw, nut shell, maize stover and oil palm empty fruit bunch (EFB) as feedstocks are favourable as they are not used for food and are not compromising food security [3–5]. Palm oil is the most widely used edible oil and the waste generated at palm oil mills is currently creating a major disposal problem particularly EFB, shells, and fruit palm kernel, in addition to waste produced in the plantations, such as palm trunks and fronds. The generations of large amounts of lignocellulosic waste are mainly in the form of EFB [5]. The valorization of this abundant waste is highly encouraged in countries like B. Many bacteria are able to decompose plant biomass by secreting lignocellulolytic enzymes, including cellulases, hemicellulases, ligninases and pectinases [6–9]. These enzymes have been further classified, based on structure in the Carbohydrate-Active Enzyme database (CAZy), into glycosyl hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs) and auxiliary activities (AAs) [10]. Many microbes have the capability to produce cellulase enzyme which in turn can break up cellulose chains liberating free sugar units (13). Common well-known bacteria such as *Bacillus spp.*, *Brevibacillus* spp., *Cellulomonas* spp., *Streptomyces* spp., *Pseudomonas* spp. and *Klebsiella* spp. have been widely studied for their biomass degrading abilities for biorefining applications [6,11,12].These actions are basically done by three types of Cellulases: Endoglucanases or carboxy methyl cellulases which attacks the internal sites of the amorphous region of cellulose [14,15]; exoglucanases or cellobiohydrlases which can degrade the crystalline structure of cellulose by attacking the reducing or non-reducing ends of polysaccharides chain and finally β-glucosidases which splits cellobiose into two glucose units [16]. Bacteria and fungi are now widely accepted source of cellulase enzyme and this enzyme is now being produced industrially from these sources [13, 14]. Fungal source such as genera of *Trichoderma*, *Aspergillus* [13] etc has already been established as cellulase producers and still an area under research. However despite the fact that bacterial production of cellulase has been scientifically proven for some of the genera such as *Clostridium*, *Cellulomonas*, *Thermomonospora* etc [17], many other

bacterial strains are still to be tested and optimized to get better enzyme production. *Klebsiella sp.* has also recently been reported in various articles to be an efficient cellulase producer and also some research have been carried out to clone the cellulase gene from *Klebsiella* [18]. So, this genera promises a great deal in terms of cellulase production. The deterioration of biomass into bioethanol by means of enzymatic hydrolysis requires several steps including pretreatment, cellulose hydrolysis, separation of sugar solution from lignin, microbial fermentation of sugar solution and distillation [16,19]. The pretreatment process can be classified into physical pretreatment involving milling, irradiation etc, physico-chemical pretreatment involving steam explosion or auto- hydrolysis, chemical pretreatment and finally biological pretreatment involving microbes such as bacteria and fungi [16, 17]. The biological pretreatment using white-rot fungi or *Bacillus sp.* has significant results considering the fact that it requires low energy and mild environmental conditions [20]. The main reason for this step is to break the protective coating that is made by lignin and hemicellulose network and also to lower the order of crystallinity in cellulose [16]. The physical pretreatment allows for easier binding of cellulase enzyme by increasing the surface area of biomass [21, 22]. It also decreases the crystallinity and degree of polymerization of cellulose [21]. Milling which is also a physical pretreatment is basically cutting the lignolocellulosic biomass into smaller pieces just like food chewing for size reduction [21; 23]. Irradiation such as gamma rays and microwaves can essentially increase the rate of enzyme hydrolysis and in combination with other treatments like dilute acid treatment hydrolysis will be more successful [23]. In terms of chemical pretreatment dilute acid treatment is one of the most common method for the pretreatment of lignocellulosic materials [16; 23]. When treating with acid two types of condition can be maintained: High temperature and low acid concentration or low temperature and high acid concentration. This method can actually be used for fermenting the sugars also [23]. Chemical pretreatment basically removes the amorphous lignin and hemicellulase to decrease the crystallinity [24]. Alkaline pretreatment is another example of chemical pretreatment which uses alkaline solutions like caustic soda or lime [25] for withdrawing lignin part and breaking the intermolecular ester bonds between lignin, hemicellulose and cellulose [26,27]. Treating with caustic soda increases the surface area as well as decreases the degree of polymerization of cellulose. However since caustic soda is expensive alternatives such as limes (calcium hydroxide) can be used for the disruption of lignin structure [28].

Biomass can also be pretreated with ozone by maintaining optimum water content in biomass feed which is by standard 30% [29].This process is called ozonolysis and it depends on the sample content, particle size and ozone concentration in the gas flow[29]. However this process itself is too expensive since a large amount of ozone is required [20]. Pretreatment of the lignocellulosic biomass can be done using both physical and chemical means simultaneously. These type of pretreatments are called physico-chemical pretreatment [30]. Now when we are talking about digesting cellulose several factors have to be considered including cellulose crystallinity, degree of polymerization, accessible surface area, particle size, lignin content, hemicellulose content, acetyl content etc [16]. Cellulose digestion depends much on the crystalline structure of cellulose [16]. Different lignocellulosic material varies in crystallinity and their Particle size presents a subsequent challenge in the digestion since increase in particle size would eventually decrease the surface area and it would be harder for the enzyme to attack lignicellulose material. The more the lignin content of a lignocellulosic sample the more it is shielded from the cellulase since lignin has the tendency to attach with cellulase and masks the actual cellulose content. Hemicellulose contain different sugar units but easy to break down because of their low degree of polymerization and amorphous nature [31]. Complex enzymes such as 1,4-β-D- xylaneses, exo-1,4-β-D-xylosidases, endo 1,4 β-D mannases, β-mannosiadse, α-glucornisidase, galactanses, endo arabinanses, α-L-arabinofuranosidases and α-galactosiadses can degrade hemicelluloses [32] .One of the major hemicellulases are xylanases which yield xyloligomers by hydrolyzing β-1,4 bond. An accessory enzyme such as feuroyl esterase aids the release of hemicellulose content from lignin by hydrolyzing ester bonds. So, hemicellulose content does matter in digesting the lignocellulosic biomass [33]. Lignin and hemicellulose are covalently linked and after pretreatment phase less hemicellulose content aids in the digestion of cellulose. Also hemicellulose contains acetylated xylan contents which can intervene with the enzyme recognition and greatly decreases the cellulose digestibility. To promote biodegradation of lignocellulosic biomass, the use of microbial consortia is becoming increasingly popular among scientists since one species can’t compensate for the diverse function that is needed for the multistep process of bioethanol production. In nature microbial community survive by resource complementarity and this is the clue that we have to take and develop unique bacterial communities that provide less complex,

cheaper and efficient way to produce bioethanol and remove the energy problem that's bothering us [13, 17].

1.2 Objectives of the research work:

1. Isolation of bacterial strains from Goat rumen contents.
2. Screening for cellulase producer strains be means of cellulose plate assay
3. Biochemical characterization of the selected isolated cellulose degrading bacteria by means of biochemical tests analysis, morphological and microscopic characteristics.
4. Molecular identification using whole genome sequencing and analysis.
5. To identify glycoside hydrolase (CAZy) genes from the genome.
6. To identify signal peptide prediction and active site determination of CAZy genes.
7. To observe binding energy of identified substrate by using molecular docking

CHAPTER-II

METHOD AND MATERIALS

This section shows all the utilized method for conducting this theses including isolation, biochemical and molecular characterization of cellulose degrading bacterial strains as well as the methods of NGS sequence analysis and identify gene of interest with investigate the mode of action of these genes.

2.1 Place of study

The experiment was designed and executed at Biochemistry and molecular biology research lab in Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur.

2.2 Handling of laboratory apparatus and glassware

All the laboratory equipment that were used in this study including conical flasks, glass pipettes, beakers, glass rods, petri dishes, micropipette tips, eppendorf tubes, falcon tubes, PCR tubes etc had been autoclaved at 121°c in 15 psi for 15 minutes before usage to make sure they were in sterile condition.

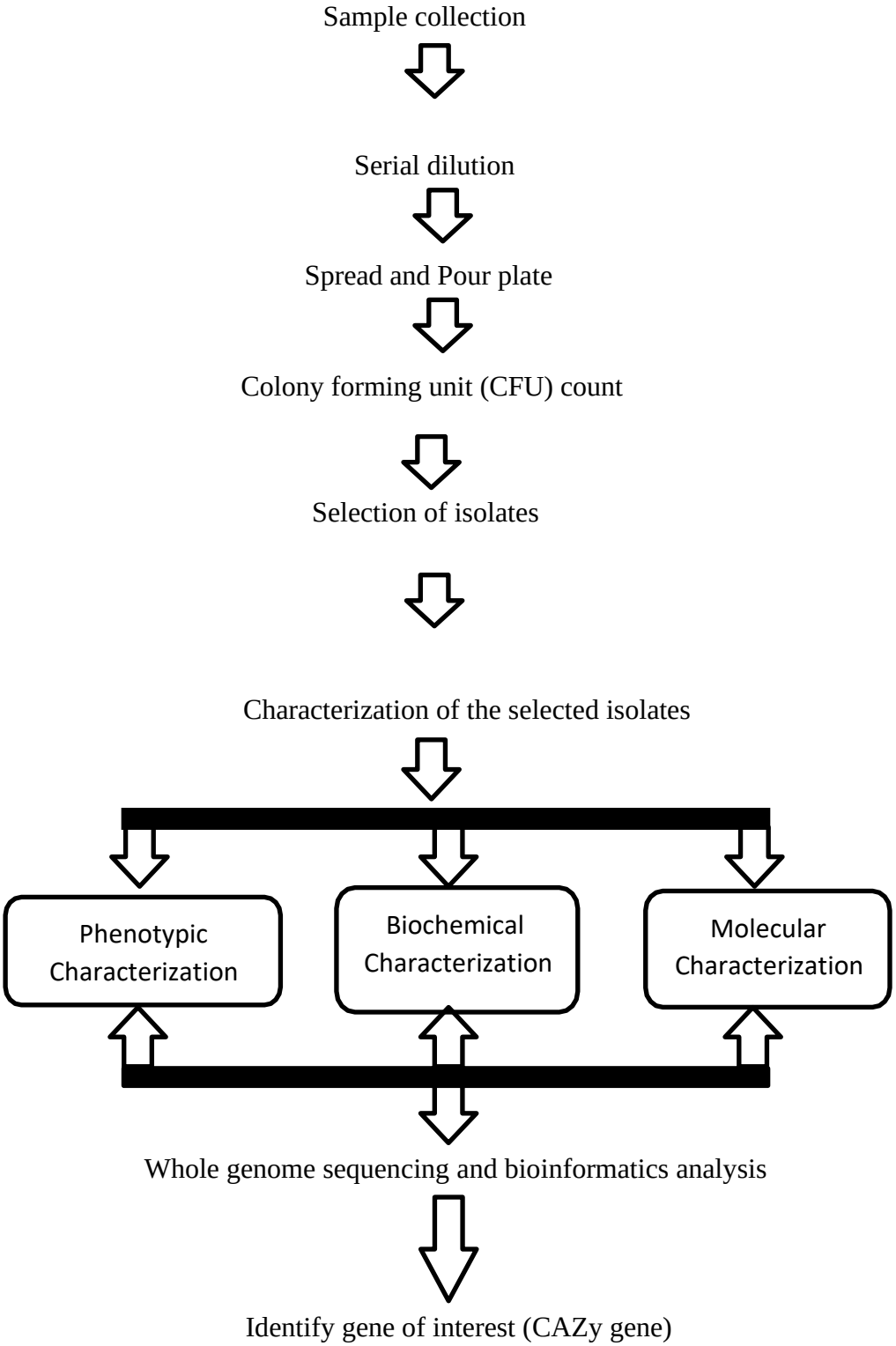
2.3 Solutions and reagents

Required solutions and reagents used in the study were freshly prepared before use. All chemicals needed were acquired from the laboratory. None of the solutions or reagents were further purified since they were of a reagent grade. The list of reagents is given in appendix A.

2.4 Media

In this study different types of media were used for selective growth, enrichment culture, and indication of specific properties. Media preparation and sterilization were done according to the protocol and standard recipe in appendix B.

2.5 Flow chart of the study



2.6 Materials and Methods

2.6.1 Chemicals and glassware

Chemicals and glassware used in this study are presented in Appendix A.

2.6.2 Sample collection

The Goat rumen sample was collected from Faculty of DVM at Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur, Bangladesh and transported to the laboratory. After that the sample underwent serial dilution and spread plate technique was used to transfer the bacteria to nutrient agar media. The plates showing distinct colonies of bacteria were chosen for further work.

2.6.3 Serial dilution

Test tubes containing 9 ml of physiological saline (0.9% NaCl) were autoclaved before use. Initially, 1 ml of fluid sample was mixed with 9 ml of saline water in a test tube in order to dilute 10^{-1} and mixed with 9 ml of saline by repeated pipetting in order to make tenfold dilution. Again, 1 ml from the 10^{-1} test tube was transferred to 10^{-2} labeled test tube and mixed with 9 ml saline solution in it by repeated pipetting. This action was repeated for the test tubes labeled as 10^{-3} to 10^{-6}.

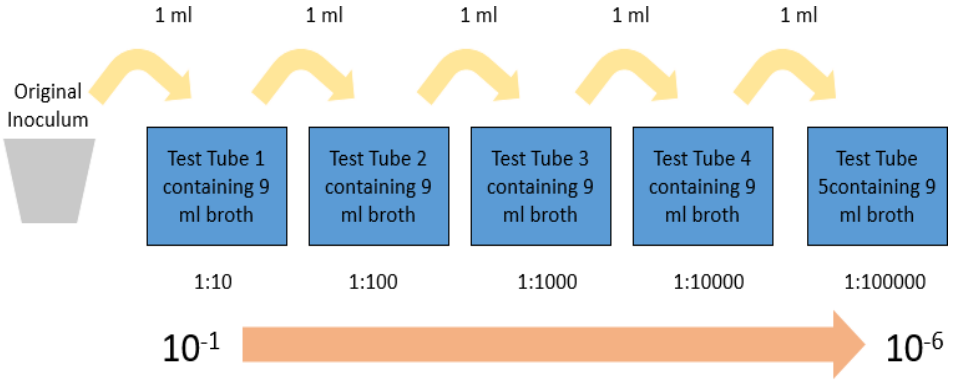


Fig 1: Serial dilution

2.6.4 Isolation of lignocellulase enzyme secreting bacteria

lignocellulose degrading bacteria were isolated and identified from Goat rumen following sequential steps. Appropriate dilutions (10^{-4}) and (10^{-6}) of bacterial samples ware spread on agar plates containing 1% yeast extract, 1.5% Tryptic Soy Broth (TSB), 1.5% agar, pH 7.0 (Gerhardt, 1994). These plates were incubated for 24 h at 37° C. After that bacterial colonies were picked from each plate and streaked on agar plates for further purification. The purified colonies were subjected to congo red agar media (composition: KH_2PO_4 0.5 g, $MgSO_4$ 0.25 g, cellulose 2 g, agar 15 g, Congo-Red 0.2 g, and gelatin 2 g distilled water 1L and at pH: 6.8–7.2) for the rapid and sensitive screening test of cellulase producers by observing clear zones after flooding with 1% congo red and 1M NaCl [34]. The grown colonies that showed discoloration of Congo- Red plate were considered as cellulose degrading bacteria.

2.7 Phenotypic identification of bacterial strain

The potential lignocellulose degrading strains were observed under microscope in order to study their properties.

2.8 Morphological & Biochemical characterization

The selected bacterial strains were differentiated based on their metabolic pattern by using different biochemical tests [35] including catalase, oxidase, MR-VP, TSI, Citrate utilization, Indole, MIU and Individual sugar fermentation tests (dextrose, lactose, maltose and sucrose) (James Cappuccino, 1998). Morphological features were identified by growing the isolated cultures on the minimal nutrient media (composition: $(NH_4)_2SO_4$, 0.1%, KH_2PO_4, 0.05%; K_2HPO_4, 0.05%; $CaCl_2$, 0.01%; NaCl, 0.01%; $MgSO_4.7H_2O$, 0.02%; yeast extract, 0.1%, and cellulose, 0.2%) [35].

2.8.1 Gram staining

The bacterial samples were subjected to gram's staining by using a primary stain crystal violet and a counter stain safranine to distinguish between gram positive and gram negative strains [36]. 95% ethyl alcohol was used as decolorizing agent. This method differentiates between gram positive and gram negative bacteria based on the presence or absence of peptidoglycan layer. Since a gram positive bacterium has a thick peptidoglycan layer it retains the primary stain, however gram negative bacteria retains the counter stain safranine [37].

2.9 Biochemical characterization

2.9.1 Catalase test

The selected bacterial isolates were tested for the presence of catalase enzyme by adding 3% H2O2 into a small amount of bacterial colony on a clean glass slide. A positive catalase test provides rapid production of bubbles indicating catalase utilization by the isolate. A negative result indicates the absence of catalase enzyme. This test identifies members of Enterobacteriaceae family which are catalase positive [35].

2.9.2 Oxidase test

Oxidase test was done on the selected isolates to see if they can utilize cytochrome c oxidase which is basically an enzyme of the electron transport chain [35]. A clean filter paper was placed on a petri dish. A drop of oxidase reagent (1% tetramethyl- p-phenylenediaminedi hydrochloride) was added onto the filter paper. Using an inoculating loop, a small amount of bacterial pure culture of the selected isolates were streaked onto the oxidase droplet. Any formation of deep blue or dark purple color within 10-30 seconds of the reagent was observed. Oxidase positive bacteria are basically aerobic and can use oxygen as a terminal electron acceptor.

2.9.3 Methyl Red-Voges-Proskauer Test/ (MR-VP) Test

Methyl Red (MR) and Voges-Proskauer (VP) broth is utilized for identifying an organism's metabolic pathway. MR is performed to detect the ability of an organism to produce stable acids end products which is achieved if the organism uses mixed acid fermentation pathway. When an organism is MR positive, it creates an acidic environment that will overcome the buffers (Potassium phosphate that resist pH) in the medium. So, after adding methyl red, the broth will stay red. The organisms that gives positive VP result further metabolizes pyruvic acid to form acetyl-methyl carbinol (acetoin). This end product, in the presence of atmospheric oxygen and 40% potassium hydroxide is converted to diacetyl. Diacetyl, under the catalytic action of alpha-naphthol and creatine, is converted into a red complex [38, 39]. MR-VP broth was prepared with autoclaved distilled water and 3.5 ml broth was transferred to each test tubes. After that the selected isolates were inoculated into the each tubes and incubated at 37°C for 24 hours. After incubation, five drops of methyl red reagent was added for detecting MR positive isolates by searching for red colors in the test tubes. For conducting VP test after inoculation and incubation at

37°C for 24 hours 0.6 ml of 5% alpha-naphthol was added to individual test tubes. Following after 0.2 ml of 40% KOH was added and the test tubes were shaken gently in atmospheric oxygen. After that the tubes were allowed to remain for 10-15 minutes, they were observed for a pink-red color development which indicates a positive result. Both MR-VP tests will produce a yellow color on the surface to indicate a negative result.

2.9.4 Triple Sugar Iron (TSI) test

TSI media is a semi-solid media containing three sugars (Lactose, Sucrose, and Glucose) and also iron which is done to confirm the ability of the bacteria to utilize these sugars [40]. After preparing the media 7 ml of the broth was distributed to each test tubes. The media was sterilized by autoclaving at 15 psi, 121°C for 15 min. After sterilization, the media was left to cool by keeping the test tubes in a slanted position in order to form a butt and a slant. The selected isolates were inoculated and the tubes were subjected to Incubation at 37°C for 24 hours. After incubation, the appearance of the media was observed by following below protocol (Table 1) [40]

Table 1: TSI result evaluation parameters

Slant/Butt condition	**Slant color/Butt color**	**Decision taken**	**Gas formation**
Alkaline slant/Alkaline butt	Red/Red	glucose, lactose an sucrose non-fermenter	Depending on the visibility of bubbles or cracks in the test tubes
Alkaline slant/acidic butt	Red/Yellow	Glucose fermentation only	Depending on the visibility of bubbles or cracks in the test tubes
Acidic slant/acidic butt	Yellow/Yellow	glucose, lactose and/or sucrose fermenter	Depending on the visibility of bubbles or cracks in the test tubes
Note: Production of black color means the bacterial isolate can utilize iron and have produced H_2S.			

2.9.5 Citrate Utilization Test (CIU)

The citrate test is utilized to identify the ability of the bacterial isolate to utilize citrate as a carbon and energy source. This test is basically necessitates in the identification of gram negative bacteria based on their metabolic product since gram negative isolates contain citrate-permease that uses citrate present in the medium as a carbon source of energy. This test is also used to distinguish between members of the Enterobacteriaceae family. A positive CIU test will rely on the generation of alkaline by-products of citrate metabolism which will subsequently increase the pH of the medium demonstrated by the color change of a pH indicator. Citrate enters into the cell and is cleaved by citrate lyase releasing oxaloacetate and acetate after which this oxaloacetate is metabolized to pyruvate and CO2. After that CO2 reacts with water and sodium carbonate to release an alkaline by-product. The ammonium salts that are present in the media are also used by the microorganisms which ultimately produces another alkaline product ammonia and turning the media from deep green to intense Prussian blue [41,42].

A sterile glass pipette was used to transfer 10 ml of the media to each of the test tube. After that, the media was left to cool until it had a semi solid consistency. The selected isolates were inoculated in individual test tubes and incubated at 37°C for 24 hours. The intense Prussian blue appearance of the media and the rise in pH above 7.6 (which is due to carbonates and bicarbonates produced as by-products of citrate catabolism) indicates a positive result. Also some bacterial growth will be visible on the surface of the test tube.

2.9.6 Indole test

The indole test screens for the ability of an organism to degrade the amino acid tryptophan by using the enzyme tryptophanase and produce indole. Indole can then react with (p)- dimethyl aminobenzaldehyde (Kovac's reagent) or (p)- dimethyl aminocinnamaldehyde produce a red colored complex [43,44]. Peptone broth was formulated and the test tube containing the media were sterilized by autoclaving at 15 psi. pressure (121°C) for 15 minutes. The selected isolates were inoculated right after and incubated at 37°C for 24 to 48 hours. After incubation 5 drops of was added to each test tubes. The cultures producing the cherry red color ring were indicated as positive and the absence of red color was indicated as a negative result.

2.9.7 Motility Indole Urease (MIU) Test

The MIU test is done for multiple reasons including to detect if the selected isolate can utilize urea, if the organism is mobile or not and if the bacteria can produce indole [36]. The media was prepared using MIU agar medium and 5 ml sterile 40% urea solution per 95 ml basal medium. After that, 6 ml solution was transferred to each sterilized test tube and autoclaved at 15 psi. pressure (121°C) for 15 minutes. The media was left to cool down completely until it had a semi solid consistency by gently placing the tubes in a horizontal manner. The selected isolates were inoculated and incubated at 37°C for 24 hours. The results were interpreted as per table 2

Table 2. MIU result evaluation parameters

Conditions	**Positive**	**negative**
Urea	A color change from yellow to orange to pink-red	No color change
Motility	A cloudy mass of bacteria is seen	Hazy medium without any accumulation
Indole	pink-mauve color on the bottom of the tube	No pink-mauve color is formed

Note: Basically the results for motility and urease are observed first and then the medium is read for indole production. But to be absolutely sure Indole and urease test has been done separately to identify the ability of the selected isolates to use urea and tryptophan.

2.9.8 Urease test

Urease test utilizes the ammonia production from urea by microorganism using urease enzyme. Organisms containing urease hydrolyses urea into ammonia and CO2. With formation of ammonia the media color changes from light orange to magenta pink (pH 8.1) since ammonia alkalines the medium [45, 46].If the organism utilizes urease, the medium will change color indicating positive result due to the presence of phenol red indicator in the urea agar media.

The media was formulated using urea agar base and a pH of 6.8 was maintained. After formulation, 6 ml solution was transferred to each sterilized test tube and autoclaved at 15 psi. pressure (121°C) for 15 minutes. The media was left to cool down completely and after reaching room temperature the selected isolates were inoculated and incubated at 37°C for 24 hours for observation of urease activity.

2.9.9Fermentation of sugar: lactose, maltose, sucrose, dextrose

Specific sugar fermentation test was done using phenol red carbohydrate broth to observe if the selected isolates can utilize these sugars. Different sugars (Lactose, Maltose, Dextrose, and Sucrose) media were made according to protocol provided by the manufacturers and 6ml of the media were transferred into each test tubes. The final pH of 7.4 ± 0.2 was maintained [47]. After that the selected isolates were inoculated and incubated at 37°C for 24 hours. The appearance and color of the media was observed after incubation. After incubation the tubes that turned yellow as a result of acid production by the bacteria were taken as a positive and the ones with no color change was indicated as negative.

2.10 Technical protocol of Whole genome Sequencing

Nextera Transposase Adapters

The transposase adapters are used for Nextera tagmentation.

Read 1 5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Read 2 5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

PCR Primers Index

1 Read 5′ CAAGCAGAAGACGGCATACGAGATGTCTCGTGGGCTCGG

2 Read 5′ AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC

Table 3: Supported Combination

Sequencing Instrument	**Illumina MiniSeq System**
Setup Option	Local Run Manager
Setup Workflow Generate	FASTQ
Library Preparation Kit:	Nextera XT DNA Sample Prep
Indexing:	Dual Indexing
Reagent Kits:	MiniSeq Kit
Analysis Workflow:	Generate FASTQ
Analysis Software:	Local Run Manager

2.10.1 Creating a Run with Local Run Manager

Set Run Parameters

- Log in to Local Run Manager.
- Select **Create Run**, and then select **Generate FASTQ**.
- Enter a run name that identifies the run from sequencing through analysis.
- The run name can contain alphanumeric characters, spaces, and the following special characters
- [Optional] Enter a run description to further identify the run.
- The run description can contain alphanumeric characters, spaces, and the following special characters

Specify Run Settings

- Select the library prep kit from the Library Prep Kit drop-down list.
- Select **2** to specify a dual-indexed run.
- Select to specify the readtype.
- Enter the number of cycles for the run.
- [Optional]For Custom Primers, specify any custom primer information to be used for the run by selecting the appropriate check boxes.
- Custom primer options vary based on your instrument or Local Run Manager implementation.

Enter Samples Manually

- Adjust the samples table to an appropriate number of rows.
- In the Rows field, use the up/down arrows or enter a number to specify the number of rows to add to the table .Select to add the rows to the table.
- Select to delete a row.
- Right-click on a row in the table and use the commands in the contextual menu.
- Enter a unique sample ID in the Sample ID field.
- Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
- [Optional] Enter a sample description in the Description field.

- Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
- Specify an Index 1sequence.
- Select **Show Index Sequence**/**Show Index Names** to toggle between showing the name of the index and the index sequence
- Specify an Index 2sequence.
- Select a manifest file from the Manifest drop-down list.
- [Optional] Enter a project name in the Sample Project field.
- Use alphanumeric characters, dashes, or underscores. Spaces are not allowed in this field.
- [Optional] Select **Export Sample Sheet** to export the sample information in *.csv format. The exported sample sheet can be used as a template, or imported when creating new runs.

2.10.2 Creating Nextera XT DNA Libraries

Tagment Genomic DNA

Preparation

Table 4: Prepare the following consumables:

Item	Storage	Instructions
gDNA	-25°C to -15°C	Thaw on ice. Invert the thawed tubes 3–5 times, and then centrifuge briefly.
ATM	-25°C to -15°C	Thaw on ice. Invert the thawed tubes 3–5 times, and then centrifuge briefly.
TD	-25°C to -15°C	Thaw on ice. Invert the thawed tubes 3–5 times, and then centrifuge briefly.
NT	15°C to 30°C	Check for precipitates. If present, vortex until all particulates are resuspended.

