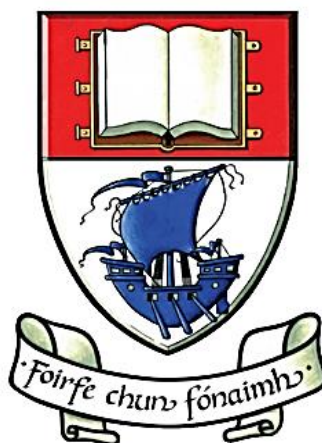# THE IDENTIFICATION OF GENES AND ENZYMES IN THE ALDOXIME-NITRILE METABOLISING PATHWAY

Submitted to Waterford Institute of Technology for the Degree of Doctor of Philosophy

By Tríona-Marie Dooley-Cullinane

Pharmaceutical and Molecular Biotechnology Research Center (PMBRC)

Waterford Institute of Technology,

Waterford,

Ireland

Under the Supervision of

Dr. Lee Coffey and Dr. Catherine O'Reilly

2019

DECLARATION

The work described in this thesis isoriginal and was solely carried out by the author, and the work of others has been duly referenced in the text. No part of this thesis has been previously submitted for a degree at this or any other institute.

Signed:_____

Date:_____

ACKNOWLEDGEMENTS

Gannon-Stephenson's, you have made me feel like part of the furniture over the last nine years and have provided more support that you will ever know, I will be forever thankful.

To Marcus, I don't know where to begin. You have been on this journey with me since my undergraduate studies. You have witnessed all the ups and downs and your support and love has never waivered throughout. You have been there every day and night through it all and knew exactly what I needed, even when I didn't. Your sense of humor and wit have made the longest days more bearable. I will forever be indebted to you for your years of guidance, support, patience and love.

Finally, to Alice, this dissertation is dedicated to your memory. I hope that one day I can become half the woman you were, and I thank you for your encouragement and support during my younger years.

ABSTRACT

Biocatalysis is gaining much attention for the production of pharmaceutical intermediates, drug compounds and, fine chemicals. Biocatalysis, as a route of synthesis, boasts many benefits over traditional chemical synthesis. It can account for enantiopure compounds, lower waste production, decreases the need for potential purification and down-stream processing while also allowing the reactions to be carried out at milder conditions, such as lower temperatures and a relatively neutral pH (Gong *et al.* 2012, Singh *et al.* 2006).

Due to the often low efficacy of racemic drugs or in fact the side effects caused by one of the enantiomers, enantiopure products are highly sought after (Albarrán-Velo *et al.* 2018). Thalidomide, a drug that was previously prescribed to treat morning sickness during pregnancy was sold as a safe and effective anti-emetic, in its racemic form. Unfortunately, years later the level of incidents of miscarriage and foetal birth defects observed amongst the women who were prescribed the drug led to the discovery that the S-enantiomer of the drug was teratogenic.

Enzymes comprising the Aldoxime-Nitrile pathway are the focus of the body of work presented in this thesis. In 2010, it was reported that there were over 30 nitrile-containing pharmaceuticals on the market with another 20 at clinical trial stage (Fleming *et al.* 2010). Many nitrile containing drugs or intermediates may have been produced via chemical synthesis coupled with biocatalysis. Focusing on the identification of aldoxime dehydratase enzymes can offer a new, improved route to nitrile synthesis. The research presented in chapter two and published by (Dooley-Cullinane *et al.* 2016) presents a new high-throughput real-time (qPCR) for the detection of novel aldoxime dehydratase genes. The potential application of these genes/enzymes was reinforced by the Food and Drug Administration (FDA) approval of 11 nitrile containing pharmaceuticals since 2010. This is an indication as to what is to come with regard to an increase in market size for bio-pharmaceuticals produced by biocatalysis, in particular nitrile containing pharmaceuticals.

The remaining enzymes which comprise the aldoxime-nitrile pathway, the nitrile-degrading enzymes such as nitrilases and nitrile hydratases (NHases), have been shown

to display a broad substrate scope and complex nitrile substrates can be converted into their corresponding amides with the final enzyme, and amidase converting the amide to the corresponding carboxylic acids. The area of nitrile metabolism has been of great interest and will be into the foreseeable future (Huisman and Collier 2013, Gong *et al.* 2012).

The body of work presented, details outputs and discoveries which form a part of the larger web of research undertaken by the molecular biocatalysis research group. The overall aim of research carried out by the group is the identification and discovery of novel genes and enzymes towards the production of industrially or biopharmaceutical relevant products. This thesis presents two novel molecular methods for the detection of genes within the aldoxime-nitrile pathway. A novel qPCR method was designed for the detection of aldoxime dehydratase genes. The second molecular method presented is a novel clade-specific touchdown conventional PCR which allows for the clade specific detection of novel nitrilase genes. Additionally, this protocol allows for the detection of nitrilase genes with predicted enzymatic activity in model substrates based on their clade classification.

A novel nitrilase previously identified from *Burkholderia* sp. LC8 was cloned, expressed and characterised with results displaying a broad substrate range and excellent enantioselectivity towards mandelonitrile.

The research also documents the discovery of a novel nitrilase partial sequence through the construction of a genomic fosmid library from genomic DNA of *Rhodococcus erythropolis* SET-1. This body of work involved the screening of fosmid clones for functional activity on 3-HGN. Further work presented a discussion on the attempts to elucidate the full nitrilase gene sequence via fosmid insert sequencing and also whole genome sequencing which led to the identification of a nitrile hydratase/amidase genomic cluster in *R. erythropolis* SET-1.

TABLE OF CONTENTS

ABBREVIATIONS

3-HBN	3-Hydroxybutyronitrile
3-HGN	3-Hydroxyglutaronitrile
3-HPPN	3-Hydroxy-3-phenylpropionitrile
AMD	Amidase
BAM	2,6-dichlorobenzamide
CNG'S	Cyanogenic glycosides
CyPy	3-cyanopyridine
EDTA	Ethylenediaminetetraacetic acid
IAA	Indole-3-acetic acid
IAN	Indole-3-acetonitrile
IS	Insertion Sequence
NHASE	Nitrile hydratase
OXD	Aldoxime dehydratase
PAA	Phenylacetate
PAAm	Phenylacetamide
PAOx	Phenylacetaldoxime
PYOx	Pyridine-3-aldoxime
PAN	Phenylacetonitrile
PFGE	Pulsed field gel electrophoresis
PyOx	E-pyridine3-aldoxime
PyOx	Pyridine-3-aldoxime
RCF	Relative centrifugal force
TAE	Tris base, acetic acid and EDTA
TBE	Tris base, borate and EDTA
3-HPPN	3-Hydroxy-3-phenylpropionitrile
FC 1	Functional Clone 1
FC 2	Functional clone 2

LIST OF FIGURES

LIST OF TABLES

PROPOSED RESEARCH

Many papers have been published regarding nitrile metabolising enzymes, their identification by novel real-time PCR methods, their gene clusters and corresponding substrate spectrums. The main scope of research carried out by the Molecular Biotechnology and Biocatalysis research group at WIT is presented in Figure A with a more detailed work-flow approach visible in Figure B. The overall work flows and research approach can be applied to any class of enzymes; however, the scope of this thesis concerns the genes involved in the microbial aldoxime-nitrile pathway.
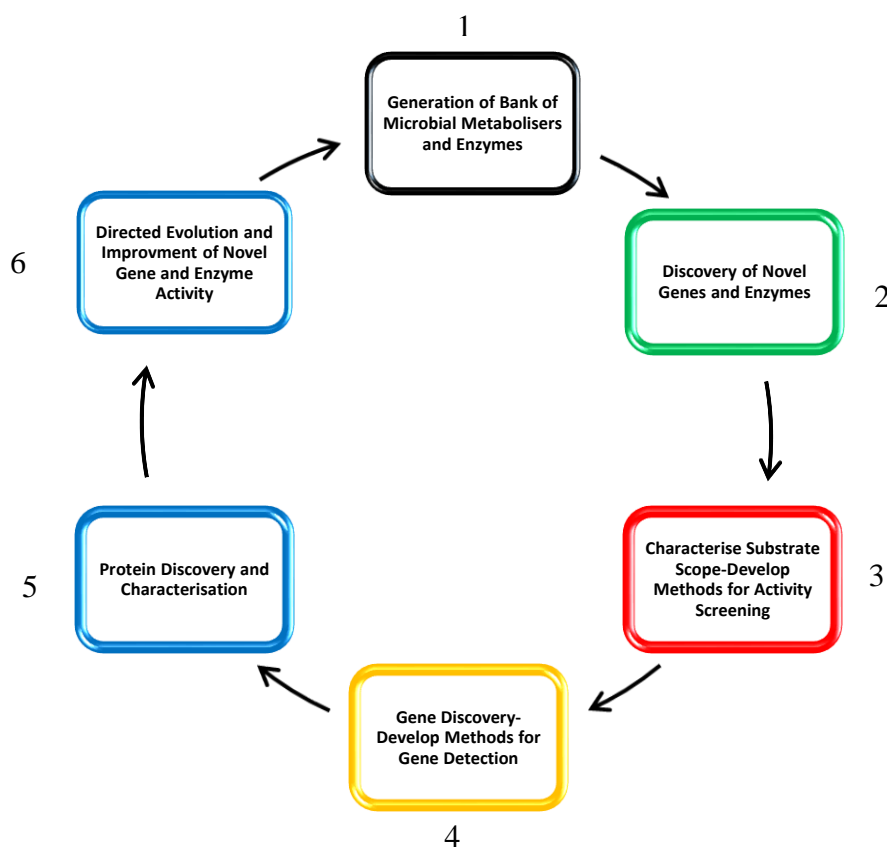
Figure A The scope of research carried out by the Molecular Biology and Biocatalysis Research group at WIT. Numerical values indicate the order of research tasks performed towards the discovery of novel biocatalysts.

This involves the generation/isolation of a biocatalysis microbial bank, characterisation of substrate spectrums, discovery of novel genes and subsequent purification and characterisation of these enzymes, with the end-goal encompassing the production of an optimised protein.

Much research has been carried out by the research group on nitrile gene clusters regarding the source and function of the genes they are comprised of. A number of novel molecular methods such as real-time PCR screening assays have been developed, to analyse the origin of genes within a specific genera or species (Coffey *et al.* 2010, Coffey *et al.* 2009). Recently (Coady *et al.* 2013) designed a 'high-throughput screening strategy for the assessment of nitrile-hydrolysing activity towards β-hydroxy acids'. More research must be dedicated to elucidating gene and protein function, to fully realise their potential applications in industry. Expanding the microbial metabolising bank will require the development of gene screening assays for new and additional gene/enzyme groups within the aldoxime-nitrile pathway.
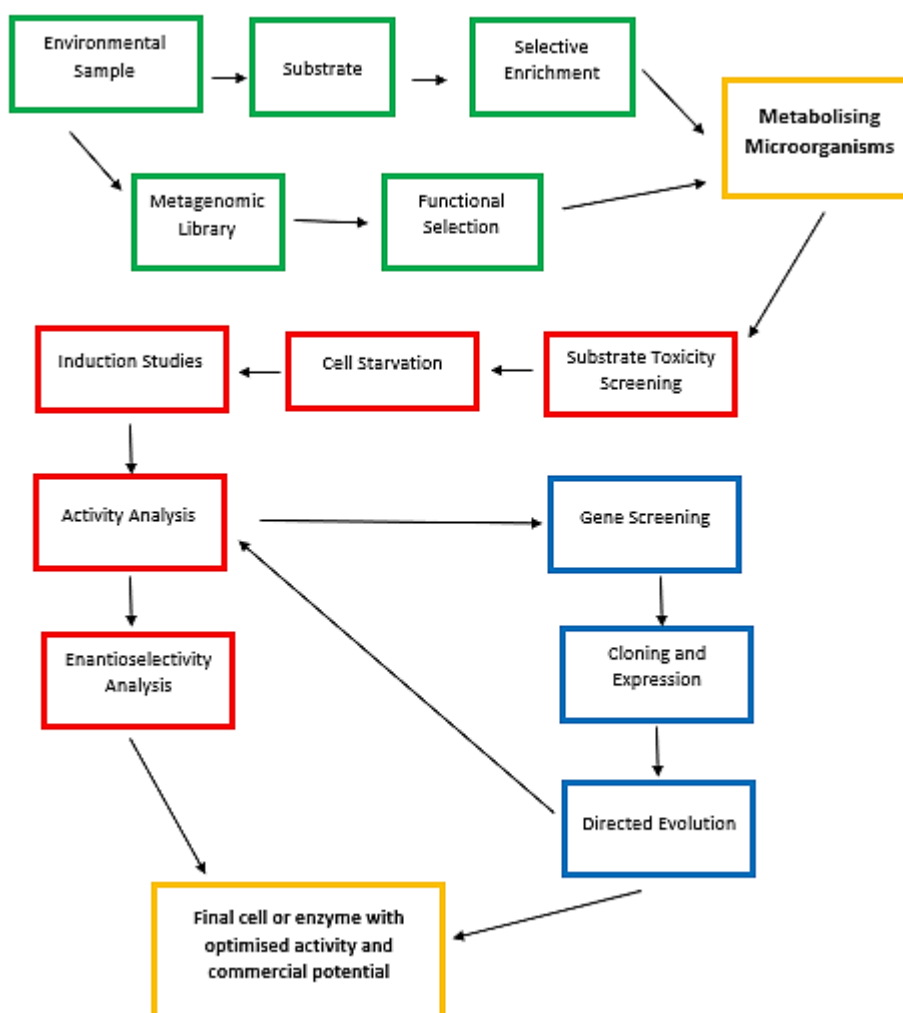
Figure B The work flow and steps involved in the discovery of microorganisms/enzymes.

PROJECT AIMS

The proposed research aimed to develop a high throughput real-time PCR (qPCR) assay for the detection of novel aldoxime dehydratase genes. Firstly, the WIT biocatalysis bank was expanded on three key nitrile substrates: 3-hydroxybutyronitrile, 3-hydroxyglutaronitrile and, 3-hydroxy-3-phenylpropionitrile. This expansion aimed to increase the pool of isolates which was thenscreened for genes of interest.

A previously identified nitrilase enzyme by the research group was re-cloned, expressed and characterised. This body of work aimed to assess the nitrilase enzyme against a broad range of substrates with a focus on the industrially relevant substrate mandelonitrile. The enzyme was then screened for enantioselectivity towards mandelonitrile.

A well-studied Rhodococcus erythropolis strain SET-1 (Coady *et al*. 2015, Coady *et al*. 2013) was subjected to gene screening to identify the gene responsible for the nitrile metabolising activity previously characterised. A fosmid gene library with R. erythropolis SET-1 genomic DNA was prepared, encompassing the genome in one library. Any identified functional clones from the fosmid genomic library were subjected to gene screening for nitrile metabolising genes of interest. In order to identify nitrile metabolising genes of interest, comparative sequence analysis between whole genome sequence of SET-1 and the full fosmid insert sequence of any identified functional clones was performed.

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 Nitriles and their occurrence

Nitrile compounds (organocyanides) have a general structure of (R-CN) Figure 1.1. The R can be either an alkyl, heterocyclic or aromatic group where their naming depends on the carboxylic acid. A triple bond joins the Carbon (C) and Nitrogen (N) and they have also been known as carbonitriles.

$$N \equiv C - R$$

Figure 1.1 General structure of nitrile compounds with the triple bonds observed between the carbon and nitrogen

The presence of nitriles in the environment is due to both biological and industrial practices. Plants and microorganisms hydrolyse naturally occurring nitriles as either cyanogenic glycosides or cyanolipids (Graham *et al.* 2000). During plant biosynthesis of cyano glycosides (CNGs) and cyanolipids, nitriles are produced as intermediates (Kobayashi and Shimizu 2000). Cyanoglycosides are amino-acid-derived secondary plant products and are found in more than 2000 species. Upon tissue injury, the cyanoglycosides are hydrolysed to either sugar, keto or aldehyde compound or HCN.

1.2 Toxicity of nitriles

Most nitriles are highly toxic, mutagenic, and carcinogenic in nature. Aliphatic nitriles have been thought to manifest their toxicity by the discharge of cyanide (Ahmed and Farooqui 1982). Aromatic or fully substituted carbons are not considered to be as toxic as aliphatic nitriles, as they do not result in the release of cyanide (Fleming *et al*. 2010) The toxicities of nitriles in humans can be presented by symptoms such as gastric problems, nausea, respiratory distress, convulsions, coma or skeletal deformities (Banerjee *et al.* 2002). Humans can be affected by nitrile toxicity by either dietary intake or physical contact (Grogan *et al.* 1992).

1.3 Nitrile degrading pathways

Cyanide and nitriles can be degraded via three enzymatic systems:

I. Cyanide oxidation- This cyanide degradation pathway has been identified in *Pseudomonas flourescens* (Rollinson *et al.* 1987)

II. Hydrogenation of nitriles- carried out by nitrogen-fixing bacteria and nitrogenase.

III. Hydrolysis- The most common and studied pathway of microbial nitrile metabolism.

This literature review and subsequent thesis will focus on the hydrolysis pathway as it is the most studied enzymatic nitrile degredation pathway in microbial cells.

1.4 Nitrile hydrolysis

A one-step metabolism of nitriles to their corresponding carboxylic acid involves the use of a nitrilase enzyme. This enzyme directly hydrolyses the nitrile into a carboxylic acid omitting the intermediate step of amide production which is seen with the use of nitrile hydratase enzymes (Figure 1.2). This one-step reaction involvesa nucleophilic attack by a reactive sulphydryl group of the nitrilase enzymeleading to the formation of an enzyme bound imine. The imine is then hydrated to form a tetrahedral intermediate. This leads to the formation of an acyl-enzyme from the intermediate with an ammonia leaving group. A carboxylic acid is then formed from the hydrolysis of a carboxylic acid (Bandyopadhyay *et al.* 1986). A two-way pathway for nitrile metabolism requires a nitrile hydratase (NHase) and amidase enzyme to convert nitriles to amides and from amides to carboxylic acids respectively.

Chemical   Enzymatic

$$RC \equiv N \qquad\qquad RC \equiv N \qquad\qquad RC \equiv N$$

H_2O H_2O $2\ H_2O$
H^+ or M^+ **Nitrile hydratase**   **Nitrilase**

$$RCONH_2 \qquad\qquad RCONH_2$$

or

H_2O H_2O
H^+ or M^+ **Amidase**

$\longrightarrow NH_3$ $\longrightarrow NH_3$ NH_3

$$RCOOH \qquad\qquad RCOOH \qquad\qquad RCOOH$$

A   B   C
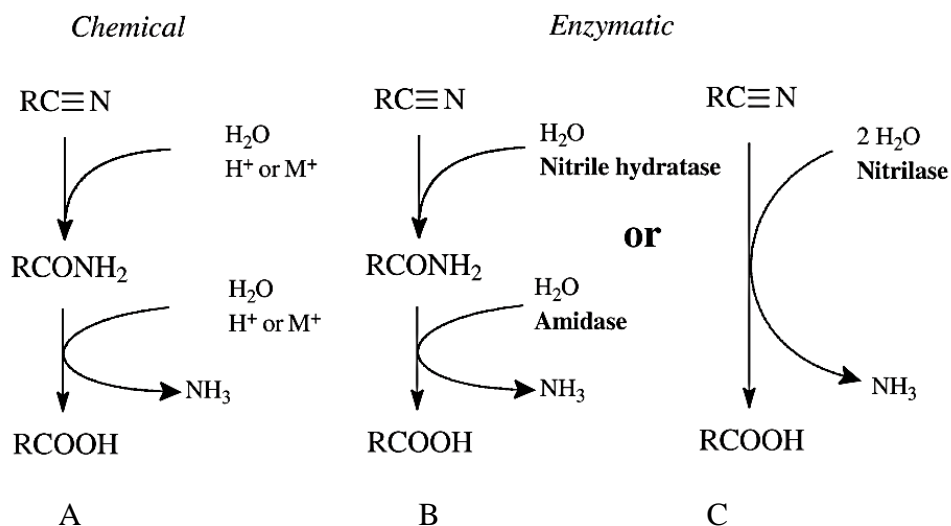
Figure 1.2 The chemical and enzymatic pathways of nitrile hydrolysis (Fournand and Arnaud 2001). (A) nitrile hydrolysis via chemical approach, (B) nitrile hydrolysis via a nitrile hydratase (EC 4.2.1.84) and amidase (EC 3.5.1.4) two-step enzymatic pathway and (C) nitrile hydrolysis via a one-step nitrilase (EC 3.5.5.1) enzymatic pathway.

1.5   Nitrile metabolising enzymes

1.5.1   Nitrilases (EC 3.5.5.1)

1.5.1.1   Sources of Nitrilases

Nitrilases have been sourced from plants, bacteria and fungi. The frequency distribution of nitrilases from these sources is not yet known.

Bacterial

From the early 1960's, bacteria as a source of nitrilases have been heavily studied. The first ricinine hydrolysing nitrilase was discovered from a soil sample and identified as a member of the *Pseudomonas* genus (Robinson and Hook 1964). From this several bacteria were assessed for their nitrilase hydrolysing activity. The generas of *Nocardia, Alcaligens, Actinobacter* and *Pseudomonas* are some of the bacteria which display nitrilase activity (Harper 1977, Mathew *et al.* 1988, Nagasawa *et al.* 1990). Each of the nitrilases has been purified, characterised, cloned, and expressed in a range of host systems. Many of these strains yielded nitrilases which were applied in an industrial process e.g. the nitrilase of *Rhodococcus rhodochrous* J1 for the production of nicotinic acid (Mathew *et al.* 1988).

Fungi

Nitrilase activity has been reported in several genera of filamentous fungi ranging from *Aspergillus* to *Giberella, Penicillium* and *Fusarium* (Gong *et al.* 2012). There is a limited amount of information available on nitrilases from filamentous fungi even after the discovery of a *Penicillium* nitrilase, which was shown to hydrolyze indole-3-acetonitrile (IAN) into indole-3-acetic acid (IAA) (Figure 1.3) (Thimann and Mahadevan 1964).
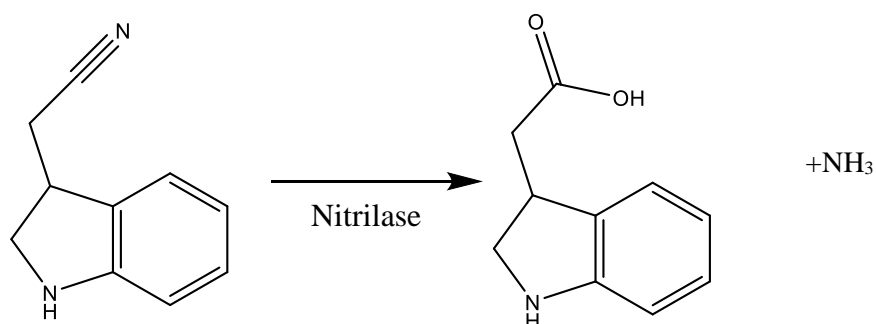
Figure 1.3 Conversion of indole-3-acetonitrile to indole-3-acetic acid mediated by the *Penicillium* nitrilase identified by Thimann and Mahadevan (1964).

Fusarium solani IMI 196840, which was isolated from bromoxynil-treated soil, displayed the ability to utilize benzonitrile as its sole nitrogen and carbon source (Harper 1977). The nitrilase from *F. solani* IMI 196840 was the only uniquely characterised fungal nitrilase over the next decade. Post discovery of the *F. solani* nitrilase, fungal nitrilases were more extensively studied. Nitrilases from sources such as *F. oxysporum f. sp. melonis, A. niger K10* and *F. solani O1* were purified and characterised (Winkler *et al.* 2009).

Nitrilase activity has seldomly been displayed by yeasts. More commonly, the nitrile hydratase/amidase system is exhibited. However, there have been nitrilases from yeast such as *Exophiala oligosperma* R1, *Cryptococcus sp.* UFMG-Y28 and *Torulopsis candida* GN405 (Rustler *et al.* 2008, Fukuda *et al.* 1973).

Plants

The nitrilases obtained from plant sources have displayed a wide substrate scope for both aliphatic and aromatic nitriles and are a significant part of the nitrilase superfamily. A study carried out by Thimann (1958) identified the first plant nitrilase from barley (*Hordeum vulgare* var. Six Row) with characterisation of the nitrilase confirming activity

towards indole-3-acetonitrile (IAN), producing the plant hormone auxin indole-3-acetic acid (IAA) (Thimann and Mahadevan 1964).

Ishikawa *et al.* (2007) cloned and characterised three isoforms of nitrilase from *Brassica rapa*. Two of the nitrilases namely BrNIT-T1 and BrNIT-T2 (80% amino acid homology to *Arabidopsis thaliana* AtNIT1 and AtNIT2 respectively*)*, displayed high activity towards IAN while one of the isoenzymes Br-NIT-T4 (90% amino acid homology to AtNIT4) displayed low activity for IAN. Three of the isoenzymes of *Arabidopsis thaliana* successfully converted IAN to IAA (AtNIT1 to AtNIT3) however, AtNIT4 which displayed 60% or less amino acid homology to the other three isoenzymes, did not accept IAN as a substrate, similar to the highly similar protein BR-NIT-T4. Further study on the most divergent nitrilase, AtNIT-4 was realised to be a β-cyano-L-alanine hydratase which demonstrated both nitrilase and nitrile hydratase (NHase) activity (Piotrowski *et al.* 2001).

Though there are many nitrilases and isoenzymes available from plant sources, such correlation of amino acid residues and activity indicates that nitrilase activity is dependent on the enzyme structure and so will display preferential hydrolysis for varying substrates (Vorwerk *et al.* 2001).

1.5.2 Nitrilase classification and structure

Nitrilase/cyanide hydratase enzymes form part of the larger 'nitrilase superfamily' of enzymes. The nitrilase/cyanide hydratase portion of the super family includes the cyanide hydratase, cyanide di-hydratase, nitrilase and the less closely related aliphatic amidase (O'Reilly & Turner 2003, Novo *et al.* 1995). The nitrilase superfamily is classified into thirteen branches based on sequence homology, however, nitrilase activity is only observed across branch one with branch two largely showing amidase activity. Substrate specificity of nitrilases is grouped based on their activity displayed towards aliphatic, aromatic or arylacetonitriles (Pace and Brenner 2001).

Branches one to three of the superfamily can be found across prokaryotes and eukaryotes, with branch one being the only group of enzymes which have been shown to act primarily on nitriles. Branches two and three have been shown to contain aliphatic amidases and amino-terminal amidases respectively (Pace and Brenner 2001).

The branch four group of amidase proteins prefer secondary amine substrates in contrast to amidase protein found in branches two and three which display activity towards more simple amines. The branch five proteins are related to the β-ureidopropionase enzymes. These enzymes are responsible for pyrimidine catabolism and the production of β-alanine (Pace and Brenner 2001). Branch six contains carbamlyase enzymes, which have been utilised for the production of broad spectrum β-lactam antibiotics such as ampicillin.

Branches seven to nine contain NAD synthetase enzymes which have been shown to contain a nitrilase related domain, as described by (Pace and Brenner 2001). This has afforded the branch seven to nine enzymes the ability to metabolise substrates such as glutamine into a simpler nitrogen source such as ammonia (NH_3).

The branch 10 group of proteins are denoted 'NIT proteins.' The NIT proteins were originally identified as an approximate 300 amino acid terminal extension on the Fhit tumour suppressor protein found in in human, worm and murine tissues. They have since been classified as Rosetta-stone proteins due to the fusion of the NIT1 and Fhit and correlating mRNA levels of both proteins in murine tissues (Pace and Brenner, 2001).

Finally, the branches 11 to 13 of the nitrilase superfamily have been shown to be rather non-classified or characterised. Brenner (2002) noted that the branch 12 enzymes display protein post-translational modification in that they are often fused with the RimI superfamily of N-terminal acetyltransferases (ribosomal proteins) with branch 11 and 13 classified as containing sequences of unknown specificity (Brenner, 2002).

The crystal structure of two nitrilase family members – NitFHit from *Caenorhabditis elegans* (Pace *et al.* 2000) and *N*-carbamyl-D-aminoacid-amidohydrolase from *Agrobacterium* sp. strain KNK712 (Nakai *et al.* 2000) confirm a four-layer sandwich structure α-β-β-α and active site consisting of the catalytic triad Glu – Cys – Lys which is presumed for all nitrilase enzymes. Atomic structure for three additional nitrilase family members are the putative CN hydrolyase from yeast (Kumaran *et al.* 2003), the proposed protein PH0642 from *Pyrococcus horikoshii* (1j31) (Sakai *et al.* 2004) and the amidase from *Geobacillus pallidus* RAPc8 (Kimani and Vinod 2007). These atomic structures display approximately 20% variance with divergent N- and C-termini (Thuku *et al.* 2007). The

monomer sandwich fold and the triad residues are centre to the nitrilase catalytic site and enhancing nitrilase performance (Pace *et al.* 2000).

Thiol binding reagents containing metals such as Ag^+ or Cu^+ appear to act as strong inhibitors of nitrilase enzymes, confirming the essential role of the cysteine residue for the enzyme activity. The nitrilase superfamily of enzymes have previously been visualised as tetramers, hexamers, octamers, tetradecamers, octadecamers however they always maintain a dimer interface (Thuku *et al.* 2009).

Nitrilases are generally present as inactive dimers in solution and are commonly found to contain six to 26 subunits which self-associate to form the active enzyme (O'Reilly and Turner, 2003). However, activity has been observed from four subunits of the *Fusarium oxysporum f.sp. melonis*, indicating that nitrilase enzymes can in fact be active between four and 26 subunits (Goldlust & Bohak 1989). Additionally, the nitrilase from *Pyrococcus abyssi* and the bromoxynil-specific nitrilase of *Klebsiella pneumoniae sp. ozaenae* are active as dimers (Mueller *et al.* 2006, Stalker *et al.* 1988). More recently monomeric nitrilases have been identified such as the 37 kDa NitM24D13 nitrilase from *Pyrococcus sp.* M24D13. The archaeal nitrilase displayed high similarity to both the molecular mass and the dimeric structure of *Klebsiella ozaenae* (38 kDa), *Pseudomonas xuorescens* DSM 7155 (40 kDa), and actinomycete *Rhodococcus rhodochrous* J1 (41 kDa) (Mueller *et al.* 2006) .

As previously noted, nitrilases maintain a dimer interface and subunit consisting of a single peptide with a molecular mass measuring between 32 to 45 kDa which aggregates to form active homoligomers (Thuku *et al.* 2009, Banerjee *et al.* 2002, O'Reilly and Turner 2003). However, there are some nitrilase enzymes which have a subunit molecular mass which is greater than the commonly observed 32- 45 kDa. Two nitrilases have been characterised from *Fusarium solani* IMI196840. The first identified nitrilase displayed a subunit molecular mass of 72 kDa, however the second and newly characterised nitrilase displayed a subunit molecular mass of 40 kDa which was more closely related to the subunit size observed in *Fusarium solani* O1 (Harper, 1977) and *F. oxysporum f. sp. melonis* (Goldlust and Bohak 1989). The active form of the *F. solani* IMI196840 nitrilase measured 620 kDa (Harper, 1977).

A study carried out by Kobayashi *et al.* (1985) characterised a nitrilase from *Nocardia sp.* (*Rhodococcus*). The nitrilase from *R. rhodochrous* J1 has been extensively studied. The nitrilase enzyme was determined to have a molecular mass of 76 kDa, with two subunits of identical molecular mass. The enzyme was shown to contain all the common amino acids with a strong prevalence of alanine, glutamic acid, aspartic acid and valine residues (Kobayashi *et al.* 1989). The nitrilase of *R. rhodochrous* J1 is unusual as generally nitrilase enzymes have one subunit molecular mass in the average range of 30-45 kDa which self-aggregate to form larger active enzymes such as the nitrilase from *Rhodococcus rhodochrous* NCIMB 11216, which had a subunit of 45 kDa and an active aggregated to form a 560 kDa (12 subunits) (Hoyle *et al.* 1998). Similarly, the nitrile from *Rhodococcus rhodochrous* K22 has a subunit mass of 41 kDa forming a multimeric aggregated enzyme of 650 kDa (15-16 subunits) (Kobayashi *et al.* 1990).

1.5.3 Nitrilase activity

Nitrilase enzymes have been well characterised since the discovery of the first plant nitrilase in 1958 and the first bacterial nitrilase in 1964 (Thimann and Mahadevan 1958, Robinson and Hook 1964). Nitrilases exhibit a range of different physicochemical characteristics along with a diverse substrate range occurring in both prokaryotes and eukaryotes (Thuku *et al.* 2007).They have been further sub-divided into three categories based on the wide substrate range they exhibit; (1) nitrilases that hydrolyse aromatic nitriles (2) nitrilases that favourably hydrolyse aliphatic nitriles and (3) nitrilases that favourably hydrolyse arylacetonitriles (Banerjee *et al.* 2002, Brady *et al.* 2004, Mueller *et al.* 2006). Nitrilases, unlike NHases, do not require any metal co-factor or prosthetic group for activity. Nitrilase enzymes are generally inducible in nature however there are some reports of constitutive nitrilase enzymes in *Klebsiella ozaenae* (Stalker *et al.* 1988)*, Brevibacterium strain* R312 (Gradley *et al.* 1994), *Bacillus subtilis* ZJB-063 (Zheng *et al.* 2008), *Alcaligenes* sp. ECU0401 *(Zheng et al.* 2008) and, *Rhodococcus rhodochrous* BX2 (Zheng *et al.* 2008). All known fungal nitrilases are inducible like the majority of bacteria nitrilases. 2-Cyanopyridine has been utilized as a common inducer for filamentous fungi, however, low activity is generally observed from the majority of fungi (Martínková *et al*. 2009). They are reported to have catalytically essential cysteine residues at, or near, the active site (Kobayashi *et al.* 1992).

Nitrilases enzymes from both plant and microbial sources have demonstrated broad substrate activity across aliphatic and aromatic nitriles. A study published by Robertson *et al*. (2004) identified 137 unique nitrilase sequences from the construction and screening of a metagenomic library. These nitrilases were then classified into clades based on their DNA sequence homologies (Figure 1.4).

Figure 1.4 Clade classification of 137 novel nitrilase sequences as per Robertson *et al*. (2004). Six clades were identified based on DNA sequence homology.

The newly discovered nitrilase enzymes were then biochemically characterised. Each of the six clades displayed correlation between the enzyme's clade classification and its activity or enantioselectivity towards a particular substrate. The enzymes within the six clades were tested towards 3-hydroxyglutaronitrile (3-HGN) and mandelonitrile.

Clade one (1A and 1B inclusive) displayed the highest diversity regarding the enzymes enantioselectivity. Activity towards both the R- and S- enantiomers is observed (Figure 1.4). The biochemical analysis of these enzymes confirmed clades one, three, four and six displayed a preference for the S-enantiomers of these substrates. In contrast to this, clades two and five displayed a preference for R-enantioselectivity towards both 3-HGN and mandelonitrile (Figure 1.4). Based on the biochemical analysis, it was evident that the mandelonitrile was proving to sub-optimal substrate for several of the identified enzymes. The majority of the clades displayed little to no activity towards the substrate, however clade two displayed high activity and also high enantioselecivity towards the production of R-mandelic acid (Figure 1.4). This was the first study that displayed such a correlation across a large number of novel nitrilase sequences.

The nitrilase from *Aspergillus niger* K10 was cloned and characterised by Kaplan *et al.* (2006) and is one of the most extensively studied fungal nitrilases alongside *F. oxysporum f. sp. melonis, F. solani* O1 and *F. solani* IMI196840 (Vejvoda *et al.* 2010, Vejvoda *et al.* 2008, Goldlust and Bohak, 1989). *A. niger* K10 was identified to be an aromatic nitrilase with little to no activity observed on aliphatic substrates, similar to the *Fusarium* genus (Kaplan *et al.* 2006). Interestingly, the recombinant nitrilase did result in the production of amide intermediates, this was more pronounced for some substrates such as heteroaromatic nitriles, this was also observed with the nitrilase of *Fusarium oxysporum* f. sp. *melonis* (Goldlust and Bohak, 1989). The formation of an amide product was also observed by the two nitrilases AtNit1 and AtNit4 from *Arabidopsis thaliana,* with the quantity and enantioselectivity found to be strongly dependent on nitrile structure (Osswald *et al.* 2002). Additionally, the bacterial nitrilase from *Rhodococcus erythropolis* ATCC 39484 when screened towards phenylacetonitrile, produced the corresponding amide which accounted for approximately two percent of the final product (Stevenson *et al.* 1992).

The nitrilase activity of *Rhodococcus rhodochrous* J1 has been extensively characterised (Nagasawa *et al.* 2000, Komeda *et al.* 1996, Nagasawa *et al.* 1990, Kobayashi *et al.* 1989) The nitrilase was first cloned and expressed by Kobayashi *et al.* (1989) where enzyme kinetic studies showed that the enzyme was capable of converting mono-substituted benzonitrile derivatives however, steric hinderance and in turn a decrease or loss in activity was observed with bulky side groups. Further study of the *Rhodococcus rhodochrous* J1 nitrilase displayed results which suggested a high affinity for an aromatic or heteroaromatic ring, for example benzonitrile (Kobayashi *et al.* 1989). Later research involving induction with ε-Caprolactam resulted in the nitrilase displaying a higher relative activity towards acrylonitrile than the aromatic benzonitrile (Nagasawa *et al.* 2000). Prior to this paper it was thought that only nitriles could induce nitrilase activity, the identification of ε-caprolactam as a suitable inducer for the *R. rhodochrous* J1 nitrilase was a positive outcome as caprolactam a cheaper, less toxic alternative than using the previous inducer isovaleronitrile. Moreover, caprolactam induction was also observed across some other *Rhodococcus sp.* strains and the arlyacetonitrilase producing *Alealigenes faecalis* JM3 (Nagasawa *et al.* 1990).

1.5.4   Nitrilase gene structure

It is common to find gene clusters in bacteria, such as operons and regulons. The organisation of genes into clusters generally reflects the involvement of these genes in a communal metabolic process or their involvement in supramolecular complexes. (Lathe *et al.* 2000, Rogozin *et al.* 2002, Lawrence 2003). There is a conserved Nit 1C cluster which contains seven genes from the nitrilase subfamily/clade 1 which was discovered by Robertson *et al.* (2004), six of these are genomic with one found on a plasmid (Podar *et al.* 2005).

A study carried out by Podar *et al.* (2005) revealed that six related coding sequences, which comprise part of the Nit1 cluster (undescribed prior to this study) were located on the same DNA strand, suggesting they are part of a regulon/operon identified in six sequenced bacterial genomes including *Synechocystis sp.* PCC 6803, *Synechococcus sp.* WH 8102 SYNW1417, *Rubrivivax gelatinosus* PM1 (Rgel01_2) lytic, *Burkholderia fungorum* (contig 416), *Photorhabdus luminescens* sub-*sp. laumondii* TTO1, *Verrucomicrobium spinosum* and the *Klebsiella pneumoniae* virulence plasmid pLVPK. They predicted the order of these genes in the operon was as follows; '(1) hypothetical protein, (2) nitrilase, (3) radical S-adenosyl methionine superfamily member, (4) acetyltransferase, (5) AIR synthase, and (6) hypothetical protein' (Podar *et al.* 2005). The seventh gene encoded a flavoprotein. It was located at either the beginning of the gene cluster but on the opposite strand, or the end of the gene cluster on the same strand, depending on the species studied.

Further research identified variations in arrangement of this cluster in *Verrucomicrobium spinosum, Burkholderia* and *Rubrivivax* (Figure 1.5 and Figure 1.6). These variations in arrangement of this cluster correlated with the major taxonomic bacterial groups, however, outside of this cluster no conservation was demonstrated between different bacterial species with regard to genes or communal metabolic processes (Podar *et al.* 2005).
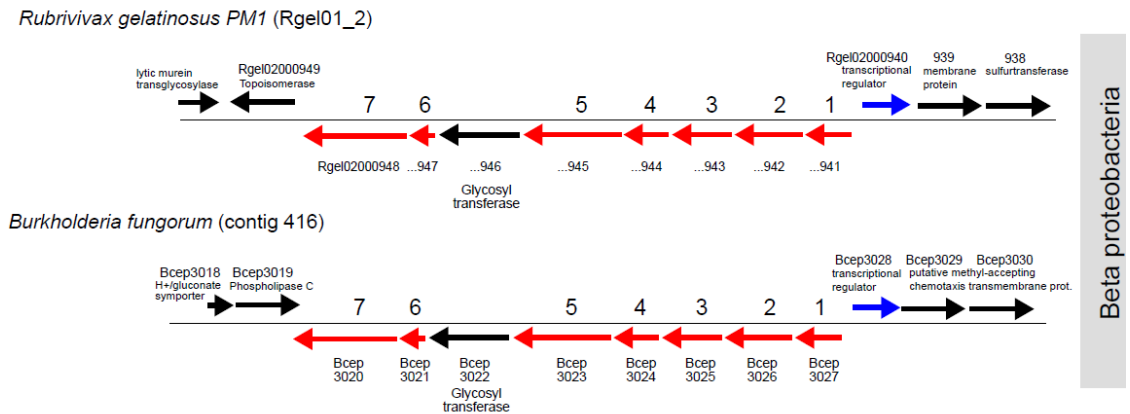
Figure 1.5 Architectural differences within the Nit 1C gene cluster between two beta proteobacterial: *Rubrivivax gelatinosus* PM1 and *Burkholderia fungorum* (Podar *et al.* 2005). (1) hypothetical protein, (2) nitrilase, (3) radical S-adenosyl methionine superfamily member, (4) acetyltransferase, (5) AIR synthase, (6) hypothetical protein and (7) flavoprotein. The conserved Nit 1C gene cluster is flanked by unrelated genes which are species dependent.
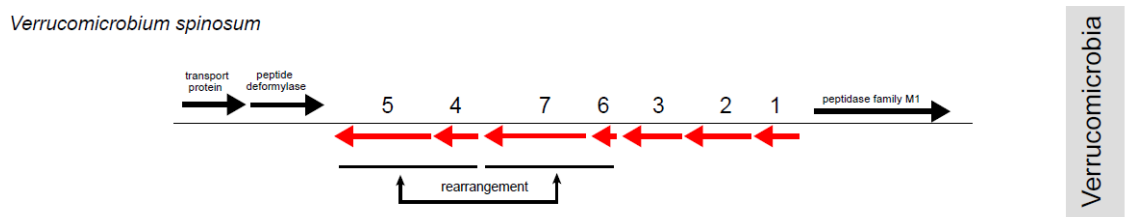
Figure 1.6 Rearrangement of genes within the conserved Nit 1C gene cluster in *Verrucomicrobium spinosum* (Podar *et al.* 2005).
(1) hypothetical protein, (2) nitrilase, (3) radical S-adenosyl methionine superfamily member, (4) acetyltransferase, (5) AIR synthase, (6) hypothetical protein and (7) flavoprotein.

The Nit 1C gene cluster was not evident in subfamily/clade 3, while the other nitrilase subfamilies, which are distantly related (4, 5 and 6), displayed no obvious connections with the Nit 1C conserved gene cluster (Podar *et al.* 2005). The described research presented instances where some genera or species e.g. *Klebsiella pneumoniae,* contained two nitrilase genes which displayed sequence homology for different subfamilies. Many studies have assigned occurrences such as horizontal gene transfer (HGT) (Coffey *et al.* 2010, Coffey *et al.* 2009, O'Mahony *et al.* 2005, Lorenz and Wackernagel 1994) coupled with neofunctionalization and ancient gene duplication events as a means of explanation (Podar *et al.* 2005).

The nitrilase gene (nitA) from *R. rhodochrous* J1 has been characterised and overexpressed in *E. coli* cells (Kobayashi *et al.* 1992). Upon sequencing analysis of a 1.4 kB region an open reading frame (nitR) of 957 bp was identified. Complete loss of nitrilase activity was observed upon deletion of both the central and 3' portion of the open reading frame. The nitR gene codes for a positive transcriptional regulator for nitrilase (nitA) expression (Komeda *et al.* 1996). NitR was essential for nitrilase activity as the protein product produced was required for the induction of nitA by isovaleronitrile (Komeda *et al.* 1996). The gene was located directly in the 3' flanking region of nitA. The amino acid sequence of nitR showed resemblance to the positive regulator families XylS and AraC of *Pseudomonas putida* and *E. coli* respectively (Komeda *et al.* 1996).

Rhodococcus erythropolis AJ270 has also been shown to contain a nitrilase gene (Coffey, 2007) with the predicted protein sequences displaying 60-75 % homology to various nitrilase protein sequences within the 2A clade of nitrilases as presented by Robertson *et al.* (2004). Low sequence homology was observed across the 3' region of the nitrilase in comparison to the 5' end upon alignment to all significantly similar nitrilases. Though the recombinant nitrilase was cloned and expressed, limited activity was observed from the AJ270 nitrilase and at present it is thought it may resemble an arylactonitrilase as the only detected activity is towards phenylacetonitrile. However, further characterisation of the enzyme is required before this can be determined conclusively. This research has not yet been continued to date as the nitrilase gene was shown to be unstable upon culturing, hindering any purification or characterisation.

1.6 Nitrile Hydratase (4.2.1.84)

The nitrile hydratase (NHase) enzyme is part of a two-step process to nitrile hydrolysis involving a second enzyme namely an amidase (Figure 1.7). Firstly, the nitrile is hydrolysed via the NHase enzyme to an amide. The amidase then converts the amide to a carboxylic acid and the corresponding ammonium ion.
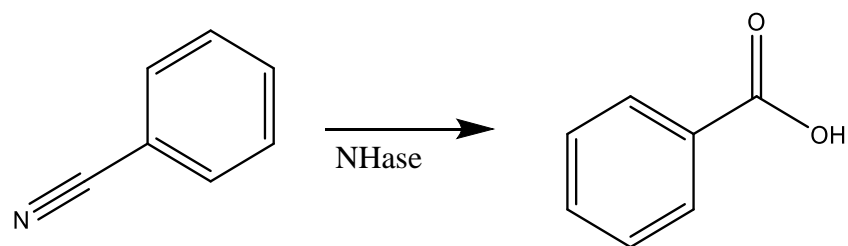
Figure 1.7 Conversion of benzonitrile to benzylamide mediated by an NHase enzyme

1.6.1   Sources

The NHase enzyme was first discovered from the bacterium *R. rhodochrous* J1, originally called *Arthrobacter sp.* J1 (Prasad 2010, Asano *et al.* 1980). It is common for some microorganisms such as *R. rhodochrous* J1 tohave both a nitrilase and NHase/amidase system. The three enzymes can be selectively induced via substitution of either the nitrile or amide in the media (Prasad, 2010). NHase/amidase systems have been recognised in many prokaryotic and eukaryotic systems.

1.6.1.1   Bacterial

Unlike nitrilases, the majority of the NHase enzymes belong to the many genera of phyla: Proteobacteria, Actionobacteria, Cyanobacteria and Firmicutes. Various species of Proteobacteria (*Acidovorax, Agrobacterium, Burkholderia, Bacillus, Bradyrhizobium, Comanonas, Klebsiella, Mezorhizobium, Moraxella, Pantoea, Pseudomonas, Rhizobium, Rhodopseudomonas, Serratia*) and   Actinobacteria (*Amycolaptosis, Arthrobacter, Corynebacterium , Microbacterium, Mircocossus, Nocardia, Pseudonocardia and Rhodococcus*) have exhibited NHase activity (Prasad and Bhalla 2010). Even though the above have reported nitrile hydratase activity, the NHase enzyme has been most extensively studied from *Rhodococcus* species.

1.6.1.2 Eukaryotic

Some fungi have exhibited NHase activity; NHase activity has been displayed by *Trichoderma sp., Myrothecium verrucaria,* and yeasts such as *Aureobasidium pullulans, Candida sps.,* and *Pichia kluyveri* (Jallageas *et al.* 1980, Maier-Greiner *et al.* 1991, Rezende *et al.* 1999). NHase activity has not been described in plants, however, phylogenetic analysis does suggest that *Ricinus communis* contains molecular evidence for the presence of NHase gene.

Phylogenetic analysis of *M. brevicollis* did indicate that the NHase gene contained both α and β subunits (Foerstner *et al.* 2008). Additional research by Marron *et al.* (2012) confirmed that the α and β subunits were expressed as a fusion protein. These may have been obtained by gene transfer from a proteobacterial species, facilitated by its phagotrophic method of feeding (Marron *et al.* 2012). Prior to work carried out by Marron *et al.* (2012), NHase genes had been described in only two distantly related eukaryotic species, *Monosiga brevicollis* and *Aurecococcus anophagefferens* (Foerstner *et al.* 2008, Gobler *et al.* 2011*).*

The NHase gene discovered in *Aurecococcus anphagefferens* (eukaryotic picoplankton) only encoded for an alpha (α) subunit of the NHase protein. Further analysis identified the NHase gene to be more similar to the *M. brevicollis* NHase gene that that of a prokaryote NHase. Further studies must be carried out on the characterisation and purification of these eukaryotic NHase enzymes before a comprehensive view of their origin, activity and frequency distribution can be gathered.

1.6.2 Nitrile Hydratase structure

NHase enzymes are metalloenzymes classified into two groups either iron (ferric NHases) or cobalt (cobalt NHases).The grouping depends on the catalytic ion contained at the catalytic centre of the enzyme. Banerjee *et al.* (2002) indicated that the metal ion contained within the enzyme's active site is either required to accommodate protein folding or to act as a catalyst for the hydrolysis of -CN.

Prokaryotic NHase enzymes are composed of an α and β subunit of varying molecular masses (kDa) (Mizunashi *et al.* 1998). Bacterially-encoded NHases have been shown to contain their active site within the α-subunit which contains 3 cysteine residues (Martinez *et al.* 2017). More recently a study by Nelp *et al.* (2014) identified a rare NHase composed of three subunits. The three subunits of the toyocamycin NHase (TNHase) displayed high homology to the α- and β- subunits of known NHase genes (Figure 1.8). The TNHase was first was purified from *Streptomyces rimosus* and used for the conversion of the nucleoside antibiotic toyocamycin to sangivamycin. This putative eukaryotic NHase displays a fused α- and β-subunit structure which is bridged by an insert with multiple His residues which can vary in number, as with recently cloned and expressed NHase from M. *brevicollis* $(His)_{17}$ (Martinez *et al.* 2017).
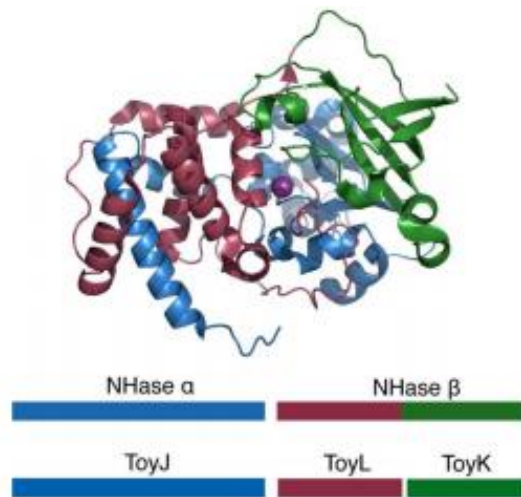
Figure 1.8 The toyocamycin NHase three subunit formation as determined by (Nelp *et al*. 2014)

Aside from the rare NHase structure observed with the TNHase, the NHases produced by *R. rhodochrous* J1 are the most studied. Most commonly one organism produces one NHase enzyme; however, *R. rhodochrous* J1 is unique as it produces two NHases, both a high molecular weight (H-NHase) measuring 530 kDa (Mizunashi *et al.* 1998) and a low molecular weight (L-NHase) measuring 94 kDa (Lan *et al.* 2017). Analysis of the amino acid composition highlighted alanine, glutamic acid, aspartic acid and valine to be the most prevalent residues (Kobayashi *et al.* 1989).

In 2017 Martinez *et al.* cloned and expressed the putative NHase from *M. brevicollis* which was previously identified in 2008 by Foerstner *et al*. The Co-type

NHase was successfully expressed and active without an activator protein or the *E. coli* molecular chaperones GroEs/EL. The NHase was expressed as a single polypeptide, which was achieved by the fusion of the α- and β-subunits using seventeen-histidines as a linker region. Though this is the first example of an expressed and active fused eukaryotic NHase,the NHase from *Pseudomonas* putida has also been expressed as an active fused α and β subunit (artificially created via PCR) (Xia *et al.* 2016).

1.6.3   Nitrile Hydratase activity

NHases have displayed a broad substrate range with some displaying preferential hydrolysis for either aliphatic or aromatic nitriles, with cobalt type NHases displaying a more varied substrate range in comparison to the iron type NHases (Desai and Zimmer 2004). The cobalt-type high-molecular mass NHase of *R. rhodochrous* J1 displayed a preferential hydrolysis for aliphatic nitriles whereas the low-molecular mass NHase displayed a preferential hydrolysis for aromatic nitriles (Wieser *et al.* 1998, Kobayashi *et al.* 1993). The induction of the *R. rhodochrous* J1 NHases is regulated with the supplementation of media with the appropriate inducer, which interestingly differs for the H-NHase and L-NHase. The H-NHase is induced with the addition of urea with the L-NHase induced via supplementation of cyclohexanecarboxamide (Kobayashi 1996). Both Co-type NHases required the supplementation of the cobalt co-factor for activity.

Cobalt type (Co) NHases have shown a preference for the hydration of aromatic nitriles and iron type (Fe) NHases have shown higher catalytic activity for aliphatic nitriles, though this may not always be the case as seen with the H-NHase of *R. rhododchrous* J1 (Table 1.1) (Wieser *et al.* 1998). The substrate specificity of a NHase enzyme can be altered by changing the metal ion, from a Co-type to Fe-type, at the centre of the active site (Miyanaga *et al.* 2001). In comparison to the NHases from *R. rhodochrous* J1 the Fe-type NHase from *R. erythropolis* AJ270 has displayed a broad substrate range encompassing aliphatic, aromatic and heterocyclic nitriles and amides (O'Mahony *et al.* 2005).

Table 1.1 NHase activity from a variety of microorganisms with their corresponding Co/Fe type and substrate specificities.

Microorganism	Co/Fe type	Substrate Specificity	Reference
Rhodococcus rhodochrous J1 (High molecular mass)	Co	aliphatic	(Wieser *et al.* 1998)
Rhodococcus rhodochrous J1 (Low molecular mass)	Co	aromatic	(Wieser *et al.* 1998)
Pseudomonas chlororaphis B23	Fe	aliphatic	(Nagasawa *et al.* 1987)
Rhodococcus sp. N771	Fe	aliphatic	(Endo *et al.* 2001)

A dissimilarity between the Co-type and Fe-type NHases is that the Co-type do not exhibit photo-reactivity, which is uniquely found among Fe-type NHases (Nojiri *et al.* 2000). The most extensively studied Fe-type NHase is that of *Rhodococcus erythropolis* N-771. This enzyme is unusual in that a loss of activity was observed upon incubation of the enzyme in the dark, with activity almost fully restored upon photoactivation (Endo *et al.* 2001). Similar photoactivity of NHase enzymes can also be observed with *Rhodococcus erythropolis* N-774 and R312 (Brandão *et al.* 2003). O'Mahony *et al.* (2005) characterised the NHase gene clusters of *R. erythropolis* AJ270, *R. erythropolis* AJ300 and *Microbacterium* sp. AJ115. Interestingly, the Fe-type NHase from AJ270 did not display photosensitivity.

It has widely been shown that the both Co- and Fe- type NHase enzymes require an activator protein to allow for functional expression of the enzyme in *E. coli*. The P47K protein observed in the *Rhodococcus erythropolis* JCM6826 gene cluster has been confirmed as an essential activator protein for the functional expression of NHase enzymes in *E. coli* from a number of *Rhodococcus* sp. (N771 and N774) and in *Pseudomonas chlororaphis* B23. In some species this P47k protein may be denoted as P44K (Song *et al.* 2008). Though Co- and Fe-type NHase enzymes share high sequence identity, their respective activator proteins are different sizes and share little sequence identity. The activator protein for Co-type NHase measures approximately 14 kDa with the Fe-type measuring approximately 47 kDa (Zhou *et al.* 2010, Cameron *et al.* 2005)

Interestingly, the Fe-type NHase from *Comamonas testosteroni* (formally *Pseudomonas testosteroni*) Ni1 did not require the activator protein for activity when cloned into *E. coli* (Kuhn *et al.* 2012). However, upon the analysis of the Fe-type NHase, two histidine residues were present at the catalytic site which are not found with other Fe-type NHases. It is thought that the two additional histidine residues alongside the active site location (located near an open solvent channel) may aid with iron incorporation without the aid of an activator protein (Kuhn *et al.* 2012) Additionally, the Co-type NHase from *M. Brevicollis* was actively expressed in *E. coli* without the aid of an activator protein or the *E.coli* molecular chaperones (Martinez *et al.* 2017).

As earlier stated, the active site of bacterially encoded NHase enzymes is found within the α-subunit which generally contains three cysteine residues (Martinez *et al.* 2017). Nelp *et al.* (2014) has shown that the α-subunit of TNHase was sufficient to build the active site complex. Activity assays also confirmed that the α-subunit displayed activity towards toyocamycin and a niacin precursor 3-cyanopyridine. However, it is important to note that the remaining two subunits did account for increased substrate specificity and catalytic efficiencies (Nelp *et al.* 2014)

1.7 Amidase (E.C. 3.5.1.4)

Amidase enzymes hydrolyse amides to carboxylic acids and ammonia (Figure 1.9). They are abundant in nature and are commonly found as part of a bi-enzymatic pathway with NHase. Amidase enzymes have potential production applications for a number of fine chemical intermediates and products (Sharma *et al.* 2009).
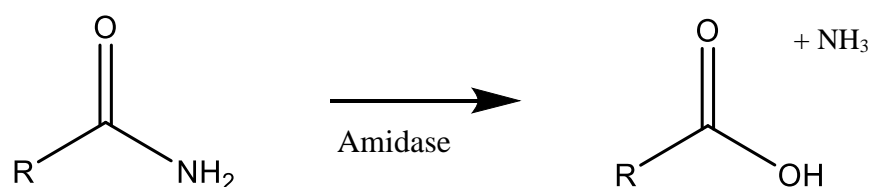
Figure 1.9 Conversion of a simple amide to the corresponding carboxylic acid via amidase.

1.7.1   Sources

1.7.1.1   Bacteria

Many of the discovered amidases are obtained from bacterial sources. *Rhodococcus, Corynebacterium,Mycobacterium, Pseudomonas, Bacillus, Micrococcus, Brevibacterium, Nocardia, Streptomyces, Blastobacter, Arthrobacter, Alcaligenes* and *Lactobacillus* are some of the bacteria which exhibit amidase production (Fournand and Arnaud 2001).

The substrate specificity of each amidase is dependent on the organism from which it is found. For instance the amidase purified and characterised from *R. rhodochrous* J1 was found to have a very wide substrate range, in contrast to this an amidase purified from *Mycobacterium avium* was specific for nicotinamide (Kobayashi *et al.* 1993, Kimura 1959).

1.7.1.2   Fungi

There is limited literature available on amidases purified and characterised from yeast. *Candida utilis* has been shown to produce an amidase which hydrolyses a wide spectrum of substrates (Brady 1969). It was shown to hydrolyse substrates such as acetamide, propionamide, butyramide, valeramide, hexanoamide and acrylamide (Brady 1969). Another amidase, purified from the yeast *Torula cremoris (*Joshi and Handler 1962), did not display a wide spectrum range like the amidase purified from *Candida utilis* - instead it displayed specificity for the substrate nicotinamide.

A fungal amidase was purified from *Aspergillus nidulans* which is the most studied fungal strain to date (Fournand and Arnaud 2001, Hynes and Pateman 1970). A formamidase, acetamidase and a wide spectrum amidase are amongst several fungal amidases which have been identified and studied (Fournand and Arnaud 2001).

1.7.1.3   Plants

A partially purified peptide amidase (Pam) from *Citrus sinensis L.* (oranges) showed high stereo- and regio- selectivity. Pam's are responsible for the hydrolysis of C-terminal amide groups in peptides and N-protected amino acids. It did not deaminate amino-acid amides and only L-amino acid amides proved to be suitable substrates, displaying

stereoselectivity (Kammermeier-Steinke *et al.* 1993). An amidase present in the *Arabidopsis thaliana* catalysed the hydrolysis of IAA, a plant growth hormone from indole-3-acetamide (Pollmann *et al.* 2003).

1.7.2 Structure

The protein structure of aliphatic amidases comprise the α/β hydrolase fold like that of the nitrilase superfamily, where signature amidases are related to aspartic proteinases (Sharma *et al.* 2009). *Pseudomonas aeruginosa* contains an inducible aliphatic amidase encoded by the amiE gene where each monomer of the homo-hexameric 38 kDa enzyme displayed a four-layer αββα sandwich domain. It also contains an additional 81-residue long C-terminal segment. Further crystallography studies show that the C-terminal residues wrap around the c-terminal segment of the adjoining monomer forming tightly packed dimers (Figure 1.10) (Andrade *et al.* 2007).
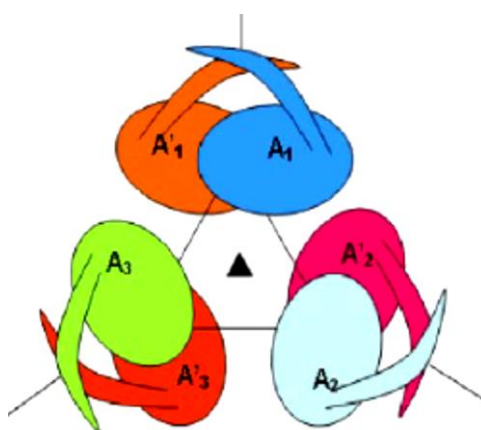
Figure 1.10 Structural representation of the homo-hexameric aliphatic amidase from P. *aeruginosa* (Andrade *et al.* 2007).

Tightly packed homodimer formation can be observed with interlinking long C-terminus residues

The amidase from *R. rhodochrous* J1 expressed as part of the L-NHase gene cluster was measured to be 55 kDa with further analysis indicating that the active enzyme was composed of two 55 kDa subunits (Kobayashi *et al.* 1993). In contrast to this the amidase from *Rhodococcus erythropolis* MP50 was suggested by the authors to be composed of eight identical subunits measuring at approximately 61 kDa each (Hirrlinger and Stolz 1996). The homodimer amidase from *Pseudomonas chlororaphis* B23 displays a molecular mass of 54 kDa (Ciskanik *et al.* 1995). The enantioselective amidase from

Brevibacterium sp. R312 is thought to be a homodimer measuring 54.6 kDa with the strain also containing an acylamide amidohydrolase (wide spectrum amidase) which presents as a tetramer with identical subunits measuring approximately 43 kDa (Mayaux *et al.* 1991).

1.7.3 Amidase activity

Amidases are known to hydrolyse a wide variety of amides ranging from short chain aliphatic amides, mid-chain amides, arylamides, α-aminoamides and α-hydroxyamides. (Sharma *et al.* 2009) The grouping of amidases is based on their preferred substrate or their catalytic site (Fournand and Arnaud 2001). Amidases with specific activity can be separated from those which are non-specific. Amidase enzymes have displayed the ability to withstand extremes of temperature and pH which is linked to the tightly packed dimers which are formed by the interlinking long C-terminus residues as in the case with the aliphatic amidase from *P. aeruginosa* (Andrade *et al.* 2007). The amidase from *R. rhodochrous* J1 also displayed stability at a broad pH range from 6.7 to 10.0 (Kobayashi *et al.* 1993). Similarly, the amidase from *P. chlororaphis* B23 displayed a broad pH range as the enzyme exhibited 100% activity between pH 5.9 and pH 9.9. The amidase enzyme which was cloned and expressed from *R. rhodochrous* J1 displayed activity towards both aliphatic and aromatic amides.

The amidase from *R. erythropolis* MP50 displayed a broad substrate scope across a number of aliphatic and aromatic nitriles when induced with naproxen nitrile or naproxen amide (Hirrlinger and Stolz 1996). The enzyme was enantioselective (99 ee%) toward the production of the S-enantiomers of 2-phenylpropionamide, naproxen amide, and ketoprofen amide. Similarly, the amidase from *P. chlororaphis* B23 displayed enantioselectivity towards 2-phenylpropionamide (100 ee%) (Ciskanik *et al.* 1995). Additionally, the amidase which was cloned and expressed from *Brevibacterium* sp. strain R312 displayed enantioselectivity to 2-phenylpropionamide.

As microbial amidases are part of the 13-branch nitrilase superfamily it is not unusual that the active site of the amidase from *P. aeruginosa* was identified to contain the same Glu-Lys-Cys triad. Single point mutations of the cystine residue saw profound changes to the substrate scope of this once limited aliphatic amidase. The newly mutated enzyme

displayed activity towards longer amides (butyramiade and valeroamide) and aromatic amides (phenylacetamide) (Andrade *et al.* 2007). The native amidase of *Klebsiella pneumoniae* NCTR 1 is unusual as upon incubation of the enzyme with ethylenediaminetetraacetic acid (EDTA) (10 mM), a 5% loss in amidase activity was exhibited which the authors proposed was due to the chelation of the iron and cobalt metals (Nawaz *et al.* 1996).

1.7.4 NHase and Amidase gene cluster

NHase and amidase coexist in a bi-enzymatic pathway in most sources. These genes plus those which are associated with nitrile metabolism and/or regulation are known as nitrile metabolism gene clusters. The gene clusters associated with the Co-type NHase of *R. rhodochrous* J1 and the Fe-type NHase of *Rhodococcus* sp. N771 have been extensively studied and characterised.

