Chapter 2

Literature review

Literature review

2.1   Introduction

2.1.1   The fossil fuel industry in South Africa

Energy to drive the South African economy is derived from various fossil fuel related sources, all of which can have a significant environmental impact. These fossil and other non-renewable fuel sources comprise oil, natural gas, coal, hydropower, nuclear power and biomass (Fig. 2). However, South Africa is unique in that it manufactures synthetic liquid fuel from coal and gas, known as synfuel (Fig. 3). Approximately 40% of South African liquid fuel requirements are met by synfuels, courtesy of Sasol (*ca*. 35%) and PetroSA (*ca*. 5%) (Surridge[1], personal communication). The synfuel industry was initially constructed to address supply security issues and this technology is now being exported, e.g. a new Sasol plant in Qatar (Surridge[1], personal communication).
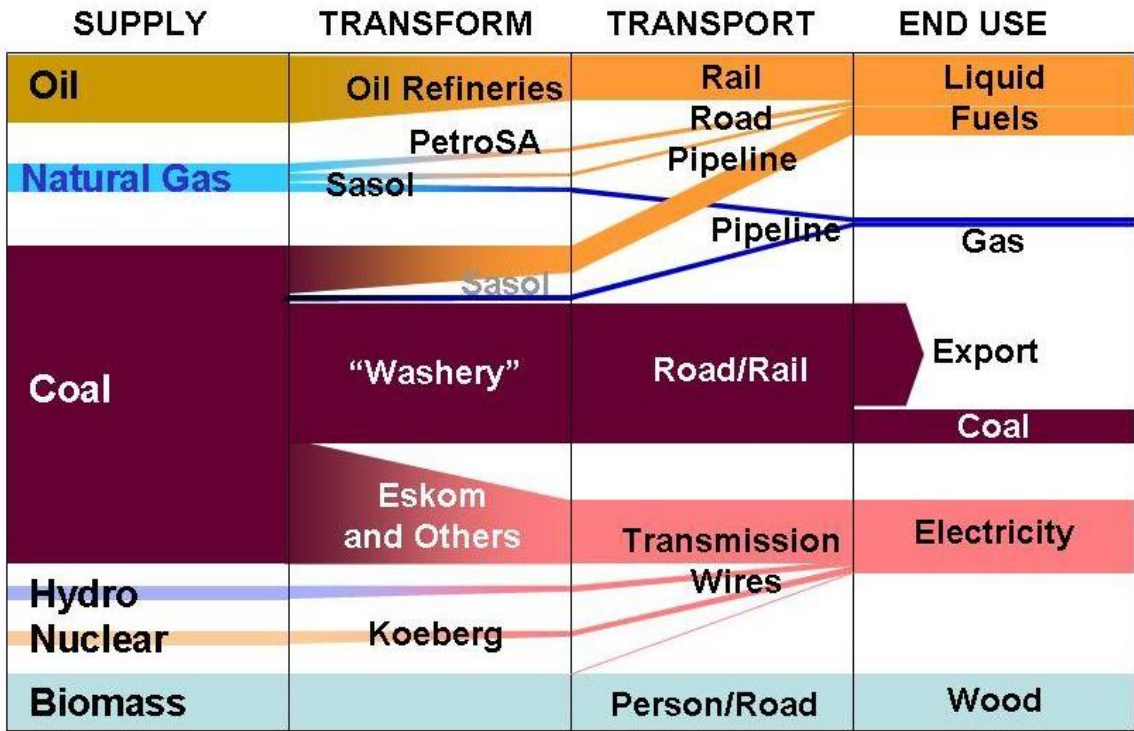
Figure 2: Energy flow from primary energy supply to final use – roughly to scale (Department of Minerals and Energy 2003).
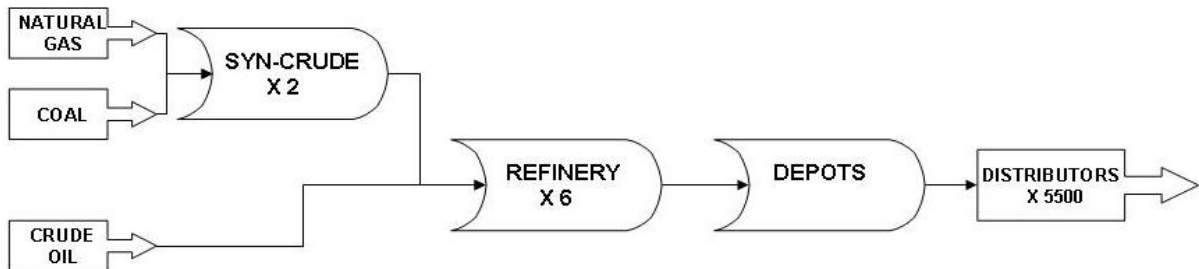

Figure 3: The petroleum product supply chain for South Africa (Surridge[1], personal communication).

The potential threat of crude oil leaks or spills from storage tanks is massive when considering that the approximate coastal crude oil storage capacity is at least 10.4 million barrels at the main storage unit in Saldanha Bay, plus operational stocks at the six refineries countrywide. Coastal and inland refined product storage must be maintained at 1.15 billion litres, a 21-day supply, since 20 billion litres of all fuel types are used annually in South

21

Africa (Surridge[1], personal communication). Possible leaks from high-pressure (maximum 100 bar petroleum products and 59 bar gas) underground petroleum and gas pipelines transporting fuel inland should also be considered, as 3000 km of pipelines ranging in diameter from 150-508 mm pose a potential threat to surrounding soil (Fig. 4). Total product throughput within the pipelines is 16 billion litres per annum liquid fuel and 450 million cubic metres of gas (Petronet SA 2005). Currently pipelines extend across five provinces of South Africa and construction is underway of a new multi-products pipeline between Durban and Gauteng (Petronet SA 2005). Approximately 5500 garages nationwide store refined fuel in underground storage facilities, hence posing a further risk of soil pollution should these tanks leak (Surridge[1], personal communication).
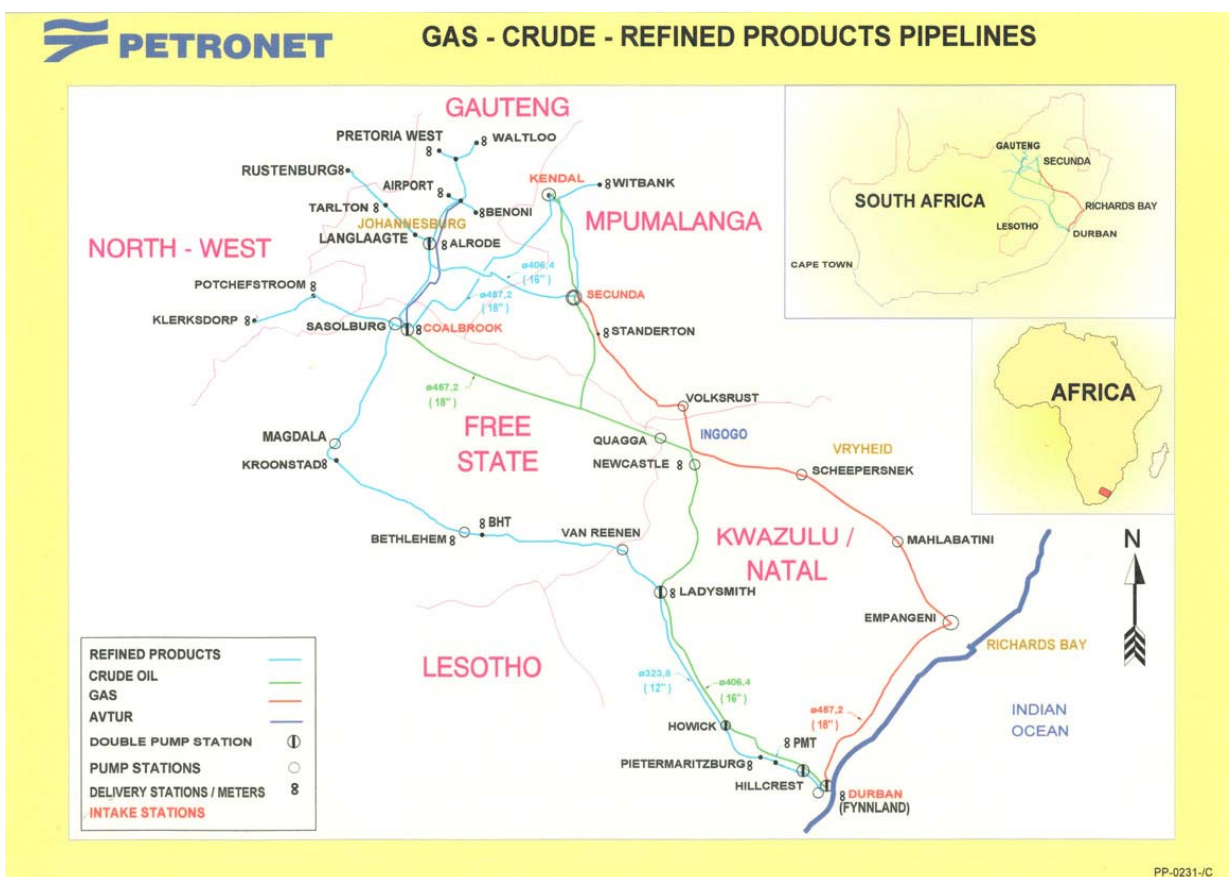

Figure 4: Areas in South Africa where refined oil products, crude oil, gas and avtur (non-jet engine aviation fuel) are delivered by pipeline, stored, transported and distributed (Petronet SA 2005).

Finally, the end-user, in the form of vehicles, is also a major potential source of pollution due to movement, during accidents and also as a result of oil and petrol leaks from engines. During 2005, South Africa was the best performing automobile market internationally, with domestic sales and production rising to all time highs. New vehicle sales amounted to 565 018 units, a 25.7% increase from 2004. During 2004 sales improved by 22.0%, reaching 449 603 vehicles compared with 368 470 units sold during 2003 (Fig. 5) (NAAMSA 2005). As a result of so many new vehicles coming onto the roads annually in South Africa, as well as the vehicles still on the roads at the end of each year, the potential for random point pollution caused by commercial and passenger vehicles can currently be assumed to increase by approximately 26% annually.
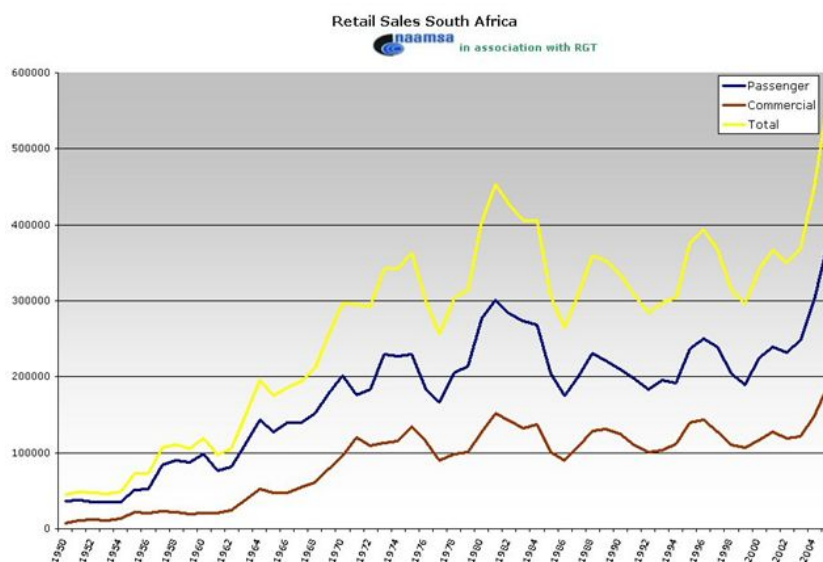

Figure 5: Passenger, commercial and total vehicle sales in South Africa from 1950 to 2005 (NAAMSA 2005).