Save the following tagmentation program on the thermal cycler:

- Choose the preheat lidoption
- 55°C for 5minutes

- Hold at10°C

Procedure

- AddthefollowingvolumesintheorderlistedtoeachwellofanewHard-ShellskirtedPCRplate. Pipette tomix.
- TD (10µl)
- Normalized gDNA (0.2 ng/µl per sample) (5µl)
- Add 5 µl ATM to each well. Pipette tomix.
- Centrifuge at 280 × g at 20°C for 1minute.
- Place on the preprogrammed thermal cycler and run the tagmentation program. When the samplereaches10°C,*immediately* proceed to step5 because the transposome is still active.
- Add 5 µl NT to each well. Pipette to mix.
- Centrifuge at 280 × g at 20°C for 1minute.
- Incubate at room temperature for 5minutes.
- The PCR plate contains 25 µl tagmented and neutralized gDNA, all of which is used in the next step.

2.10.3 Amplify Libraries

Preparation

Table 5: Prepare the following consumables:

Item	Storage	Instructions
Index adapters (i5 and i7)	-25°C to-15°C	Only prepare adapters being used. Thaw atroom temperature for 20minutes.
NPM	-25°C to-15°C	Invert each tube to mix. Centrifuge briefly. Thaw on ice for 20minutes.

Save the following program on the thermal cycler:

- Choose the preheat lidoption.
- 72°C for 3minutes
- 95°C for 30seconds
- 12 cycles of:
  - 95°C for 10seconds

- 55°C for 30seconds
- 72°C for 30seconds

➢ 72°C for 5minutes

➢ Hold at10°C

Procedure

1. [24 libraries] Arrange the index adapters in the TruSeq Index Plate Fixture as follows.
    - Arrange Index 1 (i7) adapters in columns 1–6 of the TruSeq Index Plate Fixture.
    - Arrange Index 2(i5)adapter in rows A–D of the TruSeq Index Plate Fixture.

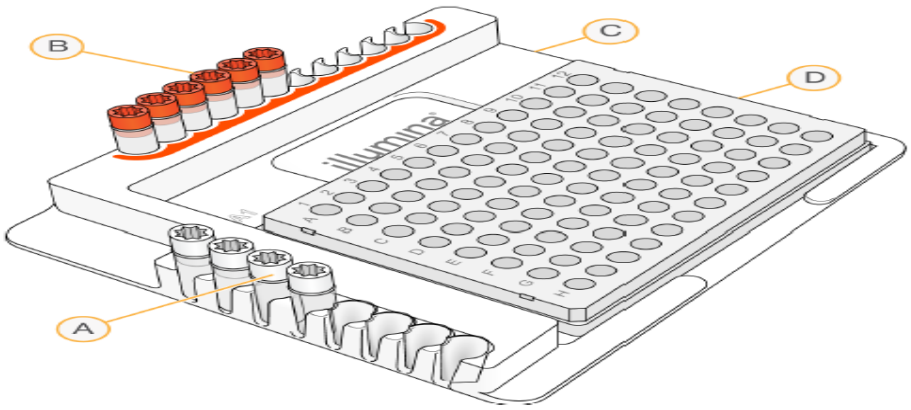


Figure 2: TruSeq Index Plate Fixture Setup for 24Libraries

A. Rows A–D: Index 2 (i5) adapters (whitecaps)
B. Columns 1–6: Index 1 (i7) adapters (orange caps)
C. TruSeq Index Plate Fixture
D. Hard-Shell PCR plat

2. 96 libraries] Arrange the index adapters in the TruSeq Index Plate Fixture asfollows.

- Arrange Index 1 (i7) adapters in columns 1–12 of the TruSeq Index Plate Fixture.
- Arrange Index2 (i5) adapter in rows A–H of the TruSeq Index Plate Fixture.

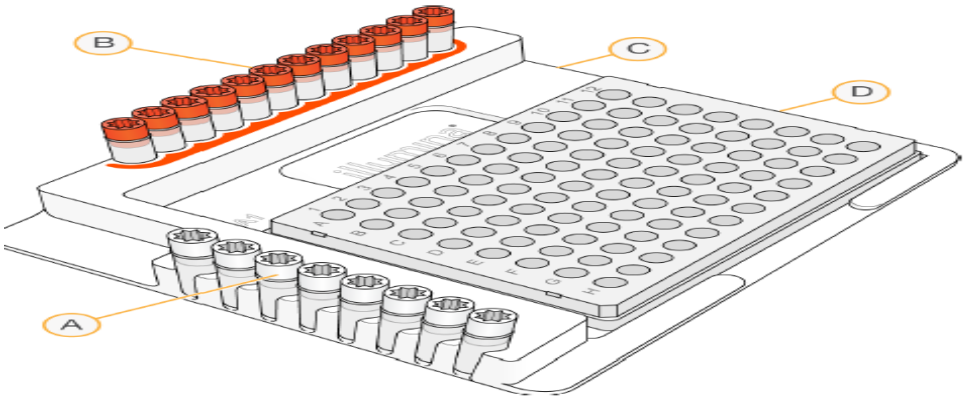


Figure 3: TruSeq Index Plate Fixture Setup for 96Libraries

A. Rows A–H: Index 2 (i5) adapters (whitecaps)
B. Columns 1–12: Index 1 (i7) adapters (orange caps)
C. TruSeq Index Plate Fixture
D. Hard-Shell PCR plate

3. Using a multichannel pipette, add 5µl of each Index 1(i7) adapter down each column. Replace the cap on each i7 adapter tube with a new orange cap.
4. Using a multichannel pipette, add 5µl of each Index2 (i5) adapter across each row. Replace the cap on each i5 adapter tube with a new whitecap.
5. Add 15 µl NPM to each well containing index adapters. Pipette to mix.
6. Centrifuge at 280 × g at 20°C for 1minute.
7. Place on the preprogrammed thermal cycle rand run the PCR program. The volume is 50µl.

2.10.4 Clean Up Libraries

Preparation

Table 6: Prepare the following consumables:

Item	Storage	Instructions
RSB	-25°C to -15°C	Thaw at room temperature. RSB can be stored at 2°C to 8°C after the initial thaw.
AM Pure XP Beads	2°C to 8°C	Let stand on the bench top for 30 minutes to bring to room temperature.

- Prepare fresh 80% ethanol from absolute ethanol.

Procedure

1. Centrifuge at 280 × g at 20°C for 1minute. Transfer 50µl PCR product from each well of the PCR plate to corresponding wells of a new midi plate.
2. Add 30 µl AM Pure XP beads to each well.
3. Smaller amplicons in Nextera XT library preps typically yield smaller insert size ranges. To maximize recovery of smaller fragments from the bead cleanup step, use the following conditions.

Input Size (bp)	AM Pure XP Recommendation	AM Pure XP Volume (µl)
300–500	1.8x AMPure XP	90
> 500	0.6x AMPure XP	30
	(0.5x AMPure XP for ≥ 2 x 250 cycles)*	(25 µl for ≥ 2 x 250 cycles)*
gDNA or other genomic input	0.6x AMPure XP	30

*Applicable only to the MiSeq™ or HiSeq™ 2500 using HiSeq Rapid v2 reagents.

4. Shake at 1800 rpm for 2minutes.
5. Incubate at room temperature for 5minutes.
6. Place on a magnetic stand and wait until the liquid is clear (~2minutes).
7. Remove and discard all supernatant from each well.

8. Wash two times as follows.
    a. Add 200 µl fresh 80% EtOH to each well.
    b. Incubate on the magnetic stand for 30seconds.
    c. Remove and discard all supernatant from each well.
9. Using a 20 µl pipette, remove residual 80% EtOH from each well.
10. Air-dry on the magnetic stand for 15minutes.
11. Remove from the magnetic stand.
12. Add 52.5 µl RSB to each well.
13. Shake at 1800 rpm for 2minutes.
14. Incubate at room temperature for 2minutes.
15. Place on a magnetic stand and wait until the liquid is clear (~2minutes).
16. Transfer 50 µl supernatant from the midi plate to a new Hard-Shell PCR plate.

2.10.5 Check Libraries

Run 1µl of undiluted library on an Agilent Technology2100 Bioanalyzer using a High Sensitivity DNA chip. The following figure shows example traces of libraries successfully sequenced on a HiSeq 2500 system. Typical libraries show a broad size distribution of ~250–1000 bp, as shown in the top panel. Various libraries can be sequenced with average fragment sizes as small as 250 bp or as large as 1500 bp.

2.10.6 Normalize Libraries

Note

Manually normalize libraries when the final library yield is less than 10–15 nM. Bead-based normalization on low yield libraries can result in overly diluted samples and low sequencing yields. For more information, see Best Practices for Standard and Bead-Based Normalization in Nextera XT DNA Library Preparation Kits (Pub.No. 470-2016-007).

Preparation

Table 7: Prepare the following consumables:

Item	Storage	Instructions
LNA1	-25°C to-15°C	Prepare under a fume hood. Bring to room temperature. Use a 20°C to 25°C water bath as needed.
LNB1	2°C to8°C	Bring to room temperature. Use a 20°C to 25°C water bath as needed.
LNW1	2°C to 8°C	Bring to room temperature. Use a 20°C to 25°C water bath as needed.
LNS1	Room temperature	Bring to room temperature.

Procedure

1. Transfer 20 µl supernatant from the Hard-Shell PCR plate to a new midi plate.
2. Add 44µl LNA1 per sample to a new 15ml conical tube. Calculate about 5% extra sample to account for sample loss due to pipetting. For example: for 96 samples, add 4.4 ml LNA1 to the tube (100 samples × 44 µl = 4.4 ml).
3. Thoroughly resuspend LNB1. Pipette to mix.
4. Transfer 8µl LNB1 per sample (including the 5% extra) to the15ml conical tube containing LNA1. Invert to mix. For example: for 96 samples, transfer 800 µl LNB1 to the tube of LNA1 (100 samples × 8 µl = 800 µl).
5. Pour the bead mixture into a trough.
6. Add 45 µl combined LNA1 and LNB1 to each well containing libraries.
7. Shake at 1800 rpm for 30minutes.
8. Place on a magnetic stand and wait until the liquid is clear (~2minutes).
9. Remove and discard all supernatant from each well.
10. Wash two times as follows.
    a. Add 45 µl LNW1 to each well.
    b. Shake at 1800 rpm for 5minutes.
    c. Place on a magnetic stand and wait until the liquid is clear (~2minutes).
    d. Remove and discard all supernatant from each well.
11. Add 30 µl 0.1 N NaOH to each well.
12. Shake at 1800 rpm for 5minutes.

13. During the 5 minute elution, label a new 96-well PCR plate SGP for storage plate.
14. Add 30 µl LNS1 to each well of the SGP plate. Set aside.
15. After the 5 minute elution, make sure that all samples in the midi plate are resuspended. If they are not, resuspend as follows.
    a. Pipette to mix or lightly tap the plate on the bench.
    b. Shake at 1800 rpm for 5minutes.
16. Place on a magnetic stand and wait until the liquid is clear (~2minutes).
17. Transfer the supernatant from the midi plate to the SGP plate.
18. Centrifuge at 1000 × g for 1minute.

Note

After denaturation, the libraries are single-stranded DNA, which resolves poorly on an agarose gel or Bioanalyzer chip. For quality control, use the double- stranded DNA saved from step 16 of the cleanup procedure

2.10.7 Pool Libraries

Preparation

To prepare for the sequencing run, begin thawing reagents according to the instructions for our instrument. If the SGP plate was stored frozen at -25°C to- 15°C, thaw at room temperature. Pipette to mix.

Procedure

1. Centrifuge at 1000 × g at 20°C for 1minute.
2. Label a new Eppendorf tube PAL.
3. Transfer 5 µl of each library from the SGP plate to the PAL tube. Invert to mix.
4. Store unused pooled libraries in the PAL tube and SGP plate at -25°C to -15°C for up to 7 days.

2.10.8 Denaturing and Diluting Libraries and PhiX Control for Sequencing

2.10.8.1 Denature and Dilute Libraries Prepare Hybridization Buffer

1. RemovethetubeofHybridizationBufferfrom-25°Cto-15°Cstorageandthawatroom temperature.
2. When thawed, store at 2°C to 8°C until you are ready to dilute denatured libraries.

3. Vortex briefly before use.

Prepare Incubator

Preheat the incubator to 98°C.

Dilute Library to Loading Concentration

1. Combine the following volumes of pooled libraries and prechilled hybridization buffer in a micro-centrifuge tube.

Table 8: Dilute Library to Loading Concentration

Library Pool	Prechilled Hybridization Buffer
2 µl	998 µl
3 µl	997 µl
4 µl	996 µl
5 µl	995 µl

The total volume is 1 ml.

2. Vortex briefly and then centrifuge at 280 × g for 1minute.
3. Transfer 250 µl diluted library to a new micro-centrifuge tube.
4. Add 250 µl prechilled Hybridization Buffer.
5. Vortex briefly and then centrifuge at 280 × g for 1minute.

Denature Diluted Library

1. Place the tube on the preheated incubator for 2minutes.
2. Immediately cool on ice.
3. Leave on ice for 5minutes.

2.10.8.2 Denature and Dilute PhiX Control

Dilute PhiX to 4 nM

1. Thaw a tube of 10 nM PhiX stock.
2. Combine the following volumes in a micro-centrifuge tube.
    a. 10 nM PhiX (10µl)
    b. RSB (15µl)

The total volume is 25 µl at 4 nM.

3. Vortex briefly and then pulse centrifuge.

Note

[Optional] Store the 4 nM PhiX at -25°C to -15°C for up to 3 months.

2.10.8.3 Denature PhiX

1. Combine the following volumes in a micro-centrifuge tube.
    a. 4 nM PhiX (5µl)
    b. 0.1 N NaOH (5µl)
2. Vortex briefly and then pulse centrifuge.
3. Incubate at room temperature for 5minutes.
4. Add 5 µl 200 mM Tris-HCl, pH7.0.
5. Vortex briefly and then centrifuge at 280 × g for 1minute.

2.10.8.4 Dilute Denatured PhiX to Loading Concentration

1. Add 985µl of prechilled Hybridization Buffer to the tube of denatured PhiX. The total volume is 1 ml at 20 pM.
2. Dilute the denatured 20 pM PhiX to 1.8 pM as follows.
    a. Denatured PhiX (45µl)
    b. Prechilled Hybridization Buffer (455µl)
    The total volume is 500 µl at 1.8pM.
3. Invert to mix and then centrifuge at 280 × g for 1minute.

Note

Stored the denatured 1.8 pM PhiX at -25°C to -15°C for up to 2 weeks. After 2 weeks, cluster numbers tend to decrease.

2.10.8.5 Combine Library and PhiX Control

Table 9: Combine the following volumes.

PhiX Control and Library	Volume
Denatured and diluted PhiX control	5 µl
Denatured and diluted library	500 µl

Set aside on ice until you are ready to load it onto the reagent cartridge

Note

The library and PhiX mixture provides a PhiX spike-in of 0.5%–2.0%. Actual PhiX percentage varies depending upon the quality and quantity of the library pool.

2.10.8.6 Preparing Sequencing Consumables

Prepare the Reagent Cartridge

1. Remove the reagent cartridge from -25°C to -15°Cstorage.
2. Thaw reagents using the following water bath options. Do not submerge the cartridge. When thawed, dry the base before proceeding.

Table 10: Preparing Sequencing Consumables

Method	Time to Thaw	Stability Limit
37°C water bath	35 minutes	Up to 2 hours
Room temperature water bath (19°C to 25°C)	90 minutes	Up to 24 hours

If multiple cartridges are thawing in the same water bath, allow additional thawing time. Alternatively, thaw reagents using the following options.

Method	Time to Thaw	Stability Limit
Room temperature air (19°C to 25°C)	5 hours	Up to 24hours
Chilled at 2°C to 8°C	18 hours	Up to 72hours

3. Invert the cartridge 5 times to mix reagents.
4. Inspect the large reservoirs from the bottom of the cartridge to make sure that reagents have thawed and the reservoirs are free of ice crystals.
5. Gently tap on the bench to reduce air bubbles.

Prepare the Flow Cell

1. Remove a new flow cell package from 2°C to 8°Cstorage.
2. Set the unopened flow cell package aside at room temperature for 30minutes.

Note

Avoid repeated cooling and warming of the flow cell.

3. Remove the flow cell container from the foil package.
4. Put on a new pair of powder-free gloves.
5. Grip the flow cell by the plastic cartridge and remove it from the container.
6. Clean the glass surface of the flow cell with a lint-free alcohol wipe.
7. Dry with a lint-free lens cleaning tissue. Use care around the black flow cell gasket.
8. Inspect the flow cell ports for obstructions. Make sure that the gasket is well-seated

2.10.9 Performing a Run on the MiniSeq

1. Load Libraries onto the Reagent Cartridge

2. Clean the foil seal covering reservoir **#16** labeled **Load library here** using a low-lint tissue.
3. Pierce the seal with a clean 1 ml pipette tip.
4. Add 500µl prepared 1.8pM libraries into reservoir#16.Avoid touching the foil seal as you dispense the libraries.

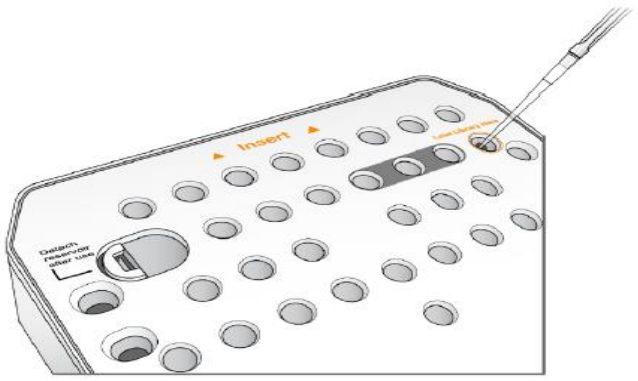


Figure 4: Load Libraries

2.10.9.1 Set Up a Run

From the Home screen, select **Sequence**.

The Sequence command releases consumables from a previous run and opens the series of run setup screens.

Log into Local Run Manager

- Enter your user name and password.
- Select **Next**.

Select Available Run

- Select a run name from the list of available runs.
- Use the up and down arrows to scroll through the list or enter a run name in the Search field.
- Select **Next**

2.10.9.2 Load the Flow Cell

- Open the flow cell compartment door.
- Press the release button to the right of the flow cell latch.
- Open Flow Cell Latch
- If present, remove the used flow cell from a previous run.
- Make sure that the flow cell stage is clean. If debris is present, clean the flow cell stage with an alcohol wipe.
- Place the flow cell on the flow cell stage over the alignment pins.

Figure 5: Place Flow Cell on Stage

- Close the flow cell latch to secure the flow cell.
- Close the flow cell compartment door.

2.10.9.3 Load the Reagent Cartridge

1. Open the reagent compartment door.
2. If present, remove the used reagent cartridge.

Note

To facilitate safe disposal of unused reagent containing form amide, the reservoir in position #9 is removable.

3. Slide the reagent cartridge into the reagent compartment until the cartridge stops.

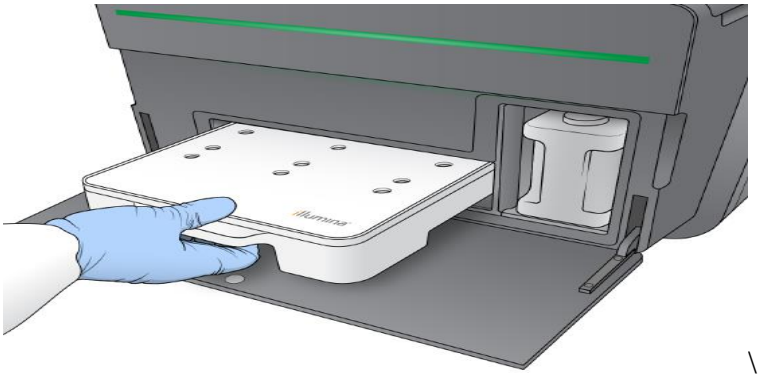


Figure 6: Load Reagent Cartridge

2.10.9.4 Confirm Run Parameters

1. Confirm run parameters.

The control software confirms the number of specified cycles using the following criteria:

- Totalcyclesdonotexceedthemaximumcyclesallowedbasedonthereagentcartridgeloaded for therun.
- Cycles for Read 1 are greater than the 5 cycles required for templategeneration.
- Index Read cycles do not exceed Read 1 and Read 2cycles.

Note

Make sure that you specify the appropriate number of Index Read cycles for the libraries that you are sequencing. For more information, see the library prep documentation.

2. [Optional] Select **Edit** to change run parameters. When finished, select **Save**.
    - **Purge consumables for this run**—Change the setting to purge consumables automatically after the current run.
    - **Run parameters**—Change the read type or number of cycles per read.
    - **Custom primers**—Change the settings for custom primers.
3. Select **Next**.

Note

Do not open the reagent compartment door or the flow cell compartment door during the automated check or during the sequencing run.

4. When the automated check is complete, select **Start**.

2.10.9.5 Review Automated Check

1. Review the automated check results.
    - To stop a check in progress, select **Cancel**.
    - For any items that do not pass, an action is required before you can proceed.
    - To restart the check, select **Retry**. The check resumes at the first incomplete or failed check.
2. To start the run, select from the following options.
    - If the system is not configured to start automatically after a successful check,

select **Start**.

- If the system is configured to start automatically after a successful check, the sequencing run begins automatically. However, if any errors occur during the check, the run does not begin automatically.

2.10.9.6 Monitor Run Progress

1 Monitor run progress, intensities, and quality scores as metrics appear on the screen.

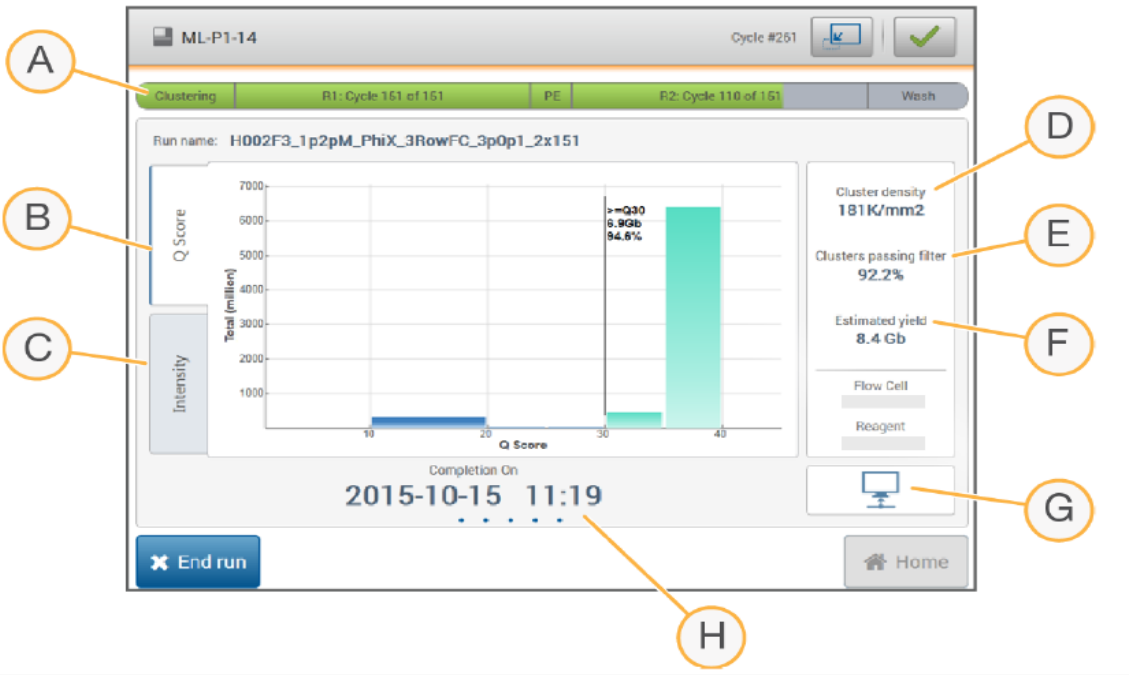


Figure 7: Sequencing Run Progress and Metrics

A. **Run progress**—Shows the current step and number of cycles completed for each read. The progress bar is not proportional to the run rate of each step.

B. **Q-Score**—Shows the distribution of quality scores (Q-scores).

C. **Intensity**—Shows the value of cluster intensities of the 90th percentile for each tile. Plot colors indicate each base: red is A, green is C, blue is G, and black is T.

D. **Cluster density (K/mm²)**—Shows the number of clusters detected for the run.

2.10.9.7 Viewing Analysis Results in Local Run Manager

1. From the Local Run Manager dashboard, select the run name.

2. From the Run Overview tab, review the sequencing run metrics.
3. To change the analysis data file location for future requeues of the selected run, select the **Edit** icon, and edit the output run folder file path. The file path leading up to the output run folder is editable. The output run folder name cannot be changed.
4. Select the **Copyto Clipboard**icon to copy the output run folder file path.
5. Select the Sequencing Information tab to review run parameters and consumables information.
6. Select the Samples & Results tab to view the analysis report.
    - If analysis was requited, select the appropriate analysis from the Select Analysis drop-down list.
    - From the left navigation bar, select a sample ID to view the report for another sample.
7. Select the **Copyto Clipboard**icon to copy the Analysis Folder file path.

2.10.10 Analysis Report

Indexing

Most Popular Index Sequences

Table 10 : Indexing Table

Column Heading	Description
Index Number	An assigned ID based on the order that samples are listed in the sample table.
Sample ID	The sample ID provided when the run was created.
Sample	The sample description provided when the run was created.
Description	The Index 1 adapter used with the sample.
Index 1 (i7)	
Index 2 (i5)	The Index 2 adapter used with the sample.

2.10.10.1 Analysis Output Files

The following analysis output files are generated for the Generate FASTQ analysis module

Table 11: Most Popular Index Sequences Table

Column Heading	Description
Sequence	Shows the sequence of the 100 most popular index sequences.
Reverse Complement	Shows the reverse complement of the original sequences.
Hit Counts	Lists the number of times the sequence was counted.

Table 12: Output Files

File Name	Description
Demultiplexing (*.demux)	Intermediate files containing demultiplexing results.
FASTQ (*.fastq.gz)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step.

2.10.11 Genome analysis using by Basespace Illumina

2.10.11.1 FastQC analysis

Most sequencers will generate a QC report as part of their analysis pipeline, but this is usually only focused on identifying problems which were generated by the sequencer itself. FastQC aims to provide a QC report which can spot problems which originate either in the sequencer or in the starting library material. FastQC can be run in one of two modes. It can either run as a stand alone interactive application for the immediate analysis of small numbers of FastQ files, or it can be run in a non-interactive mode where it would be suitable for integrating into a larger analysis pipeline for the systematic processing of large numbers of files.

It provides the length of the shortest and longest sequence in the set. If all sequences are the same length only one value is reported %GC: The overall %GC of all bases in all sequences. It seems to be optimum range of %GC (40 to 70) [48].

2.10.11.2 ASSEMBLY IN SPAdes (V. 3.9.0)

(1) Stage 1 (assembly graph construction) is addressed by every NGS assembler and is often referred to as de Bruijn graph simplification. We propose a new approach to assembly graph construction that uses the multisized de Bruijn graph, implements new bulge/tip removal algorithms, detects and removes chimeric reads, aggregates biread information into distance histograms, and allows one to backtrack the performed graph operations [49].

(2) Stage 2 (k-bimer adjustment) derives accurate distance estimates between k-mers in the genome (edges in the assembly graph) using joint analysis of distance histograms and paths in the assembly graph.

(3) Stage 3 constructs the paired assembly graph, inspired by the PDBG approach.

(4) Stage 4 (contig construction) was well studied in the context of Sanger sequencing. Since NGS projects typically feature high coverage, NGS assemblers generate rather accurate contigs (although the accuracy deteriorates for SCS) [49]. SPAdes constructs DNA sequences of contigs and the mapping of reads to contigs by backtracking graph simplifications

2.10.11.3 Prokka Genome Annotation (V. 1.11.0)

The Prokka Genome Annotation BaseSpace App compliments our suite of de novo assembly tools. Prokka expects preassembled genomic DNA sequences in FASTA format [50]. Finished sequences without gaps are the ideal input, but it is expected that the typical input will be a set of scaffold sequences produced by de novo assembly software. This sequence file is the only mandatory parameter to the software. It was used de novo assembly FASTA files as an input that can be generated by the SPAdes Genome Assembler.

