Bio-prospecting a Soil Metagenomic Library for Carbohydrate Active Esterases

A thesis submitted in fulfilment of the requirements for the degree of

MASTERS IN BIOCHEMISTRY

of

RHODES UNIVERSITY

In the

Department of Microbiology and Biochemistry

Faculty of Science

By

NTOMBIFUTHI SHEZI

Academic supervisor: Prof Brett Pletschke

Co-supervisors: Dr Konanani Rashamuse and Kgama Mathiba

January 2016

DECLARATION

I, Ntombifuthi Nomthanadazo Shezi (Student no: 15S0019) hereby declare that this thesis entitled: ‘‘Bioprospecting a soil metagenome for carbohydrate active enzymes’’ submitted for degree of Magister Scientiae: Biochemistry, at Rhodes University is my own original work and has not been previously submitted for any degree at any other university and that all the sources I have used are indicated and acknowledged by complete references

Signature: 

Ntombifuthi Shezi
Student

ACKNOWLEDGEMENTS

I would like to pass my appreciation to the following persons and organizations for their assistance in the completion of this study:

Financial assistance towards this research project was contributed by the Council of Scientific and Industrial Research (CSIR). I would also like to thank the CSIR Biosciences for providing an exceptional laboratory environment to help me carry out my research work and grow my skills and knowledge.

To Prof Brett Pletschke, Department of Microbiology and Biochemistry, Rhodes University academic supervisor: Thank you for your contribution to my career development, supervision, academic stimulation and words of encouragement. Your kindness was everything I could ever ask for from an academic supervisor, your patience and love was beyond measures. Thank you very much, God bless you!!

To Dr Konanani Rashamuse, CSIR Biosciences co supervisor: I am deeply grateful for your support, you took me under your wings and kindly showed me the way. You always challenged and inspired me to tackle through many difficulties. Thank you for managing my project and expanding my knowledge in research field. Thank you for your sacrifices, time and advice. Your support and constant push for improvement has made my experience as a student successful and more rewarding. This work is evidence that passion and working hard pays off.

To Kgama Mathiba, CSIR Biosciences co-supervisor: Thank you for your guidance, patience, encouragements, and analytical skills stimulation you gave me throughout the course of this study. You have opened up a world of knowledge and taught me an immense amount of work, which I will cherish and value for future research.

To colleagues and students at the CSIR Biosciences, thank you for your moral support, companionship, ears, smiles, chats and selfless assistance. It's true, a problem shared is a problem solved.

To my family and friends, thank you for your emotional and moral support and prayers throughout the course of the study.

To almighty God, for His love and grace without Him this study would have not been successful.

ABSTRACT

Lignocellulosic biomass is a promising renewable resource on earth. Plant biomass contains fermentable sugars and other moieties that can be converted to biofuels or other chemicals. Enzymatic hydrolysis of these biopolymers is significant in the liberation of sugars for fermentation into desired products. Owing to its complex structure, synergistic action of enzymes is required for its degradation. Enzymes that are involved in biomass degradation include cellulases, hemicellulases and the accessory enzymes acetyl xylan esterases and ferulic acid esterases. Ferulic acid esterases (FAEs, EC 3.1.1.73), represent a subclass of carboxylester hydrolases (EC 3.1.1.-) that catalyse the release of hydroxycinnamic acids (such as ferulic acid, *p*-coumaric, ferulic, sinapic and caffeic acid) that are generally found esterified to polysaccharides, such as arabinoxylans. Hydroxycinnamic acids have widespread potential applications due to their antimicrobial, photoprotectant and antioxidant properties, as well as their use as flavour precursors. Therefore, this interesting group of FAEs has a potentially wide variety of applications in agriculture, food and pharmaceutical industries. In the search for novel biocatalysts, metagenomics is considered as an alternative approach to conventional microbe screening, therefore, searching for novel biocatalysts from a soil metagenome that harbours a unique diversity of biocatalyst is significant. The aim of this study was to extract DNA from soil associated with cattle manure and construct a soil metagenomic library using a fosmid based plasmid vector and subsequently functionally screen for ferulic acid esterases using ethyl ferulate as a model substrate. A total of 59 recombinant fosmids conferring ferulic acid esterase phenotypes were identified (Hit rate 1:3122) and the two fosmids that consistently showed high FAE activities were selected for further study. Following nucleotide sequencing and translational analysis, two *fae* encoding open reading frames (FAE9 and FAE27) of approximately 274 and 322 aa, respectively, were identified. The amino acid sequence of the two ORFs contained a classical conserved esterase/lipase G-x-S-x-G sequence motif. The two genes (*fae*9 and *fae*27) were successfully expressed in *Escherichia coli* BL21 (DE3) and the purified enzymes exhibited respective temperature optima of 50 °C and 40 °C, and respective pH optima of 6.0 and 7.0. Further biochemical characterisation showed that FAE9 and FAE27 have high substrate specificity, following the fact that EFA is the preferred substrate for FAE9 (k_{cat}/K_m value of 128 $s^{-1}.mM^{-1}$) and also the preferred substrate for FAE27 (k_{cat}/K_m value of 137 $s^{-1}.mM^{-1}$). This work proves that soil is a valuable environmental source for novel esterase screening through functional based metagenomic approach. Therefore, this method may be used to screen for other valuable enzymes from environmental sources using inexpensive natural sources to encourage the screening of specific enzymes. Biochemistry of the two isolated enzymes makes these enzymes to be useful in industrial applications due to broad substrate activity that could replace the specialised enzymes to complete plant biomass degradation.

Key words: Soil metagenomic library; functional screening; ferulic acid esterases; plant biomass.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa	Amino acid
Asp	Aspartic acid
APS	Ammonium persulphate
BACs	Bacterial artificial chromosome
bp	Base pair
B-Per	Bacterial protein extraction reagent
BSA	Bovine serum albumin
BLAST	Basic local alignment search tool
C	Carbon
^{o}C	Degrees Celsius
CAZy	Carbohydrate-Active Enzymes
CEs	Carbohydrate esterases
Cfu	Colony forming units
CGA	Chlorogenic acid
C-terminus	Carboxy terminus
ddH_2O	Deionised distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EC	Enzyme Commission number
EDTA	Ethylene diamine tetraacetic acid
EF	Ethyl ferulate
FAE	Ferulic acid esterase
FAXX	O-[5-O- (transferuloyl)-α-L-arabinofuranosyl]-(1->3)-O-β-D-xylopyranosyl-(1->4)-D-xylopyranose
g	Grams
g	Gravitational force
GHs	Glycoside hydrolases
GST	Glutathione-S-transferase
HIC	Hydrophobic interaction chromatography
HPLC	High pressure liquid chromatography
His	Histidine
h	Hour(s)
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-thiogalactosidase
k_{cat}	Catalytic turnover number
kDa	Kilo Dalton
K_m	Michaelis-Menten constant
LB	Luria-Bertani
LEW	Lysis equilibration wash
MBP	Maltose binding protein
min(s)	Minute(s)
MCA	Methyl caffeate
MGA	Methyl gallate
MFA	Methyl ferulate
MpCA	Methyl p-coumarate
mM	Millimolar
MOPS	3-(N-morpholino)propanesulfonic acid

MSA	Methyl sinapate
µg	Microgram
µL	Microlitre
mL	Millilitre
MW	Molecular weight
Ng	Nanogram
N-terminus	Amino-terminus
Ni-TED	Nickel (tris-carboxymethyl ethylene diamine)
OD_{600}	Optical density at 600 nanometers
ORF	Open reading frame
PLs	Polysaccharide lyases
PCR	Polymerase chain reaction
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Ser	Serine
sp.	Species
TE	Tris/EDTA
TBA	Tributyrin agar
TBE	Tris-borate-EDTA
Tris-HCl	Tris (hydroxymethyl)methylamine hydrochloride
V_{max}	Maximum velocity
v_o	Initial velocity
v/v	Volume to volume
w/v	Weight to volume

LIST OF OUTPUTS

Some of the work described in this thesis will contribute to the following outputs;

Research paper: Ntombifuthi Shezi, Konanani Rashamuse, Kgama Mathiba, and Brett Pletschke (2016). **Bio-prospecting a Soil Metagenome Library for Carbohydrate Active Esterases**

To be submitted for publication to the Biocatalysis and Agricultural Biotechnology journal.

Poster: Ntombifuthi Shezi, Konanani Rashamuse, Kgama Mathiba, and Brett Pletschke (2016). **Bio-prospecting a Soil Metagenome Library for Carbohydrate Active Esterases**

To be presented to the South African Society for Microbiology (SASM) conference.

.

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

1. BACKGROUND

Plant cell wall polysaccharides are important components of the carbon turnover on earth. Cell wall components can be utilized for the production of numerous products ranging from food, pharmaceutical products, pulp and paper to bioethanol production. Cell wall polysaccharides provide a potential source of renewable energy that can be exploited for the development of bio-economy (Ikram-ul-Haq and Tehmina, 2006). The demand for green technologies and limited fossil fuels require the development of renewable energy resources. Understanding biological processes that can be applied to degrade the plant cell wall into renewable components is very important, in meeting the demands for biofuels and bio-based chemicals. For millennia, enzymes have been used in many industrial applications to catalyse many chemical reactions. The use of enzymes is convenient as it requires lower energy consuming pathways which makes their use environmental friendly. It is also a better alternative compared to expensive, polluting and time-consuming chemical technologies (Ikram-ul-Haq and Tehmina, 2006).

Over the years, enzymes have been used to biodegrade plant cell wall polysaccharides and to convert biomass into useful products. The ability of enzymes to biodegrade plant cell wall polysaccharides offers unique opportunities for product improvement and economic stability. Plant cell wall polysaccharides can be degraded by certain microorganisms that are capable of producing certain enzymes. However, the majority of microorganisms are not culturable using classical laboratory techniques (Handelsman, 2004) and this has led to the development of culture independent metagenomic techniques. Metagenomics is defined as the direct genetic study of genomes confined within an environmental sample (Handelsman, 2004). Metagenomic tools can be used to overcome the problem of unculturable microorganisms (Handelsman, 2004). A better understanding of the processes that allow for the exploitation of environmental samples to discover and characterise essential genes is necessary. Therefore, there is a demand for using conventional metagenomic techniques to create libraries that can be screened to discover genes of interest.

1.1. Plant cell wall

Lignocellulosic biomass is the most abundant renewable biomass on earth. Lignocellulosic biomass is a significant biopolymer of the plant cell wall. The plant cell wall is made up of a complex polymeric structure consisting of cellulose microfibrils embedded in a matrix made up of hemicelluloses, lignin and pectin (Gilbert, 2010). Cellulose is the primary component of a plant cell wall followed by hemicellulose - being the second most renewable material after cellulose (Harmsen *et al*., 2010). Lignocellulosic biomass is made up of sugar polymers. Therefore, it is a potential source of fermentable sugars which can be converted into useful products, as these are used in fuel, chemical, and bio-product production (Michelin *et al*., 2011).

Plant cell wall components interact together to provide plants with rigid cell wall (de Vries and Visser, 2001; Showalter, 1993). Cellulose consists of insoluble fibers of β-1, 4 linked glucose molecules that are noticeably crystal-like. Hemicellulose consists of β-linked sugars made up of xyloglucans, galactomannans and arabinoxylan. Pectin is made up of homogalacturonan, rhamnogalacturonan I and rhamnogalacturonan II. The structure of the plant cell wall is shown in Figure 1.1 below.

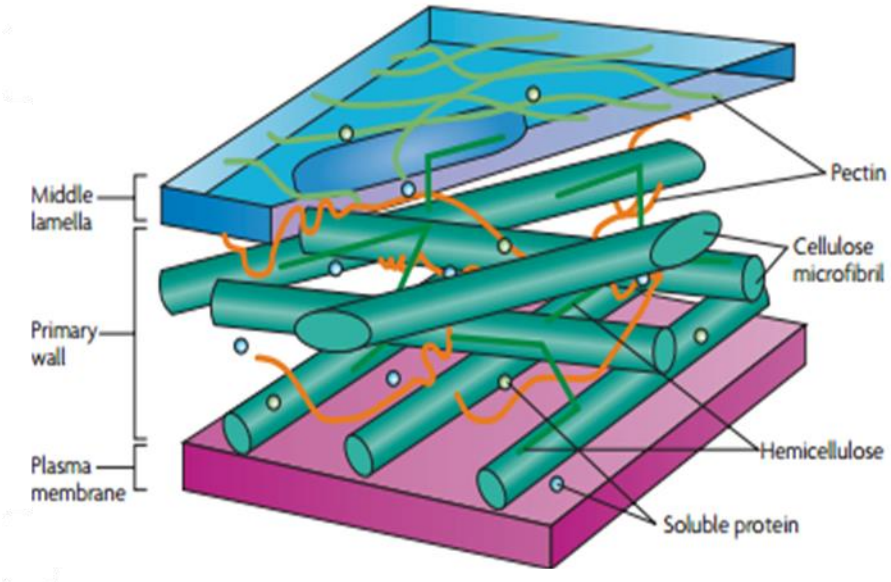


Figure 1.1: The basic structure of a plant cell wall showing cellulose, hemicellulose and pectin as the main components of plant cell wall. Lignin maybe polymerised in the spaces between these polymers. Image adapted from (Sticklen, 2008).

1.1.1. Cellulose

Cellulose is the major component of plant cell walls and is widespread in higher plants, bacteria, marine algae, and other biomass. The supramolecular structure of natural cellulose showed that it occurs in both crystalline and non-crystalline form. It is considered as a polymer of β 1, 4 linked D-glucose units. The structure of cellulose has repeated units of glucose monomers joined at a 180 ^{o} angle. A repetitive unit of glucose is called cellobiose (Himmel *et al*., 2010). The nature of bonds among glucose units allows the polymer to be arranged in long straight chains. This arrangement allows the formation of hydrogen bonds between cellulose molecules. Glucan units have wide hydrogen bonds and Van der Waals forces between each other that form the crystalline structure that is resistant to biodegradation (Himmel *et al*., 2010).

1.1.2. Hemicellulose

Hemicellulose is the second most abundant plant cell wall polysaccharide that is made up of a variety of sugar units including glucose, xylose, galactose, arabinose and mannose connected to cellulose microfibrils through hydrogen bonds. Hemicellulose polymers consist of arabino-xylan, galactans and gluco-mannan. Carbohydrates substituents of hemicelluloses are either on the main chain or become attached to the branches of carbohydrates. Xylan is the most common and significant polymer of hemicellulose. Xylan is made up of β-1,4 linked D-xylose backbone. Hemicellulose does not form a crystalline structure due to its highly branched structure. The xylan backbone can be substituted with different side groups such as L-arabinose, D-galactose, acetyl, feruloyl, *p*-coumaroyl, and glucuronic acid residues. Hemicelluloses are classified by the main residues of sugars that are found in the backbone (Wilkie and Woo, 1977). The galacto-glucomannan backbone consists of β-1,4 linked D-mannose, D-glucose units. It also contains D-galactose as side groups and xyloglucans that are made up of β-1,4-linked-D-glucose backbones substituted by D-xylose (de Vries and Visser, 2001).

1.1.3. Lignin

Lignin is an aromatic compound that forms part of the structure of plant cell walls. Lignin is made up of branched phenylpropane units: coumaryl, coniferyl and sinapyl alcohols

(Zaldivar *et al*., 2001). These lignin monomers differ in the amount of methoxylation depending on the plant species or tissues (Campbell and Sederoff, 1996). Lignin is interconnected to cellulose and hemicellulose by covalent and non-covalent bonds. The function of lignin is to maintain cell wall structural integrity by enclosing the cell wall polysaccharides.

1.1.4. Pectin

Pectin is made up of a heteropolysaccharide backbone made up of alternating homogalacturonan and rhamnogalacturonan. The pectin main backbone of α-1,4-linked D-galacturonic acid consists of hairy (homogalacturonan) and smooth (rhamnogalacturonan) regions. The smooth region consists of D-galacturonic acid residues. These residues are prone to methylation or acetylation at C-2 and C-3 positions. The hairy region consists of D-galacturonic acid that is linked to α-1,2-linked L- rhamnose residues. Acetyl groups are known for their ability to hinder sterically the cleavage of glycoside linkages. Therefore, the removal of such groups facilitates the action of lyases and hydrolases (de Vries and Visser, 2001).

1.1.5. Plant cell wall phenolics (Hydroxycinnamic acids)

Apart from the three major plant cell wall polymers (cellulose, hemicellulose and lignin), the plant cell wall also consists of minor phenolic acids. Phenolic acids that are found in the cell wall play a major role in the linkage of cell wall polymer (such as lignin) with other cell wall components through ester bonds (Yu *et al*., 2005). Phenolic acids consist of hydroxycinnamic acids and the most common and extensively distributed hydroxycinnamic acids include ferulic acid, *p*-coumaric acid, caffeic acid and sinapic acid (Figure 1.2). The major component of phenolic acids found in plant cell walls of monocots and dicots is ferulic acid (Yu *et al*., 2005). Phenolic acids play a major role in photo-protection and mechanical support in these plants.

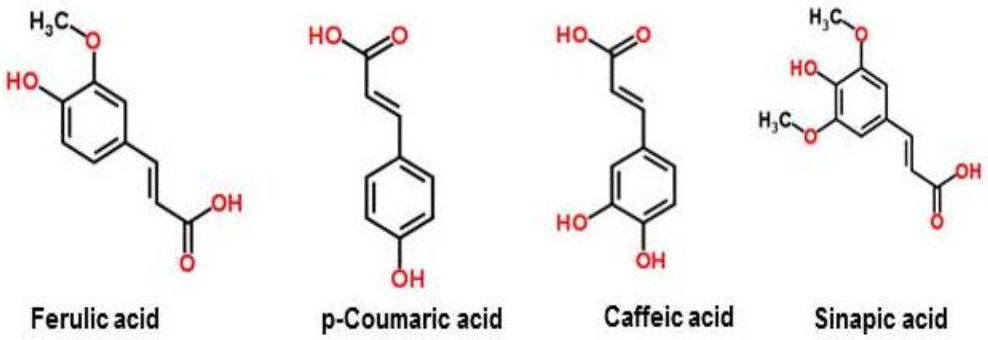


Figure 1.2: Structures of hydroxycinnamic acids commonly found in plant cell walls (Yu *et al.*, 2005).

1.2. Plant cell wall degrading enzymes

It is a challenge to convert complex lignocellulosic biomass into other substances, although its conversion offers unique opportunities for industrial bio-based chemical and energy production (Sweeney and Xu, 2012). Recently, enzymes are used in industries to collectively degrade complex biomass structures (Sweeney and Xu, 2012). The use of enzymes in industries for biomass conversion has the same output level of productivity when compared to other polluting technologies that uses chemical and physical methods. Increased application of enzymes has put microbiology and biotechnology in the spotlight (Khandeparkar and Bhosle, 2006). Exploiting industrial enzymes for biomass degradation is useful for product development and the largest group of enzymes that are used in industries are technical enzymes which consist of proteases, amylases, xylanases, cellulases, lipases and pectinases. They are used in the pulp and paper industry, laundry detergents, textile manufacturing, and the leather industry. The second group of enzymes include food enzymes that are used in dairy, brewing, wine, juice production, and in baking industries. Lastly, feed enzymes are used in animal feed. Some enzymes are used in biofuel production and in agriculture (Kirk *et al*., 2002).

Plant cell wall degrading enzymes are classified into three classes: glycoside hydrolases which hydrolyse glycosidic bonds between two sugar units, carbohydrate esterases which hydrolyse carbohydrate esters with the addition of a water molecule and polysaccharide lyases which perform non-hydrolytic cleavage of glycosidic bonds. The carbohydrate-active enzymes are divided into different families. The members that belong to one family have similar structural motifs and different substrate specificities (Henrissat, 1991). A wide range of enzymes is capable of biodegrading plant cell wall polysaccharides as indicated in Table 1.1 below. Plant cell wall degradation requires the cooperative action of carbohydrate active enzymes (CAZy). CAZy enzymes include glycoside hydrolases families (GHS), carbohydrate esterases (CEs) and polysaccharide lyases families (PLs) (Gilbert, 2010).

Table 1.1: Main CAZy Plant cell wall degrading enzymes. Adapted from Sweeney and Xu (2012).

Plant cell wall components	Enzymes	CAZy family	Functions
Cellulose	Exo and endo-glucanases	GH5,GH6,GH7,GH 8,GH9,GH12GH44, GH45,GH48,GH74, GH12	Cleaves internal or either end of cellulose chain.
	β-glucosidases	GHI, GH3, GH116	Cleaves glucose chain from non-reducing ends of cellulases.
	Xylanases	GH10, GH11	Cleaves β-1,4 bond of xylan backbone releasing xylooligomers
	β-xylosidases	GH3, GH30, GH39, GH43, GH52, GH54, GH116	Cleaves exo β-1,4 bond of xylooligomers releasing xylose
	α-glucuronidases	GH67, GH115	Cleaves α-1,2 bond between glucoronic acid side chain substitutions
	Mannanases	GH5, GH26, GH113	Cleaves β-1,4 bond of mannan releasing mannan oligomers
	Acetyl xylan esterases	CE1, CE2, CE3, CE4, CE5, CE6, CE7, CE12	Cleaves acetyl side chain substitutions releasing acetic acid
	Cinnamoyl and ferulic acid esterases	CE1	Cleaves hydroxycinnamic acids side chains releasing *p*-coumaric acid and ferulic acids
	β-mannosidases	GH1, GH2, GH5	Cleaves exo β-1,4 bond of mannan oligomers releasing mannose
Pectin	Polygalacturonases	EC4	Hydrolyse glycosidic bonds between galacturonic acid residues
	Pectin lyases	EC4	Cleaves glycosidic bonds between galacturonic residues
Lignin and other polyphenolic compounds	Laccases	EC1	Oxidise phenolic parts of lignin directly or indirectly oxidise non-phenolic parts of lignin
	Manganese peroxidases	EC1	Oxidise manganese and transfer electrons to lignin

1.2.1. Cellulose degradation

The bioconversion of cellulosic biomass requires the action of cellulases. Cellulases are enzymes that are produced by fungi, bacteria, plants and humans. They hydrolyse β-1,4 glycosidic bonds found in cellulose chains. Cellulose hydrolysis is performed by three groups of celluloses: endo-(1,4)-β-D-glucanases. (EC 3.2.1.4), exo-(1,4)-β-D-glucanases, including cellobiohydrolases (EC 3.2.1.91) and β-glucosidases (EC 3.2.1.21) (Zhang and Zhang, 2013). Endoglucanases cleave the internal O-glycosidic bonds in cellulose, resulting in glucan chains of different lengths. The ends of the cellulose chain are cleaved by exoglucanases releasing β-cellobiose. Then, β-glycosidases act specifically on the β-cellobiose disaccharides releasing glucose units (Rabinovich *et al*., 2002). Cellulases are used in different industries including the pulp and paper industry, textile industry, detergents industry, feed industry, brewing industry and in agriculture (Kuhad *et al*., 2011).

1.2.2. Hemicellulose degradation

Hemicellulose is a polymeric compound that consists of xylan, xyloglucan and mannan backbones. Complete degradation of hemicellulose requires the synergistic action of glycosyl hydrolases and carbohydrate esterases (Figure 1.3) (Juturu and Wu, 2014). Some enzymes cleave the main chain of hemicellulose internally releasing the monomeric sugars. The complete hydrolysis of polymer side chains to disaccharide and monosaccharides is performed by accessory enzymes (de Souza, 2013). Xylanases (EC 3.2.1.8) are a group of enzymes that are hemicellulolytic and are used for the hydrolysis of the xylan backbone (Kheng and Omar, 2005). The xylanolytic system consists of the key enzymes, such as endo-xylanases (endo-1,4-β-xylanase, EC 3.2.1.8), and β-xylosidases (xylan-1,4-β-xylosidase, EC 3.2.1.37). Endo-1,4-β-xylanases cleave the glycosidic bonds of the xylan backbone internally to xylo-oligosaccharides while β-xylosidases hydrolyses soluble xylo-oligosaccharides to xylose (Juturu and Wu, 2014). Accessory enzymes such as α-glucuronidase (α-glucosiduronase, EC 3.2.1.139), α-L-arabinofuranosidase (EC 3.2.1.55), arabinase (endo α-L-arabinase, EC 3 2.1.99), acetyl xylan esterase (EC 3.2.1.72), and ferulic acid xylan esterase (EC 3.2.1.73) hydrolyse linkages of side chains (de Souza, 2013; Juturu and Wu, 2014).

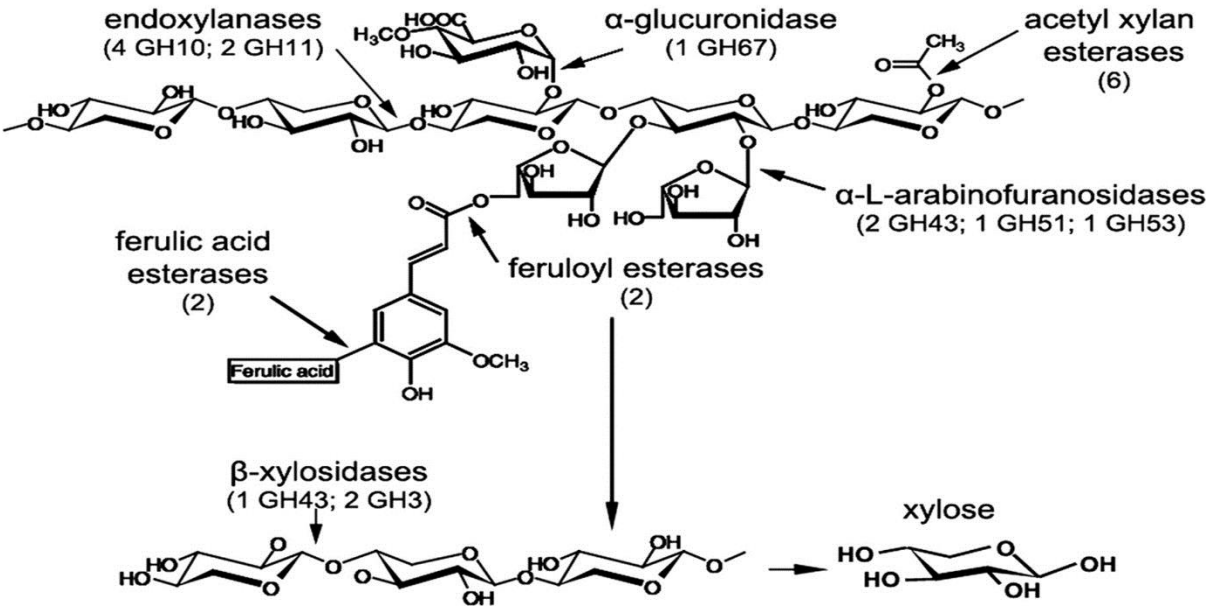


Figure 1.3: The schematic representation of arabinoxylan synergistic degradation by different enzymes. Image adapted from Sun *et al.* (2012).

1.2.3. Lignin degradation

Lignin is a highly insoluble polymer with branched phenylpropane units linked together by ester, ether and carbon–carbon linkages (Campbell and Sederoff, 1996). These linkages increase lignin recalcitrance. Degradation of lignin requires oxidative enzymatic degradation rather than hydrolytic degradation since the lignin structure consists of carbon-carbon and ether bonds. The oxidative enzymes involved in lignin degradation include lignin peroxidase, laccases, manganese peroxidase, versatile peroxidase, glyoxal oxidase and alcohol oxidase (de Souza, 2013).

1.2.4. Pectin degradation

The pectin backbone is made up of alternating hairy (homogalacturonan) and smooth (rhamnogalacturonan) regions and is hydrolysed by glycoside hydrolases and polysaccharide lyases. Glycoside hydrolases hydrolyse the smooth regions of α-1,4-linked D-galacturonic acid; these enzymes include endo-polygalacturonases and exo-polygalacturonases. The hairy region is hydrolysed by endo- and exo-

rhamnogalacturonases, xylogalacturonases, α- rhamnosidases, unsaturated glucuronyl hydrolases, and unsaturated rhamnogalacturonan hydrolases (Juturu and Wu, 2014).

1.3. Carboxylic acid esterases

Carboxylic acid esterases (EC 3.1.1.X) are hydrolytic enzymes that catalyse the formation and cleavage of ester bonds. Carboxylic esterases are a diverse group of enzymes that differ by ester bond specificity. Carboxylic esterases are a subgroup of esterases that target the hydrolysis of carboxylic esters. Carboxylic esterases consist of two groups of enzymes; the esterases (EC 3.1.1.1) and lipases (EC 3.1.1.3). These two groups differ regarding substrate specificity. Esterases hydrolyse short chains of fatty acids while lipases hydrolyse triglycerides with long chain acyl groups (Fahmy *et al*., 2008). Ferulic acid esterase, cinnamoyl esterase and cinnamic acid hydrolase are members of a subclass of carboxylic esterases that hydrolyse ester bonds between hydroxycinnamic acids found in plant cell wall. Esterases are generally found in animals, plants and microorganisms. Many carboxyl esterases from microorganisms show wide substrate specificity which allow them to be involved in many catabolic reactions. These enzymes have a number of applications in industry (Fan *et al*., 2012; Kim *et al*., 2006; Li *et al*., 2008).

1.3.1. Ferulic acid esterases (FAEs)

Ferulic acid esterases (EC 3.1.1.73) belong to the subclass of carboxylic ester hydrolases. FAEs are known for the hydrolysis of ester linkages between ferulate and polysaccharides in plant cell walls (Li *et al.,* 2011). FAEs are also known as ferulic acid esterases, cinnamoyl ester hydrolases, ferulic acid hydrolases and hydroxycinnamoyl esterases as they all hydrolyse ester linkages of ferulic acid and diferulic acids found in plant cell walls. FAEs are also considered as hemicellulose accessory enzymes as they act together with other hemicellulases. FAEs play an important role in hydrolysing ester groups that form part of hemicellulose and lignin cross-linkages (Fazary and Ju, 2007). They also hydrolyse phenolic esters into respective phenolic acids and alcohol. Plant cell walls degradability differs; some tissues of the plant cell wall are easily degradable while phenolic/aromatic compounds containing cell walls are not easily degradable.

FAEs have higher substrate specificities for aromatic moieties. Thus, they are the target for improving lignocellulose degradation, typically lignocellulose containing phenolic compounds that are ester-linked to hemicellulose (Akin, 2008). The structure of FAEs displays the α/β hydrolase-fold characteristic of a number of hydrolytic enzymes that have a catalytic triad (that is made up of serine, aspartic acid and glutamic acid) at their active sites (Figure 1.4). (Ollis *et al*., 1992). The α/β fold comprises of a main central core, eight strands of a parallel β sheet with every single second strand constituted parallel to the previous. FAEs also have a conserved pentapeptide esterase motif of glycine-x-serine-x-glycine (GxSxG), where x represents any amino acid. FAEs are considered serine esterases as they use serine as a catalytic residue for hydrolysis (Crepin *et al*., 2004).

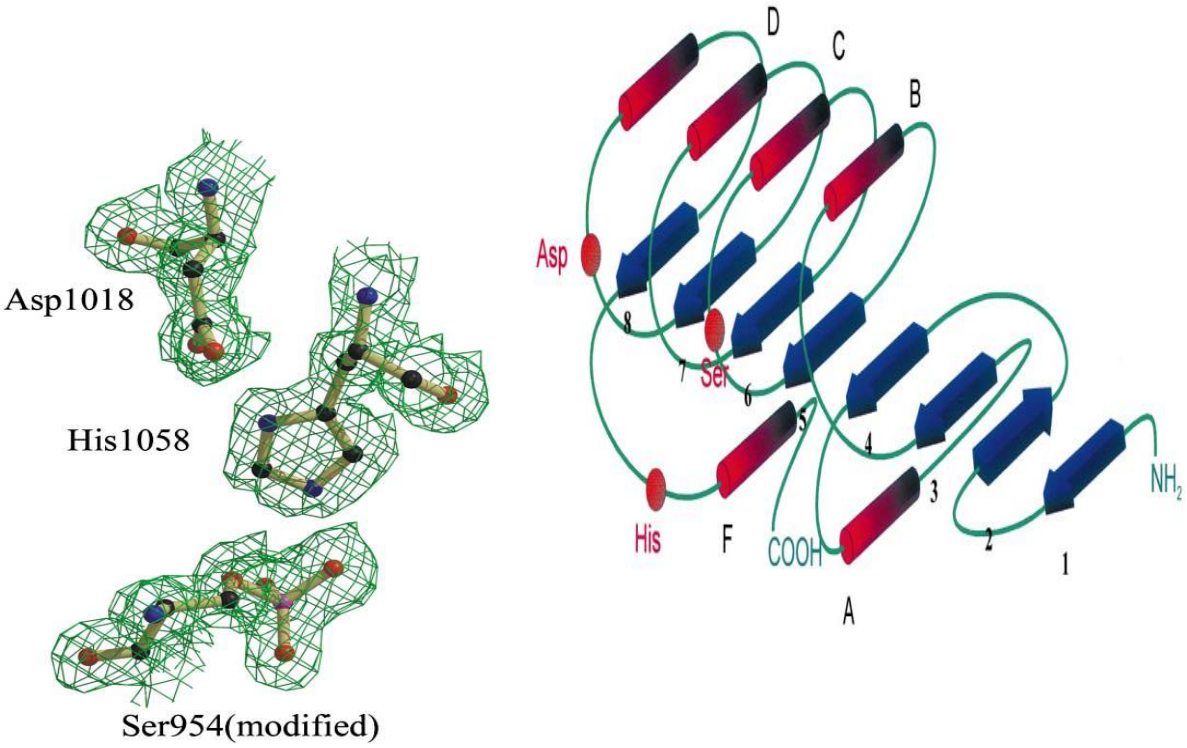


Figure 1.4: The catalytic triad of FAEs formed by residues Ser954, Asp1018, and His 1058, and the schematic presentation of the α/β-hydrolase fold. The α-sheets (1-8) are shown as blue arrows, β-helices (A-F) as red columns. The relative positions of the amino acids of the catalytic triad are indicated as red circles. Image adapted from (Bornscheuer, 2002).

1.3.2. Classification of FAEs

Ferulic acid esterases, cinnamoyl esterases and cinnamic acid hydrolases are classified based on their sequence identity/similarity and substrate utilisation. Many studies have been conducted on the characterisation and purification of FAEs from fungi, that led to the FAE classification on data collected from fungi (Udatha *et al*., 2011). Substrate specificity and sequence similarity of FAEs were the basis for classification of FAEs into four sub-groups, A, B, C, and D as shown in Table 1.2. However, Crepin *et al.* (2004) reported the fungal multiple alignment sequence that led to the suggestion of a putative type E sub-class that has unknown biochemical function currently. The four substrates that FAEs were classified upon were methyl ferulate, methyl sinapate, methyl *p*- coumarate and methyl caffeate.

Table 1.2: Summary of the ferulic acid esterases classification based on substrate specificity, sequence homology and ability to release diferulic acid from substrates.

FAE sub-groups	Microorganism	Substrate specificity				Ability to release diferulates	Sequence homology
		Methyl ferulate	Methyl sinapate	Methyl *p*-coumarate	Methyl caffeate		
A	*Aspergillus niger*[a]	✓	✓	✓		5,5' diferulic acid only	Lipase
B	*Penicillium funiculosum*[b]	✓	✓		✓	None	Cinnamoyl esterase family I and acetyl xylan esterase
C	*Talaromyces stipitatus*[c]	✓	✓	✓	✓	None	Chlorogenate esterase and Tannase
D	*Pseudomonas fluorescens*[d]	✓	✓	✓	✓	5,5' diferulic acid only	Xylanase

✓ Hydrolysable substrates

[a](Faulds and Williamson, 1995), [b](Crepin *et al*., 2003; Kroon *et al*., 1999), [c](Pa and Williamson, 1996), [d](Crepin *et al*., 2003).

Cell walls of most plants that induce FAE production contain ferulic acid which is released during plant cell wall degradation. The release of 5, 5' diferulic acid from substrates was

recognised when a panel of ferulic acid esterases were tested for the ability to release diferulic acid from given substrates, which led to the FAEs being distinguished in terms of the release of free diferulates (Udatha *et al*., 2011).

Enzymes found in Type A FAEs are capable of hydrolysing methyl ferulate, methyl sinapate and methyl *p*- coumarate esters preferably in medium containing wheat bran and oat spelt xylan, thus releasing 5,5' diferulic acid. They also show the primary amino sequence similarity of lipases (Fazary and Ju, 2007). Pa and Williamson (1996) reported that the degradation rate of arabinoxylan in wheat and barley spent grain increases in the presence of ferulic acid esterases from *Aspergillus niger* and *Pseudomonas fluorescens.* Type A FAEs act on substrates with many substituents on the benzene ring, thus they prefer to hydrolyse substrates containing phenolic/aromatic compounds with methoxy substitutions on carbon 3 and /or 5 (Crepin *et al*., 2004).

Type B FAEs are active against methyl ferulate, methyl *p*- coumarate and methyl caffeate substrates. Enzymes from this group do not show activity on methyl sinapate. They have sequence similarity to the cinnamoyl esterases family 1 acetyl xylan esterases. The production of type B ferulic acid esterase is induced by the growth on sugar beet pulp medium. In such media, they do not release 5, 5' diferulic acid; instead, they are known to release ferulic acid ester-linked to either carbon at position two of feruloylated arabinose or carbon six of feruloylated galactose residues. Type B FAEs show much activity on substrates containing aromatic compounds that have one or two hydroxyl substitutions, and that is observed in chemical structures of ρ-coumaric with one hydroxyl substitution at the benzene ring while caffeic acid shows two hydroxyl substitutions (Williamson *et al*., 1998).

Type C FAEs are active against methyl ferulate, methyl sinapate, methyl *p*- coumarate and methyl caffeate substrates. Enzymes from this group do not release 5, 5'-diferulic acid. Type C FAEs show much activity on substrates containing aromatic compounds that have hydroxyl substitutions at carbon four of the benzoic ring. The production of type C ferulic acid esterase is induced by growth on sugar beet pulp or wheat bran medium. Type C FAEs act well on water soluble feruloylated arabinoxylans. Amino acid sequence similarity shows that they have sequence similarity to chlorogenate esterase and tannase (Williamson *et al*., 1998).

Type D FAEs (FAE-D) are active against methyl ferulate, methyl sinapate, methyl *p*-coumarate and methyl caffeate substrates. Type D FAEs are capable of releasing 5, 5'-diferulic acid and their preferable induction medium is wheat bran. They also show sequence similarity to xylanases (Williamson *et al*., 1998).

There are limitations related to FAEs classification. Some FAEs from fungi were partially characterized, and little is known about the kingdom Plantae and Bacteria. Also, the classification is based on the use of four substrates which may not give the full picture of the catalytic potential of FAEs groups. The study conducted by Vafiadi *et al.* (2009) reported that FAEs from *Talaromyces stipitatus* and *Sporotrichum thermophile* were initially classified as type C FAEs but have shown specificity on different substrates. There is no classification system that has been reported yet for grouping all (Bacteria, Fungi and Plantae) kingdoms (Udatha *et al*., 2011).

