Pure

Bangor University

DOCTOR OF PHILOSOPHY

Metagenomics approaches to discover new industrially-relevant enzymes

Distaso, Marco

Award date:
2019

Awarding institution:
Bangor University

Link to publication

Bangor University

DOCTOR OF PHILOSOPHY

Metagenomics approaches to discover new industrially-relevant enzymes

Distaso, Marco

Award date:
2019

Awarding institution:
Bangor University

Link to publication

Metagenomics approaches to discover new industrially-relevant enzymes

A thesis submitted to Bangor University in
candidature for the degree of Philosophiae Doctor

by
Marco Antonio Distaso
B.Sc., M.Sc.

School of Natural Sciences
Bangor University
Bangor, Gwynedd, LL57 2UW
United Kingdom

2018

Acknowledgments

This thesis is the result of an incredible journey that deeply changed me.

I would like to express my gratitude to my supervisor, Peter Golyshin, for the great opportunity he offered me, for the great supervision and the friendly and exciting work environment he created. My thanks to Dr Tran Hai, for all the help, advice and guidance throughout these years.

A big hug to Soshila Ramayah, a precious colleague and irreplaceable friend who was always ready to help and encourage me. Thanks for your support, your reassuring words, the spicy food and the moral and spiritual lessons.

I would like to thank my fellow colleagues Rafa, Samuel and Cristina for their valuable and fundamental support.

Thanks to my family, even from thousands of miles away your encouragement and unconditional love were constantly with me.

Finally, I would like to thank all the incredible people I have met during these years. Completely different from each other, from different countries, with very different backgrounds, together we lived many amazing adventures that I will never forget. I've learned so much from all of you, and without your support I would not be here now.

Table of Contents

Chapter IV

Towards the enzymatic production of the novel herbicide sorgoleone: identification and characterization of a novel SAM-O-methyltransferase from soil metagenome 124

Chapter V

Activity screening of sorghum rhizosphere metagenomic libraries and characterization of novel esterase enzymes **171**

Chapter VI

Abstract

Biocatalysis has become an indispensable part of the industrial production of organic compounds due to the superior functional properties of enzymes over chemical synthesis and due to the need of mitigating the environmental impact of chemical processes. The demand for new biocatalysts is constantly growing and besides the numerous characterised enzymes, only few candidates meet the requirements dictated by the industrial sector. Metagenomics has made possible the exploration of the incredible microbial biochemical diversity, significantly increasing the number of novel enzymes identified thanks to naïve screening and direct sequencing of environmental DNA. By applying a metagenomics-based analysis, we investigated the rhizosphere of Sorghum plant for the identification of novel biocatalysts of industrial interest. In detail, we aimed at finding a methylation reaction for the enzymatic production of sorgoleone, an allelopathic metabolite naturally released by sorghum roots that acts as a potent weed control compound, and other hydrolytic activities with enhanced properties of common use in biotechnological applications. After enrichment with Cashew Nut Shell Liquid (CNSL), a waste product from cashew industry rich in phenolic lipids resembling our target substrate/product, *de novo* sequencing of the total metagenomic DNA and fosmid expression libraries were produced. The changes in microbial community composition and its biodegradation potential was evaluated and resulted in a reconstruction of catabolic network, which suggested the significant enrichment in bacteria known for their aromatic hydrocarbons degradation capabilities in response to CNSL addition. Metagenome analysis suggested a number of enzyme candidates that were cloned and expressed. One of them was an efficient SAM-*O*-methyltransferase from uncultured microorganism distantly related to *Mycobacterium*, able to react with the aromatic moiety of the target substrate molecule cardol triene. Preliminary experiments suggested this enzyme as a basis for the establishment and further improvement of the two-stage process for semisynthetic sorgoleone production. Furthermore, functional screening of fosmid libraries identified two highly promiscuous esterases, capable of hydrolysing a broader range of ester substrates than the most-active industrial prototypes. The work has a number of potential environmental and economic implications, namely, in moving towards the establishing of an environmentally-friendly (greener) process for production of a natural herbicide for efficient weed control. Furthermore, a notable contribution was delivered to the functional annotation of poorly described enzyme families in databases, as well as to the definition of structural determinants of the enzyme promiscuity phenomenon, feature highly valued in industrial enzymes.

List of Figures

Chapter I

Chapter II

Chapter III

Chapter V

List of Tables

Chapter V

Abbreviations

µg	Microgram
µl	Microliter
µM	Micromolar
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CDD	Conserved Domain Database
CMC	Carboxymethyl cellulose
CNSL	Cashew Nut Shell Liquid
COG	Cluster of Orthologous Group
CYP450	Cytochrome P450
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
DOE	Department of Energy
DTT	Dithiothreitol
eDNA	Environmental Deoxyribonucleic Acid
EDTA	Ethylenediamine tetra-acetic acid
EST	Esterase
g	Gram
GC-MS	Gas chromatography–mass spectrometry
HPLC	High-performance liquid chromatography
IMG/M	Integrated Microbial Genome and Microbiomes
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JGI	Joint Genome Institute
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	Litre
LB	Luria Bertani
LC-MS	Liquid chromatography–mass spectrometry
LIP	Lipase
M	Molar
mg	Milligram
ml	Millilitre
mM	Millimolar
MS	Murashige and Skoog medium
ND	Not determined
nm	Nanometer
NMR	Nuclear magnetic resonance

OD$_{600}$	Optical Density at 600 nm
OMT	*O*-methyltransferase
ORF	Open Reading Frames
PCR	Polymerase Chain Reaction
PDB	Protein Database
PMSF	phenylmethylsulfonyl fluoride
pNP	para-Nitrophenol
R$_f$	Retention factor
SAH	S-Adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SDS-PAGE	Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
TCA	*Trichloroacetic acid*
tCNSL	Technical Cashew Nut Shell Liquid
TLC	Thin Layer Chromatography
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside

CHAPTER I

Introduction

1.1. Soil microorganisms and their interactions with plants

Soil is considered the most complex and diverse habitat on Earth, holding the largest reservoir of microbial diversity. Every gram of soil harbours thousands of microbial species (Schloss & Handelsman, 2006), which are actively involved in several biological processes including organic matter decomposition and cycling of all important elements (Nannipieri et al., 2017). Currently still little is known about the soil microbial taxonomic and functional diversity, due to the impossibility to cultivate the majority of the microorganisms. Over the past decades, thanks to the advent of new culture-independent approaches, i.e., Next-Generation Sequencing (NGS) platforms and metagenomics, the scientific community is successfully filling the void in the knowledge of microbial communities' compositional and functional variabilities in different ecosystems and conditions.

Soil is a very heterogeneous and discontinuous system where microorganisms are irregularly distributed and live in discrete microhabitats, some of which have favourable growth conditions (Stotzky, 1997). The richness and evenness of the microbial community in soil is strongly affected by several environmental factors. Soil physical, chemical and biological properties, such as pH (Fierer & Jackson, 2006), soil type (Schutter et al., 2001), water content (Bossio & Scow, 1998), pressure, and mineral nutrients can affect the ecology and activity of microorganisms. Modifications in the content of bioavailable organic carbon influence the diversity of soil communities (Goldfarb et al., 2011). This includes agricultural management, with pesticide and fertilizer application (Jangid et al., 2008) and various organic amendments (de Brito et al., 1995), which significantly modify population dynamics of soil bacteria. In addition, plants are known to exercise a strong influence on soil microbial communities through root exudation and the input of lignocellulosic polymeric compounds (Nannipieri et al., 2008).

Plant roots, besides the primary functions of anchorage of the plant to the substratum and of uptake and transport of minerals and water, are involved in other complex functions, influencing biological and chemical parameters of the immediate soil surrounding, or the so-called rhizosphere (Flores et al., 1999). The rooting zone is a very nutrient-rich environment and living organisms have evolved particular systems in order to compete and proliferate. The interactions happening in the rhizosphere can be positive or negative to the plant. Positive interactions include the association of epiphytes, mycorrhizal fungi and nitrogen-fixing bacteria, the production of protective biofilms, antibiotics and iron-chelating siderophores that prevent the growth of phytopathogens, and the

degradation of soil contaminants (Li et al., 2012; Mendes et al., 2013); negative interactions include contacts with parasitic plants, pathogenic bacteria, fungi, and insects (Bais et al., 2001). The soil-plant interaction emerges as a result of roots' ability to synthesise and secrete low and high molecular weight compounds into the soil (Walker, 2003). The scale of plant exudation is huge, up to 50% of gross produced carbon is excreted by the roots; root exudates change the chemical and physical properties of the soil and provide nutrients and chemotactic molecules, provoking a variety of different responses that affect neighbouring plants, insects, and microbes (Nardi et al., 2000; Walker, 2003).

1.2. Root exudate and root-rhizosphere communication

Roots perform active exudation secreting into the soil a large variety of low and high molecular weight compounds. The impact of this process varies with the age and physiological state of the plant, type of soil and nutrient availability. Polysaccharides and mucilage are an example of high molecular weight compounds secreted by root cap cells with lubrication functions (Walker et al., 2003). Low molecular weight constituents such as organic acids, sugars, amino acids, phenolic compounds and other secondary metabolites account for the majority of the secretion (Bertin et al., 2003). Although root exudation represents an important carbon cost for the plant (Marschne, 2013), the benefits for the plant and the ecosystem are numerous; with a wide range of mechanisms, the exudate directly and indirectly targets other plant species, soil microorganisms and rhizosphere nematodes.

For example, root exudate can consist of metabolites that act as antimicrobials, phytotoxins, nematicidal and insecticidal compounds (Weir et al., 2004). Roots can produce chemical signals to increase the resistance to aphids in neighbouring plants, dissuading the proliferation of aphids or, indirectly, attracting predators and parasites of the offending herbivore (Chamberlain et al., 2001). Acting as metal chelators, root exudates can alter soil chemistry and modify the soil nutrient availability increasing metal solubility and mobility (Dakora & Phillips, 2002). The deposition of secondary metabolites into the rhizosphere can attract or inhibit the growth of specific microorganisms, influencing the microbial community structure. Chemical attraction (chemotaxis) is a familiar method of cross talk between plant roots and microbes. In response to flavonoids produced by legume roots, transcription of rhizobial nodulation (nod) genes is induced and specific *Proteobacteria* develop a symbiotic association with leguminous plants (Scheidemann & Wetzel, 1997). The result of the infection of root nodules by the soil bacteria is the fixation of atmospheric nitrogen into a plant usable form (ammonium); in return, the plant supplies the bacteria with

carbohydrates, proteins, and oxygen (Moulin et al., 2001). Microorganisms referred to as plant growth promoting rhizobacteria (PGPR) (Gray et al., 2005) are able to produce phytostimulators, enhancing plant maturation; others produce protective biofilms, antibiotics or antifungal metabolites with resulting advantages for the plant species (Bais et al., 2004). A negative form of signalling is mediated by the release of potent phytotoxins that reduce the establishment, growth and survival of susceptible neighbouring plants in a process referred to as allelopathy (Rice, 1984).

1.3. Allelopathy

The International Allelopathy Society defined allelopathy as follows: "Any process involving secondary metabolites produced by plants, micro-organisms, viruses, and fungi that influence the growth and development of agricultural and biological systems (excluding animals), including positive and negative effects" (Torres et al., 1996). Allelopathy is a phenomenon that occurs constantly in natural ecosystems; it involves the production and release of organic compounds, commonly referred to as allelochemicals, by plants into the environment, with consequent direct and indirect effects to another plants (Fig. 1.1). Root exudate, highly rich in secondary metabolites and organic compounds, can be considered the largest source of allelochemicals input into the soil environment. Allelochemicals mediate positive and negative effects in terms of crop establishment and performance (Weston & Duke, 2003), usually conferring to the plant that release them an advantage over the neighbouring crop and weed species (Rice, 1984). The most frequently reported morphological effects on plants are the inhibited or retarded seed germination, the negative impact on coleoptile elongation and on the development of radicles, shoots and roots. The allelopathic effects of these toxic compounds does not exclusively depend on their concentration in the soil compartment. Some allelochemicals are released in the soil as already active compounds, others are influenced in their accumulation and phytotoxic effect by abiotic and biotic environmental conditions (e.g. pH, temperature, light, oxygen, microorganisms) (Inderjit & Weiner, 2001).
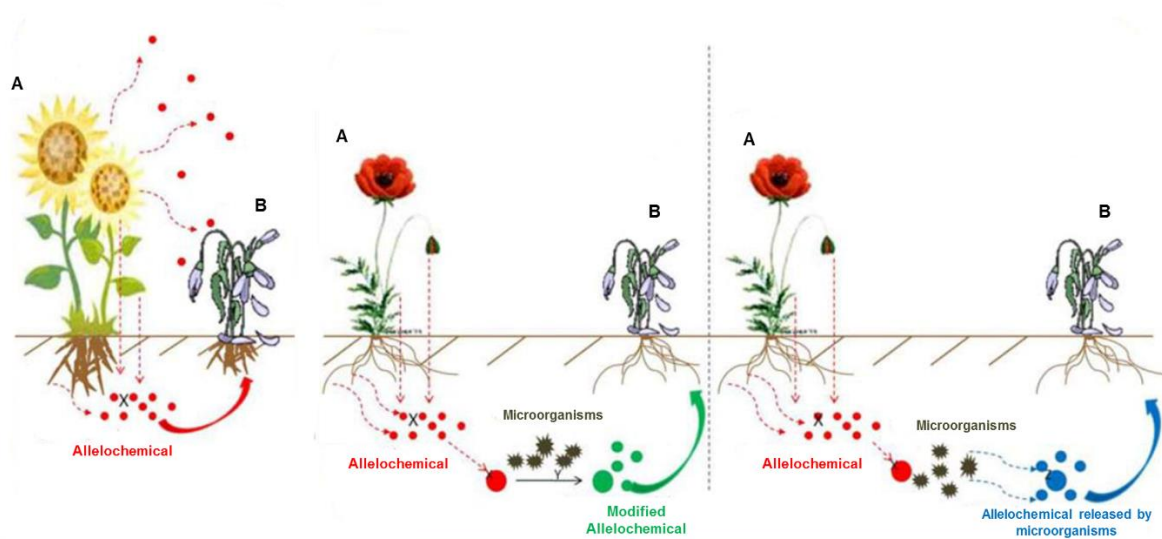
Fig. 1.1 Multi-dimensional nature of allelopathic interactions. Plant A releases allelochemicals which directly affect growth of plant B. Plant A releases allelochemicals which are modified and activated by microorganisms. Plant A releases allelochemical which stimulates microorganisms to produce new allelochemicals that affects growth of plant B (Soltys et al., 2013).

Phytotoxins can be produced or released by different plant organs including roots, rhizomes, leaves, stems, pollen, seeds and flower. Allelochemicals differ remarkably in chemical structure, modes of action and effects on plants. They are products of the secondary metabolism and belong to several chemical groups: terpens, benzoquinones, flavonoids, terpenoids, phenolic acids, fatty acids and non-protein aminoacids. Numerous crops have been investigated for allelopathic activity towards weeds or other crops species. A suppressive effect on weed has been reported for a wide range of temperate and tropic crops. These comprise alfalfa (*Medicago sativa*), barley (*Hordeum vulgare*), clovers (*Trifolium* spp., *Melilotus* spp.), oats (*Avena sativa*), pearl millet (*Pennisetum glaucum*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum spp*.), sunflower (*Helianthus annuus*), sweet potato (*Ipomoea batatas*), and wheat (*Triticum aestivum*) (Dilday et al., 2001; Rawat et al., 2017; Miller, 1996; Weston, 1996). Rye is one of the most important allelopathic crop, used to suppress weeds as a rotational and cover crop. The main allelochemicals released by rye (also found in maize and wheat) are the glucoside DIMBOA (2,4-Dihydroxy-7-methoxy-1,4-benzoxazin-3-one) and DIBOA (2,4-dihydroxy-1,4(2H)-benzoxazin-3-one), which are degraded spontaneously to the corresponding benzoxazolinones MBOA (6-methoxybenzoxazolin-2-one) and BOA (Wu et al., 2001a). Many other compounds were also identified such as p-coumaric acid, p-hydroxybenzoic acid, vanillic acid, etc. (Wu et al., 2001b; Schulz et al., 2013). Sunflower is also considered an allelopathic crop, which releases 16 different phenolic compounds. Eight sunflower cultivars were screened for their weed suppressing properties and up to 80% of weed biomass reduction was observed (Alsaadawi et al., 2012). Several

other allelopathic compounds have been identified: the respiratory inhibitor napthoquinone juglone (*Juglans nigra*, black walnut) (Jose & Gillespie, 1998a, 1998b), the sesquiterpenoid lactone artemisinin (*Artemisia annua*, wormwood) (Duke et al., 1987), numerous phenolic acids and momilactone A and B (*Oryza sativa*, rice) (Kato-Noguchi & Peters, 2013), the amino acid derivative L-DOPA (*Mucuna pruriens*, velvet bean) (Fujii et al., 1992) and the benzoquinone sorgoleone (*Sorghum* spp.) (Nimbal et al., 1996b).

The general participation of soil microbes in the allelopathic phenomenon is well established, via direct or indirect effects (Inderjit, 2005). Microbes can be affected by allelopathic plants leading to indirect effects on competing plants. Mutualistic microbes, such as nodulating bacteria and mycorrhizal fungi, are negatively affected by allelochemicals produced by plants, which gain a competitive advantage over species highly dependent on these symbiotic microorganisms (Cipollini et al., 2012). Microbes can influence the toxicity and quantitative availability of allelochemicals. It has been reported how the soil microbial community had adapted to using these compounds as an energy source, enhancing their degradation (Kong et al., 2008). Undoubtedly, allelochemical recovery is usually greater in sterile soils than non-sterile soils (Inderjit et al., 2010); nevertheless, estimating the half-life of different allelopathic compounds in soils is not straightforward due to the high dependence on soil characteristics, microbial composition, concentration of allelochemical and presence of other easier-to-digest organic compounds (Gimsing et al., 2009; Kong et al., 2008). On the other hand, degradation of allelochemicals by soil microbes sometimes can lead to the formation of more toxic compounds. Benzoxazolinones MBOA and BOA released by many cereals can be further degraded to 2-amino-(3H)-phenoxazin-3-one (APO) and other APO-related metabolites with increased phytotoxicity (Gents et al., 2005). In another example, the innocuous substance gallotannin, exuded by the common invasive reed *Phragmites australis,* is hydrolysed by microbial tannase enzymes to the toxic phenolic gallic acid that restricts the growth of native plants (Bains et al., 2009).

1.3.1. Allelopathy for weed management

The presence of weeds negatively affects the physiological activity and growth of surrounding crops, depriving them of nutrients, space and light. A reduction in yield caused by weeds ranging from 23% to 40% for some important crops was estimated (Oerke, 2006). Therefore, weed management appears as an imperative and challenging issue to address. Currently, the most used methods involve the application of synthetic herbicides and the mechanical removal of the infesting weeds. However, nowadays the chemical approach is no longer sustainable in terms of costs, environmental, human

and animal health. Additionally, the development of herbicide resistance in weed populations has become a critical phenomenon; there are currently more than 250 species of herbicide-resistant weed that have evolved resistance to 163 different herbicides (International Survey of Herbicide Resistance Weeds, 2018).

In this scenario, the exploitation of allelopathy may represent a powerful green strategy for weed management. Historically, crops with allelopathic properties have been exploited for centuries in the agronomic sector; crop rotation in field and the use of crop residues as green manure are some common examples. Research studies have allowed new and more sophisticated ways to exploit the full potential of allelopathic crops in agriculture (Cheng & Cheng, 2015).

The transmission of allelopathic traits between species is an interesting strategy to enhance weed-suppressing activity of crops. Different approaches were undertaken, such as traditional breeding methods (Kim & Shin, 2003) and genetic engineering (Duke et al., 2001). In order to improve allelopathy through breeding, researchers must first identify variations in the allelopathic potential of available germplasm. An extensive effort in the identification and characterization of competitive morphological traits and genetic markers associated with allelopathic properties is therefore necessary. As reported by surveys of crop cultivars, variation in weed suppressive ability has been observed in crop species and among genotypes and cultivars of the same species. The assessment of seedlings of 453 wheat accessions showed different levels of inhibition of root growth of ryegrass, ranging from 9.7 to 90.9% (Wu et al., 2000). Twenty-seven rice lines were tested against barnyard-grass and weeds from the family *Alismataceae* (Seal & Pratley, 2010), showing significant differences in their suppressive ability and weed specificity. The cultivar Amaroo was the most efficient, with 97% inhibition of root elongation. Differences in allelopathic potential have been studied also among accessions of barley, cucumber, oats, soybean, sunflower, and sorghum (Nimbal et al., 1996b). Important results were achieved in the identification of genetic markers and mapping of allelopathic genes (Wu et al., 2003, 2008) using RFLP and AFLP strategies. Two major quantitative trait loci (QTL) associated with allelopathy were found on wheat chromosome 2B, with one of them responsible for 29% of total phenotypic variance of wheat seedling allelopathy. QTLs controlling allelopathy in rice were also mapped (Jensen et al., 2001). An improved allelopathic rice cultivar was obtained by breeding efforts and it was released for commercial cultivation in China (Kong et al., 2011).

An interesting and prospective research area is the development of new classes of herbicides based on the chemistry of allelopathic molecules. The advantages of using compounds of biological origin are numerous. Firstly, allelochemicals are fully or partially water-soluble, which make them easier to apply, biodegradable and less soil persistent (Dayan et al., 2009a). Moreover, these molecules act at a different molecular level than synthetic herbicides, binding to new target sites and

offering the opportunity to overcome weed resistance to already commercialized compounds (Nimbal et al., 1996a; Dayan et al., 2009b). Defining the mechanisms of action and the biosynthetic pathway of these plant-synthesised secondary metabolites is therefore pivotal for their chemical or enzymatic production. Organic herbicides and plant-growth inhibitors have been manufactured from allelopathic plant materials to inhibit weed growth in fields (Ogata et al., 2008; Guillon, 2003; Selga & Kiely, 1997; Soltys et al., 2013). Cinmethylin is a monocot weed control agent developed in the early 1980s as a derivative of the natural monoterpene 1,8-cineole, component of eucalyptus oil (Dayan et al., 2003b), but its mode of action is still unknown. Sarmentine is an amide with broad-spectrum herbicide activity isolated from *Piper* species; wild mustard and pigweed were the most sensitive to sarmentine. Its mode of action involves the destabilization of plasma membranes and the interference of the photosystem II (Dayan et al., 2015). Sarmentine was patented but not commercialised yet (Huang & Asolkar, 2015).

The allelopathic potential of sorghum has been studied for years and extensively described. The p-benzoquinone sorgoleone is considered the main compound responsible for the toxicity towards various weed species. Sorgoleone proved to be more effective in inhibiting weed growth after formulation as a wettable powder (Uddin et al., 2014). The strong weed-suppressive ability of formulated sorgoleone raised interest as an effective, natural, environmentally friendly approach for weed management

1.4. Sorgoleone discovery and herbicidal activity

Sorghum (*Sorghum bicolor* L. Moench) is an important cereal grain crop, originated from Ethiopia and preferentially cultivated in Africa, United States, Australia, India and South America (Weston, 1996). The *Sorghum* genus includes approximately 25 species with a wide spectrum of applications, especially used as green manure or cover crop to suppress weed population.

The allelopathic effect of sorghum has been used in agriculture for decades and several studies documented its weed suppression effect. The foliar spray of aqueous extracts of sorghum plants (sorgaab) significantly increased the yield of tested crop including wheat (Cheema & Khaliq, 2000), maize (Ahmad et al., 2000) and mungbean (Cheema et al., 2001), due to increased weed suppression. The inhibitory effect was greatly enhanced in combination with sunflower and eucalyptus water extracts (Cheema et al., 2002). Sorghum used as cover crop reduced weed biomass up to 90%, with signs of growth inhibition of the subsequent alfalfa crop (Forney et al., 1985; Weston, 1996). Different studies reported the production and release of many primary phenolic acids, ubiquitous in

plants, that have phytotoxic activity, in concentrations that depend on the cultivar and plant age. Molecules like p-hydroxybenzoic, gallic acid, syringic acid, protocatechuic acid, vanillic acid, benzoic acid, p-coumaric acid have been identified (Cheema et al., 2007). The JS-263 cultivar had the highest content of total isolated phenolic compounds (904.6 µg g^{-1}), while extracts of the cultivar Sibbi contained the lowest phenolic content (51.35 µg g^{-1}). Nevertheless, these findings do not explain the exceptional inhibitory activity observed in sorghum (Alsaadawi et al., 1986; Panasiuk et al., 1986). The analysis conducted on 25 different sorghum cultivars resulted in the observation of a common hydrophobic, pigmented yellow droplets at the apex of the root hairs, insoluble in aqueous media, but extractable in chloroform wash solution (Czarnota et al., 2001). Major constituents of the root exudate appeared to be the lipid benzoquinone sorgoleone (2-hydroxy-5-methoxy-3-[(8'Z,11'Z)-8',11',14'-pentadecatriene]-p-hydroquinone) along with its reduced hydroquinone form dihydrosorgoleone (Fig. 1.2). These two compounds accounted for 90% of the exudate (Netzly & Butler, 1986; Czarnota et al., 2003b; Erickson et al., 2001); the remainder of the exudate consisted of sorgoleone congeners varying in the degrees of saturation of the aliphatic chain (one to three double bonds) or in the substitutions in the aromatic ring (Kagan et al., 2003). The combination of all these molecules appeared to contribute to the allelopathic potential of sorghum.
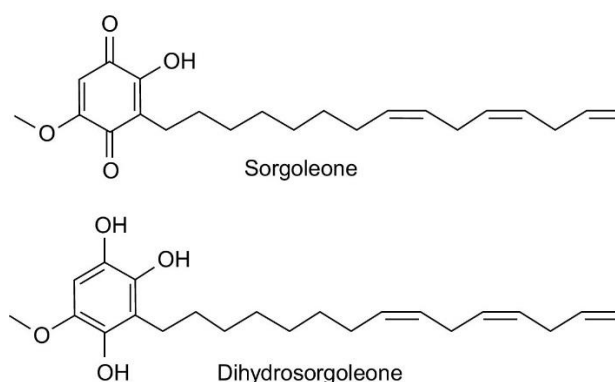
Fig. **1.2** Structures of sorgoleone and its reduced analogue dihydrosorgoleone.

Comparative analyses within different sorghum accession indicated considerable variability in the chemical composition of the exudate. The cultivation and the screening of more than 25 genetically diverse sorghum plants under identical environmental conditions produced dissimilar amounts of exudate, with differences in the purity and concentration of the sorgoleone secreted (Czarnota et al., 2003a, Nimbal et al., 1996b).

1.4.1. Activity of sorgoleone

Secondary plant products like sorgoleone may have diverse roles in the rhizosphere, effecting seedlings and mature plants, bacteria and microbes. The phytotoxic lipophilic compound sorgoleone has several interesting modes of action, interfering with different physiological processes on a wide range of plant species. Sorgoleone exhibited inhibitory activity as pre- and post-emergence herbicide in both greenhouse and field conditions (Uddin et al., 2010b). Initial bioassays that investigated the effects of sorgoleone on plant growth revealed a significant reduction of radical elongation of *Eragostis tef* by the application of 125 µM sorgoleone, and reduction in frond number and dry weight of *Lemna minor* by the application of 100 µM sorgoleone (Einhellig & Souza, 1992). Hydroponic culture bioassays showed a strong inhibition of radicle elongation and seedling growth in µmolar concentration of commonly occurring field weeds (broadleaf and grass weeds species), namely velvetleaf (*Abutilon theophrasti*), jimsonweed (*Datura stramonium*), redroot pigweed (*Amaranthus retroflexus*), ivyleaf morningglory (*Ipomea hederacea*), green foxtail (*Setaria viridis*), crabgrass (*Digitaria sanguinalis*), and barnyard grass (*Echinochloa crusgalli*) (Einhellig & Souza, 1992; Nimbal et al., 1996b). The inhibitory activity of sorgoleone appeared to be dose-dependent, with effects on leaf area and shoot/ root weights by concentration of sorgoleone as little as 10 µM. Differential tolerance to sorgoleone was observed, with strong inhibition of shoot and root growth in barnyardgrass, velvetleaf, and large crabgrass. Large crabgrass appeared the most sensitive species, showing 50% of shoot growth inhibition (Nimbal et al., 1996b). Soil impregnated with sorgoleone at concentration ranging from 10 to 80 ppmw had inhibitory effect on the growth of lettuce, common purslane, velvetleaf, sicklepod, and redroot pigweed (Czarnota et al., 2001). Growth reduction of 14-days-old seedlings was observed after foliar application of sorgoleone, with a significant inhibition of black nightshade, redroot pigweed, common ragweed, sicklepod and other grass species (Czarnota et al., 2001). Sensitive species showed signs of chlorosis, suggesting a possible inhibiting role of sorgoleone in chlorophyll biosynthesis.

The inhibitory activity of sorgoleone appears to be also time-dependent, greatly influenced by the age of leaf tissue. Younger seedlings are more sensitive to the action of sorgoleone. The exposure of leaf disc of velvetleaf seedlings to 100 µM sorgoleone resulted in a complete inhibition of ETR (electron transport rate) in 3-days-old and 4-days-old discs. Nevertheless, 50% reduction of ETR was observed in leaf discs of 7-days-old seedlings and no inhibition was measured on tissues 14-days-old or older. By contrast, inhibition with 33 µM atrazine was strong on leaves of all ages (Dayan et al., 2009b). This result was confirmed by the absence of radioactivity in the foliage of 3-weeks-old velvetleaf seedlings exposed to ^{14}C-ring labelled sorgoleone, suggesting that sorgoleone was not

translocated acropetally to the foliage by old seedlings. However, hypocotyl and cotyledonary tissues of velvetleaf emerging from the soil and treated directly with the allelochemical, showed clear signs of inhibition (Dayan et al., 2009b); this result confirmed that sorgoleone can be absorbed through the cuticle and epidermis, inhibiting the photosynthesis of young plants. Therefore, sorgoleone exerts its activity as a pre-emergence herbicide, affecting the tissues of developing seedlings that encounter sorghum root exudate in the soil.

The biological activity of sorgoleone proved to be comparable to synthetic herbicides such as diuron and atrazine, and the molecular target sites of sorgoleone have been located in the photosynthetic and mitochondrial electron transport (Einhellig et al., 1993; Nimbal et al., 1996a; Rimando et al., 1998), in the enzyme p-hydroxyphenylpyruvate dioxygenase (Meazza et al., 2002), and in the root H^+-ATPase pump (Hejl & Koster, 2004b). Sorgoleone is an effective inhibitor of the photosynthetic electron transport chain due to its structural similarity to the natural electron acceptors plastoquinone and ubiquinone. The site of action in the PSII complex was determined using isolated thylakoid membranes from spinach, comparing the competitive binding behaviour in the presence of atrazine, a potent inhibitor of photosynthesis, and other known PSII inhibitors (Nimbal et al., 1996a). Sorgoleone bound the Q_B site of D1 protein within the PSII complex, requiring half the amount of free energy of the plastoquinone as suggested by docking studies (Czarnota et al., 2001). Investigation in spinach thylakoids showed that the inhibition of sorgoleone was similar to the herbicide diuron, with I_{50} of 0.10 and 0.12 μM, respectively. Inhibition and competitive binding behaviour of diuron, atrazine, metribuzin and sorgoleone was also observed using potato and redroot pigweed thylakoids (Nimbal et al., 1996a). Interestingly, besides competing for the same site, atrazine-resistant plants showed reduction in growth in the presence of sorgoleone. In fact, atrazine is a Ser_{264} family PSII inhibitor whereas sorgoleone belongs to the His_{215} family; therefore, mutations resulting in atrazine resistance do not lead to cross resistance to sorgoleone (Dayan et al., 2009b). A decreased in the uptake of nutrient solution was observed in soybean seedlings treated with sorgoleone and linked to the decrease in the H^+-ATPase activity detected in root microsomal membranes (Hejl & Koster, 2004b). A similar mode of action was elucidated for juglone, an allelochemical exuded from the roots of black walnut (*Juglas nigra* L.) (Hejl & Koster, 2004a). Overall, the toxicity of sorgoleone may be the combination of diverse mechanisms, with the inhibition of photosynthesis and respiration in young seedlings, together with the phytotoxic activities on other molecular target sites in older plants. Herbicidal activity of sorgoleone was analysed using its formulated wettable powder form against a number of weed and crop species. Broadleaf weed species, in particular *Rumex japonicas* and *Plantago asiatic*a, showed greater susceptibility than grass weed species in both pre- and post-emergence applications (Uddin et al., 2014). In general, formulated sorgoleone exhibited 25-30%

higher inhibition compared to pure sorgoleone (Uddin et al., 2014); evidence that opens up interesting possibilities for the development of a new greener approach for weed management.