The gene arrangement of the amidase and NHase genes from *R. rhodochrous* J1 is different to that of *R. erythropolis* N-774 and *P. chlororaphis* B23, in that the amidase is downstream of the NHase α and β subunits (Figure 1.11) (Sakashita *et al.* 2008). The gene arrangement is also different as the amidase and NHase gene are not co-expressed in *R. rhodochrous* J1 (Cameron *et al.* 2005).
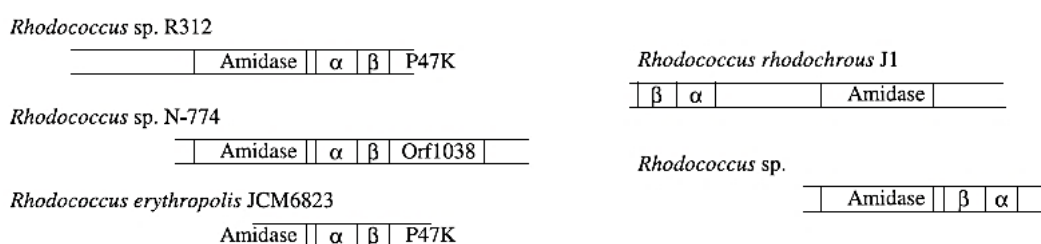
Figure 1.11 Structural organisation of the NHase and amidase genes found in the following *Rhodococcus* species; *R. rhodochrous* J1, *R. erythropolis* JCM6823, *Rhodococcus* sp. R312 and *Rhodococcus* sp. N-774. (Fournand and Arnaud 2001).

The α and β represent the alpha and beta subunits of the nitrile hydrates genes respectively, which are associated with the bi-enzymatic system. The open reading frame (ORF 1038) has no clear function to date with P47K acting as a activator protein required for NHase expression in *E. coli* (Song *et al.* 2008).

The NHase/amidase gene clusters of *R. rhodochrous J1*

As previously stated, two NHase enzymes are produced by *R. rhodochrous* J1, a high molecular mass (H-NHase) and a low molecular mass (L-NHase). Both the H-NHase and the L-NHase are Co-type metalloenzymes and contain two subunits, α and β (Mizunashi *et al.* 1998). The amidase which is coupled with the L-NHase in *R. rhodochrous* J1 consists of two subunits identical in molecular mass (55 kDa). Induction of this amidase with propionamide leads to the expression of the amidase with a preference for aliphatic substrates, however, the amidase displays a preference for aromatic amides upon induction with benzamide (Kobayashi *et al.* 1993).

The gene cluster of *R. rhodochrous* J1 as described by Komeda *et al.* (1996) has shown the H-NHase to contain an α subunit (26 kDa) and β subunit (29 kDa) encoded by the *nhh*A and *nhh*B genes respectively (Figure 1.12). The open reading frames *nhh*C and *nhh*D are required for activation of NHase, where *nhh*D is thought to be a positive regulator. NhhC displays significant homology to the amino acid sequence for the negative regulator AmiC of *P. aeruginsoa* (Wilson and Drew 1991). This negative regulator for the aliphatic amidase has been induced with low-molecular mass amides (propionamide) (Komeda *et al.* 1996).
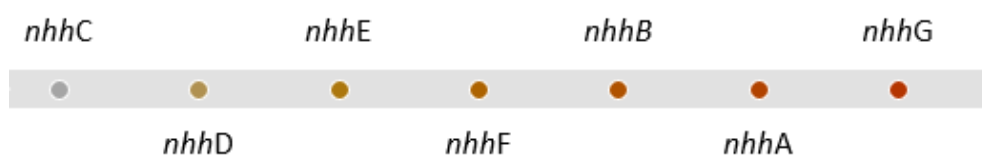
Figure 1.12 Schematic of the H-NHase gene cluster of *R. rhodochrous* J1.
The genes *nhh*C and *nhh*D code for activation of NHase with *nhh*C displaying homology to a negative transcriptional regulator and *nhh*D displaying homology to a positive regulator. *The nhh*F gene codes for an insertion sequence with *nhh*B and *nhh*A denoting the β- and α- subunits of the NHase gene respectively. *Nhh*G has been shown to code for a metallochaperone.

An insertion sequence IS*1164* is encoded by *nhh*F. This insertion sequence represents a mobile DNA element and may be the basis for the lack of the amidase gene in this cluster, regardless of the fact that H-NHase is induced by the presence of amides. *Nhh*E is not

required for NHase activity and its function remains unclear. It is suggested that *nhh*G functions a self-subunit swapping chaperone but also as a metallochaperone in a subunit specific reaction (Zhou *et al.* 2010).

The genetic structure of the L-NHase gene cluster was also determined Komeda *et al.* (1996).

Figure 1.13 Schematic of the L-NHase gene cluster of *R. rhodochrous* J1.

The gene *nhl*D denotes a gene of regulatory function, *nhl*C a proposed positive regulator for NHase expression with *nhl*D a proposed negative regulator for NHase expression. The *nhl*B and *nhl*A represent the β- and α- subunit of the NHase gene respectively with *amd*A representing an amidase. The nhlF gene has not formally been identified with nhlE representing a self-subunit swapping chaperone.

The gene cluster of L-NHase from *R. rhodochrous* J1 contains an amidase gene (*amd*A) and genes for both α and β subunits of NHase, *nhl*A and *nhl*B respectively. *nhl*D gene has shown 44.8% similarity to the regulatory genes *mer*R from *Streptomyces lividans* and 45.9% to *ars*R from *E. coli* which are of heavy metal resistance (Komeda *et al.* 1996). From this, it is thought that *nhl*D may function as a redundant Co^+ sensory protein. Komeda *et al.* (1996) indicated that there is large amount of non-coding sequence between the three genes *nhl*C, *nhl*D, and *nhl*B. A 357 bp region of non-coding sequence was identified between the *nhl*C and *nhl*D, with 884 bp identified between *nhl*C and *nhl*B suggesting that these genes are independently regulated. The *nhl*C gene is proposed to function positively for L-NHase expression and *nhl*D as a negative regulator (Komeda *et al.* 1996).

It was shown that in the presence of an amide, the *nhl*C gene inhibited the repression of the *nhl*D genes, allowing for L-NHase expression. The absence of an amide inhibited the *nhl*C gene allowing the repression of L-NHase production by the activation of the *nhl*D

gene. The amdA and *nhl*A genes are co-transcribed. The *nhl*A and *nhl*B have shown 50% and 36% protein homology to the products of *nhh*A and *nhh*B genes of the H-NHase gene cluster (Komeda *et al.* 1996). There is little known on the function of the *nhl*F gene however it is believed to be associated with the effective expression of the L-NHase gene. *Nhl*E has been identified as a self-subunit swapping chaperone which facilitates the incorporation of cobalt into the NHase (Zhou *et al.* 2010) with recombinant NHase activity dependant on its expression.

The NHase/amidase gene cluster of *Rhodococcus erythropolis* AJ270

Rhodococcus erythropolis AJ270 was isolated from industrially contaminated soil samples (Blakey *et al.* 1995). The broad substrate spectrum and the ability of the NHase to hydrolyse a number of nitriles (aromatic aliphatic and heterocyclic) irrespective of the electronic and steric effects of the substituents has led to its extensive study (Meth-Cohn and Wang 1997).

R. erythropolis AJ270 contains a single Fe-type NHase with α- and β- subunits measuring approximately 23 kDa and an amidase which were induced with the use of acetonitrile or acetamide substrates (O'Mahony *et al.* 2005). The NHase of strain AJ270 has exhibits low enantioselectivity while the amidase exhibits high enantioselectivity. Whole cells of AJ270 were used to catalyse the biotransformation of 1-arylaziridine-2-carbonitriles, 3-arylpent-4-enenitriles and 1-benzyl-2-methylazetidine-2-carbonitrile (Leng *et al.* 2009, Wang *et al.* 2007, Gao *et al.* 2006). The NHase from *Rhodococcus* sp. AJ270 is not photosensitive, despite many Fe-type NHases exhibiting this characteristic (O'Mahony *et al.* 2005).

The *R. erythropolis* AJ270 gene arrangement is similar to *R. erythropolis* JCM6823, *Rhodococcus* sp. R312, *Rhodococcus* sp. N-774 and *Bacillus pallidus* RAPc8B in that the amidase gene precedes the NHase gene. However, it also differed to that of *R. rhodochrous* J1 (L-NHase cluster) and *B. pallidus* RAPc8B in that the α- subunit of the NHase gene preceded the β- subunit (Figure 1.14).
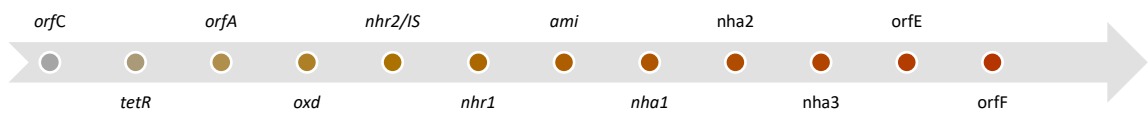
Figure 1.14 The NHase/amidase gene cluster of *R. erythropolis* AJ270 as described by Coffey (2007).

*Orf*C (putative Acyl-CoA synthetase), *tet*R (transcriptional regulator), *orf*A (regulatory protein), *oxd* (aldoxime dehydratase), *nhr*2/IS (interrupted NHase 2 gene with *IS1166*), *nhr*1 (putative amidase regulatory protein), *ami* (amidase), *nha*1 (α- subunit NHase), *nha*2 (β-subunit NHase), *nha*3 (protein required for functional expression of NHase), *orf*E (unpredicted protein) and *orf*F (putative amino acid transporter protein).

The NHase and amidase genes of AJ270 display 100% sequence identity to the corresponding genes in *R. erythropolis* N771. *Orf*E and *nha*3 from AJ270 display a 93% DNA sequence homology to the corresponding gene found in *Rhodococcus* sp. N771 however the function of this gene is not distinguished. The NHase and amidase of AJ270 are inducible with different repression/induction patterns (O'Mahony *et al.* 2005). The 'IS' represents an insertion element *IS1166* which has disrupted the *nhr*2 gene (O'Mahony *et al.* 2005).

The gene cluster of AJ270 (Figure 1.14) displayed 96% sequence homology to the *nhr*1 and *ami* genes of *Microbacterium* sp. AJ115 and *Rhodococcus erythropolis* AJ300 while displaying 94% sequence homology to the *nha*1 and *nha*2 genes of the same strains. In contrast, the gene cluster exhibits low sequence homology to *P. chlororaphis* B23 and *Pseudomonas* sp. K9 (Coffey 2007).

The NHase/amidase gene cluster of *B. pallidus* RAPc8B

The co-type NHase and amidase gene cluster of *B. pallidus* RAPc8B was cloned and sequenced by Cameron *et al.* (2005) (Figure 1.15). The gene cluster arrangement was similar to *R. erythropolis* JCM6823, *Rhodococcus* sp. R312 and *Rhodococcus* sp. N-774 in that the amidase was preceding the NHase gene. However, the β subunit preceded the α- subunit which was the opposite to that of *Rhodococcus* sp. N-774, R312 and *R.*

erythropolis JCM6823, but similar to the NHase subunit gene arrangement of. *R. rhodochrous* J1.
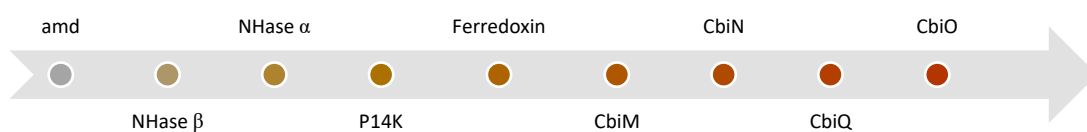
Figure 1.15 The NHase/amidase gene cluster of *B. pallidus* RAPc8B as described by Cameron *et al.*, (2005)

Amd (Amidase), NHase β (NHase β subunit), NHase α (NHase α subunit), P14K (accessory protein), CbiM-CbiQ-CbiN-CbiO (cobalt transport proteins).

B. pallidus RAPc8B was also shown to have a 122 amino acid accessory protein designated P14K which was shown to increase folding of the α subunit prior to associating with the β-subunit of the NHase gene (Cameron *et al*. 2005). Interestingly the NHase gene could be expressed independently of the P14K protein however co-expression of the P14K with the NHase gene did see an increase in the specific activity.

1.8 Aldoximes

Aldoximes have a general structure R-CHNOH (Figure 1.16). They are derived from amino acids and are thought to be intermediates in plant biosynthesis of nitriles towards the pathway involving indoleacetic acid (IAA), CNGs and glucosinolates in plants (Kobayashi and Shimizu 2000).
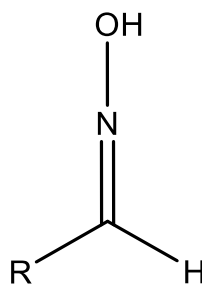
Figure 1.16 General aldoxime structure

Aldoxime dehydratase enzymes can be utilized for the conversion of aldoximes to nitriles (Figure 1.17 and Figure 1.18) potentially providing an economical route to the synthesis of nitriles of interest for the biopharmaceutical/chemical industry. Previous attempts to express the recombinant phenylacetaldoxime dehydratase of *Bacillus* sp. OxB-1 in its active form have proved difficult due to inclusion body formation, however the overexpression of OxdA from *Pseudomonas chlororaphis* B23 in *E. coli* was successful (Oinuma *et al.* 2003).
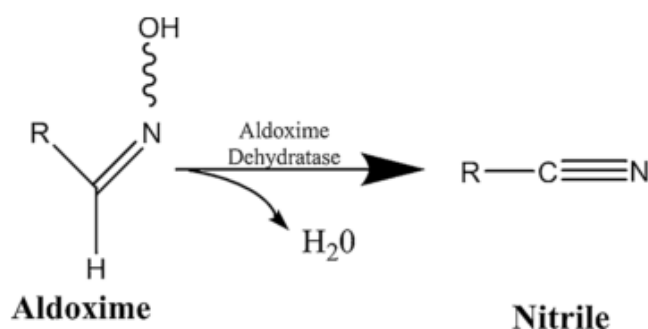
Figure 1.17 Conversion of aldoxime to nitrile mediated by an aldoxime dehydratase

1.9   Aldoxime Dehydratase (4.99.1.5)

Aldoxime degrading enzymes have been extensively studied over the past few years as another route for nitrile production (Figure 1.18). The chemical method for the dehydration of aldoximes to nitriles requires harsh conditions, however the opposite can be said for the biotransformation of aldoximes to nitriles which is carried out at neutral pH with the expectation of a pure product (Asano 2002). Many bacteria, fungi, and some yeasts have demonstrated phenylacetaldoxime-degrading abilities (Kato *et al.* 2000). It has previously been shown that the aldoxime and nitrile metabolising genes coexist due to a shared metabolic pathway of aldoxime metabolism.
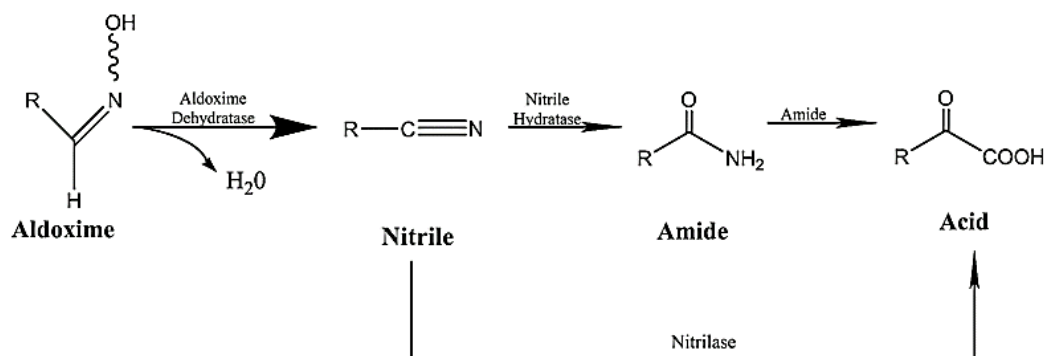
Figure 1.18 Schematic for the biocatalytic conversion of an aldoxime to nitrile via aldoxime dehydratase, with subsequent conversion to carboxylic acid via the aldoxime-nitrile metabolising pathway.

1.9.1   Bacterial sources of aldoxime degrading enzymes

Phenylacetaldoxime (PAOx) and pyridine-3-aldoxime (PyOx) have previously been utilized as model compounds of alkylaldoxime and arylaldoxime respectively, to allow for the determination of the aldoxime degrading abilities in bacteria (Kato *et al.* 2000). Numerous *Rhodococcus* spp. demonstrated aldoxime degrading abilities, with a high distribution frequency detected across the *Rhodococcus* genus (Table 1.2). Interestingly, nitrile-hydrolysing activity on either phenylacetonitrile (PAN) or 3-cyanopyridine (CyPy) was observed among the aldoxime-degrading isolates (Kato *et al.* 2000). A study by Kato *et al.* (2000) where bacteria, fungi and yeasts comprised the 975 microorganisms evaluated for aldoxime degrading activity, resulted in 98 confirmed PAOx degraders and 107 shown to degrade PyOx, conveying the wide spread distribution of the phenylacetaldoxime degrading genes (Kato *et al.* 2000).

Table 1.2 *Rhodococcus* sp. and *R. erythropolis* strains which exhibit aldoxime and nitrile degrading abilities on a number of substrates (Kato *et al.* 2000).

Bacteria	PAOx	PAN	PyOx	CyPy
Rhodococcus erythropolis BG-13	+	+	+	+
Rhodococcus erythropolis BG-16	+	+	-	-
Rhodococcus sp. YH3-3	+	+	+	+
R. rhodochrous J1	+	+	+	+

PAOx= phenylacetaldoxime, PAN= phenylacetonitrile, PyOx= E-pyridine3-aldoxime and CyPy= Cyanopyridine

1.9.2 Co-induction of aldoxime and nitrile degrading enzymes

Co-induction of aldoxime and nitrile metabolising enzymes within a microorganism has been observed with *Bacillus* sp. OxB-1. It was one of the first bacteria which was shown to exhibit both aldoxime and nitrile degrading abilities. The bacteria demonstrated the ability to catalyse the conversion of Z-phenylacetaldoxime (Z-PAOx) to phenylacetonitrile (PAN) via phenylacetaldoxime dehydratase (Kato *et al.*, 2000; Xie *et al.* 2001). A study carried out by Kato *et al.* (2000) assessed a number of microorganisms for their ability to metabolise Z-phenylacetaldoxime to phenylacetonitrile. A number of isolates displayed nitrile degrading abilities on the corresponding phenylacetonitrile which was produced. The majority of biotransformation assays resulted in the production of an amide intermediate which was indicative of a nitrile hydratase amidase system.

However, some cases saw no amide production, it was proposed that this may be due to the activity of a novel nitrilase which demonstrates amidase activity, such as the nitrilase in *Rhodococcus rhodochrous* J1 (Kobayashi *et al.* 1998) or perhaps the activity of a nitrile hydratase/amidase system where the activity of the amidase is faster than that of the nitrile hydratase (Kato *et al.* 2000).

Likewise*, Rhodococcus* sp. YH3-3 utilized 3-pyridine-3-aldoxime as a sole nitrogen source, by carrying out the dehydration of 3-pyridine-aldoxime to 3-cyanopyridine, which was then converted to nicotinamide by a NHase. Following that, an amidase converted

the nicotinamide to nicotinic acid (Kato *et al.* 1999). The aldoxime dehydratase exhibited a wide substrate range including many aryl- and alkyl-aldoximes while the nitrilase, NHase (Cobalt-type) and amidase of the strain also exhibited a wide substrate spectrum (Kato *et al.* 1999).

1.9.3   Aldoxime and nitrile degrading genes

1.9.3.1   Phenylacetaldoxime dehydratase (*oxd*) and nitrilase

An *orf* 1 upstream of the *oxd* gene was determined to encode a nitrilase in *Bacillus* sp. OxB-1. The nitrilase displayed a high homology to the predicted protein sequence of the nitrilase of *R. rhodochrous* J1. The *orf* 2 found upstream of *orf* 1 was thought to encode a transcriptional regulator for the nitrilase of *orf* 1 (Figure 1.19). Likewise, *oxd* genes have also been identified in *Pseudomonas syringae* DC300 and B728a adjacent to the nitrilase gene (Kato and Asano 2006, Kato *et al.* 2000). Further research presented by Podar *et al.* (2005) described a nitrilase in *Pseudomonas syringae* (ZP_00124692) which appears to be co-transcribed with a phenylacetaldoxime dehydratase.

Figure 1.19 Schematic of the genes present in the 'aldoxime-nitrile' pathway *Bacillus* sp. OxB-1, adapted from Kato *et al.* 2000.
*Orf*2 (nitrilase transcriptional regulator), *orf*1 (nitrilase), *oxd* (phenylacetaldoxime dehydratase).

Recent research published by Rädisch *et al.* (2018) cloned and characterised a phenylacetaldoxime enzyme from *Bradyrhizobium* sp. which appears to be part of a putative Oxd-Nitrilase operon. An experimentally confirmed nitrilase was found upstream of the OxdBr1 gene. However, upon further analysis the experimentally confirmed nitrilase (GenBank: AAR97505.1) displayed a high sequence homology to an "amidohydrolase" (94%, 100% coverage) which is indicative, according to Rädisch *et al.* (2018), that the protein may have been incorrectly termed "amidohydrolase".

1.9.3.2 Phenylacetaldoxime dehydratase (oxd) and NHase/amidase

A gene encoding aldoxime dehydratase was located upstream of the amidase gene of the NHase gene cluster in *Pseudomonas chlororaphis* B23. The *Oxd*A enzyme was more efficient towards aliphatic substrates when compared to aromatic substrates. This gene cluster in *P. chlororaphis* B23 was the first such cluster described (Oinuma *et al.* 2003). The acyl-CoA synthetase proteins are denoted by '*acs*A' and where 'nhpC' denotes a regulatory protein. The 'nhpS' portion of the gene cluster has no assigned function (Figure 1.20) (Oinuma *et al.* 2003).
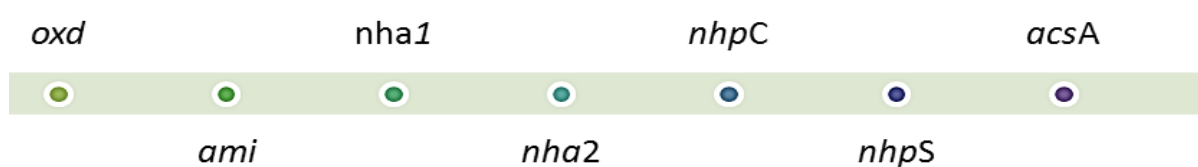
Figure 1.20 Schematic of the Phenylacetaldoxime dehydratase gene cluster for *Pseudmonas chlororaphis* B23 adapted from Oinuma *et al.* (2003).
Oxd (phenylacetaldoxime dehydratase), *ami* (amidase), *nha*1 (α- subunit NHase), *nha*2 (β-subunit NHase), *nhp*C (NHase activator like protein), *nhp*S (unknown protein), *acs*A (acetyl co-A synthetase).

Another study involving the cloning and overexpression of an aliphatic aldoxime dehydratase (*Oxd*K) from *Pseudomonas* sp. K9, a known glutaronitrile degrader, exhibited a gene cluster accountable for aldoxime-nitrile metabolism (Kato and Asano 2006).

The gene cluster contained the following genes: *Oxd*, NHase (Fe-type), amidase and acyl-CoA ligase/synthase (Figure 1.21). Regulatory genes were also present alongside genes of an unknown function. The '*orf*j' codes for acyl-CoA synthetase proteins and '*orf*i' has not been assigned a function. The cluster contained genes like that of *Rhodococcus* sp. N-771, with minor arrangement differences. A 90% sequence identity is seen between *Oxd*K and *Oxd*A genes of *Pseudomonas* sp. K9 and *P. chlororaphis* B23 with similarities also seen in the arrangement of genes within the clusters (Kato and Asano 2006) .
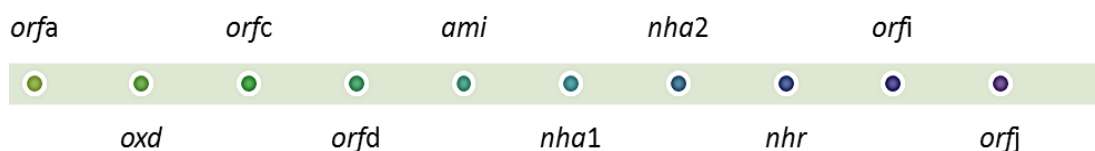
Figure 1.21 Schematic for the 'aldoxime-nitrile' pathway of *Pseudomonas* sp. K9 adapted from (Kato and Asano (2006).

*Orf*A (putative regulatory protein), *oxd* (aldoxime dehydratase), orfc (unknown function), *orf*d (unknown function), *ami* (amidase), *nha*1 (α- subunit of NHase), *nha*2 (β-subunit of NHase), *nhr* (NHase activator), *orf*i (no assigned function) and *orf*j (Acyl Co-A ligase/synthetase).

As described earlier, *R. erythropolis* AJ270 was shown to contain a NHase/amidase gene cluster with the presence of an *oxd* gene identified by Coffey (2007). A segment of the *R. erythropolis* AJ270 NHase/amidase gene cluster which was previously described in full can be seen in Figure 1.22.
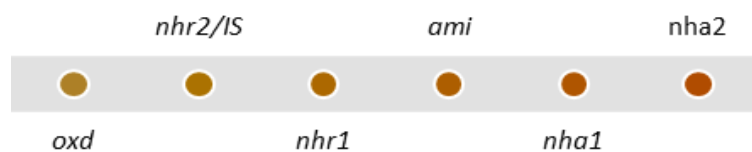
Figure 1.22 A segment of the full NHase/amidase gene cluster from *R. erythropolis* AJ270. The gene cluster shows the aldoxime dehydratase (*Oxd*) gene upstream of *nhr*2/IS (interrupted NHase 2 gene with *IS1166*), *nhr*1 (putative amidase regulatory protein), *ami* (amidase), *nha*1 (α-subunit NHase), *nha*2 (β-subunit NHase).

1.9.3.3 Acyl-CoA synthetase involved in aldoxime/nitrile metabolism

Two open reading frames *nhp*S and *acs*A were identified immediately downstream of the NHase gene cluster of *Pseudomonas chlororaphis* B23 (Figure 1.20). The 'acsA' gene product was expressed in *E. coli* and demonstrated acyl-CoA synthetase activity toward butyric acid and CoA as substrates. This led to the synthesis of butyryl-CoA. The substrate specificity was similar to the aldoxime dehydratase, NHase and amidase of *P. chlororaphis* B23 (Hashimoto *et al.* 2005).

P. chlororaphis B23 exhibited aldoxime dehydratase, nitrile dehydratase and amidase activities, as well as acyl co-A synthetase activity when butyraldoxime was used as the sole nitrogen and carbon source in the media. Hashimoto *et al.* (2005) proposed the acyl-CoA would be further used as an energy sources via the β-oxidation pathway. An open reading frame showing predicted protein homology to acyl-CoA synthetase has been identified in *Pseudomonas* sp. K9 (Figure 1.21); however, the function has not yet been determined. The 'oxd' gene of *Pseudomonas* sp. K9 displayed 76% sequence homology to *Rhodococcus globerulus ('oxdRG')* whereas the 'oxd' gene of *Pseudomonas chlororaphis* B23 displayed 77% sequence homology to *R. globerulus oxd*RG (Kato and Asano, 2006), indicating that there is moderate homology between the aldoxime dehydratase genes.

1.10 Applications of enzymes from the aldoxime-nitrile metabolising pathway

Nitrilases, NHases and amidases have been used in tri-enzymatic systems to degrade effluent from industrial processes. These enzymatic systems have been employed to degrade the following: acrylonitrile, acrylamide, acrylic acid, acrolein, cyanopyridine, fumaronitrile, succinonitrile and maleimide which are produced during acrylamide production (Nomura *et al.* 2013, Kato *et al.* 2000).

1.10.1 Bioremediation

The overuse of nitrile herbicides and pesticides has led to their abundant occurrence in the environment. These herbicides and pesticides and their corresponding metabolites are widely distributed in our environment and pose a threat to ecological systems making them a focal point for bioremediation (Kobayashi and Shimizu 2000, Banerjee *et al.* 2002). NHases are present in the environment at high levels due to their ability to degrade benzonitriles to benzamides (Holtze *et al.* 2008). The metabolites produced by the degradation of these mother compounds such as herbicides or pesticides are often more prevalent in the environment as they can be less susceptible to degradation. The degradation of dichlobenil (2,6-dichlorobenzonitrile) leads to formation of the metabolite 2,6-dichlorobenzamide (BAM) (Holtze *et al.* 2008). *R. erythropolis* AJ270 exhibited the ability to degrade dichlobenil to BAM, however it did not exhibit the ability to degrade BAM to its corresponding carboxylic acid (Holtze *et al.* 2008).

Nitrilases can also be useful for bioremediation. Nitriles are abundant in the environment due to biological and industrial practices, to be discussed in this section. The degradation of nitriles is highly important due to their toxic nature. Some toxic nitriles include acetonitrile, acrylonitrile, succinonitrile, benzonitrile and dichlobenil, many of which are employed in pharmaceutical/industrial processes, plastics, herbicides and pesticides (Gong *et al.* 2012). Soil bacteria such as *R. rhodochrous* PA-34 and *Rhodococcus* sp. NDB 1165 have exhibited excellent specificity towards benzonitrile, indicating that environmental isolates or in fact the nitrile metabolising genes responsible for the observed activity could prove useful for bioremediation of grounds/water contaminated with this nitrile (Vesela *et al.* 2010).

In April 2011, the European Commission (EC) implemented a ban on the use of herbicides and pesticides containing dichlobenil due to the extremely toxic nature of the metabolite 2,6-dichlorobenzamide (European Commission, 2011). As previously noted, some environmental isolates had developed/shown the ability to degrade the nitrile to its corresponding amide; in this case, it proved problematic when the resulting compound from the biocatalytic activity of the isolates resulted in the product 2,6-dichlorobenamide. This then resulted in the need for an amidase which is active on the amide to allow for the conversion of the amide to the corresponding, less harmful carboxylic acid. A recent study published by Pei *et al.* (2017) detailed the co-expression of a hydratase and amidase system which achieved the conversion of 2,6-dichlobenil and ioxynil to their less toxic carboxylic acids.

1.10.2 Surface modification of polymers

Nitrilases and NHases have also been employed in the surface modification of polymers. Polymerisation of acrylonitrile results in polyacrylonitrile (PAN) which is produced at approximately 2.73 million tons per annum worldwide, most of which are used in the textile processing industry (Gong *et al.* 2012). The use of enzymes such as NHase to alter the surface of these polymers is a preferred alternative to that of the harsh conditions required by chemical methods. NHases from microorganisms such as *B. imperial* CBS 49874 and *Corynebacterium nitrilophilus* ATCC 21419 have been used to increase the hydrophilic nature of the textiles to improve dye uptake (Battistel *et al.* 2001). The use of enzymes to modify the surface of these polymers superseded the chemical methods resulting in the improvedability of textiles to uptake dye while maintaining the stability of the polymer.

1.10.3 Industrial and biopharmaceutical compounds

Biotransformations with nitrile and amide metabolising enzymes have become a focal point for biopharmaceuticals. An *Acinetobacter* sp. Strain AK226 was used in the biotransformation of S-(+)-ibuprofen from racemic 2-(4'-isobutylphenyl) propionitrile (Yamamoto *et al.* 1990). The enzyme displayed high enantioselectivity producing S-(+)-ibuprofen. An immobilised and soluble NHase from *Rhodococcus erythropolis* A4 has been used for the biotransformation of nitriles such as 3-oxonitriles, 3-hydroxy-2-

methylenenitriles, 4-hydroxy-2-methylenenitriles ,3-hydroxynitriles and 3-acyloxynitrile into amides.

Iminodiacetic acid (IDA) is widely used for the production of glyphosate herbicides, electroplating solutions, chelating resin, surfactants, and anticancer drugs. A mutant of *Alcaligenes faecalis* isolated by Zhang *et al.* (2013) exhibited excellent catalytic activity for the conversion of iminodiacetonitrile to iminodiacetic acid. A culture of *R. rhodochrous* J1 cells displayed NHase activity against isonicotinamide, catalyzing the production of isonicotinic acid which is usedin the industrial production of producing a tuberculostatic drug used to treat tuberculosis (Zhou *et al.* 2005).

A nitrilase produced by *Rhodococcus* sp. NDB 1165 has been used in the production of nicotinic acids employed in pharmaceutical formulations (Velankar *et al.* 2010, Prasad *et al.* 2007). The nitrilase exhibited high enzyme substrate specificity for aromatic and unsaturated aliphatic nitriles, in comparison to saturated aliphatic nitriles (Prasad *et al.* 2007). An amidase produced by *Rhodococcus sp.* AJ270 has exhibited high enantioselectivity for the production of pharmaceutical intermediates (β-aryl-, β-alkyl- and β-hydroxy-α-methylenepropiononic acids) under mild conditions. Similarly, NHases from *R. erythropolis* AJ270 have been used in the production of intermediates involved in chemical synthesis such as 2,2-dichloro-3-phenylcyclopropane amides (Wang and Wu 2003).

An amidase from *Pseudomonas* sp. MCI3434 has displayed high substrate specificities for particular carboxamide compounds which are catalysed into non-proteinogenic amino acids. These non-proteinogenic amino acids are used as precursors for the synthesis of pharmaceutical compounds (Velankar *et al.* 2010, Komeda *et al.* 2004).

The use of these enzymes to produce carboxylic acids has become common place. Free cells of *R. rhodochrous* Jl have been shown to catalyse 3-cyanopyridine directly to nicotinic acid without the formation of an amide (Nagasawa *et al.* 1988). Acrylic acid (2-propionic acid) has an annual estimated output of 4.2 million metric tons worldwide. The derivatives of acrylic acid are commonly used for surface coatings, textiles, adhesives, and paper treatment amongst others. Interestingly, the nitrilase from *R. rhodochrous* J1, which had been previously characterised to display activity only towards aromatic

nitriles, however showed an increased relative activity towards acrylonitrile when induced by ε-caprolactam (Nagasawa *et al.* 1990). The ε-caprolactam induced resting cells displayed 128% relative activity in comparison to the aromatic substrate of benzonitrile (100%), resulting in a nitrilase which could be utilized towards the production of acrylic acid by changing the inducing substrate.

A successful example of biotransformations is the use of NHase for acrylamide production. According to Yamada and Kobayashi (1996) there are many problems associated with the chemical synthesis of acrylamide such as: polymerisation of both the substrate and product, the formation of by-products such as nitrylotrispropionamide and ethylene cyanohydrin, the rate of acrylic acid formation supersedes the rate of acrylamide formation. Whole cells of *P. chlororaphis* B23 and *Rhodococcus* sp. N774 were used as first and second-generation biocatalysts. Then *R. rhodochrous* J1 was discovered which is now used to produce over 30,000 tons of acrylamide per annum (Brady *et al.* 2004). The Lonza AG company utilize the L-NHase of *R. rhodochrous* J1 to produce nicotinamide from 3-cyanopyridine (Brady *et al.* 2004, Banerjee *et al.* 2002).

Most of the commercially available glycolic acids are at present produced using high-pressure - high-temperature reactions of formaldehyde and carbon monoxide under acid catalysis. These acids are used for a range of products from household cleaners to industrial elements and cosmetics, rendering them a valuable commodity (Yunhai *et al.* 2006). Cross-linked carrageenan immobilized cells of *Alcaligenes* sp. have been successfully employed in the enzymatic transformation of glycolonitrile to glycolic acid (He *et al.* 2010). The use of whole cells for the biocatalysis of these acids supports the move towards a green chemistry synthesis route by reducing waste (limited need for solvents), providing a reusable biological catalyst and reducing or eliminating by-product formation.

Another application of enzymes is the production of enantiopure products. Enantiopure acids and amides have been produced by whole cell catalysts of *R. erythropolis* AJ270. The whole cells of AJ270 were used to catalyse biotransformation of 1-arylaziridine-2-carbonitriles, 3-arylpent-4-enenitriles and 1-benzyl-2-methylazetidine-2-carbonitrile (Leng *et al.* 2009). The amidase of *R. erythropolis* AJ270 has been shown to exhibit high enantioselectivity unlike the associated NHase (Gao *et al.* 2006).

The reliance of the bio-pharmaceutical industry on enantiopure chiral building blocks is well documented, with examples such as the enzymatic hydrolysis of 3-hydroxyglutaronitrile into (R)-4-cyano-3-hydroxybutyric acid. This hydrolysis reaction was employed in the chemoenzymatic synthesis of Atorvastatin, better known as Lipitor. This enabled the production of a high value product from the low-cost chemical epichlorohydrin (Bergeron *et al.* 2006). Another example of the use of biocatalysis is the bioconversion of (R)-*o*-chloromandelonitrile to the corresponding acid (R)-*o*-Chloromandelic acid, a precursor to the drug Clopidogrel® a platelet aggregation inhibitor (Wang *et al.* 2013). According to the European Union (2007) the products from the biotechnology sector across Europe will account for approximately one third of the share of industrial production, with an estimated value of approximately €300 bn by 2030. The predicted market share further infers the importance of biocatalysis and the need to discover novel enzymes with catalytic abilities required by the biopharmaceutical industry.

1.11 Horizontal Gene Transfer

Horizontal gene transfer (HGT) can be defined as the movement of genetic material from one organism to another, in which the recipient is not the offspring of the former. It is an important mechanism thought to have shaped many genomes (Callier 2019, Soucy *et al.* 2015). Whole genome analysis has highlighted the horizontal gene flow between species and genera. In bacteria, gene transfer can take place by one of three methods: natural transformation, transduction, or conjugation (Davison 1999). It is suggested that many genera/ species which possess two nitrilase genes acquired them through horizontal gene transfer or ancient gene duplication as discussed in Section 1.5.4 (Podar *et al.* 2005). HGT has been seen to occur in Archaea, Bacteria and Eukarya.

1.11.1 Horizontal gene transfer in bacteria

The three mechanisms of HGT; natural transformation, conjugation and transduction have proved successful vehicles for microbial evolution. The uptake of naked DNA by a bacterial cell and the heritable incorporation of the genetic information is known as natural transformation. Natural transformation is considered to be a mechanism for the

horizontal transfer of genes in natural bacterial populations (Lorenz and Wackernagel 1994). The DNA may be extracellular, plasmid or chromosomal. Each of the three mechanisms for gene transfer have specific requirements suggesting limitations to their occurrence in natural habitats. A study published by Coffey *et al.* (2010) details the evidence for HGT observed by the detection of a number of Fe-type NHase genes in various environmental isolates from geographically diverse locations.. Further research by the group also presented evidence for HGT when a number of highly similar nitrilase genes were detected among a number of *Burkholderia* sp. isolates which were isolated from different environmental sampling locations (Coffey *et al.* 2009).

1.11.2 Natural Transformation

Transformation involves the uptake of naked DNA and its incorporation into the genome of another microorganism. *Bacillus* sp. in particular, *Bacillus licheniformis* and *Bacillus subtilis* have been well documented as displaying natural transformation abilities (Lorenz and Wackernagel 1994, Goldberg *et al.* 1966,). Some bacteria become naturally competent at some stage through-out their life cycle (*Streptococcus pneumoniae*) (Davison 1999).

1.11.3 Conjugation

The majority of HGT involves conjugation (Top *et al.* 2015). It is the most demanding method of HGT regarding its specific requirements as the donor cell must contain a conjugative element such as the plasmid or transposon. The recipient and donor cells must establish sufficient physical contact to allow for successful and stable transfer of DNA (Harrison and Brockhurst 2012, Norman *et al.* 2009) . There are several types of conjugation which allow for the exchange of genetic information. A self-transmissible plasmid contains sufficient intracellular mechanisms to account for its own transfer plasmid, such as the RP4 of *E. coli* (Soucy *et al.* 2015, Lorenz and Wackernagel 1994, Davison 1999) .

Alternatively, two circular plasmids may fuse together and co-integrate. The non-mobilizable plasmid gets co-integrated with the fused self-transmissible partner plasmid. An example is the mobilization of the IncQ plasmid RSF1010 by conjugative IncP1 plasmids such as RP4 of *E. coli* (Davison 1999). This accounts for the transfer of large

DNA fragments and HGT between closely and distantly related bacteria (Salyers *et al.* 1995).

1.11.4 Transduction

The method of transduction involves genes being encapsulated in bacteriophage particles which are then transferred to another bacterium. Transduction can either be generalised, where the genes are transferred to the other bacterium using a generalised bacteriophage such as 'Coliphage P1', or specialised where a specific bacteriophage is used such as 'coliphage λ'. The specialised transduction method means that only genes located near the site of prophage integration are transferred (Lorenz and Wackernagel 1994). However, there are restrictions with transduction as some phages have a limited host range, which can be limited to a single bacterial species. As bacteriophages are abundant in the environment, many bacterial species could build up resistance to these phage particles.

1.12 Genomic and pathogenicity islands

Genomic islands (GIs) play an important role in the evolution of many bacteria due to the dissemination of a variety of genes including antibiotic resistance, virulence and catabolic genes leading to the formation of new metabolic pathways. These GIs are discrete DNA segments differing between closely related bacterial strains. These genomic islands are acquired via HGT, some of these are mobile while others are not, or those which once were may no longer be mobile (Juhas *et al.* 2009). Once acquired these GIs combine with the bacterial plasmid or chromosome and so become integrated into the new host.

It is thought that these GIs are strongly selected for the transfer of novel genes to other bacteria for an adaptive and auxiliary function (Gaillard *et al.* 2006). GIs have been shown to carry genes for sucrose metabolism, aromatic compound metabolism, mercury resistance and siderophore synthesis (Larbig *et al.* 2002). The first GIs were detected in uropathogenic *E.coli* harbouring virulence genes (Hacker *et al.* 1990)

A mobile GI's lifecycle would begin with acquisition of the GI by the host via HGT. After this the GI would be integrated into the host chromosome via a site-specific integration

by an integrase enzyme. The GIs then begin to develop with either the loss of genes or acquirement of other mobile genetic elements. The GI would then excise from the hosts chromosome and undergo the transfer to another host (van der Meer *et al.* 2009). The integrase enzyme encoded by the '*int*' gene is most often coded by GIs themselves. There is little or no conservation between the GIs which have been studied. It is thought that the integrase is aided by an excisionase during the excision of the GI from the host chromosome (Ramsay *et al.* 2006, Lesic *et al.* 2004).

The characterisation of GIs is based on several factors such as: absence from closely related strains of the host, absence from the genomes of closely related strains, size (usually between 10 and 200 kb while genomic inlets are between 1 and 10 kb), difference in the codon usage between the GI and the rest of the genome, possession of mobile genetic elements (transposons, IS and integrases or excisionases), differences between the GC content, flanking by direct repeats (may be generated post integration into the host genome), or the presence of blocks of genes which encode a specialised function such as metabolism (Wilson and Nickerson 2006, Collyn *et al.* 2004).

If a virulence function is encoded by GIs, they may also be called pathogenicity islands (PAIs), symbiosis islands, metabolic islands, or resistance islands. PAIs contribute to rapid changes in virulence potential and are known to have contributed to genome evolution by horizontal gene transfer in many bacterial pathogens (Juhas *et al.* 2009). Many PAIs have been well studied and characterised. For example: a metabolic GI has been identified in some *Ralstonia* sp. strains. The GIs contain genes which encode for chlorobenzene and toluene degradation (Müller *et al.* 2003).

Plasmid borne GIs are common as a GI is not limited to localisation on the host's chromosome. It is suggested that the SGI1 of *Salmonella enterica* DT104 may have previously existed as an integrative conjugative plasmid, as indicated by its DNA sequence (Boyd *et al.* 2001). The PAIs of pXO1, the large *Bacillus anthracis* plasmid which harbours the anthrax toxin genes, is another documented example of plasmid borne GIs.

1.13 Insertion Sequences

Insertion sequences act as transposable elements. They are short DNA sequences which are now considered to be the most common transposable element. These genetic elements can move from one position on a chromosome to another position on the same or another chromosome. IS have been known to prevent the expression of a gene by interrupting the coding sequence. ISs are sometimes accompanied by a transcription or translation terminator sequence; this can prevent gene expression if the genes are found downstream of the promoter in the same operon. The ISs, unlike transposons, can only encode functions which are encoded in their mobility (Mahillon and Chandler 1998). When the IS become inserted into the host by recombination it is common for direct repeats to transpire in the flanking DNA of the target site. This is due to an attack on each DNA strand at the target site of the IS by one of the two transposon ends which occurs in a staggered fashion (Mahillon and Chandler 1998). IS sequences are generally found on host chromosomes and large plasmids and Siguier *et al.* (2006) proposed that small plasmids have much lower nonessential genes, and so they generally cannot withstand IS insertion. This hypothesis stemmed from that data which showed that there was a 0% occurrence of an IS in plasmids lower than 20 Kb, with contrasting data showing IS incidence in large plasmids at varying levels (15-40%).

CHAPTER 2

REAL-TIME PCR DETECTION OF NOVEL ALDOXIME DEHYDRATASE GENES IN ENVIRONMENTAL NITRILE METABOLISING ISOLATES

2 REAL-TIME PCR DETECTION OF NOVEL ALDOXIME DEHYDRATASE GENES IN ENVIRONMENTAL NITRILE METABOLISING ISOLATES

2.1 Introduction

As previously described in Section 1.9, aldoxime dehydratase genes are of growing importance to industry to allow for the biotransformation of aldoximes to nitriles, facilitating the move towards the 'green chemistry' approach for nitrile biosynthesis. As discussed in Section 1.9 , nitriles are abundant in nature and in turn, there are multiple applications for the enzymes which form a part of the aldoxime-nitrile pathway, such as aldoxime dehydratase.

According to Fleming *et al.* (2010), there were over 30 nitrile containing pharmaceuticals on the market in 2010 with another 20 at the clinical trial stage. Since 2010, 11 nitrile containing pharmaceuticals have been approved by the FDA (FDA, 2016). A patent has also been filed for a "Method for biocatalytic production of nitriles from oximes and oxime dehydratases usable therein" (Piatesi *et al.* 2013). The patent describes a protocol which allows for the dehydration of aldoximes to nitriles. The dehydration of aldoximes to nitriles via biocatalysis boasts high-yields and enantioselectivity, with H_2O as the sole by-product, advantages which are often associated with biotransformations (Xie *et al.* 2001, Metzner *et al.* 2014). The presence and increase of nitrile containing pharmaceuticals on the market indicates that there will be an increased demand for biocatalytic solutions for their production, via aldoxime dehydratase mediated biosynthesis.

Atorvastatin, commercially known as 'Lipitor' is a statin which as noted earlier was launched in 1997 and is was one the world's largest-grossing drug with 2004 sales of $12 billion and 2006 sales grossing $12.9 billion. A nitrilase mediated biotransformation was employed for the conversion of 3-HGN to the chiral intermediate (R)-ethyl-4-cyano-3-hydroxybutyric acid, a precursor for Atorvastatin as published by Bergeron *et al.* (2006).

Previously published work by the research group presented the development and utilization of high throughput assays for the detection, identification and activity screening of novel nitrilase and NHase genes (Coady *et al.* 2013, Coffey *et al.* 2010, Coffey *et al.* 2009, O'Mahony *et al.* 2005). The work in this chapter presents a high-throughput real-time (qPCR) assay for the detection of novel partial aldoxime dehydratase gene sequences from soil microorganisms.

Utilization of the aldoxime-nitrile metabolising pathway for the production of nitriles and carboxylic acids for industrial purposes has become more prevalent as the microorganisms involved are abundant in nature, with aldoxime-degrading abilities found to be widely distributed in bacteria, fungi, and some yeasts (Kato *et al.* 2000). Aldoxime dehydratase enzymes have been shown to co-exist with nitrile-metabolising enzymes, namely nitrile hydratase and nitrilase (Figure 2.1) (Kato *et al.* 2004, Kato *et al.* 1999, Asano and Kato 1998).
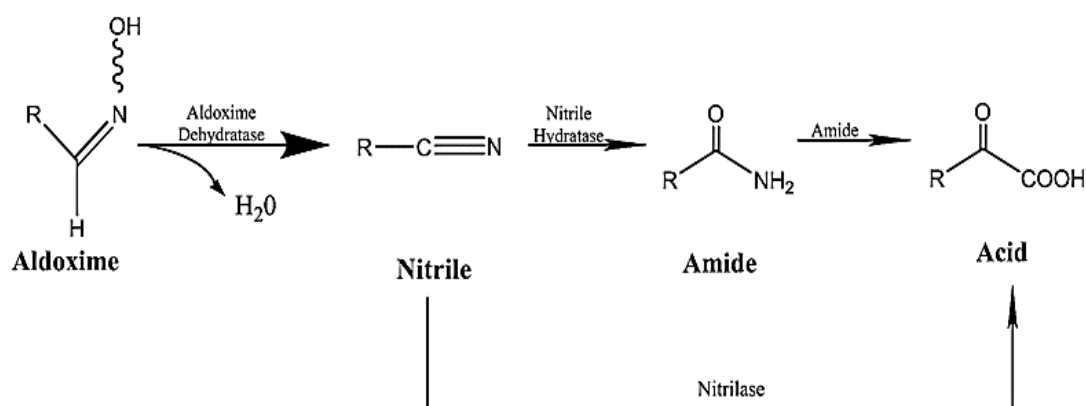
Figure 2.1 Aldoxime-nitrile metabolising pathway in microorganisms.

A study carried out by Kato *et al.* (2000) investigated the distribution of aldoxime dehydratase among a number of microorganisms, with the phenylacetaldoxime dehydratase observed among a number of bacteria, fungi and yeast with a high prevalence observed in the genera *Bacillus, Arthrobacter, Fusarium* and *Rhodococcus*. *Bacillus* sp. OxB-1 was one of the first bacteria to exhibit both aldoxime and nitrile degrading abilities. The bacteria demonstrated the ability to catalyse the conversion of Z-phenylacetaldoxime (Z-PAOx) to phenylacetonitrile (PAN), with further study

highlighting the activity of a nitrilase and amidase on phenylacetamide (PAAm) to form phenylacetate (PAA) (Asano & Kato, 1998).

Metzner *et al.* (2014)) reported the enantioselective dehydration of a racemic aldoxime with enantioselectivity of up to 98 ee% achieved via whole cell biocatalysis with *Bacillus* sp. OxB-1. A gene encoding aldoxime dehydratase (*oxd*) was located upstream of the amidase gene within the nitrile hydratase gene cluster in *Pseudomonas chlororaphis* B23. The gene cluster in *P. chlororaphis* B23 was the first such cluster described (Oinuma *et al.* 2003).

The aim of the body of work presented in this chapter was to expand the WIT microbial biocatalysis library and to design a real-time PCR (qPCR) assay for the high-throughput detection of novel aldoxime dehydratase metabolising genes in a subset of nitrile metabolising isolates across a number of genera. The assay described boasts many advantages over conventional PCR methods. Limitations of end-point PCR such as poor sensitivity and the inability to quantify results can be overcome with real-time PCR as it features high technical sensitivity (< 5 copies), high precision ($< 2\%$ standard deviation) and quantitative ability (Paiva-Cavalcanti *et al*. 2010, Klein 2002, Ginzinger 2002) while allowing for high throughput capabilities.

The detection of nitrilase and nitrile hydratase containing isolates by real-time PCR has been previously described by Coffey *et al.* (2010) and Coffey *et al.* (2009). High sensitivity was observed for the target genes with no false-positive signals observed. Another advantage is that a high throughput screening approach is aided by the 96-well format with the measurement of reaction kinetics from the early stages of the assay, omitting the need for post-PCR processing i.e. agarose gel electrophoresis.

One limitation of real-time PCR analysis utilizing SYBR green dye is that the fluorescence signal measured could be primer dimer, or non-specific amplification; this is since it binds to double stranded DNA and so care must be given during the primer design stage to ensure non-specific amplification is omitted or limited. This limitation could be avoided by using a TaqMan® assayas the signal generated would rely on the detection of a specific product/target via probes, however this would increase cost per run/sample. Additionally, utilization of a specific probe to detect *oxd* genes across a

number of environmental isolates could result in lack of detection due to the highly specific nature of TaqMan® assays and so the design of a SYBR green assay was preferred for this application. Conversely, the addition of a dissociation stage can allow the end user to distinguish between any potential primer dimer, false amplification and positive signals for the gene of interest based on melting temperatures (Tm). This cannot be achieved by end-point PCR which only separates products based on size.

2.2 Research Hypothesis and Aims

Hypothesis:

A nitrile metabolising microorganism containing a gene which codes for nitrile hydrolysis, may also contain a gene(s) which code for aldoxime dehydration.

Aim:

The aim of this research is to design a qPCR assay for the detection of aldoxime dehydratase genes in nitrile metabolising microorganisms with in line with the hypothesis of the co-existance of nitrile and aldoxime metabolising genes.

2.3 Materials and Methods

Over 200 microorganims isolates from environmental soil samples were shown to have the ability to utilise nitriles as a sole Nitrogen source. A number of these environmental nitrile metabolising soil isolates were screened for the detection of aldoxime dehydratase genes via a high throughput real-time PCR assay. A conserved region within the aldoxime dehydratase gene was first identified from the alignment of several aldoxime dehydratase genes from the database, allowing for the design of real-time primers (qPCR) and conventional PCR primers. *Rhodococcus erythropolis* AJ270 has been previously shown to contained an aldoxime dehydratase gene upstream of the NHase and amidase genes (O'Mahony *et al.* 2005) (Figure 2.2) and so was utilized as a positive control.
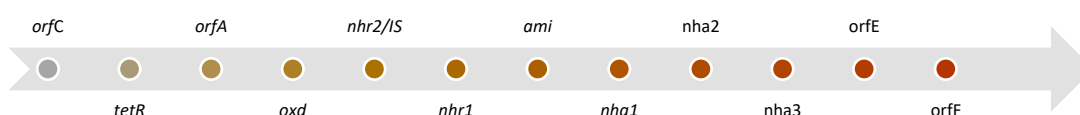
Figure 2.2 An aldoxime dehydratase *oxd* gene present upstream of the NHase (*nha1-nha3*) and amidase (*amd*) genes in *Rhodococcus erythropolis* AJ270.

2.3.1 Soil sampling locations

Soil samples were collected from numerous locations as noted in (Table 2.1). These soil samples were selected as they represented a number of diverse environmental locations. Soil environmental samples were selected based on the premise that these samples would have been exposed to industrial waste from an old tannery industry (soil I.D. 1 to 3), metal content (soil I.D. 6-9) or as unclassified soils (soil I.D. 4-5).

Table 2.1 Soil samples and their corresponding sampling locations

Soil I.D.	Soil location	GPS coordinates
1	River Clodagh, River bed, under bridge, Portlaw, Co. Waterford.	52.85381,-7.315259
2	River Clodagh, River Bank, Portlaw, Co. Waterford.	52.285357,-7.315479
3	River Clodagh, River bed, Curraghmore Estate, Portlaw, Co. Waterford.	52.289885, -7.339611
4	Curraghmore gardens, Curraghmore Estate, Portlaw, Co. Waterford.	52.290742, -7.354373
5	Roadside, Passage East, Co. Waterford.	52.234746, -7.022656
6	Passage East beach near ferry dock, Co. Waterford.	52.240267, -6.974197
7	Passage East, ferry dock, Co. Waterford.	52.23969-, -6973051
8	River Suir, Mount Congreve, under the bridge, Co. Waterford.	52.244914, -7.205009
9	Stream, Mount Congreve, Co. Waterford.	52.244652, -7.205020

2.3.2 Enrichment and isolation of putative nitrile metabolising microorganisms

Soil samples were induced in M9 minimal media (1 g of soil per 10 mL of M9 minimal broth) (Green *et al*. 1989) with modifications (O'Mahony et al. 2005) with a range of nitriles (10mM) as the N source: 3-hydroxyglutaronitrile (3-HGN), 3-hydroxybutyronitrile (3-HBN) and, 3-hydroxy-3-phenylpropionitrile (3-HPPN) Figure 2.3).

A pooled soil approach was also taken with soils 1-9 across the three substrates. To achieve this 0.5 g of each soil was added to a tube and after vigorous mixing (to ensure the soil samples were well dispersed) 1.0 g of mixed soil was taken and added to M9 minimal broth with the substrate of choice. Post incubation at 25 °C for 7 days at 200 rpm, soil supernatants were collected and stored at -70 °C with a final concentration (v/v) of 20% glycerol. A portion of the remaining soil supernatant was utilized for isolation of nitrile metabolising environmental microorganisms.

Pure isolates were obtained from enrichment cultures as per Coffey *et al.* (2010) with the following exception: agar-agar ultrapure (Merck, Darmstadt, Germany) was used for M9 agar. This involved a minimum of three rounds of streaking onto M9 agar with 10 mM of the corresponding substrate which was used during the enrichment stages. The plates were incubated at 25 °C for a maximum of 5 days. A number of pre-existing nitrile metabolising environmental isolates had previously been isolated from a selection of soil sources by Coffey *et al.* (2010) and were also utilized during this study.
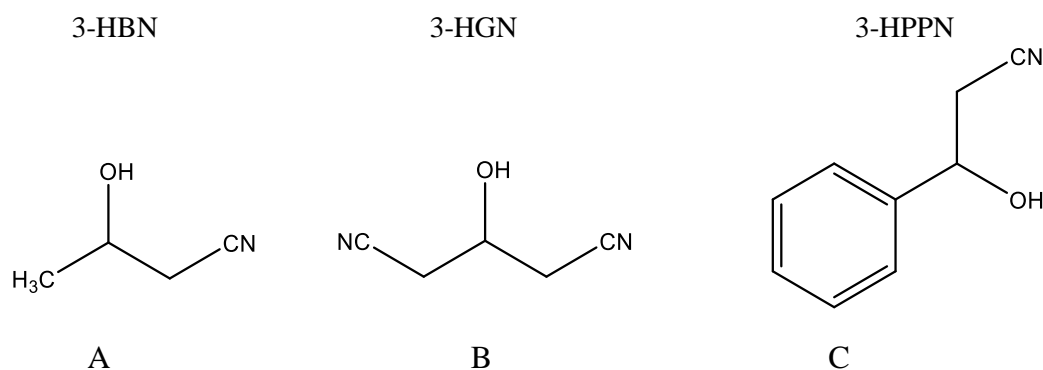
Figure 2.3 Industrially relevant nitriles used for the enrichment for soil microorganisms at a final concentration of 10 mM as the sole nitrogen source. (A) 3-HBN, (B) 3-HGN and, (C) 3-HPPN.

2.3.3 Maintenance of environmental microorganisms

M9 minimal media (recipe in Appendix I), prepared as per Green & Sambrook with modifications (Coffey *et al.* 2010) containing 10 mM nitrile as the sole source of nitrogen was utilized for the maintenance of wild-type strains. The microorganisms were incubated at 25 °C for 2 days. 20% (v/v) nutrient broth (Oxoid™, Cat. No. CM0001B) glycerol stocks were prepared from M9 broth inoculations via centrifugation at 13,000 rpm for 5 minutes. Glycerol stocks were stored at – 70 °C.

2.3.4 Detection of aldoxime dehydratase genes from nitrile metabolising environmental isolates using real-time (qPCR) and endpoint PCR

Firstly, the purified environmental nitrile metabolising isolates were screened via qPCR to see if a positive signal was observed for the presence of an aldoxime dehydratase gene as per Section 0. Following the detection of a positive signal for an aldoxime dehydratase gene via qPCR, endpoint PCR was carried out to amplify the conserved *oxd* region, as described in Section 2.3.5.2. Table 2.2 displays the primer sequences and target amplicons for the primer pairs utilized for both the qPCR and endpoint screening for the presence of a conserved region of the aldoxime dehydratase genes.

Table 2.2 Target amplicons and corresponding primer sequences for both endpoint and qPCR primer pairs.

Primer Name	Sequence 5'-3'	Target Amplicon
OXDCI-F	TACTGGGGGTTCGATGCGCGA	Conserved internal oxd region
OXDCI-R	GTTGCTGTAGCACCCGACGG	Conserved internal oxd region
OXD-SYBR-F	CGGGGACACGACAACATCG	Conserved internal oxd region
OXD-SYBR-R	GGATTTCGTCCAGGTAGAGGC	Conserved internal oxd region

A ClustalW (Thompson *et al*., 1994) alignment of homologous *Rhodococcus* spp., *oxd* gene sequences, was used for the real-time primer design. The target binding sites of these primer sets can be seen in the alignment of the conserved *oxd* region to a number of known nitrile metabolising isolates, Figure 2.4.

2.3.5 16s rRNA PCR amplification

The amplification of the 16s rRNA gene for bacterial species identification was carried out as described by Marchesi *et al.* (1998) using the 63F and 1387R primer pair. The endpoint PCR recipe used was as per Section 2.3.5.2 with the exception of the primer pairs used.
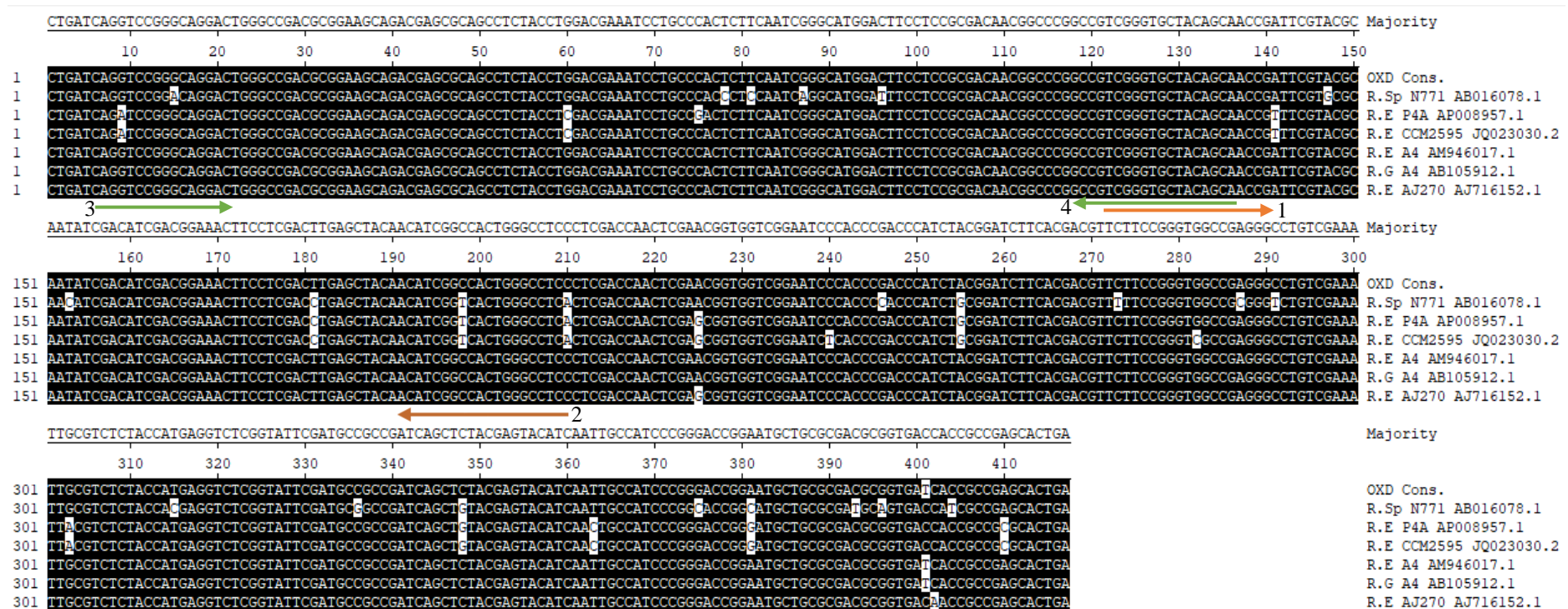
Figure 2.4 Alignment of *Rhodococcus erythropolis* spp. (AJ270 (AJ716152.1) PR4 (AP008957.1), A4 (AM946017.1) and, CCM2595 (JQ023030.2) with *R. globerulus* A4 (AB105912.1) and *R.* sp. N-771 (AB016078.1) with the conserved internal *oxd* region. The green arrows represent the primer binding sites for OXDCI-F[1] and OXDCI-R[2] respectively and with the orange arrows representing the primer binding sites for SYBR-2F[3] and SYBR-2R[4] respectively. Differences in DNA sequences within the conserved internal *oxd* gene are highlighted in white.

2.3.5.1 qPCR detection of aldoxime dehydratase genes

A 7300 real-time PCR system (Applied Biosystems, CA, USA) was used for the detection of aldoxime dehydratase genes. NCBI primer-BLAST designing tool was exploited for the primer design (Ye *et al.* 2012), with primers supplied by Eurofins, MWG operon, Germany. All assays were carried out using SYBR™ green detection method (SYBR Green PCR Master mix (Applied Biosystems, Cat. #4364346) and ROX as the passive reference dye.

To a 15 µl reaction mixture the following was added: 7.5 µl of SYBR™ master mix, 1 µl of primer mix OXD-SYBR-F and OXD-SYBR-R (5 µM each primer, final concentration of 0.33 µM/µL), 5.9 µl of sterile distilled water and 1.0 µl of cell suspension at O.D.600 $_{nm}$ = 0.2, final cell O.D.600$_{nm}$= 0.013). All samples were assayed in triplicate and replicated twice. A dissociation stage (stage 4) was included. The thermocycler conditions are displayed in (Figure 2.5).