The relatedness of primary amino acid sequence from all three kingdoms revealed that some FAEs from different FAE types belong to the same clade, thus the classification based on phylogenetic analysis does not correspond with the substrate specificity classification scheme. The evolutionary relationship of three kingdoms is different from each other (Udatha *et al*., 2011). Given the challenges related to FAEs classification, Udatha *et al.* (2011) proposed a new classification system in 2011 for FAEs. They classified FAEs into 12 different families using the properties of the whole sequence. They further divided FAEs into sub-families by taking into consideration the catalytic residue pattern in the sequences of each family.

1.3.3. Mode of action of FAEs

The active site of FAEs is formed by the serine, histidine, and aspartic acid catalytic triad (Dodson and Wlodawer, 1998; Prates *et al*., 2001). The catalytic centre of FAEs at all times entails a nucleophilic serine, a completely conserved histidine and the acidic residue aspartic acid. In order for the reaction to take place, the catalytic histidine must be positioned next to the catalytic serine, while aspartic acid must be positioned next to the catalytic histidine. The mechanism of action of FAEs occurs in two steps; acylation and deacylation (Figure 1.5). During the acylation step, the hydroxyl oxygen of the catalytic serine residue carries out the nucleophilic attack on the carbonyl of the ester substrate. After the attack, the catalytic histidine reposition itself and deprotonates the catalytic serine leading to the formation of the first tetrahedral intermediate (Warshel *et al*., 1989). The catalytic aspartic acid stabilizes the

protonated histidine by hydrogen bonding. The tetrahedral intermediate results in the oxyanion which is positioned in the direction of the oxyanion hole. This hole is a result of the hydrogen bonding between the carbonyl oxygen anion of the substrate and the two nitrogen atoms of the other catalytic residues. Histidine transfers the proton to the leaving group resulting in deprotonated histidine. The deprotonated histidine causes the protonation of ester oxygen to release the product. The first tetrahedral intermediate collapses, and the remaining substrate remains attached to serine (Prates *et al*., 2001).

During the deacylation step, a water molecule performs a nucleophilic attack on the carbonyl carbon of the remaining substrate. Histidine protonates the water molecule, which results in the formation of a second tetrahedral intermediate. The intermediate is then stabilized by the formation of an oxyanion hole. The nucleophilic serine receives a proton from the base resulting in the ester oxygen being protonated. The intermediate then collapses and the final product is released (Prates *et al*., 2001).

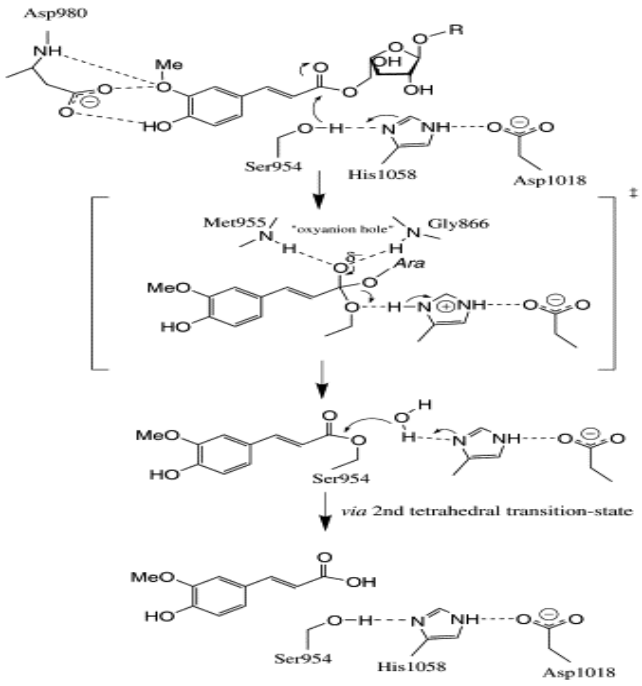


Figure 1.5: Catalytic mechanism of FAEs showing the acylation and deacylation steps (Prates *et al*., 2001).

1.3.4. Applications of FAEs

There is an increased interest in the use of FAEs in many biotechnological processes including the pulp and paper industry, biofuel, animal feed, textile and laundry, food and

agriculture, and pharmaceuticals (Panda and Gowrishankar, 2005). Ferulic acids are used in food industries as preservatives and for the production of vanillin. Vanillin is an important compound that adds flavour in foods, beverages and perfumes (Ou and Kwok, 2004). FAEs are used to produce vanillin through biotransformation processes in microorganisms. Biotransformation processes involve three steps; the first step is the decarboxylation of ferulic acid by decarboxylase enzyme to produce 4-vinylguaiacol and vanillin. The second step involves the reduction of ferulic acid to dihydroferulic acid from which vanillic acid and vanillin are made. The third step involves the production of vanillin and vanillic acid by the formation of coniferyl alcohol from ferulic acid (Ou and Kwok, 2004).

FAEs are used as food additives and as a natural antioxidant in foods. An antioxidant property of FAEs allows it to be used as food preservative. FAEs are also less affected by pH changes than other phenolic compounds such as caffeic acids, thus making them more suitable as a food preservative (Ou and Kwok, 2004). FAEs are also known to stimulate hormone secretion in humans; it can be used as an ergogenic substance in sport related foods. FAEs are also a common ingredient in cosmetics which contribute to skin protection against UV damage (Ou and Kwok, 2004). FAEs are used in the pulp and paper industry to remove fine particles thus reducing the use of chemicals during the bleaching process. Acetyl esterases improve this process by eliminating substitutions groups and linkages that are found between polymers during the pulping process, resulting in making lignin complexes more soluble (Fazary and Ju, 2008).

1.4. Biocatalysts gene discovery

Life on earth revolves around microorganisms. Microorganisms live in vast natural environments representing a large reservoir of microbial diversity (Lombard *et al*., 2011). A large diversity of microorganisms represents a large amount of genetic pool that can be exploited to recover novel genes (Culligan *et al*., 2014). Microbial genomes are reservoirs of genetic diversity and can be used to exploit and characterise new enzymes (Ferrer *et al*., 2008; Hjort *et al.*, 2010). Understanding the population dynamics within a consortium can help in understanding the genetic information of members within that consortium. The ability to capture the whole genetic pool from a complex environment can help in better understanding the microbial diversity and screening for target biomolecules and valuable products. Methods for accessing gene diversity can be divided into culture dependent or independent techniques.

1.4.1. Culture-dependent methods

Traditional cultivation methods are always based on culture-based techniques for the identification and characterization of microorganisms from environmental samples (Neelakanta and Sultana, 2013). In a culture-dependent method, novel genes are isolated from animal tissues, plants and microorganisms. Microorganisms are cultured under their suitable growth conditions in order to obtain desired traits. Culture-dependent techniques are limiting as they only allow the cultivation of a small fraction of microorganisms, leaving a huge portion of microbes unknown (Singh, 2010). Studies have been conducted to improve cultivation techniques of microorganisms that are recalcitrant to growth in a culture that simulate their natural environment, but still the ratio of uncultivated to cultivated microorganisms remain high (Neelakanta and Sultana, 2013).

1.4.2. Culture independent methods

An advancement in biotechnology has allowed the access of desired protein or biomolecules to be accessed directly from DNA through metagenomic studies. The metagenomic term was first used in the year 1998 by Handelsman *et al*. (1998). It is the genomic study of microorganisms using a non-culture based method. Since then, metagenomics has been applied in microbial ecology and biotechnology for the access of the whole microbial genome in a complex environment. It has provided information on the changes of organisms from different geographic regions (Carpi *et al*., 2011). An overview of the metagenomics approach is represented in Figure 1.6.

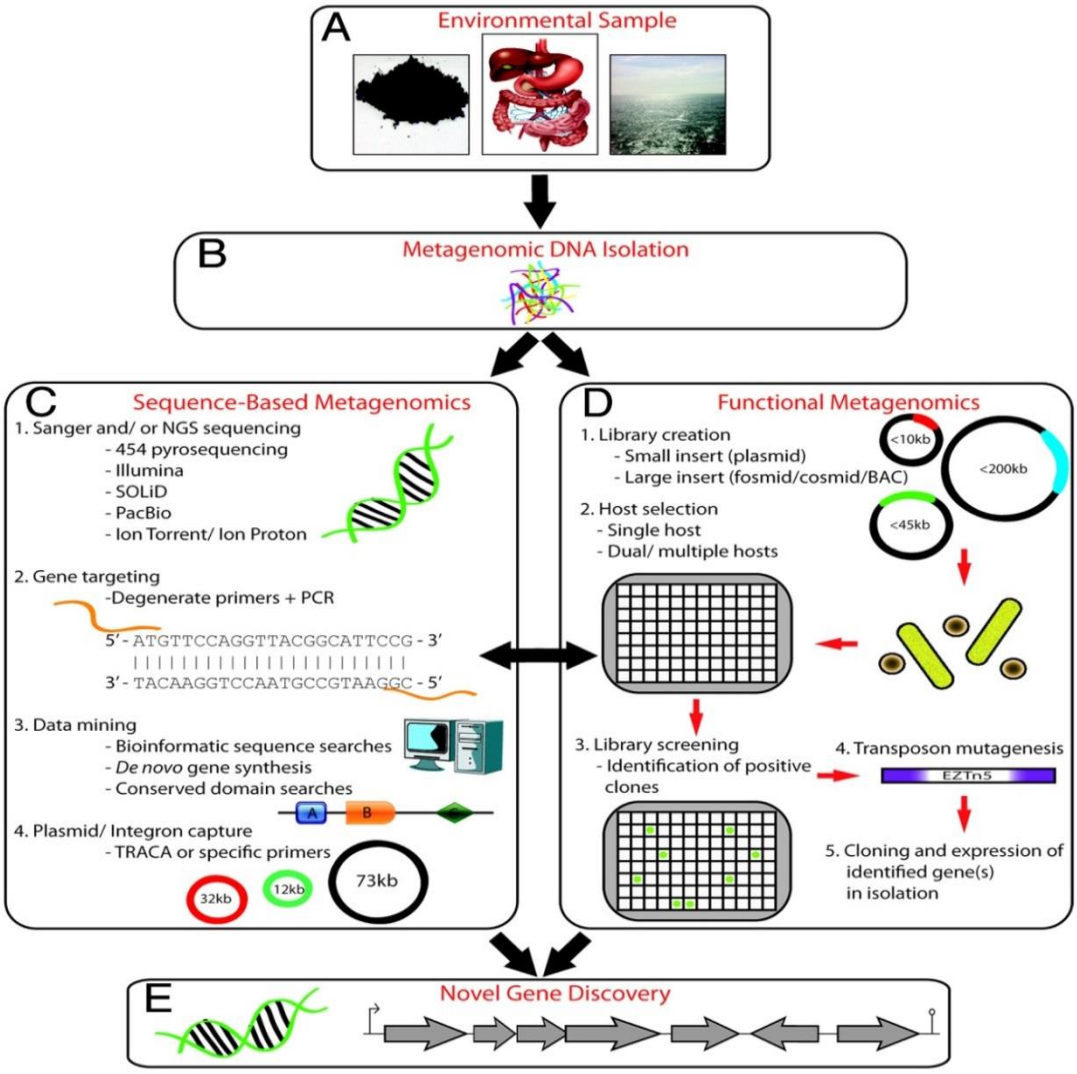


Figure 1.6: An overview process of novel gene discovery using the metagenomic approach from an environmental sample through sequenced based screening and functional screening. Image adapted from Culligan *et al*. (2014).

Metagenomic studies have been conducted on different kinds of environments. Any environment can be selected for the analysis of a metagenome. The study performed by Mathur *et al.* (2005) reported that it is crucial to access environmental samples from diverse hotspots around the world. The highly considered environments in metagenomics which are believed to have vast microbial diversity are soil (Daniel, 2005; Rondon *et al.*, 2000), marine environment (Béjà *et al*., 2000), and human gut (Lepage *et al*., 2013). The physiochemical distribution of microorganisms is different within the samples, as microorganisms have preferences in their locations (Neelakanta and Sultana, 2013).

Soil contains a large number of microbial species, and therefore, a high species diversity (Daniel, 2005). Soil is a complicated environment due to wide microbial niches, but it is preferred for gene mining as it is a vast potential resource for natural product discovery (Daniel, 2005). The human gut is another environment that has a large number of microorganisms which together form the gut microbiota. Human microbiota evolves with humans and there is a high diversity of microorganisms (Lepage *et al*., 2013). There are other extreme environments (and natural environments) used in metagenomics studies that are rich when targeting certain genes. The physico-chemical distribution of microorganisms is different within the samples. Each environment presents its own unique challenges to metagenomic investigation and requires a specifically designed method to accommodate physico-chemical factors (Lombard *et al*., 2011).

1.4.3. Metagenomic methods

Through metagenomics approach, genomic DNA can be extracted either directly or indirectly from the environment. DNA extraction method must be conducted properly in order to access the whole community within the sample and to ensure that high quality and pure DNA is obtained (Culligan *et al*., 2014). Direct DNA extraction involves direct cell lysis from the sample, while indirect DNA extraction involves extraction of cells from the sample followed by lysis and purification (Fan *et al*., 2012; Prakash and Taylor, 2012). Direct extraction of DNA is a widely used method and it is suitable for small DNA fragments; construction of large DNA inserts is not much favoured for direct extraction (Desai *et al*., 2008). DNA is accessed by using mechanical, enzymatic, and chemical lysis. In order to obtain high quality DNA during DNA isolation it is important to conserve diversity, avoid sharing and contamination of DNA and also to consider the origin of the sample, microbial composition and physico-chemical conditions (Singh *et al*., 2013). A major difficulty in DNA isolation is contamination when working with environmental samples due to humic acid and phenolic compounds. These compounds have an effect on cloning, restriction digestion, ligation and transformation of the isolated DNA (Liles *et al*., 2008).

Metagenomic libraries can be created by cloning the target genes into a suitable vector (Handelsman, 2004). Metagenomic library construction into vectors can be done using different methods. The method of use depends on the kind or size of the desired library and the type of screening method. Metagenomic library size determines the amount of captured

genetic information from the sample. Gene libraries are created using vectors. Selecting a suitable vector is important for the maintenance and expression of the cloned genes (Kakirde *et al*., 2010). The choice of vector depends on the length of DNA inserts. Various vectors used in metagenomic studies include small insert vectors, phage based vectors, cosmids, fosmids, and bacterial artificial chromosomes (BACs). High molecular weight DNA is usually cloned in cosmids, fosmid and BACs. DNA inserts that range from 25 to 45 kb are typically inserted into fosmid vectors and inserts that range between 15 to 40 kb are inserted into cosmid vectors. Large inserts of 100-200 kb need to be constructed into BACs. If the target is for small genes the DNA inserts between 2 and 10 kb can be constructed into plasmid vectors (Li *et al*., 2009). Metagenomic library construction vectors should be compatible with the host selected for screening. Functional enzymatic activity depends on more than one genetic subunit that requires the clone to have the complete gene sequence (Uchiyama and Miyazaki, 2009).

In order to obtain unbiased data, the methods on how to obtain a representative sample, how to extract DNA from the sample and how to analyse microbial DNA from the sample should be taken into consideration (Elshahed *et al*., 2008). Thirdly, the DNA extracted (especially from soil) may be contaminated by humic acid or fulvic acids that display the same DNA properties. These compounds have inhibitory activity in enzymatic reactions, thus they result in cloning inefficiency (Lombard *et al*., 2011).

There are some disadvantages in heterologous gene expression. Firstly, in order for the host to be able to express the genes of the cloned DNA, it must have a compatible expression system, otherwise the discovered activities would be low, thus requiring high-throughput screening procedures (Handelsman, 2004). Secondly, due to the difficulty of obtaining high quality DNA from the environmental sample, sometimes there is uneven distribution of microorganisms in the sample which affects the picture of microbial diversity that is obtained from the sample. Many metagenomic projects prefer *Escherichia coli* as the cloning host. Many researchers rely on the use of *E. coli* as an expression host due to high transformation efficiency and genetic manipulation. It is also deficient in restriction-modification systems and lacks genes of homologous recombination (Casali, 2003). The screening of metagenomic libraries for desired novel genes, biomolecules or enzymes involves two metagenomic approaches, function based screening of expression libraries and sequence based analysis (Lorenz *et al*., 2002). The choice of screening method depends on the type of constructed library, functional activity of interest, and the time and resources available to characterize the library.

1.4.3.1. Function based metagenomics

Function based metagenomics is aimed at identifying unknown genes and their encoded enzymatic activities from a metagenomic library through heterologous expression screening. The phenotype conferred on the cloned DNA into the expression host is detected (Kakirde *et al*., 2010). Genes are accessed by this approach without any prior knowledge of the targeted gene sequence information. This allows the discovery of unknown gene products. Metagenomic libraries are constructed and screened for the target enzyme or compound as stated by (Li *et al*., 2009). Activity detection in the function based approach is achieved by high throughput screening of library clones on indicator media. Another approach to functional screening is to use mutants of host strains that need heterologous complementation for growth under selective media. The recombinant clones that contain the gene of interest will grow in the selective media (Simon and Daniel, 2009).

Functional based metagenomics has led to the discovery of novel enzymes which include lipases and esterases (Lee *et al*., 2004; Rhee *et al*., 2005; Voget *et al*., 2003), proteases (Lee *et al*., 2007), amylases (Rondon *et al*., 2000), chitinases (Cottrell *et al*., 1999) and nitrilases (Robertson *et al*., 2004).

Functional based metagenomic screening strategies have successfully been used previously as a powerful tool to discover novel esterases from various environmental niches (Berlemont *et al*., 2011; Lämmle *et al*., 2007; Tirawongsaroj *et al.*, 2008). The success of this approach stems from a reliable plate screening assay based on tributyrin as substrate. The assay allows for the detection of positive esterase phenotypes by the presence of zone of clearance around positive phenotypes as a result of enzymatic cleavage of the ester bonds in the tributyrin ester substrate to realise butyric acid which lowers the pH of the microenvironment. A number of FAEs from various sources have also been accessed through metagenomics functional based approach using including termite (Rashamuse *et al*., 2014) and leachate (Rashamuse *et al*., 2012).

1.4.3.2. Sequenced based metagenomics

In sequenced based metagenomics the target genes are identified from conserved DNA sequences of known genes from the database. The method of sequenced based metagenomics depends on the identification of homologies amongst randomly sequenced genetic copies and previously conserved sequences that are known (Daniel, 2005). Therefore, oligonucleotide primers that need to identify genes (either by PCR or directly)

need to be designed accordingly to reveal conserved amino acids sequence motifs, in order for them to match the unknown target genes with any possible sequence. Genes of interest can be identified by direct sequencing through hybridisation or by PCR (Li and Vederas, 2009). The advantage of sequence based metagenomics is that there is independence on gene expression of the target genes (Lorenz *et al*., 2002). Limitations of this approach is that full-length gene acquisition that is required for the production of desired product is not assured, However, without the assistance of this approach some genes may have not yet been discovered (Tuffin *et al*., 2009).

Development of advanced sequencing technologies has made it easy to access genetic diversity from environmental samples (Kakirde *et al*., 2010). Traditionally, Sanger sequencing was used to sequence large metagenomes from a complex environment. However, Sanger sequencing methods have complications, especially when the environmental sample consists of a complex community which require more sequencing procedures (Tyson *et al*., 2004). Advances in biotechnology have resulted in next generation sequencing, which is a high throughput screening method coupled with low sequencing costs. The next generation sequencing methods that are widely implemented are GS-FLX 454 pyrosequencer, Illumina system, SOLID system, Helicos system and PacBio RS II (Mardis, 2008). Further development of such technologies will reduce the limitations associated with sequence based screening.

1.5. Rationale of the study

Over the past years enzymes have been part of biochemical processes in many industrial applications. Enzymes catalyze many chemical reactions converting substrates to products under controlled conditions, using lower energy consuming pathways which make their use environmental friendly and have been selected as a better alternative to polluting, expensive and time-consuming chemical technologies (Ikram-ul-Haq and Tehmina, 2006). It is very important to use alternative renewable resources for the production of everyday products. Biofuels are one of the most important products that need to be manufactured using green technologies. Recently, polysaccharides from plant cell walls have been identified as promising renewable sources that can be converted into fermentable sugars for biofuel production. Plant biomass production is massive enough to meet the demands for the production of bio-based chemicals and fuel. The ability to alter the raw material by molecular biology methods is significant towards the production of high-value bio-based products.

Recent genetic manipulation techniques provide the large scale production of enzymes which can be produced at reasonable prices. Modified biocatalysts with the specific gene of interest are now produced using protein engineering and related molecular evolution techniques. By using modern molecular techniques it is possible to increase the characterisation of genes of interest. Genes can now be modified to increase selectivity, stability and activity. The concept of directed evolution allows the creation of libraries from environmental samples and the expression of genes encoding enzymes of interests into suitable expression vectors to increase the production of specific genes of interest. Therefore, in this study libraries were created from the soil metagenome and they were subsequently screened for carbohydrate active enzymes using function based metagenomics. The genes were expressed in a suitable expression system to produce more of the desired enzymes. The enzymes were then characterised for biochemical activities.

1.6. Aims and Objectives

1.6.1. Hypothesis

The inability to cultivate the majority of microorganisms in the laboratory environment has limited the capacity to discover novel enzyme genes. It is hypothesized that the soil associated with a cattle manure environment harbours a unique diversity of carbohydrate active esterase genes that can be accessed through applied metagenomics.

1.6.2. Aim

The aim of this study was to generate a metagenomic library from soil associated with cattle manure environmental sources and subsequently screen this library for genes encoding carbohydrate active esterases.

1.6.3. Objectives

1.6.3.1 To isolate environmental DNA associated with cattle manure soil and to subsequently construct a metagenomic library.

1.6.3.2 To screen a metagenomic library for carbohydrate active esterases through a functional based approach

1.6.3.3 To recombinantly produce and purify the identified enzymes using *E. coli* as a host strain.

1.6.3.4 To biochemically characterise the identified enzyme(s) in terms of pH and temperature profiles and to determine the kinetic properties of the purified enzyme, substrate affinity (K_m) and catalytic turnover (k_{cat}) properties

In addition of the current literature review chapter, the thesis includes further three chapters., organised as follows: chapter two which covers methodological approach used in this study; followed by chapter three and chapter four which covers total DNA extraction, metagenomic library creation and screening, PCR sub-cloning, recombinant protein production and biochemical characterisation; and then culminates by overall concluding remarks.

CHAPTER TWO: MATERIALS AND METHODS

2. METHODOLOGY

2.1. Chemicals and reagents

All chemicals that were used were of molecular biology and analytical grade and were obtained from various companies specified below. Chemicals and reagents were purchased from specified companies. Bacteriological agar, Tryptone, sodium chloride and yeast extract were purchased from Merck (South Africa). ZR Soil Microbe MidiPrep™ from Zymo Research (USA), GeneJET™ gel extraction kit from Fermentas and the GeneJET™ Plasmid Miniprep Kit from Thermo Scientific (South Africa). DNA size markers, Restriction enzymes, T4 DNA ligase, and restriction enzymes were obtained from Thermo Scientific (South Africa). Kappa Hifi PCR kit was obtained from Kappa Biosystems, (South Africa). Oligonucleotide primers for polymerase chain reaction (PCR) were synthesized by Inqaba Biotec (South Africa). Ethyl ferulate, methyl esters and corresponding methyl ester acids of ferulic, ρ-coumaric, caffeic, gallate and sinapic esters were purchased from Apin chemicals (UK). Protein molecular weight markers were obtained from Thermo Scientific (South Africa). Protino® Ni-TED kit was purchased from Macherey Nagel (Germany). Antibiotics were purchased from Sigma Aldrich (USA). All molecular biology based experiments were performed using double distilled water.

2.2. Bacterial strains, growth conditions, primers and vectors

All media were cooled to 55 ºC after autoclaving. Appropriate antibiotics and inducers were added when necessary. Plates were poured and allowed to solidify at room temperature before use (Table 2.1; Table 2.2). Buffer preparations and vectors used in this study are shown in Table 2.3 and Table 2.4

Table 2.1: Growth media used in this study

LB Medium (Luria-Bertani Medium)	**Constituent**	**L^{-1}**
	Tryptone	10 g
	Yeast extract	5 g
	NaCl	10 g
For LB agar above LB constituents plus 15 g of Bacterial agar was added. Media was autoclaved at 121 ºC for 20 min.		
Tributyrin agar	**Constituents**	**250 mL**
	Bacteriological agar	250 mL
	Tryptone	3.75 g
	Yeast extract	1.25 g
	NaCl	2.5 g
The media was sonicated for 1 minute and was autoclaved at 121 ºC for 20 min.		
Ethyl ferulate stock	**Constituent**	**250 mL**
	DMSO	3.0 mL
	Distilled water	2.0 mL
The stock solution was added drop-wise while stirring to 55 ºC cooled, autoclaved (250 mL containing 3.75 g bacteriological agar, 125 mL of 50 mM phosphate buffer, pH 7.0 and 125 mL of deionized water).		

Table 2.2: Antibiotics and inducers used in this study

Antibiotics	**Preparation**
Chloramphenicol	Dissolved in 100%ethanol 12.5 µg/mL final concentration
Ampicillin	Dissolved in distilled water 100 µg/mL final concentration
Kanamycin	Dissolved in distilled water 50 µg/mL final concentration
Arabinose	Dissolved in distilled water 10% arabinose final concentration
Isopropyl-β-D-thiogalactopyranoside (IPTG)	0.1 M final concentration

Table 2.3: Buffers and solutions used in this study

Note: Unless otherwise stated, all buffers were prepared in distilled water.

Buffer	Component	pH
10× Agarose gel running buffer	L^{-1}:108 g Tris base; 55 g Boric acid; 7.45 g EDTA	
50 mM Na_2HPO_4	L^{-1}: 7.09 g of Na_2HPO_4	
50 mM NaH_2PO_4	L^{-1}: 5.9 g of NaH_2PO_4	
50 mM phosphate buffer	61.0% 50 mM Na_2HPO_4; 39.0% 50 mM NaH_2PO_4	7.0
Tris-HCl Buffer	L^{-1}: 6.1 g Trizma base adjust pH using 32% HCl	7.0
10× SDS-PAGE running buffer	L^{-1}: 30.0 g of Tris base, 144.0 g of glycine, and 10.0 g of SDS in 1000 mL of H_2O.	8.3
2× SDS-PAGE sample buffer	20 mL: 0.125 M Tris-HCl (pH 6.8); 20% (v/v) glycerol; 4% (w/v) SDS; 0.2% (w/v) bromophenol blue; 200 mM DTT	8.0
Coomassie staining solution I	10% (v/v) acetic acid; 0.125% (w/v) coomassie brilliant blue G; 25% (v/v) isopropanol	
Coomassie staining solution II	10% (v/v) acetic acid; 0.003% (w/v) coomassie brilliant blue G; 10% (v/v) isopropanol	
Coomassie staining solution III	10% (v/v) acetic acid; 0.003% (w/v) coomassie brilliant blue G	
SDS de-staining solution	10% (v/v) acetic acid and 10% (v/v) methanol	
TE buffer	10 mM Tris-HCl (pH 8.0); 1 mM EDTA (pH 8.0)	8.0
MOPS	L^{-1}:10.46g of MOPS	6.5
LEW buffer	50 mM $Na_2H_2PO_4$, 300 mM NaCl	8.0
IMAC elution buffer	50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole	8.0
HIC equilibration buffer	300 mM $(NH_4)_2SO_4$, 20 mM Tris	7.0
HIC elution buffer	20 mM Tris-HCl	7.0
1× Universal buffer	500 mM Tris, 500 mM boric acid, 330 mM citric acid, 500 mM Na_2PO_4	2.0-12.0

Table 2.4: Strains and vectors used in this study

Strains	Genotype/Features	Selective marker	Source
E. coli EPI300-*T1*^{R}	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZ ΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL nupG trfA tonA dhfr].	Chloramphenicol	Epicentre
E. coli DH5α	fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Kanamycin	Lucigen
E. coli BL21-(DE3)	F– *ompT hsdSB (rB- mB-) gal dcm lon* λ(DE3 [*lacI lac*UV5-T7 gene 1 *ind1 sam7 nin5*])	Kanamycin	Lucigen
Vectors			
pCC2FOS	Copy controlled vector, linearized and dephosphorylated at *Eco*72I restriction site. Requires EPI300™-T1R *E. coli* strain for high copy number induction, used for construction of metagenomics library.	Chloramphenicol	Epicentre
pJET1.2/blunt	Contains a lethal restriction enzyme gene that is disrupted by ligation of a DNA insert into the cloning site. As a result, only bacterial cells with recombinant plasmids are able to form colonies.	Ampicillin	Fermentas
pET28a(+)	Expression vector of N-terminally His-tagged protein	Kanamycin	Fermentas
pTZ57R	Linearized pTZ57R vector with 3'-ddT overhangs for TA cloning of PCR products with blue/white screening.	Ampicillin	Fermentas
pET30b	Expression vector of N-terminal and C-terminal His-tagged protein	Kanamycin	Fermentas
pFos_*fae3*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae9*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae14*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae27*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae71*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae83*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pFos_*fae91*	pCC2FOS™ derived constructs from the soil metagenomic library encoding ferulic acid esterase activity.	Chloramphenicol	This study
pJET_*fae3*	PJET1.2 derived vector construct containing *fae3* full length PCR gene product.	Ampicillin	This study

pJET_*fae9*	PJET1.2 derived vector construct containing *fae9* full length PCR gene product.	Ampicillin	This study
pJET_*fae14*	PJET1.2 derived vector construct containing *fae14* full length PCR gene product. product.	Ampicillin	This study
pJET_*fae27*	PJET1.2 derived vector construct containing *fae27* full length PCR gene product.	Ampicillin	This study
pJET_*fae71*	PJET1.2 derived vector construct containing *fae71* full length PCR gene product.	Ampicillin	This study
pJET_*fae83*	PJET1.2 derived vector construct containing *fae83* full length PCR gene product.	Ampicillin	This study
pJET_*fae91*	PJET1.2 derived vector construct containing *fae91* full length PCR gene product.	Ampicillin	This study
pET28_*fae3*	pET28a(+) derived vector construct containing *fae3* full length PCR gene product.	Kanamycin	This study
pET28_*fae9*	pET28a(+) derived vector construct containing *fae9* full length PCR gene product.	Kanamycin	This study
pET28_*fae14*	pET28a(+) derived vector construct containing *fae14* full length PCR gene product.	Kanamycin	This study
pET28_*fae27*	pET28a(+) derived vector construct containing *fae27* full length PCR gene product.	Kanamycin	This study
pET28_*fae71*	pET28a(+) derived vector construct containing *fae71* full length PCR gene product.	Kanamycin	This study
pET28_*fae83*	pET28a(+) derived vector construct containing *fae83* full length PCR gene product.	Kanamycin	This study
pET28_*fae91*	pET28a(+) derived vector construct containing *fae91* full length PCR gene product.	Kanamycin	This study
pET30_*fae9*	pET30b(+)derived vector construct containing *fae9* full length PCR gene product	Kanamycin	This study

2.3. Methods

2.3.1. DNA methods

2.3.1.1. Sampling

A soil sample was collected from a cattle kraal in Venda, Limpopo, South Africa by Dr Konanani Rashamuse. The sample was stored at 4 °C in the laboratory until analysed**.**

2.3.1.2. DNA extraction from the soil

DNA extraction from soil was conducted using ZR Soil Microbe MidiPrep™ (Zymo Research, USA) with some modifications. The modifications included a bead beating step for 40 seconds using Gene Disruptor (Scientific Industries, USA) for complete microbial cell disruption. The rest of the extraction protocol was followed according the manufacturers' recommended instructions. The purified DNA was eluted in a sterile 1× TE buffer (pH 8.0).

2.3.1.3. DNA Quantification

The purity and DNA concentration were determined using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, USA) by measuring absorbance at 260 nm and 280 and calculating the ($A_{260/280}$) ratio.

2.3.1.4. Determination of DNA size

The size of the isolated DNA was analysed by electrophoresis in 1% agarose gel prepared in 1× TBE buffer containing 0.5 µg/mL ethidium bromide (Sambrook and Russell, 2001). The TBE Buffer was prepared as follows [108 g Tris base, 55 g boric acid, 7.45 g EDTA and filled up to 1 L with dH_2O]. Prior to electrophoresis, DNA samples were mixed with loading dye 10× Fast digest green buffer (Fermentas). Lambda DNA molecular markers *Pst*I and *Hin*dIII were used. Samples were electrophoresed in 1× TBE buffer at 100 volts (V) for 1, 5 hrs using the PerfectBlue™ Horizontal Mini Gel System (PeqLab, UK). After electrophoresis the gel was visualised and photographed using a digital imaging system UV-transilluminator (SYNGENE G- Box).

2.3.1.5. Metagenomic DNA Purification

The DNA from either the agarose gel excisions or solution was purified using the gel extraction kit. The recovered DNA was then quantified using NanoDrop 1000 Spectrophotometer (Thermo Scientific, USA) as described in section 2.3.1.3.

2.3.1.6. Restriction enzyme digestion

Restriction endonuclease digestions of the DNA were performed in 20 µL reaction volumes; containing 16 µL (40 ng/µL DNA), 2 µL of 10× fast digest green buffer (Fermentas), and 2 µL of fast digest restriction enzyme. The reaction was incubated for 1 hour at 37 °C. The restricted product was analysed by gel electrophoresis.

2.3.2. Metagenomic library construction

2.3.2.1. Fosmid library construction

The extracted Metagenomic DNA was end-repaired to generate 5'– phosphorylated blunt-ended DNA fragments following the protocol from the DNA End-repair kit (Epicentre). The reaction mixture was incubated at room temperature for 45 minutes followed by DNA purification and quantification (method described above). The End repaired DNA (~1.581 pmol) was then ligated into CopyControl PCC2FOS Vector (0.099 pmol) following the protocol from the EpiFOS™ Fosmid Library Production Kit (Epicentre). The Ligation reactions were packaged into MaxPlax Lambda packing extract and then used to transfect into EPI300-T1^{R} phage resistant strain. A volume of 100 µL infected cells were plated on LB agar supplemented with 12.5 µg/mL chloramphenicol and incubated at 37 °C overnight to select for CopyControl fosmid clones. The number of colony forming units was determined using the following equation:

$$\text{Titre} = \frac{(\#\text{ of colonies})(\text{dilution factor})(1000\ \mu\text{L/mL})}{\text{Volume phage plated } (\mu\text{L})}$$

2.3.2.2. Library storage

The CopyControlled fosmid clones were stored by resuspending all colonies from the agar surfaces using 4 mL of LB broth supplemented with 12.5 µg/mL chloramphenicol for each plate. The mixture was centrifuged at 9 000 × *g*, 10 minutes, the supernatant was discarded and the pellet was then resuspended in 20 mL of LB broth supplemented with 12.5 µg/mL chloramphenicol and 20% final glycerol concentration. A 1 mL volume of library was stored in cryo-vial tubes and stored at - 80 °C.

2.3.2.3. Determination of the library insert size

For the determination of the library size, one tube of – 80 °C stored libraries was thawed on ice. A volume (100 µL) of 10^{-4} diluted cells was diluted in LB broth was plated on LB agar supplemented with 12.5 µg/mL chloramphenicol and the plates were incubated at 37 °C overnight. A total of 20 single colonies were randomly selected and grown in 5 mL of LB broth supplemented with 12.5 µg/mL chloramphenicol and 10% (v/v) arabinose. The plasmid DNAs were isolated and endonuclease restricted with *Bam*HI and *Hin*dIII *enzymes.* The average mean (excluding the plasmid vector fragment) size was calculated from the insert DNA fragments.

2.3.3. Activity screening for FAEs

The metagenomic library was primarily screened for total carboxyl esterases on tributyrin substrate followed by a specific secondary screening for ferulic acid esterase activity on ethyl ferulate substrate.

2.3.3.1. Screening for esterases

Carboxyl esterases were screened on tributyrin agar (TBA) prepared by supplementing LB agar with 1% (v/v) glyceryl tributyrate (Sigma, USA). The media was sonicated using the Vibra Cell™ Sonicator (Sonics and Materials, USA) until it became cloudy. The respective antibiotics were added to the autoclaved media after it has cooled to 55 °C. The library was plated onto the media and incubated at room temperature for 2 and 3 days at 28 °C. Clones with zones of clearance around the colony margins were selected and subjected for secondary screening on ethyl ferulate agar.

2.3.3.2. Screening for FAEs

Esterase positive clones identified from the TBA screening assay were further screened on ethyl ferulate agar for identification of FAE positive clones. Wells were punched on ethyl ferulate agar. The cell culture suspensions of the esterase positive clones were suspended in 50 mM phosphate buffer pH 7.0, followed by the addition of 50 µL of the suspension into the punched wells of the ethyl ferulate agar. FAE activities were identified by zones of clearance around the wells after 1-2 days of incubation at 37 °C.

2.3.4. Sequencing

FAE positive fosmids were sent for nucleotide sequencing using the Illumina next generating sequencing platform facility from Inqaba Biotech (South Africa). The sequences were analysed using Bioedit (Hall, 1999) The website (http://www.ncbi.nlm.nih.gov/) was used to predict the open reading frames (ORFs) and to also compare the sequenced gene to other proteins in the database by using the basic local alignment search tool (BLAST) for protein (Altschul *et al*., 1997). Signal peptides were predicted using SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/).

Non-redundant database sequences from the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast) were used as reference controls for the inference of phylogenetic position of the FAE sequences identified in this study. Based on the sequence identities, closely related protein sequences were identified through homology searches using the BLASTP against the NCBI non-redundant database. Sequence alignments and editing were done using the BioEdit software. The alignments were manually inspected and trimmed to equal length in BioEdit. The MEGA 4.0.2 software was used for phylogenetic inference using the neighbour-joining method.

2.3.5. Cloning of FAE gene(s)

2.3.5.1. Polymerase chain reaction (Touch down PCR)

PCR amplifications were performed using a thermo cycler (C1000 Touch^{TM} thermal cycler, Bio-Rad). Primers were designed with restriction cuts *Nde*I and *Xho*I, *Nde*I and *Hin*dIII, and *Bam*HI and *Xho*I (Table 2.5). The primer pairs were used to amplify targeted FAE genes by a touchdown PCR protocol using the PCR kit. Thermal cycling conditions were as follows: 98 °C for 3 minutes followed 2/9 cycles of (94 °C for 30 seconds; 60 °C for 30 seconds; 72 °C for 30 °C seconds and 94 °C for 30 seconds). Then followed by 6/29 cycles of (94 °C for 30 seconds; 45 °C for 30 seconds; 72 °C for 30 seconds); 72 °C for 7 minutes. The expected PCR products were excised from the gel and purified using the gel extraction kit. Purified PCR products were ligated into pJET1.2 blunt-end cloning vector at the picomoles ratio of Vector: Insert ratio of 1:3.

The ligation reaction was used to transform *E. coli* DH5α cells. The plasmid DNA from five randomly selected colonies were screened for correct PCR products by means of engineered restriction sites before ligation into pET28a vector at a picomoles ratio (1:3 vector: insert) using T4 DNA ligase. The ligation reaction was used to transform competent *E. coli* BL21 (DE3) cells. Plasmids were isolated from the transformants and screened for correct insert size by means of engineered restriction sites, resulting in pET28_*fae*3, pET28_*fae*9, pET28_*fae*14, pET28_*fae*27, pET28_*fae*71, pET28_*fae*83 and pET28_*fae*91 expression plasmids.

Table 2.5: Gene specific oligonucleotide sequences used in this study. The engineered restriction sites are underlined.

