Metagenomic approaches to gene discovery

Quinton Christian Meyer

A thesis submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Biotechnology,
University of the Western Cape
Bellville

Supervisor: Prof. D.A. Cowan

February 2006

DECLARATION

I declare that *Metagenomic approaches to gene discovery* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Quinton Christian Meyer 21 February 2006

Abstract

The classical approach to gene discovery has been to culture micro-organisms demonstrating a specific enzyme activity and then to recover the gene of interest through shotgun cloning. The realization that these standard microbiological methods provide limited access to the true microbial biodiversity and therefore the available microbial genetic diversity (collectively termed the Metagenome) has resulted in the development of environmental nucleic acid extraction technologies designed to access this wealth of genetic information, thereby avoiding the limitations of culture dependent genetic exploitation. In this work several gene discovery technologies was employed in an attempt to recover novel bacterial laccase genes (EC 1.10.3.2), a group of enzymes in which considerable biotechnological interest has been expressed. Metagenomic DNA extracted from two organic rich environmental samples was used as the source material for the construction of two genomic DNA libraries. The small insert plasmid based library derived from compost DNA consisted of approximately 10^6 clones at an average insert size of 2.7Kbp, equivalent to 2.6 Gbp of cloned environmental DNA. A Fosmid based large insert library derived from grape waste DNA consisted of approximately 44000 cfu at an average insert size of 25Kbp (1.1 Gbp cloned DNA). Both libraries were screened for laccase activity but failed to produce novel laccase genes. As an alternative approach, a multicopper oxidase specific PCR detection assay was

developed using a laccase positive *Streptomyces* strain as a model organism. The newly designed primers were used to detect the presence of bacterial multicopper oxidases in environmental samples. This resulted in the identification of nine novel gene fragments showing identity ranging from 37 to 94% to published putative bacterial multicopper oxidase gene sequences. Three clones p*MCO*6, p*MCO*8 and p*MCO*9 were significantly smaller than those typically reported for bacterial laccases and were assigned to a recently described clade of *Streptomyces* bacterial multicopper oxidases. Two PCR based techniques were employed to attempt the recovery of flanking regions for two of these genes (p*MCO*7 and p*MCO*8). The use of TAIL-PCR resulted in the recovery of 90% of the p*MCO*7 ORF. As an alternative approach the Vectorette™ system was employed to recover the 3' downstream region of p*MCO*8. The complexity of the DNA sample proved to be a considerable technical challenge for the implementation of both these techniques. The feasibility of both these approaches were however demonstrated in principle. Finally, in an attempt to expedite the recovery of full-length copies of these genes a subtractive hybridization magnetic bead capture technique was adapted and employed to recover a full – length putative multicopper oxidase gene from a *Streptomyces* strain in a proof of concept experiment. The StrepA06p*MCO* gene fragment was used as a 'driver' against fragmented *Streptomyces* genomic DNA ('tester') and resulted in the recovery of a 1215 bp open reading frame. Unexpectedly, this ORF showed only 80% identity to the StrepA06p*MCO* gene sequence at nucleotide level, and 48% amino acid identity to a putative *mco* gene derived from a *Norcardioides* sp JS614.

Acknowledgements

I would like to express my most sincere gratitude and appreciation to the following people for their assistance and support throughout this study:

My supervisor, Professor Don Cowan for his patience and unwavering support.

My family and friends for their continued encouragement.

The National Research Foundation and the Ryoichi Sasakawa young leadership fund for providing the necessary funding.

CTAB	Cetyl Trimethyl Ammonium Bromide
PVPP	Polyvinylpyrrolidone
PEG	Polyethylene glycol
SIP	Stable isotope probing
BrdUTP	5-Bromo-2'-deoxyuridine 5'-triphosphate
SH	Suppressive hybridization
DEA	Differential expression analysis
SI	Super-integron
BAC	Bacterial artificial chromosome
ABTS	2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)
HAA	3-hydroxyanthranilate
MCO	Multicopper oxidase
TB	Terrific broth
FD	Farad
DGGE	Denaturing gradient gel electrophoresis
SAP	Shrimp alkaline phosphatase
bp	base pairs
°C	degrees celsius
DNA	deoxyribonucleic acid

ddH_2O	deionized distilled water
EDTA	ethylenediamine tetra-acetic acid
et al.	*et alia* (and others)
g	gram
h	hour
IPTG	isopropyl-β-D-thiogalactopyranoside
kDa	kilodalton
kbp	kilobasepairs
l	liter
LB	Luri Bertani
M	molar
mg	milligram
min	minutes
mM	millimolar
ng	nanogram
OD	optical density
PCR	polymerase chain reaction
pmol	picomol
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
sec	second
Tris	Tris-hydroxymethyl-aminomethane

µg	microgram
µl	microlitre
V	volts
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ATP	adenosine triphosphate
BSA	bovine serum albumin
C-terminus	carboxy terminus
dATP	deoxy-adenine 5'-triphosphate
dCTP	deoxy-cytosine 5'-triphosphate
DEPC	diethylpyrocarbonate
dGTP	deoxy-guanine 5'-triphosphate
dNTP	deoxyribonucleotides
dTTP	deoxy-thymine 5'-triphosphate
i.e.	that is
MCS	multiple cloning site
mRNA	messenger RNA
MW	molecular weight
O/N	overnight
ORF	open reading frame
RFLP	restriction fragment length polymorphisms
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT	room temperature

RT-PCR	reverse transcription polymerase chain reaction
U	units of enzymatic activity
UV	ultraviolet
w/v	weight per volume
cfu	Colony forming unit

Table of contents

Table of figures

Table of tables

Chapter 1: Literature Review

Material from this chapter appears in Cowan *et al.*: Metagenomic gene discovery: past, present and future. TIBTECH 2005, 23(6):321-329.

1.1 Introduction

It is widely accepted that the microbial biodiversity present in the environment (collectively termed the Metagenome) had been severely underestimated. This was mainly because the determination of bacterial biodiversity relied on identifying bacteria by culturing environmental samples, without fully realizing the bias introduced by the chosen culturing conditions. During the past 15 years however, two studies made significant contributions to our understanding of the true extent of bacterial biodiversity, by providing evidence that demonstrates the gross underestimation.

Torsvik and co-workers (1990) employed DNA re-association kinetic analysis on microbial community DNA and estimated the bacterial diversity in the order of 10^3 taxa per sample, an impressive figure at the time. Amman and co-workers (1995) then demonstrated that direct microscopic visualization of a natural sample yielded a

population count of one to two orders of magnitude higher than estimates from culturing the same sample.

Both estimates proved to be conservative as the soil matrix obscures a significant fraction of the organisms present, which was therefore not identified by direct visualization. Also due to the phenotypic similarity of many bacterial taxa, distinguishing between taxa by morphology alone does not provide reliable data for species quantification.

Re-analysis of DNA re-association kinetic data has demonstrated that an inaccurate assumption of an equal frequency distribution for bacterial taxa in environmental samples has led to the conclusion that the true bacterial biodiversity was probably underestimated a thousand fold (Gans *et al.*, 2005). It is therefore not unreasonable to assume a prokaryotic taxa diversity of 10^6 species in a nutrient rich environmental sample. It is generally assumed that bacterial species show a log normal distribution with the majority of the taxa equally represented, a small number of taxa dominating the niche and an equal fraction of under-represented or rare species, although the true nature of the distribution frequency remains unknown at this point (Curtis *et al.*, 2002).

It is now accepted in the wider microbial ecology scientific community that as much as 99% of the microorganisms present in nature are not cultivated by the standard techniques (Amman *et al.*, 1995; Schloss and Handelsman, 2003; Daniel, 2005). As a result these microorganisms have erroneously been described as unculturable in some of the earlier literature. It has become more apparent that the limited access through culturing reflects the inadequacy of the standard culturing

techniques (Leadbetter, 2003; Dunbar *et al.*, 1999). Several hypotheses have been proposed to explain the observed poor culturability. These include inadequate knowledge of the required chemical composition of culture media needed for successful culturing, underestimation of the possible slow growth rate of many bacterial taxa or possibly that unculturable microorganisms are in fact phylogenetically similar to the culturable microbes but in a physiological state that renders them recalcitrant to culturing (Rondon *et al.*, 1999; Leadbetter, 2003).

The majority of the microorganisms on earth are located in the open ocean and soil as well as in the oceanic and terrestrial sub-surfaces. One study estimated the total number of prokaryotic cells on earth as 4-6 x 10^{30} and with a growth rate of 1.7 x 10^{30} cells per year (Whitman *et al.*, 1998). Clearly, bacteria present a widely distributed and abundant source of genetic diversity that can be exploited for the recovery of novel genes, their products and entire metabolic pathways, some of which can be of significant economic value.

The exploitation of this genetic reservoir has now been made possible by advances in our ability to recover significantly more genetic information from environmental samples in a culture independent manner. Total community DNA extraction technology is giving researchers access to the genomes of previously “unculturable” microorganisms, the result of which is a clearer and more comprehensive insight into microbial ecology and greater access to potentially novel and useful genes through the adaptation of classical molecular biology techniques on a metagenomic DNA scale.

In this chapter the current technologies employed in metagenomic gene discovery are reviewed. This includes approaches to environmental nucleic acid extraction and purification, sequence independent gene discovery strategies (*i.e.* metagenome sequencing, environmental library construction and activity based screening) as well as sequence dependent approaches such as gene-specific PCR. Other technologies relevant to metagenomic gene discovery also discussed here include gene and genome enrichment strategies. This review concludes with an overview of multicopper oxidases, in particular laccase (E.C. 1.10.3.2), a class of enzyme which is of considerable interest in a biocatalytic context and the target enzyme of this thesis.

1.2 Environmental (Metagenomic) nucleic acid extraction

1.2.1 Metagenomic DNA extraction

The approach to metagenomic DNA extraction broadly remains similar to that used in the extraction of DNA from pure cultures. Basic steps include cell lysis, separation of the DNA and cell debris, and DNA recovery and purification. Successful extraction is, however, complicated by several inefficiencies during the isolation procedure. These include incomplete cell lysis, the absorption of DNA to soil surfaces, the co-extraction of enzymatic inhibitors from soil, as well as the loss, degradation and damage of DNA (Rochelle *et al*., 1992; More *et al*., 1994; Frostegard *et al*., 1999). Despite these limitations two principle strategies exist for the recovery of microbial DNA from soil, the cell extraction and direct lysis methods. For more comprehensive reviews see Roose-Amsaleg *et al.* (2001) and Robe *et al.* (2003).

1.2.1.1 Cell extraction methods

The cell extraction method isolates intact organisms from the soil sample prior to cell lysis. Isolation is most commonly achieved by repeated grinding and differential centrifugation (Holben *et al.*, 1988) or by density centrifugation in a medium such as sucrose, Nycodenz, Percoll or metrizamide (Pillai *et al.*, 1991; Robe *et al.*, 2003). The advantage of this method is that many naturally occurring contaminants which may hamper the subsequent molecular manipulation of the recovered DNA can be removed. This method also reassures the researcher of the source of the genetic material (if his target is specifically of bacterial origin), as fungi, plant and other environmental DNA can easily be removed by washing and centrifugation or filtration. There are, however, several disadvantages to this method which have resulted in the direct lysis method being the strategy of choice.

An important disadvantage is the loss of microorganisms during isolation, resulting in the extracted community DNA not being representative of the actual biodiversity. It has been reported that only 25-35% of the microbes present in the environmental sample are actually recovered (Holben *et al.*, 1988; Steffan *et al.*, 1998). The poor recovery can be explained by the fact that different bacterial groups associate with varying strengths to soil particles (Prieme *et al.*, 1996). The loss of genetic information can skew the data of environmental population studies, and reduce the scope for discovery of novel genes. This method is also much more time consuming and labor intensive when compared with direct lysis protocols.

1.2.1.2 Direct lysis methods

Extraction methods based on the direct lysis strategy do not require the isolation of microorganisms prior to cell lysis. This strategy generally yields more DNA and is believed to provide a better representation of the environmental biodiversity (More *et al*., 1994). The most important disadvantage of direct lysis methods is the co-recovery of humic and fulvic acids with the environmental DNA, as a result of its similar physico-chemical properties (Tebbe and Vahjen, 1993; Ogram *et al*., 1987). These contaminants are visible as a dark color in the DNA extracts and have been demonstrated to act as inhibitors in both restriction endonuclease and polymerase chain reactions (Tebbe and Vahjen, 1993; Jackson *et al*., 1997; Miller *et al*., 1999).

The first step in any direct lysis isolation procedure is the disruption of the soil matrix and the lysis of the bacterial cells. To this end mechanical, chemical and enzymatic methods have been developed (see Table 1-1). An extraction protocol will often require the use of a combination of different methods (Miller *et al*., 1999). The most widely used mechanical lysis methods are thermocycling (Hugenholtz *et al*., 1998) and bead mill homogenization (Kuske *et al*., 1998; Miller *et al*., 1999), although ultrasonication (Picard *et al*., 1992), microwave heating (Dijkmans *et al*., 1993) and grinding under liquid nitrogen have also been reported (Zhou *et al*., 1996). Despite the robust nature of these methods, they have proven to be both necessary and effective for optimum DNA isolation (Miller *et al*., 1999). The physical disruption causes a more effective dislodging and separation of the micoorganisms from the soil matrix, allowing better access to the lysis buffer and therefore increased cell lysis. A disadvantage of these methods is the degradation of the isolated DNA. The mechanical forces tend to

shear the DNA which can pose a problem depending on the prospective use of the DNA.

For most applications high molecular weight DNA is required. This is particularly important for metagenomic DNA library construction, irrespective of whether phage or plasmid vectors are used. Large inserts reduce the number on clones needed to represent the community, and provide greater opportunity for recovery of full length open reading frames. Restriction enzyme digestion of excessively sheared DNA for the construction of a small insert library significantly reduces the cloning efficiency because the generated DNA fragments have incompatible ends. A potential risk when employing PCR based techniques on sheared DNA is the possibility of chimera formation (Field *et al.*, 1997). However, DNA fragment size can be controlled by the choice of extraction method. Miller and co-workers (1999) have demonstrated that by altering the duration and speed of the bead mill homoginization step, larger fragments can be selectively produced.

Chemical and enzymatic lysis methods are comparatively more gentle procedures that minimize DNA damage. However, these methods tend to be less effective at disrupting the soil matrix and exposing the cells to the lysis buffer. In most chemical lysis procedures a detergent such as sodium dodecyl sulfate (Zhou *et al.*, 1996) or sarkosyl (Holben *et al.*, 1988) is used to aid cell membrane lysis (Berthelet *et al.*, 1996). In addition buffers are complemented by the addition of other compounds such as chelating agents (EDTA, Chelex 100) to inhibit nuclease activity and disperse the soil matrix (Miller *et al.*, 1999). Humic acid complexing compounds such as PVPP (Gray and Herwig, 1996) and CTAB (Zhou *et al.*, 1996) are also used in an effort to

increase DNA purity. Chemical lysis methods are also accompanied by the addition of enzymes to promote cell lysis. Commonly used enzymes include lysozyme (Rochelle *et al*., 1992; Tebbe and Vahjen, 1993), Proteinase K (Zhou *et al*., 1996), achromopeptidase (Liu *et al*., 1997a) and pronase E (Jacobsen and Rasmussen, 1992).

To recover the DNA from the soil and cell debris (see Table 1-2), phenol and/or chloroform extraction steps are included in extraction protocols (Tebbe and Vahjen, 1993; Zhou *et al*., 1996). The use of saturated salt solutions such as sodium chloride, potassium chloride, ammonium acetate, sodium acetate and potassium acetate to separate the DNA from contaminating proteins has also been investigated (Selenska and Klingmuller, 1991; Frostegard *et al*., 1999; Miller *et al*., 1999). Once the DNA is recovered it can be concentrated via ethanol, isopropanol or PEG precipitation. It has been suggested that alcoholic precipitation favors the co-precipitation of humic acids whereas PEG does not. The latter, however, acts as a PCR inhibitor and must also be removed, resulting in a further loss of DNA. Isopropanol is considered as a good compromise since it produces a good yield without compromising DNA purity. For a more detailed review, see Roose-Amsaleg (2001). To successfully manipulate the isolated DNA, further purification is often critical. Several purification techniques exist (see Table 1-3), and are often combined to achieve a suitable level of purity. Purification strategies include cesium chloride density ultracentrifugation, adsorption chromatography, agarose gel electrophoresis and in rare instances dialysis and filtration (Tebbe and Vahjen, 1993; Jacobsen and Rasmussen, 1992; Berthelet *et al*., 1996; Stach *et al*., 2001).

1.3 Environmental RNA extraction

The technologies for recovering RNA from environmental samples are similar to those described previously for DNA isolation. DNA extraction protocols are merely modified to optimize RNA recovery. The major factor that needs to be considered for successful RNA extraction is the low stability of RNA. RNA degradation is minimized by ensuring that RNases are suitably inactivated. This involves working at low temperatures together with the addition of RNase inhibitors. The direct addition of commercial reagents, like RNAlater (Sigma) to environmental samples aids in the preservation of the RNA molecules. This is of critical importance where the gene expression profile in the environmental sample is under investigation. In most extraction procedures, DNA and RNA are co-extracted followed by DNA degradation to purify the RNA. The co-recovery and purification of both DNA and RNA from twenty diverse environmental samples has been reported (Hurt, 2001). It was showed that it was possible to purify both community DNA and RNA from a single sample in very good yield.

Table 1-1 Methods used for cell lysis in direct extraction procedures

Cell lysis method	Description	Comment	Reference
Mechanical	Freeze-thaw cycles Freeze-boil cycles Mortar mill grinding Bead-mill homogenization Bead-beating Microwave heating Sonication	May favour disruption of gram (-) bacteria Considered most effective method for cell disruption Lysis of gram (+) Uncommon/ rarely used Disperse soil	(More *et al.*, 1994) (Zhou *et al.*, 1996) (Miller *et al.*, 1999) (Dijkmans *et al.*, 1993) (Frostegard *et al.*, 1999)
Chemical	SDS Sarkosyl	Detergent Detergent	(Zhou *et al.*, 1996) (Holben *et al.*, 1988)
Enzymatic	Lysozyme Proteinase K Achromopeptidase Pronase E	Aids cell lysis	(Zhou *et al.*, 1996) (Zhou *et al.*, 1996) (Liu *et al.*, 1997a) Jacobsen and Rasmussen, (1992)

Table 1-2 Reagents included for the removal of contaminants

Contaminant removal	Description	Comment	Reference
Organic solvents	Phenol Chloroform Isoamyl alcohol	Protein extraction	(Zhou *et al.*, 1996)
Salts	Sodium chloride Potassium chloride Ammonium acetate Sodium acetate Potassium acetate	Disperses and binds soil Separation of DNA and contaminating proteins	(Frostegard *et al.*, 1999) (Miller *et al.*, 1999) (Selenska and Klingmuller, 1991)
Other	PVPP CTAB	Absorbs humic acids and other phenolic compounds	(Berthelet *et al.*, 1996) (Zhou *et al.*, 1996)

Table 1-3 Post-extraction DNA purification methods

DNA purification	Description	Comment	Reference
CsCl centrifugation	CsCl density gradient	Separates chromosomal DNA from contaminants based on differences in density	(Holben *et al.*, 1988)
Chromatography	Sephadex G50,G100 G200 Sepharose 4B	Purification via size exclusion to remove humic acids	(Eschenfeldt *et al.*, 2001; Jackson *et al.*, 1997)
Electrophoresis	Separation based on size	Used to remove humic acids due to differences in migration rate	(Chandler *et al.*, 1997)

1.4 Genome and gene enrichment

Stable isotope probing (SIP) provides the means whereby the genomes of specific metabolically active organisms in an environmental sample can be labeled and selectively isolated through density gradient centrifugation. This technology relies on the utilization of a specific labeled 'enrichment' substrate that causes the incorporation of heavier isotopes into the nucleic acid of the target microorganisms. For example, ^{13}C-labeled methanol and phenol have been used in such enrichment studies to identify novel methanol dehydrogenase gene variants belonging to Acidobacterial taxa as well as species involved in phenol degradation (Radajewski *et al.*, 2002; Manefield *et al.*, 2002). The main limitations of this technique include cross-feeding and recycling of the labeled substrate resulting in loss of specific enrichment. Another disadvantage is the commercial unavailability of labeled compounds. As an alternative, BrdUTP labeling also provides access to metabolically active organisms but lacks the same specificity as SIP (Yin *et al.*, 2000).

Suppressive Subtraction Hybridisation (SSH) can identify genetic differences between microorganisms and is therefore a powerful technique for specific gene enrichment. Adaptors are ligated to the DNA populations and subtractive hybridization is carried out to select DNA fragments unique to each DNA sample. This method has typically been applied to analyse genetic differences between two closely related bacteria (e.g., in the identification of genetic elements contributing to pathogenesis) (Bart *et al.*, 2000). It has also recently been used to identify differences between two complex DNA samples isolated from bovine rumens (Galbraith *et al.*, 2004). Clearly, these techniques might be adapted to target specific genes in related metagenomes.

For example, genes involved in the bioremediation of an environmental pollutant could be identified by the comparison of a reference metagenome with a 'perturbed' metagenome (i.e., impacted by a specific pollutant). The relatively crude nature of this approach identifies the total genetic difference between the two bacterial populations and is not specific to genes of interest or genes whose expression was up-regulated on addition of the environmental pollutant or xenobiotic.

To selectively enrich for a specific target gene within a metagenome, a more practical approach would be to use one of the differential expression technologies that rely on the isolation of mRNA to target transcriptional differences in gene expression. Several innovative methods have been developed, reviewed in Green *et al.* (2001). These techniques have so far almost exclusively been used to study patterns in eukaryotic gene expression. Differential Expression Analysis (DEA) is a particularly effective enrichment tool. This approach was successfully applied to identify bacterial genes up-regulated in the absence of iron (Bowler *et al.*, 1999). This suggests the possibility of comparing the expression profile of a metagenomic sample pre- and post-exposure to a specific substrate or xenobiotic. In this way, the expression of genes up-regulated for the specific activity can be identified.

1.5 Gene discovery technologies

1.5.1 Gene targeting: Gene specific PCR as a technique for the recovery of novel genes

The improved methods for DNA extraction from environmental sources and polymerase chain reaction have made it possible to target a specific type of gene directly from the extracted metagenome. At least four different applications using gene specific targeting in the study of environmental samples exist. These include i) the study of microbial diversity; ii) the detection of microorganisms with specific metabolic capabilities; iii) the detection of specific microorganisms (for instance pathogens) and iv) the discovery and recovery of novel genes.

By screening the metagenome of a given environmental sample, a more complete picture of the true microbial biodiversity can be obtained than through culturing alone. The most common approach is based on the amplification of a gene unique to prokaryotes, the 16S rRNA gene, using a set of universal primers that is specific to a group of organisms; *i.e.*, bacteria or archaea. The amplified products are subsequently cloned to generate up an rRNA gene library. These clones can be sequenced for further phylogentic analysis. This approach has been very widely used in the study of bacterial and archaeal diversity (Takai and Sako, 1999; Dunbar *et al.*, 1999).

Gene specific PCR has also been used to detect the presence of microorganisms with specific metabolic capabilities. This approach is especially useful in the determination of the biodegradation potential of indigenous microbial populations or for the identification of microorganisms with properties that are potentially of

commercial value. Examples of this approach include the screening for methanotrophic bacteria by targeting enzymes such as methane monooxygenase and methanol dehydrogenase (McDonald *et al*., 1995; Henckel *et al*., 2000) and the identification of chemolithotrophic ammonium-oxidizing bacteria by screening for the presence of ammonia monooxygenase genes (Sinigalliano *et al*., 2004). In another example, the biodegradation potential of an indigenous microbial population was determined by screening the specific metagenome for the presence of catechol 2,3-dioxygenase, chlorocatechol dioxygenase or phenol hydroxylase (Berthelet *et al*., 1996; Mesarch *et al*., 2000; Watanabe *et al*., 1998; Futamata *et al*., 2001). Other reported uses include the identification of denitrifying bacteria (Braker *et al*., 1998; Hallin and Lindgren, 1999) and polyhydroxyalkanoate producing bacteria (Sheu *et al*., 2000).

The detection of a specific microorganism via gene specific PCR is also of particular interest in the identification of pathogens. For example, *Vibrio hollisae* has been identified in seafood by targeting two genes presumed to be conserved within the species or genus (Vuddhakul *et al*., 2000).

Finally, gene specific PCR allows the recovery of novel genes and thereby novel enzymes directly from environmental sources that may be of significant interest to biotechnology industries. In this approach, published sequence information on the target gene is used to design degenerate primers to a conserved region within the gene. The major limitation of this approach is that, typically after amplification only a fragment of the gene is recovered rather than the complete full-length copy. To obtain the complete gene a genomic library has to be constructed and screened. The latter approach can be somewhat time consuming and several innovative strategies have been developed to

circumvent this problem. Okuta and co-workers (1998) demonstrated the recovery of novel genes encoding enzymes with the desired activity. In a method termed cassette PCR, the central fragment of catechol 2,3, dioxygenase genes was isolated from the DNA obtained from a mixed culture of phenol and crude oil degrading bacteria. The internal fragment of a previously cloned full-length copy of the catechol 2,3, dioxygenase gene was then replaced by a novel gene fragment, and in the process a partially novel hybrid catechol 2,3, dioxygenase gene was constructed.

Eschenfeldt and (2001) co-workers demonstrated the use of a semi-nested PCR strategy to isolate 2,5 Diketo – D – gluconic acid reductase genes and recovered two novel forms of the enzyme from soil DNA. Numerous other techniques have been developed to isolate DNA flanking regions. These include inverse PCR, ligation mediated PCR, panhandle PCR, universal fast walking PCR and thermal asymmetric interlaced PCR, as well as commercially available kits like the Universal Vetorette™ System (Sigma) (Ochman *et al*., 1993; Megonigal *et al*., 2000; Myrick and Gelbart, 2002; Mishra *et al*., 2002; Liu and Whittier, 1995). To the best of the author's knowledge, none of these techniques have been successfully applied on a metagenomic scale. Perhaps this reflects the technical difficulty encountered when specifically targeting low copy number genes within these tremendously complex DNA samples.

Reverse transcriptase (RT)-PCR has also been employed to recover genes from environmental samples. Using RT-PCR technology Wilson and co-workers (1999) reported the recovery of genes encoding naphthalene degrading enzymes from indigenous microorganisms present in the groundwater of a coal tar waste

contaminated site. This is a particularly useful tool in monitoring specific gene expression within an environmental sample.

An interesting and recent discovery is the existence of integrons, which may act as a reservoir for novel genes. Studies on the mechanisms involved in the acquisition of antibiotic resistance by microorganisms ultimately led to the discovery of a new type of genetic element, the integron. Integrons are naturally occurring gene capture, dissemination and expression systems that have, until recently, primarily been associated with antibiotic resistant and pathogenic bacteria [reviewed by Rowe-Magnus and Mazel (1999)]. The key structural features of an integron include a gene cassette integration site (att1), an intI gene that encodes an integrase of the tyrosine recombinase family, and two promoters which respectively drive the expression of the integrase gene and incorporated gene cassettes. The structure of integron operons has been comprehensively reviewed by Bennet (1999). Although an integron cannot initiate its own translocation, such elements has been associated with conjugative plasmids as well as transposon families Tn21 and Tn7, which can act as vehicles to facilitate integron movement within and between bacterial species (Liebert *et al.*, 1999).

The mobile element in the system is the gene cassette which is composed of a single open reading frame and a chromosomal attachment site (attC), also referred to as the 59 base element (59-be). There are considerable variations amongst gene cassettes in the length (57 to 141bp) and sequence of 59-base elements (Hall *et al.*, 1991; Collis *et al.*, 1998). Despite being essentially unique attC sites, are all characterized by the presence of an imperfect inverted repeat structure and 25 base pairs at each end of the element that conforms to a consensus sequence. Only four

bases of each consensus sequence is entirely conserved, three of which are found in what is also referred to as the inverse core site (RYYYAAC) and the core site (G↓TTRRRY, ↓ shows the point of recombination) see Rowe-Magnus and Mazel (1999). The integrase catalyze the insertion of the gene cassette into the integration site under the control of the strong promotor via site-specific recombination using att1 and attC as its substrates (Collis and Hall, 1993).

The exact mechanism whereby gene cassettes are generated remains unclear, but a hypothesis regarding the process of bacterial mRNA reverse transcription has been suggested (Recchia and Hall, 1997). It is, however, readily accepted that each bacterial species potentially has its own pool of gene cassettes that can be exchanged within, and between, species. If the extent of microbial biodiversity is taken into account this constitutes an enormous source of genetic diversity.

The view that integrons are primarily associated with antibiotic resistance has radically changed since it has become evident that integrons are widely dispersed in nature and potentially play a significant role in bacterial genome evolution (Smalla *et al.*, 2000; Rowe-Magnus and Mazel, 2001). This belief is in part due to the discovery of super-integrons (SI) and the realization that they occur widely in nature.

Super-integrons have been identified in several proteobacterial genera including *Vibrionaceae, Shewanella*, Xanthomonads and Pseudomonads (Mazel *et al.*, 1998; Rowe-Magnus *et al.*, 1999; Rowe-Magnus *et al.*, 2001). The term super-integrons was coined due to the size of the cassette clusters. In one instance, the presence of as many as 216 unidentified genes has been demonstrated (Heidelberg *et al.*, 2000).

A variety of novel genes have thus far been identified within integrons and super-integrons. In some cases, the recovered open reading frames show clear homology to the genes of known enzymes. In other instances, only hypothetical functions have been deduced for the gene cassettes. Nevertheless, it has become clear that super-integrons act as a repository of genes coding for many different enzymes. The successful use of a PCR based strategy for the direct recovery of gene cassettes from integrons has been reported. In this study, Stokes and co-workers (2001) demonstrated the recovery of several novel genes with sequences homologous to DNA glycosylase, phosphotransferase, methyl transferase and thiotransferase, as well as a number of open reading frames with no homologues in the sequence database. In this approach primers were designed to target the conserved regions within the 59 base element. These primers were then used to amplify gene cassettes directly from the extracted metagenomes of different environmental samples. The amplicons were subsequently cloned and sequenced to identify their putative function. Although this method serves to recover novel genes, it lacks the specificity required when seeking the genes of specific types of enzymes. A strategy that combines gene specific PCR and integron screening seems an obvious approach. In this method, a degenerate primer specific for the gene of interest would be combined with a flanking 59 base element primer, similar to that employed by Stokes and co-workers (2001).

Another potentially viable strategy for screening integrons is by combining direct PCR as described by Stokes *et al.* (2001) with expression screening as employed by many workers for screening multigenomic DNA libraries. The latter is discussed in more detail in the following section. In this approach the pool of amplicons generated would

be cloned and the presence of a gene of interest identified through functional expression and a suitable assay.

1.5.2 Metagenomic DNA libraries

Traditionally, genomic DNA libraries are constructed to represent the genome of a single organism (Song *et al*., 1999; Kanoh *et al*., 1998). This technique provides the means whereby an entire genome can be stored stably for further investigation. Library construction involves several basic steps: the generation of suitably sized DNA fragments, the cloning of these fragments into an appropriate vector and the screening for the desired gene. The ease with which the target gene is recovered from the library depends on the strategy followed to screen the library. Typically, the screening method can involve DNA hybridisation by probing the library with a fragment of the gene previously amplified via PCR or another homologues sequence (Kanoh *et al*., 1998). Because PCR recovers only a partial fragment of the gene, this is typically used to recover the full copy of the gene. This method is somewhat time consuming and cumbersome. In instances where the source DNA of the library is of bacterial origin, a more direct approach can be taken by using expression screening as a method to identify the gene of interest. This method is viable as bacterial genes lack introns, provided that an appropriate assay can be identified for the particular target enzyme. For example, Sullivan and co-workers (1999) reported the recovery of a lipase from a genomic DNA library constructed for *Acinetobacter calcoaceticus*. In their study they constructed a library for *Acinetobacter calcoaceticus* by partially digesting the genomic DNA with *HindIII* and cloning the generated fragments into a pUC19 vector. In order to

recover a gene encoding a lipase, a complementation assay was performed. A lipase deficient strain of *Acinetobacter calcoaceticus* was transformed with recombinant pUC19 clones. Where lipase was functionally expressed, a blue halo would be found around a clone growing on spirit blue agar. Similarly, reporter genes have also proven useful. Bereswill and co-workers (1998) isolated a specific transcriptional regulator from *Heliobacter pylori* (Fur gene). In the presence of iron this enzyme down-regulates expression. The library was constructed with pBluescript SK+ as the vector. An *E. coli* strain carrying a *lacZ* reporter gene under the control of a *fur* regulated *fhuF* promoter was transformed with recombinant clones derived from the library. In the presence of iron, a clone expressing *fur* would be expected to repress the expression of *lacZ* thereby, allowing blue/white selection screening.

The approach of screening metagenomic libraries by functional gene expression for the presence of a specific enzyme activity has been employed with some success (see Table 1-4 and Table 1-5). There have been a number of reports describing both the construction of metagenomic DNA libraries from diverse environments, and the subsequent recovery of novel genes from these libraries. Novel genes discovered this way include chitinase, 4-hydroxybutyrate dehydrogenase, enoyl-coenzyme A hydratases/isomerases, lipase, esterase, and genes encoding Na+/H+ antiporter activity (Cottrell *et al.*, 1999; Henne *et al.*, 1999; Henne *et al.*, 2000; Majernik *et al.*, 2001).

Some studies report the recovery of entire metabolic pathways through the construction of large insert libraries. Such studies use bacterial artificial chromosomes (BAC) as vector (Brady *et al.*, 2001; Courtois *et al.*, 2003). Although this approach gives significantly more coverage of the metagenome, these libraries are technically more

difficult to construct, with the isolation of good quality high molecular weight DNA being the main limiting factor (Bertrand *et al.*, 2005).