A standard curve was created by carrying out serial dilution (10^{-1} to 10^{-7}) of the positive control, *R. erythropolis* AJ270 cells with a starting O.D.600$_{nm}$ = 7.2. Subsequent real-time screening employed the same method as previously described.
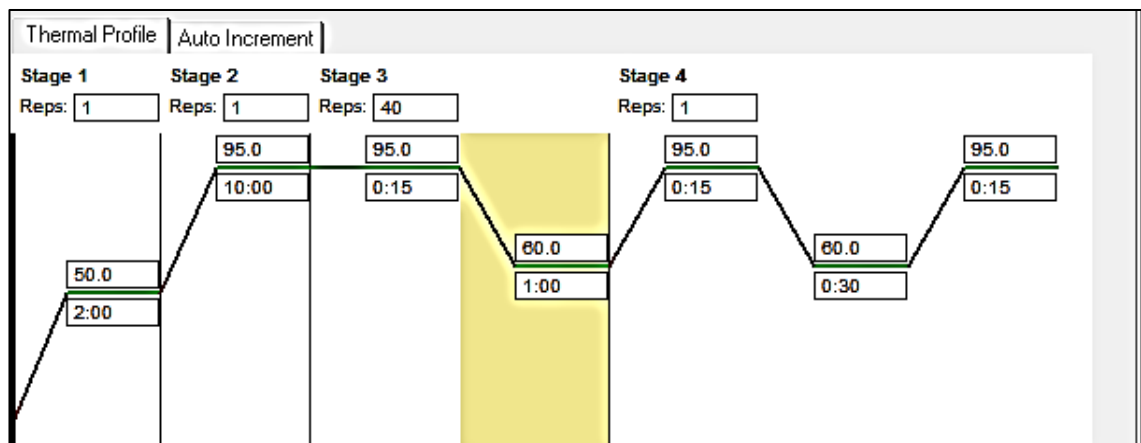
Figure 2.5 Real-time PCR thermal profile, with an additional dissociation stage (stage 4).

2.3.5.2 Endpoint PCR detection of aldoxime dehydratase genes

Endpoint PCR primers were designed using Primer-Blast (Ye *et al.* 2012) from the ClustalW (Thompson *et al.* 1994) alignment of homologous *Rhodococcus* spp., *oxd* gene sequences as seen in Figure 2.4. The primers were synthesised and supplied by MWG Biotech Ltd. or Sigma-Aldrich Ltd. Per 15 µL reaction mixture the assay contained; 50-100 ng DNA, 7.5 µL of 2X GO-Taq green mastermix (Promega, Cat. No. M7501), 1 µL of 25 µM PCR mix (final primer concentration 1.66 µM) and 5.5 µL of sterile water. For colony PCR cells were resuspended in sterile water to an $O.D.600_{nm} = 1.0$, and 1 µL was added to the reaction (final cell $O.D.600_{nm} = 0.06$). An ABI Veriti® thermocycler was used throughout with the conditions detailed in Table 2.3.

The forward primer OXDCI-F and reverse primer OXDCI-R were used to amplify a conserved region of the *oxd* gene from a subset of the environmental *oxd*$^+$ isolates (Figure 2.4). The target conserved internal oxd gene was 233 bp with the full aldoxime dehydratase gene from *Rhodococcus erythropolis* AJ270 measuring 1062 bp (AJ716152.1). In *R. erythropolis* AJ270 the forward primer bound to the conserved aldoxime dehydratase gene region from residues 511 to 531 with the reverse primer binding to residues 764 to 784. The 16S ribosomal RNA was amplified from environmental isolates using primers 63f and 1387r as per Marchesi *et al.* (1998).

Table 2.3 Thermocycler conditions for the endpoint PCR amplification of the conserved internal region of the aldoxime dehydratase genes from environmental nitrile metabolising isolates.

Stage	Temperature (°C)	Time (minutes)	Number of cycles
Initial denaturation	95	5	1
Denaturation	95	1	
Annealing	55	1	25
Extension	70	1	
Final Extension	72	8	1
Hold	4	hold	1

2.3.6 Gel analysis and photography of gels

All PCR products and DNA extractions were subjected to agarose gel electrophoresis for qualitative analysis. Agarose (Cleaver Scientific Cat. No. 9012-36-6) was added to 100 mL of 1X TAE buffer (40 mM Tris, 1 mM EDTA, 20 mM acetic acid, pH 8), to a concentration of 0.8% (w/v) unless otherwise stated. After heating the agarose solution was cooled to ~50 °C. 1 μL/100 mL ethidium bromide (10 mg/ml, Sigma, Cat. No. E-1510) was then added to the molten agarose which was allowed to set in electrophoresis trays.

The gel was placed in the electrophoresis tank containing 1X TAE buffer. 5μl aliquots of DNA samples were mixed with 1μL loading dye (Cleaver Scientific CSL-loadingdye) before being added to the wells of the gel. λ DNA *EcoR1/HinDIII* (NEB Cat No. G1731), λ Monocut mix (Cat No. N3019S) and 100 bp (NEB Cat No. N3231L) DNA size standards were used to allow for the size estimation of DNA. Electrophoresis was then applied at 6 V/cm and 120 mA until adequate migration was attained.

2.3.6.1 Visualisation of gels

All gels were analysed using Syngene Gene Genius Bio-Imaging system under U.V. illlumination.

2.4 Results

2.4.1 Expansion of microbial biocatalysis library

201 environmental putative nitrile metabolising isolates were isolated from soil enrichments. These microorganisms were purified based on their ability to grown on agar containing a nitrile of interest as their sole nitrogen source (Table 2.4). The selective enrichments were carried out with soils and nitriles as they became available, with a pooled soil approach later adapted due to resource limitation.

Table 2.4 Soil samples and the number of putative nitrile metabolising environmental microorganisms isolated with corresponding nitrogen source.

Soil	Nitrogen source (10 mM)	Number of isolates
1	3-HGN	17
2	3-HGN	14
3	3-HGN	18
4	3-HGN	3
	3-HBN	6
5	3-HBN	7
6	3-HBN	7
7	3-HBN	5
8	3-HPPN	5
9	3-HBN	15
1-9	3-HGN	55
1-9	3-HBN	43
1-9	3-HPPN	6
Total Number of Isolates		**201**

2.4.2 Real-time PCR analysis

From the 201 environmental isolates which were purified from various soil samples in Ireland (as detailed in Table 2.4) subjected to the high throughput real-time screening, a positive signal for the conserved region of the oxd gene was observed from 109 isolates, indicating that they contained an aldoxime dehydratase gene.

Figure 2.6 displays the amplification plot obtained from the screening of serially diluted *R. erythropolis* AJ270 positive control cells with the 10^{-7} dilution showing a Ct value within the cut-off limit of 28. From this a standard curve was generated (Figure 2.7).
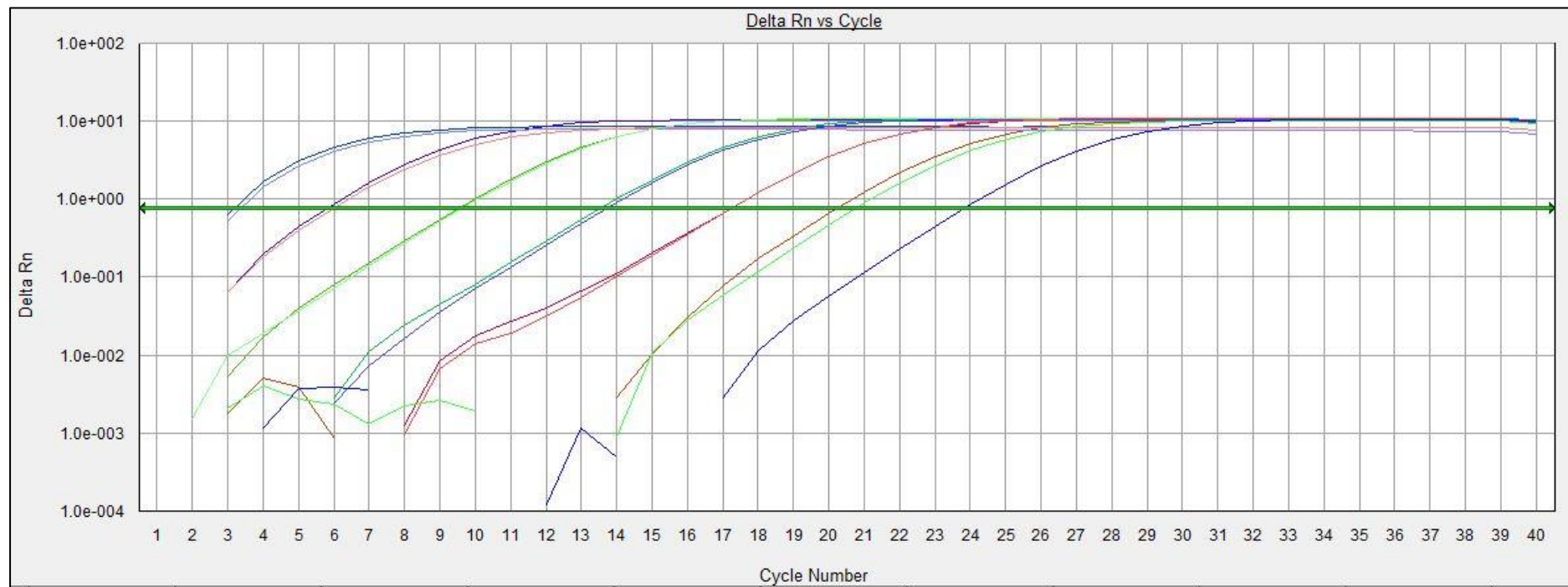
Figure 2.6 Real-time amplification analysis of the serial dilution of *Rhodococcus erythropolis* AJ270 cells, used as a positive control for the estimation of the limit of detection of the conserved region of the aldoxime dehydratase gene.

The standard curve (Figure 2.7) created displayed an R^2 value of 0.997 confirming linear correlation between the concentration of cells (O.D.$_{600nm}$) present and the Ct value obtained during the real-time screening. The equation of the line reads Y= -3.531890X + 27.532110, the slope of the standard curve measures -3.531890 resulting in an amplification efficiency of 91.93 % as determined by the following equation:

$E = -1 + 10^{\frac{-1}{slope}}$ (Agilent genomics, 2018).
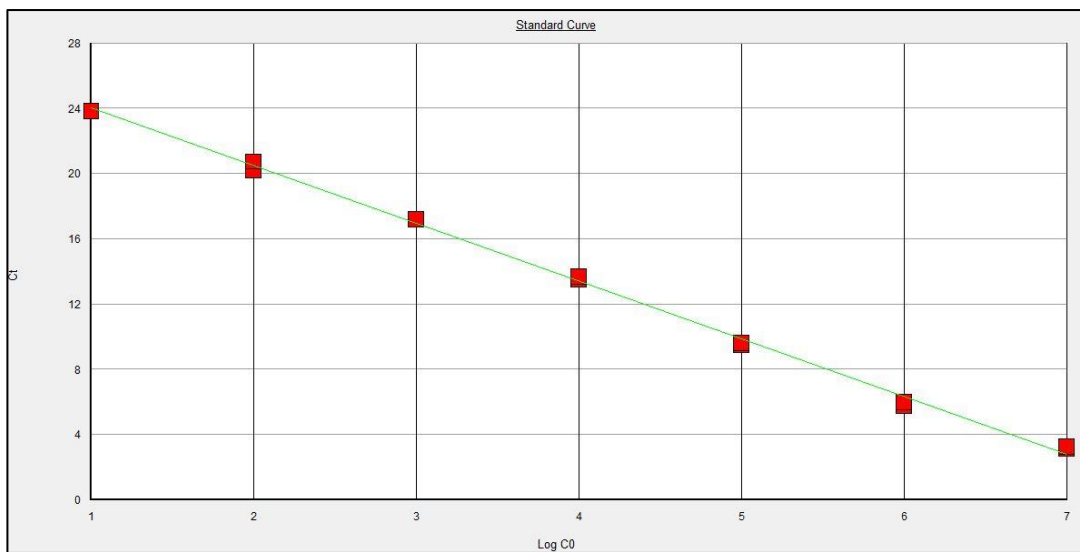
Figure 2.7 Standard curve of *R. erythropolis* AJ270 *oxd* positive DNA with equation of the line Y= -3.531890X + 27.532110 with corresponding R^2 value of 0.9991.

A sample amplification plot which was observed during the real-time PCR screening for an *oxd* gene was is shown in Figure 2.8. Table 2.5 contains the Ct values for all positive *oxd* signals identified through the screening of a sub-set of isolates from the 109.
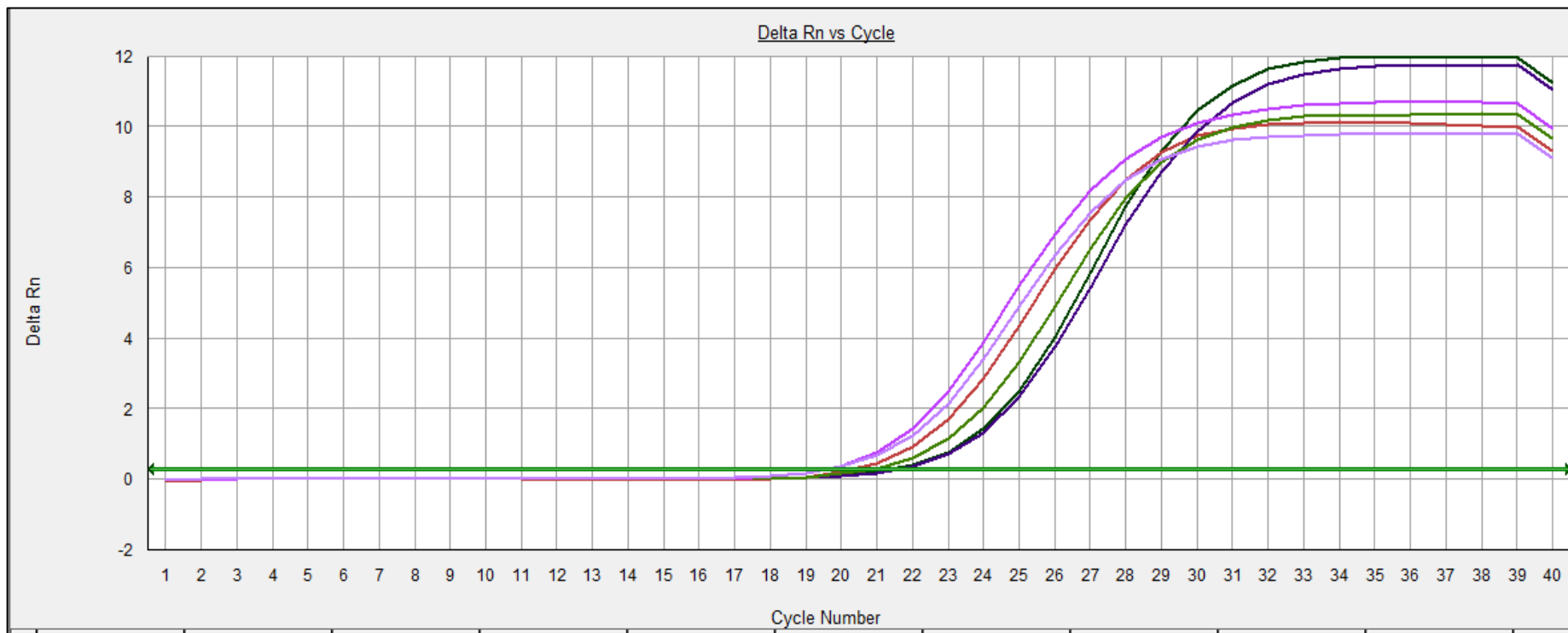
Figure 2.8 Typical amplification plot obtained during the high throughput screening of environmental nitrile metabolising isolates for the detection of the conserved aldoxime dehydratase gene region.

2.4.3 Endpoint PCR analysis

From the 201 putative nitrile metabolising isolates which were purified on 3HGN, 3-HBN and 3-HPPN, 109 displayed a positive signal for an *oxd* gene. An additional 18 nitrile metabolising microorganisms which were previously isolated by L. Coffey (WIT) also displayed a positive signal for an *oxd* gene.

Of the 109 isolates (purified within this batch of work) which displayed a positive *oxd* signal, 12 were selected for endpoint PCR. The subset of 12 isolates were chosen at random for endpoint PCR amplification of the gene and to allow for subsequent sequencing and gene identification. Detection of the conserved region of the aldoxime dehydratase gene (Figure 2.9) was detected from all 12 of the putative nitrile metabolising environmental isolates. Additionally, an *oxd* product was also obtained via endpoint PCR for the 18 pre-existing isolates. All products were confirmed as *oxd* genes by subsequent sequencing (sequences available in Appendix I). Figure 2.9 displays the amplification of the conserved internal region of the *oxd* gene and the full *oxd* gene across multiple genera. One isolate, ST53 did not display the amplification of the target gene despite the Ct and Tm values displayed during the qPCR assay. This may be due to the differences in sequence at the primer binding site or perhaps the leading sequence may have resulted in secondary structures making it difficult for the *Taq* polymerase to reach the target site and to allow for amplification. However, a gradient PCR was performed allowing for the detection of the *oxd* gene indicating that it was differences in primer binding.
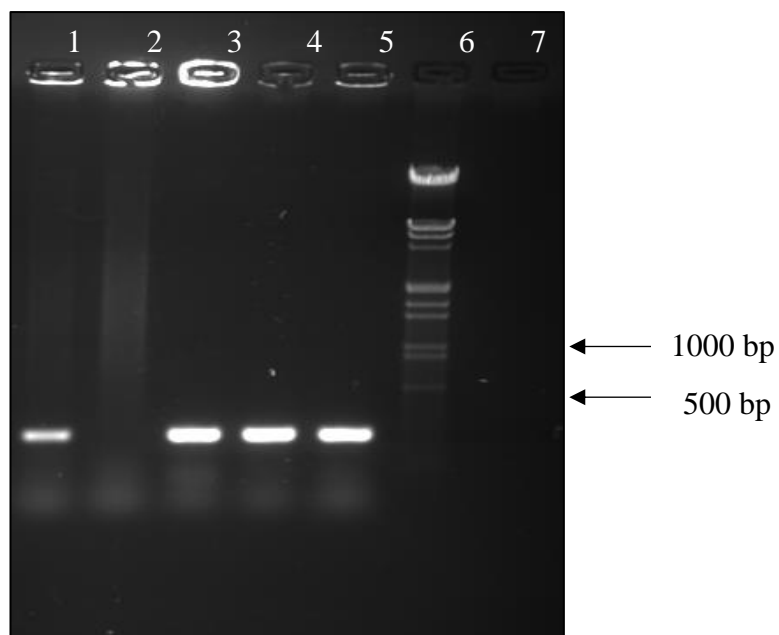
Figure 2.9 Amplification of the *oxd* int gene from putative nitrile metabolising environmental isolates. (Lane 1) ST180 *oxd* int; (lane 2), ST53 *oxd* int; (lane 3), ST111 oxdint; (lane 4), ST125 *oxd* int (lane 5), *R. erythropolis* AJ270, *oxd* int; lane 6, *Eco*R1/*Hin*DIII marker; Lane 7, *oxd* int negative control.

2.4.4   Amplification of 16s rRNA genes

To allow for the identification of the putative nitrile metabolising microorganisms which contained the *oxd* gene, 16s rRNA amplification was carried out across the 12 unidentified microorganisms, Figure 2.10.
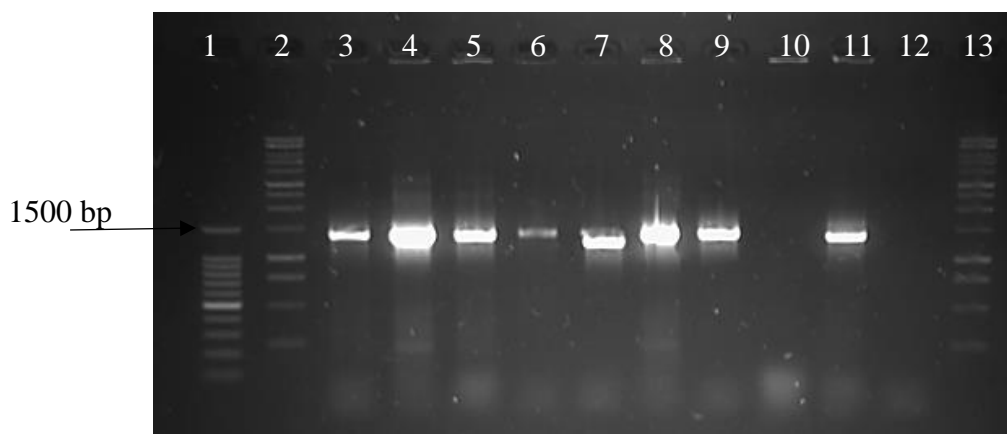
Figure 2.10 Sample electrophoresis gel displaying the amplification products of the 16s rRNA PCR on *oxd*⁺ isolates. Promega 100 bp ladder (lane 1), Promega 1 kB ladder (lane 2), ST144 (lane 3), ST111 (lane 4), ST146 (lane 5), ST196 (lane 6), ST30 (lane 7), ST125 (lane 8), ST119 (lane 9), negative control (lane 10), positive control (*R. erythropolis* AJ270) (lane 11), empty lane (lane 12), Promega 1 kB ladder (lane 13).

2.4.5   Sequence Analysis

The amplification products obtained from the endpoint PCR for an *oxd* gene were sequenced to confirm the gene was in fact an aldoxime dehydratase (sequencing data available in appendix I). Following the confirmation of *oxd* gene detection, 16s rRNA gene sequencing of the 12 unknown environmental nitrile metabolising isolates identified in this study was performed as per Section 2.3, in order to determine the genus of the unknown nitrile-metabolising isolates which were shown to be *oxd* positive. A minimum of 425 bp was used for 16s rRNA analysis with a sequence homology cut off of >98%. The 16s rRNA gene sequence was used to identify the genera of bacteria presented in Table 2.5.

Table 2.5 Taxanomic classification of previously unidentified and pre-existing putative nitrile metabolising environmental isolates shown to harbour the oxd gene with corresponding oxd Ct values and the 16s rRNA gene accession number.

Isolate	Geographical Location	N source[a]	16s rRNA Accession Number	*oxd* Ct Value
Brevundimonas sp. ST111	Ireland	3-Hydroxybutyronitrile	KX495348	23.03
Rhizobium sp. ST125	Ireland	3-Hydroxybutyronitrile	KX495352	20.44
Rhizobium sp. ST119	Ireland	3-Hydroxybutyronitrile	KX495353	18.83
Bacillus sp. ST144	Ireland	3-Hydroxy-3-phenylpropionitrile	KX495347	21.55
Enterobacter sp. ST146	Ireland	3-Hydroxy-3-phenylpropionitrile	KX495349	22.18
Pseudomonas sp. ST30	Ireland	3- Hydroxybutyronitrile	KX495351	19.40
Pseudomonas sp. ST196	Ireland	3-Hydroxyglutatonitrile	KX495350	21.19
R. erythropolis ST158	Ireland	3-Hydroxyglutaronitrile	KX495356	21.93
Rhodococcus sp. ST135	Ireland	3-Hydroxybutyronitrile	KX495355	19.66
Rhodococcus sp. ST199	Ireland	3-Hydroxyglutaronitrile	KX495354	22.55
Stenotrophomonas sp. ST180	Ireland	3-Hydroxyglutaronitrile	KX495357	22.22
Stenotrophomonas sp. ST53	Ireland	3-Hydroxybutyronitrile	KX495358	22.32
Pre-existing putative Nitrile metabolising Isolates (Coffey *et al.*, 2010)				
Bacillus sp. F10	Australia	Acetonitrile		21.04
Bacillus sp. NN1	Ireland	Acetonitrile		26.03
Burkholderia sp. F73	Poland	Acrylonitrile		26.33
Buttiauxella sp. SW2-28	Ireland	Acrylonitrile		20.05
R. erythropolis SET-1	Ireland	Adiponitrile		25.19
Rahnella sp. SS1-25	Ireland	Benzonitrile		24.37
Rahnella sp. SS1-7	Ireland	Acetonitrile		24.02
Rhodococcus sp. SS1-29	Ireland	Acrylonitrile		25.51
Rhodococcus sp. F17	Poland	Acetonitrile		20.35
Rhodococcus sp. F37	Australia	Adiponitrile		21.93
Rhodococcus sp. F52	Ireland	Mandelonitrile		22.55
Rhodococcus sp. SS1-2	Ireland	Acetonitrile		22.35
Rhodococcus sp. LC11	Ireland	Acetonitrile		22.65
Rhodococcus sp. F30	Australia	Benzonitrile		27.15

Isolate	Geographical Location	N source[a]	16s rRNA Accession Number	*oxd* Ct Value
Rhodococcus sp. NN5	Ireland	Acetonitrile		24.15
Serretia sp. SS1-18	Ireland	Benzonitrile		26.33
Streptomyces sp. F61	Poland	Mandelonitrile		27.79

[a]Sole nitrogen source used in M9 media for the selective enrichment of putative nitrile metabolising isolates from environmental sources.

A phylogenetic tree was constructed using the conserved internal region of the amplified *oxd* gene. The tree was constructed to assess sequence homology across the subset of thirty isolates, with sequence homology observed in the range of 98% to 100% (Figure 2.11). A high level of similarity was observed across the sequenced *oxd* genes with *oxd* gene sequence from *R. jostii* RHA1 as the rooting sequence.
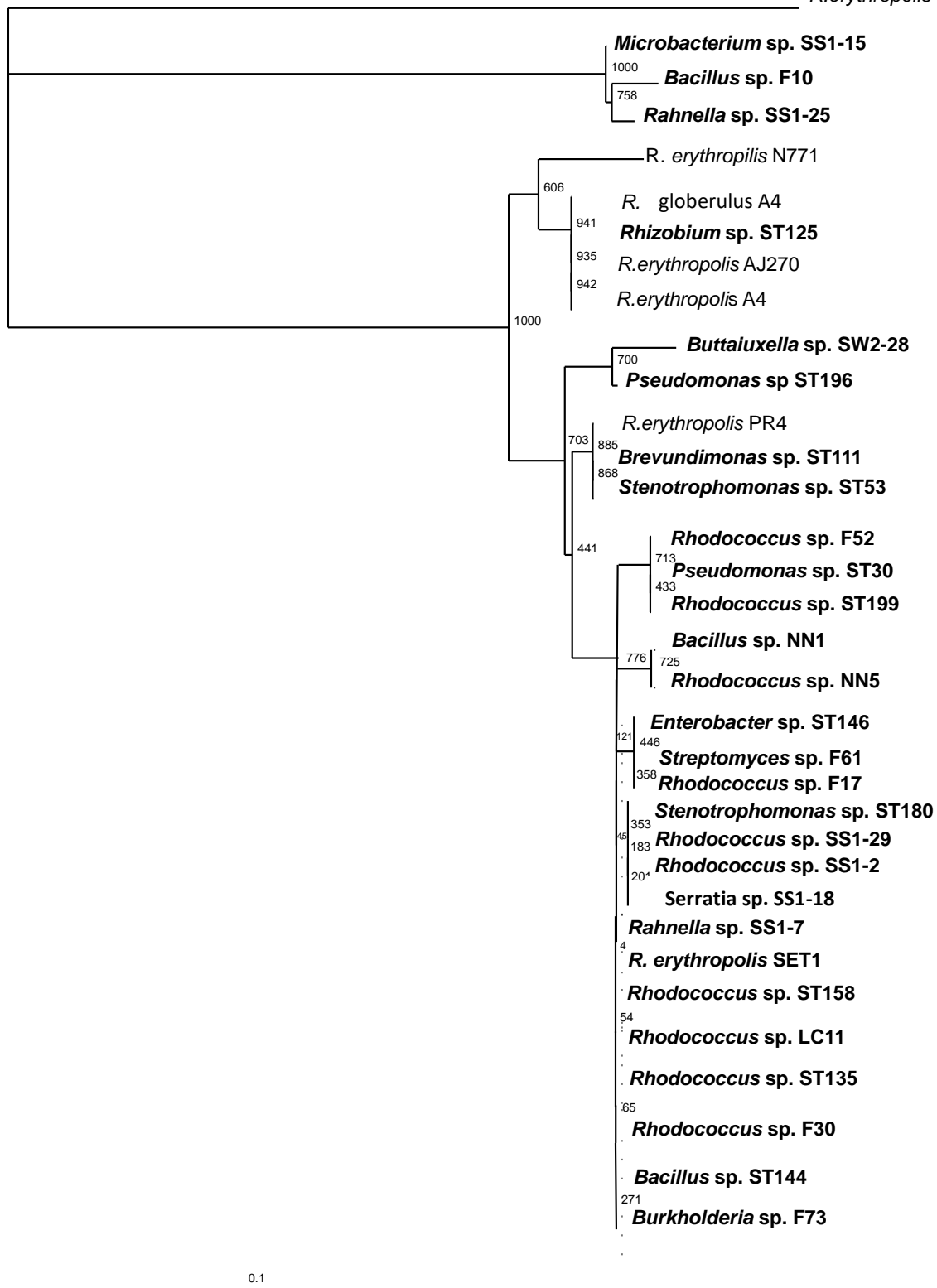
Figure 2.11 A phylogram constructed with a 140 bp region of the conserved internal oxd gene region across several environmental nitrile metabolising isolates.

R. jostii RHA1 was used to root the tree where the strains presented in this research are in bold.

2.5 Discussion

The microbial biocatalysis library was expanded to improve the chances of detecting isolates containing genes of interest and demonstrating high performance activities on substrates of interest. Three key industrially important nitrile substrates (

Figure 2.3 were used for the enrichment and isolation of putative nitrile metabolising isolates as described in Section 2.3.2. The soil samples selected for the enrichment and isolation of putative nitrile metabolising isolates were from various regions around the south east of Ireland, coordinates noted in Table 2.4. Additionally, the pre-existing soils were selected from a number of geographically diverse locations including Australia and Poland, as noted in Table 2.5. Soils were selected based on potential nitrile contamination from industrial or anthropological practices at the site of collection. One substrate of interest is 3-HGN as it holds industrial relevance for the production of statins. Statins are utilized for the treatment of hypocholesterolemia and atherosclerosis.

Many microorganisms containing genes for the aldoxime-nitrile metabolising pathway have been isolated from environmental samples. A pooled soil enrichment approach was carried out due to limited resources (nitrile substrate). However, enrichments were also carried out with individual soils as noted in Table 2.4. Early research pooled soil samples for enrichments on the key nitriles of interest, however, the enrichment of individual soils reduced the risk of isolates demonstrating higher activity outcompeting those with lower activity. This was paramount as the lower activity isolates may have a higher enantioselective (ee%).

The purpose of the described research was to design a real-time PCR assay for the high-throughput detection of novel aldoxime dehydratase metabolising genes in several environmental nitrile metabolising isolates across multiple genera. To overcome the limitations of end-point PCR, such as poor sensitivity and the inability to quantify results, a real-time PCR method was designed.

Aldoximes can be prepared by a condensation reaction of an aldehyde with hydroxylamine, as this is perceived to be an accessible route for substrate production. Aldoximes are of high value for synthetic purposes (Betke *et al.* 2018). This ensures that there would be an accessible route for aldoxime synthesis towards the production of

intermediate or key nitriles. There is an ever-increasing interest in the use of aldoxime dehydratase towards the production of nitriles (Piatesi *et al.* 2013, Metzner *et al.* 2014), and in turn there is an increase in demand for the identification of novel *oxd* genes across multiple genera as this will allow for greater selection of enzymes with optimal activity (Dooley-Cullinane *et al.* 2016).

As previously noted, Fe-type NHase genes have been previously detected in a number of the pre-existing environmental nitrile metabolising isolates which were utilised in this study (Coffey *et al.* 2010). This led to the selection of these isolates for the real-time screening due to the fact that identification of *oxd* genes upstream of the NHase gene cluster have been observed in a number of microorganisms, in particular *Pseudomonas chlororaphis* B23 (Oinuma *et al.* 2003). The remaining isolates chosen from the study were putative nitrile-metabolisers, as demonstrated by their ability to utilize either 3-hydroxybutyronitrile (3-HBN), 3-hydroxyglutaronitrile (3-HGN) or 3-hydroxy-3-phenylpropionitrile (3-HPPN) as their sole nitrogen source in minimal media.

Oxd positive signals were typically represented by Ct values in the range of 19 to 26. The range of Ct values observed can be explained by the variations in differences in primer binding specificity. If the primer binding specificity was lower for some of the environmental isolates assayed, then variations in the Ct values would yet again be observed. The standard curve designed (Figure 2.7) displayed an R^2 value of 0.997 with *R. erythropolis* AJ270, demonstrating linear correlation between the Ct value obtained and the quantity of cells or DNA (ng/µL). The amplification efficiency of the assay was calculated to be 91.93%. Efficiency values in the range of 90% to 110% are typically acceptable with an amplification efficiency measuring over 100% possibly due to the presence of primer dimer formation which is typically of greatest concern with SYBR$^®$ Green assays (Agilent Genomic, 2018).

The high-throughput screening of microbial isolates for the presence of novel *oxd* genes across several genera was aided via the designed real-time PCR assay. To ensure that the *oxd* positive signals detected during the real-time high throughput screen were not false positives, endpoint PCR was carried out to ensure that the conserved internal region of the *oxd* gene could be detected and amplified. Some isolates such as *Stenotrophomonas*

sp. ST53 did not allow for the amplification of the conserved internal region upon initial screening of the *oxd* gene via endpoint PCR, despite displaying a Ct value in the positive signal range. The lack of amplification of the conserved internal region could be due differences in residues at the various primer binding sites or an alternative intermediate sequence leading to the formation of secondary structures, thus making it difficult for the *Taq* polymerase to amplify the gene of interest.

A gradient endpoint PCR was carried out to allow for the identification of the adjusted annealing temperature which in turn allowed for the amplification of the *oxd* gene in *Stenotrophomonas* sp. ST53. No false-positive detection for an *oxd* gene in the sub-section of the screened isolates was observed. For the 30 *oxd* positive real-time PCR signal yielding isolates tested, *oxd* gene partial sequence was confirmed. The need for DNA extractions from environmental samples was omitted from the assay via the use of cells as the template for real-time PCR. The elimination of such a step further demonstrates the high-throughput capabilities of the real-time screening platform.

Sequence homology between the elucidated conserved partial *oxd* sequences was displayed by construction of a phylogenetic tree, (Figure 2.11). To root the tree, the phenylacetaldoxime dehydratase gene from *R. jostii* RHA1 was used. Sequence homology of 66.4 % can be observed between the conserved *oxd* region *of R. erythropolis* AJ270 (positive control) and *R. jostii* RHA1.

The NCBI database indicated that many identified *oxd* genes were detected among *Rhodococcus* sp. leading to database sequence bias. Prior to this research, approximately 12% of the *oxd* genes on the database were identified in *Rhodococcus* sp. with approximately 15% of publications studying the *oxd* genes from *Rhodococcus* sp. Considering the database bias towards *oxd* gene sequences in *Rhodococcus* sp. it would be inevitable that primer design based on alignments of these sequences would thus further bias the screening for *oxd* genes. Thus, implying that such methods and primers developed on the limited sequences at present would be more liable to detect other highly similar *Rhodococcus* sp. *oxd* genes. This would warrant further skewing of the database in the future with the potential future detection of *oxd* genes to be further limited/skewed towards one genera.

In this research, *oxd* genes across 11 genera were detected based on the qPCR *oxd* primers which were designed for the internal conserved region. Though the assay did detect genes from *Rhodococcus* sp. many of the *oxd* genes identified were from other genera, some of which were not previously shown to harbour *oxd* genes. *Brevundimonas* sp., *Buttiauxella* sp., and *Rahnella* sp., are some of the genera not previously shown to harbour *oxd* genes (Table 2.5). This signified that the region is not limited to conservation in *Rhodococcus* sp., ensuring that the described assay can aid the reduction in database bias earlier described.

Research published by Kato *et al.* (2005) presented an endpoint PCR assay for the detection of novel *oxd* genes, where the *oxd* gene from *R. erythropolis* N-771 and *Bacillus* sp. Ox-B 1 were utilized as probes for southern blot hybridization. The study did not demonstrate the capability of amplifying *oxd* genes from isolates comprising Co-type NHase or Nitrilase genes. This may be due to low sequence similarities between the *oxd* genes associated with each nitrile metabolising enzyme. In this study an *oxd* gene was detected in *R. erythropolis* SET-1, a known nitrile degrader as shown by Coady *et al.* (2013). Further research presented in Chapter 4 and 5 details the detection of a partial nitrilase gene, further indicating that this assay may prove beneficial for the detection of *oxd* genes associated with microbial isolates containing a confirmed nitrilase.

As aldoxime dehydratase enzymes are part of the 'aldoxime-nitrile' pathway in microorganisms, it is also worth considering that isolates on the database which contain a known aldoxime dehydratase should be explored for novel nitrile metabolising genes. As presented in this body of work, a number of the isolates selected for gene screening were selected on the premise that they contained a known nitrile metabolising gene, as published by Coffey *et al.* (2010). All of these isolates display ed a positive signal (qPCR) for an aldoxime dehydratase gene and subsequent sequencing of the end-point PCR amplicon confirmed that partial gene sequence detected was in fact a partial aldoxime dehydratase gene. This indicates that there may be numerous undetected/unidentified aldoxime dehydratase genes currently on the NCBI database, a potential untapped resource for the biocatalysis industry. It may also be worth considering that as both the aldoxime dehydratase and nitrile metabolising genes form a pathway, the substrate preferences of the detected nitrile metabolising gene may give an indication of the potential metabolising abilities of the aldoxime dehydratase, if present.

Based on the research presented by Betke *et al.* (2018) a large number of the aldoxime dehydratase enzymes screened in their study displayed some difficulty with aromatic structures/substrates, perhaps due to steric hinderance, however, this may serve as a guide for the discovery of novel aldoxime dehydratases. Based on the previous point, it may be said that nitrile metabolising isolates demonstrating good activity on aromatic nitriles may also house an aldoxime dehydratase with the capabilities to dehydrate the aldoxime precursor.

There are many benefits associated with the use of aldoxime as the precursor to nitriles of interest. Firstly, aldoxime dehydratase enzymes can be utilised to easily prepare nitriles as described by Betke *et al.* (2018). Secondly aldoxime dehydratase has displayed the ability to produce enantiospecific nitriles from racemic aldoximes (Betke *et al.* 2018). This is important for many reasons: (i) it decreases the need for purification of the enantiomer of interest (ii) it decreases the amount of downstream processing required and in turn cuts costs, (iii) it will prove beneficial for sectors such as the bio-pharmaceutical industry where only one enantiomer is required.

Additionally, a patent was filed by BASF enzymes describing the application/use of aldoxime dehydratase for the production of citroenllyl nitriles and other compounds utilized in the fragrance industry (Piatesi *et al.* 2013). In some cases Betke *et al.* (2018) noted that some aldoxime dehydratase enzymes displayed a change in preference of enantioselectivity, where the R-aldoxime resulted in the formation of the S-nitrile. Though this is not the case with all aldoxime dehydratase enzymes and seems to be substrate dependent, it is a characteristic that could prove beneficial in the synthesis of new or novel nitrile substrates for the bio-pharmaceutical industry.

Previous research carried out within the research group highlighted evidence of HGT of Fe-type NHase (Coffey *et al.* 2010, O' Mahony *et al.* 2005) and nitrilase genes (Coffey *et al.* 2009). In this study, *oxd* genes from microorganisms which have been isolated from geographically diverse regions displayed high sequence homology, which is thought to be suggestive of HGT. It is believed that changes in environmental conditions due to human activities have noticeably led to HGT-mediated evolutions, particularly evident in bacteria. High sequence homology in the range of 98-100% suggests that the *oxd* genes originated from a common ancestor/source. It is essential to document that the sequence

homology is based on the conserved internal *oxd* region and so it should be noted that the percentage homology observed between full *oxd* genes would be expected to be of a lower percentage. Differences in amino acid residues were observed despite the high DNA sequence homology between the elucidated *oxd* genes, indicating that they are in fact different *oxd* genes being detected across the various genera.

As addressed in Section 1.9.3, *oxd* genes have previously been shown to typically be part of NHase gene clusters and have been associated with nitrilase activity (Kato & Asano 2006). The distribution of a variety of genes including antibiotic resistance, catabolic and virulence genes is aided by genomic islands (GIs), in turn accounting for the evolution of many bacteria (Juhas *et al.* 2009). The GIs in question are often mobile elements and can be acquired via HGT, thus accounting for the mobility of the *oxd* genes. Likewise, conjugated transposons may facilitate plasmid mobilization and co-integrate formation between distantly or closely related bacteria as they demonstrate the ability to form covalently closed circular intermediates which can independently excise (Davison 1999, Salyers *et al.* 1995). High DNA homology has been observed within the conserved regions of the detected partial aldoxime dehydratase genes; this is expected as the primers were designed on areas of high conservations. However, should the full genes be amplified, and high-level DNA homology be observed across the full gene it could be proposed that these genes were acquired through HGT, in particular if the originating microorganisms are from geographically diverse locations.

Though high homology is observed between most of the elucidated sequences, a relatively lower sequence identity within the conserved region of *Microbacterium* sp. SS1-15, *Bacillus sp*. F10 and *Rahnella* sp. SS1-25 is evident. It is possible that this lower sequence homology percentage is due to sequence divergence or mutation events. As the *oxd* genes were detected from several environmental nitrile metabolisers, whose aldoxime degrading abilities have not been tested, it is not known if the *oxd* genes detected are active or silent. Because of this it cannot be determined if the divergence is due to HGT via functional selection due to selective pressure or if the mutations are in fact silent mutations and any potential activity remains unchanged or even lost.

2.6 Conclusion

To summarise, a novel qPCR method was developed which detected novel partial aldoxime dehydratase genes in putative and known nitrile metabolising isolates. The high-throughput real-time PCR assay designed and described above allows for the detection of *oxd* genes from multiple genera, including some of which had not been previously shown to harbour these genes, and across microorganisms isolated from environmentally and geographically diverse locations. Moreover, the detection of the *oxd* gene across many genera will contribute to the reduction in database bias, further ensuring that there will in future be a larger pool of *oxd* genes for database mining and further studies. The presented study will also allow for the detection of novel *oxd* genes which can be employed in the biotransformation and production of industrial and pharmaceutically relevant nitriles, thus assisting the move towards greener chemistry.

CHAPTER 3

SCREENING OF THE BIOCATALYSIS LIBRARY FOR NOVEL NITRILASE GENES VIA NIT CLADE TOUCHDOWN PCR PROTOCOL

Note

- The nitrilase clade touchdown PCR assay presented in this chapter was designed by Lee Coffey (WIT) and has previously been used for the successful detection of novel nitrilase genes from the pre-existing nitrile metabolising bank of environmental isolates.

- In this chapter, the nitrilase clade touchdown PCR assay was used to screen 12 environmental isolates which have been shown to contain an *oxd* gene and whose isolation was detailed in Chapter 2; however, no nitrilase gene was detected.

- This project analysed all the sequencing data available from a previously identified nitrilases against an updated database. In this study the assay was used to screen more isolates with sequencing and phylogenetic analysis, published as per the reference provided above. Work not carried out within this study is clearly identified throughout the chapter. Additionally, this assay was used to identify a nitrilase from *R. erythropolis* SET-1 in this study which will be detailed in Chapter 4.

3 SCREENING OF THE BIOCATALYSIS LIBRARY FOR NOVEL NITRILASE GENES VIA NIT CLADE TOUCHDOWN PCR PROTOCOL

In 2004 Robertson *et al.* identified 137 novel nitrilase genes from the creation and functional screening of a metagenomic library which contained 651 environmental samples. Prior to this research there were fewer than 20 nitrilases reported across the scientific and patent literature. The creation of environmental DNA (eDNA) libraries ensured that the non-culturable portion of the environment, encapsulating wholly diverse genomes, was utilized for the novel gene discovery expedition.

After analysis of the 137 novel nitrilase sequences Robertson *et al*. (2004) observed that the nitrilases could be separated into distinct clades based on sequence homology. Figure **3.1** presents the phylogenetic analysis of the 137 novel nitrilases with obvious clade classifications observed by the sequences grouping in clusters (Section 1.5.2). Figure **3.1** also displays the activity and enantioselectivity profiles of the novel nitrilases with a preference for (R)-enantioselectivity (A) and the phylogenetic display (B) displays the enantioselectivity profiles of the novel nitrilases towards mandelonitrile.

Mandelonitrile
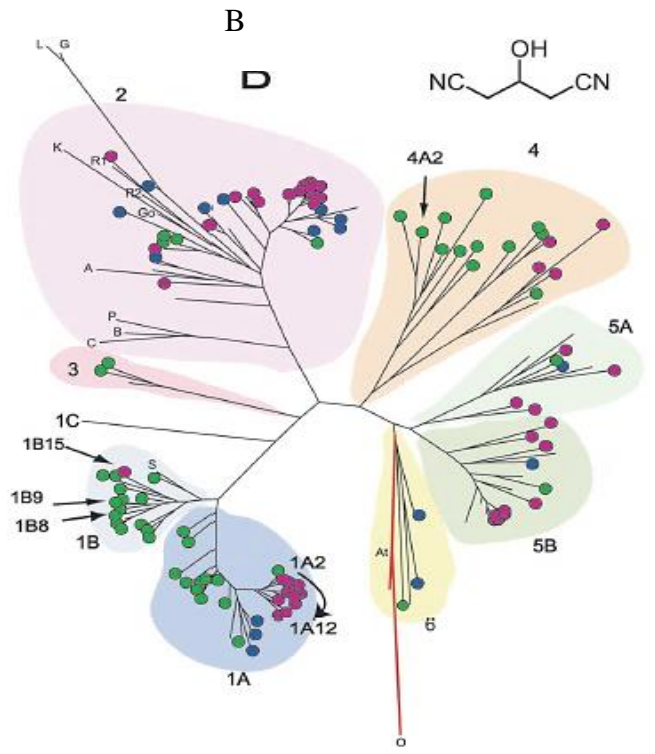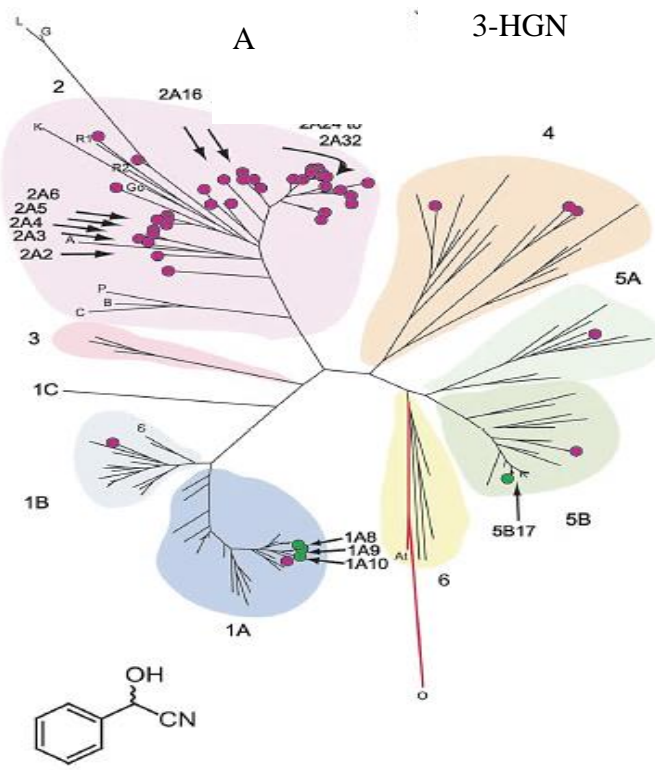
Figure 3.1 Phylogenetic analysis of the 137 discovered nitrilase sequences from eDNA library construction and screening. The schematic presents a likelihood tree as presented by Robertson *et al.* 2004.

The phylogenetic display (A) represents the enantioselectivity profile of the novel nitrilases towards 3-HGN with magenta dots representing (R)-enantioselectivity and green dots representing (S)-enantioselectivity with blue dots indicative of no activity towards the model substrate. The phylogenetic display (B) represents the enantioselectivity profiles of the novel nitrilases towards mandelonitrile with magenta dots representing (R)-enantioselectivity and green dots representing (S)-enantioselectivity with blue dots indicative of no activity towards the model substrate.

From this it was seen that the clade classification of the novel nitrilases did allow to an extent the prediction of enantioselectivity profiles. The importance of substrate specificity stems for the intended application of these enzymes. Within the biotechnological and pharmaceutical industry, significant focus is placed upon the purity of the expected products and the enantiomer which is produced, further showing the importance of gene sequences and their relationship to the predicted enzyme activity (Coady 2014, Martínková and Kren 2010, Rasor and Voss 2001). Robertson *et al.* (2004) also tested the 137 nitrilases towards the industrially relevant substrate mandelonitrile (Figure 3.2).
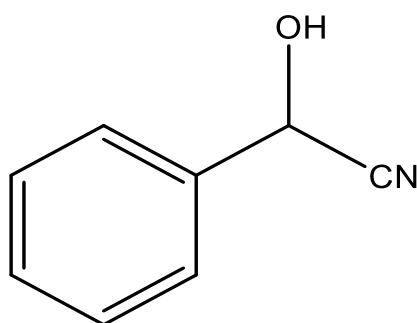
Figure 3.2 Chemical structure of mandelonitrile

The enantioselectivity of the 2A clade (37 members) was seen to demonstrate only (R)-selectivity towards mandelonitrile, as conveyed by the magenta dots, with only 4 nitrilases from the 1A (31 members) (1A8, 1A9 and 1A10) and one nitrilase from the 5B clade (21 members) (5B17) displaying (S)- enantioselectivity.

Previous research by the research group has focused on the isolation and screening for nitrile metabolising isolates (Coffey *et al.* 2010, Coffey *et al.* 2009) designing screening protocols for the aldoxime dehydratase gene (*oxd*) in the aldoxime-nitrile metabolising

pathway, as presented in chapter two and the use of applied metagenomics for the discovery of novel nitrile metabolising enzymes (Soares Bragança *et al.* 2017). This chapter presents a touchdown PCR protocol which was developed by Lee Coffey (WIT). The method was used in this body of work to screen a subset of nitrile metabolising isolates (which were detailed in Chapter 2) for novel partial nitrilase genes. The phylogenetic analysis of the previously discovered partial nitrilase sequences was performed against an updated database. All sequencing and phylogenetic analysis was performed as part of this body of work.

3.1 Research Hypothesis and Aims

Hypothesis:

A pre-exisitng nitrilase clade PCR protocol can be used to discover novel nitrilase genes in putative and confirmed nitrile metabolsing isolates.

Aim:

The aim of this research was to discover novel partial nitrilase genes from a number of environmental putative nitrile metabolisers, with a particular focus on the confirmed nitrile metabolising isolate *R. erythropolis* SET-1

3.2 Method and Materials

3.2.1 Soil sources and environmental nitrile metabolising isolates

Soil samples were collected from various locations across Ireland with soil 8 from Peoples Park Waterford, soil 9 from under the Lauryl Tree in Piltown, Co. Kilkenny and Soil 10 from under a bonfire site. Seaweed samples (A) *Ceramium codiola*, (B) *Ulva lactuca* and, (C) *Palmaria palmata* were collected from local beaches in the South of Ireland. A selective enrichment was performed using 10 g of these environmental samples in 10 mL of M9 minimal media (recipe in Appendix A) with 10 mM nitrile as the sole nitrogen source in the media. These enrichments were incubated at 25 °C at 150 rpm for 7 days prior to sampling.

Post-enrichment 100 µL of the selective enrichment supernatant was spread plated onto M9 minimal media agar (M9 minimal media containing 1.5% purified agar) containing 10 mM of the corresponding nitrile as the sole nitrogen Source. These plates were incubated at 25 °C for 7 days until visible grown was seen. Resulting microorganism were purified for a minimum of 3 consecutive rounds of screening on M9 minimal agar until isolate purify was confirmed.From the selective enrichments and subsequent purifications of these soils, Coffey *et al.* (2009) identified a number of nitrilase gene sequences which are subjected to phylogenetic analysis in this study (Table 3.1).

Table 3.1 List of putative nitrile metabolising environmental isolates and their corresponding nitrilase gene amplified by Lee Coffey (WIT) and analysed in this study.

Isolate	Clade	Soil or Seaweed	Nitrilase Accession Number
Unidentified isolate SS4	1A	10	MG383631
Unidentified isolate SS17	1A	9	MG383638
Unidentified isolate. SS17	1A	9	MG383637
Unidentified isolate SS17	1A	9	MG383636
Unidentified isolate SS9	1B	8	MG383630
Unidentified isolate SS10	1B	9	MG383634
Unidentified isolate SS6	1B	8	MG383628
Unidentified isolate SS8	1B	8	MG383629
Unidentified isolate SW2-37	1B	A	MG383640
Unidentified isolate SS34	2A	10	MG383637
Unidentified isolate SS33	2A	10	MG383635
Unidentified isolate SS10	2A	9	MG383632
Staphylococcus sp. SW2-31	2A	C	MG383639
Unidentified isolate SS33	4A	10	MG383636
Unidentified isolate SW2-10	5A	B	MG383638

Eighteen environmental putative nitrile metabolising isolates which were purified and screened for the presence of an *oxd* gene as presented in Chapter 2 were also subjected to screening for a nitrilase gene.

Table 3.2 Nitrile metabolising environmental isolates which were identified as being oxd^+ in Chapter 2, which were screened for the presence of a clade-specific nitrilase gene.

Isolate	Geographical Location	Sole nitrogen (N) source	16s rRNA Accession Number
Brevundimonas sp. ST111	Ireland	3-Hydroxybutyronitrile	KX495348
Rhizobium sp. ST125	Ireland	3-Hydroxybutyronitrile	KX495352
Rhizobium sp. ST119	Ireland	3-Hydroxybutyronitrile	KX495353
Bacillus sp. ST144	Ireland	3-Hydroxy-3-phenylpropionitrile	KX495347
Enterobacter sp. ST146	Ireland	3-Hydroxy-3-phenylpropionitrile	KX495349
Pseudomonas sp. ST30	Ireland	3- Hydroxybutyronitrile	KX495351
Pseudomonas sp. ST196	Ireland	3-Hydroxyglutatonitrile	KX495350
R. erythropolis ST158	Ireland	3-Hydroxyglutaronitrile	KX495356
Rhodococcus sp. ST135	Ireland	3-Hydroxybutyronitrile	KX495355
Rhodococcus sp. ST199	Ireland	3-Hydroxyglutaronitrile	KX495354
Stenotrophomonas sp. ST180	Ireland	3-Hydroxyglutaronitrile	KX495357
Stenotrophomonas sp. ST53	Ireland	3-Hydroxybutyronitrile	KX495358

3.2.2   Primer design

Primer sets were designed by Lee Coffey (WIT) using a ClustalW alignment in 2010 of clade-specific nitrilase genes as published by (Robertson *et al.* 2004) and all unclassified-clade nitrilase sequences available in the Genbank database (Benson *et al.* 2013), including putative sequences. Each primer pair was designed with a maximum degeneracy of 192. All primers designed and used in this study were supplied by Eurofins, MWG operon, Germany (Table 3.3).

Table 3.3 Clade-specific primer pairs sequences (5'-3') and expected amplicon size (bp)

Nitrilase Clade	Forward Primer	Reverse Primer	Expected Amplicon size (bp)
1A	atgrtctggggvcargghga	tcdccytgbccccagaycat	287
1B	caygarcgsatgrtstgggg	tccatcatbckyttkcgytt	440
2A1-20	gsvytbtgctgytgggarca	rtartgvccrgcvggrtc	487
2A2-37	gsvytbtgctgytgggarca	gartartgsccgrcsggrtc	512
3A	caycgcaarctscarccsac	ttcatsakbscstcratctg	125
4A	caycgcaarytgrwgccsac	catcywrtkytcccagca	125
5A	tgctgggarmayyayatgcc	tcsgsvcgcgmrtartgscc	412
5B	tgctgggarmayyayatgcc	tcrccdwyrttratccaytc	250
6A	caycgyaagctcrtgccvac	catdhrgttytcccagcaga	250

*Universal code for degenerate bases R = A/G, Y = C/T, M = A/C, K = G/T, S = C/G, W = A/T, B = C/G/T, D = A/G/T, H = A/C/T, V = A/C/G, and N = A/C/G/T.

3.2.3 Nitrilase gene screening approach - *Touch-down PCR protocol*

Each 15 µL PCR mixture contained 7.5 µL Gotaq™ Green Master Mix (Promega, UK), 15 pmol of each primer and (i) environmental bacterial cell suspensions adjusted to a final O.D.600_{nm} = 0.04, or (ii) 15 ng of metagenomic fosmid clone DNA. The following PCR conditions were used: 1 cycle of 95 °C for 5 min, 2 cycles of 95 °C for 1 min, 58 °C for 1 min, 72 °C for 40 seconds, followed by a 1 °C reduction in annealing temperature every 2 cycles to 51 °C inclusive; followed by 20 cycles of 95 °C for 1 min, 50 °C for 1 min and 72 °C for 40 seconds. A final extension stage of 8 minutes was utilised. Negative controls contained water in place of cells/DNA, with positive controls consisting of nitrilase DNA, supplied by Verenium Corporation (now BASF, USA), listed in Table 3.4.

3.2.4 Amplification of 16s rRNA genes

The primers were synthesised and supplied by MWG Biotech Ltd. Per 15 µL reaction mixture the assay contained: 50-100 ng DNA, 7.5 µL of 2X GO-Taq green mastermix (Promega, Cat. No. M7501), 1 µL of 25 µM PCR mix (final primer concentration 1.66 µM) and 5.5 µL of sterile water. For colony PCR cells were resuspended in sterile water to an O.D.600nm = 1.0 and 1 µL was added to the reaction (final cell O.D.600nm = 0.06).

The 63F and 1387R primer pair were used for the amplification of the 16s rRNA gene (Marchesi *et al.* 1998).

Table 3.4 Clade-specific positive controls utilized during the touchdown PCR screening with corresponding accession numbers provided by Verenium Corporation (now BASF, USA).

Clade	Clone I.D.	Accession Number
1A	1	AY487553.1
1A	3	AY487431.1
1A	21	AY487549.1
1A	23	AY487460.1
1A	26	AY487560.1
1B	7	AY487530.1
1B	8	AY487510.1
1B	10	AY487550.1
1B	15	AY487529.1
2A	4	AY487473.1
2A	13	AY487497.1
2A	14	AY487438.1
2A	20	AY487543.1
2A	24	AY487523.1
2A	25	AY487448.1
3A	2	AY487516.1
4A	2	AY487520.1
4A	9	AY487490.1
4A	14	AY487498.1
5A	2	AY487469.1
5A	5	AY487494.1
5A	8	AY487517.1
5B	1	AY487467.1
5B	9	AY487493.1
5B	20	AY487466.1
6A	2	AY487454.1
6A	3	AY487447.1

3.2.5 Gel extractions

PCR products that required purification were purified using the QIAquick Gel Extraction Kit (Qiagen Cat. No. 28704) or the Promega Wizard® SV Gel and PCR Clean-up System (Cat. No. A9281) as per the manufacturer's instructions, with the exception of sterile distilled water being used in place of elution buffer.

3.2.6 Cloning of potential nitrilase amplicons

Previously identified partial nitrilase amplicons were cloned into the pDrive vector, (Qiagen, cat. no. 231122) by Coffey *et al.* (2009). Potential products generated in this study were cloned into the pGEM®-T Easy vector (Promega, A1360) (Figure 3.3). Plasmids were transformed into Novablue Gigasingles™ (Novagen, cat. no. 71227) andpositive clones were selected via blue/white screening.
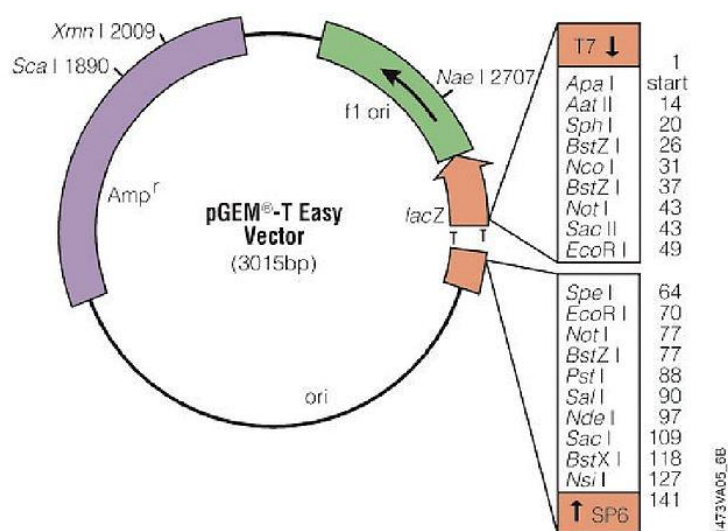
Figure 3.3 pGEM®-t Easy vector (3.015 kB) map.

The vector map displays the multiple cloning site (MCS) flanked with restriction sites, AmpR (ampicillin resistance marker), Origin denotes the origin of replication with f1 origin allowing for the preparation of single-stranded DNA (ssDNA). P_{lac} and lacZ encoding for the lac promoter and β-galactosidase respectively (Fisher Scientific, 2017).

3.2.7   Sequence Analysis

Sequencing (carried out previously by Lee Coffey, WIT) of each potentially positive nitrilase plasmid inserts as cloned previously by the research group was carried out in triplicate using the vector-specific primers M13 Forward (-20) and M13 reverse, with all sequencing carried out by GATC Biotech, Germany. Nucleotide sequences were analysed using Blastn software from the NCBI database (Altschul *et al.* 1990). The same sequencing approach was carried out with the potential positive nitrilase plasmid inserts amplified within this study with the exception of primer pairs T7 promoter primer and the Sp6 promoter primer. Phylogenetic trees were constructed using Dendroscope (Uson & Cornavacca, 2012) with a clustalW alignment of the sequenced nitrilase region with gaps allowed, bootstrap trials at 1000 and bootstrap labels on nodes.

3.3   Results

3.3.1   Nitrilase Clade Screening

Of the 12 environmental nitrile metabolising isolates (Table 3.2) which were screened, six isolates produced an amplicon of the expected size (approximately 512 bp) for the 2A clade using the 2A1-37 primer pair (Figure 3.4). No other isolate provided an amplicon of the expected size across the remaining clades.
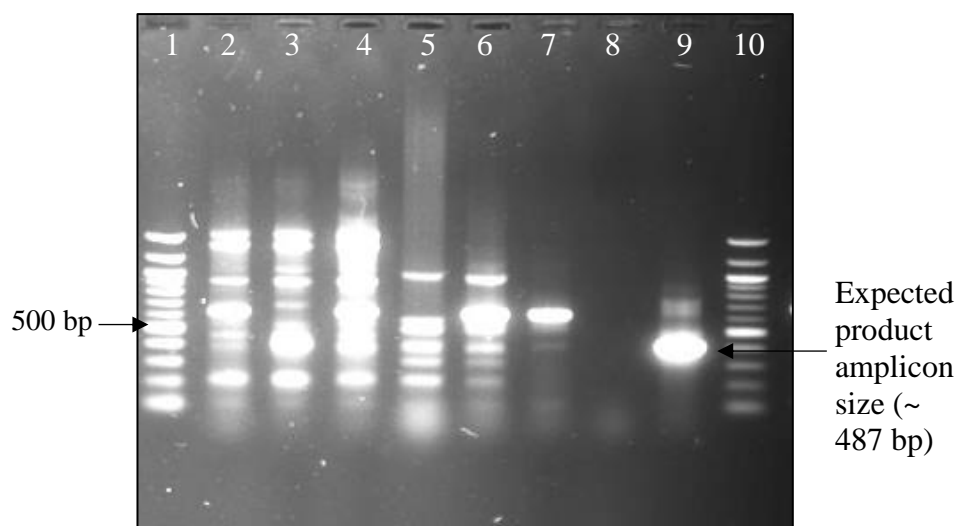
Figure 3.4 Gene screening of *oxd* positive isolates for a 2A clade nitrilase using primer pair 2A1-37.

Promega 100 bp marker (lane 1), *Brevundimonas* sp. ST111 (lane 2), *Rhizobium* sp. ST119 (lane 3), *Pseudomonas* sp. ST30 (lane 4), *Rhodococcus* sp. ST135 (lane 5), *Rhodococcus* sp. ST199 (lane 6), *Stenotrophomonas* sp. ST180 (lane 7), negative control (lane 8), positive control (Verenium nitrilase) (lane 9), Promega 100 bp marker (lane 10).

The potential nitrilase PCR amplicons were cloned into the pGEM®-t Easy vector and transformed into gigasingles. The resulting plasmid inserts were sequenced. Following sequencing analysis, it was confirmed that none of the cloned amplicons were nitrilase sequences. The *Brevundimonas* sp. ST111 amplicon displayed 98% sequence homology a non-ribosomal peptide synthetase in the *Rhodococcus* sp. H-CA8f chromosome (accession number CP023720.1). The amplicon from *Rhodococcus* sp. ST199 displayed 100% sequence homology to a decarboxylase in *Rhodococcus* sp. PBTS2 (accession number CP015220.1). The *Pseudomonas* sp. ST30 amplicon displayed 99% sequence homology to a putative hydrolase from *Rhodococcus erythropolis* YL-1 genome (accession number CP017299.1). The *Stenotropohomonas* sp. ST180 amplicon displayed 100% sequence homology to the LuxR family transcriptional regulator from *R. erythropolis* R138 genome (accession number CP007255.1). The *Rhizobium sp. ST119* amplicon displayed 100% sequence identity to a peptide synthetase in *Rhodococcus erythropolis* B343 (accession number CP011295.1). The *Rhodococcus* sp. ST135 amplicon displayed 99% homology to a class II Aldolase in multiple *Rhodococcus* sp.

Previous research by Coffey *et al.* (2009) identified four novel partial putative nitrilase sequences in the 1A clade with sequence homologies to a number of nitrilase genes across multiple genera. Four novel partial nitrilase sequences were also identified within the 2A clade encompassing *Microbacterium* sp.*, Arthrobacter* sp., *Serratia* sp., and *Staphylococcus* sp., with sequence homology observed across a number of nitrilases from multiple genera, specifically *Rhodococcus* sp. One partial novel nitrilase sequence was identified in the 4A clade in *Microbacterium* sp. SS33, with an observed sequence homology to a small number of partial nitrilase sequences available on the database. One novel 5A nitrilase was discovered from *Ochrobactrum* sp. with a high sequence homology to a number of 5A *Burkholderia* sp. nitrilase sequences.