1.4.2. Release of sorgoleone into the rhizosphere

The biosynthesis of sorgoleone is linked to the presence of living root hairs (Yang et al., 2004a), mostly constitutive and proportional to the root biomass. Two sorghum accessions, the hybrid SX-17 (*Sorghum bicolor* > *Sorghum sudanese*) and johnsongrass (*Sorghum halepense*) were investigated to identify the intracellular location of root exudate production. Root hairs are cytoplasmic dense tubular extensions of the root epidermis that appear to be the metabolic sites for the production of sorgoleone (Duckett et al., 1994). These cells are rich in mitochondria and endoplasmic reticulum, signs of a high physiological involvement. Microscopic evidences validated the intracellular production location of sorgoleone. Transmission electron microscopy images showed lipophilic vesicles associated with smooth endoplasmic reticulum (SER) throughout the cytoplasm and an accumulation of osmiophilic globules between the cell wall and the plasma membrane (Fig. 1.4) (Czarnota et al., 2003a), as previously observed also in other secretory cells (Fahn, 2000). A cellular transport mechanism or the fusion of the vesicles with the cell membrane were hypothesised as mechanisms responsible for the deposit of the material between the cell wall and the plasmalemma (Czarnota et al., 2003a). From this space, the exudate then moves to the tip of the root hairs where SEM images and light microscopy showed golden brown droplets of root exudates (Fig. 1.3) (Czarnota et al., 2001, 2003a).
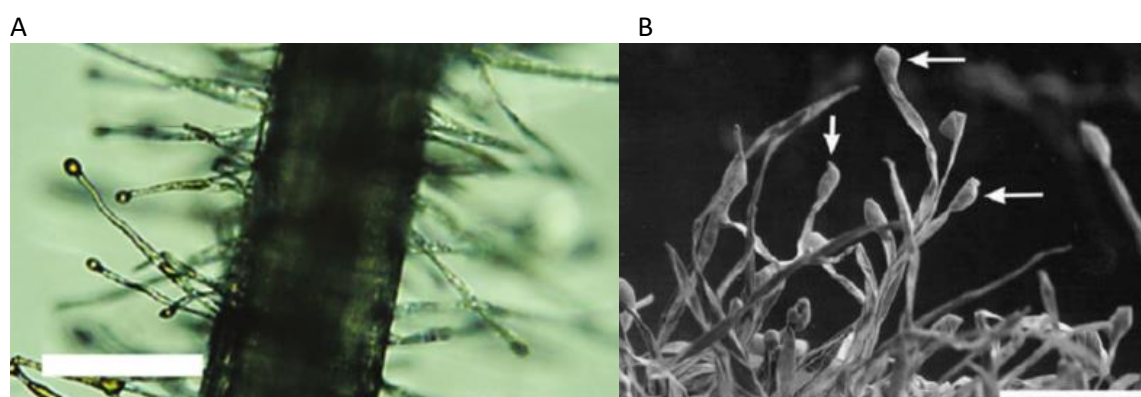
Fig. 1.3 A. SX-17 root hairs producing the golden brown droplets containing sorgoleone. Bar=50 μm. B. Cryo-SEM of johnsongrass showing the globules of root exudate on the tips of the root hairs (white arrows). Bar=100 μm (Czarnota et al., 2003a).
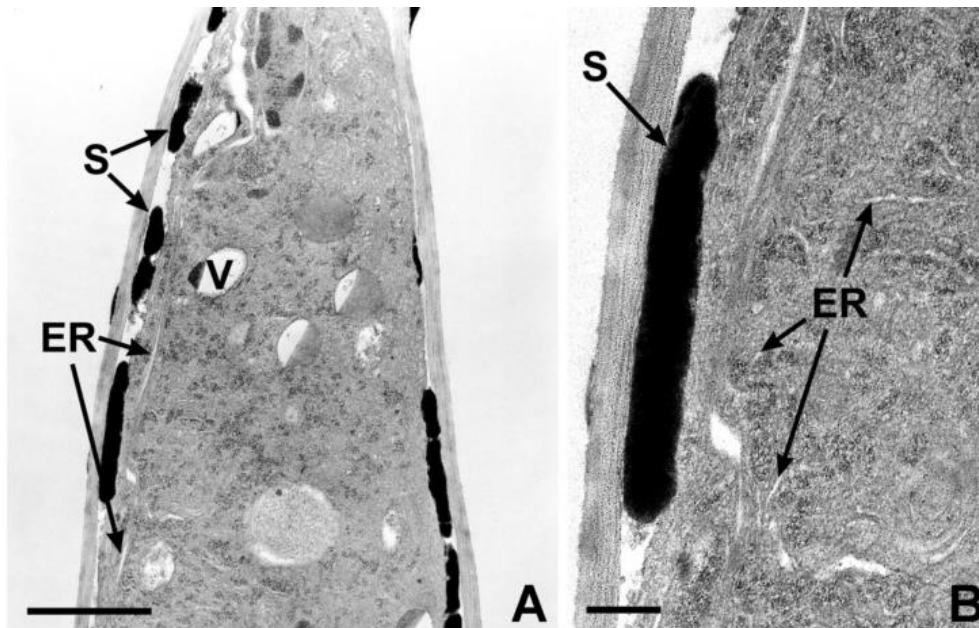
Fig. 1.4 A, B. Transmission electron micrograph showing longitudinal sections of *Sorghum bicolor x Sorghum sudanes*e cv. SX-17 root hairs. Root exudates, stained black, are deposited between the plasma membrane and cell wall. A, Bar=1 μm. B, Bar=0.2 μm. ER=endoplasmic reticulum, S=root exudates, V=vacuole (Czarnota et al., 2003a)

Production of exudate occurs shortly after germination from the apex of root hairs, right after these specialized cells have stopped elongating. As long as sorghum is actively growing and producing root hairs, sorgoleone is continually produced and secreted into the rhizosphere. Its production appears to be constant over time, reaching approximately 20 μg mg^{-1} dry weight of root (Dayan et al. 2009b). The constant amount of sorgoleone produced per root dry weight may be the result of a feedback inhibition mechanism that stops the exudation of the lipophilic compound once droplets accumulate at the root tip. The roots segments washed with water released the inhibition and restored the production of sorgoleone to typical levels within 7 days (Dayan et al. 2009b). The levels of sorgoleone exudation appeared to be affected by environmental conditions. The amount of sorgoleone exuded is sensitive to temperature, with an optimum of production at 30°C. Low temperatures inhibited root growth and sorgoleone synthesis, as well as temperatures above 35°C reduced the amount of sorgoleone produced of 95% (Dayan, 2006). Acidic pH conditions slightly stimulated sorgoleone production (Dayan, 2006). Light treatment did not affect root development but significantly repressed sorgoleone production, in particular under blue and red light. Interestingly, the treatment of sorghum seedlings with a water-soluble extract of velvetleaf generated an increase in the amount of sorgoleone produced in a dose-dependent manner (Dayan, 2006), not complemented with an increase in root biomass. A result that suggested a stimulating response to the presence of other plants in the release of the allelopathic compound.

1.4.3. Mineralization of sorgoleone and interaction with soil microbes

Sorgoleone is a hydrophobic compound with a log K_{ow} of 6.2 (Trezzi et al., 2006) and therefore is expected to strongly bind with the organic matter in soil. An investigation conducted in Danish and American soils (Gimsing et al., 2009) showed how mineralization of sorgoleone varied depending mostly on the type of soil used (decisive is the organic matter content and previous exposure to sorgoleone) and initial concentration of the allelochemical. Interestingly, all parts of sorgoleone molecule were susceptible to degradation but not to the same extent. The methoxy group was rapidly degraded after 10 days, whereas the quinone ring and the aliphatic chain showed slower degradation rate. American soils had the best degradation potential, with up to 40% of sorgoleone mineralized after 70 days, suggesting a better predisposition of the microbial community to using sorgoleone as carbon source.

1.4.4. Biosynthesis of sorgoleone

In order to determine the biosynthetic pathway of sorgoleone and identify its metabolic intermediates, retrobiosynthetic NMR studies of sorghum exudate collected from seedlings grown with isotopically labeled substrates were carried out. The first labeling study demonstrated that ^{13}C-acetate was incorporated into the aromatic ring of a dihydroquinone derivative of sorgoleone (Fate & Lynn, 1996). As postulated for structural similar compounds (Kozubek & Tyman, 1999), it was suggested that the sorgoleone pathway involves the convergence of two metabolic routes, the fatty acid pathway for the synthesis of the aliphatic tail and the polyketide synthase pathway for the constitution of the quinone head (Fig. 1.5). The biosynthetic pathway likely starts in the plastids with the generation of an unusual $\Delta^{9,12,15}$ C16:3-CoA fatty acid by fatty acid synthase and desaturases (Dayan et al., 2003a). The intermediate is then used as substrate by a specialized polyketide synthase (PKS) outside the plastids, which catalyses the condensation of 3 units of malonyl-CoA, cyclization and decarboxylation, yielding the resorcinolic lipid intermediate (5-pentadecatrienyl resorcinol). The next step is the methylation of the hydroxyl on carbon 3 of the resorcinol ring that was observed with ^{13}C-methyl-methionine through the putative action of S-adenosyl-L-methionine-dependent *O*-methyltransferase (OMT). The last step involved in the pathway is the dihydroxylation on carbons 2 and 4 catalysed by hydroxylases (likely one or two P450 monooxygenases) to produce the reduced form of sorgoleone (Dayan et al., 2003a, 2010).
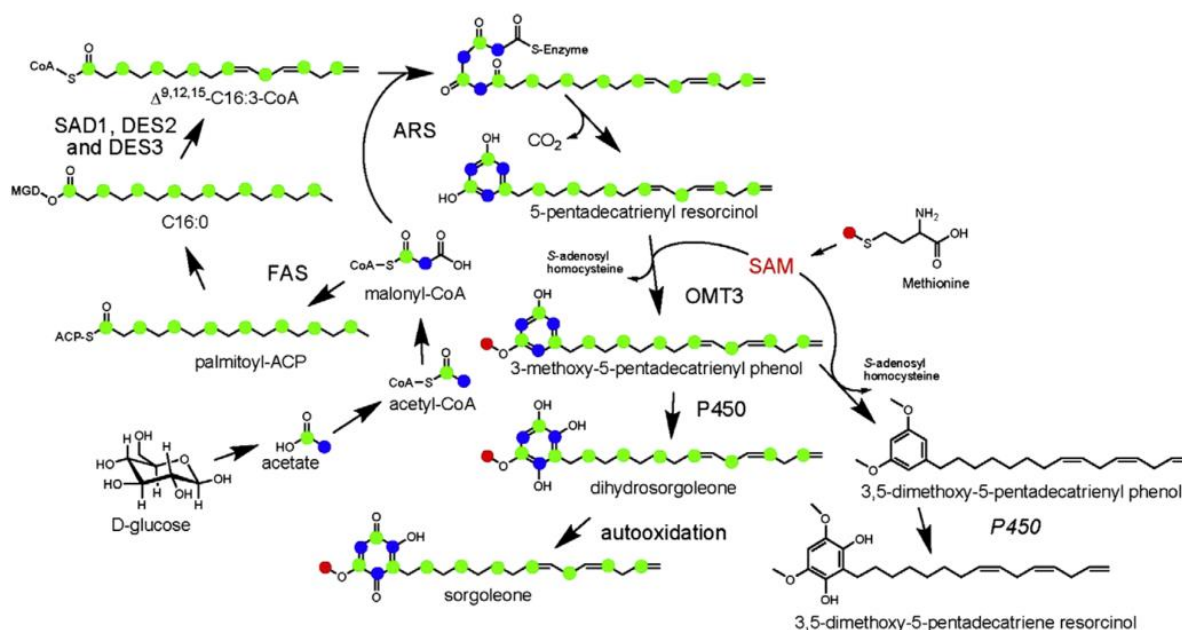
Fig. 1.5 Biosynthesis of sorgoleone showing the incorporation of ^{13}C-labeled substrates in the carbon backbone. 2-^{13}C-D-glucose (green) incorporated in carbon atoms 1, 3 and 5 of the quinone head, as well as carbon atoms 2',4',6',8',10',12', and 14' of the tail. 1-^{13}C-acetate (blue) incorporated in carbon atoms 1 and 5 of the quinone head but not in the lipid tail. ^{13}C-methyl-L-methionine (red) incorporated in the 3-methoxy group. FAS = fatty acid synthase, SAD1, DES2 and DES3 = fatty acid desaturases, ARS = acyl-resorcinol synthase (polyketide synthase), OMT3 = O-methyltransferase, P450 = P450 monooxygenase. (Dayan et al., 2010)

1.5. Genetic pathway of sorgoleone synthesis

The recognised herbicidal properties of sorgoleone attracted the attention of the scientific community that aimed to identify and characterise the genes and enzymes involved in the sorgoleone biosynthesis. The isolation of the genes is highly appealing for the agronomical sector, as this could provide new opportunities in terms of generation of transgenic crops with enhanced allelochemical characteristics, but also for the use of plant cells as bioreactors in the production of sorgoleone. Several studies were carried out and the putative biosynthetic pathway is now established (Fig. 1.6). The biosynthesis of sorgoleone appeared to require four types of enzymes: fatty acid desaturase (DES), type III polyketide synthase (PKS), S-adenosyl-L-methionine dependent *O*-methyltransferase (OMT) and cytochrome P450 monooxygenase.

Sorgoleone accumulation is restricted to root hairs (Czarnota et al., 2001, 2003a), therefore the enzymes involved in its production were considered to be predominantly expressed in these specialised cells. Thus, a root hair specific expressed sequence tag (EST) library was built and used for the investigation (Baerson et al., 2008a). The *S. bicolor* cultivar BTx623 was selected as a model system for identifying genes associated with sorgoleone biosynthesis since this cultuvar was previously used

to generate a sorghum genomics infrastructure and a completed sorghum genome sequence (Paterson et al., 2009). Total RNA was purified and used for the construction of the cDNA library from which the EST database was generated, resulting in approximately 5,500 sequences. The database was mined for the candidate enzymes performing BLAST searches using functionally characterized protein sequences (Baerson et al., 2008a; Cook et al., 2010; Pan et al., 2007, 2018). Few candidate genes with high expression levels in root hairs for all of the targeted enzyme families were shortlisted. The specific candidates were then functionally characterized by heterologous expression of the full-length open reading frames into *E. coli* or *Saccharomyces cerevisiae.*
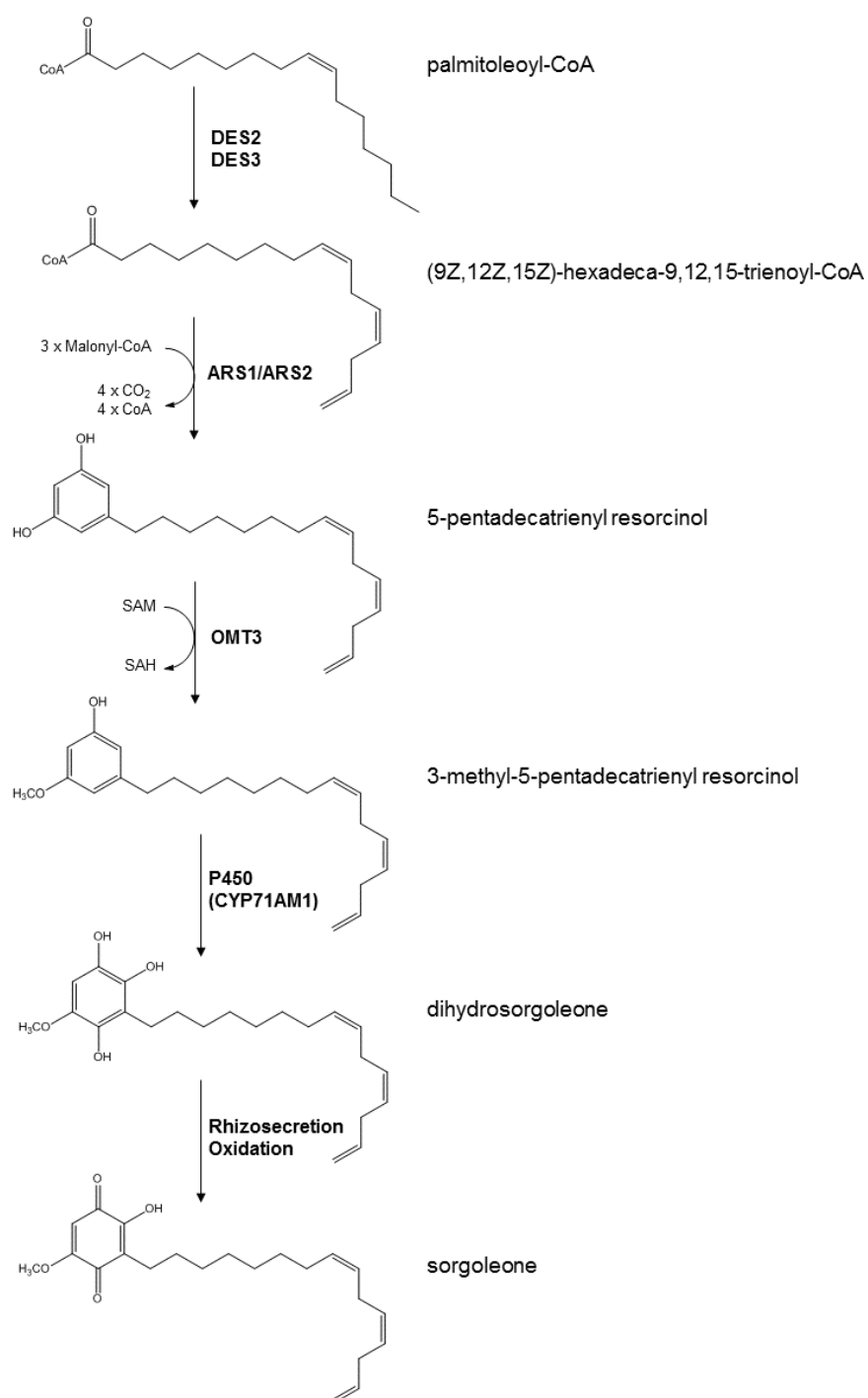
Fig. 1.6 Proposed biosynthetic pathway of the allelochemical sorgoleone in sorghum plants. Dihydrosorgoleone, the hydroquinone produced in vivo, is thought to undergo auto-oxidation once secreted into the soil to yield sorgoleone, a more stable benzoquinone. ARS, alkylresorcinol synthase; DES, fatty acid desaturase; OMT, O-methyltransferase; P450, cytochrome P450 (Baerson et al., 2008b).

1.5.1. Fatty acid desaturase from sorghum

The formation of $\Delta^{9,12}$ double bonds is quite common and ubiquitous in plants, and enzymes responsible for the synthesis of polyunsaturated fatty acids in plant have been characterised (Behrouzian & Buist, 2002; Uttaro, 2006). It is instead unusual the presence of a Δ^{15} terminal double bond in the aliphatic side chain, which is catalysed by a unique class of desaturases not yet described. The work conducted in two different studies allowed the identification of putative fatty-acid desaturases involved in the formation of the C16:3 desaturation pattern of the aliphatic tail of sorgoleone (Yang et al., 2004b; Pan et al., 2007) (Fig. 1.6). The GC-MS analysis of the fatty acids composition of various *S. bicolor* genotype BTX623 tissues showed the presence of 16:2 and 16:3 fatty acids in the lipid extracts of roots hairs, evidence that confirmed the primary role of these cells in the biosynthesis of the fatty acids precursors and sorgoleone. Two fatty acid desaturase-like sequences were identified within the EST dataset from isolated root hairs, encoding the enzymes SbDES2 and SbDES3 that revealed the ability to catalyse consecutive desaturation steps converting palmitoleic acid ($16:1\Delta^9$) to hexadecadienoic acid ($16:2\Delta^{9,12}$), and this to hexadecatrienoic acid ($16:3\ \Delta^{9,12,15}$) (Pan et al., 2007). Relative expression level determined by real-time RT-PCR with gene specific primers, showed a preferential expression of the two selected desaturases in root hair cells, displaying level of transcripts 1,000 times higher compared to other tissues (stem, immature leaf, mature leaf, panicle, root with hair removed). Based on sequence analysis, SbDES3 showed significant similarity to FAD3-type sequences, which generally catalyse the conversion of 18:2 to 18:3 in phospholipids (Browse et al., 1993). Multiple alignment analysis revealed 12 amino acid changes in SbDES3 sequence in position highly conserved, which were considered responsible for the gained ability of this enzyme of mediating the formation of the three terminal sequential double bonds in the side chain of sorgoleone. SbDES2 was instead more similar to plant FAD2-type desaturase, which typically acts on C_{16}/C_{18} acyl chains at the Δ^{12} position (Los & Murata, 1998). The co-expression in yeast cells of the two desaturases resulted in the production of the $16:3\Delta^{9,12,15}$ fatty acid with a terminal double bond, precursor of the aliphatic side chain of sorgoleone (Pan et al., 2007).

1.5.2. Alkylresorcinol synthase from sorghum

In the proposed pathway (Fig. 1.6), the polyunsaturated fatty acyl-CoA previously generated is used as starter unit for a polyketide synthase enzyme in order to produce the 5-pentadecatrienyl resorcinol intermediate. A specific subclass of type III polyketide synthases, referred to as

alkylresorcinol synthase (ARS), has been suggested to participate in this reaction (Dayan et al., 2003a). Type III polyketide synthases are homodimers that catalyse sequential condensation reactions, typically using malonyl-CoA as extender units, with differences in the number of condensation steps performed (Shimizu et al., 2017). Type III PKSs that preferentially use long-chain acyl-CoA starters are referred to as alkylresorcinol synthases which catalyse the formation of 5-n-alkylresorcinols (Funa et al., 2006, 2007). Analysis performed by GC-MS on the methanol extract prepared from sorghum tissues confirmed the significant accumulation of 5-pentadecatrienyl resorcinol in root hairs cells (Cook et al., 2010). The root hair ESTs dataset (Pan et al., 2007; Baerson et al., 2008a) was mined for candidate PKSs and five unique PKS-like sequences were identified. Two enzymes, ARS1 and ARS2, displayed root hair preferential expression as reported by real-time RT-PCR analysis, and amino acid sequence similarity to previously characterized plant type III PKSs (Austin & Noel, 2003). Both enzymes catalysed formation of 5-alkylresorcinols in activity assay performed using various fatty acyl-CoA starter units with differences in length and degree of saturation, and malonyl-CoA as extender unit (Cook et al., 2010). The preferred substrates were myristoyl-CoA (C14) and palmitoleoyl-CoA (C16:1Δ^9) for ARS1 and ARS2, respectively. Activity was registered also with the hexadecatrienoyl-CoA (C16:3$\Delta^{9,12,15}$), the physiological substrate proposed in the sorgoleone biosynthesis. Although the specificity for substrates such as palmitoyl-CoA (C16) and palmitoleoyl-CoA (C16:1) was higher, the activity of the recombinant enzymes in *in vitro* assays may not reflect the behaviour of ARS1 and ARS2 *in vivo*. An additional validation of the role of *ARS1* and *ARS2* in sorgoleone biosynthesis was obtained by the gene knockdown experiment using RNA interference in transgenic sorghum plants (Cook et al., 2010). Hairpin RNA-forming binary vectors were assembled for targeted downregulation of *ARS1* and *ARS2,* and six independently transformed *S. bicolor* events were performed and evaluated. Expression levels of the two genes were substantially reduced in hpRNA+ individuals, with a decrease in transcript accumulation from 66% to 96% of *ARS1* and from 55% to 86% of *ARS2*. A correlation was observed with respect to sorgoleone accumulation, whose levels, determined by GC-MS of root exudates prepared from 10-d-old hpRNA+ seedlings, were significantly reduced to undetectable levels.

Homology-based modelling was carried out to explain the observed substrate specificity of the enzymes for long-chain fatty acyl-CoA. Particular attention was given to the enzyme active sites which revealed differences when compared to the crystal structures of plant-specific PKSs templates, namely *M. sativa* chalcone synthase CHS2 (Ferrer et al., 1999) and *G. hybrida* 2-pyrone synthase 2-PS (Jez et al., 2000). The highly conserved Thr among plant type III PKSs (Thr-132 in *M. sativa* CHS2 and Thr-137 in *G. hybrida* 2-PS) was replaced by a Tyr in ARS1 and ARS2, with a consequent narrowing of the active site cavity. Furthermore, an Ala residue in the sorghum enzymes replaced the Met conserved in the majority of type III PKS enzymes (Met-137 in *M. sativa* CHS2, Met-142 in *G. hybrida*

2-PS), which represents the exclusive involvement of the second subunit to the active site cavity. The Ala residue in this position is rare among type III PKSs, but it is contained in bacterial enzymes that use long-chain fatty acyl-CoAs as substrates (Sankaranarayanan et al., 2004). Additional amino acid positions, known from site-directed mutagenesis studies to be relevant for defining the size of the active site cavity and the number of condensation reactions performed, appeared to contain substitutions in *S. bicolor* ARSs. In particular, it was observed an inverse relationship between residue's side chain bulk and number of condensation reactions performed. The most represented Thr-197 in CHS2 was replaced by Ala in sorghum enzymes, and Gly-256 in CHS2 and Leu-261 in 2-PS that appeared to influence the type of starter units accepted were substituted by a Met (Cook et al., 2010).

Collectively, the results obtained from ARS1/2-targeting RNAi experiments, from functional assays using recombinant ARS1 and ARS2, from tissue specific expression pattern, and structural studies strongly supported the involvement of the two alkylresorcinol synthases in the biosynthesis of sorgoleone.

1.5.3.   O-methyltransferase from sorghum

After cluster analysis, 12 *O*-methyltransferase-like sequences were identified in the root hair EST dataset. Quantitative real time RT-PCR conducted on the cDNAs prepared from total RNAs isolated from root hairs, root system and other tissues, revealed a preferential expression of three OMT-like sequences, referred to as *SbOMT1*, *SbOMT2* and *SbOMT3*, with the highest amount of transcripts found in root hairs (Baerson et al., 2008a). Sequence and structural analysis of the deduced amino acid sequences highlighted a significant sequence similarity with previously characterised plant OMTs, in particular in the residues and motifs associated with the catalysis, substrate orientation and cofactor binding (Zubieta et al., 2001). Functional characterization of the recombinant SbOMT1, SbOMT2 and SbOMT3 was performed using several benzene derivatives and a series of 5-*n*-alkylresorcinols, including the proposed sorgoleone intermediate 5-*n*-pentadecatrienyl resorcinol (Baerson et al., 2008a). SbOMT1 preferred substrate was eugenol, but significant activity was also recorded with the monomethyl ether orcinol and resorcinol. SbOMT1 was classified as a eugenol OMT (EOMT), functionally related to the previously characterized enzymes isolated from sweet basil (Gang et al., 2002) and *Clarkia breweri* (Wang et al., 1997). Recombinant SbOMT2 showed no activity against any of the substrates provided, probably due to a lack of functionality of the encoded protein. In contrast, SbOMT3 exhibited some degrees of promiscuity as activity occurred with all the substrates analysed. A preference was observed for alkylresorcinolic compounds containing from 4 to 7 carbons

in the alkyl side chain, with highest activity detected for 5-hexyl-resorcinol (Baerson et al., 2008a). Phylogenetic analysis using putative and functionally characterized plant-specific type I OMTs and the peculiar substrate utilization, categorized SbOMT3 into a novel subclass of type I OMT. The substrate specificity was studied from a structural point of view by homology modelling and molecular docking. The structure of SbOMT3 was modelled using the crystallized protein of plant type I isoflavone-7-OMT (IOMT) from *M.sativa* (Zubieta et al., 2001), which revealed the highest identity score (37%). SbOMT3 appeared to possess an N-terminal dimerization domain and a larger C-terminal domain containing the cofactor S-adenosylmethionine (SAM) and the substrate binding sites, as observed in other type I OMTs. The catalytic histidine His-279 as well as the binding motifs of the methyl donor SAM were extremely conserved, suggesting that SbOMT3 shares the same S_n2 catalytic mechanism for methyl transfer with previously characterised enzymes. The binding of the resorcinolic substrates occurred into a hydrophobic pocket able to accept the resorcinol ring and the aliphatic side chain. The orientation of the side chain of His-332 created a curved cavity particularly adapted to accommodate the 3D conformation of 5-pentadecatrienyl resorcinol (Baerson et al., 2008a).

This study, in addition to the *in silico* preference for alkylresorcinolic compounds structurally related to the physiological substrate, argued in favour of the potential role for SbOMT3 in the sorgoleone biosynthetic pathway.

1.5.4. Cytochrome P450 monooxygenases from sorghum

The final step of the sorgoleone biosynthetic pathway considers the dihydroxylation of the 3-methyl-5-pentadecatrienyl resorcinol intermediate mediated by cytochrome P450s enzymes to produce dihydrosorgoleone, the reduced form of sorgoleone that undergoes oxidation upon rhizosecretion (Dayan et al., 2003a). Recently, the transcripts of the cytochrome P450 *CYP71AM1* were found to accumulate specifically in root hairs cells, and therefore the enzyme was considered to be involved in the sorgoleone biosynthesis (Pan et al., 2018). CYP71AM1 exhibited extensive similarity to functionally characterized plant P450 enzymes and appeared to belong to plant-specific CYP71 family. Typical conserved elements were observed in the protein such as the O_2 binding site, the proline-rich region, and the E-R-R triade located upstream of the heme-binding cysteine motif (Bak et al., 2011). *In vivo* enzymatic assay carried out in *S. cerevisiae* expression system indicated dihydroxylation activity of CYP71AM1. Two hydroxyl group were added to the ortho positions of the resorcinol ring of the provided substrate 3-methyl-5-pentadecatrienyl resorcinol; GC-MS analysis confirmed that the peak with retention time of 14.16 minutes and its ion fragmentation were

comparable with the chromatogram and the mass spectrum of the standard dihydrosorgoleone (Pan & Baerson, 2014; Pan et al., 2018). Whether the dihydroxylation occurs in a sequential manner or not, it has not been elucidated yet. In addition, molecular docking model with the proposed substrate indicated the accommodation of the molecule in the hydrophobic substrate-binding pocket and a possible interaction with the acidic residue Asp-311. Downregulation of *CYP71AM1* via RNAi successfully reduced transcripts accumulation in *S. bicolor* transformants from 61 to 90%, and a consistent reduction in sorgoleone amount was recorded (Pan et al., 2018).

1.6. Chemical synthetic route for sorgoleone production

As mention before, sorgoleone biosynthesis is strongly dependent on root hair formation. Quantification of sorgoleone production is however difficult since it is strongly affected by chemical, biological, genetic and environmental conditions (Dayan, 2006). The exudation appeared to be a continuous process, as long as the droplets accumulating on the tip of root hairs are washed off, with an estimated concentration of 20-25 µg of exudate/mg root dry weight for sorghum cultivar SX17 (Dayan et al., 2009b). The investigation of different growth conditions successful enhanced sorgoleone production in grain sorghum roots (cultivar Chalsusu). The combined treatment of sorghum seeds with auxins and Hoagland solution positively stimulated root growth and sorgoleone production. In particular, the application of 5 µg/ml of indole-3-butyric acid and half-strength Hoagland solution increased the production to a concentration of 84.6 µg/mg RDW of sorgoleone (Uddin et al., 2010a). The sorgoleone levels, although susceptible to modulation, are relatively low when looked in the optic of a large-scale production of the molecule as bioherbicide. Therefore, solvent extraction of the exudate from sorghum root hairs seems to be unfeasible for the large-scale agricultural weed control.

Total chemical synthesis of sorgoleone was achieved the first time in 1990, with a process comprising 17 steps (Sargent et al., 1990). Few studies have been conducted on the synthesis of sorgoleone analogues with potential use as herbicides. Using 3,5-dimethozybenzylic alcohol as starting material, 2-acetoxy-5-methoxy-3-(pent-1-yl)-1,4-benzoquinone was chemically synthesised in seven steps. The synthesised acetoxyquinone showed inhibiting effect on the growth of tested plants, comparable to sorgoleone to some extent. However, the overall yield of the process was only 14.6% (Barbosa et al., 2001). Other aryl-substituted p-benzoquinones were also produced in yield ranging from 46 to 97%. All the synthesised compounds caused inhibition of the tested plants (*Cucumis sativus, Sorghum bicolor, Euphorbia heterophylla, Ipomoea grandifolia*) to some extent. The most active compound was 2-(4-formyl-2,5-dimethoxyphenyl)-5-[(tridecyloxy)methyl]-benzo-1,4-quinone

that proved to have a herbicidal activity superior than sorgoleone, displaying inhibitory effect on *Sorghum bicolor* itself (de Almeida Barbosa et al., 2006). However, very little is known about the effects of the above-mentioned compounds.

A considerable step forward was achieved in 2012 with the description of a 5-step chemical route for the preparation of sorgoleone (Tverezovskiy et al., 2012). The starting material was cardol, a component of the cashew nut shell liquid (CNSL), the main by-product of the cashew industry. The method considered the use of the cardol mixture extracted from CNSL, which comprised cardol congeners with different degree of saturation of the alkyl side chain. Nevertheless, the harsh conditions and the toxicity of the reagents represented a serious constraint for the application of the above synthetic methods and, as a general consideration, of every chemical route. Moreover, the state of the art for chemical production of sorgoleone appeared to be unappealing in terms of yield, costs, time and waste production.