The recovery of high quality high molecular weight DNA is critical to the efficient preparation of metagenomic libraries whereas mechanical cell lysis methods generally provide a higher DNA yield and better species representation from the sample than chemical methods (see above), and the prolonged heating that causes an increase in the co-extraction of humic substances is avoided. This is especially important when extracting DNA from heavily contaminated environmental samples. There is generally a trade-off between DNA purity, DNA fragment size and yield. This may pose a problem for library construction because digesting severely degraded template DNA with restriction enzymes is impractical. To address this problem, Wilkinson and co-workers (2002) described an efficient approach for constructing environmental libraries with mechanically sheared DNA. DNA fragments larger than 1kb were recovered via electrophoresis and treated with Vent DNA polymerase to fill in the sheared ends. This was followed by 3'-adenylation with *Taq* DNA polymerase. The modified fragments were then cloned into the pCR-XL-TOPO TA cloning vector and transformed into *E.coli*. The authors reported the construction of a library representing approximately 185Mbp of cloned environmental DNA.

Low recovery rates of positive clones via functional screening (typically less than 0.01%) are a feature of all reported metagenomic screening studies. Where, the facilities are available for high throughput screening [as is the case in industry (see Table 1-5)] this may be less of a concern. However, for the average laboratory, searching for low abundance genes this poses a serious practical limitation. Gabor and

co-workers (2004) employed an *in silico* approach to investigate the factors that influence the successful recovery of genes from the metagenome through expression screening.

There are at least two key issues influencing the probability of recovering a certain gene. Firstly, the target gene must be transcribed and secondly the transcript must be translated into enzyme at a level sufficient for detection. The three modes whereby functional expression can be facilitated are through independent expression, transcriptional fusion expression and translational fusion expression.

During independent expression, the target gene is expressed from its native promoter. This requires that the transcription machinery of the surrogate host used for screening recognizes the native promoter signals and successfully initiates expression. Although *E. coli* is known to be relatively promiscuous in its ability to recognize foreign gene control elements, a definite bias towards *Firmicutes* genes have been demonstrated (Gabor *et al*., 2004). This is typically the mode of expression employed when screening BAC libraries.

Transcriptional fusion expression relies on the use of a vector based promoter to drive expression of the target gene via read-through transcription. The key factor in this approach is the size of the DNA insert. This is because the presence of expression termination signals that interrupt gene expression seems to be the main limiting factor. Gabor and co-workers (2004) demonstrated that for a library consisting of an average insert size of 2-10 Kbp, approximately 10^5 to 10^6 clones must be screened for successful recovery of a single target gene.

Gabor and co-workers also concluded that recovering genes through translational fusion expression is practically impossible, taking into account the large number of clones that would have to be screened (typically in excess of 10^7 clones).

Other factors that may influence successful screening include the need for the presence of particular transcription factors, inducers, chaperones, cofactors, protein modifying enzymes, and proper secretion machinery (Gabor *et al.*, 2004). Despite these limitations and apparent inefficiencies, the success of metagenomic DNA libraries as a tool for gene discovery have been clearly demonstrated (Rhee *et al.*, 2005; Yun *et al.*, 2004; Ranjan *et al.*, 2005).

In an on-going attempt to expedite and improve the efficiency of screening, several alternative bacterial hosts in combination with shuttle vectors are under development. These include *Rhizobium leguminosarum* (*Alphaproetobacteria*), *Bacillus* sp. (*Firmicutes*) and *Streptomyces lividans* (*Actinobacteria*) (Kaneko *et al.*, 2005; Martinez *et al.*, 2004).

Table 1-4 Characteristics of metagenomic libraries.

Examples of libraries constructed for gene targeting are shown*.

Target gene	Host/vector systems used	Library size (no. of independent clones)	Average insert size/ (Kbp)	% Prokaryote metagenome represented[d]	Reference
Chitinase	Lambda Zap II/ GigapackIII	750,000[a]	2-10	11[b,c]	(Cottrell *et al.*, 1999)
4-hydroxybutyrate dehydrogenase; lipase, esterase; Cation/H^+ antiporters	*E. coli* DH5α/ pBluescript	930000	5-8	14[c]	(Henne *et al.*, 1999; Henne *et al.*, 2000; Majernik *et al.*, 2001)
Lipase, amylase, nuclease	*E. coli* DH10B/ pBeloBAC11	3648 24576	27 44.5	0.2[c] 3[c]	(Rondon *et al.*, 2000)
Heme biosynthesis Phosphodiesterase	*E. coli* TOP10/ pCR-XL-TOPO	37000	1-10	0.5[c]	(Wilkinson *et al.*, 2002)
Polyketide biosynthesis	*E. coli, S. lividans* shuttle cosmid	5000	50	0.7	(Courtois *et al.*, 2003)
Alcohol oxidoreductase	*E. coli*/ pSK^+	583,000 360,000 324,000	4.4 3.8 3.5	5 3 2	(Knietsch *et al.*, 2003a)

* Caution is advised in attempting to directly compare metagenomic libraries made in different laboratories using different systems.

[a] Number of clones screened

[b] 1800 genomic species were estimated for an oligotrophic open ocean environment (Venter *et al.*, 2004). Due to the coastal location of the sample used in this study (Cottrell *et al.*, 1999), we are assuming a 10-fold higher species diversity.

[c] In making these calculations, an average of 10^4 prokaryotic species per environmental sample and an average prokaryotic genome size of 4Mbp was assumed .

[d] Chemical lysis methods of DNA extraction from soil samples are relatively non-aggressive and we assume that the contribution from eukaryotic (particularly fungal) genomes is minor. It is acknowledged that this assumption might be invalid.

Table 1-5 Commercialisation of metagenomic technologies[a]

Company	Target products	Classes	Products/Market	Commercial interest
BASF www.corporate.basf.com	Enzymes	Amylase Hydratase	Acidophilic glucoamylase	Food industry, aiding with the digestion of starch
Bioresearch Italia, SpA (Italy), via R.Lepetit n.34, 21040 Gerenzano 9647-4400	Anti-infectives	N.D.	Dalbavancin	Development of human gene targeted therapeutics and novel anti-infective
B.R.A.I.N www.brain-biotech.de	Bioactive peptides and enzymes for	N.D.	Nitrile hydratases Cellulases	Degussa AG Partnership for the industrial

	pharmaceuticals and agrochemicals			processe
Cubist Pharmaceuticals **http://www.cubist.com/**	Anti-infectives	N.D.	N.D.	Various commercial relationships. Variety of products in Stage I, II and III trials.
Diversa **www.diversa.com**	Enzymes	Nitrilase	Discovery of > 100 novel nitrilases.	Drug, lowering cholesterol levels
		Glycosidase Phytase	Production of Lipitor Pyrolase™ 160 & Pyrolase™ 200 Phyzyme™ XP	Broad spectrum ß-mannanase & ß-glucanase added to animal feed to break down indigestible phytate in grains and oil seeds to release digestible
		Fluorescent Protein	DiscoveryPoint™ Green-FP* and Cyan-FP*	Novel green and cyan fluorescent proteins for potential use in drug discovery, commercial screening and academic research
	Biometabolites			
Diversa & Invitrogen **www.invitrogen.com**	Enzymes	DNA polymerase	ThermalAce™ & Replicase™ DNA for research and diagnostics	Research and Diagnostics
eMetagen	Enzymes	Polyketides	eMetagen Gene	Food, agriculture,

www.emetagen.com	Antibiotics Small active molecules		and Pathway Banks™ Large clone DNA libraries encoding biosynthetic pathways for 5000 to 20,000 secondary metabolites.	research and other commercial applications Pharmaceuticals: antimicrobial, anticancer and other bioactive properties
Kosan Technology www.kosan.com	Antibiotics	Polyketides	Adriamycin, Erythromycin, Mevacor, Rapamycin, Tacrolimus (FK506), Tetracycline, Rapamycin,	Therapeutic Drugs
Genencor www.genencor.com	Enzymes	Lipase Protease	Washing powder and alkaline tolerant protease.	Cleaning industry
Libragen www.libragen.com	Antibiotics and biocatalysis for pharmaceuticals	N.D.	Anti-infective and antibiotic discovery. Biocatalysis discovery for pharmaceuticals (partnership with Synkem)	Medicine. Synthesis of pharmaceuticals
Prokaria www.prokaria.is/	Enzymes	Rhamnosidase	Food and agricultural industry	Food industry
		ß-1,6 Gluconase		Anti-phytopathogenic fungal agent
		Single stranded	Research and	

		DNA ligase	diagnostics	
Proteus www.proteus.fr	Enzymes Antibiotics Antigens	Not specified	Products for the agricultural, environmental, food, medical and chemical industries.	Development of novel biomolecules
Xanagen www.xanagen.com	Libraries	Gene products	Unspecified	Services in library construction, screening and annotation

Note: Some of the products listed above may have been derived from metagenomic libraries with prior enrichments or from single genomes
N.D. - no details available or products still under development.

1.5.3 Metagenomic cDNA (transcriptomic) libraries

cDNA libraries have primarily been employed for the investigation of eukaryotic gene expression (Rebel *et al.*, 1995; Starkey *et al.*, 1998). The translation of this technology to a metagenomic format, although feasible, does impose significant limitations and technical challenges on its successful use (Cowan *et al.*, 2005), however, recently partial sequencing of such a library has led to the identification of several novel open reading frames (Grant *et al.*, 2005). Metagenomic cDNA libraries could provide access to novel fungal genes in environmental samples, but only those that are actively transcribed at the time of RNA isolation. In a prokaryotic context, it would be possible to identify genes expressed or up-regulated in an environmental sample that may act in response to changing environmental conditions for example, the identification of genes involved in bioremediation of phenolic contamination. This approach may be

particularly useful when combined with cDNA subtractive hybridization techniques.

Two major limitations of this technology are RNA instability and transient nature of the expression profile. Both these issues can be addressed by the addition of commercial reagents like RNAlater (Sigma), to the environmental sample at the point of collection. This not only ensures the stability of the mRNA but also arrests mRNA degradation, ensuring the preservation of the expression profile at the time of sample collection. Botero and co-workers (2005) recently demonstrated the feasibility of environmental cDNA libraries by constructing such a library and targeting 16S rRNA genes from the RNA extracted from a geothermally heated soil sample.

1.5.4 Metagenome sequencing

With the availability of automated high throughput sequencing facilities, it has now become feasible to sequence metagenomes (Venter *et al.*, 2004; Tyson *et al.*, 2004). The aim has thus far been to reconstruct the genomes present within less complex microbial niches, a task that requires powerful algorithms and computational processing power. In a gene discovery context however, metagenome sequencing holds several limitations, for the worker interested in a specific class of enzyme. Apart from the cost implications it remains a shotgun approach and is therefore limited in its ability to target specific genes. Also, although numerous novel genes can undoubtedly be identified, assigning function to these open reading frames remains a significant challenge. This is evident from the large number of genes classified as conserved hypothetical,

miscellaneous and of unknown function (nearly 50%) (Venter *et al.*, 2004). Determining enzyme function from DNA sequence is a problem that will not be solved in the near future and will require innovative and powerful new bioinformatic tools.

1.6 Multicopper oxidases

Multicopper oxidases are a group of enzymes characterized by the presence of three spectroscopically different copper centers; *i.e.*, type 1 (or blue), type 2 (or normal) and type 3 (or binuclear). Each of the copper centers play a distinctive role in the catalytic mechanism of these enzymes. It is due to the spectroscopic properties of the type 1 copper center that this group has also been termed the large blue copper proteins or blue copper oxidases. Enzymes belonging to this family include Laccase (EC 1.10.3.2), primarily found in plants and fungi although several bacterial laccases have also been described, L-ascorbate oxidase (EC 1.10.3.3) found in plants and Ceruloplasmin (1.16.3.1) (or ferroxidase), present in the serum of mammals and birds (Messerschmidt and Huber, 1990; Alexandre and Zhulin, 2000). Other proteins have also been grouped into this family based on sequence similarity. These include several blood coagulation factors and bacterial copper resistance proteins (Mann *et al.*, 1988).

1.6.1 Laccase (EC 1.10.3.2)

1.6.1.1 Natural occurrence and function

Laccase was first identified in the Japanese lacquer tree *Rhus vernicifera*, where it is presumed to play a role in lignin synthesis (Thurston, 1994; Ryden and Hunt, 1993). Subsequently, this activity has been isolated from ligninolytic fungi, including *Phanerochaete chrysosporium*, *Trametes versicolor* and *Nematoloma frowardi*, as well as from the ascomycetes *Neurospora crassa* and *Podospora anserine* (Ralph and Catcheside, 1997; Cohen *et al*., 1987; Hofrichter *et al*., 1997; Claus, 2003). The enzyme is also proposed to be widespread in prokaryotes, with reports demonstrating laccase-like activity in several bacterial taxa (Claus, 2003; Alexandre and Zhulin, 2000; Givaudan *et al*., 1993; Sanchez-Amat *et al*., 2001). The enzyme is localized both extra-and intra-cellularly in plants as well as in fungi (Mayer, 1986; Rogalski *et al*., 1991). The presence of a leader signal peptide sequence in most of the putative bacterial laccases suggests a similar extracellular localization (Alexandre and Zhulin, 2000).

Laccases catalyze the single electron oxidation of a range of phenolic and non-phenolic substrates, while reducing molecular oxygen to water (Call and Mucke, 1997; Thurston, 1994). In the process, phenoxy radicals are produced which undergo further non-enzymatic reactions (Kawai *et al*., 1988). These non-enzymatic reactions include radical – radical coupling, disproportionation, deprotonation and nucleophillic attack by water resulting in polymerization, alkyl-aryl cleavage, C_α - oxidation and the demethoxylation of phenolic reductants, and

seem to play an important role in the enzymes natural function (Kawai *et al.*, 1999).

Several functions have been proposed for laccases. These include fungal morphogenesis *i.e.* pigment synthesis, cell to cell adhesion by the oxidative polymerization of cell wall components, rhizomorph formation, fruiting body development, plant pathogenesis, ligninolysis and sexual differentiation (Thurston, 1994). Possible functions for laccase - like enzymes in prokaryotes include a role in cell division, pigmentation formation, oxidation of phenolic compounds, electron transport, sporolation, Mn oxidation, UV and hydrogen peroxide resistance, ferroxidase activity, nucleoside oxidase activity, Cu resistance and phenoxazinone synthesis (Claus, 2003).

1.6.1.2 Enzyme structure and reaction mechanism

The laccase molecular structure has primarily been deduced from fungal laccases (Piontek *et al.*, 2002), although protein modeling of a putative bacterial laccase from *E. coli* (PcoA) shows a significant similarity to the structure of a laccase isolated from *Coprinus cinereus* (Figure 1.1) (Alexandre and Zhulin, 2000). These enzymes are often present as several isoenzymes differing in molecular weights (50 to 110kDa) and isoelectric points, and can form multimeric complexes. Depending on the source organism, the optimum pH ranges between 2.5 and 7.5, with an optimum temperature for activity of 50 to 60 ℃ (Molitoris *et al.*, 1972; Saparrat *et al.*, 2002; Taniguchi *et al.*, 1982).

A further distinctive feature of fungal laccases is the presence of covalently linked carbohydrate moieties that contributes as much as 45% of the total molecular weight. This post-translational modification functions to stabilize the enzyme in the extracellular environment (Claus, 2003).

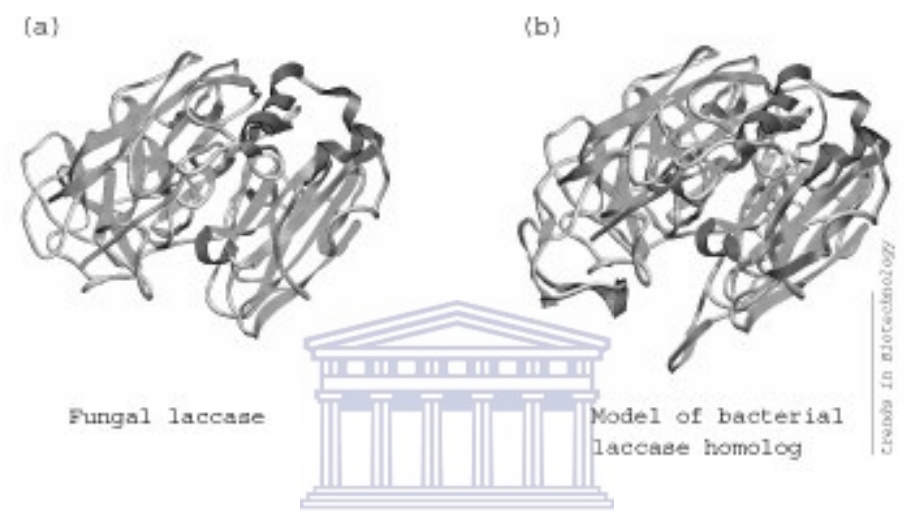

Figure 1.1 Structural comparison of a Fungal laccase and a bacterial laccase homolog. a) *Coprinus cinereus* laccase structure .b) *Escherichia coli* laccase homolog PcoA. Figure was obtained from Alexandre and Zhulin (2000).

The laccase copper binding site consists of four copper ions. Depending on spectroscopic and paramagnetic characteristics, these ions are classified as one of three types, (Messerschmidt, 1997). Type 1 copper is primarily involved in the single electron substrate oxidation. Type 2 copper functions to reoxidize the type 1 copper ion as well as to stabilize the hydrogen peroxide intermediate. In the final step, type 3 copper reduces molecular oxygen to water through the transfer of four electrons (Claus, 2003).

The type 1 copper ion reacts with the reducing substrate, whereas the three remaining copper ions belonging to types 2 and 3 form a tri-nuclear copper cluster where oxygen is reduced to water. In the catalytic mechanism, the reducing and oxidizing substrates bind sequentially to the active site, with the products being immediately released after the electron transfer has occurred. During the interaction, the type 1 ion accepts an electron from the reducing substrate and then passes it on to the tri-nuclear cluster where the dioxygen molecule is eventually reduced to water (Mester and Tien, 2000).

Due to the low redox potential of laccase, it is unable to directly oxidize non-phenolic substances. There have however, been several reports on the use of substrates which can act as mediators for the oxidation of non-phenolic aromatic substances including for example, ABTS (2,2'-azino-bis-(3-ethylebenzo-thiazole-6-sulfonic acid), 3-HAA (3-hydroxyanthranilate) and violuric acid (Fabbrini *et al.*, 2002; Kawai *et al.*, 1988). Once oxidized by the enzyme, these mediators can go on to oxidize non-phenolic substances of higher reducing potential or other larger substrates that are not accessible to the enzymes active site due to their size.

1.6.1.3 Sequence characteristics of laccase

The fungal laccase family has been comprehensively studied (Kumar et al., 2003; Thurston, 1994). The majority of the published laccase sequence information is derived from this group and has shown that laccases average in length between 500 and 600 amino acids. A similar size range has also been demonstrated for bacterial laccases, with an average between 450 and 600

amino acids (Alexandre and Zhulin, 2000). Multiple sequence alignments of fungal laccase protein sequences have identified the presence of four highly conserved histidine rich domains responsible for copper binding in the enzymes active site (Thurston, 1994). Relative to the alignment ruler these domains are located at positions 140 to 143 (HWHG), 185 to 187 (HSH), 493 to 500 (HPXHLHGH) and 558 to 564 (HCHX3H) (Figure 1.2)

These copper binding domains have also been identified in bacterial laccases (see chapter 5 for in depth discussion).

Kumar and co-workers (2003) identified four signature sequences (designated L1 to L4) common to fungal laccases and encompassing the copper binding domains (Table 1-6). A remarkable similarity exists between regions within signatures L1 (HWHG) and L3 (HLHG) as well as between L2 (HSH) and L4 (HCH). This and the fact that signatures L1 and L2 and L3 and L4 are localized in different domains within the laccase tertiary two domain structure, has been taken as evidence for a possible gene duplication event occurring during the evolution of this family of genes (Kumar et al., 2003).

Table 1-6 Fungal laccase signature sequences

Nomenclature	Position[a]	Observed signature sequences
L1	140	H-W-H-G-X_9-D-G-X_5-QCPI
L2	180	G-T-X-W-Y-H-S-H-X_3-Q-Y-C-X-D-G-L-X-G-X-(FLIM)
L3	493	H-P-X-H-L-H-G
L4	550	G-(PA)-W-X-(LFV)-HCHI-DAE-X-H-X_3-G-(LMF)-X_3-(LMF)

a Position according to alignment ruler in Figure 1.2

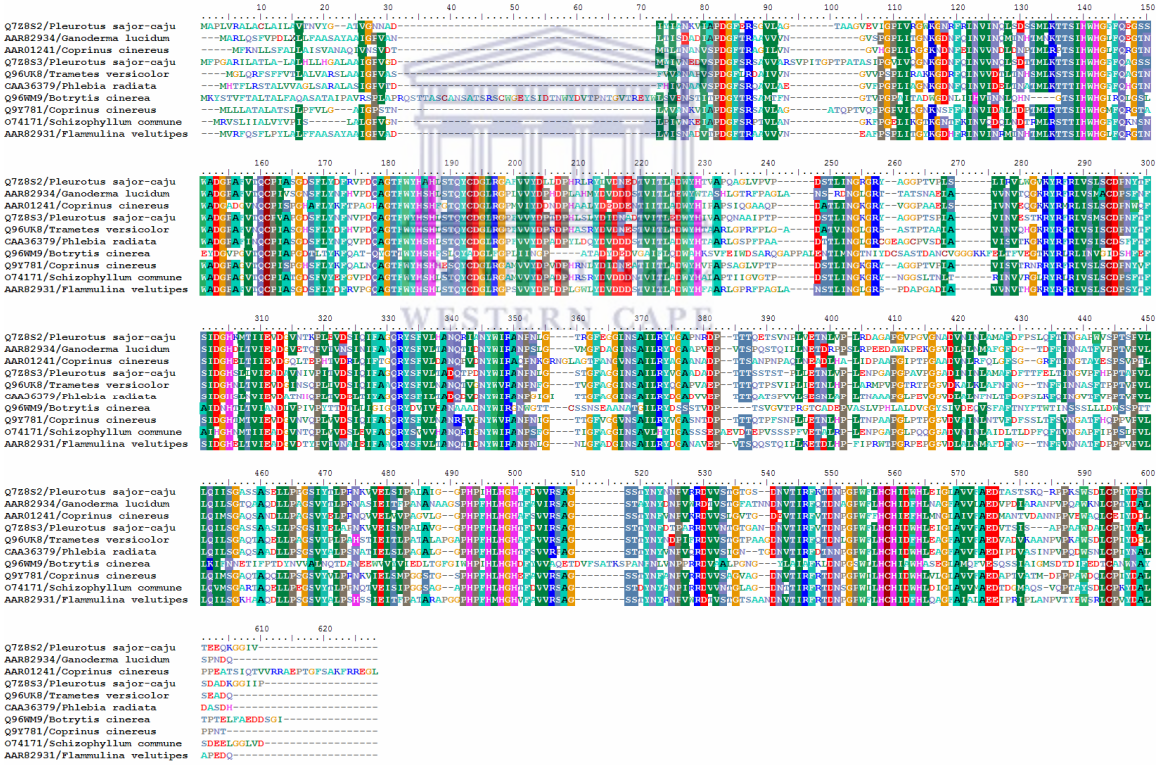

Figure 1.2 Multiple sequence alignment of fungal laccase protein sequences. Genbank accession numbers are listed.

1.7 Aims and strategy of this study

The general aim of this work was to develop metagenomic methods for the culture-independent discovery of novel genes and enzymes. For the purposes of this study, multicopper oxidases was selected as a suitable target, based on the projected diversity of genes. Specific aims were:

- To extract metagenomic DNA from two organic rich environmental samples and to determine the microbial biodiversity therein.
    - Compare three different metagenomic DNA extraction protocols with regard to DNA yield, size and purity.
    - Detect the presence of bacterial DNA by 16SrRNA gene detection.
    - Detect the presence of eukaryotic DNA by 18SrRNA gene detection.
    - Demonstrate the level of microbial biodiversity by DGGE.

This work aimed to prepare high quality DNA from environmental samples known to be heavily contaminated with phenolics.

- To construct both a small and large insert metagenomic DNA library from extracted environmental DNA and to screen for the presence of novel bacterial laccases by expression screening.
    - Construct a small insert metagenomic library (2-10Kbp) using the commercially available pCR® T7 TOPO TA Expression kit.
    - Construct a large insert metagenomic library (>20Kbp) using the commercially available CopyControl™ Fosmid library production kit.

- Screen for the functional expression of laccase with a colorimetric agar plate assay.

This work investigated the feasibility of metagenomic DNA libraries as a means to recover novel bacterial laccase genes from compost and grape waste environmental DNA.

- To isolate novel bacterial multicopper oxidases from metagenomic DNA samples through gene specific PCR.
  - Design multicopper oxidase specific degenerate primers and verify specificity.
  - Probe metagenomic DNA samples for novel multicopper oxidase gene fragments
  - Recover full- length genes by comparing the commercial Vectorette™ system and thermal asymmetric interlaced PCR for efficiency.

This work aimed to develop a bacterial multicopper oxidase gene specific PCR assay and to investigate the occurrence of bacterial multicopper oxidase genes in compost and grape waste environmental samples. The feasibility of two PCR based techniques for the recovery of full-length ORFs from the metagenome were also investigated.

- To develop a novel gene specific subtractive hybridization technique for the efficient recovery of bacterial genes from metagenomic DNA sample.

- Use *Streptomyces* multicopper oxidase positive control genomic DNA in a proof of concept experiment with biotin labeled driver (*MCO* gene fragment) in a magnetic bead capture subtractive hybridization approach.

A subtractive magnetic bead capture protocol was adapted to develop a more efficient technique for the recovery of full-length bacterial genes. As a proof of concept a multicopper oxidase ORF was recovered from a Streptomyces strain.

Chapter 2: Materials And Methods

2.1 Materials

The suppliers of materials used in this study are given in Table 2.1.

Table 2-1 Materials and suppliers

Supplier	Reagent
Promega	Restriction enzymes, JM109 (DE), pGEM®T-Easy vector system, dNTPs, LM-SIEVE Agarose
Fermentas	T4 DNA ligase, InsT/Aclone ™ PCR product cloning kit
Invitrogen	pCR® T7 TOPO® TA Expression kit
Amersham	GFX PCR DNA and Gel band purification kit, GFX Micro plasmid prep kit
Qiagen	QIAEX II Gel extraction system
Sigma	Chemical reagents, Universal Vectorette™ system
Gibco	Fungizone
Bioline	BIO-X-ACT™ DNA polymerase, Agarose
Roche	Shrimp Alkaline Phosphatase, DIG DNA labeling and Detection kit, Positively charged Nylon membranes
IDT	Oligonucleotide primers
BIO-RAD	Gene Pulser® cuvette
TaKaRa	T4 DNA polymerase
Dynal	Dynabeads® kilobaseBINDER™ kit
Merck	Chemical reagents
Pharmacia	Sephadex G100
Epicentre®	CopyControl™ Fosmid library production kit

2.2 Bacterial strains and plasmids

Bacterial strains and plasmid vectors used in this study are listed in Tables 2.2 and 2.3.

Table 2-2 Bacterial strains used in this study

Strain	Genotype	Use in this study	Source
E.coli strains			
XL1- Blue	*rec*A1 *end*A1 *gyr*A96 *thi-1 hsd*R17 *sup*E44 *rel*A1 *lac* [F´ *pro*AB *lacI*^{q}*Z*Δ M15 T*n10* (Tet^{r})]	Routine cloning	Stratagene
Top 10F'	F'[*lacI*^{q} Tn*10*(Tet^{R})] *mrc*A Δ(*mrr-hsd*RMS-*mcr*BC) phi80 *lacIZΔM15Δlac X74* *rec*A1 *deo*R *ara*D139 Δ*(ara-leu)*7697 *gal*U *gal*K *rps*L(Str^{R}) *end*A1 *nup*G	Metagenomic library construction	Invitrogen
EP300™-T1^{R}	[F^{-}mcrA Δ(*mrr-hsd*RMS-*mcr*BC)) phi80 *lacIZΔM15Δlac X74* *rec*A1 *end*A1 *ara*D139 Δ (*ara,leu*)7697 *gal*U *gal*K λ^{-} *rps*L *nup*G *trf*A *ton*A *dhfr*]	Metagenomic library construction	Epicentre®

Table 2-3 Plasmid vectors used in this study

Plasmid	Description	Use in this study	Source
pGEM –T- Easy	Size 3015 bp, T7 promoter, SP6 promoter, Amp^r, *lac* operator, *Lac*Z start codon, phage f1 region, pUC M13 priming sites, 3' - T overhangs	Routine cloning of PCR products	* Promega
pCR® T7/CT-TOPO®	Size 2702 bp, Amp^r, T7 promoter, V5 epitope, Zeocin^r, polyhistidine region, pUC origin, TOPO® cloning site	Metagenomic library construction, Routine cloning of PCR products	* Invitrogen
pTZ57R/T	Size 2888 bp, *lac* operator, *Lac*Z start codon, phage f1 region, pUC M13 priming sites, 3' - T overhangs, rep (pMB1), Amp^r	Routine cloning of PCR products	* Fermentas
pCC1FOS™	Size 8139 bp, lacZ ,Chi^R, cos, loxP, ParA, B and C, oriV	Metagenomic library construction	* Epicentre®

2.3 Analytical procedures

2.3.1 Spectrophotometry

DNA concentration was quantified spectrophotometrically as described by Sambrook and Russell (2001) using a Nanodrop ND-1000 spectrophotometer. Absorbance measurements were taken at 260 nm and 280nm and used to calculate the DNA sample concentration and purity. For double stranded DNA an optical density reading of 1 at A_{260} = 50ng/µl. DNA samples were considered largely free from protein contamination if the A_{260} to A_{280} ratio was between 1.5 and 2.0.

2.3.2 Agarose gel electrophoresis

Preparative and analytical agarose (Bioline) gels were prepared in 0.5X TBE (45mM Tris-HCl, 45mM boric acid, 1mM EDTA, pH 8) with ethidium bromide at a final concentration of 0.5µg/ml and electrophoresis was performed as described by Sambrook and Russell (2001).

2.4 Molecular biology techniques

2.4.1 Restriction endonuclease digestions

Restriction endonuclease digestions were performed for the preparation of DNA fragments during cloning procedures or as a method to characterize newly constructed recombinant DNA clones. Manufacturers (Promega) specifications

were followed where appropriate. Digestion products were typically analyzed on 1% agarose gels, with the applicable controls and molecular weight size markers.

2.4.2 Polymerase chain reaction

Unless otherwise indicated polymerase chain reaction amplification conditions were as follows. The polymerase chain reaction (50µl) included 5 µl 10X buffer (200mM Tris pH 8.8, 100mM KCl, 100mM $(NH_4)_2SO_4$, 20mM $MgSO_4$, 1% (v/v) Triton X-100), primers and dNTPs used at a final concentration of 2µM and 200µM respectively, 1U *Taq* polymerase and 20ng DNA template. PCR cycling conditions were as follows, 94℃, 3 min; 30 cycles (94℃, 30 sec, *℃, 1 min, 72℃, 1 min); 72℃, 5 min.

* For specific primer annealing temperatures, see Table 2-4

2.4.3 DNA sequencing and analysis

All sequencing was done with a MegaBACE 500 Automated Capillary DNA Sequencing System (Amersham Biosciences) by the sequencing unit at the Molecular and Cell Biology Department of the University of Cape Town. Plasmid DNA was purified using the GFX Micro plasmid prep kit (Amersham) and 1µg plasmid DNA was sent for sequencing. The primers used for sequencing are listed in Table 2-4. Sequence data were analysed using the BioEdit (version 5.0.9) freeware application (Hall, 1999).

2.4.4 DNA Ligation

Ligation reactions (15µl reaction volume) were set up according to the specifications provided by the manufacturer (Fermentas), with 1 unit T4 DNA

ligase in the appropriate ligation buffer, at 16°C for sticky end or 22°C for blunt end ligations.

2.4.5 Transformation of *E.coli* cells by electroporation

All *E.coli* strains used for transformation were prepared for electroporation as described by Sambrook and Russell (2001). For electroporation, an appropriate amount of ligation mix was added to the competent cells and kept at 4°C. The mixture was transferred to a chilled BIO –RAD gene pulser® cuvette (0.1 cm electrode). The electroporation conditions were as follows; volts 1.8KV, resistance 200Ω and capacitance 25µFD. Pre-warmed TB media (1.2% w/v tryptone, 2.4% w/v yeast extract, 0.4% v/v glycerol) was added to the transformation mixture and incubated at 37°C for one hour with agitation. The transformation mixture was plated out on LB plates containing the appropriate amount of antibiotic, X-GAL and IPTG as required.

2.4.6 Recovery of DNA fragments from agarose gels

During all cloning procedures, DNA fragments were recovered from agarose gels using the Geneclean™ III kit (Bio101). Excised gel slices containing DNA were melted at 45-55°C in 2.5 volumes of 3M NaI. After the addition of 5µl ice cold GLASSMILK® suspension, the DNA binding was facilitated by gentle agitation for 15 min. The mix was then incubated on ice for a further 3 min. After pelleting the GLASSMILK® and bound DNA, the pellet was washed three times with 500µl New Wash (Bio101). The DNA was eluted twice in ddH_2O at 45-55°C

for 8 min to a final volume of 16µl. Alternatively a GFX PCR DNA and Gel band purification kit was used as per the manufacturers instructions.

2.4.7 Plasmid DNA isolation and purification

The alkaline lysis method (Sambrook and Russell, 2001), first described by Birnboim and Doly (1979), was used to extract plasmid DNA. Selected colonies were used to inoculate 3ml LB broth supplemented with the appropriate antibiotics and grown overnight. Cells were harvested by centrifugation at 10000g for 1 min. The pellets were resuspended in 100µl of solution 1(50mM glucose, 10mM EDTA, 25mM Tris (pH 8.0)) and incubated at room temperature for 5 min. Complete lysis of the cells and denaturation of DNA was achieved by adding 200µl solution 2 (0.2M NaOH, 1% SDS) and placing the tubes on ice for 5 min. Genomic DNA, RNA and protein was precipitated by the addition of 150µl 3M NaAc (pH 4.8). After 10 min on ice, cell debris and precipitates were pelleted by centrifugation at 15000 rpm (10 min, 4°C). The DNA was precipitated from the supernatant in two volumes of 96% ethanol for 30 min. After the final centrifugation and 80% ethanol wash step the DNA was resuspended in ddH_2O.