3.3.2   16s rRNA identification

The unidentified isolates yielding nitrilase sequence were identified within this study by 16S rRNA PCR amplification and subsequent sequencing as per Marchesi *et al.* (1998) (Figure 3.5 and Figure 3.6), with the exception of *Staphylococcus* sp. SW2-31 which was previously identified by Lee Coffey (WIT).
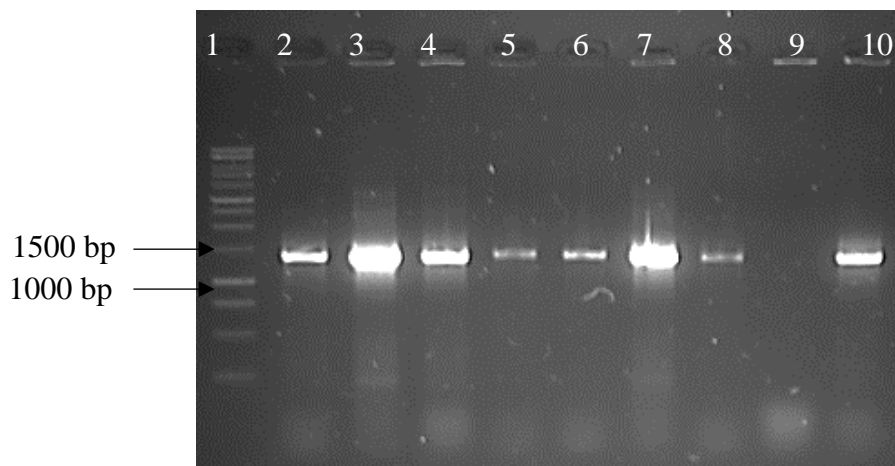
Figure 3.5 Sample electrophoresis gel displaying the amplification products of the 16s rRNA PCR nitrilase containing isolates.

Promega 1 kB ladder (lane 1), unidentified isolate SS4 (lane 2), unidentified isolate SS17 (lane 3), unidentified isolate SS9 (lane 4), unidentified isolate SS10 (lane 5), unidentified isolate SS6 (lane 6), unidentified isolate SS8 (lane 7), unidentified isolate SW2-37 (lane 8), negative control (lane 9), positive control (*R. erythropolis* AJ270) (lane 10).
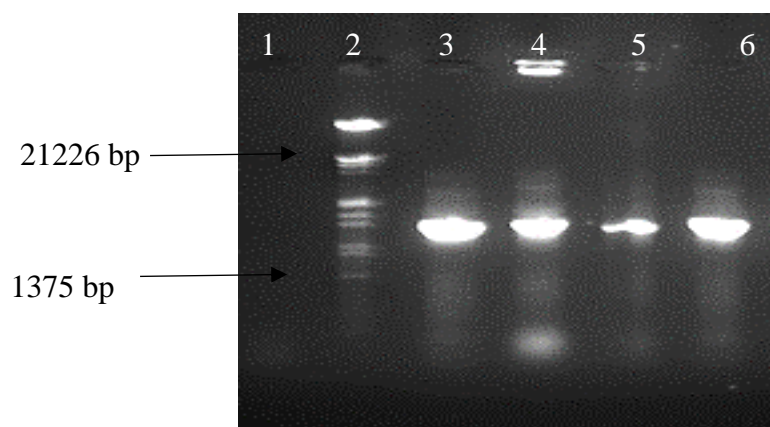
Figure 3.6 Sample electrophoresis gel displaying the amplification products of the 16s rRNA PCR nitrilase containing isolates.

Negative control (lane 1), λ DNA *Eco*R1/*Hin*DIII marker (lane 2), unidentified isolate SS34 (lane 3), unidentified isolate SS33 (lane 4), unidentified isolate SW2-10 (lane 5), positive control (*R. erythropolis* AJ270) (lane 6).

Table 3.5 List of nitrile metabolising environmental isolates and their corresponding nitrilase gene amplified by Lee Coffey (WIT) and analysed in this study.

16s rRNA amplification and sequencing was performed (where accession numbers are noted below) and analysed in this study.

Isolate	Clade	16s rRNA Accession No.	Nitrilase Accession No.
Rhodococcus sp. SS4	1A	MG132107	MG383631
Serratia sp. SS17	1A	MG132112	MG383638
Serratia sp. SS17	1A	MG132112	MG383637
Serratia sp. SS17	1A	MG132112	MG383636
Erwinia sp. SS9	1B	MG132104	MG383630
Serratia sp. SS10	1B	MG132111	MG383634
Rhodococcus sp. SS6	1B	MG132108	MG383628
Rhodococcus sp. SS8	1B	MG132109	MG383629
Serratia sp. SW2-37	1B	MG132113	MG383640
Arthrobacter sp. SS34	2A	MG132103	MG383637
Microbacterium sp. SS33	2A	MG132105	MG383635
Serratia sp. SS10	2A	MG132111	MG383632
Staphylococcus sp. SW2-31	2A	N/A	MG383639
Microbacterium sp. SS33	4A	MG132105	MG383636
Ochrobacterium sp. SW2-10	5A	MG132106	MG383638

3.3.3 Phylogenetic analysis of the previously identified nitrilase sequences

As previously noted, 15 nitrilase partial sequences were identified by the wider research group. The DNA sequences of the nitrilases were compiled and a Blast search was carried out to identify any new nitrilases which were since added to the database. A threshold of 80% DNA similarity was set when selecting sequences from the databases for phylogenetic tree construction. From this the pre-existing novel nitrilases were combined with the positive control sequences supplied by Verenium corporation (now part of BASF) and compared to nitrilase sequences from the database resulting in the phylogenetic trees as seen in Figure 3.7 to Figure 3.11.
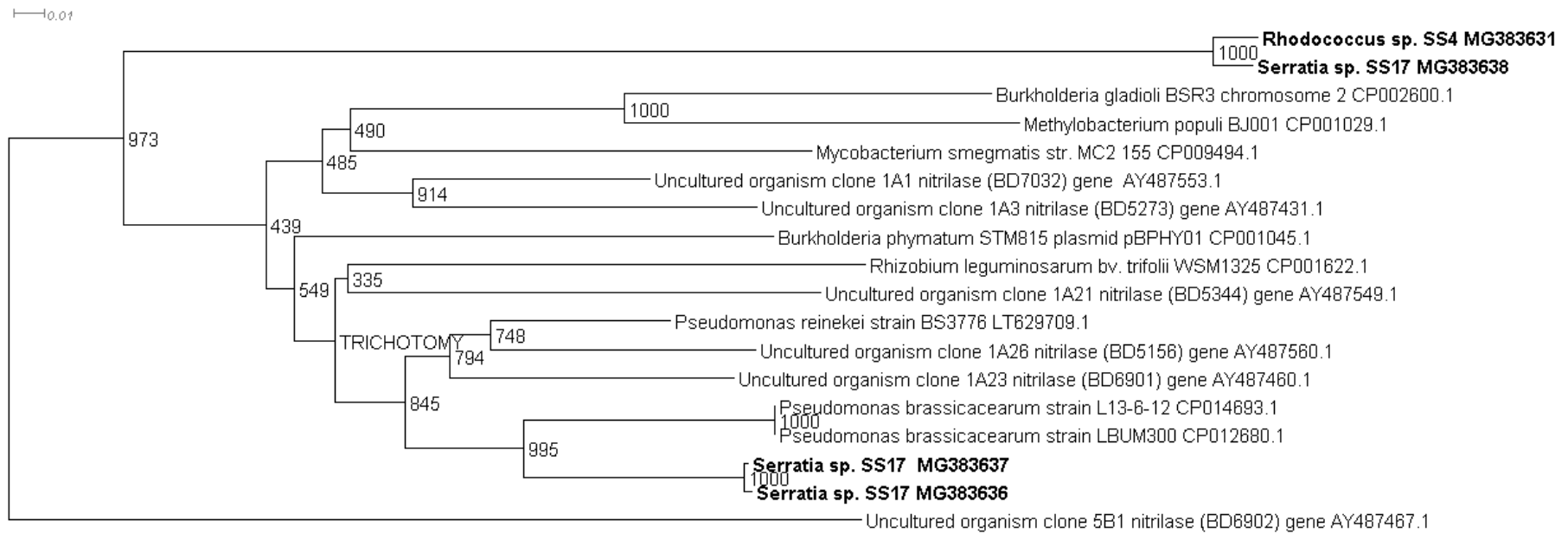
Figure 3.7 Alignment of partial 1A clade nitrilase sequences presented in this study highlighted in bold) with 1A clade nitrilase sequences from the database.
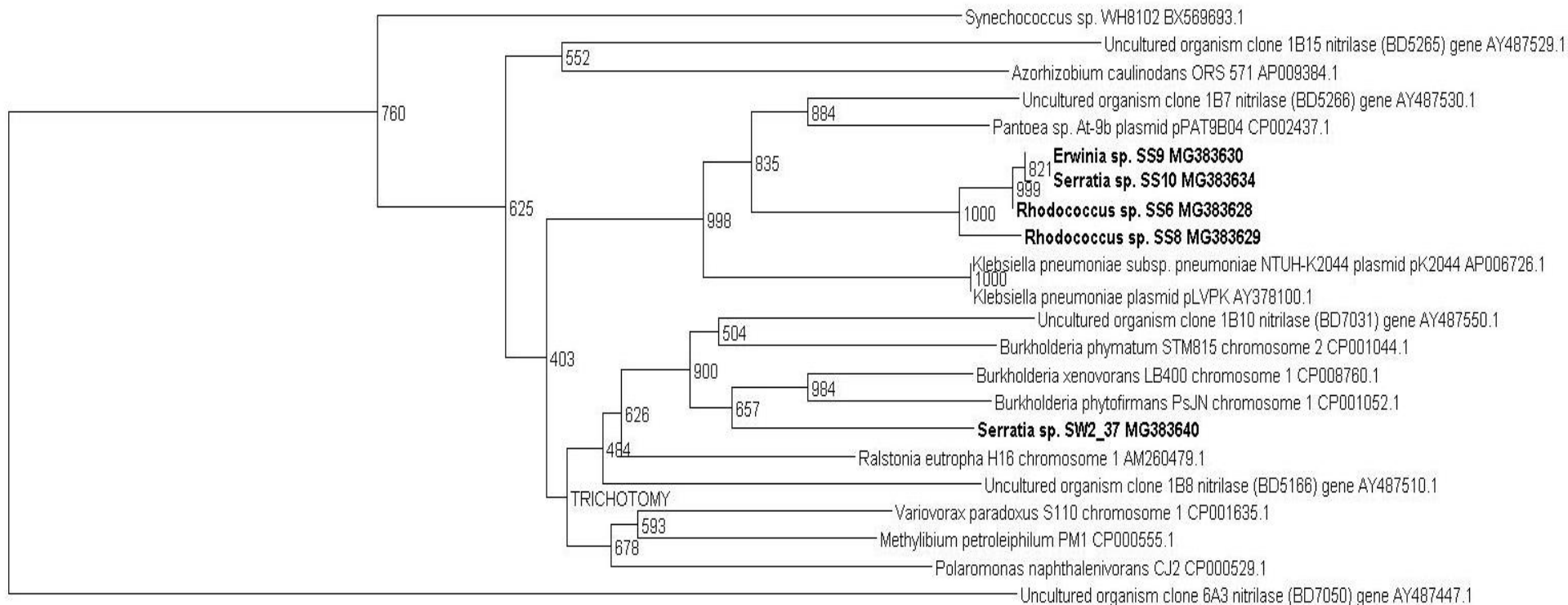
Figure 3.8 Alignment of partial 1B clade nitrilase sequences presented in this study (highlighted in bold) with 1B clade nitrilase sequences from the database.
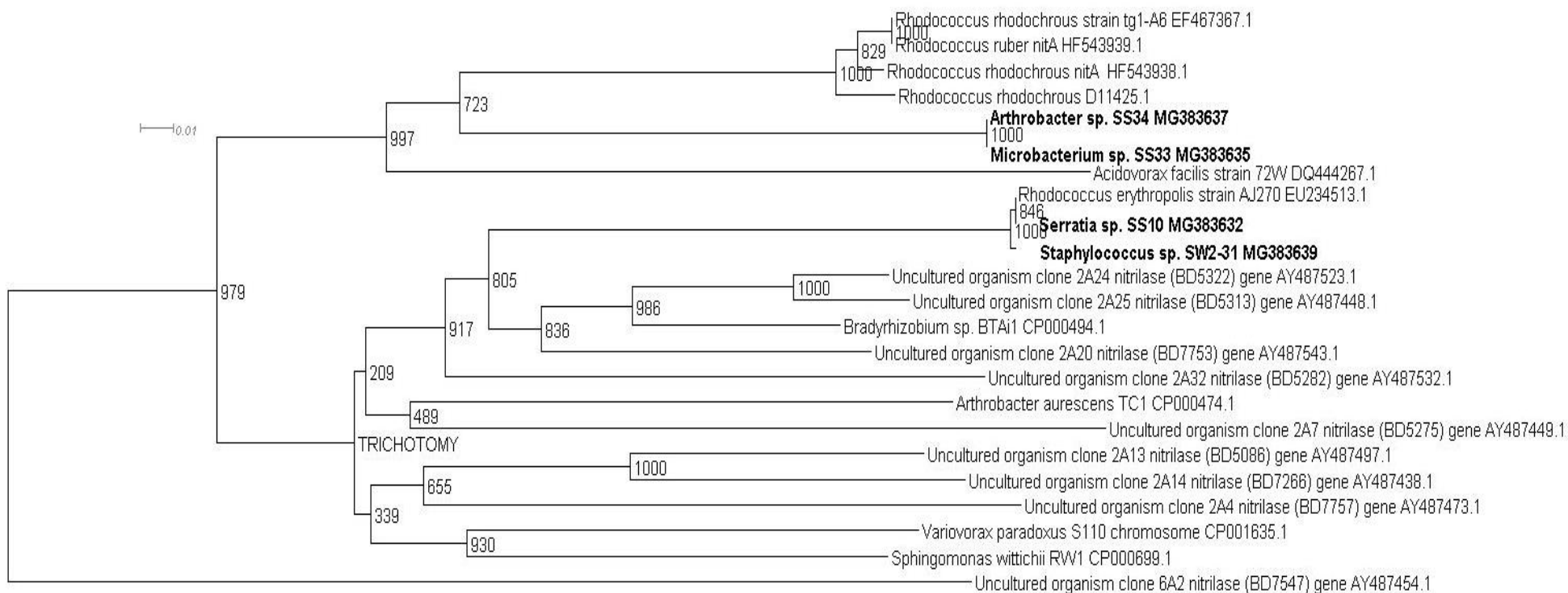
Figure 3.9 Alignment of partial 2A clade nitrilase sequences presented in this study (highlighted in bold) with 2A clade nitrilase sequences from the database.
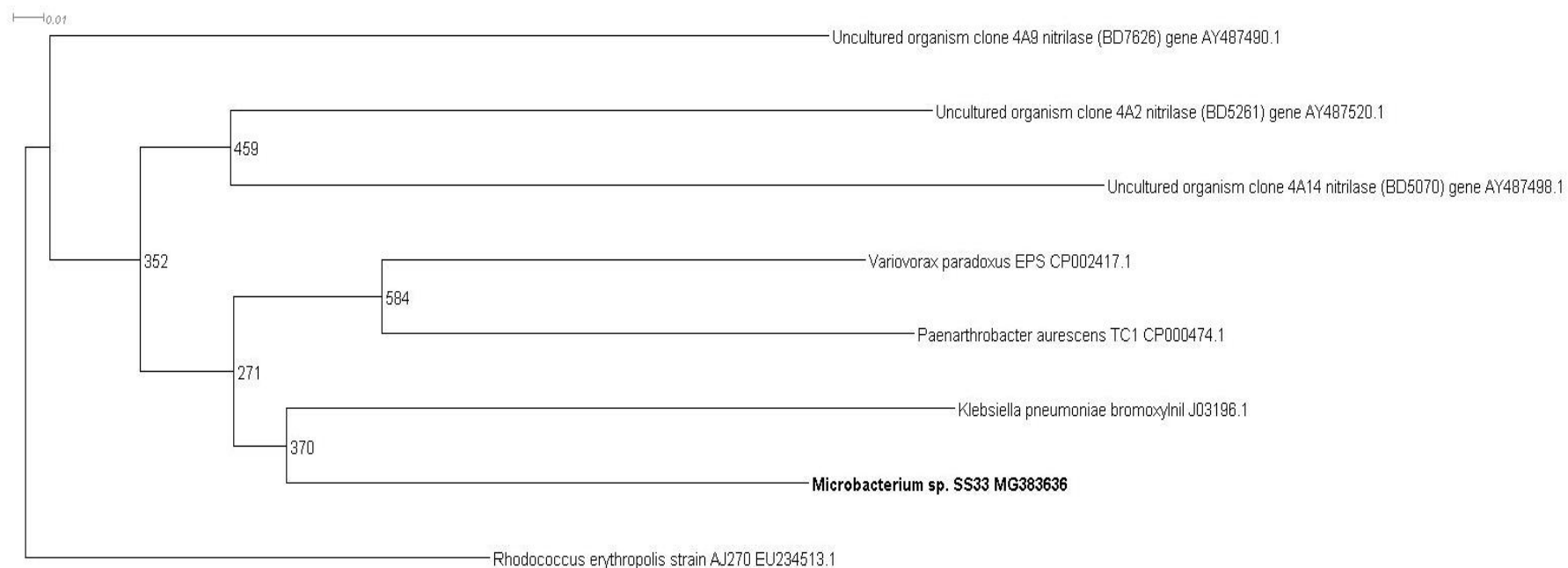
Figure 3.10 Alignment of partial 4A clade nitrilase sequences presented in this study (highlighted in bold) with 4A clade nitrilase sequences from the database.
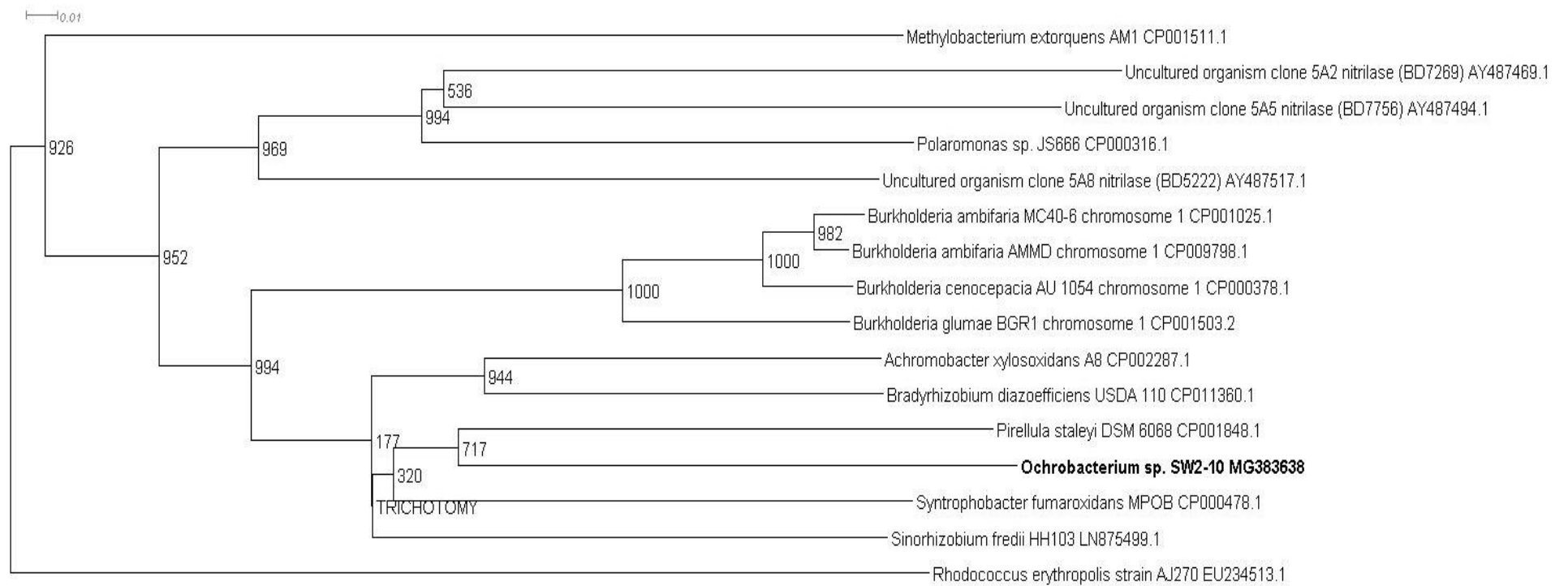
Figure 3.11 Alignment of partial 5A clade nitrilase sequences presented in this study (highlighted in bold) with 5A clade nitrilase sequences from the database.

3.4 Discussion

Enzymes or whole cell biotransformations are heavily utilized to produce biotechnological and pharmaceutical intermediates. The microbial nitrile metabolising pathway is of particular interest, with nitrilase enzymes affording the ability to produce enantiopure carboxylic acids (Martínková *et al.* 2017, Coady *et al.* 2013). As stated previously, Fleming *et al.* (2010) noted that there were over 30-nitrile containing pharmaceuticals on the market and another 11 nitrile-containing drugs were approved since 2011 (Dooley-Cullinane *et al.* 2016). One well known example of chemoenzymatic synthesis relates to the API Atorvastatin, for the cholesterol lowering drug Lipitor® (Solano *et al.* 2012). Nitrilases have also been extensively utilized in synthesis of some key carboxylic acids including p-amino benzoic acid, acrylic acid, nicotinic acid, mandelic acid, glycolic acid and 4-hydroxyphenyl acetic acid etc (Nigam *et al.* 2017, Rustler *et al.* 2007, Kaul *et al.* 2004).

As the primer design was originally carried out by Lee Coffey (WIT) in 2010, it was paramount that attention was given to the level and location of degeneracy afforded, to ensure that the detected gene would not lend to sequence bias, while taking an updated database into consideration. Confirmed nitrilase gene sequences from nitrile metabolising microorganisms and cloned recombinant enzymes available on the NCBI database were utilized during primer design and again in this study for updated database/primer comparisons. The primer degeneracy for the 1A clade was afforded the maximum value observed across the primer pairs to reduce detection bias amongst samples.

The same approach was applied during the primer design for each clade, with differences observed in the number of sequences exploited and the degree of degeneracy. Sequence homologies between the exploited nitrilase genes permitted the identification of conserved regions within the genes, with the degeneracy accounting for the variation evident within each clade.

Previous research by the group identified four 1A clade partial nitrilases, as seen in the phylogram presented in Figure 3.7. Two of these partial genes, identified in isolate *Rhodococcus* sp. SS4 (MG383627) and *Serratia* sp. SS17 (MG383638) displayed high sequence homology, despite being isolated from different soil sources and on different

nitriles, acetonitrile and benzonitrile respectively. They display the highest sequence similarity to the uncultured gene 1A1 (AY 487553.1). The remaining two nitrilases discovered in *Serratia* sp. SS17 (MG383633 and MG383634) displayed the highest sequence homology to uncultured nitrilase 1A23 (AY487460.1) as discovered by Robertson *et al.* (2004), where the majority of enzymes clustered in the 1A clade displayed S-enantioselectivity. Though the above partial nitrilase sequences were isolated on achiral substrates, clade categorisation could suggest that the bacterial cells or nitrilase enzymes may permit enantioselective hydrolysis of a suitable substrate.

Five previously discovered novel partial nitrilase sequences were elucidated across the 1B clade. Four of these partial nitrilase sequences were identified in microorganisms which were isolated from selective enrichments with soil samples and one from a microorganism, *Serratia* sp. SW2-37 (MG383640), which was isolated from a selective enrichment of a *Ceramium codicola* sample. The constructed phylogram can be seen in Figure 3.8. These partial sequences coding the nitrilase genes detected in isolate *Erwinia* sp. SS9 (MG383630) and *Serratia sp*. SS10 (MG383631) displayed the highest homology despite being isolated from different soil sources on alternative nitriles during the enrichments, acetonitrile and adiponitrile respectively, similar to isolate *Rhodococcus* sp. SS8 (MG383629) and *Rhodococcus* sp. SS6 (MG383628) The aforementioned nitrilase displayed the highest sequence homology to the uncultured clone 1B10 (AY 487550.1) (Robertson *et al.* 2004). The majority of the 1B clade, as conferred by Robertson *et al.* (2004), display S-enantioselectivity. This also applies to the uncultured clone 1B8 (AY 487510.1) which displays the highest sequence homology to the partial identified nitrilase sequence from *Serratia* sp. SW2-37 (MG383640).

The 2A clade, as discussed by Robertson *et al.* (2004) was largely R-selective. The four-elucidated partial nitrilase sequences from this current study displayed high sequence homology to each other but also to a number of *Rhodococcus* sp. nitrilase genes. Three of the four partial nitrilases, *Serratia* sp. SS10 (MG383632), *Microbacterium* sp. SS33 (MG383635) and *Arthrobacter* sp. SS34 (MG383637) were isolated from selective enrichments with soil; in contrast, the *Staphylococcus* sp. SW2-31 (MG383639) partial nitrilase sequence isolated from a selective enrichment of *Palmaria palmata* samples. Interestingly, the phylogram for the 2A clade, presented in Figure 3.9, displays the highest number of *Rhodococcus* sp. nitrilase genes. This may be due to sequence bias in previous

methods used to discover these nitrilase genes or it may hypothesise that *Rhodococcus* sp. could in fact harbour a higher ratio of R-selective versus S-selective enzymes, which would be indicated by their clade categorisation. From another perspective, it could signify that there is a bias in the culture conditions from environmental sources which favour *Rhodococcus* sp. or, that they account for a high percentage of the culturable portion the soil samples utilized in studies.

One partial sequence for a 4A novel nitrilase gene was detected by the research group in *Microbacterium* sp. SS34 (MG383636), displaying the highest sequence similarity to *Klebsiella pneumoniase* bromoxynil as presented in the phylogram in Figure 3.10. Published research by Robertson *et al*. (2004) accounted for the detection of 17 novel 4A nitrilases. The comparatively low number of partial nitrilase genes detected in this study is not unexpected as a section of environmental isolates were screened, in comparison to the screening of a metagenomic library. It may be a preliminary indication that the nitrilase genes associated with the 4A clade are not commonly found in the culturable portion of environmental isolates involved in the study. The available literature is suggestive that the majority of enzymes which are categorised into the 4A clade could be S-selective (Robertson *et al.* 2004).

One partial sequence for a 5A novel nitrilase was identified by the research group from seaweed isolate *Ochrobacterium* sp. SW2-10 (MG383638) (refer to phylogenetic tree in Figure 3.11). It displayed the highest sequence homology to *Piruella staleyi* DSM6068 (CP 001848.1) and, based on the findings published by Robertson *et al.* (2004), it could potentially encode an (R)-selective enzyme.

The touchdown PCR method resulted in non-specific product amplification due to the use of degenerate primers and the touch-down PCR approach, resulting in the need for gel extraction as described in Section 3.3.1. The 5A phylogram displays a cluster of *Burkholderia* sp. which may indicate that these predicted R-selective 5A clade nitrilase genes are common amongst this genera, perhaps obtained by horizontal gene transfer, as in the the published findings by Coffey *et al.* (2009). It is not surprising that the ratio of 5A partial nitrilase genes to either the 1A, 1B, or 2A clades was lower. Based on the published findings by Robertson *et al.* (2004) only nine 5A clade nitrilases were

discovered from the metagenomic library of a total of 137 nitrilases, indicating that they may in fact not be common in culturable microorganisms.

For the clades 3A, 5B, and 6A, no phylogenetic analysis is available as no partial sequences for nitrilase enzymes were identified in the subset of screened isolates. This is not surprising as only seven novel nitrilases were identified by Robertson *et al.* (2004) across the 3A and 6A clades, with 3 and 4 clones respectively. While there were 21 clones identified by Robertson *et al.* within the 5B clade, our lack of detection of a 5B clade nitrilase within our subset of isolates may be due to the fact they originate from the non-culturable portion of the environmental sample, or it may be a case that they were simply not present in our samples. It is also worth noting that the seven aforementioned nitrilase genes were identified from the construction of a metagenomic library by Robertson *et al.* (2004). In contrast, the approach used in this study relied on screening environmental cultured isolates which only represented a fraction of the culturable portion of the selected environmental samples. The designed primers and nitrilase screening approach did yield positive amplification of the nitrilase gene from all the positive controls associated with the 3A, 5B, and 6A clades utilised in the study, inferring that the lack of detection is due to sample selection rather than primer and method design.

The touchdown PCR method described allows for the screening of clade-specific novel nitrilase genes. The development of such a protocol allows for the targeted identification of enzymes with a predicted activity based on clade characterisation, as reported by Robertson *et al*. 2004. The novel method described in this study will enable researchers to prioritise samples for further analysis such as full gene sequencing, via primer-walking or bridge PCR, or whole genome sequencing on next generation sequencing platforms. Nitrilases from metagenomic sources, or from wild-type organisms displaying interesting metabolising abilities can be detected, thus omitting the need for whole genome sequencing or for fosmid insert sequencing. Various methods can be employed to obtain the full nitrilase gene sequence once identified via this assay, such as primer-walking, inverse/bridge PCR etc. Indeed, if whole genome sequencing is feasible, at the very least this assay enables the prioritisation of sample sequencing and pre-validates the elucidation of complete nitrilase sequence discovery.

The described body of work allowed for the phylogenetic analysis of 15 clade-specific partial nitrilase genes (previously detected by the research group) and genus identification of the unidentified nitrilase containing microorganisms. This novel touchdown PCR assay will allow for the detection of clade-specific novel nitrilases with predicted enantioselective profiles as determined by their clade classifications. As noted previously, it is estimated that by 2030, the "products of white biotechnology and bioenergy will account for 30 % of industrial production worth €300 billion" (European Union, 2007). The drive toward a 'green chemistry' approach is increasing the demand for the discovery of novel genes and enzymes (Rasor and Voss, 2001). As the body of research increases, the true extent to which these enzymes can be utilized becomes more evident.

3.5 Conclusion

In summary, the novel nitrilase clade-specific PCR protocol resulted in the detection of a 2A nitrilase gene from *R. erythropolis* SET-1. The work describes the screening of 12 putative nitrile metabolising isolates (presented in Chapter 2). The chapter further describes the phylogenetic analysis of previously identified 15 nitrilases (partial sequences) (previously discovered by the research group) against an updated database. 16s rRNA gene identification of these nitrile metabolising isolates was also performed in this study. These newly designed clade-specific PCR assays will allow for the efficient search for nitrilase genes based on clade-classifications enabling the identification and elucidation of novel nitrilase genes with a predicted enantioselectivity profile. The method will prove to be valuable to the ongoing search for novel nitrile metabolising enzymes towards industrial and biopharmaceutical applications.

CHAPTER 4

GENETIC CHARACTERISATION OF *RHODOCOCCUS ERYTHROPOLIS* SET-1- A KNOWN NITRILE METABOLISER

4   GENETIC CHARACTERISATION OF *RHODOCOCCUS ERYTHROPOLIS* SET-1, A KNOWN NITRILE METABOLISER

As discussed earlier, enzymes which are involved in the aldoxime-nitrile metabolising pathway hold great promise for pharmaceutical and industrial applications. Prior to the research presented in Chapter 2, a nitrile-metabolising bank consisting of 265 microorganisms was previously isolated from various environmental sources across several nitriles by the research group. This study involved the expansion of this microbial bank utilizing three nitriles which are of interest to the pharmaceutical industry: 3-hydroxyglutaronitrile (3-HGN), 3-hydroxybutayronitrile (3-HBN), and 3-hydroxy-3-phenylpropionitrile (3-HPPN). The impact of biotransformations on such substrates has been evident with the blockbuster drug Lipitor, where the drug intermediate is produced by the enzymatic desymmetrization of 3-HGN (Patel *et al.* 2009). Increasing the nitrile-metabolising bank enabled a greater microbial/enzyme pool, thus increasing the chance of identifying promising microorganisms/enzymes towards pharmaceutical/industrial applications.

Work within the research group identified a promising nitrile metabolising isolate from the 265 isolate bank preceding this study, namely *R. erythropolis* SET-1 (Coady *et al.* 2013). Whole cell biotransformations with this isolate demonstrated > 99.9 ee% on 3-HBN for the (S)-enantiomer product of 3-hydroxybuytric acid (3-HBA) (Coady *et al.* 2013). As detailed in the publication, no amide intermediate of the substrate 3-HBN was detected during the whole cell reactions and from this it was proposed that the enzyme was a nitrilase, specifically an enantioselective nitrilase.

However, during the previous HPLC analysis of the biotransformations reactions it was observed that both the (R)- and (S)-enantiomer of the substrate 3-HBN were not detectable. Should the nitrilase be enantiospecific and be the sole enzyme responsible for the observed nitrile metabolising activity then it would have been expected that the (R)-enantiomer of 3-HBN would be observed. The biotransformation also never displayed a yield greater than 42% for the (S)-enantiomer, which suggested that the remaining (S)-nitrile was not being utilised by the nitrilase to produce the acid as (i) it enantioselectively produced the (S)-enantiomer of 3HBA and (ii) the yield never exceeded 50% for (S)-

3HBA. From this the group proposed that the nitrile substrate may have also been utilised by other enzymes within the cell as it was a whole cell biotransformation.

Following this body of work, additional research on *R. erythropolis* SET-1 was carried out (Coady *et al.* 2015). Whole cell biotransformations with *R. erythropolis* SET-1 displayed amide formation in cases where the β-hydroxy nitriles possessed a vinyl group at the α position to the β-hydroxy nitrile. It was proposed that the amide formation which was observed with a variety of structural analogues of 3-hydroxybutyronitrile could be due to the presence of a NHase/amidase cluster or perhaps a nitrilase which may also demonstrate amide formation such as the nitrilases from *Rhodococcus.* sp. ATCC 39484 (Stevenson *et al.* 1992) *A. thaliana* AtNit1 (Osswald *et al.* 2002), and the nitrilase from *A. niger* K10 (Kaplan *et al.* 2006) as earlier discussed in Section 1.5.3.

To fully realise the industrial potential of the nitrile metabolising activity the proposed nitrile metabolising gene(s) would have to be elucidated, cloned, and expressed. This chapter describes the steps which were taken towards achieving this.

4.1 Research Hypothesis and Aims

Hypothesis:

A confirmed nitrile metabolising isolate *R. erythropolis* SET-1 contains a nitrile metabolising gene(s) responsible for the enantioselective hydrolysis of industrially informed nitriles such as 3-HGN, 3HBN and 3-HPPN.

Aim:

The aim of this research was to elucidate the nitrile metabolsing gene(s) in *R. erythropolis* SET-1 by using a combination of both fosmid insert sequencing and whole genome sequencing and subsequent analysis.

4.2   Materials and Methods

4.2.1   Genomic DNA extractions

The Anachem Keyprep nucleic acid extraction kit (Anachem, Cat. No. AM111113) was used to extract genomic DNA from all microorganisms, as per the manufacturer's guidelines with DNA eluted in sterile deionised water.

4.2.2   Circular plasmid DNA extractions

The Zymo research Zyppy™ plasmid miniprep kit (Cambridge Bioscience, Cat. No. D4019) was used to extract circular DNA as per the manufacturer's guidelines with DNA eluted in sterile deionised water.

4.2.3   Pulsed field gel electrophoresis

Pulsed field gel electrophoresis (PFGE) was carried out using a CHEF (Contour-clamped Homogenous Electric Field) Drive II system (BioRad) with Yeast chromosome marker (NEB Cat. No. N0345S), mid-range 1 (NEB Cat. No. N3551S) and the low-range marker (NEB Cat No. N0350S) as size standards and Agarose, SeaKem® Gold (VWR, Cat. No. 733-1552). 10 X TBE buffer stock solution (Biorad Cat. No. 1610770) was diluted to a 0.5 X TBE working solution. *R. erythropolis* SET-1 genomic samples were run under two migration profiles: (i) 30 to 60 s switch time at 6V for 22 hours and (ii) 40 to 80 s switch time at 6V for 22 hours.

4.2.3.1   Plug preparation

R. erythropolis SET-1 cell stocks were removed from -70 °C storage, thawed on ice and glycerol removed by centrifuging at 10,000 rpm followed by the washing of cells with 0.85 % (sodium chloride) NaCl solution. The final $O.D._{600nm}$ was adjusted to 2.0 by resuspension in TE buffer @ pH 8.0. A 1 % agarose solution prepared in 0.5% TBE buffer

and added to the cell solution, and the plug mould was filled. 600 μl of the cell/agarose mixture was dispensed into wells of a reusable plug mould (BioRad) and was allowed to cool for 15 mins at 4 °C. The agarose plugs were gently transferred from the wells and into 25 ml sterile bottles containing 0.5% TBE buffer.

4.2.3.2   Cell lysis

The cell lysis solution was prepared as follows: 10 mM Tris @ pH 7.2, 50 mM NaCl, 0.2% (w/v) sodium deoxycholate (Sigma Cat No. D6750), 0.5 % (w/v) N-Lauroylsarcosine sodium salt solution (Fluka Biochemika, Cat No. 61747). The above solution was autoclaved at 121 °C for 15 minutes. After autoclaving 0.8 % (w/v) Lysozyme (Sigma-Aldrich, Cat No. L6876) and 10 U mutanolysin (Sigma, Cat. No. M4782) were filter sterilised and added. The plugs were incubated in this mixture for 2 hours at 37 °C in a waterbath (Grant SUB Aqua 26).

4.2.3.3   Proteinase K treatment of lysed *R. erythropolis* SET-1 embedded plugs

The Proteinase K solution was composed of: 7 % (w/v) of Sodium dodecyl sulfate (SDS) (Sigma-Aldrich, Cat. No. L3771), 0.2 % (w/v) of Sodium deoxycholate (Sigma-Aldrich, Cat. No. D6750), 0.005 % (w/v) Proteinase K (Sigma-Aldrich, Cat. No. P2308) and 20 % (v/v) of 500 mM EDTA. The plugs were incubated for 2 days at 55 °C in a water bath (Grant SUB Aqua 26).

4.2.3.4   Plug storage

After cell lysis and proteinase K treatment the plugs were washed three times for 1-hour intervals in TE buffer pH 8.0 and then stored at 4 °C.

4.2.4   S1 nuclease digestion of *R. erythropolis* SET-1 embedded clones

S1 nuclease digestion has been utilized to degrade circular plasmids or megaplasmids to their linearized form (Rose and Fetzner 2006, Overhage *et al.* 2005, Barton *et al.* 1995) *R. erythropolis* SET-1 embedded agarose plugs were subjected to S1 nuclease (Sigma-Aldrich, Cat. No. 5661) digestion as per Overhage *et al.* (2005). Post cell lysis and

proteinase K treatment, a 1 mm slice of the *R. erythropolis* SET-1 embedded plugs were washed twice in S1 nuclease buffer (30 mM sodium acetate (Sigma-Aldrich, Cat. No. S2889), 50 mM NaCl (Sigma-Aldrich, Cat. No. S3014), 1 mM $ZnCl_2$ (Fisher Scientific, Cat. No. 10172830), 50% (v/v) glycerol (Fisher Scientific, Cat. No. G33-1) and incubated with 50 Units S1 nuclease per mL for 1 hour at 37 °C. The reaction was stopped by adding 10 mM EDTA (Fisher Scientific, Cat. No. 10618973) and cooling to 4 °C. All control experiments were performed without S1 nuclease.

4.2.5   Fosmid library construction

A DNA fosmid library of SET-1 was prepared using the Lucigen Copyright® V2.0 fosmid library production kit as per the manufacturers guidelines with the exception of omitting the size selection of end-repair DNA. The packaging and subsequent transfection of the pSMART® FOS fosmid vector (Figure 4.1) was carried out using the Stratagene gigapack III XL packaging extracts (Agilent Cat No. 200207) as per the Lucigen Copyright® V2.0 fosmid library production kit packaging guidelines using the host strain *E. coli* replicator FOS. Transformants were selected via blue/white screening as per manufacturers guidelines.
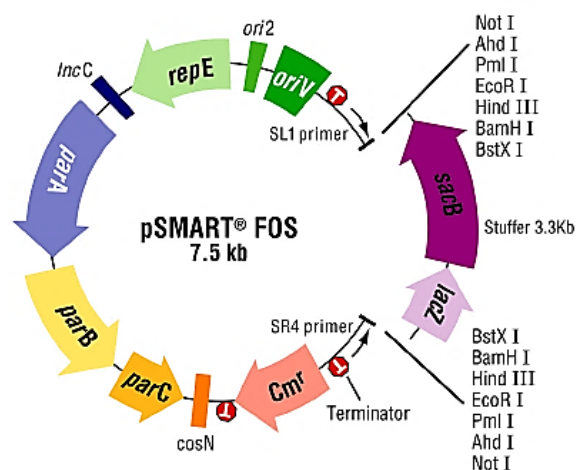
Figure 4.1 pSMART® FOS vector used for the creation of the *R. erythropolis* SET-1 genomic DNA (Lucigen Corporation, 2018.)

4.2.5.1 Growth and maintenance of fosmid clones

Fosmid clones were grown in TB media containing 12.5 µg/ mL chloramphenicol. TB media was prepared as follows (1.18% (w/v) Bacto-tryptone (Oxoid, Cat. No. LP0042), 2.36% (w/v) yeast extract (Fisher Scientific, Cat. No. BP9727), 0.94% (w/v) dipotassium hydrogen phosphate (Merck millipore, Cat. No. M1051090100), 0.22% (w/v) potassium dihydrogen phosphate (Merck Millipore, Cat. No. M1051080500). Post-autoclaving of the TB media 0.8 mL of 50% (v/v) glycerol (Fisher Scientific, Cat. No. G33-1) was added per 100 ml to a final concentration of 0.04%. 1,920 clones were stored separately in a 96 well format and 66,080 clones were stored combined in a single tube format. All clones were stored in TB broth containing 12.5 µg/ mL chloramphenicol with glycerol to a final concentration of 20% at -70 °C. For induction, fosmid clones were grown overnight in TB medium plus 12.5 µg/ mL chloramphenicol in the presence of 1 X Copycontrol™ fosmid auto-induction solution (Epicenter, Cat No. AIS10F7) at 37 °C and 225 rpm.

4.2.5.2 Functional screening of fosmid clones

The fosmid library was screened for potential functional clones using M9 minimal media with 5 mM 3-HGN or 3-HBN as the sole nitrogen source and Glucose (0.4% (v/v final concentration) as the sole carbon source. An initial screening stage involved the streaking of 1,920 clones over a period of 10 weeks. This was followed by a 'bulk screening' method which involved transfecting the replicator FOS cells with the remaining phage particles and subsequent auto-induction in TB broth with 12.5 µL/mL chloramphenicol and auto-induction solution. A pre-determined dilution (titering the transfected cells) of the auto-induced 'bulk-clone' library was spread plated with functional clones identified over background growth.

4.2.5.3 Fosmid DNA extraction

Fosmid DNA was extracted from *R. erythropolis* SET-1 fosmid clones using the FosmidMAX™ DNA Purification Kit (Lucigen, FMAX046 in 1.5 mL format as per the manufacturer's guidelines with the exception of solubilisation of fosmid DNA in sterile water instead of elution buffer.

4.2.6 Restriction enzyme digestion of fosmid DNA extractions

Restriction enzyme digestions were carried out on any potential functional clones to eliminate duplicates by comparison of the restriction profiles. The 10 µL restriction digest reaction was as follows: X µL (150 ng) DNA, 20U Enzyme, 1 µL 10 X restriction digest buffer, Y µL sterile deionised water. All restriction enzymes used were supplied by Sigma-Aldrich or New England Biolabs (Table 4.1).

Table 4.1 Restriction enzymes used during the restriction digestion of functional clones with their corresponding supplier and Cat. No.

Enzyme	Supplier	Cat. No.
EcoR1	Sigma-Aldrich	R6265
HinDIII	Sigma-Aldrich	R1137
BamH1	Sigma-Aldrich	R0260
S1 Nuclease	NEB	M5761
NOT1	NEB	R0189S
XBAI	NEB	R0145

4.2.7 Nessler's microscale ammonia assay

Induced functional clones were initially screened for activity towards 3-HGN and 3-HBN using the Nessler's colorimetric assay as per Coady *et al.* (2013). The biotransformations were carried out in 150 µL format containing potassium phosphate buffer pH 7, a final cell O.D.$_{600nm}$ of 0.5 and a final substrate concentration of 10 mM. The biotransformation was carried out over a 24 hr, 48 hr, 72 hr and 168 hr period. To quench the biotransformation 37.5 µl of 250 mM HCl (Sigma-Aldrich, Cat. No. 435570) was added to stop the reaction. Cell biomass was removed at 500 x g for 10 minutes at 4 °C. 20 µL of the quenched reaction supernatant was transferred to a microtiter plate and to this 181 µL of the Nessler's master mix was added (151 µL deionised H$_2$O, 1.0 µL 10N NaOH (Sigma-Aldrich, Cat. No. 765429) and 25 µL Nessler's reagent (Sigma-Aldrich, 72190). The reaction was incubated at room temperature (22 °C) for 15 minutes and the absorbance was read at 425 nm.

4.2.8 Endpoint PCR

All primers were synthesised and supplied by MWG Biotech Ltd. or Sigma-Aldrich Ltd. Each 15 µL reaction mixture contained 50-100 ng genomic DNA or 15 ng of metagenomic fosmid DNA, 7.5 µL of 2X GO-Taq green mastermix (Promega, Cat. No. M7501), 1 µL of 25 µM PCR mix (final concentration of 1.6 µM) and 5.5 µL of sterile water. For colony PCR, cells were resuspended in sterile water to an $O.D.600_{nm}$ of 1.0, 1 µL was added to the reaction (final cell $O.D\ 600_{nm} = 0.06$). An ABI Veriti® thermocycler was used for all endpoint PCR reactions.

4.2.8.1 Nitrilase, nitrile hydratase and amidase PCR assays

Gene screening of the functional clones for the detection of Fe-type nitrile hydratase and amidase gene(s) was carried out using primers and thermocycler conditions (ABI Veriti®) as per Coffey *et al*. (2010). The nitrilase clade PCR gene (presented in Chapter 3), NHase genes α- and β- subunits and amidase genes were screened for via endpoint PCR using primer pairs presented in Table 4.2. The Nitrilase clade specific PCR was performed as per Section 3.2.3.

The NHase subunits were screened for as follows; primer pair NHase-αF and NHase-αR for the α-subunit and primer pair NH-βF (ATGGATGGAGTACACGATCT) and NH-βR (TCAGGCCGCAGGTCGAGGT) as per Soares Bragança *et al*. (2017) and Coffey *et al*. (2010). An amplicon of approximately 600 bp was the expected size for the NHase α and β subunits. For the NHase gene screening the thermocycler conditions were as follows: 95 °C for 5 min (1 cycle), 95 °C for 1 min, 56 °C for 1 min, 72°C for 40 s (30 cycles), 72 °C for 5 min (1 cycle) and a hold of 4 °C.

The amidase gene was screened for with primer pair AMD1-F and AMD1-R. The thermocycler conditions were as follows: 95 °C for 5 min (1 cycle), 95 °C for 1 min, 56°C for 2 min, 72°C for 40 s (30 cycles), 72°C for 5 min (1 cycle) and a hold of 4 °C as per Soares Bragança *et al.* (2017). The expected amplicon size for the amidase gene was approximately 1780 bp.

Primer pair TMDCNit-AmpF and TMDCNit-ampR were used for the detection of the TMDC-Nit, a partial nitrilase detected from FC1 of the *R. erythropolis* SET-1 fosmid library. The thermocycler conditions for TMDC-Nit were as follows: 95 °C for 5 min (1 cycle), 95 °C for 1 min, 57 °C for 30 seconds, 72°C for 1 min (30 cycles), 72 °C for 5 min (1 cycle) and a hold of 4 °C.

Table 4.2 Primer pairs utilized for the attempted amplification of NHase, amidase and a nitrilase specific gene from *R. erythropolis* SET-1.

Target Amplicon	Primer Sequence (5'-3')
TMDCNIT-Amp-F	TACTCGATGTACGCCAAGGG
TMDCNIT-Amp-R	CGATCTCCGCGTAGAGGATG
NHase-αF	ATGTCAGTAACGATCGACCAC
NHase-αR	(AGGCAGTCCTTGGTG ACGAT)
AMD1-F	ATACGCGTGAATTCGTGGCGACAAT-CCGACCTGAC
AMD1-R	GGTGTTGAGTCGGAGTGGATCTTCGAAACT-TCCTAG

4.2.9 Long Range PCR

The Qiagen Long range PCR kit (Qiagen, Cat. No. 206401) was utilized for the attempted amplification of a 10 kb, 20 kb and 40 kb (maximum potential fosmid insert size) fragment within a fosmid insert (approximately one quarter, half, and also the complete insert respectively) (Table 4.3).The PCR reactions and thermocycler conditions were as per the manufacturer's guidelines with the exception of reducing the reaction volume from 50 µl to 10 µL. Each template was assessed at 1 ng, 10 ng and 100 ng of fosmid DNA per 10 µL reaction for each primer pair.

Table 4.3 Primer pairs utilized for the attempted amplification of a 10 kB, 20 kB and, 40 kB using the Qiagen long range PCR kit, with target amplicons noted, see Figure 4.21 for proposed primer binding sites.

Primer 1	Primer 2	Target Amplicon
SL1	SR4	Fosmid vector, (entire insert)
SL1-62	SR5-60	Fosmid Vector, (entire insert)
SL1	TMDCNIT-amp F	Fosmid vector to mid-insert[1]
SL1	TMDCNIT- amp R	Fosmid vector to mid-insert[1]
SR4	TMDCNIT-amp F	Fosmid vector to mid- insert[1]
SR4	TMDCNIT- amp R	Fosmid vector to mid- insert[1]

[1] the location and orientation of the known partial nitrilase sequence is unknown; the 'mid-insert' refers to the middle of the known nitrilase partial sequence.

4.2.10 Cloning of PCR products and transformation of ligated DNA

PCR products were cloned using the pGEM®-T easy vector system I (Promega, Cat No. A1360) or the Qiagen PCR cloning kit (Qiagen, Cat No. 231122), as per the manufacturer's guidelines. Ligated products (4 µL) were transformed into competent cells (100 µL) prepared by the Zymo Research Mix & Go™ *E. coli* Transformation Kit (Cat. No. T3001) and mixed gently. The reaction mixture was then stored on ice for 30 seconds before plating onto pre-warmed LB plates (37 °C) containing 100 µg/mL ampicillin (Fisher Scientific, Cat. No. BP1760). The competent cells and transformation protocol were as per the manufacturer's instructions which involved the omission of heat-shock step during the transformation process.

4.2.11 Gel extractions

PCR products that required purification were purified using the QIAquick Gel Extraction Kit (Qiagen Cat. No. 28704) or the Promega Wizard® SV Gel and PCR Clean-up System (Cat. No. A9281) as per the manufacturer's instructions, with the exception of sterile distilled water being used in place of elution buffer.

4.2.12 Ethidium bromide staining of gels post-electrophoresis

Gel electrophoresis was carried out as per Section 2.3.6 with the exception of agarose gels being stained post gel electrophoresis after overnight runs. In this instance all gels were stained in 100 mL deionised water containing 10 μL of 10 mg/ mL Ethidium Bromide (Sigma Cat. No. E1510) for 1 hour on a platform shaker at 10 rev/min (Stuart Scientific). Gels were then subjected to de-staining twice with 100 mL deionised water for 1 hour under the same conditions.

4.2.13 DNA sequencing of PCR products

DNA sequencing was performed in W.I.T. using an ABI Prism 310 Genetic Analyzer as per the manufacturer's instructions. Sequencing extension reactions were performed using BigDye® V3.1 terminator chemistry as per the manufacturer's guidelines. Commercial DNA sequencing was performed by Source Bioscience and GATC Biotech.

Table 4.4 A list of sequencing primers used for attempted elucidation of the novel nitrilase sequence from SET-1 by direct fosmid sequencing. Figure 4.21 displays the proposed primer binding sites.

Primer Set	Primer Sequence	Target Sequence
TMDCNIT- seq F	GGCTGTACGTGATCGGCG	nitrilase
TMDCNIT- seq R	CGCGACATCCAGTCGTCA	nitrilase
SL1	CAGTCCAGTTACGCTGGAGTC	Insert end sequence
SR4	TTTAGTATGGTTGTACC	Insert end sequence
SL1-62	CAGTCCAGTTACGCTGGAGT	Insert end sequence
SR5-60	GCGACTTCAAGTCACGTGAA	Insert end sequence

4.2.14 Biotransformation with amidase inhibitor Diethyl Phosphoramidate (DEPA)

Biotransformations with whole cells of *R. erythropolis* SET-1 (O.D.600$_{nm}$= 1.0) were carried out in a 6 mL volume in potassium phosphate buffer (0.1M) pH 7.5 containing 10 mM DEPA (Sigma-Aldrich, Cat. No. 384011) and 3-HBN at 10 mM. A negative control assay was also performed as above with DEPA omitted from the reaction. The biotransformations were carried out at 25 °C, 150 rpm for 24 hours. The reactions were quenched with the addition of HCL until a pH of 2.0 was achieved. Cell biomass was removed via centrifugation at 15,000 rpm for 2 minutes at 4 °C. The supernatant was derivatized as per Coady *et al.* (2014) (Figure 4.2).
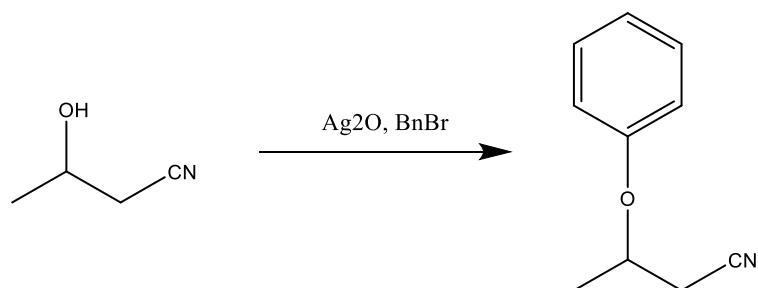
Figure 4.2 Benzylation reaction schematic of 3-HBN achieved by the reaction of 3-HBN in DCM (2 mL) with benzyl bromide (4 equiv) and Ag$_2$O (1 equiv) at room temperature (°C) for 24 hours with no UV exposure as per Coady, (2014). The benzylation of the biotransformation reaction results in 3 main products as displayed in Figure 4.3.
.
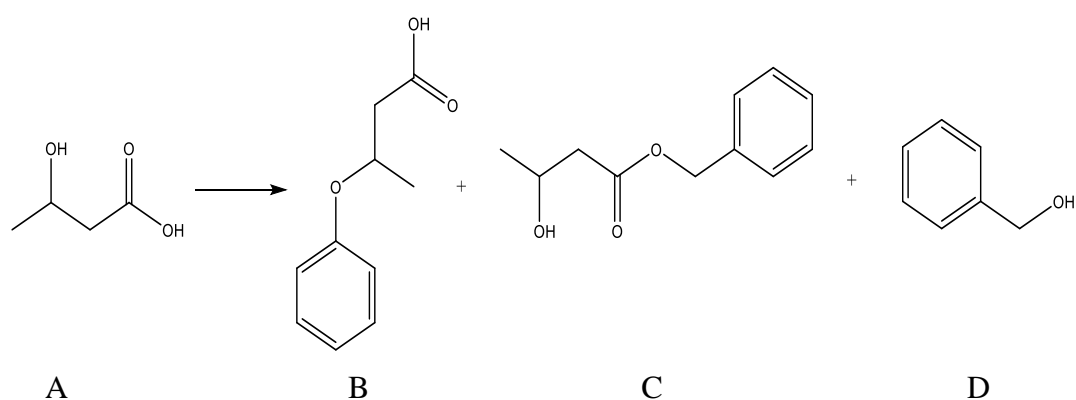
Figure 4.3 Three main benzylation products achieved from the benzylation reaction of 3-HBN biotransformations. (A) 3-HBA, (B) 3-Benzyloxybutyric acid, (C) benzyl-3-hydroxybutonate by-product and (D) Benzyl bromide, adapted from Coady (2014).

119

4.2.15 Biotransformation reaction analysis by HPLC

To assess the biotransformation reactions for the production of 3-HBA and remaining 3-HBN, the sample biotransformation reactions were subjected to NP-HPLC analysis as per Coady (2015) and Coady *et al.* (2014). Biotransformation reactions were benzylated as per Figure 4.2. Post benzylation the sample was filtered and evaporated with subsequent dilution in Hexane:IPA (90:10). The diluted sample was injected onto a ChiralPAK® A-DH (Daicel, Cat. No. 19324) column (150 mm X 4.6 mm) with a particle size of 5 µm and a mobile phase composition Hexane:IPA (90:10) and 0.1 % TFA to assess for acid production. The run conditions were set to a flow rate of 0.8 mL/min and detection set to 215 nm. For remaining nitrile detection, the sample was injected onto a ChiralPAK® O-JH (Daicel, Cat. No. 17324) 150 mm X 4.6 mm with a particle size of 5 µm. The mobile phase consisted of 90:10 Hexane:IPA with no TFA and a flow rate of 0.8 mL/min and detection at 215 nm.

4.3 Results

4.3.1 Genetic characterisation of *R. erythropolis* SET-1

4.3.1.1 Plasmid Characterisation

The potential presence of large plasmids in SET-1 was assessed via pulsed field gel electrophoresis (PFGE) as per Kulakov *et al.* (2005) to ensure that there were no circular plasmid(s) contained within SET-1, as if they were present they would not be included in the genomic DNA extraction and in turn omitted from the subsequent genomic library. A plasmid was identified that migrated according to a linear size of 555 kB. To ensure that the plasmid was linear, the *R. erythropolis* SET-1 embedded plug was digested with S1 nuclease. The plasmid measured at 555 kB pre- and post-digestion confirming it was a linear plasmid (Figure 4.4 (A)). In addition to this the plasmid migrated to 555 kB under two different PFGE conditions also confirming it was linear as per Rose & Fetzner (2006). Figure 4.4 (B) displays an *Eco*R1 digested and undigested plasmid extraction product. From the resulting gel it can be seen that the undigested DNA does not represent the typical banding pattern typically displayed by plasmid DNA. Post-digestion it would be

expected to observe numerous bands should this DNA represent a plasmid; however, this was not the case. The smeared DNA profile evident in Figure 4.4 (B) was more representative of carry over chromosomal DNA.
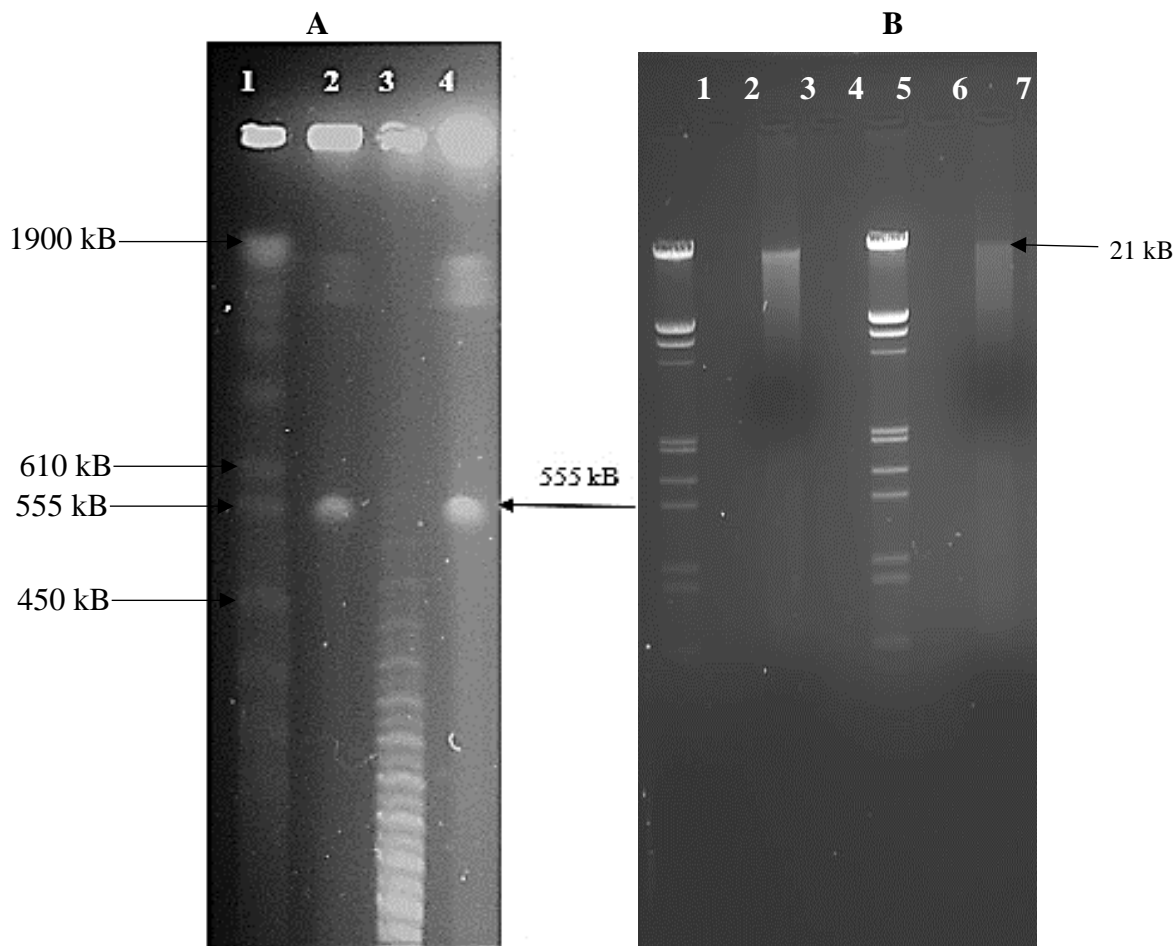
Figure 4.4 (A) Plasmid characterisation of SET-1 via Pulsed Field Gel Electrophoresis. Lane 1, Yeast chromosome marker (NEB, Cat. No. N0345), Lane 2 *R. erythropolis* SET-1 pre-S1 nuclease digestion; Lane 3, Mid-range marker (NEB, Cat. No. N3551), lane 4 *R. erythropolis* SET-1 post- S1 nuclease digestion. (B) Lane 1, λ DNA *Eco*R1/HinDIII (Promega, Cat. No. G1731), Lane 3 Zyppy™ plasmid extraction from *R. erythropolis* 3HBN induced cells pre EcoR1 digestion, Lane 5, λ DNA *Eco*R1/*HinD*III (Promega, Cat. No. G1731), Lane 7 Zyppy™ plasmid extraction from 3HBN induced *R. erythropolis* SET-1 cells post *Eco*R1 digestion.

4.3.2   Fosmid library construction

A fosmid library was created using *R. erythropolis* SET-1 genomic DNA (Figure 4.5) to allow for the identification of the novel nitrile metabolising gene(s) in SET-1 by functional selection, as attempts undertaken were unsuccessful. As the detected plasmid

in SET-1 was shown to be linear in Section 4.3.1.1 it would be included in the genomic DNA extraction and in turn in the genomic library.
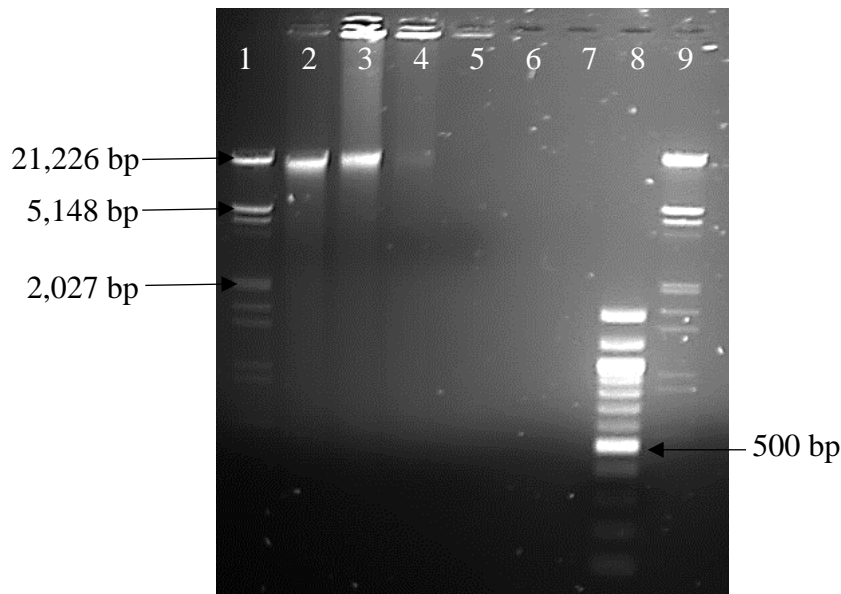
Figure 4.5 *R. erythropolis* SET-1 genomic DNA gel analysis prior to fosmid library construction. Lane 1, λ DNA *Eco*R1/*Hin*DIII (NEB Cat No. G1731), lane 2, *R. erythropolis* SET-1 genomic DNA prep 1, lane 3, *R. erythropolis* SET-1 genomic DNA prep 2, lane 8 100 bp DNA marker (NEB Cat No. N3231L), lane 9 λ DNA *Eco*R1/*Hin*DIII (NEB Cat No. G1731).

Following the gel analysis of *R. erythropolis* SET-1 genomic DNA the DNA was end-repaired prior to ligation to the pSMART® FOS vector. Following successful ligation, the vector was transfected into the Stratagene packaging extracts as per Section 4.2.5. The packaged phage preparation was titred and estimated to contain enough phage/DNA to transfect/produce 68,000 clones.