Naphthalene and toluene, found in petroleum and diesel products, are two of the most common PAHs that are subject to biodegradation. Naphthalene is a crystalline, aromatic, white, solid hydrocarbon, it is volatile and forms a flammable vapour. The name is derived from the Latin *naphtha*, meaning liquid bitumen, and is of Semitic origin. It consists of two fused benzene rings, is classified as a benzenoid PAH, and is manufactured from coal-tar.

When converted to the phthalic anhydride, it is used in the manufacture of plastics, dyes and solvents, and as antiseptic and insecticide (Wikipedia 2005b):
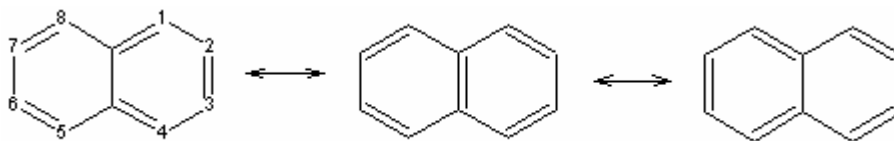
Toluene, also referred to as methylbenzene or phenylmethane, is a clear, water-insoluble liquid. The name is derived from *toluol*, referring to tolu balsam, an aromatic extract from the tree *Myroxylon balsamum* (L.) Harms from which it was first obtained (Wikipedia 2005c). It is an aromatic hydrocarbon with a methyl side-chain, widely used as an industrial feedstock, octane booster in fuel, solvent in paints, rubber, printing, adhesives, lacquers, in leather tanning, disinfectants, and in the production of phenol, polyurethane foams and TNT (Wikipedia 2005c):
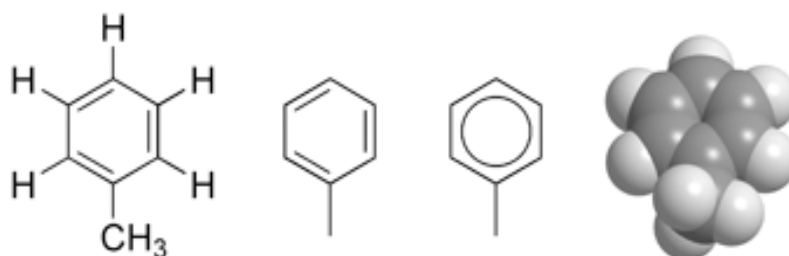
2.1.2 Soil health

Soil health can be defined as "the capacity of soil to function as a vital living system to sustain biological productivity, promote environmental quality and maintain plant and animal health" (Doran and Zeiss 2000). Productivity of conventional agricultural systems largely depends on the functional process of soil microbial communities (Girvan *et al.* 2003). These communities' structure and diversity are influenced by the soil structure and spatial distribution as well as the relationship between abiotic and biotic factors of microbial communities (Torsvik and Øvereås 2002). With the advent of various types of industries over the past 200 years, the ecology of earth's ecosystems has been severely disrupted. The commercialisation, extraction, refining, transportation, distribution and storage of petroleum

products have led to oil, petrol and diesel pollution of soils. In petrol-polluted water that may seep into soil, benzene, toluene, ethylbenzene and xylene (BTEX) isomers are present in the water-soluble fraction (Prenafeta-Boldú *et al.* 2002). This disruption has decreased biodiversity and selected for cosmopolitan microbial species better adapted to survive in the changed environment (Kozdrój and Van Elsas 2001). Not least impacted by these changes are the microbiota inhabiting the soil.

2.1.3 Pollution

Hydrocarbons are currently the main source of the world's energy resources due to the energy they produce when combusted. This also makes them the world's main source of pollution in the case of spills and waste products. There are essentially three types of hydrocarbons, viz. (i) aromatic hydrocarbons that have at least one aromatic ring, (ii) saturated hydrocarbons, including n-alkanes, branched alkanes and cycloalkanes that do not have double-, triple- or aromatic-bonds, and (iii) unsaturated hydrocarbons with one or more double- or triple-bonds between carbon atoms, referred to as alkenes and alkynes, respectively (Atlas 1981; Wikipedia 2006a). The most notorious class of hazardous compounds found in petrol, diesel, oil, as well as in coal-tar and its derivatives, are the PAHs. Polyphenols and PAHs are common industrial pollutants and are found as co-contaminants in the environment. They are hydrophobic organic compounds consisting of two or more benzene rings fused into a single aromatic structure. They may form naturally from burning of organic matter or from production and partial combustion of fossil fuels (Joner *et al.* 2002). Hopanes, complex alicyclic compounds, are of the most environmentally persistent components of petroleum spillage (Atlas 1981). Mammalian liver enzymes (cytochrome P-450 and epoxide hydrolase) oxidise certain PAHs to fjord- and bay-region diol-epoxides which, in turn, form covalent adducts with DNA (Bogan *et al.* 2001). Due to this, many PAHs promote effects similar to other carcinogens, once taken up by the body (Guerin 1999; Bogan *et al.* 2001). Sixteen

PAHs have been included in the United States Environment Protection Agency's priority pollutant list (Bogan *et al.* 2001).

2.1.4   Plants and phytoremediation

The presence of plant rhizospheres in hydrocarbon-polluted soils facilitates an increase in microbial numbers and metabolic activity within the soil. Studies have shown that root length, surface area, volume and diameter play a role in the rehabilitative effect of plants in crude oil-polluted soil (Merkl *et al.* 2005). Roots can also improve the physical and chemical properties of pollutant-stressed soil, besides increasing contact between microbes associated with plant roots and pollutants in the soil (Aprill *et al*. 1990). This effect was first described by Hiltner (1904), who defined the rhizosphere as the zone of soil in which microbes are influenced by plant root systems and where soil organisms have an impact on plants. Microbes isolated from the rhizosphere may have root growth-promoting or growth-inhibiting properties (Kuiper *et al*. 2004). Studies of plant species involved in phytoremediation have indicated that various grass species and leguminous plants are suitable for biodegradation. It is known that gram-negative rods such as *Pseudomonas* species dominate the rhizosphere (Kuiper *et al*. 2004). Some success in rehabilitation of hydrocarbon-polluted soils has been achieved by phytoremediation. It is defined as the use of plants to remove, destroy or sequester hazardous substances from the environment (Glick 2003). It has been documented that remediation of hydrocarbon-polluted sites is enhanced by cultivation of plants (Merkl *et al.* 2005).

Plants can reduce hydrocarbon levels in the soil, although the mechanism by which this happens is not yet entirely understood. Phytoremediation depends greatly on the stimulation of rhizosphere microorganisms by plant roots (Tesar *et al.* 2002). However, hydrocarbon uptake is limited by the lipophilicity of the hydrocarbons in question, which affects their

passage through the cell membrane. This uptake is thought to be attributed to increased microbial activity in polluted soils, as supported by community levels of degrading bacteria increasing during phytoremediation (Wünsche *et al.* 1995; Siciliano *et al.* 2003). BTEX isomers are the most amenable to elimination from the environment by indigenous microorganisms though degradation can be impeded by the natural ecological system (Koizumi *et al.* 2002). Most polluted environments are anoxic, and since aerobic degradation of hydrocarbons is faster than anaerobic processes, their removal can be less efficient in a polluted environment (Koizumi *et al.* 2002).

A variety of grass species, legumes and fast-growing trees such as poplar, alder and willow, with high transpiration rates, have been used in phytoremediation (Jordahl *et al.* 1997). Such plants have extensive root systems that provide large root surface areas available for soil contact. Merkl *et al.* (2005) proved that larger root surface areas are proportionately related to petroleum hydrocarbon degradation levels in the plant genera *Brachiaria*, *Cyperus* and *Eleusine*. Plant roots provide attachment sites to microbes and a source of nutrients, consisting mainly of organic acids, including amino acids, as well as sugars and complex carbohydrates, in the form of exudates (Mehmannavaz *et al*. 2002; Tesar *et al*. 2002). By way of example, Jordahl *et al*. (1997) reported that the number of microbes degrading benzene, toluene and xylene are five times higher in the rhizosphere of poplar trees than in surrounding soil. Successful rhizoremediation by plants depends on factors such as primary and secondary metabolites, colonisation, survival and ecological interactions with other organisms. In addition, the mucigel secreted by root cells, lost root cap cells, starvation of root cells and the decay of complete roots also provides nutrients (Reilley *et al*. 1996). Thus, plant roots have been suggested as a substitute for tilling of soil to incorporate additives and to improve aeration as a method of remediation (Kuiper et al. 2004). A broad phylogenetic range of bacteria, including the genera *Achromobacter*, *Acidovorax*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Corynebacterium*, *Flavobacterium*, *Micrococcus*, *Mycobacteium*, *Norcardia*,

Pseudomonas, *Rhodococcus*, *Sphingomonas* and *Xanthomonas* are involved in the breakdown of hydrocarbons (Tesar *et al.* 2002).

Soil microbial communities are influenced by plant roots in various ways, e.g. excretion of organic compounds and competition for nutrients and attachment surfaces. Kuiper *et al.* (2004) reported that *Cyperus esculentus* L., *Eleusine coracana* (L.) Geartn. and *Brantha serratia* L. rhizospheres accommodate a large variety of bacteria. This probably is due to their ability to harbour large numbers of microorganisms on their highly-branched root systems. Plants with extensive root systems provide larger root-soil surface areas for attachment of microbes (Tesar *et al.* 2002). Plants influence soil pH, moisture and oxygen content by secretion of substances into the surrounding rhizosphere (Schroth and Hilderbrand 1964). Root exudates are common to all higher plants and are known to influence the abiotic and biotic environment of the rhizosphere (Schroth and Hilderbrand 1964). Studies characterising the culturable rhizosphere bacteria showed that plants have specific effects on communities. However, these bacteria represent only a very small component of those actually present in soil (Duineveld *et al.* 2001).