2.4.8 Large scale bacterial genomic DNA extraction

For the extraction of bacterial genomic DNA from pure cultures, the hot CTAB method was used (Sambrook and Russell, 2001). Cultures were grown overnight in 100ml of the appropriate growth medium and harvested at 6000 rpm for 10 min. The pellet was resuspended in 9.5ml TE buffer (10mM Tris-HCl pH 7.4, 1mM EDTA) and 0.5ml of a 20% SDS solution together with 50µl proteinase

K (20mg/ml) was added and left to incubate at 37℃ for one hour. 1.8ml of a 5M NaCl solution was then added and the tube was inverted gently. The tube was left to incubate at 65°C for 20 min after the addition of 1.5ml of a 1% w/v CTAB solution (made up in 0.7M NaCl). The cell debris was pelleted at 5000 rpm for 10 min. Contaminating proteins were removed by extracting the supernatant twice with an equal volume of chloroform/isoamyl alcohol at 7000 rpm for 10 min. The genomic DNA was precipitated from the aqueous phase with a standard isopropanol precipitation step (Sambrook and Russell, 2001). The DNA pellet was resuspended in 500µl ddH_2O.

Table 2-4 Oligonucleotide primers used in this study

Primer name	Sequence	Tm	Use in this study
Lac1F	GGGCCACGAAACCAGCNTNCAYTGGCA	64	Amplification of multicopper protein gene fragments
Lac1R	AACCACATGCCGTGCAGRTGNATNGGRTG	64	
Lac1F-5'-Bio	Biotin- GGGCCACGAAACCAGCNTNCAYTGGCA	64	Driver production
G4NF1	CTAGTCTAGAATATGATGACGGCCGGGAGG	63	Recovery of p*MCO*7 3' flanking region
G4NF2	CTAGTCTAGAACCCTCAACTATGCCATGCTG	63	
G4NF3	CTAGTCTAGACAAAGCCGGTGAGAATATCCG	63	
G4NR3	CTAGTCTAGATCGTACTTGTGCGTGCTGTG	63	Recovery of p*MCO*7 5' flanking region
G4NR2	CTAGTCTAGACGTGCATCGCGATCATAGATC	63	
G4NR1	CTAGTCTAGAGATCATGTACCAGTCGGTGG	64	
G7NF1	AGCACGGCACCCACATGTA	67	Recovery of p*MCO*8 3' flanking region
G7NF2	GCCATCCTGCTGCACAACT	67	
G7NF3	TCAACAGCAAGGTCTTCCCG	67	
G7NR1	CCACATCGACAGATTGCCG	67	Recovery of p*MCO*8 5' flanking region
G7NR2	TGAACATGCCCATCATTCCC	67	
G7NR3	GCATAGGTTTCACCCGGCT	67	
AD1	TGWGNAGWANCASAGA	44	Arbitrary degenerate primers for random priming in semi-nested random primed PCR
AD2	AGWGNAGWANCAWAGG	44	
AD3	CAWCGICNGAIASGAA	53	
AD4	TCSTICGNACITWGGA	55	
AD5	CGCGGATCCNNNNNNNNN	58	
NAD1	SWGANAWGAA	40	
NAD2	GTNCGASWCANAWGTT	40	
NAD3	WGTGNAGWANCANAGA	40	
NAD4	TGWGNAGWANCASAGA	40	
pUC/M13F	GTTTTCCCAGTCACGAC	55	Sequencing
pUC/M13R	CAGGAAACAGCTATGAC	55	Sequencing
U1510R	GGTTACCTTGTTACGACTT	50	16S rDNA PCR
E9F	GAGTTTGATCCTGGCTCAG	50	16S rDNA PCR
18SUR	CYGCAGGTTCACCTACRG	50	18S rDNA PCR

18SUF	CTGGTTGATYCTGCCAGT	50	18S rDNA PCR
341FGC	CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG CCTACGGGAGGCAGCAG	TD	DGGE bacteria
534r	ATTACCGCGGCTGCTGG	TD	DGGE bacteria
NS7GC	CGC CCG GGG CGC GCC CCG GGC GGG GCG GGG GCA CGG GGGGAGGCAATAACAGGTCTGTGATGC	TD	DGGE Fungi
F1Ra	CTTTTACTTCCTCTAAATGACC	TD	DGGE Fungi
T7 Adapter F	TTCTAATACGACTCACTATAGGACGCACCGATA CCGATACCGTTACTCGAT	n/a	Adaptor for priming site introduction and re-amplification
T7 Adapter R	TCGAGTAACGGTATCGGTGC	n/a	
T7 (short)	AATACGACTCACTATAGG	54	

(TD) Touchdown

2.5 Metagenomic DNA extraction from phenolic rich environmental samples

2.5.1 Environmental soil sample collection

Compost derived from plant matter was purchased from a commercial compost vendor. Decomposed grape waste was collected from the wine land district in Stellenbosch, Cape Town, South Africa. The environmental samples were stored at -80℃ until further use.

2.5.2 Direct DNA extraction

2.5.2.1 Chemical method (Zhou)

Genomic DNA was extracted from environmental samples as described by Zhou (1996), followed by a SephadexG100/PVPP purification step as described in section 2.4.3.

2.5.2.2 Bead beating method (modified Miller method)

The extraction protocol is a modification of the Miller protocol (Miller *et al.*, 1999). Up to 0.5g of sample was added to a beat-beater vial containing 0.5g white quartz sand (Sigma) and resuspended in 300µl phosphate buffer (100mM NaH_2PO_4 pH 8). Equal volumes of SDS lysis buffer (100mM NaCl, 500mM Tris-HCl pH 8, 10% SDS) and chloroform – isoamyl alcohol were added and the suspension was gently mixed. The vials were shaken at 4.5 m/s for 40 sec in a FastPrep FP120 machine (B101). The samples were then centrifuged at 13000

rpm for 5 min to pellet cell debris and soil particles. The supernatant (650µl) was transferred to a fresh tube and 360µl 7M NH_4OAc added. The tube was inverted several times and re-centrifuged at 13000 rpm for 5 min. The supernatant (580µl) was again transferred to a fresh tube and the genomic DNA precipitated with 315µl isopropanol for 15 min at room temperature. The DNA was recovered by centrifugation at 13000 rpm for 5 min. The DNA was finally washed with 70% ethanol, air dried and resuspended in 100µl ddH_2O.

2.5.2.3 Chemical/ enzymatic/mechanical cell lyses (Lamontagne method)

Genomic DNA was extracted from environmental samples as described by LaMontagne (2002).

2.5.3 Sephadex G100/PVPP column purification

To remove humic acid contamination from environmental DNA samples, a slightly modified Sephadex G100/PVPP method was employed (Stach *et al*., 2001). The end of a 1ml syringe was plugged with glass wool and packed with a 50% Sephadex G100 slurry (Pharmacia). Once the matrix had settled to approximately 400µl, a 10% (w/v) PVPP solution was added to a final volume of 800µl. Up to 100 µl of sample was loaded and the column was placed inside a 15ml falcon tube and centrifuged for 3 min at 1500 rpm in an EppendORF 5810R bench top centrifuge. The DNA was precipitated using a standard ethanol DNA precipitation method (Sambrook and Russell, 2001).

2.5.4 Microbial diversity analysis

2.5.4.1 Denaturing gradient gel electrophoresis (DGGE)

To estimate the bacterial and fungal diversity present in the extracted metagenomes, denaturing gradient gel electrophoresis (DGGE) was employed as described by Watanabe (Watanabe *et al*., 2001) in a SciPlas V20-HCDC system (BIO-RAD). A 9% polyacrylamide gel with a 30% - 60% urea/formamide gradient was cast (BIO-RAD 475 gradient former), 100% UF = 7M Urea, 40% formamide. PCR reactions were electrophoresed at 60℃ for 16H at 100V in 0.5X TAE (20mM Tris-HCl, 2.9% v/v Glacial acetic acid, 1mM $Na_2EDTA.2H_2O$ pH 8.5) and then stained for 10 min in a 0.5µg/ml ethidium bromide solution. Standard PCR reactions were set up as described in section 2.3.2 (see Table 2.4 for primer information). PCR cycling conditions were as follows,

18S rDNA cycling conditions

94℃, 4 min; 10 cycles (94℃, 30 sec, 60℃*, 30 sec, 72℃, 1.5 min); 30 cycles (94℃, 30 sec, 50℃, 30 sec, 72℃, 1.5 min); 72℃, 10 min.

*1℃ decrease/cycle

16S rDNA cycling conditions

95℃, 1 min; 94℃, 4 min; 20 cycles (94℃, 45 sec, 65℃*, 45 sec, 72℃, 2 min); 20 cycles (94℃, 30 sec, 55℃, 30 sec, 72℃, 2 min); 72℃, 10 min.

*0.5℃ decrease/cycle.

2.6 Metagenomic DNA library construction and expression screening

2.6.1 Metagenomic DNA preparation

2.6.1.1 Blunt end polishing of DNA ends

Genomic DNA fragments were treated with T4 DNA polymerase (TaKaRa) in order to remove any sheared ends generated during the beat beating extraction procedure. Polishing of the sheared ends was done in two steps. Firstly, the 3' to 5' exonuclease activity of T4 DNA polymerase was used to remove any 3' overhangs. The reaction was set up to a final volume of 50µl and included the following; approximately 5µg DNA, 5µl 10X T4 DNA polymerase buffer (330mM Tris-acetate pH 7.9, 660mM potassium acetate, 100mM magnesium acetate, 5mM DTT), 0.01% BSA and T4 DNA polymerase (1U). The reaction was incubated for 30 min at 37°C.

DNA polymerase 5' to 3' polymerization activity was used to fill in 5' overhangs. The reaction mixture was adjusted as follows; volumes were adjusted to 60µl, a dNTP mix was added to a final concentration of 30µM, the buffer and BSA concentrations were adjusted appropriately. The reaction was left to incubate for a further 30 min after which the DNA was recovered from the reaction mixture via the GFX DNA and gel band purification kit (Amersham); DNA was eluted in 50µl ddH_2O.

2.6.1.2 3' A-tailing of blunt ended DNA

The terminal transferase activity of *Taq* DNA polymerase was used for the addition 3' A- overhangs. The reaction mixture was made up to a final volume of 60µl and included 44µl of DNA (see section 2.5.1.1), 10X NEB PCR buffer (0.2M Tris-HCl pH 8.8, 0.1M KCl, 0.1M $(NH_4)_2SO_4$, 20mM $MgSO_4$, 1% v/v Triton X-100), 300µM dATP, 80µM dNTP mix and 1.5 U *Taq* polymerase. The reaction mixture was incubated at 72°C for 30 min. The 3' A-tailed DNA was recovered from the reaction using the GFX DNA and gel band purification kit (Amersham), DNA was eluted in 12µl ddH_2O.

2.6.1.3 Dephosphorylation of metagenomic DNA

Dephosphorylation of the 5' ends is a critical step in the preparation of genomic DNA for efficient cloning into the pCR®TOPO® TA expression vector (Shuman, 1994). The dephosphorylation reaction was set up to a final volume of 10µl. This included 6µl 3' – A tailed DNA, 1µl 10X shrimp alkaline phosphatase buffer (0.5M Tris-HCl, 50mM $MgCl_2$ pH 8.5) and 1U shrimp alkaline phosphatase (SAP). The reaction was incubated at 37°C for 60 min after which SAP was inactivated by incubation at 65°C for 20 min.

2.6.1.4 Cloning in pCR®TOPO® TA Expression plasmid

The ligation reaction was set up as described in the pCR®TOPO® TA Expression kit instruction manual (Invitrogen). Dephosphorylated DNA (4µl) was added to 1µl (10ng) pCR®T7/CT -TOPO® vector and 1µl salt solution (300mM

NaCl, 15mM $MgCl_2$). The ligation mix (2µl) was used to transform 40µl One Shot® TOP10F' electrocompetent *E. coli* cells (transformation efficiency 10^9 CFU/1µg control plasmid DNA). Electroporation was performed using a BIO-RAD Gene Pulser™. The transformation mix was added to a chilled Gene Pulser® cuvette (0.1 cm electrode gap). The electroporation conditions were as follows; 1.5Kv, 200Ω and 25µFD. After electroporation, 960µl warm TB (1.2% w/v tryptone, 2.4% w/v yeast extract, 0.4% v/v glycerol) was added and the cells were allowed to recover by incubation at 37°C for one hour.

2.6.2 Library amplification and storage

The transformed cells were pooled and the volume was adjusted to 50ml with TB medium. Ampicillin (100 µg/ml) and Tetracycline (12.5 µg/ml) was added to the appropriate concentrations. The cells were then incubated for 4 hours at 37°C in a shaking incubator. In order to determine the number of amplifications, a dilution series was set up using 10µl of the cell suspension, plated on LB – ampicillin agar plates and incubated overnight at 37°. The cell suspension was divided appropriately and individual copies of the library were stored at -80 °C.

2.6.3 Fosmid library production

High molecular weight metagenomic DNA was extracted as described in section 2.5.2.1. Approximately 10µg DNA was prepared for blunt end cloning, gel purified and cloned into the pCC1 FOS™ vector according to the manufacturers instructions (Epicentre®).

2.6.4 Expression screening

2.6.4.1 Laccase assay

Metagenomic DNA libraries were screened for laccase activity by adapting the chromogenic oxidative coupling reaction between ABTS and *p*-hydroxybenzoic acid (Shin *et al.*, 1987) to a solid agar support. The composition of laccase screening plates was as follows, LB agar plates containing ampicillin (100µg/ml) or chloramphenicol (12.5µg/ml), 0.1% (w/v) ABTS and 0.1% *p*-hydroxybenzoic acid. The pH was adjusted to 6.

2.7 Sequence specific approach – Semi nested random PCR

2.7.1 Degenerate primer design

Fungal laccase gene sequences were obtained from the BRENDA enzyme database (**http://www.brenda.uni-koeln.de**) and combined with the sequences of putative bacterial laccases identified by Alexandre and Zhulin (2000). These sequences were aligned using ClustalW (Thompson *et al.*, 1994) and targets for degenerate primer design were identified. Degenerate primers (listed in Table 2-4) were designed to two of these target sites using the online COnsensus DEgenerate Hybrid Oligonucleotide Primers (CODEHOP) program (Rose *et al.*, 1998). The choice of program parameters included a maximum degeneracy of 128 - fold and an optimum annealing temperature of 64°C. All other program parameters were set at the default settings. The polymerase chain reaction included; 10X NEB buffer (200mM Tris -HCl pH 8.8, 100mM KCl, 100mM $(NH_4)_2SO_4$, 20mM $MgSO_4$, 1% (v/v) Triton X-100), primers and dNTPs

used at a final concentration of 2µM and 200µM, respectively, 1U *Taq* polymerase and 20ng DNA template. A touchdown PCR was done and the cycling conditions were as follows; 94°C, 3 min; 5 cycles (94°C, 30 sec, 64°C, 1 min, 72°C, 1 min); 16 cycles (94°C, 30 sec, 64°C (1°C decrease/cycle, 1 min), 72°C, 1 min); 10 cycles (94°C, 30 sec, 50°C, 1 min, 72°C, 1 min); 72°C, 5 min.

2.7.2 In-gel ligation

In instances where the PCR product concentration was too low to allow the use of conventional DNA purification methods prior to cloning, an in-gel ligation was performed. The agarose gel was prepared by melting 1g low melting point agarose (Promega) in 100ml 1X TA (40mM Tris-HCl, 5.7% v/v Glacial acetic acid pH 8.5) buffer, with ethidium bromide added to a final concentration of 0.05µg/ml**.** PCR products were separated by agarose gel electrophoresis in 1X TA buffer at the appropriate voltage. The DNA band was excised from the gel and melted at 55°C. Ligation reactions included 10X T4 DNA ligation buffer (Roche) and 1U T4 DNA ligase. The appropriate volume of DNA in the melted gel and vector was added to a final volume of 10µl. The ligation reaction was incubated at 16°C overnight. The ligation reaction was melted at 55°C prior to transformation and 2µl was typically transformed.

2.7.3 Thermal asymmetric interlaced PCR (TAIL-PCR)

For the recovery of 5' and 3' flanking regions of target ORFs, a thermal asymmetric interlaced PCR (TAIL–PCR) was performed (Liu and Whittier, 1995). This typically involved three rounds of amplification using nested gene specific

primers combined with a set of arbitrary degenerate primers. The product of each reaction served as the template in the successive reaction. Due to the potentially large size of the amplicons (>2kbp) the BIO-X-ACT ™ long DNA polymerase system was used (Bioline). The final concentrations of the PCR reaction components were as follows; 1X OptiBuffer™, 2.5mM MgCl2, 0.2mM dNTP mix, 0.4U BIO-X-ACT™- long polymerase, 0.5µM gene specific primer, 0.5µM arbitrary degenerate primer and 1µl template (a 1:50 dilution of the previous PCR product). Reactions were taken up to a final volume of 20µl with ddH2O. Polymerase chain reactions were performed in a GeneAmp® PCR system 2700, with the cycling conditions slightly modified from those described by Liu and Whittier (1995) as shown in Table 2-5.

Table 2-5 A summary of TAIL-PCR cycling conditions

Reaction	No. of cycles	Thermal conditions
1st Round	1	94℃(3')
	10	94℃(15''), *℃ (1'), 72℃(3')
	20	94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), 40℃ (1'), 72℃ (3')
	1	72℃ (5')
2nd Round	1	94℃(3')
	10	94℃(5''), *℃ (1'), 72℃ (3')
	20	94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), 40℃ (1'), 72℃ (3')
	1	72℃ (5')
	1	94℃(3')
3rd Round	10	94℃(5''), *℃ (1'), 72℃ (3')
	20	94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), *℃ (1'), 72℃ (3')
		94℃(5''), 40℃ (1'), 72℃ (3')
	1	72℃ (5')

* See Table 2-4 for the annealing temperature of the gene specific primers.

2.7.4 Universal Vectorette™ system

High molecular weight metagenomic DNA extracted as described in section 2.5.2.1 was used for the construction of a metagenomic Vetorette™ library. The protocol for library construction was followed as described in the manufacturers instruction manual (SIGMA).

2.8 Sequence specific approach – Gene specific subtractive hybridization

2.8.1 Adaptor oligonucleotide reconstruction and ligation

The adapter oligonucleotide was produced by re-annealing two overlapping oligonucleotide primers (see Table 2-4) in a 10X annealing buffer (1M Tris-HCl pH 7.5, 5M NaCl, 0.5M EDTA), 200 nmole of each oligonucleotide was added and the volume adjusted to 50µl. The reaction was heated to 94°C for 5 min, then kept at 65°C for 10 min before being allowed to cool to 25°C for 1 to 2 hours. The reconstructed adapter oligonucleotide was stored at -20°C until further use.

2.8.2 Genomic tester DNA preparation

Genomic DNA (10µg) was partially digested for 15 min with 0.5 U *Rsa* I (Fermentas). DNA fragments ranging from 2-3Kbp were purified from a 1% TBE agarose gel using the GFX PCR DNA and Gel band purification kit according to the manufacturer's instruction (Amersham). The terminal transferase activity of *Taq* DNA polymerase was used for the addition 3' A- overhangs. The reaction was set up as described in section 2.5.1.2. The 3' A-tailed DNA was recovered from the reaction using the GFX DNA and Gel band purification kit (Amersham) and the DNA was eluted in 10µl ddH_2O. Reconstructed adapter was ligated to the fragmented DNA in a 1 to 6 molar ratio. The ligation reaction was set up as per the manufacturer's instructions (Fermentas) and incubated overnight at 22°C. Prior to the final purification step, 2U of Klenow DNA polymerase and 5µl dNTP

mix (1mM stock) was added and left to incubate for 30 min at 37℃. The tester DNA was finally purified directly from the ligation reaction with the GFX PCR DNA and Gel band purification kit.

2.8.3 Biotin labeled gene specific driver production

Biotinylated gene specific driver DNA was prepared by amplification from a cloned target gene fragment using a 5' – biotinylated gene specific forward degenerate primer and an unlabeled reverse degenerate primer (see Table 2-4). The amplification was performed as described in section 2.6.1. Labeled driver DNA was gel purified as described in the previous section. Driver DNA was denatured and immobilized onto 50 µl of the streptavidin coated magnetic beads as described by the manufacturer (Dynal).

2.8.4 Biotin/magnetic bead purification of DNA

Genomic tester DNA (1µg) was added to the prepared beads and subtractive hybridization was performed as described by Jacobson (Jacobsen, 1995). The magnetic beads were finally resuspended in 10µl ddH_2O and stored at 4℃ until further use.

2.8.5 Tester re-amplification

To avoid the introduction of random mutations during reamplification the high fidelity BIO-X-ACT ™ long DNA polymerase system was used (Bioline). The final concentrations of the PCR reaction components were as follows; 1X OptiBuffer™, 2.5mM $MgCl_2$, 0.2mM dNTP mix, 0.4U BIO-X-ACT™- long

polymerase, 0.5µM T7 primer. Cycling conditions were as follows: 94℃, 3 min; 30 cycles (94℃, 30 sec, 56℃, 1 min, 72℃, 3 min); 72℃, 15 min.

Chapter 3: Metagenomic DNA Extraction From Organic Rich Environmental Samples

3.1 Introduction

Previous studies have shown that one of the mechanisms implicated in the microbial degradation of phenolic molecules is through the production of oxidoreductases such as laccase (Gianfreda and Rao, 2004). The metagenomes of organic rich environmental samples can therefore be considered as a good source of genetic material for the recovery of novel forms of these types of enzymes. Due to their high organic content, compost and decomposed grape skins (grape waste) were chosen as sources of metagenomic DNA for this study.

Numerous extraction and purification protocols have been developed for the direct extraction of metagenomic DNA, (for review see (Roose-Amsaleg *et al*., 2001). Due to the high organic content of the environmental samples and as co-purification of humic acids has been shown to have a deleterious effect on the downstream molecular manipulation of DNA (Jackson *et al*., 1997), several extraction protocols were evaluated for the recovery of high quality metagenomic DNA, with purity, yield and fragment size as the investigation criteria. These included methods as described by Zhou (1996), Miller (1999) and LaMontagne (2002).

3.2 Aims

To extract high quality metagenomic DNA from the target environmental samples and to demonstrate the extent of microbial diversity present in the isolated DNA. A complete phylogenetic analysis of the environmental samples is beyond the scope of this study.

3.3 Experimental strategy

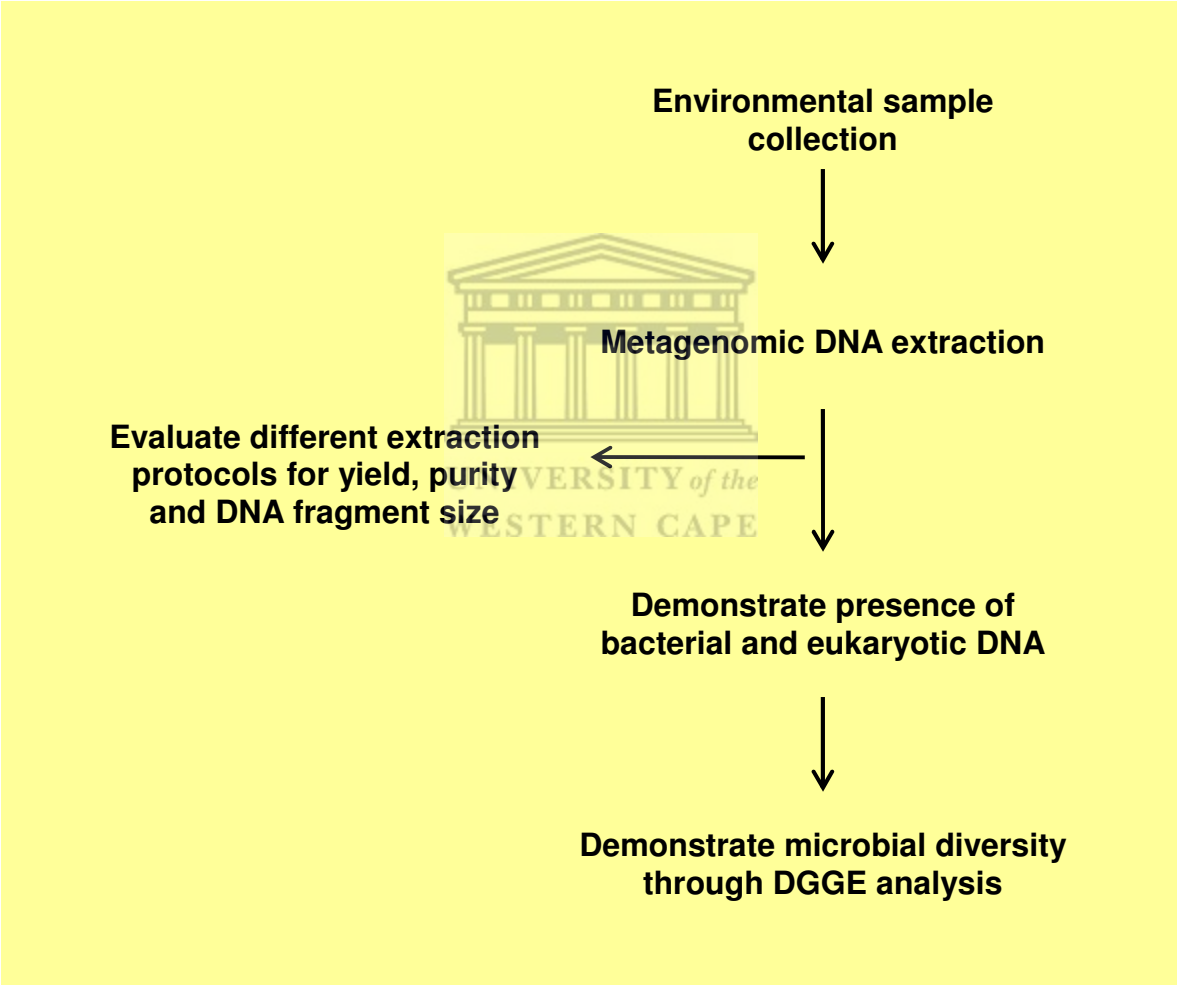

Figure 3.1 Experimental strategy. Environmental DNA was extracted from environmental samples and the quality of the DNA was assessed. The presence of eukaryotic and bacterial DNA was determined and microbial diversity was demonstrated quantatively by denaturing gradient gel electrophoresis (DGGE).

3.4 Results

3.4.1 Direct extraction of metagenomic DNA from compost and grape waste

A compost sample (9 kg) manufactured from plant material was commercially obtained. This compost sample was presumably in the maturation phase of decomposition as it was well degraded with no visibly identifiable plant material present.

A sample of decomposed grape waste (1kg) was obtained from a farm in the Stellenbosch wine farm district. It was evident that this sample was in the early phases of decomposition as significant grape skins, seeds and stalks were still present. Both environmental samples were subdivided into 50g samples and stored at -80ºC until further use.

3.4.1.1 Chemical/ enzymatic cell lysis: Zhou method

5 grams of each environmental sample was used as the starting material in the extraction protocol as described in section 2.5.2.1. Cell lysis, using proteinase K and SDS was performed at a relatively high temperature (65℃). The hot detergent treatment resulted in significant cell lysis, but also had the effect of releasing humic acids into solution, as has been observed previously (Tebbe and Vahjen, 1993; Roose-Amsaleg *et al.*, 2001). Due to the lower humic acid content in the grape waste sample, this effect was less significant as

compared to compost. The addition of CTAB did not reduce the co-purification humic acid from compost, as demonstrated for other environmental samples (Zhou *et al.*, 1996). Contaminating proteins were removed by several phenol/chloroform extractions, and the extracted DNA was ultimately recovered by alcohol precipitation. For further purification, the extracted DNA was passed through a Sephadex G100/PVPP column as described in section 2.5.3. In extracting DNA from compost an excessive amount of humic acid was co-purified, such that the DNA pellet could not be fully re-dissolved. Several passages through Sephadex G100/PVPP columns removed a significant amount of humic acid contamination but resulted in excessive DNA loss (results not shown). The problems were not encountered in extraction of grape waste DNA and the results are summarized in Table 3.1. Approximately 20 µg DNA was extracted per gram of sample. Standard deviations of less than 25% demonstrated the reproducibility of the extraction method. An A_{260}/A_{280} ratio of 1.8 (0.07 SD) was determined for DNA extracted from grape waste after Sephadex G100/PVPP purification. The quality of DNA extracted by the Zhou method was the highest of all the extraction protocols tested, with the majority of the DNA fragments exceeding 14kbp with very little visible degradation (Figure 3.2).

3.4.1.2 Chemical/ enzymatic/mechanical cell lysis: Lamontagne method

Due to the failure of the Zhou protocol to yield good quality DNA from the compost sample a method specifically modified for the extraction of DNA from organic rich samples was tested (section 2.5.2.3). In particular, procedures known to be effective in reducing humic acid contamination were employed. These included lower concentrations of EDTA and SDS as well as high salt concentrations in the lysis buffer. The low detergent concentration was compensated for by utilizing the cell lysis action of chloroform in combination with mechanical shearing (Miller *et al.*, 1999). The results of three extractions are summarized in Table 3-1. A DNA yield of 6.5 µg (SD 0.06) per gram compost was obtained and was found to be sufficiently pure for PCR amplification. The extracted DNA had an A_{260}/A_{280} absorbance ratio of 2.0 (SD 0.07). However the DNA was significantly sheared with the majority of the fragments being smaller than 10 Kbp (Figure 3.2). The extensive DNA fragmentation may be attributed to the harsh nature of the mechanical cell lysis in combination with the low EDTA concentration which results in the inefficient inactivation of nucleases (Miller *et al.*, 1999; LaMontagne *et al.*, 2002).

3.4.1.3 Mechanical/ chemical cell lysis: modified Miller method

Mechanical cell lysis methods commonly used for direct DNA extraction from environmental samples include bead mill, bead beating homogenization and freeze thawing. Mechanical treatments disrupt soil particles and release encapsulated microorganisms, allowing chemical reagents to access and lyse

the cells. Mechanical treatments also result in cell membrane disruption. Using a bead – beating protocol (section 2.5.2.2) DNA was recovered from both the compost and grape waste samples at mean concentrations of 29.4 µg and 52.9 µg per gram sample, respectively (Table 3-1). The DNA yields from repeated extractions were more reproducible from compost samples (SD 3.1) than from grape waste (SD 17.8). This result is not entirely unexpected considering the heterogeneity of the grape waste and the small amounts of sample used per extraction.

In each case the DNA was significantly sheared with the majority of the DNA fragments smaller than 10 Kbp (Figure 3.2). The two main factors which influence both the yield and quality of the DNA are the speed and the duration for which homogenization is performed. Although an increase in either of these parameters can result in an increase in DNA yield, it generally also results in poorer DNA quality. Environmental DNA extracts for both samples were suitably pure (see Table 3-1) and did not require any down-stream purification prior to manipulation.

Table 3-1 Comparison of the yield, quality and purity of the metagenomic DNA extracted from environmental samples

Sample	DNA extraction method	DNA yield (µg/g sample)	Fragment size (Kbp)	Purity A_{260}/A_{280}
Compost	Miller *et al.* (1999)	29.4 (n=10, SD 3.1)	<10	1.7 (SD 0.01)
	Lamontagne *et al.* (2002)	6.5 (n=3, SD 0.06)	<10	2.0 (0.07)
Grape waste	Zhou *et al.* (1996)	20.0 (n=5, SD 3.9)	>14	1.8 (0.07)
	Miller *et al.* (1999)	52.9 (n=10, SD 17.8)	<10	1.7 (SD 0.03)

(n) Number of extractions performed

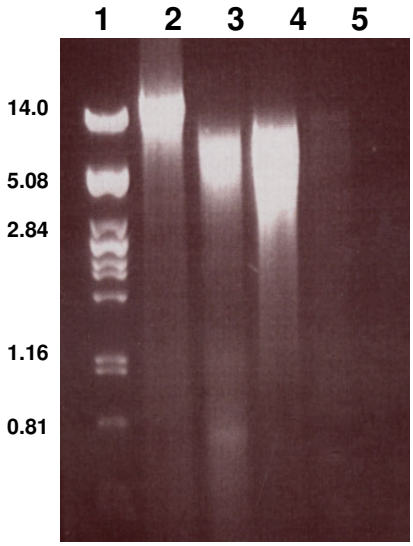

Figure 3.2 Quality of DNA extracted with three different extraction protocols. Equal volumes of the DNA extracts were separated on a 1% agarose gel. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Grape waste DNA (Zhou method). Lane 3) Grape waste DNA (Miller method). Lane 4) Compost DNA (Miller method). 5) Compost DNA (Lamontagne method).

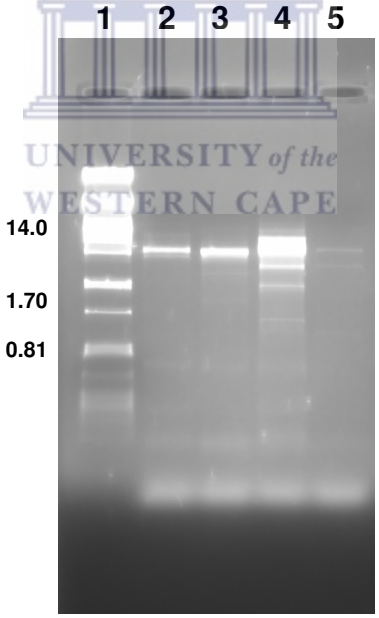

Figure 3.3 Detection of eukaryotic DNA in compost and grape waste. Products of 18S rDNA PCR amplification were separated on a 2% agarose gel. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lane 2) Grape waste DNA (Zhou method). Lane 3) Grape waste DNA (Miller method). Lane 4) Compost DNA (Miller method). 5) Compost DNA (Lamontagne method).