4.3.3 Restriction enzyme digestion of fosmid DNA preparations to identify successful DNA ligation and transfection

To ensure the fosmid clones selected contained inserts from the *R. erythropolis* SET-1 genome, restriction digestions were carried out on 9 fosmid clones (randomly selected). The fosmid clones were subjected to a double digest with *EcoR1* and *HinDIII*. From Figure 4.8, it is evident that each fosmid clone contained an insert from the SET-1 genome as the vector size is 7.5 kB and all restriction digestions contained an estimated combined vector and insert size of up to 52 kB.
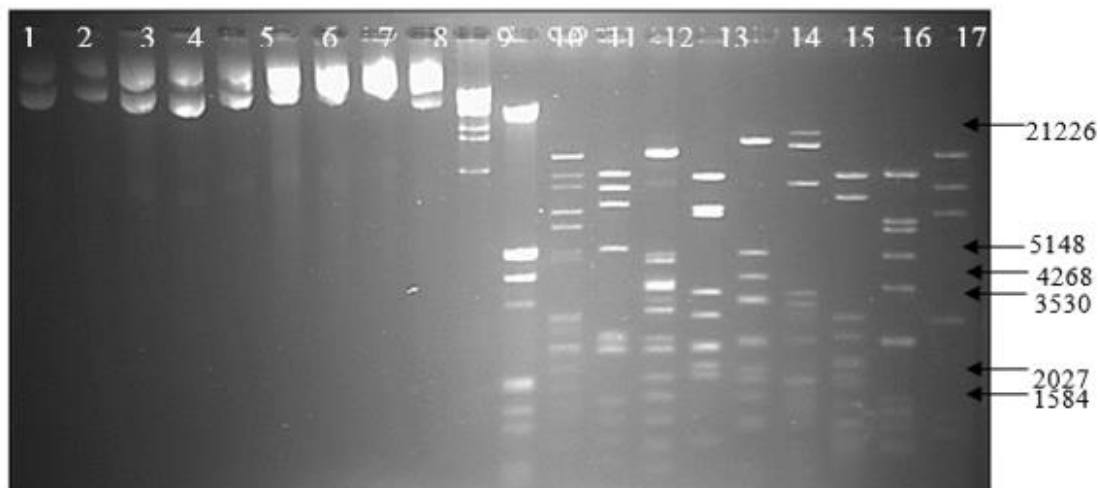
Figure 4.6 Restriction digestion of nine fosmid clones to confirm successful ligation of DNA and transfection of cells.

Lanes 1-9, fosmid extractions 1-9 undigested; lane 10, λ monocot mix marker; lane 11, *Eco*R1/*Hin*DIII marker; lanes 12-20 digested fosmid extractions 1-9 respectively.

The vector map in Figure 4.7 displays multiple restriction sites flanking the stuffer fragment. The vector is supplied pre-linearized and dephosphorylated at the *pmlI* site. While performing the restriction digestions of successfully transformed fosmid clones with *Eco*R1, *Hin*DIII and *Bam*HI it is notable that a 7.5 kB band (representative of the linearized vector) will not be evident as the restriction sites of the three aforementioned enzymes are removed during vector preparation.

Figure 4.7 Vector sequence map of the pSMART® FOS vector.

4.3.4 Functional screening

Initial and bulk screening approaches were assumed as discussed in Section 4.2.5.2.

<u>Initial clone selection</u>

1,920 successful transformants were initially selected for the functional screening process in a 96-well format (Figure 4.8). These clones were streaked weekly over a period of 10 weeks identifying two functional clones. The ability of the clones to grow on the M9 minimal media containing the nitrile (3HBN or 3HGN) as the sole nitrogen source was thought to be due to the low-level leaked expression of the genes assuming the required promoters and regulators were also cloned within the ~ 40 kB insert. Based on the formula provided in the Lucigen copyright manual, it is estimated that 822 clones are required to cover an average *Rhodococcus* genome, from this it was estimated that the 1,920 clones correspond to approximately two times genome coverage.

Functional clone 1 (FC1) identified via growth demonstrated on minimal media with nitrile of interest.

Figure 4.8 Sample image of round 10 functional screening in 96 sample format on M9 minimal media containing 5 mM nitrile, functional clone one (FC1) identified.

Bulk library screening

The remaining fosmid library was plated in a 'bulk screening' method (Section 4.2.5.2) which involved transfecting, auto-inducing, and plating the remaining ~ 66,080 clones over a number of M9 minimal media plates with the appropriate nitrile at 5 mM. Four clones were evident amongst the background growth (Figure 4.9). This approach was taken to ensure the remainder of the genomic library (~ 66,000 clones) was screened as selection of the ~ 66,000 individual clones was not feasible.
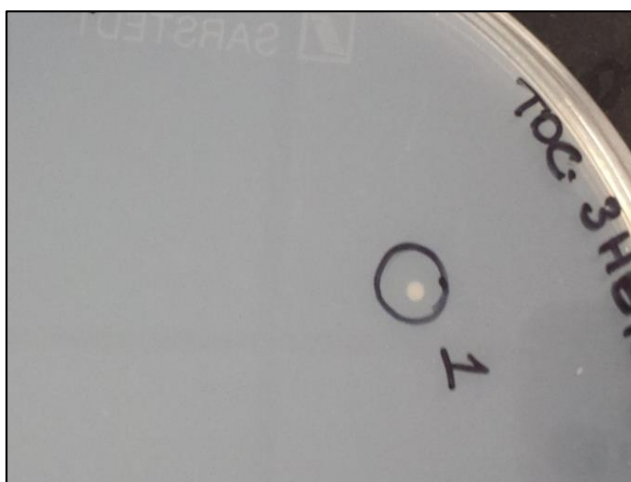
Figure 4.9 One of four functional clones isolated via 'bulk screening' on M9 minimal media with 3-HBN at 5 mM.

4.3.4.1 Restriction enzyme digestion of functional clones

From the initial clone selection and the bulk screening approach a total of six functional clones were identified from the full *R. erythropolis* SET-1 genomic library. Two clones were identified from the initial individual clone selection of 1,920 with the remaining four identified from the bulk screening approach. To ensure that the functional clones identified in Section 4.3.2 were unique, double digestions and subsequent gel electrophoresis was performed. From the initial analysis it was assumed functional clones one, three and four contained the same DNA insert and two, five and six contained the same DNA insert based on the restriction profiles (Figure 4.10).
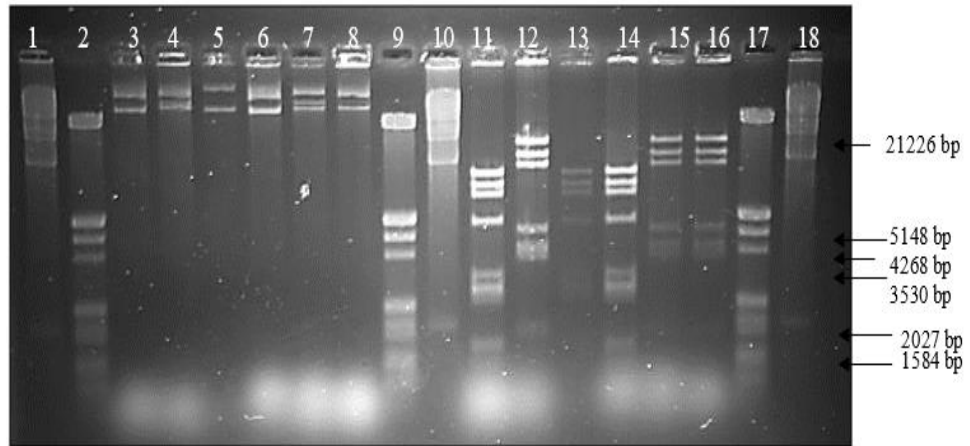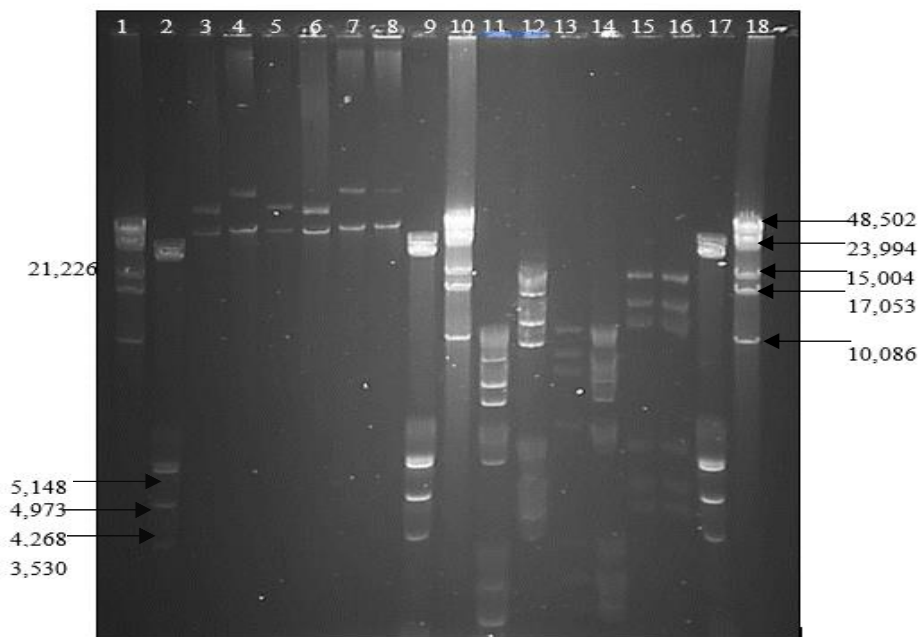
Figure 4.10 (A) Agarose gel analysis of the restriction digestion analysis of 6 functional fosmid clones.

Lane 1, λ DNA monocot mix; Lane 2, *Eco*R1*/HinD*III marker; Lane 3-8; Functional clones one-six; Lane 9, *EcoR1/HinDIII* marker; Lane 10, λ monocot mix; Lane 11-16, Digested functional clones one-six respectively; Lane 17, *EcoR1/HinDIII* marker; Lane 18, λ monocot mix. (B) displays the same gel as figure 4.9 (A) with an extended run time.

4.3.4.2 Fosmid end sequencing of functional clones

All six functional clones were subjected to fosmid end sequencing to confirm the gel assumption that functional clones one, three and four contained the same insert with two, five and six also containing the same insert based on the restriction digestion profiles evident in Figure 4.10. The replication of fosmid inserts during library preparation is not unexpected due to the cell/clone duplication which occurs during an incubation stage within the protocol post transfection.

Sequencing with SR 4 and SL 1 primers

The SL 1 primer was used to sequence the forward direction of the DNA insert in the functional clones, with the data confirming the results aforementioned. Sequencing with the SR 4 primer confirmed that functional clones one, three and four were in fact the same insert, displaying 99% DNA sequence homology to a hypothetical protein from the Aldolase II superfamily from *R. erythropolis* BG 43 complete genome. Functional clones two, five and six were shown to contain the same insert, displaying 98% DNA sequence homology to an ATP dependent DNA helicase from *Rhodococcus* spp. and *R. erythropolis* BG 43 complete genome.

Further fosmid end sequencing was carried out with the reverse primer SL 1. Functional clones one, three and four displayed 99-100 % sequence homology to *R. erythropolis* BG 43, complete genome. The sequence was recognised as AraC family transcriptional regulator. Functional clones two and five displayed 98-99 % sequence homology to *R. erythropolis* BG 43, complete genome. From the restriction profiles and fosmid end-sequencing of each of the six functional clones it was confirmed that there were two unique functional clones, namely FC 1 and FC 2.

4.3.5 Nessler's colourimetric assay

The Nessler's colourimetric assay was performed with FC1 and FC2 as per Section 0 using 3-HGN and 3-HBN as substrates. Both induced and uninduced cells were assessed. No activity was evident.

4.3.6 Gene screening of functional clones for a novel nitrilase gene

R. erythropolis SET-1 was well-studied to date by the research group (Coady *et al.* 2015, Coady *et al.* 2013). All data generated and presented in the 2013 publication by Coady *et al.* indicated that a nitrilase was enzyme gene responsible for the nitrile metabolising activity observed from SET-1. Based on the foundation of that research, the search for the novel nitrilase commenced.

Nitrilase clade screening was performed on the two functional clones identified from the fosmid genomic library of *R. erythropolis* SET-1. A product was displayed at the expected size for the 2A clade, (Figure 4.11). A product at the expected size was also amplified using the 5A clade primers on FC 2 and using the 5B primers on FC 1 (Figure 4.12). Multiple non-specific bands are evident due to the high level of degeneracy of the primers.
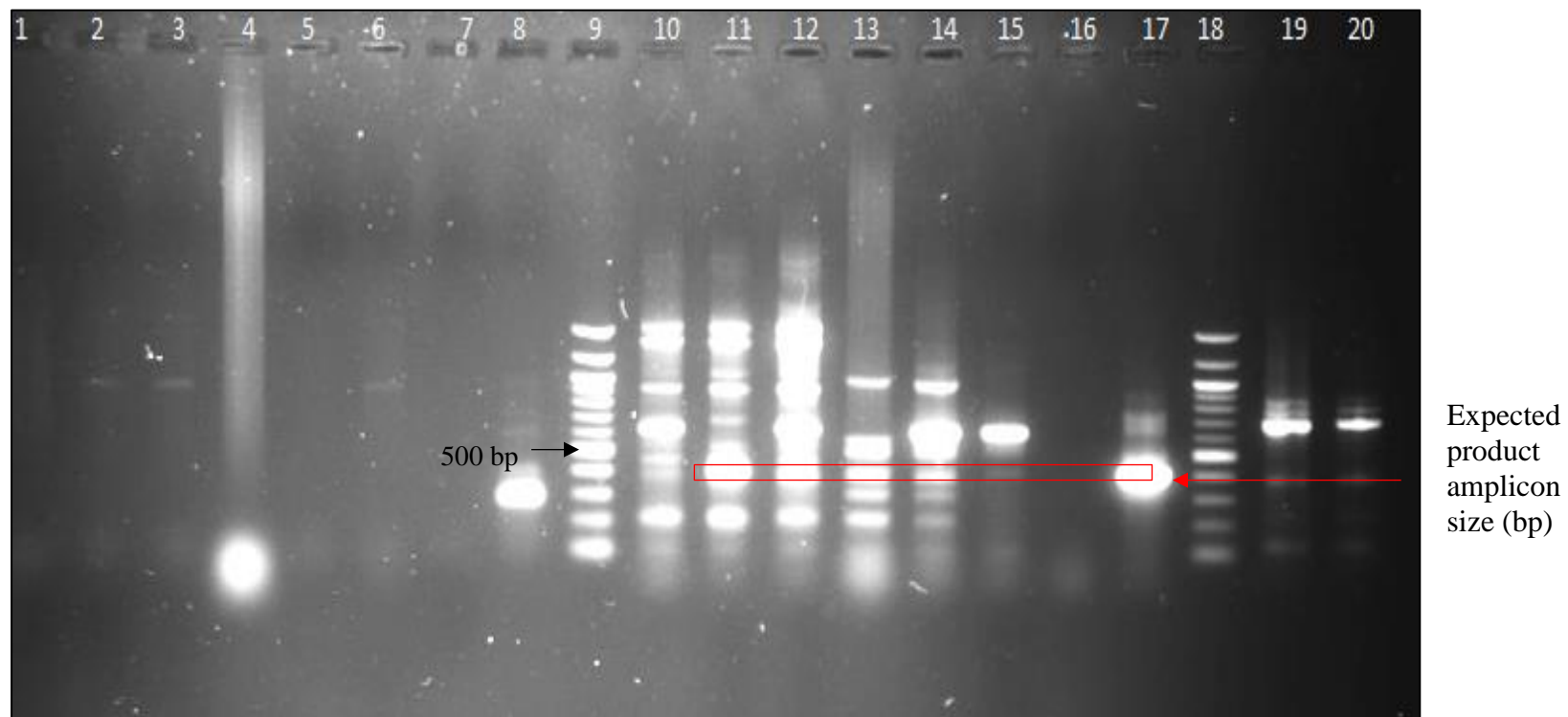
Figure 4.11 Gene screening of FC 1 and FC 2 for nitrilase clades 1A, 2A1-37 and 3A.

Lane 1-3, FC 1(1A clade); lane 4-6, FC 2 (1A clade); lane 7, negative control; lane 8, positive control (Verenium nitrilase 1A8- 287 bp); lane 9, 100 bp marker; lane 10-12 FC 1 (2A1-37 clade); lane 13-15 FC 2 (2A1-37 clade); lane 16, negative control; lane 17, positive control (2A7 Verenium nitrilase-512 bp); lane 18, 100 bp marker; lane 19, FC 1 (3A clade); lane 20, FC 1 (3A clade).
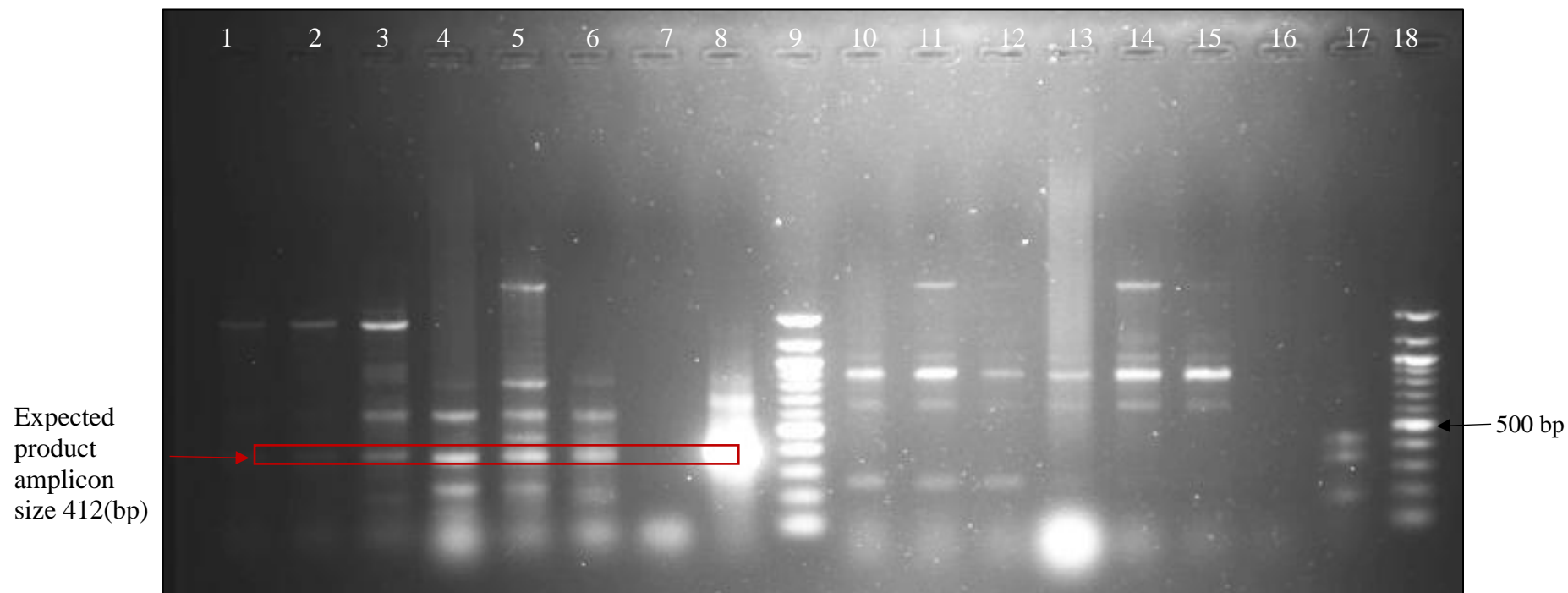
Figure 4.12 Gene screening of FC 1 and FC 2 for nitrilase clades 5A and 5B.

Lane 1-3, FC 1 (5A clade); Lane 4-6, functional clone 2 (5A clade); Lane 7, negative control; lane 8, positive control (5A8 Verenium nitrilase-412 bp); Lane 9, 100 bp marker; lane 10-12, FC 1 (5B clade); lane 13-15, functional clone 2 (5B clade); lane 16, negative control; lane 16, positive control (5B2 Verenium nitrilase- 250 bp); lane 18, 100 bp marker.
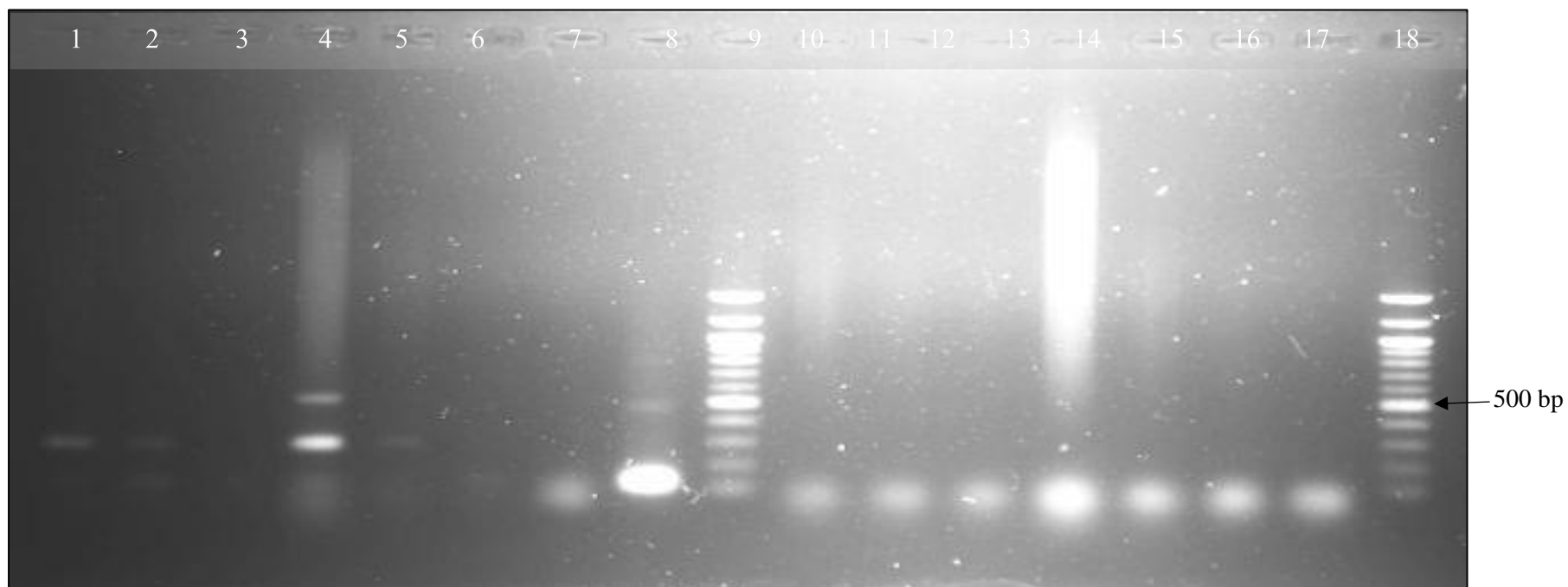
Figure 4.13 Gene screening of FC 1 and FC 2 for nitrilase clade 6A and *Burkholderia* conserved nitrilase region.

Lane 1-3, FC 1 (6A clade); lane 4-6, FC 2 (6A clade); lane 7, negative control; lane 8, positive control (6A4 Verenium nitrilase- 250 bp); lane 9, 100 bp marker; lane 10-12, FC 1 (*Burkholderia*); lane 13-15, FC 2 (*Burkholderia*); lane 16, negative control; lane 18, 100 bp marker.

4.3.7 Cloning of the potential nitrilase clade amplicons

The potential nitrilase amplicons from both functional clones were gel extracted and cloned into the p-GEM®-T easy vector and subjected to sequencing. A partial nitrilase sequence (TMDC-nit) was elucidated from an FC 1 cloned amplicon displaying 98 % DNA sequence homology to a nitrilase from the 5A clade; 5A8 (Figure 4.14) (Robertson *et al.* 2004).

4.3.7.1 Alignment of TMDC-nit with the 5A8 nitrilase
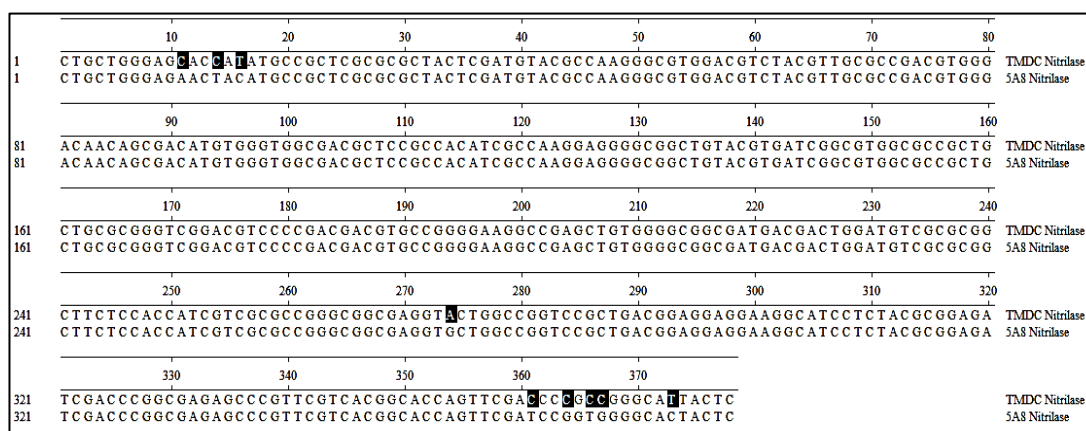
Figure 4.14 Alignment report of the nitrilase partial sequence (TMDC-nit) to the 5A8 nitrilase (Blast accession number AY487517.1) displaying 98% DNA sequence homology with the differences in bases shaded.

This novel nitrilase was suspected to be responsible for nitrile metabolising activity observed in *R. erythropolis* SET-1, however, the cell of course may harbour more than one nitrilase within its genome. No NHase or amidase genes were detected within the FC 1 or FC 2 gene screening.

4.3.8 Attempts to identify the complete nitrilase gene sequence

4.3.8.1 Direct fosmid sequencing

Direct fosmid sequencing was attempted on fosmid extractions with various primers as per Table 4.4 . The TMDCNIT-amp primers, which were designed to detect the partial identified nitrilase sequence within FC1, did result in detection of the target gene in the fosmid extractions utilized for direct sequencing. No sequence was obtained with the TMDCNIT-seq primers (to read across the known internal region of the partial nitrilase

133

gene) or TMDCNIT-amp primers (to amplify the known partial nitrilase gene). Based on this it known that the partial TMDCNit sequence is present in the FC1 due to the PCR amplification of the target region and confirmation with subsequent sequencing analysis, however the lack of sequence across the region from direct fosmid sequencing is indicative of sequencing inhibition, perhaps due to conformation issues. The identified TMDCNit displayed 69.58% GC; based on the high GC % it is likely that there may be runs and repeats across the full gene sequence (or in regions flanking the full gene) potentially leading to the formation of secondary structures and thus in turn preventing sequencing.

4.3.8.2   Restriction digestion of fosmid DNA

The fosmid DNA was subjected to restriction digestion to allow for linearization of the fosmid insert. Linearization of some FC1 DNA extractions was in some cases necessary to allow for the detection of the TMDCNit product, this also indicated that there may be conformation issues with the fosmid or surround DNA to the target sequence. Only some of the FC1 DNA extractions required linearization to allow for the detection of the TMDCNit partial target, which was potentially linked to the varying recovery ratio of different fosmid forms (open, nicked, covalent closed) across each fosmid DNA extraction. Three restriction digestions were carried out (*Eco*R1, *Eco*R1/*Hin*DIII, *Eco*RI/*Bam*H1) to allow for restriction profile comparison and also to allow for various digestion products to be obtained(Figure 4.16). Figure 4.15 displays the amplification of the nitrilase from the fosmid DNA post-digestion using the TMDCNIT-amp primers.
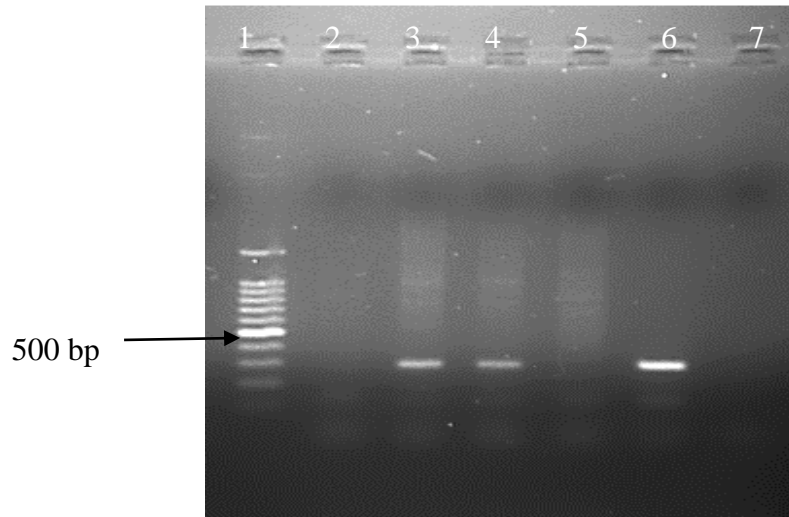
Figure 4.15 Agarose gel analysis of the amplification of the TMDCNit from restriction enzyme digested FC1 fosmid DNA**.**

Lane1, 100 bp marker; lane 2, N/A, lane 3, *EcoR1* digestion, lane 4, FC1 *EcoR1/HinDIII* digestion, lane 5, FC1 no restriction digestion; lane 6, positive control; lane 7, negative control
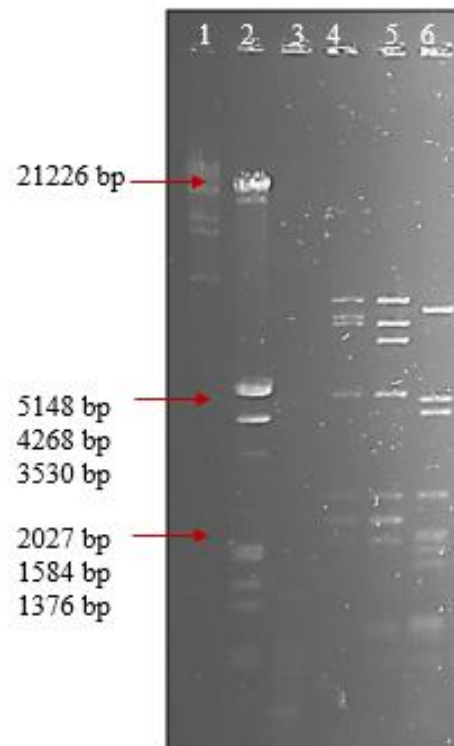
Figure 4.16 Restriction enzyme digestion of FC 1 with single and double enzyme digestions.
Lane 1, Lambda monocut mix marker; lane 2, *Eco*R1/*Hin*DIII marker; lane 3, 100 bp marker; lane 4 FC 1 restriction digestion *Eco*R1; lane 5, FC 1 restriction digestion *Eco*R1/*Hin*DIII; lane 6 FC 1 restriction digestion *Eco*R1/*Bam*H1.

Variability between some of the FC1 DNA extractions was observed upon gel electrophoresis analysis regarding extraction profiles (Figure 4.19- lane 4 and 6). To assess if the cause for the variance was an excision event versus conformation and plasmid forms the fosmid DNA from two FC1 fosmid extractions, representing both profiles, was exposed to three restriction digestion conditions (Figure 4.17).
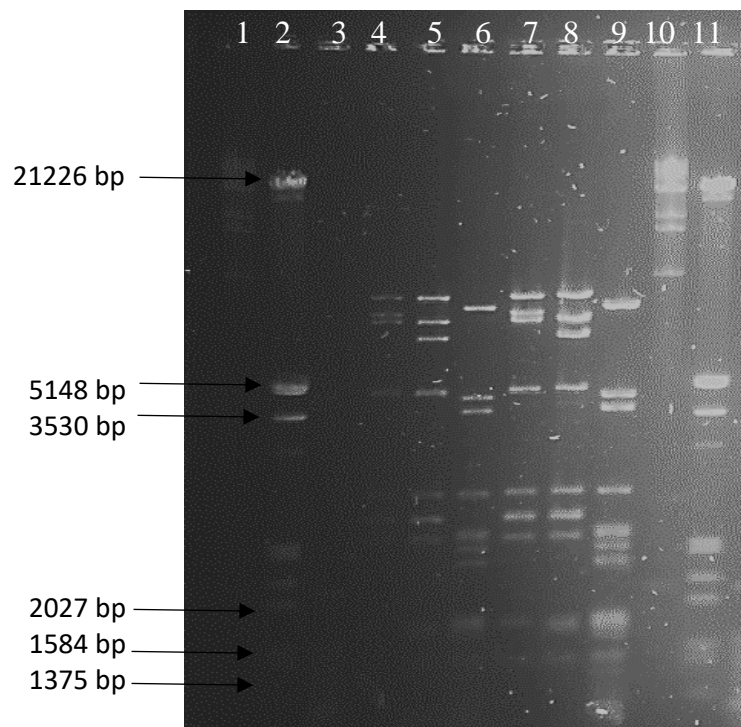
Figure 4.17 Overnight agarose gel analysis of two FC1 fosmid DNA extractions exposure to 3 various restriction digestion conditions.

Lane 1, λ DNA monocut mix marker; lane 2, *EcoR1/HinDIII* marker; lane 3, 100 bp marker; lane 4, *Eco*R1 digestion of FC1 extraction b; lane 5, *Eco*R1/*Hin*DIII digestion of FC1 extraction b; lane 6, *Eco*R1/*Bam*H1 digestion of FC1 extraction b; lane 7, *Eco*R1 digestion of FC1 extraction e; lane 8, *Eco*R1/*Hin*DIII digestion of FC1 digestion e; lane 9, *Eco*R1/*Bam*H1 digestion of FC1 extraction e; lane 10, λ DNA monocut mix marker; lane 11, *Eco*R1/*Hin*DIII marker.

As the *Eco*R1/*Bam*H1 restriction profile was not fully visible on the overnight gel,the restriction digestion was re-analysed on a 1.5% TAE gel to ensure no potential small excision events were missed (Figure 4.18).
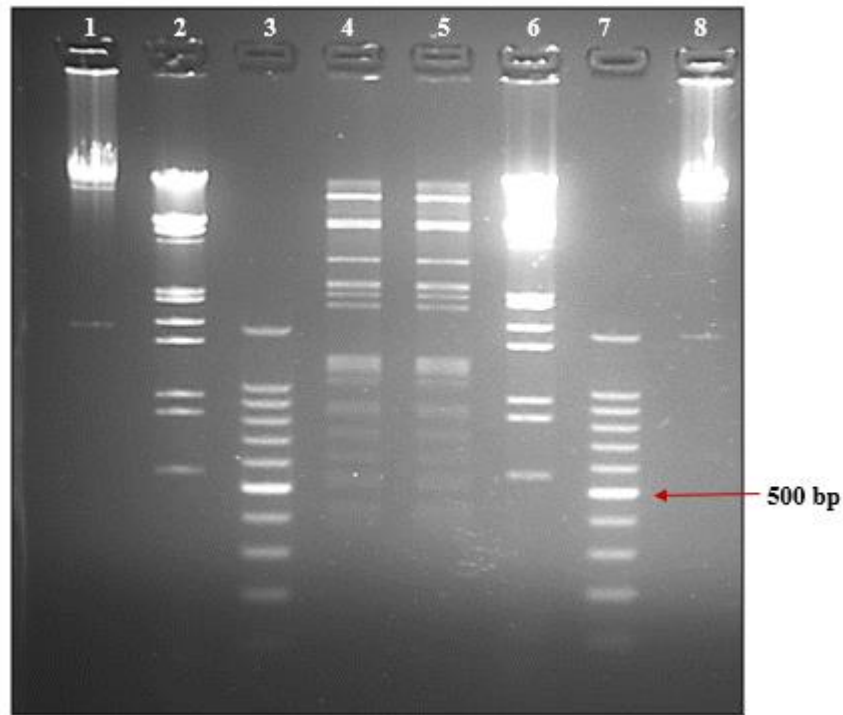
Figure 4.18 Agarose gel analysis of *Eco*R1/BamH1 digested FC 1 fosmid DNA extractions. Lane 1, Lambda monocut mix marker; lane 2, *Eco*R1/HinDIII marker; lane 3, 100 bp marker; lane 4, *Eco*R1/BamH1 digestion on FC 1 extraction 1b; lane 5 *Eco*R1/BamH1 digestion on FC 1 extraction 1e; lane 6, *Eco*R1/*Hin*DIII marker; lane 7, 100 bp marker; lane 8, Lambda monocut mix marker.

Upon comparison of the three restriction profiles for both of the FC1 fosmid clones it was concluded that there was no excision event and that lack of amplification from these fosmid DNA extractions was likely conformational.

4.3.8.3 Confirmation of the nitrilase in the fosmid clone DNA extractions

To confirm that the nitrilase gene was still present in the fosmid extractions (i.e. not lost due to an excision/transposition event), endpoint PCR was used to amplify the known partial nitrilase. Confirmation of detection of the TMDCNit was carried out prior to linearization or restriction digestion of fosmid DNA and prior to any sequencing or long-range PCR attempts. The endpoint PCR was also carried out with any linearized or digested fosmid DNA. Intermittent detection of the gene was observed with undigested fosmid DNA as discussed in Section 4.3.8.2 as displayed in Figure 4.20. Figure 4.19 displays intact fosmid DNA measuring above the 21226 bp. Due to the intermittent

detection of the TMDCNit from the FC1 fosmid DNA, linearization and restriction digestion of the FC1 DNA extractions was performed to try and counteract this.
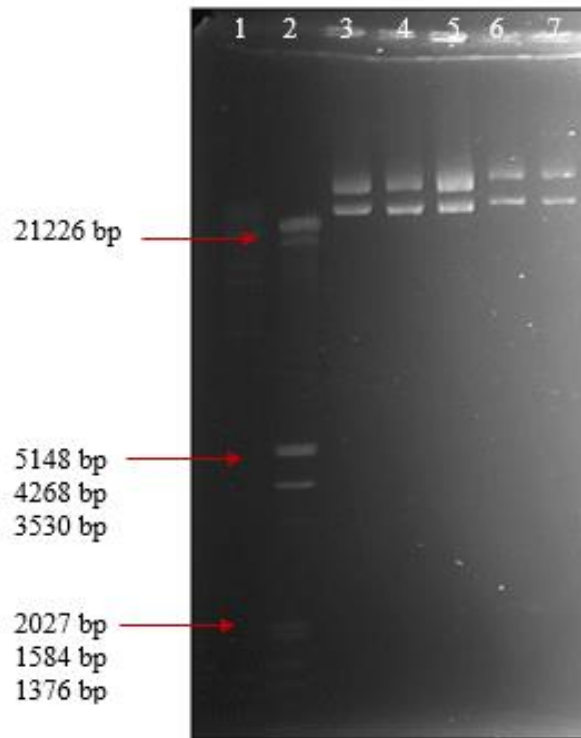
Figure 4.19 Fosmid DNA extractions of FC 1.

Lane 1 Lambda monocut mix marker; lane 2, *Eco*R1/*Hin*DIII marker; lane 3, FC 1 extraction a; lane 4, FC 1 extraction b; lane 5, FC 1 extraction c; lane 6, FC 1 extraction d; lane 7, FC 1 extraction e.
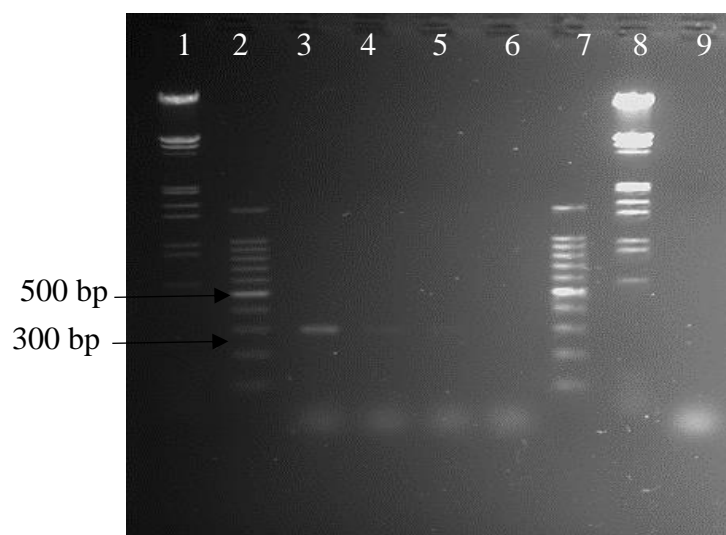
Figure 4.20 Agarose gel analysis of the attempted endpoint PCR amplification of the TMDCNit from fosmid extractions of FC1 (a-d) as presented in Figure 4.19.

Lane 1, *Eco*R1/*Hin*DIII marker, lane 2, 100 bp marker, lane 3 FC1 extraction a, lane 4, FC1 extraction b, lane 5 FC1 extraction c, lane 6 fosmid extraction d, lane 7, 100 bp marker, lane 8 *Eco*R1/*Hin*DIII marker, lane 9, negative control.

4.3.8.4 Long-Range PCR

Attempts were made to amplify a 10 kB, 20 kB and 40 kB fragment from FC 1 fosmid DNA using the Qiagen long range PCR kit as per Section 4.2.9. It was hoped that a linear 40 kB (maximum fosmid insert size), 20 kB or 10 kB fragment would be amplified containing the full TMDCNit nitrilase gene. The aim of this sub-section of research was to amplify and sub-clone the large fragment containing the nitrilase gene, with the hope of bypassing the direct fosmid sequencing limitations. As suspected, fosmid conformation was responsible, therefore it was decided to overcome this by amplifying a linear DNA fragment via the long-range PCR approach.

No consistent or specific amplification was observed. A combination of primer pairs was utilized to try and achieve amplification of the nitrilase gene as the location of the partial gene sequence obtained within FC1 insert sequence was unknown (Figure 4.21). The primer pair combinations involved the use of the TMDCNIT-seq primers (originally intended for direct fosmid sequencing/walking, but in this case to read across the gene), TMDCNIT-amp primers (originally intended to amplify the known partial nitrilase sequence, but in this instance, they were utilised to bind to a known region of the gene).

139

To elucidate the full nitrilase gene the following combinations were employed; TMDCNIT-ampR/SL1, TMDCNIT-ampF/SL4, TMDCNIT-seqR/SR4 and TMDCNIT-seqF/SL1.
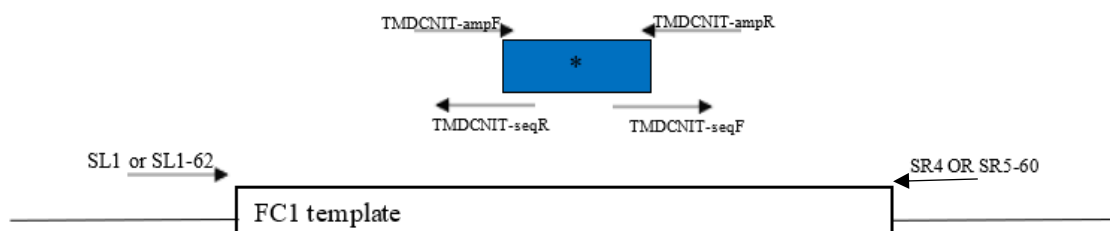
Figure 4.21 Schematic representing the primer pairs utilized for the TMDCNit long-range assays. *hypothetical nitrilase location

4.3.8.1 FC1 DNA additional analysis

Restriction digestion with *Eco*R1 and double digestion with *Eco*R1/*Hin*DIII of the fosmid DNA from FC 1 has previously allowed for the amplification of the TMDCNit known partial nitrilase gene. However, it cannot be guaranteed that the restriction digestions do not cut the remaining unknown sequence. FC1 DNA was subjected to restriction digestion with *Eco*R1 and subjected to gel electrophoresis. Seven bands were gel extracted as per Section 3.2.5, visible in Figure 4.22.
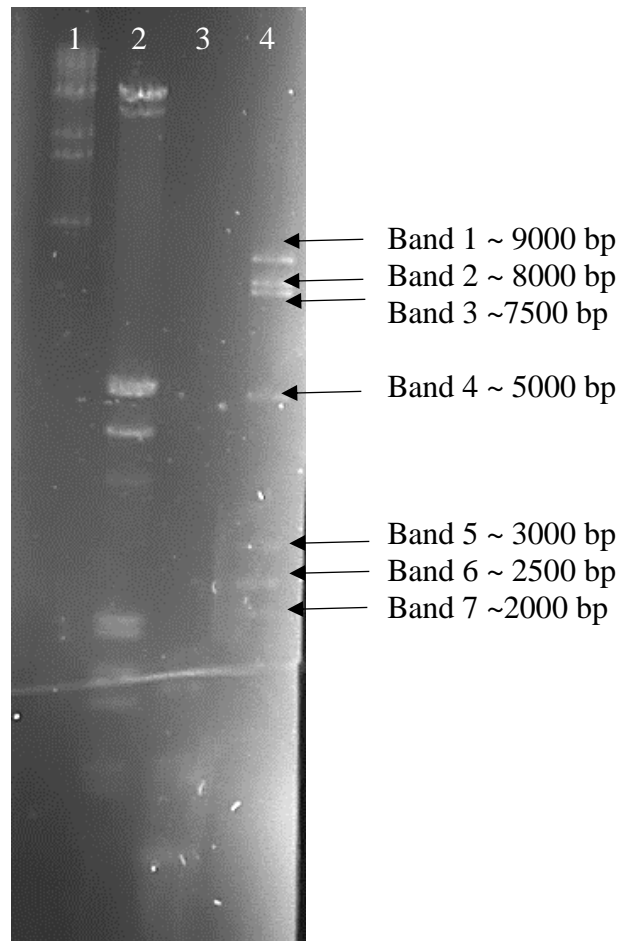
Figure 4.22 Gel electrophoresis 0.55% TAE overnight gel run of digested FC1 fosmid DNA. EtBR stained 'control lane' containing 10 μL of *Eco*R1 FC1 fosmid DNA restriction digestion (uncleaned) which was used to measure the distance (cm) for gel extraction of bands from 'sample' lanes. Lane 1; λ DNA monocut mix marker; lane 2, EcoR1/HindIII marker; lane 3, 100 bp maker; lane 4; control lane (*Eco*R1 digested FC1 fosmid DNA) (300 ng).

Following the successful gel extraction of the restriction enzyme digested (*Eco*R1) FC1 DNA, endpoint PCR was carried out for the detection of the TMDCNit partial nitrilase gene (Figure 4.23). A product at the expected size for the target gene (expected product size of 297 bp) was amplified from all gel extracted bands alongside non-specific products. For this reason, sub-cloning was not persued.
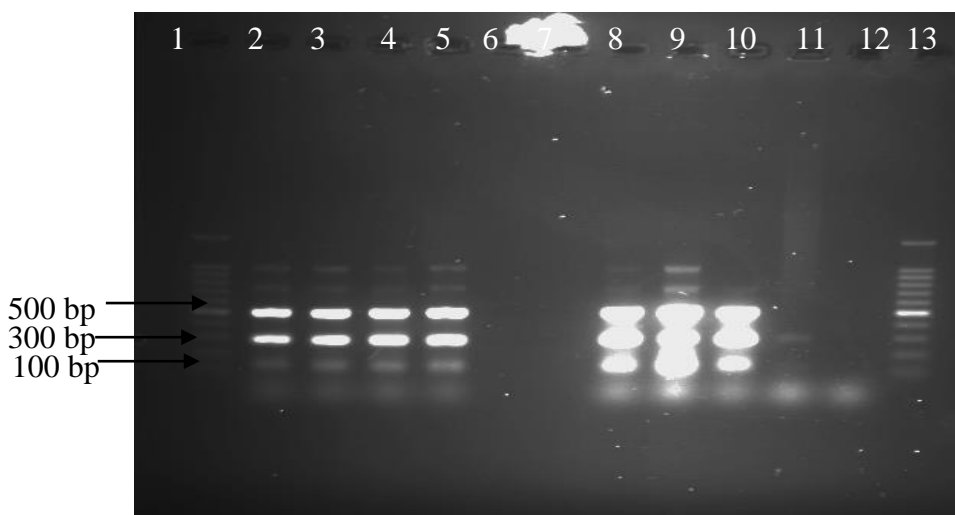
Figure 4.23 Attempted amplification of the known nitrilase using TMDCNIT-amp primers from the gel extracted bands of *Eco*R1 digestion of FC 1.

Lane 1, 100 bp marker; lane 2, band 1; lane 3, band 2; lane 4, band 3; lane 5, band 4; lane 6, n/a; lane 7, n/a; lane 8, band 5; lane 9, band 6; lane 10, band 7; lane 11, positive control; lane 12, negative control; lane 13, 100 bp marker.

4.3.9 Biotransformations with the amidase inhibitor DEPA

As earlier stated, *R. erythropolis* SET-1 was already under study by the wider research group (Coady *et al.* 2015, Coady *et al.* 2013). Prior research to the 2015 article had focused on the identification of a nitrilase based on outcomes and data available from the 2013 paper, which indicated that the nitrile metabolising activity was characteristic of a nitrilase enzyme. By the 2015 research article, the partial TMDCNit was discovered and attempts were underway to elucidate the full sequence, as earlier described. In 2015, it was discovered that *R. erythropolis* SET-1 did in fact also display the ability to produce an amide intermediate at various ratio to acid product which was heavily dependent on the substrate structure.

Based on the findings of the 2015 study and while the attempts to elucidate the full TMDCNit gene sequence continued, the whole cells of *R. erythropolis* SET-1 were subjected to a biotransformation with the known amidase inhibitor DEPA. The whole-cell biotransformations were carried out as per Section 4.2.14 with the resulting reaction analysis performed as per Section 4.2.15. Figure 4.24 displays the retention times for the (R) and (S)-enantiomers of 3-HBN with Figure 4.25 displaying the retention time for the corresponding carboxylic acids.

142
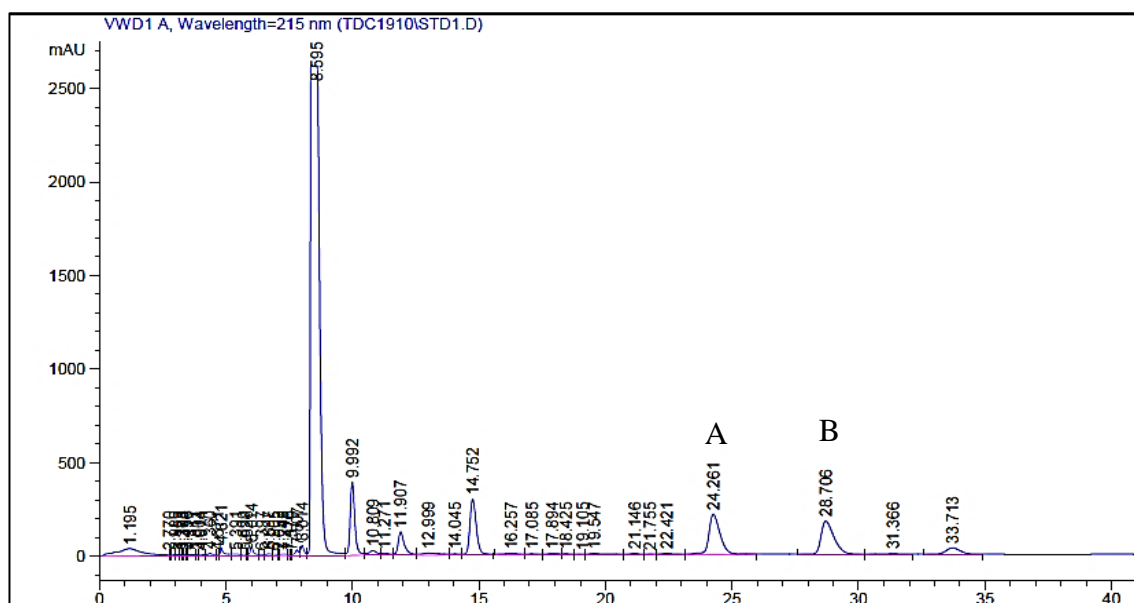
Figure 4.24 Separation of 3-HBN standard enantiomers on a chiracel OJH column.

Mobile phase 90:10 Hexane: Isopropyl alcohol at a flow rate of 0.8 mL/min and 254 nm. (A)-(R)-3-hydroxybutyronitrile eluted at 24.261 minutes and (B)- (S)-3-hydroxybutyronitrile eluted at 28.708 minutes.
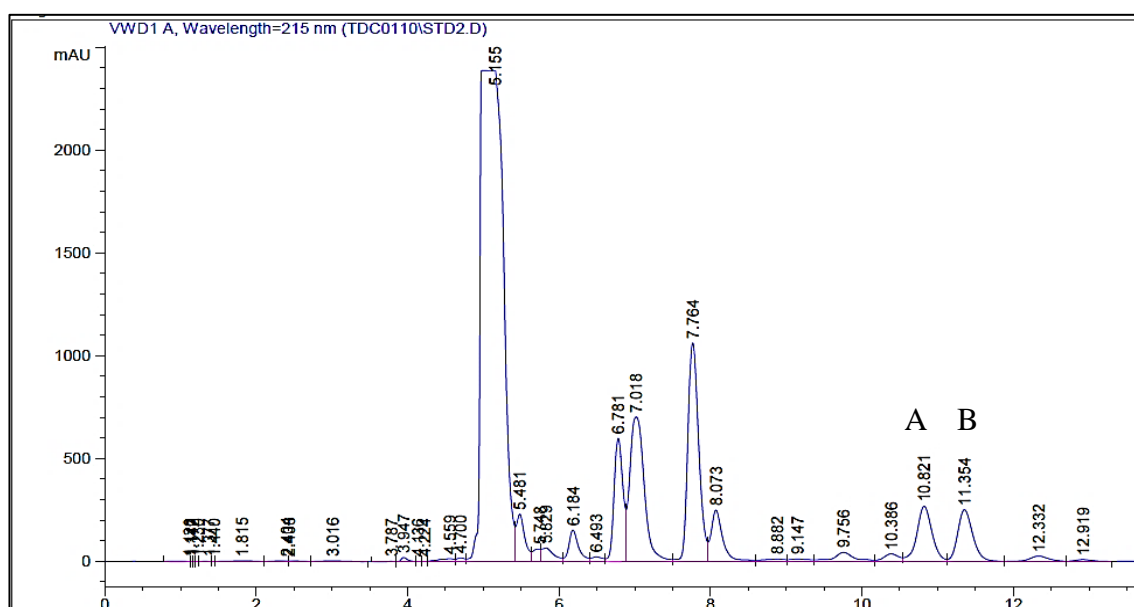
Figure 4.25 Separation of 3-HBA standard enantiomers on a chiracel ADH column.

Mobile phase 90:10 Hexane: Isopropyl alcohol with 0.1% TFA at a flow rate of 0.8 mL/min and 215 nm. Peak (A)- (R)-3-hydroxybutyric acid eluted at 10.8 minutes and Peak (B)- (S)-3-hydroxybutyric acid eluted at 11.3 minutes.

Upon the addition of the amidase inhibitor a loss of production of (S)-3-hydroxybutyric acid was observed (Figure 4.26). The nitrile, as seen previously (Coady *et al.* 2013) was not recovered after the 24-hour biotransformation.
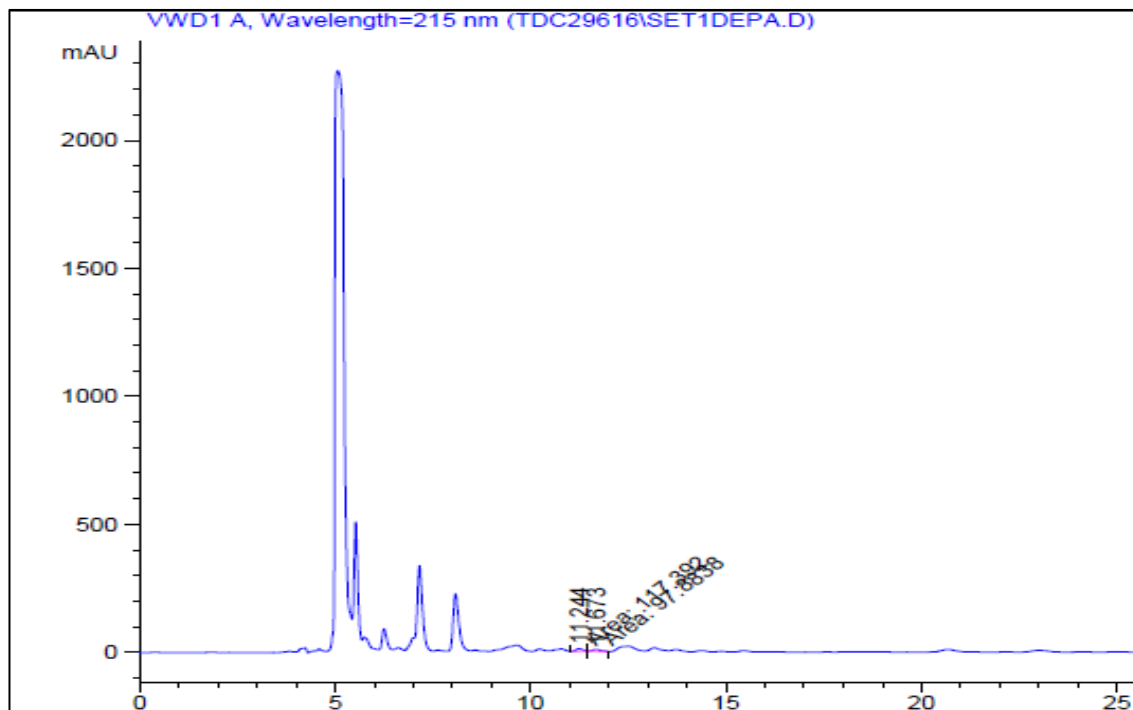
Figure 4.26 10 mM DEPA spiked to the whole cell biotransformation of *R. erythropolis* SET-1 with 10 mM 3-hydroxybutyronitrile. No 3-hydroxybutyric acid was evident and no nitrile was recovered. Chromatographic conditions as per Figure 4.25.

The biotransformation control assay which contained no amidase inhibitor (DEPA) displayed the production of the (S)-enantiomer of 3-hydroxybutyric acid with a retention time of 11.8 minutes (Figure 4.27).
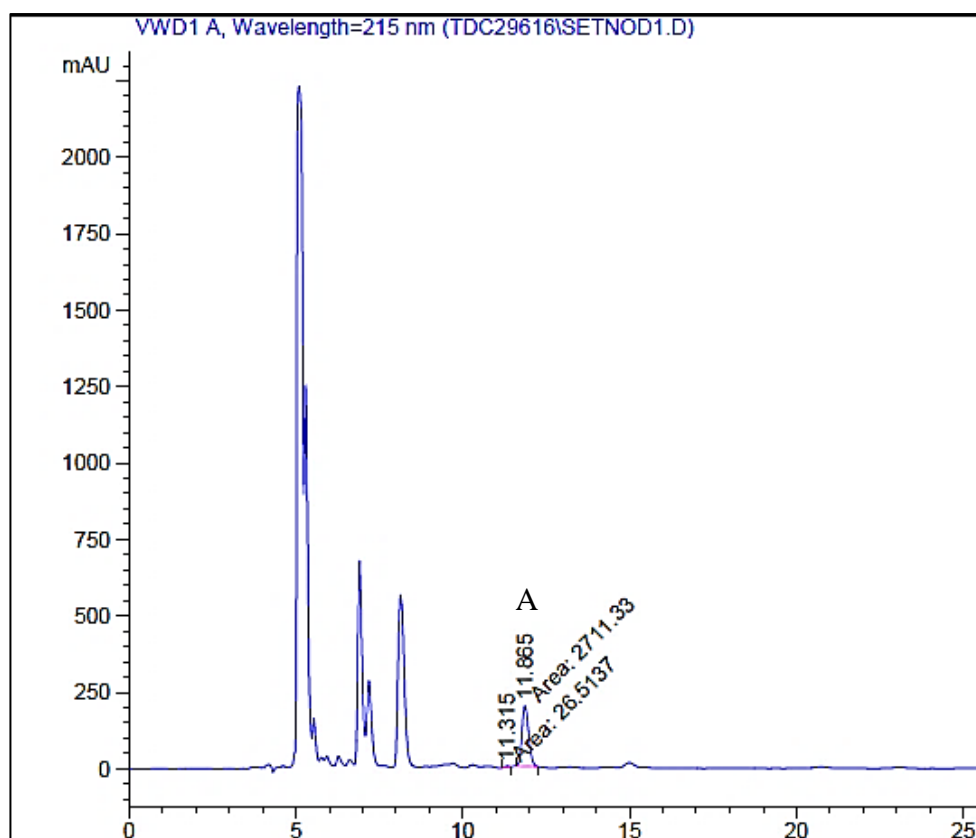
Figure 4.27 Biotransformation of SET-1 whole cells with 10 mM 3-hydroxybutyric acid with the omission of DEPA. Chromatographic conditions: as per Figure 4.25. Peak (A)- (S)-enantiomer of 3-hydroxybutyric acid with a retention time of 11.8 minutes

4.4 Discussion and Conclusion

Plasmid characterisation of *R. erythropolis* SET-1 was carried out to ensure no circular plasmids were detected as such plasmids would likely have been omitted from the fosmid library preparation if they were not subjected to the appropriate linearization treatment. Smaller circular plasmids were screened for using a plasmid miniprep kit for the extraction of plasmids up to 21 kB, with extractions analysed by agarose gel electrophoresis (Section 4.2.2). Two potential plasmid bands were recovered from the circular plasmid preparation and they were subjected to a double digestion with *Eco*R1 and *Hin*DIII, post digest the plasmid displayed a smear rather than the typical banding patters representative of chromosomal DNA carryover, confirming that there were no small circular plasmids in *R. erythropolis* SET-1(Figure 4.4 (B)).

The migration of large circular plasmids differs to that of linear plasmids in that they migrate more slowly and not in a pulsed-size dependent manner i.e. the apparent size by comparison of migration to the linear DNA size standard is not a true reflection of the circular plasmid size. This does mean, however, that PFGE can not only be used to detect plasmids but also to permit characterisation.

Post PFGE analysis of *R. erythropolis* SET-1 under two different migrating patterns, it was evident from the size standard that the plasmid migrated to 555 kB under both electrophoresis conditions, indicating that it is a linear plasmid. S1 nuclease has been shown to digest supercoiled plasmids to their linear forms (Iwaki *et al.* 2013, Basta *et al.* 2004, Bai *et al.* 2003). Comparison of two SET-1 plugs pre- and post-treatment with S1 nuclease confirmed that the plasmid was linear due both samples displaying the same migration pattern (Figure 4.4 (A)). It would be expected that if the plasmid was circular, post-treatment with S1 nuclease would have resulted in a different migration pattern to the pre-treated sample. Additionally, it is common for linear plasmids to have 5' covalently bound proteins (Meinhardt *et al.* 1997). These covalently bound proteins can act as an anchor and prevent plasmid detection. Proteinase K treatment, which was incorporated as part of the *R. erythropolis* SET-1 genomic DNA PFGE plug preparation, has been shown to remove any covalently bound terminal proteins if present on the large linear plasmid thus allowing for correct sizing and detection.

As described, a fosmid library was constructed to elucidate the SET-1 gene(s) of interest by functional screening as previous attempts by the research group were unsuccessful. The number of clones required for a complete fosmid library (representing the entire genome ligated) was estimated to be 822, where the packaged phage preparation was titred and estimated to contain enough phage/DNA to transfect/produce 68,000 clones. The initial clone selection involved plating almost 2,000 clones and subjecting them to functional selection, yielding two clones.

The bulk screening approach, which involved the enrichment of the remaining titre and functional selection resulted in four functional clones. This ensured the full potential of the fosmid genomic library was screened, taking into the account potential clone loss due to competition and outgrowth in the enrichment stage. The decision to screen the full potential library and cover the genome as many times as possible was also to combat any potential risk of excision or ORF interruption. Simply, if only 822 clones were isolated, then there would be a high risk of encountering an excision even, particularly if the nitrilase was unstable or if it was present in an unstable segment of DNA, such as an area of DNA flanked by an IS as detailed above.

The six clones were then subjected to double digestion with *Eco*R1/*Hin*DIII to characterise the fosmid inserts. The restriction profiles indicated that fosmid clones one, three and four contained the same insert and that two, five and six also contained the same insert. To confirm this, fosmid end sequencing was carried out as described in Section 4.3.4.2. As stated, in Section 4.3.4.2, of the six functional clones only two were unique and renamed FC 1 and FC 2. The generation of functional clones with identical inserts during fosmid library construction was not unexpected. Firstly, during the library preparation stages the transfected cells are incubated in a complex media for up to one hour prior or plating for selection. During this stage the transfected cells are diving and so many copies of the clone and in turn the fosmid and insert are created. Additionally, the bulk screening approach which was used during this study also involved the incubation of the transfected cells for a period of one hour, also accounting for clone duplication.

Upon identification of the two unique clones, nitrilase gene screening commenced using the nitrilase clade PCR detailed in chapter three. A nitrilase partial sequence (TMDC-Nit) was elucidated displaying 98 % DNA sequence homology to a nitrilase from the 5A clade (Figure 4.12). It would be expected that upon elucidation of the full nitrilase sequence, the percentage homology to the 5A8 nitrilase would decrease as the primers were designed on conserved regions of known nitrilase sequences. No amidase or nitrile hydratase genes were detected in FC 1 or FC 2 using the primer pairs noted in Table 4.2. This does not indicate that there are no nitrile hydratase or amidase genes present within the functional clone inserts. The nitrile hydratase primers used are specific of Fe-type nitrile hydratase genes and so it is possible that either of the functional clones contained a Co-type nitrile hydratase or a novel gene which could not be detected by this primer pair. Additionally, it is also possible that the functional clones contain an amidase gene but a novel sequence which also could not be detected by the primers utilised in this study.

Based on the identified nitrilase partial sequence in FC 1, a primer pair was designed (TMDCNit-amp) to allow for the amplification of the gene from all DNA templates to be used in the study. The amplification of the TMDCNit was used as a control to ensure that all FC 1 fosmid DNA which would be utilised for the identification of the complete gene, did in fact contain the known region. The identified TMDCNit displayed 69.58 % GC; based on the high GC % and repetitive regions of DNA across the full gene sequence, it is a possibility that there may be aformation of secondary structures and in turn preventing sequencing.