Gene	Primers (5'-3')	Length bp	T_M	Restriction sites
Fae3	*Fae*3F: 5-ATCCATATGAAAATGAAAATATTGAAGGTATTC-3	33	59 °C	*Nde*I
	*Fae3*R:5-GATCTCGAGTTACTTTATTTTTTCGAGAAACGCTTTAG-3	38	65 °C	*Xho*I
Fae9	*Fae9*F 5-ATCCATATGGCTCAGGCCTTTGATTCCCGGGATG-3	25	27 °C	*Nde*I
	*Fae9*R -GATCTCGAGTCATTGTCCCATAGGCAGTTTCGGTTTTC-3	31	22 °C	*Xho*I
Fae14	*Fae14*F: 5-ATCCATATGCCGTTCTTCGCCAGGGGTAAC-3	30	70 °C	*Nde*I
	*Fae14*R: 5-GATAAGCTTTCAGCCCGAGTGGGCCGCGATGAAG-3	34	74 °C	*Hin*dIII
Fae27	*Fae27*F: 5-ATCCATATGCGGGCGGCGAGGATCGATATGAAC-3	33	72 °C	*Nde*I
	*Fae27*R: 5-GATCTCGAGTCACAGGCTGCGCTCCAACGTG-3	31	73 °C	*Xho*I
Fae71	*Fae71*F:5-ATCGGATCCATGTCGCCATCGTAGACCGCCCATTC-3	35	74 °C	*Bam*HI
	*Fae71*R:5-GATCTCGAGCTAGCGATGAAGACATTGCTGGCAC-3	34	71 °C	*Xho*I
Fae83	*Fae83*F:5-ATCCATATGCAGGACAGCGCGCGGATCGAGCTG-3	33	74 °C	*Nde*I
	*Fae83*R: 5-GATCTCGAGTCACTCCGCCTCCGGATGCTGGGCGAAG-3	37	77 °C	*Xho*I
Fae91	*Fae91*F:5-ATCCATATGTCCTTGGATATTGGCGGCCAG-3	30	68 °C	*Nde*I
	*Fae91*R:5-GATCTCGAGTCATCCAGCAGAAGGCGGCAGACGTG-3	35	75 °C	*Xho*I
M13 forward	5- GTAAAACGACGGCCAGT	17	57 ºC	None
M13 reverse	5-CAGGAAACAGCTATGAC	17	54 ºC	None
T7 promoter	TAATACGACTCACTATAGGG	20		None
T7 terminator	GCTAGTTATTGCTCAGCGG	19		None

2.3.6. Protein methods

2.3.6.1. Protein expression

Expression of His-tagged proteins was carried out using *E. coli* BL21 (DE3) as the host strain. Expression plasmids were used to transform *E. coli* BL21 (DE3) resulting in *E. coli* BL21 (DE3)/(pET28_*fae*3, pET28_*fae*9, pET28_*fae*14, pET28_*fae*27, pET28_*fae*71, pET28_*fae*83 and pET28_*fae*91) expression strains. The two expression strains were freshly inoculated into 5 mL LB broth containing 50 µg/mL kanamycin and grown overnight at 37 °C and subsequently sub-cultured into 50 mL LB broth supplemented with kanamycin (50 µg/mL). Cells were grown until they reached an optical density (OD_{600} nm) of 0.5-0.6 using the DU 800 spectrophotomer and followed by induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. The expression was monitored hourly for 8 hours at 25 °C, shaking at 250 rpm. A corresponding culture without IPTG was used at the uninduced control. In order to improve expression levels, various growth temperatures were tested at 37, 25, 20 and 16 °C. The inducer concentration was also varied between 0.1-1 mM final concentrations.

2.3.6.1.1. Immobilized metal ion affinity chromatography (IMAC)

His-tagged proteins were purified by immobilized metal ion affinity chromatography (IMAC) using the Protino® Ni-TED packed 2000 column (Machenery-Nagel, Germany). For larger culture volumes, the purification chromatography was monitored automatically using Akta Avant 150 purifier based on the Protino® Ni-TED resin. The Protino® column of approximately 25 mL was first packed using Protino® Ni-TED resin. The column was equilibrated with a LEW buffer and 15 mL of the soluble crude fraction containing a targeted His-tagged protein was loaded. The unbound proteins were washed with three column volumes of LEW buffer followed by elution of the targeted protein with the elution buffer. Eluted targeted proteins were concentrated using a VIVASpin 10 kDa cut-off membrane spin column (Sartorius Stedim, France) and the recovered proteins were resuspended in 50 mM phosphate buffer (pH 8.0). Fractions were analysed using SDS-PAGE.

2.3.6.1.2. Ammonium sulphate precipitation

Solid ammonium sulphate was added to crude soluble protein fractions to 10% ammonium sulphate saturation. The suspension was incubated at 4 °C for 16 h with continuous mixing using a Roller Drum machine (New Brunswick Scientific, USA). Precipitated proteins were removed by centrifugation (27 000 *x g*, 30 min, 4 °C) and the recovered supernatant fraction (8 mL) was loaded onto the HIC column.

2.3.6.1.3. Hydrophobic interaction chromatography (HIC)

The recovered supernatant fraction from section 2.3.6.1.2 above was loaded onto the pre-packed phenyl Sepharose 26/10 HIC column (GE Healthcare, USA), equilibrated with HIC equilibration buffer. The unbound proteins were eluted with the three column volumes of equilibration buffer followed by elution of the His-tagged recombinant protein with HIC elution buffer. The collected fractions were analysed on SDS-PAGE.

2.3.6.1.4. Size exclusion chromatography

Proteins were loaded onto the pre-packed Sephacryl S200 size exclusion (SE) chromatography column (GE Healthcare, USA), pre-equilibrated with 300 mM NaCl, 50 mM Tris, pH 7.0. The purity of the collected fractions were analysed on SDS-PAGE and the fractions showing FAE activity were pooled and concentrated in 50 mM phosphate buffer (pH 8.0) using the VIVASpin 10 kDa cut-off membrane spin column (Sartorius Stedim, France). The globular size of proteins obtained from the SE purification steps were determined by loading the protein fractions onto the column. A standard curve was constructed using the following proteins as standards: γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.35 kDa).

2.3.7. SDS-PAGE Analysis

Sodium dodecyl sulphate page (SDS-PAGE) was used to analyse different protein fractions (Laemmli, 1970). SDS-PAGE gel preparation is represented in Appendix B3. Proteins were mixed with loading dye and boiled at 95 °C for 5 minutes before loading onto the gel. The appropriate amount of sample and dye to load on the gel was determined by using the normalisation table in Appendix B1. Electrophoresis was performed in 1× SDS-PAGE

running buffer at 180 V for 40 minutes. Gels were stained with Coomassie Brilliant Blue staining solution and de-stained with destaining solution (Table 2.3).

2.3.8. Biochemical characterisation of enzymes

Unless otherwise stated all reactions were performed in triplicate. Appropriate experimental controls were also included in all experiments.

2.3.8.1. Protein quantification

Protein concentration was estimated based on the Quick Start™ Bradford dye reagent (Bio-Rad, SA) using bovine serum albumin (BSA) at concentrations of 0.1-0.5 mg/mL. An appropriately diluted volume of the sample (10 µL) was mixed with 240 µL Bradford dye reagent followed by incubation at room temperature for 5 min and the absorbance was then measured using the Bio-Tek® spectrophotometer at 595 nm using the KC4 software. A typical BSA standard curve is shown in Appendix B4

2.3.8.2. General esterase assay based on p-nitrophenyl butyrate

Reactions were performed in 1 ml volumes containing 50 mM phosphate buffer, with 1 mM *p*-nitrophenyl butyrate (dissolved in isopropanol) as the substrate and an appropriate amount of enzyme extract. The reaction was carried out at 25 °C. The reaction rates were monitored by measuring the release of *p*-nitrophenol at 405 nm in 1 cm path length cuvettes. The background hydrolysis of the substrate was deducted using reference samples without enzyme under identical conditions. The molar co-efficient of extinction under these conditions was 0.0096 $M^{-1}cm^{-1}$. One unit (1 U) of the esterase activity was defined as the amount of esterase catalyzing the release of 1 µmol of *p*-nitrophenol per min from *p*-nitrophenyl butyrate at 25 °C. The specific activity was defined as units per mg of protein ($U.mg^{-1}$). A calibration curve of *p*-nitrophenol under these conditions is shown in Appendix B5.

Substrate specificity profiles of the purified protein were performed using *p*-NP esterified with fatty acid of different carbon chain lengths. The release of *p*-NP was measured by determining the absorbance in a DU-800 spectrophotometer at 405 nm. The reaction mixture consisted of 10 µL of diluted enzyme and 990 µL of lipase assay buffer containing 50 µL *p*-NP substrate.

2.3.8.3. FAE activity assay

The FAE activity assay was carried out in 1 mL volumes containing 3-(N-morpholino) propanesulfonic acid (MOPS) buffer, 1 mM methyl ferulate as substrate and an appropriate amount of enzyme extract. The reaction was started by addition of 10 µL diluted enzyme extract at 40 °C. After 5 minutes of incubation the reaction was terminated by adding 200 µL of 32% (v/v) HCl solution. Samples (500 µL) were transferred to HPLC vials and analysed by HPLC. The release of FA was quantified by calculating the area of the FA peak at 320 nm. The background hydrolysis of the substrate was deducted using reference samples without enzyme under otherwise identical conditions. One unit of the enzyme activity was defined as the amount of enzyme required to release 1 µmol of FA min^{-1} from methyl ferulate substrate at 40 °C. The specific activity was expressed in units per milligram of protein (U. mg^{-1}).

2.3.8.3.1. HPLC conditions

The hydrolytic activity of FAEs was determined by HPLC using a Phenomex Luna 5µ C_{18} (2) 150 x 4.6 mm reverse phase column at 320 nm and a mobile phase-0.1% (v/v) trifluoroacetic acid: 100% (v/v) acetonitrile at 60:40(v/v)-with a flow rate of 1 $mL.min^{-1}$.

2.3.8.3.2. Temperature profile

The temperature optima of the purified enzymes were determined at temperatures ranging between 20 and 70 °C at 10 °C intervals. Reactions were carried out in 50 mM MOPS buffer consisting of 1 mM methyl ferulate substrate and 10 µL of appropriately diluted solutions enzymes in a total volume of 1 mL. Reactions were incubated at 40 °C for 5 min, before termination by the addition of 200 µL HCl (36% (v/v)). The control reaction mixture (excluding the enzyme) was included to correct for auto hydrolysis of the substrate. The reaction was analysed by HPLC as described above.

2.3.8.3.3. pH profile

The effect of pH on enzyme activity was analysed using 1× universal buffer (Table 2.3). The pH of the buffer was adjusted to between 2 and 12 using 1 M NaOH or 1 M HCl. The purified enzyme samples were incubated in buffers at varying pH. FAE enzyme samples (10 µL) were incubated at 40 °C for 5 min with 1 mM MFA in 1× universal buffer at varying pH. The

reaction was performed in triplicate and the release of FA was analysed using HPLC as described previously.

2.3.8.3.4. Substrate specificities on hydroxycinnamic esters

Substrate specificities of the enzymes towards hydroxycinnamic acid derivatives were determined using the following hydroxycinnamic acid ester substrates: MpCA, MCA, EFA, MFA, MGA and MSA. FAE enzymes samples (10 µL) were incubated at 40 °C for 5 min with 1 mM hydroxycinnamic acids in MOPS buffer pH (6.5) and the release of acids was analysed using HPLC as described previously. Standard curves and peaks of the hydroxycinnamic acids are shown in Appendix B6.

2.3.8.3.5. Determination of kinetics constants

Experimental data of initial velocities versus substrate concentrations of the hydroxycinnamic acid ester substrates (0.1-10 mM stock solutions dissolved in dimethyl sulfoxide (DMSO) and water) were obtained. Standard curves of respective substrates and products were constructed in triplicate at concentrations ranging from 0.1-1.6 mM. FAE enzyme samples (10 µL) were incubated at 40 °C for 5 min with the respective concentration of hydroxycinnamic acids in 3-(N-morpholino)propanesulfonic acid (MOPS) buffer, pH 6.5. Data were fitted to the Michaelis-Menten equation using Graph-Pad Prism v6.0 (GraphPad Software, San Diego, CA, USA) to generate estimates of K_m and k_{cat} values. Kinetics graphs are shown in Appendix B7.

CHAPTER THREE: METAGENOMIC LIBRARY CONSTRUCTION, SEQUENCING AND SCREENING

3. METAGENOMICS

3.1. BACKGROUND

Soil is a complex environment that consists of a large number of bacteria, archaea, viruses and eukaryotic organisms. Due to the widespread microbial diversity found in soil, this environment is greatly preferred for gene mining as it is potentially a vast resource for natural product discovery (Daniel, 2005). Microbial diversity present in different environments has been underestimated due to the bias associated with culturing methods (Gans *et al.,* 2005). Metagenomic DNA extraction allows for the access of genomes from microorganisms that are otherwise unamenable to culturing (Handelsman, 2004).

Each environmental sample presents its own unique challenges regarding metagenomic DNA isolation and a specifically designed extraction method needs to be optimised for each targeted sample, taking into account the physico-chemical factors of the sample (Lombard *et al*., 2011). All metagenomic approaches rely heavily on the quality of the DNA and for some applications (e.g. functional metagenomics approach) the size of the metagenomic DNA is an important consideration. In principle, a functional metagenomics approach involves a number of steps: (i) extraction of high molecular weight quality DNA, (ii) end-repair of the isolated DNA (iii) ligation into an appropriate vector (which could be a plasmid, cosmid or fosmid), (iv) transformation of an appropriate host strain which is usually *E.coli* to create a gene library and screening. Depending on the approach, screening of the metagenomic library can be done either through sequence based or activity based screening.

Most of the functional libraries created to date have been based on fosmid vectors, particularly the PCC2FOS™ vector series from Epicentre (www.epibio.com). This PCC2FOS™ vector features include the use of chloramphenicol as an antibiotic selectable marker, *E. coli* F factor-based partitioning and a single copy origin of replication, an *oriV* high copy origin of replication, a bacteriophage lambda *cos* site for lambda packaging or lambda

terminase cleavage, a bacteriophage P1 loxP site for cre-recombinase cleavage and a bacteriophage T7 RNA polymerase promoter flanking the cloning site. They are suitable for DNA sizes between 25 and 45 kb. The aim of the study was to create high quality metagenomic library that would subsequently be screened for ferulic acid esterase activities.

3.2. RESULTS

3.2.1. DNA extraction from soil

Metagenomic DNA was extracted and purified from soil rich in cattle manure using a combination of physical and chemical steps. A total DNA concentration of 4 ± 0.7µg/g of starting soil material was obtained from triplicate DNA extraction attempts. The size of the extracted DNA was ≥11 kb (Figure 3.1 A and B). The extracted metagenomic DNA was of high quality as evidenced by the $A_{260/230nm}$ and $A_{260/280nm}$ ratios which were at 2.30 and 1.89. The purity of extracted DNA was of further tested by using different restriction enzymes (*Bam*HI, *Sma*I*, Hin*dIII*, and Sau*3A*)* to restrict DNA and the ability of these selected endonucleases to restrict DNA provided further proof that the DNA was sufficiently good quality for downstream processing (Figure 3.1 C).

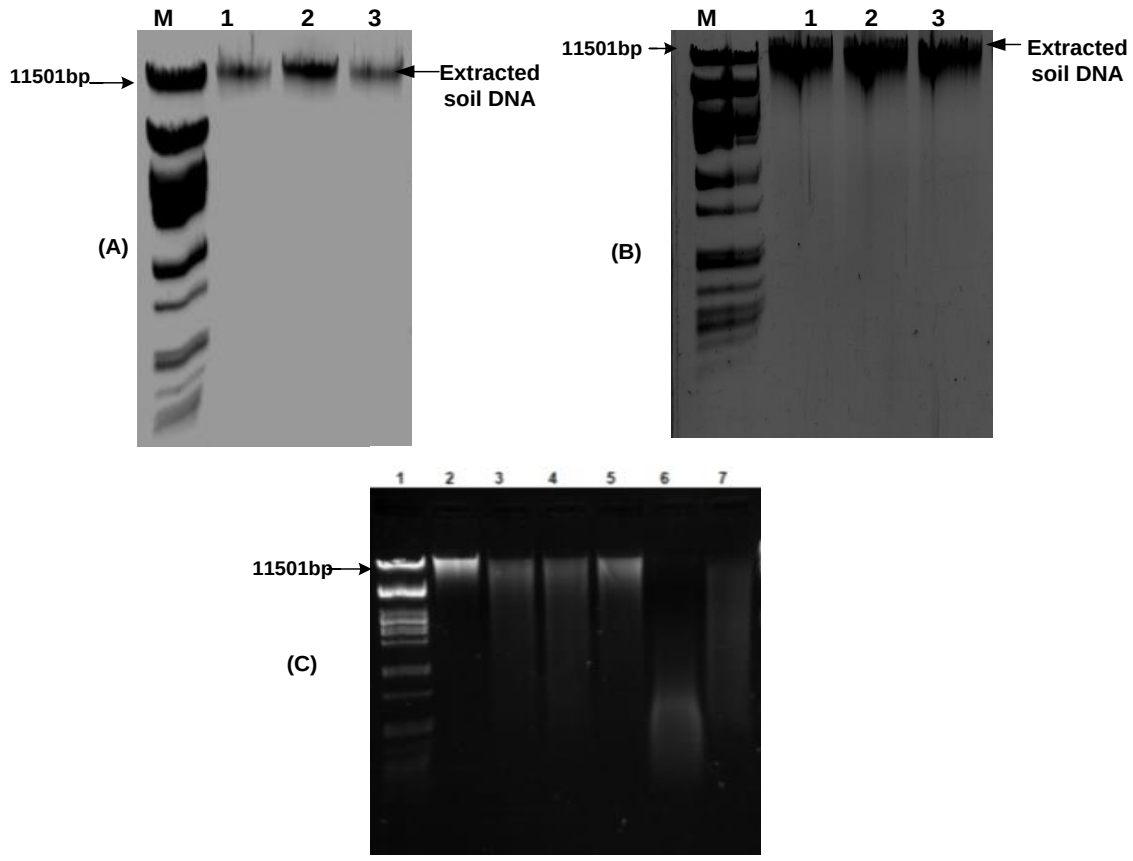


Figure 3.1: Analysis of metagenomics DNA

(A) Agarose gel electrophoresis of extracted metagenomic DNA from soil. M: λ *Pst*I DNA marker. Lane1-3: extracted DNA in triplicate using the standard ZymoResearch protocol. **(B)** M: λ *Pst*I DNA marker, Lane1-3: extracted DNA in triplicate after modification of the protocol **(C)** Purified metagenomic DNA restricted using restriction enzymes**.** Lane 1: λ *Pst*I DNA marker, Lane 2: unrestricted DNA, Lane 3-6: *Bam*HI, *Sma*I*, Hin*dIII*, and Sau*3A restricted DNA respectively*.* Lane: 7 (Collectively *Bam*HI, *Sma*I, *Hin*dIII restricted DNA).

3.2.2. Metagenomic library construction

A fosmid library was constructed using a CopyControl pCC2FOS^{TM} vector which resulted in a library size of approximately 1.4x10^5 colony forming units (cfu). The average insert size of the library was calculated to be around 28 kb and no redundant restriction pattern was observed based on *Bam*HI *and Hin*dIII restriction enzymes as shown in (Figure 3.2 and Figure 3.3).

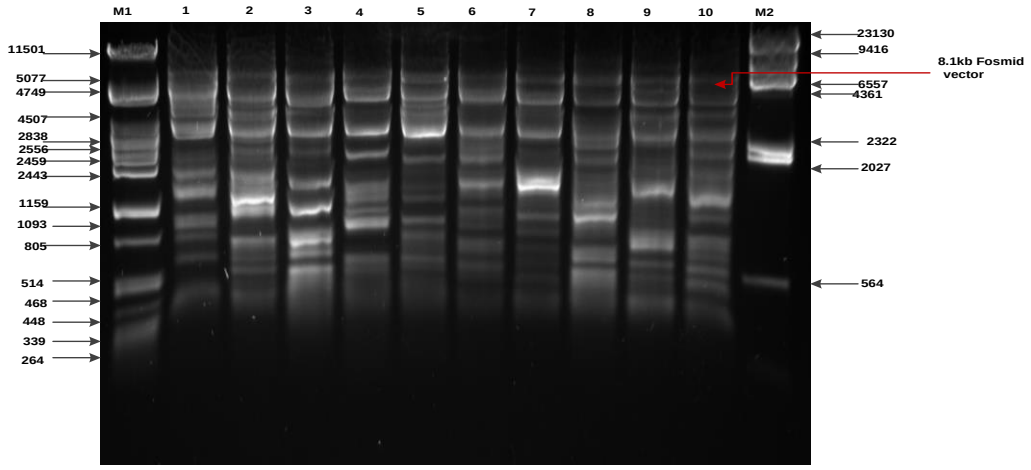


Figure 3.2: Agarose gel electrophoresis of *Bam*HI and *Sma*I resctricted randomly selected fosmid clones (Sample 1).

M1: *Pst*I; M2: λ *Hin*dIII DNA markers; Sample 1: lane1-10 represents fosmid DNA restriction (from 10 randomly selected clones) with *Bam*HI and *Sma*I restriction endonucleases.

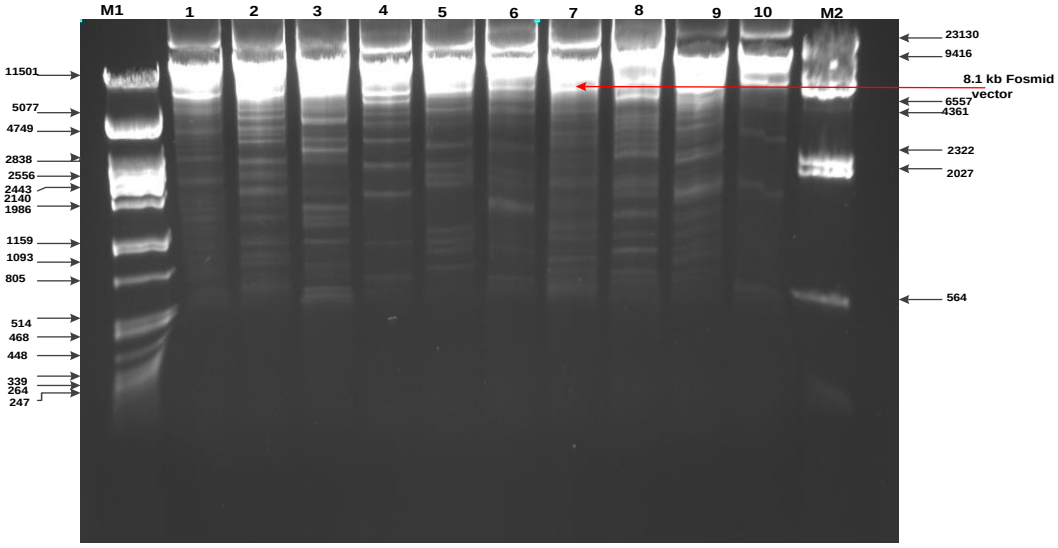


Figure 3.3: Agarose gel electrophoresis of *Bam*HI and *HindIII* restricted randomly selected fosmid clones (Sample 2).

M1: *Pst*I; M2: λ *Hin*dIII DNA markers; Sample 2: lane 1-10 represents fosmid DNA digestion (from 10 randomly selected clones) with *Bam*HI and *HindIII* restriction endonucleases.

3.2.3. Functional screening for carbohydrates active enzymes

For the purpose of this study, a total of 58410 cfu were functionally screened for esterase activity on glyceryl tributyrate substrate, resulting in 59 clones showing esterase activity equating a hit rate of 1:900. All 59 positive esterase clones were further screened for specific ferulic acid esterase (FAE) activity on ethyl ferulate agar. A total of 17 clones from 59 positive esterase clones (1:3122 hit rate) showed zones of clearance on ethyl ferulate agar (Figure 3.4). The 17 FAE clones were double restricted with *Bam*HI and *Hin*dIII restriction enzymes in order to see the digestion pattern of the clones (Figure 3.5). Clones of identical restriction patterns were considered to be the same resulting in 7 (pFos_fae3, pFos_fae9, pFos_fae14, pFos_fae27, pFos_fae71, pFos_fae83, pFos_fae91) clones with unique patterns.

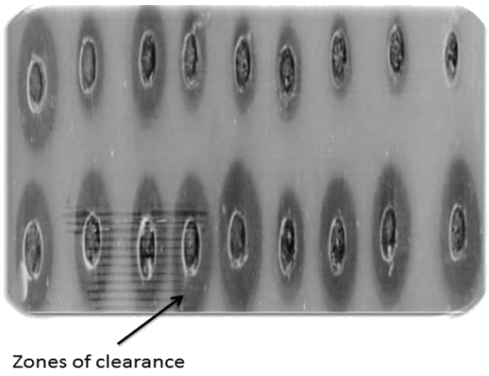


Figure 3.4: Putative FAEs producing fosmids clones screened on ethyl ferulate agar with zones of clearance on ethyl ferulate agar.

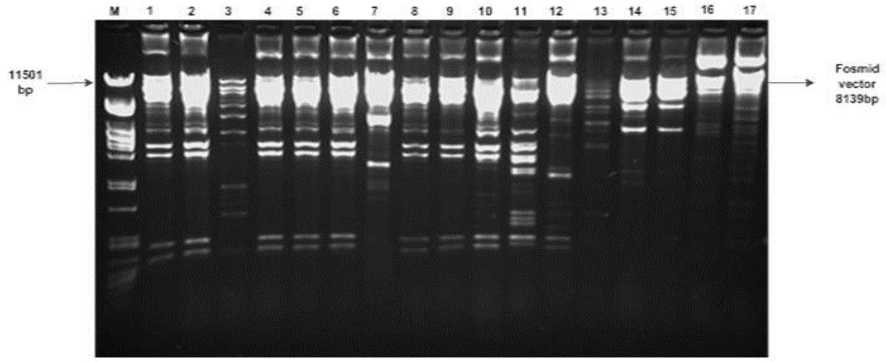


Figure 3.5 Restriction pattern of 17 putative FAEs that produced zones of hydrolysis on ethyl ferulate agar

Lane M: λ DNA *Pst*I marker, Lane 1-17: putative 17 FAEs restricted with *Bam*HI and *Hin*dIII restriction enzymes.

3.2.4. Nucleotide Sequencing

3.2.4.1. Sequence analysis of putative FAEs

Equimolar concentrations of the fosmid DNA from 7 fosmids were pooled together and sequenced using Illumina next generating sequencing platform. After sequence assembly and annotation a total of 7 putative FAEs were identified based on the presence of typical esterase motifs (G-x-S-x-G) (Dodson and Wlodawer, 1998; Warshel *et al.,* 1989) on homology searches based on BLASTp.

Table 3.1 summarises the BLASTp search results of the putative FAEs. Two ORFs FAE3 and FAE14 did not resemble typical serine motifs, although the catalytic domain scan indicated a great match to the esterase family (Figure 3.6 and Figure 3.7). The SignalP server 4.0 revealed that only ORFs for FAE71 and FAE83 contained N-terminal signal peptide with greater than 99% cut off probability, at the following positions aa 23 and aa 24 and aa 28 and aa 29, respectively.

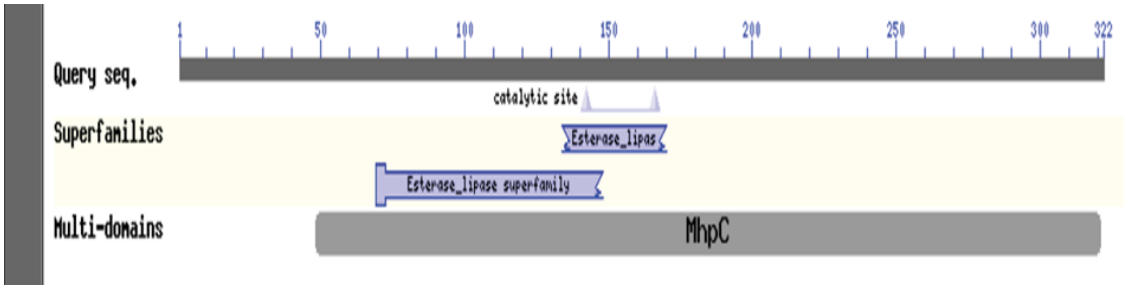


Figure 3.6: The central domain scan of FAE3 which shows that it belongs to an esterase-lipase superfamily although it does not have any known serine motif.

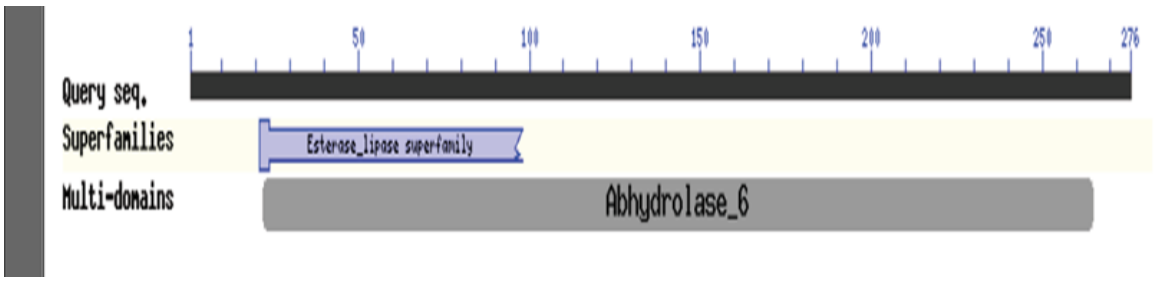


Figure 3.7: The central domain scan of FAE14 which shows that it belongs to an esterase-lipase superfamily although it does not have any known serine motif.

The FAE3 ORF was 969 bp in length and encoded a polypeptide of 322 amino acids with a predicted polypeptide of 37 kDa molecular mass predicted molecular mass. Sequence analysis of FAE3 revealed that it has 29%-56% identity at the amino acid level with several esterases homologues from *Desulfococcus oleovorans* (55%), *Bermanella marisrubri* (38%), *Pseudomonas syringae* (29%), *Pseudomonas syringae* (29%) and *Pseudomonas syringae* (30%).

The FAE9 ORF was 822 bp in length and encoded a polypeptide of 274 amino acids with 32 kDa predicted molecular mass. FAE9 has 57% highest sequence identity scores to esterase homologues from an uncultured bacterium (57%), *Subdoligranulum* sp. (54%), *Caldicellulosiruptor owensensis OL* (52%), *Clostridiales bacterium* oral taxon 876 (52%), *Clostridium papyrosolvens* (52%), and an *uncultured bacterium* (54%).

The FAE14 ORF was 831 bp in length and encoded a polypeptide of 276 amino acids with 32 kDa predicted molecular mass. Sequence analysis of FAE14 revealed that it has 29%-55% identity at the amino acid level with several homologous esterases from *Desulfococcus oleovorans* Hxd3 (55%), *Bermanellamarisrubri* (38%)*, Pseudomonas syringae* (29%), *Pseudomonas syringae* (29%) and *Pseudomonas syringae* (30%)*.*

The FAE27 ORF was 969 bp in length and encoded a polypeptide of 322 amino acids with 34 kDa predicted molecular mass. Sequence analysis of FAE27 revealed that it has 57%-59% identity at the amino acid level with several homologous esterases from *Erythrobacter* sp*.* SD-21 (57%), *Novosphingobium aromaticivorans* (59%), *Citromicrobium bathyomarinum* (59%), *Citromicrobium* sp*.* JLT1363 (59%) and *Novosphingobium nitrogenifigens* (58%).

The FAE71 ORF was 971 bp nucleotides in length and encoded a polypeptide of 321 amino acids with 36 kDa predicted molecular mass. A signal sequence of 23 amino acids was estimated at the N-terminal of the FAE71 sequence (Figure 3.8). Cleavage site between position 23 and 24: AAA-QN. Sequence analysis of FAE71 revealed that it has 48%-49% identity at the amino acid level with several homologous esterases from *Sphingobium* sp. KK22 (49%), *Sphingomonas* (48%)*, Sphingobium ummariense* (48%), *Sphingobium* sp*.* SYK-6 (49%) and *Caulobacter* sp. JGI 0001013-D04 (48%).

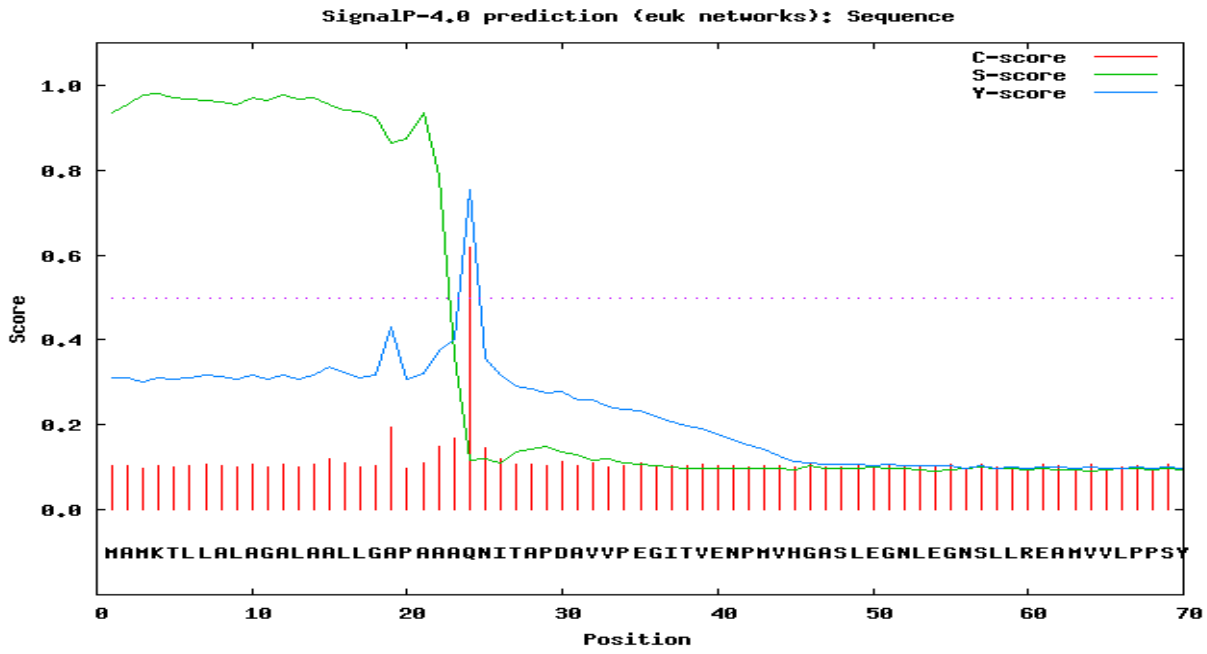


Figure 3.8: Prediction of N-terminal signal peptide cleavage site in polypeptide FAE71.

The FAE83 ORF was 951 bp in length and encoded a polypeptide of 316 amino acids in which 28 amino acids encoded a signal peptide (Figure 3.9). Cleavage site between position 28 and 29: AMA-QD. Predicted molecular mass of FAE83 was 36 kDa. Sequence analysis of FAE83 revealed that it has 43%-51% identity at the amino acid level with several homologous esterases from *Sphingobium* sp. SYK-6 (51%), *Novosphingobium aromaticivorans* (46%), *Sphingomonas* (44%) *Sphingobium* sp. Ant17 (43%) and *Sphingobium* sp. KK22 (51%).

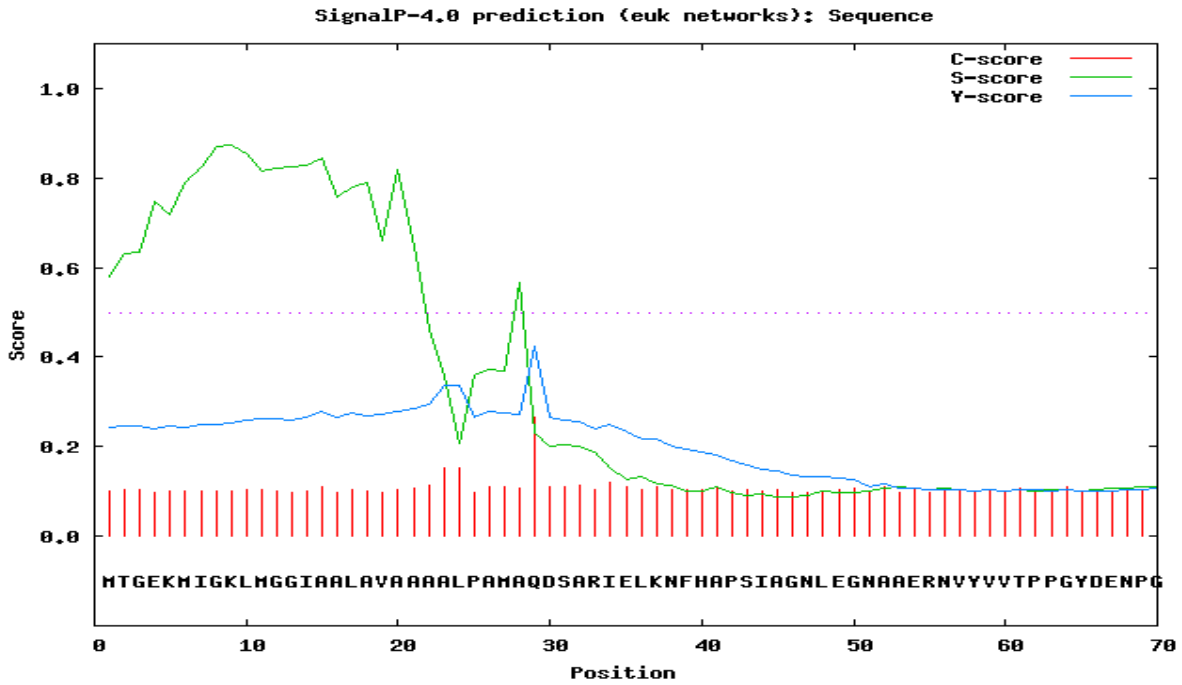


Figure 3.9: Prediction of N-terminal signal peptide cleavage site in polypeptide FAE83.

The FAE91 ORF was 1062 bp nucleotides in length and encoded a polypeptide of 353 amino acids with 38 kDa predicted molecular mass. Sequence analysis of FAE91 revealed that it has 47%-49% identity at the amino acid level with several homologous esterases from *Acidobacteriaceae bacterium* KBS 89 (50%), *Sphingobium* sp. KK22 (49%), *Sphingomona* (48%), *Acidobacteria* b*acterium* KBS 146 (47%) and *Caulobacter* sp. JGI 0001013-D04 (49%).

Table 3.1: Summary for a BLAST search of putative FAEs.