3.4.2 Diversity analysis

3.4.2.1 Detection of eukaryotic DNA in extracted DNA samples by 18S rDNA PCR

The presence of eukaryotic DNA was verified by the amplification of 18S rRNA genes using primers 18SUF and 18SUR (see Table 2-4). In each case an amplicon in the expected size range of approximately 1800bp was amplified (Figure 3.3).

3.4.2.2 Detection of bacterial DNA in extracted DNA samples by 16S rDNA PCR

The presence of prokaryotic DNA was verified by the amplification of 16S rRNA genes using primers E9F and U1510R (see Table 2-4). For each DNA extract an amplicon in the expected size range of 1500bp was amplified (Figure 3.4).

3.4.2.3 Assessment of microbial diversity present in compost and grape waste DNA by denaturing gradient gel electrophoresis

The extent of the microbial diversity present in both environmental samples was investigated by denaturing gradient gel electrophoresis of the mechanically extracted DNA as described in section 2.5.4.1 (Watanabe et al., 2001; de Souza et al., 2004). Bacterial diversity was assessed by the amplification of a 190 bp 16S rDNA gene fragment using primers 341FGC and 534r specifically designed for DGGE (Table 2-4). The 16S gene fragments were subsequently separated on a denaturing gradient polyacrylamide gel as described in section 2.5.4.1. From

the gradient profile, it is evident that both environmental samples showed significant bacterial diversity (Figure 3.5). In an attempt to identify some of the apparently well represented organisms, dominant amplicons were extracted, T/A cloned (section 2.3.4) and sequenced (section 2.4.4). The sequences were compared by BLAST analysis against the NCBI nucleotide database. All sequences showed significant similarity to previously identified but uncultured bacteria present in composted material (Table 3-2).

Similarly, the fungal diversity present was quantitatively determined by the separation of 18S rRNA gene fragments amplified using primers NS7GC and F1Ra (Table 2-4). Both samples demonstrated the presence of significant diversity, with the grape waste sample far exceeding that of the compost (Figure 3.5). This result is in line with the relative degree of decomposition of the two samples. Dominant 18S rDNA amplicons were also cloned, sequenced and compared by BLAST analysis in an attempt to identify some of the eukaryotic organisms present (Table3-2). Two phylotypes showing significant similarity to fungi typically present in degrading wood and plant material *Pseudallescheria angusta* and *Cyphellopsis anomala* were identified in the compost sample. Putative species, *Dipodascus capitatus*, was identified in the grape waste sample.

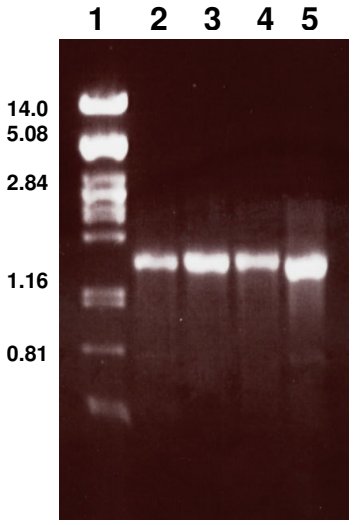

Figure 3.4 The detection of bacterial DNA in compost and grape waste. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lane 2) Grape waste DNA (Zhou method). Lane 3) Grape waste DNA (Miller method). Lane 4) Compost DNA (Miller method). 5) Compost DNA (Lamontagne method).

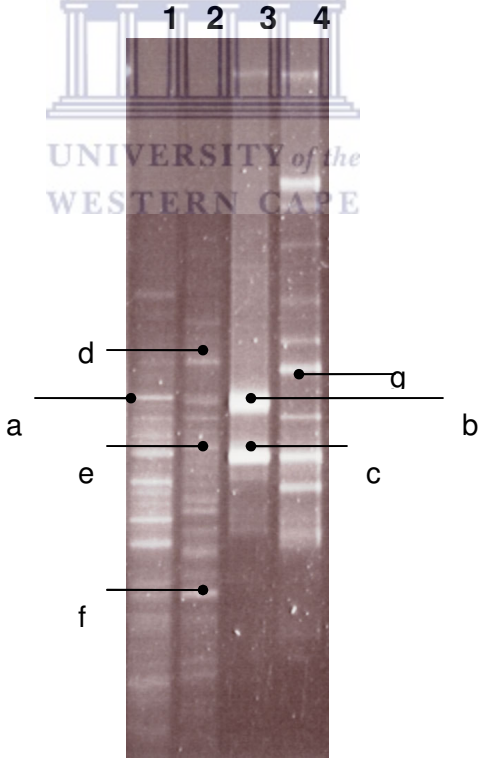

Figure 3.5 DGGE results demonstrating microbial diversity in compost and grape waste samples. Lane1) Compost 16S. Lane 2) Grape waste 16S. Lane 3)

Compost 18S. Lane 4) Grape waste 18S. Arrows indicate fragments sequenced and listed in Table 3.2.

Table 3-2 Organisms identified in compost and grape waste by denaturing gradient gel electrophoresis.

Sample	Organism	E-value/Score	NCBI Accession no.
Compost	Bacteria		
	Uncultured delta proteobacterium[a]	1e-97/363	AJ318168
	Fungi		
	Cyphellopsis anomala[b]	e-170/606	AF426949
	Pseudallescheria angusta[c]	0.0/678	AF275528
Grape waste	Bacteria		
	Uncultured bacterium[d]	8e-65/254	AB064703
	Uncultured bacterium[e]	5e-92/344	AY209424
	Uncultured bacterium[f]	1e-74/286	AY592094
	Fungi		
	Dipodascus capitatus[g]	e-180/369	AB083080

(a-g) See Figure 3.5 for position on DGGE gel

3.5 Discussion

Three different DNA extraction protocols were compared with regard to DNA yield, purity and fragment size. Two organic rich environmental samples, compost and grape waste, were selected as the source material. DNA extracted from these types of samples tend to be contaminated with phenolic substances like humic acids as a result of co-purification due to their similar chemical properties to nucleic acids. It was demonstrated that the hot chemical lysis method as described by Zhou (1996) was suitable for DNA extraction from grape waste, yielding suitably pure DNA with a molecular weight exceeding 14 Kbp and an average yield of 20µg DNA/ gram sample. However, this method was not considered appropriate for use with the compost sample as excessive amounts of humic acids were co-purified, necessitating several additional purification steps resulting in high DNA shearing and a significant reduction in yield.

A protocol developed specifically for metagenomic DNA extraction from compost was subsequently tested (LaMontagne *et al*., 2002). Although acceptably pure DNA could be extracted, this method was much more laborious, and yielded less DNA (6.5µg DNA/ gram sample) of lower molecular weight (less than 10Kbp).

The final protocol evaluated on both the compost and grape waste samples was the mechanical shearing method as described by Miller (1999). Although this method resulted in significant shearing of the extracted DNA with a mean fragment size of less than 10 Kbp, pure DNA in a very high yield was obtained (29µg DNA/gram of compost and 52µg DNA/gram of grape waste).

The Miller protocol was selected as the method of choice for all subsequent DNA extractions. DNA extracts using this method were screened for the presence of bacterial and eukaryotic DNA by 16S and 18S gene PCR detection, respectively. In both environmental samples bacterial and eukaryotic DNA could be detected. Bacterial and fungal diversity was analysed by denaturing gradient gel electrophoresis. DGGE results showed that both samples possessed a high degree of microbial diversity. The identification of dominant amplicons was undertaken by extraction, cloning and sequencing. BLAST analysis of sequences indicated that bacterial phylotypes could be assigned to known but uncultured taxa. Dominant fungal amplicons were identified as *Cyphellopsis anomala, Pseudallescheria angusta* and *Dipodascus capitatus*.

Chapter 4: Metagenomic DNA Library Construction And Expression Screening

4.1 Introduction

Small insert metagenomic DNA libraries have typically been constructed from partially digested metagenomic DNA cloned into high copy number plasmid vectors like pSK+ and pUC (Henne *et al*., 1999; Knietsch *et al*., 2003b). This very simple approach combined with activity based screening has resulted in the discovery of numerous novel enzymes (Cowan *et al*., 2005). This approach however is not suitable for cloning highly fragmented DNA obtained from mechanical DNA extraction methods like beat beating. Restriction enzyme digestion of short DNA fragments results in poor cloning efficiency due to the formation of non-compatible sticky ends which ultimately hampers library production. The increased DNA yield and bacterial taxa representation offered by mechanical extraction protocols merits the investigation of alternative cloning strategies for the construction of high quality small insert libraries. Wilkinson (2002) described a new strategy for cloning sheared metagenomic DNA through TOPO® TA cloning and in this chapter the technique was employed for the construction of a library derived from compost environmental DNA. The TOPO® TA cloning system provides a quick and efficient means that allow cloning of mechanically sheared metagenomic DNA. The speed and coupled efficiency is due to the innovative nature of the cloning strategy. The TOPO® vector is constructed by the integration of topoisomerase I from *Vaccinia* virus.

The enzyme recognizes a specifically engineered sequence (5'-CCCTT-3') which flanks the vector multiple cloning site (Shuman, 1991). The enzyme cleaves a phosphodiester bond on one of the DNA strands after this sequence and forms a covalent bond between the cleaved 3'- phosphate and a tyrosyl residue of the enzyme located at position 274. This results in the formation of a vector/ topoisomerase I complex with a protruding 3'- thymidine (T) overhang that can be employed for TA cloning. The phospho-tyrosyl bond is attacked by the 5'-hydroxyl group of the incoming DNA insert and thereby releases the topoisomerase to form the recombinant vector (Shuman, 1994).

As an alternative approach, the CopyControl™ Fosmid Library Production Kit was also employed for the production of a larger insert library derived from grape waste environmental DNA. Both libraries were screened for bacterial laccase activity.

4.2 Aims

To construct and expression screen a small insert genomic DNA library derived from compost metagenomic DNA for laccase activity.

To construct and expression screen a large insert genomic DNA library derived from grape waste metagenomic DNA for laccase activity.

4.3 Experimental strategy

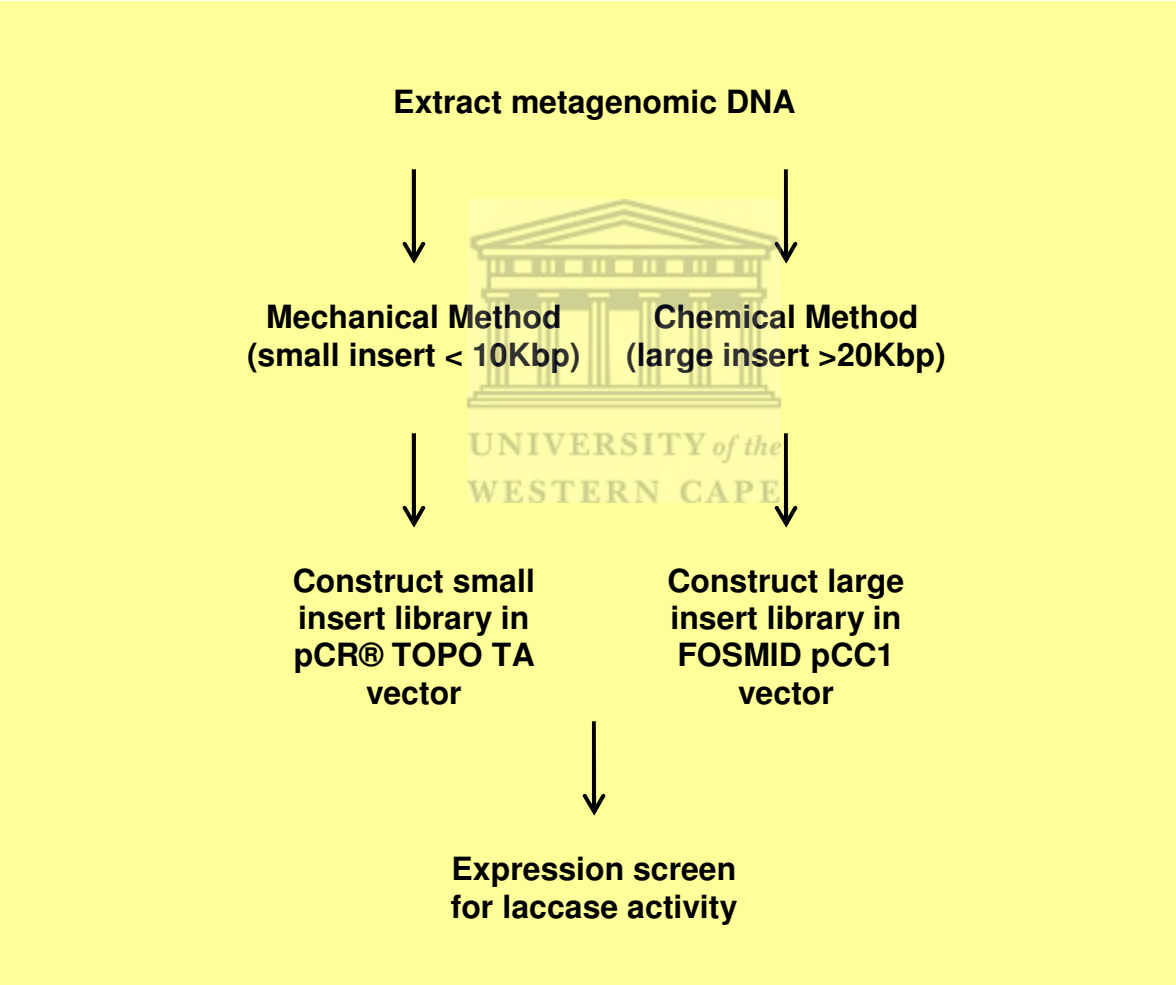

Figure 4.1 Experimental strategy followed for the activity based discovery of bacterial laccases genes.

4.4 Results

4.4.1 Small insert metagenomic library construction and characterization

Metagenomic DNA extracted from compost using a mechanical extraction method (section 3.4.1.3) was used as the source material for library construction. The DNA was already randomly sheared in the appropriate size range (2-10Kbp) and therefore did not require any additional enzymatic cleavage which may have biased the library. DGGE analysis results described in chapter three (section 3.4.2.3) demonstrated that the mature compost sample was likely to be a better choice for the source material than the grape waste as there was significantly lower eukaryotic DNA diversity present in this sample. The presence of eukaryotic DNA would have a deleterious effect on the quality of the metagenomic DNA library as any cloned eukaryotic 'junk DNA' would decrease the coverage of bacterial taxa present and therefore reduce the chances of identifying bacterial laccases through expression screening.

The TOPO® cloning strategy requires insert DNA with protruding 3'- adenine (A) tails for cloning. Due to random shearing, sticky ends were created and therefore the ends of the metagenomic DNA fragments had to be blunt-ended by removing/ or filling in any 5' and 3' protruding ends. This was done by employing the exonuclease and polymerase functions on T4 DNA polymerase on approximately 5µg template DNA as described in section 2.6.1.1. 3'- adenine overhang tails were introduced by the non-template dependent terminal transferase activity of *Taq* polymerase.

The presence of 5'- hydroxyl groups are of critical importance for efficient cloning of incoming insert DNA and was demonstrated in the early work describing the

development of this cloning strategy (Shuman, 1994). The compost DNA was therefore dephosporylated as described in section 2.6.1.3. To ensure optimal cloning efficiency the DNA was gel purified by preparative agarose gel electrophoresis. During this step DNA fragments in the 2 to 10Kbp size were recovered for cloning as described in section 2.4.6 (Figure 4.4).

Approximately 50ng of the prepared metagenomic DNA was ligated into 10ng pCR®T7/CT-TOPO® vector (section 2.6.1.4) and then transformed into electrocompetent *E.coli* TOP10F' cells, a 1% of the transformation volume was plated on LB-ampicillin agar plates to determine the library size. An average transformation efficiency of approximately 2,6 X 10^7 cfu/µg plasmid DNA (SD 4,6 X 10^6, n=4) was observed and the constructed compost metagemonic DNA library ultimately consisted of approximately 1,05 X 10^6 cfu. The *E.coli* TOP10F' strain provides high transformation efficiencies (≈ 10^9 cfu/µg DNA) under optimal conditions. This strain also allows long term stable storage of the plasmid DNA.

The library was amplified by pooling the transformation reactions and adjusting the volume with fresh TB – ampicillin media. Amplification was monitored by allowing growth for 30 minute increments and monitoring the state of amplification by plate counts. After two divisions (4 copies), each copy of the library was stored at – 80^0C as glycerol stocks until further use. To estimate the average insert size and recombination efficiency, 50 colonies were selected at random and inoculated in LB-ampicillin media for subsequent plasmid extraction (section 2.4.7). To estimate the insert size, 300ng of each plasmid was digested with 1U *EcoR*I (section 2.4.1). After digestion non – recombinant pCR®T7/CT-TOPO® vector is visible as a 2702 bp linear fragment. Even

though the sample is to small to provide statistically significant data it does provide some indication of the library quality. A recombination efficiency of approximately 90% was estimated with an average insert size of 2.7 Kbp (SD ± 1.5 Kbp**,** n=50). This constitutes approximately 2.6 Gbp cloned metagenomic DNA. Figure 4.3 shows the *EcoR*I digestion results of some of the larger clones obtained (5 – 8 Kbp in size).

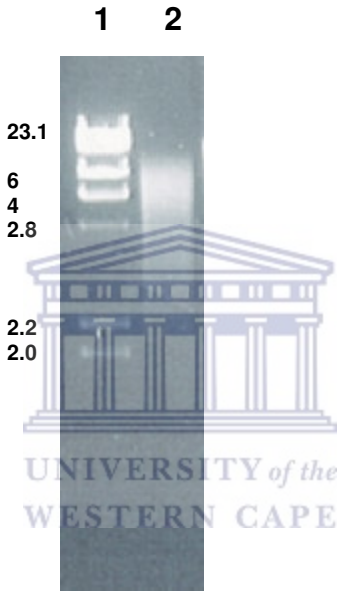

Figure 4.2 Compost metagenomic DNA prepared for T/A cloning into the pCR®T7/CT-TOPO® vector. Lane 1) Lambda *Hind*III mw size marker (Kbp). Lane 2) Compost metagenomic DNA (approximately 150ng).

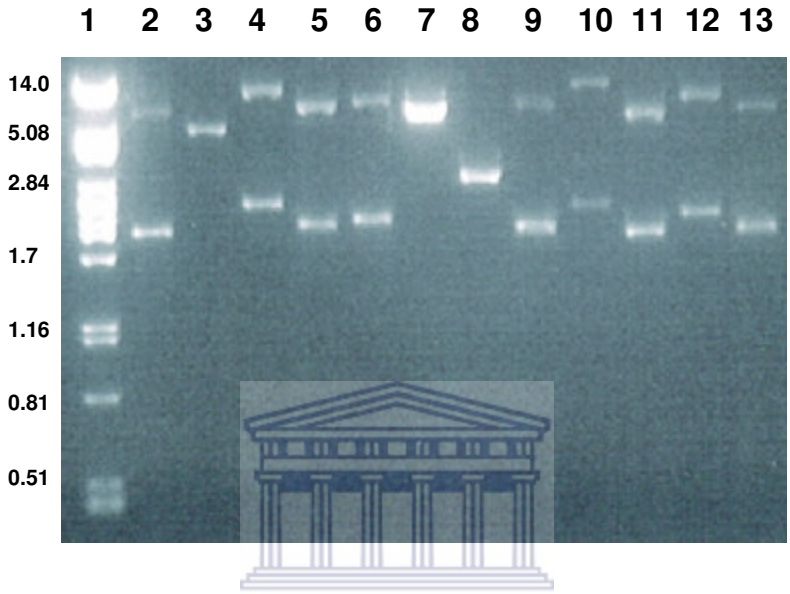

Figure 4.3 Randomly selected library plasmid clones digested with 1U *EcoR*I. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lanes 2-13) Randomly selected library clones.

4.4.2 CopyControl™ Fosmid library production

The CopyControl™ Fosmid Library Production Kit (Epicentre®) was employed for the construction of a large insert library. This cloning system developed by Wild and co-workers (2002), provides the advantage of a typical single copy fosmid vector by allowing the efficient and stable blunt end cloning of large DNA inserts (≈ 45 Kbp) while also offering the induction of clones to high copy number for subsequent downstream isolation and manipulation. The CopyControl™ Fosmid Library cloning strategy is depicted in Figure 4.6.

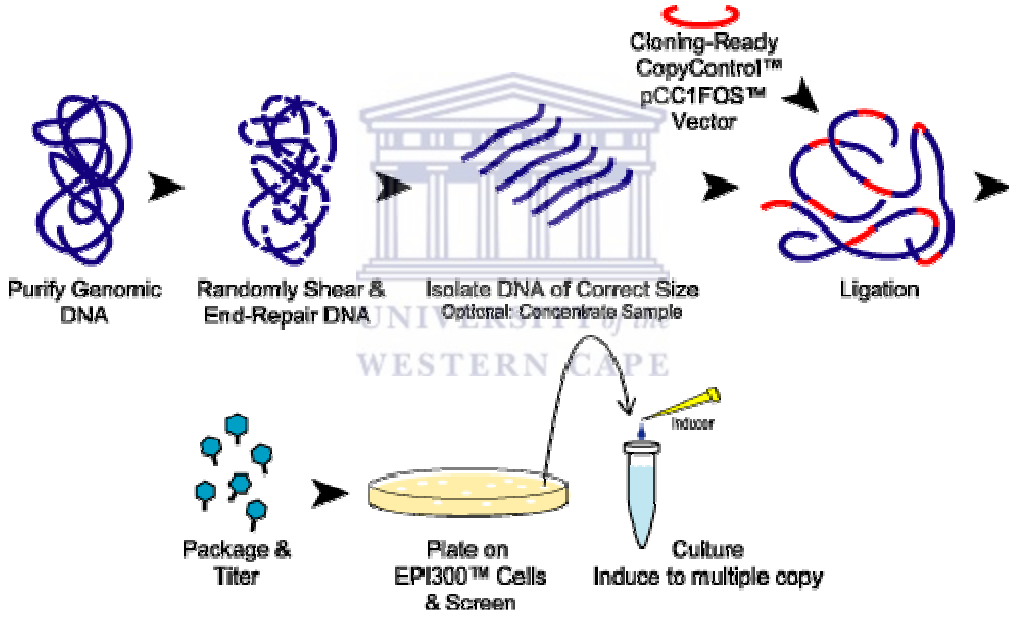

Figure 4.4 CopyControl™ Fosmid Library production strategy for blunt end cloning of large DNA fragments, (from the manufacturers manual).

4.4.3 Large insert metagenomic library construction and characterization

For the construction of a large insert metagenomic DNA library, grape waste metagenomic DNA extracted by a chemical/ enzymatic DNA extraction method was used as source material (section 3.4.1.1). As discussed earlier, during chemical extraction methods the genomic DNA is subjected to less harsh handling conditions. This ultimately favor the recovery of high molecular weight DNA which is required for efficient fosmid library construction as the system requires insert DNA in the 40 Kbp size range for optimal cloning.

High eukaryotic DNA content was demonstrated previously in the grape waste sample (3.4.2.3), probably due to the early stage of decomposition of the sample. Although the compost sample was likely to be a better choice due to its mature state, difficulties experienced with the extraction of high quality DNA by chemical methods (3.4.1.1) necessitated the use of grape waste DNA for this application.

Grape waste DNA was prepared for blunt end cloning (section 2.6.3). DNA fragments (>14Kbp) were recovered from a low melting point agarose gel after overnight separation at 30V (Figure 4.5). Approximately 100ng grape waste DNA was ligated and then packaged into the supplied MaxPlax Lambda Packaging extracts. In order to estimate the library size, the packaged fosmid clones were titered by infection of the EPI300-T1^{R} plating strain. From the phage titration results it was calculated that a library consisting of approximately 44000 cfu was constructed. To estimate the average cloned insert size, single colonies were selected at random from the titering plates and induced to produce a high copy number (section 2.6.3). The plasmids were extracted and digested by 1U *Eco*RI (section 2.4.1). Based on the DNA fragments obtained, the

average insert size was estimated at 25 Kbp (SD ± 5.7 Kbp, n= 20) (Figure 4.6). This constituted approximately 1.1 Gbp of cloned grape waste DNA.

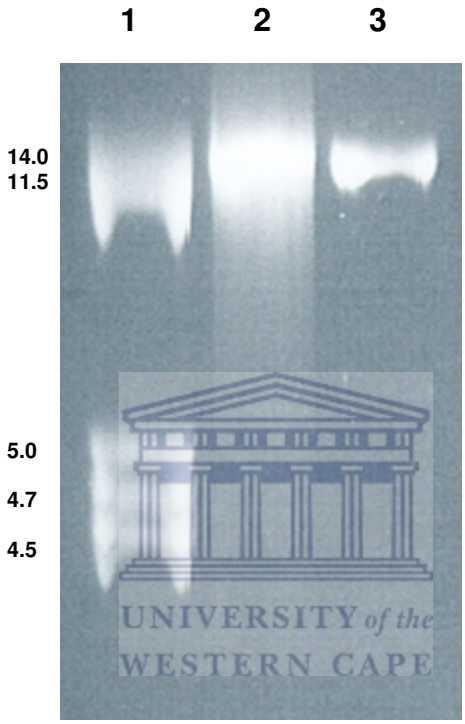

Figure 4.5 Grape waste environmental DNA preparation for fosmid cloning. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lane 2) Grape waste metagenomic DNA. Lane 3) Fosmid control DNA (37Kbp).

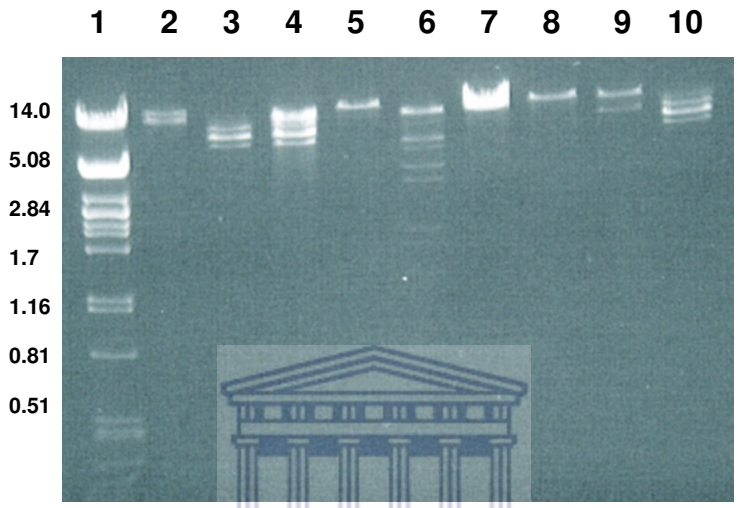

Figure 4.6 Randomly selected library fosmid clones digested with 1U *EcoR*I. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lanes 2-10) Randomly selected library clones.

4.4.4 **Expression screening for laccase activity**

Functional screening for laccase activity was performed on a solid agar support using a chromogenic oxidative coupling reaction between 2,2'-azino-di(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) and *p*-hydroxybenzoic acid (Shin *et al*., 1987). Laccase activity is detectable a as stable deep purple colored product (3-(6'sulfo-3'—ethylbenzothiazol-2'-ylidenehydrazono)-4-oxo-1,5'-cyclohexadiene-1-carboxylic acid).

The small insert compost DNA library was screened on LB-agar indicator plates augmented with 100µg/ml ampicillin (section 2.6.4.1). To ensure adequate clone separation a single copy of the library was diluted to 10000 cfu/ml, and 1000cfu plated per indicator plate and incubated at room temperature. The screening plates were continuously monitored for the presence of laccase activity for up to 15 days. No positive clones were identified. As a positive control each batch of indicator plates were tested for color development by spotting 50 µl laccase isolated from *Trametes versicolor* (0.83 U/ml) onto an indicator plate.

In order to determine whether the failure to detect laccase activity was due to the absence of expressible laccase genes in the library or whether laccase activity had to be induced, plasmid DNA from a single copy of the library was extracted (section 2.4.7) and screened for the presence of laccase by gene specific PCR using *mco* degenerate primers (section 2.4.2). The development of the laccase PCR detection assay is described in chapter 5 (section 5.4.1). No amplicons were obtained, thus suggests that bacterial laccase genes were not present in the small insert library (Figure 4.7).

The grape waste DNA fosmid library was screened by infection of *E.coli* EPI300-T1® cells which were subsequently plated on indicator plates (12,5µg/ml chloramphenicol) at approximately 1000 cfu/plate. The indicator plates were continuously monitored for laccase activity for up to 20 days post infection. No laccase activity was detected.

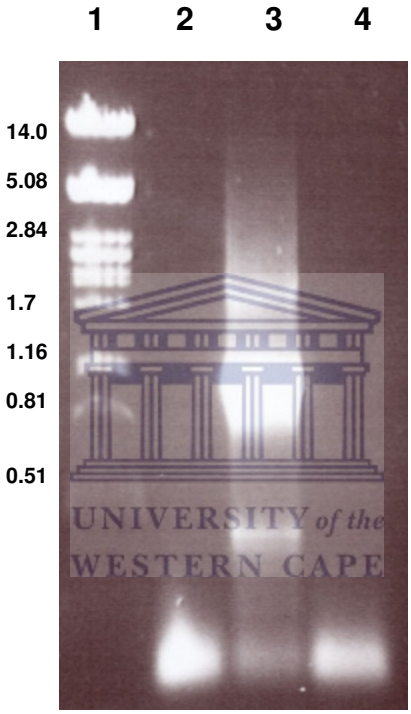

Figure 4.7 Laccase specific gene PCR detection. Lane 1) Lambda *Pst*I mw size marker (Kbp). Lane 2) Negative PCR control (no DNA). Lane 3) *Streptomyces* laccase positive control. Lane 4) Compost metagenomic DNA library.

4.5 Discussion

Metagenomes are virtually inexhaustible reservoirs for the recovery of novel and potentially economically valuable genes. Various technologies have been developed to target specific genes within environmental samples. One of the earliest and still most popular approaches is activity based screening of metagenomic DNA libraries. Most reports describe the construction of small insert DNA libraries (2-10Kbp) and the subsequent recovery of various novel genes (Henne *et al*., 1999; Henne *et al*., 2000; Cottrell *et al*., 1999). Other workers constructed large insert libraries that could be used to identify entire metabolic pathways and other biomolecules (Brady *et al*., 2001; Rondon *et al*., 2000; Gillespie *et al*., 2002).

Although this strategy was feasible it also had some limitations as evidenced by the low rate of gene discovery, typically less than 0.01% of the clones screened (Cowan *et al*., 2005). This has proved to be less of a problem where high throughput screening technology is employed such as in industry, but high operating cost limits its use in most laboratories (Table 1-5).

The process for small insert library construction has been to partially digest high molecular weight metagenomic DNA and then, following appropriate size selection, clone the DNA fragments into a high copy number plasmid carrying an inducible promoter. It is accepted that the harsher mechanical DNA extraction methods not only provides higher DNA yield but also gives more comprehensive access to the metagenome (Roose-Amsaleg *et al*., 2001). The DNA is however typically of lower molecular weight which results in inefficient cloning due to the formation of non-compatible ends. To address this problem, Wilkinson (2002) proposed a cloning

strategy for T/A cloning mechanically sheared metagenomic DNA into a high copy number TOPO® plasmid vector. This approach holds the advantage of cloning randomly sheared DNA with high efficiency using the novel TOPO® cloning mechanism. This is particularly useful as large libraries are required to ensure successful activity based screening. A disadvantage of the strategy, however, is the DNA manipulation necessary to blunt end, A- tail and dephosphorylate the DNA for efficient cloning. Apart from being somewhat time consuming, each manipulation step could also potentially introduce an inefficiency that ultimately influences the success of the library production.

In activity based screening, three modes of gene expression are possible; independent expression, transcriptional fusion and translational fusion (Gabor *et al*., 2004). The TOPO® T/A cloning system was designed for cloning and over-expression of single genes under the control of a strong vector based T7 promoter. To prevent leaky expression of genes that may be toxic or harmful to the cells, clones are constructed in the *E.coli* TOP10F' strain lacking the lambda DE3 lysogen and therefore the T7 RNA polymerase. This enables highly efficient cloning and stable storage of the cloned gene. When expression is required the clone is transformed into the *E.coli* BL21 (DE3) expression host. This means that libraries can be constructed efficiently, stably maintained and stored within the cloning strain as well as screened for independent expression. To employ the other possible modes of gene expression *i.e*. transcriptional and translational fusion, the library must be extracted in plasmid form and re-transformed into the expression host.

In this study a small insert metagenomic DNA library was constructed from compost-derived environmental DNA. The sample was selected based on its apparently lower eukaryotic DNA content (see section 3.4.2.3). The constructed library, consisting of approximately 1,05 X 10^6 cfu with an average insert size of 2.7 Kbp, was screened for laccase activity but yielded no positive clones.

Gabor and co-workers (2004) proposed a formula describing the number of clones necessary (Np) to detect an enzymatic activity with a certain probability (P) when using independent expression.

$$Np = \frac{\ln(1-P)}{\ln\left(1-\frac{l-X}{G \cdot c \cdot z}\right)} \qquad (1)$$

l is the insert size, X the gene size of interest, G the average genome size, z the number of different taxa present assuming an even distribution and c, a constant equaling 1 for independent expression.

The *in sillico* data reported by Gabor and coworkers (2004) seem to suggest that independent expression screening of a library with an average insert size of 2.7Kbp would require screening in excess of 10^7 clones. This is clearly much larger than the constructed compost library. However, the equation described by Gabor and co-workers (2004) has some limitations. For example, it fails to take into account the influence of DNA from other sources present in metagenomic DNA samples. Cloning eukaryotic,

archaeal and viral DNA undoubtedly has a negative influence on the quality of the library and therefore the number of clones needed for successful gene detection.