Aside from conformational concerns, excision was another possibility, especially if the nitrilase gene was unstable or within an unstable region of DNA as discussed earlier. Restriction digestions of two fosmid extractions from FC 1 were carried out. These clones were selected based on their different conformation profiles visible via gel electrophoresis (Figure 4.19).The restriction digestions of these two functional clones were carried out with three enzyme digestions (Figure 4.17). The digestions of both functional clones appeared to display the same restriction profile, however, the digestions with *Eco*R1/*Bam*H1 were not clear on the overnight gel electrophoresis with a higher concentration agarose gel (Figure 4.18). The gel analysis and comparison of restriction profiles alongside the detection of the gene post-digestion eliminated excision as the root

cause for the unsuccessful sequencing attempts. Based on the data available at this stage it was still considered possible that the gene was present, albeit perhaps in a difficult to sequence region, as noted above.

Once it was confirmed that the gene had not excised and that it was detectable in all DNA and cell samples prior to further investigations, colorimetric activity assays were carried out (Section 4.3.5). As described, the clones did not display any measurable activity. The pSMART® FOS vector is a cloning vector and so expression of the nitrile metabolising gene of interest in theory would not occur as (i) the cloning vector is not capable of expression and, (ii) no inducer was added to the media. Any nitrile metabolising activity which is observed (such as the functional FC 1 displaying the ability to grow on M9 minimal media with 3-HBN as the sole nitrogen source) is hypothesised to be related to the low-level leaked activity from the regulators and promoters which could be cloned within the ~ 40 kB insert.

The replicator FOS strain supplied with the Lucigen Copyright® V2.0 fosmid library production kit did not display the ability to grown on the M9 minimal media or broth containing 5 mM or 10 mM nitrile. From this, it was determined that the growth which was evident on the M9 minimal media plates during the functional screening process was 'real' growth and was due to the DNA insert encoding a nitrile metabolising gene of interest.

The functional clones did not display the ability to grow in the M9 minimal media broth. It is thought that the ultrapure agar used may have supplemented the media with a trace element/metal which was required for growth/functional activity. Furthermore, the pSMART® FOS vector is not an expression vector and so any growth demonstrated would be due to the expressional leakage of the insert genes. This very low level of enzyme per cell may also account for a lack of visible growth in the M9 minimal broth vs the slow, visible growth displayed on the M9 minimal agar. This would further account for the lack of activity detected by the colorimetric assay, as the Limit of detection (LOD) may not have been sensitive enough.

Variation was observed amongst the levels of amplification between each restriction digestion, which may be due to discrepancies within the ratio of fragment recovery during

the clean-up process. A lack of amplification has been noted on many occasions with the *EcoR1*/*Bam*H1 digestion; thus, this restriction digestion has not been employed hereinafter. It may be that the sequence obtained from sequencing of the amplified conserved nitrilase region has some base pair errors leading to a *Bam*H1 restriction site which is not identified by the potential 'incorrect' sequence. More likely is that the *Bam*H1 site exists but access to the target site is limited due to the secondary structure formation. If the target site is unreachable, the enzyme would be unable to restrict the site in question.

A long-range PCR kit was utilized as per Section 4.3.8.4. As noted, no consistent amplification was observed based on the primer binding sites and extension times permitted. The primer pair SL 1 and SR 4, which were used for the PCR, exhibited melting temperatures with 10 °C variation and so the lack of amplification was attributed to this. A second subset of primers were designed based on the known fosmid vector sequence, to replace SL 1 and SR 4, labelled SL1-62 and SR5-60 respectively. These primers displayed similar melting temperatures and so allowed for a standardised PCR protocol. Gradient PCR was carried out with all primer pairs; however, only non-specific amplification was observed.

The lack of amplification may be attributed to conformational issues of the fosmid DNA of the insert in that the *Taq* polymerase may have not been able to gain access to the primer binding sites. The location and direction of the nitrilase within the fosmid insert is unknown and so it may be a case that the long-range PCR conditions which were employed were not optimised to allow for the amplification of the DNA region containing the nitrilase gene. Attempts to amplify the nitrilase gene involved coupling the primers to ensure one primer bound to the known partial nitrilase sequence and one bound to the known vector sequence. This was tried in both directions with the primers as shown in Figure 4.21 however it may still be possible that the coupling and reaction conditions utilized were not optimised to achieve this. The replacement primers ensured that an appropriate annealing temperature could be selected for all the primer pairs and when coupled with gradient PCR non-specific amplification, and in some cases no amplification was observed.

Due to the unsuccessful attempts towards the elucidation of the remaining nitrilase sequence, the FC1 extract was digested. The resulting bands from the digestion of fosmid DNA from FC 1 with *Eco*R1 were gel extracted. The extracted DNA was subjected to PCR with TMDCNIT-amp primers to try and identify the band which contained the known nitrilase sequence. In theory, one of the gel extracted bands should contain the target gene and so it would be expected to see one product amplified in one of the extracted products. Figure 4.18 displays the multiple banding observed for all of the gel extracted bands (two non-specific bands, 1 band at the expected size). For this reason, sub-cloning was not carried out with the gel extracted bands.

As previously discussed, the known sequence of the nitrilase displays a high GC % content and repetitive DNA sequence, this could be a cause for the amplification of 'non-specific' products. It may also be the case that the sequence within the ~40 kb insert within FC 1 consists of multiple copies of this gene, possibly caused by the presence of a transposase or duplication events within *R. erythropolis* SET-1. Based on the literature available and a study presented by Siguier *et al.* (2006) it is thought that IS are more common in larger plasmids. As stated, *R. erythropolis* SET-1 was shown to contain a large linear plasmid measuring 555 kB. It has been documented that attempts to elucidate the full nitrilase gene sequence were unsuccessful, potentially due to the locations of the partial nitrilase within the SET-1 genome. Insertion sequences are known to cause direct repeats to transpire in the flanking DNA of the target site. If the nitrilase gene was present on the plasmid and was preceded or followed by an IS sequence, then the direct repeats which are commonly associated with IS as noted above, could in fact cause secondary structure formation leading to a difficult to sequence region, or if the region could be sequenced successfully, contig generation post sequencing with Illumina technology (as per the fosmid insert sequencing) may have resulted in false contig formation, thus omitting a contig section containing the full nitrilase gene.

Following the failed attempts to elucidate the full nitrilase gene by the approaches discussed above, fosmid insert sequencing was performed. The aim of this batch of work was to sequence the full approximate 40 kB fosmid insert, in which the partial nitrilase gene was detected. The full fosmid insert sequencing was unsuccessful for the elucidation of the remaining nitrilase sequence; this further highlighted the possibility that the fosmid insert may contain difficult to sequence regions. The NGS sequencing using Illuma

technology relied on the construction of a library and subsequent sequencing with data analysis involving contig construction. If the sequence within the 40 kb (maximum fosmid insert size) region was highly repetitive or contained a high GC % content, then it is likely that this 'difficult to sequence region' was by-passed and so the nitrilase gene could in fact be present but omitted from resulting contigs due to 'false' contig construction.

Based on the available data presented to date, it is possible that the full nitrilase gene is present with the full fosmid insert but it cannot be detected. Firstly, the nitrilase gene could be preceded by or followed by an insertion sequence site. As described earlier this would lead to repetitive regions within the flanking DNA, thus leading to secondary structures and so in turn creating a difficult to sequence region, as the *Taq* polymerase may not be able to gain access to the target site or a primer may not be able to bind to the target sequence. Secondly, the nitrilase gene itself may be unstable or in an unstable region of DNA, particularly if the gene is present in a segment of DNA which was acquired by HGT, potentially leading to areas of repetition with the gene sequence and secondary structure formation, resulting in the same problems as above. Thirdly, it is possible that the partial nitrilase gene sequence detected is just that, a partial gene. It could be possible that this partial gene sequence is present in multiple areas within the *R. erythropolis* SET-1 genome. This could be possible if the segment of DNA was acquired by HGT as noted above, and if the segment acquired contained a transposable element or perhaps a transposase which may be responsible for the 'cut and paste' mechanism of the partial gene or part of the full gene which may not be accessible for any of the reasons described above.

The activity of the environmental isolate *R. erythropolis* SET-1 has previously been described (Coady *et al.* 2015, Coady *et al.* 2013). Upon commencement of this body of work, the 2013 publication describing the *R. erythropolis* SET-1 activity proposed a nitrilase enzyme was responsible for the nitrile metabolising activity observed on β-hydroxynitriles. On the foundation of the pre-existing research and activity demonstrated in the biotransformations with whole cells, it was decided that the initial focus should be placed on the search for a nitrilase gene.

However, due to the inability to elucidate the full nitrilase gene from functional FC1 or in fact from SET-1 genomic DNA, a research question was raised- was the partial nitrilase part of a full nitrilase gene, a partial sequence (as earlier discussed), or perhaps not responsible for the activity in question? Sustaining this research question was the finding of the 2015 paper published by Coady *et al*. (2015). The paper proposes that the range of activity observed by SET-1, across several nitrile substrates is potentially due to a combination of nitrilase/nitrile hydratase systems due to the presence of amide in the completed biotransformations.

This dual metabolism system is a possibility since the biotransformations involved the use of the whole cells vs a recombinant enzyme. However, an inhibitor assay using thiol binding agents which are known nitrilase inhibitors, resulted in a decrease or complete loss of activity, thus indicating that a nitrilase may still be responsible for the activity observed on the substrate of interest 3-HBN (Coady *et al*. 2015).

Supplementary support for this hypothesis is the body of research (though minimal) which has been published, detailing various reports of amide formation by nitrilase enzymes, which support the hypothesis that a nitrilase enzyme may still be exclusively responsible for the activity observed. As reported by Kobayashi *et al.* (1998), the nitrilase from *R. rhodochrous* J1 displayed the ability to produce both amide (though in small amounts) and carboxylic acid from the starting nitrile.

Diethyl phosphoramidate (DEPA), is a known amidase inhibitor (Angelini *et al.* 2015, Brady *et al.* 2004, Bauer *et al.* 1998). Based on this, biotransformations were performed containing DEPA using the same substrate, 3-HBN as noted above. The resulting chromatogram displayed no carboxylic acid with the addition of DEPA, this was indicative that a nitrile hydratase amidase system would have been responsible for the desired activity from *R. erythropolis* SET-1 (Figure 4.26).

The reactions carried out by Coady *et al.* (2013) did note that no amide was present in the reactions and so it was thought to be a one-step hydrolysis by the nitrilase, however, there was no (R)- nitrile detected (post 3-hour incubation), as would be expected due to the 42% yield of (S)-carboxylic acid. The lack of nitrile was discussed as being due to potential isomerase present in the whole cells (this is unlikely as the maximum yield for

the S-carboxylic acid was never more than 42%) or perhaps as noted by Coady *et al.* (2013), it may have been a case where the remaining nitrile was perhaps utilised by other genes/enzymes within the *R. erythropolis* SET-1 genome. As neither the (R)- or (S)-nitrile were recovered with or without the addition of DEPA, it could also be argued that the nitrile metabolising gene of interest, if a nitrile hydratase, would use both the (R)- and (S)- nitrile with the amidase displaying enantioselectivity for the (S)-amide, in turn producing the (S)-acid, namely (S)-3-hydroxybutyric acid.

Previous studies on *R. erythropolis* SET-1 noted that no amide was detected during the biotransformation with 3-hydroxybutyonitrile under the biotransformation conditions as described in Coady *et al.* (2013).

The 2015 publication by Coady *et al.* however, confirmed that amide production was observed depending on substrate structure. The publication also noted that the addition of a docking group at the hydroxy position in β-hydroxyaromatic nitriles displayed an increase in activity, a characteristic which is more commonly observed with a nitrile hydratase amidase system. At present, based on the data available and the data generated in previous studies (Coady *et al.* 2015, Coady 2014, Coady *et al.* 2013), it is still ambiguous as to whether the conversion of the nitrile to the corresponding acid is a dual system involving both microbial pathways of nitrile metabolism or a nitrile hydratase amidase system. Based on the data available at present and prior to whole genome and fosmid insert sequencing, it was thought that *R. erythropolis* SET-1 did in fact harbour a nitrilase and nitrile hydratase system which worked in tandem, depending on the structure of the nitrile which was provided.

Parallel to the above, whole genome sequencing was carried out with *R. erythropolis* SET-1, as presented in Chapter 5. The genomic DNA was confirmed as containing the TMDC-nit partial gene prior to the genome sequencing. The comparative analysis of the fosmid insert sequencing to the whole genome sequencing of SET-1 is described in chapter five.

CHAPTER 5

COMPARATIVE ANALYSIS- FOSMID INSERT SEQUENCING VS WHOLE GENOME SEQUENCING OF *R. ERYTHROPOLIS* SET-1

5 COMPARATIVE ANALYSIS- FOSMID INSERT SEQUENCING VS WHOLE GENOME SEQUENCING OF *R. ERYTHROPOLIS* SET-1

Chapter 4 documents the attempts undertaken to elucidate the full nitrilase sequence of the nitrilase gene which was detected from FC1,a functional clone from the SET-1 genomic library, denoted TMDCNit. Failure to elucidate the full nitrilase sequence via the methods previously described led to the preparation of fosmid DNA from FC1 and genomic DNA from the nitrile metabolising isolate *Rhodococcus erythropolis* SET-1 for insert sequencing and whole genome sequencing respectively.

Fosmid DNA was extracted from FC 1 containing the TMDCNit partial nitrilase for complete fosmid sequencing by GATC Biotech. The known nitrilase sequence was not detected via direct fosmid Sanger sequencing as earlier noted therefore it was decided to attempt to sequence the full fosmid insert. Illumina sequencing technology was used for subsequent sequencing of the FC1 insert. Next-generation sequencing (NGS) has many benefits over the traditional Sanger-based approach including: faster sequencing runs, higher level of data generation, ability to sequence longer nucleotides/regions, and the ability to fine tune the resolution achieved. However, there are some limitations with NGS when a DNA template with a high GC % or repetitive regions are encountered; this will be addressed in this chapter.

5.1 Research Hypothesis and Aims

Hypothesis:

A combination of fosmid insert sequencing and whole genome sequencing can be used to elucidate novel nitrile metabolising genes of interest from a known nitrile metabolising isolate, *R. erythropolis* SET-1.

Aim:

The aim of this research was to use a combination of fosmid insert sequencing and whole genome sequencing to identify nitrile metabolising genes from the nitrile metabolising isolate *R. erythropolis* SET-1.

5.2 Material and Methods

5.2.1 Fosmid DNA extraction

Fosmid DNA was extracted using the Epicentre FosmidMax- DNA purification kit (Lucigen, Cat. No. FMAX046) as per manual, with the exception that DNA elutions were performed in Tris-HCl buffer pH 7.5. The fosmid DNA extraction was screened for the presence of TMDCNit as per Section 4.3.8.3.

5.2.2 Genomic DNA extraction

R. erythropolis SET-1 was harvested on M9 minimal broth with 10 mM 3-HBN as the sole nitrogen source for 3 days at 25 °C at 150 rpm. Prior to genomic DNA extractions the whole cells of R. erythropolis SET-1 were screened for ee % towards 3-HBN as per Section 4.3.9. Genomic DNA was extracted using the Anachem Keyprep nucleic acid kit (Anachem, Cat. No. AM111113), as per the manufacturer's guidelines with DNA eluted in 0.1 M Tris-HCL pH 7.5. The genomic DNA of *R. erythropolis* SET-1 was screened for TMDCNit prior to sequencing, as per Section 4.3.8.3.

5.2.3 Fosmid Insert and whole genome sequencing

Fosmid insert sequencing was carried out by GATC Biotech (Konstanz, Germany). Whole genome sequencing of *R. erythropolis* SET-1 was carried out by Microbes NG (University of Birmingham), both companies utilising Illumina technology.

5.2.4 Amplification of the nitrile metabolising genes

Gene screening for Fe-type nitrile hydratase and amidase genes in SET-1 genomic DNA was carried out using conventional PCR and primers designed for the full and conserved regions of the genes as per Coffey *et al.* (2010).

5.2.5 Sequence analysis

Nucleotide sequences were analysed using BLASTn software from the NCBI database (Altschul *et al.* 1990). The functional clone insert was analysed utilising the RAST software (Brettin *et al.* 2015, Overbeek *et al.* 2014, Aziz *et al.* 2008) to allow for annotation of a 36,526 bp fragment, identified as the FC1 insert.

5.3 Results & Discussion

5.3.1 Fosmid insert sequence analysis

Fosmid inserts range from 30-40 kB as determined by the Lucigen Copyright® V2.0 fosmid library production kit. Prior to fosmid insert sequencing, the fosmid DNA extractions were shown to contain the TMDCNit nitrilase partial amplicon (Figure 5.1). From Figure 5.1 it is evident that three FC 1 fosmid DNA extractions displayed variation in the amplification levels of TMDCNit. This was also observed and discussed in Chapter 4. It is proposed that the variation in levels of amplification relates to the ratios of forms of fosmid DNA recovered. For instance, FC1 extraction (a) displays the highest level of amplification, perhaps indicating that the FC 1extraction (a) has a larger ratio of linearized fosmid, favouring amplification. Based on the resulting amplification, FC 1 fosmid DNA extraction (a) was used as the DNA template for fosmid DNA sequencing.

Following identification of the partial nitrilase sequence, attempts were made to elucidate the full sequence. Primers were designed based on the known sequence to allow for direct fosmid sequencing. Linearization of the fosmid DNA was required in most cases to allow for the amplification of the partial nitrilase gene via conventional PCR. This would suggest that the conformation of the fosmid or conformation of the DNA insert within the fosmid is preventing amplification of the partial nitrilase amplicon and in turn sequencing. Though the gene could be detected in all template samples, direct fosmid sequencing was unsuccessful. As earlier noted, the nitrilase partial sequence obtained from FC 1 displays 69.58% GC content alongside repetitive sequence regions. This may be leading to the formation of secondary structures such as hairpin loops, thus making sequencing extremely difficult. Earlier experiments permitted successful fosmid-end sequencing i.e. reading from the fosmid vector into the cloned region, this indicated that

failure to obtain sequence from direct fosmid sequencing commencing at the known partial nitrilase sequence could be due to a difficult to sequence region.

The *R. erythropolis* SET-1 whole cells from which genomic DNA was extracted did display the expected activity from the biotransformation analysis (>99.9% ee on 3-HBN), however, no TMDCNit product was amplified from the genomic DNA (Figure 5.1). The lack of detection of the TMDCNit nitrilase from the genomic DNA of SET-1 was not surprising. Throughout this study it was evident that detection of the TMDCNit varied between DNA extraction and templates. It is thought that the high GC% content and the runs and repeats with the known sequence are responsible for the formation of secondary structures, in turn limiting or in some cases hindering the amplification of the gene.
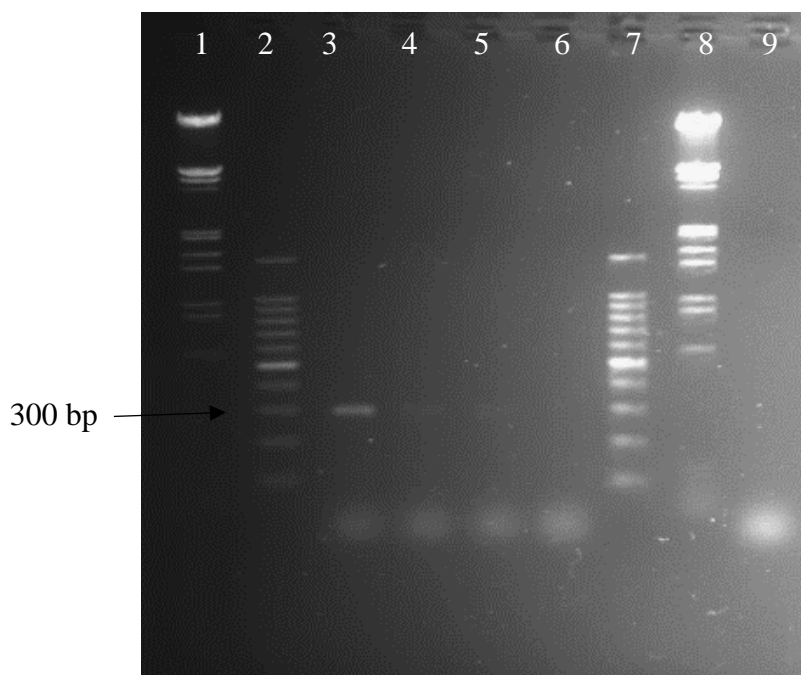
Figure 5.1 Agarose gel analysis TMDCNit amplification from FC 1 fosmid DNA and *R. erythropolis* SET-1 genomic DNA extractions.
Lane 1, *EcoR1*/*HinDIII* marker; lane 2, 100 bp marker; lane 3, FC 1 extraction (a); lane 4, FC 1 extraction (b), lane 5, FC 1 extraction (c), lane 6, *R. erythropolis* SET-1 genomic, lane 7, 100 bp marker; lane 8, *EcoR1*/*HinDIII* marker, lane 9, negative control.

Upon analysis of the contigs generated from the fosmid insert sequencing, a contig measuring 36,526 bp was identified as being the FC 1 insert, based on the fosmid end reading as described in Section 4.3.4.2.

This insert was aligned utilizing BLASTn software from the NCBI database (Altschul *et al.* 1990), to the reference genome for *Rhodococcus erythropolis* CCM2595 (CP 003761.1). The microorganism *R. erythropolis* CCM2595 was selected as the reference genome within this study as it is noted as the representative genome for *Rhodococcus erythropolis* by the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/genome/1638). The FC1 insert displayed 98% DNA similarity to a region within the CCM2595 reference genome (Figure 5.3)

A region of 14,762 bp did not align to the reference genome, with an additional 1,757 bp gap evident on the CCM2595 genome within the alignment, corresponding to a 509 bp region (hereafter referred to as Region X) of the FC1 SET-1 insert. Region X of the *FC1 R. erythropolis* SET-1 insert was not aligned to the reference genome, with a Blastn query of this 509 bp resulting in a 100% DNA homology to a hypothetical protein (CP007255.1) in the genome of *R. erythropolis* R138. Full screening of the fosmid insert sequence revealed no DNA homology to the partial TMDC-nit from SET-1 or the 5A8 nitrilase (Figure 5.2).

Further to this, no region of interest with regard to a nitrile metabolising gene of interest was evident with the aligned region of *R. erythropolis* CCM2595 genome. No nitrilase, nitrile hydratase or amidase genes of interest were identified. Additionally, no regulatory genes which would suggest the region of DNA contained in FC 1 was preceding or perhaps following a nitrile metabolising cluster or close to genes contained in the aldoxime-nitrile pathway were detected.

Figure 5.2 Alignment report of the 509 bp region from FC 1 (which did not align to the reference genome of CCM2595) to *R. erythropolis* R138 region and the TMDCNit nitrilase partial sequence amplified from FC 1 fosmid DNA.

The base pairs shaded in black represent the nucleotides which match across the three DNA sequences with nucleotides of difference highlighted in white.

Figure 5.3 Schematic displaying the alignment of the FC1 SET-1 fosmid insert to the *R. erythropolis* CCM2595 (CP0037651.1) genome.

The initial 14,762 bp of the FC1 insert (36, 526 bp) did not align to any region of the reference CCM2595 genome, and so a BLASTn query was carried out with this region. The 14,762 bp region displayed 99% DNA homology to a region of the *R. erythropolis* BG43 genome (CP011295.1), featuring a Ferredoxin NADP (4FE-4S) binding protein followed by a putative ferredoxin-NADP reductase (Figure 5.4). Due to the fact that the reference genome did not contain this region it was decided to compare to FC1 insert region of 36, 526 bp to both the reference genome of CCM2595 and also BG43 (Figure 5.4 and Figure 5.5).
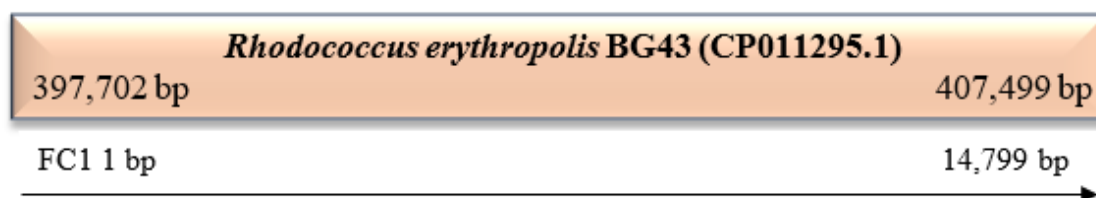
Figure 5.4 Schematic displaying the alignment of first 14,799 bp from the FC1 SET-1 insert to the *R. erythropolis* BG43 (CP011295.1) genome.

A ferredoxin NADP binding protein (4FE-4S) and a putative ferredoxin NADP reductase were identified within the first 14,799 bp of the FC1 insert when aligned to the R. erythropolis BG43 genome. The alignment of the FC1 *R. erythropolis* SET-1 insert to the *R. erythropolis* BG43 genome displayed 99 % DNA homology with 94% query cover as presented in Figure 5.5. The *R. erythropolis* BG43 genome did not display alignment or DNA homology to the partial TMDCNit from *R. erythropolis* SET-1 or to the uncultured 5A8 nitrilase.
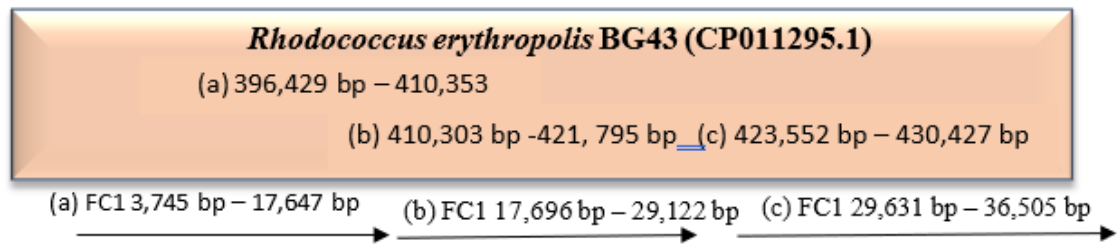
Figure 5.5 Schematic displaying the alignment of the FC1 SET-1 insert to the *R. erythropolis* BG43 (CP011295.1) genome. Three ranges were identified displaying 99 % DNA homology to the FC insert from *R. erythropolis* SET-1.

Conserved Domains within the FC1 *R. erythropolis* SET-1 insert were analysed to see if there was any indication of a nitrilase superfamily conserved domain (Figure 5.6). This analysis was carried out using the NCBI CD finder (Marchler-Bauer and Bryant, 2004). The NCBI's ORF finder was used to search for open reading frames within the *R. erythropolis* SET-1 FC1 insert. The resulting ORFs were constructed as in Figure 5.7.
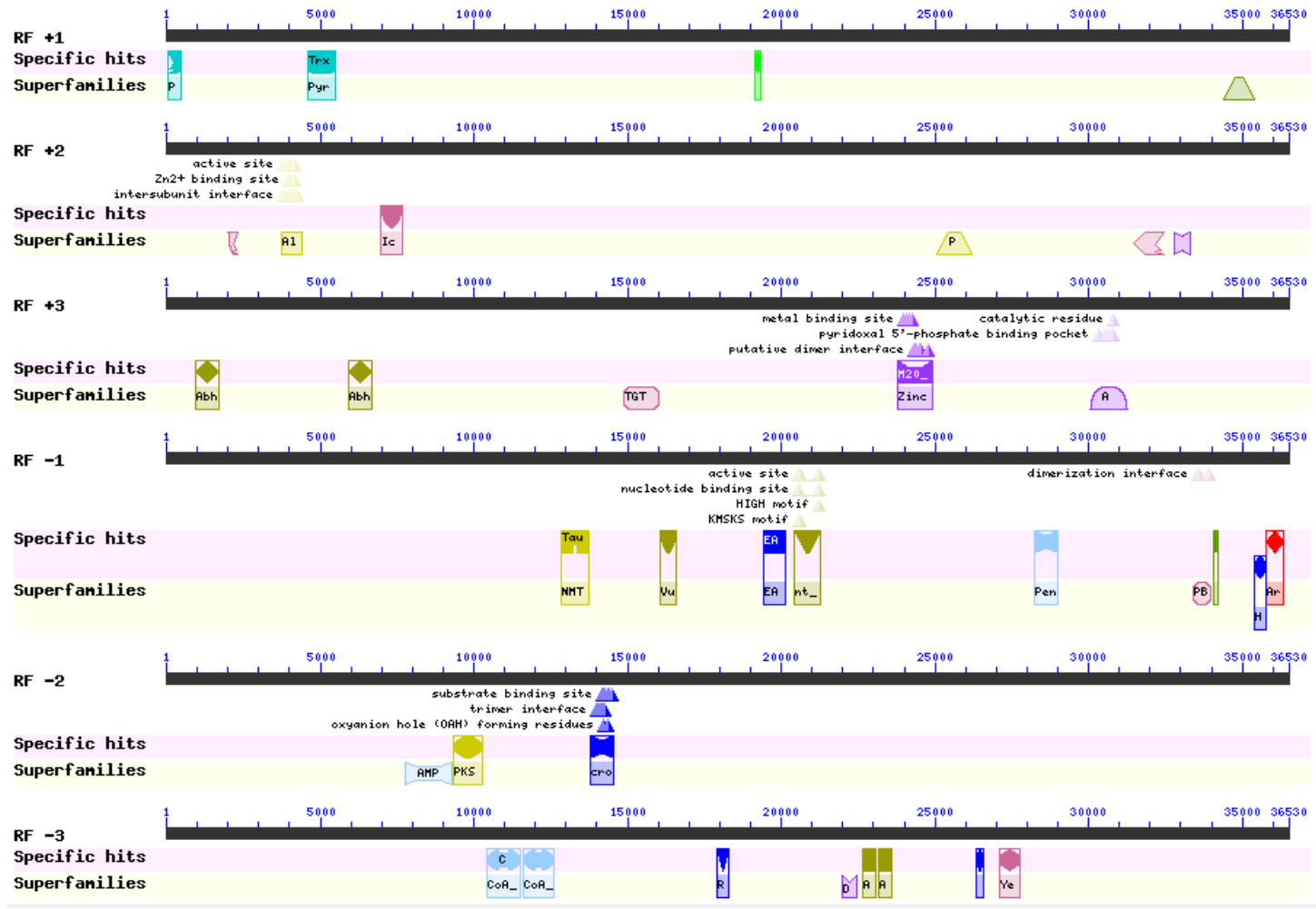
Figure 5.6 Conserved domain search of the FC-1 SET-1 insert displaying numerous active and primer binding sites.
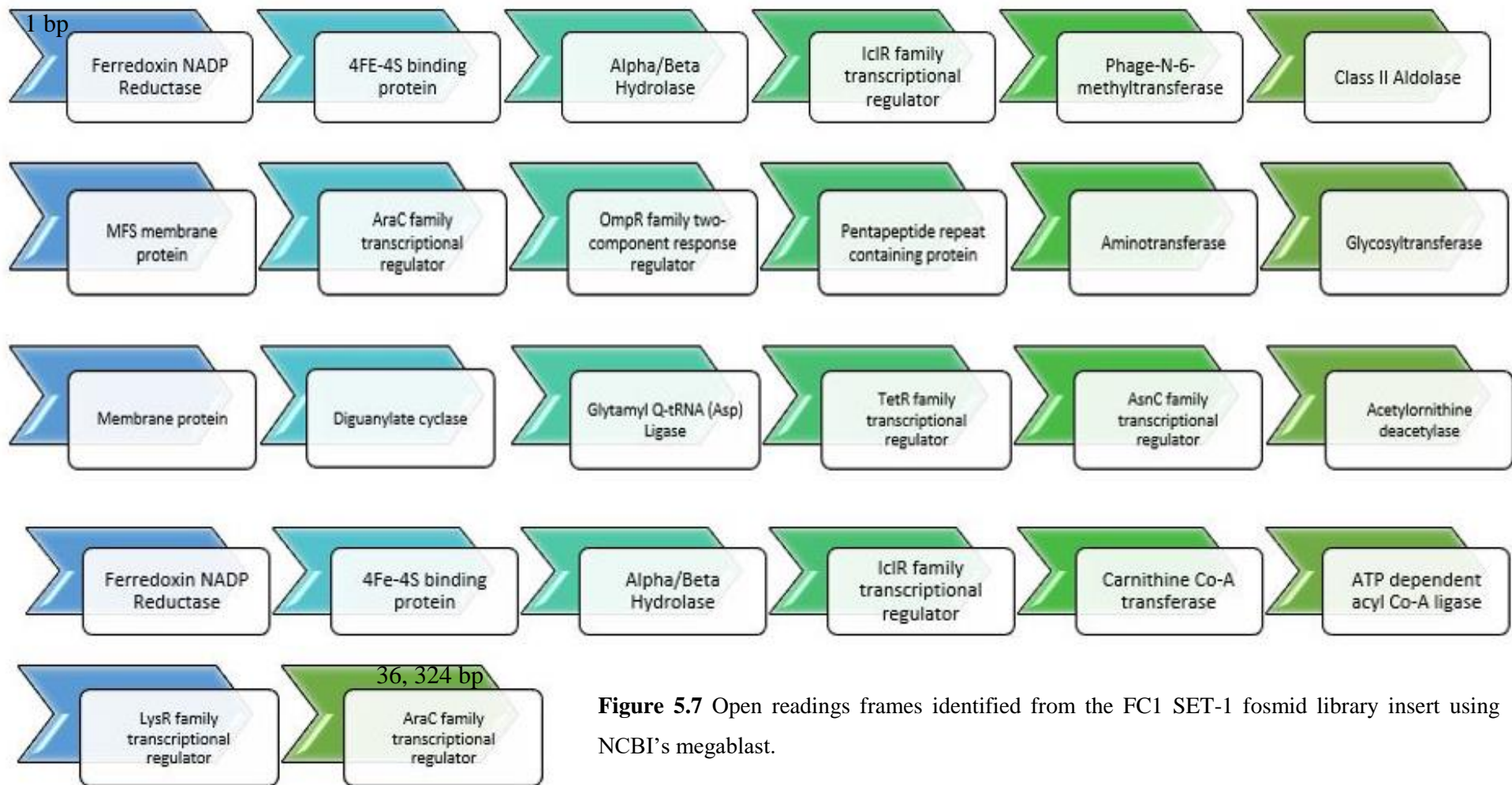
Figure 5.7 Open readings frames identified from the FC1 SET-1 fosmid library insert using NCBI's megablast.

The FC1 SET-1 insert contained a 509 bp region which, as earlier stated, did not align to either the reference genome of CCM2595 or BG43. This region was analysed using the NCBI's BLAST tool searching for regions of high homology. The region matched the *R. erythropolis* R138 (CP007255.1) genome at 100% query and 100% identity (BLASTn), featuring a hypothetical protein H351-01300 (BLASTp). The hypothetical protein showed no conserved domains when analysed with the NCBI's Conserved domain identifier, however, identical hypothetical proteins were observed across many *Rhodococcus* sp. with no function identified. Region X did display a 46 bp match at the 3' end of the *Rhodococcus* sp. BH4 cell filamentation protein and low homology was observed when compared to the cell filamentation protein in *R. erythropolis* BG43 (XU06_31540).

Where the FC1 SET-1 contained Region X, CCM2595 and BG43 displayed a 1754 bp gap in the alignment homology as seen in Figure 5.3.

Figure 5.8 This region was analysed for regions of high homology to the NCBI's database.

Within this region two hypothetical proteins (AKD95961.0 and AKD95961.1) displaying alignment to the BG43 genome were identified, with 99 % homology. No features or function were identified. A conserved domain search was carried out with domains relating to two superfamilies identified: an Abhydrolase superfamily conserved domain was identified followed by RNA recognition Motif (RRM) family protein conserved domain.

The FC1 SET-1 sequenced region was analysed for palindromic sequences, with particular attention paid to the flanking region of 26,138 bp – 26,647 bp due to the 510 bp region X which was evident in the FC1 SET-1 sequenced region which displayed no homology to CCM2595 or BG43. Forms of genome rearrangement are represented by the

introduction or loss of sequence from the region of interest. With regard to the FC1 SET-1 insert, Region X could be indicative of an insertion in SET-1, or excision from the other *Rhodococcus* sp. suggesting perhaps the gene/associated DNA was no longer required for survival.

There were 8 palindromic sequences evident in the FC1 DNA insert (Table 5.1), each of which contained no spacer nucleotides or mismatched bases, as determined by the DNA Analyzer software as published by Brázda *et al.* (2016). Palindromic sequences can lead to the formation of secondary structures such as hairpins, which can lead to difficulties when trying to analyse or use DNA containing them as a template for studies. This supports the earlier hypothesis, that *R. erythropolis* SET-1 did in fact contain difficult to sequence regions, perhaps owing to secondary structure and the isolate may in fact contain the sequence of interest (the remaining nitrilase sequence) but in a difficult to sequence region. However, any FC1 DNA extractions which were utilised as DNA templates for fosmid insert sequencing were first shown to contain the TMDCNit gene.

Table 5.1 List of palindromic sequences evident in the FC1 SET-1 insert with zero gaps or mismatches as analysed by the DNA Analyzer software (Brázda *et al.* 2016).

Length (bp)	Position (bp)	Sequence
6	13873	CTCGAATTCGAG
6	15601	GCACAATTGTGC
6	20735	GTCGTC GACGAC
6	20789	GACGACGTCGTC
6	23055	GCGGTCGACCGC
7	28185	CTCCGTCGACGGAG
6	35008	GCTGCCGGCAGC
7	30989	GATCCTCGAGGATC

5.3.2 Whole Genome Sequencing Analysis

Whole genome sequencing (WGS) of *R. erythropolis* SET-1 resulted in 24.258 MB of data over 8188 contigs of varying length. *Rhodococcus* sp. genomes are known to be high in GC content and have been strongly linked with both gene duplication events and horizontal gene transfer (Letek *et al.* 2010). Letek *et al.* (2010) also discussed how

Rhodococcus sp. genomes can range from 4.3 megabases (Mb) to over 10 Mb, with the majority measuring over 5 Mb.

Firstly, a Blast alignment between the full WGS contig library and the FC1 SET-1 insert was carried out using the Blast Alignment software. Two nodes were highlighted as having significant homology to the FC-1 insert, namely node 16 and node 30. Node 16 consisted of a 157.85 kB region with node 30 measuring 93.069 kB. FC1 displayed 100 % homology with 51% and 41% coverage respectively to the node 16 and node 30 sequences. These nodes did not show any overlap in sequence and did in fact show sequence homology to different regions of the CCM2595 reference genome, indicating that they are not two contigs overlapping or covering the same region.

As node 16 displayed one continuous alignment to the FC-1 insert it was selected as the area of interest. The FC 1 insert displayed homology in the reverse direction from bp 17,719 to 35,505 matching node 16 at 157,850 to 139,064 bp. The full FC 1 insert was assessed for ORFs as previously described and shown in Figure 5.7. The full sequence comprising node 16 was aligned to the reference genome of CCM2595. The aim of this alignment was to assess the regions upstream and downstream of area of high homology to the FC 1 sequence. It was hoped that a region displaying gene of interest (genes comprising the aldoxime-nitrile pathway) would be annotated. No genes associated with the aldoxime-nitrile metabolising pathway were evident.

Parallel to this a conserved domain search was carried out on the full node 16 region as with the full FC 1 insert, visible in Figure 5.6. The resulting conserved domains were scanned for a nitrilase superfamily conserved domain; however, none were detected. Surprisingly, a conserved domain "PBP1-NHase" denoted as a type I periplasmic-binding protein of the nitrile hydratase (NHase) system was identified; this displayed 98% DNA homologyto a pentapeptide repeat containing protein of CCM2595, with no noted function.

Node 16 and node 30 did not display any DNA homology to the partial nitrilase amplified from the fosmid insert library despite displaying high homology to the remainder of the FC-1 SET-1 insert. As *R. erythropolis* CCM2595 is the reference genome within the analysis, the nucleotide sequence for the α-subunit of the nitrile hydratase gene (6182038

to 6182661 bp) of CCM2595 was aligned to the full WGS library constructed from *R. erythropolis* SET-1 genomic DNA. One node, namely node 89 contained a DNA sequence which displayed 99% DNA homology to the alpha subunit of the nitrile hydratase gene from CCM2595.

Node 89 which consists of 30.086 kB was then aligned to the CCM2595 genome(Figure 5.9). This resulted in an alignment at 99% DNA homology from bases 6165773 to 6197678 on the CCM2595 genome. The gene cluster within the aligned region contained a nitrile hydratase gene cluster (Figure 5.10).
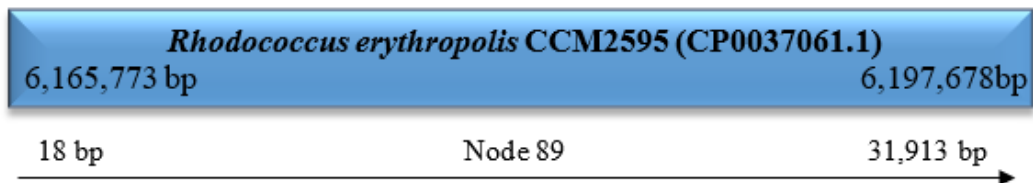
Figure 5.9 Alignment of Node 89 from the WGS of SET-1 genomic DNA to CCM2595 genome. A DNA homology of 99% was observed across the displayed alignment.
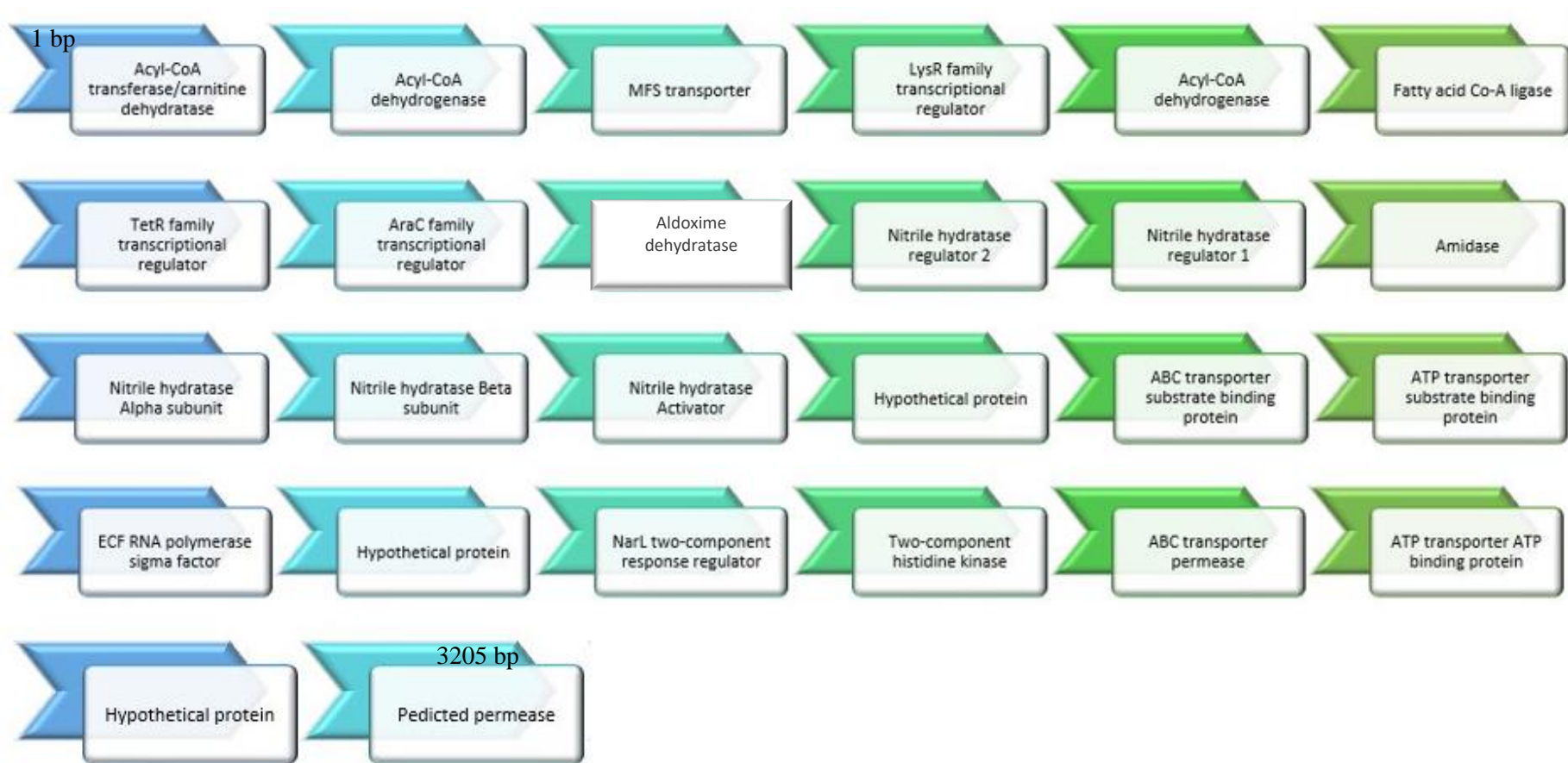
Figure 5.10 Node 89 from the WGS of *R. erythropolis* SET-1 genomic DNA. Nitrile metabolising genes identified on the direct strand displayed 99% DNA homology to the CCM2595 genome. The hypothetical protein preceding the nitrile hydratase regulator 2 has been identified as an aldoxime dehydratase based on accession number JQ023030.2.

Post identification of the nitrile hydratase in the whole genome library, the *R. erythropolis* SET-1 cell (genomic DNA and colony PCR) and *R. erythropolis* SET-1 FC1 was subjected to gene screening for a nitrile hydratase (Figure 5.12 and Figure 5.13) and amidase gene(s) (Figure 5.11) as per Section 4.2.8.1.

The fosmid DNA from FC 1 was also subjected to amplification of these genes, however no product was amplified. It was not expected that the FC1 from the fosmid library would display an amplicon for the nitrile hydratase or amidase genes as it aligned to a different region in the reference genome CCM2595. However, due to the uncertainty of potential structural variation in the fosmid DNA recovery and potential secondary structure formation, screening was performed.

The *R. erythropolis* CCM2595 genome contains a 1061 bp aldoxime dehydratase gene upstream of the cluster presented. The gene is located at 6,177,287 to 6,178,348 bp (Genbank accession CP0037061.1). This aldoxime dehydratase displays 99% homology to a region within node 89 from the *R. erythropolis* SET-1 genomic library, indicating that the partial aldoxime dehydratase sequence detected in *R. erythropolis* SET-1 is a full gene. The α (MH257756) and β (MH257757) subunits of the nitrile hydratase gene in *R. erythropolis* SET-1 have successfully been amplified from genomic DNA.
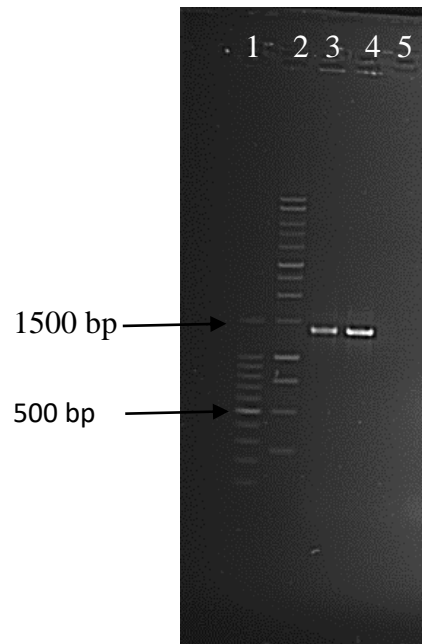
Figure 5.11 Agarose gel analysis of amidase PCR products from *R. erythropolis* SET-1 genomic DNA.

Lane 1- NEB 100 bp DNA marker, Lane 2- 1 kB marker, Lane 3- 50 ng *R. erythropolis* SET-1 genomic DNA, Lane 4- *R. erythropolis* AJ270 positive control (50 ng genomic DNA), Lane 5- negative control.
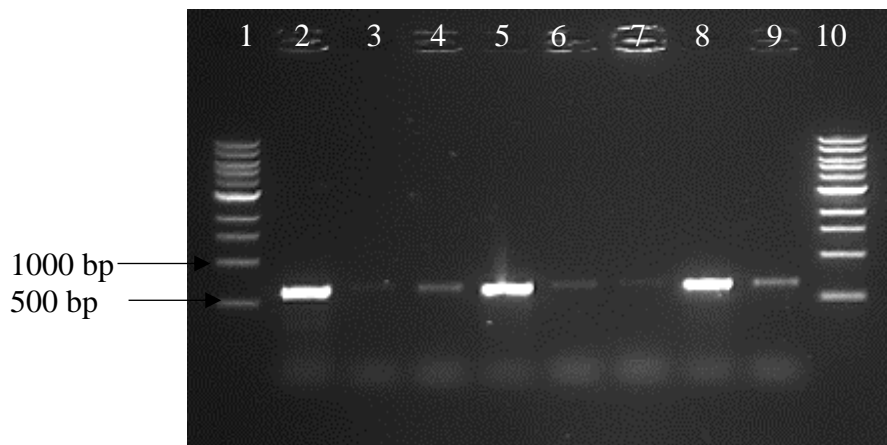
Figure 5.12 Agarose gel analysis of NHase-α PCR.

 Lane 1- NEB 1 kB DNA marker, lane 2- *R. erythropolis* AJ270 positive control (50 ng genomic DNA), lane 3- negative control, lane 4- 5 ng *R. erythropolis* SET-1 genomic DNA, lane 5- 50 ng *R. erythropolis* SET-1 genomic DNA, lane 6- 10 ng *R. erythropolis* SET-1 genomic DNA, lane 7-9 colony PCR *R. erythropolis* SET-1, lane 10- NEB 1 kB DNA marker.
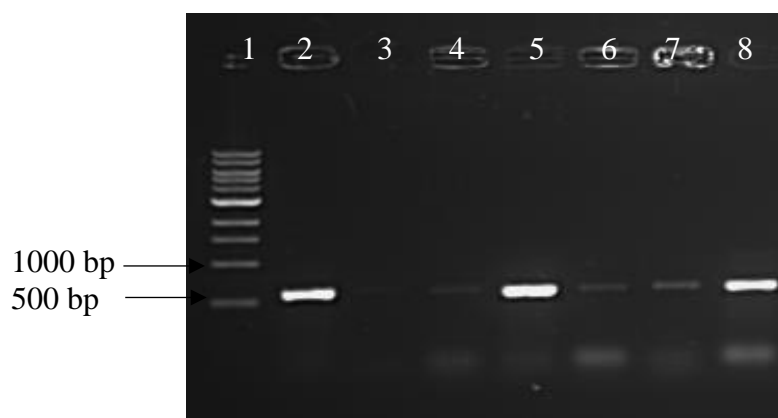
Figure 5.13 Agarose gel analysis of NHase-β PCR.

Lane 1- NEB 1 kB DNA marker, lane 2- *R. erythropolis* AJ270 positive control (50 ng genomic DNA), lane 3- negative control, lane 4- 5 ng *R. erythropolis* SET-1 genomic DNA, lane 5- 50 ng *R. erythropolis* SET-1 genomic DNA, lane 6- 10 ng *R. erythropolis* SET-1 genomic DNA, lane 7-8 colony PCR *R. erythropolis* SET-1.

Both the NHase α and β subunit displayed 99% DNA homology to the α and β subunit from the *R. erythropolis* CCM2595 reference genome, across the sequenced region. The amidase gene (MH257755) detected also displays 99% DNA homology to the amidase detected in the reference genome of CCM2595, across the sequence region.

In conclusion, from the data generated from the whole genome sequencing it is evident that SET-1 does in fact contain a nitrile hydratase gene cluster, with the α- and β- subunits of nitrile hydratase and the amidase gene from SET-1 genomic DNA successfully amplified and sequenced. The lack of detection of the nitrilase gene in the *R. erythropolis SET-1* genome does not confirm that it is not present in the *R. erythropolis SET*-1 cells, perhaps just that it is in a difficult to sequence region of the genome or length of DNA. Though 24 MB of data was generated through WGS, it is possible that though there is high coverage some of the genome may not have been sequenced. It is also possible that the WGS did not provide enough coverage and so some regions of the genome may not be adequately covered. Furthermore, it may also be a limitation of the Illumina sequencing technology in that the construction of scaffolds and contigs from a high GC % isolate may in fact result in a 'false contig' forming, or perhaps an area of overlap of sequence may in fact be an area of high GC %, runs and repeats and so falsely omitted from the library generated.

CHAPTER 6

GENE CLONING, OVEREXPRESSION AND CHARACTERISATION OF THE NITRILASE FROM *BURKHOLDERIA* SP. LC8

6 GENE CLONING, OVEREXPRESSION AND CHARACTERISATION OF THE NITRILASE FROM *BURKHOLDERIA* SP. LC8

6.1 Introduction

As discussed in Chapter 1, nitriles retain the general structure (R-CN) and they have been utilized for the production of high-value added products such as carboxylic acids and amides for biopharmaceutical and industrial applications (Solano *et al.* 2012, Singh *et al.* 2006, Brady *et al.* 2004, Banerjee *et al.* 2002).

The aforementioned advantages of biotransformations with nitrile metabolising enzymes as discussed in Chapter 1 certify that they hold great potential for the production of numerous biopharmaceutical and industrially relevant amide/acid products (Solano *et al.* 2012, Singh *et al.* 2006, Brady *et al.* 2004, Banerjee *et al.* 2002). According to Fleming *et al*. in 2010 there were over 30-nitrile containing pharmaceuticals on the market and anothereleven nitrile-containing drugs were approved since 2011 (Dooley-Cullinane *et al.* 2016). One well known example of chemoenzymatic synthesis relates to the API Atorvastatin, for the cholesterol lowering drug Lipitor® (Gong *et al.* 2013, Solano *et al.* 2012, Singh *et al.* 2006).

Previous work in the research group has revealed two very similar nitrilase genes, *nit*1 and *nit*2. The research presented by Coffey *et al.* in 2009 revealed that these genes are present in numerous bacteria with evidence indicating the nitrilase genes are in fact widespread in the environment. Additionally, published work by the research group has focused on the identification of clade-specific nitrilase genes towards biopharmaceutical and industrial applications (Dooley-Cullinane *et al.* 2018) This would prove useful in the detection of novel nitrilase genes with a proposed predicted activity based on their clade classification as detailed by Robertson *et al.* (2004).

This chapter details the cloning, overexpression and characterisation of the *nit*1 gene, originally identified by Lee Coffey *et al.* (2009).

6.2 Research Hypothesis and Aims

Hypothesis:

A novel nitrilase enzyme (Nit1) can be used for the hydrolysis of a range of industrially informed nitriles to their corresponding carboxylic acids.

Aim:

The aim of this research was to characterise the Nit1 enzyme with a focus on the enzyme's activity towards a number of industrially informed nitrile substrates.

6.3 Materials and Methods

6.3.1 Bacterial strain and culture conditions

The nitrilase gene (*nit*1) from *Burkholderia* sp. LC8 was previously ligated to the p-Drive vector by L. Coffey (WIT) (Qiagen, Cat. No. 231122) and transformed to Novablue Gigasingle® cells. The *nit*1 containing cells were cultivated in Luria-Bertani (LB) broth (Sigma-Aldrich, Cat. No. L3522) with 100 µg/mL ampicillin (Sigma-Aldrich, Cat. No. A0166) at 37 °C overnight at 200 rpm to allow for amplification of the *nit*1 gene and subsequent cloning into the pRSF-2 Ek/LIC vector (Novagen, Cat. No. 71364) expression vector.

6.3.2 PCR amplification and cloning

Each 15 µL PCR reaction mixture contained 7.5 µL Platinum™ SuperFi™ Green PCR Master Mix (ThermoFisher Scientific, Cat. No. 12359010), 15 $_\text{P}$mol of each primer and 1 µL *nit*1 containing p-drive clone cell suspension with a final cell O.D.600_{nm}= 0.04. The following PCR conditions were used: 1 cycle of 95 °C for 5 min, 30 cycles of 95 °C for 1 min, 56 °C for 1 min, 72 °C for 2 min, followed by 1 cycle of 72 °C for 5 min. The PCR product was cleaned using the Zymo Research clean and concentrator™ -5 (Zymo Research, Cat. No. D4013) as per the manufacturer instructions with elution in water. The resulting PCR product was incubated at 72 °C for 20 minutes in a 10 µL reaction containing 1U Taq polymerase (Amresco, Cat. No. N555) 0.2 mM dATP, 1.5 mM $MgCl_2$, 1X buffer and 1.0 µg DNA product for the addition of an A-overhang prior to ligation.

The pRSF-2 Ek/LIC vector (Novagen, Cat. No. 71364) was used for the N-terminal His-tagged expression of the nitrilase. Cloning procedure were followed as per manufacturer's instructions. The resulting clone was transformed into *E. coli* BL21 (DE3) as per manufacturer's guidelines for heat shock transformations.
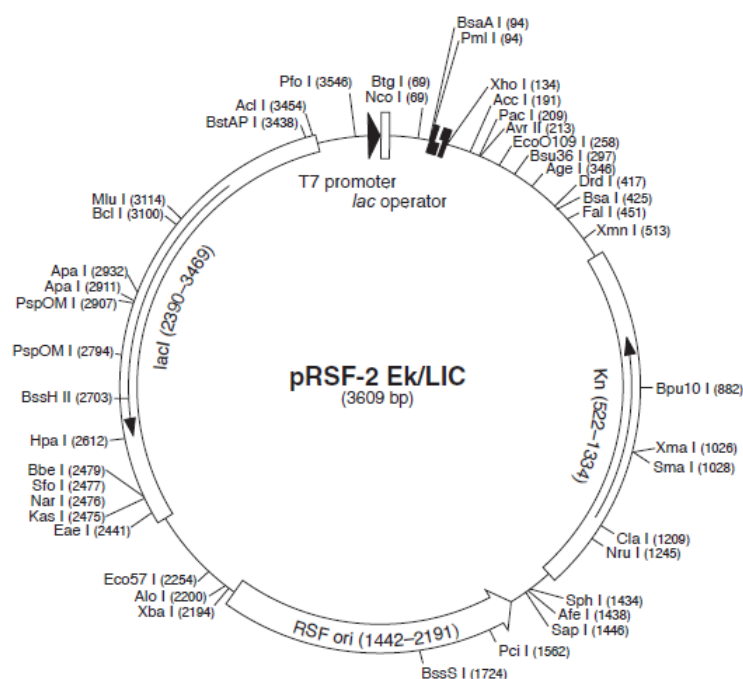
Figure 6.1 pRSF-2 EK/LIC (3609 bp) map displaying the multiple cloning site (MCS). The MCS is flanked with restriction sites, KanR (kanamycin resistance marker, RSF ori denotes the origin of replication derived from the RSF1030 plasmid. T7*lac* promoter to control transcription and, a lacI coding region i.e. the lac operon oppressor. Multiple restriction sites are evident throughout the vector map. (Novagen, 2018)

6.3.3 Protein induction, purification and gel analysis

6.3.3.1 Induction

The *E. coli* BL21 (DE3) clone containing the *nit*1 gene in the pRSF-2 Ek/LIC expression vector (Novagen, Cat. No. 71364) was grown overnight in LB broth with a kanamycin concentration of (w/v) of 30 µg/mL at 37 °C and 200 rpm. Following overnight cultivation, the cells were centrifuged at 3200 x g for 20 minutes at 4 °C. The cell pellet was washed in sterile LB broth and re-centrifuged as per the conditions noted above. The resulting cell pellet was resuspended in fresh LB broth with a kanamycin concentration of (w/v) of 30 µg/mL. To 200 mL of LB (in a 2 L flask) broth containing a kanamycin

concentration of (w/v) of 30 µg/mL, 200 µL of uninduced cells were added. This culture was grown at 37 °C and 200 rpm until an O.D.600$_{nm}$ of 0.6 to 0.8 was reached. Upon reaching the required O.D.600$_{nm}$ the flask was set on ice for 5 minutes. Post-cooling of the culture IPTG was added to a final concentration of 0.05 mM. The resulting culture was incubated overnight at 16 °C at 200 rpm.

6.3.3.2 His-tagged protein purification

The *E. coli* induced cells containing the His-tagged recombinant *nit*1 were centrifuged at 3200 x g for 20 minutes at 4 °C. The cells were lysed using the B-PER complete bacterial protein extraction reagent product (Fisher Scientific, Cat. No. 89821) as per the manufacturer's instructions. The His-tagged soluble protein was purified using the Ni–NTA purification system as per the manufacture's guidelines (Thermofisher Scientific, Cat. No. K95001). The purified protein was then desalted with Zeba™ Spin Desalting Columns, 30K MWCO (ThermoFisher Scientific, Cat. No. 87766) as per the manufacturer guidelines. The purified desalted nitrilase was concentrated by ultrafiltration with the Corning® Spin-X® UF concentrators (Sigma-Aldrich, Cat. No. CLS431485). Protein concentration was determined by the UV measurement at 280 nm using the Nanodrop® ND-1000 spectrophotometer. All purification procedures were carried out at 4 °C.

6.3.3.3 SDS-PAGE analysis

To confirm purity and determine protein size of the His-tagged recombinant Nit1 protein, SDS-PAGE analysis was carried out using 12% polyacrylamide gel. The resolution gel (15 mL) was prepared as follows:5.65 mL deionized water, 6.0 mL 30% acrylamide/bis solution (Biorad, 1610154), 3.04 mL 1.87 M Tris-HCl buffer pH 8.8, 150 µL SDS (Sigma-Aldrich 862010), 150 µL 10% ammonia persulfate (Fisher Scientific, Cat. No. BP179-100), 6.0 µL Temed (Sigma-Aldrich, Cat. No. T9281). The stacking gel consisted of 3.0 mL of distilled water, 850 µL of 30 % acrylamide/bis solution, 1.04 mL of 0.6 M Tris-HCl buffer pH 6.8, 50 µL 10% SDS, 50 µL 10% ammonia persulfate and5.0 µL Temed. Sample buffer (20 mls) consisted of 2 mls 0.5% bromophenol blue Sigma-Aldrich, Cat. No. 114391), 2 mls 0.6 M Tris-HCL pH 6.8, 2 g sucrose (Sigma-Aldrich, Cat. No. 84097), 0.2 mls 10% SDS and 100 µl β-mercaptoethanol (Sigma-Aldrich, Cat. No. M6250).

For some protein analysis the Bolt™ system (ThermoFisher, Cat. No. NW0412AIB2) was utilized. The SDS- PAGE gel analysis was performed on the premade 4- 16 % Bis-Tris plus gels (ThermoFisher Scientific, Cat. No. NW04120BOX) using 1X Bolt® MES running buffer (ThermoFisher Scientific, Cat. No. B0002). The samples were prepared using the 4X Bolt® LDS sample buffer (ThermoFisher Scientific, Cat. No. B0007) and the Bolt® 10X sample reducing agent (ThermoFisher Scientific Cat. No. B0009) both at a 1X concentration. The SeeBlue™ Plus2 (Thermo Fischer Scientific, Cat No. LC5925) pre-stained protein standard was used as a protein ladder (Figure 6.2).
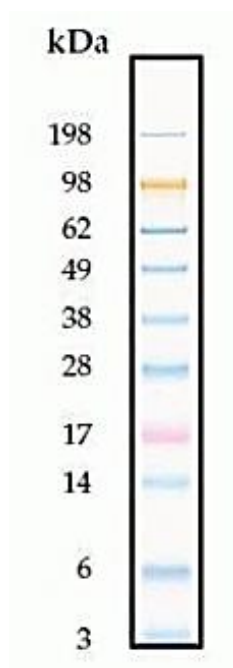
Figure 6.2 SeeBlue™ Plus2 pre-stained protein standard displaying a protein separation range of 3 to 198 kDa (ThermoFisher Scientific, 2018).

The resulting Bolt™ gels were stained and destained utilising the InstantBlue™ stain (Sigma-Aldrich, Cat. No. ISB1L-1L) as per the manufacturer's guidelines.

6.3.3.4 Native-PAGE™ analysis

For Native-PAGE™ protein analysis a mini gel tank system (ThermoFisher, Cat. No. A25977) was utilized. The Native-PAGE™ gel analysis was performed using the premade 4- 16% Bis-Tris gels (ThermoFisher Scientific, Cat. No. BN1002BOX) using the NativePAGE™ Running Buffer kit (ThermoFisher Scientific, Cat. No.BN2007). The protein samples were prepared using the NativePAGE™ Sample Buffer (4X) (ThermoFisher Scientific, Cat. No. BN2003) and the Bolt® 10X sample reducing agent (ThermoFisher Scientific Cat. No. B0009) and the NativePAGE™ 5% G-250 Sample Additive (ThermoFisher Scientific Cat. No. BN2004). The NativeMark™ Unstained Protein Standard (ThermoFisher Scientific Cat. No. LC0725) was used as a protein ladder (Figure 6.3).
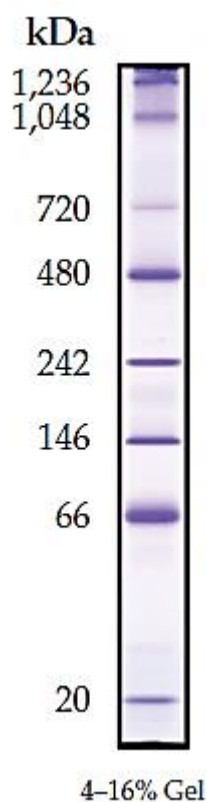
Figure 6.3 The NativeMark™ Unstained Protein Standard displaying a protein separation range of 20 to 1,236 kDa.