Output:

- **.gff** This is the master annotation in GFF3 format, containing both sequences and annotations. It can be viewed directly in Artemis or IGV
- **.gbk** This is a standard Genbank file derived from the master .gff. If the input to prokka was a multi-FASTA, then this will be a multi-Genbank, with one record for each sequence
- **.fna** Nucleotide FASTA file of the input contig sequences
- **.faa** Protein FASTA file of the translated CDS sequences
- **.ffn** Nucleotide FASTA file of all the annotated sequences, not just CDS
- **.sqn** An ASN1 format "Sequin" file for submission to Genbank. It needs to be edited to set the correct taxonomy, authors, related publication etc
- **.fsa** Nucleotide FASTA file of the input contig sequences, used by "tbl2asn" to create the .sqn file. It is mostly the same as the .fna file, but with extra Sequin tags in the sequence description lines
- **.tbl** Feature Table file, used by "tbl2asn" to create the .sqn file
- **.err** Unacceptable annotations - the NCBI discrepancy report
- **.log** Contains all the output that Prokka produced during its run. This is a record of what settings you used, even if the --quiet option was enabled
- **.txt** Statistics relating to the annotated features found

2.10.11.4 GENIUS Metagenomics (V.1.1.0)

The GENIUS® Metagenomics application uses CosmosID's acurated genome database and high performance algorithms to provide rapid, accurate, and actionable bacterial identification at the species, subspecies, and/or strain level.

GENIUS identifies multiple bacteria in a whole genome shotgun metagenomic sample and is able to discriminate pathogens from their 'near neighbors' using statistical and computational methods, with no prior assumptions as to what is present in the sample. GENIUS uses raw, unassembled reads as input and matches the sequence against our curated reference database of known pathogens and microbes.

2.10.11.5 Bacterial Analysis Pipeline (V.1.0.4)

The Bacterial Analysis Pipeline app initially predicts the species of bacterial input genomes using a kmer-based approach . Futher, acquired antimicrobial resistance genes are identified using a BLAST-based approach, where the nucleotide sequence of the input genome is compared to the genes in the ResFinder database . Depending on the identified species, Multilocus Sequence Typing (MLST) is performed, also using a BLAST-based approach [51, 52]. If the input genome is recognised as belonging to Enterobacteriaceae or the gram positive bacteria, BLAST is used to search for plasmid replicons using the PlasmidFinder database (4). Identified plasmids of the incF, IncH1, IncH2, IncI1, IncN, or IncA/C type are further subtyped by plasmid MLST (4). Finally, identified *Escherichia coli*, *Enterococcus* spp., *Staphylococcus aureus*, and *Listeria* are compared to the Virulence Finder database containing known virulence genes.

2.10.11.6 Annotating a Genome Using RAST (Rapid Annotation using Subsystem Technology)

The basic steps used to annotate a genome using RAST are described in the subsections below. Input to the process is a prokaryotic genome in the form of a set of contigs in FASTA format. As described below, the actual RAST server will allow a user to specify a set of gene calls, but in the usual case RAST will make its own calls.The service is freely available for the annotation

of prokaryotic genomes. The genomes may be "complete" or they may be in hundreds of contigs (which does impact the quality of the derived annotations). A new user must *register* for the service, which involves giving us contact information and acquiring a password. By registering users, we can create a framework in which users have access to only those genomes that they have submitted. It allows us also to contact the user once the automatic annotation has finished or in case user intervention is required. After login the user can monitor his/her submitted job/ jobs on the Job Overview page. After the annotation is complete the user can choose to download the annotated genome in a variety of export formats (e.g. GenBank, FASTA, GFF3, Excel) or browse the genome in the comparative environment of the SEED Viewer without having the data actually installed in the SEED[53].

The SEED-Viewer environment presents the user with a variety of options for the immediate analysis of the annotated genome. The Organism Overview page contains basic information on the Genome such as Taxonomy, Size, The number of Contigs, the Number of Coding Sequences and RNAs and counts of non-hypothetical and hypothetical gene annotations. In addition it contains the Number of Subsystems that were automatically determined to be present in the genome[53].

2.10.11.7 KEGG Automatic Annotation Server (KAAS)

KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways [54].

The KAAS provides three views of the analyzed data. ‘KO list’ is the flat list of query genes with the K numbers given by the KAAS. ‘KO hierarchy’ is the hierarchical list of annotated genes, which is categorized according to the BRITE database. ‘Pathway map’ is the list of pathways with links to graphical pathway maps. The annotated query genes are highlighted in the maps. Each box in the map is linked to functional information in the KO database. The user can re-compute the KO annotation with a different BLAST threshold from ‘threshold change’ option. ‘Download’ is an option to download the text file of KO annotation and reconstructed graphical pathway maps.

The KAAS is implemented using two methods: the bi-directional best hit (BBH) information method, and the single-directional best hit information method (SBH method). A complete set

of genes in a genome is preferable as the query because the KAAS works best with BBH method. On the other hand, the SBH method can also be used for a limited number of ORFs or ESTs [54]. The computation time of the BBH method is about twice that of SBH.

Bi-directional hit rate

Given a genome to be annotated, it is compared against each genome in the reference set of the KEGG GENES database by BLAST searches in both forward and reverse directions, taking each gene in genome A as a query compared against all genes in genome [55]. Those BLAST hits with bit scores less than 60 are removed. Because the bit scores of a gene pair a and b from two genomes A and B, respectively, can be different in forward and reverse directions, and because the top scores do not necessarily reflect the order of the rigorous Smith–Waterman scores, we define the BHR as:BHR ¼ Rf _ Rr*:*

Here, R¼So/Sb where So is the bit score of a against b, and Sb is the score of a against the best-hit gene in genome B (which may not necessarily be b). Rf refers to the score from the forward BLAST (A against B), and Rr refers to the score from the reverse BLAST (B against A). We select those genes whose BHR is greater than 0.95.

2.10.11.8 Average nucleotide identity (ANI)

Average nucleotide identity (ANI) is a category of computational analysis that can be used to define species boundaries of Archaea and Bacteria. Calculating ANI usually involves the fragmentation of genome sequences, followed by nucleotide sequence search, alignment, and identity calculation. The original algorithm to calculate ANI used the BLAST program as its search engine. An improved ANI algorithm was developed to accommodate the concept of orthology. Here compared four algorithms to compute ANI, namely ANIb (ANI algorithm using BLAST), ANIm (ANI using MUMmer), OrthoANIb (OrthoANI using BLAST). By comparing values to the ANIb that is considered a standard, OrthoANIb and OrthoANIu exhibited good correlation in the whole range of ANI values. ANIm showed poor correlation for ANI of <90% [56].

For average nucleotide identity based on MUMmer (ANIm) calculation, the dnadiff script of the MUMmer tool version 3.0 is used instead of a self-written parser used by the original JSpecies stand-alone implementation.

2.10.11.9 Venn diagram analysis

Venn Diagram is a diagram used in set theory, showing the distribution of the number of elements between 3 sets (4 or more sets can be drawn) It is elements belonging to each set are shown in the circle, and overlapping parts are shown as common elements. It is used for biological applications to illustrate the number of orthologous or unique genes between closely related genomes. In Venn diagram analysis usually used Genebank, fasta file format for genomic comparison. This analysis was done by (http://bioinformatics.psb.ugent.be/webtools/Venn/) web base software [57].

2.10.11.10 Pangenome analysis

The CGView Server generates graphical maps of circular genomes that show sequence features, base composition plots, analysis results and sequence similarity plots. Sequences can be supplied in raw, FASTA, GenBank or EMBL format. Additional feature or analysis information can be submitted in the form of GFF (General Feature Format) files. The server uses BLAST to compare the primary sequence to up to three comparison genomes or sequence sets. The BLAST results and feature information are converted to a graphical map showing the entire sequence, or an expanded and more detailed view of a region of interest. Several options are included to control which types of features are displayed and how the features are drawn [58]. The CGView Server can be used to visualize features associated with any bacterial, plasmid, chloroplast or mitochondrial genome, and can aid in the identification of conserved genome segments, instances of horizontal gene transfer, and differences in gene copy number. Because a collection of sequences can be used in place of a comparison genome, maps can also be used to visualize regions of a known genome covered by newly obtained sequence reads. The CGView Server can be accessed at http://stothard.afns.ualberta.ca/cgview_server/ [59]. GC skew was defined as the normalized excess of C over G in a given sequence, (C −□G)/(C + G), which is calculated with sliding windows along the genome. GC skew is defined to be 0 when the amount of C equals that of G. To eliminate the use of window slides, cumulative skew can be calculated as the cumulative sum of the walker graph score at each nucleotide position along

the genome, with scores A = 0, T = 0, G = 1, and C = −1 [60]. In this work, however, the cumulative GC skew was calculated by taking the cumulative sum of the GCskew in each of the windows, to normalize the cumulative skew strength without it being affected by the length of the genome. GC skew = (G − C)/(G + C)

2.10.12 Signal peptide prediction analysis

The graphical output from SignalP (neural network) comprises three different scores, *C*, *S* and *Y*. Two additional scores are reported in the SignalP3-NN output, namely the *S-mean* and the *D-score*, but these are only reported as numerical values [61].

For each organism class in SignalP; Eukaryote, Gram-negative and Gram-positive, two different neural networks are used, one for predicting the actual signal peptide and one for predicting the position of the signal peptidase I (SPase I) cleavage site. The *S-score* for the signal peptide prediction is reported for every single amino acid position in the submitted sequence, with high scores indicating that the corresponding amino acid is part of a signal peptide, and low scores indicating that the amino acid is part of a mature protein.

The *C-score* is the ``cleavage site'' score. For each position in the submitted sequence, a C-score is reported, which should only be significantly high at the cleavage site. Confusion is often seen with the position numbering of the cleavage site. When a cleavage site position is referred to by a single number, the number indicates the first residue in the mature protein, meaning that a reported cleavage site between amino acid 26-27 corresponds to that the mature protein starts at (and include) position 27.

Y-max is a derivative of the C-score combined with the S-score resulting in a better cleavage site prediction than the raw C-score alone. This is due to the fact that multiple high-peaking C-scores can be found in one sequence, where only one is the true cleavage site. The cleavage site is assigned from the Y-score where the slope of the S-score is steep and a significant C-score is found [61].

The *S-mean* is the average of the S-score, ranging from the N-terminal amino acid to the amino acid assigned with the highest Y-max score, thus the S-mean score is calculated for the length of the predicted signal peptide. The S-mean score was in SignalP version 3.0 used as the criteria for discrimination of secretory and non-secretory proteins [61].

The *D-score* is introduced in SignalP version 3.0 and is a simple average of the S-mean and Y-max score. The score shows superior discrimination performance of secretory and non-secretory proteins to that of the S-mean score which was used in SignalP version 1 and 2. For

non-secretory proteins all the scores represented in the SignalP3-NN output should ideally be very low.

2.10.13 Enzymatic Active site determination

At first the template was determined using by SWISS-MODEL algorithms to ensure the active site of the identified amino acid sequence [62] .Those template was generated in PDB format .After identification of the amino acid residues (PDB) responsible for substrate delivery, binding and orientation in the active site. Substrate binding, as a rule, takes place in the so-called structural pockets and cavities on the protein's surface. Various amino acid residues forming the active site area are not involved directly into the catalytic machinery but interact with the substrate's functional groups, while diffusion and orientation ensure a productive binding and reactive conformation of the enzyme-substrate complex. The CASTp structural analysis algorithm can be used to complete this step [63]. Then Pymol program was used for structural analysis and visualization of active site [64].

2.10.14 Molecular docking analysis

A ligand can be selected either by specifying its identifier from the ZINC database or by uploading structure files (mol2). The former possibility allows users who are not familiar with 3D structure files to start a docking assay with only a ZINC accession code (AC). The latter allows uploading several ligands at once or uploading ligands that are not present in the ZINC database. As for the target protein, SwissDock not only supports the widely used Mol2 format[65], but also the direct upload of CHARMM input files describing the ligand: a coordinate file (PDB), extra topology (RTF) and parameter files (PAR). The ligand is immediately set up after its definition and its prepared structure can be downloaded [66]. The structure of the target protein, as well as that of the ligand, can be automatically prepared for docking. In addition, the cumbersome syntax of the docking engine is hidden behind a clean web interface providing reasonable alternative sets of parameters as well as sample input files. All calculations are performed on the server side, so that docking runs do not require any computational power from the user. The interpretation of docking results and their integration into existing research pipelines is greatly facilitated by the seamless visualization of docking predictions in the UCSF Chimera molecular viewer, which can be launched directly from the web [65].

CHAPTER-III

RESULT AND DISCUSSION

3.1 Isolation of lignocellulase enzyme secreting bacterial

Lignocellulose degrading bacteria were collected and screened from Goat rumen (GR) contents. The appropriate serial dilutions of samples were made. The bacteria acquired from 10^{-4} and 10^{-6} of Goat rumen (GR) contents dilutions were selected and inoculated on cellulose, xylan, pectin agar plates. Isolated bacterial colonies with lignocellulose degrading activity were further screened to obtain the pure culture [34]. The colony number 1, 8,16,15,18 and 24 from GR 10^{-6} showed lignocellulase secreting activity and colony number were named as xylan-1, xylan-8, xylan-16, CMC-15, CMC-18 and Pectin-24, respectively.

The xylan-1, xylan-8, xylan-16, CMC-15, CMC-18 and Pectin-24 produced clear zone of hydrolysis when it was flooded with 1% congo red and 1M NaCl respectively due to hydrolysis of lignocellulose.

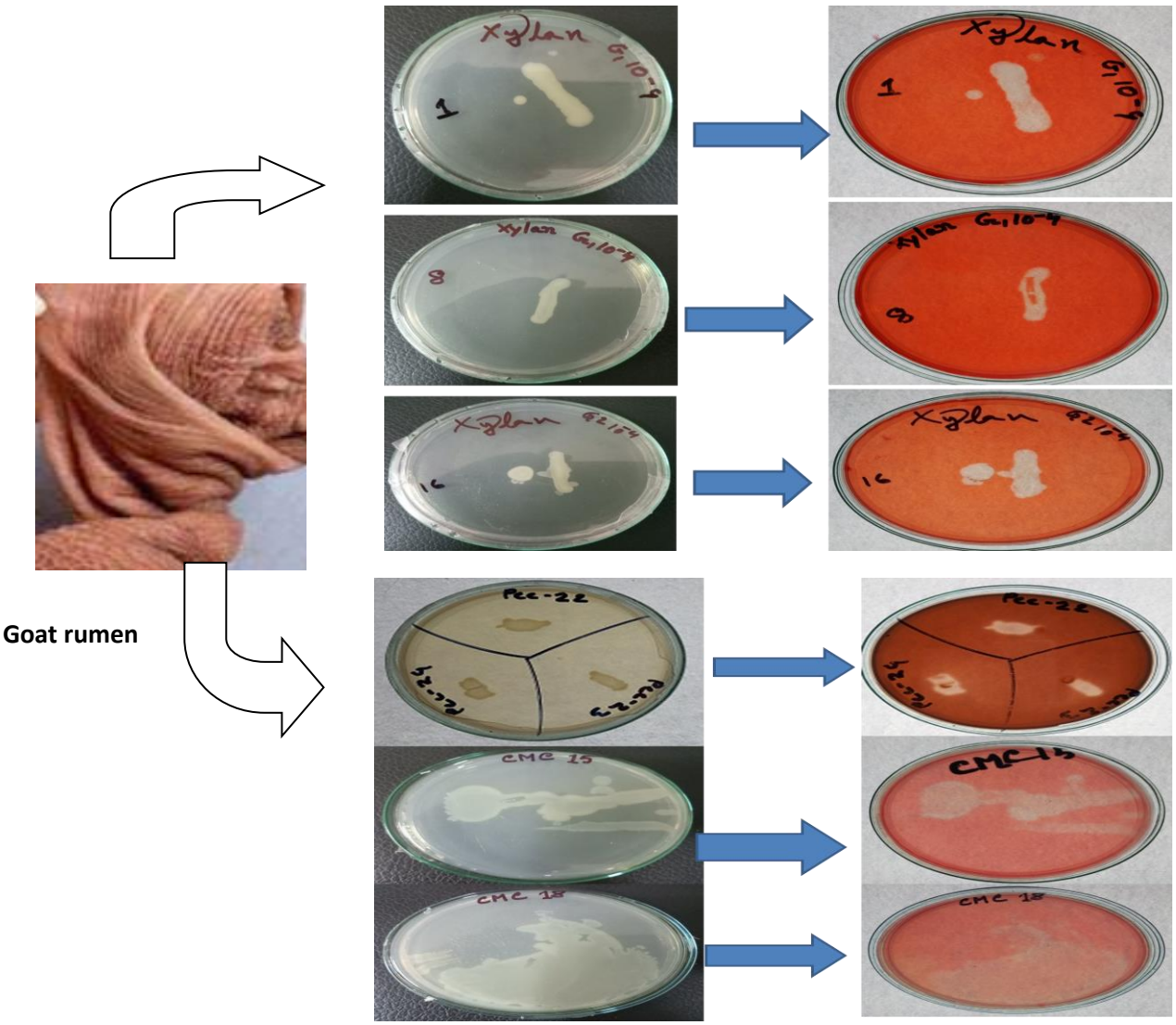


Fig 8: Screening of lignocellulase enzyme secreting bacteria from Goat rumen (GR) contents xylan-1, xylan-8, xylan-16, CMC-15, CMC-18 and Pectin-24.

3.2 Biochemical test

Biochemical test is performed to know some characteristic of bacteria and for the identification of different types of bacteria according to the references that are referred by biochemist.

Several testes were done for biochemical characterization to reach this goal. These are:

1. Oxidase test
2. Catalase test
3. Motility Indole Urease (MIU) test
4. Indole test
5. Triple sugar iron test (TSI) test
6. Citrate Utilization Test
7. MR-VP test
8. Carbohydrate utilization test

The results are presented in Table 2 were a positive result is represented by (+)Ve and a negative result is represented by (-)Ve.

Table 13: Biochemical tests performed and results

Sample Name	MIU	Indol test	TSI	Citrate	MR	VP	Oxidase	Catalase	Dextose	Maltose
Xylan-1	**+++**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(--)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**
Xylan-8	**++**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(--)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**
Xylan-16	**+++**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**
CMC-15	**++**	**(-)Ve**	**(+)Ve**	**(--)Ve**	**(+)Ve**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**
CMC-18	**++**	**(-)Ve**	**(+)Ve**	**(--)Ve**	**(+)Ve**	**(-)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**
Pectin-24	**__**	**(-)Ve**	**(+)Ve**	**(-)Ve**	**(-)Ve**	**(-)Ve**	**(-+Ve**	**(+)Ve**	**(+)Ve**	**(+)Ve**

3.2.1 Catalase test

The catalase test facilitates the detection of catalase enzyme in bacteria through media. Catalase enzyme produced by microorganisms that live in oxygenated environments to neutralize toxic forms of oxygen metabolites such as H_2O_2..The catalase enzyme in bacteria decomposes hydrogen peroxide to oxygen and water. It is a very important enzyme in protecting the cell from oxidative damage by reactive oxygen species. Identified Strain no.CMC,15,18, Xylan-1,8,16 and Pectin-24 showed lots of bubble formation on each slide (Fig 9). So, It indicates the positive result after addition of 30% H_2O_2 on the slide.

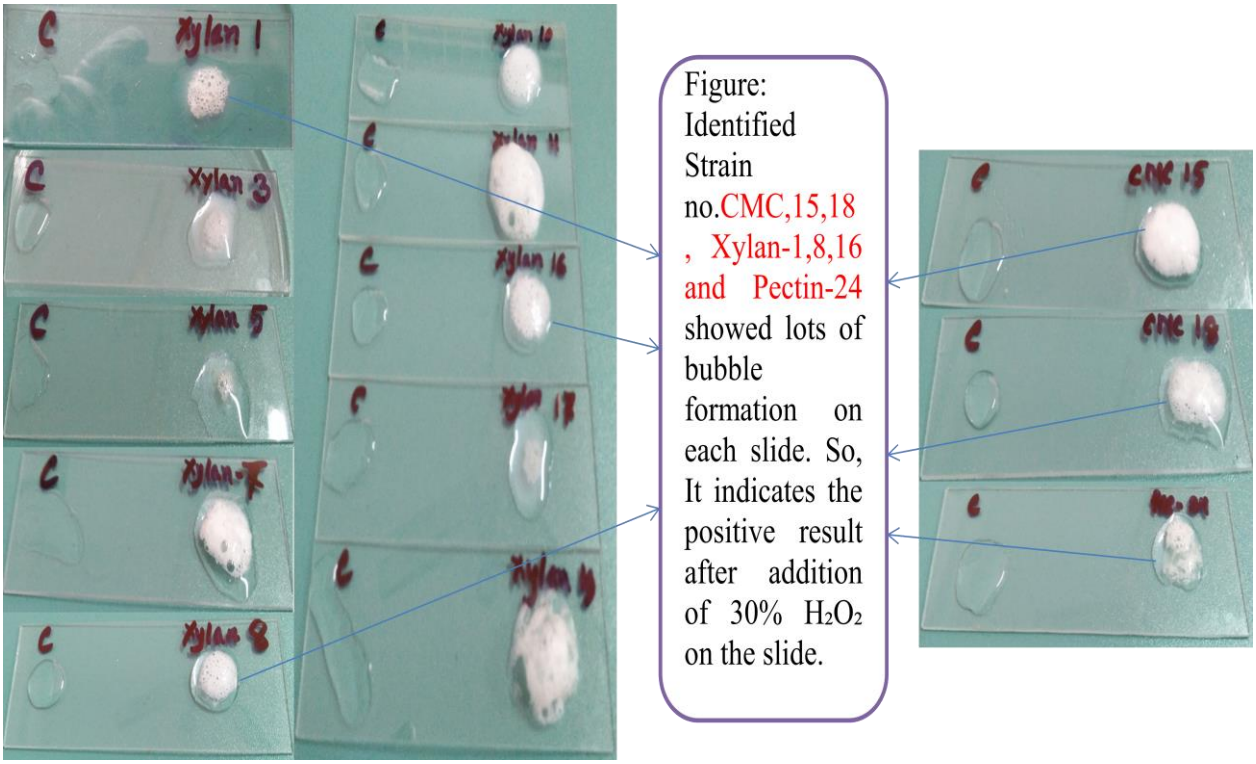


Figure 9: Catalase test

3.2.2 Oxidase Test results

The **oxidase test** is used to determine if a bacterium produces certain cytochrome c oxidases. The oxidase test result was observed after filter paper was soaked with the substrate *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine dihydrocholoride and then the isolated bacterial strains were smeared on the oxidase reagent soaked filter paper with the loop in a sterile condition. Isolated bacterial strain named identified Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24 strain show oxidase activity by exhibiting dark blue color.The bacteria which is oxidase positive produce the

enzyme named cytochrome oxidase or indophenol oxidase which have the contribution on transportation of electrons from donor compounds to electron acceptors and the test reagent, *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine dihydrocholoride works as a artificial electron donor for the enzyme oxidase. The oxidized reagent forms the colored compound indophenol blue. The cytochrome system is usually only present in aerobic organisms that are capable of using oxygen as the terminal electron acceptor. The end-product of this metabolism is either water or hydrogen peroxide (broken down by catalase).

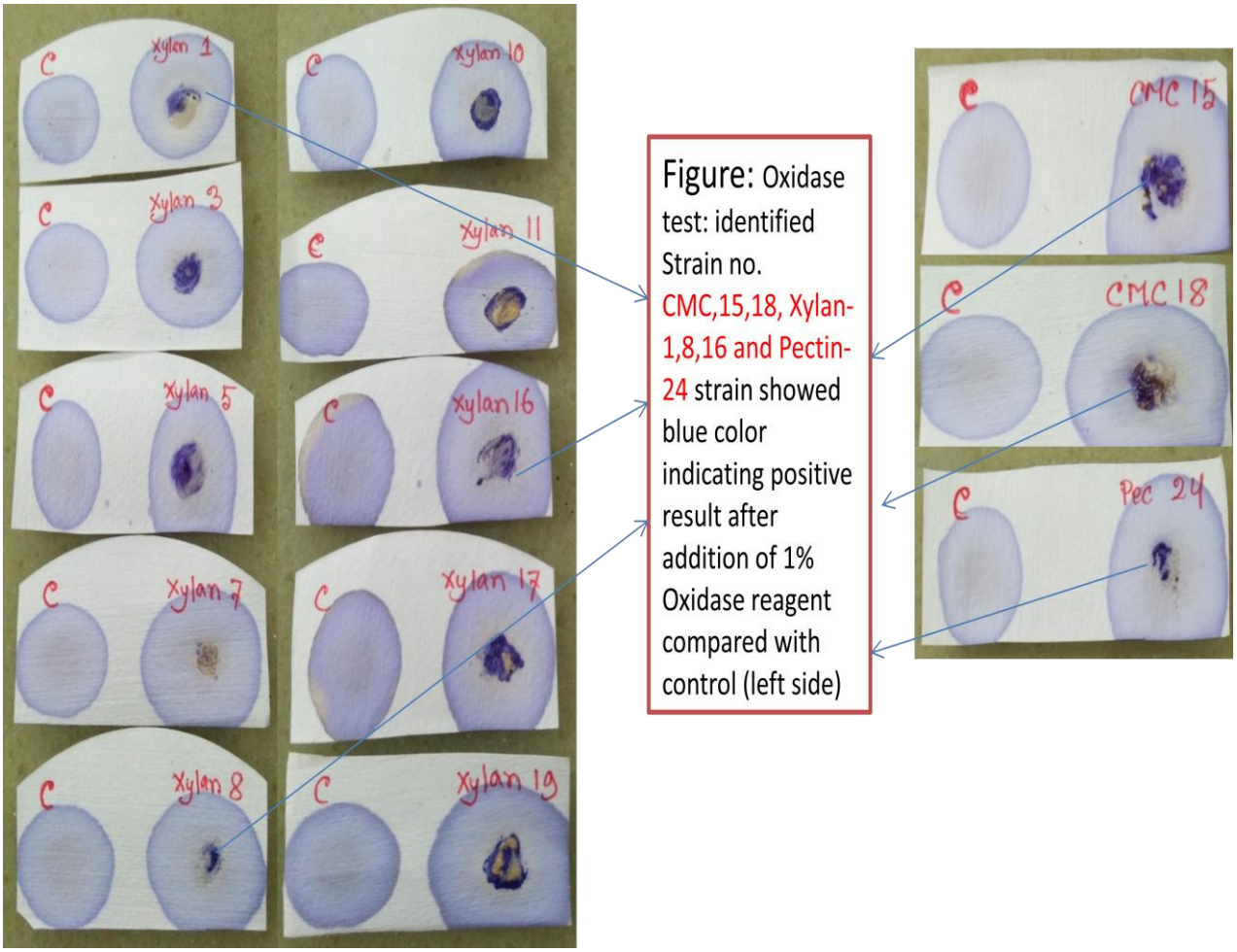


Figure 10: Oxidase test

3.2.3 Methyl Red (MR) Test results

Methyl red test is conducted to detect the bacteria those can produce acidic end product using mixed acid fermentation pathway through color changing from yellow to red after addition of methyl red indicator in MR-VP broth. If bacteria produce a large amount of acidic end product it will overcome the phosphate buffer supplied by MR-VP broth When methyl red is added, if acidic end products are present, the methyl red will stay red within 5 minutes on the other hand if there is

no acidic end product present the solution will stay yellow after addition of methyl red indicator. As seen in Fig. 4.5,identified Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24, Strain no. CMC-15,18 showed reddish color on the surface of the medium compared with control within 15 minutes(Fig 11). That means it showed positive activity. Xylan-1,8,16 and Pectin-24 don't showed reddish ring after addition of methyl red indicator that indicates negative results Actually, the isolated CMC,15,18, Xylan-1,8,16 and Pectin-24 had used the mixed acid pathway, and produced acidic end products such as lactic, acetic, and formic acid. These acidic end products are stable and will remain acidic. In previous study, Wilson et al. (1935), who used plating methods and single cell isolations, found that only one of thirty strains giving MR positive results.