1.7. Novel enzymatic route for sorgoleone production

The adoption of biocatalysts is the primary strategy for the set-up of green and effective processes. In fact, enzymes of microbial origin are extensively used in various application areas, comprising food, detergents and textile processing, synthesis of pharmaceutical and agrochemical molecules, pesticides and insecticides (Demain & Adrio, 2008; Gupta et al., 2002; Monsan & O'Donohue, 2010; Davis & Boyer, 2001; Koeller & Wong, 2001). As part of this PhD work, the enzymatic synthesis of sorgoleone has been proposed with the intent of replacing the existing chemical process (Fig. 1.7). The putative starting material is the above-mentioned cardol triene, constituent of the widely available natural feedstock CNSL. The route comprises two steps: the methylation of the 3' hydroxyl group of cardol triene by a microbial S-adenosylmethionine dependent *O*-methyltransferase, followed by its dihydroxylation performed by cytochrome P-450 monooxygenase, to produce dihydrosorgoleone. The benefits of the designed process are economic and environmental. The transformation of an abundant low-value chemical into a high-value compound is possibly the ambition of every biotransformation process. The use of enzymes in biotransformation represents an evident greener alternative to toxic chemicals. Furthermore, microbial enzymes are relatively cheap to produce and they offer a number of advantages in terms of product titre, energy demand, substrate specificity, and fine-tuning possibilities. A number of biocatalytic routes have been already developed that revealed great advantages over the traditional chemical pathways (Bornscheuer et al., 2012).
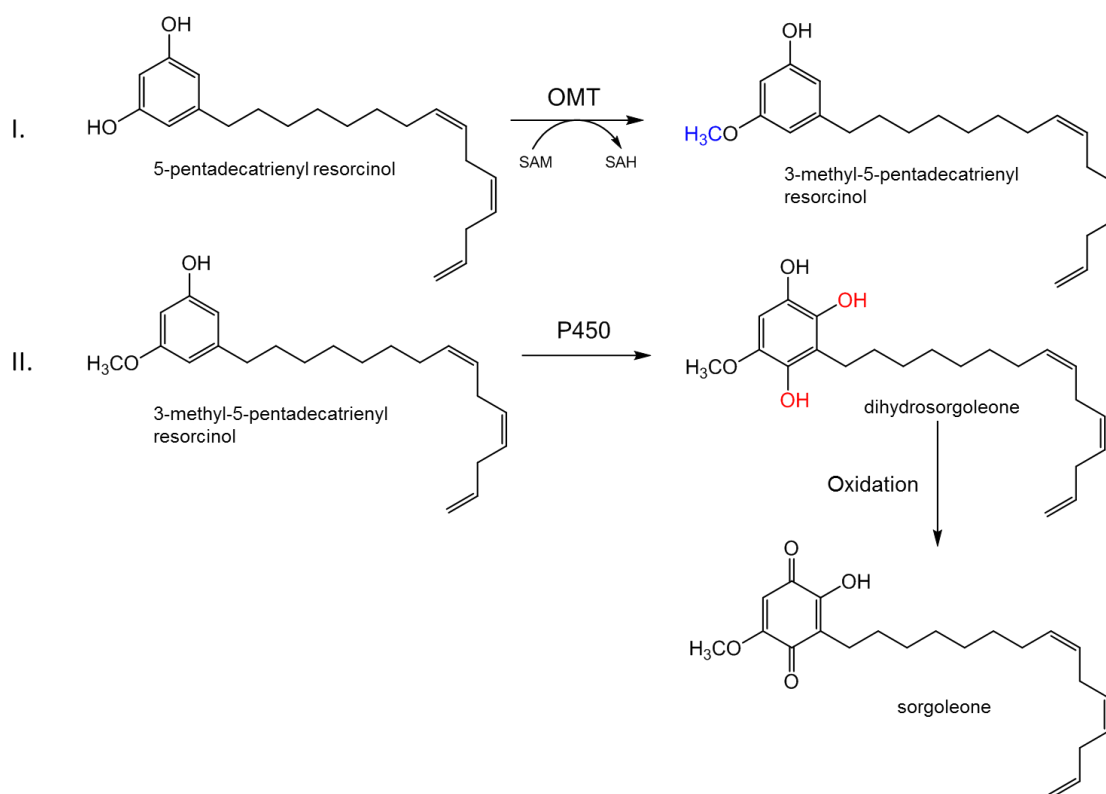
Fig. 1.7 Proposed biosynthetic pathway for sorgoleone synthesis. In the first step, the substrate cardol triene (5-pentadecatrienyl resorcinol) is converted to monomethoxycardol triene (5-pentadecatrienyl resorcinol-3-methyl ether) by an *O*-methyltransferase. S-adenosylmethionine (SAM) is used as methyl donor and S-adenosylhomocysteine (SAH) is released. In the second step, monomethoxycardol triene is dihydroxylated to dihydrosorgoleone by cyt-P450 monooxygenases. The product of the reaction auto-oxidases to sorgoleone.

1.8. Cashew Nut Shell Liquid as a potential feedstock for sorgoleone production

Cashew tree (*Anacardium occidentale* L.) is a native plant of Brazil and a representative of the *Anacardiaceae* family. Nowadays, the cashew tree is cultivated in many countries around the world, especially in tropical and subtropical regions. The cashew tree is mainly cultivated for its cashew apple, a pseudocarp used for production of beverages and food, and for the nut, which is the primary commercial product of cashew plantation (Duarte & Paull, 2015). The global annual production of cashew nut (with shell) consistently increased more than 370% in the past 25 years, rising from 1,004,017 tonnes in 1991 to 3,713,467 in 2014 (FAOSTAT, 2017). Considering a period of 10 years, from 2004 to 2014, the top 6 producers of cashew nut were Vietnam, Nigeria, India, Côte d'Ivoire, Brazil, and Indonesia, responsible for 80% of the global production of 41,393,467 tonnes (FAOSTAT, 2017). The cashew nut is covered by a hard shell which is filled with cashew nut shell liquid (CNSL), a

dark oily substance obtained as a by-product during the processing of the nuts. CNSL represents approximately 30-35% of the nut shell weight (Mazzetto et al., 2009), and it is composed by a mixture of different non-isoprenoid phenolic lipids.
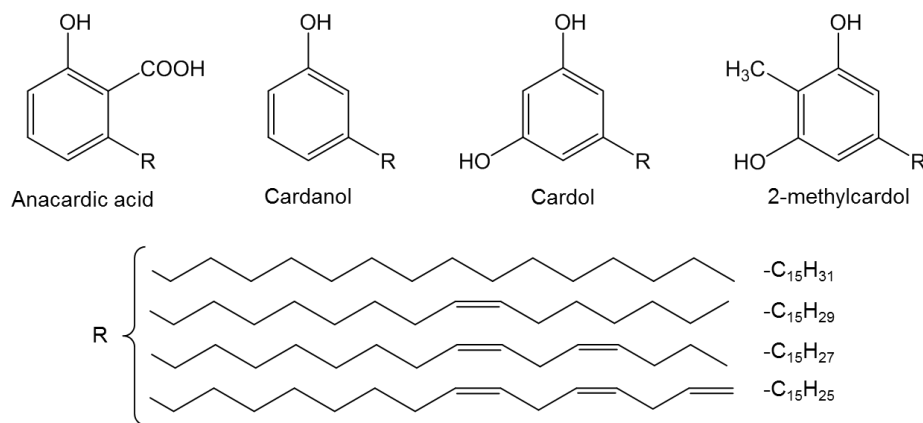
Fig. **1.8** Main constituents of Cashew Nut Shell Liquid.

CNSL contains four major components: 6-pentadecenyl salicylic acid (anacardic acid), 3-pentadecenyl phenol (cardanol), 5-pentadecenyl resorcinol (cardol) and 2-methyl-5-pentadecenyl resorcinol (2-methyl cardol) (Fig 1.8). Each of the constituents occurs as four meta-alkyl phenols (Tyman, 1996) differing in the degree of saturation in the side chain: 3-pentadecyl, 3-8Z-pentadecenyl, 3-8Z,11Z-pentadecadienyl, and 3-8Z,11Z,14-pentadecatrienyl. Differences in relative abundance of the components were observed according to the geographical origin of the cashew, plant age, and other extraction parameters (Tyman et al., 1981; Yuliana et al., 2012). The chemical composition of CNSL is greatly affected by the extraction method applied. The extraction processes can be classified into those that involve heating and those that are done in cold or room temperature. The solvent extraction and pressing methods are carried out at cold/room temperature producing natural CNSL; its typical composition comprises anacardic acid (60-70%), cardol (10-20%), cardanol (3-10%), and 2-methylcardol (2-5%). Technical CNSL (tCNSL) instead is the result of the roasting process of the nut shells at high temperatures (180-200 °C), which promotes the decarboxylation of anacardic acid into cardanol; as a result, tCNSL contains mainly cardanol (60-70%), cardol (10-20%), polymeric material (5-10%), and other minor constituents (Kumar et al., 2002).

The industrial applications of CNSL are numerous and include the manufacture of polymers and resins (such as polyurethanes, acrylics, polyesters, phenolic resins) (Balgude & Sabnis, 2014; Bhunia et al., 1999), specialty coatings, anticorrosive paints, and flame retardants (Menon et al., 1985;

Lubi & Thachil, 2000). CNSL is also useful in fuel blends and mixtures and for producing diesel oil (Vijayakumar et al., 2016; Pushparaj & Ramabalan, 2013). The biological activity of its components has been explored and they have been reported to have antimicrobial (Kubo et al., 1993a; Parasa et al., 2011), larvicidal (Lomonaco et al., 2009), antioxidant (Trevisan et al., 2006; Oliveira et al., 2011), antifungal (Kannan et al., 2009), and anti-tumour activity (Kubo et al., 1993b). Cardol triene was found to be an irreversible inhibitor of mushroom tyrosinase for the oxidation of L-DOPA in a dose-dependent manner (Zhuang et al., 2010). A number of applications of cardanol derived polymers has been described in literature (Calo et al., 2007; Pillai, 2010; Lubi & Thachil, 2000). Cardanol and cardol are considered very attractive molecule due to their structural features, which make them acceptable precursors for multiple chemical modifications. The long aliphatic chain containing up to three double bonds increases the hydrophobicity and the boiling point of the compounds, allowing several conversions via isomerization (Mgaya et al., 2015), polymerization (Kanehashi et al., 2013), metathesis (Julis et al., 2014), and carbonylation reactions (Mgaya et al., 2016). The reactive sites of the phenolic and the hydroxyl groups allow a series of additional modifications (Shukla & Srivastava, 2014).

Toxicity of CNSL was reported in different studies. CNLS was documented as responsible of contact dermatitis, eczematous dermatitis and vesicant reactions (Diogenes et al., 2005; Marks et al., 1984; Hirao et al., 2008), and weak promoter of carcinogenesis (George & Kuttan, 1997). Wastewater and effluents from the cashew industry showed toxic effects in *Artemia* bioassays. In general, the toxicity of CNSL could be ascribed to the high content in phenolic compounds (Pimentel et al., 2009; Leite et al., 2015). Modification of the composition of microbial communities after CNSL amendment was reported and putative CNSL-degrading microorganisms were identified (Sabna Prabha et al., 2011; Sindhu & Potty, 2015, 2016). Efficient treatment strategies are needed in order to reduce the environmental impact of the cashew industry.

Different methods have been reported in literature for the isolation of cardanol and cardol from CNSL. The distillation under reduced pressure was one of the first method described but still in use (Harvey, 1937; Sanjeeva et al., 2014). Other methods involve column chromatography, solvent extraction, and supercritical carbon dioxide (Kumar et al., 2002; Marriott & Gould, 2013). The development of simple and cost-effective methods for the isolation and purification of the constituents is crucial for their use in industrial applications.

1.9. Enzyme discovery from metagenomes

Frequently, the catalyst of choice for a designed reaction is already present in nature and its discovery through screening methods is faster and more efficient than engineering mediocre enzymes. In this context, the establishment of the sorgoleone enzymatic process relies upon finding and optimizing suitable enzymes. Microbes represent the most diverse and abundant group of living organisms and an exceptional resource of biocatalysts. The extreme diversity of microbial protein-encoded functions represents an appealing reservoir for novel biocatalysts for industrial exploitation. Metagenomics, as a culture-independent approach, makes accessible in principle the entire genetic fraction of a microbial community extracted directly from environmental samples. The possibilities offered by metagenomics allow us to explore the great microbial enzymatic diversity, identifying candidate catalysts required by the market. Currently, metagenomic libraries are established from a variety of habitats and a number of new enzymatic activities have been isolated and biochemically characterised (Ferrer et al., 2016). Metagenomic techniques for the screening of microbial enzyme diversity are comprehensively reviewed in the following chapter (see Chapter II).

1.10.   Aims and objectives of the project

The work conducted for the fulfilment of this thesis was partially performed in conjunction with the Innovate UK-funded project "Biotechnology for anti-weeds". It involved the collaboration with a team in the BioComposites Centre of Bangor University (UK), the industrial partners Almac Group (Craigavon, UK) and Hockley International (Manchester, UK). The overall aim of this PhD project was to identify and characterise through metagenomics approach novel biocatalysts of industrial interest that can constitute new enzyme-based "greener" processes.

The main hypotheses of the study were:

- The microorganisms from *Sorghum bicolor* rhizosphere, or/and soil microorganisms exposed to sorghum exudate and to sorgoleone-like compounds do possess an enzymatic arsenal which can be useful for modification of sorgoleone precursors and/or synthesis of sorgoleone.
- CNSL-, or other oleophilic compounds, exposed soil microbiome is enriched in microorganisms with hydrolytic enzymes of industrial interest, *inter alia*, in lipases and esterases.

To validate these, both sequence- and function-based metagenomic approaches were applied to identify enzymes of interest, which were cloned, expressed, characterised, and assessed for their applicability in potential applications.

The project was structured as follow:

- Enrichment of the soil bacterial community of *S. bicolor* in microcosm environment with CNSL, extraction of the environmental DNA and establishment of metagenomic libraries.
- Shotgun metagenome sequencing and analysis of abundance of relevant groups of enzymes, taxonomic and functional study of the changes in the soil microbial community structure in response to CNSL enrichment.
- Investigation into the possibility of the enzymatic synthesis of the bioherbicide sorgoleone: efforts towards the identification of SAM-*O*-methyltransferases capable of methylating selectively the resorcinolic compound cardol triene into the sorgoleone precursor. Biochemical characterization and optimization of the enzymatic reaction and evaluation of the feasibility of an industrial application.
- Screening of metagenomic libraries for novel industrially relevant enzymes (lipolytic activities) and their biochemical characterization.

The project allowed the investigation into the possibility of an eco-friendly, scalable and economic enzymatic synthesis of the bioherbicide sorgoleone. Inherent environmental and economic advantages would result from the establishment of an efficient synthetic process alternative to toxic chemicals, allowing the transformation of a low value chemical into a high-value material. Additionally, metagenomic screening allowed the discovery of new very active esterases that could outperform best industrial counterparts.

1.11.   References

Ahmad, S.A., Cheema, Z.A. & Ahmad, R. (2000). Evaluation of sorgaab as natural weed inhibitor in maize. *J. Plant Anim. Sci.* 10. p.pp. 141–146.

Alsaadawi, I.S., Al-Uqaili, J.K., Alrubeaa, A.J. & Al-Hadithy, S.M. (1986). Allelopathic suppression of weed and nitrification by selected cultivars ofSorghum bicolor (L.) moench. *Journal of Chemical Ecology*. 12 (1). p.pp. 209–219.

Alsaadawi, I.S., Sarbout, A.K. & Al-Shamma, L.M. (2012). Differential allelopathic potential of sunflower (Helianthus annuus L.) genotypes on weeds and wheat (Triticum aestivum L.) crop. *Archives of Agronomy and Soil Science*. 58 (10). p.pp. 1139–1148.

Austin, M.B. & Noel, J.P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports*. 20 (1). p.pp. 79–110.

Baerson, S.R., Dayan, F.E., Rimando, A.M., Nanayakkara, N.P.D., Liu, C.-J., Schröder, J., Fishbein, M., Pan, Z., Kagan, I. a, Pratt, L.H., Cordonnier-Pratt, M.-M. & Duke, S.O. (2008a). A functional genomics investigation of allelochemical biosynthesis in Sorghum bicolor root hairs. *The Journal of biological chemistry*. 283 (6). p.pp. 3231–47.

Baerson, S.R., Rimando, A.M. & Pan, Z. (2008b). Probing allelochemical biosynthesis in sorghum root hairs. *Plant signaling & behavior*. 3 (9). p.pp. 667–70.

Bains, G., Sampath Kumar, A., Rudrappa, T., Alff, E., Hanson, T.E. & Bais, H.P. (2009). Native Plant and Microbial Contributions to a Negative Plant-Plant Interaction. *Plant physiology*. 151 (4). p.pp. 2145–2151.

Bais, H.P., Loyola-Vargas, V.M., Flores, H.E. & Vivanco, J.M. (2001). Root-specific metabolism: The biology and biochemistry of underground organs. *In Vitro Cellular & Developmental Biology - Plant*. 37 (6). p.pp. 730–741.

Balgude, D. & Sabnis, A.S. (2014). CNSL: An environment friendly alternative for the modern coating industry. *Journal of Coatings Technology Research*. 11 (2) p.pp. 169–183.

Barbosa, L.C. de A., Ferreira, M.L., Demuner, A.J., Silva, A.A. da & Pereira, R. de C. (2001). Preparation and phytotoxicity of sorgoleone analogues. *Química Nova*. 24 (6). p.pp. 751–755.

Behrouzian, B. & Buist, P.H. (2002). Fatty acid desaturation: variations on an oxidative theme. *Current Opinion in Chemical Biology*. 6 (5). p.pp. 577–582.

Bertin, C., Yang, X. & Weston, L.A. (2003). The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil*. 256 (1). p.pp. 67–83.

Bhunia, H.P., Nando, G.B., Basak, A., Lenka, S. & Nayak, P.L. (1999). Synthesis and characterization of polymers from cashewnut shell liquid (CNSL), a renewable resource III. Synthesis of a polyether. *European Polymer Journal*. 35 (9). p.pp. 1713–1722.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C. & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*. 485 (7397). p.pp. 185–194.

Bossio, D.A. & Scow, K.M. (1998). Impacts of Carbon and Flooding on Soil Microbial Communities: Phospholipid Fatty Acid Profiles and Substrate Utilization Patterns. *Microbial Ecology*. 35 (3). p.pp. 265–278.

de Brito, A.M., Gagne, S. & Antoun, H. (1995). Effect of compost on rhizosphere microflora of the

tomato and on the incidence of plant growth-promoting rhizobacteria. *Applied and environmental microbiology*. 61 (1). p.pp. 194–9.

Browse, J., McConn, M., James, D. & Miquel, M. (1993). Mutants of Arabidopsis deficient in the synthesis of alpha-linolenate. Biochemical and genetic characterization of the endoplasmic reticulum linoleoyl desaturase. *The Journal of biological chemistry*. 268 (22). p.pp. 16345–51.

Calo, E., Maffezzoli, A., Mele, G., Martina, F., Mazzetto, S.E., Tarzia, A. & Stifani, C. (2007). Synthesis of a novel cardanol-based benzoxazine monomer and environmentally sustainable production of polymers and bio-composites. *Green Chemistry*. 9 (7). p.p. 754.

Chamberlain, K., Guerrieri, E., Pennacchio, F., Pettersson, J., Pickett, J.A., Poppy, G.M., Powell, W., Wadhams, L.J. & Woodcock, C.M. (2001). Can aphid-induced plant signals be transmitted aerially and through the rhizosphere? *Biochemical Systematics and Ecology*. 29 (10). p.pp. 1063–1074.

Cheema, Z., Khaliq, A., Abbas, M. & Farooq, M. (2007). *Allelopathic potential of sorghum (Sorghum bicolor L. Moench) cultivars for weed management*.

Cheema, Z., Khaliq, A. & Akhtar, S. (2001). *Use of Sorgaab (Sorghum Water Extract) as a Natural Weed Inhibitor in Spring Mungbean*.

Cheema, Z.A. & Khaliq, A. (2000). Use of sorghum allelopathic properties to control weeds in irrigated wheat in a semi arid region of Punjab. *Agriculture, Ecosystems & Environment*. 79 (2–3). p.pp. 105–112.

Cheema, Z.A., Khaliq, A. & Ali, K. (2002). Efficacy of sorgaab for weed control in wheat grown at different fertility levels. *Pakistan Journal of Weed Science Research*. 8 p.pp. 33–38.

Cheng, F. & Cheng, Z. (2015). Research Progress on the use of Plant Allelopathy in Agriculture and the Physiological and Ecological Mechanisms of Allelopathy. *Frontiers in plant science*. 6. p.p. 1020.

Cipollini, D., Rigsby, C.M. & Barto, E.K. (2012). Microbes as targets and mediators of allelopathy in plants. *Journal of chemical ecology*. 38 (6). p.pp. 714–27.

Cook, D., Rimando, A.M., Clemente, T.E., Schröder, J., Dayan, F.E., Nanayakkara, N.P.D., Pan, Z., Noonan, B.P., Fishbein, M., Abe, I., Duke, S.O. & Baerson, S.R. (2010). Alkylresorcinol synthases expressed in Sorghum bicolor root hairs play an essential role in the biosynthesis of the allelopathic benzoquinone sorgoleone. *The Plant cell*. 22 (3). p.pp. 867–87.

Czarnota, M.A., Paul, R.E.X.N., Dayan, F.E., Chandrashekhar, I. & Weston, L.A. (2001). *Mode of Action , Localization of Production , Chemical Nature , and Activity of Sorgoleone : A Potent PSII Inhibitor in Sorghum spp . Root Exudates Symposium Mode of Action , Localization of Production , Chemical Nature , and Activity of*. 15 (4). p.pp. 813–825.

Czarnota, M.A., Paul, R.N., Weston, L.A. & Duke, S.O. (2003a). Anatomy of Sorgoleone-Secreting Root Hairs of Sorghum Species. *International Journal of Plant Sciences*. 164 (6). p.pp. 861–866.

Czarnota, M.A., Rimando, A.M. & Weston, L.A. (2003b). Evaluation of seven Sorghum (Sorghum sp.) accessions. *Journal of Chemical Ecology*. 29 (9). p.pp. 2073–2083.

Dakora, F.D. & Phillips, D.A. (2002). Root exudates as mediators of mineral acquisition in low-nutrient environments. *Plant and Soil*. 245 (1). p.pp. 35–47.

Davis, B.G. & Boyer, V. (2001). Biocatalysis and enzymes in organic synthesis. *Natural product reports*. 18 (6). p.pp. 618–40.

Dayan, F.E. (2006). Factors modulating the levels of the allelochemical sorgoleone in Sorghum bicolor. *Planta*. 224 (2). p.pp. 339–46.

Dayan, F.E., Cantrell, C.L. & Duke, S.O. (2009a). Natural products in crop protection. *Bioorganic & medicinal chemistry*. 17 (12). p.pp. 4022–34.

Dayan, F.E., Howell, J. & Weidenhamer, J.D. (2009b). Dynamic root exudation of sorgoleone and its in planta mechanism of action. *Journal of experimental botany*. 60 (7). p.pp. 2107–17.

Dayan, F.E., Kagan, I. a & Rimando, A.M. (2003a). Elucidation of the biosynthetic pathway of the allelochemical sorgoleone using retrobiosynthetic NMR analysis. *The Journal of biological chemistry*. 278 (31). p.pp. 28607–11.

Dayan, F.E., Owens, D.K., Watson, S.B., Asolkar, R.N. & Boddy, L.G. (2015). Sarmentine, a natural herbicide from Piper species with multiple herbicide mechanisms of action. *Frontiers in plant science*. 6. p. p. 222.

Dayan, F.E., Rimando, A.M., Pan, Z., Baerson, S.R., Gimsing, A.L. & Duke, S.O. (2010). Sorgoleone. *Phytochemistry*. 71 (10). p.pp. 1032–9.

Dayan, F.E., Romagni, J.G. & Duke, S.O. (2003b). Herbicides, Cinmethylin. In: *Encyclopedia of Agrochemicals*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

de Almeida Barbosa, L.C., de Alvarenga, E.S., Demuner, A.J., Virtuoso, L.S. & Silva, A.A. (2006). Synthesis of New Phytogrowth-Inhibitory Substituted Aryl-p-Benzoquinones. *Chemistry & Biodiversity*. 3 (5). p.pp. 553–567.

Demain, A.L. & Adrio, J.L. (2008). Contributions of microorganisms to industrial biology. *Molecular biotechnology*. 38 (1). p.pp. 41–55.

Dilday, R.H., Mattice, J.D., Moldenhauer, K.A. & Yan, W. (2001). Allelopathic Potential in Rice Germplasm Against Ducksalad, Redstem and Barnyard Grass. *Journal of Crop Production*. 4 (2). p.pp. 287–301.

Diogenes, M.J.N., Vilar, M.D.O.L.L., Diogenes, J.A.C.N., Morais, S.M.D.E., Serrer, M.A.R., Andrade, A.N.A.R.C.D.E., Costa, E.B. & Alchorne, A.D.O.A. (2005). Detection of sensitivity to cardol in north-eastern and south-eastern Brazil. *Contact Dermatitis*. 52 (5). p.pp. 293–294.

Duarte, O. & Paull, R.E. (2015). *Exotic fruits and nuts of the New World*. O. Duarte & R. Paull (eds.). Wallingford: CABI.

Duke, S.O., Scheffler, B.E., Dayan, F.E., Weston, L.A. & Ota, E. (2001). Strategies for Using Transgenes to Produce Allelopathic Crops. *Weed Technology*. 15 (4). p.pp. 826–834.

Duke, S.O., Vaughn, K.C., Edward M. Croom, J. & Elsohly, H.N. (1987). Artemisinin, a Constituent of Annual Wormwood (Artemisia annua), Is a Selective Phytotoxin. *Weed Science*. 35 (4). p.pp. 499–505.

Einhellig, F.A., Rasmussen, J.A., Hejl, A.M. & Souza, I.F. (1993). Effects of root exudate sorgoleone on photosynthesis. *Journal of Chemical Ecology*. 19 (2). p.pp. 369–375.

Einhellig, F.A. & Souza, I.F. (1992). Phytotoxicity of sorgoleone found in grain Sorghum root exudates. *Journal of Chemical Ecology*. 18 (1). p.pp. 1–11.

Erickson, J., Schott, D., Reverri, T., Muhsin, W. & Ruttledge, T. (2001). GC-MS Analysis of Hydrophobic Root Exudates of Sorghum and Implications on the Parasitic Plant Striga asiatica. *Journal of Agricultural and Food Chemistry*. 49 (11). p.pp. 5537–5542.

Fahn, A. (2000). Structure and Function of Secretory Cells. *Advances in Botanical Research*. 31. p.pp. 37–75.

FAOSTAT (2017). *No Title*. [Online]. 2017. Food and Agriculture Organization of the United Nations. Available from: http://fao.org/faostat.

Fate, G.D. & Lynn, D.G. (1996). Xenognosin Methylation Is Critical in Defining the Chemical Potential Gradient That Regulates the Spatial Distribution in Striga Pathogenesis. *Journal of the American Chemical Society*. 118 (46). p.pp. 11369–11376.

Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A. & Noel, J.P. (1999). Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology*. 6 (8). p.pp. 775–784.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O. V. & Golyshin, P.N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology*. 9 (1). p.pp. 22–34.

Fierer, N. & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*. 103 (3).

Flores, H.E., Vicanco, J.M. & Loyola-Vargas, V.M. (1999). 'Radicle' biochemistry: The biology of root-specific metabolism. *Trends in Plant Science*. 4 (6) p.pp. 220–226.

Forney, D.R., Foy, C.L. & Wolf, D.D. (1985). Weed Suppression in No-Till Alfalfa (Medicago sativa) by Prior Cropping of Summer-Annual Forage Grasses. *Weed Science*. 33 (4). p.pp. 490–497.

Fujii, T., Shibuya, T. & Yasuda, T. (1992). Allelopathy of velvetbean [Mucuna pruriens]: Its discrimination and identification of L-DOPA as a candidate of allelopathic substances. *JARQ*. 25 p.pp. 238–247.

Funa, N., Awakawa, T. & Horinouchi, S. (2007). Pentaketide resorcylic acid synthesis by type III polyketide synthase from Neurospora crassa. *The Journal of biological chemistry*. 282 (19). p.pp. 14476–81.

Funa, N., Ozawa, H., Hirata, A. & Horinouchi, S. (2006). Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in Azotobacter vinelandii. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (16). p.pp. 6356–61.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P. & Pichersky, E. (2002). Characterization of phenylpropene O-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. *The Plant cell*. 14 (2). p.pp. 505–19.

Gents, M.B., Nielsen, S.T., Mortensen, A.G., Christophersen, C. & Fomsgaard, I.S. (2005). Transformation products of 2-benzoxazolinone (BOA) in soil. *Chemosphere*. 61 (1). p.pp. 74–84.

George, J. & Kuttan, R. (1997). Mutagenic, carcinogenic and cocarcinogenic activity of cashewnut shell liquid. *Cancer Letters*. 112 (1). p.pp. 11–16.

Gimsing, A.L., Bælum, J., Dayan, F.E., Locke, M.A., Sejerø, L.H. & Jacobsen, C.S. (2009). Mineralization of the allelochemical sorgoleone in soil. *Chemosphere*. 76 (8). p.pp. 1041–1047.

Goldfarb, K.C., Karaoz, U., Hanson, C.A., Santee, C.A., Bradford, M.A., Treseder, K.K., Wallenstein, M.D. & Brodie, E.L. (2011). Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical Recalcitrance. *Frontiers in Microbiology*. 2. p.pp. 94.

Guillon, M. (2003). *Herbicidal composition comprising an allelopathic substance and method of use thereof*.

Gupta, R., Beg, Q. & Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and

industrial applications. *Applied Microbiology and Biotechnology*. 59 (1). p.pp. 15–32.

Harvey, M.T. (1937). *Process of destructively distilling cashew nut shell liquid*.

Hejl, A.M. & Koster, K.L. (2004a). Juglone Disrupts Root Plasma Membrane H + -ATPase Activity and Impairs Water Uptake, Root Respiration, and Growth in Soybean (Glycine max) and Corn (Zea mays). *Journal of Chemical Ecology*. 30 (2). p.pp. 453–471.

Hejl, A.M. & Koster, K.L. (2004b). The Allelochemical Sorgoleone Inhibits Root H+ -ATPase and Water Uptake. *Journal of Chemical Ecology*. 30 (11). p.pp. 2181–2191.

Hirao, A., Oiso, N., Matsuda, H., Kawara, S. & Kawada, A. (2008). Occupational allergic contact dermatitis due to cashew nut oil. *Contact Dermatitis*. 59 (2). p.pp. 131–132.

Huang, H. & Asolkar, R. (2015). *Use of sarmentine and its analogs for controlling plant pests*.

Inderjit (2005). Soil Microorganisms: An Important Determinant of Allelopathic Activity. *Plant and Soil*. 274 (1–2). p.pp. 227–236.

Inderjit, Bajpai, D. & Rajeswari, M.S. (2010). Interaction of 8-hydroxyquinoline with soil environment mediates its ecological function S. J. Johnson (ed.). *PLoS ONE*. 5 (9). p.pp. 1–7.

Inderjit & Weiner, J. (2001). Plant allelochemical interference or soil chemical ecology? *Perspectives in Plant Ecology, Evolution and Systematics*. 4 (1). p.pp. 3–12.

International Survey of Herbicide Resistance Weeds (2018). *http://www.weedscience.org*. [Online]. 2018. Available from: http://www.weedscience.org.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Sanderlin, J.S., Reeves, J.H., Jenkins, M.B., Endale, D.M., Coleman, D.C. & Whitman, W.B. (2008). Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biology and Biochemistry*. 40 (11). p.pp. 2843–2853.

Jensen, L.B., Courtois, B., Shen, L., Li, Z., Olofsdotter, M. & Mauleon, R.P. (2001). Locating Genes Controlling Allelopathic Effects against Barnyardgrass in Upland Rice. *Agronomy Journal*. 93 (1). p.p. 21.

Jez, J.M., Austin, M.B., Ferrer, J.-L., Bowman, M.E., Schröder, J. & Noel, J.P. (2000). Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry & Biology*. 7 (12). p.pp. 919–930.

Jose, S. & Gillespie, A.R. (1998a). Allelopathy in black walnut (Juglans nigra L.) alley cropping. I. Spatio-temporal variation in soil juglone in a black walnut–corn (Zea maysL.) alley cropping system in the midwestern U. *Plant and Soil*. 203 (2). p.pp. 191–197.

Jose, S. & Gillespie, A.R. (1998b). Allelopathy in black walnut (Juglans nigraL.) alley cropping. II. Effects of juglone on hydroponically grown corn (Zea maysL.) and soybean (Glycine max. *Plant and Soil*. 203 (2). p.pp. 199–206.

Julis, J., Bartlett, S.A., Baader, S., Beresford, N., Routledge, E.J., Cazin, C.S.J. & Cole-Hamilton, D.J. (2014). Selective ethenolysis and oestrogenicity of compounds from cashew nut shell liquid. *Green Chem.* 16 (5). p.pp. 2846–2856.

Kagan, I. a, Rimando, A.M. & Dayan, F.E. (2003). Chromatographic Separation and in Vitro Activity of Sorgoleone Congeners from the Roots of Sorghum bicolor. *Journal of Agricultural and Food Chemistry*. 51 (26). p.pp. 7589–7595.

Kanehashi, S., Yokoyama, K., Masuda, R., Kidesaki, T., Nagai, K. & Miyakoshi, T. (2013). Preparation

and characterization of cardanol-based epoxy resin for coating at room temperature curing. *Journal of Applied Polymer Science*. 130 (4). p.pp. 2468–2478.

Kannan, V.R., Sumathi, C.S., Balasubramanian, V. & Ramesh, N. (2009). Elementary Chemical Profiling and Antifungal Properties of Cashew (Anacardium occidentale L .) Nuts. *Botany Research Intrenational*. 2 (4). p.pp. 253–257.

Kato-Noguchi, H. & Peters, R.J. (2013). The Role of Momilactones in Rice Allelopathy. *Journal of Chemical Ecology*. 39 (2). p.pp. 175–185.

Kim, K.-U. & Shin, D.-H. (2003). The importance of allelopathy in breeding new cultivars. *FAO Plant Production and Protection Paper (FAO)*.