6.3.4 Biotransformations with the Nit1 purified His-tagged recombinant enzyme

The standard nitrilase assay was conducted in a 200 µl reaction mix in microtiter plates containing sodium phosphate buffer (100 mM, pH 7.0), 10 mM nitrile, final concentration of nitrilase at 0.015 mg at 40 °C for 2 hours. To quench the reaction 50 µL of 250 mM HCl was added and acetonitrile (3:1) was also added to 200 µL of the reaction. Samples were centrifuged at 15,000 x g for 10 minutes to pellet the denatured protein. The production of the corresponding carboxylic acid was determined by high-performance liquid chromatography (HPLC). Activity screening towards mandelonitrile was the exception to the above where 150 uL of 1 M HCl was added to the biotransformation and no ACN was used.

6.3.5 Optimum temperature and pH determination

To characterise the enzyme's substrate range, the optimum temperature and pH were first determined. Biotransformations were carried out as per Section 6.3.5. To determine the optimum temperature for the recombinant Nit1 recombinant enzyme the biotransformation reaction was carried out as described in Section 6.3.5 in the temperature range of 0 °C to 70 °C at the following with 10 mM mandelonitrile (Sigma-Aldrich, Cat. No. 116025) as the substrate.

A pH range of 4.0 to 9.0 was assessed at the following: 0.1 M sodium acetate buffer (4.0, 4.6, 5.0 and, 6.0), 0.1 M potassium phosphate buffer (6.0, 6.4, 7.0, 7.5 and, 8.0) and 0.1 M tris-HCl buffer (8.0, 8.5 and 9.0) with enzyme assays performed in triplicate.

6.3.6 Analytical methods

The nitriles and corresponding acids were quantified by HPLC using an xTerra C18 column 4.6 mm X 150 mm, 5µm (Waters, 186000490) on an Agilent 1100 HPLC, with the exception of mandelonitrile which was quantified using a Hyperclone ODS (150 x 4.6 mm; 3 µm) (Phenomenex, Cat. No. 00D-4419-E0) on an Agilent 1200 HPLC as per the described conditions below.

6.3.6.1 Analysis of mandelonitrile and mandelic acid (method developed by Andrew Quigley, WIT)

Separation of mandelonitrile and mandelic enantiomers was performed on a C18 Hyperclone ODS (150 mm x 4.6 mm; 3 µm) (Phenomenex, Cat. No. 00D-4419-E0) with a mobile phase containing 5 mM hydroxypropyl-β-cyclodextrin (TCI, Cat. No. H0979) of (A) H2O with 0.1% formic acid and (B) ACN with 0.1 % formic acid (98:2) at 210 nm.

The separation of total nitrile and total acid products was performed using gradient separation on a C18 Hyperclone ODS (150 x 4.6 mm; 3 µm) with mobile phase (A) H20 with 0.1% formic acid and (B) ACN with 0.1% formic acid with a gradient of 10% to 50% over a run time of 20 minutes. The flow rate was set to 1.0 mL/min with a 20 µL injection volume and column temperature set to 10 °C with UV detection at 210 nm. A standard curve for mandelonitrile and mandelic acid were created (Figure 6.4 and Figure 6.5).
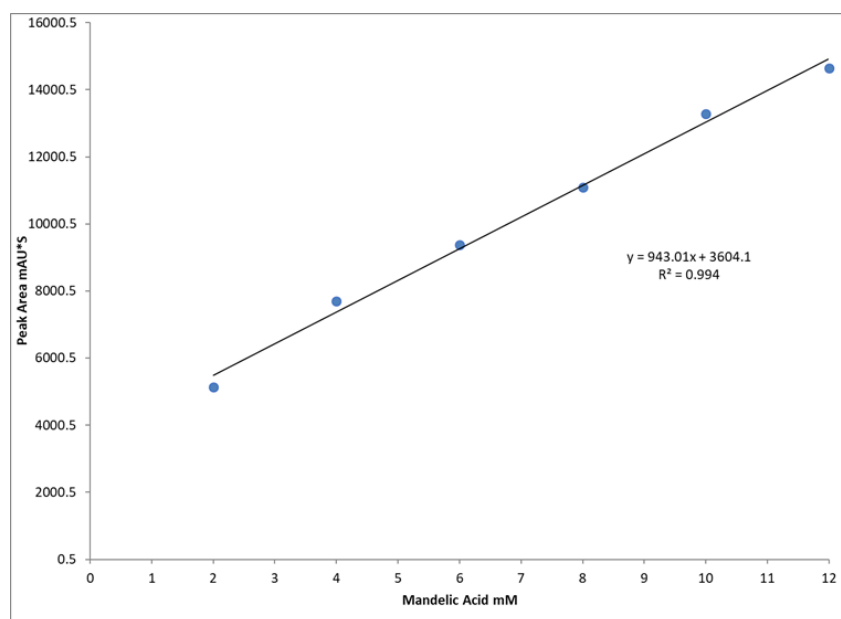
Figure 6.4 Standard curve of mandelic acid detected on C18 Hyperclone ODS (150 x 4.6 mm; 3 µm) with mobile phase (A) $H_2$0 with 0.1% formic acid and (B) ACN with 0.1% formic acid with a gradient of 10% to 50% over a run time of 20 minutes at 210 nm.
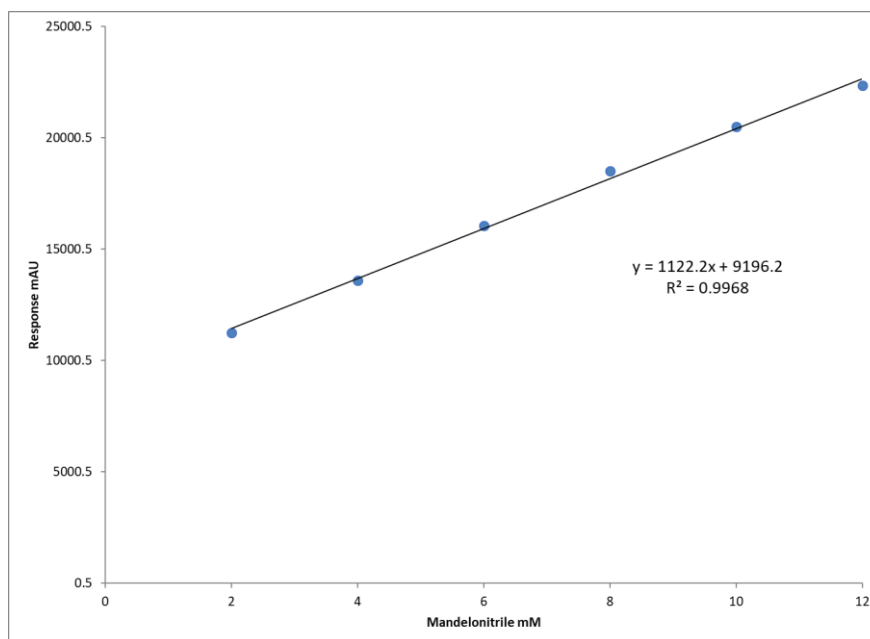
Figure 6.5 Standard curve of mandelonitrile detected on C18 Hyperclone ODS (150 x 4.6 mm; 3 µm) with mobile phase (A) H_2O with 0.1% formic acid and (B) ACN with 0.1% formic acid with a gradient of 10% to 50% over a run time of 20 minutes at 210 n

6.3.6.2 Analysis of 4-Hydroxyphenylacetonitrile (4-HPAN) and 4-Hydroxyphenylacetic acid (4-HPAA)

Separation of 4-HPAN and 4-HPAA was carried out with (20:80) (A) ACN and (B) H_2O with 0.3 % phosphoric acid (Sigma-Aldrich, Cat. No. V800287) and a flow rate of 0.8 mL/min, 20 µL sample injection with UV detection at 270 nm. Standard curves were prepared at a concentration of 2, 4, 6, 8, 10 mM of 4-HPAN (Figure 6.6) and 4-HPAA (Figure 6.7). The R^2 value for the nitrile and corresponding acid standard curves were > 0.99, allowing for quantitative analysis for subsequent biotransformations.

184
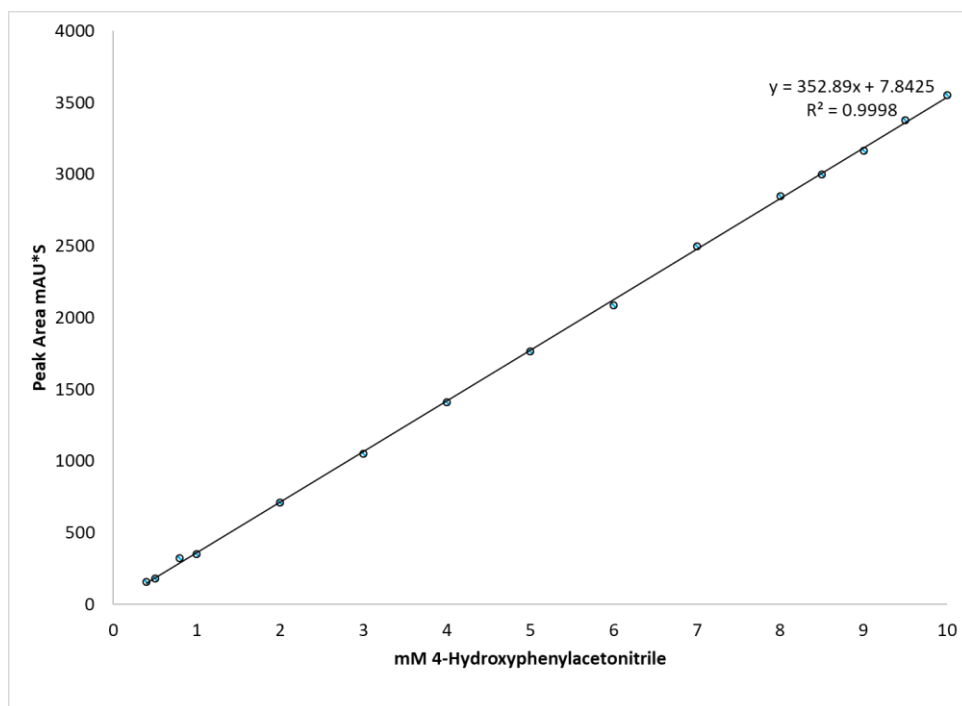
Figure 6.6 Standard curve of 4-hydroxyphenylacetonitrile detected on an xTerra C18 column (4.6 mm X 150 mm, 5µm) on an Agilent 1100 HPLC with a flow rate of 0.8 mL/min of mobile phase (20:80) (A) ACN and (B) H_2O and 0.3% phosphoric acid at 270 nm.
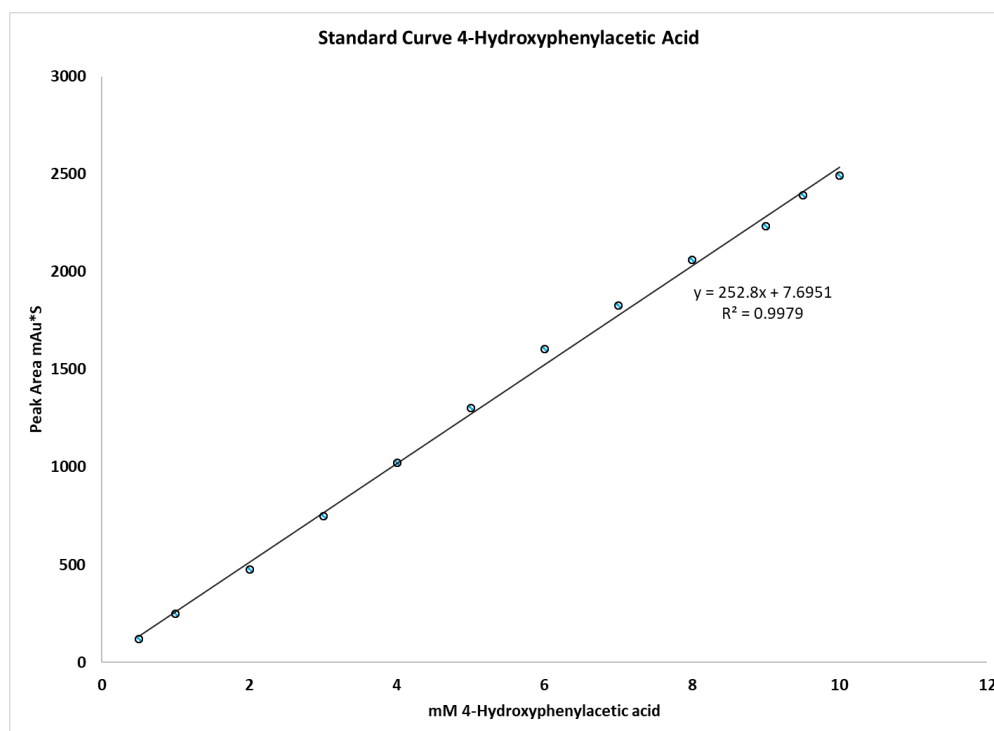
Figure 6.7 Standard curve of 4-hydroxyphenyl acetic acid detected on an xTerra C18 column (4.6 mm X 150 mm, 5µm) on an Agilent 1100 HPLC conditions as per Figure 6.3.

6.4 Results

6.4.1 Endpoint PCR amplification of *nit*1

The full *nit*1 gene was amplified from an *E. coli* Novablue Gigasingle™ (Novagen, Cat. No. 71227) clone which contained the nitrilase gene previously cloned into the p-Drive vector (Qiagen, cat. no. 231122). The resulting PCR product was analysed via gel electrophoresis on a 1.0 % agarose gel with a NEB 1 kB marker (NEB, N3232S) and displayed a band at the expected size of 1020 bp (Figure 6.8).
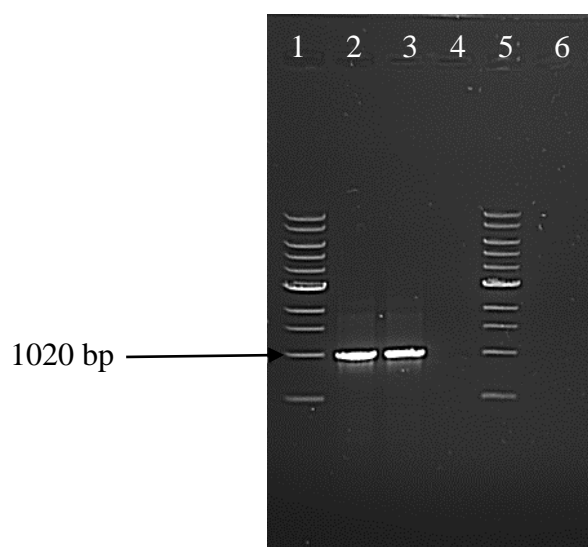
Figure 6.8 Agarose gel analysis of the complete *nit*1 gene PCR product.
Lane 1- NEB 1kB marker, lane 2- *nit*1 PCR, lane 3- *nit*1 PCR, lane 4- negative control, lane 5- NEB 1Kb marker.

6.4.2 Re-cloning of *nit*1 into an expression vector

The amplified *nit*1 gene was successfully ligated into the pRSF-2 Ek/LIC vector (Novagen, Cat. No. 71364) (Figure 6.1) and transformed into *E. coli* BL21 (DE3) competent cells. To confirm that the *nit*1 gene was successfully transformed, a plasmid extraction of transformed clones was performed (Figure 6.9) for confirmation PCR and subsequent sequencing of a selection of transformants.
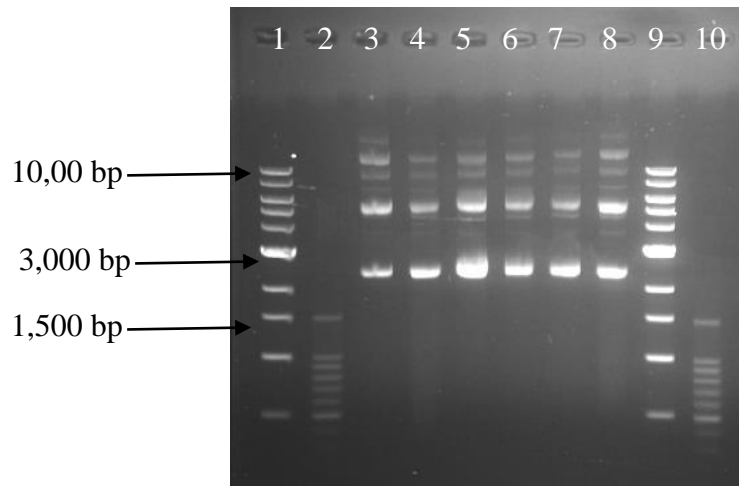
Figure 6.9 Agarose gel analysis of plasmid extractions of *E. coli* BL21 (DE3) clones transformed with *nit*1 ligated pRSF-2 Ek/LIC vector.
Lane 1- NEB 1 kB marker, lane 2-100 bp marker (Promega); lane 3- *nit*1 BL21 clone 1, lane 4-*nit*1 BL21 clone 2, lane 5- *nit*1 BL21 clone 3, lane 6- *nit*1 BL21 clone 4, lane 7- *nit*1 BL21 clone 5, lane 8- *nit*1 BL21 clone 6, lane 9- NEB 1 kB marker and, lane 10- 100 bp marker (Promega).

The plasmid extractions of the BL21 (DE3) clones were then used as template DNA for an endpoint PCR amplification with vector-specific primers to confirm successful ligation of the full *nit*1 gene in the pRSF-2 Ek/LIC vector.
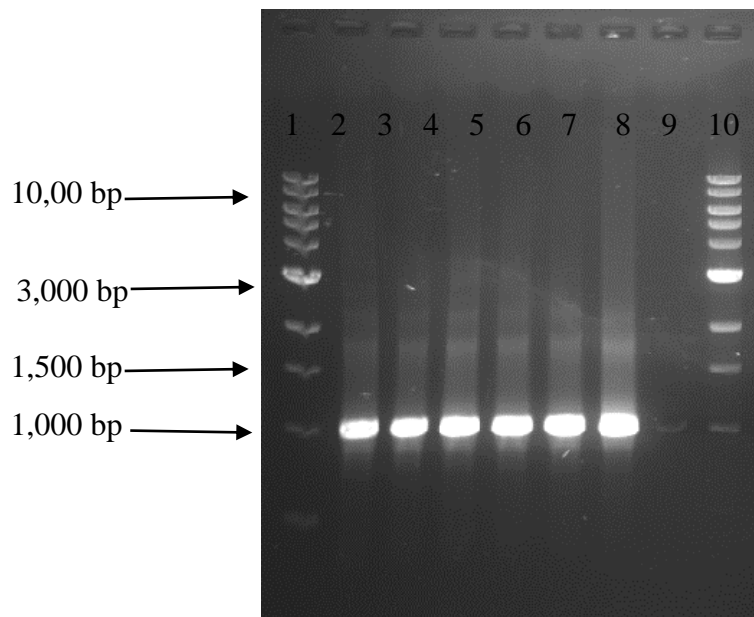
Figure 6.10 Agarose gel analysis of the full *nit*1 PCR product with plasmid extraction from *E. coli* BL21 (DE3) transformants as template DNA

Lane 1- NEB 1 kB marker, lane 2- *nit*1 BL21 clone 1, lane 3- *nit*1 BL21 clone 2, lane 4- *nit*1 BL21 clone 3, lane 5- *nit*1 BL21 clone 4, lane 6- *nit*1 BL21 clone 5, lane 7- *nit*1 BL21 clone 6, lane 8-N/A, lane 9- NEB 1 kB marker.

6.4.3   Induction of the Nit1 protein

To assess if the induction of the Nit1 protein was successful, SDS-PAGE gel analysis of the uninduced and induced cultures was performed using the BOLT™ system as per Section 6.3.3 (Figure 6.11).
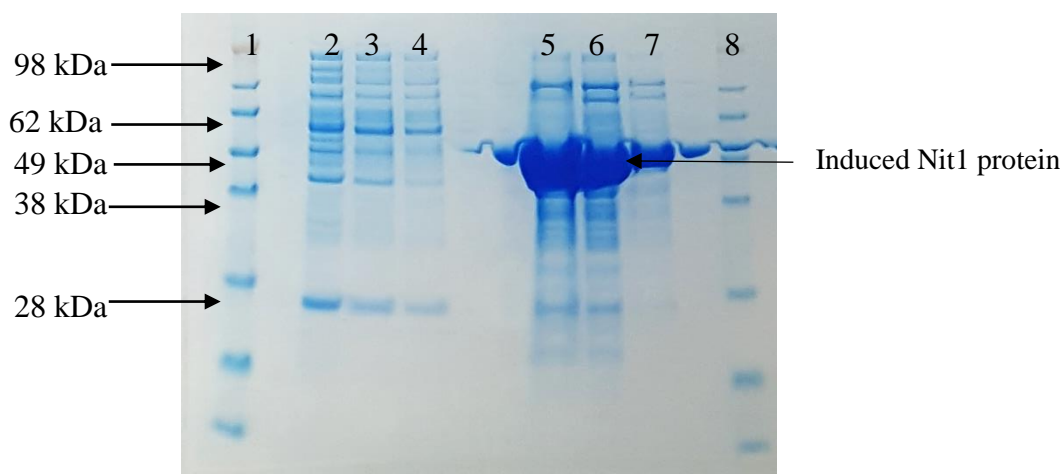
Figure 6.11 Bolt™ SDS-PAGE gel analysis of soluble portion of the uninduced and induced cultures of NIT1 post staining with InstantBlue®.

Lane 1- SeeBlue™ Plus2 pre-stained protein standard, lane 2- uninduced Nit1 culture 1 neat, lane 3- uninduced Nit1 culture 10^{-1}, lane 4- uninduced Nit1 culture 10^{-2}, lane 5- induced Nit1 culture 1 neat, lane 6- induced Nit1 culture 10^{-1}, lane 7- induced Nit1 culture 10^{-2}, lane 8- induced Nit1 culture 10^{-3}, SeeBlue™ Plus2 pre-stained protein standard.

6.4.4 Purification of the recombinant His-tagged Nit1 protein

The His-tagged recombinant Nit1 was purified as per Section 6.3.3.2. SDS-PAGE analysis displayed a purified protein post-affinity chromatography and de-salting. During the analysis it was evident that there were some *E. coli* proteins present in the purified batch of Nit1 protein (Figure 6.12).
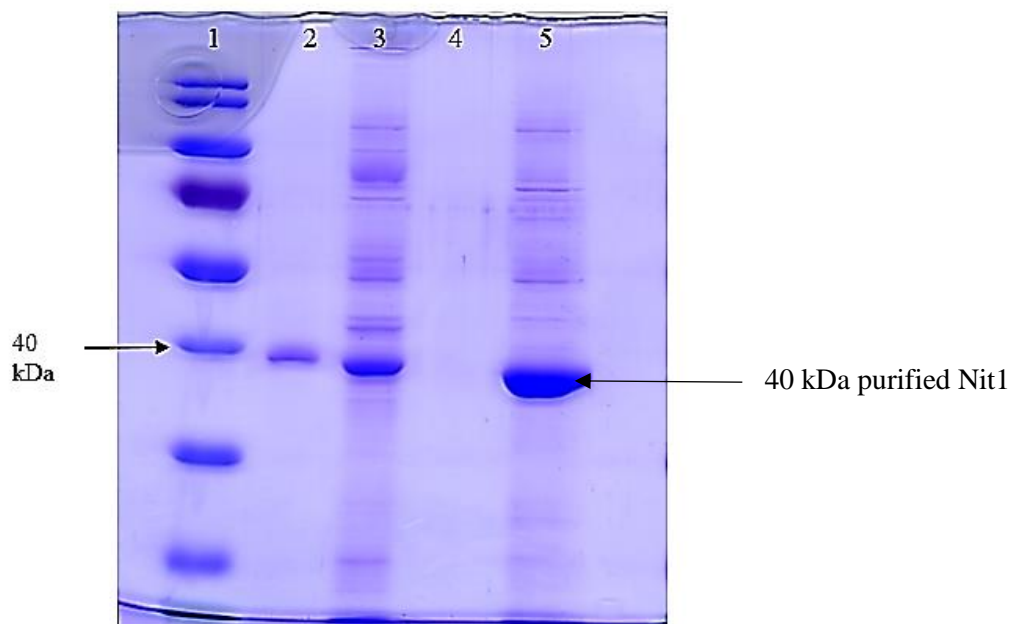
Figure 6.12 SDS- gel analysis of the Nit1 enzyme post purification and de-salting.
The enzyme measured at 40 kDa as expected. Lane 1; Protein marker, lane 2; Desalted purified
*Nit*1 (His-tagged); lane 3; *Nit*1 induced cell lysate 0.03 mM IPTG overnight at 16 °C at 200 rpm;
lane 5; *Nit*1 pre-purification 0.05mM IPTG overnight at 16 °C at 200 rpm.

6.4.4.1 Native-PAGE™ analysis

The Native-PAGE™ gel analysis of the purified Nit1 enzyme displayed a ladder profile
within the size range of 146 kDa to 720 kDa (Figure 6.13). It is proposed that each subunit
is approximately 44 kDa including the His-tag. From the banding profile observed it is
hypothesised that the typical nitrilase dimer subunit formation is observed with the
smallest 4 protein bands measuring approximately 156 kDa, 264 kDa, 352 kDa, 440 kDa
and 528 kDa respectively. It is also thought that the largest visible band, measuring
approximately 704 kDa, is 16 subunits. There are some faint bands evident within the
proposed 528 kDa and 704 kDa range. It is hypothesised that these bands represent
monomer subunits which may have formed due to degradation of the dimer subunits
within the cell, or during protein purification.
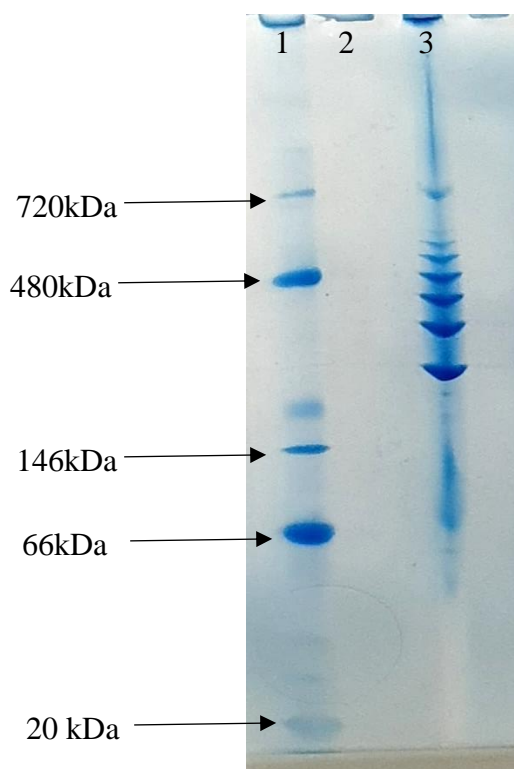
Figure 6.13 Native-PAGE™ gel analysis of the purified NIT1 post staining with InstantBlue®. Lane 1; NativeMark™ Unstained Protein Standard, lane 2; N/A, lane 3; Purified Nit1 enzyme

6.4.5 Temperature optimization

The optimum temperature of the Nit1 purified protein was determined. Nit1 enzymatic activity was measured at various temperatures (0–70 °C) in 100 mM potassium phosphate buffer (pH 7.0) with 10 mM mandelonitrile as the substrate. The optimal temperature was found to be 40 °C. When the temperature was above 40 °C, the activity of Nit1 decreased sharply (Figure 6.14 and Table 6.1). Assays were performed in triplicate.

Table 6.1 Temperature optimization of the Nit1 recombinant protein (peak area).

Temperature (°C)		0	4	10	20	25
Average 1		720.12	1012.13	1585.32	2446.36	2502.65
Average 2		721.23	1010.23	1585.33	2445.32	2503.18
Average 3		720.13	1011.23	1586.23	2445.38	2501.68
Standard Deviation		0.52	0.78	0.43	0.48	0.62
Average triplicate of assays		720.49	1011.20	1585.63	2445.69	2502.50
Relative Activity (%)		12.6	17.7	27.7	42.8	43.8
Temperature (°C)		30	40	50	60	70
Average 1		3538.32	5717.4	3359.8	2258.72	736.25
Average 2		3536.32	5714.52	3358.54	2259.62	736.22
Average 3		3539.31	5717.42	3356.52	2260.25	736.23
Standard Deviation		1.24	1.36	1.35	0.63	0.01
Average triplicate of assays		3537.98	5716.45	3358.29	2259.53	736.23
Relative Activity (%)		61.9	100.0	58.7	39.5	12.9
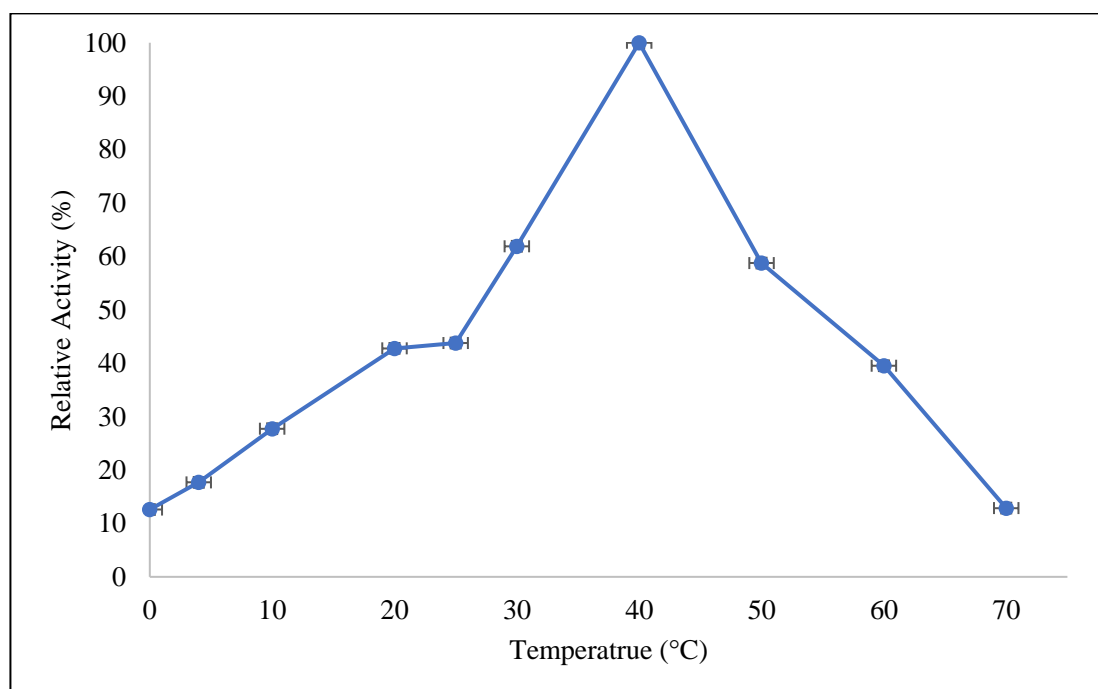
Figure 6.14 Temperature activity assay of Nit1 enzyme towards mandelonitrile (10 Mm) in 0.1 M potassium phosphate buffer pH 7.0. Error bars represent the standard error across the mean of three replicates.

6.4.6 pH optimization

The optimal pH of the Nit1 purified protein was determined in 100 mM sodium acetate buffer, 100 mM potassium buffer and 100 mM Tris-HCl buffer at various pH conditions as per Section 6.3.5, with 10 mM mandelonitrile as the substrate. Nit1 exerted a higher activity between pH 6 and 7, and the highest activity was observed at pH 7.0 in 100 mM potassium phosphate buffer (chromatograms in appendix I) (Figure 6.15 and Table 6.2). To the best of our knowledge, nitrilases typically have the highest activity in a weakly basic environment. Assays were performed in triplicate.

Table 6.2 pH optimization of the Nit1 recombinant enzyme (peak area).

	0.1 M Sodium Acetate			
	4	**4.6**	**5**	**6**
Average 1	2286.76	2986.99	3346.3	3355.36
Average 2	2286.89	2985.91	3346.27	3358.63
Average 3	2286.35	2985.96	3345.94	3358.36
Standard Deviation	0.23	0.50	0.16	1.48
Average of triplicate assays	2286.67	2986.29	3346.17	3357.45
Relative Activity (%)	43.5	57.2	64.1	64.3

	0.1 M Potassium Phosphate				
	6	**6.4**	**7**	**7.5**	**8**
Average 1	4274.95	4660.15	5252.16	5222.57	5160.99
Average 2	4273.15	4659.94	5250.15	5223.23	5161.19
Average 3	4272.91	4660.47	5252.32	5222.35	5160.87
Standard Deviation	0.91	0.22	0.99	0.37	0.13
Average of triplicate assays	4273.67	4660.19	5251.54	5222.72	5161.03
Relative Activity (%)	81.8	89.2	100.0	99.5	98.3

	0.1 M Tris-HCL		
	8	**8.5**	**9**
Average 1	4810.12	3407.25	3103.12
Average 2	4813.69	3408.25	3103.25
Average 3	4812.28	3408.15	3101.32
Standard Deviation	1.47	0.45	0.88
Average of triplicate assays	4813.06	3407.88	3102.56
Relative Activity (%)	91.7	64.9	59.1
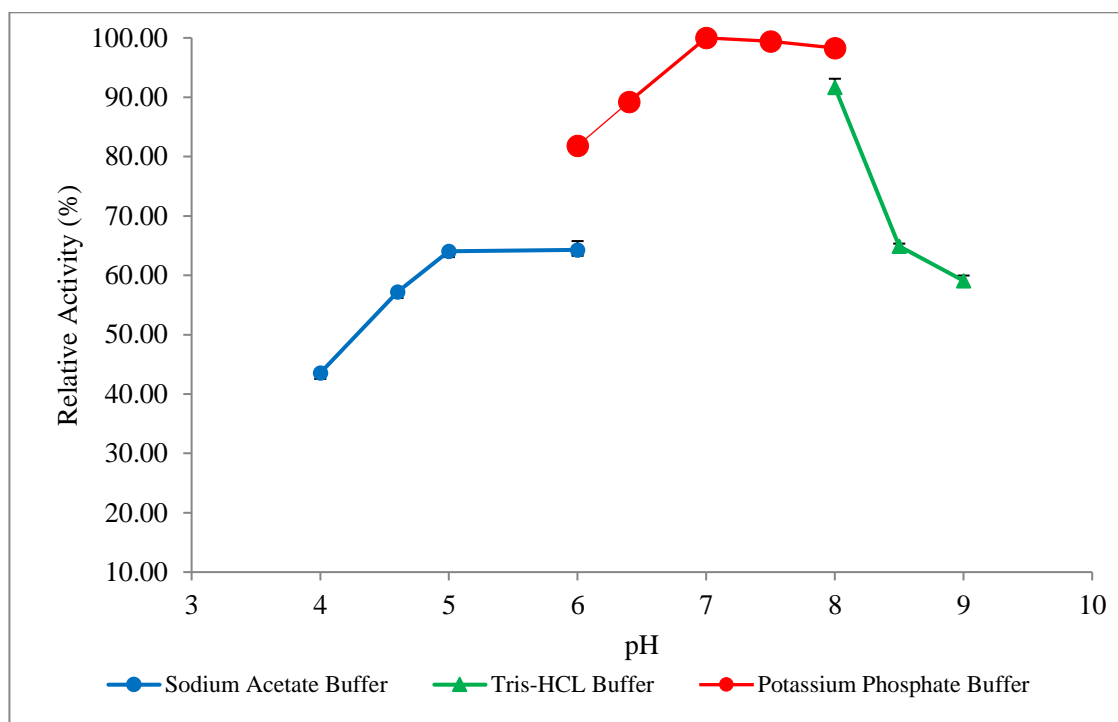
Figure 6.15 pH activity assay of Nit1 enzyme towards mandelonitrile (10 mM) in a pH range of 4.0 to 9.0 at 40 °C in 0.1 M potassium phosphate buffer pH 7.0. Error bars represent the standard error across the mean of three replicates.

6.4.7 Activity Analysis

The Nit1 enzyme displayed activity across a broad range of substrates. The activity of Nit1 towards the following substrates was determined via the Nessler's colorimetric assay: 3-HBN, 3-HGN, 3-HPPN, Benzonitrile (BENZO), 2-Phenylpropionitrile (2-PPN) and 2-Methylbutyronitrile (2-MBN) as detailed in Section 0 (Table 6.3). The Nit1 enzyme displayed the highest activity towards 3-Hydroxyphenylacetonitrile (4-HPAN) and 2PPN as determined by the 9.83 mM and 8.37 mM of acid formed.

Table 6.3 Activity analysis of the purified Nit1 enzyme towards a range of nitriles using the Nessler's ammonia assay.

Substrate	3HBN	3HGN	3HPPN	BENZO	2PPN	2MBN
Average 1	1.064	1.027	0.754	0.678	1.079	0.998
Average 2	1.064	1.025	0.748	0.702	1.098	0.987
Average 3	1.071	1.028	0.748	0.687	1.134	0.995
Standard Deviation	0.003	0.001	0.003	0.010	0.023	0.005
Average	1.07	1.03	0.75	0.69	1.10	0.99
mM Acid	8.02	7.72	5.65	5.19	8.37	7.47

Further activity analysis was performed using the analytical methods as detailed in Section 6.3.6. The activity of the Nit1 enzyme was determined against 4-Hydroxyphenylacetonitrile (Table 6.4). The Nit1 enzyme displayed almost full conversion of the enzyme over a 2-hour period.

Table 6.4 Activity analysis of the purified Nit1 enzyme towards 4-HPAN (peak area response) using the analytical method described in Section 6.3.6.2.

4-HPAN	Peak Area	mM Acid
Average 1	623.21	9.83
Average 2	622.25	9.82
Average 3	623.28	9.83
Standard Deviation	0.47	0.01

The Nit1 enzyme has displayed a high activity towards 4-HPAN, 3-HBN and 2-PPN. In addition to the above substrates, the enzyme's affinity for mandelonitrile was also assessed. The Nit1 enzyme displayed activity towards mandelonitrile with 7.32 mM of mandelic acid produced over 2 hours (Table 6.5).

Table 6.5 Activity analysis of the purified Nit1 enzyme towards mandelonitrile (peak area response) using the analytical method described in section 6.3.6.1.

	Mandelic acid	**Mandelonitrile**
Average 1	5252.16	8955.24
Average 2	5252.15	8955.59
Average 3	5252.81	8955.31
Standard Deviation	0.31	0.19
Total Amended Average	10504.75	13433.07
mM	7.32	3.81

Enantioselectivity screening was carried out on mandelonitrile (10 mM) and analysed via HPLC using a chiral mobile phase additive to facilitate the separation of mandelic acid enantiomers as described in Section 6.3.6.1. The purified Nit1 enzyme displayed the production of (R)-mandelic acid > 99.9 ee% Figure 6.16).
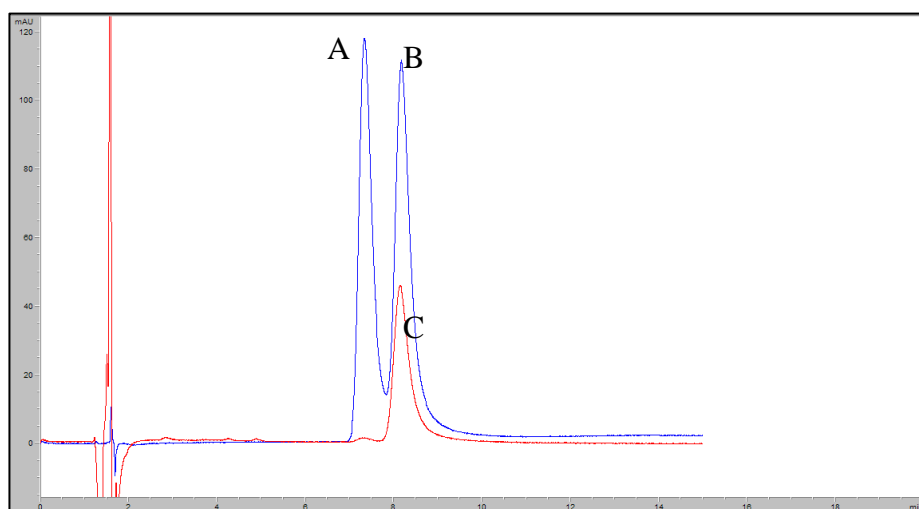
Figure 6.16 Chiral separation of mandelic acid standards (blue) with (S)-mandelic acid (A) eluting prior to (R)-mandelic acid (B) on a C18 column. Chromatogram displays (R)-mandelic acid (C) as the product of the product of a biotransformations with purified NIT1 enzyme with an ee % of > 99.9.

Mobile phase (A) H_2O with 0.1 % formic acid and (B) ACN with 0.1% formic acid (98:2) and 5 mM hydroxypropyl-β-cyclodextrin, 1.0 mL/min flowrate with UV detection at 210 nm. Chromatogram displays (R)-mandelic acid (C) as the product of the product of a biotransformations with purified NIT1 enzyme with an ee % of > 99.9.

The Nit1 enzyme displayed activity towards a range of aliphatic and aromatic nitriles. The relative activity of the Nit1 enzyme is shown in

Figure **6.17**.
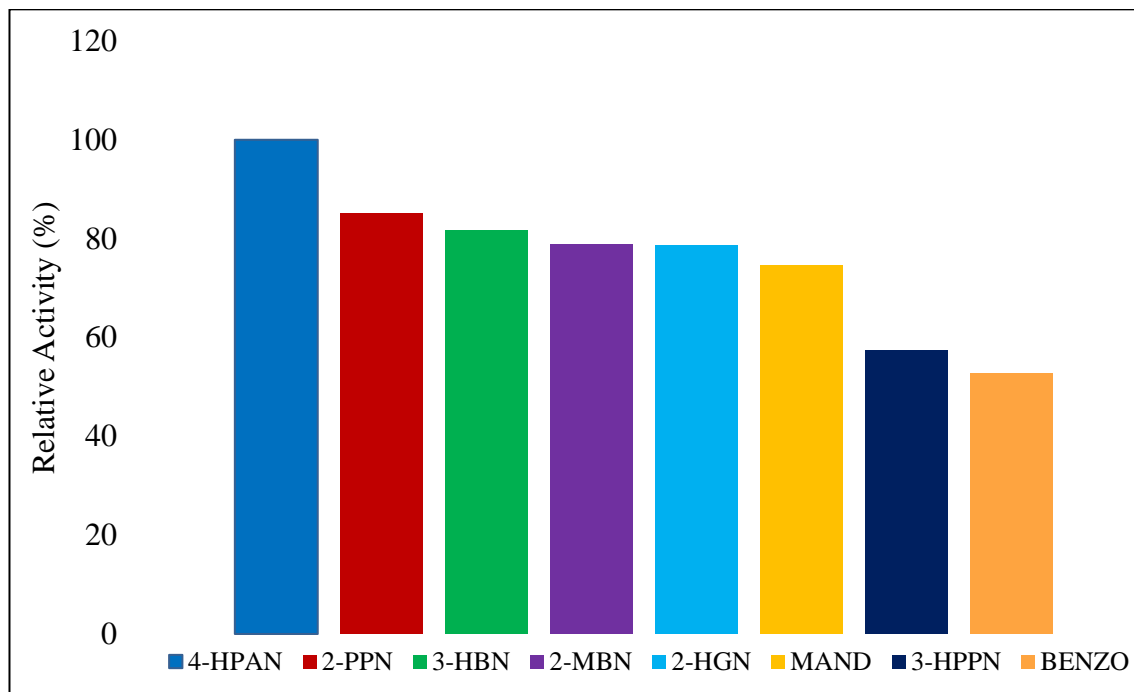
Figure 6.17 Relative activity of Nit1 towards a range of nitrile substrates.

6.5   Discussion and Conclusion

The Nit1 enzyme was cloned into the pRSF-2 EK/LIC expression vector to facilitate induction and subsequent His-tag purification. The recombinant Nit1 clone was induced overnight at a previously optimized temperature of 16 °C to facilitate a slower level of induction of protein expression. *E. coli* is commonly used for the expression and production of non-native *E. coli* proteins. *E. coli* has an optimum temperature of 37 °C, which can be used for a fast level of expression of non-native *E. coli* proteins. However, this often causes issues relating to the formation of inclusion body formation as the *E. coli* cell is producing the protein faster than the cell can accommodate correcting/folding, thus resulting in inactive, in-soluble protein. Decreasing the temperature during induction slows down the rate at which *E. coli* cells produce the protein. As the level of production is slowed so too is the rate at which the *E. coli* cell folds the non-native protein which it is producing. This results in a slower level of protein production, but a higher level of soluble protein. Based on this and the previous optimisation study, the induction was performed at 16 °C in order to increase the amount of soluble protein by reducing the level of inclusion body formation.

The recombinant clone was induced with both 0.03 mM and 0.05 mM of IPTG with the soluble fractions analysed by SDS-gel (Figure 6.11). The 0.05 mM IPTG saw increased induction and production of soluble protein and so it was utilized as the induction concentration throughout. The purified protein was analysed via SDS-page gel electrophoresis with the His-tagged protein measuring approximately 44 kDa (including the histidine tag), with nitrilase enzymes commonly found to measure 32 to 45 kDa which aggregate to form active homoligomers (Thuku *et al.* 2009, Banerjee *et al.* 2002, O'Reilly and Turner 2003). The nitrilase from *Labrenzia aggregate* (LaN) displayed measured approximately 37 kDa via SDS-Page (Zhang *et al*. 2012) with *the Burkholderia cenocepacia* J2315 also measuring 37 kDa (Wang *et al.* 2013). The purification of the Nit1 recombinant protein did in some cases result in the co-purification of *E. coli* proteins at low levels. It is proposed that these low-level contaminating proteins are from the *E. coli* proteome, as some endogenous *E. coli* proteins are naturally rich in histidine residues (Anderson *et al.* 2014), thus facilitating their binding to the purification resin and in turn co-purification.

Native-PAGE analysis was also performed to assess the native form of the Nit1 recombinant protein. From Figure 6.13, it can be seen that various subunit sizes are evident on the Native-PAGE gel. Based on the gel, it is proposed that the 5 subunits within the range of 200 kDa to 528 kDa are composed of multiples of dimers, with each monomer measuring approximately 44 kDa each. There are also 3 faint subunit bands evident above the 480 kDa band on the Nativemark™ protein ladder. It is proposed that these 3 faint bands once formed one of the larger aggregated forms of the Nit1 enzyme, however through purification or inability to self-associate within the cell during induction it may have lost one of the 44 kDa subunits from the larger dimer units. It is unknow if this 'ladder' of oligomers is present in the cell pre-lysis. If this 'laddering' is in the cell pre-lysis this would be due to a lack of formation and aggregation. It is also possible that the 'laddering' is due to the breakdown of the larger intact form of the enzyme upon his-tag purification. Whole cell-lysates could be used for assessment to analyse the cellular contents pre-his-tag purification; however, this may still lead to the 'laddering' effect.

It is not known if each band of the laddered native enzyme is active, or if perhaps only one band is the active subunit. Previous studies have shown nitrilase enzymes to be active with as little as four subunits, such as the nitrilase from *Fusarium oxysporum f.sp. melonis* (Goldlust and Bohak 1989). In order to assess if the Nit1 enzyme is active in each of the subunit formations each of the bands could be assessed for activity. This would involve re-analysing the purified Nit1 protein on a higher percentage gel to allow for sufficient separation of the bands. Once sufficient separation was achieved the subunit bands would be gel extracted and an in-gel activity assay would be performed.

In addition to this further analysis should be performed in order to determine the true form of the native Nit1 enzyme using X-ray crystallography. The first nitrilase crystal structure elucidated was the Nit formation of the NitFhit Rosetta Stone protein (Pace *et al.* 2000). The Fhit is encoded as a fusion protein to the Nit nitrilase superfamily member. Upon crystallisation and x-ray analysis, it was seen that the Nit portion of the fusion protein displayed the α–β–β–α sandwich fold with a Cys–Glu–Lys catalytic triad (Pace *et al.* 2000). To date, the crystal structure of bacterial nitrilases are not commonly studied due to difficulties encountered. One such difficulty is referred to as 'twinning' *(*Raczynska *et al.* 2011).

More recently, Raczynska *et al*. (2011) analysed several crystal structures of the thermoactive nitrilase from *Pyrococcus abyssi* (PaNit), the first microbial nitrilase which was shown to be active as a dimer and not a larger complex. Though crystal structures were generated and analysed, the authors noted the fractions of the crystal structure could not be de-twinned and so the high-resolution range for the data is largely incomplete. However, it was shown that the PaNit did contain the typical α–β–β–α sandwich fold with a Cys–Glu–Lys catalytic triad (Raczynska *et al*. 2011). Based on the limited research available on the analysis of microbial nitrilases via X-ray crystallography it was clear that the body of work required to elucidate the true crystal structure would be extensive. It is also believed that this work would need to precede enzyme kinetics. However, it is proposed that this research would be carried out in future to further kinetics and advance the performance of the enzyme.

To assess the optimum activity conditions for the purified Nit1 enzyme a biotransformation was performed in 0.1 M potassium phosphate buffer at pH 7.5. The purified recombinant enzyme displayed the highest activity towards mandelonitrile at an optimum temperature of 40 °C, within the range of 30 °C to 50 °C as typically seen with nitrilase enzymes from mesophilic organisms (Zhang *et al*. 2012). More commonly, nitrilase enzymes have been shown to be active within the temperature range of 30 °C to 65 °C (O'Reilly and Turner 2003). In 2016, a thermophilic nitrilase, NitM24D13, isolated from *Pyrococcus* sp. M24D13 was shown to have an optimum activity at 85 °C (Dennett *et al.* 2016). Following the identification of the Nit1 optimum temperature, the optimum pH as assessed. Nitrilase enzymes generally display activity with the pH range of 5.5 to 8.0, wwith the purified Nit1 showing an optimum pH of 7.0 in 0.1 M potassium phosphate buffer. Nitrilase enzymes commonly display an optimum activity in the range or 7.0 to 8.5. The purified nitrilase from *Labrenzia aggregate* (LaN) displayed an optimum pH of 8.0 with an optimum temperature of 50 °C. The optimum temperature and pH of the Nit1 enzyme, makes it largely suitable for the production of chiral building blocks and pharmaceutical intermediates, such as mandelic acid.

The activity of the Nit1 protein was measured towards four industrially relevant nitriles namely 3HGN, 3HBN, 3-HPPN and mandelonitrile. The Nit1 enzyme displayed a higher affinity towards mandelonitrile than that of the other three nitriles aforementioned.

However, the Nit1 enzyme displayed a broad substrate scope, with activity demonstrated towards aliphatic, aromatic and di-nitriles (Figure 6.17).

Mandelic acid and in particular (R)-mandelic acid is an important pharmaceutical intermediate. The production of anti-tumour agents and anti-obesity drugs rely on the availability of the (R)-enantiomer (Surivet and Vatèle, 1998) (R)-mandelic acid has also been employed in the production of semi-synthetic cephalosporins (a class of β-lactam antibiotics) (Terreni *et al.* 2001). The Nit1 enzyme produced the (R)-mandelic acid (> 99.9 ee %), there are few published nitrilase enzymes which demonstranstrate a high level of enantioselectivity for (R)-mandelonitrile. The most notable purified or recombinant nitrilases for (R)-mandelic acid production are as follows: *Alcaligenes* sp. ECU0401, *Burkholderia cenocepacia* J2315 and *Pseudomonas fluorescens* EBC191 at 97%, 97.4% and 31 % enantiomeric excess respectively (Wang *et al*. 2015, Zhang *et al*. 2011, Kiziak *et al*. 2005). The high level of enantioselectivity exhibited by Nit1 results in very low levels of the (R)-mandelonitrile remaining, meaning that the Nit1 has commercial potential for the production of the pharmaceutically relevant (R)-mandelic acid intermediate. Nitrilases with preferential hydrolysis for (S)-mandelonitrile are less common, are seen by the phylogenetic display constructed by Robertson *et al*. (2004) in Figure **3.1**. Mandelonitrile is known to be unstable in certain conditions and so it cycles through a process of forming benzaldehyde and cyanide once the mandelonitrile is removed from a reaction. In simple terms, this results in the formation of benzaldehyde and cyanide within the reaction. Once nitrile is used to make the corresponding (R)-mandelic acid by the Nit1 enzyme, the benzaldehyde and cyanide react and form more racemic nitrile. The nitrile and benzaldehyde/cyanide complex are always, in turn, in an equilibrium.

In conclusion, the Nit1 enzyme has large industrial/pharmaceutical potential for the production of intermediates. Now that the industrial potential has been realised, further analysis is warranted on the Nit1 enzyme. Firstly, enzyme kinetics should be performed to determine the rate of activity of the Nit1 protein. Additionally, it would be beneficial to perform both solvent and metal tolerance testing of the Nit1 enzyme. It would also be recommended that the enzyme undergoes further analysis to determine the enzyme native structure.

CHAPTER 7

FUTURE WORK

7 FUTURE WORK

In summary, a segment of the body of work presented details novel molecular methods for the detection of aldoxime dehydratase genes and clade-specific detection of nitrilase genes. The search for either whole cell or purified biocatalysts continues to facilitate the move towards a green chemistry approach for the production of bio-pharmaceutical compounds and fine chemicals. The development of such molecular methods will allow for the detection of these novel genes/enzymes towards these applications. Additionally, the research detailed the cloning, expression and characterisation of a novel nitrilase which displays enantioselectivity towards the industrially relevant mandelonitrile, displaying >99.9 ee% for the (R)-enantiomer of mandelic acid. Within the scope of the research was the aim to identify the novel nitrile metabolising genes of interest from *R. erythropolis* SET-1. A novel nitrilase partial sequence was discovered alongside a full nitrile hydratase/amidase gene cluster. Whole genome sequencing did not aid in the elucidation of the nitrilase discovered in *R. erythropolis* SET-1, however some potential routes to identification are discussed below.

The novel molecular method designed for the detection of novel aldoxime dehydratase gene allowed for the identification of 30 partial novel aldoxime dehydratase sequences. The next stage in the research would be to assay the whole cells on aldoxime substrates of interest to allow for the production of their corresponding nitriles. Upon the identification of a microbial cell with the desired activity, the complete aldoxime dehydratase gene would be amplified and cloned. As the partial sequences are known at present, it would be wise to design primers which would accommodate primer walking to elucidate the full sequence. As these known sequences are designed on a conserved region across the aldoxime dehydratase genes, these primers may also work across the larger batch of isolates screened by the real-time PCR assay. Post-cloning, the enzymes would be expressed and characterised towards the production of nitriles of interest. The best performing enzyme could then be exposed to directed evolution to produce an enzyme which is optimised towards the production of a pharmaceutically relevant precursor or product. As the body of work would increase the sample data and the sequence data available, it would be of interest to see if the aldoxime sequences elucidated

could perhaps be grouped in clades based on DNA sequence homology, allowing the design of a clade-specific PCR assay.

An in-house nitrilase clade specific PCR previously allowed for the identification of 15 novel nitrilase partial sequences. The phylogenetic analysis of the novel 15 nitrilase partial sequences to the database was performed in this body of work. Future work based on the foundation of this assay would involve the design of a real-time PCR assay with a similar application and a greater level of sensitivity. The partial nitrilases should have their full sequence elucidated and also be cloned and expressed with purifications allowing for full enzyme characterisation. Following the characterisation of these enzymes, directed evolution would be carried out. Post-generation of a mutant library a number of mutants with desired activity would be selected by subjecting the clones to a continuous culture allowing for tight control over the clones and allowing the tailoring of the desired conditions in which the clone would perform (pH, temperature, specific enantiomer, solvent tolerance etc).

This thesis also facilitated the discovery of a novel nitrilase partial sequence from a fosmid genomic library of *R. erythropolis* SET-1. Following the identification of the nitrilase partial sequence, attempts to elucidate the complete nitrilase gene were described. All attempts were unsuccessful; however, the following could be undertaken to elucidate the full nitrilase sequence. Firstly, a transposon mutagenesis library could be constructed (either using the fosmid DNA from the functional clone or *R. erythropolis* SET-1 genomic DNA). The transposon may interrupt the nitrilase at a different position to that of the known partial sequence, which allow for additional sequence to be elucidated by identifying mutants with knocked out activity. Should this sequence then be discovered, the newly identified sequence could be used to design a primer and then when coupled with the pre-existing primers perhaps an amplicon could be amplified, containing the unknown sequence (Figure 7.1). One limitation with using genomic DNA is that the transposon may in fact interrupt the nitrile hydratase which was detected and sequenced in Chapter 6, possibly leading to a false knockout regarding activity. This of course can also be a potential limitation should these genes be contained there.
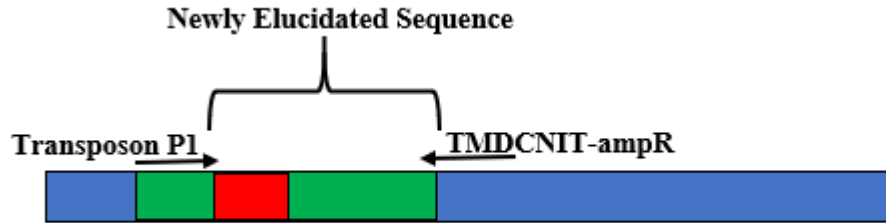
Figure 7.1 The above schematic details a proposed transposon library construction with the aim to elucidate the full nitrilase sequence as identified in chapter 3 from SET-1.

The blue represents either the fosmid insert from the SET-1 genomic library or SET-1 genomic DNA. The green region represents the known partial nitrilase sequence with the red indicting the transposon site.

Secondly, a DIG labelled probe could be constructed from either the novel nitrilase partial sequence identified or perhaps the newly identified sequence if the described transposon mutagenesis approach was successful. The genomic DNA from *R. erythropolis* SET-1 could then be digested by *Sau*3 or double digested with *Eco*RI and *Hin*DIII to produce a fragmented library. Should the digested band fragment containing the target sequence be identified by Southern blotting and probe hybridisation it would be gel extracted and ligated to a vector, such as the pGEM®-T easy vector or the TOPO-XL cloning vector. Attempts could then be made to sequence the nitrilase. If the flanking DNA fragments which contain the target DNA have been decreased in nucleotide length (bp) by the digestion with restriction enzymes, then the opportunity for secondary structures could also be reduced, perhaps aiding with sequencing.

Thirdly, pulsed field gel electrophoresis could be carried out with *R. erythropolis* SET-1. Research by the group has demonstrated that there is evidence of HGT of nitrilase genes from environmental isolates (Coffey *et al.* 2009), future work could also assess the feasibility that the partial nitrilase gene could be present on the plasmid detected in *R. erythropolis* SET-1, a plasmid potentially acquired via HGT event. To do this the large linear plasmid identified in this body of work would be gel extracted with the agarose subjected to β-agarase digestion. Once the large linear fragment DNA is recovered, the sample linear plasmid could be digested as described above and probed with the DIG labelled probe to confirm which band, if any, contained the target nitrilase. Additionally, the large linear plasmid DNA could be used to create a fosmid library which could then

be screened in the same approach as the fosmid genomic library. This workflow would take the same approach as described previously in that the clones would be subjected to functional selection. Parallel to this, plasmid curing could be attempted with *R. erythropolis* SET-1 cells to identify if loss of activity (i.e. loss of the active nitrile metabolising genes) can be correlated to loss of the plasmid. Should the plasmid be shown to contain the target gene, then the remainder of the genomic DNA could be omitted reducing the amount of screening required. Should this linear plasmid DNA contain the target gene then perhaps this sample could be used as the template for whole genome sequencing allowing for higher coverage of the DNA fragment. It would also reduce the amount of DNA used as the template for the sequencing, which in the case of high GC % microorganisms such as *Rhodococcus* sp. could reduce the chance of contig overlap and so reduce the chance of sequence being omitted from the final constructed library.

The data generated this body of work also identified a nitrile hydratase-amidase gene cluster. The nitrile hydratase gene cluster was identified with the α and β subunits of the nitrile hydratase and the amidase gene confirmed by gene sequencing. The nitrile hydratase gene, both alpha and beta subunit, could be amplified and cloned alongside the amidase. Following successful cloning, these enzymes would be characterised to see if their individual and/or combined activity profiles correlate with the previous activity observed with the whole cells of *R. erythropolis* SET-1. Cloning and expression of the enzymes would allow for the generation of a bi-enzymatic pathway, with potential enzyme immobilization allowing for a one-pot synthesis approach for compounds of interest. These enzymes would be subjected to directed-evolution post enzyme purification and characterisation, possibly allowing for an optimised enzyme system towards the production of either amides or carboxylic acid.

The additional data generated from WGS could also be utilized to identify any additional genes of interest either within or outside of the aldoxime-nitrile pathway. The whole genome sequencing also provided full sequence for an aldoxime dehydratase. Based on this *R. erythropolis* SET-1 could be tested on several aldoximes, and if it demonstrated promising activity, the aldoxime dehydratase can readily be cloned with the sequence elucidated.

Furthermore, there is also a high-demand for transaminase enzymes, towards the production of unnatural amino acids, which are common chiral building blocks in the pharmaceutical industry. Numerous transaminases were identified during the sequence analysis of the whole genome library. Provided *R. erythropolis* SET-1 displayed tolerance to and utilization of both substrates required by transaminase enzymes, the enzyme could be identified, cloned and expressed with directed evolution producing an enhanced enzyme, as proposed above for the nitrile metabolising enzymes.

Further research on the Nit1 enzyme characterised in the body of work presented would involve the study of enzyme kinetics alongside solvent and metal tolerance testing. A major undertaking would be to further analyse the enzyme native structure, perhaps via crystallography. In addition to the above it would also be interesting to subject the *nit*1 gene to rounds of directed evolution to improve enzyme characteristics such as rate of conversion, optimum temperature or pH to facilitate the use of the Nit1 protein for increased conversion of the industrially relevant mandelonitrile or perhaps one of the other nitriles such as 3-HGN, 3-HBN or 3-HPPN under alternative conditions. This would be achieved through error prone PCR, followed by the generation of mutant libraries which could easily been analysed via a high throughput screening method such as the Nessler's colorimetric assay (Coady *et al.* 2013) or pre-existing analytical methods (Coady *et al.* 2015) to identify improved enantioselectivity profiles.

BIBLIOGRAPHY

Albarrán-Velo, J., González-Martínez, D. and Gotor-Fernández, V. (2018) 'Stereoselective biocatalysis: A mature technology for the asymmetric synthesis of pharmaceutical building blocks', *Biocatalysis and Biotransformation*. Informa UK Limited, trading as Taylor & Francis Group, **36**(2), pp. 102–130. doi: 10.1080/10242422.2017.1340457.

Altschul, S. F., Gish W, Miller W, Myers, E W. Lipman (1990) 'Basic local alignment search tool.', *Journal of Molecular Biology*, **215**(3), pp. 403–10. doi: 10.1016/S0022-2836(05)80360-2.

Anderson, K. R., Leksa, N. C. and Schwartz, T. U. (2014) 'Optimized *E. coli* epression strain LOBSTR eliminates common contaminants from His-tag purification', *Proteins*, **81**(11), pp. 1857–1861. doi: 10.1002/prot.24364.Optimized.

Andrade, J., Karmali Amin, Carrondo Maria A., Frazão Carlos (2007) 'Structure of amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state', *Journal of Biological Chemistry*, **282**(27), pp. 19598–19605. doi: 10.1074/jbc.M701039200.

Angelini L. M, L Ngelini, Leticia Mara Limada Silva, Amanda Ribeiro Martins Rocco, Lucas de Freitas Coli, Milagre, Cintia Duarte de Freitas. (2015) 'A high-throughput screening assay for distinguishing nitrile hydratases from nitrilases', *Brazilian Journal of Microbiology*, **46**(1), pp. 113–116. doi: 10.1590/S1517-838246120130851.

Asano, Y. and Kato, Y. (1998) 'Z-phenylacetaldoxime degradation by a novel aldoxime dehydratase from *Bacillus* sp. strain OxB-1', *FEMS Microbiology Letters*. Blackwell Publishing Ltd, **158**(2), pp. 185–190. doi: 10.1111/j.1574-6968.1998.tb12818.x.

Aziz R. K, Bartels, Daniela Best, Aaron A DeJongh, Matthew Disz, Terrence Edwards, Zagnitko Olga *et al.* (2008) 'The RAST server: rapid annotations using subsystems technology', *BMC Genomics*. BioMed Central, **9**(1), p. 75. doi: 10.1186/1471-2164-9-75.

Bai, Y.-L. Yang, Zhi-Long Qiao, Ming-Qiang Zhang, Xiu-Ming Zhou, Jing Gao, Cai-Chang *et al.* (2003) 'The action of S1 nuclease and a cloning strategy for microcircular DNAs.', *Sheng wu gong cheng xue bao = Chinese journal of biotechnology*, **19**(2), pp. 240–243.

Banerjee, A., Sharma, R. and Banerjee, U. (2002) 'The nitrile-degrading enzymes: current status and future prospects', *Applied Microbiology and Biotechnology*. Springer-Verlag, **60**(1–2), pp. 33–44. doi: 10.1007/s00253-002-1062-0.

Barton, B. M., Harding, G. P. and Zuccarelli, A. J. (1995) 'A general method for detecting and sizing large plasmids', *Analytical Biochemistry*, **226**(2), pp. 235–240. doi: 10.1006/abio.1995.1220.

Basta T., Keck A., Klein J., Stolz A. (2004) 'Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains.', *Journal of bacteriology*, **186**(12), pp. 3862–72. doi: 10.1128/JB.186.12.3862-3872.2004.

Bauer, R., Knackmuss, H. J. and Stolz, A. (1998) 'Enantioselective hydration of a arylpropionitriles by a nitrile hydratase from *Agrobacterium tumefaciens* strain d3', *Applied Microbiology and Biotechnology*, **49**(1), pp. 89–95. doi: 10.1007/s002530051142.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I. Lipman, D.J. *et al.* (2013) 'GenBank', *Nucleic Acids Research*, **41**(D1), pp. 36–42. doi: 10.1093/nar/gks1195.

Bergeron, S., Chaplin, D., Edwards, J., Ellis, B., Hill C., Holt-Tiffin K. *et al.* (2006) 'Nitrilase-catalysed desymmetrisation of 3-hydroxyglutaronitrile: preparation of a statin side-chain intermediate', *Organic Process Research & Development*. American Chemical Society, **10**(3), pp. 661–665. doi: 10.1021/op050257n.