The product of G.z in equation (1) describes the bacterial DNA complexity within the environmental sample. However, it fails to account for any contribution made to the total DNA by eukaryotes, archaea and DNA viruses. To obtain more realistic estimation of the number of clones required for successful screening, equation (1) may be transformed to reflect the total DNA complexity.

$$Np = \frac{\ln(1-P)}{\ln\left(1-\frac{l-X}{Ct}\right)} \tag{2}$$

Ct is the total DNA complexity of the metagenome which includes the contributions made by bacteria (Cb), eukaryotes (Ce), archaea, (Ca) and DNA viruses (Cv) see equation 3.

$$Ct = Cb + Ce + Ca + Cv \tag{3}$$

The fraction of bacterial DNA present (fb) in a metagenomic DNA sample can therefore be described as :

$$fb = \frac{Cb}{Ct} \tag{4}$$

Equation 2 can now be transformed to account for the impact of DNA from other sources by substituting Nt .

$$Np = \frac{\ln(1-P)}{\ln\left(1-\frac{(l-x)fb}{G \cdot c \cdot z}\right)} \tag{5}$$

To illustrate the effect of the bacterial DNA fraction present (f_b) on Np the following model calculation was performed (Figure 4.8). Assuming l = 2.7 Kbp, X = 1.5 Kbp, G = 4 Mbp, c=1, P =0.9 and z =10^4, the resulting curve demonstrates the relationship between the number of clones to be screened in relation to the bacterial DNA contribution.

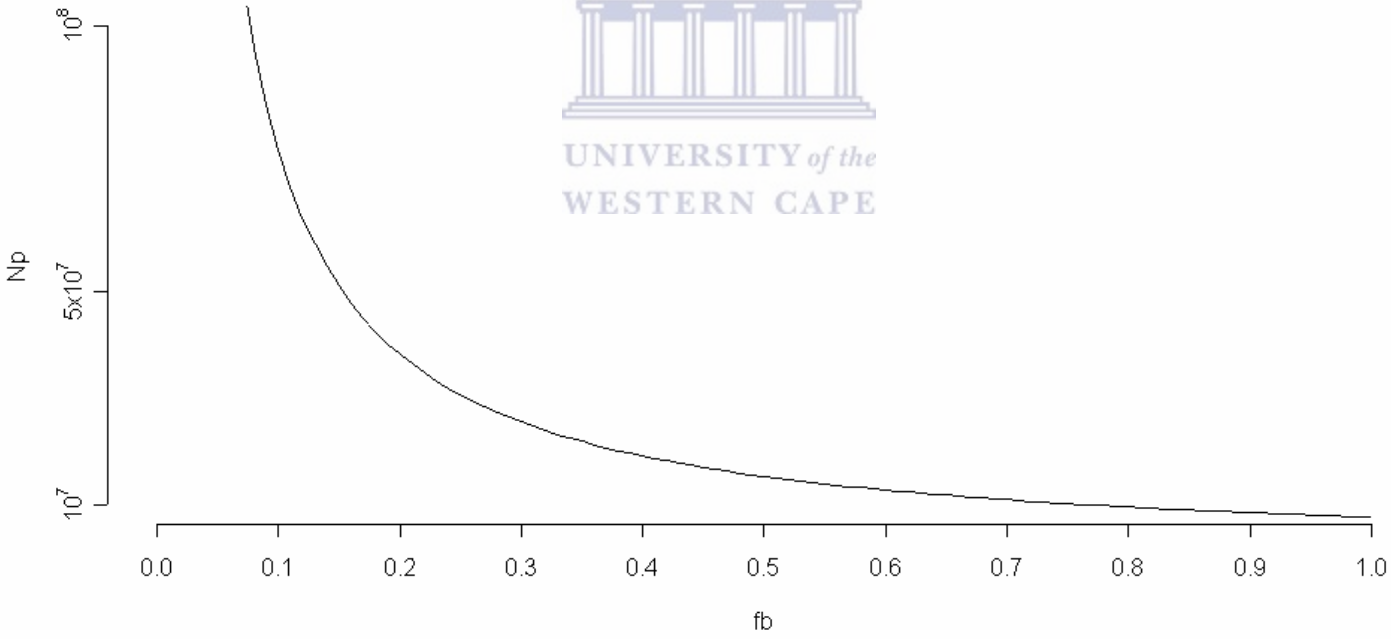

Figure 4.8 Dependence of Np on the bacterial fraction (fb) for independent expression. The curve was calculated with Equation 5.

Although the compost sample demonstrated significantly less eukaryotic diversity than the grape waste based on DGGE results (section 3.4.2.3), the exact contribution of eukaryotic DNA to the overall DNA extraction is unknown. The presence of this 'junk' DNA is believed to have contributed to the failure in identifying laccase genes, as the model predicted the screening of a library at least hundred fold larger than was constructed.

Gabor and co-workers (2004) also suggested an equal frequency distribution for the bacterial taxa present, an assumption that was shown to significantly underestimate bacterial diversity (Gans *et al*., 2005). Application of this assumption therefore underestimates the number of clones to be screened for successful identification of novel single copy genes. Constructing small insert libraries from highly complex metagenomic DNA samples therefore results in very weak coverage of the metagenome in libraries consisting of less than 10^7 clones.

To determine whether any laccase genes were cloned as part of the metagenomic library, plasmid DNA was extracted from an amplified copy of the library and used as template in a laccase gene specific PCR detection assay. The development of this assay is described in chapter 5. No laccase amplification was obtained and it was therefore concluded that due to poor coverage no laccase genes were cloned initially (Figure 4.10). Further expression screening of the library through other modes of gene expression was therefore considered inappropriate.

As an alternative approach, a large insert fosmid library was constructed from grape waste metagenomic DNA. DGGE analysis of the sample demonstrated high prokaryotic and eukaryotic diversity (section 3.4.2.3). As discussed earlier, the presence

of eukaryotic DNA affects the quality of an environmental library by diminishing bacterial taxa coverage which inevitably results in a lower gene recovery rate. The material used for library construction was chemically extracted to obtain high molecular weight DNA fragments. It was assumed that due to the comparatively gentler approach eukaryotic DNA contamination would be kept to a minimum (Roose-Amsaleg *et al*., 2001), although its exact contribution was undetermined.

In the absence of pulsed field gel electrophoresis technology, separation of large DNA fragments were difficult as evidenced by the poor separation obtained on a 1% agarose gel between the 14Kbp Lambda molecular weight size marker, the 37Kbp size marker provided by the manufacturers and the grape waste DNA sample (Figure 4.5). The library was therefore constructed from DNA fragments within this size range and yielded approximately 1.1 Gbp of cloned metagenomic DNA. Efforts to recover novel bacterial laccase genes through independent expression failed and the cause was again assumed to be due to poor coverage of the complex DNA sample.

Chapter 5: PCR Detection Of Bacterial Multicopper Oxidases In Metagenomic DNA

5.1 Introduction

A number of biotechnological applications using laccases have been identified (Duran *et al.*, 2002). For example, the class of chemicals known as catechols (dihidroxybenzene derivatives) is important in the chemical industry for the production of agrichemicals, pharmaceuticals, paints, stains, photographic developers, synthetic flavors and dyes. Catechols can be synthesized chemically by the oxidation of salicylaldehyde and the demethylation of guaiacol (Kodama *et al.*, 1997). For economic reasons there is considerable interest in producing catechols with the aid of microorganisms or biocatalysts. Burton and co-workers (1998) demonstrated the production of catechols and *o*-quinone products from a range of phenolic substrates by using immobilized and soluble forms of a laccase isolated from *Agaricus bisporu*s.

Laccase activity has primarily been demonstrated in wood degrading fungi where it plays a role in lignin degradation (reviewed in section 1.6.1.1). However, there has been increasing interest in identifying and isolating bacterial laccase genes. The presence of laccase activity has been shown in a number of diverse bacterial taxa, including *Azospirillum lipoferum* ((Givaudan *et al.*, 1993), *Marinomonas mediterranea* (Sanchez-Amat *et al.*, 2001), *Escherichia coli* (Grass and Rensing, 2001) and *Bacillus*

halodurans (Ruijssenaars and Hartmans, 2004) the widespread presence of laccases in bacteria has also been suggested based on bacterial genome sequenceing results (Alexandre and Zhulin, 2000).

Due to the advances in metagenomic DNA extraction technology, it is now feasible to access much of the extant bacterial diversity and therefore investigate the occurrence of laccase genes in bacteria. To the best of the author's knowledge, no degenerate primers have been reported in the literature for the specific detection of bacterial laccases at the time this study was initiated. To this end, such a set of primers were designed and its specificity verified using a laccase positive *Streptomyces* strain. Metagenomic DNA extracted from compost and grape waste samples (see chapter 3) were screened for the presence of bacterial laccases. Two different PCR based strategies were investigated for the possible recovery of full-length copies of these bacterial multicopper oxidase genes. These included a semi- nested random primed PCR technique (Eschenfeldt *et al*., 2001) as well as the universal Vectorette™ system.

Finally, to improve the efficiency with which genes can be recovered from bacterial genomes, existing magnetic bead capture technology (Jacobsen, 1995) was adapted to enable the specific and rapid recovery of single copy bacterial genes. As a proof of concept, a putative *mco* gene was recovered from a *Streptomyces* strain.

5.2 Aims

To design a set of bacterial multicopper oxidase specific degenerate primers from published *mco* sequence information, and to detect the presence of novel *mco* genes in compost and grape waste metagenomic DNA.

To investigate the possible use of a semi-nested random primed PCR approach for the recovery of full-length *mco* genes from a complex environmental DNA sample.

To investigate the possible application of the commercial Universal Vectorette™ System for the recovery of full-length *mco* genes.

To develop a subtractive hybridization magnetic bead capture technique (SH-MBC) for the recovery of full-length bacterial genes.

5.3 Experimental strategy

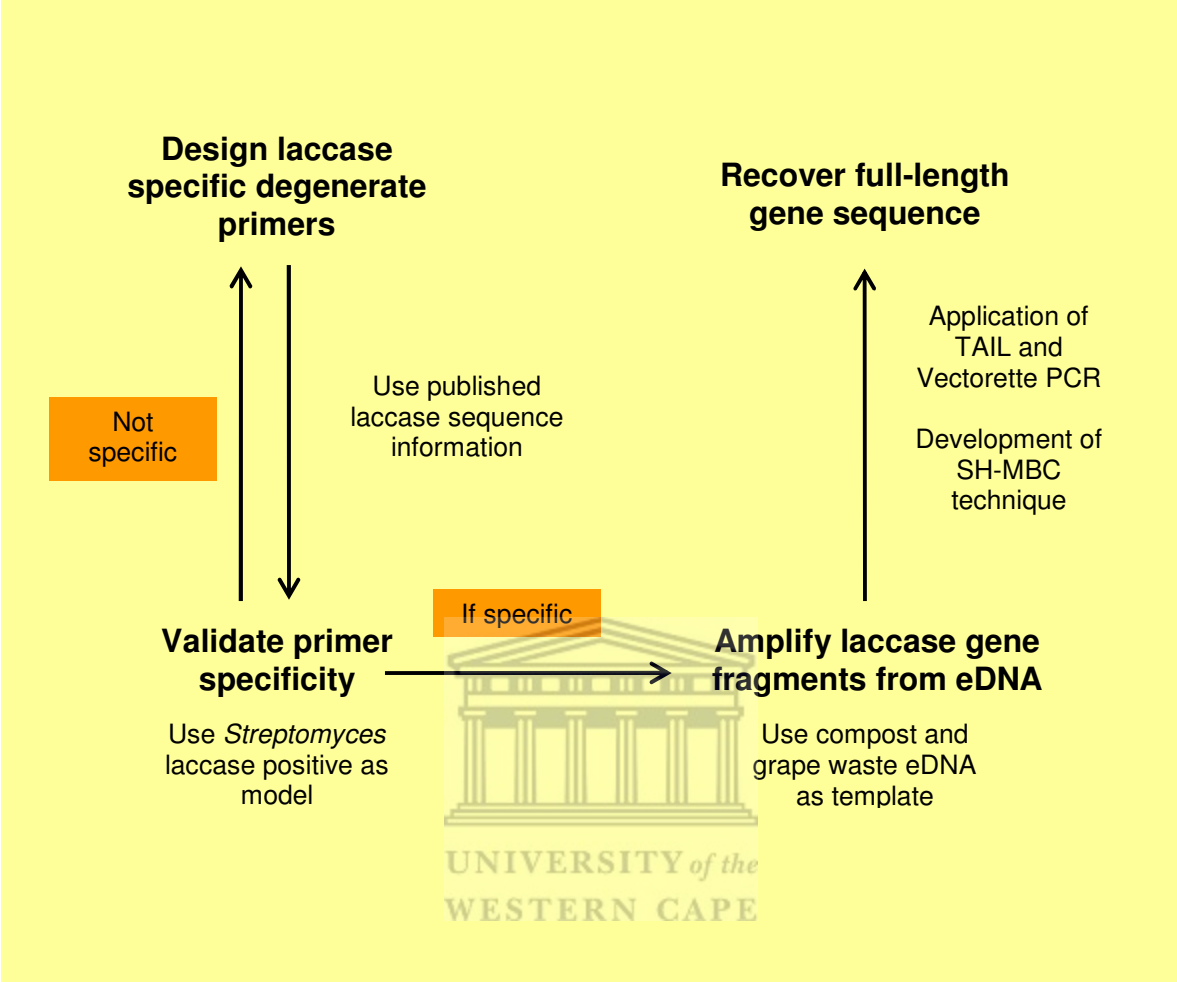

Figure 5.1 The flowchart illustrates the experimental design followed for the gene specific PCR detection of novel bacterial *mco* genes from metagenomic DNA samples.

5.4 Results

5.4.1 Gene specific detection of bacterial laccase genes

5.4.1.1 The design of laccase specific degenerate primers

Protein sequences identified *in silico* as putative bacterial laccases (Alexandre and Zhulin, 2000) as well as published fungal laccase protein sequences were obtained from GenBank (see Table 5-1). The sequences were aligned in the BioEdit software package (Hall, 1999) using the ClustalW algorithm for multiple sequence alignment (Thompson *et al.*, 1994). The result of the alignment is depicted in Figure 5.2. The alignment shows the presence of four highly conserved regions previously identified as copper-binding domains (Solano *et al.*, 2001). Relative to the alignment ruler, these domains are located at positions 126 to 129 (HWHG); 171 to 173 (HSH); 495 to 498 (HLHG) and 549 to 551 (HCH). Due to the high degree of sequence conservation, these sites were regarded as potentially good targets for degenerate primer design.

A survey of other protein sequences in the multicopper oxidase (*mco*) family including L-ascorbate oxidases (plant), copper resistance proteins (bacterial) and ceruloplasmin (mammalian) suggested that these domains were highly conserved across most members of the *mco* gene family except ceruloplasmin (results not shown). Degenerate primers designed to any of these sites would therefore not result in the exclusive amplification of laccase gene fragments. Due to the presence and size of introns in eukaryotic DNA it is unlikely that these *mco* genes could be amplified from a metagenomic DNA sample. However, bacterial copper resistance genes were likely targets for the primer set.

To design a *mco* specific primer set, the COnsensus DEgenerate Hybrid Oligonucleotide Program (CODEHOP) was used (Rose *et al.*, 1998). Conventional degenerate primer design results in a primer pool of which only a small fraction of the primers can participate in the amplification reaction, only a fraction of the primers having the sequence required to recognize the target gene. In the CODEHOP methodology ,clamping consensus sequences are included on both the 5' and 3' ends with the primer displaying a degenerate core. As all the primers within the degenerate pool have the same clamping sequences, primer specificity is higher resulting in improved amplification (Rose *et al.*, 1998).

The CODEHOP program requires protein 'blocks' as input data in order to generate degenerate primers based on selected program parameters provided by the user. Unaligned laccase protein sequences (Table 5-1) were submitted to the online BlockMaker program (Henikoff *et al.*, 1995) and several blocks were generated. Related protein sequences tend to have several conserved regions in common as shown in Figure 5.3. BlockMaker uses two different algorithms to identify these conserved regions separated by unaligned sequence, and these are referred to as un-gapped blocks (Henikoff *et al.*, 1995). The two motif-finding algorithms used, MOTIF and GIBBS are based on different computational principles and therefore a block identified by both algorithms validates the integrity of such a block. Examples of two blocks identified by MOTIF and GIBBS is shown in Table 5-2. Two un-gapped blocks identified by MOTIF were selected and submitted to the online CODEHOP program. The chosen blocks included copper-binding sites (HWHG) located at position 126 to 129 as the forward primer target and (HWHG) at position 495 to 498 as the reverse primer target (Figure

5.2). A maximum degeneracy of 128-fold was selected and all other program parameters were kept at the default settings. The designed primers were designated LacF1 (32 fold degeneracy) and LacR1 (64 fold degeneracy) (see Table 2.4).

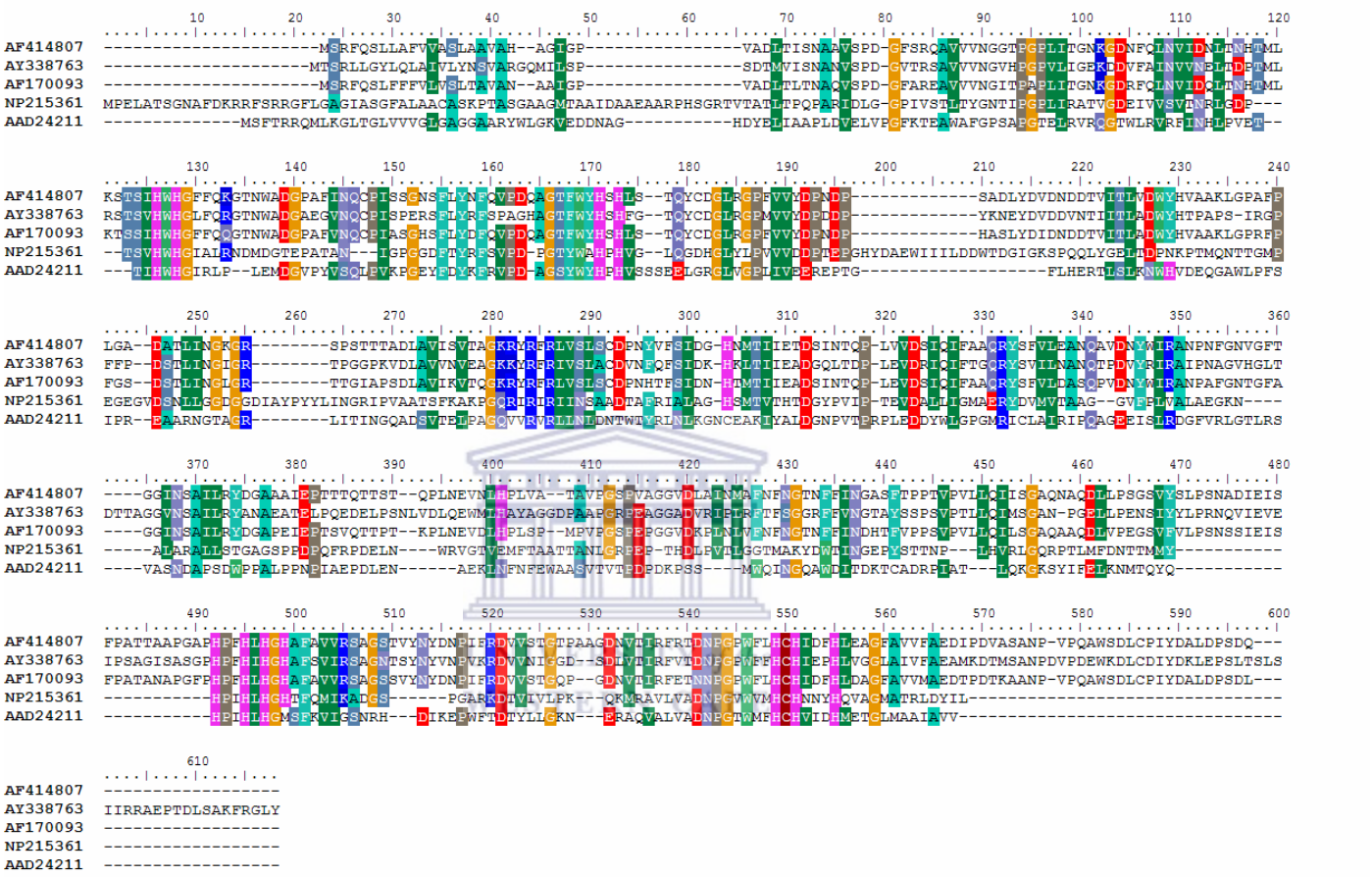

Figure 5.2 Target sites for degenerate primer design identified by aligning putative bacterial (Alexandre and Zhulin, 2000) and fungal laccase protein sequences obtained from GenBank (see Table 5.1).

Table 5-1 Laccase sequences used for the generation of protein blocks

Species	NCBI Accession no.	E-value/score	No. of AA	Function	Function experimentally verified	Reference
T. pubescens	AF414807	0.0/951	520	laccase	yes	(Galhaup *et al.*, 2002)
C.cinerea	AY338763	0.0/1085	549	laccase	yes	(Hoegger *et al.*, 2004)
P. cinnabarinus	AF170093	0.0/1009	518	laccase	yes	(Otterbein *et al.*, 2000)
M. tuberculosis	NP215361	0.0/954	504	Probable oxidase	no	(Cole *et al.*, 1998)
P. putida	AAD24211	0.0/897	460	CumA precursor protein	yes	(Brouwers *et al.*, 1999)

Table 5-2 Blocks identified by MOTIF and GIBBS

NCBI Accession no.	MOTIF	GIBBS
AF170093	492 HPFHLHGHAF	492 HPFHLHGHAFAVVRSAGS
AF414807	492 HPFHLHGHAF	492 HPFHLHGHAFAVVRSAGS
AY338763	492 HPFHIHGHAF	492 HPFHIHGHAFSVIRSAGN
NP215361	492 HPIHLHGHTF	492 HPIHLHGHTFQMIKADGS
AAD24211	492 HPIHLHGMSF	492 HPIHLHGMSFKVIGSNRH

5.4.1.2 Verification of mco degenerate primer specificity using a laccase positive *Streptomyces* strain as a model organism

To verify the specificity of the multicopper oxidase degenerate primer set, and to determine the optimum conditions under which these primers can amplify the target DNA fragments with reasonably low background and a good yield, a *Streptomyces* strain displaying laccase activity (Dr. Stenico, personal communication) was used as a model organism. The organism designated *Streptomyces* RUA06 was provided by Prof. Burton (Dept. of Chemical engineering, University of Cape Town). Laccase activity was measured according to the guaiacol laccase assay as described by Guresir and co-workers (2005).

Genomic DNA was extracted from an overnight culture as described in section 2.4.8. 20ng genomic DNA was used as template in a touchdown PCR (section 2.7.1)

and a DNA fragment of the expected size (approximately 1000 bp) was detected (Figure 5.3). The amplicon was cloned into the pGEM® -T easy vector (this clone is referred to as StrepA06p*MCO* from here on).

The amplicon was sequenced using M13 forward and reverse primers (see section 2.4.3 and Table 2-4). Sequence analysis demonstrated that the size of the amplicon was 900 bp. The primer sequences for both LacF1 and LacR1 (see Table 2-4) could be identified at the ends of the amplicon, and as the amino acid sequences for the degenerate cores were known, the open reading frame (ORF) could be identified. The obtained 300 amino acid sequence encoded by the ORF was compared by BLAST analysis against the NCBI database using Blastp (Altschul *et al*., 1990). The BLAST search results are summarized in Table 5-3, with the cloned gene fragment showing the most significant match to a Mn (II)-oxidation-associated multicopper oxidase from a *Pseudomonas* species (94% identity at amino acid level). Manganese dependent laccase activity was demonstrated for this enzyme (Francis and Tebo, 2001). As no other amplicons were obtained from the *Streptomyces* genomic DNA, the primers were considered multicopper oxidase specific.

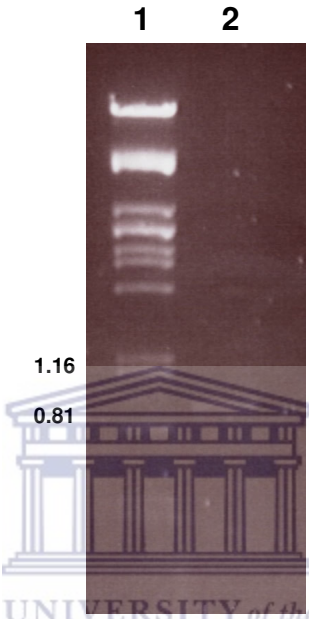

Figure 5.3 The detection of a multicopper oxidase gene fragment in the laccase positive model organism *Streptomyces* RUA06. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) 900bp *mco* fragment amplified from 20ng *Streptomyces* RUA06 genomic DNA.

5.4.1.3 Detection of bacterial mco gene fragments in grape waste and compost environmental DNA

To detect the presence of bacterial *mco* genes, the DNA samples were screened by PCR using the newly designed *mco* degenerate primers under the conditions as described in section 2.7.1. In both environmental DNA samples, two distinct DNA fragments were amplified, a fragment in the expected size range of 1Kbp (based on published putative bacterial *mco* gene sequences), as well as a smaller fragment of approximately 500bp in size (Figure 5.4). All the amplicons were cloned into the pGEM® -T easy vector as described in section 2.7.2.

As the degenerate primers were thought likely to detect a number of different *mco* gene fragments, 36 white colonies were picked from each plate after blue white selection. After replica plating, the colonies were directly screened by PCR (section 2.4.2) using the vector primers; m13 forward and m13 reverse (Table 2-4).

Unique amplicons were identified prior to sequencing by restriction fragment length polymorphism (RFLP) analysis. Amplicons were digested with *Rsa*I and the products of the digestion were resolved by agarose gel electrophoresis (Figure 5.5). Based on the restriction profiles, 5 unique clones were identified from the compost DNA sample and 4 from grape waste DNA. All clones were sequenced and the translated amino acid sequences were compared by BLAST analysis against the NCBI protein database (Altschul *et al*., 1990). Based on the BLAST results obtained the clones were designated as p*MCO*1 - 5 (compost) and p*MCO*6 to 9 (grape waste). The results are summarized in Table 5-3..

5.4.1.4 Characterization of putative bacterial mco gene fragments

BLAST results identified all clones (p*MCO*1 - 9 listed in Table 5-3) as putative bacterial multicopper oxidases, with the percentage identity at amino acid level ranging from 34% (p*MCO*4) to 92% (p*MCO*9). Based on the low E-values obtained for each of the best alignments, the BLAST results were considered statistically significant and therefore reliable and informative. It is particularly noteworthy that a strong similarity exists between bacterial multicopper oxidases and clones p*MCO*6, p*MCO*8 and p*MCO*9. This was an unexpected result as these gene fragments are significantly smaller than expected for bacterial multicopper oxidases (and laccases in particular) based on the published sequence information. As discussed previously, the designed degenerate primers were likely to detect bacterial copper resistance proteins due to the conservation of the copper binding domains amongst multicopper oxidases. This was verified by the seemingly strong similarity of p*MCO*7 to a copper resistance protein (Table 5-3).

A multiple alignment was generated from clones p*MCO*1 to 9 and laccase protein sequences used in the primer design (Table 5-1), using the ClustalW algorithm (Thompson *et al.*, 1994). The alignment illustrates the diversity amongst the isolated fragments at amino acid level with the histidine rich domains showing the only significant degree of conservation (Figure 5.6). These highly conserved residues have been identified as playing a central role in copper ion binding (Solano *et al.*, 2001). The alignment further demonstrates the novelty of the isolated gene fragments, with very little sequence homology between the isolated putative multicopper oxidases and other published *mco* protein sequences.

To investigate the phylogenetic relationship of the newly isolated *mco* gene fragments, a distance tree was constructed by employing the neighbor joining method and using the Phylowin software package (Figure 5.7). The topography of the tree shows the presence of two distinct clades. It demonstrates the division of p*MCO*3, p*MCO*4, p*MCO*StrepA06 and the putative bacterial laccase identified in *P. putida* from the rest of the multicopper oxidase genes. The isolated *mco* gene fragments seem to be more closely related to fungal laccase (laccase-Q12541, Figure 5.7), but is not supported by a significant bootstrap value. Several significant subgroups are, however evident, including an apparent copper resistance clade (CRP-yp191102, putative *mco* genes identified in *X. campestris*, *P. syringae, E. coli* and p*MCO*7) and a clade grouping clones p*MCO*6, 8 and 9. The grouping and apparent isolation of the members of this clade from the rest of the putative bacterial multicopper oxidases was expected due to their small size (see 5.4.1.3). To the best of the author's knowledge, no reports have described the detection and identification of this group of small putative bacterial *mco* genes. Characterization of the function of these genes requires the isolation and expression of the full-length open reading frames. The isolation of the open reading frame for p*MCO*8 is described in section 5.4.2.1.

To further clarify the putative function of the isolated gene fragments, the protein sequences were investigated for the presence of four laccase specific signature sequences (Kumar *et al*., 2003). Table 5-4 shows the comparison between laccase signature sequences (L1 to L3) to those identified in the clones. A comparison could not be made for signature sequence L4 (Table 1-6) as this region was not amplified due to its downstream location with respect to the reverse degenerate primer binding site. The

identified signature L1 sequences show very strong similarity to the expected laccase pattern with the only exception. The intervening sequence designated as X is five amino acids long in all clones, two amino acids shorter than that typical for fungal laccases. This same pattern is observed in the putative bacterial laccase sequences identified by Alexandre and Zhulin (2000) .

Signature sequence L2 also showed remarkable similarity between the clone and laccase sequences. With the exception of clones 1,2,4 and 7, the intervening residues denoted as X were two residues longer in all other clones as compared to the typical laccase pattern. The L3 signature sequences identified in all clones conformed perfectly to that of the conserved laccase pattern.

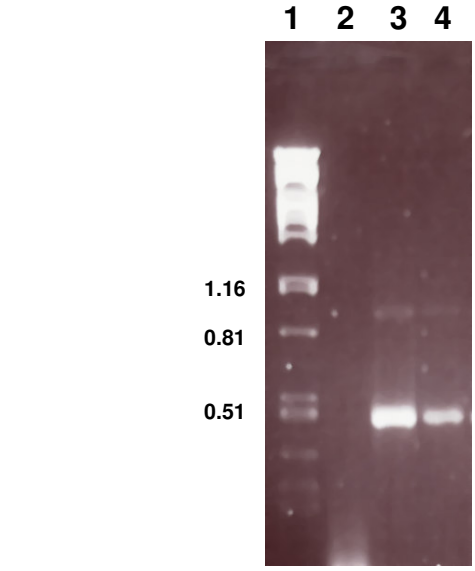

Figure 5.4 The detection of bacterial multicopper oxidases in environmental DNA samples. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Negative PCR control (no DNA). Lane 3) Grape waste 4) Compost.

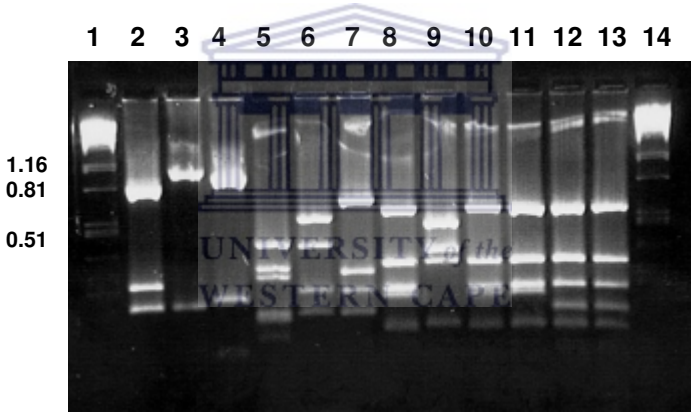

Figure 5.5 RFLP profiles of randomly selected clones screened by *Rsa*I digestion (5U) and separated on a 2% agarose gel. Lane 1 and 14) Lambda PstI molecular weight size marker (Kbp). Lanes 2 to 13) Compost 1Kbp fragment clones.

Table 5-3 BLAST results of multicopper oxidase gene fragments

BLAST Results							
Clone	**Fragment length (bp)**	**Sequence (no AA)**	**Organism**	**Putative Function**	**Bits**	**E-value**	**% Identity**
p*MCO*1	1032	344	*Nocardia farcinica*	putative multicopper oxidase	312	9e-84	48
p*MCO*2	1017	339	*Rhodobacter capsulatus*	putative multicopper oxidase	294	2e-78	47
p*MCO*3	998	332	*Mesorhizobium sp.* BNC1	putative multicopper oxidase	347	2e-94	50
p*MCO*4	1014	338	*Rubrobacter xylanophilus*	putative multicopper oxidase	186	5e-46	34
p*MCO*5	987	329	*Mesorhizobium loti*	metallo-oxidoreductase	206	5e-52	37
p*MCO*6	423	141	*Burkholderia cepacia*	putative multicopper oxidase	207	6e-53	62
p*MCO*7	1146	382	*Ornithobacterium rhinotracheale*	copper resistance protein	444	e-123	55
p*MCO*8	423	141	*Burkholderia cepacia*	putative multicopper oxidase	144	9e-34	63
p*MCO*9	429	143	*Rubrivivax gelatinosus*	putative multicopper oxidase	292	1e-78	92
StrepA06p*MCO*	900	300	*Pseudomonas* sp. ISO6	Mn(II)-oxidation-associated multicopper oxidase	603	1e-171	94

Figure 5.6 Alignment of isolated putative multicopper oxidase and laccase amino acids sequences.