Genes	Nucleotide length (bp)	Signal peptide	Amino acid length	Mw (kDa)	Identity (similarity) %	Amino acid sequence	Serine motif	Best Hit
Fae3	969	No	322	37	56	MKMKILKVFKTIALVLFAIIAALLIAST INHQTLLKNEFKEYYPSDGMVEVDGNKF HVYSEGAGDLTLVFMSGHGTNSPTLDFK PLWMKMVDDYRIAIVEKSGYGWSETSNS PRDIDTILEETRKALELSGEKGPYVLFP HSMSGLEAIYWAQKYPDEVKAIIGLDPC TPETIKILPEPEKTQLHLMYSISRMGIT RFMPDFDFESNFPLMKSEDLTEENKQEY LAVFYKSAFSKDMLREVNYLYDNAKTVA ENEVPIHTPMYFFISDDQEAIAIGWKEA SSGYLSKITNGKHMQFATGHYVHYDKSD IIAEEAKAFLEKIK	No known serine motif	Alpha/beta hydrolase fold protein [Dethiobacter alkaliphilus] WP_008514622.1
Fae9	822	No	274	32	57	MATFQIDFYSNALKGITPLTAILPVEIP ENIPGIEARERKEPFRTIYLLHGYSGSS NDWLHGSRIDFLARIFQVAVVMPSGRNS FYLDDAIKDEYYERLVSEEIVEFSRKVF PLSQKREDTTIGGLSMGGYGAMRNGLKH SDVFGSIFAFSSALITDKTAQMKEGDSN PVMAPFSYYRHVFGDLDTLIGSDKDPKA LAKRLVDNGSDIPKIFMACGTEDFLIGE NRDFHRHLTDLGIEHEYRESPGIHDWVF WDAYIEKAMEWLYGKPKLPMGQ	DTTIG**GLSMG**GYG (125-129)	FAE1 [uncultured bacterium] AGJ83839.1
Fae14	831	No	276	32	55	MPFFARGNARIYYEDRGSGEPVIAVHGL IGNTRYWKPLTDLIGERCRFISMDMRGH GYTEVNGEPHGYDVQTVGRDIVALADHL GIPRFHLLTHSTGGFAGVRHAMDDCSRF KSLILTNTASATSVVPGDERTIRDYHDR FAAWFQRFDWDQIMGVLKMTPGPFFRGV VESPSSGELLVLAREVVRIGNRDLIAAF IRSFYTDPDPRVDGLRRISCPVLIVTGE KDDLMLEPSRLMAREIPGARLLEYEGVG HMSALEAPDRLARDVMDFIAAHSG	No known serine motif	alpha/beta hydrolase fold protein [Desulfococcus oleovorans]WP_012175009.1

Fae27	969	No	322	34	57	MRAARIDMNGNDNSMRPDVRALLDMMAA MQQPKLYELPLEEARSPALVALVDAPPR DLPVIRDLSCPGPAGDIPLRYYDARDRR EAGPAILFLHGGGFVIGNIDTYHSLCAE IAARTGLPVVSVDYRLAPEHPFPAAPDD CEAAARWLAASPTELGFDVTGLIPMGDS AGGNLTIVTTLALLDRAAQVPVVMQVPI YPIASAIEDHASMREFGAGHFLEAEIMG WFTHCYAPDPASPRNYPLLREDHSQTPP TVLLTAGLDPLRDSGREYGAALASAGVD LTLIEARGMIHGFLQMRKALPSTSRDMD ALFAAMAATLERSL	ILFLH**GGGF**VIGNI (95-98) GLIPM**GDSAG**GNLTI (166-170)	Lipase [Erythrobacter sp. SD-21] WP_006832276.1
Fae71	971	Yes (23 and 24: AAA-QN D=0.842 D-cutoff= 0.450)	321	36	49	MNSDMLQITKKLALSAFALIIFAGCTQK DEQNTFQNDGNPLIKNKFTADPAPLVHD GTLYLYVGHDEYYEGQDTASGGKEFNIT EWLCYSTEDMQKWTDHGSVLKPTDFEWG VGEAWASQVIEHEGKFYYYTTVQAGEPY NSKVIGVAVSDSPTGPFTDAIGKPLITD EMTSNGARGWWNDIDPTFIMDEEGQAWL SWGNGTCFLARLKPNLIELDGEIETVDL PLFVEGPWLHQRGDLYYLTYASMGEGRE TISYATAPSMEGPWTPQGELTGMAENSF TIHPGIVEFNGSWYLFYHNATLTLDGIE GAIGRRSVCVEELHYNPDGTMQFVEQTT EGVASR	GRALV**GHSMG**GYG TW (166-170)	Esterase [Sphingobium sp. SYK-6]WP_014074884.1
Fae83	951	Yes (28 and 29: AMA-QD D=0.557 D-cut-off=0.450)	316	36	51	MQDSARIELKNFHAPSIAGNLEGNAAER NVYVVTPPGYDENPGKRYPVVVFLHGYW ATPQMYQETMKFEEAVDIAAEAGNEVIM VIPDGHSKLRGGFYSSGPTVGDYESFVA RDLVGWVDANYRTLAKPESRGLAGHSMG GYGTIRIAMKNPGVFSSIYMMSACCLDP MPINAETAQRIEAMSAEEAANADFGQLA PVSTLATWSPDPTAEGWLKADTGLKEDG TVDPLVNYRLAANSPVVILPQYLPALNG LRAFAMDIGDKDFLLEGNRIFREELDRF GVKYDFELYEGDHGNRIPERIRAEVLPF FAQHPEAE	SRGLA**GHSMG**GYG TI (136-140)	Esterase [Sphingobium sp. SYK-6]WP_014074884.1

Fae91	1062	No	353	38	50	MSLDIGGQGFVLAGNGVFREGRDRCGVQ CDCERYEGDQGNRMPERTRRRRAALLRP ASGGGVMAMKKLLALAGALAALLGAPAA AQNITAPDAVVPEGIAVENPMVHGASLE GNLEGNNLLREAMVVLPPSYATSPERHY PVVYYLHGFAISGRDFYDFMQVPTAVAD NAAAGREFIVVVPDTLTRMGGSMYSNSV TVGDFQTFVARDLVAYIDSHYRTIPERE GRALVGHSMGGYGTWRIGMRYPDVFNAI WAQSACCISPRQETAESAAAMAAVPFEG VDESGFGMRAGLASMTAWSPNPLNPPFH ADFPLGEDGEVDTLVIAQWAAPSCLLSP SPSPRASWPARLPPSAG	GRALV**GHSMG**GYG T (230-234)	Hypothetical protein [Acidobacteriaceae bacteriumWP_0207130 12.1

3.2.4.2. Analysis of deduced amino acid sequences

The deduced amino acid sequence of FAE71, FAE83 and FAE91 contained GHSMG sequence motif, while FAE9 and FAE27 contained a GLSMG and GDSAG sequence, respectively. These sequence motifs corresponded well to the G-x-S-x-G signature motif that is conserved in esterase families. Multiple sequence analysis of FAE3 and FAE14 with other two closely related sequence homologues did not reveal any conserved pentapeptide serine motifs (Figure 3.10), although the catalytic domain scan indicated a great match to the esterase family. Sequence homology of FAE9 was seen when compared with other five closely related esterase amino acid sequences from a BLAST search. The conserved motif GLSMG was identified at position 129-133 (Figure 3.11) and the oxyanion hole at position (49-50). The possible catalytic triad of FAE9 is also indicated in Figure 3.11, where serine is at position 131, histidine at position 183 and aspartic acid at position 223. These amino acids involved in a catalytic triad are highly conserved in all regions.

Multiple sequence alignment of FAE27 was performed with related esterase sequences and it showed the GDSAG pentapeptide motif at position 175-179 (Figure 3.12). The FAE27 catalytic triad is at serine 176, histidine 299 and aspartic acid 273. FAE27 also had the GGGF motif at position 102-105 motif and an oxyanion hole at the same position. The multiple sequence alignment of FAE71, FAE83 and FAE91 revealed conserved pentapeptide motifs at position Ser 232, Asp 311, and His 147 (Figure 3.13). The oxyanion hole denoted by the His-Gly (HG) dipeptide was conserved amongst all aligned putative FAEs at position 47-48 (Figure 3.13).

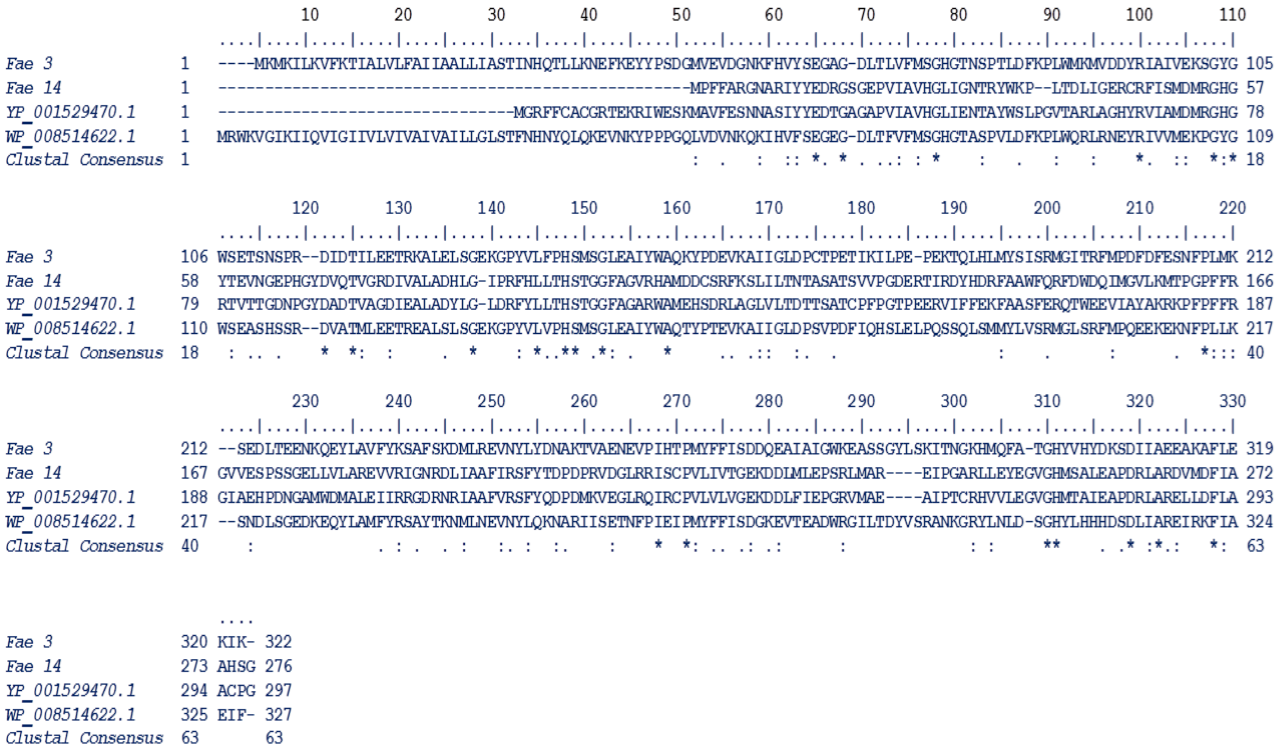


Figure 3.10: Multiple sequence alignment of FAE3 and FAE14 performed with ClustalW. FAE3 and FAE14 were aligned with two esterases homologues and they did not show any known esterase motif.

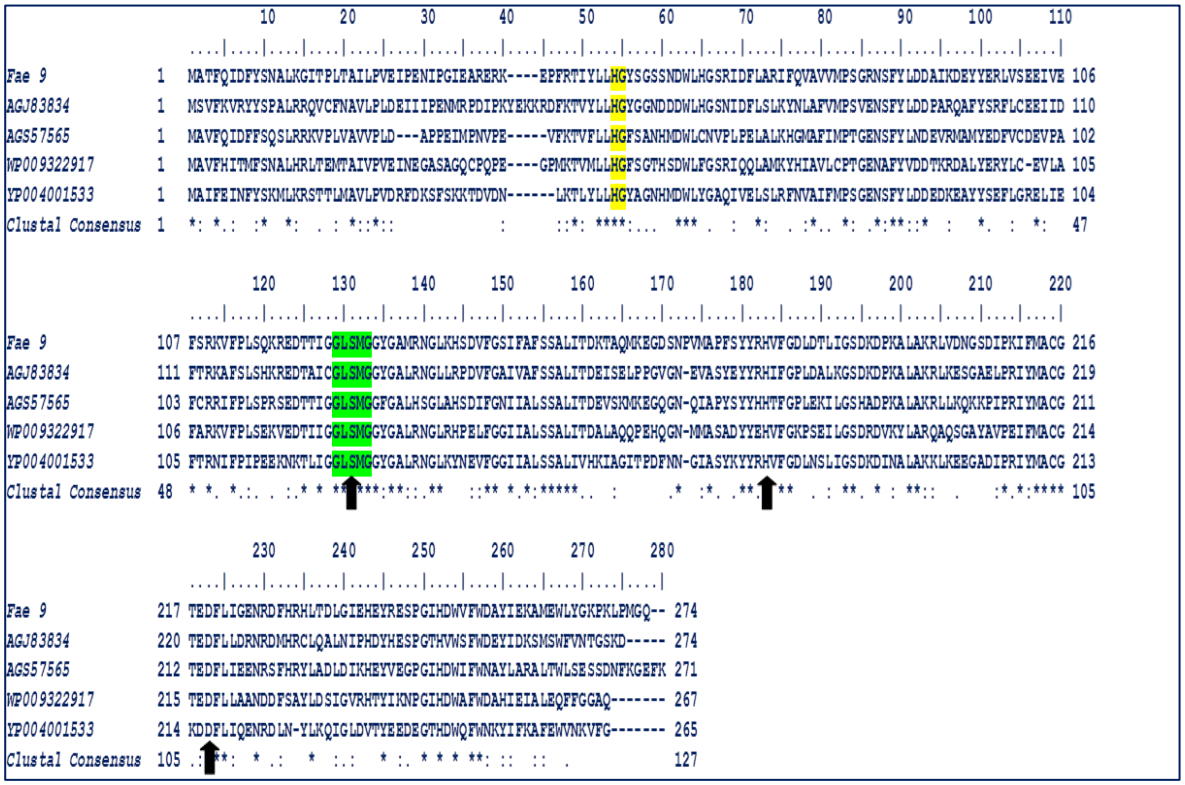


Figure 3.11: Multiple sequence alignment of FAE9 performed with ClustalW. FAE9 was aligned with five putative esterases that showed similarity with FAE9. The pentapeptide motif is highlighted in green and the oxyanion hole HG is highlighted in yellow. The amino acids involved in the catalytic triad are indicated with a black arrow with Ser 131, Asp 223 and His 183.

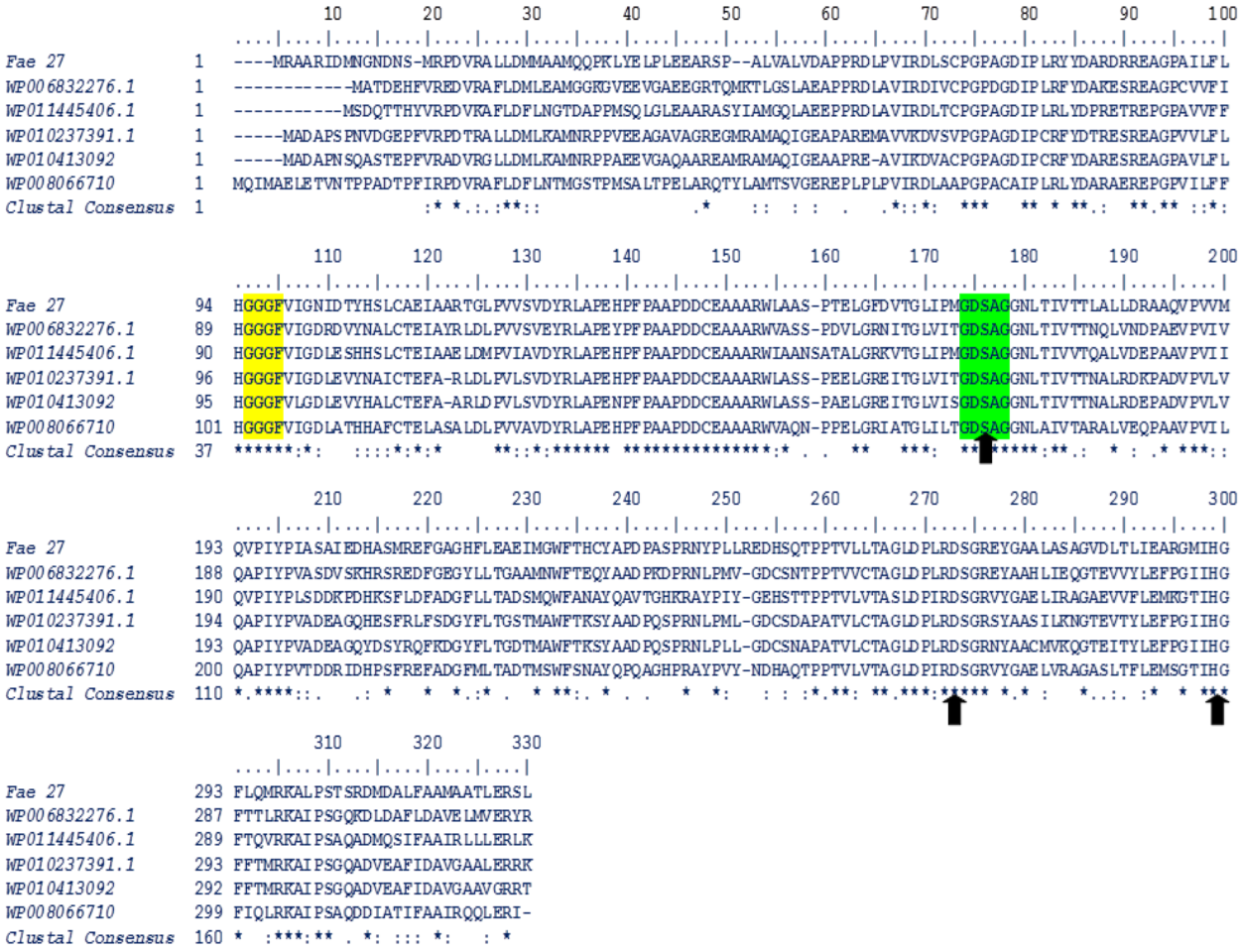


Figure 3.12: Multiple sequence alignment of FAE27 performed with ClustalW. FAE27 was aligned with five putative esterases that showed similarity with FAE27. The pentapeptide motif is highlighted in green and the GGGF oxyanion region motif highlighted in yellow. The amino acids involved in the catalytic triad are indicated with a black arrow with Ser 176, Asp 273 and His 299.

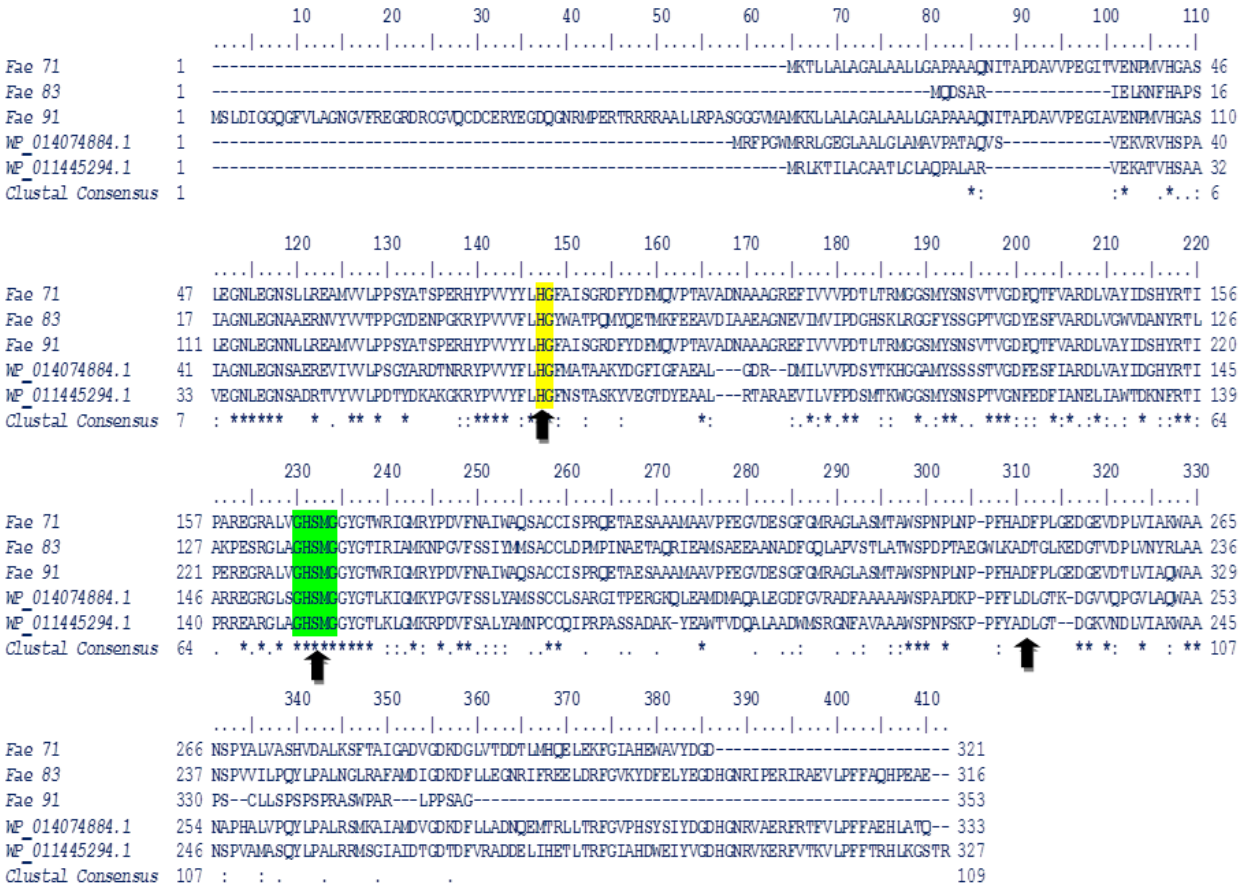


Figure 3.13: Multiple sequence alignment of FAE71, FAE83 and FAE91 performed with ClustalW. The pentapeptide GHSMG motif is highlighted in green and the oxyanion hole (HG) are highlighted in yellow. The amino acids involved in the putative catalytic triad of FAE71, FAE83 and FAE91 are indicated by a black arrow with Ser 232, Asp 311 and His 147.

3.2.5. Phylogenetic tree analysis

Primary structures of FAE3, FAE9, FAE14, FAE27, FAE71, FAE83 and FAE91 were compared with other FAE sequences of (Type A, B, C, D and E) esterases as assigned by Crepin *et al.* (2004). A phylogenetic tree (Figure 3.14) was achieved by using neighbouring-joining analysis from the MEGA 4.0.2 software programme. The phylogenetic tree indicated that all the seven FAE clusters are independent of the four classes; therefore, this shows that this could be new FAE sequences.

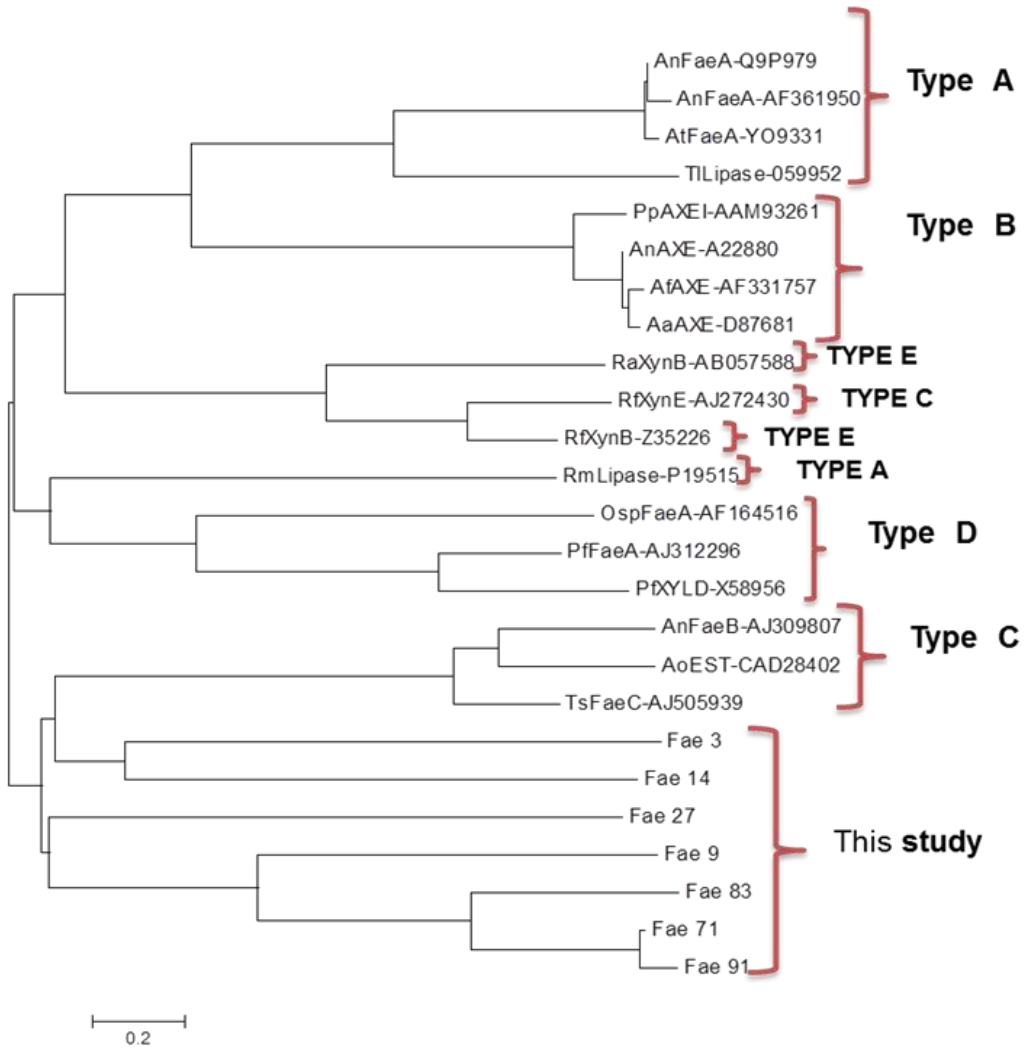


Figure 3.14: Phylogenetic analysis of FAE3, FAE9, FAE14, FAE27, FAE71, FAE83 and FAE91.

The classification scheme of Type A, B, C, D and E (as represented by Crepin *et al*., 2004) are specified on the figure. Sequence names with their accession numbers are shown on the right of the tree with Type A having AaFaeA *Aspergillus awamori* FaeA (**Q9P979**), AtFaeA *Aspergillus turbingensis* FaeA (**Y09331**), RmLipase *Rhizomucor miehei* lipase (**P19515**), AnFaeA *Aspergillus niger* FaeA (**AF361950**) and TlLipase *Thermomyces lanuginosus* lipase (**O59952)**. Type B: PpAXEI *Penicillium purpurogenum* AXE I (**AAM93261**), AaAXE *Aspergillus awamori* AXE (**D87681**), AfAXE *Aspergillus ficuum* AXE (**AF331757**) and AnAXE *Aspergillus niger* AXE (**A22880**). Type C: RfXynE *Ruminococcus flavefaciens* XynE (**AJ272430**), AoEST *Aspergillus oryzae* selective esterase (**CAD28402**), TsFaeC *Talaromyces stipitatus* FaeC (**AJ505939**). Type D: PfFaeA *Penicillium funiculosum* FaeA (**AJ312296**), PfXYLD *Pseudomonas fluorescens* XYLD (**X58956**). Type E: RaXynB *Ruminococcus albus* XynB (**AB057588**) and RfXynB *Ruminococcus flavefaciens* XynB (**Z35226**), PeESTA *Piromyces equi* ESTA (**AF164516**).

3.3. DISCUSSION

Soil is one of the key sources that is rich in antibiotics and biocatalysts. Soil metagenomics has yielded many diverse and novel enzymes for a broad range of applications. Other factors such as soil contamination also favor the occurrence of esterases and the appearance of certain biochemical properties in the enzymes isolated from contaminated samples (López-López *et al*., 2014). Soil metagenomic studies assist with the accessibility of novel enzymes from soil that has many potential applications in industries (Daniel, 2005). In this study, a metagenomic library was constructed from soil. DNA was extracted and purified since direct extraction of DNA can result in co-extraction of humic acids which interfere with cloning efficiency and transformation efficiency (Daniel, 2005). Problems associated with DNA extracted from acidic soils have also been reported (He *et al.*, 2005). For each soil type the protocol for DNA extraction needs to be modified (de Castro *et al*., 2011). Initial attempt of DNA extraction using the standard protocol from ZymoResearch resulted in a DNA size of ≥ 11 kb. Following the optimisation using a bead-beating step resulted in a DNA size greater than that 11 kb which was suitable for metagenomic DNA extraction. High quality and high molecular weight DNA was extracted (Figure 3.1). The extraction of high molecular weight DNA is compatible with fosmid library construction.

In this study, a fosmid library was constructed using a CopyControl pCC2FOS^{TM} vector which resulted in a library size of approximately $1.4x10^5$ cfu. These results are in line with the metagenomic libraries constructed in other studies, demonstrating the capability of cloning environmental genomic DNA into fosmids as an approach that is independent of culturing (Rondon *et al*., 2000). Pang *et al*. (2008) constructed a fosmid library from forest topsoil containing 3624 fosmid clones with insert sizes ranging from 23 kb. Fosmids are useful for constructing stable metagenomic libraries from complex environmental samples. However, during the process of fosmid library construction a large amount of DNA can be lost due to handling and processing (Rondon *et al.*, 2000).

FAE71 and FAE83 sequence analysis revealed a leader peptide at the N-terminal consisting of 23 and 28 amino acids, respectively (Figure 3.8 and Figure 3.9 above). The presence of the leader peptide indicates that FAE71 and FAE83 are extracellular enzymes that are exported through the membrane with the help of the leader sequence. The activities of extracellular FAEs have been reported by Faulds and Williams. (1991).

The catalytic triad of esterases consists of Ser-Asp-His and the consensus pentapeptide sequence (Gly-x-Ser-x-Gly), where x represents any amino acid (Bornscheuer, 2002). Analysis of the FAEs primary structures identified in this study revealed a classical GxSxG

pentapeptide motif near the serine active site. This motif has been reported in other FAEs primary structures (Rashamuse *et al*., 2009). The nucleophilic serine is positioned by the nucleophilic elbow to allow easy access on one side by the active site histidine residue and on the other side by the substrate. FAE9 has a pentapeptide motif at position (129-133) and the oxyanion hole HG at position (49-50). The oxyanion hole plays an important role in the stabilization of the substrate in the transition state, whereby its two residues donate their backbone amide protons for stabilization (Sumby *et al*., 2009). FAE27 comprises of the N-terminal oxyanion GGGX signature motif at position (102-105) and the nucleophilic elbow GDSAG motif at position (175-179) (Figure 3.12). The GGGX motif is known to be associated with the specificity of tertiary alcohols (Henke *et al*., 2002).

The catalytic triad for FAE27 is at Ser 176, His 299 and Asp 223. The serine that is involved in the catalytic triad is located at the centre of the nucleophilic elbow. The location of the catalytic serine on the nucleophilic elbow is one of the conserved features of these α / β fold subfamily (Schubot *et al.,* 2001*)*. This has been presented by the crystal assembly of FAEs from EstIM 1 and Xyn10B ferulic acid module (Jeon *et al*., 2012; Prates *et al*., 2001). The multiple sequence alignment of FAE71, FAE83 and FAE91 revealed conserved pentapeptide motifs at position Ser 232, Asp 311, and His 147 (Figure 3.13). The oxyanion hole denoted by the His-Gly (HG) dipeptide was conserved amongst all aligned putative FAEs at position 47-48 (Figure 3.13). The results from these sequence analysis of these esterases shows that they belong to serine esterases

FAEs have been classified into (Type A,B,C,D and E) based on the identity of primary amino acid sequence, substrate specificity and the ability to release diferulic acids from model and complex substrates (Crepin *et al*., 2004). Phylogenetic analysis of seven FAEs identified in this study suggests that they do not belong to any classification system assigned by Crepin *et al*. (2004) (Figure 3.14). This classification system is only limited to primary amino acid similarity, there is no advance connection that could be recognized due to the absence of comparable data of enzyme activity. However, Crepin *et al.* (2004) also reported that comparison of protein sequences and corresponding enzyme activity data using synthetic methyl esters will assist in the validation of enzyme classes in phylogenetic analysis. Another classification system for FAEs has been developed by Udatha *et al*. (2011) based on a computer algorithm. This system could be used to further classify these enzymes.

CHAPTER FOUR: EXPRESSION, PURIFICATION AND BIOCHEMICAL CHARACTERISATION OF FAE9 AND FAE27

4. EXPRESSION, PURIFICATION AND CHARACTERISATION

4.1. BACKGROUND

Systems for recombinant production of proteins have been developed to meet the demands of the biotechnology industry (Demain and Vaishnav, 2009). The desired gene is cloned and the protein is amplified into suitable expression host systems such as bacteria, yeasts, insects and plants (Demain and Vaishnav, 2009). The best expression system is selected taking into consideration the protein quality, function, yield and speed of production. Recombinant protein production in *E. coli* expression systems continues to be the dominant method for laboratory investigation (Chen, 2012). Advantages of using an *E. coli* expression system include rapid growth, rapid expression, ease of culture and high productivity (Swartz, 1996). *E. coli* has been used for the cost proficient production of many commercialized proteins (Jonasson *et al*., 2002). *E. coli* recombinant expression systems have been developed to facilitate maximum protein recovery, soluble protein production and easy protein purification. However, expression in *E. coli* does not guarantee that a protein will correctly fold or will accumulate in a significant amount, or will be functional, even if present in high levels.

Over the years many strategies have been developed that promote solubility of the targeted recombinant proteins. Some of these strategies include the addition of fusion tags. Some expression vector systems allow the expression of the protein of interest as a fusion partner to improve both solubility and purification. Solubility tags can sometimes be combined with 6-histidine tag for both solubility and affinity function (Esposito and Chatterjee, 2006).

Most expression vectors compatible with the *E. coli* expression systems provide a number of advantages aimed at improvising solubility and easy purification of the target recombinant protein(s). Some of the more common fusion partners include a histidine tag (Smith *et al*., 1988), maltose-binding protein (MBP) (Maina *et al*., 1988), thioredoxin (Lu *et al*., 1996), glutathione-S-transferase (GST) (Smith and Johnson, 1988) and pectate lyase (pelB) (LaVallie and McCoy, 1995). Recombinant proteins produced in *E. coli* systems can also be purified using classical chromatographic methods based on size exclusion, ion exchange

and hydrophobic interaction that separate proteins according to size, charge and hydrophobicity, respectively (Rosano and Ceccarelli, 2014).

FAEs have significant differences in biochemical and biophysical characteristics. FAE molecular masses range from 27-210 kDa and most FAEs show maximum activity at temperature and pH values between 30 and 65 °C and 3.0 and 9.5, respectively (Fazary and Ju, 2007). Typically substrate specificity of FAEs is characterised using synthetic methyl esters of phenyl alkanoic acids, where applicable the specificity of FAEs to cleave 1,5 ester bond is usually done using O-[5-O- (transferuloyl)-α-L-arabinofuranosyl]-(1->3)-O-β-D-xylopyranosyl-(1->4)-D-xylopyranose (FAXX). FAXX is not commercially available, hence few reports are on their specificity (Rumbold *et al*., 2003). The aim of this chapter was to develop expression plasmids for the previously identified seven putative *fae* genes through PCR sub-cloning and subsequent purification and characterization of the recombinant FAEs.

4.2. Expression strategy

In order to design the expression plasmids for heterologous expression of putative FAEs, gene specific primers were designed and used to amplify nucleotide sequences matching to the putative *fae* sequences. The expression strategy adopted in this study is shown in Figure 4.1. Restriction enzymes that did not cut within the respective gene sequence and designed into the primer sequences used are shown in Table 2.5.

The primers were designed with restriction sites that allow cloning into pET28a vector. The reverse primers were also designed without a stop codon to allow recombinant proteins to be expressed in-frame with the C-terminus tag. Correct PCR products (FAE 3, FAE 9, FAE 14, FAE 27, FAE 71, FAE 83 and FAE 91) were obtained (Figure 4.2) using touch down PCR.

The purified PCR products were ligated into a blunt end pJET1.2 vector and transformed into *E. coli* DH5α and stored as glycerol stocks at -80 °C. Screens harbouring the correct insert sizes were excised, purified and ligated into the expression vector pET28a. The pET28a expression plasmids were transformed into *E. coli* BL21 (DE3) competent cells for expression studies. The ligation reaction was confirmed by restriction digestion. Plasmids of three and five randomly selected recombinants of pJET1.2 and pET28a, respectively, were restricted with respective restriction enzymes and analysed on agarose gel (Figure 4.3 and Figure 4.4).

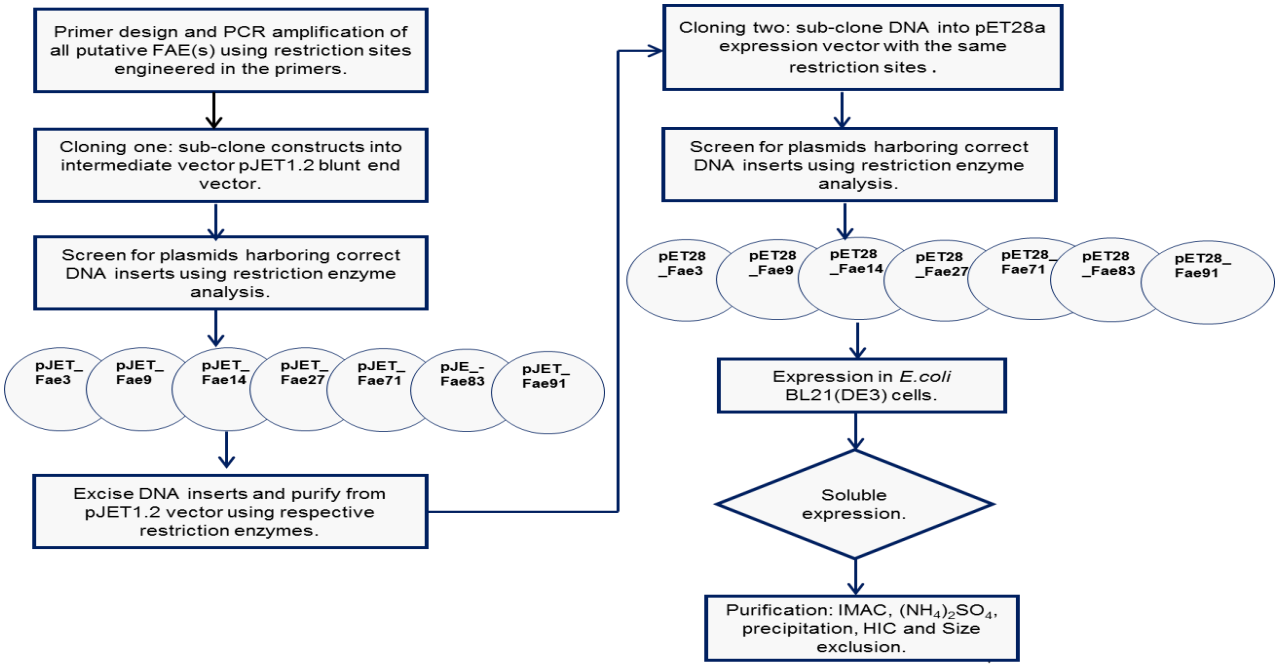


Figure 4.1: The expression strategy adopted in this study.