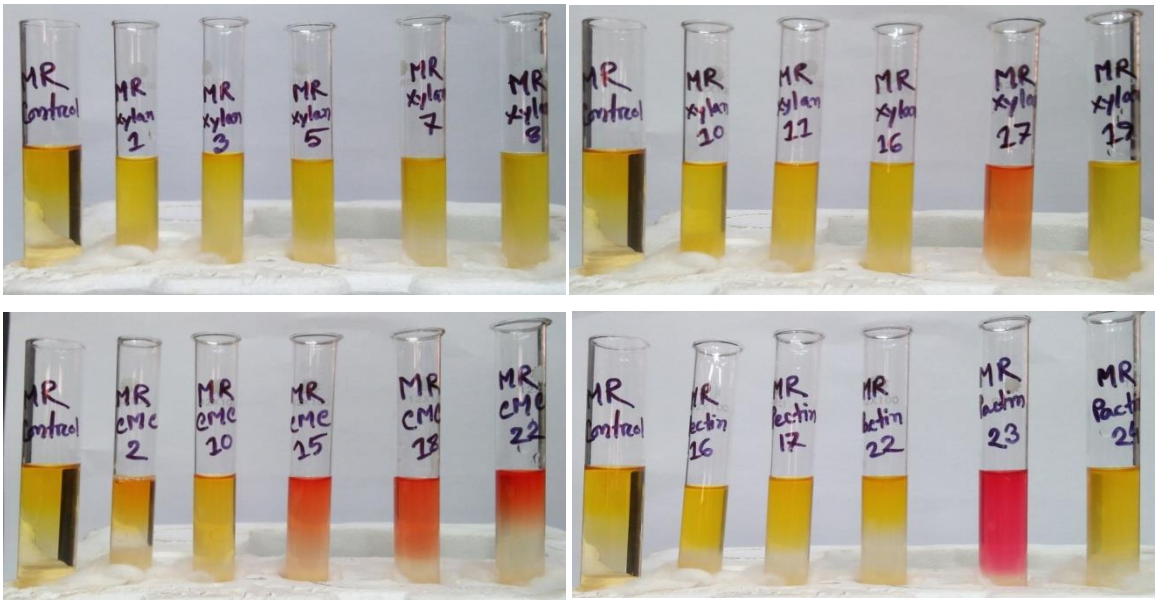


Figure 11: Methyl red (MR) test

3.2.4 Voges-Proskauer(VP) Test

The VP test detects organisms that utilize the butylene glycol pathway and produce acetoin. When the VP reagents are added to MR-VP broth that is inoculated with an organism that uses the butylene glycol pathway, the acetoin end product is oxidized in the presence of potassium hydroxide (KOH) to diacetyl. Diacetyl then reacts to produce a pink red color or reddish color indicates positive resulton the contrary, presence of yellow or copper like color indicates the negative result. Another point is that, when the same bacterial strain shows MR positive result on the same medium(MR-VP broth) these strain will not show VP positive result because large amount of acidic solution will not overcome the phosphate buffer so there is no chance of giving both MR and VP results by the same organisms on the same medium. According to statement the

isolated Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24 showed VP negative results by showing copper like color that indicates the strains used mixed acid pathways rather than butylene glycol pathway [38, 39].

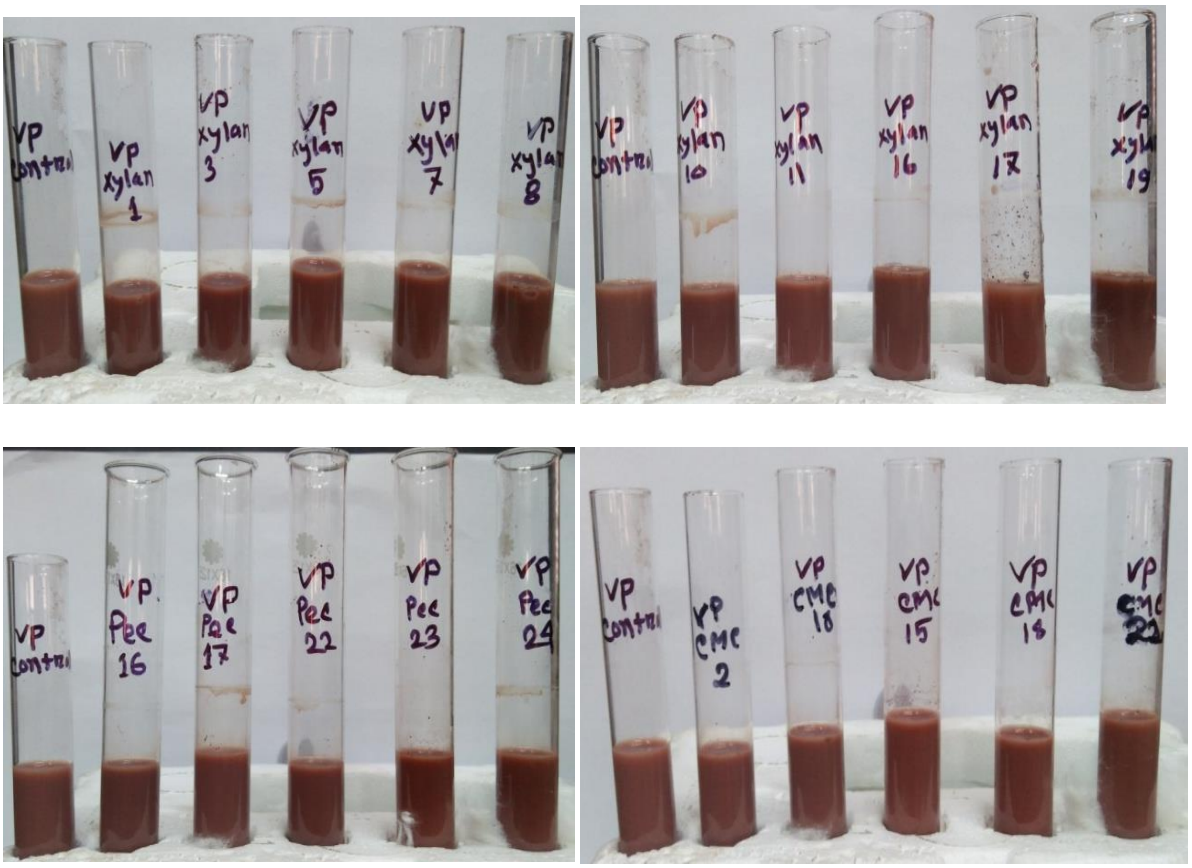


Figure 12: Voges-Proskauer(VP) Test

3.2.5 Triple Sugar Iron Agar (TSI) Test

Indentified Strain no.CMC, 15, 18, Xylan-1,8,16 and Pectin-24. Yellow colored butt and slant indicate the fermentation of lactose and fructose among three sugars and iron have been found in strain no. CMC, 15, 18, Xylan-1,8,16 and Pectin-24. Strain no. Xylan-16 and cmc-18 also indicates production of CO2 gas. These results showed yellow color in both, the phenol red indicator would turn into yellow color both in butt and slant due to the formation of large amount of acids by the fermentation of lactose and sucrose. If both slant and butt would red indicate no fermentation of sucrose, lactose, glucose and iron in that test. The production of black colored FeS indicates the fermentation of iron in the test. No black colored FeS in both slant and butt was seen. According to the statement, the isolate CMC,15,18, Xylan-1,8,16 and Pectin-24 are capable of fermenting glucose only.

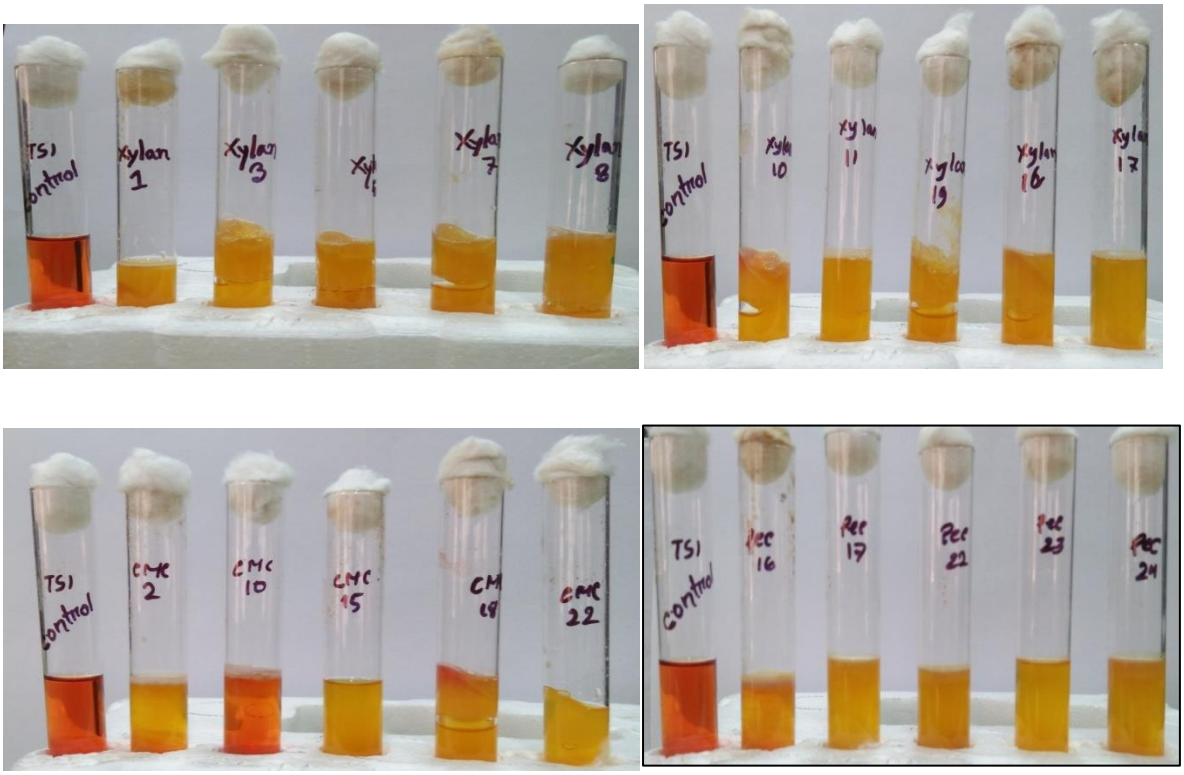


Figure 13: Triple Sugar Iron Agar (TSI) Test

3.2.6 Citrate Utilization test

Citrate test is often used to identify Gram negative bacteria and distinguish between *Enterobacteriace* family and environmental isolates based on their metabolic pathway. Gram negative bacteria can utilize citrate as a carbon source and inorganic salts as source of nitrogen served in media by the producing of citrate –perm ease enzyme which has contribution to the conversion of citrate to pyruvate ultimately formation of alkaline carbonates and bicarbonates and ammonium salts to ammonia ultimately formation of ammonium hydroxide and the color of the media will be changed. The shift in pH turns the bromothymol blue indicator in the medium from green to blue above pH 7.6 [42]. If this reaction occurs, the medium will be blue in color indicating positive result. As seen in Fig.4. 8, Indentified Strain no. CMC, 15, 18, Xylan-1, 8, 16 and Pectin-24. Blue color formation after incubation from green colored smple indicates the positive color compare with control. Strain no. Xylan-1, 8, 16 indicates positive result and CMC-15,18 , pectin-24 that indicates negative results(Fig 14). In general, citrate positive results are demonstrated by the *Salmonella, Edwardsiella, Citrobacter, Klebsiella, Enterobacter Serratia,Providencia* species, and *Bacillus subtilis*, while citrate negative results are exhibited by the *Escherichia, Shigella, Morganella,* and *Yersinia* species etc.

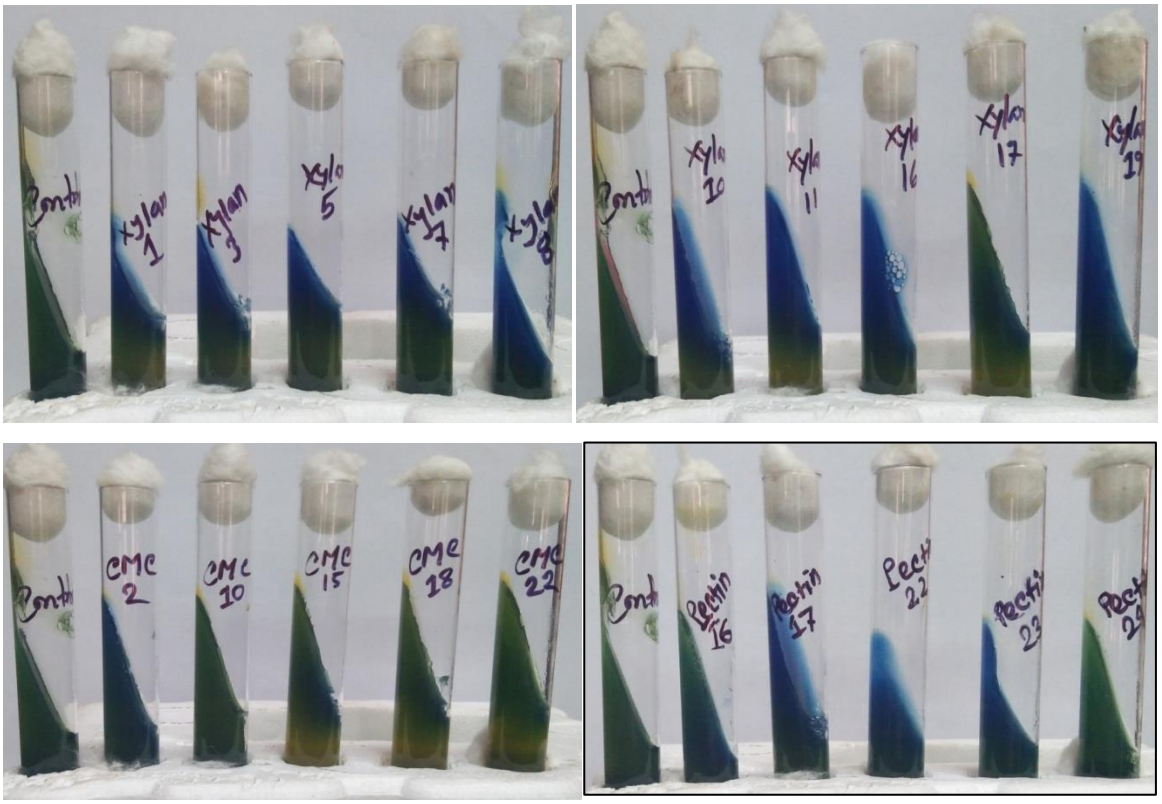


Figure 14: Citrate Utilization test

3.2.7 Motility Indole Urease (MIU) Test

The MIU (motility, indole, and Urease) medium base is formulated to detect motility, urease and indole production in a single tube. As seen in (Fig. 15), a red or red-violate ring appeared in the top surface of the MIU culture medium containing bacterial strains. Indole test is used to determine the ability of an organism to split amino acid tryptophan to form the indole compound [43]. In fact, tryptophan is hydrolyzed by tryptophanase to produce three possible in products such as indole, pyruvate, and ammonium. Most strains of *E.coli*,*P. vulgaris*, *M. morganii* and *Providenica* are indole positive.Indole test can also aid in species differentiation.It is important to mention that *Klebsiella oxytoca* is indole positive where as *Klebsiella pneumoniae*is indole negative. In addition, Citrobacter Koseriis indole positive where as *Citrobacter freundii* is indole negative [43]. On the other hand, Appearance of red ring indicates the indole and Urease positive , and white colored turbity indicates the motility positive result. Comparing with the indole test indentified Strain no.CMC-15, 18 and Xylan-8, 16 showed motility and urease. Urease test is

useful to identify *Cryptococcus* species, Brucella, *Helicobacter pylori.Presence of white colored turbidity indicates the motility positive.*

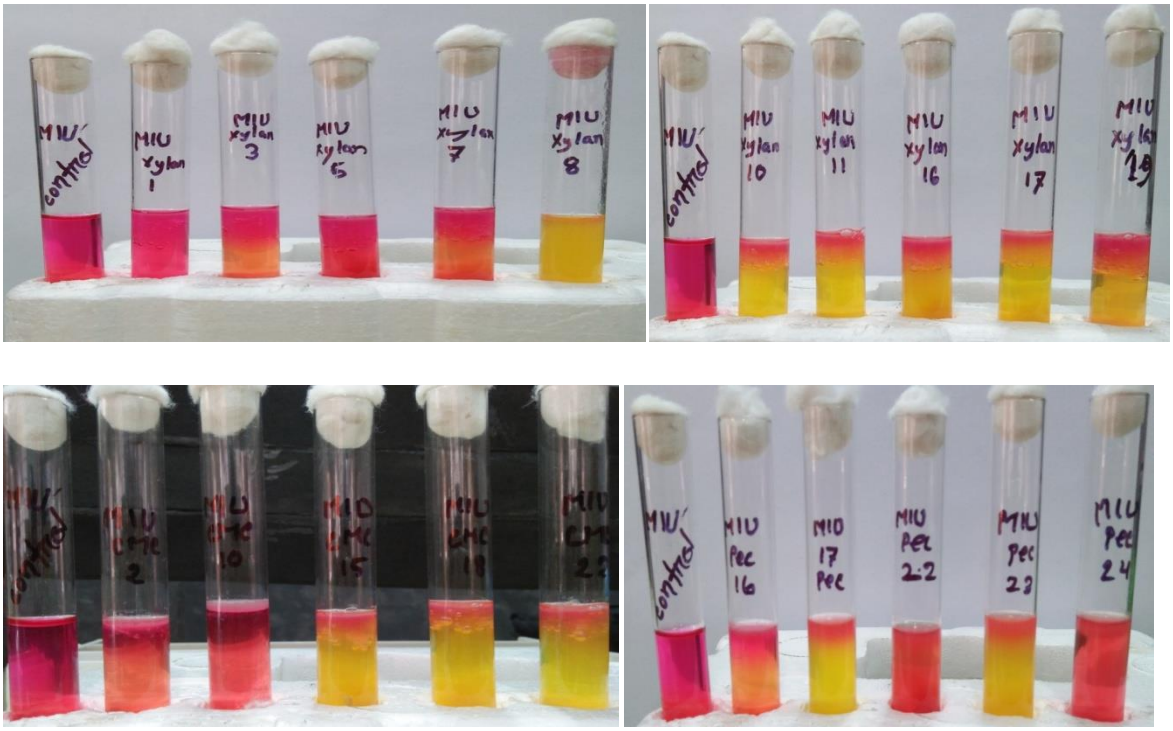


Figure 15: Motility Indole Urease (MIU) Test

All stains showed that indole are negative results

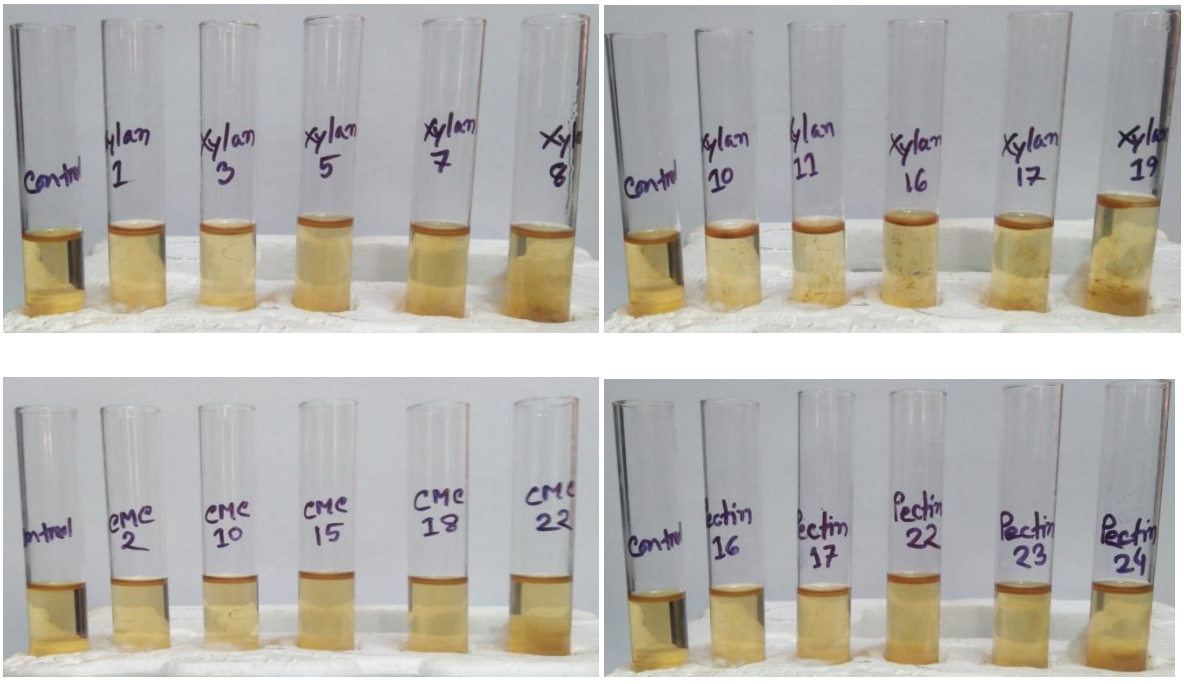


Figure 16: Indole test

3.2.8 Fermentation of sugar: lactose, maltose, sucrose, dextrose Results

The carbohydrate fermentation test is used to determine whether or not bacteria can ferment a specific carbohydrate. When microorganisms ferment carbohydrate an acid or acid with gas are produced, which is detected by the color change of the pH indicator. Yellow color formation indicates the fermentation of carbohydrate and red colored medium indicates that bacteria cannot ferment carbohydrates [47].

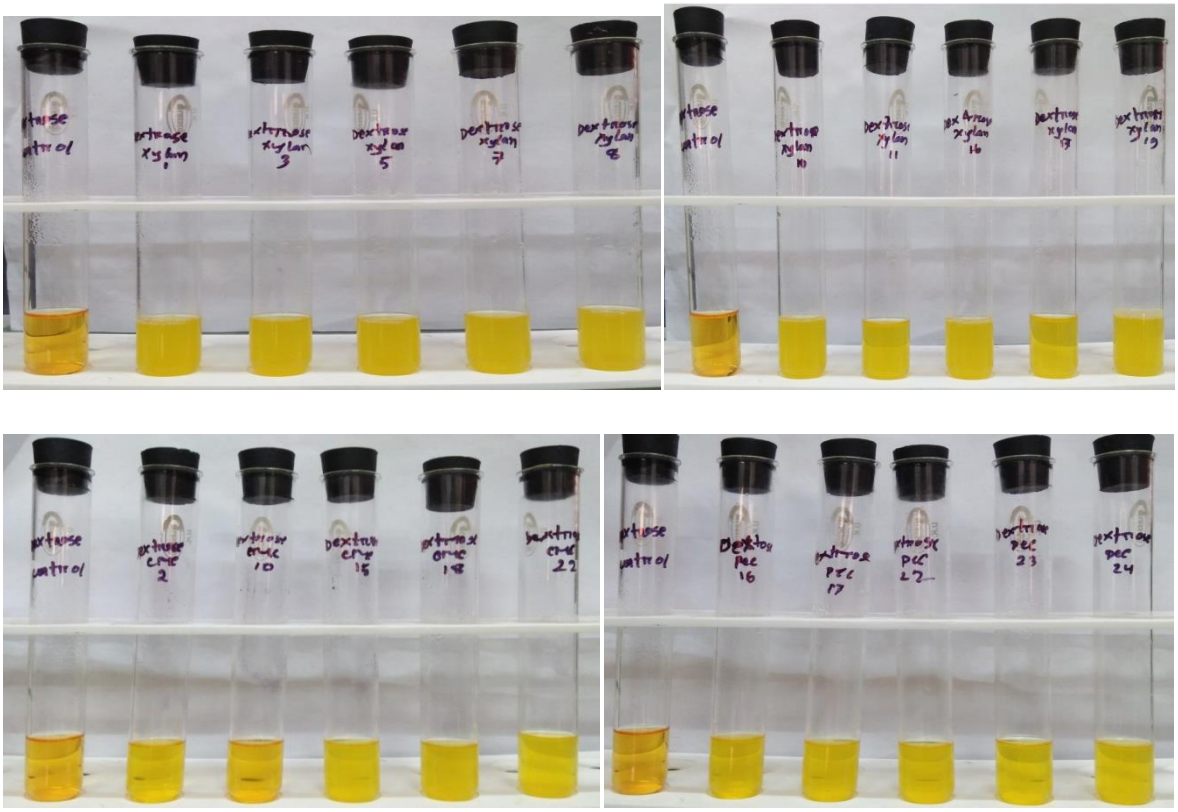


Figure 17 A: Dextose

In this study, following to the color changing indication isolated identified Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24 fermented Dextose compared with control (Fig.17 A). In the case of maltose fermentation showed identified Strain no. CMC, 15,18, Xylan-1,8,16 and Pectin-24 showed yellow color indicating positive result (Fig. 17B). Lactose fermentation showed identified Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24 showed yellow color indicating positive result. (Fig. 17C). On the otherhand, Sucrose fermentation showed identified Strain no. CMC,15,18, Xylan-1,8,16 and Pectin-24 showed yellow color indicating positive result (Fig. 17D.). So yellow color indicating the strains are capable of utilize and ferment dextose ,maltose, lactose and sucrose compared with each control.

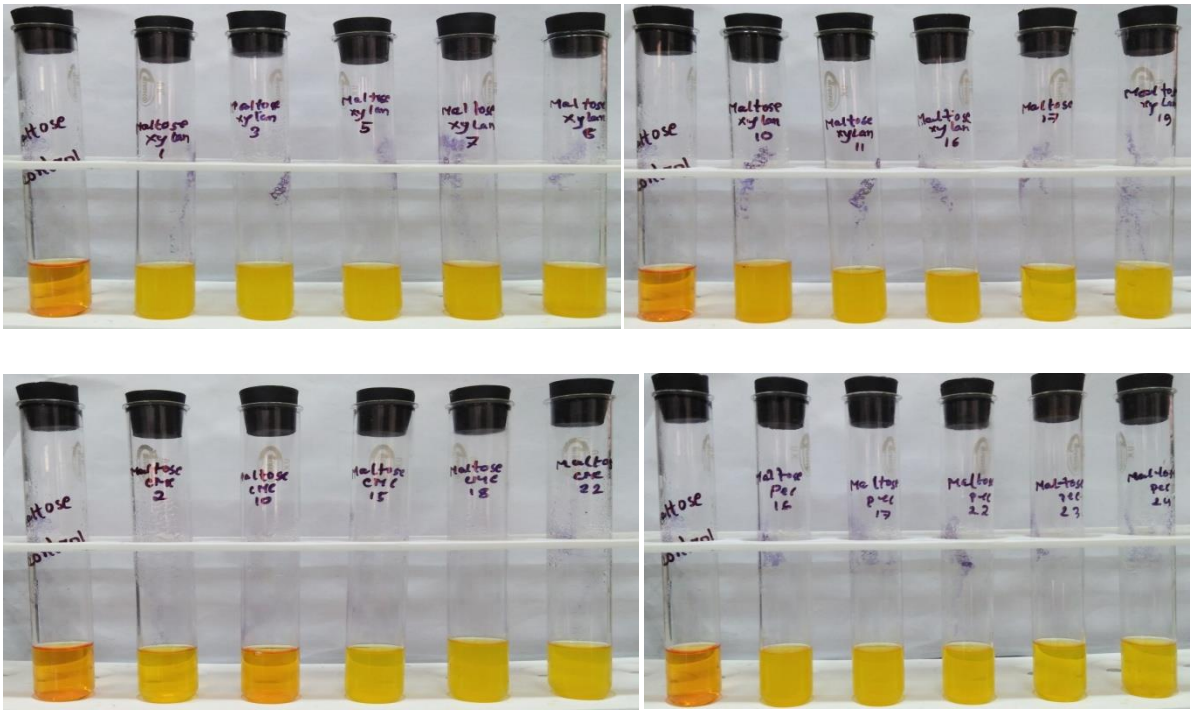


Figure 17B: Maltose

Many bacteria use glucose, a monosaccharide or simple sugar, because they possess the enzymes required for the degradation and oxidation of this sugar. Fewer bacteria are able to use complex carbohydrates like disaccharides (lactose or sucrose) or polysaccharides (starch). Disaccharides and polysaccharides are complex sugar that are linked by glycosidic bonds; bacteria must produce enzymes to cleave these bonds such that the simple sugars that result can be transported into the cell. If the bacteria cannot produce these enzymes then the complex carbohydrate is not used.