Koeller, K.M. & Wong, C.-H. (2001). Enzymes for chemical synthesis. *Nature*. 409 (6817). p.pp. 232–240.

Kong, C.-H., Chen, X.-H., Hu, F. & Zhang, S.-Z. (2011). Breeding of commercially acceptable allelopathic rice cultivars in China. *Pest Management Science*. 67 (9). p.p. n/a-n/a.

Kong, C.H., Wang, P., Gu, Y., Xu, X.H. & Wang, M.L. (2008). Fate and Impact on Microorganisms of Rice Allelochemicals in Paddy Soil. *Journal of Agricultural and Food Chemistry*. 56 (13). p.pp. 5043–5049.

Kozubek, A. & Tyman, J.H.P. (1999). Resorcinolic Lipids, the Natural Non-isoprenoid Phenolic Amphiphiles and Their Biological Activity. *Chemical Reviews*. 99 (1). p.pp. 1–26.

Kubo, I., Muroi, H., Himejima, M., Yamagiwa, Y., Mera, H., Tokushima, K., Ohta, S. & Kamikawa, T. (1993a). Structure-antibacterial activity relationships of anacardic acids. *Journal of Agricultural and Food Chemistry*. 41 (6). p.pp. 1016–1019.

Kubo, I., Ochi, M., Vieira, P.C. & Komatsu, S. (1993b). Antitumor agents from the cashew (Anacardium occidentale) apple juice. *Journal of Agricultural and Food Chemistry*. 41 (6). p.pp. 1012–1015.

Kumar, P.P., Paramashivappa, R., Vithayathil, P.J., Rao, P.V.S. & Rao, a. S. (2002). Process for isolation of cardanol from technical cashew (Anacardium occidentale l.) Nut shell liquid. *Journal of Agricultural and Food Chemistry*. 50. p.pp. 4705–4708.

Leite, A. de S., Dantas, A.F., Oliveira, G.L. da S., Gomes Júnior, A.L., de Lima, S.G., Citó, A.M. das G.L., de Freitas, R.M., Melo-Cavalcante, A.A. de C. & Dantas Lopes, J.A. (2015). Evaluation of toxic, cytotoxic, mutagenic, and antimutagenic activities of natural and technical cashew nut shell liquids using the Allium cepa and Artemia salina bioassays. *BioMed research international*. 2015. p.p. 626835.

Li, H.-Y., Wei, D.-Q., Shen, M. & Zhou, Z.-P. (2012). Endophytes and their role in phytoremediation. *Fungal Diversity*. 54 (1). p.pp. 11–18.

Lomonaco, D., Pinheiro Santiago, G.M., Ferreira, Y.S., Campos Arriaga, Â.M., Mazzetto, S.E., Mele, G. & Vasapollo, G. (2009). Study of technical CNSL and its main components as new green larvicides. *Green Chem.* 11 (1). p.pp. 31–33.

Los, D.A. & Murata, N. (1998). Structure and expression of fatty acid desaturases. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*. 1394 (1). p.pp. 3–15.

Lubi, M.C. & Thachil, E.T. (2000). Cashew nut shell liquid (CNSL) - a versatile monomer for polymer synthesis. *Designed Monomers and Polymers*. 3 (2). p.pp. 123–153.

Marks, J.G., DeMelfi, T., McCarthy, M.A., Witte, E.J., Castagnoli, N., Epstein, W.L. & Aber, R.C. (1984). Dermatitis from cashew nuts. *Journal of the American Academy of Dermatology*. 10 (4). p.pp.

627–631.

Marriott, R.J. & Gould, M.S. (2013). *The extraction of cardol and cardanol from cashew nut shell liquid*.

Marschne, H. (2013). *Mineral nutrition of higher plants*. Academic Press.

Mazzetto, S.E., Lomonaco, D. & Mele, G. (2009). Óleo da castanha de caju: oportunidades e desafios no contexto do desenvolvimento e sustentabilidade industrial. *Química Nova*. 32 (3). p.pp. 732–741.

Meazza, G., Scheffler, B.E., Tellez, M.R., Rimando, A.M., Romagni, J.G., Duke, S.O., Nanayakkara, D., Khan, I.A., Abourashed, E.A. & Dayan, F.E. (2002). The inhibitory activity of natural products on plant p-hydroxyphenylpyruvate dioxygenase. *Phytochemistry*. 60 (3). p.pp. 281–288.

Mendes, R., Garbeva, P. & Raaijmakers, J.M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*. 37 (5). p.pp. 634–663.

Menon, A.R.R., Pillai, C.K.S., Sudha, J.D. & Mathew, A.G. (1985). Cashew nut shell liquid-its polymeric and other industrial products. *Journal of Scientific and Industrial Research*. 44. p.pp. 324–338.

Mgaya, J.E., Bartlett, S.A., Mubofu, E.B., Mgani, Q.A., Slawin, A.M.Z., Pogorzelec, P.J. & Cole-Hamilton, D.J. (2016). Synthesis of Bifunctional Monomers by the Palladium-Catalyzed Carbonylation of Cardanol and its Derivatives. *ChemCatChem*. 8 (4). p.pp. 751–757.

Mgaya, J.E., Mubofu, E.B., Mgani, Q.A., Cordes, D.B., Slawin, A.M. & Cole-Hamilton, D.J. (2015). Isomerization of anacardic acid: A possible route to the synthesis of an unsaturated benzolactone and a kairomone. *European Journal of Lipid Science and Technology*. 117 (2). p.pp. 190–199.

Miller, D.A. (1996). Allelopathy in Forage Crop Systems. *Agronomy Journal*. 88 (6). p.p. 854.

Monsan, P. & O'Donohue, M.J. (2010). *Industrial Biotechnology in the Food and Feed Sector*. W. Soetaert & E. J. Vandamme (eds.). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Moulin, L., Munive, A., Dreyfus, B. & Boivin-Masson, C. (2001). Nodulation of legumes by members of the β-subclass of Proteobacteria. *Nature*. 411 (6840). p.pp. 948–950.

Nannipieri, P., Ascher-Jenull, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G. & Valori, F. (2008). *Effects of Root Exudates in Microbial Diversity and Activity in Rhizosphere Soils*.

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G. & Renella, G. (2017). Microbial diversity and soil functions. *European Journal of Soil Science*. 68 (1). p.pp. 12–26.

Nardi, S., Concheri, G., Pizzeghello, D., Sturaro, A., Rella, R. & Parvoli, G. (2000). Soil organic matter mobilization by root exudates. *Chemosphere*. 41 (5). p.pp. 653–658.

Netzly, D. & Butler, L. (1986). Roots of sorghum exude hydrophobic droplets containing biologically active components. *Crop Science*. 26 (10377). p.pp. 775–778.

Nimbal, C., Yerkes, C., Weston, L. & Weller, S. (1996a). Herbicidal activity and site of action of the natural product sorgoleone. *Pesticide Biochemistry and …*. 83. p.pp. 73–83.

Nimbal, C.I., Pedersen, J.F., Yerkes, C.N., Weston, L. a. & Weller, S.C. (1996b). Phytotoxicity and Distribution of Sorgoleone in Grain Sorghum Germplasm. *Journal of Agricultural and Food Chemistry*. 44 (5). p.pp. 1343–1347.

Oerke, E.C. (2006). Crop losses to pests. *The Journal of Agricultural Science*. 144 (01). p.p. 31.

Ogata, T., Hamachi, M. & Nishi, K. (2008). *Organic Herbicide for Paddy Field*.

Oliveira, M.S.C., Morais, S.M. de, Magalhães, D.V., Batista, W.P., Vieira, Í.G.P., Craveiro, A.A., de Manezes, J.E.S.A., Carvalho, A.F.U. & de Lima, G.P.G. (2011). Antioxidant, larvicidal and antiacetylcholinesterase activities of cashew nut shell liquid constituents. *Acta Tropica*. 117 (3). p.pp. 165–170.

Pan, Z. & Baerson, S.R. (2014). *Novel cytochrome P450 enzymes from Sorghum bicolor*.

Pan, Z., Baerson, S.R., Wang, M., Bajsa-Hirschel, J., Rimando, A.M., Wang, X., Nanayakkara, N.P.D., Noonan, B.P., Fromm, M.E., Dayan, F.E., Khan, I.A. & Duke, S.O. (2018). A cytochrome P450 CYP71 enzyme expressed in *Sorghum bicolor* root hair cells participates in the biosynthesis of the benzoquinone allelochemical sorgoleone. *New Phytologist*. 218 (2). p.pp. 616–629.

Pan, Z., Rimando, A.M., Baerson, S.R., Fishbein, M. & Duke, S.O. (2007). Functional Characterization of Desaturases Involved in the Formation of the Terminal Double Bond of an Unusual 16:3Δ 9, 12, 15 Fatty Acid Isolated from Sorghum bicolor Root Hairs. *Journal of Biological Chemistry*. 282 (7). p.pp. 4326–4335.

Panasiuk, O., Bills, D.D. & Leather, G.R. (1986). Allelopathic influence ofSorghum bicolor on weeds during germination and early development of seedlings. *Journal of Chemical Ecology*. 12 (6). p.pp. 1533–1543.

Parasa, L.S., Tumati, S.R., Kumar, C., Chicurupati, S. & Rao, G. (2011). In vitro antimicrobial activity of cashew (Anacardium occidentale, L.) nuts shell liquid against methicillin resistant Staphylococcus aureus (MRSA) clinical isolates. *International Journal of Pharmacy Science*. (3). p.pp. 436–440.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I. V, Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J. & Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*. 457 (7229). p.pp. 551–6.

Pillai, C.K.S. (2010). Challenges for Natural Monomers and Polymers: Novel Design Strategies and Engineering to Develop Advanced Polymers. *Designed Monomers & Polymers*. 13 (2). p.pp. 87–121.

Pimentel, M.R., De Lima, D.P., Martins, L.R., Beatriz, A., Santaella, S.T. & Lotufo, L.V.C. (2009). Ecotoxicological analysis of cashew nut industry effluents, specifically two of its major phenolic components, cardol and cardanol. *Pan-American Journal of Aquatic Sciences*. 4 (3). p.pp. 363–368.

Pushparaj, T. & Ramabalan, S. (2013). Green fuel design for diesel engine, combustion, performance and emission analysis. In: *Procedia Engineering*. 2013, Elsevier, pp. 701–709.

Rawat, L.S., Maikhuri, R.K., Bahuguna, Y.M., Jha, N.K. & Phondani, P.C. (2017). Sunflower allelopathy for weed control in agriculture systems. *Journal of Crop Science and Biotechnology*. 20 (1). p.pp. 45–60.

Rice, E.L. (1984). *Allelopathy.* Elsevier Science.

Rimando, A.M., Dayan, F.E., Czarnota, M.A., Weston, L.A. & Duke, S.O. (1998). A New Photosystem II Electron Transfer Inhibitor from Sorghum bicolor. *Journal of Natural Products*. 61 (7). p.pp. 927–930.

Sabna Prabha, S., Sisu Pramod, S. & Potty, V. (2011). Isolation, Identification and Methanogenesis of

CNSL Degrading Bacteria and Immobilized Bioremediation Techniques of CNSL and its Contaminated Water from Cashew Industry. *Middle-East Journal of Scientific Research*. 10 (3). p.pp. 327–331.

Sanjeeva, S.K., Pinto, M.P., Narayanan, M.M., Kini, G.M., Nair, C.B., SubbaRao, P.V., Pullela, P.K., Ramamoorthy, S. & Barrow, C.J. (2014). Distilled technical cashew nut shell liquid (DT-CNSL) as an effective biofuel and additive to stabilize triglyceride biofuels in diesel. *Renewable Energy*. 71. p.pp. 81–88.

Sankaranarayanan, R., Saxena, P., Marathe, U.B., Gokhale, R.S., Shanmugam, V.M. & Rukmini, R. (2004). A novel tunnel in mycobacterial type III polyketide synthase reveals the structural basis for generating diverse metabolites. *Nature Structural & Molecular Biology*. 11 (9). p.pp. 894–900.

Sargent, M. V., Wangchareontrakul, S., Ma, C.-Y., Turner, C.E., Guerrero, A., Meinwald, J. & McPhail, A.T. (1990). The synthesis of the first natural host germination stimulant for Striga asiatica(witchweed). *Journal of the Chemical Society, Perkin Transactions 1*. 45 (5). p.p. 1429.

Scheidemann, P. & Wetzel, A. (1997). Identification and characterization of flavonoids in the root exudate of Robinia pseudoacacia. *Trees*. 11 (5). p.pp. 316–321.

Schloss, P.D. & Handelsman, J. (2006). Toward a census of bacteria in soil. *PLoS computational biology*. 2 (7). p.p. e92.

Schulz, M., Marocco, A., Tabaglio, V., Macias, F.A. & Molinillo, J.M.G. (2013). Benzoxazinoids in Rye Allelopathy - From Discovery to Application in Sustainable Weed Control and Organic Farming. *Journal of Chemical Ecology*. 39 (2). p.pp. 154–174.

Schutter, M.E., Sandeno, J.M. & Dick, R.P. (2001). Seasonal, soil type, and alternative management influences on microbial communities of vegetable cropping systems. *Biology and Fertility of Soils*. 34 (6). p.pp. 397–410.

Seal, A.N. & Pratley, J.E. (2010). The specificity of allelopathy in rice (Oryza sativa). *Weed Research*. 50 (4). p.pp. 303–311.

Selga, J. & Kiely, W.A. (1997). *Herbicidal composition and method*.

Shimizu, Y., Ogata, H. & Goto, S. (2017). Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem*. 18 (1). p.pp. 50–65.

Shukla, P. & Srivastava, D. (2014). Reaction kinetics of esterification of phenol-cardanol based epoxidized novolac resins and methacrylic acid. *International Journal of Plastics Technology*. 18 (1). p.pp. 1–15.

Sindhu, M.R. & Potty, V.P. (2015). Biodegradation of Cashew Nut Shell Liquid by Delftia Acidovorans and Pseudomonas Aeruginosa Isolated from Marine Environment. *International Journal of Agriculture, Environment and Biotechnology*. 8 (4). p.p. 837.

Sindhu, M.R. & Potty, V.P. (2016). Isolation of Novel Fungal Strains from Marine Environment capable of Cashew Nut Shell Liquid Degradation. *International Journal of Biotechnology and Biochemistry ISSN*. 12 (1). p.pp. 55–61.

Soltys, D., Krasuska, U., Bogatek, R. & Gniazdowsk, A. (2013). Allelochemicals as Bioherbicides — Present and Perspectives. In: *Herbicides - Current Research and Case Studies in Use*. InTech.

Stotzky, G. (1997). Soil as an Environment for Microbial Life. In: van Elsas, J.D., Trevors, J.T. and Wellington, E.M.H., Eds.,. In: *Modern soil microbiology*. pp. 1–20.

Torres, A., Oliva, R.M., Castellano, D. & Cross, P. (1996). *First World Congress on Allelopathy : A Science*

for the Future. p.p. 278.

Trevisan, M.T.S., Pfundstein, B., Haubner, R., Würtele, G., Spiegelhalder, B., Bartsch, H. & Owen, R.W. (2006). Characterization of alkyl phenols in cashew (Anacardium occidentale) products and assay of their antioxidant capacity. *Food and Chemical Toxicology*. 44 (2). p.pp. 188–197.

Trezzi, M.M., Vidal, R.A., Dick, D.P., Peralba, M.C.R. & Kruse, N.D. (2006). Sorptive behavior of sorgoleone in ultisol in two solvent systems and determination of its lipophilicity. *Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes*. 41 (4). p.pp. 345–356.

Tverezovskiy, V. V, Braganca, R.M., Howard, F.A., Gould, M.S., Fowler, P.A. & Baird, M.S. (2012). *Method for preparing a herbicidal compound*.

Tyman, J.H.P. (John H.P.. (1996). *Synthetic and natural phenols*. Elsevier.

Tyman, J.H.P., Tychopoulos, V. & Colenutt, B.A. (1981). Long-chain phenols. *Journal of Chromatography A*. 213 (2). p.pp. 287–300.

Uddin, M.R., Park, K.W., Kim, Y.K., Park, S.U. & Pyon, J.Y. (2010a). Enhancing Sorgoleone Levels in Grain Sorghum Root Exudates. *Journal of Chemical Ecology*. 36 (8). p.pp. 914–922.

Uddin, M.R., Park, S.U., Dayan, F.E. & Pyon, J.Y. (2014). Herbicidal activity of formulated sorgoleone, a natural product of sorghum root exudate. *Pest Management Science*. 70 (2). p.pp. 252–257.

Uddin, M.R., Won, O.-J. & Pyon, J.-Y. (2010b). Herbicidal Effects and Crop Selectivity of Sorgoleone, a Sorghum Root Exudate under Greenhouse and Field Conditions. *Korean Journal of Weed Science*. 30 (4). p.pp. 412–420.

Uttaro, A. (2006). Biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)*. 58 (10). p.pp. 563–571.

Vijayakumar, C., Ramesh, M., Murugesan, A., Panneerselvam, N., Subramaniam, D. & Bharathiraja, M. (2016). Biodiesel from plant seed oils as an alternate fuel for compression ignition engines—a review. *Environmental Science and Pollution Research*. 23 (24). p.pp. 24711–24730.

Walker, T.S. (2003). Root Exudation and Rhizosphere Biology. *PLANT PHYSIOLOGY*. 132 (1). p.pp. 44–51.

Wang, J., Dudareva, N., Bhakta, S., Raguso, R.A. & Pichersky, E. (1997). Floral scent production in Clarkia breweri (Onagraceae). II. Localization and developmental modulation of the enzyme S-adenosyl-L-methionine:(iso)eugenol O-methyltransferase and phenylpropanoid emission. *Plant physiology*. 114 (1). p.pp. 213–21.

Weir, T.L., Park, S.-W. & Vivanco, J.M. (2004). Biochemical and physiological mechanisms mediated by allelochemicals. *Current Opinion in Plant Biology*. 7 (4). p.pp. 472–479.

Weston, L.A. (1996). Utilization of Allelopathy for Weed Management in Agroecosystems. *Agronomy Journal*. 88 (6). p.p. 860.

Weston, L.A. & Duke, S.O. (2003). Weed and Crop Allelopathy. *Critical Reviews in Plant Sciences*. 22 (3–4). p.pp. 367–389.

Wu, H., An, M., Liu, D.L., Pratley, J. & Lemerle, D. (2008). Recent Advances in Wheat Allelopathy. In: *Allelopathy in Sustainable Agriculture and Forestry*. New York, NY: Springer New York, pp. 235–254.

Wu, H., Haig, T., Pratley, J., Lemerle, D. & An, M. (2001a). Allelochemicals in Wheat (Triticum aestivum

L.): Production and Exudation of 2,4-Dihydroxy-7-Methoxy-1,4-Benzoxazin-3-One. *Journal of Chemical Ecology*. 27 (8). p.pp. 1691–1700.

Wu, H., Haig, T., Pratley, J., Lemerle, D. & An, M. (2001b). Allelochemicals in wheat (Triticum aestivum L.): variation of phenolic acids in shoot tissues. *Journal of chemical ecology*. 27 (1). p.pp. 125–35.

Wu, H., Pratley, J., Lemerle, D. & Haig, T. (2000). Evaluation of seedling allelopathy in 453 wheat (Triticum aestivum) accessions against annual ryegrass (Lolium rigidum) by the equal-compartment-agar method. *Australian Journal of Agricultural Research*. 51 (7). p.p. 937.

Wu, H., Pratley, J., Ma, W. & Haig, T. (2003). Quantitative trait loci and molecular markers associated with wheat allelopathy. *TAG Theoretical and Applied Genetics*. 107 (8). p.pp. 1477–1481.

Yang, X., Owens, T.G., Scheffler, B.E. & Weston, L. a. (2004a). Manipulation of Root Hair Development and Sorgoleone Production in Sorghum Seedlings. *Journal of Chemical Ecology*. 30 (1). p.pp. 199–213.

Yang, X., Scheffler, B.E. & Weston, L. a (2004b). SOR1, a gene associated with bioherbicide production in sorghum root hairs. *Journal of experimental botany*. 55 (406). p.pp. 2251–9.

Yuliana, M., Tran-Thi, N.Y. & Ju, Y.-H. (2012). Effect of extraction methods on characteristic and composition of Indonesian cashew nut shell liquid. *Industrial crops and products*. 35 p.pp. 230–236.

Zhuang, J.-X., Hu, Y.-H., Yang, M.-H., Liu, F.-J., Qiu, L., Zhou, X.-W. & Chen, Q.-X. (2010). Irreversible Competitive Inhibitory Kinetics of Cardol Triene on Mushroom Tyrosinase. *Journal of Agricultural and Food Chemistry*. 58 (24). p.pp. 12993–12998.

Zubieta, C., He, X.Z., Dixon, R.A. & Noel, J.P. (2001). Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransf.pdf. *Nature structural biology*. 8 (3). p.pp. 271–279.

CHAPTER II

Metagenomic Mining of Enzyme Diversity[1]

Abstract

In the present there is a great necessity of suitable biocatalysts with high process performance, as a "greener" complementary alternative to the chemical synthesis. It is expected that in the coming decade, up to 40% of bulk chemical synthesis processes could be substituted by enzymatic catalysis. The identification and optimization of an appropriate enzyme represent important requirements to obtain a successful and efficient enzymatic process. In this context, the establishment of enzymatic processes in the industry is mainly a problem of finding and optimizing new enzymes. In this sense, Nature is the richest reservoir from which enzymes can be isolated because they are continuously changing and evolving as a consequence of natural processes of selection. We are now taking advantages of sequencing and extensive screening technologies to develop enzyme discovery strategies and to identify microbial enzymes with improved and unusual activities and specificities. These approaches, in combination with modern protein engineering methods and distinct combinatorial and rational methods, will increase our chances to generate new stabilized biocatalysts that fit industrial requirements. Here, we review the methodologies, obstacles and solving problems around metagenomics investigations to screen for enzymes with activities of interest.

2.1. Introduction

Microbes represent the most diverse and abundant group of organisms on Earth, making up to 60% of the total biomass of all living organisms. They dominate all ecosystems from soils and oceans to the habitats most hostile to life, such as polar environments or acidic streamers. They play a crucial role in biogeochemical cycling of all chemical elements, are responsible for a primary production and environmental cleanup (Phale et al., 2007). In human bodies, it has been estimated that there are 10 times more microbial cells than human cells (Curtis & Sloan, 2005; Turnbaugh et al., 2007), which collectively points at the necessity of understanding the microbial community structure and diversity

---

[1] Distaso M.A., Tran H., Ferrer M., Golyshin P.N. (2017) Metagenomic Mining of Enzyme Diversity. In: Lee S. (eds) Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals. Handbook of Hydrocarbon and Lipid Microbiology. Springer, Cham

for health and disease, microbe-host interactions, processes dynamics at small-to-global scales, and ultimately for understanding the evolution of life on Earth.

In the modern industrial era, microbes are exploited as a rich source of valuable industrial products with applications across all major areas. Microbial secondary metabolites are extremely important to our health and nutrition. For example, microbial products are used as antibiotics, anti-tumour agents and immunosuppressants in the pharmaceutical industry, and as biopesticides, anti-parasite agents and food-processing agents in the agricultural sector (Demain & Adrio, 2008). Just as an example, there are at least about 17000 bioactive natural products with antibiotic properties found in Bacteria, 8700 natural antibiotics in *Actinomycetales* and 4,900 in *Fungi* (Bérdy, 2005). Similarly, microbial products are widely used by the chemical industry for the production of amino acids, vitamins, organic acids, detergents, bio-catalysts and bioconversion agents, and by environmental industries for bio-remediation and the production of bioenergy.

The preference for the microbial production of a variety of compounds that can otherwise be isolated from plants or chemically synthesized, is due to: (1) the wide range of reactions that microorganisms are capable of perform; (2) their ability to adapt to different environmental settings and thus the possibility to produce them in inexpensive media; (3) easiness to genetically manipulate them for increased production; (4) the intrinsic great diversity, which leads different species to produce enzymes catalysing the same reactions, but with a rich flexibility with respect to catalytic efficiency and stability to different physical and chemical conditions (Demain & Adrio, 2008).

Enzymes of microbial origin are an exceptional natural resource of powerful biocatalysts capable of performing a wide range of reactions and accepting an extensive collection of complex molecules as substrates. In fact, from the wide range of sources of commercial enzymes used industrially, 88% are from microorganisms and the remaining are divided between animal (8%) and plant (4%) sources (Gurung et al., 2013). Currently there is a great demand for suitable enzymatic biocatalysts with high process performances to use as a greener alternative to chemical synthesis (Davis & Boyer, 2001; Koeller & Wong, 2001).

Due to the enormous natural diversity of the microbial community, which suggests a world of possibilities in terms of new catalytic activities, it is of a high interest to assess new natural environments for the isolation and characterization of novel enzymes. We have so far explored only a small fraction of this enormous diversity. Although prokaryotes cover the largest fraction of individual living organisms, with an estimated amount of 10^3-10^5 microbial species in just 1 g of soil (Schloss & Handelsman, 2006), almost 99% of bacteria are recalcitrant to culturing. This technical limitation has over the years stimulated the development of new culture-independent tools to disclose and characterize microbial genomes. However, to date, only about 11000 Bacteria and Archaea species

have been classified, and each year, at least 600 new species are described (Kyrpides et al., 2014; Yarza et al., 2014). Metagenomics represents a powerful approach for accessing and examining the biological and molecular diversity existing in different natural environments. Metagenomics is currently thought to be one of the likely technologies to provide the candidate molecules required by the market (Ferrer et al., 2016).

In this chapter, we focus on the importance and the possible application of metagenomic studies, and we review the methodologies for the sequence homology and functional screening of libraries to discover and characterise new enzymes with activities of interest.

2.2. An industrial perspective

The market of enzymes is distributed over various application areas, comprising food (45%), detergents (34%), textile processing (11%), leather (3%) and pulp and paper (1.2%) (Demain & Adrio, 2008). Their versatility allows their use in many processes, such as the degradation of natural polymers including cellulose and proteins, as well as in fine chemical industry for the regioselective or enantioselective synthesis of pharmaceuticals and agrochemicals molecules (Nguyen et al., 2008; Gupta et al., 2002; Monsan & O'Donohue, 2010). Enzymes can also offer an efficient strategy in the field of bioremediation by the degradation of polluting chemicals (Phale et al., 2007). For this reason, the annual demand for enzymes, which currently is growing almost 7% per year, is expected to be close to 10% in 2030. At present, "white" and "red" biotechnological sectors are facing an important request for new enzymes and metabolites. The ideal biocatalyst for any industrial application needs to function sufficiently well according to several performance parameters (Table 2.1).

Table 2.1. Set of criteria used to evaluate enzyme candidates from an industrial point of view. Most promising candidate enzymes can be selected for further process development (Lorenz & Eck, 2005).

Activity
pH profile
Temperature profile
Specific activity (U/mg)
Turnover frequency (k_{cat})

Efficiency
Space-time yield
Product inhibition
Byproduct/ingredient inhibition
Producibility/expression yield

Stability
Temperature stability
pH stability
Ingredient/byproduct stability
Solvent stability

Specificity
Substrate range
Substrate specificity (K_M, k_{cat}/K_M)
Substrate regioselectivity and enantioselectivity
Substrate conversion (%)

Food and beverage enzymes constitute the largest segment of industrial enzymes. Next in the list are lipases, valuable enzymes utilised in the digestion and processing of dietary lipids (e.g., triglycerides, fats, oils). Largely diversified in their enzymatic properties and substrate specificity, microbial lipases are more stable than animal or plant lipases. Moreover, lipases are widely used in a variety of other applications such as baking, detergents or in the conversion of vegetable oil into fuel (Guo & Xu, 2005; Gupta et al., 2004). A common industrial implementation of enzymes is their use in as detergent additives. A suitable enzyme must be active under alkalophilic and thermophilic conditions. Proteases, along with amylases, oxidases, peroxidases and cellulases are commonly used in detergents, enhancing the ability to remove tough stains and to make detergents eco-friendly (Mitidieri et al., 2006; Adrio & Demain, 2014). In the starch industry, thermostable amylases are used for the enzymatic hydrolysis of starch into glucose, fructose or maltose (Gomes et al., 2003). Other

relevant applications of amylases comprise food industry (baking, brewing, fruit juices), paper industry (α-amylases for the modification of starch of coated paper), textiles, and fuel ethanol from starches (Gurung et al., 2013). In the textile industry, enzymes play an important role in the development of cleaner processes and in reducing the use of raw material and production of waste; for example, cellulases for denim finishing and laccases for decolorization of textile effluents and bleaching (Araújo et al., 2008).

The replacement of existing chemical processes with enzymatic routes becomes a successful application when it contributes to an overall cost reduction, with an increased product titre, lower energy demand, lower volumes of wastewater streams and amount of byproducts. A number of biocatalytic routes have been fine-tuned for pharmaceutical manufacturing, showing pre-eminence above the traditional chemical pathways (Bornscheuer et al., 2012). The market of antibiotics, e.g. semi-synthetic penicillins and cephalosporins, is ruled by enzymatic processes. Another important application in biocatalysis is the generation of enantiomerically pure intermediates, especially when just one of the two isomers of a compound shows the desired activity; in this context, esterases, lipases, proteases are widely applied in the preparation of chiral compounds (Demain & Adrio, 2008). Most screening requests at the industrial scale for chemical synthesis address aldo-keto reductases, followed by transaminases and lipases (Ferrer et al., 2015).

Designing a process to fit a common/mediocre enzyme can be expensive and not profitable; on the contrary, it is reasonable to find a more efficient suitable natural enzyme, by the production of pre-characterized enzyme libraries using generic substrates. Once interesting enzymes or metabolites are found, they can be used as backbone for enzyme engineering to supply a biocatalyst able to optimally match process requirements. Scale-up and process optimization should lead to a viable industrial application.

2.2.1. Mining biocatalysts from extreme environments

Extreme environments provide a great source of microbes (extremophiles) adapted to a wide range of extreme conditions, both physical such as temperature (-2 to 110 °C), pressure or radiation, and geochemical such as salinity (2-5M NaCl) and extreme pH values (<2 and >9). Corresponding microbes have been isolated from samples collected in hot springs, near volcanoes, in the deep sea, in the Arctic and Antarctic ice, in deserts and arid areas, and in other extreme locations (Wilson et al., 2009). The ability of these microorganisms to survive under extreme conditions strictly depend on peculiar enzymes that not only support microbial growth but also can constitute robust biocatalysts exhibiting enhanced features merging the industrial requirements for a variety of applications (Van den Burg, 2003; Ferrer et al., 2005a; Herbert, 1992).

Thermophilic and hyperthermophilic enzymes are stable and active at high temperatures (Vieille & Zeikus, 2001; Hough & Danson, 1999) and are therefore of biotechnological interest thanks to their adaptability to harsh industrial reaction conditions. The conveniences of conducting reactions at high temperature are well known and include an increased solubility of polymeric substrates, increased bioavailability, faster reaction rates, the decreased risk of contamination and the cost saving due to the unnecessary cooling system. Extreme thermophiles, growing optimally at 60–80 °C, are widely distributed among the genera *Bacillus*, *Clostridium*, *Thermoanaerobacter*, *Thermus*, *Fervidobacterium*, *Rhodothermus*, *Thermotoga* and *Aquifex.* Most of hyperthermophiles (with a growth optimum of 80-110 °C) belong to the archaea, for example *Sulfolobus, Pyrolobus, Thermofilum* for the Crenarchaeota phylum, *Thermococcus, Pyrococcus, Thermoplasma* for the Euryarchaeota (Gomes and Steiner 2004). Thermophilic proteases, lipases, cellulases and amylases are being identified by screening metagenomic libraries and used in many different industrial application: detergent, food, starch, textile, pulp and paper industries are the major users (Cherry & Fidantsef, 2003; Atomi et al., 2011; Kumar et al., 2011; de Carvalho, 2011). The thermostable DNA polymerase is the most successful application of a product extracted from an extremophile, *Thermus aquaticus,* which was found in the Lower Geyser Basin of Yellowstone National Park, USA (Saiki et al., 1988).

High temperatures and alkaline media usually denote processes in the paper and pulp industry, and the identification of thermostable and alkali stable enzymes represents a major biotechnological goal in this field (Kumar et al., 2016). Cellulases and xylanases are used as bio-bleaching agents with economic and environmental advantages over chemical alternatives; moreover, they have biotechnological relevance in a number of other fields, including food, textile industries, as well as for saccharification of pre-treated lignocellulosic biomass for the production of biofuels (Hess, 2008).

Enzymes from halophiles are able to deal with high concentration of salts (up to 4 M KCl and over 5 M NaCl) (Hough & Danson, 1999). Proteins from halophilic organisms, with an excess of negatively charged amino acid residues on their surface, remain stable and active at high ionic strength (Madern et al., 2000). The property of low solubility of halophilic enzymes has been exploited by applying them in organic solvents and non-aqueous media (Klibanov, 2001). The production of halophilic enzymes, such as xylanases, amylases, proteases and lipases, has been reported for some halophiles belonging to the genera *Acinetobacter, Haloferax, Halobacterium, Halorhabdus, Marinococcus, Micrococcus, Natronococcus, Bacillus, Halobacillus* and *Halothermothrix* (Gomes & Steiner, 2004).

Microorganisms that can survive under extreme pH values could be good sources of thermoalkaliphilic or acidophilic enzymes (Golyshina et al., 2016; Méndez-García et al., 2015).