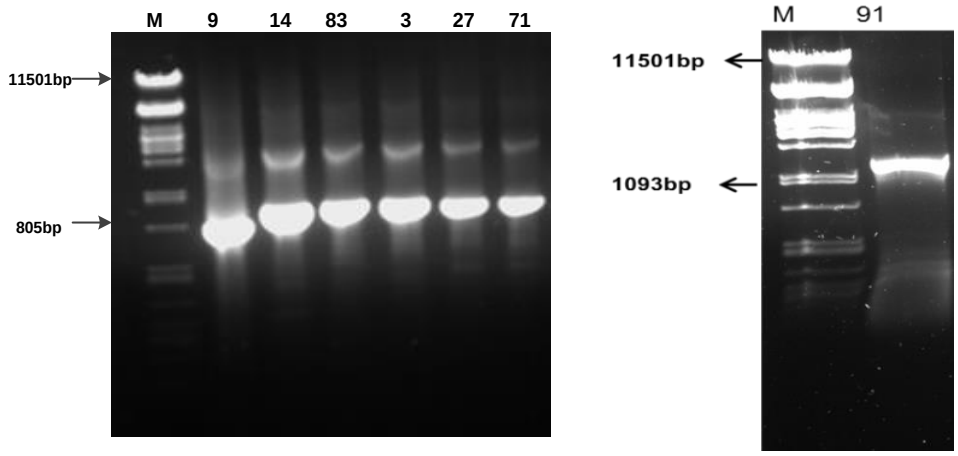


Figure 4.2: Amplified PCR product of seven FAEs.

Lane M: λ *Pst*I marker. Labels: 9, 14, 83, 3, 27, 71 and 91 represents the seven identified FAEs.

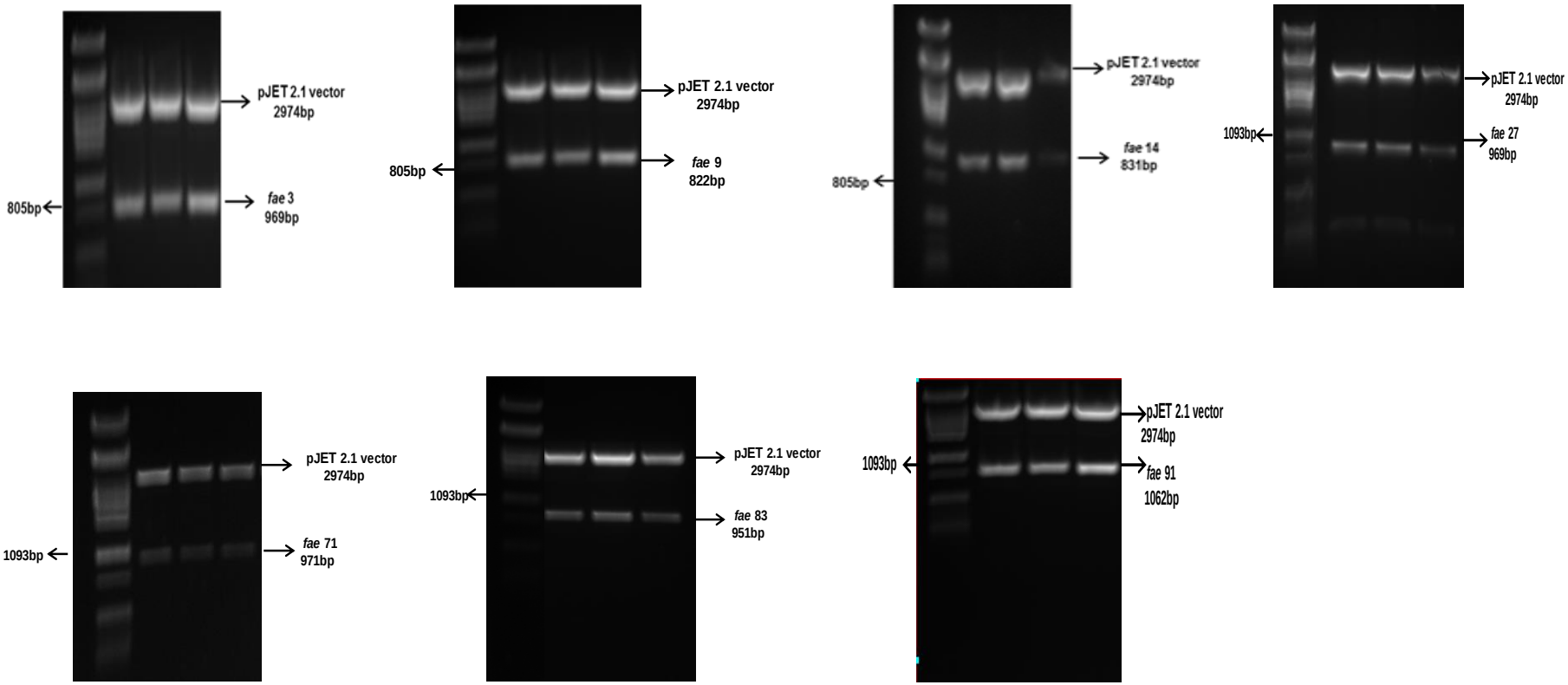


Figure 4.3: Agarose gel images confirming the successful cloning of seven FAEs into pJET1.2 vector

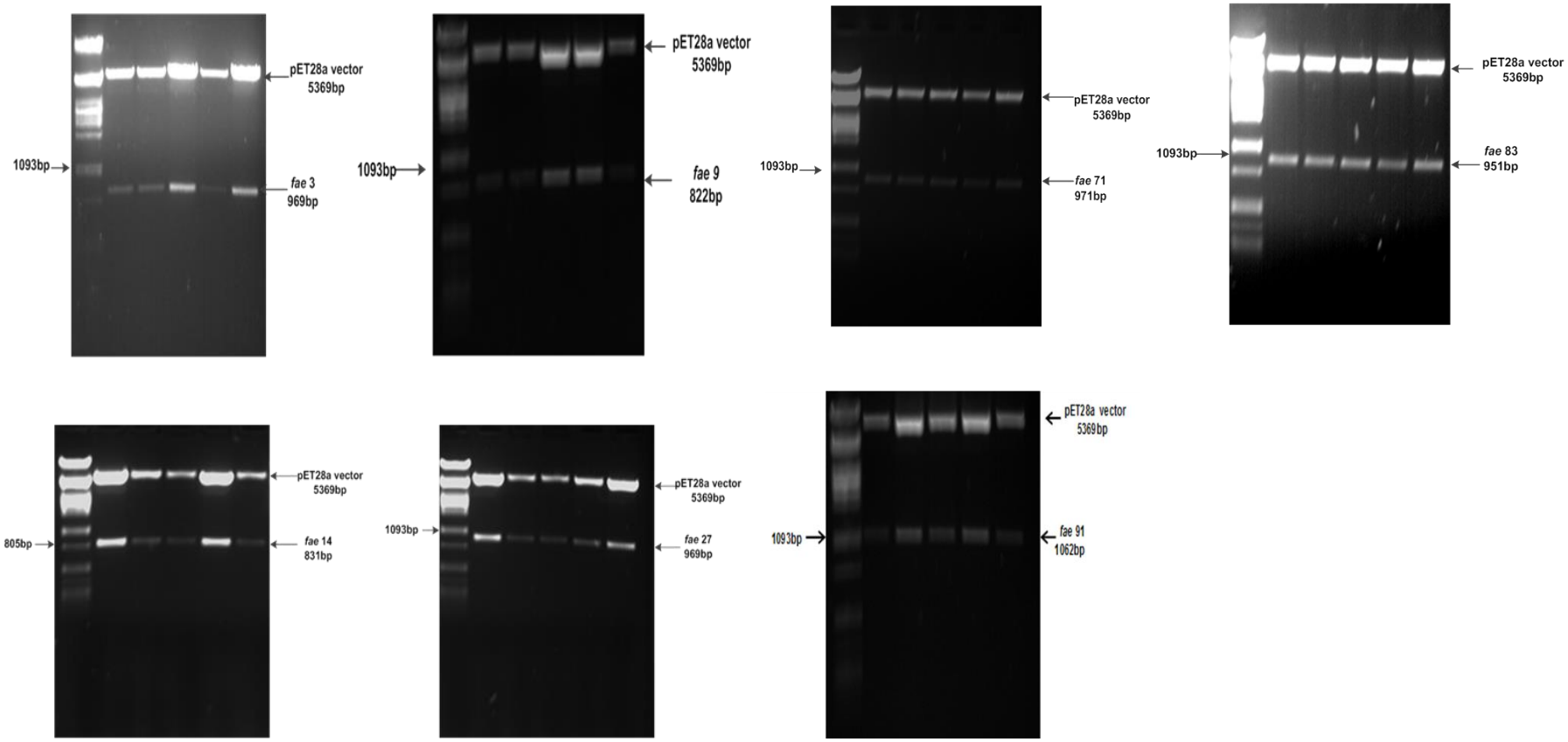


Figure 4.4: Agarose gel images confirming the successful cloning of seven FAEs into pET28a vector

Out of seven *fae* genes only two were expressed (FAE9 and FAE27) and the rest failed to express even after a number of attempts to optimise the growth temperature (16 °C -37 °C) and inducer concentration (0.1mM-1mM) and also trying different growth media. As a result only FAE9 and FAE27 were used for further study. SDS-PAGE gel images of pET28_*fae3*, pET28_*fae14,* pET28_*fae71*, pET28_*fae83* and pET28_*fae91*, which showed no expression at the expected protein size are shown in Appendix B2.

4.2.1. Expression and purification of FAE9 and FAE27

Soluble fractions of cell extracts FAE9 and FAE27 proteins were analysed on SDS-PAGE to determine the expression levels of the two proteins. The estimated subunit molecular masses were compared with the theoretical subunit molecular masses calculated from the respective amino acid sequences. Cells were grown at temperatures of 25 °C and 37 °C in order to optimise the expression of the proteins.

Gel images for induced and uninduced FAE9 are shown in Figure 4.5. A protein band was observed for FAE9 with a subunit molecular mass of 29 kDa in the cells that were induced with 0.1 mM IPTG. The time expression profile over 24 hours at 25 °C showed an increase in targeted protein band intensity. Since the ORF encoding FAE9 was subcloned into pET28a in such a way that the recombinant proteins contained a 6× His-tag at their C-termini, purification using nickel affinity chromatography was attempted. FAE9 purification was not successful with this method (Figure 4.5 C). Therefore, it was sub-cloned into pET30b and the 6× His-tag was moved from C-terminal to the N-terminal using *Nde*I and *Xho*I restriction sites but still not successful, therefore, we used classical purification methods. FAE9 was first subjected to ammonium precipitation and then purified using hydrophobic interaction chromatography.

FAE27 was over-expressed over 24 hours and the induced and uninduced FAE27 SDS-PAGE analysis is shown in Figure 4.6 A and B. A protein band was observed for FAE27 with a molecular mass of 34 kDa in cells that were induced with 1 mM IPTG. The time expression profile at 25 °C showed an increase in band intensity or an increased expression as the cells were grown up to 24 hours. There was not much difference in the expression profile of FAE27 at different temperatures. FAE27 protein was purified using the His-tag Ni-TED purification column, and SDS-PAGE gel analysis showed the ability of FAE27 to bind well to a Ni-TED column (Figure 4.6 C).

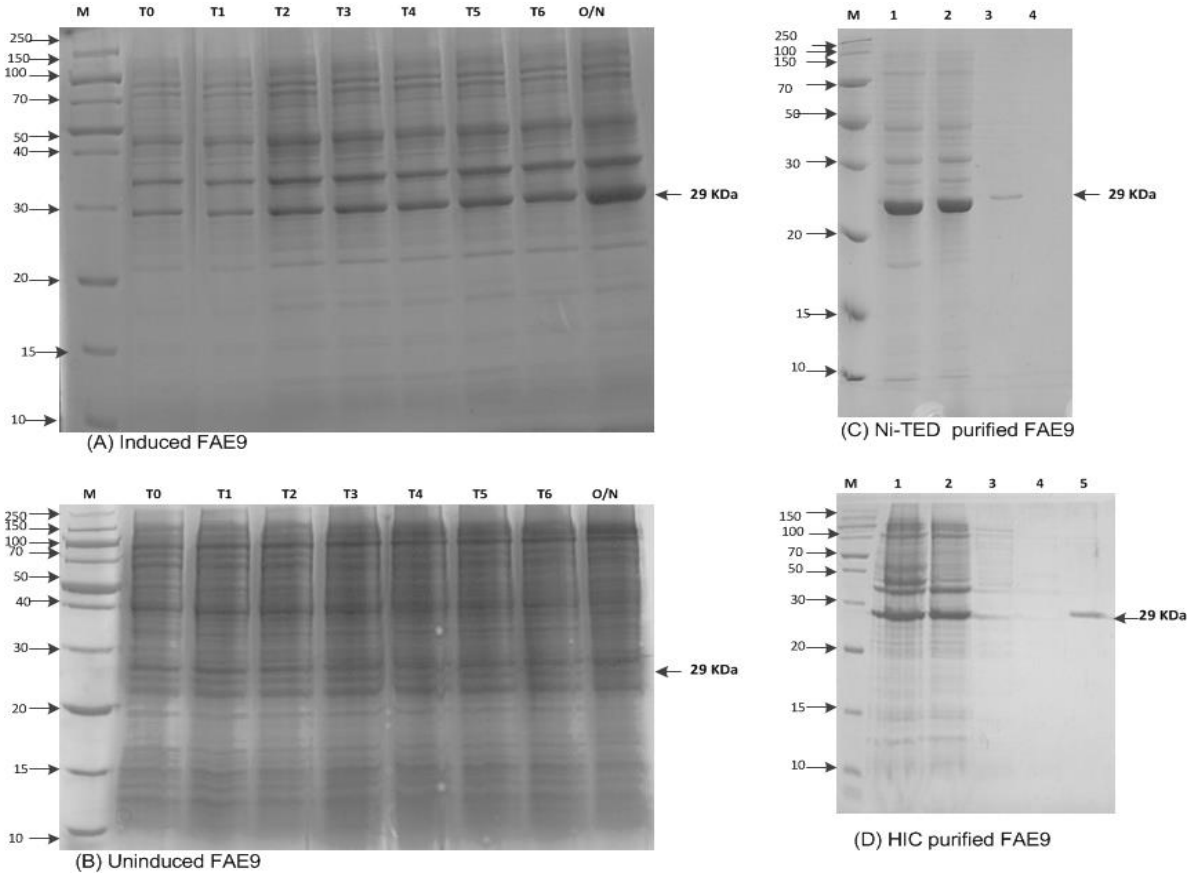


Figure 4.5: SDS-PAGE analysis of FAE9 with the protein band corresponding to a size of 29 kDa as indicated.

(A) Induced FAE9; Lane 1: protein molecular weight marker; Lane T_0 -T_6: 0.1 mM IPTG induced FAE9. (B) Uninduced FAE9; Lane 1: protein molecular weight marker; Lane T_0 -T_6: uninduced FAE9. (C) Ni-TED purified FAE9; Lane M: protein ladder, Lane 1: crude extract, Lane 2: flow through, Lane 3: wash, Lane 4: eluted protein. (D) HIC purified FAE9; Lane M: protein ladder, Lane 1: crude extract; Lane 2: $(NH_4)_2SO_4$ soluble protein, Lane 3: flow through, Lane 4: wash, Lane 5: HIC purified protein.

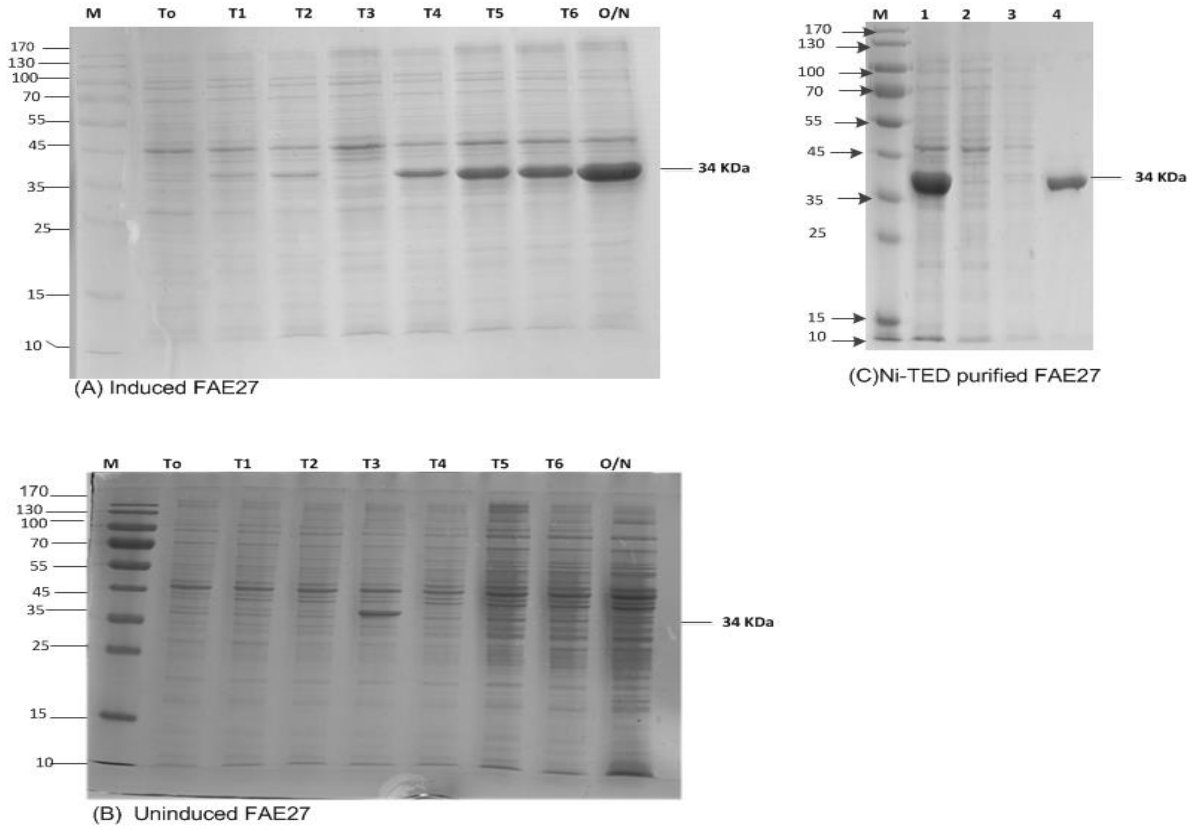


Figure 4.6: SDS-PAGE analysis of FAE27 with the protein band corresponding to a size of 34 kDa as indicated.

(A) Induced FAE27; Lane 1: protein molecular weight marker; Lane T_0 -T_6: 1 mM IPTG induced FAE27. (B) Uninduced FAE27; Lane 1: protein molecular weight marker; Lane T_0 -T_6 uninduced FAE27. (C) Ni-TED purified FAE27: Lane M: protein ladder, Lane 1: crude extract, Lane 2: flow through, Lane 3: wash, Lane 4: eluted protein.

Table 4.1 below summarises the purification steps of FAE9 and FAE27. The pellet of FAE9, after 8% $(NH_4)_2SO_4$ saturation, was then subjected to hydrophobic interaction chromatography. The purification fold increased from 1 to 1.44 and there was an apparent yield of 37%. This yield signifies the removal of unwanted proteins. FAE27 purification fold increased from 1 to 1.5 and the yield was 37.

Table 4.1: Purification table of FAE9 and FAE27

Protein	Purification method	Fraction	Volume	Total protein (mg)[b]	Total activity (Units)[c]	Specific activity ($U.mg^{-1}$)	% Yield	Purification fold
FAE9	Hydrophobic interaction (HIC)	Crude[a]	8	22.80	676 ±1.36	29.65 ±0.48	100	1
		$(NH_4)_2SO_4$	10	15.37	554 ±2.31	36.04 ±1.49	81	1.22
		HIC	38	5.44	183 ±0.15	33.76 ±1.02	27	1.14
FAE27	Affinity(IMAC)	Crude	10	15.05	1951 ±4.83	129.67 ±3.21	100	1
		Elute	10	3.58	728 ±1.18	203.06 ±3.29	37	1.57

a-From 800 mg and 910 mg of FAE9 and FAE27, respectively of wet weight *E. coli* cell pellet (from 50mL of bacterial culture).

b-Protein concentration determined by Bradford assay using BSA as a standard protein.

c-Enzyme activity measured by para-nitrophenol assay.

d-Crude extract B-PER solubilized (precipitates at 8% saturated ammonium sulphate for FAE9).

4.2.2. Size exclusion and determination of globular size

Further separation of FAE9 and FAE27 by SEC was performed. The homogeneity of the purified protein was analysed on a SDS-PAGE. The chromatogram shown in Figure 4.7 revealed several peaks of the eluent. The fractions of each peak observed were pooled together and analysed on a SDS-PAGE, the protein band corresponding to FAE9 was observed at peak 2 (Figure 4.8), while other peaks did not contain any relevant band. The elution profile of FAE27 after the SEC purification step showed a single band of FAE27 obtained from peak 1, after all the pooled fractions of peak 1 were analysed on a SDS-PAGE (Figure 4.10). The chromatogram of FAE27 SEC is shown in Figure 4.9. Both FAE9 and FAE27 appeared at 50-60 mL elution volume.

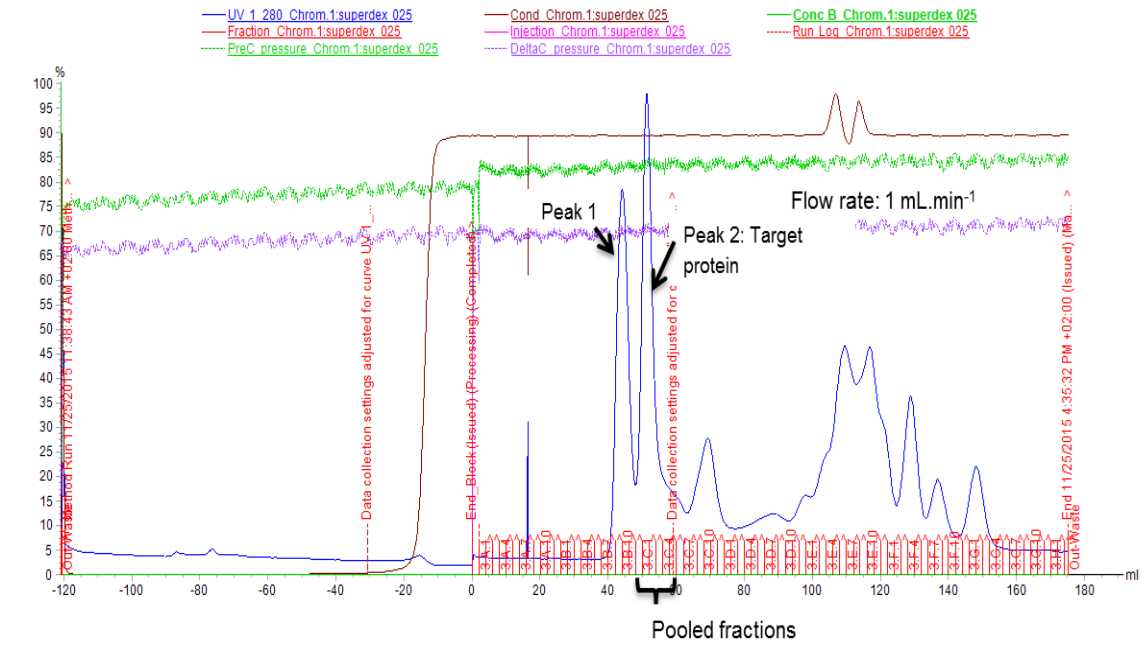


Figure 4.7: Size exclusion chromatogram profile of FAE9.

FAE9 eluent fraction obtained from the $(NH_4)_2SO_4$ step was loaded onto the size exclusion column. Fraction numbers are displayed in red; volumes are in black (mL). The blue line refers to the protein absorbance; brown line is the conductivity of the buffer.

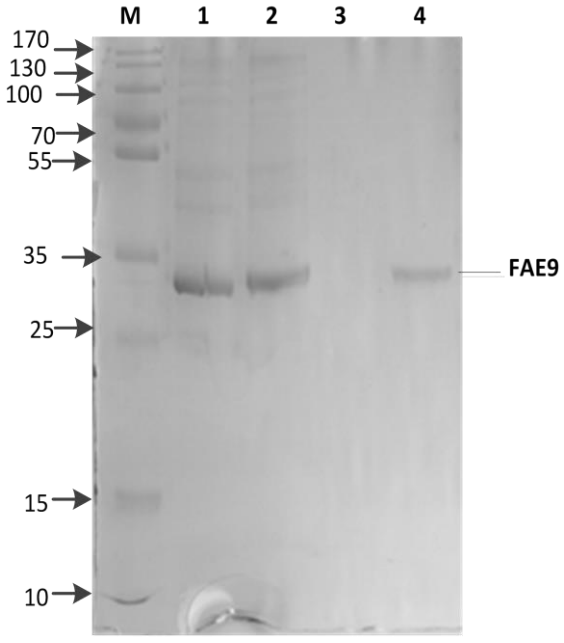


Figure 4.8: SDS-PAGE analysis of the SEC purified FAE9

SDS-PAGE analysis of the protein peaks achieved from the SEC purification. Lane M: protein molecular weight marker; Lane 1: crude extract; Lane 2: $(NH_4)_2SO_4$ soluble protein; Lane 3: peak 1 fractions; Lane 2: peak 2 fractions containing target protein.

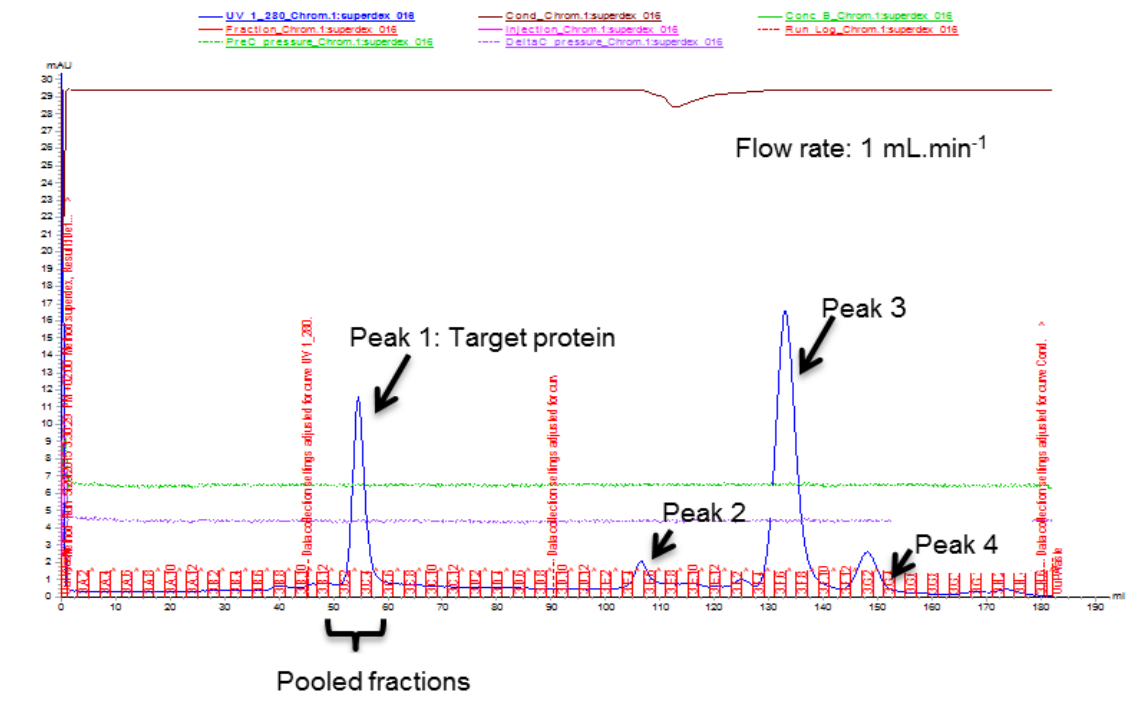


Figure 4.9: Size exclusion chromatogram profile of FAE27.

FAE27 eluent fraction obtained from the Ni-TED step was loaded onto the size exclusion column. Fraction numbers are displayed in red; volumes are in black (mL). The blue line refers to the protein absorbance; brown line is the conductivity of the buffer.

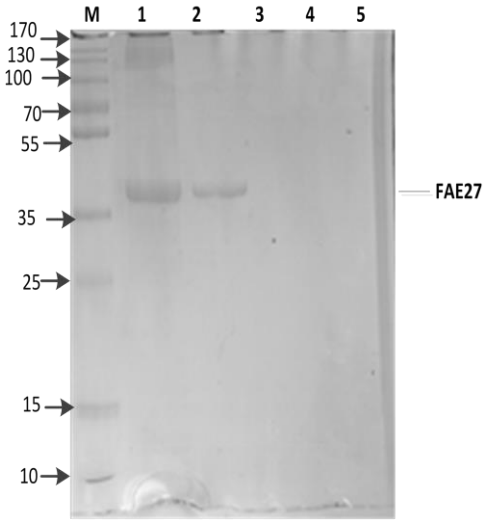


Figure 4.10: SDS-PAGE analysis of the SEC purified FAE27

SDS-PAGE analysis of the protein peaks achieved from the SEC purification. Lane M: protein molecular weight marker, Lane 1: crude extract; Lane 2: peak 1 fractions containing target protein; Lane 3: peak 2 fractions; Lane 4: peak 3 fractions; Lane 5: peak 4 fractions.

The globular sizes of FAE9 and FAE27 were deduced from the standard curve constructed using proteins of known structures and sizes. As previously identified, FAE9 and FAE27 enzymes had subunit molecular mass of 29 kDa and 34 kDa, respectively, under denaturing conditions. This was observed by a single band in SDS-PAGE. A protein molecular mass of 125 and 144 kDa under non-denaturing conditions suggests that both FAE9 and FAE27 may exist as a homotetramer. A schematic graph for FAE9 and FAE27 that relates the log molecular weights of FAE9 and FAE27 to the protein standards is shown in Figure 4.11.

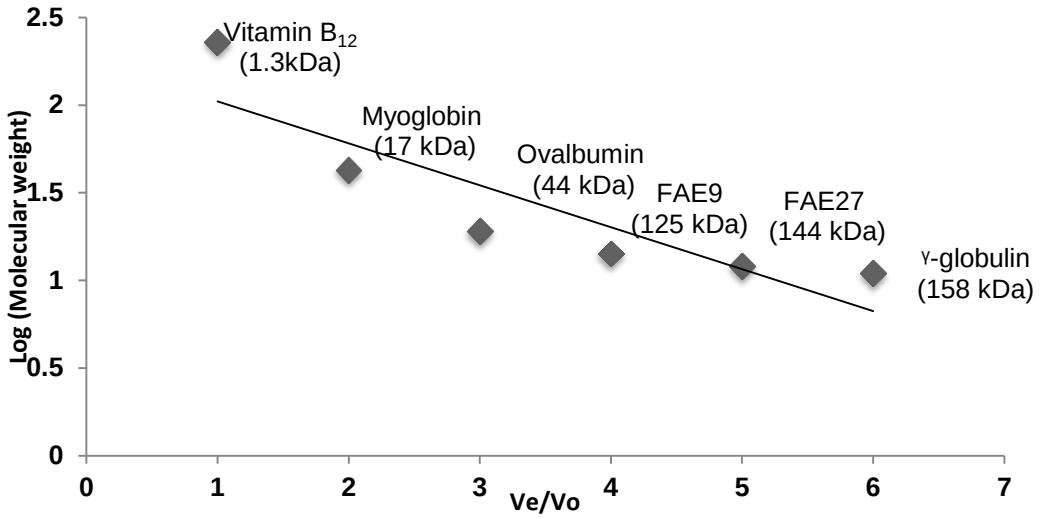


Figure 4.11: Globular molecular weights of FAE9 and FAE27 together with the protein standards.

4.2.3. Effect of temperature and pH on enzyme activity

The effect of temperature and pH on FAE9 and FAE27 activities was determined at various temperatures and pH, respectively, as shown in Figure 4.12 below. Activities of the two enzymes were presented as relative activity with the highest activity taken as 100%. The effect of temperature on enzyme activity was tested at 10 °C increments and the effect of pH at 1 pH unit increments. Activity was assayed for a 5 minutes period. The highest activity for FAE9 was at 50 °C, while for FAE27 was at 40 °C. There was a steady linear increase in activity from 20 °C up to the temperature where highest activity was observed for both enzymes. This shows that there is reasonable activity at low temperatures. However, activity drops drastically above both enzymes' highest point of activity-90% to less than 20%.

The highest activity of the purified FAE9 was at pH 6.0 and the enzyme showed more than 90% of maximum activity at the pH range of 6.0-7.0. FAE27 exhibited maximal activity in slightly acidic to neutral pH, retaining 100% activity at pH 6.5 and 9. Activity dropped drastically from above 80% to 20% between pH 9 and pH10.

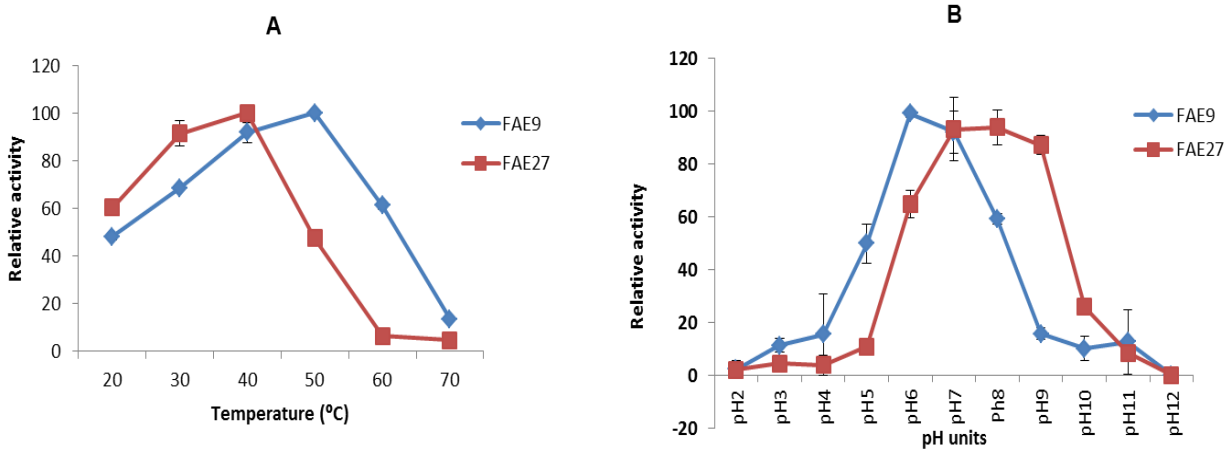


Figure 4.12: The effect of temperature (A) and pH (B) of purified FAE9 and FAE27 determined by measuring enzyme activity using methyl ferulate as a substrate. The highest activity obtained was defined as 100%.

4.2.4. Esterase activity with p-nitrophenyl substrates

Based on substrate preference, enzymes that attack ester bonds may be classified either as a lipase or esterase, based on their preference for carbon chain length. Substrate specificities of FAE9 and FAE27 against short to long chain pNP-esters (C2-C16) were investigated (Figures 4.13 A and B). FAE9 and FAE27 displayed maximum activity (100%) with pNP butyrate (C4) and pNP caprylate (C8), respectively. The activities of the two enzymes decreased with an increase in the chain length of the substrate (Figure 4.13).

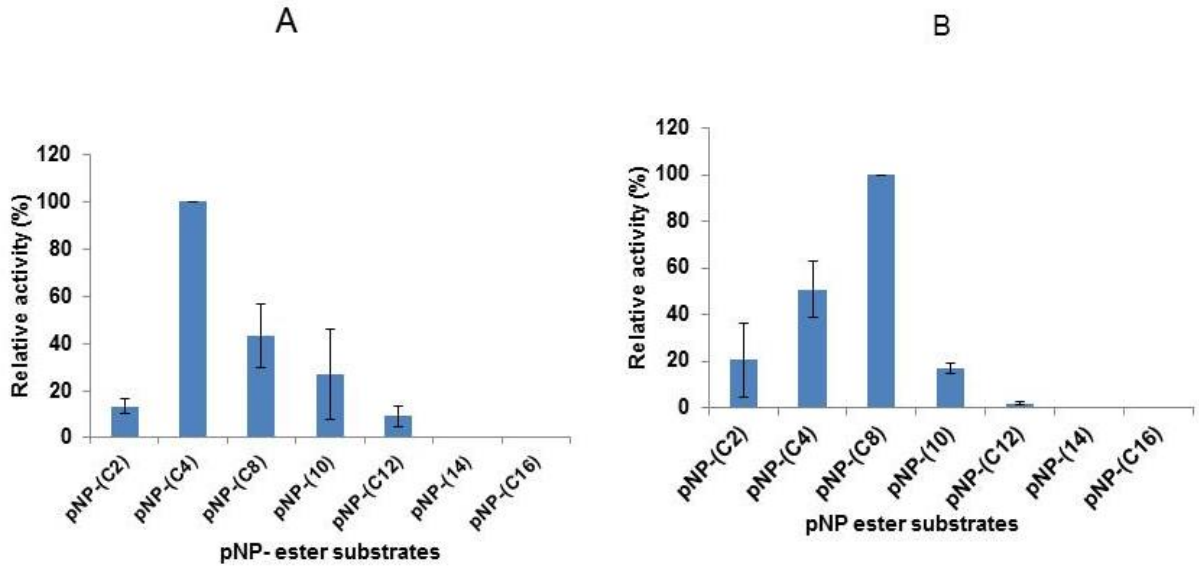


Figure 4.13: Substrate specificity of the purified (A) FAE9 and (B) FAE27 enzymes toward *p*-nitrophenyl esters of varying chain lengths. Relative activity was shown as the percentage of the activity compared to the substrate with the highest activity.

4.2.5. Substrate specificity towards hydroxycinnamic acids and kinetics

Substrate specificities of FAE9 and FAE27 towards hydroxycinnamic acids were determined by measuring enzyme activity against methyl cinnamate (MCN), methyl ferulate (MFA), ethyl ferulate (EFA), methyl caffeate (MCA), methyl sinapate (MSA), methyl p-coumarate (M*p*CA), methyl gallate (MGA) and chlorogenic acid (CGA) as substrate (Figure 4.14 A and B). Both enzymes showed highest activity (100%) against MCN, followed by MFA and EFA. Furthermore, FAE27 also showed reasonable activity against MCA (54 %) and M*p*CA (51%), as opposed to FAE9 which only showed insignificant activity against MCA (4%) and no activity toward M*p*CA. The highest activity of more than 90% was observed against MCN, while MFA and EFA had activity above 60% for both FAE9 and FAE27, when compared to other substrates. There was no significant activity observed on CGA for both enzymes.