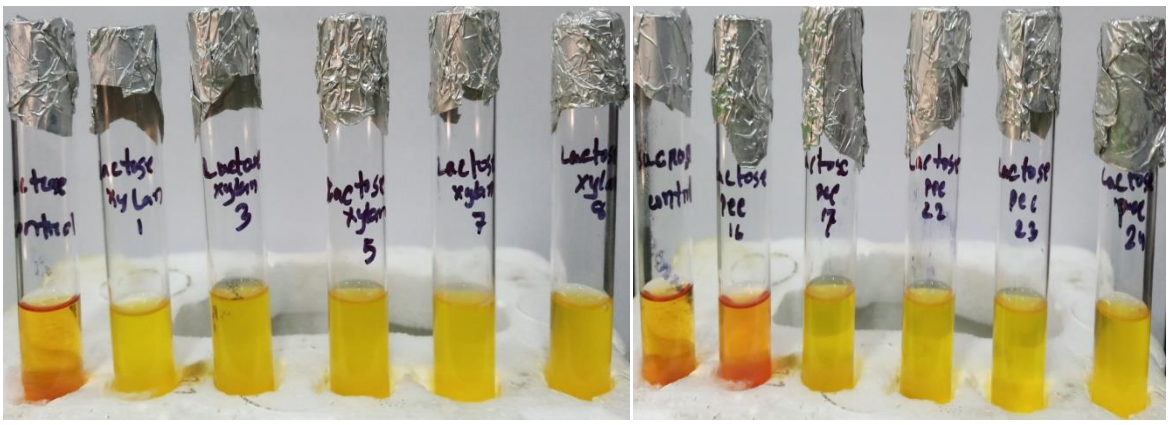


Figure 17C: Lactose

The lactose is a disaccharide consisting of monomeric glucose and monomeric galactose linked by a glycosidic bond. Bacteria that use lactose must first produce the enzyme *lactase* (*beta-galactosidase*) to break the glycosidic bond between these monomers.

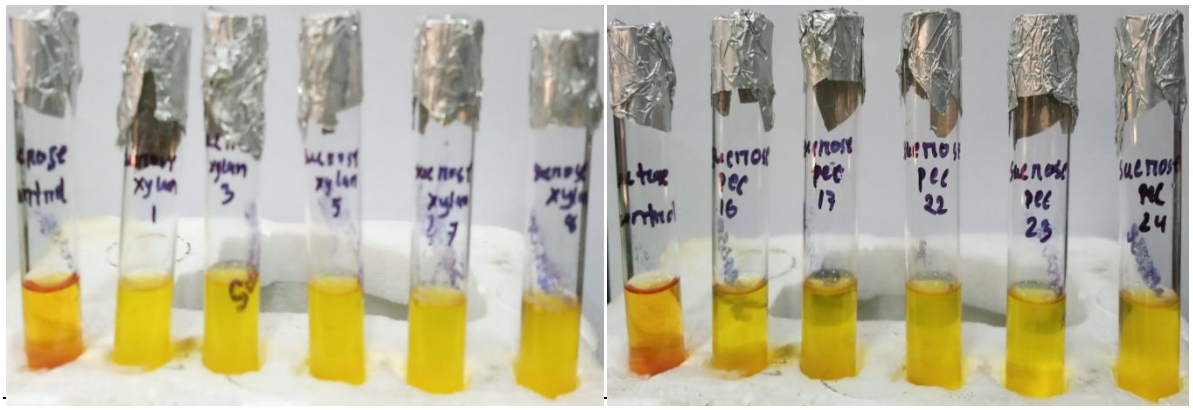


Figure 17D: Sucrose

3.3 Identification of 16s rRNA from identified 6 strains

Strain name	Genus	Gene accession number
Xylan-1	Klebsiella sp. strain HSTU-AMS44	MK691477
Xylan-8	Enterobacter sp. strain HSTU-ASM46	MK691478
Xylan-16	Enterobacter sp. strain HSTU-ASM47	MK691479
CMC-15	Shigella sp. strain HSTU-ASM49	MK691480
CMC-18	Escherichia sp. strain HSTU-ASM50	MK691481
Pectin-24	Klebsiella sp. strain HSTU-ASM51	MK691482

After finishing all experiment and molecular identification of those bacteria, one strain (Pectin-24 genomic strain name HSTU-AAM51) have been taken for NGS (Next Generation Sequencing) to identification of lignocellulase enzyme producing gene or Carbohydrate active enzymes (CAZy) gene .

3.4 NGS analysis using by list of some bioinformatics tools

3.4.1 FastQC analysis

FastQC was analyzed using Basespace (Illumina) web base server. After completing WGS, 2 fastq file have been gotten which name R1 and R2 file. Those file was using check their quality by fastQC analysis, quality depend on GC%. Previous report says the optimum range of GC% (40-70). So HSTU-AAM51 strain was showed 57.2%that indicate sequence quality is good [48].

Table 14 : FastQC analysis using Illumina V.1.9

Filename	**51_S7_L001_R2_001.fastq.gz**
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	838592
Filtered Sequences	0
Sequence length	35-151
%GC	57.2

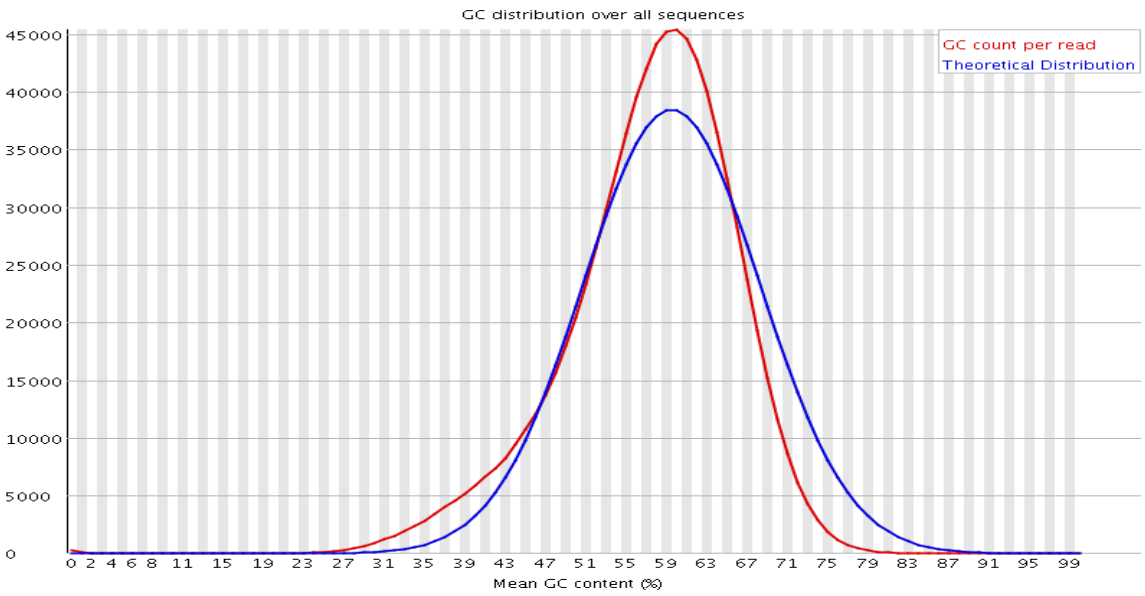


Figure 18: FastQC analysis

3.4.2 Assembly in SPAdes

NGS projects typically feature high coverage, NGS assemblers generate rather accurate contigs (although the accuracy deteriorates for SCS). SPAdes constructs DNA sequences of contigs and the mapping of reads to contigs by backtracking graph simplifications [49]. Using SPAdes V.3.9.0 algorithm HSTU-AAM51 strain have been generated for construct mapping and contigs sequence which showed 290 contigs sequence and number of read sequence 5506938bp

Table15: List in details assembly report

Assembly	contigs	scaffolds
contigs (>= 0 bp)	290	284
contigs (>= 1000 bp)	49	43
Total length (>= 0 bp)	5564045	5563774
Total length (>= 1000 bp)	5504381	5504110
contigs	54	48
Largest contig	657932	657932
Total length	5507209	5506938
GC (%)	57.32	57.32
N50	207699	279071
N75	108653	158704
L50	9	7
L75	18	14
N's per 100 kbp	0	0.18

3.4.3 GENIUS Metagenomics analysis

The GENIUS Metagenomics application have been used to acurated genome database and high performance algorithms to provide rapid, accurate, and actionable bacterial identification at the species, subspecies, and/or strain level [52]. The result of strain HSTU-AAM51 (Figure 19) was showed 98% genus conformation.

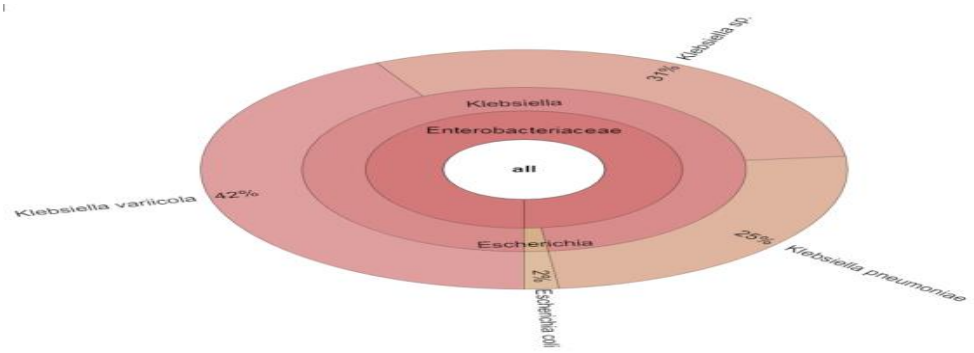


Figure 19: Metagenomics analysis of HSTU-AAM51 strain

3.4.4 ANI value calculation

Average nucleotide identity (ANI) is a category of computational analysis that can be used to define species boundaries of Archaea and Bacteria. Calculating ANI usually involves the fragmentation of genome sequences, followed by nucleotide sequence search, alignment, and identity calculation [56]. *K. variicola* HSTU-AAM51 genome have been compared with 8 same genuses where as highest result showed 99.16 % nucleotide sequence identity. ANI calculation was done by MUMmer tool version 3.0 web base sarver. All result shows below Table no 16.

Table 16: ANI calculating result

	***Klebsiella variicola* HSTU-AAM 51**	***Klebsiella aerogenes* strain AR 0009**	***Klebsiella grimontii* strain SS141**	***Klebsiella huaxiensis* strain WCHKl090001**	***Klebsiella michiganensis* strain K518**	***Klebsiella oxytoca* strain P620**	***Klebsiella pneumoniae* str. Kp52.145**	***Klebsiella pneumoniae* strain SB3432**	***Klebsiella variicola* strain 13450.**
***Klebsiella variicola* HSTU-AAM 51**	*	87.48	85.85	85.14	85.94	85.53	94.68	94.53	**99.16**
***Klebsiella aerogenes* strain AR 0009**	87.48	*	85.85	85.29	85.81	85.54	87.90	87.53	87.62
***Klebsiella grimontii* strain SS141**	85.85	85.86	*	88.75	93.97	90.71	85.98	85.93	85.98
***Klebsiella huaxiensis* strain WCHKl090001**	85.13	85.29	88.76	*	88.81	88.25	85.21	85.17	85.25
***Klebsiella michiganensis* strain K518**	85.94	85.81	93.98	88.81	*	90.72	86.00	85.90	86.11
***Klebsiella oxytoca* strain P620**	85.53	85.55	90.71	88.25	90.72	*	85.46	85.48	85.59
***Klebsiella pneumoniae* str. Kp52.145**	94.67	87.90	85.98	85.21	86.00	85.45	*	**99.07**	94.70
***Klebsiella pneumoniae* strain SB3432**	94.52	87.53	85.93	85.17	85.90	85.47	**99.06**	*	94.56
Klebsiella variicola strain 13450.	**99.16**	87.62	85.97	85.26	86.12	85.58	94.70	94.56	*

3.4.5 Phylogenetic tree analysis

The evolutionary history was inferred by using the Maximum Likelihood method based on the REALPHY online tool (fig.21). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. *Klebesiella variicola* HSTU-AAM51 strain was showed 0.006 distances from nearest species and made cluster form.

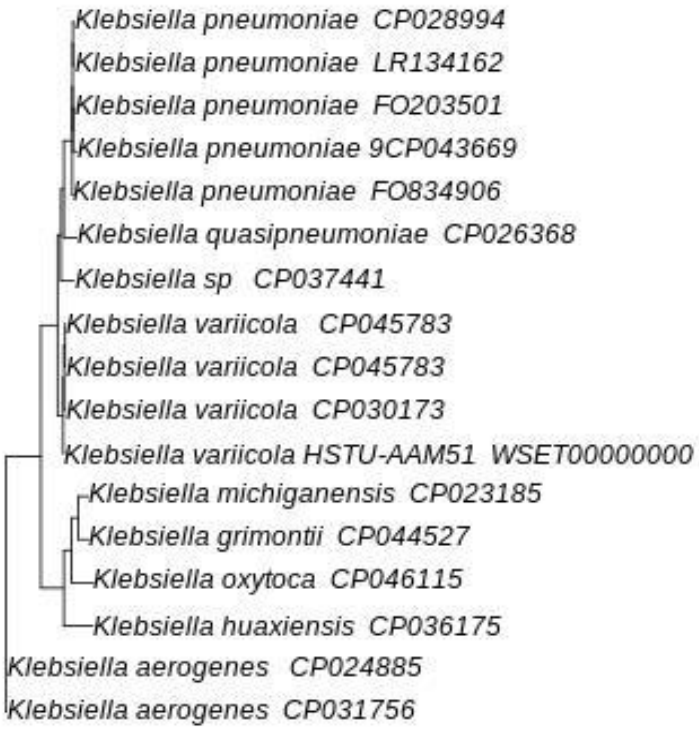


Figure 20: Phylogenetic analysis of HSTU-AAM51

3.4.6 Venn diagram analysis

Venn diagram was showing the *Klebsiella variicola* HSTU-AAM51 predicted essential amino acid sequence in different studies. HSTU-AAM51 strain was showed total 156575 numbers of amino acid sequence present in this genome in which 118833 amino acid sequences were showed unique and 23633 numbers of amino acid sequence were common among each other (Fig 21). In addition to maximum 22388 amino acid sequence were matched with reference genome (CP030173)

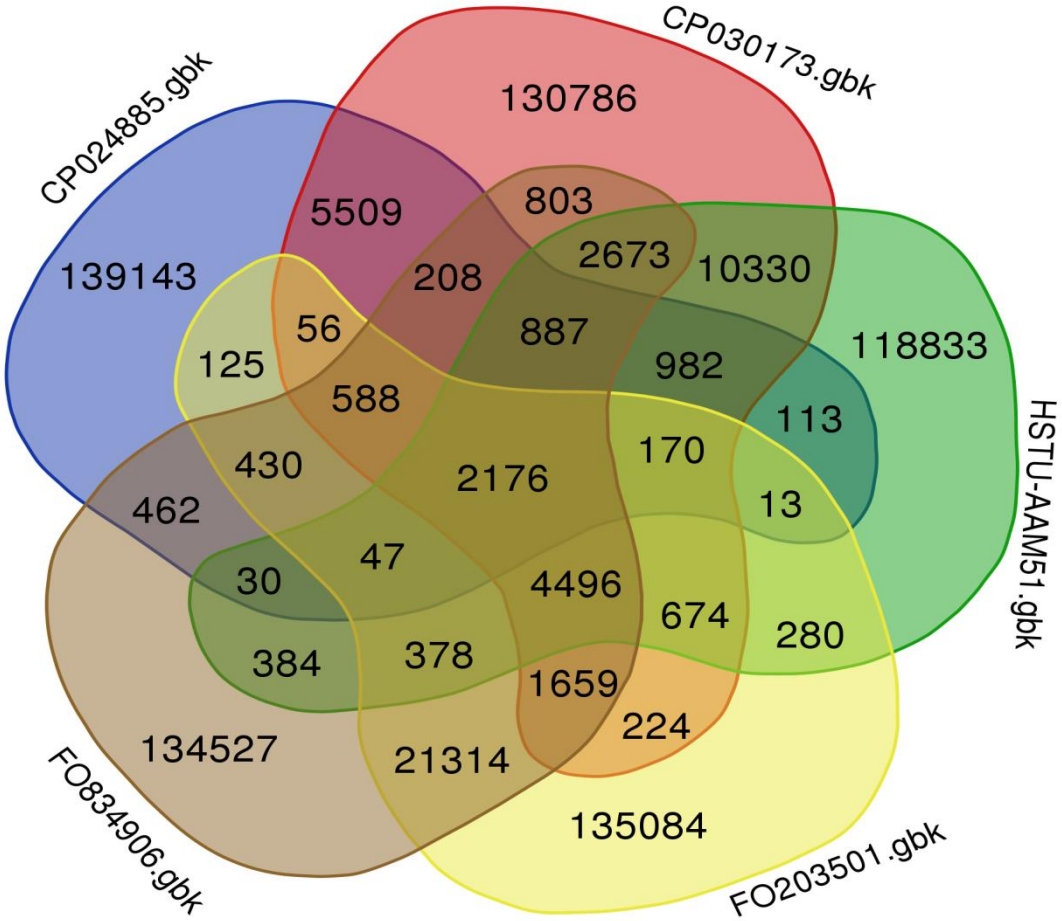


Figure 21: Venn diagram of *Klebsiella variicola* HSTU-AAM51 predicted essential amino acid sequence

3.4.7 Pan-genome analysis

CGview shows all the proteins in a bacterium in a proper way, it's called genomic map. We hereby transcribed the *Klebsiella variicola* HSTU-AAM51 bacterial protein into a genomic map where 5187 CDS, 88 tRNAs were shown in 5.5mbp length sequences. GC skew is observed in local

genomic regions primarily introduced by RNA synthesis, but the overall genomic polarity due to replication is present regardless of these local effects, and the GC skew is thus observed in intergenic regions as well as in the third nucleotide positions in codons. Because only a few *origin* and *terminus* positions had been identified by experimental means, analysis of GC skew was first utilized for the computational prediction of *ori* and *ter* positions by examining available genome sequences. Here GC-skew positive indicates that the protein coding sequence (CDS) is present in downstream and GC-skew negative indicates that the protein coding sequence (CDS) is present in upstream (Fig22).

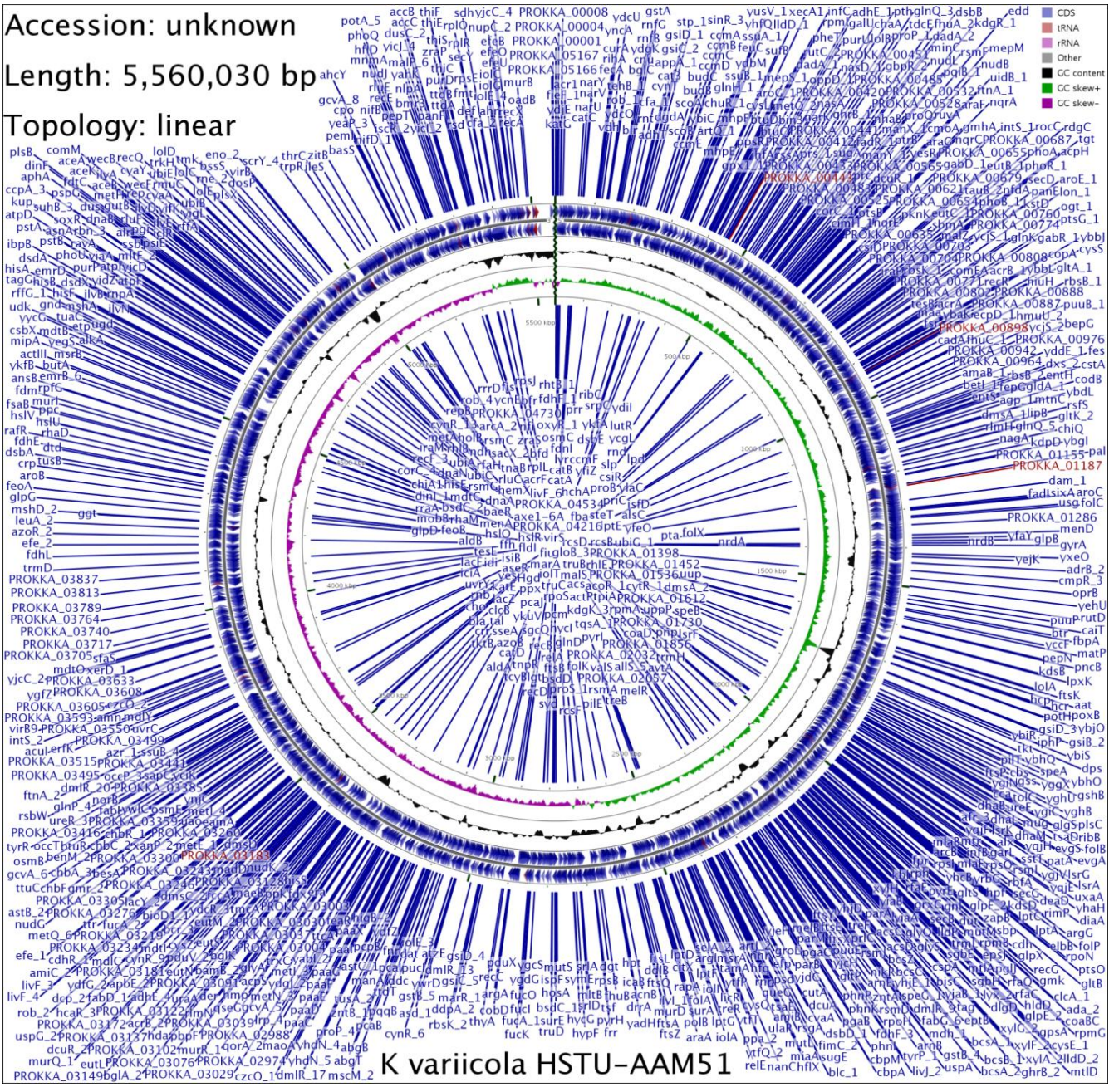


Figure 22: Genomic map of *Klebsiella variicola* HSTU-AAM51

3.4.8 KAAS Analysis

KAAS (KEGG Automatic Annotation Server Ver. 2.1) provides functional annotation of genes by BLAST or GHOST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways. After generating *K. variicola* HSTU-AAM51 strain result was showed (table 17 and table18) in this pathway.

KAAS works best when a complete set of genes in a genome is known. Prepare query amino acid sequences and use the BBH (bi-directional best hit) method to assign orthologs.

Table 17 : Genes database feature using BBH method

query name	***klebsiella variicola* HSTU-AAM51**
query type :	nuc
program :	BLAST
method :	BBH
GENES data set :	hsa, dme, ath, sce, pfa, eco, sty, hin, pae, nme hpy, rpr, mlo, bsu, sau, lla, spn, cac, mge, mtu ctr, bbu, syn, aae, mja, afu, pho, ape

Table 18 : List some KEGG Orthology(KO) result of *K. varricola* HSTU-AAM51

NODE Number	**KO (KEGG Orthology)**
NODE_1_length_657932_cov_19.9232	K00370
NODE_2_length_579060_cov_21.4723	K00164
NODE_3_length_341026_cov_23.0038	K00362
NODE_4_length_332684_cov_19.6963	K07309
NODE_5_length_325515_cov_21.3814	K03583
NODE_6_length_281156_cov_20.9314	K03632
NODE_7_length_279071_cov_21.4704	K00336

NODE_8_length_246311_cov_22.768	K00265
NODE_9_length_228282_cov_22.5105	K09800
NODE_10_length_220361_cov_22.3763	K01955
NODE_11_length_180057_cov_21.9723	K01870
NODE_12_length_168421_cov_19.4533	K03578
NODE_13_length_162960_cov_21.4953	K06894
NODE_14_length_158704_cov_20.1073	K01681
NODE_15_length_132074_cov_20.269	K03703
NODE_16_length_120002_cov_22.437	K00281
NODE_17_length_98807_cov_23.7312	K00123
NODE_18_length_98317_cov_20.3667	K07180
NODE_19_length_93212_cov_20.7668	K07788
NODE_20_length_88506_cov_23.2951	K02470
NODE_21_length_87544_cov_22.5081	K00548
NODE_22_length_86442_cov_23.494	K05851
NODE_23_length_78068_cov_21.1847	K03723
NODE_25_length_60492_cov_21.3764	K03702
NODE_26_length_52692_cov_19.1026	K03737
NODE_27_length_47922_cov_21.2351	K01756
NODE_28_length_45987_cov_21.9494	K01811
NODE_30_length_34688_cov_24.3871	K03043
NODE_31_length_26479_cov_27.1349	K03499
NODE_32_length_19294_cov_21.9445	K07787
NODE_33_length_17505_cov_20.7942	K03336
NODE_36_length_7309_cov_23.3946	K01198
NODE_37_length_4295_cov_25.6562	K00867
NODE_38_length_3357_cov_38.9747	K01571
NODE_40_length_1926_cov_47.2737	K00615
NODE_54_length_388_cov_0.945338	K02494
NODE_89_length_274_cov_0.746193	K05875
NODE_109_length_266_cov_0.783069	K00249
NODE_114_length_265_cov_0.765957	K01916
NODE_159_length_253_cov_0.818182	K18118

NODE_208_length_239_cov_0.907407	K02377
NODE_228_length_233_cov_0.942308	K03586

3.4.9 Annotating a Genome Using RAST

The basic steps was used to annotate a genome using RAST are described in the subsections below. Input to the process is a prokaryotic genome in the form of a set of contigs in FASTA format. After annotating contig fasta file showed in seed viewer result below table and also showed genomic feature in subsystem about 59% genome coverage .In subsystem –total 3038 proteins are involved where the strain HSTU-AAM51 is predicted to encode for about 858 proteins involved in carbohydrate metabolism and 594 proteins involved in amino acid metabolism (Fig 22). In addition to identify CAZy (carbohydrate active enymes) and lignin degrading protein sequence from RAST seed viewer server, identified protein shows table 20 and table 21 respectively.

Table 19: Annotating features of HSTU-AAM51.