Proteases, lipases, cellulases, peroxidase or oxidoreductases, are used as additives in detergents, bio-bleaching of pulp and paper, or in the clean-up of effluent streams of the textile processing industry (Adrio & Demain, 2014; Gomes & Steiner, 2004).

The complexity in the cultivation of extremophiles, because of the longer generation times, lower biomass yields and strict cultivation conditions than mesophilic microorganisms, requires different strategies to overcome the problems (Schiraldi & De Rosa, 2002). Metagenomic approaches allow the screening of enzymes from environmental gene pools by cloning their sequences into suitable hosts such as *E. coli* or *P. pastoris* without the need to cultivate the original strains (Ferrer et al., 2005c).

2.3. Metagenomics and its approaches

Metagenomics is a powerful approach to unravel the biodiversity of microorganisms and activities, regardless of whether or not they can be cultured in laboratory. All metagenomic studies start with the isolation of genomic DNA from an environmental sample, bypassing the need for culturing the organisms that may serve as a source for enzymes. The following step is usually the construction of clone libraries, through the cloning of the environmental DNA into a suitable vector, transforming the cloned DNA into a host organisms (typically *E. coli*) and subsequent analysis of the DNA using both bioinformatics and experimental methods (Handelsman, 2004). Nowadays though, the environmental DNA is directly subjected to the de novo sequencing and data analysis, bypassing the intermediate (and obsolete) cloning step. The discovery of enzymes in metagenomes can be conducted using two different approaches. First, sequence-based metagenomic approaches that look for enzymes homologous to known biocatalysts in the (meta)genomic data and may include PCR-based methods that use primers designed according to conserved regions of known enzymes. Second, functional metagenomics approaches, where metagenomic libraries are established and clones screened with enzyme substrates (Lam et al., 2015; Sabree et al., 2009) (Fig. 2.1)
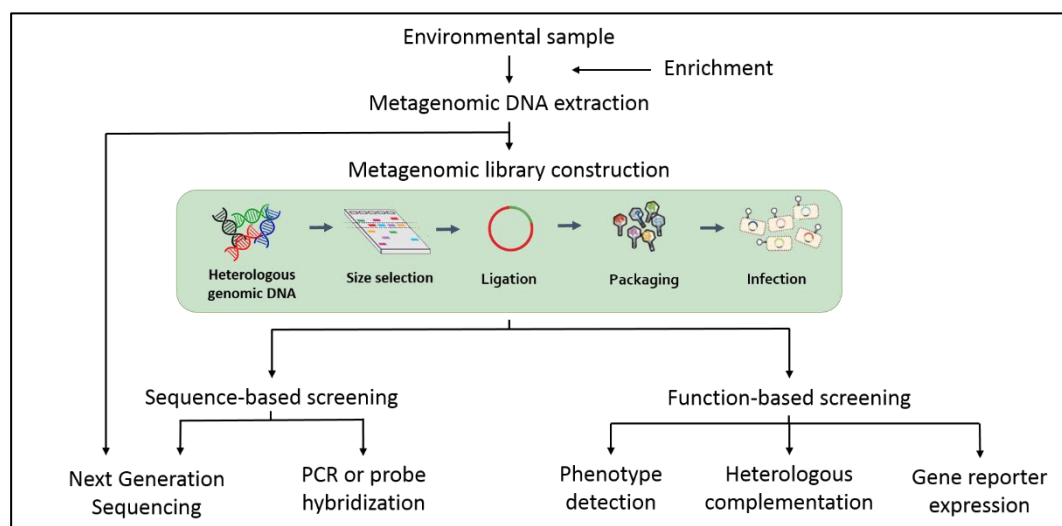
Fig. 2.1. A simplified scheme of metagenomic enzyme discovery pipeline, which highlights the difference in sequence- and function-based approaches.

The latter approach allows the discovery of genuinely novel enzymes, since bioinformatics analyses are homology-based and thus may only identify enzymes that are similar to characterised counterparts. Recently, the data on the discovery of new enzymes using metagenomic platforms has been reviewed for a period since 1998 (Ferrer et al., 2016). During that period, genetic materials from microbial communities from approximately 2200 sites around the world, with an emphasis on microbes from extreme habitats, have been examined. Environments of every kind have been included such as terrestrial (soil, plant rhizosphere) and marine habitats (superficial and deep seawater, hydrothermal vents), alkaline lakes, acid mine drainage systems, eukaryotes-associated microbiomes (rumen, gut). However, only a small fraction of the sites (11.6%) have been subjected to enzyme discovery studies with only seven thousand enzymes or clones containing enzymatic activities of interest for the industry identified and characterized (Ferrer et al., 2015, 2016), although in most cases rather superficially. This underlines that in spite of a number of success stories the biosphere is undersampled.

2.3.1. Screening Methods

2.3.1.1. Sequence-based Enzyme Discovery

Sequence-based (homology-driven, or "genome gazing") screening is based on the search for genes predicted to code for particular enzyme classes. This approach involves the design of PCR

primers or hybridization of probes on conserved regions and motifs of known protein families, and subsequent sequencing and comparison with homologous known enzymes (Ferrer et al., 2009). The approach is therefore only suitable to the discovery of members belonging to known gene families, since it completely relies on the available genome annotations. A large proportion of the open reading frames of newly-sequenced genomes have little sequence homology with known enzymes and this prevents the discovery of new protein classes with sequence elements that differ from the conserved sequences of the primers (Harrington et al., 2007)

The anchor used for the amplification or hybridization can be a 16S rRNA gene for a quick analysis providing a snap-shot of the diversity within the community (Béjà et al., 2000), but can also be a gene encoding an enzyme of interest. Sequence-based approached have led to the identification of genes encoding enzymes, such as lipases (Bell et al., 2002), glycerol dehydratases (Knietsch et al., 2003), chitinases (Hjort et al., 2010), nitrilases (Gong et al., 2013), and esterases (Ferrer et al., 2015).

Moreover, the advent of Next-Generation Sequencing platforms had a great impact. Nowadays we can analyse a huge set of sequence data, and many screening projects bypass the cloning step in favour of direct sequencing of the extracted environmental DNA. The analysis of the resulting large datasets allows the exploration of the taxonomic and functional biodiversity and of the system biology of diverse ecosystems. On the other hand, the fast increase of sequences deposited in public databases was not followed by accurate expetimental characterization, therefore the majority of the genes have been poorly or mis-annotated, and the encoded proteins labelled as hypothetical or protein of unknown functions (Schnoes et al., 2009; Fernández-Arrojo et al., 2010). A further disadvantage for the homology-based approach is that its efficiency is based on the quality and completeness of genome annotations in current databases (Hallin et al., 2008). Additionally, many of the identified sequences did not correspond to full-length proteins due to low sequence coverage.

Homology-based approaches mainly suffer from the long computation time required to search for homologs for each of the sequences within the typically massive metagenomic data sets. Recently, web-based metagenomic annotation platforms, such as the metagenomics RAST (MG-RAST) server, the IMG/M server, or JCVI Metagenomics Reports (METAREP) have been designed to analyse metagenomic data sets. The uploaded environmental data sets can be compared vs both protein and nucleotide databases. In this way, multiple metagenomic data sets derived from various environments can be compared at various functional and taxonomic levels (Meyer et al., 2008; Goll et al., 2010).

2.3.1.2.   Function-based Enzyme Discovery

Unfortunately, the generation of vast metagenome sequencing data is not followed, proportionately, by the discovery of new activities or functional characterization of proteins. Fewer than 2 out of 10,000 hits have led to the discovery of industrially relevant biocatalyst (Lorenz & Eck, 2005). Activity-based metagenomics provides an opportunity to circumvent this limitation. The functional approach is not dependent on previous genomic knowledge and allows for discovery of novel enzymes with unexpected peptide sequences with classical or new activities that would not be predicted based on DNA sequence alone. Thus, functional metagenomics complements sequence-based metagenomics, adding functional information or correcting incorrect functional assignments to nucleic acid and protein databases.

Function-driven screens usually begins with the construction of genomic expression libraries and the use of three different strategies for identifying clones of interest: direct phenotype detection using a specific colorimetric or fluorimetric substrate (Handelsman, 2004), heterologous complementation of host strains (Mirete et al., 2007; Kazimierczak et al., 2009), and detection following induction of biosensors or reporter gene expression (Uchiyama et al., 2005). The probability (hit rate) of identifying a certain gene depends on multiple factors that are intricately linked to each other: the DNA extraction method, the host-vector system, the size of the target gene, its abundance in the source metagenome, the assay method, and the efficiency of heterologous gene expression in a surrogate host.

2.3.1.3.   Cloning vectors/vehicles and library construction

DNA is first harvested from environmental samples, then size-selected, end-repaired and ligated to a vector, allowing packaging by lambda phage for subsequent infection of *E. coli.* According to the primary goal, small or large-insert libraries can be preferred, each of them offering particular advantages and disadvantages.

To obtain a function encoded by a single gene, small insert expression libraries can be built by cloning small-sized fragments (<10 kb, equivalent to 4-8 genes) into standard cloning vectors (e.g., pUC derivatives, pTOPO-XL, pBluescript SK$^+$ and pCF430) or lambda phage vectors (Sabree et al., 2009). Therefore, most genes are present in the appropriate orientation and under the influence of strong vector promoters, and thus they have a good chance of being expressed and detected by activity screens (Ferrer et al., 2009). Moreover, due to the lysis of *E. coli* cells at the end of the phage infection cycle, translated proteins are released to the extracellular matrix, which means an earlier detection of the screened activity and a negligible impact of the toxic effects associated with the expression of

lethal genes. The clear disadvantage of above plasmid vectors is in their high (up to 700) copy numbers per cell, which may facilitate a leaky expression of some genes that are incompatible with the host wellbeing.

On the other hand, to obtain targets encoded by multiple genes, large fragments must be cloned into fosmids, cosmids, or bacterial artificial chromosomes (BACs), which can harbour fragments up to 300 kbp. The commercially available pCC1FOS vector has become a frequently used tool for cloning large DNA fragments (ca. 40 kbp) from various microbial communities. One of the advantages of using pCC1FOS vector, when transformed into the appropriate host (e.g. *E. coli* Epi300), is that there is a possibility to induce its copy number, which allows a higher gene dosage and eventually a better expression than from a single-copy uninduced vector (Wild et al., 2002). Multicloning sites also allow the Sanger sequencing of termini of cloned fragments and reveal important information about the native metabolic role of the enzyme in question or can help to identify the source organism of the DNA fragment (Popovic et al., 2015).

Enzyme activities are usually assayed on agar plates supplemented with enzyme substrates. By cultivating a metagenomic library on the plates, one can identify positive clones through visual screening for the appearance of a clear zone (halo) or colour. Thousands of clones can be analysed in a single screen, with the potential to identify novel classes of proteins with known or unknown functions. Agar plate-based screening has been successfully applied to screen different metagenomic gene libraries to mine industrially useful enzymes such as lipases, esterases, cellulases, proteases, laccases and other activities (Ferrer et al., 2009, 2010; Placido et al., 2015; Tchigvintsev et al., 2015).

The hydrolysis of emulsified triacylglycerols, such as tributyrin, triolein and olive oil is generally applied to assess lipase activity, while soluble short-chain fatty acid esters are applied to study esterase activity (Placido et al., 2015). Seven lipase-producing thermophilic bacteria were identified from a Malaysian hot spring streaking the cultures on an agar plate containing olive oil, showing lipolytic activity up to 4.58 U/ml for the catalyst extracted from *Bacillus* (Sheikh Abdul Hamid et al., 2003).

The diversity of rumen hydrolytic enzymes was for the first time investigated by screening a metagenomic phage library of the rumen content of a dairy cow for hydrolase activity (Ferrer et al., 2005b). A total of 22 clones with hydrolytic activities were identified (12 esterases, 9 endo-β-1,4-glucanases and one cyclodextrinase) and characterized, among them eight were entirely new, showing no sequence similarity to enzymes deposited in public databases.

If the target activities can be linked to the survivability of the host organism, which requires the addressed gene to grow under selective conditions, the screen becomes highly sensitive and high throughput. This method is applied frequently to screen for resistance genes to toxic compounds, such

as antibiotics or heavy metals. The approach has been successful applied in the detection of enzymes such as racemases (Chen et al., 2010), DNA polymerases (Simon et al., 2009) and β-lactamases. For example, 5.4 Gbp of soil DNA led to the identification of nine unique clones containing resistance to aminoglycosidase antibiotics and one clone expressing resistance to tetracycline (Riesenfeld et al., 2004).

A functional study conducted on 12 marine environments that included extreme anoxic deep sea sites (deep hypersaline anoxic basins *Urania, Kryos,* and *Medee* in the Eastern Mediterranean Sea, and epibionts of the gut and gill of the deep sea shrimp *Rimicaris*), extremely low pH sites (Vulcano Island hydrotherms) and regions characterized by heavy industrialization and oil contamination (seawater of Messina harbour, Milazzo and Priolo refineries (Sicily), MT Haven shipwreck (near Genoa, Italy), superficial petroleum-polluted seawater near Kolguev Island and Port of Murmansk (Russia)), have yielded to the identification of several putative enzymes (Popovic et al., 2015). The frequency of positive hits was of 1 per 9 Mbp of screened DNA for esterase/lipase, 1 hit per 28.4 Mbp for glycosyl hydrolases and per 23.9 Mbp for dehalogenase screens (Popovic et al., 2015). Among identified biocatalysts, several esterases belonged to uncharacterized families or to proteins annotated to have alternate functions, proving the unique advantage of the functional screen over the sequence-based one.

2.4. Limitation in the metagenomic enzyme discovery process

2.4.1. Coverage, representation and need for enrichment

Some bottlenecks in the metagenomic enzyme discovery process lower the transition from the discovery stage of an enzyme to its application. The library representativeness of the diversity present in the original environmental DNA is an important element that must be taken in consideration if we want to access the full potential of environmental metagenomes. Community coverage is possible only in simple or high specialized environments characterized by low species richness, such as acid mine drainage at Iron Mountain, California (Tyson et al., 2004; Ram et al., 2005). The relatively simple structure of its microbial community led to the assembly of five complete genomes. In contrast, the metagenome of a very complex community, such as that from the soil, cannot be exhaustively sequenced at reasonable costs, and the reconstruction of genomes or metabolic pathways is only possible for few predominant organisms. Some sequences are less likely to be captured in libraries, and this inevitably represents a barrier to the screening.

A comparative study of shotgun sequencing of original sample to corresponding metagenomic libraries from human faeces (Lam & Charles, 2015) and corn field soil (Cheng et al., 2014) highlighted

how the relative abundance of phyla can differ in the library compared to its corresponding extract. Interestingly, the analysis indicated a bias in the relative abundance of each Operation Taxonomic Units (OTU), with some being 1000-fold overrepresented and other underrepresented in the library; AT-rich sequences appeared to be inadequately represented (Lam & Charles, 2015). The sequencing of extremely large libraries is necessary for a complete coverage of species-rich samples. In addition, the enormous volume of data generated by environmental sequencing requires advances in bioinformatics and the development of new algorithms specifically designed for metagenomic data analysis (Wooley et al., 2010)

Due to the high diversity within the microbial communities, target genes encoding for novel enzymes represent a tiny fraction of the total nucleic acid sample extracted. However, microbial community composition can be actively manipulated prior to metagenomic library construction, in order to enrich for desired activities. For functional screening, pre-enrichment of environmental samples is usually obtained by amending with specific nutrients or substrates, favouring the growth of the microbes that express the desired biocatalyst. Selection pressure for enrichment can obtain upon chemical or physical criteria. For example, a size-selective filtration was used for enrichment in the Sargasso Sea genome sequencing project (Venter et al., 2004). The main drawback of the artificial enrichment is the danger of enriching fast-growing microorganisms that do not utilize the supplied nutrients. A recent success example of the pre-enrichment method related to the research of new amidases for the enzymatic conversion of D-phenylglycine into intermediates for the synthesis of β-lactam antibiotics conducted on soil (Gabor et al., 2004). To increment the likelihood of finding useful enzymes, enrichment cultures were performed with D-phenylglycine amide as exclusive nitrogen source. 16S rRNA analysis showed a 4 times lower bacterial diversity in the enrichment cultures than the original sample. Clones exhibiting amidase activity were selected on a medium containing phenylacetyl-L-leucine or D-phenylglycine-L-leucine, allowing only the growth of recombinants capable of hydrolysing the amide compounds. Amide-positive clones were identified and after an extensive substrate profiling it was isolated a single clone (pS2) which supported a more than 2-fold higher maximal level of penicillin G accumulation than *E. coli* penicillin amidase (Gabor et al., 2004). In another study, the efficiency of screening metagenomic libraries for cellulose activities was improved more than 3 times by incubating the soil sample with crystalline cellulose before DNA extraction (Mori et al., 2014).

Similarly, microbial communities within contaminated ecosystems tend to be dominated by the organisms capable of consuming and/or tolerating toxic organic compounds. Phylogenetic analysis conducted on sediment from Milazzo bay showed how the structure and composition of the bacterial community capable of growing in hydrocarbon-contaminated superficial sediments dramatically

change in response to nutrient load and to addition of various hydrocarbons (Yakimov et al., 2005). Environmental samples can dynamically vary in phylogenetic and process equilibrium: the population of a closed niche may be modified until the system reach a new equilibrium where the key microorganisms have the working capacity to make the overall community work (Yakimov et al., 2005; Bargiela et al., 2015).

New PCR-independent amplification techniques that use multiple displacement amplification (MDA) with Φ29 DNA polymerase (Blanco et al., 1989) for whole genome amplification have allowed access to rare microbes, present in low numbers in bacterial communities or from single cells (Spits et al., 2006).

A method for enrichment is the so-called 'stable isotope probing' (SIP), an approach that selectively enriches metabolically active microbes that consume a specific labelled substrate. The DNA that incorporates high amounts of heavy isotope (such as ^{13}C) belongs to metabolically active microorganisms and it can be separated by density centrifugation. The labelled nucleic acid can be therefore used for the generation of a metagenomic enriched library (Radajewski et al., 2000; Neufeld et al., 2008). This methodology was successfully used to specifically enrich and label methylotrophic populations in sample of lake sediments able to use labelled C_1 compounds (Kalyuzhnaya et al., 2008). SIP was also useful for the isolation of various enzymes, such as novel biphenyl dioxygenase (Sul et al., 2009). Without pre-existing knowledge on the identity of the microbes within the microbial community, it is thus possible to have a direct access to their genome. However, the stable isotope probing approach can result in a loss of novel catalysts that are only present in a small number of microbes, which can be superseded by fast-growing and more abundant bacteria.

Differential display (DD) is an alternate technique that can be used for the discovery of bacterial genes and it can be applied to identify differentially expressed genes. DD is used to compare the mRNA pools: cells are grown under different physiological conditions, messenger RNA is then amplified at arbitrary sites by reverse transcription (RT), followed by RT-PCR (Liang et al., 1993; Galvão et al., 2005). Levels of gene expression are thus compared and genes expressed only under a specific condition will give rise to RT-PCR bands. These are further analysed by DNA sequencing to discover new ORFs (Brzostowicz et al., 2003). The application of this approach to the prokaryotic system requires the use of arbitrary oligonucleotide primers to initiate RT of the mRNA at random sites, while for eukaryotic mRNA it is possible to take advantage of the poly(A) tails. Differential display has been successful used to discover genes for cyclohexanone monooxygenase in mixed cultures from a wastewater bioreactor (Brzostowicz et al., 2003), to discover a putative operon for 2,4-dinitrophenol degradation in *Rhodococcus erythropolis* (Walters et al., 2001), and to explore the gene expression

patterns in marine microbial communities using an adaptation of this technique called Transcriptome Fingerprinting Analysis (TFA) (Coll-Lladó et al., 2011).

2.4.2. Need for multiple hosts

Currently, the bacterium *E. coli* is the most common host organism used for screening metagenomic libraries, although the heterologous expression remains a barrier for the extraction of the maximum information from functional analyses. The *E. coli* transcription-translation machinery is not completely compatible with the expression of genes harvested from environmental microbes: this can result in a low proportion of positive clones or in inactive enzymes after expression (Gabor et al. 2007; Loeschcke et al. 2013)*.

An alternative host is *Bacillus subtilis*, which is endotoxin-free and secretes proteins into the extracellular medium; unfortunately there is an insufficient number of expression vectors conceived for this bacterium that furthermore presents plasmid instability, misassembled proteins and active proteases (de Carvalho, 2016). Other alternative organisms for library construction and screening are being used, including strains of *Streptomyces* and *Pseudomonas* (Ferrer et al., 2007) or *Rhizobium leguminosarum* (Wexler et al., 2005)*. The adoption of a broader host range vectors capable of replication in several hosts will positively expand the range of detectable activities, although further optimization of the conditions for high transformation efficiency and for expression of, for instance, metagenomes derived from archaea-dominated communities, is still necessary. While expression systems in further bacteria (e.g. *Rhodobacter, Burkholderia),* archaea *(Haloferax, Sulfolobus*) and yeasts *(Pichia, Saccharomyces, Schizosaccharomyces)* (Craig et al., 2010; Liebl et al., 2014) are relatively well established for expression of individual proteins, none of these hosts have so far been utilised for metagenomic libraries construction and screening.

2.5. High-throughput and single cell genomics

The impact of NGS technologies on metagenomics has been very profound; currently a typical metagenomic project is sequence-based and generates large amounts of sequence data. In fact, conventional function-based screening of metagenomic libraries faces challenges such as low hit rate of positive clones, labour intensiveness, low throughput, and excessive time. In functional metagenomic screenings, the low frequency of positive clones is a significant limitation. As we already discussed, common screening methods are based on the conversion of easy-to-screen general assays in solid or liquid media. The lack of relevant substrates and screening methods for rare enzymatic

activities, as well as the poor performance of enzymes under non-natural conditions, represent a major technological limitation (Fernández-Arrojo et al., 2010; Singh, 2010). The development of multi-substrate approaches for high-throughput functional screening and the design of new chromogenic compounds that can mimic the real complex target substrates should be of high interest. High-throughput screening methods are inevitably necessary: new assays have been established in recent years and they are in constant development along with the progress in robotics and functional assays.

2.5.1. Single-cell genomics

Single-cell genomic (SCG) is a novel technology that is paving the way to new methods of analysing and understanding our environment. Specific organisms of interest can be targeted by processing a small number of cells that are relevant to a specific function in the environment. A complex sample can be enriched for a target cell population using fluorescence-activated cell sorting (FACS): it enables the physical separation of a user-defined sub-population of cells of interest from a complicated mixture for further analysis, such as downstream genomics or proteomics (Bergquist et al., 2009; Kodzius, 2016). Therefore, this approach allows the identification of biological activity within a single cell matching specific criteria, such as size, shape, or presence of specific nucleic acids or activities, using fluorescent dyes (Rinke et al., 2014a; Woyke et al., 2009). Among the many advantages of FACS are its high throughput, high sorting speed, and ability to sort live cells. In addition, FACS can detect the presence of cells in a droplet, rejecting the empty ones, and directing the cells with desired properties into multiwell/microtitre plates (Kodzius, 2016). Droplet microfluidic-based screening is recognized as an efficient tool for the assessment of cellular behaviour at the single-cell level. FACS can be used to detect expression of certain types of genes by regulation of a fluorescent biosensor present in the same cell as the metagenomic DNA (Rinke et al., 2014b). The co-encapsulation of cells expressing target molecules (e.g., enzymes and antibodies) with a reporter molecule, such as a fluorescent substance, provides a rapid detection of droplets that contain cells producing molecules of interest (Brouzes et al., 2009; Agresti et al., 2010; Scanlon et al., 2014).

In a recent work, genes coding for lipolytic enzymes were retrieved from a soil metagenomic library through a gel microdroplet (GMD) technique combined with fluorescent-activated cell sorting (FACS) (Hosokawa et al., 2015). Directly after the library construction into the fosmid vector and the infection of EPI300-T1[R] *E. coli* cells, the clones were encapsulated in agarose GMD with fluorogenic substrates (fluorescein dicaprylate) and IPTG via a microfluidic droplet generator. After incubation, the substrate diffused into the periplasm and was enzymatically hydrolysed; microscopic observation enabled detection and counting of lipolytic-active clones from the library pools in a digital format, based on the presence or absence of green fluorescence products in the GMDs. A hit rate of 0.62%

was estimated with this method compared to the 0.28% of the conventional agar-plate assay, thanks to a more clear identification of the fluorescent signal rather than the hydrolysis halo. Moreover, the GMD based assay has the potential of reducing significantly the reagent volume required (approximately 100 pL/cell) and the time required to assess the activity of clones from 3 days to 12 hours.

2.5.2. SIGEX

An efficient screening tool is the substrate-induced gene expression (SIGEX) which, in combination with FACS technology, allows the high-throughput screening of 30,000 clones/s of metagenomic libraries constructed in an operon-trap vector (Uchiyama & Miyazaki, 2010). The operon-trap vector includes a co-expressed reporter gene, such as *gfp*. The GFP fluorescence is used to detect clones that contain metagenomic fragments containing genes that are expressed in response to induction substrates. This approach was successful applied to identify transcriptional activators responsible for aromatic compound degradation (Uchiyama et al., 2005; Uchiyama & Miyazaki, 2013) using different aromatic inducing compounds, such as benzoate, naphthalene, salicylate, 3-methylcatechol and 4-chlorocatechol.
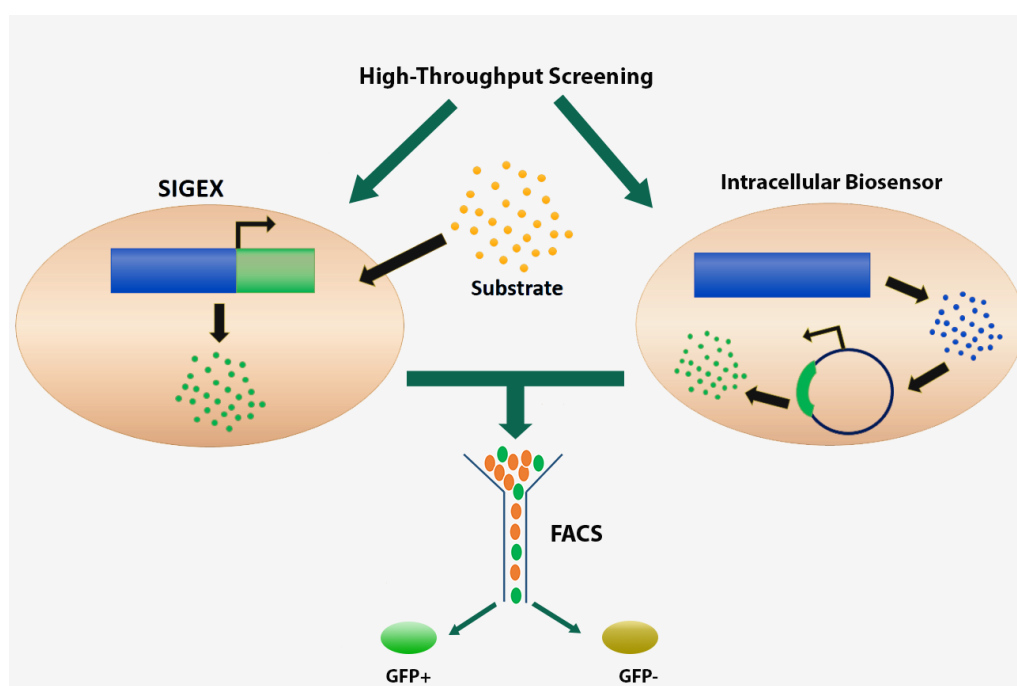
Fig. **2.2**. Substrate-induced gene expression (SIGEX). Metagenomic DNA is cloned into a promoter-trap vector containing a *gfp*-reporter downstream of the cloning site. Promoters in the metagenomic DNA that respond to specific carabolites induce expression of GFP. An intracellular biosensor detects biologically active metagenomic gene products. GFP expression is dependent on the interaction of those small molecules with the biosensor proteins. Finally, FACS is used to sort the GFP+ and GFP- cells separately.

2.5.3. Microarrays

The DNA microarray technology is an alternative tool with the potential to investigate simultaneously multiple genes. Functional gene arrays (FGAs) are microarrays that contain probes for key genes involved in microbial functional processes, such as biodegradation of environmental contaminants or biogeochemical cycling of C, N, S, P and metals (He & Deng, 2012). The development of a FGA implies a certain workflow: the selection of functional genes, the retrieval and verification of all the sequences from public databases; the design of oligonucleotide probes, usually with specific software tools; finally the "in-house" or commercially microarray construction spotting the probes onto glass slides (He & Deng, 2012). Distinct FGAs were developed immobilizing PCR amplicons or oligonucleotides to target specific functional processes. For example, FGAs of 50-mer oligonucleotides were developed to monitor microbial communities in acidic environment and bioleaching systems (Yin et al., 2007). Yin et al. investigated acid mine drainage environments and they designed the microarray containing 1071 probes using 16S rRNA sequences and a diverse set of functional genes involved in carbon, nitrogen, sulphur and iron metabolisms, metal resistance. Acidophilic microorganisms known to be in the acidic environments were used as keywords for identifying appropriate genes in the GenBank database to use as probe (*A. ferrooxidans, Leptospirillum* sp*., Acidiphilum* spp.)*. The established microarray was used as a generic profiling tool that revealed differences among various microbial communities. A total of 80 16S rRNA genes probes and 150 functional genes probes were detected, proving that the 50-mer oligonucleotide array can be used as a specific and quantitative tool.

Park et al. (2008) developed another format of DNA-microarray technology for rapidly identifying clones from metagenome libraries: fosmid clones (targets) were spotted on a slide and specific gene probes were labelled and used for hybridization. This approach was used to investigate fosmid libraries obtained from marine sediments (Park et al., 2008). However, specificity, high hybridization efficiency and detection sensitivity are the most critical parameters of this technique, and the target sequences extracted from the conserved regions of already known protein families radically reduce the chances of obtaining new proteins (Gabor et al., 2007).

2.5.4. The way forward – from enzyme discovery to the preparation of ready-to-use biocatalysts

It is clear that the establishment of enzymatic processes in the industry is mainly a problem of finding new enzymes. As mentioned above, one can take advantage of sequencing and extensive screening technologies to develop enzyme discovery strategies and to identify microbial enzymes with

improved and unusual activities and specificities. However, novel and potentially more efficient enzymes obtained from metagenomics are likely to require further modification by protein or enzyme engineering to obtain increased enzyme properties and catalytic activities. In addition to that, enzymes are delicate materials that need to be stabilized to survive a range of challenges conditions typically used in industrial processes. Accordingly, new formulations are needed that can stabilise and protect enzymes from adverse conditions, including those originated from washing, leaching and solvent attack. Here, distinct combinatorial and rational methods to generate such stabilized biocatalysts are required, regardless of the apparent relative improvements at the end of the process (Tran & Balkus, 2011; Bommarius et al., 2013). The material may not be just an inert component that allow the reuse of the enzyme but rather may play a role in the enzyme activity and stability.

Nanotechnology has emerged as a promising tool to overcome problems of using enzymes for industrial purposes, providing improved stability, higher activity, protection against protease attack and minimising solubility-related issues (Kim et al., 2010). Various nanobiocatalytic approaches, using nanostructured materials, such as magnetic nanoparticles, polymers nanofibers and nanoporous materials, have been successfully used and tested as nanoscale reactors (Lei et al., 2002; Ge et al., 2009). Uniform and well-controlled nanopores provide a large pore volume and high surface area, which can be used for improved enzyme and peptides loading. The covalent bounds between polymer and enzyme, in addition to the enzyme crosslinking via glutaraldehyde (GA) treatment, results in an aggregate coating with exceptional success in enzyme stabilization (Singh, 2010). Moreover, this system is also resistant to protease activity as the nanoporous media prevents the enzyme aggregate from leaching out and from being attacked by proteases in the fermenter (Kim et al., 2010), which enable it to be reused. Other outstanding example include the enzyme immobilization and stabilization in superparamagnetic silica/iron oxide nanocomposites with structured porosity (Valdés-Solís et al., 2009), maghemite (γ-Fe2O3) nanoparticles functionalized with a reactive mulfunctional polymer (Shukoor et al., 2008) and dendronized manoparticles (Gustafsson et al., 2015), to cite some.

Adsorption of enzymes on nanomaterials can improve enzyme activity and thermal stability. Magnetic nanoparticles, with an iron oxide core, covered by different types of polymers, allow an easy recovery of immobilized proteins. Xylanase from *Aspergillus niger* immobilized onto Fe_3O_4-coated chitosan magnetic nanoparticles showed increased thermal stability compared to the free enzyme and retained 87.5 % activity after seven successive reactions by magnetic separation (Liu et al., 2014).

2.6. Conclusions and Research Needs

The demand for industrial enzymes is on a continuous rise driven by a growing need for sustainable and "greener" economy. Microorganisms play part as the largest source of an extraordinary amount of biocatalysts with a potential wide range of applications across several industries. Although numerous microbes inhabit the biosphere, the majority of them are reluctant to be cultivated under standard laboratory condition. Recent developments in metagenomics, in combination with other tools such as proteomics and recombinant DNA techniques for gene synthesis, have facilitated the discovery of new microbial enzymes from nature and their evolution to variants with improved catalytic properties.