Kinetic parameters were determined using the substrates that showed high substrate preference (MCN, MFA, EFA, and MSA for FAE9) and (MCN, MFA, EFA, MCA, and M*p*CA for FAE27). The K_m, $V_{max,}$ k_{cat} and k_{cat}/K_m values were calculated and are shown in Table 4.2 for FAE9 and Table 4.3 for FAE27 below. FAE9 showed high substrate affinity (K_m) against EFA, while FAE27 displayed high affinity for MCA. Catalytic turnover (k_{cat}) of FAE9 and FAE27 was high with MCN and MFA as substrates, respectively. Catalytic efficiency values

(k_{cat}/K_m) showed that the EFA and M*p*CA were the preferred substrates for both FAE9 and FAE27.

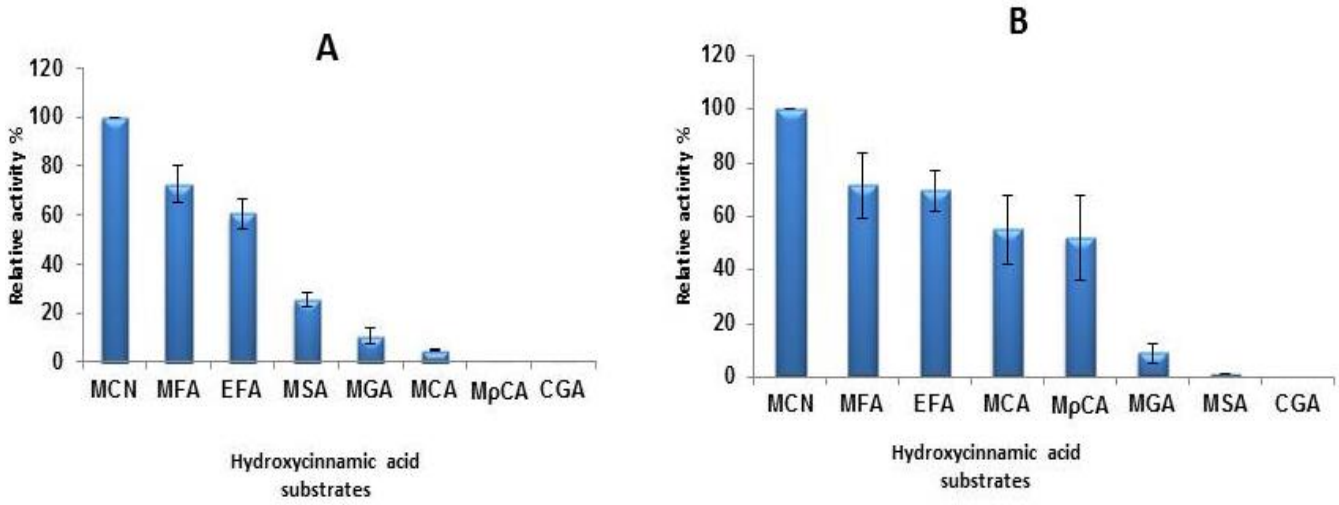


Figure 4.14: Substrate preference of the purified (A) FAE9 and (B) FAE27 enzymes toward hydroxycinnamic acids. Relative activity was shown as the percentage of the activity compared to the substrate with the highest activity.

Table 4.2: Kinetic parameters of FAE9 for various hydroxycinnamic acids

Substrates	V_{max}(µmol.min^{-1}.mg^{-1})	K_m(mM)	k_{cat}(s^{-1})	k_{cat}/K_m(s^{-1}.mM^{-1})
MCN	36.05	0.16	17.66	110.34
MFA	29.87	0.13	14.60	112.31
EFA	28.86	0.11	14.10	128.18
MSA	18.25	0.35	6.10	17.43

Table 4.3: Kinetic parameters of FAE27 against four hydroxycinnamic acids

Substrates	V_{max}(µmol.min^{-1}.mg^{-1})	K_m (mM)	k_{cat}(s^{-1})	k_{cat}/K_m (s^{-1}.mM^{-1})
MCN	24.31	0.21	13.97	66.52
MFA	36.79	0.53	21.13	39.87
EFA	35.11	0.15	20.58	137.20
MCA	3.84	0.07	1.25	17.86
M*p*CA	27.76	0.19	15.95	83.95

4.3. DISCUSSION

Two FAEs were successfully over-expressed in *E. coli* BL21 (DE3) as an expression host. In this study FAE9 and FAE27 over-expression was observed after 3 hours of induction with IPTG. A protein band migrating at ~29 kDa and 34 kDa for FAE9 and FAE27, respectively, was present in cells which had been induced with IPTG (Figure 4.5 A and Figure 4.6 A). Ferulic acid esterases from different sources have molecular weights that range from 11-210 kDa (Fazary and Ju, 2007). Ferulic acid esterases from *Aspergillus niger* were also reported to have molecular weights that range from 29-132 kDa (Topakas *et al*., 2012). FAEs have significant differences in biochemical and biophysical characteristics. After purification using nickel affinity chromatography was attempted, SDS-PAGE of the fractions of FAE27 eluted from the column, showed no additional protein band suggesting that the protein was fairly purified by this method (Figure 4.6 C). FAE9 purification was not successful with this method (Figure 4.5 C). Therefore, it was sub-cloned into pET30b and the 6× His-tag was moved from C-terminal to the N-terminal using *Nde*I and *Xho*I restriction sites but still not successful, therefore, we used classical purification methods. FAE9 and FAE27 were successfully purified to near homogeneity with specific activities of 33 ($U.mg^{-1}$) and 203 ($U.mg^{-1}$), respectively. Percentage recoveries of 27% and 37%, and purification fold of 1.12 and 1.57 were observed for FAE9 and FAE27, respectively (Table 4.1).

SDS-PAGE analysis of FAE9 and FAE27 gave a single band corresponding to a protein band of 29 kDa and 34 kDa, respectively. In addition, size exclusion chromatography (Figure 4.7 and Figure 4.9) resulted in elution of these proteins that corresponds to the estimated globular masses of FAE9 and FAE27 to be 125 kDa and 144 kDa, respectively. This suggests that both proteins exist as homotetramers. In contrast, Asther *et al*. (2005) reported a chlorogenic acid hydrolase dimer with the globular molecular mass of 170 kDa. Most FAEs have been reported as dimers, the multimeric protein structures of FAE9 and FAE27 could bring some more information on the structures of ferulic acid esterases.

Most FAEs show maximum activity at temperatures and pH between 30 and 65 °C, and 3.0 and 9.5, respectively (Fazary and Ju, 2007). In this study, FAE9 and FAE27 showed highest activity at 50 °C and 40 °C, respectively (Figure 4.12 A). Temperature parameters of an enzyme are critical as they determine the industrial capability of an enzyme. At temperatures above optimal temperature an enzyme undergoes denaturation, thus executing inefficient catalysis (Faiz *et al*., 2007). Fae34 isolated from termite hindgut also had optimal activity at 50 ^{o}C (Rashamuse *et al*., 2009). Many FAEs reported to date are active at optimal pH

between 4.5 and 6.5, However, some have a broad optimum pH range of up to pH 9. FAE9 showed an optimum pH of 6.5 while FAE27 showed a pH optimum of pH 6 to 8 (Figure 4.12B). EstF27 isolated from a soil metagenomic library also displayed a broad pH range and optimum activity at a pH of 6.5 (Sang *et al*., 2011). More than 70% of FAE27 activity was observed under alkaline pH up to pH 9 and low activity at pH <5, indicating that the FAE27 is stable in both alkaline and acidic conditions. Optimal activity under acidic and alkaline conditions is advantageous in many biotechnological applications. A broad pH optimum range has also been observed for other family VII carboxyl esterases (Kim *et al.*, 2008).

Esterases and lipases belong to the same family of serine hydrolases. They both hydrolyse ester bonds producing alcohol and carboxylic acid. But they are differentiated by their substrate preference. Lipases prefer water insoluble substrates such as triglycerides with long chain fatty acids, high substrate hydrophobicity, high solvent stability and generally have high enantioselectivity and interfacial activation. The two proteins strongly hydrolysed the ester bonds of pNP butyrate (C4) and pNP caprylate (C8), but were not able to strongly hydrolyse long chain fatty acids (Figure 4.13), thus proving them to be esterases. Hydrolysis of short chain fatty acids by FAEs was also reported in the study conducted by Rashamuse *et al*. (2014). Esterases prefer triglycerides with short chain fatty acids, usually shorter than C6, do not have interfacial activation and have high to low substrate hydrophobicity, high to low or no enantioselectivity (Bornscheuer, 2002).

Two enzymes were tested for substrate preference on hydroxycinnamic acids. Both enzymes displayed more than 90% relative activity towards MCN and more than 80% relative activity towards MFA and EFA (Figure 4.14). The activities of both FAE9 and FAE27 against methyl caffeate were higher than that of chlorogenic acid. Chlorogenic acid is a substrate that is made up of caffeic acid that is ester-linked to quinic acid and the metabolism of chlorogenic acid occurs via hydrolysis to caffeic and quinic acids. Therefore, these results suggest that the presence of quinic acid has an effect in substrate recognition and inhibits the positioning of the substrate ester bond in the FAE9 and FAE27 active sites essential for optimum activity (Kroon *et al*., 1997). There was no activity against chlorogenic acid from both enzymes. Apart from an influence by quinic acid, chlorogenic acid chemical structure consist of two benzene rings, while other substrates have one benzene ring. In the study conducted by Williamson *et al*. (1998), it was reported that substrate accessibility was governed by the substitutions to allow the association of the substrate ester bond at their

respective active sites. These results show that there is a distance necessity between the aromatic ring and the ester bond of the substrate to allow hydrolysis by these esterases.

FAE9 exhibited close to 30% relative activity while FAE27 exhibited less than 2-5% relative activity detected against MSA. FAEs from *Aspergillus niger* revealed reduced substrate preference against MSA (Pa and Williamson, 1996). FAE9 showed high preference to MCN, MFA, EFA and MSA which might suggest that it belongs to the Type A ferulic acid esterases. FAE27 showed broad substrate preference to MCN, MFA, EFA, MCA and M*p*CA. This hydrolytic pattern against substrates is similar to that observed for type C ferulic acid esterases (Liu *et al*., 2001).

The K_m value (0.16 mM) of FAE9 for methyl cinnamate is slightly lower than that of FAE27 (0.21 mM) for the same substrate, suggesting that they have same substrate affinity. However, the catalytic turnover number of FAE9 (k_{cat}=17.66) is higher than that of FAE27 (k_{cat}=13.97). The K_m value increases when the enzyme specificity is reduced due to the presence of inhibitors. The substrate analogue binds to the free enzyme or the enzyme substrate complex resulting in catalytically inactive enzyme substrate inhibitor complex. As a result, both specificity and catalytic properties of the enzyme are modified resulting in higher K_m and a decrease of the catalytic rate. The catalytic efficiency (k_{cat}/K_m) values (128 and 137$s^{-1}.mM^{-1}$) of FAE9 and FAE27 for EFA, respectively, are higher than all other substrates tested in this study, indicating that EFA was indeed the best substrate for both enzymes. The kinetic parameters of FAE9 and FAE27 are shown in Table 4.2 and 4.3.

5. GENERAL CONCLUSIONS

5.1. Research study in perspective and conclusions

The aim of this research project was to bio-prospect the metagenome obtained from the soil associated with cattle manure, and discover possible new genes that might code for carbohydrate active esterases. Discovery of carbohydrate active enzymes is essential as they play a major role in the degradation of lignocellulosic material found in plant cell walls. Lignocellulosic material from different natural sources could be a rich source for renewable energy due to the sugars trapped in the cell walls. Plant cell wall material is recalcitrant to degradation and it requires the action of multiple enzymes to be hydrolysed completely. Soil is richer in microbial diversity than any other ecosystem and the majority of microorganisms from soil are not well categorized due to the inability to culture them in cultivation media. The soil ecological unit is an attractive reservoir for the discovery of novel enzymes. Bio-prospecting a soil metagenome is the great platform to overcome the inability to culture microorganisms and search for novel enzymes.

Due to the importance of finding novel enzymes, the soil associated with cattle manure was used in this study to search for carbohydrate active esterases. Esterases are hydrolytic enzymes that hydrolyse the ester bonds with the release of water. Esterases are divided into 31 subgroups based on the ester bond cleavage specificity. There are carboxylic ester hydrolases that target carboxylic esters and there are thiolester hydrolases that target the thiolester bonds. Ferulic acid esterases belong to the carboxylic ester class and they hydrolyse ester bonds found in phenolic compounds into phenolic acids and respective alcohols. Plant cell walls consist of inter-linked phenolic compounds between the cell wall polysaccharides and lignin. Ferulic acid esterases hydrolyse the ester bonds found in hemicellulose into its sugar components.

In this study, total metagenomic DNA from soil was successfully extracted and due to the quality of extracted DNA, a metagenomic library was successfully constructed using the fosmid based metagenomic library construction system. Functional based screening was adopted to screen for esterases from the metagenomic library and also to subsequently screen for ferulic acid esterases using agar based screening. A total of seven clones (FAE3, FAE14, FAE27, FAE71, FAE83, and FAE91) containing *fae* encoding ORFs were discovered in this study. Sequence analysis of primary structures showed a (49-57 %) identity/similarity of FAEs discovered in this study with proteins from the NCBI-Blast-P

database. This percentage identity shows that these enzymes are novel and unique when compared to the ferulic esterases identified to date. Analysis of a classical esterase-lipase serine motif revealed that FAEs discovered in this study belong to the esterase-lipase family with the GxSxG signature motif. However, FAE3 and FAE14 had no known classical serine motif although they exhibit FAE activity. This proves that these enzymes are novel and that they might constitute a new type of esterase family that has not yet been identified.

Heterologous gene expression was used to express all discovered FAEs using a pET28a expression vector and *E. coli* strain. The histidine tagged recombinants were over-expressed under the control of the T7 expression promoter and IPTG at varying temperatures. Only two FAEs (FAE9 and FAE27) successfully expressed and the others did not show expression of protein, even under different optimisations. However, the application of IMAC affinity purification did not support FAE9 purification by this method. It was suspected that perhaps the His tag was folded inside and thus was not exposed to bind to the IMAC resin. The His-tag was therefore moved from the C-terminal to the N-terminal and pET30b was used as the expression vector. The putative FAE9 was then purified using ammonium sulphate precipitation and hydrophobic interaction chromatography.

FAE9 and FAE27 were reasonably pure and were biochemically characterised for temperature and pH optima which revealed highest activity at acidic to neutral pH and at 40-50 °C optimal temperature. The application of FAE9 and FAE27 was demonstrated by the analysis of the activity of these two enzymes on hydroxycinnamic acid synthetic substrates. The two proteins displayed high substrate specificity towards methyl cinnamate and methyl ferulate.

In conclusion, the outcome of this work proves soil is a valuable source for screening of novel enzymes through functional a based-metagenomic approach. Therefore, this method may not only be employed for screening esterases from soil it can also be used to screen for other carbohydrate active enzymes that are involved in the degradation of plant cell wall. The biochemical properties, including broad substrate specificity displayed by the two isolated enzymes, make these enzymes useful in industrial applications.

5.2. Future work

5.2.1. Investigation into the release of phenolic acids from natural substrates.

Plant cell walls consist of a complex network of polysaccharides that are linked to lignin polymers and in nature; their degradation requires the synergistic action of glycoside hydrolases, carbohydrate esterases, and polysaccharide lyases. Ferulic acids are involved in the ester/ether linkages of lignin components. Ferulic acid esterases have been identified and characterised for their ability to hydrolyse natural and synthetic substrates. There has been an increase in the demand of natural products; therefore obtaining enzymes that are capable of producing phenolic acids from natural compounds is significant. In this study, FAE9 and FAE27 were only tested against synthetic hydroxycinnamic acid substrates. Future work may include the investigation of enzymes for their ability to release phenolic acids from natural compounds which may include agro-industrial wastes and wheat bran.

5.2.2. Regulatory machinery of ferulic acid esterase gene expression

Since other FAEs isolated in this study were not successfully expressed, a more in-depth investigation is needed to understand the regulatory factors involved in ferulic acid esterase gene expression. Additional work is also required to understand the active site and catalytic mechanism of FAE3 and FAE14 isolated in this study, as there was no known serine motif detected, although the encoding ORFs blasted them as enzymes belonging to the esterase lipase family.

6. REFERENCES

Akin, D. E. (2008). Plant cell wall aromatics: influence on degradation of biomass: *Biofuels*, *Bioproducts and Biorefining*, 2, pp. 288-303.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs: *Nucleic Acids Research*, 25, pp. 3389-3402.

Asther, M., M. I. E. Alvarado, M. Haon, D. Navarro, M. Asther, L. Lesage-Meessen, and E. Record. (2005). Purification and characterization of a chlorogenic acid hydrolase from *Aspergillus niger* catalysing the hydrolysis of chlorogenic acid: *Journal of Biotechnology*, 115, pp. 47-56.

Berlemont, R., D. Pipers, M. Delsaute, F. Angiono, G. Feller, M. Galleni, and P. Power. (2011). Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements: *Revista Argentina de microbiologia*, 43, pp. 94-103.

Bornscheuer, U. T. (2002). Microbial carboxyl esterases: classification, properties and application in biocatalysis: *FEMS Microbiology Reviews*, 26, pp. 73-81.

Béjà, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, and S. B. Jovanovich. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage: *Environmental Microbiology*, 2, pp. 516-529.

Campbell, M. M., and R. R. Sederoff. (1996). Variation in Lignin Content and Composition (Mechanisms of Control and Implications for the Genetic Improvement of Plants): *Plant physiology*, 110, pp. 3.

Carpi, G., F. Cagnacci, N. E. Wittekindt, F. Zhao, J. Qi, L. P. Tomsho, D. I. Drautz, A. Rizzoli, and S. C. Schuster. (2011). Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks: *PloS One*, 6, pp. e25604.

Casali, N. (2003). *Escherichia coli* host strains, *E. coli* Plasmid Vectors, *Springer*, pp. 27-48.

Chen, R. (2012). Bacterial expression systems for recombinant protein production: *E. coli* and beyond: *Biotechnology Advances*, 30, pp. 1102-1107.

Cottrell, M. T., J. A. Moore, and D. L. Kirchman. (1999). Chitinases from uncultured marine microorganisms: *Applied and Environmental Microbiology*, 65, pp. 2553-2557.

Crepin, V. F., C. B. Faulds, and I. F. Connerton. (2003). Production and characterization of the *Talaromyces stipitatus* feruloyl esterase FAEC in *Pichia pastoris*: identification of the nucleophilic serine: *Protein Expression and Purification*, 29, pp. 176-184.

Crepin, V. F., C. B. Faulds, and I. F. Connerton. (2004). Functional classification of the microbial feruloyl esterases: *Applied Microbiology and Biotechnology*, 63, pp. 647-652.

Culligan, E. P., R. D. Sleator, J. R. Marchesi, and C. Hill. (2014). Metagenomics and novel gene discovery: Promise and potential for novel therapeutics: *Virulence*, 5, pp. 399-412.

Daniel, R. (2005). The metagenomics of soil: *Nature Reviews Microbiology*, 3, pp. 470-478.

de Castro, A. P., B. F. Quirino, H. Allen, L. L. Williamson, J. Handelsman, and R. H. Krüger. (2011). Construction and validation of two metagenomic DNA libraries from Cerrado soil with high clay content: *Biotechnology Letters*, 33, pp. 2169-2175.

de Souza, W. R. (2013). Microbial Degradation of Lignocellulosic Biomass: Sustainable Degradation of Lignocellulosic Biomass-Techniques, A*pplications and Commercialization*, 9, pp. 207-247.

de Vries, R. P., and J. Visser. (2001). *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides: *Microbiology and Molecular Biology Review*s, 65, pp. 497-522.

Demain, A. L., and P. Vaishnav. (2009). Production of recombinant proteins by microbes and higher organisms: *Biotechnology Advances*, 27, pp. 297-306.

Desai, C., K. Jain, and D. Madamwar. (2008). Evaluation of *In vitro* Cr (VI) reduction potential in cytosolic extracts of three indigenous *Bacillus* sp. isolated from Cr (VI) polluted industrial landfill: *Bioresource Technology*, 99, pp. 6059-6069.

Dodson, G., and A. Wlodawer. (1998). Catalytic triads and their relatives: *Trends in Biochemical Sciences,* 23, pp. 347-352.

Elshahed, M. S., N. H. Youssef, A. M. Spain, C. Sheik, F. Z. Najar, L. O. Sukharnikov, B. A. Roe, J. P. Davis, P. D. Schloss, and V. L. Bailey. (2008). Novelty and uniqueness patterns of rare members of the soil biosphere: *Applied and Environmental Microbiology*, 74, pp. 5422-5428.

Esposito, D., and D. K. Chatterjee. (2006). Enhancement of soluble protein expression through the use of fusion tags: *Current Opinion in Biotechnology*, 17, pp. 353-358.

Fahmy, A. S., A. Z. Abo-Zeid, T. M. Mohamed, H. M. Ghanem, I. H. Borai, and S. A. Mohamed. (2008). Characterization of esterases from *Cucurbita pepo* cv.“Eskandrani”: *Bioresource Technology*, 99, pp. 437-443.

Faiz, O., A. Colak, N. Saglam, S. Canakci, and A. O. Belduz. (2007). Determination and characterization of thermostable esterolytic activity from a novel thermophilic bacterium *Anoxybacillus gonensis* A4: J*ournal of Biochemistry and Molecular Biology*, 40, pp. 588-594.

Fan, X., X. Liu, R. Huang, and Y. Liu. (2012). Identification and characterization of a novel thermostable pyrethroid-hydrolyzing enzyme isolated through metagenomic approach: *Microbial Cell Factories*, 11, pp. 33.

Faulds, C. B., and G. Williamson. (1991). The purification and characterization of 4-hydroxy 3-methoxycinnamic (ferulic) acid esterase from *Streptomyces olivochromogenes*: *Journal of General Microbiology*, 137, pp. 2339-2345.

Faulds, C. B., and G. Williamson. (1995). Release of ferulic acid from wheat bran by a ferulic acid esterase (FAE-III) from *Aspergillus niger*: *Applied Microbiology and Biotechnology*, 43, pp. 1082-1087.

Fazary, A. E., and Y. H. Ju. (2007). Feruloyl esterases as biotechnological tools: current and future perspectives: *Acta Biochimica et Biophysica Sinica*, 39, pp. 811-828.

Fazary, A. E., and Y. H. Ju. (2008). The large-scale use of feruloyl esterases in industry: *Biotechnology and Molecular Biology Reviews*, 3, pp. 95-110.

Ferrer, M., A. Beloqui, K. N. Timmis, and P. N. Golyshin. (2008). Metagenomics for mining new genetic resources of microbial communities: *Journal of Molecular Microbiology and Miotechnology*, 16, pp. 109-123.

Gans, J., M. Wolinsky, and J. Dunbar. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil: *Science*, 309, pp. 1387-1390.

Gilbert, H. J. (2010). The biochemistry and structural biology of plant cell wall deconstruction: *Plant Physiology*, 153, pp. 444-455.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT: *Nucleic Acids Symposium Series,* 41, pp. 95-98.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, 4, pp. 669-685.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products: *Chemistry and Biology,* 5, pp. 245-249.

Harmsen, P., W. Huijgen, L. Bermudez, and R. Bakker. (2010). Literature review of physical and chemical pretreatment processes for lignocellulosic biomass, 1st ed. Wageningen UR*, Food and Biobased Research,* pp.1-49.

He, J., Z. Xu, and J. Hughes. (2005). Pre-lysis washing improves DNA extraction from a forest soil: *Soil Biology and Biochemistry*, 37, pp. 2337-2341.

Henke, E., J. Pleiss, and U. T. Bornscheuer. (2002). Activity of lipases and esterases towards tertiary alcohols: insights into structure–function relationships: *Angewandte Chemie International Edition*, 41, pp. 3211-3213.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities: *Biochemical Journals*, 280, pp. 309-316.

Himmel, M. E., Q. Xu, Y. Luo, S.Y. Ding, R. Lamed, and E. A. Bayer. (2010). Microbial enzyme systems for biomass conversion: emerging paradigms: *Biofuels*, 1, pp. 323-341.

Hjort, K., M. Bergström, M. F. Adesina, J. K. Jansson, K. Smalla, and S. Sjöling. (2010). Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil: *FEMS Microbiology Ecology*, 71, pp. 197-207.

Ikram-ul-Haq, M. M. J., and S. Tehmina. (2006). An innovative approach for hyperproduction of cellulolytic and hemicellulolytic enzymes by consortium of *Aspergillus niger* MSK-7 and *Trichoderma viride* MSK-10: *African Journal of Biotechnology*, 5, pp. 609-614.

Jeon, J. H., H. S. Lee, J. T. Kim, S. J. Kim, S. H. Choi, S. G. Kang, and J. H. Lee. (2012). Identification of a new subfamily of salt-tolerant esterases from a metagenomic library of tidal flat sediment: *Applied Microbiology and Biotechnology*, 93, pp. 623-631.

Jonasson, P., S. Liljeqvist, P. A. Nygren, and S. Ståhl. (2002). Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*: *Biotechnology and Applied Biochemistry*, 35, pp. 91-105.

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

Kakirde, K. S., L. C. Parsley, and M. R. Liles. (2010). Size does matter: application-driven approaches for soil metagenomics: *Soil Biology and Biochemistry*, 42, pp. 1911-1923.

Khandeparkar, R. D. S., and N. B. Bhosle. (2006). Isolation, purification and characterization of the xylanase produced by *Arthrobacter* sp. MTCC 5214 when grown in solid-state fermentation: *Enzyme and Microbial Technology*, 39, pp. 732-742.

Kheng, P. P., and I. C. Omar. (2005). Xylanase production by a local fungal isolate, *Aspergillus niger* USM AI 1 via solid state fermentation using palm kernel cake (PKC) as substrate: *Songklanakarin Journal of Science and Technology*, 27, pp. 325-336.

Kim, Y. J., G. S. Choi, S. B. Kim, G. S. Yoon, Y. S. Kim, and Y. W. Ryu. (2006). Screening and characterization of a novel esterase from a metagenomic library: *Protein Expression and Purification*, 45, pp. 315-323.

Kim, S. J., C. M. Lee, B. R. Han, M. Y. Kim, Y. S. Yeo, S. H. Yoon, B. S. Koo, and H. K. Jun. (2008). Characterization of a gene encoding cellulase from uncultured soil bacteria: *FEMS microbiology letters,* 282, pp. 44-51.

Kirk, O., T. V. Borchert, and C. C. Fuglsang. (2002). Industrial enzyme applications: *Current Opinion in Biotechnology*, 13, pp. 345-351.

Kroon, P. A., C. B. Faulds, C. Brezillon and G. Williamson. (1997). Methyl p/lenylalkanoales as substrates to probe the active sites of esterases: *European Journal of Biochemistry*, 248, pp. 245-251

Kroon, P. A., M. T. Garcia-Conesa, I. J. Fillingham, G. P. Hazlewood, and G. Williamson. (1999). Release of ferulic acid dehydrodimers from plant cell walls by feruloyl esterases: *Journal of the Science of Food and Agriculture*, 79, pp. 428-434.

Kuhad, R. C., R. Gupta, and A. Singh. (2011). Microbial cellulases and their industrial applications: *Enzyme Research,* 2011.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4: *Nature*, 227, pp. 680-685.

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

LaVallie, E. R., and J. M. McCoy. (1995). Gene fusion expression systems in *Escherichia coli: Current Opinion in Biotechnology*, 6, pp. 501-506.

Lee, D. G., J. H. Jeon, M. K. Jang, N. Y. Kim, J. H. Lee, J. H. Lee, S. J. Kim, G. D. Kim, and S. H. Lee. (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library: *Biotechnology Letters*, 29, pp. 465-472.

Lee, S. W., K. Won, H. K. Lim, J. C. Kim, G. J. Choi, and K. Y. Cho. (2004). Screening for novel lipolytic enzymes from uncultured soil microorganisms: *Applied Microbiology and Biotechnology*, 65, pp. 720-726.

Lepage, P., M. C. Leclerc, M. Joossens, S. Mondot, H. M. Blottière, J. Raes, D. Ehrlich, and J. Doré. (2013). A metagenomic insight into our gut's microbiome: *Gut*, 62, pp. 146-158.

Li, J., S. Cai, Y. Luo, and X. Dong. (2011). Three feruloyl esterases in *cellulosilyticum ruminicola* H1 Act synergistically to hydrolyze esterified polysaccharides: *Applied and Environmental Microbiology*,77, pp. 6141-6147.

Li, L. L., S. R. McCorkle, S. Monchy, S. Taghavi, and D. van der Lelie. (2009). Bioprospecting metagenomes: glycosyl hydrolases for converting biomass: *Biotechnology for Biofuels*, 2.

Li, J. W. H., and J. C. Vederas. (2009). Drug discovery and natural products: end of an era or an endless frontier?: *Science*, 325, pp. 161-165.

Li, G., K. Wang, and Y. H. Liu. (2008). Molecular cloning and characterization of a novel pyrethroid-hydrolyzing esterase originating from the Metagenome: *Microbial Cell Factories*, 7, pp. 38.

Liles, M. R., L. L. Williamson, J. Rodbumrer, V. Torsvik, R. M. Goodman, and J. Handelsman. (2008). Recovery, purification, and cloning of high-molecular-weight

DNA from soil microorganisms: *Applied and Environmental Microbiology*, 74, pp. 3302-3305.

Liu, S. Q., R. Holland, and V. L. Crow. (2001). Purification and properties of intracellular esterases from *Streptococcus thermophiles*: *International Dairy Journal*, 11, pp. 27–35.

Lombard, N., E. Prestat, J. D. Van Elsas, and P. Simonet. (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics: *FEMS Microbiology Ecology*, 78, pp. 31-49.

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

Lu, Z., E. A. DiBlasio-Smith, K. L. Grant, N. W. Warne, E. R. LaVallie, L. A. Collins-Racie, M. T. Follettie, M. J. Williamson, and J. M. McCoy. (1996). Histidine Patch Thioredoxins mutant forms of thioredoxin with metal chelating affinity that provide for convenient purifications of thioredoxin fusion proteins: *Journal of Biological Chemistry*, 271, pp. 5059-5065.

López-López, O., M. E. Cerdán, and M. I. G. Siso. (2014). New extremophilic lipases and esterases from metagenomics: *Current Protein and Peptide Science*, 15, pp. 445-455.

Maina, C. V., P. D. Riggs, A. G. Grandea, B. E. Slatko, L. S. Moran, J. A. Tagliamonte, and L. A. McReynolds. (1988). An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein: *Gene*, 74, pp. 365-373.

Mardis, E. R. (2008). Next-generation DNA sequencing methods: A*nnual Review of Genomics and Human Genetics*, 9, pp. 387-402.

Mathur, E. J., G. Toledo, B. D. Green, M. Podar, T. H. Richardson, M. Kulwiec, and H. W. Chang. (2005). A biodiversity-based approach to development of performance

enzymes: Applied metagenomics and directed evolution: *Industrial Biotechnology*, 1, pp. 283-287.

Michelin, M., M. d. L. T. de Moraes, D. P. da Silva, D. S. Ruzene, A. A. Vicente, J. A. Jorge, H. F. Terenzi, and J. A. Teixeira. (2011). Production of xylanolytic enzymes by *Aspergillus terricola* in stirred tank and airlift tower loop bioreactors: *Journal of Industrial Microbiology and Biotechnology*, 38, pp. 1979-1984.

Neelakanta, G., and H. Sultana. (2013). The Use of Metagenomic Approaches to Analyze Changes in Microbial Communities: *Microbiology Insights*, 6, pp. 37-48.

Ollis, D. L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, and J. Schrag. (1992). The α/β hydrolase fold: *Protein Engineering*, 5, pp. 197-211.

Ou, S., and K. C. Kwok. (2004). Ferulic acid: pharmaceutical functions, preparation and applications in foods: *Journal of the Science of Food and Agriculture*, 84, pp. 1261-1269.

Pa, K., and G. Williamson. (1996). Purification and characterization of a novel esterase induced by growth of *Aspergillus niger* on sugar-beet pulp: *Biotechnology and Applied Biochemistry*, 23, pp. 255-262.

Panda, T., and B. S.Gowrishankar. (2005). Production and applications of esterases. *Applied Microbiology and Biotechnology*, 2, pp. 160-169.

Pang, M., N. Abdullah, C. W. Lee, and C. C. Ng. (2008). Isolation of high molecular weight DNA from forest topsoil for metagenomic analysis: *Asia Pacific Journal of Molecular Biology and Biotechnology*, 16, pp. 35-41.

Prakash, T., and T. D. Taylor. (2012). Functional assignment of metagenomic data: challenges and applications: *Briefings in Bioinformatics*, 13, pp. 711-727.

Prates, J. A. M., N. Tarbouriech, S. J. Charnock, C. M. G. A. Fontes, L. S. Ferreira, and G. J. Davie. (2001). The Structure of the Feruloyl Esterase Module of Xylanase 10B from *Clostridium thermocellum* Provides Insights into Substrate Recognition: *Structure*, 9, pp. 1183-1190.

Rabinovich, M. L., M. S. Melnick, and A. V. Bolobova. (2002). The structure and mechanism of action of cellulolytic enzymes: *Biochemistry* (Moscow), 67, pp. 850-871.

Rashamuse, K., V. Magomani, T. Ronneburg, and D. Brady. (2009). A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous β-lactamase activity on nitrocefin: *Applied Microbiology and Biotechnology*, 83, pp. 491-500.

Rashamuse, K., T. Ronneburg, W. Sanyika, K. Mathiba, E. Mmutlane, and D. Brady. (2014). Metagenomic mining of feruloyl esterases from termite enteric flora: *Applied Microbiology and Biotechnology*, 98, pp. 727-737.

Rhee, J. K., D. G. Ahn, Y. G. Kim, and J. W. Oh. (2005). New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library: *Applied and Environmental Microbiology*, 71, pp. 817-825.

Robertson, D. E., J. A. Chaplin, G. DeSantis, M. Podar, M. Madden, E. Chi, T. Richardson, A. Milan, M. Miller, and D. P. Weiner. (2004). Exploring nitrilase sequence space for enantioselective catalysis: *Applied and Environmental Microbiology*, 70, pp. 2429-2436.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, and C. Minor. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms: *Applied and Environmental Microbiology*, 66, pp. 2541-2547.

Rosano, G. L., and E. A. Ceccarelli. (2014). Recombinant protein expression in microbial systems : *Frontiers in Microbiology*, 5, pp. 1-2.

Rumbold, K., P. Biely, M. Mastihubová, M. Gudelj, G. Gübitz, K. H. Robra, and B. A. Prior. (2003). Purification and properties of a feruloyl esterase involved in lignocellulose degradation by *Aureobasidium pullulans*: *Applied and Environmental Microbiology*, 69, pp. 5622-5626.

Sambrook, J., and D. W. Russell. (2001). Molecular Cloning: A Laboratory Manual, 3rd ed. New York: *Cold Spring Harbor Laboratory Press*.

Sang, S. L., G. Li, X. P. Hu, and Y. H. Liu. (2011). Molecular cloning, overexpression and characterization of a novel feruloyl esterase from a soil metagenomic library: *Journal of Molecular Microbiology and Biotechnology*, 20, pp. 196-203.

Schubot, F. D., I. A. Kataeva, D. L. Blum, A. K. Shah, L. G. Ljungdahl, J. P. Rose, and B. C. Wang. (2001). Structural basis for the substrate specificity of the feruloyl esterase domain of the cellulosomal xylanase Z from *Clostridium thermocellum*: *Biochemistry,* 40, pp. 12524-12532.

Showalter, A. M. (1993). Structure and function of plant cell wall proteins: *The Plant Cell,* 5, pp. 9-23.

Simon, C., and R. Daniel. (2009). Achievements and new knowledge unraveled by metagenomic approaches: *Applied Microbiology and Biotechnology*, 85, pp. 265-276.

Singh, B. K. (2010). Exploring microbial diversity for biotechnology: The way forward: *Trends in Biotechnology*, 28, pp. 111-116.

Singh, S. P., K. Sagar, and B. K. Konwar. (2013). Strategy in metagenomic DNA isolation and computational studies of humic acid: *Current Research in Microbiology and Biotechnology*, 1, pp. 9-11.

Smith, D. B., and K. S. Johnson. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase: *Gene*, 67, pp. 31-40.

Smith, M. C., T. C. Furman, T. D. T. Ingolia, and C. Pidgeon. (1988). Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity

chromatography for recombinant proteins: *Journal of Biological Chemistry,* 263, pp. 7211-7215.

Sticklen, M. B. (2008). Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol: *Nature Reviews Genetics*, 9, pp. 433-443.

Sumby, K. M., A. H. Matthews, P. R. Grbin, and V. Jiranek. (2009). Cloning and characterization of an intracellular esterase from the wine-associated lactic acid bacterium *Oenococcus oeni*: *Applied and Environmental Microbiology*,75, pp. 6729-6735.

Sun, J., C. Tian, S. Diamond, and N. L. Glass. (2012). Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*: *Eukaryotic Cell*, 11, pp. 482-493.

Swartz, J. R. (1996). *Escherichia coli* recombinant DNA technology. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. Washington, DC: A*merican Society of Microbiology Press*, pp. 1693-1711.

Sweeney, M. D., and F. Xu. (2012). Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments: *Catalysts*, 2, pp. 244-263.

Tirawongsaroj, P., R. Sriprang, P. Harnpicharnchai, T. Thongaram, V. Champreda, S. Tanapongpipat, K. Pootanakit, and L. Eurwilaichitr. (2008). Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library: *Journal of Biotechnology*, 133, pp. 42-49.

Topakas, E., M. Moukouli, M. Dimarogona, and P. Christakopoulos. (2012). Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila*: *Applied Microbiology and Biotechnology*, 94, pp. 399-411.

Tuffin, M., D. Anderson, C. Heath, and D. A. Cowan. (2009). Metagenomic gene discovery: how far have we moved into novel sequence space: *Biotechnology Journal*, 4, pp. 1671-1683.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment: *Nature*, 428, pp. 37-43.

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

Udatha, D., I. Kouskoumvekaki, L. Olsson, and G. Panagiotou. (2011). The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases: *Biotechnology Advances*, 29, pp. 94-110.

Vafiadi, C., E. Topakas, V. R. Nahmias, C. B. Faulds, and P. Christakopoulos. (2009). Feruloyl esterase-catalysed synthesis of glycerol sinapate using ionic liquids mixtures: *Journal of Biotechnology*, 139, pp. 124-129.

Voget, S., C. Leggewie, A. Uesbeck, C. Raasch, K. E. Jaeger, and W. R. Streit. (2003). Prospecting for novel biocatalysts in a soil metagenome: *Applied and Environmental Microbiology*, 69, pp. 6235-6242.

Warshel, A., G. Naray-Szabo, F. Sussman and J. K. Hwang, J. K. (1989). How do serine proteases really work?: *Biochemistry,* 28, pp. 3629-3637.

Wilkie, K. C. B., and S. L. Woo. (1977). A heteroxylan and hemicellulosic materials from bamboo leaves, and a reconsideration of the general nature of commonly occurring xylans and other hemicelluloses: *Carbohydrate Research*, 57, pp. 145-162.

Williamson, G., C. B. Faulds and P. A. Kroon. (1998). Specificity of ferulic acid (feruloyl) esterases: *Biochemical Society Transactions,* 26, pp. 205-209.