2.1.5 Biodegradation

Indigenous microorganisms, including bacteria and fungi, are able to degrade PAHs in soil, leading to *in situ* rehabilitation of the soils. Bioremediation of hydrocarbon-polluted soils using microbes for detoxification and rehabilitation is an efficient, economic and versatile environmental treatment. PAH-degrading microbes are pervasive in ecosystems where pollutants may serve as carbon sources, and seem to establish themselves soon after pollution occurs (Margesin *et al.* 2000). The reclamation of polluted land reduces the possibility that groundwater will become polluted, and enhances the rate of biodegradation (Gibson and Parales 2000; Mishra *et al.* 2001). It has been shown that hydrocarbon-degrading bacteria are

ubiquitously distributed in natural pristine environments. Wünsche *et al.* (1995), for instance, reported a 3.6% baseline community of hydrocarbon utilising bacteria that increased on addition of hydrocarbon pollutants. Thus, natural degradation of pollutants in low-risk oil-polluted sites is a cost-effective rehabilitation alternative to more traditional clean-up procedures (Gibson and Parales 2000; Margesin and Schinner 2001). Microbes have also been shown to use BTEX compounds as electron-donors in their metabolism, thereby facilitating pollution remediation in affected sites (Stephen *et al*. 1999). Supporting this, Wünsche *et al.* (1995) reported that substrate utilisation patterns in the Biolog system changed upon addition of hydrocarbons. Previously pristine soil bacterial communities shifted to a predominantly *Pseudomonas* population with hydrocarbon degradation capability, thus demonstrating a natural bioremediation adaptation potential. Similarly, Maila *et al*. (2005b), using a combination of Biolog™ and molecular methods, found that pollution removal by indigenous microbial communities at different soil levels was 48% in topsoil, 31% at 1m deep and 11% at 1.5m deep. Thus, PAHs and phenols have been shown to be biodegradable under appropriate conditions (Guerin 1999). However, the most readily degraded hydrocarbons are the n-alkanes with a relative molecular mass of up to n-C_{44} (Atlas 1981). Biodegradation of these n-alkanes commences via a mono-terminal attack, forming a primary alcohol, an aldehyde and a monocarboxylic acid. Further degradation is via β-oxidation forming a two-carbon unit, shorter fatty acids, acetyl co-enzyme A and CO_2 (Atlas 1981). Various bacteria are known to catabolise two- to four-ring PAHs as sole source of carbon, thus rendering them good candidates for site remediation applications (Bogan *et al.* 2001). This catabolism takes place using aromatic hydrocarbon dioxygenases within multicomponent enzyme systems (Samanta *et al*. 2002). Dioxygen is added to the aromatic nucleus of the PAH in question, forming an arene cis-diol as follows:
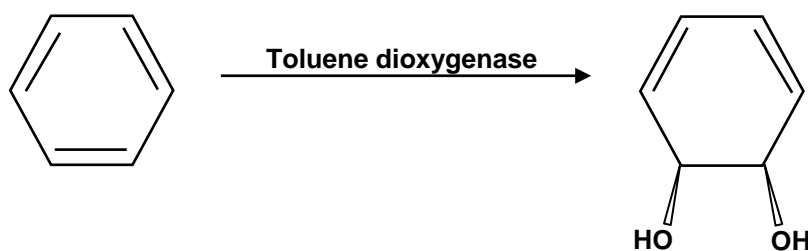
(Gibson and Parales 2000)

It has been hypothesised that metabolic engineering may improve microbial capacity for degradation of toxic compounds. However, the efficiency of naturally occurring organisms capable of this metabolism could be enhanced by optimising bioavailability, adsorption and mass transfer (Samanta *et al*. 2002). Widada *et al*. (2002) isolated 19 PAH-degrading bacterial species belonging to the genera *Ralstonia*, *Sphingomonas*, *Burkholderia*, *Pseudomonas*, *Comamonas*, *Flavobacterium* and *Bacillus* from environmental samples in Kuwait, Indonesia, Thailand and Japan. Enrichment cultures from these samples were supplemented with either naphthalene or phenanthrene as sole carbon source and multiple phenotypes, in terms of utilisation and degradation metabolism, were observed. Tesar *et al.* (2002) listed a broad range of bacterial genera capable of hydrocarbon breakdown, including *Achromobacter*, *Acidovorax*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Corynebacterium*, *Flavobacterium*, *Micrococcus*, *Mycobacterium*, *Nocardia*, *Pseudomonas*, *Rhodococcus*, *Sphingomonas* and *Xanthomonas*. In addition to this, Riis *et al.* (2003) found certain bacteria capable of bioremediation of diesel-polluted soils under high salinities. Bacteria from the genera *Cellulomonas*, *Bacillus*, *Dietzia* and *Halomonas* rehabilitated soils with a salinity of up to 15% (Riis *et al*. 2003). Recently, Kleinsteuber *et al*. (2006) determined that salinity affects the dominant species in diesel-polluted soils differently, low salinity favouring *Sphingomonas* spp., higher salinities *Ralstonia* spp. and very high salinities the halophilic genera *Halomonas*, *Dietzia* and *Alcanivorax*. Some bacteria have been described to degrade specific PAHs in culture. Willison (2004), for instance, found a species designated *Sphingomonas* sp. CHY-1

capable of degrading chrysene as sole carbon source in culture after enrichment. More specifically, members of the *Providencia* genus are known to completely break down hexahydro-1,3,5-trinitro-1,3,5-tiazine (RDX) and nitroso-RDX, and have been used for this purpose in bioremediation (Kitts *et al.* 1994).

Ecto- and endomycorrhizal fungi are cosmopolitan and form symbiotic associations with the roots of plants (Linderman 1988). These endophytic fungi, particularly the ectomycorrhizae, aid plants in the absorption of nutrients from soil, especially immobile elements such as zinc, copper, sulphur, calcium, potassium, iron, magnesium, manganese, chlorine, boron and nitrogen. Absorption of phosphorus is enhanced by both ecto- and endomycorrhizae (Linderman 1988). Mycorrhizal fungi have been reported to reduce plant responses to other stresses such as high salt levels and noxious compounds associated with mine pollution, landfills, heavy metals and micro-element toxicity (Linderman 1988).

Bioremediation, by virtue of biodegradation, depends primarily on overcoming any nutrient limitations in the soil to be rehabilitated. Remediation of hydrocarbon-polluted soils is usually limited by the amount of free carbon, phosphorus and/or nitrogen present (Bogan *et al.* 2001; Margesin and Schinner 2001; Röling *et al.* 2002). However, Struthers *et al.* (1998) found that the herbicide atrazine could be degraded by *Agrobacterium radiobacter* in soil without addition of extra carbon or nitrogen sources, although inoculated cell numbers did not increase, indicating a state of survival rather than growth. Microbial community numbers can be increased by the injection of soluble nutrients just below the surface of the soil. This can, however, lead to excessive localised microbial growth in nutrient-injected areas, resulting in "biofouling" (Bogan *et al.* 2001). The use of gaseous formulations has been demonstrated to better distribute nutrients throughout a system for bioremediation purposes (Bogan *et al.* 2001). Rather than injecting nutrients, nutrient supplementing, particularly with nitrogen and phosphorus fertilisers, is known to enhance biodegradation of oil released into a marine

31

environment (Kasai *et al.* 2002). However, amendments to rectify nutrient deficiencies must be optimal as too high amounts may lead to eutrophication and too little may result in suboptimal biodegradation (Röling *et al.* 2002). Triethylphosphate (TEP) and tributylphosphate (TBP) are the safest phosphorus compounds that can readily be gasified and forced through deficient soil, whereas gaseous nitrous oxide has been used to supply nitrogen (Bogan *et al.* 2001). While not enhancing remediation of PAH-polluted soil, delivery of gaseous nutrients has been shown to expedite *in situ* remediation of soils polluted with chlorinated solvents, volatile organic compounds, C_4-C_{10} alkanes and monoaromatic hydrocarbons (Bogan *et al.* 2001). Lee *et al.* (2003) found that adding pyruvate at optimal levels to PAH-polluted soils as an additional carbon source, aided in the breakdown of PAHs (naphthalene used in model). They were able to determine the optimal concentrations of carbon sources for complete degradation of naphthalene by *Pseudomonas putida* G7.

Microorganisms intended for inoculation into polluted soils can be carried on various materials. Agricultural by-products are most commonly used to transfer microbes without affecting their degradative capacity (Mishra *et al.* 2001). In this respect, the rate and intensity of pollutant degradation in influenced by environmental factors such as the original indigenous microbial community, nutrient availability, oxygen levels, pH, temperature, moisture content, quality, quantity and bioavailability of pollutants, and soil properties (Margesin *et al*. 2000). Although bioremediation is the primary mechanism involved in removal of soil pollutants, other processes such as dispersion, dilution, sorption, volatilisation and abiotic transformation are also instrumental in the rehabilitation process (Margesin and Schinner 2001).

2.2   Rhizosphere

The rhizosphere is a niche that maintains indigenous soil microbial communities involved in the plant-soil nutrient cycle. It also plays a vital part in the survival of plants under adverse chemical soil conditions (Izaguirre-Mayoral *et al.* 2002). Phytoremediation uses rhizosphere technology in biodegradation enhancement. Plant health can be influenced by the promotion of production of phytohormones, furnishing of nutrients, nitrogen fixation, and the suppression of microbes detrimental to plants through antagonism (Da Silva *et al.* 2003). Siciliano *et al.* (2003) demonstrated that effective TPH phytoremediation systems promote the increase in numbers of bacteria with hydrocarbon catabolic genes. PAHs may be removed by volatilisation, photo-oxidation, sorption and leaching. This is enhanced by the presence of plants (Joner *et al.* 2002).

2.2.1   Exudates

Rhizosphere soil is modified with respect to pH, O_2, CO_2 and nutrient availability. Plants exude readily degradable substances into the soil that augment microbial activity in the rhizosphere (Schroth and Hildebrand 1964; Joner *et al.* 2002). These substances are released via volatilisation, leaching, exudation or decomposition and can influence the growth of other organisms in the soil, including that of nearby plants (Meissner *et al.* 1986).

The exact composition of root exudates in soil is unknown, mainly as a result of sloughing and autolysis of epidermal cells constantly affecting the environment (Schroth and Snyder 1961). However, three aspects of modified soil characteristics in the rhizosphere contribute to phytoremediation of organic pollutants, viz. higher microbial activity, higher oxidation potential, and modified microbial community (Joner *et al.* 2002). Plant secondary compounds (exudates) found in rhizosphere soil can include polyphenols and flavanoids. Some of these

compounds are suppressive to microbial growth while others enhance it (D'Arcy Lameta and Jay 1987). Thus, microbial communities within the rhizosphere are definitely affected by the type of root exudates produced by plants. In combination with bacterial PAH-degradative ability, plant roots contain soluble and wall-bound oxidative enzymes that are directly implicated in PAH-degradation (Joner *et al.* 2002). Phytoremediation systems, including the plant and its microbial rhizosphere community, can therefore be implemented as a means of increasing the hydrocarbon degradation potential of soil, but fertilisation is required for maximum results (Siciliano *et al.* 2003).

2.2.2   Microbial communities

The "population concept" is central to the fields of ecology, evolutionary biology and conservation biology. Krebs (1994) defined a population as "a group of organisms of the same species occupying a particular space at a particular time". Waples and Gaggiotti (2006) recently reviewed the definition of a population when considered in the context of ecological and evolutionary paradigms, and suggested several criteria for determining when groups of individuals are different enough to be considered separate communities. A natural population is bounded by ecological or genetic barriers only, for example within a local population individuals interact ecologically and reproductively. Based on this interaction, Waples and Gaggiotti (2006) concluded that a cluster of individuals without using locality sampling information detects true communities only under moderate to low gene flow. Therefore, for the purposes of this thesis, studying a large number of different species interacting within an environment will be referred to as studying a community. Thus, due to gene flow between communities within a community, it follows that the fairly recent advent of DNA markers has led to a great interest in studying natural communities genetically.

Soil microbial communities are relatively evenly distributed in unpolluted environments. However, Smalla *et al.* (2001) proved that there is a reduced evenness in the rhizosphere compared to unplanted soil. Zhou *et al. (*2002) examined microbial communities in 29 different soil types. They found that in low-carbon soils the diversity pattern of the surface soil was evenly distributed, while subsurface samples exhibited a distinct pattern. High-carbon soils, by contrast, displayed uniform diversity throughout the soil layers examined, indicating that spatial isolation differences in community structure could be overcome when the carbon content of a soil is high.