Functional metagenomics has proven to be highly successful for mining the enormous microbial diversity and thus providing biomolecules that merge industrial criteria. However, the refinement of methods in metagenomics will be crucial for further success in areas of industrial biocatalysis. The identification of obstacles and solving problems to cloning and screening will aid in the development of new tools and technologies for functional metagenomics. Improving *E. coli* (and other bacteria, fungi and archaea) as screening host and exploring the potential of different organisms will help overcoming limits at the level of transcription and translation, and likely improve future hits rates. High-throughput techniques for the screening of millions of clones and for the further evaluation of enzyme performance, will allow identifying highly active, efficient and promiscuous biocatalysts.

Once a new or improved activity has been successfully identified by metagenomics, the process of integrating them at industrial scale may still be interfered if the expression of the pure proteins does not provide sufficient amounts of enzyme at reasonable costs. Further, it is recognized that the final catalytic properties of a biocatalyst are mainly the result of substrate arrangement in its active site driven by its natural or evolved polypeptide sequence and structure, and the influence that the immobilization process, needed to reuse and stabilize the enzymes, additionally exerts. For this reason, when starting a metagenomics investigation all these factors need to be taken into consideration.

2.7. References

Adrio, J.L. & Demain, A.L. (2014). Microbial enzymes: tools for biotechnological processes. *Biomolecules*. 4 (1). p.pp. 117–39.

Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.-C., Marquez, M., Klibanov, A.M., Griffiths, A.D. & Weitz, D.A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences*. 107 (9). p.pp. 4004–4009.

Araújo, R., Casal, M. & Cavaco-Paulo, A. (2008). Application of enzymes for textile fibres processing. *Biocatalysis and Biotransformation*. 26 (5). p.pp. 332–349.

Atomi, H., Sato, T. & Kanai, T. (2011). Application of hyperthermophiles and their enzymes. *Current Opinion in Biotechnology*. 22 (5). p.pp. 618–626.

Bargiela, R., Gertler, C., Magagnini, M., Mapelli, F., Chen, J., Daffonchio, D., Golyshin, P.N. & Ferrer, M. (2015). Degradation Network Reconstruction in Uric Acid and Ammonium Amendments in Oil-Degrading Marine Microcosms Guided by Metagenomic Data. *Frontiers in microbiology*. 6. p.p. 1270.

Béjà, O., Aravind, L., Koonin, E. V, Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., DeLong, E.F., Baldwin, J.M., Schertler, G.F., Unger, V.M., Oesterhelt, D., Stoeckenius, W., Hoff, W.D., Jung, K.H., Spudich, J.L., Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., Henderson, R., Lanyi, J.K., Bieszke, J.A., Spudich, E.N., Scott, K.L., Borkovich, K.A., Spudich, J.L., Mullins, T.D., Britcshgi, T.B., Krest, R.L., Giovannoni, S.J., Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R., Koonin, E. V., Doolittle, W.F., Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P., Lanyi, J.K., Sippl, M.J., Guex, N., Peitsch, M.C., Heberle, J., Felsenstein, J., Mukohata, Y., Ihara, K., Tamura, T., Sugiyama, Y., Elish, M.E., Pierce, J.R., Earhart, C.F., Shimono, K., Iwamoto, M., Sumi, M., Kamo, N., Stoeckenius, W., Lozier, R.H., Bogomolni, R.A., Kaback, H.R., Prossnitz, E., Gee, A., Ames, G.F., Robertson, D.E., Kaczorowski, G.J., Garcia, M.L., Kaback, H.R., Yao, V.J., Spudich, J.L., Pace, N.R., Fuhrman, J.A., McCallum, K., Davis, A.A., Schmidt, T.M., DeLong, E.F., Pace, N.R., Rappe, M.S., Kemp, P.F., Giovannoni, S.J., Acinas, S.G., Anton, J., Rodriguez-Valera, F., Eilers, H., Pernthaler, J., Glockner, F.O., Amann, R., Sasaki, J. & Spudich, J.L. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science (New York, N.Y.)*. 289 (5486). p.pp. 1902–6.

Bell, P.J.L., Sunna, A., Curach, N.C., Bergquist, P.L., Gibbs, M.D. & Nevalainen, H. (2002). Prospecting for novel lipase genes using PCR a. *Microbiology*. 148 (8). p.pp. 2283–2291.

Bérdy, J. (2005). Bioactive Microbial Metabolites. *The Journal of Antibiotics*. 58 (1). p.pp. 1–26.

Bergquist, P.L., Hardiman, E.M., Ferrari, B.C. & Winsley, T. (2009). Applications of flow cytometry in environmental microbiology and biotechnology. *Extremophiles*. 13 (3) p.pp. 389–401.

Bommarius, A.S., Paye, M.F., Bornscheuer, U.T., Pohl, M., Lutz, S., Ma, S.K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G.W., Sheldon, R.A., Martinez, C.A., Hu, S., Dumond, Y., Tao, J., Kelleher, P., Tully, L., Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., Hughes, G.J., Nakamura, C.E., Whited, G.M., Polizzi, K.M., Bommarius, A.S., Broering, J.M., Chaparro-Riggers, J.F., Bommarius, S., Broering, J.M., Eijsink, V.G.H., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., Burg, B. van den, Vriend, G., Eijsink, V.G.H., Gaseidnes, S., Borchert, T. V., Burg, B. van den, Winkler, C.K., Clay, D., Davies, S., O'Neill, P., McDaid, P., Debarge, S., Steflik, J.,

Karmilowicz, M., Wong, J.W., Faber, K., Breithaupt, C., Kurzbauer, R., Schaller, F., Stintzi, A., Schaller, A., Huber, R., Macheroux, P., Clausen, T., Lepore, B.W., Liu, D., Peng, Y., Fu, M., Yasuda, C., Manning, J.M., Silverman, R.B., Ringe, D., Winn, M., Foulkes, J.M., Perni, S., Simmons, M.J.H., Overton, T.W., Goss, R.J.M., Rosche, B., Li, X.Z., Hauer, B., Schmid, A., Buehler, K., Hudlicky, T., Reed, J.W., Schrewe, M., Julsing, M.K., Buhler, B., Schmid, A., Leresche, J.E., Meyer, H.-P., Littlechild, J.A., Guy, J., Connelly, S., Mallett, L., Waddell, S., Rye, C.A., Line, K., Isupov, M., Secundo, F., Reetz, M.T., Wu, S., Newhouse, T., Baran, P.S., Hoffmann, R.W., Wang, M., Si, T., Zhao, H., Wistuba, D., Schurig, V., Wang, M., Roberts, D.L., Paschke, R., Shea, T.M., Masters, B.S.S., Kim, J.-J.P., Yi, D., Devamani, T., Abdoul-Zabar, J., Charmantray, F., Helaine, V., Hecquet, L., Fessner, W.-D., Asztalos, P., Parthier, C., Golbik, R., Kleinschmidt, M., Hübner, G., Weiss, M.S., Friedemann, R., Wille, G., Tittmann, K., Goo, K.S., Sim, T.S., Roach, P.L., Clifton, I.J., Hensgens, C.M.H., Shibata, N., Schofield, C.J., Hajdu, J., Baldwin, J.E., Weber, S.S., Bovenberg, R.A.L., Driessen, A.J.M., Bloom, J.D., Arnold, F.H., Bommarius, A.S., Blum, J.K., Abrahamson, M.J., Hernandez, K., Fernandez-Lafuente, R., Foo, J.L., Ching, C.B., Chang, M.W., Leong, S.S.J., Robic, S., Burioni, R., Cassi, D., Cecconi, F., Vulpiani, A., Li, W.F., Zhou, X.X., Lu, P., Koudelakova, T., Chaloupkova, R., Brezovsky, J., Prokop, Z., Sebestova, E., Hesseler, M., Khabiri, M., Plevaka, M., Kulik, D., Smatanova, I.K., Rezacova, P., Ettrich, R., Bornscheuer, U.T., Damborsky, J., Koudelakova, T., Bidmanova, S., Dvorak, P., Pavelka, A., Chaloupkova, R., Prokop, Z., Damborsky, J., Blum, J.K., Bommarius, A.S., Blum, J.K., Ricketts, M.D., Bommarius, A.S., Harano, Y., Singh, R., Tiwari, M., Singh, R., Lee, J.-K., Barends, T.R.M., Polderman-Tijmes, J.J., Jekel, P.A., Hensgens, C.M.H., Vries, E.J. de, Janssen, D.B., Dijkstra, B.W., Simone, G. De, Menchise, V., Manco, G., Mandrich, L., Sorrentino, N., Lang, D., Rossi, M., Pedone, C., Cipolla, A., Delbrassine, F., Lage, J.-L. Da, Feller, G., Camps, M., Herman, A., Loh, E., Loeb, L.A., Bershtein, S., Goldin, K., Tawfik, D.S., Bershtein, S., Tawfik, D.S., Peisajovich, S.G., Tawfik, D.S., Barros, D.P.C. de, Fernandes, P., Cabral, J.M.S., Fonseca, L.P., Rogers, T.A., Bommarius, A.S., Johnson, P., Whateley, T.L., Niesen, F.H., Berglund, H., Vedadi, M., Senisterra, G.A., Finerty, P.J., Rodrigues, J. V., Prosinecki, V., Marrucho, I., Rebelo, L.P.N., Gomes, C.M., Levenspiel, O., Jenkins, R.O., Rogers, T.A., Daniel, R.M., Bommarius, A.S., Daniel, R.M., Danson, M.J., Eisenthal, R., Peterson, M.E., Eisenthal, R., Danson, M.J., Spence, A., Daniel, R.M., Ema, T., Kageyama, M., Korenaga, T., Sakai, T., End, N., Schoning, K.U., Zhang, B.H., Weng, Y.Q., Xu, H., Mao, Z.P., Cowan, D.A., Fernandez-Lafuente, R., Grazú, V., López-Gallego, F., Montes, T., Abian, O., González, R., Hermoso, J.A., García, J.L., Mateo, C., Guisán, J.M., Cecchini, D.A., Pavesi, R., Sanna, S., Daly, S., Xaiz, R., Pregnolato, M., Terreni, M., Brun, N., Garcia, A.B., Deleuze, H., Achard, M.F., Sanchez, C., Durand, F., Oestreicher, V., Backov, R., Backov, R., Kim, H.S., Lee, S.J., Lee, E.Y., Fishman, A., Levy, I., Cogan, U., Shoseyov, O., Asuri, P., Karajanagi, S.S., Yang, H.C., Yim, T.J., Kane, R.S., Dordick, J.S., Asuri, P., Karajanagi, S.S., Vertegel, A.A., Dordick, J.S., Kane, R.S., Wilson, L., Illanes, A., Pessela, B.C.C., Abian, O., Fernandez-Lafuente, R., Guisan, J.M., Bolivar, J.M., Wilson, L., Ferrarotti, S.A., Fernandez-Lafuente, R., Guisan, J.M., Mateo, C., Setschenow, J.Z., Wong, C.H., Drueckhammer, D.G., Sweers, H.M., Bommarius, A.S., Karau, A., Hofmeister, F., Peter, H.V.H., Wong, K.-Y., Sinanogl, O., Abdulnur, S., Grigorjev, P.A., Bezrukov, S.M., Melander, W., Horvath, C., Collins, K.D., Collins, K.D., Collins, K.D., Collins, K.D., Washabaugh, M.W., Jones, G., Dole, M., Broering, J.M., Bommarius, A.S., Broering, J.M., Bommarius, A.S., Collins, K.D., Washabaugh, M.W., Collins, K.D., Lumry, R., Eyring, H., Broering, J.M., Bommarius, A.S., Yeh, V., Broering, J.M., Romanyuk, A., Chen, B.X., Chernoff, Y.O., Bommarius, A.S., Hall, M., Rubin, J., Behrens, S.H., Bommarius, A.S., Kaushik, J.K., Bhat, R., Boerema, D.J., Tereshko, V.A., Kent, S.B.H., Nasiripourdori, A., Naderi-Manesh, H., Ranjbar, B., Khajeh, K., Sathish, H.A., Kumar, P.R., Prakash, V., Tiwari, A., Bhat, R., Arroyo, M., Torres-Guzman, R., Mata, I. de la, Castillon, M.P., Acebal, C., Arroyo, M., Torres-Guzman, R., Mata, I. de la, Castillon, M.P., Acebal, C., Suresh, C.G., Pundle, A. V., SivaRaman, H., Rao, K.N., Brannigan, J.A., McVey, C.E., Verma, C.S., Dauter, Z., Dodson, E.J., Dodson, G.G., Estell, D.A., Graycar, T.P., Miller, J. V., Powers, D.B., Burnier, J.P., Ng, P.G., Wells, J.A., Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P., Estell, D.A., Colletier, J.P., Aleksandrov, A., Coquelle, N.,

Mraihi, S., Mendoza-Barbera, E., Field, M., Madern, D., Soskine, M., Tawfik, D.S., Arnold, F.H., Wintrode, P.L., Miyazaki, K., Gershenson, A., Schoemaker, H.E., Mink, D., Wubbolts, M.G., Amara, A.A.-A.-F., Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S.K., Gall, M., Bottcher, D., Joosten, H.J., Bornscheuer, U.T., Saven, J.G., Kang, S.G., Saven, J.G., Lorenzo, M. Di, Hidalgo, A., Molina, R., Hermoso, J.A., Pirozzi, D., Bornscheuer, U.T., Cesarini, S., Bofill, C., Pastor, F.I.J., Reetz, M.T., Diaz, P., Opperman, D.J., Reetz, M.T., Park, J.T., Hirano, J.-I., Thangavel, V., Riebel, B.R., Bommarius, A.S., Katzberg, M., Skorupa-Parachin, N., Gorwa-Grauslund, M.-F., Bertau, M., Ito, Y., Ikeuchi, A., Imamura, C., Park, H., Joo, J., Park, K., Yoo, Y., Qian, Z., Horton, J.R., Cheng, X., Lutz, S., Badoei-Dalfard, A., Khajeh, K., Asghari, S.M., Ranjbar, B., Karbalaei-Heidari, H.R., Liang, J., Lalonde, J., Borup, B., Mitchell, V., Mundorff, E., Trinh, N., Kochrekar, D.A., Cherat, R.N., Pai, G.G., Kim, M.-S., Lei, X., Gonzalez-Blasco, G., Sanz-Aparicio, J., Gonzalez, B., Hermoso, J.A., Polaina, J., Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X.Q., Tomlinson, G.L., Verruto, J., Wong, V.W.K., Mathur, E.J., Short, J.M., Robertson, D.E., Steer, B.A., Chaparro-Riggers, J.F., Polizzi, K.M., Bommarius, A.S., Rathi, P.C., Radestock, S., Gohlke, H., Radestock, S., Gohlke, H., Hirose, S., Kawamura, Y., Mori, M., Yokota, K., Noguchi, T., Goshima, N., Synstad, B., Gaseidnes, S., Aalten, D.M.F. van, Vriend, G., Nielsen, J.E., Eijsink, V.G.H., Macdonald, J.M., Tarling, C.A., Taylor, E.J., Dennis, R.J., Myers, D.S., Knapp, S., Davies, G.J., Withers, S.G., Tews, I., vanScheltinga, A.C.T., Perrakis, A., Wilson, K.S., Dijkstra, B.W., Gao, D., Narasimhan, D.L., Macdonald, J., Brim, R., Ko, M.-C., Landry, D.W., Woods, J.H., Sunahara, R.K., Zhan, C.-G., Turner, J.M., Larsen, N.A., Basran, A., Barbas, C.F., Bruce, N.C., Wilson, I.A., Lerner, R.A., Larsen, N.A., Turner, J.M., Stevens, J., Rosser, S.J., Basran, A., Lerner, R.A., Bruce, N.C., Wilson, I.A., Rogers, C.J., Eubanks, L.M., Dickerson, T.J., Janda, K.D., Zheng, F., Zhan, C.-G., Huang, X., Gao, D., Zhan, C.-G., Wei, Y., Swenson, L., Castro, C., Derewenda, U., Minor, W., Arai, H., Aoki, J., Inoue, K., Servin-Gonzalez, L., Derewenda, Z.S., Pikkemaat, M.G., Linssen, A.B.M., Berendsen, H.J.C., Janssen, D.B., Feng, X.J., Sanchis, J., Reetz, M.T., Rabitz, H., Reetz, M.T., Soni, P., Acevedo, J.P., Sanchis, J., Augustyniak, W., Brzezinska, A.A., Pijning, T., Wienk, H., Boelens, R., Dijkstra, B.W., Reetz, M.T., Jochens, H., Aerts, D., Bornscheuer, U.T., Kawasaki, K., Kondo, H., Suzuki, M., Ohgiya, S., Tsuda, S., Gumulya, Y., Reetz, M.T., Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., Loon, A. van, Wyss, M., Lehmann, M., Pasamontes, L., Lassen, S.F., Wyss, M., Lehmann, M., Wyss, M., Amin, N., Liu, A.D., Ramer, S., Aehle, W., Meijer, D., Metin, M., Wong, S., Gualfetti, P., Schellenberger, V., Watanabe, K., Ohkuri, T., Yokobori, S., Yamagishi, A., Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., Pluckthun, A., DiTursi, M.K., Kwon, S.J., Reeder, P.J., Dordick, J.S., Polizzi, K.M., Chaparro-Riggers, J.F., Vázquez-Figueroa, E., Bommarius, A.S., McVey, C.E., Walsh, M.A., Dodson, G.G., Wilson, K.S., Brannigan, J.A., Steipe, B., Schiller, B., Pluckthun, A., Steinbacher, S., Vazquez-Figueroa, E., Chaparro-Riggers, J., Bommarius, A.S., Gomes, J., Steiner, W., Podar, M., Reysenbach, A.L., Reetz, M.T., Carballeira, J.D., Vogel, A., Gunasekaran, K., Ramakrishnan, C., Balaram, P., Serov, A.E., Odintzeva, E.R., Uporov, I. V., Tishkov, V.I., Besenmatter, W., Kast, P., Hilvert, D., Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H., Riechmann, L., Winter, G., Gilbert, W., Doolittle, W.F., Blake, C.C.F., Bogarad, L.D., Deem, M.W., Silberg, J.J., Endelman, J.B., Arnold, F.H., Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L., Arnold, F.H., Yuan, L., Kurek, I., English, J., Keenan, R., Clarke, N.D., Heinzelman, P., Snow, C.D., Smith, M.A., Yu, X.L., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., Arnold, F.H., Heinzelman, P., Snow, C.D., Wu, I., Nguyen, C., Villalobos, A., Govindarajan, S., Minshull, J., Arnold, F.H., Romero, P.A., Stone, E., Lamb, C., Chantranupong, L., Krause, A., Miklos, A.E., Hughes, R.A., Fechtel, B., Ellington, A.D., Arnold, F.H., Georgiou, G., Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C., Arnold, F., Smith, M.A., Rentmeister, A., Snow, C.D., Wu, T., Farrow, M.F., Mingardon, F., Arnold, F.H., Parsiegla, G., Reverbel-Leroy, C., Tardif, C., Belaich, J.P., Driguez, H., Haser, R., Määttä, J.A.E., Eisenberg-Domovich, Y., Nordlund, H.R., Hayouka, R., Kulomaa, M.S., Livnah, O., Hytönen, V.P., Repo, S., Paldanius, T.A., Hytönen, V.P., Nyholm, T.K.M., Halling, K.K., Huuskonen, J., Pentikäinen, O.T., Rissanen, K., Slotte, J.P., Airenne, T.T., Salminen, T.A., Kulomaa, M.S., Johnson, M.S., Baugh,

L., Trong, I. Le, Cerutti, D.S., Mehta, N., Gülich, S., Stayton, P.S., Stenkamp, R.E., Lybrand, T.P., Livnah, O., Bayer, E.A., Wilchek, M., Sussman, J.L., Otey, C.R., Landwehr, M., Endelman, J.B., Hiraga, K., Bloom, J.D., Arnold, F.H., Cerdobbel, A., Winter, K. De, Aerts, D., Kuipers, R., Joosten, H.-J., Soetaert, W., Desmet, T., Osorio, N.M., Fonseca, M.M.R. da, Ferreira-Dias, S., Cowan, D.A., Owusu, R.K., Cowan, D.A., Hao, J.J., Berry, A., Reetz, M.T., Soni, P., Fernandez, L., Gumulya, Y., Carballeira, J.D., Adalbjörnsson, B. V., Toogood, H.S., Fryszkowska, A., Pudney, C.R., Jowitt, T.A., Leys, D., Scrutton, N.S., Brenna, E., Gatti, F.G., Monti, D., Parmeggiani, F., Serra, S., Burg, B. Van den, Kreij, A. de, Veek, P. Van der, Mansfeld, J., Venema, G., Mansfeld, J., Ulbrich-Hofmann, R., Sharma, S., Mittal, A., Gupta, V.K., Singh, H., Slusarczyk, H., Felber, S., Kula, M.R., Pohl, M., Vazquez-Figueroa, E., Yeh, V., Broering, J.M., Chaparro-Riggers, J.F. & Bommarius, A.S. (2013). Stabilizing biocatalysts. *Chemical Society Reviews*. 42 (15). p.p. 6534.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C. & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*. 485 (7397). p.pp. 185–194.

Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J.B., Rothberg, J.M., Link, D.R., Perrimon, N. & Samuels, M.L. (2009). Droplet microfluidic technology for single-cell high-throughput screening. *Proceedings of the National Academy of Sciences*. 106 (34). p.pp. 14195–14200.

Brzostowicz, P.C., Walters, D.M., Thomas, S.M., Nagarajan, V. & Rouvière, P.E. (2003). mRNA differential display in a microbial enrichment culture: simultaneous identification of three cyclohexanone monooxygenases from three species. *Applied and environmental microbiology*. 69 (1). p.pp. 334–42.

Van den Burg, B. (2003). Extremophiles as a source for novel enzymes. *Current Opinion in Microbiology*. 6 (3) p.pp. 213–218.

de Carvalho, C.C.C.R. (2011). Enzymatic and whole cell catalysis: Finding new strategies for old processes. *Biotechnology Advances*. 29 (1). p.pp. 75–83.

de Carvalho, C.C.C.R. (2016). Whole cell biocatalysts: essential workers from Nature to the industry. *Microbial Biotechnology*.

Chen, I.-C., Thiruvengadam, V., Lin, W.-D., Chang, H.-H. & Hsu, W.-H. (2010). Lysine racemase: a novel non-antibiotic selectable marker for plant transformation. *Plant Molecular Biology*. 72 (1–2). p.pp. 153–169.

Cheng, J., Pinnell, L., Engel, K., Neufeld, J.D. & Charles, T.C. (2014). Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *Journal of Microbiological Methods*. 99. p.pp. 27–34.

Cherry, J.R. & Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. *Current Opinion in Biotechnology*. 14 (4). p.pp. 438–443.

Coll-Lladó, M., Acinas, S.G., Pujades, C. & Pedrós-Alió, C. (2011). Transcriptome fingerprinting analysis: an approach to explore gene expression patterns in marine microbial communities. *PloS one*. 6 (8). p.p. e22950.

Craig, J.W., Chang, F.-Y., Kim, J.H., Obiajulu, S.C. & Brady, S.F. (2010). Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and environmental microbiology*. 76 (5). p.pp. 1633–41.

Curtis, T.P. & Sloan, W.T. (2005). Microbiology. Exploring microbial diversity--a vast below. *Science (New York, N.Y.)*. 309 (5739). p.pp. 1331–1333.

Davis, B.G. & Boyer, V. (2001). Biocatalysis and enzymes in organic synthesis. *Natural product reports*. 18 (6). p.pp. 618–40.

Demain, A.L. & Adrio, J.L. (2008). Contributions of microorganisms to industrial biology. *Molecular biotechnology*. 38 (1). p.pp. 41–55.

Fernández-Arrojo, L., Guazzaroni, M.-E., López-Cortés, N., Beloqui, A. & Ferrer, M. (2010). Metagenomic era for biocatalyst identification. *Current opinion in biotechnology*. 21 (6). p.pp. 725–33.

Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V. & Golyshin, P.N. (2015). Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatalysis and Biotransformation*. 33 (5–6). p.pp. 235–249.

Ferrer, M., Beloqui, A. & Golyshin, P.N. (2010). Screening metagenomic libraries for laccase activities. *Methods in molecular biology (Clifton, N.J.)*. 668. p.pp. 189–202.

Ferrer, M., Beloqui, A., Timmis, K.N. & Golyshin, P.N. (2009). Metagenomics for mining new genetic resources of microbial communities. *Journal of molecular microbiology and biotechnology*. 16 (1–2). p.pp. 109–23.

Ferrer, M., Golyshina, O., Beloqui, A. & Golyshin, P.N. (2007). Mining enzymes from extreme environments. *Current opinion in microbiology*. 10 (3). p.pp. 207–14.

Ferrer, M., Golyshina, O. V, Chernikova, T.N., Khachane, A.N., Martins dos Santos, V.A.P., Yakimov, M.M., Timmis, K.N. & Golyshin, P.N. (2005a). Microbial Enzymes Mined from the Urania Deep-Sea Hypersaline Anoxic Basin. *Chemistry & Biology*. 12 (8). p.pp. 895–904.

Ferrer, M., Golyshina, O. V, Chernikova, T.N., Khachane, A.N., Reyes-Duarte, D., Santos, V.A.P.M. Dos, Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N. & Golyshin, P.N. (2005b). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology*. 7 (12). p.pp. 1996–2010.

Ferrer, M., Martínez-Abarca, F. & Golyshin, P.N. (2005c). Mining genomes and 'metagenomes' for novel catalysts. *Current opinion in biotechnology*. 16 (6). p.pp. 588–93.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O. V. & Golyshin, P.N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology*. 9 (1). p.pp. 22–34.

Gabor, E., Liebeton, K., Niehaus, F., Eck, J. & Lorenz, P. (2007). Updating the metagenomics toolbox. *Biotechnology journal*. 2 (2). p.pp. 201–6.

Gabor, E.M., de Vries, E.J. & Janssen, D.B. (2004). Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology*. 6 (9). p.pp. 948–958.

Galvão, T.C., Mohn, W.W. & de Lorenzo, V. (2005). Exploring the microbial biodegradation and biotransformation gene pool. *Trends in Biotechnology*. 23 (10). p.pp. 497–506.

Ge, J., Lu, D., Liu, Z. & Liu, Z. (2009). Recent advances in nanostructured biocatalysts. *Biochemical Engineering Journal*. 44 (1). p.pp. 53–59.

Goll, J., Rusch, D.B., Tanenbaum, D.M., Thiagarajan, M., Li, K., Methé, B.A. & Yooseph, S. (2010). METAREP: JCVI metagenomics reports--an open source tool for high-performance comparative metagenomics. *Bioinformatics (Oxford, England)*. 26 (20). p.pp. 2631–2.

Golyshina, O. V., Lünsdorf, H., Kublanov, I. V., Goldenstein, N.I., Hinrichs, K.U. & Golyshin, P.N. (2016). The novel extremely acidophilic, cell-wall-deficient archaeon Cuniculiplasma divulgatum gen. Nov., sp. nov. represents a new family, Cuniculiplasmataceae fam. nov., of the order Thermoplasmatales. *International Journal of Systematic and Evolutionary Microbiology*. 66 (1). p.pp. 332–340.

Gomes, I., Gomes, J. & Steiner, W. (2003). Highly thermostable amylase and pullulanase of the extreme thermophilic eubacterium Rhodothermus marinus: production and partial characterization. *Bioresource Technology*. 90 (2). p.pp. 207–214.

Gomes, J. & Steiner, W. (2004). The Biocatalytic Potential of Extremophiles and Extremozymes. *Food Technology and Biotechnology*. 42 (4). p.pp. 223–225.

Gong, J.-S., Lu, Z.-M., Li, H., Zhou, Z.-M., Shi, J.-S. & Xu, Z.-H. (2013). Metagenomic technology and genome mining: emerging areas for exploring novel nitrilases. *Applied microbiology and biotechnology*. 97 (15). p.pp. 6603–11.

Guo, Z. & Xu, X. (2005). New opportunity for enzymatic modification of fats and oils with industrial potentials. *Organic & biomolecular chemistry*. 3 (14). p.pp. 2615–9.

Gupta, R., Beg, Q. & Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*. 59 (1). p.pp. 15–32.

Gupta, R., Gupta, N. & Rathi, P. (2004). Bacterial lipases: An overview of production, purification and biochemical properties. *Applied Microbiology and Biotechnology*. 64 (6) p.pp. 763–781.

Gurung, N., Ray, S., Bose, S. & Rai, V. (2013). A broader view: Microbial enzymes and their relevance in industries, medicine, and beyond. *BioMed Research International*. 2013 p.pp. 329121.

Gustafsson, H., Küchler, A., Holmberg, K., Walde, P., Mateo, C., Palomo, J.M., Fernandez-Lorente, G., Guisan, J.M., Fernandez-Lafuente, R., Hudson, S., Cooney, J., Magner, E., Brady, D., Jordaan, J., Fernandez-Lafuente, R., Garcia-Galan, C., Berenguer-Murcia, Á., Fernandez-Lafuente, R., Rodrigues, R.C., Tran, D.N., Balkus, K.J., Magner, E., Sheldon, R.A., Pelt, S. van, Cantone, S., Ferrario, V., Corici, L., Ebert, C., Fattor, D., Spizzo, P., Gardossi, L., DiCosimo, R., McAuliffe, J., Poulose, A.J., Bohlmann, G., Schoffelen, S., Hest, J.C.M. van, Costantini, F., Tiggelaar, R., Sennato, S., Mura, F., Schlautmann, S., Bordi, F., Gardeniers, H., Manetti, C., Burchardt, M., Wittstock, G., Jia, F., Narasimhan, B., Mallapragada, S., Kang, W., Liu, J., Wang, J., Nie, Y., Guo, Z., Xia, J., Mayer, S.F., Kroutil, W., Faber, K., Betancor, L., Luckarift, H.R., Kim, S.-H., Lee, S.-M., Kim, D.-U., Cui, J.-Z., Kang, S.-W., Gustafsson, H., Johansson, E.M., Barrabino, A., Odén, M., Holmberg, K., Gustafsson, H., Thörn, C., Holmberg, K., Thörn, C., Carlsson, N., Gustafsson, H., Holmberg, K., Åkerman, B., Olsson, L., Thörn, C., Gustafsson, H., Olsson, L., Thörn, C., Gustafsson, H., Olsson, L., Lee, C.-H., Lin, T.-S., Mou, C.-Y., Hartmann, M., Jung, D., Hartmann, M., Kostrov, X., Carlsson, N., Gustafsson, H., Thörn, C., Olsson, L., Holmberg, K., Åkerman, B., Küchler, A., Adamcik, J., Mezzenga, R., Schlüter, A.D., Walde, P., Fornera, S., Bauer, T., Schlüter, A.D., Walde, P., Guo, Y., Beek, J.D. van, Zhang, B., Colussi, M., Walde, P., Zhang, A., Kröger, M., Halperin, A., Schlüter, A.D., Grotzky, A., Nauser, T., Erdogan, H., Schlüter, A.D., Walde, P., Onda, M., Ariga, K., Kunitake, T., Pescador, P., Katakis, I., Toca-Herrera, J.L., Donath, E., Keighron, J.D., Keating, C.D., Guedidi, S., Yurekli, Y., Deratani, A., Déjardin, P., Innocent, C., Altinkaya, S.A., Roudesli, S., Yemenicioglu, A., Sakr, O.S., Borchard, G., Palazzo, G., Colafemmina, G., Iudice, C.G., Mallardi, A., Solomon, B., Lotan, N., Katchalski-Katzir, E., Swoboda, B.E.P., Massey, V., Aibara, S., Yamashita, H., Mori, E., Kato, M., Morita, Y., Nandiyanto, A.B.D., Kim, S.-G., Iskandar, F., Okuyama, K., Grotzky, A., Altamura, E., Adamcik, J., Carrara, P., Stano, P., Mavelli, F., Nauser, T., Mezzenga, R., Schlüter, A.D., Walde, P., Dirksen, A., Dawson, P.E., Rodahl, M., Höök, F., Krozer, A., Brzezinski, P., Kasemo, B., Childs, R.E., Bardsley, W.G., Brown, K.C., Corbett, J.F., Loveless, N.P., Fornera, S., Walde, P., Rosenholm, J.M., Lindén, M., Lavery, C.B., MacInnis, M.C., MacDonald, M.J., Williams, J.B.,

Spencer, C.A., Burke, A.A., Irwin, D.J.G., D'Cunha, G.B., Pazur, J.H., Kleppe, K., Wang, Y., Caruso, F., Wang, Y., Caruso, F., Li, J., Yin, G., Ding, Y., Liao, X., Chen, X., Huang, Z., Yao, Y., Pu, X., Kandel, K., Althaus, S.M., Pruski, M., Slowing, I.I., Zhang, X., Guan, R.-F., Wu, D.-Q., Chan, K.-Y., Lee, D., Lee, J., Kim, J., Kim, J., Na, H.B., Kim, B., Shin, C.H., Kwak, J.H., Dohnalkova, A., Grate, J.W., Hyeon, T., Kim, H.S., Lei, C., Shin, Y., Magnuson, J.K., Fryxell, G., Lasure, L.L., Elliott, D.C., Liu, J., Ackerman, E.J., Dai, Z.H., Ni, J., Huang, X.H., Lu, G.F., Bao, J.C., Dai, Z., Bao, J., Yang, X., Ju, H., Zhou, Z., Hartmann, M., Luz, R.A.S., Pereira, A.R., Souza, J.C.P. de, Sales, F.C.P.F. & Crespilho, F.N. (2015). Co-immobilization of enzymes with the help of a dendronized polymer and mesoporous silica nanoparticles. *J. Mater. Chem. B*. 3 (30). p.pp. 6174–6184.