Yu, P., J. J. McKinnon, and D. A. Christensen. (2005). Hydroxycinnamic acids and ferulic acid esterase in relation to biodegradation of complex plant cell walls: *Canadian Journal of Animal Science*, 85, pp. 255-267.

Zaldivar, J., J. Nielsen, and L. Olsson. (2001). Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration: *Applied Microbiology and Biotechnology*, 56, pp. 17-34.

Zhang, X. Z., and Y. H. P. Zhang. (2013). Cellulases: Characteristics, Sources, Production, and Applications: *Bioprocessing Technologies in Biorefinery for Sustainable Production of Fuels, Chemicals and Polymers*, pp. 131-146.

7. APPENDIX

Appendix A: Molecular methods supplementary data

Appendix A1: Summary of BLAST search results involving families, nucleotides sequences and amino acid sequences

Genes	Family	Sequence	Amino acid sequence	Signal peptide	Best Hit
*Fae*3	Esterase lipase family	ATGAAAATGAAAATATTGAAGGTATTCAAAACAATTGCACTTGTTCTATTTG CTATTATAGCAGCACTATTGATTGCGTCTACGATTAATCACCAAACTCTATT AAAAAATGAGTTCAAAGAGTACTATCCATCTGATGGCATGGTTGAAGTAGAT GGTAATAAGTTTCATGTATATTCGGAGGGTGCAGGCGACCTTACCCTTGTAT TTATGTCCGGACATGGTACCAATAGTCCTACATTGGATTTCAAGCCCTTGTG GATGAAGATGGTAGATGACTACAGGATTGCAATAGTAGAAAAATCAGGTTAC GGCTGGAGTGAGACATCAAATAGTCCAAGGGACATAGATACCATTCTTGAAG AAACACGAAAGGCATTGGAACTGTCGGGAGAGAAAGGTCCCTATGTACTTTT CCCTCATTCCATGTCAGGGCTTGAGGCGATATATTGGGCTCAAAAATACCCA GATGAAGTAAAAGCAATTATAGGATTGGATCCATGTACTCCTGAAACTATTA AGATTCTTCCTGAACCTGAAAAAACACAATTGCACCTTATGTATAGTATATC AAGAATGGGAATTACCAGATTTATGCCAGATTTTGACTTTGAAAGTAATTTC CCTTTGATGAAGTCAGAGGACTTGACAGAAGAAAATAAACAAGAATATTTAG CGGTTTTCTATAAAAGTGCTTTTTCAAAAGATATGTTAAGAGAAGTCAATTA TCTATACGATAATGCCAAAACAGTAGCAGAAAACGAAGTTCCTATCCACACC CCCATGTACTTTTTTATTTCCGACGACCAAGAAGCTATCGCAATTGGCTGGA AAGAAGCATCATCGGGTTATCTATCCAAGATTACTAATGGAAAGCATATGCA GTTTGCTACTGGTCATTATGTCCATTATGATAAATCAGATATAATTGCTGAG GAGGCTAAAGCGTTTCTCGAAAAAATAAAGTAA	MKMKILKVFKTIALVLFAIIAALL IASTINHQTLLKNEFKEYYPSDGM VEVDGNKFHVYSEGAGDLTLVFMS GHGTNSPTLDFKPLWMKMVDDYRI AIVEKSGYGWSETSNSPRDIDTIL EETRKALELSGEKGPYVLFPHSMS GLEAIYWAQKYPDEVKAIIGLDPC TPETIKILPEPEKTQLHLMYSISR MGITRFMPDFDFESNFPLMKSEDL TEENKQEYLAVFYKSAFSKDMLRE VNYLYDNAKTVAENEVPIHTPMYF FISDDQEAIAIGWKEASSGYLSKI TNGKHMQFATGHYVHYDKSDIIAE EAKAFLEKIK	No	Alpha/beta hydrolase fold protein [Dethiobacter alkaliphilus] WP_008514622.1

*Fae*9	Esterase lipase family with FAEs	ATGGCTCAGGCCTTTGATTCCCGGGATGTCTGTCTATTTTGTGGACGCCCAA GAACCATTTATGTACTTATAATAAAAAACGAAGAGAGGATTATATATGGCTA CGTTTCAGATCGATTTTTATTCCAACGCACTGAAGGGCATCACACCCCTTAC GGCCATACTTCCGGTGGAGATACCGGAGAATATCCCCGGGATAGAGGCCAGG GAGAGAAAAGAGCCTTTTAGGACCATCTATCTGCTGCACGGGTATTCCGGAA GCAGTAACGACTGGCTCCACGGATCGAGGATCGATTTCCTGGCCCGTATCTT TCAGGTGGCGGTGGTCATGCCCTCGGGCAGAAACAGCTTTTATCTGGACGAT GCGATCAAGGATGAATATTATGAGAGGCTGGTCAGTGAAGAGATCGTGGAAT TCTCCCGCAAGGTGTTTCCCCTGTCTCAAAAGAGAGAGGACACGACCATCGG CGGTCTCTCCATGGGCGGGTACGGCGCGATGCGCAATGGCCTCAAGCACAGC GATGTATTCGGCAGTATATTCGCGTTTTCATCGGCTCTTATTACCGACAAAA CTGCGCAGATGAAAGAAGGCGATAGCAACCCTGTCATGGCGCCTTTTAGCTA TTACAGGCATGTGTTTGGAGATCTGGATACGCTCATAGGCAGTGACAAGGAC CCCAAGGCATTGGCGAAAAGACTCGTGGATAATGGATCGGATATCCCCAAGA TATTTATGGCCTGCGGAACCGAAGACTTTTTGATCGGGGAAAACCGTGACTT CCACCGGCACCTAACCGATCTGGGGATCGAGCACGAGTACAGGGAGAGTCCC GGTATTCATGATTGGGTTTTTTGGGATGCGTATATAGAAAAAGCGATGGAGT GGCTGTACGGAAAACCGAAACTGCCTATGGGACAATGA	MATFQIDFYSNALKGITPLTAILP VEIPENIPGIEARERKEPFRTIYL LHGYSGSSNDWLHGSRIDFLARIF QVAVVMPSGRNSFYLDDAIKDEYY ERLVSEEIVEFSRKVFPLSQKRED TTIGGLSMGGYGAMRNGLKHSDVF GSIFAFSSALITDKTAQMKEGDSN PVMAPFSYYRHVFGDLDTLIGSDK DPKALAKRLVDNGSDIPKIFMACG TEDFLIGENRDFHRHLTDLGIEHE YRESPGIHDWVFWDAYIEKAMEWL YGKPKLPMGQ	No	FAE1 [uncultured bacterium] AGJ83839.1
*Fae*14	Alpha/beta hydrolase	ATGCCGTTCTTCGCCAGGGGTAACGCGCGGATCTATTACGAGGATCGGGGCA GCGGGGAGCCGGTGATCGCGGTCCACGGGCTCATCGGCAACACCCGCTACTG GAAGCCGCTCACCGACCTTATCGGCGAGAGGTGCAGGTTCATCTCCATGGAC ATGCGCGGCCACGGGTACACGGAGGTAAACGGCGAGCCCCATGGCTACGACG TGCAGACCGTGGGCAGGGACATCGTTGCCCTGGCCGATCACCTGGGCATACC GCGCTTTCACCTCCTCACCCACTCCACCGGGGGTTTTGCGGGCGTGAGGCAC GCCATGGACGACTGCAGCCGCTTCAAGTCGCTCATCCTCACCAACACCGCCT CGGCCACCTCCGTGGTCCCGGGCGACGAGCGAACGATCAGGGATTACCACGA CAGGTTCGCCGCGTGGTTTCAGCGCTTCGACTGGGACCAGATCATGGGGGTG CTCAAGATGACGCCGGGCCCGTTTTTCCGGGGCGTCGTGGAGTCGCCGTCAT CCGGGGAGCTGCTCGTTCTCGCCCGCGAGGTGGTTCGGATCGGCAACCGGGA CCTCATTGCGGCGTTCATCCGGTCGTTCTACACCGACCCCGACCCCCGCGTG GATGGGCTCCGCCGGATCAGCTGCCCGGTGCTCATAGTCACCGGGGAAAAGG ACGACCTCATGCTGGAGCCCAGCAGGCTCATGGCCCGCGAAATCCCGGGGGC CCGGCTTCTCGAGTATGAGGGCGTCGGGCACATGAGCGCACTGGAGGCCCCT GACCGCCTGGCCCGCGACGTGATGGACTTCATCGCGGCCCACTCGGGCTGA	MPFFARGNARIYYEDRGSGEPVIA VHGLIGNTRYWKPLTDLIGERCRF ISMDMRGHGYTEVNGEPHGYDVQT VGRDIVALADHLGIPRFHLLTHST GGFAGVRHAMDDCSRFKSLILTNT ASATSVVPGDERTIRDYHDRFAAW FQRFDWDQIMGVLKMTPGPFFRGV VESPSSGELLVLAREVVRIGNRDL IAAFIRSFYTDPDPRVDGLRRISC PVLIVTGEKDDLMLEPSRLMAREI PGARLLEYEGVGHMSALEAPDRLA RDVMDFIAAHSG	No	Alpha/beta hydrolase fold protein [Desulfococcus oleovorans] YP_001529470.1

*Fae*27	Esterase lipase family	ATGCGGGCGGCGAGGATCGATATGAACGGCAACGACAATTCCATGCGCCCCG ACGTGCGGGCGCTGCTCGACATGATGGCGGCGATGCAGCAGCCCAAGCTTTA TGAACTGCCGCTGGAGGAAGCGCGCTCGCCCGCGCTGGTGGCGCTGGTGGAT GCGCCGCCACGCGACCTGCCGGTGATCCGCGATCTCTCCTGCCCCGGGCCTG CGGGCGACATTCCGCTGCGATATTACGACGCGCGGGATAGGCGCGAGGCGGG CCCGGCGATCCTGTTCCTGCATGGCGGCGGTTTCGTGATCGGGAATATCGAC ACCTATCACAGCCTCTGCGCCGAGATCGCGGCGCGTACCGGCCTTCCGGTGG TCTCGGTGGATTACCGGCTGGCGCCCGAACATCCCTTCCCCGCCGCGCCCGA CGATTGCGAAGCGGCGGCGCGCTGGCTGGCCGCTTCTCCCACCGAGCTTGGC TTCGATGTGACGGGGCTCATCCCCATGGGGGACAGCGCCGGTGGCAATCTCA CCATCGTCACGACGCTGGCGCTGCTTGACCGCGCTGCCCAGGTTCCGGTGGT GATGCAGGTGCCGATCTACCCCATCGCCAGCGCGATCGAGGATCATGCCAGC ATGCGCGAATTCGGTGCCGGGCACTTCCTGGAAGCGGAGATCATGGGCTGGT TCACGCACTGCTACGCGCCCGATCCGGCGAGCCCGCGCAATTATCCGCTGCT GCGCGAGGATCATTCGCAGACCCCGCCGACAGTGCTACTGACCGCCGGGCTG GACCCGCTGCGCGATTCCGGACGGGAATATGGCGCCGCGCTGGCGAGCGCAG GCGTGGACCTGACCTTGATCGAAGCGCGCGGAATGATCCATGGCTTCCTGCA AATGCGCAAGGCGCTGCCCAGCACGTCGCGCGACATGGATGCACTTTTCGCC GCAATGGCCGCCACGTTGGAGCGCAGCCTGTGA	MRAARIDMNGNDNSMRPDVRALLD MMAAMQQPKLYELPLEEARSPALV ALVDAPPRDLPVIRDLSCPGPAGD IPLRYYDARDRREAGPAILFLHGG GFVIGNIDTYHSLCAEIAARTGLP VVSVDYRLAPEHPFPAAPDDCEAA ARWLAASPTELGFDVTGLIPMGDS AGGNLTIVTTLALLDRAAQVPVVM QVPIYPIASAIEDHASMREFGAGH FLEAEIMGWFTHCYAPDPASPRNY PLLREDHSQTPPTVLLTAGLDPLR DSGREYGAALASAGVDLTLIEARG MIHGFLQMRKALPSTSRDMDALFA AMAATLERSL	No	Lipase [Erythrobacter sp. SD-21] WP_006832276.1
*Fae*71	Esterase lipase family	ATGTCGAGATGCCGGGCCATGAAGGGCAACACCACCTCGTTGAATCGCTGGC CGATCCGGTTGACGTGGTCGCCATCGTAGACCGCCCATTCATGCGCGATGCC GAACTTCTCCAGCTCCTGGTGCATCAGCGTGTCGTCCGTGACCAGCCCGTCC TTGTCGCCGACATCGGCGCCGATGGCAGTAAAGCTTTTGAGCGCATCGACAT GGCTGGCGACCAGCGCGTAGGGCGAATTGGCCGCCCATTTGGCGATCACCAG CGGATCGACTTCGCCATCCTCGCCCAGCGGGAAATCGGCATGGAATGGCGGG TTGAGCGGGTTGGGCGACCACGCCGTCATCGAGGCGAGACCCGCGCGCATCC CGAAGCCGGATTCGTCGACGCCCTCGAACGGCACGGCGGCCATGGCGGCGGC GCTTTCCGCCGTCTCCTGCCGCGGGCTGATGCAGCAGGCACTCTGCGCCCAG ATCGCGTTGAAGACATCCGGATAGCGCATGCCGATCCGCCAGGTGCCATATC CGCCCATCGAATGGCCGACGAGCGCGCGCCCTTCGCGCGCCGGGATCGTGCG GTAATGGCTGTCGATATAGGCGACGAGGTCGCGGGCGACGAAGGTCTGGAAG TCGCCGACGGTGACCGAATTCGAATACATGCTGCCGCCCATCCTGGTGAGCG TGTCGGGCACCACCACGATGAATTCGCGCCCGGCCGCCGCGTTATCCGCGAC AGCGGTCGGCACCTGCATGAAGTCGTAGAAGTCGCGGCCCGAGATCGCGAAG CCGTGGAGGTAATAGACTACCGGATAGTGCCGCTCCGGGCTGGTCGCATAGC TCGGCGGCAGCACCACCATGGCTTCGCGCAGCAGGCTATTGCCCTCGAGATT GCCTTCGAGCGAAGCGCCGTGAACCATCGGGTTCTCGACCGTGATGCCCTCG GGAACGACCGCGTCGGGCGCGGTGATATTCTGCGCCGCGGCCGGCGCGCCCA GTAGCGCAGCCAGGGCTCCGGCCAGTGCCAGCAATGTCTTCATCGCTAG	MKTLLALAGALAALLGAPAAAQNI TAPDAVVPEGITVENPMVHGASLE GNLEGNSLLREAMVVLPPSYATSP ERHYPVVYYLHGFAISGRDFYDFM QVPTAVADNAAAGREFIVVVPDTL TRMGGSMYSNSVTVGDFQTFVARD LVAYIDSHYRTIPAREGRALVGHS MGGYGTWRIGMRYPDVFNAIWAQS ACCISPRQETAESAAAMAAVPFEG VDESGFGMRAGLASMTAWSPNPLN PPFHADFPLGEDGEVDPLVIAKWA ANSPYALVASHVDALKSFTAIGAD VGDKDGLVTDDTLMHQELEKFGIA HEWAVYDGD	Yes (23 and 24: AAA-QN D=0.842 D-cut-off=0.450)	Esterase [Sphingobium sp. SYK-6] WP_014074884.1

*Fae*83	Esterase lipase family	ATGCAGGACAGCGCGCGGATCGAGCTGAAGAATTTTCATGCGCCGTCGATCG CCGGCAATCTCGAAGGCAACGCGGCAGAGCGTAATGTCTATGTCGTGACGCC GCCGGGCTATGATGAGAATCCGGGCAAGCGCTATCCGGTGGTGGTTTTCCTC CATGGCTACTGGGCGACGCCGCAAATGTACCAGGAGACGATGAAGTTCGAAG AGGCGGTCGATATCGCTGCCGAGGCCGGCAACGAAGTGATTATGGTGATCCC CGACGGCCATTCGAAATTGCGCGGCGGCTTCTATTCGAGCGGACCGACCGTC GGCGATTATGAGAGCTTCGTCGCCCGCGACCTCGTCGGCTGGGTCGATGCGA ATTACCGCACGCTGGCGAAGCCGGAATCGCGCGGCCTCGCCGGCCACTCGAT GGGCGGATACGGCACCATCCGCATCGCGATGAAGAATCCGGGCGTGTTCTCC AGCATTTACATGATGAGCGCCTGCTGCCTCGACCCGATGCCGATCAATGCGG AGACCGCGCAGCGGATCGAGGCGATGAGCGCGGAAGAGGCCGCCAATGCCGA TTTCGGGCAGCTGGCGCCGGTCTCGACGCTGGCCACGTGGTCGCCGGATCCG ACGGCGGAAGGCTGGCTCAAGGCGGATACGGGGCTGAAGGAAGACGGCACGG TCGATCCGCTGGTGAATTACCGGCTGGCGGCCAATTCGCCGGTGGTGATCCT GCCGCAATATCTCCCCGCGCTGAACGGGCTGCGCGCCTTCGCCATGGATATC GGCGACAAGGACTTCCTGCTCGAAGGCAACCGGATCTTCCGCGAGGAGCTCG ACCGGTTCGGCGTCAAATATGATTTCGAGCTCTACGAGGGCGACCATGGCAA CCGCATTCCGGAGCGGATCCGCGCCGAAGTGCTGCCCTTCTTCGCCCAGCAT CCGGAGGCGGAGTGA	MQDSARIELKNFHAPSIAGNLEGN AAERNVYVVTPPGYDENPGKRYPV VVFLHGYWATPQMYQETMKFEEAV DIAAEAGNEVIMVIPDGHSKLRGG FYSSGPTVGDYESFVARDLVGWVD ANYRTLAKPESRGLAGHSMGGYGT IRIAMKNPGVFSSIYMMSACCLDP MPINAETAQRIEAMSAEEAANADF GQLAPVSTLATWSPDPTAEGWLKA DTGLKEDGTVDPLVNYRLAANSPV VILPQYLPALNGLRAFAMDIGDKD FLLEGNRIFREELDRFGVKYDFEL YEGDHGNRIPERIRAEVLPFFAQH PEAE	Yes (28 and 29: AMA-QDD=0.557 D-cut-off=0.450)	Esterase [Sphingobium sp. SYK-6] WP_014074884.1
*Fae*91	Esterase-lipase family	ATGTCCTTGGATATTGGCGGCCAGGGTTTCGTGCTCGCAGGAAACGGTGTTT TCCGCGAGGGGCGCGACCGCTGCGGCGTCCAATGTGATTGCGAGCGTTACGA GGGCGACCAGGGCAACCGCATGCCGGAGCGGACGCGGCGCCGAAGGGCGGCC CTTCTTCGCCCAGCATCTGGAGGCGGAGTGATGGCGATGAAGAAACTGCTGG CACTGGCCGGAGCCCTGGCGGCGCTACTGGGCGCGCCGGCCGCGGCACAGAA TATCACCGCGCCCGACGCGGTCGTTCCCGAGGGCATCGCGGTCGAGAACCCG ATGGTTCACGGAGCTTCGCTCGAAGGCAATCTCGAAGGCAACAACCTGCTGC GCGAAGCGATGGTGGTGCTGCCGCCGAGCTATGCGACCAGCCCGGAGCGGCA CTATCCGGTGGTCTATTACCTCCACGGCTTCGCGATCTCGGGCCGCGACTTC TACGACTTCATGCAGGTGCCGACCGCTGTCGCGGATAACGCGGCGGCCGGGC GCGAATTCATCGTCGTGGTGCCCGACACGCTCACCAGGATGGGCGGCAGCAT GTATTCGAATTCGGTCACGGTGGGCGACTTCCAGACCTTCGTCGCCCGCGAC CTCGTCGCCTATATCGACAGCCATTACCGCACGATCCCGGAGCGCGAAGGGC GCGCGCTCGTCGGCCATTCGATGGGCGGCTACGGCACCTGGCGGATCGGGAT GCGCTATCCGGATGTCTTCAACGCGATCTGGGCGCAGAGTGCCTGCTGCATC AGCCCGCGGCAGGAGACGGCGGAAAGCGCCGCCGCCATGGCCGCCGTGCCGT TCGAGGGCGTCGATGAATCCGGCTTCGGGATGCGCGCGGGTCTCGCCTCGAT GACGGCGTGGTCGCCCAACCCGCTCAACCCGCCCTTCCATGCCGATTTCCCG CTGGGCGAGGATGGTGAAGTCGATACGCTGGTGATCGCCCAATGGGCGGCCC CTTCCTGTCTTTTATCCCCCTCTCCGAGCCCACGCGCCTCCTGGCCAGCACG TCTGCCGCCTTCTGCTGGATGA	MSLDIGGQGFVLAGNGVFREGRDR CGVQCDCERYEGDQGNRMPERTRR RRAALLRPASGGGVMAMKKLLALA GALAALLGAPAAAQNITAPDAVVP EGIAVENPMVHGASLEGNLEGNNL LREAMVVLPPSYATSPERHYPVVY YLHGFAISGRDFYDFMQVPTAVAD NAAAGREFIVVVPDTLTRMGGSMY SNSVTVGDFQTFVARDLVAYIDSH YRTIPEREGRALVGHSMGGYGTWR IGMRYPDVFNAIWAQSACCISPRQ ETAESAAAMAAVPFEGVDESGFGM RAGLASMTAWSPNPLNPPFHADFP LGEDGEVDTLVIAQWAAPSCLLSP SPSPRASWPARLPPSAG	No	Hypothetical protein [Acidobacteriaceae bacterium WP_020713012.1

Appendix B: Protein Methods supplementary data

Appendix B1: Normalisation table for Induced and uninduced FAE9 and FAE27

FAE 9 Induced

Time (hr)	Dilution Factor (DF)	OD_{600} of diluted Sample	OD_{600} at harvest (DF * OD_{600} of diluted sample)	Sample Conc. Factor	Z (conc Factor * OD_{600})	Volume to load (15-Well mini gel) (µl)	Stock Sample	
							Protein sample (µl)	Dye sample (µl)
0	1	0.8592	0.86	500	430	0.42	8.38	11.62
1	10	0.1905	1.91	500	953	0.19	3.78	16.22
2	10	0.1952	1.95	500	976	0.18	3.69	16.31
3	10	0.2209	2.21	500	1105	0.16	3.26	16.74
4	10	0.1941	1.94	500	971	0.19	3.71	16.29
5	10	0.2685	2.69	500	1343	0.13	2.68	17.32
6	10	0.2791	2.79	500	1396	0.13	2.58	17.42
O/N	10	0.4840	4.84	500	2420	0.07	1.49	18.51

FAE 9 Uninduced

Time (hr)	Dilution Factor (DF)	OD_{600} of diluted Sample	OD_{600} at harvest (DF * OD_{600} of diluted sample)	Sample Conc. Factor	Z (conc Factor * OD_{600})	Volume to load (15-Well mini gel) (µl)	Stock Sample	
							Protein sample (µl)	Dye sample (µl)
0	1	0.7432	0.74	500	372	0.48	9.69	10.31
1	10	0.1336	1.34	500	668	0.27	5.39	14.61
2	10	0.2377	2.38	500	1189	0.15	3.03	16.97
3	10	0.2264	2.26	500	1132	0.16	3.18	16.82
4	10	0.2690	2.69	500	1345	0.13	2.68	17.32
5	10	0.3034	3.03	500	1517	0.12	2.37	17.63
6	10	0.3294	3.29	500	1647	0.11	2.19	17.81
O/N	10	0.6002	6.00	500	3001	0.06	1.20	18.80

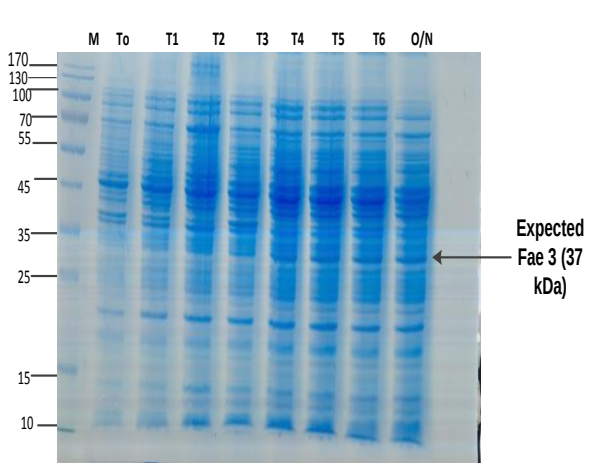
FAE 27 Induced

Time (hr)	Dilution Factor (DF)	OD_{600} of diluted Sample	OD_{600} at harvest (DF * OD_{600} of diluted sample)	Sample Conc. Factor	Z (conc Factor * OD_{600})	Volume to load (15-Well mini gel) (µl)	Stock Sample	
							Protein sample (µl)	Dye sample (µl)
0	1	0.7311	0.73	500	366	0.49	9.85	10.15
1	10	0.1414	1.41	500	707	0.25	5.09	14.91
2	10	0.1337	1.34	500	669	0.27	5.39	14.61
3	10	0.2095	2.10	500	1048	0.17	3.44	16.56
4	10	0.2000	2.00	500	1000	0.18	3.60	16.40
5	10	0.2717	2.72	500	1359	0.13	2.65	17.35
6	10	0.2674	2.67	500	1337	0.13	2.69	17.31
O/N	10	0.3999	4.00	500	2000	0.09	1.80	18.20

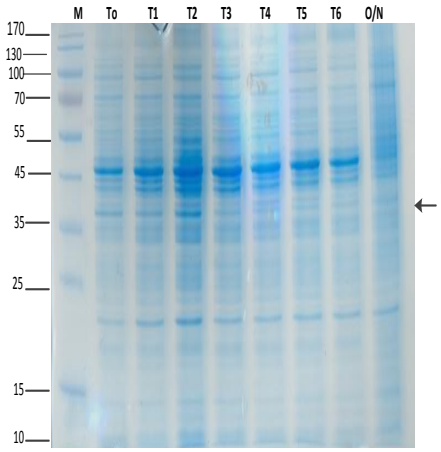
FAE 27 Uninduced

Time (hr)	Dilution Factor (DF)	OD_{600} of diluted Sample	OD_{600} at harvest (DF * OD_{600} of diluted sample)	Sample Conc. Factor	Z (conc Factor * OD_{600})	Volume to load (15-Well mini gel) (µl)	Stock Sample	
							Protein sample (µl)	Dye sample (µl)
0	1	0.7043	0.70	500	352	0.51	10.22	9.78
1	10	0.1301	1.30	500	651	0.28	5.53	14.47
2	10	0.1791	1.79	500	896	0.20	4.02	15.98
3	10	0.1860	1.86	500	930	0.19	3.87	16.13
4	10	0.2553	2.55	500	1277	0.14	2.82	17.18
5	10	0.2841	2.84	500	1421	0.13	2.53	17.47
6	10	0.3149	3.15	500	1575	0.11	2.29	17.71
O/N	10	0.5459	5.46	500	2730	0.07	1.32	18.68

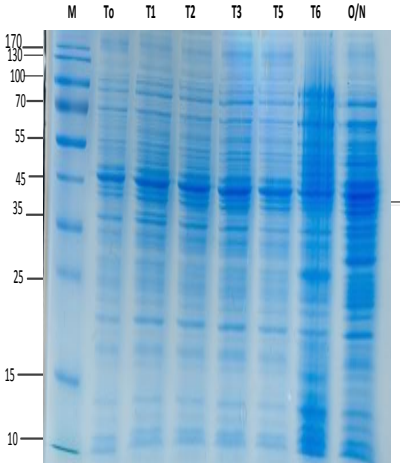
Appendix B2: SDS-PAGE gels of putative FAE(s) that did not express.



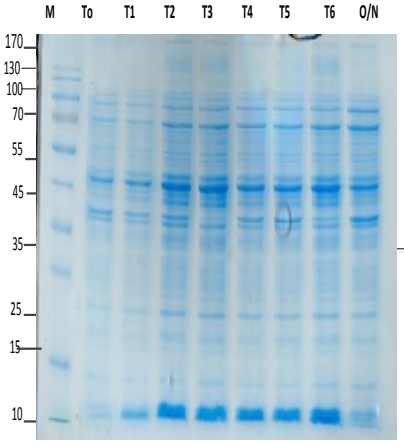
Expected Fae 3 (37 kDa)



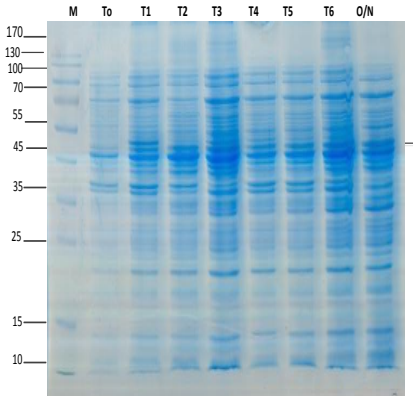
Expected Fae 14 (32 kDa)



Expected Fae 71 (36 kDa)



Expected Fae 83 (36 kDa)

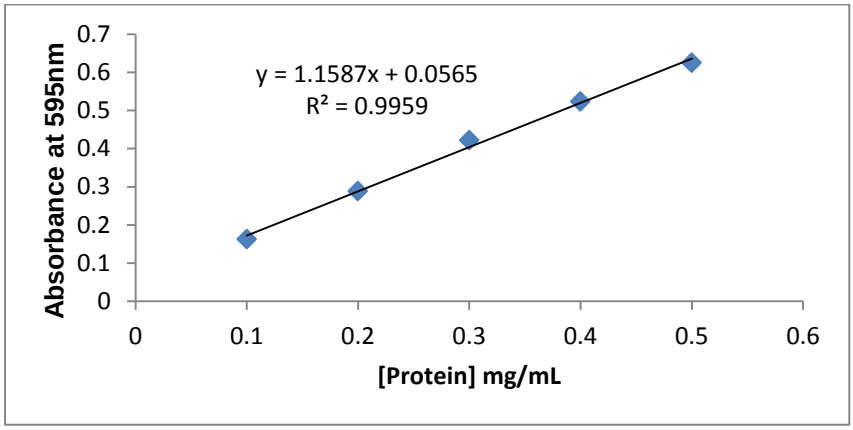


Expected Fae 91 (38 kDa)

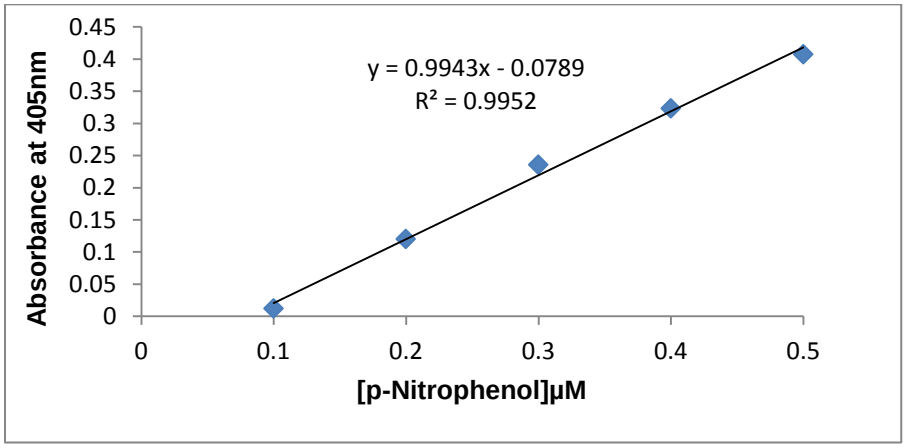
Appendix B3: Preparation of 12% separating gels and 4% stacking gels for SDS-PAGE

12% Separating gel		**4% Stacking gel**	
Reagents	Volume(mL)	Reagents	Volume(mL)
Distilled water	4.3	Distilled water	2.92
40% Acrylamide: 19.1 Bis-acrylamide	3	40% Acrylamide: 19.1 Bis-acrylamide	0.5
1.5M Tris-HCl (pH 8.8)	2.5	0.5M Tris-HCl (pH 6.8)	0.5
10% SDS	01	10% SDS	0.04
10% Ammonium sulphate	0.1	10% Ammonium sulphate	0.04
TEMED	0.004	TEMED	0.004

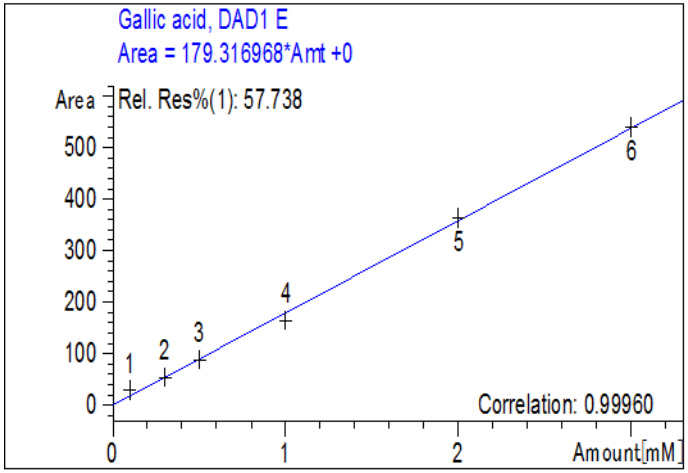
Appendix B4: BSA standard curve for protein assay

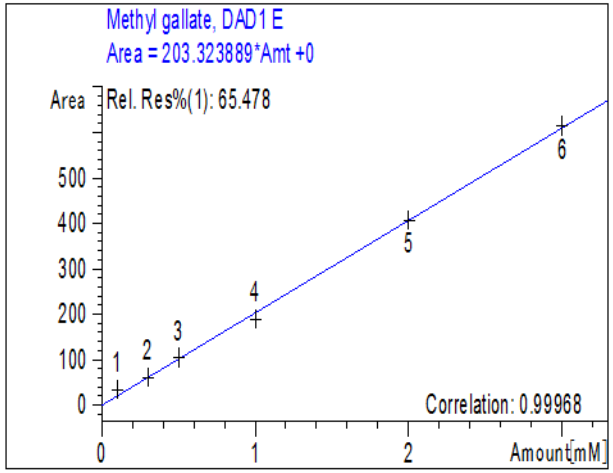


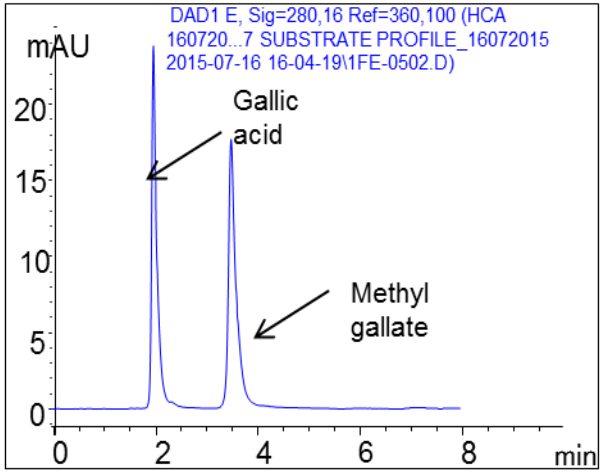
Appendix B5: Standard curve of p-Nitrophenol

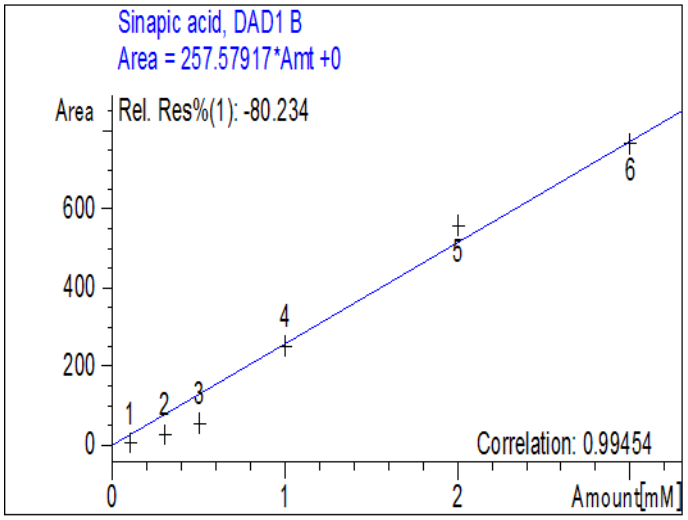


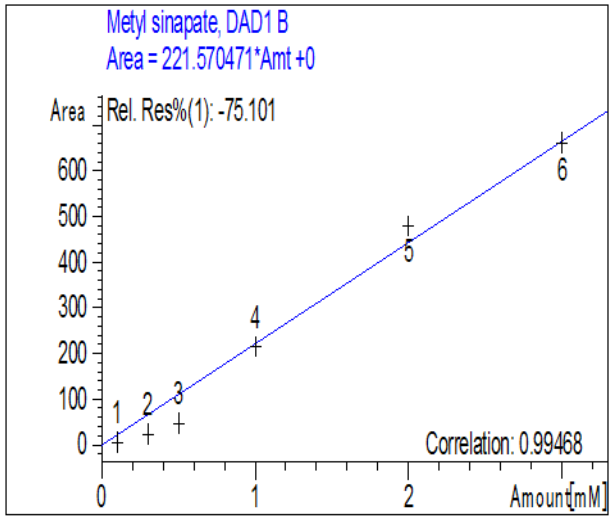
Appendix B6: Standard curves and chromatograms of hydroxycinnamic acids