The general assumption stands that higher microbial diversity is proportional to an increased catabolic potential (Dejonghe *et al.* 2001). This can be extrapolated to imply that high species diversity leads to more effective removal of metabolites and pollutants from a substrate. Improving the bioremoval capacity of the soil by inoculating specific strains or consortia of microorganisms is referred to as bioaugmentation (Halden *et al.* 1999; Dejonghe *et al.* 2001). Two components constitute diversity in an environment, viz. total number of species present (species richness/abundance) and species distribution (species equitability) (Dejonghe *et al.* 2001). To promote and increase the degradative potential of a microbial community, competence for certain reactions under the conditions is required, implying that genes within the system need to be activated to participate in the energy flux of the environment (Dejonghe *et al.* 2001).

2.2.3   Assessment of species richness and diversity

Several methods are available to determine the richness of diversity in an environment, including different plating methods, light and fluorescence microscopy, and DNA and RNA analysis (Dejonghe *et al.* 2001; Duineveld *et al.* 2001; Torsvik and Øvereås 2002). There are some general limitations to be taken into account when studying microbial diversity. Spatial

heterogeneity is noteworthy since most environmental replicates consist of 1-5g of sample material, which does not give a true reflection of the spatial distribution of microorganisms (Kirk *et al*. 2004). Culturing colony-forming units (cfu) on different media was the most popular method for investigating microbial diversity. However, most bacteria targeted for isolation from environmental samples are difficult to culture due to constraints imposed by artificial media on which they are to be grown (Sekiguchi *et al.* 2002). Culture-based methods are tedious and certain organisms, e.g. mycobacteria, can take a long time before starting to grow. Only 1-10% of global bacterial species are culturable due to the selectivity of growth media and conditions (McCaig *et* al. 1999; Von Wintzingerode *et al.* 2002; Kirk *et al.* 2004). Less than 1% of microbes from soils in polluted environments are culturable (Stephen *et al*. 1999). Respiration analysis of individual cells within soil samples indicated higher numbers of metabolically active bacteria than the number of culturable bacteria (McCaig *et* al. 1999). Thus, both microscopy and plating lack the capacity to discriminate between multiple bacterial communities and to assess their diversity (Duineveld *et al.* 2001). Furthermore, should an organism be cultured on an artificial medium, substances produced by the organism in culture can either inhibit or stimulate growth of other microbes. These substances may have a markedly reduced effect once introduced into soil as an ameliorant due to pH, adsorption by clay and microbial utilisation, all of which can influence the rhizosphere (Schroth and Hilderbrand 1964).

Molecular methods have provided a more accurate view of species richness within diversity. Initially, random fragments of environmental genomic DNA were cloned and those containing rRNA genes were selected for sequencing (Dejonghe *et al.* 2001). The next advance in molecular analysis came when PCR was used to selectively amplify these rRNA genes from total microbial community DNA, using different sets of primers to amplify the genes from all types of organisms (Archaea, Bacteria, Eukarya) (Dejonghe *et al.* 2001; Torsvik and Øvereås 2002). Ahn *et al*. (1999) probed DNA from PAH-polluted soil for

naphthalene and other PAH metabolism. They found that most PAH-degrading bacteria had a NAH7-like genotype using the *nah*A probe, and only 15% were not detected using this probe. New gene probes were thus suggested for enumeration of PAH-degrades. The next logical step from this technology was that mixed PCR fragments could be cloned and sequenced or be separated and visualised by various fingerprinting techniques, e.g. DGGE (Dejonghe *et al.* 2001; Duineveld *et al.* 2001). However, these techniques are only as efficient as their methodologies, i.e. efficient cell lysis, maximum unsheared DNA extraction, unbiased PCR amplification and effective downstream analysis (Kirk *et al*. 2004).

2.2.4   Remediation

Several methods are available for determining the level of remediation in polluted soils. Screening for the disappearance of pollutants can be achieved by monitoring toxicity in a test organism for product or change. Classically, species used for toxicity response have been *Ceriodaphnia* (crustacean of the family Daphniidae) and *Pimephales promelas* Rafinesque (a fish, commonly known as "Fathead minnow", of the family Cyprinidae) in water, and several invertebrates in soils (White *et al.* 1998). However, analysis of microbial communities have since proved to be a far more comprehensive indicator of residual pollutants. Monitoring the return of a baseline community is used to indicate that the biological community of a soil is returning to normal (White *et al.* 1998). Li *et al*. (2006) found that species of tolerant bacteria in heavy metal-polluted soils increase in numbers with time and further pollution and can consequently be indicative of the level of heavy metal pollution and thus of soil quality.

Rhizosphere microflora are not easily destabilised due to the buffering effect of the biotic and abiotic surroundings they inhabit (Bahme *et al.* 1988). Research has shown, however, that the rhizosphere microflora can be altered by inoculation of plant roots with specific rhizobacteria. The capacity of the shift in microflora depends on several factors, e.g. the nature of the

introduced strain, the effectiveness of its colonisation and its ability to persist on root systems for a prolonged period (Bahme *et al.* 1988). The inoculum size and mode of delivery affects the community dynamics within the soil, i.e. community density declines proportionately to the distance from the point/source of inoculation (Bahme *et al.* 1988). Two delivery systems for applying rhizobacteria to underground plant organs have been described by Bahme *et al.* (1988), namely bacteria-impregnated granules that are mechanically incorporated into soil before planting, and low-pressure drip-irrigation systems containing the desired bacterial strain.

Burkholderia species are regularly isolated from plant rhizospheres, thus making them good potential agents for rhizoremediation. O'Sullivan and Mahenthiralingam (2005) found *Burkholderia* to be the predominant genus isolated from PAH-polluted soils. Of the various *Burkholderia* strains, six (CSV90, EML1549, K712, RASC, TFD2 and TFD6) also capable of 2,4-dichlorophenoxyacetate degradation. *B. xenovorans* strain LB400 is an aerobic degrader of polychlorinated biphenyls (PCBs) using the enzyme biphenyl-2,3-dioxygenase. This species can break down up to hexachlorinated biphenyls when supplemented with maltotriose esters to increase water solubility and hence bioavailability (O'Sullivan and Mahenthiralingam 2005). *B. vietnamiensis* strain G4 is able to co-metabolise trichloroethylene (TCE), which is an organic pollutant in groundwater aquifers, and toluene or phenol, using the enzyme toluene *o*-monooxygenase. Strain G4 has been extensively studied and is subject to two US patents, 4925802 and 5543317 (O'Sullivan and Mahenthiralingam 2005). Strain G4 preferentially degrades toluene in culture and therefore toluene levels have to be maintained to achieve maximum (100%) TCE biodegradation. Since toluene and phenol cannot be used during *in situ* environmental rehabilitation, a mutant of the G4 strain, PR1, which does not require additional nutrients, has been engineered to remove most TCE within a few weeks. Despite this, the G4 strain still proved to be a more efficient bioremediator. A mutant toluene *o*-monooxygenase gene was therefore spliced from G4 into *Escherichia coli* to

yield an organism with a higher rate of TCE degradation and with an enhanced PAH as well as naphthalene degradation capacity (O'Sullivan and Mahenthiralingam 2005).

Petrol and diesel, as well as crude oil spills in soils at fuel stations, have been found to be bioaugmented to a certain extent by members of the genera *Micrococcus, Corynebacterium, Flavobacterium, Bacillus* and *Pseudomonas* (Rahman *et al.* 2002). More specifically, pentachlorophenol was remediated with *Flavobacterium* and *Arthrobacter,* whereas augmentation of 2,4,5-trichlorophenooxyacetic acid with *Rhodococcus chlorophenolicus* and *Pseudomonas cepacia* accelerated its removal (Halden *et al.* 1999). Petroleum PAHs in a marine environment are known to be biodegraded by bacteria belonging to the genus *Cycloclasticus* (Kasai *et al.* 2002). Less species-specifically, Da Silva *et al.* (2003a) found a number of *Paenibacillus* species to have agricultural importance due to their ability to degrade several PAHs.

There has been much focus on the use of bacteria for bioremediation purposes in recent research. However, fungi may also play an important role in the rehabilitation process. In general, fungi are capable of tolerating harsher environmental conditions than bacteria and could well be involved in the degradation of petroleum hydrocarbons in soil (Prenafeta-Boldú *et al.* 2002). Da Silva *et al.* (2003b) isolated filamentous fungi from estuarine sediments in Brazil and monitored their ability to degrade PAHs, particularly pyrene, in culture. They found a *Cyclothyrium* sp. to be the most efficient, simultaneously degrading 74, 70, 59 and 38% of pyrene, phenanthrene, anthracene and benzo[a]pyrene, respectively. Additionally, toluene, ethylbenzene and xylene have been shown to be degraded by a *Cladophialophora* sp. (Prenafeta-Boldú *et al.* 2002).

2.3 Techniques for culture-independent assessment of microbial communities

Culturable proportions of bacterial communities from the environment are negligible compared with the number of species that are present. Thus, culture techniques for environmental bacterial community diversity analysis are becoming obsolete. Øvereås and Torsvik (1998) compared culturable bacterial diversity of agricultural soil communities with diversity obtained by molecular means. They found that molecular methods revealed a much higher bacterial diversity than classical isolation techniques, and concluded that bacterial diversity studies should embrace entire communities, not only the culturable portion.

Several molecular techniques have been developed to identify and determine species diversity of microorganisms without isolation (Kawai *et al.* 2002). PCR-based techniques are becoming increasingly popular for research ranging from diagnostic work to genome fingerprinting and probing (Torsvik and Øvereås 2002). PCR is regularly applied to assay environmental samples due to the ability of the technique to detect relatively small numbers of target organisms without requiring cell culture (Volossiouk *et al.* 1995). Thus, PCR can be used to target certain types of genes expected within specific communities and performing specialised functions. Sei *et al.* (2003) developed a set of primers for detecting and monitoring alkane-degrading bacteria. The primers were designed to target the homologous regions of alkane hydroxylase genes (alk genes) and thus assess the alkane-degrading potential of a particular environment. These primers were tested on communities capable of degrading n-alkanes, the major component of crude oil. According to Sei *et al*. (2003) it was found that shorter n-alkane chains were degraded first by Group I alkane-degrading bacteria, whereas Group III alkane-degrading bacteria degraded longer chains later. However, as with most techniques there are some drawbacks to using PCR, e.g. preferential amplification of certain types of sequences, chimeric sequence generation and false results due to pollution (Osborne *et al.* 2005). Despite this, PCR remains reliable and forms the base-technique for

most molecular work.

Ribosomal RNA (rRNA) molecules are used as molecular chronometers due to their high degree of structural and functional conservation. Consequently, domains within rRNA molecules harbour independent rates of sequence change (Kent and Triplett 2002). Phylogenetic relationships can be determined by examining these changes over time (Kent and Triplett 2002).

Initial assessment of soils, using culture-independent methodologies, revealed the presence of three main bacterial divisions, viz. *Proteobacteria*, *Fibrobacter* and low GC gram-positive bacteria (Kent and Triplett 2002). Specific genes coding for enzymes that are known to be involved in hydrocarbon catabolism have been identified. Widmer *et al.* (1998), realising the potential of environmental microorganisms, specifically *Pseudomonas* species, developed a PCR protocol for selective detection of *Pseudomonas* (*sensu stricto*) in the environment. They designed a highly-selective primer pair for the 16 rRNA genes of *Pseudomonas* species that was used with 91.7% efficacy for bacterial identification from the environment based on sequence phylogeny. Following this, Milcic-Terzic *et al.* (2001) and Whyte *et al.* (2001) combined culture-dependent methods and molecular analysis using hydrocarbon catabolic gene probes *alkB* (C_6-C_{32} n-paraffin degradation), *xylE* (toluene and xylene degradation) and *ndoB* (naphthalene degradation) to demonstrate the presence of hydrocarbon-degrading microbes in polluted soils.