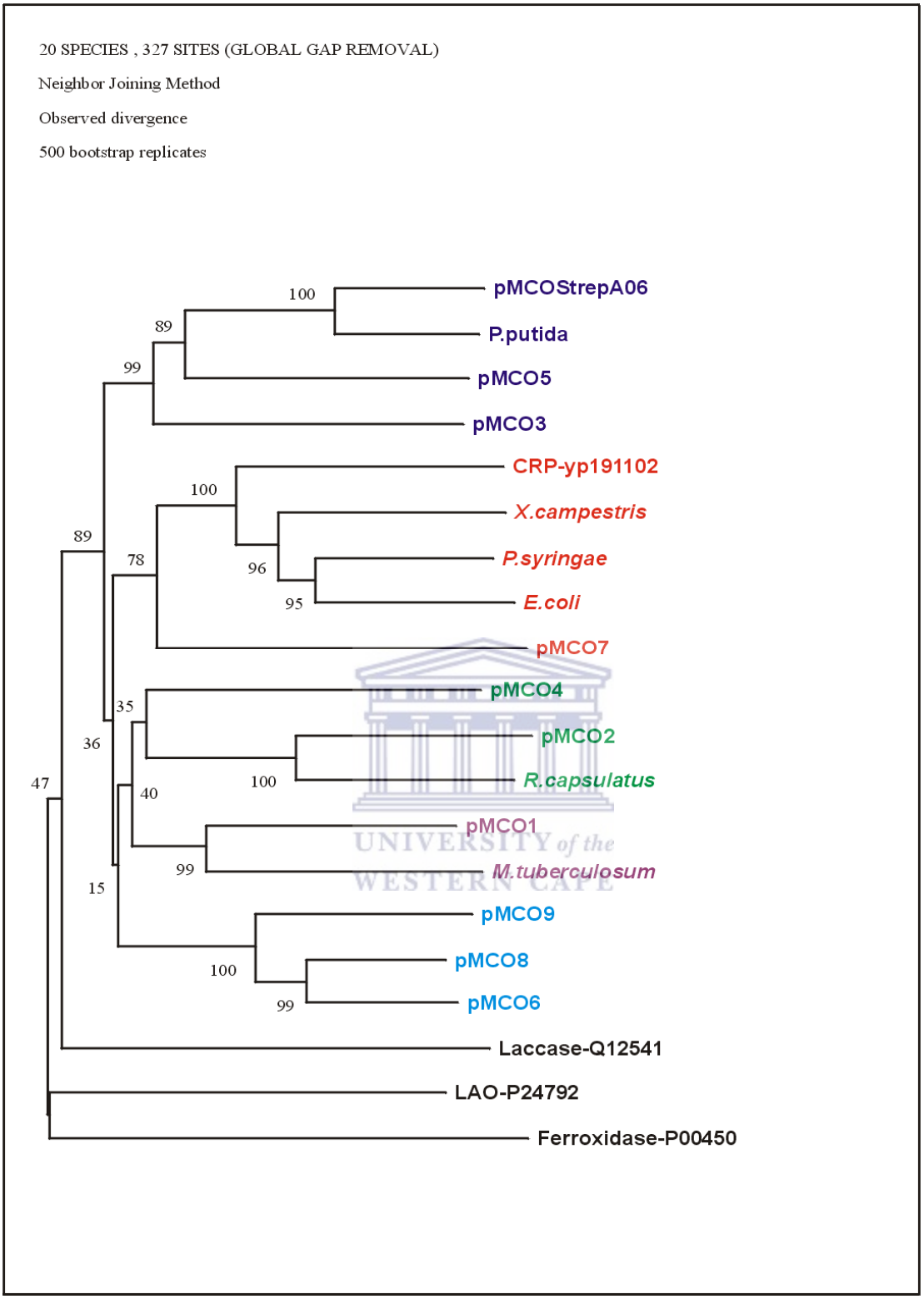

Figure 5.7 The phylogenetic relationship between newly isolated *mco* gene fragments and representative members of the multicoper oxidase gene family.

Table 5-4 Laccase signature sequences L1 to L3 identified in isolated multicopper oxidase fragments

Laccase signature sequences*			
Clone	L1	L2	L3
Laccase	H-W-H-G-X_9-D-G-X_5-QCPI	G-T-X-W-Y-H-S-H-X_3-Q-Y-C-X-D-G-L-X-G-X-(FILM)	H-P-X-H-L-H-G-H
p*MCO*1	H-W-H-G-X_7-D-G-X_5-QAPI	G-T-X-W-F-H-P-H-X_3-Q-L-D-R-A-L-X-A-X-(LIIE)	H-P-X-H-L-H-G-H
p*MCO*2	H-W-H-G-X_7-D-G-X_5-IPLI	G-T-X-W-M-H-S-H-X_3-Q-E-Q-A-L-L-X-A-X-(LIVR)	H-P-X-H-L-H-G-H
p*MCO*3	H-W-H-G-X_7-D-G-X_5-QKPI	G-T-X-W-Y-H-P-H-X_5-Q-V-G-X-G-L-X-G-X-(LILE)	H-P-X-H-L-H-G-H
p*MCO*4	H-W-H-G-X_7-D-G-X_5-QAPM	G-T-X-W-Y-H-S-H-X_3-Q-E-Q-T-G-I-X-G-X-(LIIE)	H-P-X-H-L-H-G-H
p*MCO*5	H-W-H-G-X_7-D-G-X_5-QPPV	G-T-X-F-F- H-S-H- X_5-Q-I-G-X-G-I-X-G-X-(LIIE)	H-P-X-H-L-H-G-H
p*MCO*6	H-W-H-G-X_7-D-G-X_5-QPPI	G-T-X-M-Y-H-P-H-X_5- Q-L-A-M-G-M-X-G-X-(FIIH)	H-P-X-H-L-H-G-H
p*MCO*7	H-W-H-G-X_7-D-G-X_5-TAPV	G-T-X-W-Y-H-S-H-X_3-Q-E-Q-V-G-M-X-G-X-(IVIY)	H-P-X-H-L-H-G-H
p*MCO*8	H-W-H-G-X_7-D-G-X_5-QPHI	G-T-X-M-Y-H-P-H-X_5-Q-M-X-L-G-M-X-G-X-(FIIH)	H-P-X-H-L-H-G-H
p*MCO*9	H-W-H-G-X_7-D-G-X_5-QPAI	G-T-X-M-Y-H-P-H-X_5-Q-M-X-M-G-M-X-G-X-(WYTH)	H-P-X-H-L-H-G-H
p*MCO*StrepA06	H-W-H-G-X_7-D-G-X_5-QLPV	G-S-X-W-Y-H-P-H-X_5-E-L-G-R-G-I-V-G-L-X-(LIVE)	H-P-X-H-L-H-G-H

* Laccase signature L4 not shown due to the length of obtained gene fragments.

5.4.2 Semi-nested random primed PCR as a method for the recovery of the full-length genes from metagenomic DNA

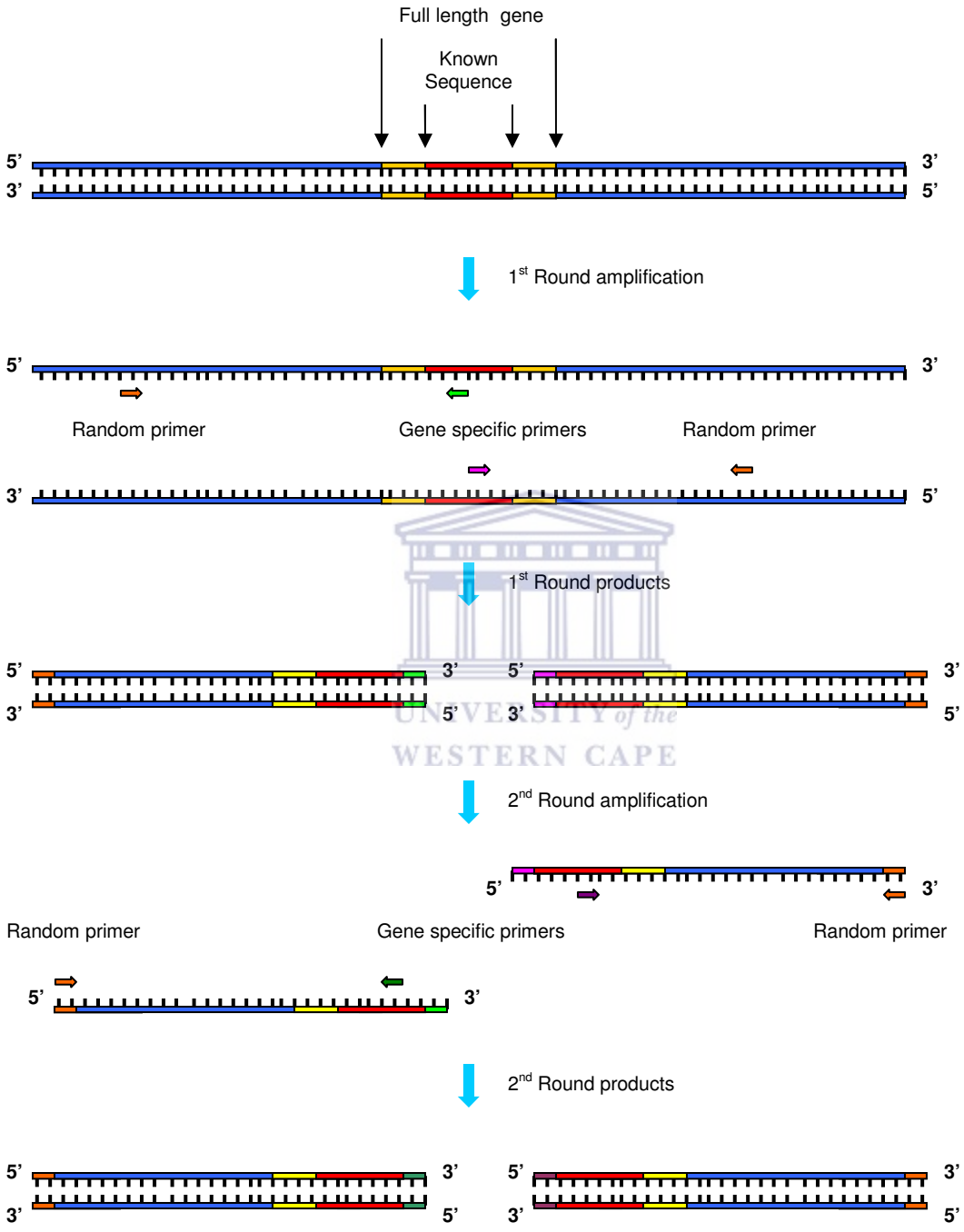

Figure 5.8 Schematic diagram of the semi-nested random primed PCR strategy employed for the recovery of *mco* gene fragment flanking regions from environmental DNA.

5.4.2.1 The amplification of 5' and 3' flanking regions of pMCO7 via thermal asymmetric interlaced PCR

To experimentally verify and demonstrate the function of the detected putative multicopper oxidases, full-length copies of the genes had to be isolated for expression and characterization purposes. In an attempt to recover a full-length copy of p*MCO7*, a semi-nested random primed PCR technique was employed. The approach is schematically depicted in Figure 5.8. In summary, to recover the 5' and 3' flanking regions from an isolated p*MCO7* gene fragment, a set of nested forward (recovers 3' end) and reverse primers (recovers 5' end) are used in combination with arbitrary random primers in successive reactions with the product of each round of amplification acting as the template for the next. In this approach the aim is to obtain random priming up and downstream from the full-length gene via the arbitrary degenerate primers. This allows the amplification of the target flanking regions. Typically after several rounds of amplification discrete amplicons are obtained. These fragments can then be cloned and sequenced to verify the recovery of the flanking regions. This approach was first employed on a metagenomic scale by Eschenfeldt and co-workers (2001) as discussed in chapter 1 (section 1.5.1).

Due to the technical difficulties associated with PCR amplification from excessively sheared DNA, clone p*MCO7* was chosen as a target for recovery because of the availability of high molecular weight grape waste genomic DNA (section 3.4.1.1).

Six nested p*MCO7* specific primers were designed and designated G4NF1 to 3, and G4NR1 to 3 (see Table 2-4). For random priming, nine arbitrary degenerate primers designated AD1 to 5, and NAD1 to 4 (see Table 2-4) were used, as described by Liu

and Whittier (1995). All attempts to employ the protocol as described by Eschenfeldt and co-workers (2001) failed, with no discrete amplicons observed, even after three successive amplifications (results not shown). This was presumed to be due to the complexity of the grape waste metagenomic DNA sample cuasing excessive miss priming.

In an effort to reduce non-specific background amplification in the second and third round amplifications, a thermal asymmetric interlaced PCR (TAIL-PCR) approach was modified and employed (Liu and Whittier, 1995). In this protocol, two high temperature gene specific annealing and amplification steps are performed for each low temperature annealing randomly primed step, which typically results in increased specificity (see section 2.7.3).

In the first round amplification, gene specific primers G4NF1 and G4NR1 were used in parallel reactions in combination with random primers AD1 to 5, and NAD1 to 4 (see Table 2-4). After amplification a 1 to 50 dilution was made from the PCR product and used as template in the second round amplification with primers G4NF2 and G4NR2, respectively in combination with the same random primer used during the first round. After two rounds of amplification a discrete amplicon in the 500bp size range was obtained from reactions G4NR2/NAD1 and G4NR2/NAD3. Reactions G4NR2/NAD2 and G4NR2/NAD4 resulted in the specific amplification of an amplicon approximately 200bp in size (Figure 5.9). Both fragments were cloned into the pTZ57R/T vector (Table 2-3) and sequenced with the m13f primer (Table 2-4). The sequences obtained from amplicons G4NR2/NAD1 and G4NR2/NAD3 were 466bp in size and were primed from G4NR2. This sequence was aligned to the p*MCO7* internal fragment and a 193bp

overlapped sequence was generated (Figure 5.11). Surprisingly, the amplicons derived from G4NR2/NAD2 and G4NR2/NAD4 showed no significant similarity to p*MCO*7, and are likely to have resulted from the non-specific priming of G4NR2.

A similar approach was followed for the recovery of the 3' flanking region, except for an additional third round amplification that was required to generate discrete amplification. All reactions with G4NF3 showed the amplification of the same sized amplicon (419bp) which was subsequently cloned and sequenced as described previously (Figure 5.10). This amplicon also resulted from the singular priming of G4NF3 and a 51bp overlap with p*MCO*7 was generated (Figure 5.11). The largest open reading frame from the reconstructed p*MCO*7 gene was translated. BLAST analysis demonstrated that p*MCO*7 was significantly similar to a copper resistance protein identified in *Ornithobacterium rhinotracheale,* with 54% identity at amino acid level. An alignment between these two protein sequences showed that the TAIL-PCR was unsuccessful in recovering the full-length copy of p*MCO*7 (Figure 5.12). This conclusion was based on the failure to identify the start codon and ribosomal binding site as well as the termination sequence could not be identified in the reconstructed p*MCO*7 open reading frame (Figure 5.11). From these results it is clear that both gene specific primers missprimed within the gene and therefore did not recover the entire open reading frame.

During the second round amplification using nested forward primer G4NF2, several larger discrete amplicons were generated (Figure 5.10), from sequencing results it was concluded that these amplicons were also due to non-specific miss-priming of G4NF2 (results not shown). In summary, it is concluded that although

successive rounds of amplification were essential to reduce non-specific mispriming, they also resulted in the preferential amplification of smaller DNA fragments that did not necessarily includes the flanking regions (Figure 5.9).

In an attempt to recover the remainder of the p*MCO7* gene the TAIL-PCR approach was slightly modified. Based on the experimental evidence accumulated it was suggested that recovery of gene fragments large enough to constitute the full-length gene would be obtained during the first round amplification provided that a significant amount of the background misspriming could be reduced. The main source of background miss-priming is a result of the complexity of the metagenomic DNA sample and the nature of arbitrary degenerate primers binding randomly. Although the use of random priming is integral to the principle of TAIL-PCR, the propensity for background amplification seems to be counter-productive for flanking region recovery from a metagenomic DNA sample with such a high degree of complexity.

To reduce background misspriming and increase specificity, a simpler approach using the same amplification conditions as described in section 2.7.3 was used with nested primers G4NF3 and G4NR3 during a single round amplification on the grape waste metagenomic DNA. Although a number of distinct amplicons were generated for both reactions (Figure 5.13), sequencing results showed no similarity to p*MCO7* for any of the amplicons (results not shown).

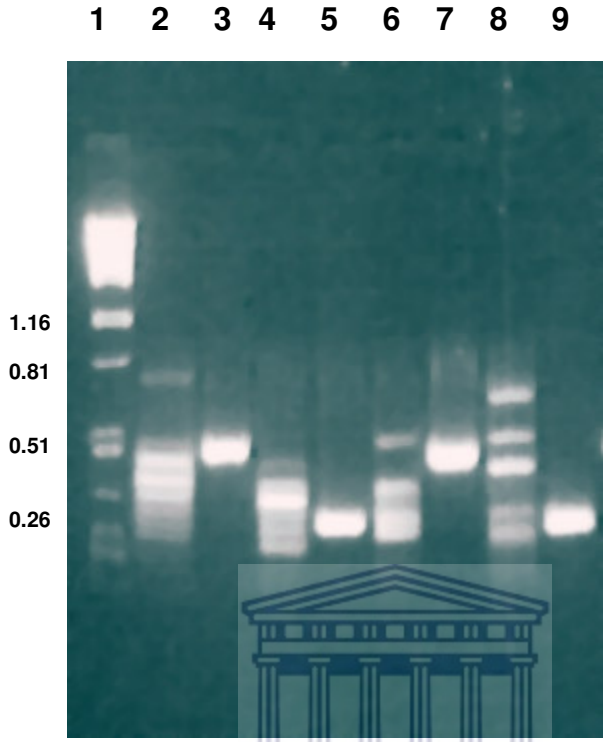

Figure 5.9 Thermal asymmetric interlaced PCR for the recovery of p*MCO7* 5' end. PCR reactions were separated on a 2% agarose gel. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) G4NR1/NAD1. Lane 3) G4NR2/NAD1. Lane 4) G4NR1/NAD2. Lane 5) G4NR2/NAD2. Lane 6) G4NR1/NAD3. Lane 7) G4NR2/NAD3. Lane 8) G4NR1/NAD4. Lane 5) G4NR2/NAD4.

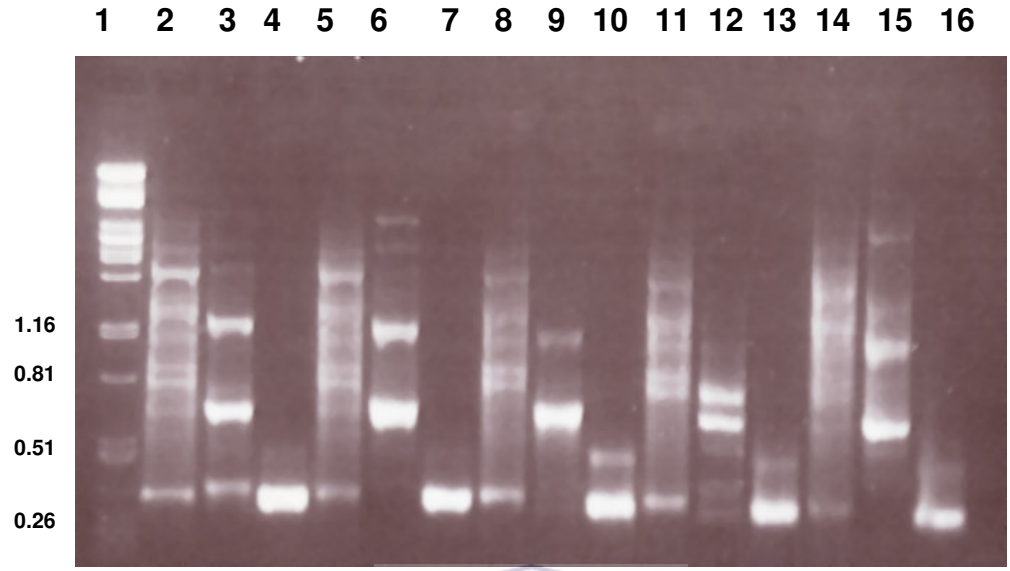

Figure 5.10 Thermal asymmetric interlaced PCR for the recovery of p*MCO7* 3' end. PCR reactions were separated on a 2% agarose gel. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) G4NF1/AD1. Lane 3) G4NF2/AD1. Lane 4) G4NF3/AD1. Lane 5) G4NF1/AD2. Lane 6) G4NF2/AD2. Lane 7) G4NF3/AD2. Lane 8) G4NF1/AD3. Lane 9) G4NF2/AD3. Lane 10) G4NF3/AD3. Lane 11) G4NF1/AD4. Lane 12) G4NF2/AD4. Lane 13) G4NF3/AD4. Lane 14) G4NF1/AD5. Lane 15) G4NF2/AD5. Lane 16) G4NF3/AD5.

Figure 5.11 Reconstruction of p*MCO7*.

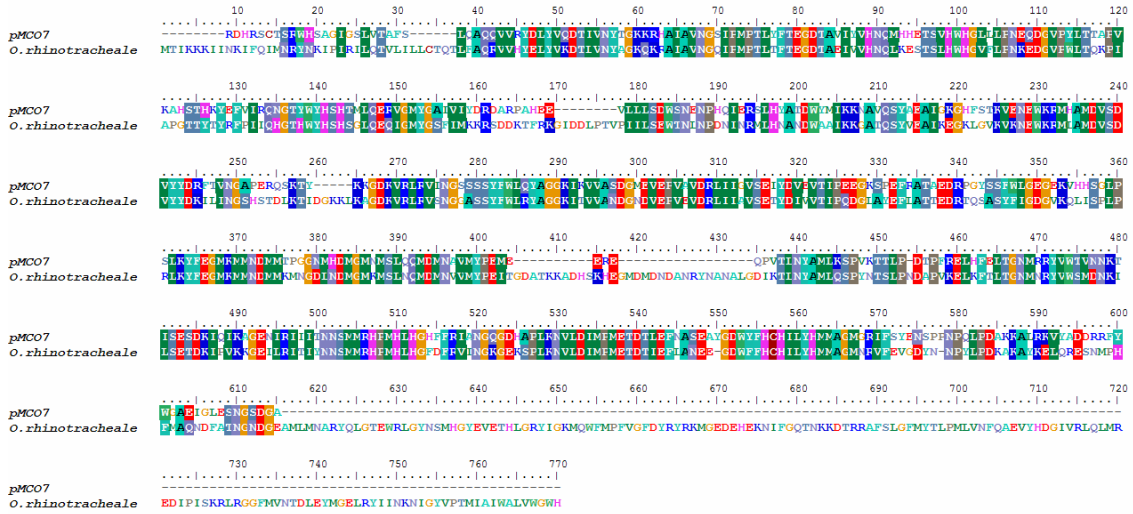

Figure 5.12 Alignment of p*MCO7* and *Ornithobacterium rhinotracheale* copper resistance protein (genbank Accession number AAT09359).

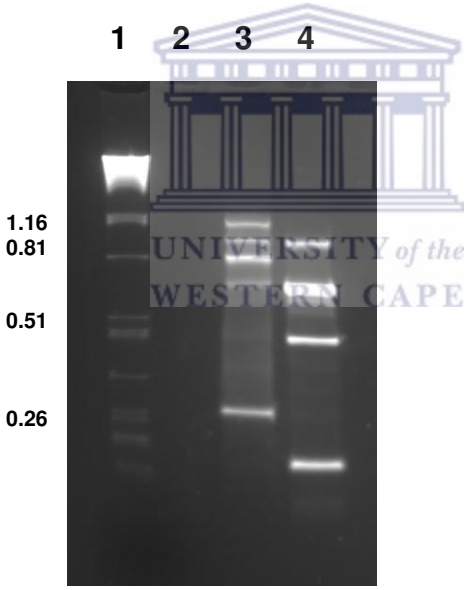

Figure 5.13 Single round thermal asymmetric interlaced PCR of G4NF3 and G4NR3. PCR reactions were separated on a 2% agarose gel. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) PCR negative control Lane 3) G4NF3. Lane 4) G4NR3.

5.4.3 Investigation of the universal Vectorette™ system as a method for the recovery of the full-length genes from metagenomic DNA

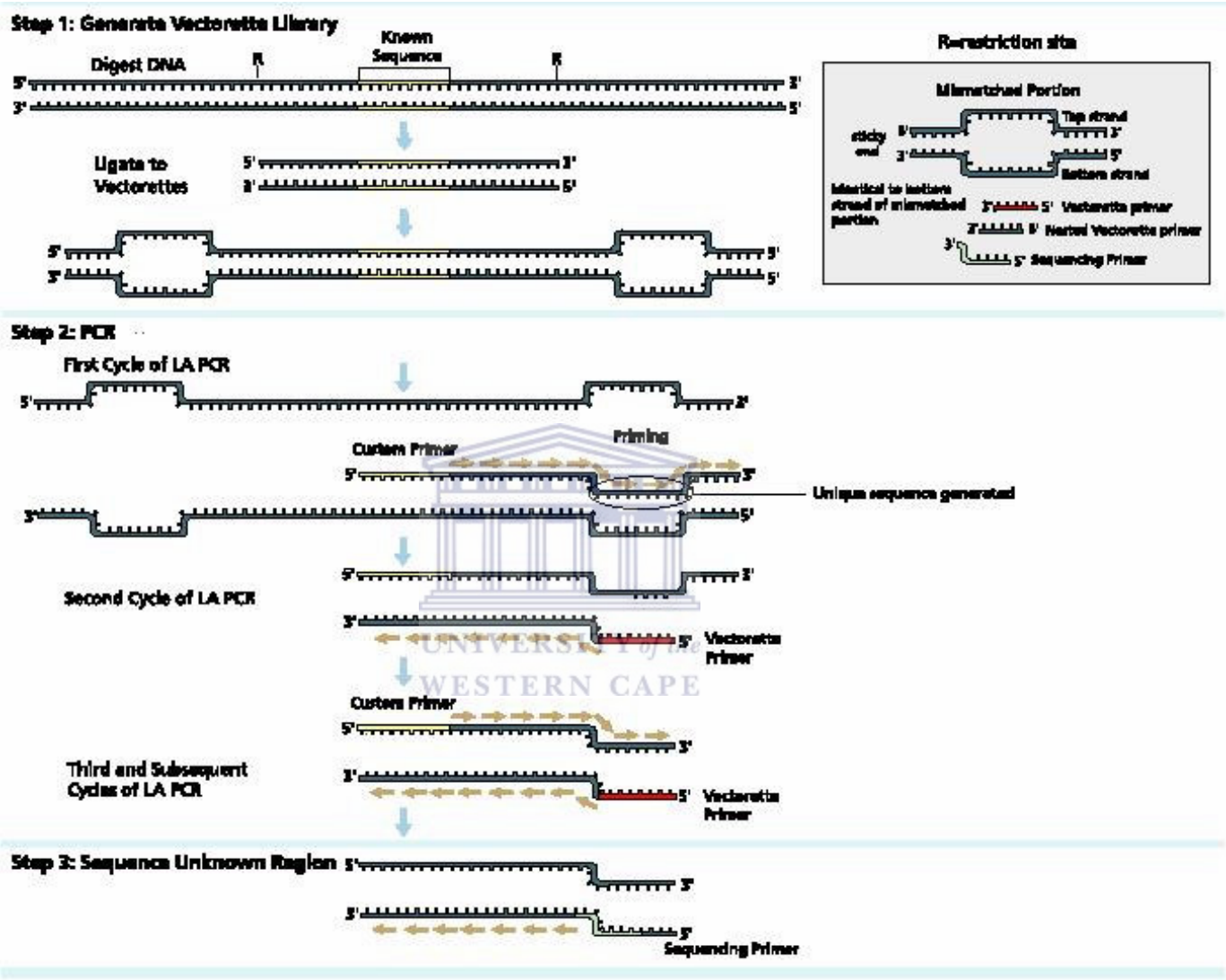

Figure 5.14 Schematic of the Universal Vecorette™ system strategy employed for gene walking. Figure was obtained from the Vectorette™ manual.

As an alternative approach to attempt the recovery of full-length *MCO* genes from the grape waste metagenomic DNA sample the universal Vectorette™ system (Sigma) was investigated. The schematic diagram depicted in Figure 5.14 demonstrates the principle of this strategy. This approach is essentially similar to gene walking through ligation mediated PCR. However a significant difference is the increased specificity obtainable as a result of the specifically designed mismatched region within the adapter sequence. This prevents non-specific amplification from the Vectorette ™ primer, and only once the gene specific primer has primed and reconstructed the Vectorette™ primer priming site can bi-directional amplification occur.

5.4.3.1 Grape waste metagenomic DNA Vectorette™ library construction

Prior to the construction of the Vectorette™ DNA library, the sequences of the recovered mco gene fragments were analysed for the presence of the restriction enzyme recognition sites listed in the manufacturers manual as suitable for use in the digestion of the target genomic DNA (see 2.7.4). This was done to prevent choosing a restriction enzyme that had a recognition site within a mco gene fragment, as this could hamper attempts to recover the full-length gene. HindIII was identified as the most appropriate choice. Approximately 500ng metagenomic DNA was digested overnight with HindIII as described in section 2.4.1. As a digestion and subsequent ligation control ,1 µg phage lambda genomic DNA was also digested. The completion of the digestion reactions were monitored by agarose gel electrophoresis (see 2.3.2). The commercially provided adapter with HindIII compatible sticky ends was ligated to grape waste metagenomic DNA as described in section 2.7.4. The results of the ligations are shown

in Figure 5.15. From these results it is evident that the adapter ligation reactions reached completion and that the majority of the DNA fragments should have acquired adapters at each end. This is evident from the similarity of the phage lambda ligation control to the HindIII digested phage lambda DNA. If adapter ligation was incomplete, high molecular phage lambda DNA would likely have been generated during the ligation reaction. After adapter ligation, DNA fragments of the grape waste metagenomic DNA Vectorette™ library ranged from 1 to 6 Kbp (Figure 5.15).

5.4.3.2 The amplification of 3' flanking regions of pMCO8 using Vectorette™ PCR

Based on the sequence information obtained for pMCO8 (section 5.4.1.3) a set of three nested forward and reverse primers were designed and designated as G7NF1 to G7NF 3 and G7NR1 to G7NR 3 (see Table 2-4). As the first round amplification reaction is generally the most critical step in the successful recovery of the flanking regions (Liu and Whittier, 1995), the specificity of primers G7NF1 and G7NR1 were validated by PCR with high molecular weight grape waste metagenomic DNA (20ng) as template. Based on the sequence information, amplification of a 296 bp DNA fragment was expected. Figure 5.16 shows the specific amplification of a DNA fragment in the expected size range. Sequence analysis of the DNA fragment confirmed its origin as part of pMCO8. As no other amplicons were visible this primer pair was considered specific and suitable for use in the vecorette PCR.

The first round Vectorette™ amplification was performed with the prepared library DNA as template using G7FN1 (for 3' end recovery) and G7NR1 (for 5' end recovery) in combination with the provided Vectorette™ primer (supplied by the

manufacturer). As controls, separate reactions were set up that included either the gene specific or Vectorette™ primers. This was done to enable the identification of non-specific background amplification and to identify probable amplicons containing the flanking regions as those amplions unique to the gene specific/ Vectorette™ primer combination were likely to be good targets for analysis. The result of the first round amplification is shown in Figure 5.17. No background amplification was generated by the Vectorette™ primer (lane 5) or G7NF1 (lane 4). Furthermore, no amplification was obtained in the combined reaction (lane 3). However, G7NR1 missprimed significantly and produced excessive background (lane 5). Although two seemingly unique amplicons were generated in the combined reaction with the G7NR1 and Vectorette™ primers, due to the large degree of G7NR1 background a nested second round PCR was performed. 1 in 50 dilutions of both reactions were prepared and used as the templates for the nested reaction. Primers G7NF2 and G7NR2 were used in combination with the Vectorette™ nested primer (supplied by the manufacturer) and the same controls were included as described previously. Figure 5.18 shows the result of the nested PCR.

In the control reaction for primer G7NR2 a significant degree of non-specific background amplification was again produced (lane 3). Although several amplicons were obtained in the G7NR2/Vectorette™ nested primer reaction (lane 2), it was impossible to distinguish possible targets for further analysis as a result of the high degree of G7NR2 misspriming. A third round amplification was therefore performed using G7NR7 and the Vectorette™ nested primers. Several seemingly unique

amplicons were generated Figure 5.19). These were cloned, sequenced and analysed but showed no significant similarity to pMCO8

A single amplicon was, however, generated in a PCR using G7FN2 and the Vectorette™ nested primer (Figure 5.18, lane 6). This amplicon was also cloned and sequenced. Further analysis showed that this amplicon (473 bp in size) did indeed constitute the 3' downstream region of fragment p*MCO*8, with a 177 bp overlap as shown in Figure 5.20**.** From the sequence information obtained the stop codon was identified as well as the proposed fourth laccase specific signature sequence constituting the copper binding site HCH located at position 545 relative to the alignment ruler (Figure 5.20).

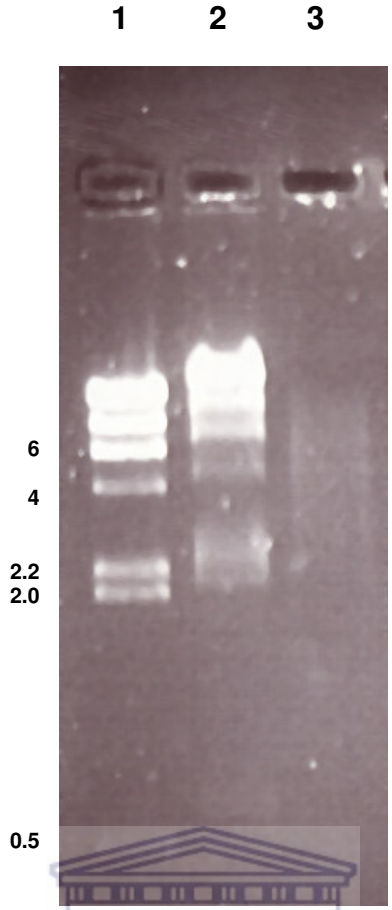

Figure 5.15 Grape waste metagenomic DNA Vectorette™ library construction. A sample of each ligation reaction was separated on a 1% agarose gel. Lane 1) Lambda *Hind*III molecular weight size marker (Kbp). Lane 2) Lambda *Hind*III ligation positive control. Lane 3) Grape waste metagenomic DNA Vectorette™ library.