Genome	*Klebsiella variicola* HSTU-AAM51
Domain	Bacteria
Taxonomy	Bacteria; *Klebsiella variicola* HSTU-AAM51
Size	5,564,045
GC Content	57.2
N50	207699
L50	9
Number of Contigs (with PEGs)	290
Number of Subsystems	588
Number of Coding Sequences(CDS)	5187
Number of RNAs	110

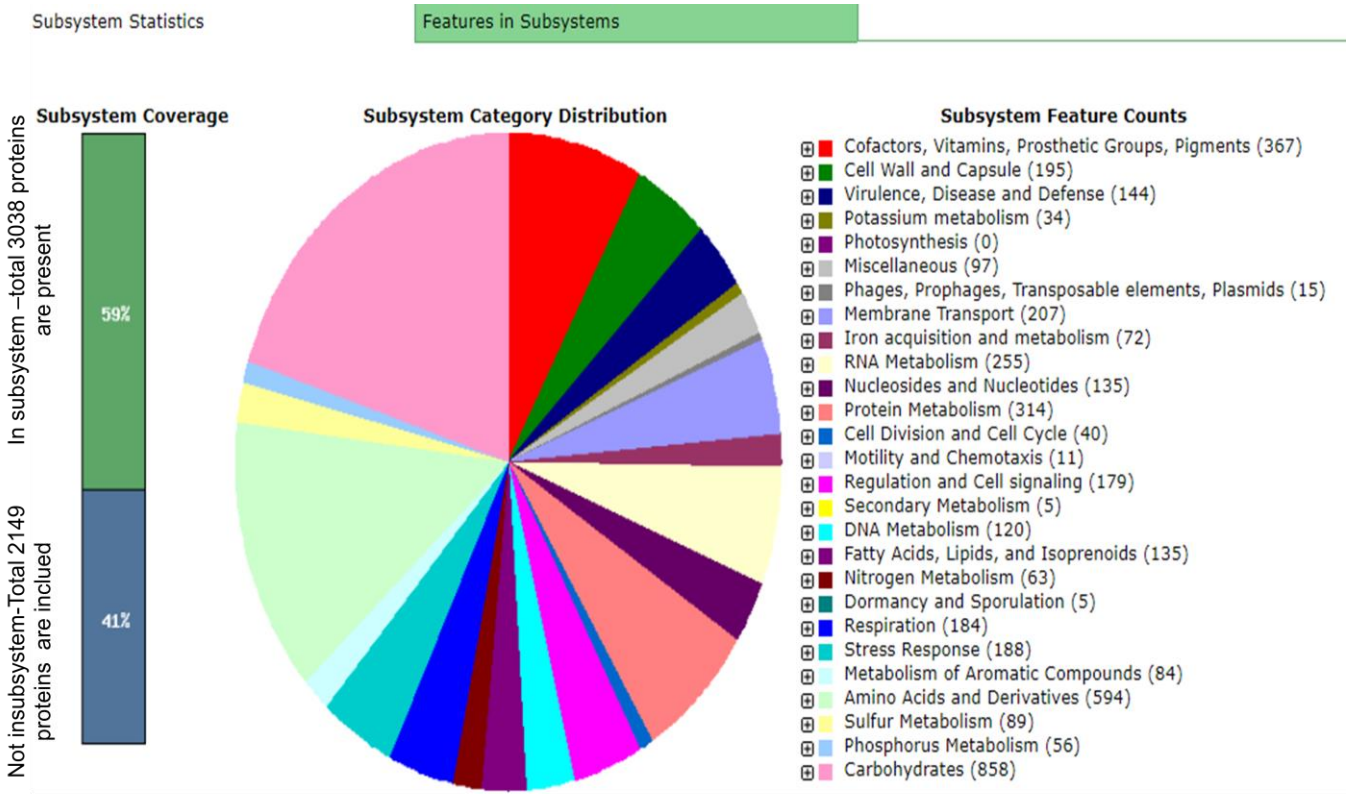


Figure 23: **Genes connected to subsystems and their distribution in different categories**

3.5 Identified lignocelluloses and lignin degrading enzymes from HSTU-AAM51 Strain.

The process of lignocellulose degradation is of great research interest, especially for biotechnology, due to its potential as a sustainable resource for biofuels and biomaterial production [67]. Lignocellulose is composed of cellulose, hemicellulose, and lignin. The degradation of these polymers requires the synergistic action of multiple Carbohydrate-Active enZyme (called CAZymes) families, typically acting together as an enzyme cocktail with coordinated oxidative, hydrolytic and non-hydrolytic activities. Since lignin, a complex heteropolymer providing strength and resistance to plant tissues [68], protects carbohydrate polymers against enzymatic digestion, its degradation is a critical first step in lignocellulose degradation enabling the liberation of cellulose and hemicellulose [69]. Many enzymes are known as lignin-modifying enzymes (LMEs): lignin peroxidases, manganese peroxidases, versatile peroxidases, laccases and cellobiose dehydrogenases [70]. The lignocelluloses and lignin degrading bacteria *Klebsiella variicola* HSTU-AAM51 was isolated from goat rumen contents. We were found 45 genes are involved in the degradation of lignocellulose and other polysaccharides, including 6 cellulose degrading genes (GH3, GH5, GH13,

GH94) , 4 xylanose degrading genes (GH10,CE4,GH43), 1 mannose degrading gene (GH5), 4 hemicellulose degrading genes (GH27, GH2, GH13, GH53) , 1 pectin degrading gene (PE1) and 29 lignin degrading genes. Some lignin degrading genes were found in different locus of the genome (Table 21).

Table 20: Identified lignocellulose degrading enzymes

Catagory	CAZy Family	Activities	Locus tag (NODE: gene position)
Cellulase	GH3	Beta-glucosidase	NODE _16: 112694-114121
		Beta-glucosidase	NODE_18: 45650-48094
	GH5	Endoglucanase precursor	Node_6: 154138-155139
		Endoglucanase precursor	NODE_6: 164751-165857
	GH13 type	1,4-alpha-glucan (glycogen) branching enzyme	NODE_19: 22756-24942
	GH94	Cyclic beta-1,2-glucan synthase	NODE_27: 34624-43200
Xylanase	GH10 Glycoside Hydrolase Family 10	Endo-1,4-beta-xylanase A precursor	NODE_23 : 50961-49810
	CE4	Polysaccharide deacetylase	NODE_8: 134469- 135722
	GH43	Beta-xylosidase	NODE_42: 3575-1896
Mannase	GH5	Endo-b1,4-mannanase 5C	NODE_9: 37773-35581
Hemicellulase	GH27	Alpha-galactosidase	NODE_28: 37864-39987
	GH2	Beta-galactosidase	NODE_13: 85763-82656

	GH Family 13	Family 13 glycosyl hydrolase, row 724	NODE_27: 64452- 62770
	GH53	Arabinogalactan endo-1,4-beta-galactosidas	NODE_2: 214155- 215357
Pectinase	GH family	Pectinesterase	NODE_29: 34218-32935

Table 21: Identified lignin degrading enzymes

Lignin degrading gene	Locus tag (NODE: gene position)	Enzyme name
RpfF	NODE_25: 82154-79965	Enoyl-CoA hydratase/isomerase family protein RpfF
SAUSA300_0226	NODE_3: 61343-63487	3-hydroxyacyl-CoA dehydrogenase
Amico_1318	NODE_6: 27217-26471	Carboxyvinyl-carboxy phosphonate phosphoryl mutase
Dgeo_0969	NODE_2: 4708-5478	SAM-dependent methyltransferase
CTP10_5409	NODE_13: 29411-30313	LysR family transcriptional regulator YnfL
STM2281	NODE_10: 118452- 119327	Putative LysR family transcriptional regulator

yneJ	NODE_13: 113780- 112902	LysR family transcriptional regulator
ydcI	NODE_1: 143424- 144347	LysR family transcriptional regulator
STM3121	NODE_1: 261720- 260848	Putative LysR family transcriptional regulator
B4417_2085	NODE_20: 74411- 75319	LysR family transcriptional regulator YeiE
lysR	NODE_33: 29056- 28127	Transcriptional regulator, LysR family
paaY	NODE_11: 85507- 84911	Phenylacetic acid degradation protein PaaY
nahG	NODE_13: 140882-139728	Salicylate hydroxylase
TMU3MR103_0947	NODE_11: 145100-144066	Zinc-containing alcohol dehydrogenase/quinone oxidoreductase
N505_0106110	NODE_17: 69412- 68648	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
ypeA	NODE_12: 141045-141470	Acetyltransferase YpeA
CRYZ	NODE_14: 4468-5448	Quinone oxidoreductase
HPD	NODE_1: 358460- 356607	4-hydroxyphenylpyruvate dioxygenase
PHYHD1	NODE_11:109812- 110684	Phytanoyl-CoA dioxygenase domain-containing protein 1
GOX1095	NODE16:16147-17190	Aminomethyltransferase (Glycine cleavage system T

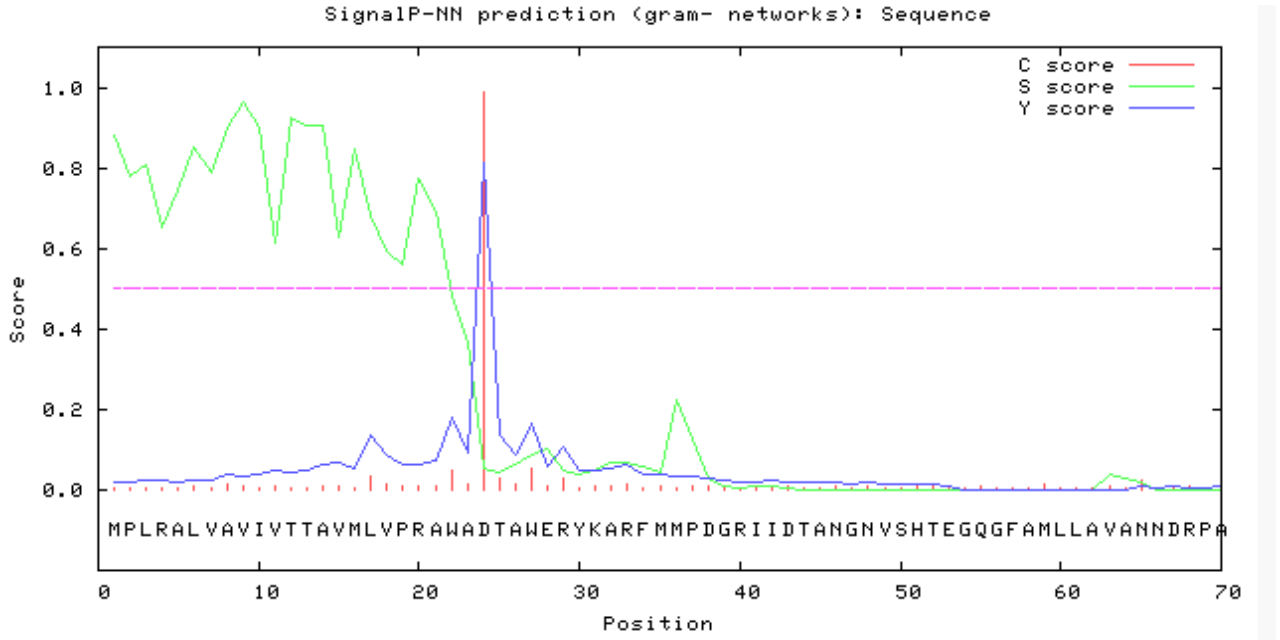
		protein)
ligB	NODE_32:5242- 6504	Protocatechuate 4,5-dioxygenase beta chain
pcaH2	NODE_1:141891-142631	Protocatechuate 3,4-dioxygenase beta chain (Ec 1.13.11.3) (3,4-pcd)
pcaG	NODE_1: 142628-143248	Protocatechuate 3,4-dioxygenase alpha chain
tcbE	NODE_25:67703-66744	Carboxymethylenebutenolidase
sodB	NODE_1:184360-184941	Superoxide dismutase [Fe]
trxA	NODE_25: 21591-21920	Thioredoxin 1
TRX2	NODE_12: 7722-7297	Thioredoxin-2
trxB	NODE_4:145309-146277	Thioredoxin reductase
SPJ_0097	NODE_18: 54203-53283	Glyoxalase family protein
ahpC	NODE_2:476462-477025	Alkyl hydroperoxide reductase
ahpF	NODE_2:477214-478779	Alkyl hydroperoxide reductase subunit F

3.6 Signal peptide prediction analysis of CAZy genes

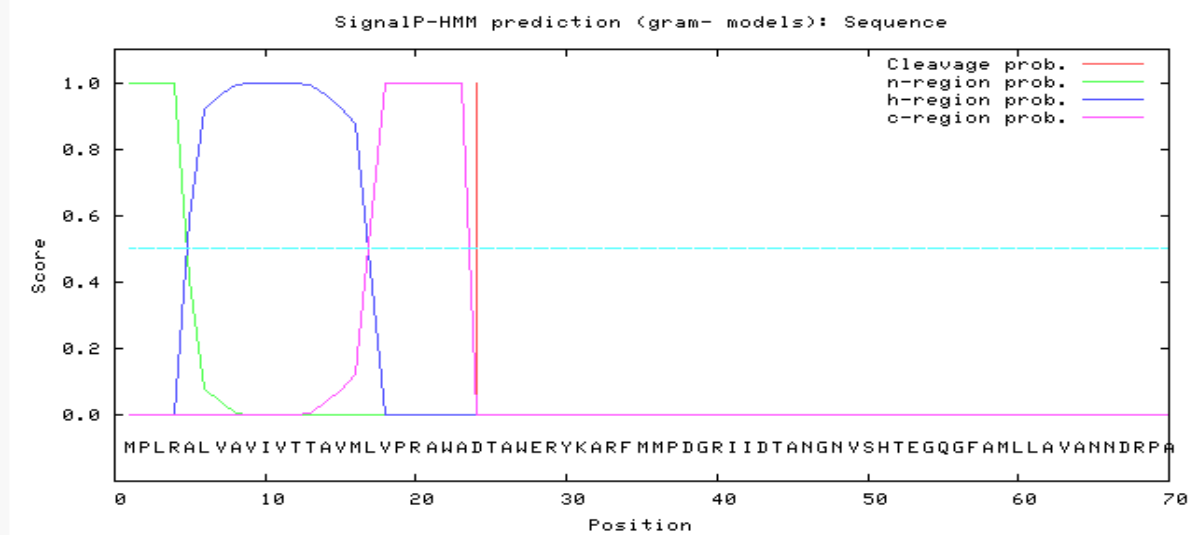
Signal peptides are found in proteins that are targeted to the endoplasmic reticulum and eventually destined to be either secreted/extracellular/periplasmic/etc., retained in the lumen of the endoplasmic reticulum, of the lysosome or of any other organelle along the secretory pathway. A number of computational tools are available for detecting signal peptides, but their abilities to locate the signal peptide cleavage sites vary significantly. We characterized a set of 7 (GH5,

GH94, GH10,CE4,GH5, GH53 and PE1) secreted CAZy proteins from the genome *Klebsiella variicola* HSTU-AAM51 by using signal peptide prediction tools (SignalP 3.0) and also found to be SignalP 3.0-NN with highly accuracy for cleavage site recognition. The extracellular accessibility of these proteins makes them ideal targets for protein therapeutics. Secreted proteins and a majority of cell-surface proteins possess an N-terminal signal peptide. The result of signal peptide prediction shows (Fig 24-31)

Endoglucanase precursor (GH5)




```
>Sequence                 length = 70
 Measure  Position  Value  Cutoff  signal peptide
  max. C    24       0.990   0.52   YES
  max. Y    24       0.811   0.33   YES
  max. S     9       0.963   0.92   YES
  mean S     1-23    0.750   0.49   YES
       D     1-23    0.780   0.44   YES
Most likely cleavage site between pos. 23 and 24: AWA-DT
```



>Sequence

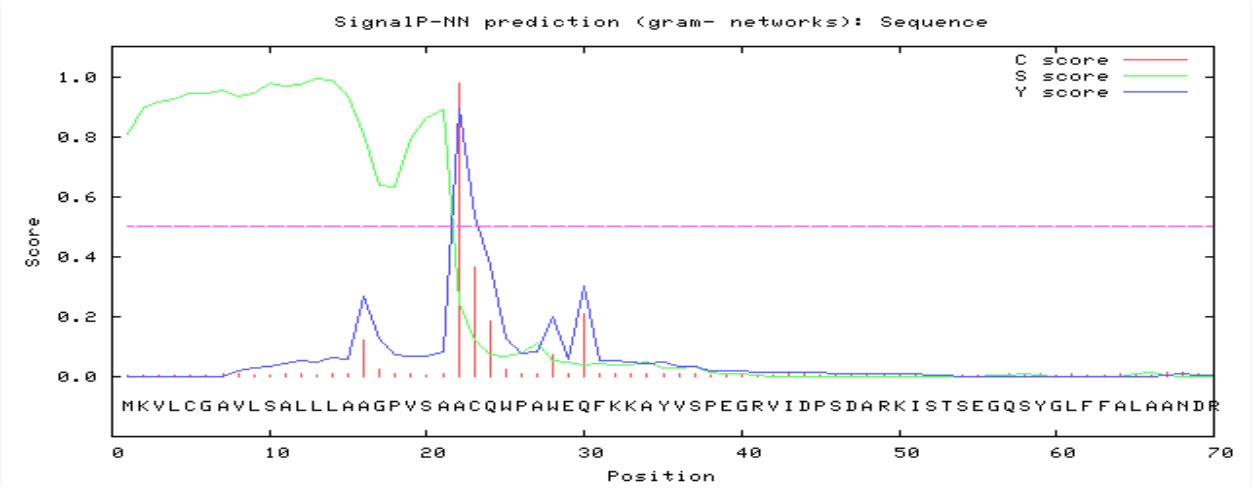
Prediction: Signal peptide

Signal peptide probability: 1.000

Max cleavage site probability: 0.999 between pos. 23 and 24

Figure 24: Signal peptide prediction Endoglucanase precursor (GH5)

Endoglucanase precursor (GH5)

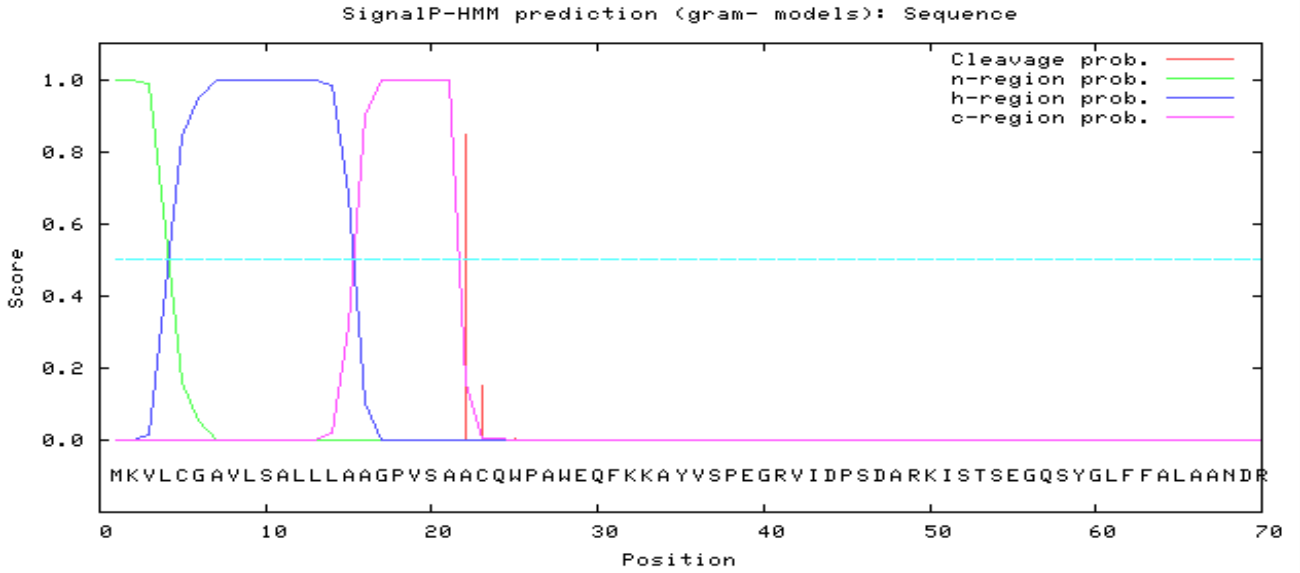


>Sequence length = 70

Measure	Position	Value	Cutoff	signal peptide
max. C	22	0.978	0.52	YES
max. Y	22	0.896	0.33	YES
max. S	13	0.991	0.92	YES
mean S	1-21	0.892	0.49	YES

D 1-21 0.894 0.44 YES

Most likely cleavage site between pos. 21 and 22: VSA-AC



>Sequence

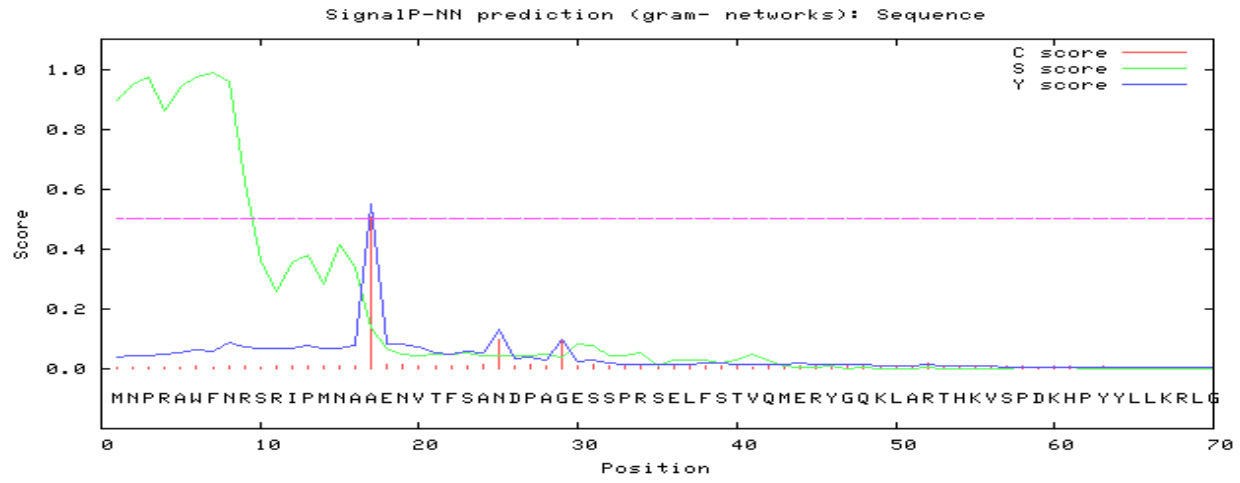
Prediction: Signal peptide

Signal peptide probability: 1.000

Max cleavage site probability: 0.845 between pos. 21 and 22

Figure 25: Signal peptide prediction Endoglucanase precursor (GH5)

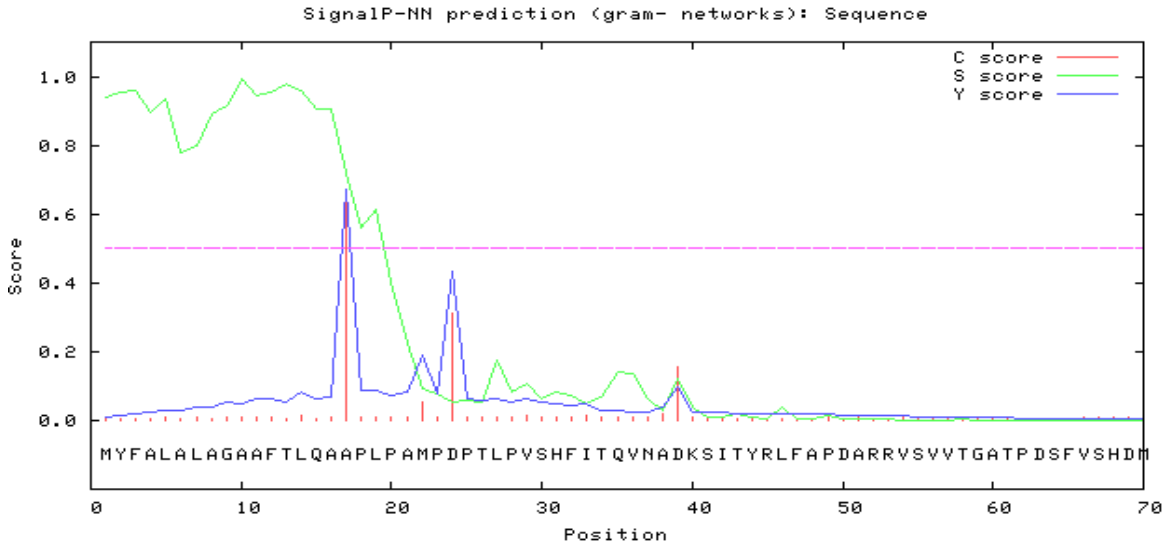
Cyclic beta-1,2-glucan synthase(GH94)



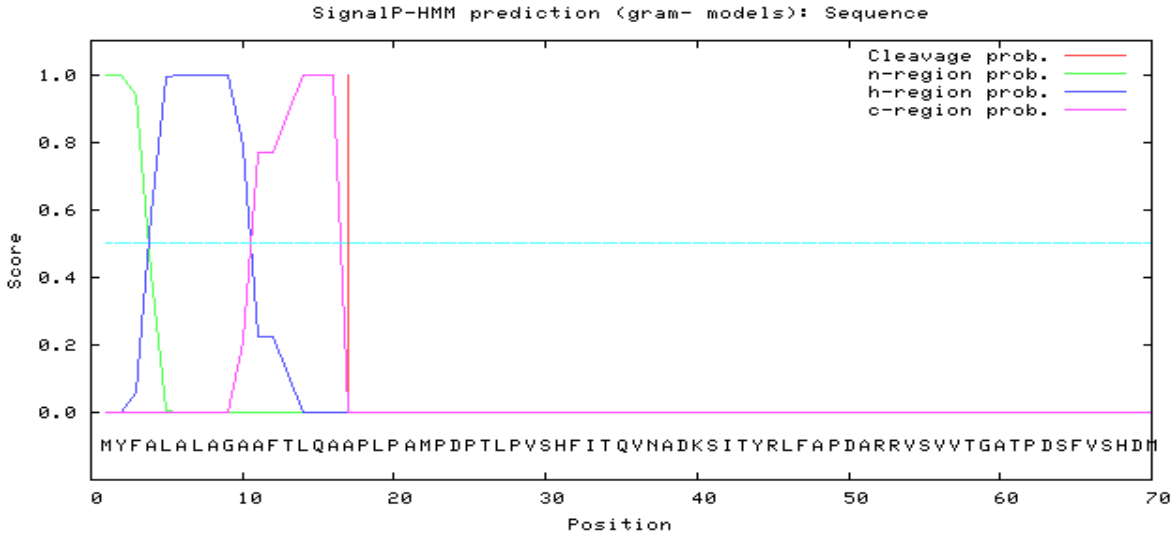
```
>Sequence                 length = 70
 Measure   Position   Value   Cutoff   signal peptide
  max. C     17        0.505    0.52    NO
  max. Y     17        0.551    0.33    YES
  max. S      7        0.987    0.92    YES
  mean S     1-16      0.658    0.49    YES
       D     1-16      0.605    0.44    YES
 Most likely cleavage site between pos. 16 and 17: MNA-AE
```

Figure 26: Signal peptide prediction of Cyclic beta-1,2-glucan synthase (GH94)

Endo-1,4-beta-xylanase A precursor (GH10)




```
>Sequence                 length = 70
 Measure   Position   Value   Cutoff   signal peptide
  max. C     17        0.644    0.52    YES
  max. Y     17        0.673    0.33    YES
  max. S     10        0.993    0.92    YES
  mean S      1-16     0.919    0.49    YES
       D      1-16     0.796    0.44    YES
 Most likely cleavage site between pos. 16 and 17: LQA-AP
```