Nitrogen-fixing microorganisms can be instrumental in hydrocarbon pollution bioremediation (see 2.1.4). However, they are difficult to culture due to their different growth requirements and physiology limiting simultaneous cultivation of separate species (Widmer *et al.* 1999). Molecular methods for identifying the presence of nitrogen-fixing *Bacteria* and *Archaea* are now available through the design of broad-spectrum highly degenerate primers. *nifH* is the

general marker gene in nitrogen-fixing bacteria and encodes the enzyme nitrogen reductase. It has an extensive database of sequences available for comparative purposes. Rosado *et al*. (1998) studied the diversity of *nifH* gene sequences in *Paenibacillus azotofixans* and found sequence divergence at DNA level, but more conserved sequence at protein level, hence the design of degenerate primers. Widmer *et al*. (1999) followed suit and designed two universal sets of degenerate primers for nested PCR, based on the amino acid sequence of the conserved *nifH* gene.

2.3.1   Microbial community analysis

Microbial community analysis, independent of culturing the organisms, involves the extraction of signature biochemicals from the environmental samples (Blackwood *et al*. 2003). The first culture-independent estimate of prokaryotic organisms in soil indicated 4600 distinct genomes in one gram of soil (Torsvik *et al*. 1990a). Extracted DNA or RNA can, via molecular genetic techniques, facilitate microbial community analysis to be coupled with phylogeny. The uncultured diversity will reflect species closely related to known cultured organisms and also species from virtually uncultured lineages (Blackwood *et al*. 2003).

Characterisation of genes of microbes involved in the degradation of organic pollutants has led to the application of molecular techniques in microbial ecology of polluted areas (Milcic-Terzic *et al*. 2001). Molecular methods usually involve the separation of PCR amplicons on the basis of DNA nucleotide sequence differences, most often the 16S rRNA gene. However, taxonomic resolution of 16S rDNA sequences can be insufficient for discriminating between closely-related organisms in e.g. cyanobacteria, where the rRNA 16S to 23S internal transcribed spacer (rRNA-ITS) provided better distinction between species (Janse *et al.* 2003). Molecular methods include DGGE, ribosomal intergenic spacer analysis (RISA), single strand conformation polymorphism (SSCP), amplified ribosomal DNA restriction analysis

(ARDRA) and terminal restriction fragment length polymorphism (T-RFLP). Several of these methods, such as SSCP, ARDRA and T-RFLP, do not reveal diversity unless the community is very simple, due to only a very small number of species indicated in rehybridisation or sequence analysis being visualised on a gel (Nakatsu *et al.* 2000; Blackwood *et al.* 2003). However, catabolic gene probes can be used in nucleic acid hybridisations to characterise sequences (Milcic-Terzic *et al.* 2001). Laurie and Lloyd-Jones (1999) probed a set of genes isolated from *Burkholderia* Sp. RP007 involved in PAH catabolism. They found that the *phn* locus, containing nine open-reading-frames, codes for enzymes degrading naphthalene and phenanthrene.

A rapid means of determining the relative abundance of common species present in a given sample, which do not need to be culturable, is provided by molecular techniques. Gelsomino *et al.* (1999) found after extensive molecular fingerprinting that similar soil types (clay, sand, loam, etc.) tend to contain similar dominating bacteria. Thus it is evident that soil type affects the microbial community present and not only the type of pollution to which they are exposed. Bundy *et al.* (2002) found that comparative bioremediation experiments on different soil types, all polluted with diesel, did not lead to the eventual development of a similar microbial community. They concluded that different soils have different inherent microbial potentials to degrade hydrocarbons. Molecular methods also allow for the elucidation of major differences between communities for testing of hypotheses on the basis of sample comparison (Blackwood *et al.* 2003). However, they do not always reveal the organisms primarily involved in the main energy flux of the system. Soil microbial ecologists suggest that only a few organisms are directly significant at a particular site (Dejonghe *et al.* 2001). If these organisms are targeted for non-culture analysis, more information could be revealed. For example, Leys *et al.* (2005) characterised fast-growing mycobacteria in PAH-polluted soils by means of PCR primers that targeted 16S regions of the *Mycobacterium* genome. PCR-DGGE was then used to distinguish between different species and ultimately in

elucidating the phylogeny (genetic relatedness) of the PAH-degrading species.

2.3.2 Denaturing gradient gel electrophoresis

Denaturing gradient gel electrophoresis (DGGE) is a most appropriate molecular method for monitoring microbial community ecology. Wamberg *et al.* (2003) utilised DGGE to map the bacterial component in the pea (*Pisum sativum* L.) rhizosphere community, and observed that chemical changes in the rhizosphere during plant growth were mirrored by concomitant changes within the bacterial community present. MacNaughton *et al.* (1999) used DGGE to identify community members responsible for bioremediation of a crude oil spill and to monitor community changes and pollution level reduction over time. DGGE relies on variation in genetic sequence of a specific amplified region to differentiate between species within microbial communities (Banks and Alleman 2002; Koizumi *et al.* 2002). PCR product is electrophoresed through a polyacrylamide gel containing a linear DNA-denaturing gradient. The band pattern on the gel forms a genetic fingerprint of the entire community being examined (Gillan 2004). Most commonly, 16S rRNA genes are used to give an overall indication of the species composition of a sample. Partial sequence of this gene has been analysed from environments as complex as soil (Throbäck *et al.* 2004). Bodelier *et al.* (2005) screened the methane-oxidising bacteria from freshwater marshlands using combinations of existing 16S primers. They found that, when combined, direct PCR of universal and specific primers yielded community profiles identical to those obtained from nested amplification.

Although 16S gene analyses presently are the most informative for broad community analyses, other genes can also be examined for community diversity. Functional genes have more sequence variation and can be used to discriminate between closely-related but ecologically different communities. Throbäck *et al.* (2004) exploited the *nirS, nirK* and *nosZ* genes involved in denitrification as more discerning community biomarkers. DGGE has even

44

been extrapolated to applications in plant protection research, including analysis of gut flora of several insect pest species, phylloplane and rhizosphere communities associated with different plant varieties, and the impact of biopesticides on natural microflora (O'Callaghan *et al.* 2003).

2.3.3   Single-strand-conformation polymorphism

SSCP of DNA is used in mutation detection and analysis. It involves the separation of single-stranded PCR rDNA products with the same number of base-pairs but a different conformational structure, on a polyacrylamide gel (Dejonghe *et al.* 2001). This technique has been adapted for the analysis of, and differentiation between, cultivated pure-culture soil microorganisms and non-cultivated rhizosphere microbial communities (Schwieger and Tebbe 1998). Under non-denaturing conditions, single-stranded DNA folds into sequence-dependent secondary conformations. These structures render different electrophoretic motilities to the molecules that can then be separated on a non-denaturing polyacrylamide gel. SSCP can be used in conjunction with an automated DNA sequencer to differentiate between species using PCR products of 16S rRNA (Schwieger and Tebbe 1998). A limitation of using this technique for community DNA analysis is the high rate of reannealing after denaturation, especially at high DNA concentrations. Another constraint of SSCP is the appearance of two bands on electrophoretic gels as a result of only double-stranded PCR product being obtained. Characteristically, three bands are observed on gels, one of a double-stranded product and two of the single-stranded DNA molecules from PCR. In some instances, there may be four or more bands visible on the gel due to differing structural conformations, e.g. hairpin folding due to palindromic sequences. Likewise, physical conformation of products may be similar, causing them to overlap in the gel, resulting in fewer bands being visualised on a gel. Finally, heteroduplex DNA strands with a similar sequence adhere together, forming breathing heteroduplexes of two or more PCR products (Schwieger and Tebbe 1998).

2.3.4 Amplified ribosomal DNA restriction analysis

Another PCR-based DNA-fingerprinting technique, which makes use of restriction of amplified fragments, is amplified ribosomal DNA restriction analysis (ARDRA). This technique yields a high number of bands per species, and therefore cannot provide reliable genotypic characterisation at community level (Dejonghe *et al.* 2001). It is, however, particularly suitable for monitoring communities and assessing microbial diversity, and can focus on specific sub-groups within a community (Dejonghe *et al.* 2001). Lagacé *et al.* (2004) made use of 16S rDNA sequencing of ARDRA fragments for identifying bacterial communities in maple sap. The ARDRA profiles yielded a dendogram illustrating relationships between bacterial strains, and γ-proteobacteria were found to be dominant throughout the year.

2.3.5 Reverse transcription PCR

RT-PCR involves the extraction of RNA instead of DNA, and profiles the metabolically active microorganisms in a system (Dejonghe *et al.* 2001). It is a dual-step process. The first step entails the production of complementary DNA (cDNA) from a messenger RNA (mRNA) template using dNTPs and an RNA-dependent reverse transcriptase at 37°C. The second step involves the use of a thermostable transcriptase and a set of upstream and downstream DNA primers. Temperatures fluctuating between 38-95°C facilitate sequence-specific binding of the primers to the cDNA and allow transcriptase to produce double-stranded DNA. After approximately 30 cycles, the original RNA template is degraded by RNAse H, leaving pure cDNA in solution. It is now possible to simplify this process into a single step by using wax beads, containing the required enzymes, that melt at the higher temperatures releasing their contents.

Exponential amplification via RT-PCR provides a highly sensitive technique that can detect very low copy number RNAs. This technique is widely used in the diagnosis of genetic diseases and in the quantitative determination of the abundance of specific different RNA molecules within a cell or tissue as a measure of gene expression, e.g. Northern blot.

2.3.6 Base-specific fragmentation and mass spectrometry

Base-specific fragmentation of PCR-amplified 16S rDNA followed by mass spectrometry of the fragment pattern is being used for rapid identification of bacteria (Von Wintzingerode *et al.* 2002). This method is inherently accurate and rapid, making it attractive as a tool for high-throughput microbe identification in pharmaceutical and industrial applications.

2.3.7 Signature lipid biomarker analysis/environmental nucleic acid probes

Signature lipid biomarkers can be used in biomas shift monitoring. Signature lipid biomarker analysis/environmental nucleic acid probes (SLB/ENAP) are relatively inexpensive molecular fingerprinting techniques used to ascertain a quantitative measurement of the microcosm. Chemical extraction of phospholipid fatty acids from the soil can be useful in determining the diversity within the soil and in estimation of the microbial biomass (Banks and Alleman 2002). It determines when community ecology becomes analogous to a known community that is considered to be safe (White *et al.* 1998). Total cellular phospholipid fatty acids (PLFAs) are not stored in cells and thus have a rapid turnover in communities. These make ideal markers for monitoring viable biomass within a community viz. an increase in cis/trans monoenic PLFAs in cells is indicative of toxic stress within bacterial communities and thus results in a change in their growth phase (Stephen *et al.* 1999). Specific PLFA biomarkers can be used to indicate broad microbial community diversity encompassing bacteria, fungi,

algae, gram-negative and -positive organisms, sphingomonads, actinomycetes and sulphate-reducing bacteria. Limitations of PFLA analysis include shortcomings in analysis of gram-negative communities.   These profiles are dominated by monoenoic, saturated and cyclopropane fatty acids that are broadly distributed and thus fairly uninformative with regard to gram-negative population structure. This method has been combined with nucleic acid-based analysis such as DGGE to allow for better community elucidation (Stephen *et al*. 1999).