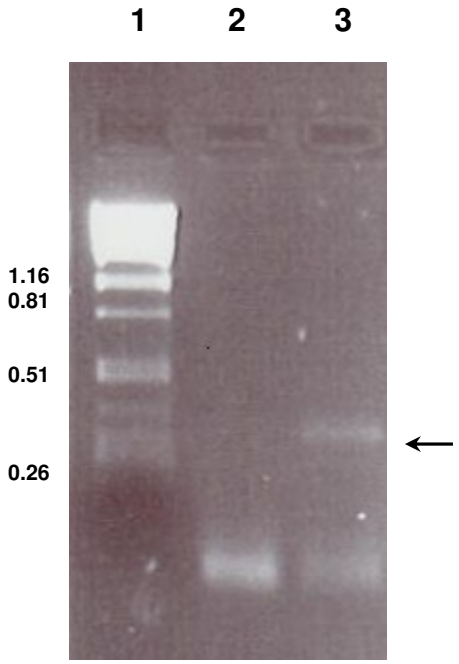

Figure 5.16 Validation of G7NF1 and G7NR1 gene specific primer specificity. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) PCR negative control Lane 3) G7NF1/G7NR1. Arrow indicated the position of the target amplicon.

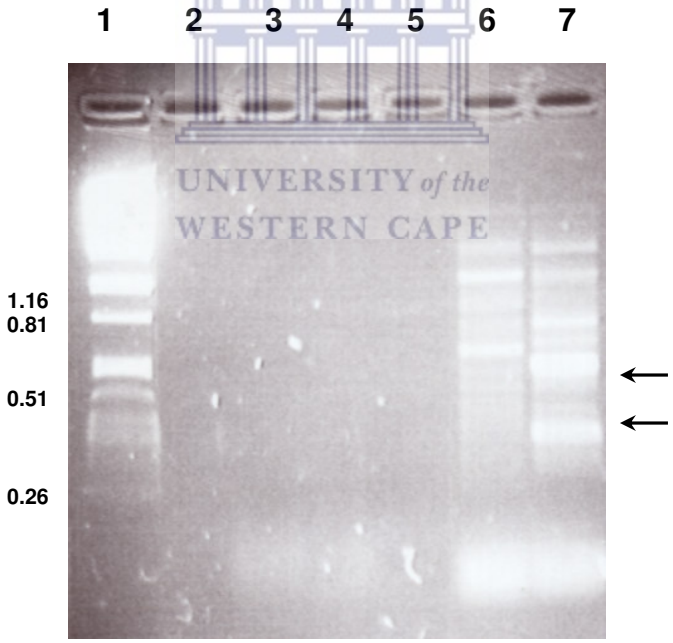

Figure 5.17 First round Vectorette™ PCR. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) empty. Lane 3) G7NF1/ Vectorette™ primer. Lane 4) G7NF1. Lane 5) Vectorette™ primer. Lane6) G7NR1. Lane 7) G7NR1/ Vectorette™ primer. Arrows indicate the position of possible target amplicons.

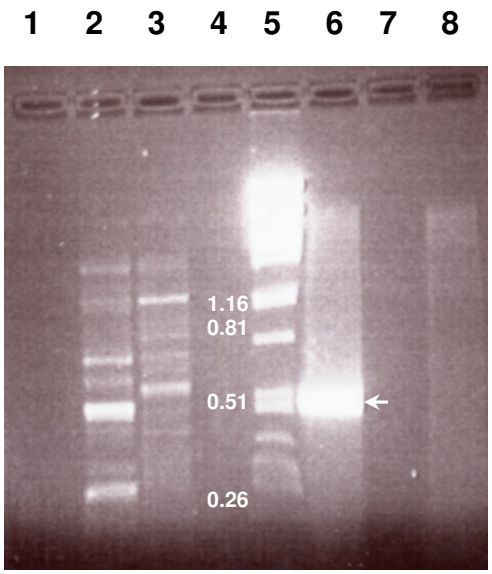

Figure 5.18 Second round Vectorette™ PCR. Lane 1) Empty. Lane 2) G7NR2 . Lane 3) G7NR2/ Vectorette™ nested primer. Lane 4) Empty. Lane 5) Lambda *Pst*I molecular weight size marker (Kbp). Lane 6) G7NF2/ Vectorette™ nested primer. Lane 7) G7NF2. Lane 8) Vecto**rette™** nested primer. Arrows indicate the position of possible target amplicons.

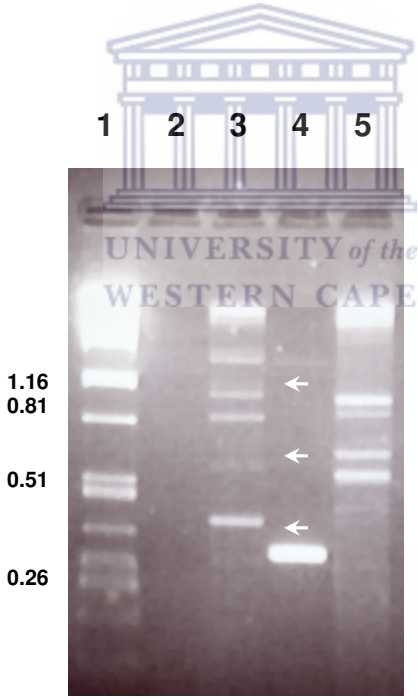

Figure 5.19 Third round Vectorette™ PCR. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Empty. Lane 3) G7NR3/ Vectorette™ nested primer. Lane 4) G7NR3. Lane 5) Vectorette™ nested primer. Arrows indicate the position of possible target amplicons.

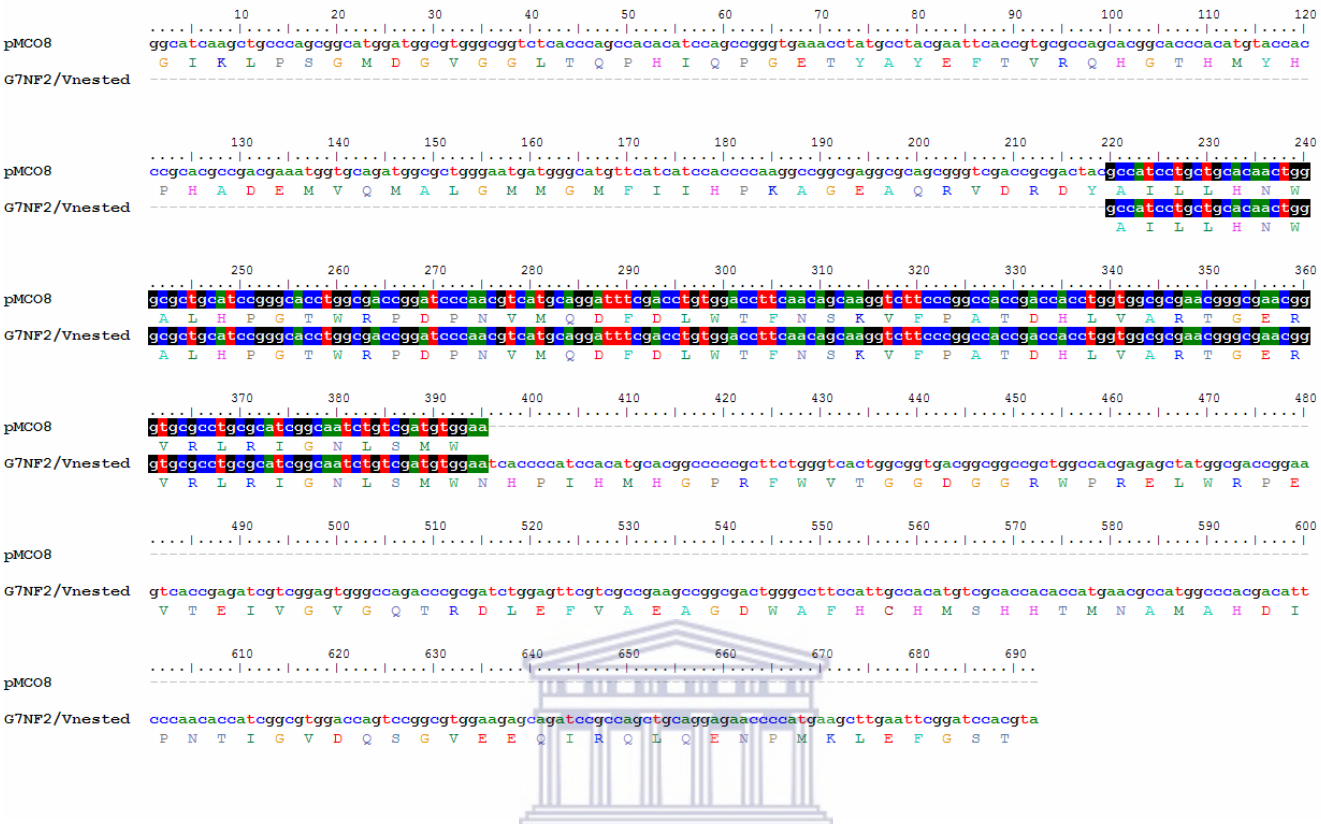

Figure 5.20 Reconstruction of p*MCO*8

5.4.4 Recovery of a putative full-length *mco* gene from a *Streptomyces* strain using subtractive hybridization magnetic bead capture technology (SH-MBC)

To expedite the recovery of bacterial genes, magnetic bead capture methodology (Jacobsen, 1995) was adapted to allow the rapid and efficient recovery of a specific target gene. As a proof of concept, a putative *mco* gene was recovered from a *Streptomyces* strain. Briefly, genomic 'tester' DNA was prepared by random fragmentation of the target genomic DNA into the appropriate size (1.5-5 Kbp). The DNA was prepared for TA cloning and adaptor priming sites were introduced to the ends to allow re-amplification. The *mco* gene fragment DNA ('driver') was prepared by biotin labeling of the 5' ends, and the driver was immobilized onto the magnetic beads through the irreversible streptavidin/biotin interaction. The *mco* gene-specific capture was performed by hybridization of the tester DNA to the prepared magnetic beads. After two stringency washes the beads were used as template in a PCR reaction and the full-length putative *mco* ORF was recovered.

5.4.4.1 Tester and driver DNA production

Streptomyces genomic tester DNA was prepared by partial *Rsa*I digestion as described in section 2.8.2. DNA fragments ranging between 1.5 and 5 Kbp (Figure 5.21) were recovered from the agarose gel after separation and prepared for TA cloning. These sized fragments were selected due to the fact that they were large enough to contain a full-length open reading frame but short enough to allow easy re-amplification. The first step in the DNA preparation was the introduction of adapter priming sites.

Adapter sequences were reconstructed and ligated to the 3'-A-tailed DNA fragments as described in section 2.8.1 (Table 2.4). To reconstruct the priming site on the adapter, it was necessary to first blunt-end the adaptor 5' overhang. The effective ligation of adapter priming sites to the tester DNA was verified by re-amplification of 20 ng of the prepared tester DNA (section 2.8.5) (Figure 5.22).

The biotin labeled *mco* gene fragments ('driver') were prepared by re-amplification of the StrepA06p*MCO* fragment using the biotin labeled LacF1-5'-bio primer as described in section 2.8.3 (Table 2.4) (Figure 5.23). To prepare the gene specific magnetic beads, 1µg of the driver DNA was denatured and immobilized onto the beads (section 2.8.3). As any unbound driver molecules present could have a deleterious effect on the efficiency of tester DNA hybridization to the beads, the excess driver DNA was also removed by repeated washing steps prior to hybridization.

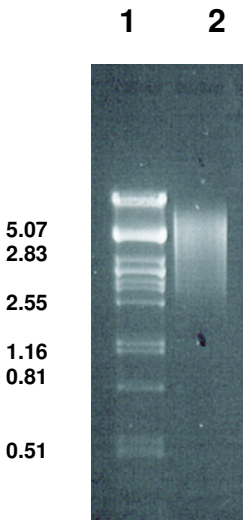

Figure 5.21 *Streptomyces* genomic DNA *Rsa*I partial digestion. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) *Streptomyces* genomic DNA digested with 0.5 U *Rsa*I for 10 minutes.

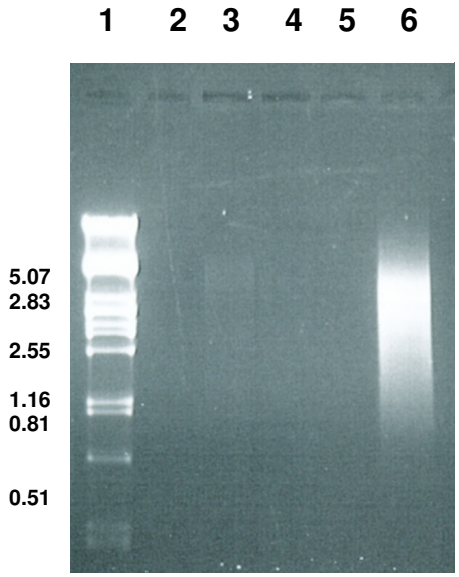

Figure 5.22 *Streptomyces* genomic tester DNA re-amplification using varying concentrations as template. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Negative control (no DNA). Lane 3) 10ng tester DNA. Lane 4) 5ng tester DNA. Lane 5) 1ng Tester DNA. Lane 6) 20ng tester DNA.

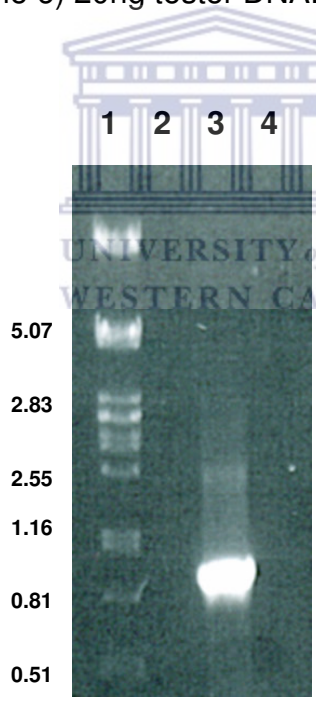

Figure 5.23 *Streptomyces mco* driver DNA production. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Negative control (no DNA). Lane 3) Biotin labeled *mco* driver DNA.

5.4.4.2 Gene specific capture and ORF characterization

Denatured *Streptomyces* 'tester' DNA was hybridized overnight to the *mco* probe labeled magnetic beads. To ensure maximum specificity, hybridization was performed at 68°C and all unbound 'tester' DNA was thoroughly removed by two high stringency washes. This was done to prevent any non-specific background amplification. The magnetic beads were resuspended in ultra high quality water and used directly in a PCR reaction as template. A T7 primer recognizing the adapter priming site was employed for re-amplification (Table 2.4). To ensure accurate and efficient amplification of the target gene fragments, a high fidelity thermal DNA polymerase (Bioline) was employed. A 1.4 Kbp amplicon was obtained after re-amplification (Figure 5.24).

Sequence analysis revealed that the 1418 bp amplicon carried a full-length putative *mco* ORF, 1215 bp. Curiously, this recovered gene was not the StrepA06p*mco* full-length ORF as expected, with the two genes showing only 80% identity at the nucleotide level (Figure 5.25). The results suggested that another putative *mco* gene was recovered from the *Streptomyces* genome. This is perhaps not as surprising if the high degree of DNA similarity is taken into account. The open reading frame could be translated into a 405 amino acid polypeptide that showed the presence of all four laccase signature sequences described in section 1.6.1.3 (Figure 5.26).

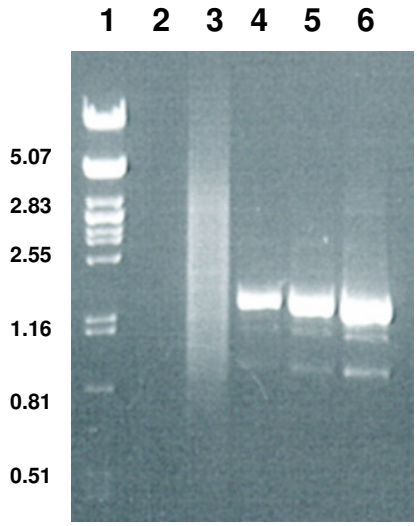

Figure 5.24 *Streptomyces* genomic tester DNA re-amplification after subtractive hybridization using varying amounts of magnetic beads as template. Lane 1) Lambda *Pst*I molecular weight size marker (Kbp). Lane 2) Negative control (no DNA). Lane 3) Positive control (20ng tester DNA). Lane 4) 1 µl beads. Lane 5) 2 µl beads Lane 6) 5 µl beads.

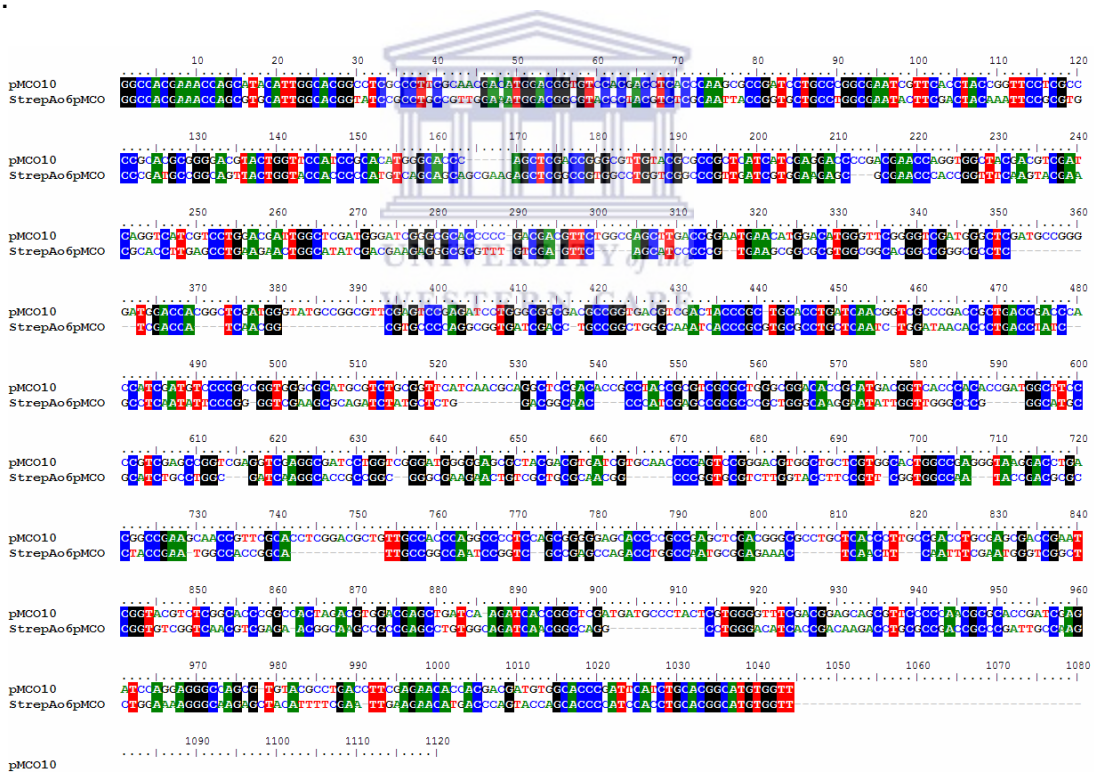

Figure 5.25 DNA sequence comparison of StrepA06p*MCO* and p*MCO*10

Figure 5.26 p*MCO*10 ORF recovered by SH-MBC technology

5.5 Discussion

The versatility and increasing importance of laccase in the biotechnology industry is demonstrated by its various applications. Some of the popular uses of the enzyme include pulp delignification, bleaching of textile dyes, addition as washing powder components, detoxification of industrial effluents and the removal of phenolics during the wine making process (Alexandre and Zhulin, 2000). To date, this has exclusively been performed by fungal laccases. This is not surprising considering the high yield and ease of purification that fungi offer as the majority of this enzyme is excreted into the growth media (Thurston, 1994).

Laccase activity has only recently been described in bacteria (Givaudan *et al.*, 1993; Sanchez-Amat *et al.*, 2001; Grass and Rensing, 2001; Ruijssenaars and Hartmans, 2004). However, mining of a non-redundant protein database by BLAST searching for bacterial proteins that show significant similarity to fungal laccases has identified several putative bacterial laccases, and the diverse bacterial taxa in which they were identified seem to suggest that these genes are perhaps widespread in bacteria (Alexandre and Zhulin, 2000).

Given the availability of sequence information for putative bacterial as well as fungal laccase genes, we aimed to design a set of degenerate primers that could be used for the specific detection of bacterial laccase genes in both isolated strains and in environmental DNA samples. One objective of the work was to experimentally verify their widespread occurrence in nature.

Using the CODEHOP program, a set of degenerate primers was designed. Each of the primers were designed to yield moderate degeneracy, ranging from 32 and 64

fold. In combining the primers in a single detection reaction, the cumulative degeneracy rose to 2048 fold (64 x 32) which should provide significant coverage of the bacterial laccase gene sequence space.

To verify the effectiveness and specificity of the primer set, a laccase positive *Streptomyces* strain was used as model organism for the PCR detection of the laccase gene. A single amplicon was generated and BLAST analysis confirmed its probable function as a Mn (II)-oxidation-associated multicopper oxidase, which has been previously demonstrated to possess manganese dependent laccase activity (Francis and Tebo, 2001). As no other amplicons were obtained from the *Streptomyces* genomic DNA, the primers were considered multicopper oxidase specific.

The PCR screening assay for bacterial laccase genes has several useful applications. Apart from the obvious detection of laccase genes in individual bacterial isolates, it could also be used to determine laccase gene distribution in the metagenome. To address this question, metagenomic DNA extracted from compost and grape waste was screened. In both cases amplification was specific with only two amplicons being generated, both of which were shown to consist of several putative bacterial multicopper oxidase gene fragments. In both samples, *mco* gene fragments somewhat smaller than expected were also amplified. Based on published sequence information available for putative bacterial *mco* genes as well as bacterial copper resistance proteins in the non-redundant protein databases, this is, to the best of the author's knowledge, the first isolation and description of this smaller type of bacterial multicopper oxidase gene. The true function of these as well as the 'typical' *mco* genes detected can only be determined by recovering the full-length genes and subsequent

expression and functional analysis. The *mco* gene fragments showed significant similarity to the proposed laccase signature sequences (Kumar *et al*., 2003). Although there were some variation in the length of intervening sequences within the signatures as compared to fungal laccases, these variations corresponded perfectly to the sequences of other putative bacterial laccases identified previously (Alexandre and Zhulin, 2000).

Taken together, the results confirmed the specificity of the PCR detection assay as all of the analysed fragments proved to be putative bacterial *mco* genes.

A further useful application for this detection assay would be for the monitoring of laccase expression within natural environments or as part of a bioremediation scheme by probing cDNA libraries prepared from these environments. As fungal laccase sequence information was used during the primer design, the expression of fungal laccases could therefore also be monitored in this way. The construction of an environmental cDNA library was reported recently (Grant *et al*., 2005). This approach may therefore well be an option in those instances where the direct enzymatic measurement of laccase is impeded by contaminants within the environmental sample.

The phylogenetic relationships (see Figure 5.7) between the detected *mco* gene fragments and known *mco* genes suggests that these gene fragments are evolutionarily closely related to fungal laccases.

The recovery of flanking DNA regions of a known DNA fragment is commonly undertaken by constructing a genomic DNA library and then screening by colony hybridization for a clone consisting of the full length sequence, often using the internal known DNA fragment as a probe (Sambrook and Russell, 2001). Unfortunately, this

approach is impractical when trying to recover a single copy gene from a metagenome, simply due to the tremendous DNA complexity inherent to metagenomic DNA samples. To construct a metagenomic DNA library suitably large to ensure complete coverage of a nutrient rich environmental sample (as in this case), would require the generation of libraries in excess of 10^5 clones with large inserts, and at least 10^7 clones using small insert libraries. Constructing and manually screening libraries of this size through hybridization would not only be very costly but also unrealistically time consuming.

To expedite the recovery of the full-length *mco* genes, two PCR based approaches were investigated; *i.e.* semi-nested random primed PCR and the commercially available Vectorette™ system.

The semi-nested random primed PCR technique had been employed previously in a metagenomic context to recover two novel forms of a 2,5 diketo –D- gluconic acid reductase (Eschenfeldt *et al*., 2001). All attempts to apply the protocol, as described by Eschenfeldt and co-workers, on the grape waste metagenomic DNA sample failed. Even after three successive rounds of semi-nested PCR amplification the degree of non-specific background was so high that no discrete amplicons could be identified. This problem might be ascribed to the complexity of this particular DNA sample, which contained both prokaryotic DNA and a significant amount of eukaryotic DNA (see chapter 3). Eschenfeldt and co-workers (2001), although not reporting on the exact level of diversity within their chosen DNA sample, did however suggest that the sample was not representative of the metagenome. They suggested that the DNA extraction protocol used may have produced a DNA sample representing a subset of the true biodiversity. The greater complexity of the grape waste metagenomic DNA sample is a

significant difference, which may explain the observed difficulty in reducing the non-specific background amplification.

In an attempt to reduce the lack of specificity, a thermal asymmetric interlaced PCR (TAIL – PCR) method was employed. This technique works on the same principle as the semi-nested random primed PCR approach, but differs in that two high temperature gene specific amplifications are introduced for every low temperature random primed annealing and amplification step (Liu and Whittier, 1995). In this way the specific amplification 'out-performs' non-specific background amplification and more discrete PCR products can be obtained. This technique was initially developed with great success to recover unknown flanking regions from P1 and YAC clones (Liu and Whittier, 1995). This method was further modified to recover the genomic sequences flanking T-DNA insertions in *Arabidopsis thaliana* (Liu *et al*., 1997b). No other reports thus far have described the use of this technique in metagenomic gene discovery.

This approach drastically improved the quality of the amplification and ultimately resulted in the recovery of significant portions of the flanking regions for clone pMCO7 (Figure 5.11). Although the full length open reading frame could not be recovered it was, however, demonstrated that this technique does provide a method for gene walking to recover the flanking regions of a single copy gene from such a complex metagenome. With the design of new gene specific primers closer to the ends and repeating the process it is quite likely that the full-length copy could have been recovered. From the results obtained, it was evident that both the 5' and 3' flanking regions were recovered by the sole amplification of the gene specific primer acting asymmetrically and not in combination with the random primer as expected. This may again reflect the effect of

the complexity of the DNA sample, by masking the specific amplicons present at a low concentration amongst the rest of the non-specific background that was inevitably generated.

The involvement of the random primer during the initial first round amplification was clear as evidenced by the generation of a high degree of background in the initial semi nested random primed PCR experiments as well as in the TAIL –PCR (see Figure 5.9 and Figure 5.10). Figure 5.13 illustrates the importance of the random primer acting to initiate amplification during the first stage of this procedure as elimination of this primer during the first round amplification failed to produce flanking region amplification. It seems likely that due to the nature of the amplification conditions, which favor priming by the gene specific primers, flanking region amplicons generated by the combination of gene specific and random primers play an important role by acting as template in the subsequent nested amplification reactions, even if present below the visible detection limit. The exact reason why these amplicons remain below the visible detection limit is uncertain. It is quite possible that the degree of degeneracy of the random primers was too high and as a result the concentration of the primer pool responsible for initiating specific amplification was too low to provide amplification of a visible signal. Experiments to increase the random primer concentration and thereby effectively raising the concentration of each primer in the degenerate pool resulted, as expected, in a significant increase in the non-specific amplification such that no discrete amplification could be obtained. A more viable approach would perhaps be to employ less degenerate random primers.

As an alternative approach, the universal Vectorette™ system for gene walking was also investigated. Due to the unexpected discovery of the smaller bacterial *mco* gene fragments (p*MCO*8, p*MCO*9 and p*MCO*6) and interest in their possible function, p*MCO*8 was chosen as a target for recovery. This system introduces priming sites sequences to the ends of the target genomic DNA by the ligation of specifically designed adaptors. As for the semi-nested random primed PCR approach, several rounds of nested amplifications are typically necessary to generate discrete amplicons. No amplification was obtained after the first round amplification in an attempt to recover the 3' end (Figure 5.17). However, according to the manufacturers, this is common occurrence when dealing with highly complex DNA samples. A second round amplification resulted in the generation of a single amplicon that proved to be the complete 3' end of p*MCO*8 (Figure 5.18 and Figure 5.20).

A similar attempt to recover the 5' end of pMCO8 failed, with the generation of significant non-specific background even after three successive rounds of nested amplification (Figure 5.17, Figure 5.18 and Figure 5.19). This was an unexpected result as the specificity of the gene specific primers used in the first round amplification was adequately demonstrated (Figure 5.16). The reason for this sudden lack of specificity is unknown. An unpurified fraction of the Vectorette™ library was used as template in the PCR as described by the manufacturer. It is possible that components within the reagents used during the library construction could be responsible for this observed loss of specificity. Nevertheless, it was demonstrated that the Vectorette™ system does offer a relatively simple means of recovering flanking regions from DNA samples as complex as metagenomic DNA.

It is apparent that it requires a considerable degree of effort and patience to recover full-length open reading frames from metagenomic DNA via PCR based approaches simply due to the inherent complexity of these types of DNA samples and the ease with which non-specific priming can occur in PCR as a result. This work however has demonstrated that despite these difficulties, in principle these techniques can be successfully applied to metagenomic gene discovery.

A more direct and rapid gene recovery technique was developed in an attempt to avoid some of the problems associated with gene specific PCR, particularly for metagenomic samples. Employing gene specific labeled magnetic beads in a subtractive hybridization assay was first demonstrated by Jacobsen (1995). This technique allowed the microscale detection of a *Pseudomonas fluorescens* genomic DNA, in soil, with a *lux* gene fragment as immobilized probe.

In this work, the technique was slightly modified to allow the specific recovery of the target gene and not the entire genome. The SH-MBC technique proved to be the most efficient and reliable gene discovery approach investigated. Although the putative *mco* gene was recovered from a single genome, and therefore validated on a much less complex DNA sample than those attempted in the PCR based technologies, the principle and practical feasibility of the approach was demonstrated. A natural evolution of this technology would be to employ a mixed pool of gene fragments derived from a metagenome as probe. This could potentially result in the recovery of numerous full or partial ORFs in a single step. It is anticipated, however, that the protocol would require some adaptation to compensate for the increased DNA complexity of a metagenomic sample. The possible sequence space that may be accessible in this way was

demonstrated by the cross reactive nature of DNA hybridization that resulted in the recovery of a different putative *mco* gene from the *Streptomyces* genome than that initially targeted.

Chapter 6: Concluding Remarks

6.1 Summary of findings

Chapter 3

- High quality genomic DNA was extracted from a grape waste sample using chemical and mechanical DNA extraction methods. Although the mechanical extraction resulted in higher yield and purity, the DNA was more degraded. The DNA from both extraction procedures were of suitable quality for further downstream molecular manipulation.
- High quality genomic DNA was extracted from a compost sample using chemical and mechanical DNA extraction methods. The hot detergent treatment employed during the Zhou chemical extraction method resulted in excessive humic acid contamination which could not be adequately removed. This led to the investigation of a chemical DNA extraction protocol specifically developed for DNA extraction from compost. Although DNA extracted through this method had significantly less humic acid contamination it also resulted in a marked decrease in DNA yield. DNA was finally extracted by a mechanical bead-beating method which provided DNA of suitable quality for further downstream molecular manipulation.

- It was concluded that the mechanical DNA extraction protocol was a more efficient method for obtaining metagenomic DNA from these organic rich environmental samples.
- The bacterial and fungal biodiversity within both samples were demonstrated by DGGE analysis. As expected from differences in the state of decomposition of the samples, the grape waste showed significantly more eukaryotic diversity. It was therefore concluded that the compost sample was a better choice for DNA library construction as the amount of cloned eukaryotic 'junk DNA' would be less. However, failure to extract high molecular weight DNA from the compost sample resulted in the grape waste DNA being the only source of material for the construction of a large insert fosmid library.

Chapter 4

- The TOPO-TA cloning procedure proved to be a quick and efficient method for cloning highly sheared metagenomic DNA for small insert library production. A library derived from compost metagenomic DNA was constructed consisting of approximately 2.6 Gbp cloned DNA.
- A large insert grape waste metagenomic DNA library was constructed using the CopyControl™ fosmid library production kit. This library consisted of approximately 1.1 Gbp cloned DNA.
- No bacterial laccases could be identified by expression screening of these libraries. It was concluded that the failure to recover genes was due to the limited metagenome coverage obtained in both libraries as the target environmental samples proved to be highly complex.

- An existing mathematical model which describes the impact of clone insert size on the required library size that is needed for successful gene discovery was developed take account of the DNA contribution made by sources other than bacteria to the metagenome. The impact of the bacterial DNA contribution was reflected by the increase in number of clones needed for gene discovery as total bacterial DNA decreased.

Chapter 5

- A new *mco* gene specific PCR detection assay resulted in the recovery of nine novel putative bacterial multicopper oxidase gene fragments from the target metagenomes.
- Three of these fragments were significantly smaller than other published bacterial *mco* sequences and may represent a new and as yet undescribed group of *mco* enzymes.

- Thermal asymmetric interlaced PCR was employed to recover partial flanking regions of putative bacterial *mco* p*MCO*7. Although the full-length open reading frame was not recovered, the viability of applying the TAIL-PCR technique to a metagenome has been demonstrated.
- The Vectorette™ system was employed to recover the 3' flanking region of p*MCO*8. This demonstrated the feasibility of this technique for the recovery of full-length open reading frames from metagenomic DNA samples.
- A magnetic bead capture protocol was adapted for the gene specific recovery of bacterial genes. The efficacy of this technique (referred to as subtractive hybridization magnetic bead capture (SH-MBC)), was demonstrated as a *mco*

gene was recovered from a *Streptomyces* strain in a proof of concept experiment.

6.2 Future work

Despite their limited coverage, both environmental libraries prepared in this study provide a resource for the potential discovery of other novel genes. This could be effected either through activity based screening or through direct sequencing. The quality of both libraries is somewhat questionable as the eukaryotic DNA content is unknown. An expensive sequencing approach may therefore not be advisable.

A theoretical study using the newly derived mathematical model and estimates for the complexity of eukaryotes, archaea and viruses in metagenomic DNA samples may shed more light on the observed low hit rate in gene discovery.