Hallin, P.F., Binnewies, T.T., Ussery, D.W., Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, J., Sutton, G., FitzHugh, W., Fields, C., Gocyne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K. V., Fraser, C.M., Smith, H.O., Venter, J.C., Jensen, L.J., Skovgaard, M., Sicheritz-Ponten, T., Jorgensen, M.K., Lundegaard, C., Pedersen, C.C., Petersen, N., Ussery, D., Hallin, P.F., Binnewies, T.T., Ussery, D.W., Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., Sebaihia, M., Peck, M.W., Minton, N.P., Thomson, N.R., Holden, M.T., Mitchell, W.J., Carter, A.T., Bentley, S.D., Mason, D.R., Crossman, L., Paul, C.J., Ivens, A., Wells-Bennik, M.H., Davis, I.J., Cerdeno-Tarraga, A.M., Churcher, C., Quail, M.A., Chillingworth, T., Feltwell, T., Fraser, A., Goodhead, I., Hance, Z., Jagels, K., Larke, N., Maddison, M., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., White, B., Whithead, S., Parkhill, J., Pedersen, A.G., Jensen, L.J., Brunak, S., Staerfeldt, H.H., Ussery, D.W., Shpigelman, E.S., Trifonov, E.N., Bolshoy, Skovgaard, M., Jensen, L.J., Friis, C., Stærfeldt, H.H., Worning, P., Brunak, S., Ussery, D., Jensen, L.J., Friis, C., Ussery, D.W., Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.-U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., Karl, D.M., Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., Yaschenko, E., Nolling, J., Breton, G., Omelchenko, M. V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E. V., Smith, D.R., Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeno-Tarraga, A.M., Wang, H., Holden, M.T., Wright, A., Churcher, C., Quail, M.A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B., Parkhill, J., Bettegowda, C., Huang, X., Lin, J., Cheong, I., Kohli, M., Szabo, S.A., Zhang, X., L. A. Diaz, J., Velculescu, V.E., Parmigiani, G., Kinzler, K.W., Vogelstein, B., Zhou, S., Myers, G.S., Rasko, D.A., Cheung, J.K., Ravel, J., Seshadri, R., DeBoy, R.T., Ren, Q., Varga, J., Awad, M.M., Brinkac, L.M., Daugherty, S.C., Haft, D.H., Dodson, R.J., Madupu, R., Nelson, W.C., Rosovitz, N.J., Sullivan, S.A., Khouri, H., Dimitrov, G.I., Watkins, K.L., Mulligan, S., Benton, J., Radune, D., Fisher, D.J., Atkins, H.S., Hiscox, T., Jost, B.H., Billington, S.J., Songer, J.G., McClane, B.A., Titball, R.W., Rood, J.I., Melville, S.B., Paulsen, I.T., Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., Hayashi, H., Bruggemann, H., Baumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A., Gottschalk, G., Sakaguchi, Y., Hayashi, T., Kurokawa, K., Nakayama, K., Oshima, K., Fujinaga, Y., Ohnishi, M., Ohtsubo, E., Hattori, M., Oguma, K., Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., Arellano, A., Coleman, M.,

Hauser, L., Hess, W.R., Johnson, Z.I., Land, M., Lindell, D., Post, A.F., Regala, W., Shah, M., Shaw, S.L., Steglich, C., Sullivan, M.B., Ting, C.S., Tolonen, A., Webb, E.A., Zinser, E.R., Chisholm, S.W., Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E. V., Gall, F. Le, Makarova, K.S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I.B., Scanlan, D.J., Marsac, N.T. de, Weissenbach, J., Wincker, P., Wolf, Y.I., Hess, W.R., Benham, C.J., Bi, C., Liolios, K., Tavernarakis, N., Hugenholtz, P., Kyrpides, N.C., Rood, J.I. & Cole, S.T. (2008). The genome BLASTatlas—a GeneWiz extension for visualization of whole-genome homology. *Molecular BioSystems*. 4 (5). p.p. 363.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews : MMBR*. 68 (4). p.pp. 669–685.

Harrington, E.D., Singh, A.H., Doerks, T., Letunic, I., von Mering, C., Jensen, L.J., Raes, J. & Bork, P. (2007). Quantitative assessment of protein function prediction from metagenomics shotgun sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (35). p.pp. 13913–8.

He, Z. & Deng, Y. (2012). Development of functional gene microarrays for microbial community analysis. *Current Opinion in Biotechnology*. 23 (1). p.pp. 49–55.

Herbert, R.A. (1992). A perspective on the biotechnological potential of extremophiles. *Trends in Biotechnology*. 10. p.pp. 395–402.

Hess, M. (2008). Thermoacidophilic proteins for biofuel production. *Trends in Microbiology*. 16 (9). p.pp. 414–419.

Hjort, K., Bergström, M., Adesina, M.F., Jansson, J.K., Smalla, K. & Sjöling, S. (2010). Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS microbiology ecology*. 71 (2). p.pp. 197–207.

Hosokawa, M., Hoshino, Y., Nishikawa, Y., Hirose, T., Yoon, D.H., Mori, T., Sekiguchi, T., Shoji, S. & Takeyama, H. (2015). Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosensors and Bioelectronics*. 67. p.pp. 379–385.

Hough, D.W. & Danson, M.J. (1999). Extremozymes. *Current Opinion in Chemical Biology*. 3 (1). p.pp. 39–46.

Kalyuzhnaya, M.G., Lapidus, A., Ivanova, N., Copeland, A.C., McHardy, A.C., Szeto, E., Salamov, A., Grigoriev, I. V, Suciu, D., Levine, S.R., Markowitz, V.M., Rigoutsos, I., Tringe, S.G., Bruce, D.C., Richardson, P.M., Lidstrom, M.E. & Chistoserdova, L. (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature biotechnology*. 26 (9). p.pp. 1029–34.

Kazimierczak, K.A., Scott, K.P., Kelly, D. & Aminov, R.I. (2009). Tetracycline Resistome of the Organic Pig Gut. *Applied and Environmental Microbiology*. 75 (6). p.pp. 1717–1722.

Kim, J., Kim, B.C., Lopez-Ferrer, D., Petritis, K. & Smith, R.D. (2010). Nanobiocatalysis for protein digestion in proteomic analysis. *Proteomics*. 10 (4). p.pp. 687–99.

Klibanov, A.M. (2001). Improving enzymes by using them in organic solvents. *Nature*. 409 (6817). p.pp. 241–246.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A. & Daniel, R. (2003). Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on Escherichia coli. *Journal of molecular microbiology and biotechnology*. 5 (1). p.pp. 46–56.

Kodzius, R. (2016). Single-cell technologies in environmental omics. *Gene*. 576 (2). p.pp. 701–707.

Koeller, K.M. & Wong, C.-H. (2001). Enzymes for chemical synthesis. *Nature*. 409 (6817). p.pp. 232–240.

Kumar, L., Awasthi, G. & Singh, B. (2011). Extremophiles: A Novel Source of Industrially Important Enzymes. *Biotechnology*. 10 (2). p.pp. 121–135.

Kumar, V., Marín-Navarro, J. & Shukla, P. (2016). Thermostable microbial xylanases for pulp and paper industries: trends, applications and further perspectives. *World Journal of Microbiology and Biotechnology*. 32 (2) p.pp. 1–10.

Kyrpides, N.C., Hugenholtz, P., Eisen, J.A., Woyke, T., Göker, M., Parker, C.T., Amann, R., Beck, B.J., Chain, P.S.G., Chun, J., Colwell, R.R., Danchin, A., Dawyndt, P., Dedeurwaerdere, T., DeLong, E.F., Detter, J.C., De Vos, P., Donohue, T.J., Dong, X.-Z., Ehrlich, D.S., Fraser, C., Gibbs, R., Gilbert, J., Gilna, P., Glöckner, F.O., Jansson, J.K., Keasling, J.D., Knight, R., Labeda, D., Lapidus, A., Lee, J.-S., Li, W.-J., Ma, J., Markowitz, V., Moore, E.R.B., Morrison, M., Meyer, F., Nelson, K.E., Ohkuma, M., Ouzounis, C.A., Pace, N., Parkhill, J., Qin, N., Rossello-Mora, R., Sikorski, J., Smith, D., Sogin, M., Stevens, R., Stingl, U., Suzuki, K.-I., Taylor, D., Tiedje, J.M., Tindall, B., Wagner, M., Weinstock, G., Weissenbach, J., White, O., Wang, J., Zhang, L., Zhou, Y.-G., Field, D., Whitman, W.B., Garrity, G.M. & Klenk, H.-P. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS biology*. 12 (8). p.p. e1001920.

Lam, K.N. & Charles, T.C. (2015). Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. *Microbiome*. 3. p.p. 22.

Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D. & Charles, T.C. (2015). Current and future resources for functional metagenomics. *Frontiers in microbiology*. 6. p.p. 1196.

Lei, C., Shin, Y., Liu, J. & Ackerman, E.J. (2002). Entrapping Enzyme in a Functionalized Nanoporous Support. *Journal of the American Chemical Society*. 124 (38). p.pp. 11242–11243.

Liang, P., Averboukh, L. & Pardee, A.B. (1993). Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic acids research*. 21 (14). p.pp. 3269–75.

Liebl, W., Angelov, A., Juergensen, J., Chow, J., Loeschcke, A., Drepper, T., Classen, T., Pietruzska, J., Ehrenreich, A., Streit, W.R. & Jaeger, K.E. (2014). Alternative hosts for functional (meta)genome analysis. *Applied Microbiology and Biotechnology*. 98 (19) p.pp. 8099–8109.

Liu, M., Dai, X., Guan, R. & Xu, X. (2014). *Immobilization of Aspergillus niger xylanase A on Fe3O4-coated chitosan magnetic nanoparticles for xylooligosaccharide preparation*.

Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K.-E. & Drepper, T. (2013). TREX: A Universal Tool for the Transfer and Expression of Biosynthetic Pathways in Bacteria. *ACS Synthetic Biology*. 2 (1). p.pp. 22–33.

Lorenz, P. & Eck, J. (2005). Metagenomics and industrial applications. *Nature*. 3 (June). p.pp. 510–516.

Madern, D., Ebel, C. & Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles : life under extreme conditions*. 4 (2). p.pp. 91–8.

Méndez-García, C., Peláez, A.I., Mesa, V., Sánchez, J., Golyshina, O. V & Ferrer, M. (2015). Microbial diversity and metabolic networks in acid mine drainage habitats. *Frontiers in microbiology*. 6. p.p. 475.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R., Venter, J., Remington, K., Heidelberg, J., Halpern, A.,

Rusch, D., Eisen, J., Wu, D., Paulsen, I., Nelson, K., Nelson, W., Tyson, G., Chapman, J., Hugenholtz, P., Allen, E., Ram, R., Richardson, P., Solovyev, V., Rubin, E., Rokhsar, D., Banfield, J., Huse, S., Huber, J., Morrison, H., Sogin, M., Welch, D., Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Diaz, N., Chuang, H.-Y., Cohoon, M., Crécy-Lagard, V. de, Disz, T., Edwards, R., McNeil, L., Reich, C., Aziz, R., Bartels, D., Cohoon, M., Disz, T., Edwards, R., Gerdes, S., Hwang, K., Kubal, M., Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., Berka, J., Braverman, M., Chen, Y., Chen, Z., Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Field, D., Morrison, N., Selengut, J., Sterk, P., Altschul, Sf., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Jarvie, T., DeSantis, T., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E., Keller, K., Huber, T., Dalevil, D., Hu, P., Andersen, G., Cole, J., Chai, B., Farris, R., Wang, Q., Wuyts, J., Peer, Yv. de, Winkelmans, T., Wachter, R. De, Leplae, R., Hebrant, A., Wodak, S., Toussaint, A., Meyer, F., Overbeek, R., Rodriquez, A., Tringe, S., Mering, C. von, Kobayashi, A., Salamov, A., Chen, K., Chang, H., Podar, M., Short, J., Mathur, E., Detter, J., Rodriguez-Brito, B., Rohwer, F., Edwards, R., McHardy, A., Martin, H., Tsirigos, A., Hugenholtz, P., Rigoutsos, I., Edwards, R., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D., Saar, M., Alexander, S., Alexander, E., Rohwer, F., Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R., Felts, B., Rayhawk, S., Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., Rohwer, F., Mou, X., Edwards, R., Hodson, R., Moran, M., Dinsdale, E., Edwards, R., Hall, D., Angly, F., Breitbart, M., Brulc, J., Furlan, M., Desnues, C., Haynes, M., Li, L., Krause, L., Diaz, N., Bartels, D., Edwards, R., Puhler, A., Rohwer, F., Meyer, F., Stoye, J., Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S., Quackenbush, J. & Rohwer, F. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9 (1). p.p. 386.

Mirete, S., de Figueras, C.G. & Gonzalez-Pastor, J.E. (2007). Novel Nickel Resistance Genes from the Rhizosphere Metagenome of Plants Adapted to Acid Mine Drainage. *Applied and Environmental Microbiology*. 73 (19). p.pp. 6001–6011.

Mitidieri, S., Souza Martinelli, A.H., Schrank, A. & Vainstein, M.H. (2006). Enzymatic detergent formulation containing amylase from Aspergillus niger: A comparative study with commercial detergent formulations. *Bioresource Technology*. 97 (10). p.pp. 1217–1224.

Monsan, P. & O'Donohue, M.J. (2010). *Industrial Biotechnology in the Food and Feed Sector*. W. Soetaert & E. J. Vandamme (eds.). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Mori, T., Kamei, I., Hirai, H. & Kondo, R. (2014). Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. *SpringerPlus*. 3. p.p. 365.

Neufeld, J.D., Chen, Y., Dumont, M.G. & Murrell, J.C. (2008). Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environmental microbiology*. 10 (6). p.pp. 1526–35.

Nguyen, D., Zhang, X., Jiang, Z.-H., Audet, A., Paice, M.G., Renaud, S. & Tsang, A. (2008). Bleaching of kraft pulp by a commercial lipase: Accessory enzymes degrade hexenuronic acids. *Enzyme and Microbial Technology*. 43 (2). p.pp. 130–136.

Park, S.-J., Kang, C.-H., Chae, J.-C., Rhee, S.-K., Allgaier, M., Uphoff, H., Felske, A., Wagner-Dobler, I., Bae, J., Rhee, S., Park, J., Chung, W., Nam, Y., Lee, I., Kim, H., Park, Y., Bodrossy, L., Stralis-Pavese, N., Murrell, J., Radajewski, S., Weilharter, A., Sessitsch, A., Grzymski, J., Carter, B., DeLong, E., Feldman, R., Ghadiri, A., Murray, A., Hallam, S., Konstantinidis, K., Putnam, N., Hallam, S., Mincer, T., Schleper, C., Preston, C., Roberts, K., Richardson, P., DeLong, E., Henne, A., Schmitz, R., Bomeke, M., Gottschalk, G., Daniel, R., Kim, B., Kim, S., Park, J., Park, W., Hwang, K., Yoon, Y., Oh, W., Kim, B., Ahn, J., Kowalchuk, G., Speksnijder, A., Zhang, K., Goodman, R., Veen, J. Van, Liu, Y.,

Zhou, J., Omelchenko, M., Majernik, A., Gottschalk, G., Daniel, R., Park, S., Kang, C., Rhee, S., Quaiser, A., Ochsenreiter, T., Klenk, H., Kletzin, A., Treusch, A., Meurer, G., Eck, J., Sensen, C., Schleper, C., Rhee, S., Liu, X., Wu, L., Chong, S., Wan, X., Zhou, J., Richmond, C., Glasner, J., Mau, R., Jin, H., Blattner, F., Roh, C., Villatte, F., Kim, B., Schmid, R., Rondon, M., August, P., Bettermann, A., Schmeisser, C., Steele, H., Streit, W., Sebat, J., Colwell, F., Crawford, R., Treusch, A., Leininger, S., Kletzin, A., Schuster, S., Klenk, H., Schleper, C., Hellemond, E. Van, Janssen, D., Fraaije, M., Verdnick, D., Handran, S., Pickett, S., Ward, B., Eveillard, D., Kirshtein, J., Nelson, J., Voytek, M., Jackson, G., Williamson, L., Borlee, B., Schloss, P., Guan, C., Allen, H., Handelsman, J., Wu, L., Thompson, D., Liu, X., Fields, M., Bagwell, C., Tiedje, J., Zhou, J., Ye, R., Tao, W., Bedzyk, L., Young, T., Chen, M. & Li, L. (2008). Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS microbiology letters*. 284 (1). p.pp. 28–34.

Phale, P.S., Basu, A., Majhi, P.D., Deveryshetty, J., Vamsee-Krishna, C. & Shrivastava, R. (2007). Metabolic Diversity in Bacterial Degradation of Aromatic Compounds. *OMICS: A Journal of Integrative Biology*. 11 (3). p.pp. 252–279.

Placido, A., Hai, T., Ferrer, M., Chernikova, T.N., Distaso, M., Armstrong, D., Yakunin, A.F., Toshchakov, S. V., Yakimov, M.M., Kublanov, I. V., Golyshina, O. V., Pesole, G., Ceci, L.R. & Golyshin, P.N. (2015). Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Applied Microbiology and Biotechnology*. 99 (23). p.pp. 10031–10046.

Popovic, A., Tchigvintsev, A., Tran, H., Chernikova, T.N., Golyshina, O. V, Yakimov, M.M., Golyshin, P.N. & Yakunin, A.F. (2015). Metagenomics as a Tool for Enzyme Discovery: Hydrolytic Enzymes from Marine-Related Metagenomes. *Advances in experimental medicine and biology*. 883. p.pp. 1–20.

Radajewski, S., Ineson, P., Parekh, N.R. & Murrell, J.C. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature*. 403 (6770). p.pp. 646–9.

Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., Shah, M., Hettich, R.L. & Banfield, J.F. (2005). Community proteomics of a natural microbial biofilm. *Science (New York, N.Y.)*. 308 (5730). p.pp. 1915–20.

Riesenfeld, C.S., Goodman, R.M. & Handelsman, J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*. 6 (9). p.pp. 981–989.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R. & Woyke, T. (2014a). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols*. 9 (5). p.pp. 1038–1048.

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R. & Woyke, T. (2014b). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols*. 9 (5). p.pp. 1038–1048.

Sabree, Z.L., Rondon, M.R. & Handelsman, J. (2009). Metagenomics. In: *Encyclopedia of Microbiology 3rd Edn*. pp. 622–632.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (New York, N.Y.)*. 239 (4839). p.pp. 487–91.

Scanlon, T.C., Dostal, S.M. & Griswold, K.E. (2014). A high-throughput screen for antibiotic drug discovery. *Biotechnology and Bioengineering*. 111 (2). p.pp. 232–243.

Schiraldi, C. & De Rosa, M. (2002). The production of biocatalysts and biomolecules from extremophiles. *Trends in Biotechnology*. 20 (12). p.pp. 515–521.

Schloss, P.D. & Handelsman, J. (2006). Toward a census of bacteria in soil. *PLoS computational biology*. 2 (7). p.p. e92.

Schnoes, A.M., Brown, S.D., Dodevski, I. & Babbitt, P.C. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS computational biology*. 5 (12). p.p. e1000605.

Sheikh Abdul Hamid, N., Zen, H.B., Tein, O.B., Halifah, Y.M., Saari, N. & Bakar, F.A. (2003). Screening and identification of extracellular lipase-producing thermophilic bacteria from a Malaysian hot spring. *World Journal of Microbiology and Biotechnology*. 19 (9). p.pp. 961–968.

Shukoor, M.I., Natalio, F., Therese, H.A., Tahir, M.N., Ksenofontov, V., Panthöfer, M., Eberhardt, M., Theato, P., Schröder, H.C., Müller, W.E.G. & Tremel, W. (2008). Fabrication of a Silica Coating on Magnetic γ-Fe $_2$ O $_3$ Nanoparticles by an Immobilized Enzyme. *Chemistry of Materials*. 20 (11). p.pp. 3567–3573.

Simon, C., Herath, J., Rockstroh, S. & Daniel, R. (2009). Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Applied and environmental microbiology*. 75 (9). p.pp. 2964–8.

Singh, B.K. (2010). Exploring microbial diversity for biotechnology: the way forward. *Trends in biotechnology*. 28 (3). p.pp. 111–6.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M. & Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (32). p.pp. 12115–20.

Sul, W.J., Park, J., Quensen, J.F., Rodrigues, J.L.M., Seliger, L., Tsoi, T. V, Zylstra, G.J. & Tiedje, J.M. (2009). DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Applied and environmental microbiology*. 75 (17). p.pp. 5501–6.

Tchigvintsev, A., Tran, H., Popovic, A., Kovacic, F., Brown, G., Flick, R., Hajighasemi, M., Egorova, O., Somody, J.C., Tchigvintsev, D., Khusnutdinova, A., Chernikova, T.N., Golyshina, O. V., Yakimov, M.M., Savchenko, A., Golyshin, P.N., Jaeger, K.-E. & Yakunin, A.F. (2015). The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Applied Microbiology and Biotechnology*. 99 (5). p.pp. 2165–2178.

Tran, D.N. & Balkus, K.J. (2011). Perspective of Recent Progress in Immobilization of Enzymes. *ACS Catalysis*. 1 (8). p.pp. 956–968.

Turnbaugh, P.J., Turnbaugh, P.J., Ley, R.E., Ley, R.E., Hamady, M., Hamady, M., Fraser-Liggett, C.M., Fraser-Liggett, C.M., Knight, R., Knight, R., Gordon, J.I. & Gordon, J.I. (2007). The human microbiome project. *Nature*. 449 (7164). p.pp. 804–10.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V. V., Rubin, E.M., Rokhsar, D.S. & Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 428 (6978). p.pp. 37–43.

Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*. 23 (1). p.pp. 88–93.

Uchiyama, T. & Miyazaki, K. (2013). Metagenomic screening for aromatic compound-responsive transcriptional regulators. *PloS one*. 8 (9). p.p. e75795.

Uchiyama, T. & Miyazaki, K. (2010). Substrate-induced gene expression screening: a method for high-throughput screening of metagenome libraries. *Methods in molecular biology (Clifton, N.J.)*. 668. p.pp. 153–68.

Valdés-Solís, T., Rebolledo, A.F., Sevilla, M., Valle-Vigón, P., Bomatí-Miguel, O., Fuertes, A.B. & Tartaj, P. (2009). Preparation, Characterization, and Enzyme Immobilization Capacities of Superparamagnetic Silica/Iron Oxide Nanocomposites with Mesostructured Porosity. *Chemistry of Materials*. 21 (9). p.pp. 1806–1814.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H. & Smith, H.O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, N.Y.)*. 304 (5667). p.pp. 66–74.

Vieille, C. & Zeikus, G.J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and molecular biology reviews : MMBR*. 65 (1). p.pp. 1–43.

Walters, D.M., Russ, R., Knackmuss, H.-J. & Rouvière, P.E. (2001). High-density sampling of a bacterial operon using mRNA differential display. *Gene*. 273 (2). p.pp. 305–315.

Wexler, M., Bond, P.L., Richardson, D.J. & Johnston, A.W.B. (2005). A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environmental microbiology*. 7 (12). p.pp. 1917–26.

Wild, J., Hradecna, Z. & Szybalski, W. (2002). Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome research*. 12 (9). p.pp. 1434–44.

Wilson, Z.E., Brimble, M.A., Pikuta, E. V., Hoover, R.B., Tang, J., MacElroy, R.D., Woese, C.R., Kandler, O., Wheelis, M.L., DeLong, E.F., Rothschild, L.J., Mancinelli, R.L., Pakchung, A.A.H., Simpson, P.J.L., Codd, R., Hendry, P., Darland, G., Brock, T.D., Samsonoff, W., Conti, S.F., Eichler, J., Niehaus, F., Bertoldo, C., Kähler, M., Antranikian, G., Hough, D.W., Danson, M.J., Maheshwari, R., Bharadwaj, G., Bhat, M.K., Craveri, R., Coronelli, C., Pagani, H., Sensi, P., Moppett, C.E., Dix, D.T., Johnson, F., Coronelli, C., Johnson, F., Chandra, B., Iden, C.R., Naiksatam, P., Kahen, R., Okaya, Y., Lin, S.-Y., Pyrek, J. St., Achmatowicz, O., Zamojski, A., Maehr, H., Cuellar, H. V., Smallheer, J., Williams, T.H., Sasso, G.J., Berger, J., Hafenbradl, D., Keller, M., Stetter, K.O., Hammann, P., Hoyer, F., Kogler, H., Temirov, Y. V., Esikova, T.Z., Kashparov, I.A., Balashova, T.A., Vinokurov, L.M., Alakhov, Y.B., Craveri, R., Aragozzi, F., Manachin, P. l., Aragozzini, F., Scolasti, C., Craveri, R., Manachin, P. l., Rindone, B., Bagli, J.F., Kluepfel, D., Jacques, M.S., Omura, S., Suzuki, Y., Kitao, C., Takahashi, Y., Konda, Y., Konda, Y., Suzuki, Y., Omura, S., Onda, M., Onda, M., Konda, Y., Ivanova, V., Kolarova, M., Aleksieva, K., Graffe, U., Dahse, H.M., Laatsch, H., Loginova, L.G., Egorova, L.A., Golovacheva, R.S., Seregina, L.M., Burgess, M.L., Barrow, K.D., Gao, C., Heard, G.M., Glenn, D., Rainey, F.A., Stackebrandt, E., Kohama, Y., Teramoto, T., Iida, K., Murayama, N., Semba, T., Tatebatake, N., Kayamori, Y., Tsujikawa, K., Mimura, T., Kohama, Y., Iida, K., Semba, T., Mimura, T., Inada, A., Tanaka, K., Nakanishi, T., Kohama, Y., Iida, K., Tanaka, K., Semba, T., Itoh, M., Teramoto, T., Tsujikawa, K., Mimura, T., Morita, K., Kobayashi, S., Ritzau, M., Keller, M., Wessels, P., Stetter, K.O., Zeeck, A., Tabor, C.W., Tabor, H., Marton, L.J., Pegg, A.E., Terui, Y., Ohnuma, M., Hiraga, K., Kawashima, E., Oshima, T., Oshima, T., Oshima, T., Imahori, K., Oshima, T., Oshima, T., Oshima, T., Oshima, T., Kawahata, S., Oshima, T., Hamasaki, N., Senshu, M., Kakinuma, K., Kuwajima, I., Hamana, K., Niitsu, M., Samejima, K., Matsuzaki, S., Ohnuma, M., Terui, Y., Tamakoshi, M., Mitome, H., Niitsu, M., Samejima, K., Kawashima, E., Oshima, T., Hamana, K., Matsuzaki, S., Niitsu, M., Samejima, K., Zarilla, K.A., Perry, J.J., Zarilla, K.A., Perry, J.J., Hamana,

K., Niitsu, M., Matsuzaki, S., Samejima, K., Igarashi, Y., Kodama, T., Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B., Stetter, K.O., Zellner, G., Kneifel, H., Schenk, A., Aragno, M., Kowalak, J.A., Dalluge, J.J., McCloskey, J.A., Stetter, K.O., Zillig, W., Stetter, K.O., Wunderl, S., Schulz, W., Priess, H., Scholz, I., Rosa, M. de, Gambacorta, A., Millonig, G., Bu'Lock, J.D., Zillig, W., Stetter, K.O., Schafer, W., Janekovic, D., Wunderl, S., Holz, I., Palm, P., Stetter, K.O., Edmonds, C.G., Crain, P.F., Hashizume, T., Gupta, R., Stetter, K.O., McCloskey, J.A., McCloskey, J.A., Crain, P.F., Edmonds, C.G., Gupta, R., Hashizume, T., Phillipson, D.W., Stetter, K.O., McGee, D.P.C., Martin, J.C., Isaac, B.G., Ayer, S.W., Letendre, L.J., Stonard, R.J., Reddy, D.M., Crain, P.F., Edmonds, C.G., Gupta, R., Hashizume, T., Stetter, K.O., Widdel, F., McCloskey, J.A., Lentzen, G., Schwarz, T., Roeßler, M., Müller, V., Kanodia, S., Roberts, M.F., Seely, R.J., Fahrney, D.E., Zeikus, J.G., Wolfe, R.S., Jannasch, H.W., Huber, R., Belkin, S., Stetter, K.O., Martins, L.O., Carreto, L.S., Costa, M.S. da, Santos, H., Martins, L.O., Huber, R., Huber, H., Stetter, K.O., Costa, M.S. da, Santos, H., Stetter, K.O., Lauerer, G., Thomm, M., Neuner, A., Achenbach-Richter, L., Stetter, K.O., Woese, C.R., Schleifer, K.H., Ehrmann, M., Krusch, U., Neve, H., Farrow, J.A.E., Collins, M.D., Delorne, C., Doco, T., Wieruszeski, J.-M., Fournet, B., Lemoine, J., Chirat, F., Wieruszeski, J.M., Strecker, G., Favre, N., Neeser, J.R., Bubb, W.A., Urashima, T., Fujiwara, R., Shinnai, T., Ariga, H., Faber, E.J., Haak, M.J. van den, Kamerling, J.P., Vliegenthart, J.F.G., Marshall, V.M., Dunn, H., Elvin, M., McLay, N., Gu, Y., Laws, A.P., Silipo, A., Molinaro, A., Castro, C.D., Ferrara, R., Romano, I., Nicolaus, B., Lanzetta, R., Parrilli, M., Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E. V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V. V., Osipov, G.A., Belyaev, S.S., Ivanov, M. V., Leone, S., Molinaro, A., Romano, I., Nicolaus, B., Lanzetta, R., Parrilli, M., Holst, O., Zarilla, K.A., Perry, J.J., Rosa, M. de, Gambacorta, A., Gliozzi, A., Hanford, M., Peeples, T.L., Langworthy, T.A., Mayberry, W.R., Smith, P.F., Rosa, M. de, Rosa, S. de, Gambacorta, A., Bu'Lock, J.D., Sugai, A., Sakuma, R., Fukuda, I., Itoh, Y.H., Kon, K., Ando, S., Itoh, T., Untersteller, E., Fritz, B., Blériot, Y., Sinaÿ, P., Blériot, Y., Untersteller, E., Fritz, B., Sinaÿ, P., Rosa, M. de, Rosa, S. de, Gambacorta, A., Bu'Lock, J.D., Rosa, M. de, Rosa, S. de, Gambacorta, A., Minale, L., Bu'Lock, J.D., Langworthy, T.A., Rosa, M. de, Gambacorta, A., Nicolaus, B., Chappe, B., Albrecht, P., Thurl, S., Schäfer, W., Rosa, M. de, Gambacorta, A., Nicolaus, B., Sodano, S., Bu'Lock, J.D., Gräther, O., Arigoni, D., Heathcock, C.H., Finkelstein, B.L., Aoki, T., Poulter, C.D., Rosa, M. de, Esposito, E., Gambacorta, A., Nicolaus, B., Bu'Lock, J.D., Morii, H., Eguchi, T., Nishihara, M., Kakinuma, K., König, H., Koga, Y., Stetter, K.O., Thomm, M., Winter, J., Wildgruber, G., Huber, H., Zillig, W., Janecovic, D., Koenig, H., Palm, P., Wunderl, S., Rosa, M. de, Gambacorta, A., Huber, R., Lanzotti, V., Nicolaus, B., Stetter, K.O., Trincone, A., Kristjánsson, J.K., Hjörleifsdóttir, S., Marteinsson, V.T., Alfredsson, G.A., Hudson, J.A., Morgan, H.W., Daniel, R.M., Wait, R., Carreto, L., Nobre, M.F., Ferreira, A.M., Costa, M.S. da, Gambacorta, A., Trincone, A., Nicolaus, B., Lama, L., Rosa, M. de, Gambacorta, A., Gliozzi, A., Rosa, M. de, Williams, R.A.D., Smith, K.E., Welch, S.G., Micallef, J., Lu, T.-L., Chen, C.-S., Yang, F.L., Fung, J.-M., Chen, M.-Y., Tsay, S.-S., Li, J., Zou, W., Wu, S.-H., Carreto, L., Wait, R., Nobre, M.F., Costa, M.S. da, Yang, F.L., Lu, C.-P., Chen, C.-S., Chen, M.-Y., Hsiao, H.-L., Su, Y., Tsay, S.-S., Zou, W., Wu, S.-H., Nobre, M.F., Trüper, H.G., Costa, M.S. da, Chung, A.P., Rainey, F., Nobre, M.F., Burghardt, J., Costa, M.S. da, Tenreiro, S., Nobre, M.F., Costa, M.S. da, Ferreira, A.M., Wait, R., Nobre, M.F., Costa, M.S. da, Morii, H., Yagi, H., Akutsu, H., Nomura, N., Sako, Y., Koga, Y., Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T., Maruyama, T., Rosa, M. de, Gambacorta, A., Trincone, A., Basso, A., Zillig, W., Holz, I., Zillig, W., Holz, I., Janekovic, D., Schafer, W., Reiter, W.D., Uda, I., Sugai, A., Kon, K., Ando, S., Itoh, Y.H., Itoh, T., Oshima, T., Imahori, K., Oshima, M., Yamakawa, T., Leone, S., Molinaro, A., Lindner, B., Romano, I., Nicolaus, B., Parrilli, M., Lanzetta, R., Holst, O., Romano, I., Lama, L., Moriello, V.S., Poli, A., Gambacorta, A., Nicolaus, B., Chen, M.-Y., Lin, G.-H., Lin, Y.-T., Tsay, S.-S., Huang, Y., Anderson, R., Yang, Y.-L., Yang, F.L., Jao, S.-C., Chen, M.-Y., Tsay, S.-S., Zou, W., Wu, S.-H., Manca, M.C., Nicolaus, B., Lanzotti, V., Trincone, A., Gambacorta, A., Peter-Katalinic, J., Egge, H., Huber, R., Stetter, K.O., Hanada, S., Takaichi, S., Matsuura, K., Nakamura, K., Meer, M.T.J. van der, Schouten, S., Hanada, S., Hopmans, E.C., Damsté, J.S.S., Ward, D.M., Huber, R., Wilharm, T.,