2.3.8   Terminal restriction fragment length polymorphism

Terminal restriction fragment length polymorphism (T-RFLP) is a culture-independent method used to obtain a genetic fingerprint of a microbial community and has been shown to be effective in discriminating between microbial communities in various environments (Blackwood *et al.* 2003). Automation increases sample throughput and accelerates analysis of bacterial communities (Kent and Triplett 2002). PCR product of 16S rDNA is used for analysis (Dejonghe *et al.* 2001). One end of the PCR product is tagged with a primer carrying a fluorescent dye. It is then cut with a restriction enzyme to form terminal restriction fragments (T-RFs) that are separated by gel electrophoresis and visualised by excitation of the fluor (Dejonghe *et al.* 2001; Blackwood *et al.* 2003). A banding pattern is obtained, each band corresponding to one species or "ribotype" (Dejonghe *et al.* 2001). This provides quantitative data on each of the T-RFs in the form of size of base-pairs and intensity of fluorescence (peak height) (Blackwood *et al.* 2003). T-RF sizes can then be compared with a theoretical database obtained from sequence information (Blackwood *et al.* 2003), thus providing the species richness as well as community structure of the ecosystem (Dejonghe *et al.* 2001).

2.3.9   Other techniques

A method for detecting extracellular DNA in environmental samples has been developed by England *et al*. (2004). This method circumvents disruption of cell membranes by not employing the use of harsh chemicals or physical disruption of whole cells within samples. England *et al*. (2004) hypothesised that the persistence of extracellular DNA in the environment is partially due to the formation of soil-DNA complexes, whereby the naked DNA released upon cell death and lysis is protected from nuclease degradation by the soil particles to which it adheres. Extracellular DNA serves two purposes in the environment, that of a nutrient source and of a gene pool. This DNA was extracted by using a gentle relatively fast extraction method involving suspension and shaking of a 0.5g sample of leaf litter in 4ml of sodium pyrophosphate (pH 8) followed by several filtration and cleaning steps resulting in application-ready extracellular DNA.

Other techniques such as ribosomal intergenic spacer analysis (RISA), ITS-restriction fragment length polymorphism (ITS-RFLP) and random amplified polymorphic DNA (RAPD) provide complex community profiles that can be analysed for community composition studies (Kent and Triplett 2002). Detection and resolution of fragment analysis can be approached with a number of methods, including automated ribosomal intergenic spacer analysis (ARISA) and length heterogeneity PCR (LH-PCR) (Kent and Triplett 2002). Most probable number (MPN) is a specialised enrichment technique using relevant substrates to estimate the number of organisms in an environment capable of degrading specific pollutants (Banks and Alleman 2002).

A widely used approach to studying bacterial diversity is using clone libraries of 16S rRNA genes. The genes are collected from naturally occurring bacteria through PCR with universal 16S rRNA gene primers (Cottrell and Kirchman 2000). Cottrell and Kirchman (2000) studied

in situ marine microbial communities and found that data from a PCR-based clone library indicate that novel, uncultivated species are widespread in global oceans. However, clone libraries are effected by biases at each step of the method (including sample collection, cell lysis, nucleic acid extraction, PCR amplification, and cloning) and can deviate from the compositions of actual communities (Cottrell and Kirchman 2000). During PCR, using controlled mixtures of 16S ribosomal DNA, the relative abundance of targeted DNA molecules in the final PCR product can be affected by biases. Several precautions have been proposed for minimizing these biases during PCR; however, the amount of bias is not known for natural habitats.

2.3.10 Possible molecular pitfalls

Due to the low number of cultured microorganisms compared to the large numbers of unculturable microbes, microbial diversity cannot be implied by cultured diversity. Therefore PCR-based molecular techniques are favoured to give a better understanding of microbial communities in mixed samples. However, a review by Von Wintzingerode *et al*. (1997) indicated pitfalls of PCR-based genomic analyses. Briefly, they concluded that after initial sample collection several difficulties could be encountered during cell lysis, DNA/RNA extraction, PCR, separation of genes and sequence data analysis. These difficulties include the following:

- Insufficient cell lysis will skew an analysis if not all microbial DNA is released from cells in the sample.

- DNA/RNA can shear into fragments after release from cells during cleaning steps and may impact on post-extraction steps thereafter.

- PCR can be inhibited by co-extracted contaminants such as humic acids from soil that hamper the reaction of template and enzyme. Amplification efficiencies should be the same across molecules, thus assumptions must be made that:

- all molecules are equally accessible to primer hybridisation, the primer-template hybrids form with equal efficacy.

- the extension efficiency of the DNA polymerase is the same across templates.

- exhaustion of reaction components affects all templates equally.

Furthermore, the formation of PCR artefacts can occur due to the creation of chimeras between two homologous molecules, deletion mutants as a result of stable secondary structures, and point mutants because of misincorporation of bases by the DNA polymerase. In addition to this the possibility of contamination as a result of foreign DNA introduced into the reaction due to experimental error must be negated, this is monitored by the incorporation of negative control reactions containing no template DNA.

- Sequence analysis of 16S rDNA is usually done by comparison with previously identified sequences deposited on global databases. However, whether environmental sequences represent uncultured or novel organisms or remain unassigned to known taxa is yet to be determined. Many sequences on the database may be of low quality due to their length (only partial) or taxonomic ambiguity (Kirk *et al*. 2004).

In order to prevent these possible inaccuracies during molecular sample analysis, Von Wintzingerode *et al*. (1997) suggested that results of different extraction methods, PCR and cloning techniques be explored simultaneously to provide the most accurate results possible.

2.4   DGGE technique and application

Muyzer *et al.* (1993) introduced DGGE as a new genetic fingerprinting technique. This method is often preferred due to its capacity to provide rapid visual indications of community changes within a sample (Anderson *et al*. 2003). Bands can then be excised and sequenced. Sequence variation in rRNA has been used for elucidating phylogenetic relationships between

organisms and in designing probes for detecting microbial taxa (Muyzer *et al.* 1993). DGGE is used to determine the microbial genetic diversity and particularly the predominant communities in a sample (Muyzer 1999; Coclin *et al.* 2001; Stamper *et al.* 2003). Janse *et al.* (2003) concluded that it can also be used to determine the purity and uniqueness of isolated strains.

Denaturing gradient gels are used for the detection of non-RFLP polymorphisms (Helms 1990). Double-stranded fragments (200-700 basepairs), the products of PCR of rRNA genes (rDNA) with the same length but differing in base-pair sequences, are separated on an increasing denaturant gradient gel (Ferris *et al.* 1996; Nakatsu *et al.* 2000; Dejonghe *et al.* 2001; Kawai *et al.* 2002). A portion of DNA can be deemed suitable for DGGE analysis if it can be specifically amplified from the target organism, has adequate heterogeneity for good resolution and is part of a gene that has a large database of sequences already available (Janse *et al.* 2003). A factor that limits DGGE efficacy is the primer design. Sequences targeted should not yield a fragment much longer than 500 basepairs (bp) for successful analysis (Throbäck *et al.* 2004). At present 16S rDNA sequences form the ever-increasing, largest gene-specific data set available on internationally accessible databases (> 30 000), making tentative identification of unknown bacteria possible (Von Wintzingerode *et al.* 2002). Øvereås *et al.* (1997) were the first to analyse archaeal rDNA with DGGE. Using domain-specific sets of primers on samples from a meromictic lake in Norway, they found an increase in *Archaea* and a decrease in *Bacteria* the deeper they sampled.

Double-stranded DNA products that undergo electrophoresis through a DGGE gel are halted when they split into single strands due to a linearly increasing gradient of denaturants (Muyzer *et al.* 1993; Curtis and Craine 1998). The denaturants most commonly used are heat (constant 60°C), formamide (0-40%) and urea (0-7M) (Helms 1990). Initially, fragments move according to relative molecular mass. However, as the denaturation gradient increases

the fragments start separating as the hydrogen bonds between the double helix beging to break, this is known as melting (Helms 1990). This partial melting retards the progress of the DNA molecule through the gel, the resultant mobility shift differing for different sequences (Muyzer *et al.* 1993). The sequence of the PCR product separation on the gel determines the denaturant concentration at which this occurs (Ferris *et al.* 1996; Curtis and Craine 1998; Nakatsu *et al.* 2000). As denaturant concentrations increase, the DNA will dissociate completely into two separate strands (Helms 1990). Fragments do not partially melt in a zipper-like fashion, and specific portions of DNA fragment become single-stranded suddenly within a narrow denaturant range (Helms 1990; Muyzer *et al.* 1993). After double-stranded DNA dissociation the gel is stained with a DNA-intercalating dye that fluoresces under ultra-violet light. For the purposes of this review and work, SYBR gold nucleic acid gel stain was used. This stain is an asymmetrical cyanine dye with two fluorescence excitation maxima, *ca*. 300 and 495nm, when bound to DNA (Tuma *et al*. 1999). When used with 300nm transillumination and Polaroid black and white photography, SYBR gold is more sensitive during intercalation than ethidium bromide, forms dye-nucleic acid complexes *ca*. 70% higher than current counterpart dyes, produces up to a 1000-fold fluorescence enhancement, is as sensitive as silver staining but requires only one step, and does not influence subsequent molecular biology protocols (Tuma *et al*. 1999).

Narrowing the denaturant range can increase the sensitivity of DGGE, hence yielding fast, reliable and reproducible results (Fromin *et al*. 2002; Temmerman *et al*. 2003). Mobility rate in the polyacrylamide gel is determined by the physical shape of the fragment, which in turn depends on the denaturant gradient and fragment sequence, with partially melted fragments moving more slowly than those that are still double-stranded (Helms 1990). During analysis of a complex microbial community, a ladder of bands forms on the gel, each corresponding to an individual PCR-product of a specific sequence (Curtis and Craine 1998, Fromin *et al*. 2002). This allows for simultaneous detection of multiple 16S rRNA sequences (Ferris and

Ward 1997; Sekiguchi *et al.* 2002). The resulting gels can be probed with diagnostic oligonucleotides to identify specific sequences or bands, and may be excised, reamplified and sequenced (Ferris *et al.* 1996). The technique is sufficiently sensitive to detect as little as one base-pair difference in a sequence (Helms 1990). However, Gillan (2004) found that changes to the DGGE protocol can result in less robust results and thus should be standardised across particular sets of experiments. Alternatively, "markers" can be constructed from known species sequences and run alongside test samples to determine the identity of bands within the sample. Theunissen *et al.* (2005) demonstrated this when analysing probiotic microorganisms from yoghurt and lyophilised capsule and tablet preparations. Two markers with known lactobacilli and *Bifidobacterium* PCR-product were run adjacent to test samples and band patterns were then used for accurate and rapid species identification. Similarly, but more complex, Keyser *et al.* (2006) used a marker composed of five known methanogenic bacterial species to determine DGGE bands from an upflow anaerobic sludge blanket bioreactor that did not match the marker. These bands were then excised and sequenced, and a DGGE marker to monitor archeal members of the microbial consortium developed based on the sequence results.