The feasibility of the TAIL-PCR and Vectorette™ approaches on a metagenomic DNA scale provides a means to recover specific target genes from these highly complex DNA samples. With the design of nested gene specific primers further downstream and therefore closer to the end of the gene, the full ORF for p*MCO*7 can be recovered. The 5' flanking region of p*MCO*8 could possible be recovered by the construction of more vectorette™ libraries using different restriction enzymes for library construction.

The proof of concept subtractive hybrization magnetic bead capture protocol (SH-MBC) could be adapted to a metagenomic scale. Immobilization of a mixed pool of internal gene fragments on the beads should result in the simultaneous recovery of several different full-length or partial target genes in one step. The concentrations of

driver and tester DNA required for successful recovery would need to be optimized in order to account for the high complexity of metagenomic DNA.

6.3 Conclusion

In this work the recovery of novel bacterial multicopper oxidases was investigated through activity based as well as sequence specific methods. The shotgun approach of environmental library construction and expression screening proved to be less reliable, particularly due the low coverage and the absence of high through-put screening technology. It has been suggested that sequence specific gene discovery methods will limit the access to sequence space, but in this study it was convincingly demonstrated that this direct approach is more reliable and useful for the detection and recovery of novel genes. One area of concern, however, is the technical difficulty associated with recovering full-length genes from metagenomic DNA samples using PCR based methods. It is undeniable that these methods are time consuming and require significant optimization. The development of a new subtractive hybridization technique, as described here, could potentially solve many of these problems and pave the way for more efficient metagenomic gene discovery.

National Contributions

Meyer, Q., Harrison, S.T.L., Burton, S.G., Soloman, M. and D.A. Cowan. The recovery of novel genes encoding oxidative enzymes from microbial community DNA. Cape Biotech. Somerset West. November 2002.

Meyer, Q., Smith, J.J., Harrison, S.T.L., Burton, S.G., Soloman, M., and Cowan, D.A. 2003. Recovering novel genes from environmental DNA samples via a gene specific PCR method. 18th congress of the South African Society of Biochemistry and Molecular Biology SASBMB, Pretoria, July 2003.

Publications

Cowan, D.A., **Meyer, Q.**, Stafford, W., Muyanga, S., Cameron, R, Wittwer, P. (2005) Metagenomic Gene Discovery: Past, Present and Future. Trends In Biotechnology 23(6): 321-329.

Cowan, D.A., Arslanoglu, A., Burton, S.G., Baker, G.C., Smith, J.J., **Meyer, Q.C**., and Cameron, R. 2004. Metagenomics, gene discovery and the ideal biocatalyst. Biochemical Society Transactions 32(2): 298-301.

Meyer, Q.C., Cowan, D.A. and S.G. Burton. 2006. Detection of bacterial laccase genes in the metagenome and the recovery of full length ORFs by subtractive hybridization magnetic bead capture. Submitted to BMC Microbiology.

Alexandre,G. and Zhulin,I.B. (2000). Laccases are widespread in bacteria. Trends In Biotechnology *18*, 41-42.

Altschul,S.F., Gish,W., Miller,W., Meyers,E.W., and Lipman,D.J. (1990). Basic Local Alignment Search Tool. Journal of Molecular Biology *215*, 403-410.

Amman,R.L., Ludwig,W., and Schleifer,K. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiology Reviews *59*, 143-169.

Bart,A., Dankert,J., and van der Ende,A. (2000). Representational difference analysis of *Neisseria meningitides* identifies sequences that are specific for the hyper-virulent lineage III clone. FEMS Microbiology Letters *188*, 111-114.

Bennet,P.M. (1999). Integrons and gene cassettes: a genetic construction kit for bacteria. Journal of Antimicrobial Chemotherapy *43*, 1-4.

Bereswill,S., Lichte,F., Vey,T., Fassbinder,F., and Kist,M. (1998). Cloning and characterization of the fur gene from Helicobacter pylori. FEMS Microbiology Letters *159*, 193-200.

Berthelet,M., Whyte,L.G., and Greer,C.W. (1996). Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. FEMS Microbiology Letters *138*, 17-22.

Bertrand,H., Poly,F., Van,V.T., Lombard,N., Nalin,R., Vogel,T.M., and Simonet,P. (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. Journal of Microbiological Methods *62*, 1-11.

Birnboim,H.C. and Doly,J.A. (1979). Rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research *7*, 1513-1523.

Botero,L.M., D'Imperio,S., Burr,M., McDermott,T.R., Young,M., and Hassett,D.J. (2005). Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. Applied And Environmental Microbiology *71*, 1267-1275.

Bowler,L.D., Hubank,M., and Spratt,B.G. (1999). Representational difference analysis of cDNA for the detection of differential gene expression in bacteria: development using a model of iron-regulated gene expression in *Neisseria meningitidi*s. Microbiology *145*, 3529-3537.

Brady,S.F., Chao,C.J., Handelsman,J., and Clardy,J. (2001). Cloning and heterologous expression of a natural product biosynthetic gene cluster from cDNA. Organic Letters *3*, 1981-1984.

Braker,G., Fesefeldt,A., and Witzel,K. (1998). Development of PCR primer systems for amplification of nitrite reductase genes *(nirk* and *nir*s) to detect denitrifying bacteria in environmental samples. Applied And Environmental Microbiology *64*, 3769-3775.

Brouwers,G.J., de Vrind,J.P., Corstjens,P.L., Cornelis,P., Baysse,C., and de Vrind-de Jong,E. (1999). cumA, a Gene Encoding a Multicopper Oxidase, Is Involved in Mn2+ Oxidation in Pseudomonas putida GB-1. Applied And Environmental Microbiology *65*, 1762-1768.

Burton,S.G., Boshoff,A., Edwards,W., and Rose,P.D. (1998). Biotransformation of phenols using immobilised polyphenol oxidase. Journal of Molecular Catalysis B: Enzymatic *5*, 411-416.

Call,H.P. and Mucke,I. (1997). History, overview and applications of mediated lignolytic systems, especially laccase-mediator-systems (Lignozym(R)-process). Journal of Biotechnology *53*, 163-202.

Chandler,D.P., Schreckhise,R.W., Smith,J.L., and Bolton,J. (1997). Electroelution to remove humic compounds from soil DNA and RNA extracts. Journal of Microbiological Methods *28*, 11-19.

Claus,H. (2003). Laccases and their occurrence in prokaryotes. Archives Of Microbiology *179*, 145-150.

Cohen,M.S., Bowers,W.C., Aronson,H., and Gray,E.T. (1987). Cell- free solubilization of coal by Polyporus versicolor. Applied And Environmental Microbiology *53*, 2840-2843.

Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., and Barry et,a. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature *393*, 537-544.

Collis,C.M. and Hall,R.M. (1993). Site specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. Journal Of Bacteriology *174*, 1574-1585.

Collis,C.M., Kim,M.J., Stokes,H.W., and Hall,R.M. (1998). Binding of the purified integron DNA integrase IntI1 to integron- and cassette-associated recombination sites. Molecular Microbiology *29*, 477-490.

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

Courtois,S., Cappellano,C.M., Ball,M., Francou,F.X., Normand,P., Helynck,G., Martinez,A., Kolvek,S.J., Hopke,J., Osburne,M.S., August,P.R., Nalin,R., Guerineau,M., Jeannin,P., Simonet,P., and Pernodet,J.L. (2003). Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Applied And Environmental Microbiology *69*, 49-55.

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

Curtis,T.P., Sloan,W.T., and Scannell,J.W. (2002). Estimation prokaryotic diversity and its limits. Proceedings Of The National Academy Of Sciences Of The United States Of America *99*, 10494-10499.

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

de Souza,F.A., Kowalchuk,G.A., Leeflang,P., van Veen,J.A., and Smit,E. (2004). PCR-denaturing gradient gel electrophoresis profiling of inter- and intraspecies 18S rRNA gene sequence heterogeneity is an accurate and sensitive method to assess species diversity of arbuscular mycorrhizal fungi of the genus Gigaspora. Applied And Environmental Microbiology *70*, 1413-1424.

Dijkmans,R., Jagers,A., Kreps,S., Collard,J.M., and Mergeay,M. (1993). Rapid method for purification of soil DNA for hybridization and PCR analysis. Microbial Releases: Viruses, Bacteria, Fungi *2*, 29-34.

Dunbar,J., Takala,S., Barns,S.M., Davis,J.A., and Kuske,C.R. (1999). Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Applied And Environmental Microbiology *65*, 1662-1669.

Duran,N., Rosa,M.A., Annibale,A., and Gianfreda,L. (2002). Applications of laccases and tyrosinases (phenoloxidases) immobilized on different supports: a review. Enzyme and Microbial Technology *31*, 907-931.

Eschenfeldt,W.H., Stols,L., Rosenbaum,H., Khambatta,Z.S., Quaite-Randall,E., and Donnely,M.I. (2001). DNA from Uncultured Organisms as a Source of 2,5-Diketo-D-Gluconic
Acid Reductases. Applied And Environmental Microbiology *67*, 4206-4214.

Fabbrini,M., Galli,C., and Gentili,P. (2002). Radical or electron-transfer mechanism of oxidation with some laccase/mediator systems. Journal of Molecular Catalysis B: Enzymatic *18*, 169-171.

Field,K.G., Gordon,D., Wright,T., Rappe,M., Urback,E., Vergin,K., and Giovannoni,S.J. (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. Applied And Environmental Microbiology *63*, 63-70.

Francis,C.A. and Tebo,B.M. (2001). cumA multicopper oxidase genes from diverse Mn(II)-oxidizing and non-Mn(II)-oxidizing Pseudomonas strains. Applied And Environmental Microbiology *67*, 4272-4278.

Frostegard,A., Courtois,S., Ramisse,V., Clerc,S., Bernillon,D., Le Gall,F., Jeannin,P., Nesme,X., and Simonet,P. (1999). Quantification of bias related to the extraction of DNA directly from soils. Applied And Environmental Microbiology *65*, 5409-5420.

Futamata,H., Harayama,S., and Watanabe,K. (2001). Group-specific monitoring of phenol hydroxylase genes for a functional assessment of phenol-stimulated trichloroethylene bioremediation. Appl. Environ. Microbiol. *67*, 4671-4677.

Gabor,E.M., Alkema W.B.L., and Janssen D.B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. Environmental Microbiology *6*, 879-886.

Galbraith,E.A., Antonopoulos,D.A., and White,B.A. (2004). Suppressive subtractive hybridisation as a tool for identifying genetic diversity in an environmental metagenome: the rumen as a model. Environmental Microbiology *6*, 928-937.

Galhaup,C., Goller,S., Peterbauer,C.K., Strauss,J., and Haltrich,D. (2002). Characterization of the major laccase isoenzyme from Trametes pubescens and regulation of its synthesis by metal ions. Microbiology (Reading, England) *148*, 2159-2169.

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

Gianfreda,L. and Rao,M.A. (2004). Potential of extra cellular enzymes in remediation of polluted soils: a review. Enzyme and Microbial Technology *35*, 339-354.

Gillespie,D.E., Brady,S.F., Bettermann,A.D., Cianciotto,N.P., Liles,M.R., Rondon,M.R., Clardy,J., Goodman,R.M., and Handelsman,J. (2002). Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Applied And Environmental Microbiology *68*, 4301-4306.

Givaudan,A., Effosse,A., Faure,D., Potier,P., Bouillant,M.L., and Bally,R. (1993). Polyphenol oxidase in Azospirillum lipoferum isolated from rice rhizosphere: Evidence for laccase activity in non-motile strains of Azospirillum lipoferum. FEMS Microbiology Letters *108*, 205-210.

Grant,S., Grant,W., Cowan,D.A., Jones,B., Yanhe,M., Ventosa,A., and Heapy,S. (2005). Construction of metagenomic
cDNA libraries from environmental samples. Applied And Environmental Microbiology.

Grass,G. and Rensing,C. (2001). CueO Is a Multi-copper Oxidase That Confers Copper Tolerance in Escherichia coli. Biochemical And Biophysical Research Communications *286*, 902-908.

Gray,J.P. and Herwig,R.P. (1996). Phylogenetic analysis of the bacterial communities in marine sediments. Applied And Environmental Microbiology *62*, 4049-4059.

Green,C.D., Simons,J.F., Taillon,B.E., and Lewin,D.A. (2001). Open systems: panoramic views of gene expression. Journal of Immunological Methods *250*, 67-79.

Guresir,M., Aktas,N., and Tanyolac,A. (2005). Influence of reaction conditions on the rate of enzymic polymerization of pyrogallol using laccase. Process Biochemistry *40*, 1175-1182.

Hall,R.M., Brookes,D.E., and Stokes,H.W. (1991). Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. Molecular Microbiology *5*, 1941-1959.

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*, 95-98.

Hallin,S. and Lindgren,P. (1999). PCR detection of genes encoding nitrite reductase in denitrifying bacteria. Applied And Environmental Microbiology *65*, 1652-1657.

Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Clayton,R.A., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D., Nierman,W.C., and White,O. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature *406*, 477-483.

Henckel,T., Jackel,U., Schnell,S., and Conrad,R. (2000). Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. Applied And Environmental Microbiology *66*, 1801-1808.

Henikoff,S., Henikoff,J.G., Alford,W.J., and Pietrokovski,S. (1995). Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene *163*, GC17-GC26.

Henne,A., Schmitz,R.A., Bomeke,M., Gottschalk,G., and Daniel,R. (1999). Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Applied And Environmental Microbiology *65*, 3901-3907.

Henne,A., Schmitz,R.A., Bomeke,M., Gottschalk,G., and Daniel,R. (2000). Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli.* Applied And Environmental Microbiology *66*, 3113-3116.

Hoegger,P.J., Navarro-Gonzalez,M., Kilaru,S., Hoffmann,M., Westbrook,E.D., and Kues,U. (2004). The laccase gene family in Coprinopsis cinerea (Coprinus cinereus). Current Genetics *45*, 9-18.

Hofrichter,M., Bublitz,F., and Fritsche,W. (1997). Fungal attack on coal: I. Modification of hard coal by fungi. Fuel Processing Technology *52*, 43-53.

Holben,W.E., Jansson,J.K., Chelm,B.K., and Tiedje,J.M. (1988). DNA probe method for the detection of specific microorganisms in the soil bacterial community. Applied And Environmental Microbiology *54*, 703-711.

Hugenholtz,P., Goebel,B.M., and Pace,N.R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. Journal Of Bacteriology *180*, 4765-4774.

Hurt,R.A. (2001). Simultaneous recovery of RNA and DNA from soils and sediments. Applied And Environmental Microbiology *67*, 4495-4503.

Jackson,C.R., Harper,J.P., Willoughby,D., Roden,E.E., and Churchill,P.F. (1997). A Simple, Efficient Method for the Separation of Humic Substances and DNA from Environmental Samples. Applied And Environmental Microbiology *63*, 4993-4995.

Jacobsen,C.S. (1995). Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. Applied And Environmental Microbiology *61*, 3347-3352.

Jacobsen,C.S. and Rasmussen,O.F. (1992). Development and Application of a New Method To Extract Bacterial DNA from Soil Based on Separation of Bacteria from Soil with Cation-Exchange Resin. Applied And Environmental Microbiology *58*, 2458-2462.

Kaneko,S., Akioka,M., Tsuge,K., and Itaya,M. (2005). DNA shuttling between plasmid vectors and a genome vector: systematic conversion and preservation of DNA libraries using the Bacillus subtilis genome (BGM) vector. Journal of Molecular Biology *349*, 1036-1044.

Kanoh,H., Nakashima,S., Zhao,Y., Sugiyama,Y., Kitajima,Y., and Nozawa,Y. (1998). Molecular cloning of a gene encoding phospholipase D from the pathogenic and dimorphic fungus, Candida albicans. Biochimica Et Biophysica Acta *1398*, 359-364.

Kawai,S., Nakagawa,M., and Ohashi,H. (1999). Aromatic ring cleavage of a non-phenolic beta-O-4 lignin model dimer by laccase of Trametes versicolor in the presence of 1-hydroxybenzotriazole. FEBS Letters *446*, 355-358.

Kawai,S., Umezawa,T., Shimada,M., and Higutchi,T. (1988). Aromatic ring cleavage of 4,6-di(tert-butyl)guaiacol, a phenolic lignin model compound, by lactase of *Coriolus versicolor*. FEBS Letters *235*, 309-311.

Knietsch,A., Waschkowitz,T., Bowien,S., Henne,A., and Daniel,R. (2003a). Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. Journal Of Molecular Microbiology And Biotechnology *5*, 46-56.

Knietsch,A., Waschkowitz,T., Bowien,S., Henne,A., and Daniel,R. (2003b). Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on Escherichia coli. Applied And Environmental Microbiology *69*, 1408-1416.

Kodama,N., Takenaka,S., Murakami,S., Shinke,R., and Aoki,K. (1997). Production of methyl, dimethyl, ethyl, chloro-, fluoro-, and hydroxyl derivatives of catechol from thirteen aromatic amines by the transpositional mutant B-9 of the aniline-assimilating Pseudomonas species AW-2. Journal of Fermentation and Bioengineering *84*, 232-235.

Kumar,S.V.S., Phale,P.S., Durani,S., and Wangikar,P.P. (2003). Combined sequence and structure analysis of the fungal laccase family. Biotechnology and Bioengineering *83*, 386-394.

Kuske,C.R., Banton,K.L., Adorada,D.L., Stark,P.C., Hill,K.K., and Jackson,P.J. (1998). Small-Scale DNA Sample Preparation Method for Field PCR Detection of Microbial Cells and Spores in Soil. Applied And Environmental Microbiology *64*, 2463-2472.

LaMontagne,M.G., Michel,F.C., Jr., Holden,P.A., and Reddy,C.A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. Journal of Microbiological Methods *49*, 255-264.

Leadbetter,J.R. (2003). Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory. Current Opinion in Microbiology *6*, 274-281.

Liebert,C.A., Hall,R.M., and Summers,A.O. (1999). Transposon Tn21, flagship of the floating genome. Microbiology And Molecular Biology Reviews: MMBR *63*, 507-522.

Liu,W.T., Marsh,T.L., Cheng,H., and Forney,L.J. (1997a). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Applied And Environmental Microbiology *63*, 4516-4522.

Liu,Y., Mitsukawa,N., Oosumi,T., and Whittier,R.F. (1997b). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. The Plant Journal *8*, 457-463.

Liu,Y.G. and Whittier,R.F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics *25*, 674-681.

Majernik,A., Gottschalk,G., and Daniel,R. (2001). Screening of environmental DNA libraries for the presence of genes conferring Na+(Li+)/H+ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. Journal Of Bacteriology *183*, 6645-6653.

Manefield,M., Whiteley,A.S., Griffiths,R.I., and Bailey,M.J. (2002). RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. Applied And Environmental Microbiology *68*, 5367-5373.

Mann,K.G., Jenny,R.J., and Krishnaswamy,S. (1988). Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. Annual Review Of Biochemistry *57*, 915-956.

Martinez,A., Kolvek,S.J., Yip,C.L.T., Hopke,J., Brown,K.A., MacNeil,I.A., and Osburne,M.S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting

heterologous natural products in multiple expression hosts. Applied And Environmental Microbiology *70*, 2452-2463.

Mayer,A.M. (1986). Polyphenol oxidases in plants-recent progress. Phytochemistry *26*, 11-20.

Mazel,D., Dychinco,B., Webb,V.A., and Davies,J. (1998). A distinctive class of integron in the *Vibrio cholerae* genome. Science *280*, 605-608.

McDonald,I.R., Kenna,E.M., and Murrell,J.C. (1995). Detection of methanotrophic bacteria in environmental samples with the PCR. Applied And Environmental Microbiology *61*, 116-121.

Megonigal,M.D., Rappaport,E.F., Wilson,R.B., Jones,D.H., Whitlock,J.A., Ortega,J.A., Slater,D.J., Nowell,P.C., and Felix,C.A. (2000). Panhandle PCR for cDNA: a rapid method for isolation of MLL fusion transcripts involving unknown partner genes. Proceedings of the National Academy of Sciences *97*, 9597-9602.

Mesarch,M.B., Nakatsu,C.H., and Nies,L. (2000). Development of catechol 2,3-dioxygenase-specific primers for monitoring bioremediation by competitive quantitative PCR. Applied And Environmental Microbiology *66*, 678-683.

Messerschmidt,A. and Huber,R. (1990). The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin. Modelling and structural relationships. European Journal Of Biochemistry / FEBS *187*, 341-352.

Mester,T. and Tien,M. (2000). Oxidation mechanism of ligninolytic enzymes involved in the degradation of environmental pollutants. International Biodeterioration & Biodegradation *46*, 51-59.

Miller,D.N., Bryant,J.E., Madsen,E.L., and Chiorse,W.C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples.
6. Applied And Environmental Microbiology *61*, 2122-2126.

Mishra,R.N., Singla-Pareek,S.L., Nair,S., Sopory,S.K., and Reddy,M.K. (2002). Directional genome walking using PCR. Biotechniques *33*, 830-834.

Molitoris,H.P., Van Breemen,J.F.L., Van Bruggen,E.F.J., and Esser,K. (1972). The phenoloxidases of the ascomycete Podospora anserina X. Electron microscopic studies on the structure of laccases I, II and III. Biochimica et Biophysica Acta (BBA) - Protein Structure *271*, 286-291.

More,M.I., Herrick,J.B., Silva,M.C., Ghiorse,W.C., and Madsen,E.L. (1994). Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Applied And Environmental Microbiology *60*, 1572-1580.

Myrick,K.V. and Gelbart,W.M. (2002). Universal fast walking for direct and versatile determination of flanking sequence. Gene *284*, 125-131.

Ochman,H., Ayala,F.J., and Hartl,D.L. (1993). Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. Methods In Enzymology *218*, 309-321.

Ogram,A., Sayler,G.S., and Barkay,T. (1987). The extraction and purification of microbial DNA from sediments. Journal of Microbiological Methods *7*, 57-66.

Okuta,A., Ohnishi,K., and Harayama,S. (1998). PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. Gene *212*, 221-228.

Otterbein,L., Record,E., Longhi,S., Asther,M., and Moukha,S. (2000). Molecular cloning of the cDNA encoding laccase from Pycnoporus cinnabarinus I-937 and expression in Pichia pastoris. European Journal Of Biochemistry / FEBS *267*, 1619-1625.

Picard,C., Ponsonnet,C., Paget,E., Nesme,X., and Simonet,P. (1992). Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Applied And Environmental Microbiology *58*, 2717-2722.

Pillai,S.D., Josephenson,K.L., Bailey,R.L., Gerba,C.P., and Pepper,I.L. (1991). Rapid method for processing soil samples for polymerase chain reaction amplification of specific gene sequences. Applied And Environmental Microbiology *57*, 2283-2286.

Piontek,K., Antorini,M., and Choinowski,T. (2002). Crystal Structure of a Laccase from the Fungus Trametes versicolor at 1.90-A Resolution Containing a Full Complement of Coppers. J. Biol. Chem. *277*, 37663-37669.

Prieme,A., Bonilla Sitaula,J.I., Klemedtsson,A.K., and Bakken,L.R. (1996). Extraction of methane-oxidizing bacteria from soil particles. FEMS Microbiology Ecology *21*, 59-68.

Radajewski,S., Webster,G., Reay,D.S., Morris,S.A., Ineson,P., Nedwell,D.B., Prosser,J.I., and Murrell,J.C. (2002). Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. Microbiology *148*, 2331-2342.

Ralph,J.P. and Catcheside,D.E.A. (1997). Transformations of low rank coal by Phanerochaete chrysosporium and other wood-rot fungi. Fuel Processing Technology *52*, 79-93.

Ranjan,R., Grover,A., Kapardar,R.K., and Sharma,R. (2005). Isolation of novel lipolytic genes from uncultured bacteria of pond water. Biochemical And Biophysical Research Communications *335*, 57-65.

Rebel,F., Renard,J.P., and Duranthon,V. (1995). PCR-generated cDNA libraries from reduced numbers of mouse oocytes. Zygote *3*, 241-250.

Recchia,G.D. and Hall,R.M. (1997). Origins of the mobile gene cassettes found in integrons. Trends In Microbiology *5*, 389-394.

Rhee,J.K., Ahn,D.G., Kim,Y.G., and Oh,J.W. (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*, 817-825.

Robe,P., Nalin,R., Capellano,C., Vogel,T.M., and Simonet,P. (2003). Extraction of DNA from soil. European Journal of Soil Biology *39*, 183-190.

Rochelle,P.A., Fry,J.C., Parkes,R.J., and Weightman,A.J. (1992). DNA extraction for 16S rRNA gene analysis to determine genetic diversity in deep sediment communities. FEMS Microbiology Letters *79*, 59-65.

Rogalski,J., Wojtas-Wasilewska,M., Apalovic,R., and Leonowicz,A. (1991). Affinity chromatography as a rapid and convenient method for purification of fungal laccases. Biotechnology and Bioengineering *37*, 770-777.

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

Rondon,M.R., Goodman,R.M., and Handelsman,J. (1999). The Earth's bounty: assessing and accessing soil microbial diversity. Trends In Biotechnology *17*, 403-409.

Roose-Amsaleg,C.L., Garnier-Sillam,E., and Harry,M. (2001). Extraction and purification of microbial DNA from soil and sediment samples.
8. Applications Soil Ecology *18*, 47-60.

Rose,T.M., Schultz,E.R., Henikoff,J.G., Pietrokovski,S., McCallum,C.M., Henikoff, and S. (1998). Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. Nucleic Acids Research *26*, 1628-1635.

Rowe-Magnus,D.A., Guérout,A., and Mazel,D. (1999). Super-integrons. Research in Microbiology *150*, 641-651.

Rowe-Magnus,D.A., Guerout,A.M., Ploncard,P., Dychinco,B., Davies,J., and Mazel,D. (2001). The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. Proceedings Of The National Academy Of Sciences Of The United States Of America *98*, 652-657.

Rowe-Magnus,D.A. and Mazel,D. (1999). Resistance gene capture. Current Opinion in Microbiology *2*, 483-488.

Rowe-Magnus,D.A. and Mazel,D. (2001). Integrons: natural tools for bacterial genome evolution. Current Opinion in Microbiology *4*, 565-569.

Ruijssenaars,H.J. and Hartmans,S. (2004). A cloned Bacillus halodurans multicopper oxidase exhibiting alkaline laccase activity. Applied Microbiology And Biotechnology *65*, 177-182.

Ryden,L.G. and Hunt,L.T. (1993). Evolution of protein complexity: the blue copper-containing oxidases and related proteins. Journal Of Molecular Evolution *36*, 41-66.

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

Sanchez-Amat,A., Lucas-Elio,P., Fernandez,E., Garcia-Borron,J.C., and Solano,F. (2001). Molecular cloning and functional characterization of a unique multipotent polyphenol oxidase from Marinomonas mediterranea. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology *1547*, 104-116.

Saparrat,M.C.N., Guillen,F., Arambarri,A.M., Martinez,A.T., and Martinez,M.J. (2002). Induction, isolation, and characterization of two laccases from the white rot basidiomycete Coriolopsis rigida. Applied And Environmental Microbiology *68*, 1534-1540.

Schloss,P.D. and Handelsman,J. (2003). Biotechnological prospects from metagenomics. Current Opinion In Biotechnology *14*, 303-310.

Selenska,S. and Klingmuller,W. (1991). Direct detection of nif-gene sequences of Enterobacter agglomerans in soil. FEMS Microbiology Letters *80*, 243-245.

Sheu,D., Wang,Y., and Lee,C. (2000). Rapid detection of polyhydroxyalkanoate-accumulating bacteria isolated from the environment by colony PCR. Microbiology *146*, 2019-2025.

Shin,T., Murao,S., and Matsumura,E. (1987). A Chromogenic Oxidative Coupling Reaction of Laccase and Angiotensin I Converting Enzyme Assay. Analytical Biochemistry *166*, 380-388.

Shuman,S. (1994). Noval Approach to Molecular Cloning and Polynucleotide Synthesis Using Vaccinia DNA Topoisomerase. J. Biol. Chem. *269*, 32678-32684.

Shuman,S. (1991). Recombination Mediated by Vaccinia Virus DNA Topoisomerase I in Escherichia coli is Sequence Specific. Proceedings Of The National Academy Of Sciences Of The United States Of America *88*, 10104-10108.

Sinigalliano,C.D., Kuhn,D.N., and Jones,R.D. (2004). Amplification of the *amoa* gene from diverse species of ammonium-oxidizing bacteria and from an indigenous bacterial population from seawater. Applied And Environmental Microbiology *61*, 2702-2706.

Smalla,K., Krogerrecklenfort,E., Heuer,H., Dejonghe,W., Top,E., Osborn,M., Niewint,J., Tebbe,C., Barr,M., Bailey,M., Greated,A., Thomas,C., Turner,S., Young,P., Nikolakopoulou,D., Karagouni,A., Wolters,A., van Elsas,J.D., Dronen,K., Sandaa,R., Borin,S., Brabhu,J., Grohmann,E., and Sobecky,P. (2000). PCR-based detection of mobile genetic elements in total community DNA. Microbiology *146*, 1256-1257.

Solano,F., Lucas-Elio,P., Lopez-Serrano,D., Fernandez,E., and Sanchez-Amat,A. (2001). Dimethoxyphenol oxidase activity of different microbial blue multicopper proteins. FEMS Microbiology Letters *204*, 175-181.

Song,J.K., Kim,M.K., and Rhee,J.S. (1999). Cloning and expression of the gene encoding phospholipase A1 from Serratia sp. MK1 in Escherichia coli. Journal of Biotechnology *72*, 103-114.

Stach,J.E.M., Bathe,S., Clapp,J.P., and Burns,R.G. (2001). PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. FEMS Microbiology Ecology *36*, 139-151.

Starkey,M.P., Umrania,Y., Mundy,C.R., and Bishop,M.J. (1998). Reference cDNA library facilities available from European sources. Molecular Biotechnology *9*, 35-57.

Steffan,R.J., Goksoyr,J., Bej,A.K., and Atlas,R.M. (1998). Recovery of DNA from soils and sediments. Applied And Environmental Microbiology *54*, 2908-2915.

Stokes,H.W., Holmes,A.J., Nield,B.S., Holley,M.P., Nevalainen,K.M., Mabbutt,B.C., and Mabbutt,B.C. (2001). Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. Applied And Environmental Microbiology *67*, 5240-5246.

Sullivan,E.R., Leahy,J.G., and Colwell,R.R. (1999). Cloning and sequence analysis of the lipase and lipase chaperone-encoding genes from Acinetobacter calcoaceticus RAG-1, and redefinition of a proteobacterial lipase family and an analogous lipase chaperone family. Gene *230*, 277-286.

Takai,K. and Sako,Y. (1999). A molecular view of archaeal diversity in marine and terrestrial hot water environments. FEMS Microbiology Ecology *28*, 177-188.

Taniguchi,V.T., Malmstrom,B.G., Anson,F.C., and Gray,H.B. (1982). Temperature Dependence of the Reduction Potential of Blue Copper in Fungal Laccase. Proceedings Of The National Academy Of Sciences Of The United States Of America *79*, 3387-3389.

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

Thompson,J.D., Higgins,D.G., and Gibson,T.J. (1994). CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Research *22*, 4673-4680.

Thurston,C.F. (1994). The structure and function of fungal laccases. Microbiology *140*, 19-26.

Torsvik,V., Goksoyr,J., and Daae,F.L. (1990). High diversity in DNA of soil bacteria. Applied And Environmental Microbiology *56*, 782-787.

Tyson, J. W., Chapman, J., Hugenholtz, P., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37-43. 2004.
Ref Type: Generic

Venter,J.C., Remington,K., Heidelberg,J.F., Halpern,A.L., Rush,D., Eisen,J.A., Wu,D., Paulsen,I., Nelson,K.E., Nelson,W., Fouts,D.E., Sevy,S., Knap,H., Lomas,M.W., Nealson,K., White,O., Peterson,J.D., Hoffman,J., Parsons,R., Baden-Tillson,H., Pfannkoch,C., Rogers,Y., and Smith,H.O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science *304*, 66-74.

Vuddhakul,V., Nakai,T., Matsumoto,C., Oh,T., Nishino,T., Chen,C.H., Nishibuchi,M., and Okuda,J. (2000). Analysis of gyrB and toxR gene sequences of Vibrio hollisae and development of gyrB- and toxR-targeted PCR methods for isolation of V. hollisae from the environment and its identification. Applied And Environmental Microbiology *66*, 3506-3514.

Watanabe,K., Teramoto,M., Futamata,H., and Harayama,S. (1998). Molecular detection, isolation, and physiological characterization of functionally dominant phenol-degrading bacteria in activated sludge. Applied And Environmental Microbiology *64*, 4396-4402.

Watanabe,K., Kodama,Y., and Harayama,S. (2001). Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. Journal of Microbiological Methods *44*, 253-262.

Whitman,W.B., Coleman,D.C., and Wiebe,W.J. (1998). Prokaryotes: the unseen majority. Proceedings Of The National Academy Of Sciences Of The United States Of America *95*, 6578-6583.

Wild,J., Hradecna,Z., and Szybalski,W. (2002). Conditionally Amplifiable BACs: Switching From Single-Copy to High-Copy Vectors and Genomic Clones. Genome Res. *12*, 1434-1444.

Wilkinson,D.E., Jeanicke,T., and Cowan,D.A. (2002). Efficient molecular cloning of environmental DNA from geothermal sediments. Biotechnology Letters *24*, 155-161.

Wilson,M.S., Bakermans,C., and Madsen,E. (1999). In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. Applied And Environmental Microbiology *65*, 80-87.

Yin,B., Crowley,D., Sparovek,G., De Melo,W.J., and Borneman,J. (2000). Bacterial functional redundancy along a soil reclamation gradient. Applied And Environmental Microbiology *66*, 4361-4365.

Yun,J., Kang,S., Park,S., Yoon,H., Kim,M.J., Heu,S., and Ryu,S. (2004). Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. Applied And Environmental Microbiology *70*, 7229-7235.

Zhou,J., Bruns,M.A., and Tiedje,J.M. (1996). DNA recovery from soils of diverse composition. Applied And Environmental Microbiology *62*, 316-322.