Betke, T. Higuchi, J., Rommelmann, P., Oike, K., Nomura, T., Kato, Y. *et al.* (2018) 'Biocatalytic synthesis of nitriles through dehydration of aldoximes: the substrate scope

of aldoxime dehydratases', *ChemBioChem*. **33**(7), pp. 1443-1454. doi: 10.1002/cbic.201700571.

Blakey, A., Colby, J., Williams, E., O'Reilly, C. (1995) 'Regio- and stereo-specific nitrile hydrolysis by the nitrile hydratase from *Rhodococcus* AJ270', *FEMS Microbiology Letters*, **129**(1), pp. 57–61. doi: http://dx.doi.org/10.1016/0378-1097(95)00135-R.

Brady, D., Beeton, A. Zeevaart, J. Kgaje, C. Van Rantwijk, F. *et al.* (2004) 'Characterisation of nitrilase and nitrile hydratase biocatalytic systems', *Applied Microbiology and Biotechnology*. Springer-Verlag, **64**(1), pp. 76–85. doi: 10.1007/s00253-003-1495-0.

Brandão, P. F. B., Clapp, J. P. and Bull, A. T. (2003) 'Diversity of nitrile hydratase and amidase enzyme genes in *Rhodococcus erythropolis* recovered from geographically distinct habitats', *Applied and Environmental Microbiology*, **69**(10), pp. 5754–5766. doi: 10.1128/AEM.69.10.5754-5766.2003.

Brázda, V. Kolomazník, J., Lýsek, J., Hároníková, L., Coufal, J., Št'astný, J., *et al.* (2016) 'Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences', *Biochemical and Biophysical Research Communications*, **478**(4), pp. 1739–1745. doi: 10.1016/j.bbrc.2016.09.015.

Brenner, C. (2002) 'Catalysis in the nitrilase superfamily', *Current Opinion in Structural Biology*, **12**(6), pp. 775–782. doi: http://dx.doi.org/10.1016/S0959-440X(02)00387-1.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J. *et al.* (2015) 'RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes', *Scientific Reports*. Nature Publishing Group, **5**. doi: 10.1038/SREP08365.

Callier, V. (2019) 'Core Concept: Gene transfers from bacteria and viruses may be shaping complex organisms', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, **116**(28), pp. 13714–13716. doi: 10.1073/PNAS.1909030116.

Cameron, R. A., Sayed, M. and Cowan, D. A. (2005) 'Molecular analysis of the nitrile catabolism operon of the thermophile Bacillus pallidus RAPc8', *Biochimical et Biophysical Acta - General Subjects*, **1725**(1), pp. 35–46. doi: 10.1016/j.bbagen.2005.03.019.

Ciskanik, L. M., Wilczek, J. M. and Fallon, R. D. (1995) 'Purification and characterization of an enantioselective amidase from *Pseudomonas chlororaphis* B23', *Applied and Environmental Microbiology*, **61**(3), pp. 998–1003.

Coady, T. (2014) 'Biotransformations Using Nitrile Hydrolysing Enzymes For', (May)-Thesis.

Coady, T. M. *et al.* (2013) 'A high throughput screening strategy for the assessment of nitrile-hydrolyzing activity towards the production of enantiopure β-hydroxy acids', *Journal of Molecular Catalysis B: Enzymatic*, **97**(0), pp. 150–155. doi: http://dx.doi.org/10.1016/j.molcatb.2013.08.001.

Coady, T. M. *et al.* (2015) 'Substrate evaluation of rhodococcus erythropolis SET1, a nitrile hydrolysing bacterium, demonstrating dual activity strongly dependent on nitrile sub-structure', *European Journal of Organic Chemistry*, **5**, pp. 1108–1116. doi: 10.1002/ejoc.201403201.

Coffey, L., Clarke, A., Duggan, P., Tambling, K., Horgan, S., Dowling, D., O'Reilly, C. *et al.* (2009) 'Isolation of identical nitrilase genes from multiple bacterial strains and real-time PCR detection of the genes from soils provides evidence', *Archives of Microbiology*, **191**(10), pp. 761–71. doi: 10.1007/s00203-009-0507-6.

Coffey, L. Owens, E., Tambling, K., O'Neill, D., O'Connor, L., O'Reilly, C., *et al.* (2010) 'Real-time PCR detection of Fe-type nitrile hydratase genes from environmental isolates suggests horizontal gene transfer between multiple genera.', *Antonie van Leeuwenhoek*, **98**(4), pp. 455–463. doi: 10.1007/s10482-010-9459-8.

Coffey, L. V. (2007) *Molecular analysis of genes involved in nitrile metabolism in Microbacterium Sp. AJ115, Rhodococcus Erythropolis AJ270, AJ300 and ITCBP*. Wateford Institute of Technology.

Corporation, Lucigen. (2018) 'Fosmid Cloning Kits', (12), pp. 1–17.

Saores Braganca, C.R.., Dooley-Cullinane, T.M., O'Reilly, C., Coffey, Lee et al. (2017) 'Applying functional metagenomics to search for novel nitrile-hydrolyzing enzymes using environmental samples', **1**(2), pp. 17–20. doi: 10.15761/BTT.1000108.

Davison, J. (1999) 'Genetic exchange between bacteria in the environment', *Plasmid*. 1999/09/18, **42**(2), pp. 73–91. doi: 10.1006/plas.1999.1421.

Dennett, G. V., Blamey, J. M. *et al.* (2016) 'A new thermophilic nitrilase from an antarctic hyperthermophilic microorganism', *Frontiers in Bioengineering and Biotechnology;* **4**, pp. 1–9. doi: 10.3389/fbioe.2016.00005.

Dooley-Cullinane, T. M., O'Reilly, C. and Coffey, L. (2016) 'Real-time PCR detection of aldoxime dehydratase genes in nitrile-degrading microorganisms', *Antonie van Leeuwenhoek*. Springer International Publishing. doi: 10.1007/s10482-016-0786-2.

Endo, I., Nojiri, M., Tsujimura, M., Nakasako, M., Nagashima, S., Yohda, M., *et al.* (2001) 'Fe-type nitrile hydratase', *Journal of Inorganic Biochemistry*, **83**(4), pp. 247–253. doi: http://dx.doi.org/10.1016/S0162-0134(00)00171-9.

European Commission (2011) 'Form for notification of final regulatory action to ban or severly restrict a chemical - 2,6-Dichlorobenzonitrile'.

European Union (2007) 'En Route to the Knowledge-Based Bio-Economomy', pp. 1–23.

Fleming, F. F. Yao, L., Ravikumar, P. C., Funk, L., Shook, B. C. (2010) 'Nitrile-containing pharmaceuticals: efficacious roles of the nitrile pharmacophore.', *Journal of Medicinal Chemistry*, **53**(22), pp. 7902–17. doi: 10.1021/jm100762r.

Goldberg, I. D., Gwinn, D. D. and Thorne, C. B. (1966) 'Interspecies transformation between Bacillus subtilis and *Bacillus licheniformis.*', *Biochemical and Biophysical Research Communications*, **23**(4), pp. 543–8.

Goldlust, A. and Z Bohak (1989) 'Induction, purification, and characterization of the nitrilase of *Fusarium oxysporum* f. sp. melonis', *Biotechnology and Applied Biochemistry*. pp. 89

Gong, J. S., Lu, Z. M., Li, H., Zhou, Z. M., Shi, J. S. *et al.* (2013) 'Metagenomic technology and genome mining: Emerging areas for exploring novel nitrilases', *Applied Microbiology and Biotechnology*, **97**(15), pp. 6603–6611. doi: 10.1007/s00253-013-4932-8.

Gradley, M. L., Deverson, L., Knowles, C., Kent C. T., *et al.* (1994) 'Asymmetric hydrolysis of R-(-),S(+)-2-methylbutyronitrile by *Rhodococcus rhodochrous* NCIMB 11216', *Archives of microbiology*, pp. 246–251.

Green, M. R. and Sambrook, J. (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Harper, D. B. (1977) 'Fungal degradation of aromatic nitriles. Enzymology of C-N cleavage by *Fusarium solani*', *Biochemical Journal*. 1977/12/01, **167**(3), pp. 685–692.

Harrison, E. and Brockhurst, M. A. (2012) 'Plasmid-mediated horizontal gene transfer is a coevolutionary process', *Trends in Microbiology*. doi: 10.1016/j.tim.2012.04.003

Hirrlinger, B. and Stolz, A. (1996) 'Purification and properties of an amidase from *Rhodococcus erythropolis* MP50 which enantioselectively hydrolyzes These include : Purification and Properties of an Amidase from *Rhodococcus erythropolis* MP50 Which Enantioselectively Hydrolyzes 2-Arylpropion', *Microbiology*, **178**(12), pp. 3501–3507. doi: 10.1128/JB.178.12.3501-3507.1996.

Hoyle, A. J., Bunch, A. W. and Knowles, C. J. (1998) 'The nitrilases of *Rhodococcus rhodochrous* NCIMB 11216', *Enzyme and Microbial Technology*, **23**(7–8), pp. 475–482. doi: 10.1016/S0141-0229(98)00076-3.

Iwaki, H., Grosse, S., Bergeron, H., Leisch, H., Morley, K., Hasegawa, Y., *et al.* (2013) 'Camphor pathway redux: functional recombinant expression of 2,5- and 3,6-diketocamphane monooxygenases of *Pseudomonas putida* ATCC 17453 with their cognate flavin reductase catalyzing Baeyer-Villiger reactions.', *Applied and environmental microbiology*, **79**(10), pp. 3282–93. doi: 10.1128/AEM.03958-12.#

Juhas, M., van der Meer, J R., Gaillard, M., Harding, R M., Hood, D W., Crook, D W., *et al.* (2009) 'Genomic islands: tools of bacterial horizontal gene transfer and evolution', *FEMS Microbiol Rev*. 2009/01/31, **33**(2), pp. 376–393. doi: 10.1111/j.1574-6976.2008.00136.x.

Kaplan, O., Vejvoda, V., Plíhal, O., Pompach, P., Kavan, D., *et al.* (2006) 'Purification and characterization of a nitrilase from *Aspergillus niger* K10', *Applied Microbiology and Biotechnology*, 73(3), pp. 567–575. doi: 10.1007/s00253-006-0503-6.

Kato, Y. Nakamura, K., Sakiyama, H., Mayhew, S G., Asano, Y (2000) 'Novel heme-containing lyase, phenylacetaldoxime dehydratase from *Bacillus* sp. strain OxB-1: purification, characterization, and molecular cloning of the gene', *Biochemistry*. 2000/01/29, **39**(4), pp. 800–809.

Kato, Y. and Asano, Y. (2006) 'Molecular and enzymatic analysis of the "aldoxime-nitrile pathway" in the glutaronitrile degrader *Pseudomonas* sp. K-9. Appl Mi', *Applied Microbiology and Biotechnology*. Berlin: Springer Science & Business Media, **70**(1), pp. 92–101. doi: http://dx.doi.org/10.1007/s00253-005-0044-4.

Kato, Y., Ooi, R. and Asano, Y. (2000) 'Distribution of aldoxime dehydratase in microorganisms', *Appl Environ Microbiol*. 2000/06/01, **66**(6), pp. 2290–2296. doi: 10.1128/AEM.66.6.2290-2296.2000.Updated.

Kato, Y., Yoshida, S. and Asano, Y. (2005) 'Polymerase chain reaction for identification

of aldoxime dehydratase in aldoxime- or nitrile-degrading microorganisms', *FEMS Microbiology Letters*, **246**(2), pp. 243–249. doi: http://dx.doi.org/10.1016/j.femsle.2005.04.011.

Kaul, P., Banerjee, A., Mayilraj, S., Banerjee, U. C. (2004) 'Screening for enantioselective nitrilases: kinetic resolution of racemic mandelonitrile to (R)-(−)-mandelic acid by new bacterial isolates', *Tetrahedron: Asymmetry*, **15**(2), pp. 207–211. doi: 10.1016/j.tetasy.2003.10.041.

Kimani, S. W. and Vinod, B. (2007) 'Structure of an aliphatic amidase from Geobacillus pallidus RAPc8. *Research Papers* , pp. 1048–1058. doi: 10.1107/S090744490703836X.

Kiziak, C., Conradt, D., Stolz, A., Mattes, R., Klein, J., (2005) 'Nitrilase from *Pseudomonas fluorescens* EBC191: cloning and heterologous expression of the gene and biochemical characterization of the recombinant enzyme.', *Microbiology (Reading, England)*, **151**(Pt 11), pp. 3639–48. doi: 10.1099/mic.0.28246-0.

Kobayashi, M., Yanaka, N., Nagasawa, T., Yamada, H. (1990) 'Purification and characterization of a novel nitrilase of *Rhodococcus rhodochrous* K22 that acts on aliphatic nitriles', *Journal of Bacteriology*, **172**(9), pp. 4807–4815. doi: 10.1128/jb.172.9.4807-4815.1990.

Kobayashi, M., Komeda, H., Nagasawa, T., Nishyamma, M., (1993) 'Amidase coupled with low-molecular-mass nitrile hydratase from *Rhodococcus rhodochrous* J1: Sequencing and expression of the gene and purification and characterization of the gene product', *European Journal of Biochemistry*, **217**(1), pp. 327–336. doi: 10.1111/j.1432-1033.1993.tb18250.x.

Kobayashi, M. (1996) 'A novel gene cluster including the *Rhodococcus rhodochrous* J1 nhlBA genes encoding a low molecular mass Nitrile Hydratase (L-NHase) Induced by Its Reaction Product', *Journal of Biological Chemistry*, **271**(26), pp. 15796–15802. doi: 10.1074/jbc.271.26.15796.

Kobayashi, M., Goda, M. and Shimizu, S. (1998b) 'Nitrilase catalyzes amide hydrolysis

as well as nitrile hydrolysis', *Biochemical and Biophysical Research Communications*, **253**(3), pp. 662–666. doi: 10.1006/bbrc.1998.9834.

Kobayashi, M., Nagasawa, T. and Yamada, H. (1989) 'Nitrilase of Rhodococcus rhodochrous J1. Purification and characterization.', *European journal of biochemistry / FEBS*, **182**(2), pp. 349–56.

Komeda, H., Komeda, H., Hori, Y., Kobayashi, M., Shimizu, S*., et al.* (1996) 'Transcriptional regulation of the *Rhodococcus rhodochrous* J1 nitA gene encoding a nitrilase', *Proceedings of the National Academy of Sciences of the United States of America.* 1996/10/01, **93**(20), pp. 10572–10577.

Komeda, H., Kobayashi, M. and Shimizu, S. (1996) 'Characterization of the gene cluster of high-molecular-mass nitrile hydratase (H-NHase) induced by its reaction product in *Rhodococcus rhodochrous* J1', *Proceedings of the National Academy of Sciences of the United States of America.* 1996/04/30, **93**(9), pp. 4267–4272.

Kuhn, M. L., Martinez, S., Gumataotao, N., Bornscheuer, U., Liu, D., *et al.* (2012) 'The Fe-type nitrile hydratase from *Comamonas testosteroni* Ni1 does not require an activator accessory protein for expression in *Escherichia coli*', *Biochemical and Biophysical Research Communications.* Elsevier Inc., **424**(3), pp. 365–370. doi: 10.1016/j.bbrc.2012.06.036.

Kulakov, L. A., Chen, S., Allen, C.R., Larkin, M.J., *et al.* (2005) 'Web-type evolution of *Rhodococcus* gene clusters associated with utilization of naphthalene', **71**(4), pp. 1754–1764. doi: 10.1128/AEM.71.4.1754.

Lan, Y., Zhang, X., Liu, Z., Zhou, Li Shen, R., *al.* (2017) 'Overexpression and characterization of two types of nitrile hydratases from *Rhodococcus rhodochrous* J1', *PLoS ONE*, **12**(6), pp. 1–14. doi: 10.1371/journal.pone.0179833.

Letek, M., González, P., Arthur, I., Rodríguez, H., Freeman, T., *et al.* (2010) 'The genome

of a pathogenic *Rhodococcus*: Cooptive virulence underpinned by key gene acquisitions', *PLoS Genetics*, **6**(9). doi: 10.1371/journal.pgen.1001145.

Lorenz, M. G. and Wackernagel, W. (1994) 'Bacterial gene transfer by natural genetic transformation in the environment', *Microbiological Reviews* 1994/09/01, **58**(3), pp. 563–602.

Marchesi, J. R., Sato, T., Weightman, A. J., Martin, A., *et al.* (1998) 'Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA', **64**(2), pp. 795–799.

Marchler-Bauer, A. and Bryant, S. H. (2004) 'CD-Search: protein domain annotations on the fly', *Nucleic Acids Research*, **32**, pp. W327–W331. doi: 10.1093/nar/gkh454.

Martinez, S., Yang, X., Bennett, B., Holz, R. C. (2017) 'A cobalt-containing eukaryotic nitrile hydratase', *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., **1865**(1), pp. 107–112. doi: 10.1016/j.bbapap.2016.09.013.

Martínková, L. *et al.* (2009) 'Fungal nitrilases as biocatalysts: recent developments', *Biotechnology Advances*, **27**(6), pp. 661–670. doi: http://dx.doi.org/10.1016/j.biotechadv.2009.04.027.

Martínková, L., Vejvoda, V., Kaplan, O., Kubáč, D., Malandra, A., *et al.* (2017) 'Recent advances and challenges in the heterologous production of microbial nitrilases for biocatalytic applications', *World Journal of Microbiology and Biotechnology*, **33**(1), p.p. 8. doi: 10.1007/s11274-016-2173-6.

Martínková, L. and Kren, V. (2010) 'Biotransformations with nitrilases.', *Current Opinion in Chemical Biology*, **14**(2), pp. 130–7. doi: 10.1016/j.cbpa.2009.11.018.

Mathew, C. D., Nagasawa, T., Kobayashi, M., Yamada, H (1988) 'Nitrilase-catalyzed production of Nicotinic Acid from 3-Cyanopyridine in *Rhodococcus rhodochrous* J1',

Applied Environmental Microbiology 1988/04/01, **54**(4), pp. 1030–1032.

Mayaux, J. F., Cerbelaud, E., Soubrier, F., Yeh, P., Blanche, F. *et al.* (1991) 'Purification, cloning, and primary structure of a new enantiomer-selective amidase from a *Rhodococcus* strain: structural evidence for a conserved genetic coupling with nitrile hydratase', *Journal of Bacteriology*, **173**(21), pp. 6694–6704. doi: 10.1128/jb.173.21.6694-6704.1991.

Meinhardt, F., Schaffrath, R. and Larsen, M. (1997) 'Microbial linear plasmids', *Applied Microbiology and Biotechnology*, **47**(4), pp. 329–336. doi: 10.1007/s002530050936.

Metzner, R., Okazaki, S., Asano, Y., Gröger, H., *et al.* (2014) 'Cyanide-free enantioselecive synthesis of nitriles: synthetic proof of a biocatalytic concept and mechanistic insights', *ChemCatChem.* **6**(11), pp. 3105–3109. doi: 10.1002/cctc.201402612.

Mizunashi, W., Nishiyama, M., Horinouchi, S., Beppu, T., (1998) 'Overexpression of high-molecular-mass nitrile hydratase from *Rhodococcus rhodochrous* J1 in recombinant *Rhodococcus* cells', *Applied Microbiology and Biotechnol*ogy 1998/07/03, **49**(5), pp. 568–572.

Mueller, P., Egorova, K., Vorgias, C., Boutou, E., Trauthwein, H., Verseck, S., Antranikian, G., *et al.* (2006) 'Cloning, overexpression, and characterization of a thermoactive nitrilase from the hyperthermophilic archaeon *Pyrococcus* abyssi', *Protein Expression and Purification*, **47**(2), pp. 672–681. doi: http://dx.doi.org/10.1016/j.pep.2006.01.006.

Muñoz Solano, D., Hoyos, P., Hernáiz, M. J., Alcántara, a R., *et al.* (2012) 'Industrial biotransformations in the synthesis of building blocks leading to enantiopure drugs.', *Bioresource Technology*, **115**, pp. 196–207. doi: 10.1016/j.biortech.2011.11.131.

Nagasawa, T., Wieser, M., Nakamura, T., Iwahara, H., Yoshida, T., Gekko, K., (2000) 'Nitrilase of *Rhodococcus rhodochrous* J1: Conversion into the active form by subunit association', *European Journal of Biochemistry*, **267**(1), pp. 138–144. doi: 10.1046/j.1432-1327.2000.00983.x.

Nagasawa, T., Nakamura, T. and Yamada, H. (1990) 'Production of acrylic acid and methacrylic acid using *Rhodococcus rhodochrous* J1 nitrilase', *Applied Microbiology and Biotechnology*. Springer-Verlag, **34**(3), pp. 322–324. doi: 10.1007/BF00170051

Nelp, M. T., Astashkin, A., Breci, L. A., Mccarty, R. M., Bandarian, V., *et al.* (2014) 'The alpha subunit of nitrile hydratase is sufficient for catalytic activity and post-translational modification. *Biochemistry,* pp. 3990-3994.

Nigam, V. K., Arfi, T., Kumar, V., Shukla, P., (2017) 'Bioengineering of Nitrilases towards its use as green catalyst: applications and perspectives', *Indian Journal of Microbiology*. Springer India, **57**(2), pp. 131–138. doi: 10.1007/s12088-017-0645-5.

Nomura, J., Hashimoto, H., Ohta, T., Hashimoto, Y., Wada, K., (2013) 'Crystal structure of aldoxime dehydratase and its catalytic mechanism involved in carbon-nitrogen triple-bond synthesis'. *PNAS.* **110** (8), pp. 2810-2815. doi: 10.1073/pnas.12003 38110/-/DCSupplemental. www.pnas.org/cgi/doi/10.1073/p nas.1200338110.

Norman, A., Hansen, L. H. and Sørensen, S. J. (2009) 'Conjugative plasmids: vessels of the communal gene pool', *Philosophical Transcations of the RoyalSocitey B*, (364), pp. 2275–2289. doi: 10.1098/rstb.2009.0037.#

Novo, C., Tata, R., Clemente, A., Brown, P. R. (1995) '*Pseudomonas aeruginosa* aliphatic amidase is related to the nitrilase/cyanide hydratase enzyme family and Cys166is predicted to be the active site nucleophile of the catalytic mechanism', *FEBS Letters*, **367**(3), pp. 275–279. doi: 10.1016/0014-5793(95)00585-W.

O'Mahony, R. *et al.* (2005) 'Characterisation of the nitrile hydratase gene clusters of *Rhodococcus erythropolis* strains AJ270 and AJ300 and *Microbacterium* sp. AJ115

indicates horizontal gene transfer and reveals an insertion of IS1166', *Antonie Van Leeuwenhoek*. 2005/04/02, 87(3), pp. 221–232. doi: 10.1007/s10482-004-3721-x.

O'Reilly, C. and Turner, P. D. (2003) 'The nitrilase family of CN hydrolysing enzymes - a comparative study', *Journal of Applied Microbiology*, **95**(6), pp. 1161–1174. doi: 10.1046/j.1365-2672.2003.02123.x.

Oinuma, K., Hashimoto, Y., Konishi, K., Goda, M., Noguchi, T., *et al.* (2003) 'Novel aldoxime dehydratase involved in carbon-nitrogen triple bond synthesis of *Pseudomonas chlororaphis* B23. Sequencing, gene expression, purification, and characterization', *Journal of Biological Chemistry*. 2003/05/30, **278**(32), pp. 29600–29608. doi: 10.1074/jbc.M211832200.

Osswald, S., Wajant, H. and Effenberger, F. (2002) 'Characterization and synthetic applications of recombinant AtNIT1 from Arabidopsis thaliana', *European Journal of Biochemistry*, **269**(2), pp. 680–687. doi: 10.1046/j.0014-2956.2001.02702.x.

Overbeek, R., Olson, R., Pusch, G., Olsen, Gary., Davis, J., *et al.* (2014) 'The seed and the rapid annotation of microbial genomes using subsystems technology (RAST)', *Nucleic Acids Research*. Oxford University Press, 42, pp. 206. doi: 10.1093/NAR/GKT1226.

Overhage, J., Sielker, S., Homburg, S., Parschat, K., Fetzner, S., (2005) 'Identification of large linear plasmids in *Arthrobacter* sp. encoding the degradation of quinaldine to anthranilate.', *Microbiology (Reading, England)*, **151**(Pt 2), pp. 491–500. doi: 10.1099/mic.0.27521-0.

Pace, H. C., Hodawadekar, S., Draganescu, A., Huang, J., Bieganowski, P., *et al.* (2000) 'Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers', *Current Biology*. 2000/08/26, **10**(15), pp. 907–917.

Pace, H. C. and Brenner, C. (2001) 'The nitrilase superfamily: classification, structure and function.', *Genome biology*, **2**(1), pp. 1–9. doi: 10.1186/gb-2001-2-1-reviews0001.

Patel, J. M. (2009) 'Biocatalytic synthesis of atorvastatin intermediates', *Journal of Molecular Catalysis B: Enzymatic*, **61**(3–4), pp. 123–128. doi: http://dx.doi.org/10.1016/j.molcatb.2009.07.004.

Pei, X., Wang, J., Guo, W., Miao, J., Wang, A., (2017) 'Efficient biodegradation of dihalogenated benzonitrile herbicides by recombinant *Escherichia coli* harboring nitrile hydratase-amidase pathway', *Biochemical Engineering Journal*. Elsevier, **125**, pp. 88–96. doi: 10.1016/J.BEJ.2017.05.021.

Piaetsi (2013) A method for the biocatalytic production of nitriles from oximes and oxime dehydratases and usable therein. USA patent application US 2013/0244298 A1. pp. 1-50.

Piotrowski, M., Schönfelder, S. and Weiler, E. W. (2001) 'The *Arabidopsis thaliana* isogene NIT4 and its orthologs in tobacco encode Cyano-L-alanine hydratase/nitrilase', *Journal of Biological Chemistry*, **276**(4), pp. 2616–2621. doi: 10.1074/jbc.M007890200.

Podar, M., Eads, J. R. and Richardson, T. H. (2005) 'Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study.', *BMC evolutionary biology*, **5,** pp. 42. doi: 10.1186/1471-2148-5-42.

Raczynska, J. E., Vorgias, C. E., Antranikian, G., Rypniewski, W., *et al.* (2011) 'Crystallographic analysis of a thermoactive nitrilase', *Journal of Structural Biology*. Elsevier Inc., **173**(2), pp. 294–302. doi: 10.1016/j.jsb.2010.11.017.

Rädisch, R., Chamátal, M., Rucká, L., Novotný, P., Petrásková, L., *et al.* (2018) 'Overproduction and characterization of the first enzyme of a new aldoxime dehydratase family in *Bradyrhizobium* sp.', *Biological Macromolecules*. Elsevier B.V., **115,** pp. 746–753. doi: 10.1016/j.ijbiomac.2018.04.103.

Rasor, J. P. and Voss, E. (2001) 'Enzyme-catalyzed processes in pharmaceutical industry', *Applied Catalysis A: General*, **221**(1–2), pp. 145–158. doi: http://dx.doi.org/10.1016/S0926-860X(01)00804-3.

Robertson, D. E., Chaplin, J., Desantis, G., Podar, M., Madden, M., *et al.* (2004) 'Exploring nitrilase sequence space for enantioselective catalysis †', *Journal of Biological Chemistry*, **70**(4), pp. 2429–2436. doi: 10.1128/AEM.70.4.2429.

Robinson, W. G. and Hook, R. H. (1964) 'Ricinine Nitrilase: I. reaction product and substrate specificity', *Journal of Biological Chemistry*, 239(12), pp. 4257–4262.

Rose, K. and Fetzner, S. (2006) 'Identification of linear plasmid pAM1 in the flavonoid degrading strain *Actinoplanes missouriensis*(T) (DSM 43046).', *Plasmid*, **55**(3), pp. 249–54. doi: 10.1016/j.plasmid.2005.10.003.

Rustler, S., Müller, A., Windeisen, V. Chmura, A. *et al.* (2007) 'Conversion of mandelonitrile and phenylglycinenitrile by recombinant *E. coli* cells synthesizing a nitrilase from *Pseudomonas fluorescens* EBC191', *Enzyme and Microbial Technology*, **40**(4), pp. 598–606. doi: 10.1016/j.enzmictec.2006.05.013.

Sakashita, T. Hashimoto, Y., Oinuma, K. I., Kobayashi, M. *et al.* (2008) 'Transcriptional regulation of the nitrile hydratase gene cluster in *Pseudomonas chlororaphis* B23', *Journal of Bacteriology*, **190**(12), pp. 4210–4217. doi: 10.1128/JB.00061-08.

Salyers, A., Shoemaker, N., Stevens, A., (1995) 'Conjugative transposons: an unusual and diverse set of integrated gene transfer elements', *Microbiological Reviews.* 1995/12/01, **59**(4), pp. 579–590.

Sharma, M., Sharma, N. and Bhalla, T. (2009) 'Amidases: versatile enzymes in nature', *Reviews in Environmental Science and Bio/Technology*. Springer Netherlands, **8**(4), pp. 343–366. doi: 10.1007/s11157-009-9175-x.

Siguier, P., Filée, J. and Chandler, M. (2006) 'Insertion sequences in prokaryotic genomes', *Current Opinion in Microbiology*, **9**(5), pp. 526–531. doi: 10.1016/j.mib.2006.08.005.

Singh, R., Sharma, R., Tewari, N., Geetanjali, R., Diwan S., (2006) 'Nitrilase and Its Application as a "Green" Catalyst', *Chemistry & Biodiversity*. **3**(12), pp. 1279–1287. doi: 10.1002/cbdv.200690131.

Song, L., Yuan, H., Coffey, L., Doran, J., Wang, M., O'Reilly, C., *et al.* (2008) 'Efficient expression in *E. coli* of an enantioselective nitrile hydratase from *Rhodococcus erythropolis*', *Biotechnology Letters*, **30**(4), pp. 755–762. doi: 10.1007/s10529-007-9611-3.

Soucy, S. M., Huang, J. and Gogarten, J. P. (2015) 'Horizontal gene transfer: building the web of life', *Nature Reviews Genetics*, **16**(8), pp. 472–482. doi: 10.1038/nrg3962.

Stalker, D. M., Malyj, L. D. and McBride, K. E. (1988) 'Purification and properties of a nitrilase specific for the herbicide bromoxynil and corresponding nucleotide sequence analysis of the *bxn* gene', *Journal of Biological Chemistry*, **263**(13), pp. 6310–6314.

Stevenson, D. E., Feng, R., Dumas, F., Groleau, D., Mihoc, A., *et al.* (1992) 'Mechanistic and structural studies on *Rhodococcus* ATCC 39484 nitrilase.', *Biotechnology and Applied Biochemistry*, **15**(3), pp. 283–302.

Surivet, J.-P. and Vatèle, J.-M. (1998) 'First total synthesis of (−)-8-epi-9-deoxygoniopypyrone', *Tetrahedron Letters*. Pergamon, **39**(52), pp. 9681–9682. doi: 10.1016/S0040-4039(98)02269-2.

Terreni, M., Pagani, G., Ubiali, D., Fernández-Lafuente, R., Mateo, C., Guisán, J., (2001) 'Modulation of Penicillin acylase properties via immobilization techniques: One-pot chemoenzymatic synthesis of Cephamandole from Cephalosporin C', *Bioorganic and Medicinal Chemistry Letters*, **11**(18), pp. 2429–2432. doi: 10.1016/S0960-894X(01)00463-2.

Thimann, K. V. and Mahadevan, S. (1964) 'Nitrilase', *Archives of Biochemistry and Biophysics*, **105**(1), pp. 133–141. doi: 10.1016/0003-9861(64)90244-9.

Thimann, K. V and Mahadevan, S. (1958) 'Enzymatic hydrolysis of indoleacetonitrile.', *Nature*, **181**(4621), pp. 1466–7.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, **22**(22), pp. 4673–4680.

Thuku, R. N., Brady, D., Benedik, M J., Sewell, B T., *et al.* (2009) 'Microbial nitrilases: versatile, spiral forming, industrial enzymes.', *Journal of applied microbiology*, **106**(3), pp. 703–27. doi: 10.1111/j.1365-2672.2008.03941.x.

Top, E. M., Mizrahi, I., Oliveira, C., Johnson, T., Shintani, M., *et al.* (2015) 'Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy', *Frontiers in Microbiology*. 6, pp. 244.

Uson, D. A. H. H. and Cornavacca, C. E. S. (2012) 'Dendroscope 3 : An interactive tool for rooted phylogenetic trees and networks', **61**(6), pp. 1061–1067. doi: 10.1093/sysbio/sys062.

Vejvoda, V., Kubáč, D., Davidová, A., Kaplan, O., Šulc, M., *et al.* (2008) 'Purification and characterization of a nitrilase from *Fusarium solani* O1', *Journal of Molecular Catalysis B: Enzymatic*, **50**(2–4), pp. 99–106. doi: 10.1016/j.molcatb.2007.09.006.

Vorwerk, S., Biernacki, S., Hillebrand, H., Janzik, I., Müller, A., *et al.* (2001) 'Enzymatic characterization of the recombinant *Arabidopsis thaliana* nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster', *Planta*, **212**(4), pp. 508–516. doi: 10.1007/s004250000420.

Wang, H., Sun, H. and Wei, D. (2013) 'Discovery and characterization of a highly efficient enantioselective mandelonitrile hydrolase from *Burkholderia cenocepacia* J2315 by phylogeny-based enzymatic substrate specificity prediction'.
Biotechnology. BMC Biotechnology, **13**(1), p. 1. doi: 10.1186/1472-6750-13-14.

Wang, J. (2015) 'Process development for the production of (R) - (−)-Mandelic Acid by recombinant *Escherichia coli* Cells harboring Nitrilase from *Burkholderia cenocepacia* J2315'. *Organic Process, Research and Development*. doi: 10.1021/acs.oprd.5b00269.

Xia, Y., Cui, W., Liu, Z., Zhou, L., Cui, Y., *et al.* (2016) 'Construction of a subunit-fusion nitrile hydratase and discovery of an innovative metal ion transfer pattern', *Scientific Reports*. Nature Publishing Group, **6**, pp. 1–13. doi: 10.1038/srep19183.

Xie, S. X., Kato, Y. and Asano, Y. (2001) 'High yield synthesis of nitriles by a new enzyme, phenylacetaldoxime dehydratase, from *Bacillus* sp. strain OxB-1.', *Bioscience, Biotechnology, and Biochemistry*, **65**(12), pp. 2666–72. doi: 10.1271/bbb.65.2666.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction.', *BMC Bioinformatics*, **13**(1), p. 134. doi: 10.1186/1471-2105-13-134.

Zhang, C., Zhang, Z. and Li, C. (2012) 'Efficient production of (R) - o -chloromandelic acid by deracemization of o- chloromandelonitrile with a new nitrilase mined from Labrenzia aggregata', pp. 91–99. doi: 10.1007/s00253-012-3993-4.

Zhang, Z. J. *et al.* (2011) 'Cloning and biochemical properties of a highly thermostable and enantioselective nitrilase from *Alcaligenes* sp. ECU0401 and its potential for (R)-(2)-mandelic acid production', *Bioprocess and Biosystems Engineering*, **34**(3), pp. 315–322. doi: 10.1007/s00449-010-0473-z.

Zheng, Y. G., Chen, J., Liu, Z. Q., Wu, M. H., Xing, L. Y., *et al.* (2008) 'Isolation, identification and characterization of *Bacillus subtilis* ZJB-063, a versatile nitrile-converting bacterium', *Applied Microbiology and Biotechnology*, **77**(5), pp. 985–993. doi: 10.1007/s00253-007-1236-x.
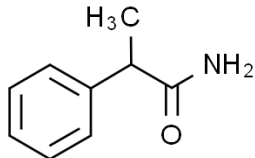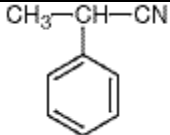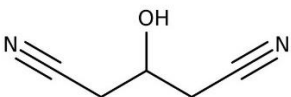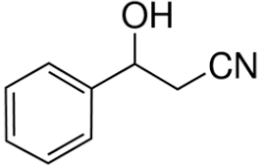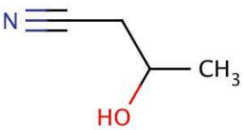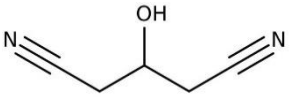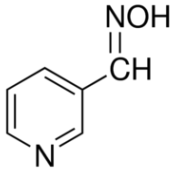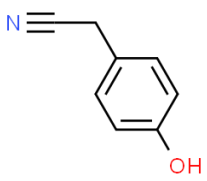
Zhou, Z. Hashimoto, Y., Cui, T., Washizawa, Y *et al.* (2010) 'Unique biogenesis of high-molecular mass multimeric metalloenzyme nitrile hydratase: intermediates and a
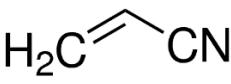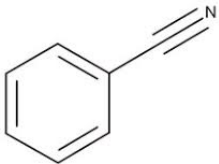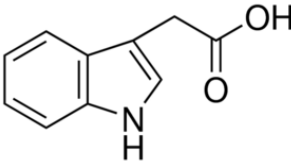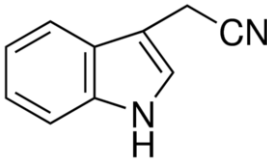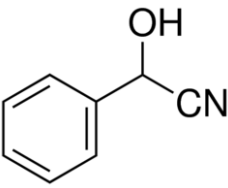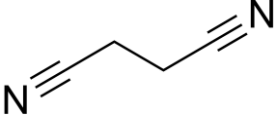
proposed mechanism for self-subunit swapping maturation', *Biochemistry*. 2010/10/05, **49**(44), pp. 9638–9648.

APPENDIX I

SUPPLEMENTAL INFORMATION

Table A1: A list of chemicals used as substrates for cell and enzyme assays with corresponding CAS number and Structure.

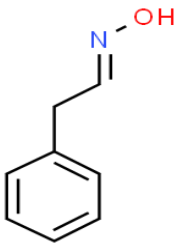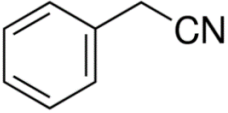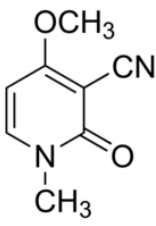| Compound	Structure	CAS Number	Reference
2-Phenylpropionamide		1125-70-8	Ciskanik *et al.*, 1995
2-Phenylpropionitrile		1823-91-2	N/A
3-Hydroxuglutaronitrile		13880-89-2	Coady *et al.* 2013
3-Hydroxy-3-phenylpropionitrile		17190-29-3	Coady *et al.* 2013
3-Hydroxybutyronitrile		4368-06-3	Coady *et al.* 2013
3-Hydroxyglutaronitrile		13880-89-2	Robertson *et al.*, 2004)
3-Pyridinealdoxime		1193-92-6	(Kato *et al.*, 1999).
4-Hydroxyphenylacetonitrile		14191-95-8	N/A

227

Acrylonitrile		107-12-1	(Gong *et al.* 2012).
Compound	**Structure**	**CAS Number**	**Reference**
Benzonitrile			
Indole-3-acetic acid		87-51-4	(Thimann and Mahadevan, 1964)
Indole-3-acetonitrile		771-51-7	(Thimann and Mahadevan, 1964)
Mandelonitrile		532-28-5	Robertson *et al.*, 2004)
Phenylacetaldoxime		7028-48-0	(Kato *et al.*, 2000).
Phenylacetonitrile		140-29-4	(Coffey, 2007)
Ricinine		524-40-3	(Robinson and Hook, 1964)
Succinonitrile		110-61-2	(Gong *et al.* 2012).

228

M9 MINIMAL MEDIA AND RECIPE INSTRUCTIONS

Preparation of 5X M-9 basis:

Add the following to 1000 mL of deionised water:

- 0.24 M (85.99 g) Sodium Phosphate dibasic dodecahydrate
- 0.11 M (14.97 g) Potassium dihydrogen ortho-phosphate
- 43.0 mM (2.54 g) Sodium Chloride

Autoclave at 121 °C for 15 minutes.

Preparation of trace elements

Add the following to 100 mL of deionised water;

- 0.1mmol (28.7 mg) Zinc sulphate heptahydrate*
- 1mmol (198 mg) Manganese chloride tetrahydrate*
- 1.0 mmol (129 mg) Cobalt chloride hydrate*
- 0.1 mmol (17 mg) Copper chloride dihydrate
- 1.0 mmol (147 mg) Calcium chloride dihydrate

Autoclave at 121 °C for 15 minutes

.
Preparation of 20 % (w/v) glucose

- Add 20 g of glucose to 100 mL of deionised water. Autoclave at 121 °C for 15 minutes.

Preparation of 0.1M Calcium chloride dihydrate

- Add 100 mM (1.47 g) to 100 mL of deionised water. Autoclave at 121 °C for 15 minutes.

Preparation of Magnesium sulfate heptahydrate

#NB must be prepared freshly, on the day, cannot be autoclaved and must be filtered using a sterile syringe and 0.22 µm filter disc.

- Add 1 M (24.65 g) to 100 mL of deionised water

Preparation of Ferrous sulfate heptahydrate

#NB must be prepared freshly, on the day, cannot be autoclaved and must be filtered using a sterile syringe and 0.22 µm filter disc.

- Add 139 mg to 100 mL of deionised water to make a 5 mM solution.

Preparation of M-9 minimal media (up to 100 mL with sterile deionised water)

- 20 mL of 5X M-9 basis;
- 100 µL of Trace elements
- 100 µL of Iron sulphate heptahydrate
- 2 mL of glucose
- 100 µL of Calcium chloride dehydrate
- 100 µL of Magnesium sulphate heptahydrate

CHAPTER 2 ALDOXIME DEHYDRATASE PARTIAL SEQUENCES

Bacillus sp. NN1

CCTCGGGCGACTACGGGTCGTCGCCGGAGACCCCGCCGCAGGTGGACGCGT
AGTATGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGG
GCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCC
ACTCTTCAATCGGGCATGGACTTCCTCCGCGACAACGGCCCGGCCGTCGGGT
GCTAC

Bacillus sp. ST144

CGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAGGCCTCGGGCGAA
CTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGTAGTAGTGCGG
GGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGGCCGACGCG
GAAGCGGACGAGCGCAGCCTCTACCTGGA

Brevundimonas sp. ST111

TCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGTA
GTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGG
GCCGACGCGGAAGCAGACGAGCGCAGCCTCTACCTCGACGAAATCCTGCCG
ACTCTTCAAT

Burkholderia sp. F73

GGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACG
CGTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGA
CTGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCT
GCCCACTCT

Buttiauxella sp. SW2-28

GGATGCAGGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAG
GTGGGCGGGTAGTAGTGCGGGGACACGACAACATCGCACTGATCAGATCCG
GGCAGGACTGGGCCGACGCGGAAGCAGACGAGCGCAGCCTCTACCTCGAC
GAAATCCTGCCGACTCTCAATC

Pseudomonas sp. ST30

CTGGGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAG
GCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGC
GTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGAC
TGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTG
CCCACTCTCAATC

Rahnella sp. SS1-7

TACTGGGGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGC
AGGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGAC
GCGTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGG
ACTGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCC
TGCCCACTCTCAATCG

Rhodococcus sp. F30

GGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAGGC
CTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGT
AGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTG
GGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCC
CACTCTCA

 Rhodococcus sp. SS1-2

ACTGGGGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCA
GGCCTCGGGCGAACTACGGGTCGTCGCCGTGACCCCGCCGCAGGTGGACGC
GTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGAC
TGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCC

Rhodococcus sp. ST135

CGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGTAG
TAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGG
CCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCA
CTCT

Enterobacter sp. ST146

AGAGTGGGCAGGATTTCGTCCAGGTAGAGGCTGCGCTCGTCCGCTTCCGCG
TCGGCCCAGTCCTGCCCGGATCTGATCAGTGCGATGTTGTCGTGCCCCCGCA
CTACTACGCGTCCACCTGCGGCGGGGTCACCGGCGACGACCCGTAGTTCGC
CCGAGGCC

Pseudomonas sp. ST196

CGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGGGTAG
TAGTGCGGGGACACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGG
CCGACGCGGAAGCAGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCA
CTCTC

Rhodococcus sp. F37

GGGGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAGG
CCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGTCGCAGGTGGACGCG
TAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACT
GGGCCGACGCGGAAGCGGACGAGCGCA

Serretia sp. SS1-18

GGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACG
CGTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGA
CTGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCT
GCCCACTCTTCAATC

Bacillus sp. F10

ACCCGGCCAAGGGCGGACGCGTCGTGGTGCTGGGGCATGACAACCTCACAC
TGATTCGTTCCGGCCAGGACTGGGCGGATGCCGAGGCAGAGGAACGCTCGC
TCTACCTCGACGAAATTTTACCGACGTTGCAGGACGGCATGGACTTCCTGCG
GGACAACGGCCAGCCGTCGGGTGCT

Rhodococcus sp. F17

GCAGGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGG
ACGCGTAGTAGTGCGGGGACACGACAACATCGCACTGATCAGATCCGGGCA
GGACTGGGCCGACGCGGAAGCAGACGAGCGCAGCCTCTACCTCGACGAAAT
CCTGCCCACTCTTCAATC

Rhodococcus sp. F52

TACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGTAGTAGTGCGGG

GGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGGCCGACGCGG

AAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCACTCTTCAAT

C

Rhodococcus erythropolis SET-1

TGGGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAG
GCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGC
GTAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGAC
TGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTG
CCCAC

Rhodococcus sp. LC11

CGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAGGCCTCGG
GCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGTAGTAG
TGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGGCCG
ACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCACT
Stenotrophomononas sp. ST180

CCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCG
TAGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACT
GGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGC
CCACTCTTCAATC

Rhodococcus sp. NN5

TACGGGTCGTCGCCGGAGACCCCGCCGCAGGTGGACGCGTAGTAGTGCGGG
GGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGGCCGACGCGG
AAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCACTCTTCAAT
CGGGCAT

Rhodococcus sp. SS1-29

GGTGACCCCGCCGCAGGTGGACGCGTAGTAGTGCGGGGGGCACGACAACATC
GCACTGATCAGATCCGGGCAGGACTGGGCCGACGCGGAAGCGGACGAGCG
CAGCCTCTACCTGGACGAAATCCTGCCCACTCTTCAATCGGGCATGGACTTC
CTCCGCGACAACGGCCCGGCCGTCGGGTGCTAC

Rahnella sp. SS1-25

TCGGTTCGATGCGCGACCGTTTCCCCATCTCCCAAACAGACTGGATGAAGCC
CACGAGCGAGCTGGAAGTGGTCGCGGGCGACCCGGCCAAGGGCGGACGCG
TCGTGGTGCTGGGGCATGACAACCTCACACTGATTCGTTCCGGCCAGGACTG
GGCGGATGCCGAGGCAGAGGAACGCTCGCTCTACCTCGACGAAA

Microbacterium sp. SS1-15

CGAGCGAGCTGCAAGTGGTCGCGGGCGACCCGGCCAAGGGCGGACGCGTC
GTGGTGCTGGGGCATGACAACCTCACACTGATTCGTTCCGGCCAGGACTGG
GCGGATGCCGAGGCAGAGGAACGCTCGCTCTACCTCGACGAAA

Streptomyces sp. F61

GATGCGCGAGCACCCGACGACCGCGCAGACCGACCGGAGGAAGGCCACGC
CCGAACGAAGAGTCGGCGCCGAGGACCCCGTCGCAGGTAGACGCTGCACTA
GTCCGCGGCCACGACAACACAGCACTGACCAGATCCGAGCAGGACGAGGC
CGACGCGCAACCGCACGACCACAGCCTCCACCTGCGACGAAATCACCGCCC
ACGACCAATGGACTTCCTCCGCGACAACGGCCCGGCCG

Rhodococcus sp. ST158

GGCCTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACG
CGTAGTAGTGCGGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGA
CTGGGCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCT
GCCCACTCTTCAATCGGGCATGGACTTCCTCCGCGACAACGGCCCGGCCGTC
GGGTGCTACAGCAACCGTTTCGTACGCAATATCGA

Rhizobium sp. ST119

TGCGGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGGGCCG
ACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCCACTC
TTCAATCGGGCATGGACTTCCTCCGCGACAACGGCCCGGCCGTCGGG

Rhodococcus sp. ST199

TCGGGCGACTACGGGTCGTCGCCGGTGACCCCGCTCGCAGGTGGACGCGTA
GTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTGG
GCCGACGCGGAAGCGGACGAGCGCAGCCTCTACCTGGACGAAATCCTGCCC
ACTCTTCAATCGGGCATGGACTTCCTCCGCGACAACGGCCCGGCCGTCGGG

Stenotrophomonas sp. ST53

GGGTTCGATGCGCGAGCGCTTTCCGATCTCTCAGACCGACTGGATGCAGGC
CTCGGGCGAACTACGGGTCGTCGCCGGTGACCCCGCCGCAGGTGGACGCGT
AGTAGTGCGGGGGCACGACAACATCGCACTGATCAGATCCGGGCAGGACTG
GGCCGACGCGGAAGCAGACGAGCGCAGCCTCTACCTCGACGAAATCCTGCC
GACTCT

Rhizobium sp. ST125

GCCGGCTTCGACTTGCCGGACGGACCGGCACACCACGATCTCACCCATCAC
ATCGACAACCAGGGCTACGAGAACCTGATCGTGGTCGGTTACTGGAAAGAT
GTTTCTTCCCAACATCGTTGGAGCACATCACCTCCAGTGTCCTCCTGGTGGG
AGTCCGAAGACCGCCTGTCGGACGGATTGGGGTTCTTCCGCGAGATCGTGG
CACCGAGAGCCGAACAATTCGAAACGCTCTACGCGTTCCAAGACGACCTCC
CCGGGGTGGGAGCTGTCATGGACGGTGTCAGCGGCGAAATCAAC

CHAPTER 3 POTENTIAL NITRILASE AMPLICONS

Brevundimonas sp. ST111

GTGGTCAGCACACACTTCGGCGCGCAGTCACGCAGCACGAAGTCCACCCGG
TCGACCGGAGAAGTCAGGTCCACCGGTATGTATCCGCCACCCGCCTTCACC
ACCGCCAGCAATGCCACCACCAGGTCGATCGAGCGGGTCATCGCGACTGCC
ACCAGGGTTTCCGGCACCACACCCTGAGCCACCAACAGTCGAGCGAGGCGG
TTTGCTCGCTCGTCGAGTTCGCGGTAGGTGACGGACAATCCGTCGCAGCGCA
CCGCCACGGCATCACCGGACGTCAGCACCGCACGATCGAAGAGATCCGACA
GCGAACCCTCGTGGGTTCCGACGCCCGAGACCGAGACCGAGGCCGAGCCCG
AGGCCGAGGCCGAT

Rhizobium sp. ST119

ATATGTTCTGGCCGGGTCTTACCCGCGATTTCTGAAATCCTCGAACCTTCCG
GCCCCGAACCTTCCGGCCCCGAACCTCGCAGCACAGCTTGTAGCTGATCGAT
CGATGCCGCCAGATCAGCGACCACGGGAGTGTCGAACAAGGCCTTCACACC
GACATCGTGCTCGATTCGGATCTGATGCGCGCTGCGACGCGAGTGGCGCTG
AGCGAGTCTCCACCCAATTCGAAGAAATTGTCGTCGGCGCCCACCTGGTCG
AG

Pseudomonas sp. ST30

AGCGCACCGAGACGGTGCCGAACCACCAGAACTTTCAGAGATCCACCGCTG
TACGTCAGCACCGCGACGTCCACTGCCACAGACGGACGCGGATAGTCCGTC
AATCCCTTGCCGCTCGTGTCTGTCCATTTATCCACCCTTGCATGTTATCTCAA
ACATGATCTATGGTTAGCGCAAAGTGATCAAACGATCATCGCGAACCGAAG
AGGTGGCTATGAAACTCACCGCACAGCAGTCCGATCGAGCGGCCGGCGTTC
TGCTCGGCACCGCAGCAGGAGACGCACTCGGCGCCGGCTACGAATTCACCT
AC

Rhodococcus sp. ST135

GGAGACGGACCGCTCAACCGAGCAGCGTTTGTCATCCACAGCAAGGTTCAC
GAGGCTCGTCCCGACGTCATCGGAGCTGTTCACGCTCACGGTTTGAACGGTA
AGGCGTTCTCCTCGCTGCACACCCCGCCTCACGCCGATCACTCAGGACGCGT
GCGCCTTTCACGAGGATCACTCGATCTACGAGGAGTATCACGGGGTGGCGC
TGGACGAAGAGGAGGGGATTCGCATCGCGGATGCCCTCGGATCGAACAAG
GCCGTGATCCTCACCAACCACGGGTCATCTGACCGTCGGGAACTTCCCGTGG
AGAGCGCAGTGTGGTG

Rhodococcus sp. ST199

CAGGCGTCCTGGGTGATCGGCTGCAGTAGCCGGTGCAGGGACGAGAACGCT
TTTCCGTGCAGCCCGTGTGCGTGCACGGCTCCGATCACGTCCGGGCGGGCTT
CGTGCACCGTGGAGTGGATGACGAAGGCTGCTCGATTCAACGGTCCCGATC
CGGACAGAACGGTGCCGTCGGCACCGACGTGCAACAGCTCCTCGGGCCGCA
CCTGGGAGAAGTGGCGTCCGAACGGCGCGGTCCAGAAGGTATCGGG

Stenotrophomonas sp. ST180

TGCGTTGGTCGAGAAGTCCATCCTCATTCGTGAGGAGATCGACGGCACCGTT
CGCTATCACCTGCTCGACCTCCTGCGGGAGTACGGACGAGAGAAGGCCGAG
ATCTCGAACGACTACGTGGAACTTCAACGGCGCCACCGTGATTGGTTCGAA
GCGCTGGTTGTTCGTGCGGAAGCCGAGTGGATCAGTGACCGGCAAGCGGCG
TGGATCGCACGGCTGGGTCGTGAACAGCACAACGTCTACGACGCGCTGGA

CHAPTER 4 TMDC NITRILASE PARTIAL SEQUENCE

CTGCTGGGAGCACCATATGCCGCTCGCGCGCTACTCGATGTACGCCAAGGG
CGTGGACGTCTACGTTGCGCCGACGTGGGACAACAGCGACATGTGGGTGGC
GACGCTCCGCCACATCGCCAAGGAGGGGCGGCTGTACGTGATCGGCGTGGC
GCCGCTGCTGCGCGGGTCGGACGTCCCCGACGACGTGCCGGGGAAGGCCGA
GCTGTGGGGCGGCGATGACGACTGGATGTCGCGCGGCTTCTCCACCATCGT
CGCGCCGGGCGGCGAGGTACTGGCCGGTCCGCTGACGGAGGAGGAAGGCA
TCCTCTACGCGGAGATCGACCCGGCGAGAGCCCGTTCGTCACGGCACCAGT
TCGACCCCGCCGGGCATTACTC

ADDITIONAL CHROMATOGRAMS

pH Study

Below are sample chromatograms observed from the gradient pH screen for *NIT*1 enzyme on 10 mM mandelonitrile.
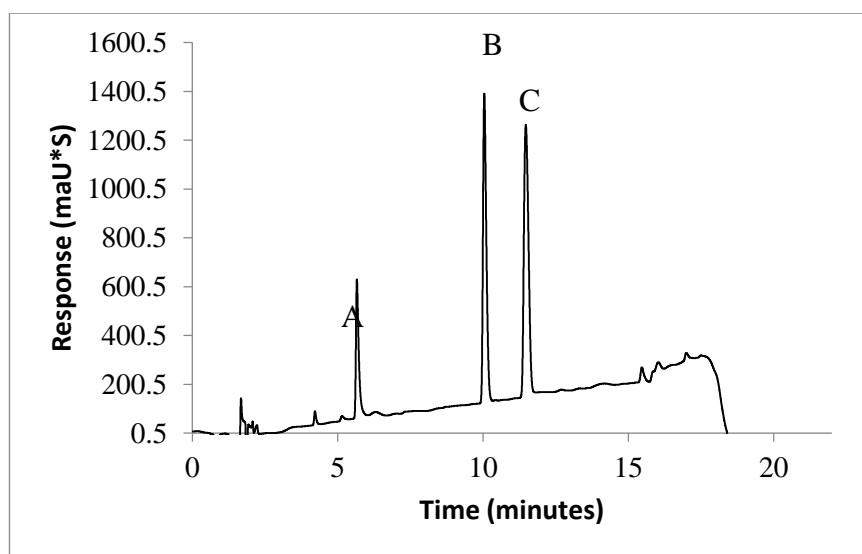

Figure A1 Separation of total nitrile and total acid products from biotransformation of *NIT*1 enzyme on 10 mM mandelonitrile in 100 mM sodium acetate buffer pH 4.0. (A) mandelic acid (B) mandelonitrile, (C) benzaldehyde. Analysis was performed using gradient separation on a C18 Hyperclone ODS (150 x 4.6 mm; 3 µm) with mobile phase (A) H_2O with 0.1 % formic acid and (B) ACN with 0.1% formic acid with a gradient of 10% to 50% over a run time of 20 minutes. The flow rate was set to 1.0 mL/min with a 20 µL injection volume and column temperature set to 10 °C with UV detection at 210 nm.
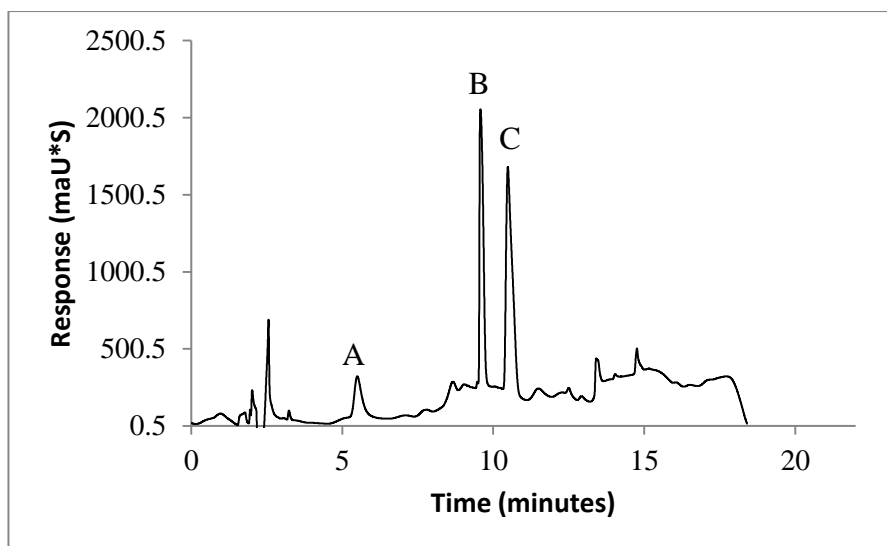
Figure A2 Separation of total nitrile and total acid products from biotransformation of *NIT*1 enzyme on 10 mM mandelonitrile in 100 mM sodium acetate buffer pH 5.0. (A) mandelic acid (B) mandelonitrile, (C) benzaldehyde. Chromatographic conditions as per Figure A1.
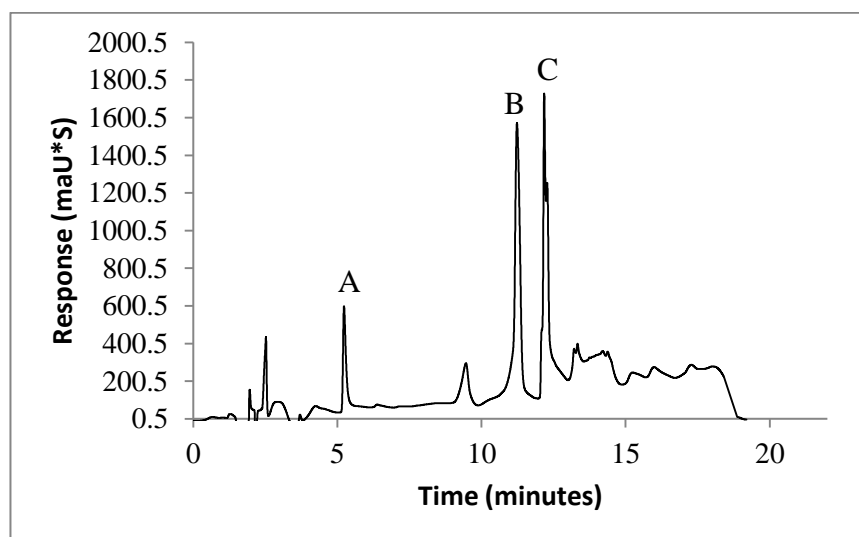

Figure A3 Separation of total nitrile and total acid products from biotransformation of *NIT*1 enzyme on 10 mM mandelonitrile in 100 mM sodium acetate buffer pH 6.0. (A) mandelic acid (B) mandelonitrile, (C) benzaldehyde. Chromatographic conditions as per Figure A1.
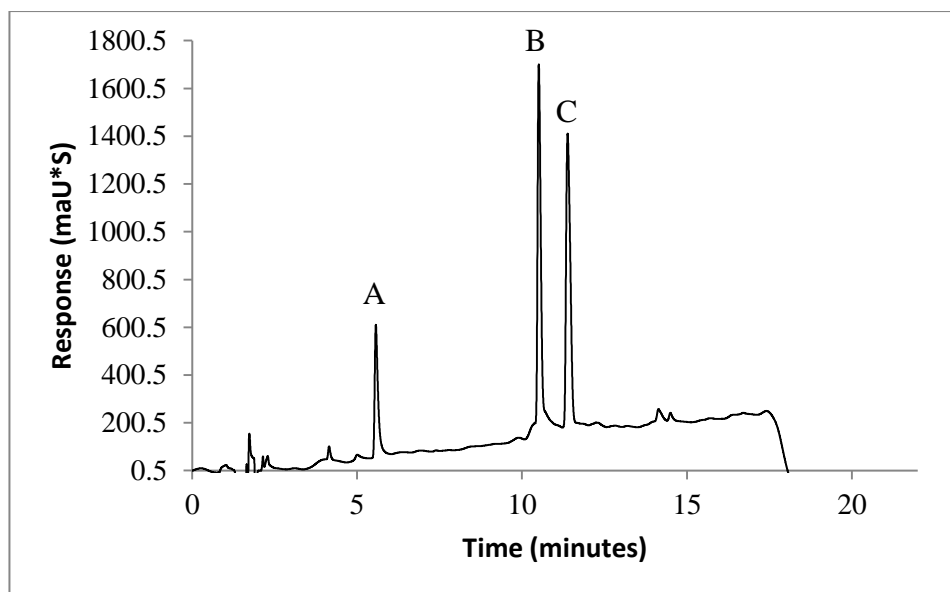
Figure A4 Separation of total nitrile and total acid products from biotransformation of *NIT*1 enzyme on 10 mM mandelonitrile in 100 mM potassium phosphate buffer pH 7.0. (A) mandelic acid (B) mandelonitrile, (C) benzaldehyde. Chromatographic conditions as per Figure A1.

Figure A5 Separation of total nitrile and total acid products from biotransformation of *NIT*1 enzyme on 10 mM mandelonitrile in 100 mM potassium phosphate buffer pH 8.0. (A) mandelic acid (B) mandelonitrile, (C) benzaldehyde. Chromatographic conditions as per Figure A1.
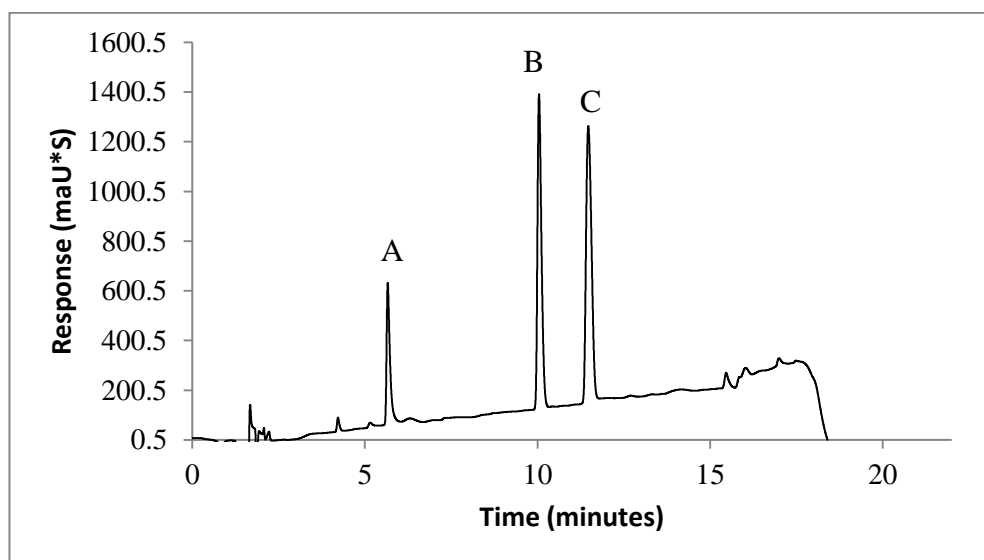
PUBLICATIONS

Title: The use of clade-specific PCR assays to identify novel nitrilase genes from environmental isolates- Online Article- in press

Reference: Dooley-Cullinane T-M, O'Reilly C, Aslam B, Coffey L, *et al.* The use of clade-specific PCR assays to identify novel nitrilase genes from environmental isolates. Microbiology Open. 2018;700. https://doi.org/10.1002/mbo3.700

Title: Real-time PCR detection of aldoxime dehydratase genes in nitrile-degrading microorganisms

Reference: Dooley-Cullinane T-M., OReilly C, Coffey L, Real-time PCR detection of aldoxime dehydratase genes in nitrile-degrading microorganisms, *Antonie Van Leeuwenhoek* 110 (2017) 217–279.