>Sequence

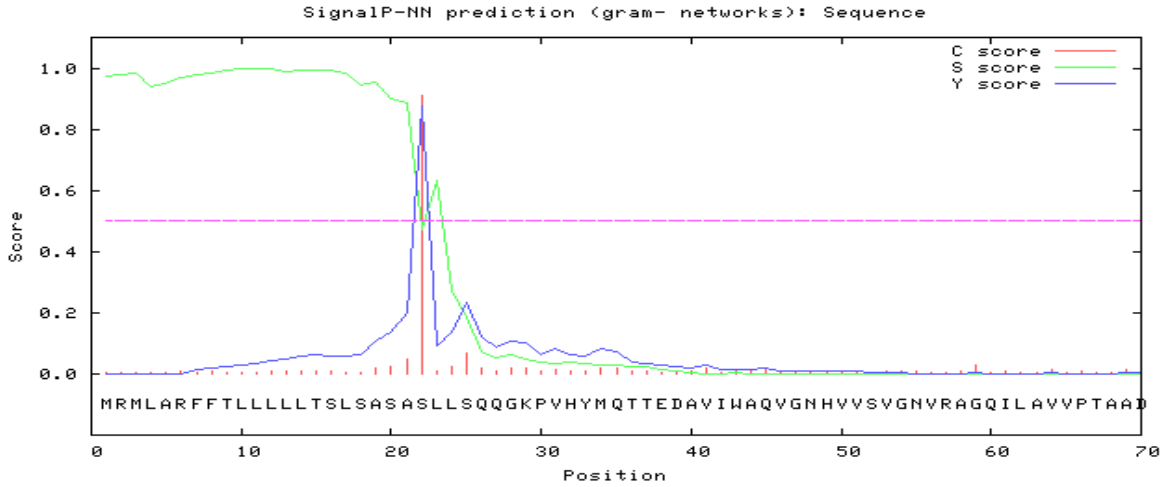
Prediction: Signal peptide

Signal peptide probability: 0.997

Max cleavage site probability: 0.997 between pos. 16 and 17

Figure 27: Signal peptide prediction of Endo-1,4-beta-xylanase A precursor (GH10)

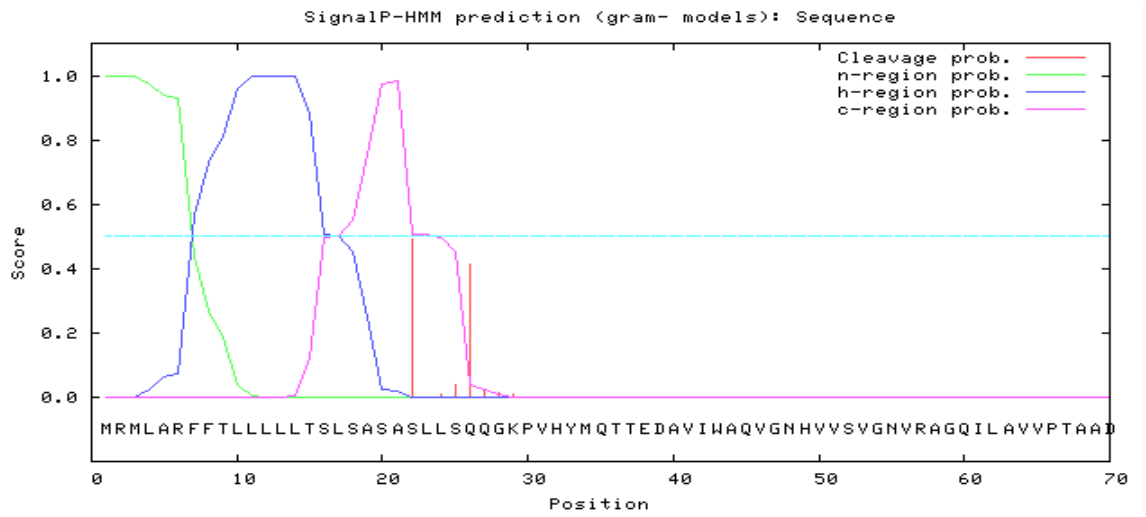
Polysaccharide deacetylase (CE4)



>Sequence length = 70

# Measure	Position	Value	Cutoff	signal peptide
max. C	22	0.910	0.52	YES
max. Y	22	0.876	0.33	YES
max. S	10	0.996	0.92	YES
mean S	1-21	0.970	0.49	YES
D	1-21	0.923	0.44	YES

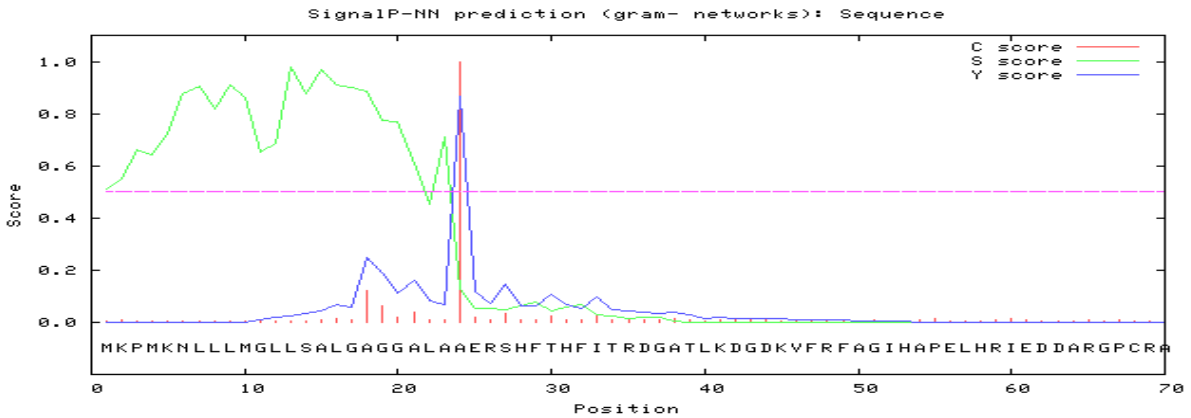
Most likely cleavage site between pos. 21 and 22: ASA-SL




```
>Sequence
Prediction: Signal peptide
Signal peptide probability: 1.000
Max cleavage site probability: 0.492 between pos. 21 and 22
```

Figure 28: Signal peptide prediction of Polysaccharide deacetylase (CE4)

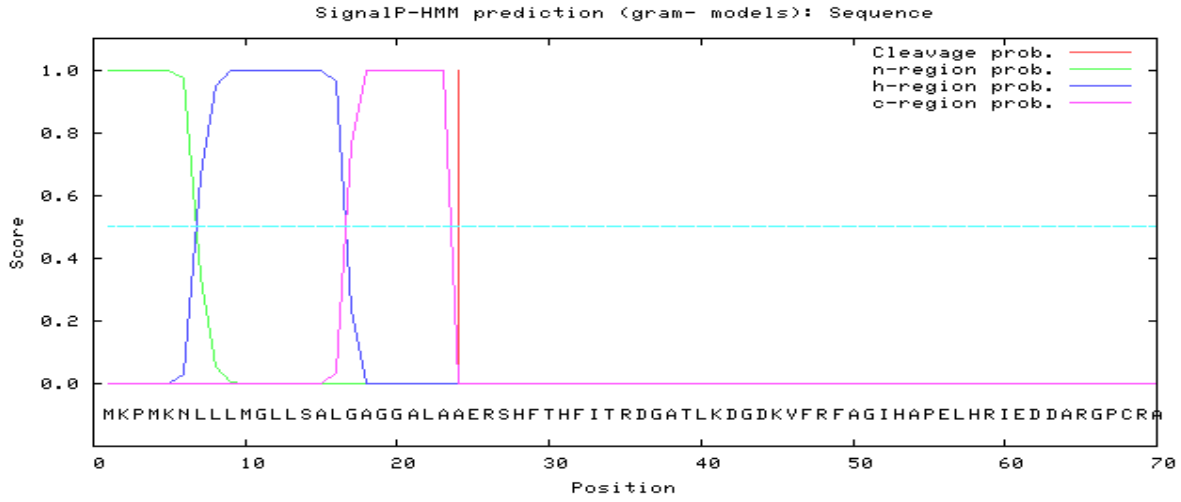
Endo-b1,4-mannanase 5C (GH5)




```
>Sequence                  length = 70
  Measure   Position   Value   Cutoff   signal peptide
  max. C     24        1.000    0.52    YES
  max. Y     24        0.868    0.33    YES
```

max.	S	13	0.976	0.92	YES
mean	S	1-23	0.767	0.49	YES
	D	1-23	0.817	0.44	YES

Most likely cleavage site between pos. 23 and 24: ALA-AE



>Sequence

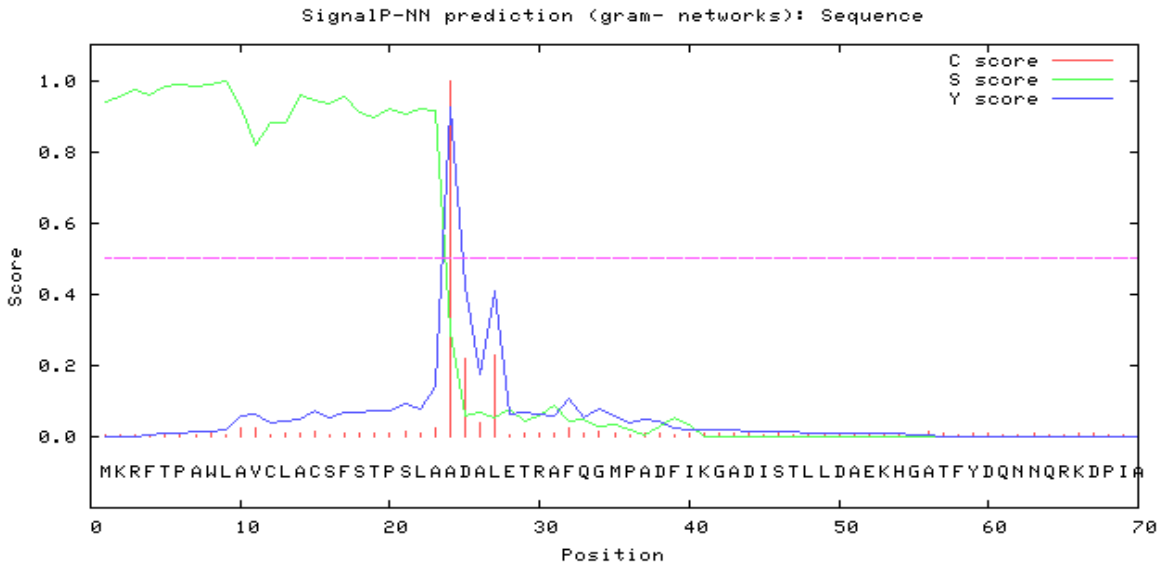
Prediction: Signal peptide

Signal peptide probability: 1.000

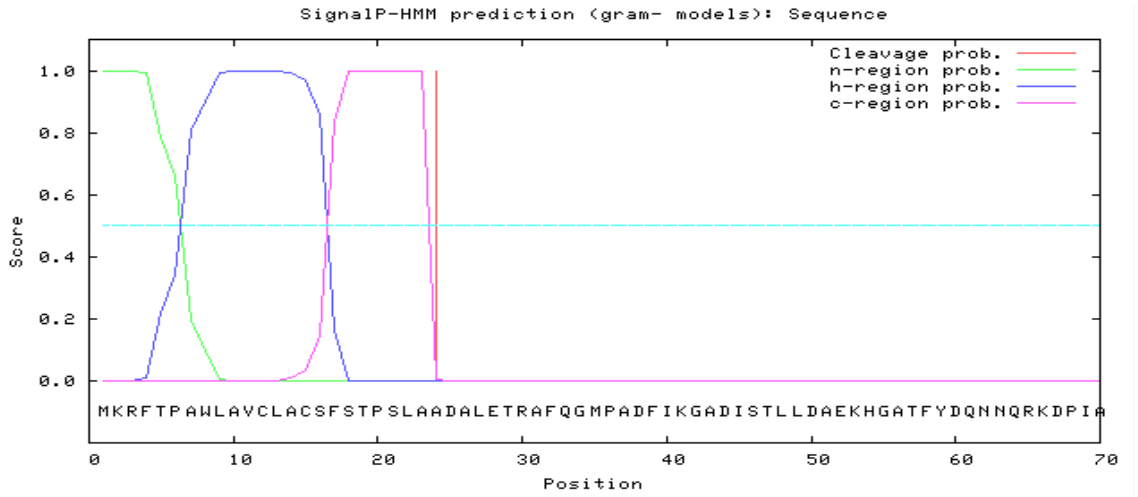
Max cleavage site probability: 0.999 between pos. 23 and 24

Figure 29: Signal peptide prediction of Endo-b1,4-mannanase 5C (GH5)

Arabinogalactan endo-1,4-beta-galactosidase(GH53)



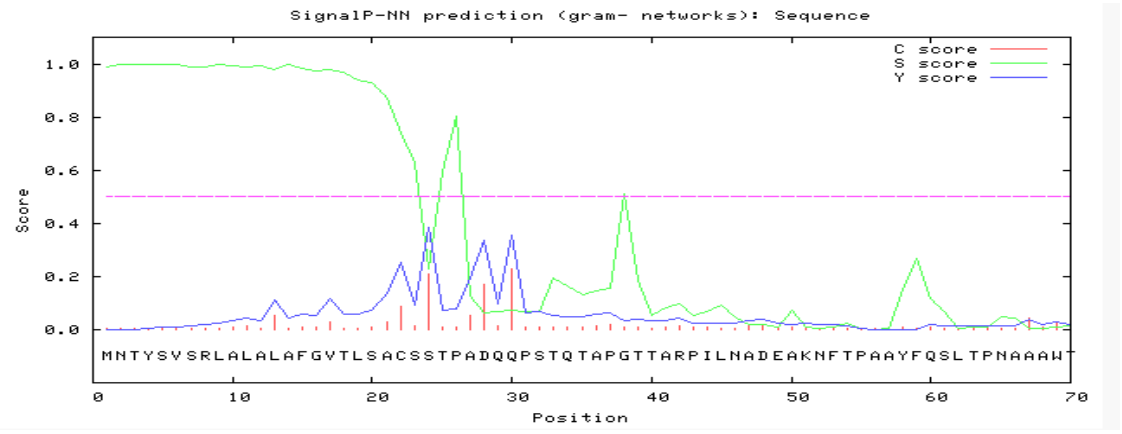
```
>Sequence                length = 70
  Measure    Position Value  Cutoff  signal peptide
  max. C     24       1.000   0.52   YES
  max. Y     24       0.926   0.33   YES
  max. S      9       0.996   0.92   YES
  mean S     1-23     0.935   0.49   YES
       D     1-23     0.931   0.44   YES
  Most likely cleavage site between pos. 23 and 24: SLA-AD
```




```
>Sequence
Prediction: Signal peptide
Signal peptide probability: 1.000
Max cleavage site probability: 0.997 between pos. 23 and 24
```

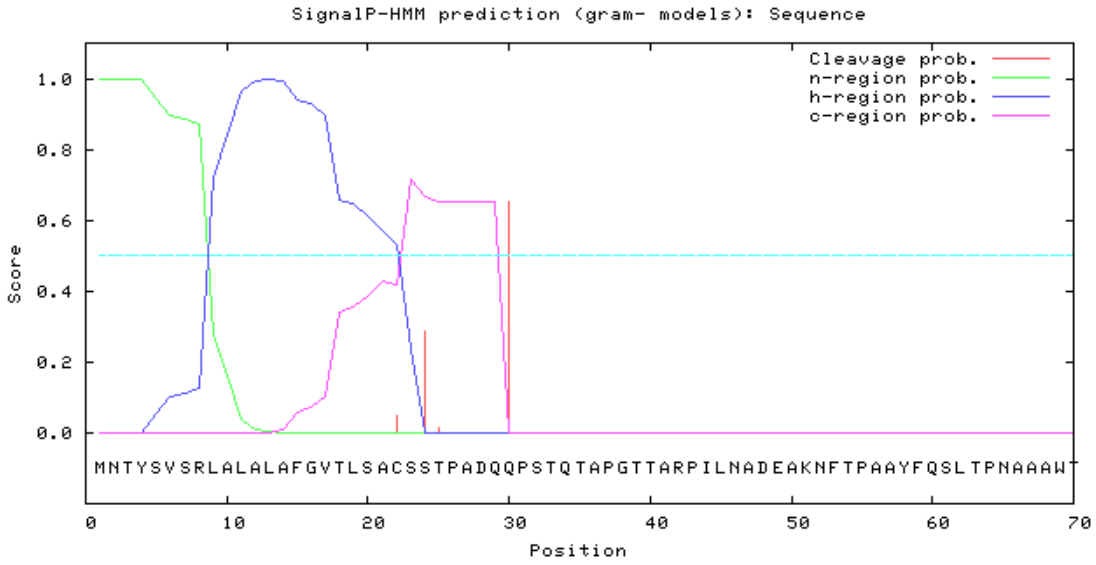
Figure 30: Signal peptide prediction of Arabinogalactan endo-1,4-beta-galactosidase(GH53)

Pectinesterase (PE1)




```
>Sequence                length = 70
```

```
  Measure   Position   Value   Cutoff   signal peptide?
  max. C     30        0.229    0.52    NO
  max. Y     24        0.385    0.33    YES
  max. S      9        0.998    0.92    YES
  mean S      1-23     0.952    0.49    YES
       D      1-23     0.669    0.44    YES
 Most likely cleavage site between pos. 23 and 24: ACS-ST
```




```
>Sequence
Prediction: Signal peptide
Signal peptide probability: 0.999
Max cleavage site probability: 0.652 between pos. 29 and 30
```

Figure 31: Signal peptide prediction of Pectinesterase (PE1)

3.7 Molecular docking analysis of CAZy genes

Cellulase is an enzyme that degrades the linear polysaccharide like cellulose into glucose by breaking the-1,4-glycosidic bonds [68]. These enzymes are the third largest enzymes with a great potential towards the ethanol production and play a vital role in degrading the biomass. The production of ethanol depends upon the ability of the cellulose, xylanose, hemicelluloses and petin to utilize the wide range of substrates. In this study, the 3D structure of cellulose, xylanose, hemicelluloses and pectinose from *Klebsiella variicola* HSTU-AAM51 was modeled by using SwissDock web server. The accuracy of the predicted 3D structure was checked using Chimera V

1.11 showed that the favored region, compatibility of an atomic model (3D) with amino acid sequence (1D) [63, 64, 65]. As the binding efficacy with the substrate might suggests the choice of the substrate as carbon and nitrogen sources were employed in the docking studies. The docking of β -D-glucopyranose, alpha & beta –D-Glucose and cellopentaose with cellulase (GH3, GH94 and GH5) exhibited the binding energy of ΔG=-10.51K cal/mol, ΔG=-6.8 K cal/mol and ΔG=-10.3 K cal/mol respectively (Fig32). The docking of β-D-xylopyranose (XYP), 4-O- β -D-xylopyranosyl-β -D-xylopyranose (BXP) and 2-(N-morpholino)-ethanesulfonic acid(MES), 2-(acetylamino)-2-deoxy-A-D-glucopyranose (NDG) with xylanase (GH10 and CE4) exhibited the binding energy of ΔG=-7.3, -6.3 K cal/mol and ΔG=-7.3, -6.46 K cal/mol respectively (Fig33). The docking of imidazole(IMD), N-acetyl-D-glucoseamine(NAG), **β**-D-glactose (GAL) and 2-[bis-(2-hydroxy-ethyl)-amino]-2-hydroxymethyl-propane-1,3-diol (BTB) with hemicellulase (GH53 and GH13) exhibited the binding energy of ΔG=-9.4, -6.7, -7.0 K cal/mol, ΔG=-7.74 K cal/mol respectively(Fig34) . The docking of **β**-D-galactopyranuronic acid (Pectin) with pectinase (PE1) exhibited the binding energy of ΔG=-6.23 K cal/mol (Fig35).

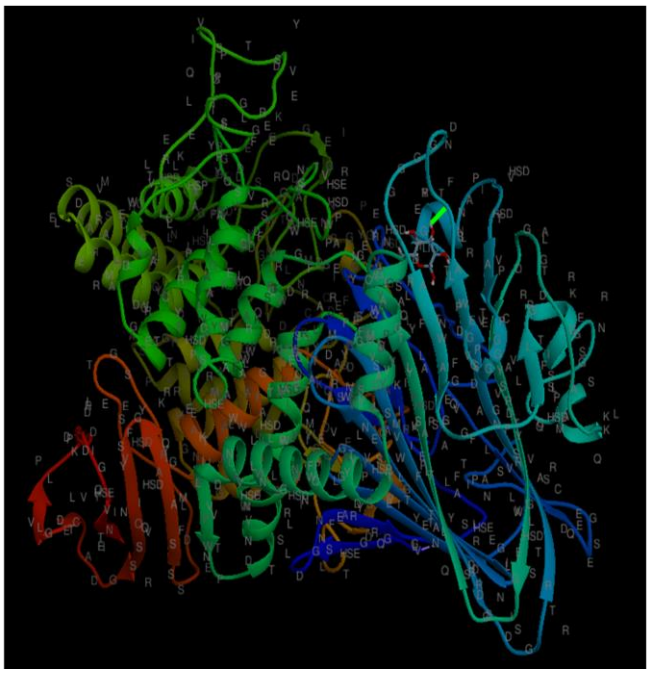
Cyclic beta-1,2-glucan synthase(GH94) + Alpha-D-Glucose

B1

REMARK Energy: -1.98809
REMARK SimpleFitness: -1.98809
REMARK FullFitness: -2912.6516
REMARK InterFull: -34.6832
REMARK IntraFull: 71.2334
REMARK solvFull: -3417.38
REMARK surfFull: 468.178
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -3417.38
REMARK deltaGcompsolvnonpol: 468.178
REMARK deltaGprotsolvpol: -3421.44
REMARK deltaGprotsolvnonpol: 468.838
REMARK deltaGligsolvpol: -14.4765
REMARK deltaGligsolvnonpol: 6.51488
REMARK deltaGvdw: -34.6832
REMARK deltaGelec: 0.0
REMARK deltaG: -6.8153377
REMARK Cluster: 0
REMARK ClusterRank: 0

Cyclic beta-1,2-glucan synthase(GH94) + Beta-D-Glucose

B2



REMARK Energy: -4.71274
REMARK SimpleFitness: -4.71274
REMARK FullFitness: -2913.5735
REMARK InterFull: -35.5191
REMARK IntraFull: 69.7835
REMARK solvFull: -3416.14
REMARK surfFull: 468.302
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -3416.14
REMARK deltaGcompsolvnonpol: 468.302
REMARK deltaGprotsolvpol: -3421.44
REMARK deltaGprotsolvnonpol: 468.838
REMARK deltaGligsolvpol: -13.4755
REMARK deltaGligsolvnonpol: 6.47544
REMARK deltaGvdw: -35.5191
REMARK deltaGelec: 0.0
REMARK deltaG: -6.8643265
REMARK Cluster: 0
REMARK ClusterRank: 0

Endoglucanase (GH5)+ CELLOPENTAOSE

C

REMARK Energy: 114.85
REMARK SimpleFitness: 114.85
REMARK FullFitness: -1269.293
REMARK InterFull: -102.372
REMARK IntraFull: 396.281
REMARK solvFull: -1781.65
REMARK surfFull: 218.448
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -1781.65
REMARK deltaGcompsolvnonpol: 218.448
REMARK deltaGprotsolvpol: -1818.59
REMARK deltaGprotsolvnonpol: 217.539
REMARK deltaGligsolvpol: -45.4234
REMARK deltaGligsolvnonpol: 18.7098
REMARK deltaGvdw: -102.372
REMARK deltaGelec: 0.0
REMARK deltaG: -10.317711
REMARK Cluster: 0
REMARK ClusterRank: 0

Figure 32: A, (B1,B2), and C indicates image of substrate binding cellulase enzymes.

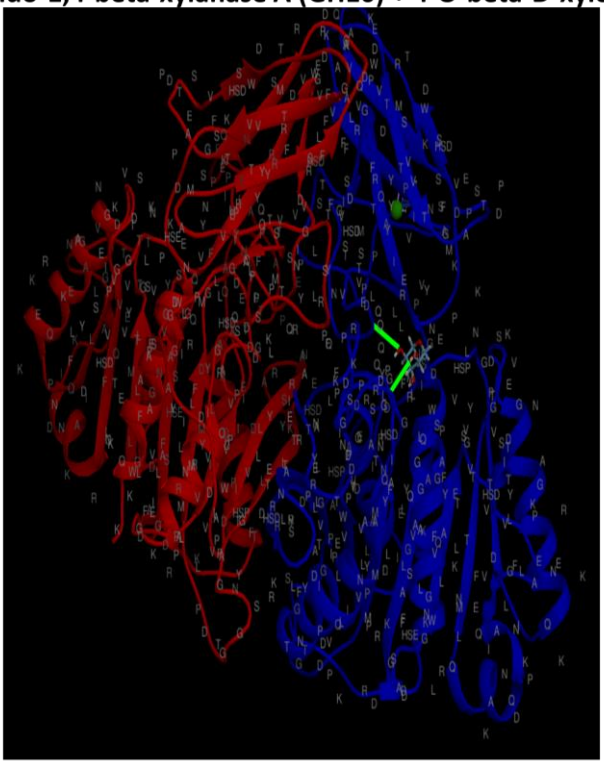
Endo-1,4-beta-xylanase A (GH10) + BETA-D-XYLOPYRANOSE(XYP)

D1

REMARK Energy: -2.73671
REMARK SimpleFitness: -2.73671
REMARK FullFitness: -3414.941
REMARK InterFull: -49.0966
REMARK IntraFull: 73.0936
REMARK solvFull: -3919.4
REMARK surfFull: 480.462
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -3919.4
REMARK deltaGcompsolvnonpol: 480.462
REMARK deltaGprotsolvpol: -3937.33
REMARK deltaGprotsolvnonpol: 481.501
REMARK deltaGligsolvpol: -15.4305
REMARK deltaGligsolvnonpol: 6.52213
REMARK deltaGvdw: -49.0966
REMARK deltaGelec: 0.0
REMARK deltaG: -7.308148
REMARK Cluster: 0
REMARK ClusterRank: 0

Endo-1,4-beta-xylanase A (GH10) + 4-O-beta-D-xylopyranosyl-beta-D-xylopyranose(BXP)

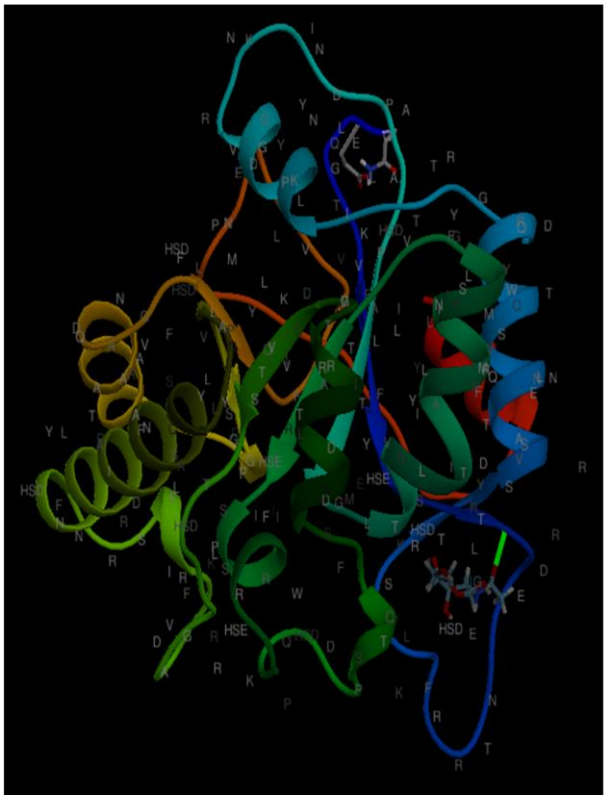
D2



REMARK Energy: -2.62819
REMARK SimpleFitness: -2.62819
REMARK FullFitness: -3423.8
REMARK InterFull: -31.8044
REMARK IntraFull: 60.5525
REMARK solvFull: -3933.06
REMARK surfFull: 480.512
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -3933.06
REMARK deltaGcompsolvnonpol: 480.512
REMARK deltaGprotsolvpol: -3937.33
REMARK deltaGprotsolvnonpol: 481.501
REMARK deltaGligsolvpol: -15.4737
REMARK deltaGligsolvnonpol: 5.70157
REMARK deltaGvdw: -31.8044
REMARK deltaGelec: 0.0
REMARK deltaG: -6.42331
REMARK Cluster: 0
REMARK ClusterRank: 0

Polysaccharide deacetylase(CE4) + 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (MES)

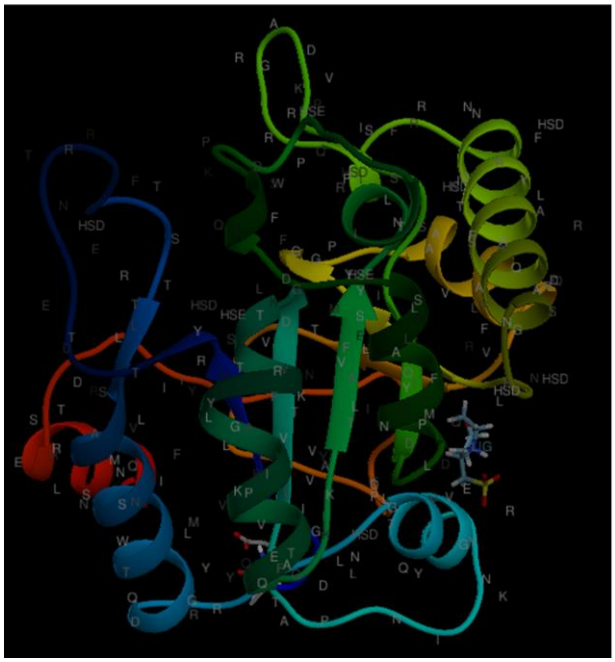
E1



REMARK Energy: -3.46629
REMARK SimpleFitness: -3.46629
REMARK FullFitness: -1197.9524
REMARK InterFull: -50.6324
REMARK IntraFull: 60.526
REMARK solvFull: -1397.37
REMARK surfFull: 189.524
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -1397.37
REMARK deltaGcompsolvnonpol: 189.524
REMARK deltaGprotsolvpol: -1412.43
REMARK deltaGprotsolvnonpol: 189.459
REMARK deltaGligsolvpol: -18.8174
REMARK deltaGligsolvnonpol: 6.97593
REMARK deltaGvdw: -50.6324
REMARK deltaGelec: 0.0
REMARK deltaG: -7.3602986
REMARK Cluster: 0
REMARK ClusterRank: 0

Polysaccharide deacetylase (CE4) + 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE (NDG)

E2

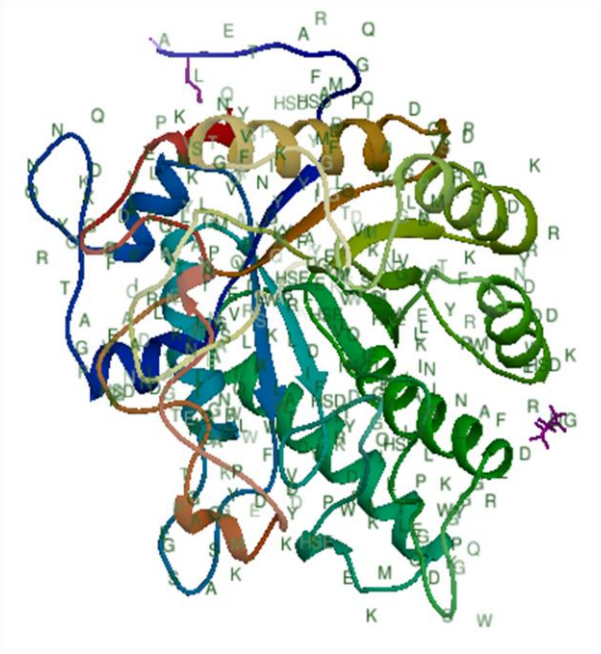


REMARK Energy: -11.6733
REMARK SimpleFitness: -11.6733
REMARK FullFitness: -1274.208
REMARK InterFull: -32.4287
REMARK IntraFull: -3.63935
REMARK solvFull: -1427.51
REMARK surfFull: 189.37
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -1427.51
REMARK deltaGcompsolvnonpol: 189.37
REMARK deltaGprotsolvpol: -1412.43
REMARK deltaGprotsolvnonpol: 189.459
REMARK deltaGligsolvpol: -34.2955
REMARK deltaGligsolvnonpol: 6.06814
REMARK deltaGvdw: -32.4287
REMARK deltaGelec: 0.0
REMARK deltaG: -6.4642296
REMARK Cluster: 0
REMARK ClusterRank: 0

Figure 33: (D1,D2) and (E1,E2), indicates image of substrate binding xylanase enzymes.