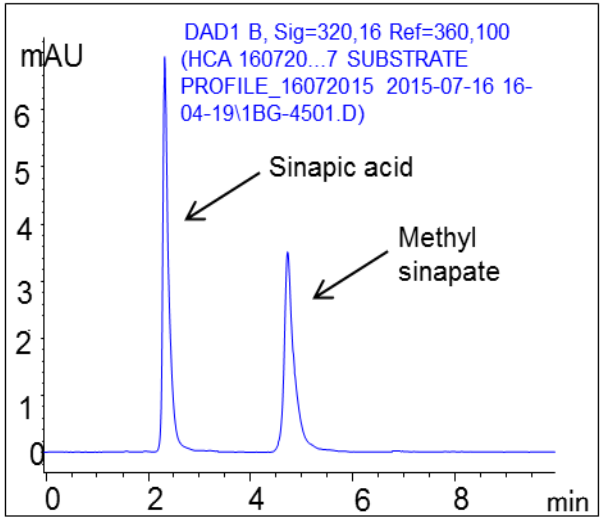


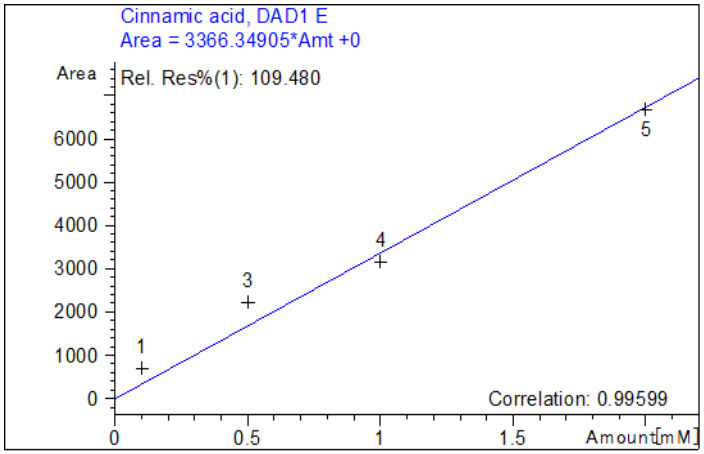


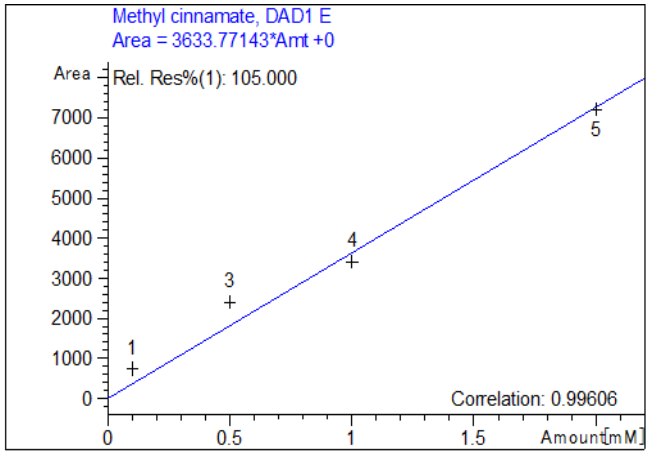


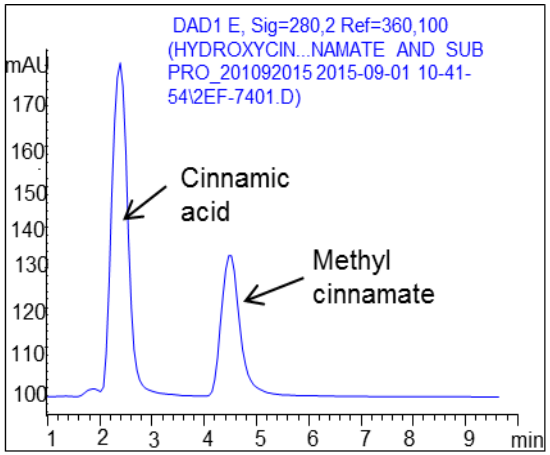


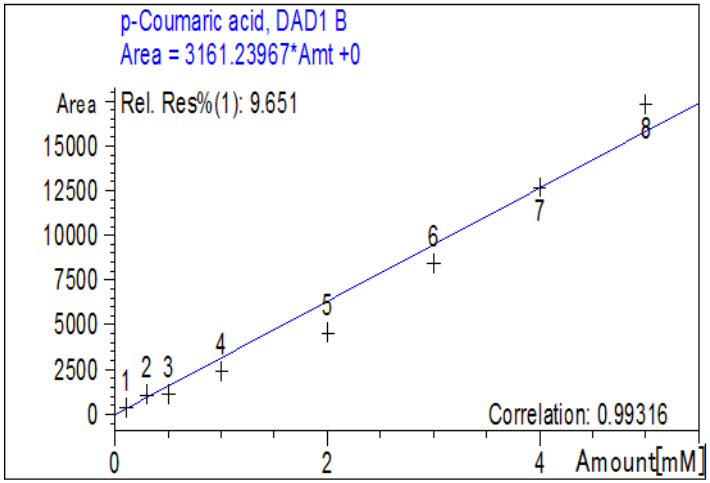


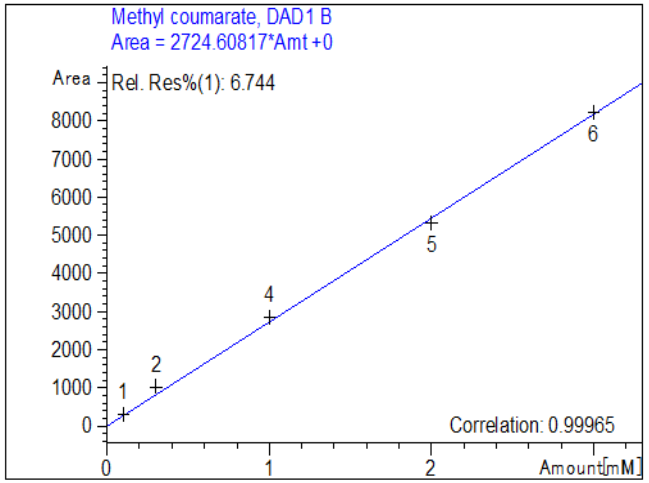


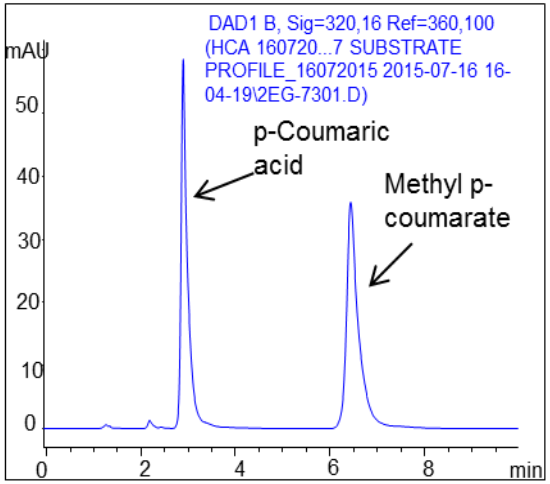

Cinnamic acid, DAD1 E
Area = 3366.34905*Amt +0
Area
Rel. Res%(1): 109.480
Correlation: 0.99599
Amount[mM]

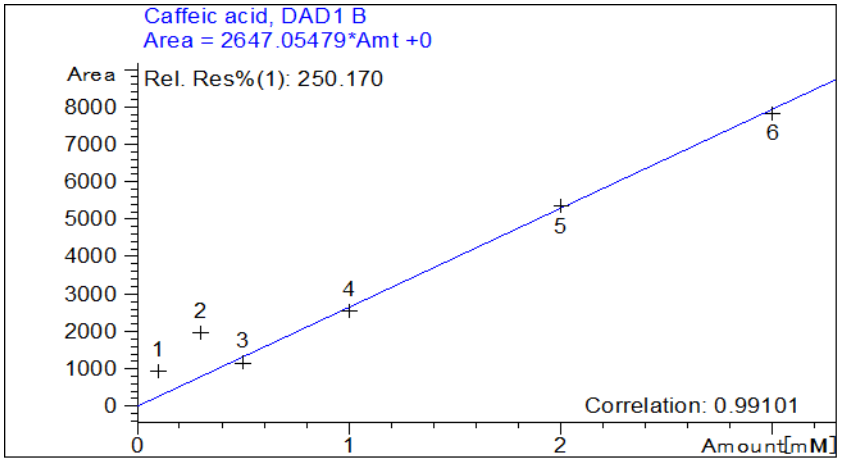

Methyl cinnamate, DAD1 E
Area = 3633.77143*Amt +0
Area
Rel. Res%(1): 105.000
Correlation: 0.99606
Amount[mM]

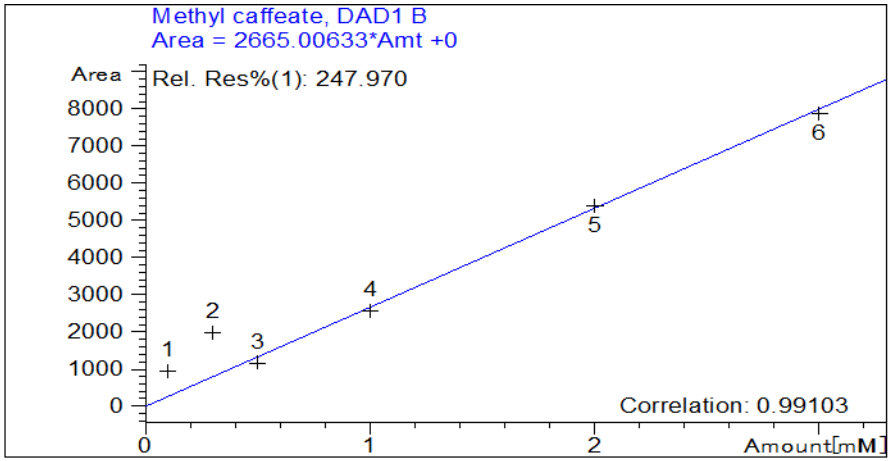

DAD1 E, Sig=280,2 Ref=360,100
(HYDROXYCIN...NAMATE AND SUB
PRO_201092015 2015-09-01 10-41-
54\2EF-7401.D)
mAU
Cinnamic
acid
Methyl
cinnamate
min

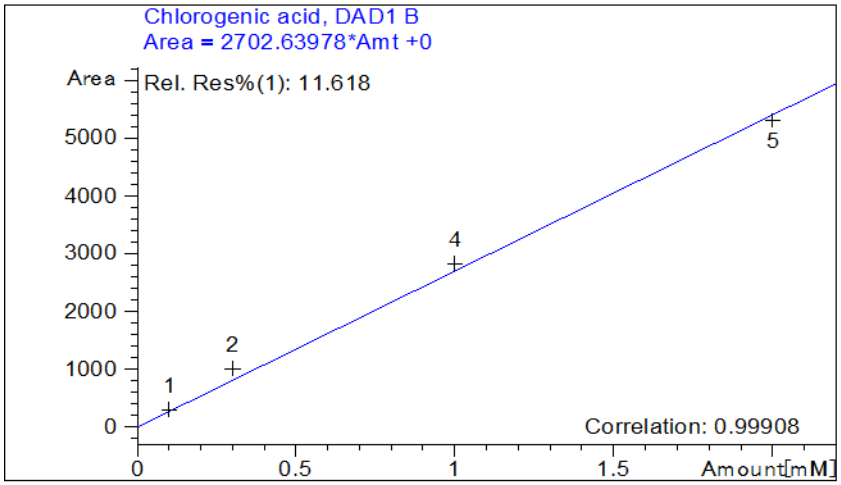

p-Coumaric acid, DAD1 B
Area = 3161.23967*Amt +0
Area
Rel. Res%(1): 9.651
Correlation: 0.99316
Amount[mM]

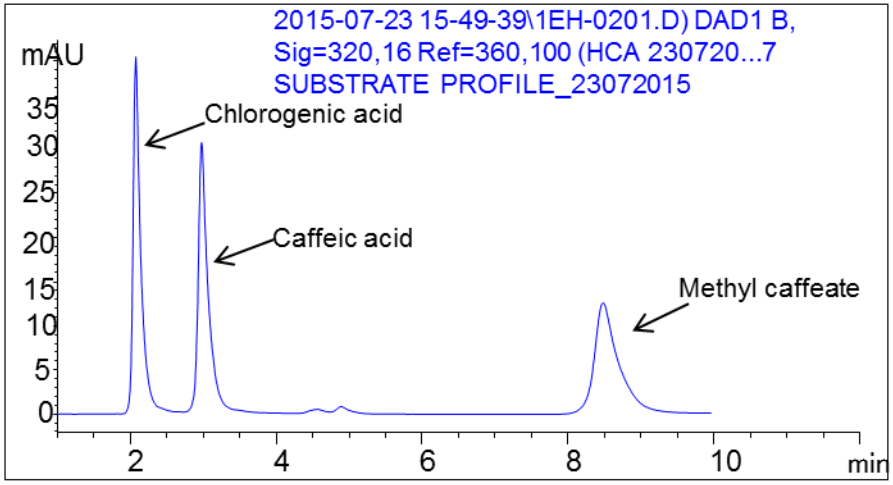

Methyl coumarate, DAD1 B
Area = 2724.60817*Amt +0
Area
Rel. Res%(1): 6.744
Correlation: 0.99965
Amount[mM]

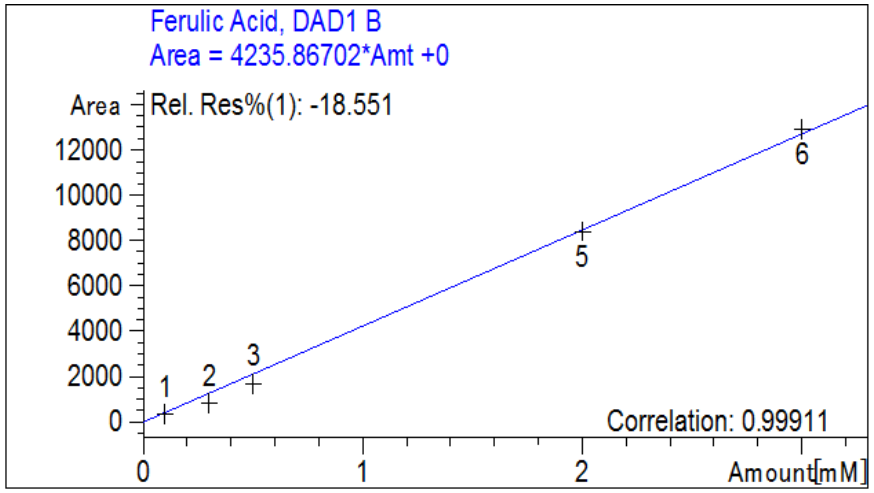

DAD1 B, Sig=320,16 Ref=360,100
(HCA 160720...7 SUBSTRATE
PROFILE_16072015 2015-07-16 16-
04-19\2EG-7301.D)
mAU
p-Coumaric
acid
Methyl p-
coumarate
min

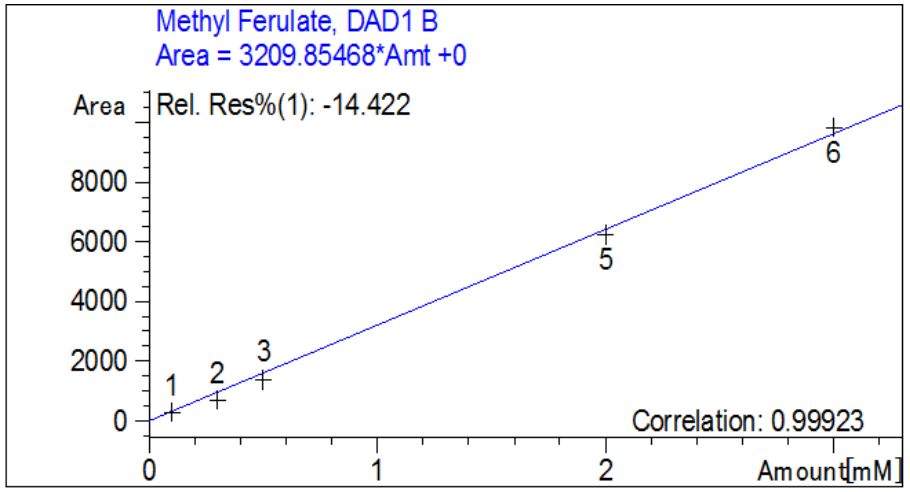

Caffeic acid, DAD1 B
Area = 2647.05479*Amt +0
Area
Rel. Res%(1): 250.170
8000
7000
6000
5000
4000
3000
2000
1000
0
1
2
3
4
5
6
Correlation: 0.99101
0
1
2
Amount[mM]

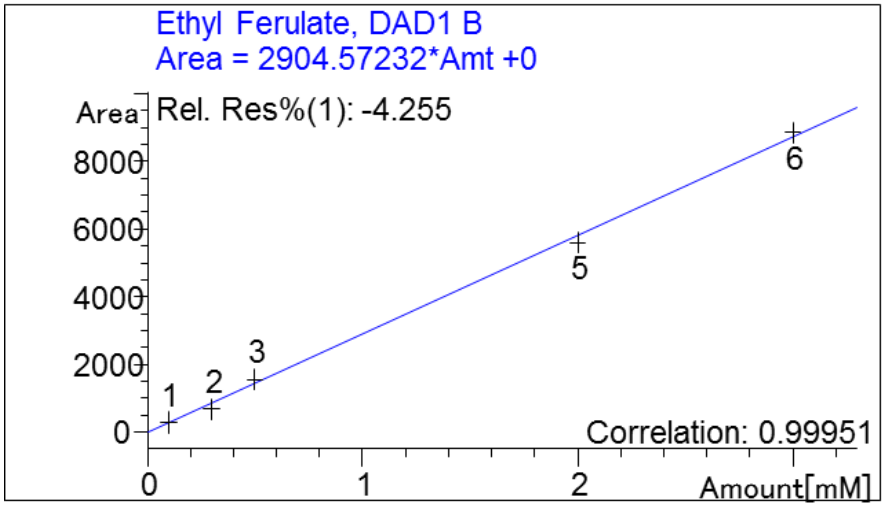

Methyl caffeate, DAD1 B
Area = 2665.00633*Amt +0
Area
Rel. Res%(1): 247.970
8000
7000
6000
5000
4000
3000
2000
1000
0
1
2
3
4
5
6
Correlation: 0.99103
0
1
2
Amount[mM]

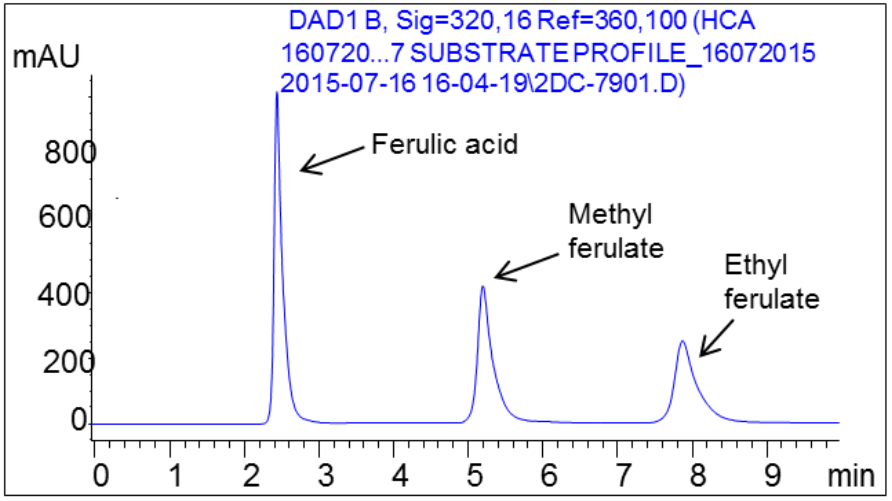

Chlorogenic acid, DAD1 B
Area = 2702.63978*Amt +0
Area
Rel. Res%(1): 11.618
5000
4000
3000
2000
1000
0
1
2
4
5
Correlation: 0.99908
0
0.5
1
1.5
Amount[mM]


2015-07-23 15-49-39\1EH-0201.D) DAD1 B,
Sig=320,16 Ref=360,100 (HCA 230720...7
SUBSTRATE PROFILE_23072015
mAU
35
30
25
20
15
10
5
0
Chlorogenic acid
Caffeic acid
Methyl caffeate
2
4
6
8
10
min


Ferulic Acid, DAD1 B
Area = 4235.86702*Amt +0
Area
Rel. Res%(1): -18.551
12000
10000
8000
6000
4000
2000
0
1
2
3
5
6
Correlation: 0.99911
0
1
2
Amount[mM]

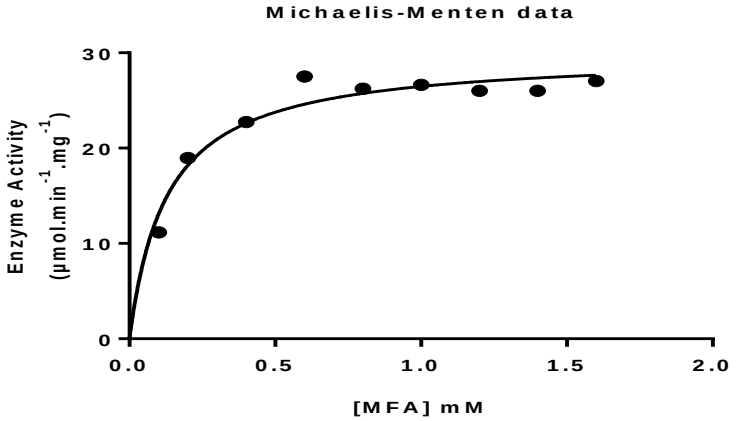

Methyl Ferulate, DAD1 B
Area = 3209.85468*Amt +0
Area
Rel. Res%(1): -14.422
8000
6000
4000
2000
0
1
2
3
5
6
Correlation: 0.99923
0
1
2
Amount[mM]

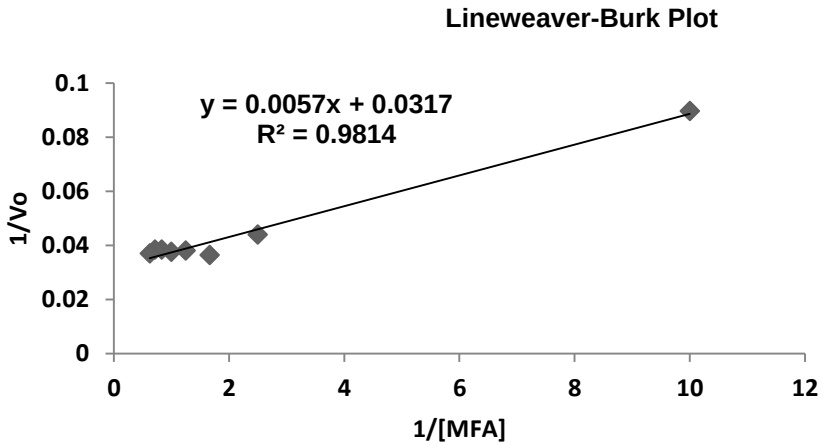

Ethyl Ferulate, DAD1 B
Area = 2904.57232*Amt +0
Area
Rel. Res%(1): -4.255
8000
6000
4000
2000
0
1
2
3
5
6
Correlation: 0.99951
0
1
2
Amount[mM]


DAD1 B, Sig=320,16 Ref=360,100 (HCA 160720...7 SUBSTRATE PROFILE_16072015 2015-07-16 16-04-19\2DC-7901.D)
mAU
800
600
400
200
0
Ferulic acid
Methyl ferulate
Ethyl ferulate
0
1
2
3
4
5
6
7
8
9
min

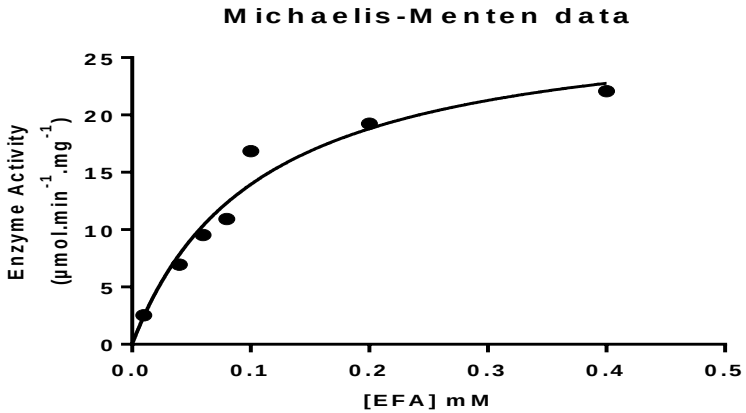
Appendix B7: Michaelis Menten curves and Lineweaver-Burk plots for substrates used in kinetics.

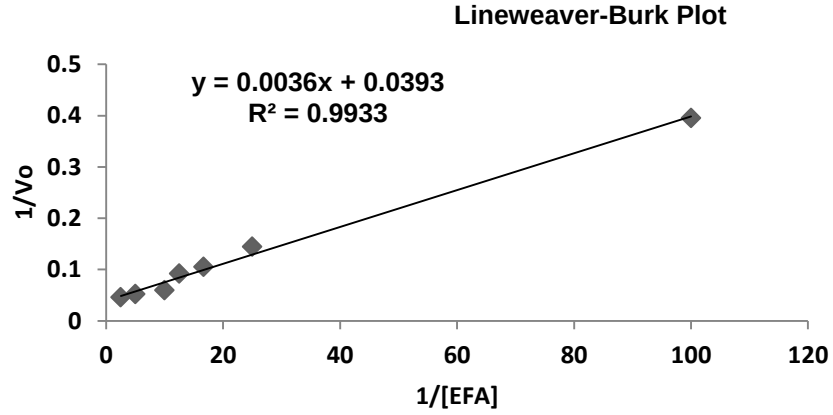
FAE9-MFA



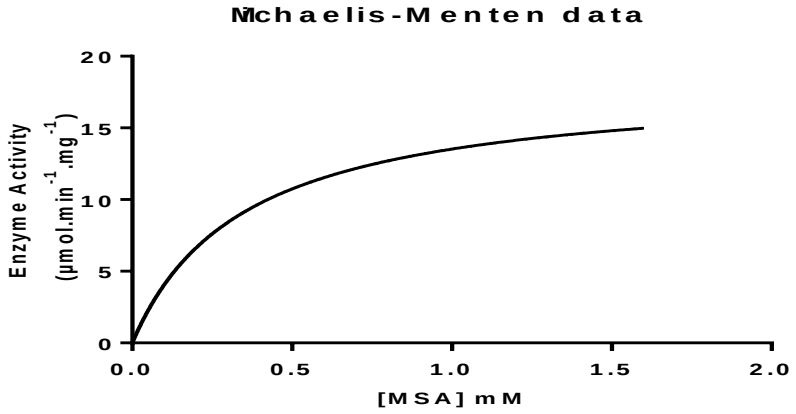


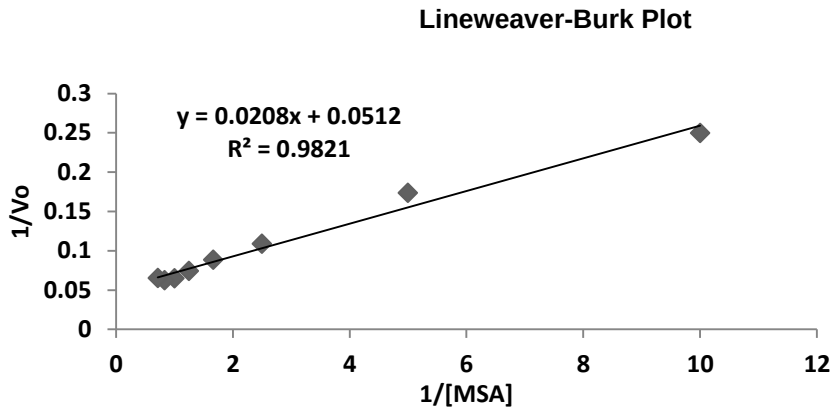
FAE9-EFA



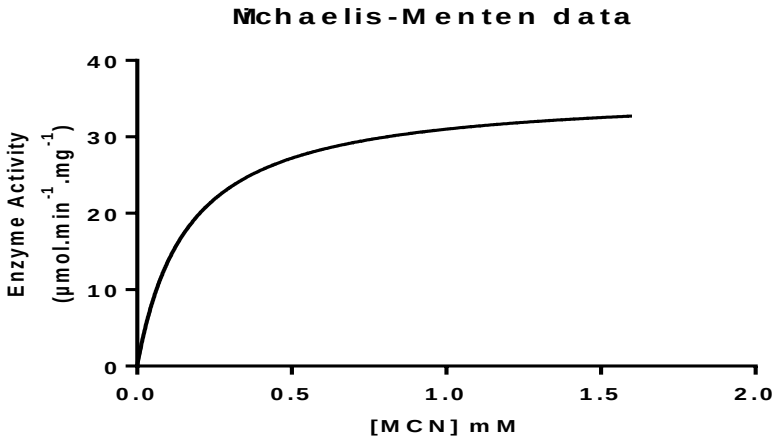


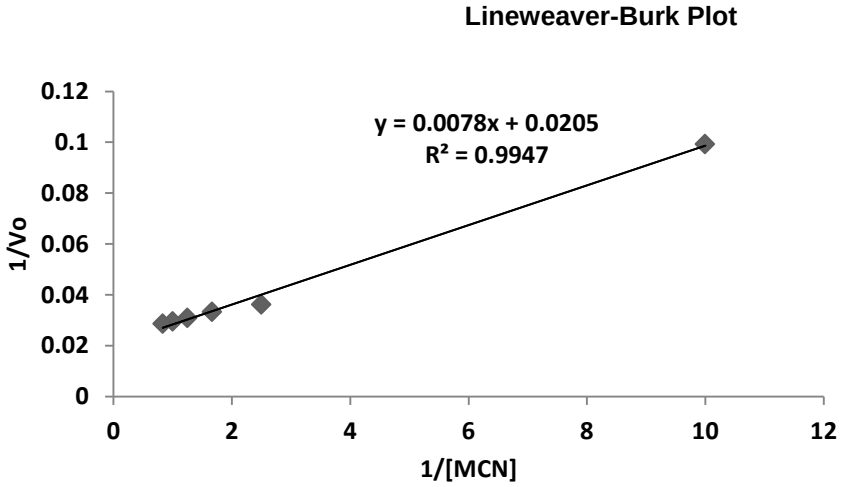
FAE9-MSA



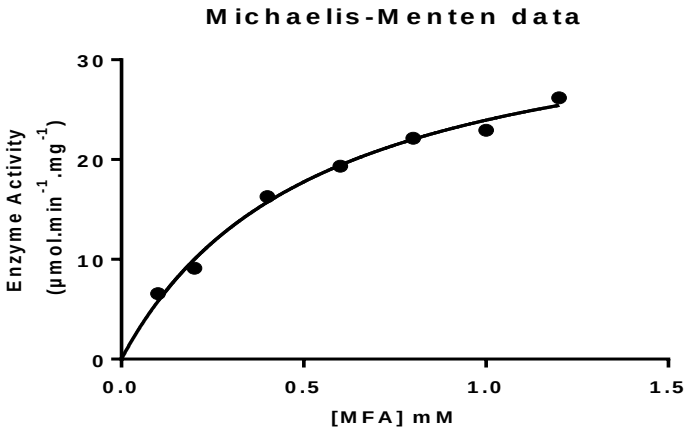


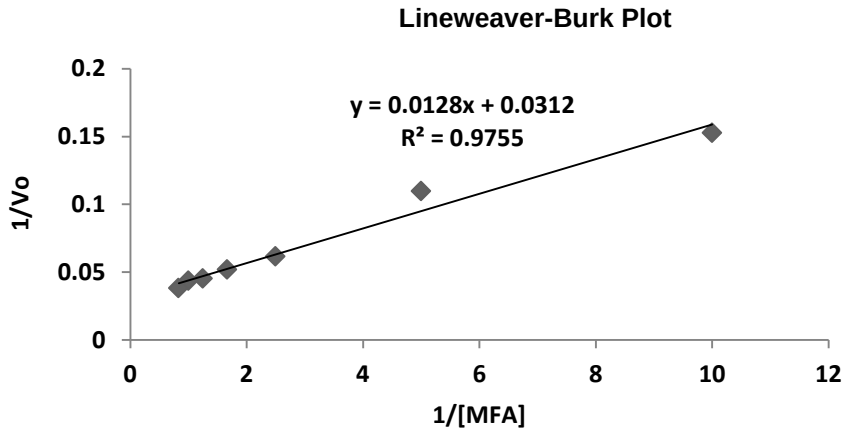
FAE9-MCN



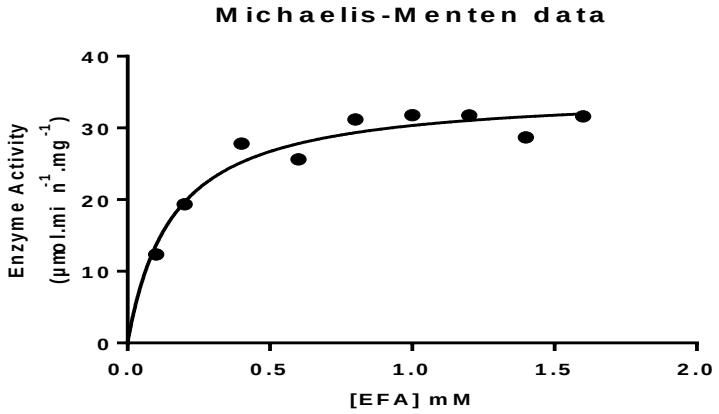


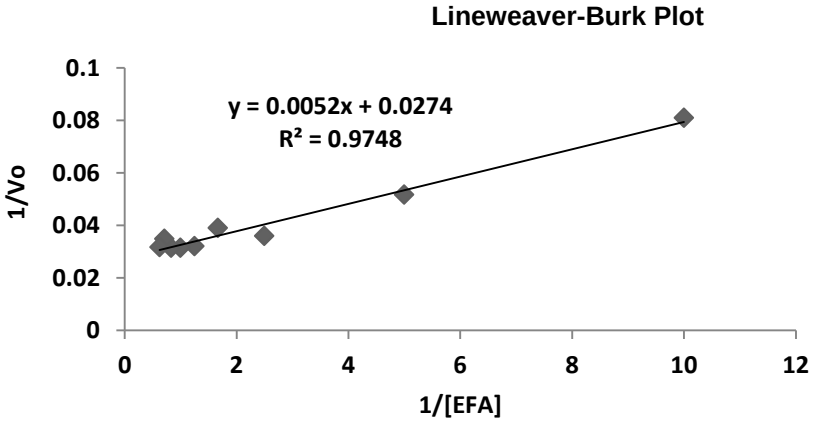
FAE27-MFA



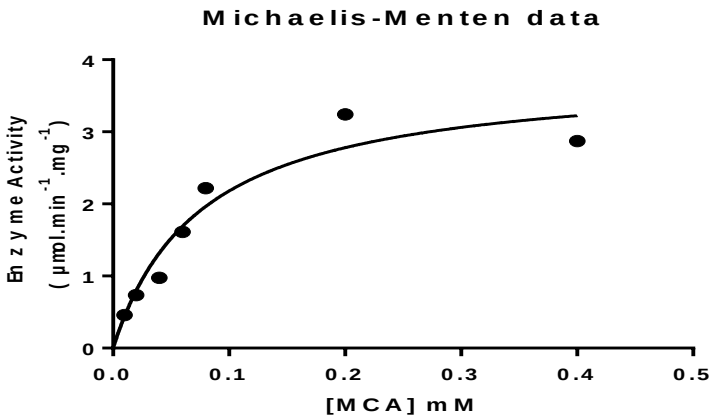


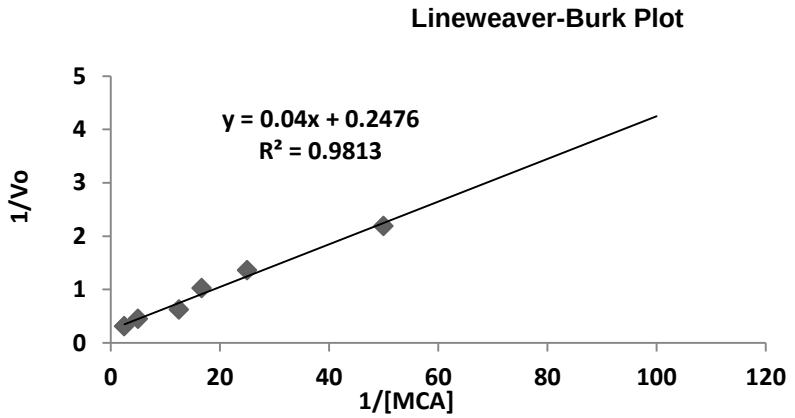
FAE27-EFA



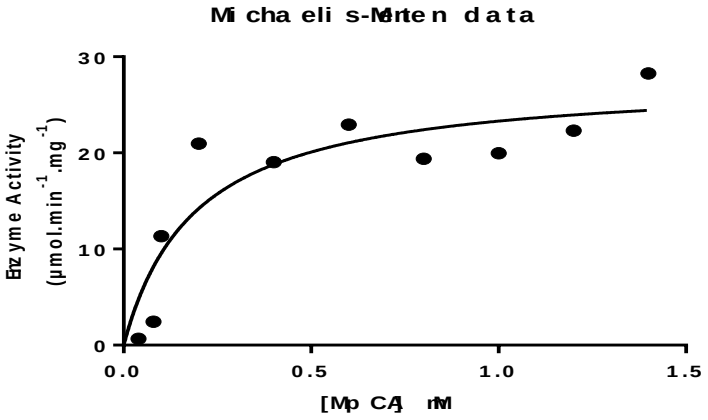


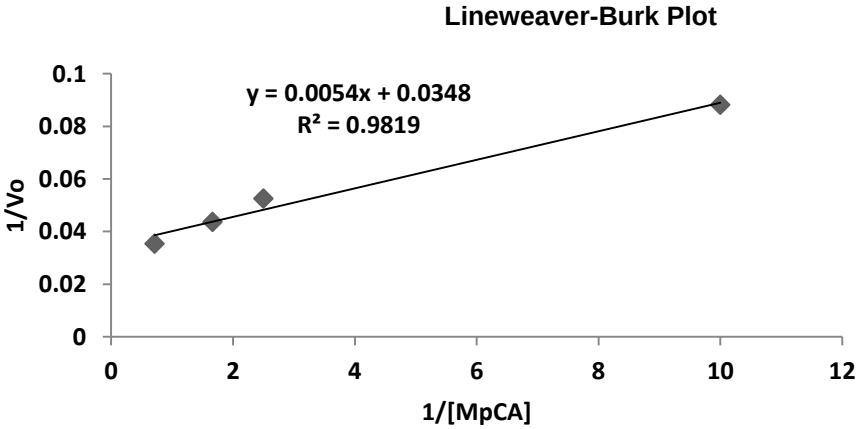
FAE27-MCA



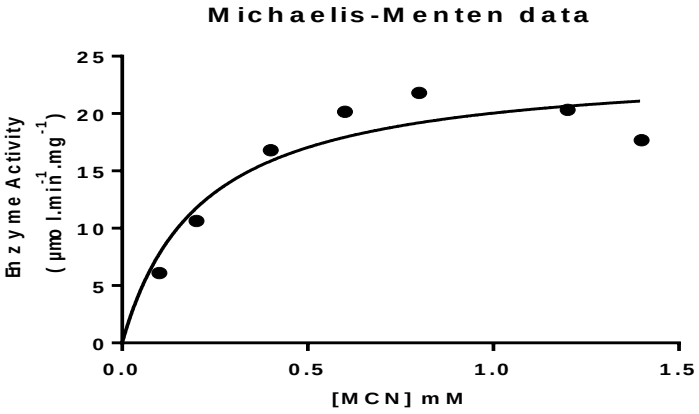


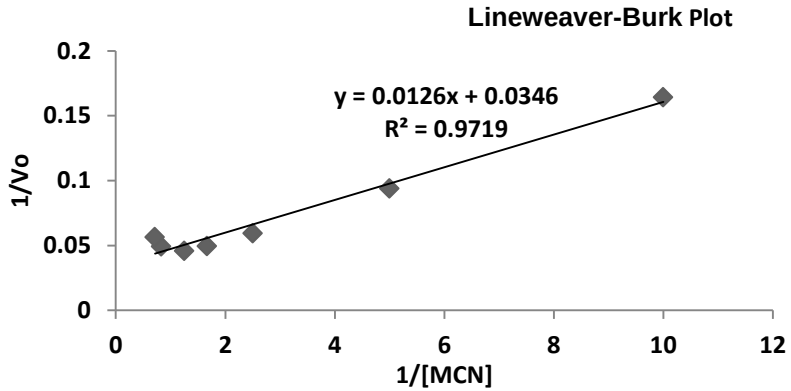
FAE27-MpCA





FAE27-MCN


Michaelis-Menten data
Enzyme Activity
(µmol.min-1.mg-1)
25
20
15
10
5
0
0.0
0.5
1.0
1.5
[MCN] mM


Lineweaver-Burk Plot
y = 0.0126x + 0.0346
R² = 0.9719
1/Vo
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12
1/[MCN]