Resolution of DGGE can be enhanced by incorporation of a GC-rich sequence into one of the primers to modify the melting behaviour of the fragment and allow for the majority of sequence variation to be detected in the denaturing gel (Ferris *et al.* 1996; Curtis and Craine 1998). A GC-clamp attached to the 5' end of a PCR product prevents complete melting during fragment separation in a denaturing gradient, and sensitises the technique enough to detect all single base changes in PCR fragments of 500bp (Heuer *et al.* 1997). Sheffield *et al.* (1989) found that attaching a GC-clamp of 40-45bp to primers allowed for the determination of single-base-mutations, previously only 40% distinguishable in DGGE analysis, to increase to 100%. Furthermore, Boon *et al.* (2002) included a GC-clamp to stabilise large fragments in all final reactions during nested PCR intended for DGGE analysis. However, despite the

advocation of the inclusion of a GC-clamp for melting stability during PCR-DGGE analysis under certain conditions the clamp can be disregarded. In this case, if no GC-clamp is added, it is recommended that the PCR product must have at least two melting domains (Chang bioscience 2004). Wu *et al.* (1997) found that GC-clamped products with a perfect melting curve yielded distorted smeared results when subjected to DGGE. They found that fragments containing a "high melting domain" provided better DGGE results when run without a GC-clamp, and concluded that if melting analysis of a PCR product predicts a high melting domain of <40bp, and differs by not more than 5°C melting temperature, then the fragment is suitable for DGGE analysis without a 5' GC-clamp.

Lanes of bands can be analysed utilising gel image software for more accurate results, using known pure culture isolates as standards for well-characterised environmental samples. Thus, gel images resulting from DGGE analysis can be digitally captured and used for species identification when samples are run against these known standards (Temmerman *et al.* 2003). These images can also be compared when samples are collected and analysed over a period of time, hence allowing monitoring of community structural changes with time (Van Hannen *et al.* 1999). Manual fine-tuning of the gel image completes the initial analysis and dendograms can be drawn to relate band pattern parallels (Fromin *et al*. 2002; Stamper *et al*. 2003). Software also calculates band densities necessary for determining the Shannon diversity index, where each band represents one species and the band intensity is proportional to the species abundance (Fromin *et al*. 2002; Stamper *et al.* 2003; Andreoni *et al*. 2004). Nübel *et al.* (1999) quantified diversity of oxygenic phototrophs within hypersaline microbial mats. The amount of bands per sample indicated species richness, whereas species abundance/"evenness" was determined by band intensity.

Limitations of DGGE include similar electrophoretic mobilities of phylogenetically related species sharing analogous sequences in the amplified area, and similar melting behaviour

between phylogenetically unrelated species (Smalla *et al.* 2001). Consequently, there may be more than one species represented by a single band on the DGGE gel. This has been demonstrated by Jackson *et al.* (2000) making use of site-directed mutagenesis to create *E. coli* 16S rDNA fragments differing by 1-4 base-pairs. Migration on DGGE gels consistently determined single base-pair changes, but multiple base differences proved to be more difficult to distinguish. Two of the sequences tested differing by two base-pairs only, showed identical migration patterns and could not be separated when run in a mixed sample. Furthermore, Vallaeys *et al.* (1997) reported that DGGE analysis of a 200bp fragment of 16S rDNA from rhizobia and methantrophs was difficult to elucidate due to low and high sequence polymorphism, respectively.

One also needs to take into account the method used for DNA extraction and purification when screening DGGE samples. Niemi *et al.* (2001) tested five different DNA extraction methods and three purification methods on rhizosphere soil samples destined for DGGE analysis. They found that the isolation and purification methods both had an effect on the final bacterial DGGE community structures of the samples. In addition to this, O'Callaghan *et al.* (2003) concluded that extracted DNA should be representative of the habitat, PCR bias must be taken into account as preferential amplification may occur due to inefficient primer annealing, and species determination should not be based on 16S rDNA sequences alone, although this is becoming increasingly more efficient as databases expand continually. There are, however, means of incorporating internal standards into the DNA extraction and PCR-DGGE process. Petersen and Dahllöf (2005) developed a protocol known as Internal Standards in Molecular Analysis of Diversity (ISMAD) that can monitor, and thus account for, experimental variability. A fluorescent 510bp PCR product is included in each sample prior to DNA extraction and recovered afterwards. PCR is monitored by adding non-competitive primers coding for a 140bp section of *Drosophila melanogaster* DNA to the same PCR as the sample. Together these internal controls reduced variation between replicate

samples during DGGE analysis. Despite these minor pitfalls, DGGE is still considered to be a reliable, reproducible, rapid and relatively inexpensive method for the simultaneous analysis of multiple samples and to map community changes over time (Muyzer 1999; Fromin *et al*. 2002).

2.4.1   Community diversity analysis

Most microbial diversity indices are based on plant and animal models, e.g. the Shannon and Simpson indices. As such, there is some difficulty in applying these indices to microbial models since they need a clear definition of species and unambiguous individual identification. This level of identification is difficult in bacteriology. An ideal bacterial index should encompass the following (Watve and Gangal 1996):

  ➢ Three important diversity dimensions, viz. species diversity, species richness/abundance and taxonomic distance between biotypes.

  ➢ Be based on a statistically justified parameter.

  ➢ Be insensitive to possible errors and variability of test results.

  ➢ Not be too sensitive for sample size.

According to this the use of Shannon algorithms to calculate microbial diversity according to DGGE gel fingerprints is acceptable. Dimensions such as diversity and richness/abundance can be determined from the number of bands and their intensity on the gel, respectively. Sequencing of each band on the gel can indicate taxonomic distance between biotypes. Diversity within the 16S rDNA is statistically well-documented and does account for possible errors and variability within the region that can be guarded by incorporating internal control standards. DGGE can be used for assessing anything from one sample individually to a large numbers of samples simultaneously.

DGGE allows for determining community as well as specific population diversity without

further analysis and without elucidating particular individuals (Muyzer 1999). It has also been used for the simultaneous identification of sequence variations in multiple genes among several organisms (Muyzer *et al.* 1993). Identity of community members can be further resolved by hybridisation of the gel with species/taxon-specific oligonucleotide probes to hypervariable regions of the sequence or by cloning and sequencing (Muyzer 1999). The gel can be used for direct analysis of genomic DNA by transferring separation patterns to hybridisation membranes, using capillary- or electro-blotting, and analysis with DNA-probes (Muyzer *et al.* 1993). PCR, with GC-clamp primers, can also be selectively employed to amplify sequences of interest, e.g. 16S, before DGGE is performed (Muyzer *et al.* 1993). Essentially, DGGE allows a high number of samples to be screened simultaneously, thus facilitating more broad-spectrum analysis. Kowalchuk *et al*. (1997) used DGGE to assess variation between different pathogenic fungal species within a taxon attacking the roots of *Ammophila arenaria* L. (marram grass). They amplified a 569bp region of the 18S rDNA gene by means of nested PCR with a GC-clamp on the final PCR. Upon assessing experimental and field/wild plants they were able to distinguish between species of fungi and detect a much higher level of diversity than in previous culture-based surveys.

2.4.2 Community dynamics studies

Due to multiple sample screening, DGGE allows for monitoring of the dynamics that microbial communities undergo during seasonal and environmental fluctuations in their habitat (Muyzer 1999). Ward *et al.* (1998) made use of 16S rDNA fragments in DGGE to study seasonal community changes of microbial communities within hot spring microbial mats. Subsequently, PCR-DGGE has been used to monitor seasonal changes in communities of bacterioplankton, the rhizosphere of chrysanthemum, post-viral bacterial lysis communities, diurnal behaviour of sulphate-reducing and phenol-degrading bacteria in activated sludge, as well as the impact of pesticide and herbicide applications on microbial

communities (Muyzer 1999). DGGE has even been applied in the mapping of communities of bacteria utilising organic-wastewater/sludge as fuel for a microbial electrochemical cell (Kim *et al*. 2004). Results indicated that microbial communities within the cell electrode differed from those in the sludge.

2.4.3   Molecular community mapping across varied environments

Culture techniques are important for the understanding of the physiology and function of microbes isolated from their natural environment. However, molecular tools can be used for monitoring enrichment cultures and facilitating the isolation of target communities from the environment (Muyzer 1999). Smalla *et al.* (1998) made use of DGGE and TGGE (temperature gradient gel electrophoresis) in the analysis of BIOLOG substrate utilisation patterns of two bacterial communities from potato rhizosphere and activated sludge. Both DGGE and TGGE showed enrichment of specific bacterial communities not evident from BIOLOG results. Prokaryotic communities are not the only type to be mapped. Foucher *et al.* (2004) determined nematode diversity in soil samples using 18S rDNA PCR-DGGE, and found a significant relationship between morphological and DGGE estimates of species richness. Marshall *et al.* (2003) tested PCR-DGGE primers for compost fungi, finding an α-elongation factor primer set targeting a portion of the 18S rDNA best for fungal community amplification. Similarly, Zuccaro *et al.* (2003) demonstrated the use of four sets of 18S primers in DGGE analysis for the identification of ascomycetes associated with algae in lichens on ferns.

2.4.4   Niche differentiation

Molecular microbial ecology is becoming more specialised, thus allowing analysis of specific functional communities within communities. Enzyme-coding genes are now being targeted

for ecological studies. They tend to display a higher level of sequence variation than the conserved 16S rDNA genes, which makes them more efficient molecular markers for phylogenetically similar but ecologically distinct communities (Muyzer 1999). In addition, targeting functional genes facilitates the study of specific activities within microbial communities. Milcic-Terzic *et al.* (2001) used genes of microbes involved in the degradation of organic pollutants for the application of molecular techniques in the microbial ecology of polluted areas. As more sequences of functional genes become available on databases worldwide, PCR-DGGE undoubtedly would deliver considerably more information regarding community structure and function.

2.4.5   Determining species diversity

Banding patterns on DGGE gels give an indication of species diversity when analysed using a visual gel analysis software package. For the purposes of the studies included in this thesis, DGGE gel image analysis was performed using the Gel2K program and fingerprints were analysed in a cluster investigation using CLUST (Norland 2004).

Bands excised from DGGE gels can be sequenced. The resulting sequences can then be used for comparative phylogenetic analysis to determine the evolutionary relationships between organisms in the community being analysed. Anderson *et al.* (2003) investigated a soil fungal community by DGGE of the ITS region (ITS1-F with a GC-clamp and ITS2 yielding a 300bp fragment), sequencing of bands, and BLAST result phylogeny of the resulting sequences. Phylogeny gives and indication of species diversity and not richness, since only one band is produced and picked from the gel per species (Van Hannen *et al.* 1999). By determining the closest relatives of unknown organisms the known characteristics can be inferred upon them (Ueda *et al.* 1995). The sequence data can also be used in the design of primers and probes for *in situ* identification of selected organisms.

2.5   Alternatives to PCR-based analyses

Microscopy and plate counts are traditional methods that are quick and inexpensive. Selective plating and direct viable counts can be used for providing information on the active heterotrophic portion of a community (Kirk *et al*. 2004). Methods are available that focus on physiological/metabolic characteristics of microbial communities, e.g. fatty acid methyl ester (FAME) profiles and phospholipid fatty acid analysis (Kent and Triplett 2002). Fluorescent *in situ* hybridisation (FISH) utilises fluorescent oligonucleotides to target rRNA sequences (Dejonghe *et al.* 2001). FISH can be used in conjunction with DAPI (4',6'-diamidino-2-phenylindole), 2-(*p*-iodophenyl)-3-(*p*-nitrophenyl)-5-phenyltetrazolium chloride (INT)-formazan, or 5-cyano-2,3-ditolyl tetrazolium chloride (CTC) staining for determining the contribution made by communities of interest to the total abundance or active cell count (Kent and Triplett 2002). However, FISH has a low throughput and this limits its application for comparison of high numbers of samples (Kent and Triplett 2002).