Endo-1,4-beta-galactosidase(GH53) + IMIDAZOLE(IMD)

F1



REMARK Energy: -22.7195
REMARK SimpleFitness: -22.7195
REMARK FullFitness: -1937.0956
REMARK InterFull: -133.928
REMARK IntraFull: 3.35536
REMARK solvFull: -2051.05
REMARK surfFull: 244.527
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -2051.05
REMARK deltaGcompsolvnonpol: 244.527
REMARK deltaGprotsolvpol: -2064.29
REMARK deltaGprotsolvnonpol: 244.539
REMARK deltaGligsolvpol: -106.126
REMARK deltaGligsolvnonpol: 3.03015
REMARK deltaGvdw: -133.928
REMARK deltaGelec: 0.0
REMARK deltaG: -9.642178
REMARK Cluster: 0
REMARK ClusterRank: 0

Endo-1,4-beta-galactosidase(GH53) + N-ACETYL-D-GLUCOSAMINE(NAG)

F2

REMARK Energy: -5.28049
REMARK SimpleFitness: -5.28049
REMARK FullFitness: -1796.0424
REMARK InterFull: -32.0094
REMARK IntraFull: 53.4541
REMARK solvFull: -2061.58
REMARK surfFull: 244.093
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -2061.58
REMARK deltaGcompsolvnonpol: 244.093
REMARK deltaGprotsolvpol: -2064.29
REMARK deltaGprotsolvnonpol: 244.539
REMARK deltaGligsolvpol: -13.875
REMARK deltaGligsolvnonpol: 7.10195
REMARK deltaGvdw: -32.0094
REMARK deltaGelec: 0.0
REMARK deltaG: -6.717126
REMARK Cluster: 0
REMARK ClusterRank: 0

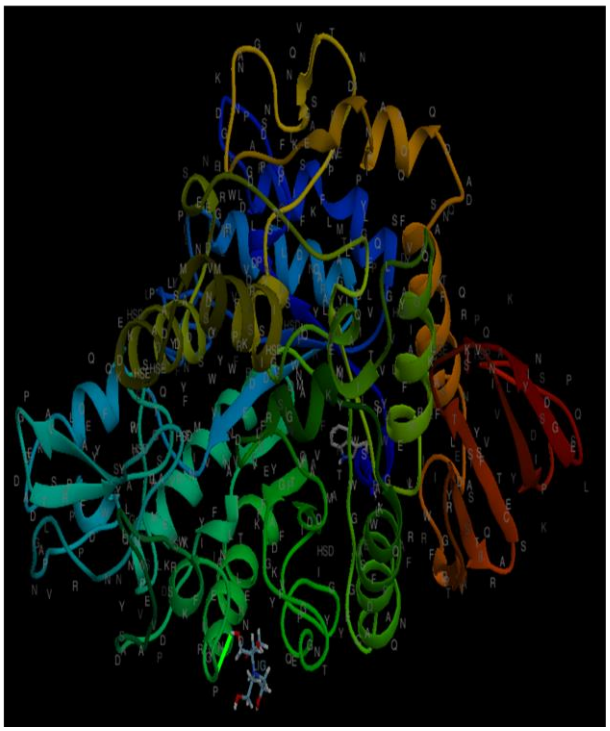
Endo-1,4-beta-galactosidase (GH53) + BETA-D-GALACTOSE (GAL)

F3

REMARK Energy: 4.16707
REMARK SimpleFitness: 4.16707
REMARK FullFitness: -1774.6725
REMARK InterFull: -35.3551
REMARK IntraFull: 80.2797
REMARK solvFull: -2063.78
REMARK surfFull: 244.183
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -2063.78
REMARK deltaGcompsolvnonpol: 244.183
REMARK deltaGprotsolvpol: -2064.29
REMARK deltaGprotsolvnonpol: 244.539
REMARK deltaGligsolvpol: -15.3371
REMARK deltaGligsolvnonpol: 6.51227
REMARK deltaGvdw: -35.3551
REMARK deltaGelec: 0.0
REMARK deltaG: -7.014488
REMARK Cluster: 0
REMARK ClusterRank: 0

Glycoside Hydrolase + 2-[BIS-(2-HYDROXY-ETHYL)-AMINO]-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (BTB)

G

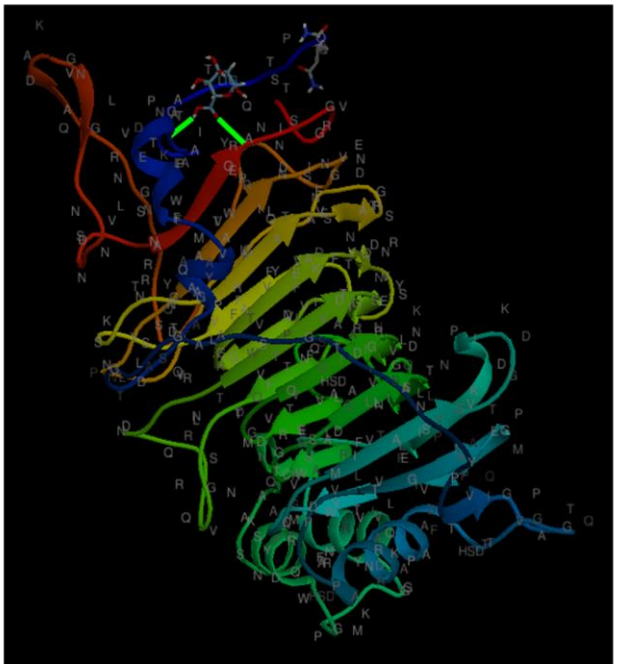


REMARK Energy: 11.0927
REMARK SimpleFitness: 11.0927
REMARK FullFitness: -2464.8303
REMARK InterFull: -66.3234
REMARK IntraFull: 130.129
REMARK solvFull: -2871.91
REMARK surfFull: 343.274
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -2871.91
REMARK deltaGcompsolvnonpol: 343.274
REMARK deltaGprotsolvpol: -2891.93
REMARK deltaGprotsolvnonpol: 343.824
REMARK deltaGligsolvpol: -34.0026
REMARK deltaGligsolvnonpol: 7.74289
REMARK deltaGvdw: -66.3234
REMARK deltaGelec: 0.0
REMARK deltaG: -7.7425184
REMARK Cluster: 0
REMARK ClusterRank: 0

Figure 34: (F1, F2, F3) and G indicates image of substrate binding hemicellulase enzymes

Pectinesterase (PE1) + BETA-D-GALACTOPYRANURONIC ACID (Pectin)

H



REMARK Energy: 9.45545
REMARK SimpleFitness: 9.45545
REMARK FullFitness: -1184.7937
REMARK InterFull: -27.7861
REMARK IntraFull: 68.2894
REMARK solvFull: -1474.62
REMARK surfFull: 249.323
REMARK extraFull: 0.0
REMARK deltaGcompsolvpol: -1474.62
REMARK deltaGcompsolvnonpol: 249.323
REMARK deltaGprotsolvpol: -1473.61
REMARK deltaGprotsolvnonpol: 248.519
REMARK deltaGligsolvpol: -16.2004
REMARK deltaGligsolvnonpol: 6.62473
REMARK deltaGvdw: -27.7861
REMARK deltaGelec: 0.0
REMARK deltaG: -6.2395544
REMARK Cluster: 0
REMARK ClusterRank: 0

Figure 35: H indicate image of substrate binding pectinase enzyme.

CHAPTER-IV

SUMMARY AND CONCLUSION

Conclusion

Lignocellulose distortion is a groundbreaking work that requires the usage of cellulase, xylanase and pectinase enzymes. A big portion of these enzymes come from bacteria. This part represents the different enzymatic assay analysis along with biochemical analysis for the partial characterization of the isolated strains from goat rumen. After which one of the strains HSTU-AAM51 have been subjected to NGS whole genome sequencing in order to fully characterize the strains and find the subsequent enzymes needed for lignocellulose breakdown. Different bioinformatics tools have been used to annotate the whole genomic sequence, the enzyme characterization, probability of finding signal peptides, pan genomic analysis, active site determination and specific substrate docking for ensuring that the strain in question has the capability to break down lignocelluloses. This data that we generated will serve as a valuable source for future research in environmental biochemistry. As It has been ensured that the strain in question has the enzyme capable of degrading the lignocellulose structure, It is definite that the source of this bacterial community will serve greatly in this area of research. it is now a matter of time when the world will be changing its focus from conventional process to bio-ethanol production. By doing so, we can have an unlimited supply of natural bio-fuel that can solve the world's energy crisis.

REFERENCES

1. R.A. Batista-García, M. del Rayo Sánchez-Carbente, P. Talia, S.A. Jackson, N.D. O'Leary, A.D. Dobson, J.L. Folch-Mallol, From lignocellulosic metagenomes to lignocellulolytic genes: trends, challenges and future prospects, Biofuels Bioprod. Biorefin. 10 (2016) 864–882.
2. Z. Anwar, M. Gulfraz, M. Irshad, Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: a brief review, J. Radiat. Res. Appl. Sci. 7 (2014) 163–173.
3. P.E. Marriott, L.D. Gómez, S.J. McQueen-Mason, Unlocking the potential of lignocellulosic biomass through plant science, New Phytol. 209 (2016) 1366–1381.
4. D. Kumari, R. Singh, Pretreatment of lignocellulosic wastes for biofuel production: a critical review, Renew. Sust. Energ. Rev. 90 (2018) 877–891.
5. S.N. Shafawati, S. Siddiquee, Composting of oil palm fibres and Trichoderma spp. as the biological control agent: a review, Int. Biodeterior. Biodegradation 85 (2013) 243–253.
6. V. Juturu, J.C. Wu, Microbial cellulases: engineering, production and applications, Renew. Sust. Energ. Rev. 33 (2014) 188–203.
7. V. Juturu, J.C. Wu, Microbial xylanases: engineering, production and industrial applications, Biotechnol. Adv. 30 (2012) 1219–1227.
8. S. Malgas, J.S. van Dyk, B.I. Pletschke, A review of the enzymatic hydrolysis of mannans and synergistic interactions between β-mannanase, β-mannosidase and α- galactosidase, World J. Microbiol. Biotechnol. 31 (2015) 1167–1175.
9. G. de Gonzalo, D.I. Colpa, M.H. Habib, M.W. Fraaije, Bacterial enzymes involved in lignin degradation, J. Biotechnol. 236 (2016) 110–119.
10. V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) D490–D495.
11. N. Kamsani, M.M. Salleh, A. Yahya, C.S. Chong, Production of lignocellulolytic enzymes by microorganisms isolated from Bulbitermes sp. termite gut in solid-state fermentation, Waste Biomass Valorization 7 (2016) 357–371.
12. H.K. Sharma, C. Xu, W. Qin, Biological pretreatment of lignocellulosic biomass for biofuels and bioproducts: an overview, Waste Biomass Valorization 10 (2019) 235–251.
13. Puentes-téllez, P. E., & Salles, J. F. (2018). *Construction of Effective Minimal Active Microbial Consortia for Lignocellulose Degradation*. 419–429.
14. de Souza, W. (2013). *Microbial Degradation of Lignocellulosic Biomass*. https://doi.org/10.5772/54325.
15. Rabinovich, M.L., M. S. M. and A. V. B. (2002). The structure and mechanism of action of cellulolytic enzymes. *Biochemistry (Moscow)*, *67*, 850–871.

16. Noura EL-Ahmady EL Naggar, S. D. and A. khalil. (2014). *bioethanol production from lignocellulosic feedstock.pdf*.
17. Kuhad, R. C., Gupta, R., & Singh, A. (2011). Microbial cellulases and their industrial applications. *Enzyme Research*, *2011*(1). https://doi.org/10.4061/2011/280696
18. Gaur, N., Narasimhulu, K., & Pydi Setty, Y. (2018). *Extraction of ligninolytic enzymes from novel Klebsiella pneumoniae strains and its application in wastewater treatment. Applied Water Science, 8(4).* doi:10.1007/s13201-018-0758-y
19. Zhu, J.Y., X.J. Pan, G. S. W. and R. G. (2009). Sulfite pretreatment (SPORL) for robust enzymatic saccharification of spruce and red pine. *Bioresour. Technol*, *100*, 2411– 2418.
20. Sun, Y. and J. C. (2002). Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresour. Technol*, *Bioresour.*, 1–11.
21. Mais, U., A.R. Esteghlalian, J. N. S. and S. D. M. (2002). Enhancing the enzymatic hydrolysis of cellulosic materials using simultaneous ball milling. *Applied Biochem. Biotechnol.*, *98*, 815–832.
22. Palmowski, L. and J. M. (1999). *Influence of the size reduction of organic waste on their anaerobic digestion.Proceedings of the 2nd International Symposium on Anaerobic Digestion of Solid Waste*. Influence of the size reduction of organic waste o.
23. Taherzadeh, M. J. and K. K. (2007). Acid-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *BioResources*, *2*, 472–499.
24. Gharpuray, M.M., Y. H. L. and L. T. F. (1983). Structural modification of lignocellulosics by treatment to enhance enzymatic hydrolysis. *Biotechnol. Bioeng.*, *25*, 157–172.
25. Yang, B. and C. E. W. (2004). Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. *Biotechnol. Bioeng.*, *86*, 88–98.
26. Fan, L.T., Y. H. L. and D. H. B. (1980). Mechanism of the enzymatic hydrolysis of cellulose: Effects of major structural features of cellulose on enzymatic hydrolysis. *Biotechnol. Bioeng.*, *22*, 177–199.
27. Tarkow, H. and W. C. F. (1969). A Mechanism for Improving the Digestibility of Lignocellulosic Materials with Dilute Alkali and Liquid Ammonia. In: Cellulases and Their Applications. *American Chemical Society, Washington, DC.*, 197–218.
28. Chang, V.S., B. B. and M. T. H. (1997). *Lime Pretreatment of Switchgrass. Proceedings of the 18th Symposium on Biotechnology for Fuels and Chemicals*. 3–19.

29. Neely, W. C. (1984). Factors affecting the pretreatment of biomass with gaseous ozone. *Biotechnol. Bioeng., 26*, 59–65.
30. Chandra, R.P., R. Bura, W.E. Mabee, A. Berlin, X. P. and J. N. S. (2007). Substrate pretreatment the key to effective enzymatic hydrolysis of lignocellulosics. *Adv. Biochem. Eng. Biotechnol., 108*, 67–93.
31. Bai, S., Ravi kumar, M., Mukesh kumar, D., Balashanmugam, P., kumaran, B., & Kalaichelvan, P. (2012). Cellulase Production by Bacillus subtilis isolated from Cow Dung. *Archives of Applied Science Research*, *4*(1), 269–279.
32. Jorgensen, H., T. Eriksson, J. Borjesson, F. T. and L. O. (2003). Purification and characterization of five cellulases and one xylanase from Penicillium brasilianum IBT 20888. *Enzyme Microbial Technol., 32*, 851–861.
33. Prates, J.A., N. Tarbouriech, S.J. Charnock, C.M. Fontes, L. F. and G. J. D. (2001). *The structure of the feruloyl esterase module of xylanase 10B from Clostridium thermocellum provides insights into substrate recognition*. *9*, 1183–1190.
34. Mishra, V. K., Rawat, V., Rana, S., & Dubey, J. (2015). *Isolation and Screening of Cellulose Degrading Microorganisms and Evaluation of its Cellulolytic Activity Isolation and Screening of Cellulose Degrading Microorganisms and Evaluation of its Cellulolytic Activity*. (November).
35. Williams, J. G. H. N. R. K. P. H. A. S. J. T. S. S. T. (1994). *Bergey's Manual of Determinative Bacteriology (9th ed.)*.
36. Colco, R. (2005). *Gram Staining*. 3–4. https://doi.org/10.1002/9780471729259.mca03cs00
37. Balows, A. (1975). Bergey's Manual of Determinative Bacteriology. Eighth Edition. *American Journal of Public Health*, *65*(3), 315. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1778052/
38. Voges, O. and B. P. (1898). *Voges Proskauer test*.
39. WERKMAN, C. H. (1930). AN IMPROVED TECHNIC FOR THE VOGESPROSKAUER TEST. *Bacteriology Section, Iowa Agricultural Experiment Station, Ames*.
40. Warrack and Clarke , 1947 Oakley et aZ . concluded that the results of carboh. *3*(July 1969), 283–289.
41. Coulter, E. W. (1931). *COLI-AEROGENES DIFFERENTIATION IN WATER*. (1925), 125–181.

42. KOSER, S. A. (1923). *CORRELATION OF CITRATE UTILIZATION BY MEM- BERS OF THE COLON-AEROGENES GROUP WITH OTHER DIFFERENTIAL CHARACTERISTICS AND WITH HABITAT*.
43. J., K. (1982). Ehrlich's benzaldehyde reaction (with urobilinogen) 80 years later. 20, 424–428.
44. Turner, B. Y. J. M. (1961). *A New Reagent for the Assay of Indole in the Tryptophanase Reaction E0-5*. 790–792.
45. Christensen, W. B. (1946). Urea decomposition as a means of differentiating Proteus and paracolon cultures from each other and from Salmonella and Shigella types. *J. Bacteriol., 52*, 461-466.
46. Vuye, A., & Pijck, J. (1973). *Urease Activity*. *26*(6), 850–854.
47. Rutter, J. M. (1969). *STUDY OF THE CARBOHYDRATE FERMENTATION REACTIONS OF is based on the detection of different permutations of the soluble antigens in culture products of the organism (Oakley , Warrack and Clarke , 1947). Oakley et aZ . concluded that the results of carboh*. *3*(July 1969), 283–289.
48. Rückert, C., Albersmeier, A., Winkler, A., & Tauch, A. (2015). *Complete Genome Sequence of Corynebacterium kutscheri DSM 20755, a Corynebacterial Type Strain with Remarkably Low G+C Content of Chromosomal DNA .Genome Announcements, 3(3).* doi:10.1128/genomea.00571-15
49. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). *SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455–477.* doi:10.1089/cmb.2012.0021
50. Seemann, T. (2014). *Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14), 2068–2069.* doi:10.1093/bioinformatics/btu153
51. Thomsen MC et al. A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance. PLoS One. 2016;11(6):e0157718
52. Carattoli A et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
53. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A.,Zagnitko, O. (2008). *The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics, 9(1), 75.* doi:10.1186/1471-2164-9-75

54. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). *KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research, 35(Web Server), W182–W185.* doi:10.1093/nar/gkm321
55. Kanehisa,M., Goto,S., Hattori,M., Aoki-Kinoshita,K.F., Itoh,M., Kawashima,S., Katayama,T., Araki,M. and Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res., 34, D354–D357.
56. Richter M, Rosselló-Móra R, Glöckner FO, and Peplies J (2015) *JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison.* Bioinformatics. 2015 Nov 16. pii: btv681
57. http://bioinformatics.psb.ugent.be/webtools/Venn/
58. Petkau, A., Stuart-Edwards, M., Stothard, P., & Van Domselaar, G. (2010). *Interactive microbial genome visualization with GView. Bioinformatics, 26(24), 3125–3126.* doi:10.1093/bioinformatics/btq588
59. Stothard,P. and Wishart,D.S. (2005) Circular genome visualization and exploration using CGView. Bioinformatics, 21, 537–539.
60. Arakawa K, Tomita M. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evol Bioinform Online*. 2007;3:159–168. Published 2007 Sep 6.
61. Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., & Brunak, S. (2004). *Improved Prediction of Signal Peptides: SignalP 3.0. Journal of Molecular Biology, 340(4), 783–795.* doi:10.1016/j.jmb.2004.05.028
62. Schwede, T. (2003). *SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research, 31(13), 3381–3385.* doi:10.1093/nar/gkg520
63. Binkowski, T. A. (2003). *CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Research, 31(13), 3352–3355.* doi:10.1093/nar/gkg512
64. *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*
65. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J, UCSF Chimera--a visualization system for exploratory research and analysis, Comput Chem.* 2004 Oct;25(13):1605-12.
66. Grosdidier, A., Zoete, V., & Michielin, O. (2011). *SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Research, 39(suppl), W270–W277.* doi:10.1093/nar/gkr366

67. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al.Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315:804–7.
68. Argyropoulos DS, Menachem SB. Lignin. Biotechnol Pulp Pap Ind. 1997;57: 127–58.
69. Anderson WF, Akin DE. Structural and chemical properties of grass lignocelluloses related to conversion for biofuels. J Ind Microbiol Biotechnol. 2008;35:355–66.
70. Pollegioni L, Tonin F, Rosini E. Lignin-degrading enzymes. FEBS J. 2015;282: 1190–213.

APPENDICES

APPENDIX A

CHEMICALS

Table A.1. Chemicals Used in Experiments

Sr.	Chemical Name	Code
1	Agar-Agar Merck	Merck 1.01613
2	D-Glucose	AppliChem A3666
3	Yeast Extract	Merck 1.03753
4	Glycerol	AppliChem A2926
5	NaCl	Biochem/ and NIB
6	K_2HPO_4	AppliChem A2942
7	$MgSO_4.7H_2O$	Merck 1.05886
8	Polygalacturonic acid	Fluka 81325
9	H_2SO_4	Merck 1.0735
10	Carboxymethyl cellulose sodium salt	Fluka Biochemical 21902 14 Soluble Starch Merck 1.01252.0250
11	Ammonium sulfate	AppliChem A3485
12	Potassium Di-hydrogen Phosphate	Merck 1.04871
13	Immersion oil	AppliChem A0699
14	Calcium chloride	AppliChem A3652
15	Crystal violet	Sigma C3886
16	Safranin	Merck 1.15948
17	Nutrient broth	Merck1.05443
18	Congo Red	Sigma C6767
19	Potassium Iodide	Sigma P8256
20	DNA isolation kit	AppliChem A3421
21	N, N, N`, N`-Tetra methyl-p-phenyl	enedione Sigma T3134
22	Tris Base	Sigma T6066
23	EDTA	AppliChem A2937

24	Isopropanol	AppliChem A3928
25	Proteinase	AppliChem A3830
26	Ethidium bromide	AppliChem A1151
27	Ethanol	AppliChem A3678
28	Taq DNA polymerase MBI,	Ferments EP0401
29	dNTP set MBI,	Fermatas R0181
30	Taq I	Fermatas ER0671
31	Primers	Ege 1, Ege 2, L1 Promega
32	Chloroform	AppliChem A3633
33	Standard agarose	AppliChem A2114

APPENDIX B

MEDIA

Media Used for Isolation (*Bacillus subtilis* Medium) (Broth)

Solution A

Yeast extract	1 g
$(NH_4)_2SO_4$	0.2 g
$MgSO_4 \times 7H_2O$	0.5 g
$CaCl_2 \times 2H_2O$	0.25 g
KH_2PO_4	0.6 g
Distilled water	500 ml
Adjust pH to	3.7 with 1 M H_2SO_4

Media Used for Extracellular Enzyme Screening

Media used for cellulase Screening

Medium I	**g/l**
Ammonium sulfate	1
Na_2HPO_4	6
KH_2PO_4	3
Agar	20

Ingredients (except agar) were dissolved in distilled water. The pH of the medium was adjusted to 4.0 with 1M H_2SO_4. Agar was added. Medium was sterilized by autoclaving at 121 ºC for 15 min.

Medium II	**g/l**
Yeast extract	1
Ammonium sulfate	2
Na_2HPO_4	6
KH_2PO_4	3
Commercial grade cellulose	5
Agar	20

Ingredients (except agar) were dissolved in distilled water. The pH was adjusted to 4.0 with 1M H_2SO_4. Agar was added and the medium was sterilized by autoclaving at 121º C for 15 min.

Medium III	**g/l**
Nutrient Broth	13
$CaCl_2$ x $2H_2O$	0.1

Ingredients were dissolved in deionized water. The pH of the medium was adjusted to 4.0 with 1 M H_2SO_4. Agar was added and the Medium was sterilized by autoclaving at 121º C for 15 min. Tween 20 was autoclaved separately and added to the medium.