Various tests are also available for bacterial identification based on physiological reactions. Among these are the catalase reaction test, the oxidative-fermentative Hugh-Leifson test, Biolog and API, a standardised, miniaturised version of existing biochemical test techniques that is simple, rapid and reliable when used in conjunction with numerical identification with or without computer software programmes.

2.5.1   Morphology

Prudent morphological analysis of bacterial cells can yield important information about diversity, microbial abundance and two-dimensional spatial distribution of microbial community members. Computer-aided systems such as CMEIAS (Centre for Microbial Ecology Image Analysis System), is a semi-automated analytic tool that uses processing and

pattern recognition techniques (with microscopy) to gather information on size and shape of digital images of organisms and classify them into their morphotypes (Kent and Triplett 2002).

2.5.2   Catalase reaction

This is a test for production of the enzyme catalase by bacterial species. Hydrogen peroxide is a harmful by-product of metabolic processes, catalase catalyses its breakdown to water and oxygen. The enzyme has one of the highest turnover rates since one molecule of catalase can convert 83 000 molecules of hydrogen peroxide to water and oxygen per second (Wikipedia 2007). Although the catalase test alone cannot identify bacteria, combined with other tests it can aid in identification (Krieg *et al*. 1984). The test is performed by picking bacterial cells from pure cultures on agar plates, using sterile wooden toothpicks, and placed on clean microscope slides. One or two drops of 3% hydrogen peroxide are added to the bacteria and the formation of bubbles within 1min is regarded as a positive reaction.

2.5.3   Aerobic and anaerobic bacteria

The fermentative or oxidative nature of bacteria is determined using the Hugh-Leifson test (Hugh and Leifson 1953). Colonies from pure culture on agar plates are stab-inoculated in duplicate into sterile test tubes containing oxidative fermentative base medium (OFBM) with added glucose. The medium in one tube of each duplicate is covered with 1cm sterile liquid paraffin. Tubes are incubated at 37°C for 48h and a colour change from green to yellow is deemed a positive test result. Bacteria can be considered fermentative when the colour changes from green to yellow in both test tubes. Oxidative bacteria induce a colour change only in the test tube containing no liquid paraffin.

2.5.4   Identification using API and Biolog

API is a series of miniaturised metabolic tests deemed instrumental in bacterial species identification. Pure isolates from agar plates are subcultured on fresh agar medium for 48h. A sterile inoculation loop is then used to suspend cells in test tubes containing 0.85% NaCl. API strips are loaded with this suspension according to the manufacturer's instructions (OMNIMED (Pty.) Ltd.). Several different tests are available for use, e.g. API 50CH, API 20NE, API 20E, etc., based on different characteristics of bacterial species.

Garland and Mills (1991) developed a technique to assess the potential functional diversity of bacterial communities through sole carbon utilisation (SSCU) metabolic patterns. From this arose the gram-negative and gram-positive Biolog plate system that contained 95 different carbon sources and a control well for metabolic bacterial identification (Kirk *et al*. 2004). Biolog EcoPlate™ is specifically tailored for microbial community and ecological studies. Its development was initially prompted when Biolog GN microplates were inoculated with a mixture of microbes in culture and the community fingerprint characteristics were measured over time. Known as community-level-physiological-profiling this method proved to be effective in distinguishing spatial and temporal microbial community changes. The plates proved to be useful in assays of the normal community and to detect changes based on an introduced variable. These studies have been conducted with communities from soil, wastewater, activated sludge, compost and industrial waste. The Biolog EcoPlate contains the 31 most utilised carbon sources for soil community analysis, each of which is repeated in triplicate for data purposes. Communities of organisms yield a characteristic reaction pattern or "metabolic fingerprint". These patterns can be statistically analysed by computer software at defined intervals over 2-5 days, hence providing data about microbial community changes over time. This method has also been compared with other methods such as PLFA and proved to be more sensitive to important factors for instance temperature and water.

2.5.5   DNA reassociation

A non-PCR-based molecular technique has also been established on the basis of DNA melting and reassociation measurements. Comparative chemistry of genomes between species gives an indication of species diversity during DNA-DNA and DNA-RNA reassociation (Sanderson 1976). Purified DNA is split into fragments and thermally denatured so that the double-helix strands separate or "melt" and, by slowly cooling the DNA, reassociate or reanneal again. Following this, the rate at which the double-helix renatures is measured spectrophotometrically (Curtis and Sloan 2005). This rate is affected by the size and complexity of DNA, with large complex DNA reannealing the slowest. Originally, this method was used to estimate size and complexity of genomes from individual organisms. However, Torsvik *et al.* (1990a) reasoned that pooled genomic DNA from a microbial community might reanneal like the DNA from a large genome. They placed sheared total soil DNA in a French press to yield fragments with an average molecular mass of 420 000 daltons. It was then hypothesised that the heterogeneity of the DNA was a measure of genetic diversity of bacteria within the soil. Indeed, they showed that DNA extracted from soil reassociated so slowly that it resembled a genome 7000 times as large as the genome of a single bacterium (Curtis and Sloan 2005). It follows that there could have been at least 7000 different prokaryotic taxa in the sample of soil analysed.

Renaturation of the homologous single-stranded DNA follows second-order reaction kinetics (Torsvik *et al.* 1990a). The renatured DNA fraction is expressed as a product of the nucleotide concentration in moles per litre ($C_o t$), and time is measured in seconds. $C_o t_{1/2}$ under defined conditions is directly proportional to the complexity or genome size of the DNA, complexity being defined as the number of nucleotides in the genome of a haploid cell, excluding repetitive DNA. Based on this, $C_o t_{1/2}$ can be considered to be a diversity index measurement of bacterial communities, which would equate to indices based on phenotypic

analysis or species diversities.

DNA-reassociation has been used in combination with other molecular techniques such as DGGE to give a more complete idea of bacterial diversity within specific communities. Torsvik *et al*. (1998) investigated the community structure of natural, polluted and agriculturally perturbed environments. They compared DGGE diversity analysis of rRNA genes with total DNA reassociation to draw parallels between community diversity techniques. Their study indicated that total soil microbial diversity was 200 times higher than bacterial isolate diversity from the same samples and that farming and pollution played a significant role in reducing bacterial diversity.

2.6 Use of 16S rDNA sequences for parsimony and distance analysis.

Certain regions of rDNA sequences are highly conserved across all organisms whereas other regions may vary. The variability within these regions increases proportionately to the increase in the evolutionary distance between organisms, thus allowing for the determination of phylogenetic relationships between microorganisms (Nakatsu *et al*. 2000). Due to their usefulness as markers in phylogenetic studies, 16S rRNA genes have been the main target for prokaryotic ecological molecular surveys (Osborne *et al*. 2005).

Ribosomal RNA (rRNA) molecules are used as molecular chronometers because of their high degree of structural and functional conservation. As a result of this, domains within rRNA molecules harbour independent rates of sequence change. Phylogenetic relationships can be determined by examining these changes over time (Kent and Triplett 2002).

A large number of genes are available for phylogenetic studies on databases worldwide. Selected sequences should be appropriate, and can be affected by the following:

➢ Structural regions in the small and large subunit rRNA genes evolve at differing rates.

➢ Non-synonymous substitution rates at codon positions 1 and 2 are often slower than synonymous substitutions at position 3.

➢ Transitions occur more frequently than transversions.

Different substitution rates result in different levels of phylogenetic resolution in different areas of DNA. This should be taken into account when examining phylogenetic relationships at different taxonomic levels.

Patterns in sequence affect the suitability of data to be used in various phylogenetic tests:

➢ Phylogenetic signal: the level of conservation of sequence data.

➢ Saturation: multiple changes at the same site due to lineage splitting. Over time two sequences saturate due to multiple changes at certain sites. Increasing substitutions will have a diminishing effect on the sequences in question. A non-linear relationship develops between sequence divergence and time, leading to information loss to the phylogeny being examined.

➢ Base/codon composition.

At present 16S rDNA sequences form the ever-increasing, largest gene-specific data set available on internationally accessible databases (> 30 000), making tentative identification of unknown bacteria more possible (Von Wintzingerode *et al.* 2002). However, they are not always the most informative genes to select for study. Dauga (2002) investigated 16S and *gyrB* phylogenetic gene trees showing relatedness between *Enterobacteriaceae*. *gyrB* is a single-copy gene present in all bacteria. It has been proposed as a suitable genetic marker for identification of bacteria and encodes ATPase within the DNA-gyrase domain. Dauga (2002) found that *gyrB* trees proved to be more reliable determinants between closely-related species

than the 16S trees. 16S has nevertheless been used in the comparison and resolution of closely-related species. Anzai *et al*. (1997) found a 93.9% homology in 16S rRNA sequence homology between *Chryseomonas*, *Flavimonas* and *Pseudomonas* and on this basis proposed them synonymous. Similarly, Warwick *et al.* (1994) proposed that *Amycolata* and *Pseudonocardia* be classified in an emended *Pseudonocardia* genus on the basis of mixed clades emerging continuously from analysis of 16S data. *Pseudonocardia* has also, based on 16S sequence data, been observed to form a monophyletic unit with *Actinobispora* and it has been suggested that the latter genus be also incorporated into *Pseudonocardia* (Lee *et al*. 2000).

2.6.1   Characterisation of 16S region

The 16S gene of the bacterial genome holds the rDNA genetic code for the 16S subunit of the ribosome. Ribosomes are organelles in which translation of the genetic code (RNA to protein) takes place, and consist of two subunits of RNA and proteins (Tamarin 1996). Ribosome size is measured on the basis of its sedimentation rate during centrifugation in a sucrose density gradient. The unit of sedimentation is S, so designated after T. Svedberg, the developer of the method in the 1920s (Tamarin 1996). The 30S subunit of an *E. coli* ribosome comprises a 16S molecule of rRNA and 21 proteins (Tamarin 1996). This subunit of rRNA is encoded on the DNA of the bacterial cell and contains sequences that are highly conserved, thus allowing for sufficient resolution to distinguish between genera and species.

Advantages of using 16S rRNA gene sequences for analysis of microbial communities include the following:

  ➢ Essential component of ribosomes.

  ➢ Universal to all cell types.

- Universally conserved and variable taxon-specific sequences where the primary structure consists of conserved and variable sequences allowing for comparison of homologous positions of different species.

- Horizontal gene transfer not likely.

- Extensive databases (e.g. GenBank) of rRNA gene sequences exist.

- rRNA sequence-based "Tree of Life" provides a scaffold for comparison of unknown sequences from natural samples.

- Acts as a molecular chronometer.

- Allows for culture-independent analysis of unknown communities.

2.6.2   Characteristic base-pairs

There are two types of sequence data generated, viz. genomic DNA and expressed sequence tags (ESTs). Genomic DNA represents the genetic material of entire organisms in the form of genomes. The genomes are constructed from multiple experiments of high accuracy. However, ESTs are short pieces of DNA, usually 400-800bp, which are transcribed into mRNA and later translated into proteins. ESTs comprise 62% of the 38.9 million genetic sequences on GenBank, they are fairly easy to sequence and can be used to locate genes and their splice sites (Wu *et al*. 2005). Mapping of ESTs to known genomes has become more important in recent years for finding genes, EST clustering, alternative splice-sites and gene function. Wu *et al*. (2005) developed new computer software (EST mapper) which is 3-1000 times faster than current market software for aligning and clustering DNA sequences, and produces alignments of better quality.