Huber, D., Trincone, A., Burggraf, S., König, H., Rachel, R., Rockinger, I., Fricke, H., Stetter, K.O., Plötz, B.M., Lindner, B., Stetter, K.O., Holst, O., Langworthy, T.A., Mayberry, W.R., Langworthy, T.A., Mayberry, W.R., Smith, P.F., Darland, G., Brock, T.D., Förster, H.J., Biemann, K., Haigh, W.G., Tattrie, N.H., Colvin, J.R., Rohmer, M., Ourisson, G., Collins, M.D., Jones, D., Rosa, M. de, Gambacorta, A., Minale, L., Rosa, M. de, Rosa, S. de, Gambacorta, A., Minale, L., Thomson, R.H., Worthington, R.D., Trincone, A., Gambacorta, A., Lanzotti, V., Rosa, M. de, Lanzotti, V., Trincone, A., Gambacorta, A., Rosa, M. de, Breitmaier, E., Collins, M.D., Langworthy, T.A., Collins, M.D., Shimada, H., Shida, Y., Nemoto, N., Oshima, T., Yamagishi, A., Zillig, W., Yeats, S., Holz, I., Bock, A., Rettenberger, M., Gropp, F., Simon, G., Fuchs, T., Huber, H., Burggraf, S., Stetter, K.O., Thurl, S., Witke, W., Buhrow, I., Schäfer, W., Zillig, W., Yeats, S., Holz, I., Böck, A., Gropp, F., Rettenberger, M., Lutz, S., Tindall, B.J., Wray, V., Huber, R., Collins, M.D., Huber, R., Kristjansson, J.K., Stetter, K.O., Lebar, M.D., Heimbegner, J.L., Baker, B.J., Jayatilake, G.S., Thornton, M.P., Leonard, A.C., Grimwade, J.E., Baker, B.J., Medigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P.N., Cheung, F., Cruveiller, S., D'Amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E.P.C., Rouy, Z., Sekowska, A., Tutino, M.L., Vallenet, D., Heijne, G. von, Danchin, A., Mitova, M., Tutino, M.L., Infusini, G., Marino, G., Rosa, S. de, Zhang, H.L., Hua, H.M., Pei, Y.H., Yao, X.S., Frisvad, J.C., Larsen, T.O., Dalsgaard, P.W., Seifert, K.A., Louis-Seize, G., Lyhne, E.K., Jarvis, B.B., Fettinger, J.C., Overy, D.P., Dalsgaard, P.W., Larsen, T.O., Frydenvang, K., Christophersen, C., Dalsgaard, P.W., Blunt, J.W., Munro, M.H.G., Frisvad, J.C., Christophersen, C., Dalsgaard, P.W., Larsen, T.O., Christophersen, C., Smith, R.I.L., Ovstedal, D.O., Seo, C., Sohn, J.H., Park, S.M., Yim, J.H., Lee, H.K., Oh, H., Seo, C., Yim, J.H., Lee, H.K., Park, S.M., Sohn, J.-H., Oh, H., Macherla, V.R., Liu, J.N., Bellows, C., Teisan, S., Nicholson, B., Lam, K.S., Potts, B.C.M., Yayanos, A.A., Gautschi, J.T., Amagata, T., Amagata, A., Valeriote, F.A., Mooberry, S.L., Crews, P., Trischman, J.A., Jensen, P.R., Fenical, W., Schumacher, R.W., Davidson, B.S., Montenegro, D.A., Bernan, V.S., Jeong, S.-Y., Shin, H.J., Kim, T.S., Lee, H.-S., Park, S.-K., Kim, H.M., Li, D.H., Wang, F., Xiao, X., Zeng, X., Gu, Q.-Q., Zhu, W., Cui, C.-B., Kakeya, H., Okada, G., Onose, R., Ubukata, M., Takahashi, I., Isono, K., Osada, H., Cui, C.-B., Kakeya, H., Okada, G., Onose, R., Osada, H., Cui, C.-B., Kakeya, H., Osada, H., Cui, C.-B., Ubukata, M., Kakeya, H., Onose, R., Okada, G., Takahashi, I., Isono, K., Osada, H., Cui, C.B., Kakeya, H., Osada, H., Cui, C.-B., Kakeya, H., Osada, H., Kobayashi, H., Nogi, Y., Horikoshi, K., Nogi, Y., Kato, C., Horikoshi, K., Davidson, B.S., Schumacher, R.W., Kameyama, T., Takahashi, A., Kurasawa, S., Ishizuka, M., Okami, Y., Takeuchi, T., Umezawa, H., Takahashi, A., Nakamura, H., Kameyama, T., Kurasawa, S., Naganawa, H., Okami, Y., Takeuchi, T., Umezawa, H., Gustafson, K., Roman, M., Fenical, W., Rychnovsky, S.D., Skalitzky, D.J., Pathirana, C., Jensen, P.R., Fenical, W., III, A.B.S., Ott, G.R., Li, X., Dobretsov, S., Xu, Y., Xiao, X., Hung, O.S., Qian, P.-Y., Shin, J., Seo, Y., Lee, H.-S., Rho, J.-R., Mo, S.J., Trischman, J.A., Jensen, P.R., Fenical, W., Batrakov, S.G., Rodionova, T.A., Esipov, S.E., Polyakov, N.B., Sheichenko, V.I., Shekhovtsova, N. V., Lukin, S.M., Panikov, N.S., Nikolaev, Y.A., Raguénès, G., Pignet, P., Gauthier, G., Peres, A., Christen, R., Rougeaux, H., Barbier, G., Guezennec, J., Rougeaux, H., Talaga, P., Carlson, R.W., Guezennec, J., Cambon-Bonavita, M.A., Raguénès, G., Jean, J., Vincent, P., Guezennec, J., Raguénès, G., Cambon-Bonavita, M.A., Lohier, J.F., Boisset, C., Guezennec, J., Dubreucq, G., Domon, B., Fournet, B., Bozzi, L., Milas, M., Rinaudo, M., Bozzi, L., Milas, M., Rinaudo, M., Guezennec, J., Qin, G., Zhu, L., Chen, X., Wang, P.G., Zhang, Y., Rougeaux, H., Guezennec, J., Carlson, R.W., Kervarec, N., Pichon, R., Talaga, P., Vincent, P., Pignet, P., Talmont, F., Bozzi, L., Fournet, B., Guezennec, J., Jeanthon, C., Prieur, D., Raguénès, G., Christen, R., Guezennec, J., Pignet, P., Barbier, G., Rougeaux, H., Kervarec, N., Pichon, R., Guezennec, J., Galinski, E.A., Trüper, H.G., Socha, A.M., Long, R.A., Rowley, D.C., Levin, R.E., Bell, R., Carmell, S., Sar, N., Kamal, A., Qureshi, A.A., Zhungietu, G.I., Chukhril, F.N., Pindur, U., Muller, J., Veluri, R., Oka, I., Wagner-Döbler, I., Laatsch, H., Lu, Z.-Y., Lin, Z.-J., Wang, W.-L., Du, L., Zhu, T.-J., Fang, Y.-C., Gu, Q.Q., Zhu, W.M., Oren, A., Volkmann, M., Gorbushina, A.A., Kedar, L., Oren, A., Hamana, K., Niitsu, M., Samejima, K., Matsuzaki, S., Pang, H., Ihara, M., Kuchino, Y., Nishimura, S., Gupta, R., Woese, C.R., McCloskey, J.A., Roberts, M.F., Desmarais, D., Jablonski, P.E., Fedarko, N.S.,

Roberts, M.F., Tindall, B.J., Ross, H.N.M., Grant, W.D., Huber, H., Thomm, M., König, H., Thies, G., Stetter, K.O., Robertson, D.E., Lesage, S., Roberts, M.F., Henrichs, S.M., Farrington, J.W., Henrichs, S.M., Farrington, J.W., Henrichs, S.M., Cuhel, R., Sowers, K.R., Robertson, D.E., Noll, D., Gunsalus, R.P., Roberts, M.F., Sowers, K.R., Gunsalus, R.P., Zinder, S.H., Sowers, K.R., Ferry, J.G., Romesser, J.A., Wolfe, R.S., Mayer, F., Spiess, E., Walther-Mauruschat, A., Zillig, W., Holz, I., Klenk, H.-P., Trent, J., Wunderl, S., Janekovic, D., Imsel, E., Haas, B., Scholz, S., Sonnenbichler, J., Schäfer, W., Hensel, R., Lamosa, P., Martins, L.O., Costa, M.S. da, Santos, H., Neuner, A., Jannasch, H.W., Belkin, S., Stetter, K.O., Lamosa, P., Goncalves, L.G., Rodrigues, M. V., Martins, L.O., Raven, N.D.H., Santos, H., Davey, M.E., Wood, W.A., Key, R., Nakamura, K., Stahl, D.A., Jorge, C.D., Lamosa, P., Santos, H., Carreto, L., Moore, E., Nobre, M.F., Wait, R., Riley, P.W., Sharp, R.J., Costa, M.S. da, Empadinhas, N., Mendes, V., Simões, C., Santos, M.S., Mingote, A., Lamosa, P., Santos, H., Costa, M.S. da, Kondo, S., Zähringer, U., Seydel, U., Sinnwell, V., Hisatsune, K., Rietschel, E.T., Hashii, N., Isshiki, Y., Iguchi, T., Kondo, S., Hashii, N., Isshiki, Y., Iguchi, T., Hisatsune, K., Kondo, S., Tamura, T., Kato, K., Iwata, S., Kotani, S., Kitaura, T., Kato, K., Iwata, S., Suginaka, H., Namba, K., Kotani, S., Tamura, T., Rodriguez-Valera, F., Ruiz-Berraquero, F., Ramos-Cormenzana, A., Rodriguez-Valera, F., Juez, G., Kushner, D.J., Torreblanca, M., Rodriguez-Valera, F., Juez, G., Ventosa, A., Kamekura, M., Kates, M., Antón, J., Meseguer, I., Rodríguez-Valera, F., Parolis, H., Parolis, L.A.S., Boán, I.F., Rodríguez-Valera, F., Widmalm, G., Manca, M.C., Jansson, P.-E., Sutherland, I.W., Pond, J.L., Langworthy, T.A., Holzer, G., Jackson, T.J., Ramaley, R.F., Meinsche, W.G., Carballeira, N.M., Miranda, C., Lozano, C.M., Nechev, J.T., Ivanova, A., Ilieva, M., Tzvetkova, I., Stefanov, K., Kates, M., Moldoveanu, N., Stewart, L.C., Kates, M., Sastry, P.S., Yengoyan, L.S., Kates, M., Yengoyan, L.S., Sastry, P.S., Kates, M., Corcelli, A., Colella, M., Mascolo, G., Fanizzi, F.P., Kates, M., Lobasso, S., Lopalco, P., Lattanzio, V.M.T., Corcelli, A., Onishi, H., Fuchi, H., Konomi, K., Hidaka, O., Kamekura, M., Matsubara, T., Iida-Tanaka, N., Kamekura, M., Moldoveanu, N., Ishizuka, I., Onishi, H., Hayashi, A., Kates, M., Sprott, G.D., Larocque, S., Cadotte, N., Dicaire, C.J., McGee, M., Brisson, J.R., Mullakhanbhai, M.F., Larsen, H., Lopalco, P., Lobasso, S., Babudri, F., Corcelli, A., Collins, M.D., Tindall, B.J., Horikoshi, K., Manfio, G.P., Atalan, E., Zakrzewska-Czerwinska, J., Mordarski, M., Rodríguez, C., Collins, M.D., Goodfellow, M., Höltzel, A., Dieter, A., Schmid, D.G., Brown, R., Goodfellow, M., Beil, W., Jung, G., Fiedler, H.-P., Dieter, A., Hamm, A., Fiedler, H.-P., Goodfellow, M., Müller, W.E.G., Brun, R., Beil, W., Bringmann, G., Qiao, G.G.H., Meutermans, W., Wong, M.W., Träubel, M., Wentrup, C., Schumacher, J.N., Green, C.R., Best, F.W., Newell, M.P., Castro, C. de, Molinaro, A., Nunziata, R., Grant, W., Wallace, A., Parrilli, M., Castro, C. de, Molinaro, A., Wallace, A., Grant, W.D., Parrilli, M., Duckworth, A.W., Grant, W.D., Jones, B.E., Meijer, D., Márquez, M.C., Ventosa, A., Romano, I., Nicolaus, B., Lama, L., Manca, M.C., Gambacorta, A., Corsaro, M.M., Gambacorta, A., Iadonisi, A., Lanzetta, R., Naldi, T., Nicolaus, B., Romano, I., Ummarino, S., Parrilli, M., Corsaro, M.M., Gambacorta, A., Lanzetta, R., Nicolaus, B., Pieretti, G., Romano, I., Parrilli, M., Pieretti, G., Corsaro, M.M., Lanzetta, R., Parrilli, M., Nicolaus, B., Gambacorta, A., Lindner, B., Holst, O., Silipo, A., Sturiale, L., Garozzo, D., Castro, C. de, Lanzetta, R., Parrilli, M., Grant, W.D., Molinaro, A., Gwynn, M.N., Box, S.J., Brown, A.G., Gilpin, M.L., Veen, A.G. van, Mertens, W.K., Damme, P.A. van, Johannes, A.G., Cox, H.C., Berends, W., Veen, A.G. van, Mertens, W.K., Bruijn, J. De, Frost, D.J., Nugteren, D.H., Gaudemer, A., Lijmbach, G.W., Cox, H.C., Berends, W., Box, S.J., Brown, A.G., Gilpin, M.L., Gwynn, M.N., Spear, S.R., Imada, A., Kitano, K., Kintaka, K., Muroi, M., Asai, M., Kintaka, K., Haibara, K., Asai, M., Imada, A., Asai, M., Haibara, K., Muroi, M., Kintaka, K., Kishi, T., Kintaka, K., Kitano, K., Nozaki, Y., Kawashima, F., Imada, A., Nakao, Y., Yoneda, M., Imada, A., Kintaka, K., Nakao, M., Shinagawa, S., Shinagawa, S., Maki, M., Kintaka, K., Imada, A., Asai, M., Shinagawa, S., Kasahara, F., Wada, Y., Harada, S., Asai, M., Rosa, M. de, Gambacorta, A., Minale, L., Bu'Lock, J.D., Grogan, D., Palm, P., Zillig, W., Kull, D.R., Pfander, H., Stierle, A.A., Stierle, D.B., Goldstein, E., Parker, K., Bugni, T., Baarson, C., Gress, J., Blake, D., Stierle, A.A., Stierle, D.B., Kemp, K., Stierle, A.A., Stierle, D.B., Kelly, K., Stierle, D.B., Stierle, A.A., Hobbs, J.D., Stokken, J., Clardy, J., Stierle, D.B., Stierle, A.A., Patacini, B., Stierle, A.A., Stierle, D.B., Patacini, B., Robijn, G.W., Gallego, R.G., Berg, D.J.C. van

den, Haas, H., Kamerling, J.P., Vliegenthart, J.F.G., Hood, S.K., Zottola, E.A., Poralla, K., König, W.A., Allgaier, H., Poralla, K., Jung, G., Golyshina, O. V., Pivovarova, T.A., Karavaiko, G.I., Kondrat'eva, T.F., Moore, E.R.B., Abraham, W.-R., Lünsdorf, H., Timmis, K.N., Yakimov, M.M., Golyshin, P.N., Batrakov, S.G., Pivovarova, T.A., Esipov, S.E., Sheichenko, V.I., Karavaiko, G.I., Javaux, E.J., Ferrer, M., Golyshina, O., Beloqui, A., Golyshin, P.N., Bode, H.B. & Müller, R. (2009). Molecules derived from the extremes of life. *Nat. Prod. Rep.* 26 (1). p.pp. 44–71.

Wooley, J.C., Godzik, A. & Friedberg, I. (2010). A primer on metagenomics. *PLoS computational biology*. 6 (2). p.p. e1000667.

Woyke, T., Xie, G., Copeland, A., González, J.M., Han, C., Kiss, H., Saw, J.H., Senin, P., Yang, C., Chatterji, S., Cheng, J.-F., Eisen, J.A., Sieracki, M.E. & Stepanauskas, R. (2009). Assembling the Marine Metagenome, One Cell at a Time N. Ahmed (ed.). *PLoS ONE*. 4 (4). p.p. e5299.

Yakimov, M.M., Denaro, R., Genovese, M., Cappello, S., D'Auria, G., Chernikova, T.N., Timmis, K.N., Golyshin, P.N. & Giluliano, L. (2005). Natural microbial diversity in superficial sediments of Milazzo Harbor (Sicily) and community successions during microcosm enrichment with various hydrocarbons. *Environmental Microbiology*. 7 (9). p.pp. 1426–1441.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R. & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*. 12 (9). p.pp. 635–645.

Yin, H., Cao, L., Qiu, G., Wang, D., Kellogg, L., Zhou, J., Dai, Z. & Liu, X. (2007). Development and evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in Acid Mine Drainages and bioleaching systems. *Journal of Microbiological Methods*. 70 (1). p.pp. 165–178.

CHAPTER III

Taxonomic and functional annotation of CNSL-enriched microbial community from Sorghum bicolor rhizosphere

3.1. Introduction

Soil is a complex and one of the most biodiverse habitats with estimated thousands of microbial species residing in just one gram of soil (Schloss & Handelsman, 2006), which are involved in organic matter turnover and mineral nutrient cycling (Nannipieri et al., 2017). The initial microbial community structure in soil is a function of its physical and chemical properties, such as pH (Fierer & Jackson, 2006; Lauber et al., 2009), soil type (Schutter et al., 2001), water content (Bossio & Scow, 1998). Changes in carbon substrates content, addition of nutrients or specific treatments influence the predominant microbial populations (Goldfarb et al., 2011). Agricultural management including pesticide and fertilizer application (Jangid et al., 2008), organic amendments (de Brito et al., 1995), and plant root exudation (Nannipieri et al., 2008) are some examples of events that significantly modify the composition of soil microbial communities.

Soil and sediments are often the recipient of many pollutants originated from anthropogenic activities. Main soil pollutants include polycyclic aromatic hydrocarbons (PAHs), petroleum and related products, pesticides, chlorophenols and heavy metals. Among the principal soil pollutants, hydrocarbons play a special role due to their wide-scale distribution and hazardous physicochemical and biological properties. Hydrocarbons enter into the environment through waste disposal, accidental spills, and losses during transport and storage. Polycyclic aromatic hydrocarbon compounds are among the most toxic components to plants and animals (Samanta et al., 2002; Shuttleworth & Cerniglia, 1995; Xue & Warshawsky, 2005). Due to the great concern for environmental and human health, polluted environments have been the target of the attention of the scientific research with the intent of setting up bioremediation processes to detoxify and remove the pollutants. Although the effectiveness of bioremediation processes is influenced by numerous abiotic and biotic factors (such as pH, temperature, nutrient availability), the presence of microorganisms with the appropriate metabolic capabilities remains the fundamental requirement. For this reason, numerous metagenomic studies have been carried out to elucidate the modification occurring in the indigenous microbial structure of contaminates sites in order to characterize their catabolic activities (Guazzaroni et al., 2013; Yang et al., 2016; Jung et al., 2016). It appears how indigenous microbes gradually adapt

to long-term contamination and develop a superior community which can make use of the contaminants (Mohammed et al., 2007). A number of bacterial genera have been reported to have evolved specific mechanisms for metabolising phenol, alkylaromatic and petroleum-derived aromatic hydrocarbons (Hazen et al., 2010; Lu et al., 2012; Brooijmans et al., 2009). Members of the bacterial genera *Pseudomonas*, *Sphingomonas*, *Bacillus*, *Ralstonia*, *Cupriavidus*, *Burkholderia*, *Acinetobacter*, *Rhodococcus*, *Mycobacterium* and *Arthrobacter* were isolated from marine and terrestrial contaminated sites (Mittal & Singh, 2009; Nešvera et al., 2015; Yan et al., 2013). Specific microorganisms have evolved to feed exclusively on hydrocarbons, including obligate hydrocarbon degraders of the genera *Oleispira*, *Oleiphilus*, *Thalassolituus*, *Alcanivorax* and *Cycloclasticus (Yakimov et al., 2007).*

The isolation and the study of these microbes helped in the definition of the metabolic pathways of aromatic compounds degradation (Ladino-Orjuela et al., 2016; Nešvera et al., 2015). Interestingly, microbial ability to degrade recalcitrant organic compounds frequently arise from the promiscuous activity of previously existing enzymes. The concept of promiscuity comprises the idea of catalytic promiscuity (catalysing distinctly different chemical transformations), and/or substrate promiscuity (ability to bind with and convert structurally diverse molecules that resemble the normal substrate) (Copley, 2003). This phenomenon may explain the origin and evolution of bacterial enzymes that degrade new synthetic compounds that had been introduced into the environment in recent times (Kivisaar, 2009; Wackett, 2004; Copley, 2009). In addition, it has been reported that microorganisms participate in the biodegradation of recalcitrant organic molecules by a process known as co-metabolism (or fortuitous metabolism). Co-metabolism was defined as "the transformation of a non-growth substance in the obligate presence of a growth-substrate or another transformable compound" (Dalton et al., 1982); neither energy nor carbon are derived from the metabolised fortuitous compound. This mechanism has been associated with the degradation of xenobiotics, such as pesticides and PAHs by naturally occurring microbial populations (Rentz et al., 2005; Van Herwijnen et al., 2003; Nzila, 2013). The pre-exposure to lower molecular weight aromatic compounds and to appropriate primary carbon sources can trigger the co-oxidation of other PAHs and improve in situ biodegradation (Bauer & Capone, 1988; Heitkamp & Cerniglia, 1988). Commonly used growth-substrates (e.g. glucose, yeast extract, peptone) promoted co-metabolism (Kim et al., 2003). Stimulation of high molecular weight PAH degradation was obtained in the presence of salicylate (Chen & Aitken, 1999). Bacteria that make use of monoaromatic and low molecular weight polyaromatic compounds as source of carbon showed co-metabolism of pollutants of similar chemical structure. Commonly reported examples include members of *Sphingomonas*, *Pseudomonas*, *Burkholderia*, *Rhodococcus* and *Mycobacterium* (Ye et al., 1996; Husain, 2008; Nzila, 2013).

Interesting examples of co-metabolism have been reported in the rhizosphere, a unique ecological environment where the degradation of environmental contaminants is generally enhanced (Pizarro-Tobías et al., 2015), with microorganisms removing PAH while growing on root exudate (Rentz et al., 2005). It has been extensively observed how plants shape the rhizosphere microbial community composition through the production of root exudates, recruiting microorganism harbouring the necessary catabolic functions (Siciliano et al., 2001) and stimulating a broader diversity of metabolic activities if compared to bulk soil (Li et al., 2014b). Plant root exudates are mainly composed by phenolic compounds that have proved to enhance microbial degradation of polychlorinated biphenyls (PCBs) and PAH by inducing the expression of oxygenases that trigger the aerobic degradation pathways (Anderson et al., 1993; Kamath et al., 2004). Moreover, the higher availability of organic matter in rhizosphere soil increase the microbial population size, as well as the contaminants bioavailability (Reilley et al., 1996; Miya & Firestone, 2000, 2001). The impact of endophytes and plant growth-promoting rhizobacteria (PGPR) in the transformation of pollutants is crucial. Endophytes are bacteria that inhabit living plant tissue and, besides the many processes they mediate, they can decrease phytotoxicity (with the production of sidephores and iron chelators) and promote the degradation of many common hydrocarbon contaminants (Li et al., 2012). Therefore, phytoremediation, or plant-assisted bioremediation, is seen as an attractive option for the management of polluted soils (Glick, 2010; Martin et al., 2014).

In light of the considerations expressed above, we postulated that the addition of Cashew Nut Shell Liquid (CNSL) into rhizosphere soil would affect the indigenous microbial community, promoting the enrichment of microorganisms capable of tolerating and/or metabolising CNSL constituents. At the same time, a possible enrichment in activities associated with the metabolism of other aromatic and polyaromatic compounds was hypothesized. *Sorghum bicolor* rhizosphere was chosen as starting material in the enzyme discovery investigation conducted for the fulfilment of this work. Sorghum spp. are well-studied allelopathic plants, known to exudate from the root hairs phenolic compounds that confer them a competitive advantage over the surrounding plants and microbial species (Dayan et al., 2010). Major constituents of the sorghum exudate are molecules that structurally resembles the components of CNSL (Dayan et al., 2003). On the contrary, the biological effects of cashew oil in the environment have been only partially investigated and ascribed to the high content in phenolic compounds. Cytotoxicity and mutagenicity were observed in prokaryotic cells (George & Kuttan, 1997; Polasa & Rukmini, 1987). *In vitro* studies on rumen microbiota showed CNSL altering rumen fermentation towards less methane fermenting bacteria. The surfactant action of CNSL acted against the bacterial cell surface breaking it, whereas gram-negative propionate-producing bacteria that possessed an outer membrane exhibited tolerance (Watanabe et al., 2010; Oh et al., 2017). An

investigation conducted on sites contaminated by effluents of the cashew industry suggested the presence of putative CNSL-degrading microorganisms such as *Pseudomonas stutzeri*, *Enterobacter cloacae*, *Enterobacter sakazakii*, *Sphingomonas paucimobilis* and *Pseudomonas pseudoalcaligenes* (Sabna Prabha et al., 2011). *Pseudomonas aeruginosa* and *Delftia acidovorans* appeared to be potent CNSL degraders, with 85% and 79% of estimated degradation after 4 hours of incubation, respectively (Sindhu & Potty, 2015). Degradation capacity of fungal species was also evaluated (Sindhu & Potty, 2016).

Here, we report the results of the comparative analysis of CNSL-enriched rhizosphere microcosms at different intervals of time. A metagenomics approach was applied for the investigation aiming to highlight the major modifications at phylogenetic and functional level. Annotated protein-coding genes were cross-linked to manually curated databases of aromatic compounds-degrading enzymes, and a potential network of catabolic reactions implicated in the poly/monoaromatics and alkenes degradation was reconstructed.

3.2. Materials and Methods

3.2.1. Germination of Sorghum plants

Soil germination of *Sorghum bicolor* plants was carried out in soil collected from the Henfaes Research Centre (53°14'21.0"N, 4°01'06.5"W, Gwynedd, Wales) in the greenhouse (Bangor University, Deiniol site). Seeds of *Sorghum bicolor* genotype BTx623 were obtained as a gift from the Agricultural Research Service of United States Department of Agriculture (Georgia, US). The soil at Henfaes consists of a fine loamy brown topsoil overlying gravel (Rheidol series) and is classified in the United Nations Food and Agriculture Organization (FAO) structure as a Dystric Cambisol (Teklehaimanot et al., 2002). In the location where the soil was collected, there was no history of sorghum being grown. The field soil sample consisted of topsoil collected from five randomly selected positions that were subsequently pooled to yield one sample. Soil was air-dried, mixed thoroughly, and stored at room temperature for use in subsequent experiments. 2L-pots were filled with soil and two seeds of *S. bicolor* BTx623 were sown per pot. Plants were cultivated in a greenhouse at 20°C and soil moisture content was maintained with tap water.

3.2.2. Enrichment with tCNSL

The amount of 3 g of technical Cashew Nut Shell Liquid (tCNSL) (Fig. 3.1) dissolved in 70% ethanol was added into the pot of 20-days-old plants and the soil was mixed thoroughly. After sixty days from tCNSL addition, plants were pulled out of the pot and soil was shaken off; samples of rhizosphere soil attached to the plant roots were then brushed off and collected. Soil slurry enriched microcosms were set up in conical 1L Erlenmeyer shaking flasks by mixing 10 g of the collected rhizosphere soil with 300 mL of sterile Murashige Skoog Basal medium (Sigma) and 10 mg/L of cycloheximide antibiotic (Sigma). Three biological replicates were performed. tCNSL was dissolved in 70% ethanol and added to the medium to a final concentration of 0.1 g/L. Flasks were kept under orbital shaking conditions at 20 °C; soil slurry was sampled every 7 days and fresh tCNSL was added for a total of 35 days. Samples used in this study were labelled as follows: 0W for the not enriched timepoint zero, and 1W, 3W and 5W for samples of tCNSL enriched culture collected after 1, 3 and 5 weeks respectively.

Fig. 3.1. Technical Cashew Nut Shell Liquid

3.2.3. DNA extraction

Metagenomic DNA was isolated from the fractions collected at timepoint zero and after 1, 3 and 5 weeks of tCNSL enrichment. Nucleic acid extraction from the collected triplicate samples of soil suspension was performed using the Meta-G-Nome DNA Isolation Kit (Epicentre Biotechnologies; WI, USA). Briefly, 50 ml of the soil suspension from the flask enrichment were centrifuged at 400 x g for 5 min. The supernatant was filtered through 0.45 µm and 0.22 µm membrane filters. This procedure was repeated with the initial soil sample four times: remaining soil was resuspended in PBS, centrifuged and the supernatant filtered as before. Filter membranes were combined and the sediment on the filter was resuspended with extraction buffer and collected. DNA extraction was carried out according to the protocol described by the manufacturer. The quality of the extracted DNA was evaluated on agarose gel and quantified with the Quant-iT dsDNA Assay Kit (Invitrogen) on a Cary Eclipse fluorimeter (Varian/Agilent) according to the manufacturer's instructions. Duplicate DNA samples corresponding to timepoint zero (0W), week 1 (1W), week 3 (3W) and week 5 (5W) were sequenced by Illumina HiSeq 2500 platform (150bp pared-end reads) at Fidelity Systems, Inc. (Gaithersburg, MD) and assembled as described previously (Golyshina et al., 2017).

3.2.4. IMG/MER annotation

The dataset obtained from the sequencing was assembled and uploaded to the online metagenomic analysis tool Integrated Microbial Genomes and Microbiomes (IMG/M: https://img.jgi.doe.gov/m/) and annotated according to the Metagenome Annotation Pipeline of the Joint Genome Institute (JGI) of U.S. Department of Energy (DOE) (Huntemann et al., 2016; Chen et al., 2017; Markowitz et al., 2014). The IMG/M is a genome browsing and annotation platform that includes archaea, bacteria, eukarya, plasmids, viruses, metagenomes and metatranscriptome

datasets. The database of IMG/M is comprised of sequence data being created by DOE Joint Genome Institute, uploaded by independent researchers, or fetched from public domain. The web interface provides comparative analysis tools for analysing their metagenome datasets. The "Compare Genome" menu included tools such as Genome Statistics, Phylogenetic Distribution, Abundance Profiles, Distance Tree, Function Profile, and Genome Clustering, which were used for the purposes of this study. Every measurement (number of hits for each functional of taxonomic group) was expressed as "estimated gene copies", through which the number of genes is multiplied by the average sequence coverage of the contigs on which these genes were predicted. Values were then normalised on the total number of annotated sequences in order to avoid artefacts due to differences in sample size.

Phylogenetic composition analysis of the samples was based on the distribution of best BLAST hits of protein-coding genes. Count values for each taxon corresponded to the estimated copies of protein coding-genes that had best BLASTP hit (Altschul et al., 1997) to proteins in the indicated taxon, with a minimum amino acid sequence identity of 30%. The counts were then expressed as percentage of the total estimated gene copies of each sample, and average and standard deviation of duplicates was calculated. Metagenomic samples were then compared in terms of their phylogenetic composition at the phylum, class, order and family level.

Functional annotation for metagenomes consisted of associating protein-coding genes with COGs, Pfams, KO terms, EC numbers (Tatusov et al., 1997; Kanehisa et al., 2016; Finn et al., 2016). Functional capabilities of the metagenomes were investigated by the Abundance Profile tool that allowed comparing the relative abundance of pre-defined groups of protein families as represented by different types of functional classifications (COGs, KEGG pathways) and functional families (Enzymes). Homology searches were performed using RPS-BLAST (Reverse Position Specific BLAST) 2.2.31 with an e-value of 1e-2 against NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al., 2017). Genomes were subjected to cluster analysis using the clustering tools available under IMG Compare Genomes main menu option. The results of hierarchical clustering are displayed in a tree format. The computed distance on the tree was based on the similarity of the functional characterization of genomes in terms of a specific protein/functional family. The metagenomic data were also annotated against SEED subsystems database in MG-RAST (Meyer et al., 2008; Overbeek et al., 2014) at a cut-off of E-value < 10^{-5}, 60% minimum identity and 15 bp minimum alignment length.

3.2.5. Biodegradation network reconstruction

Functional assignment of predicted proteins involved in degradation was performed based on BLASTP comparison against two different sequence datasets. The first dataset used was AromaDeg, a manually curated database which exploits a phylogenomic approach for functional classification of

key proteins involved in aerobic degradation of aromatic compounds (Duarte et al., 2014); in addition, an in-house database (Guazzaroni et al., 2013) containing experimentally validated protein sequences involved in intradiol cleavage of aromatics and alkene biodegradation was also used. The validated enzymes included Rieske non-heme iron oxygenases, type I extradiol dioxygenases of the vicinal oxygen chelate superfamily, type II or LigB superfamily extradiol dioxygenases, type III extradiol dioxygenases belonging to the cupin superfamily, intradiol dioxygenases and alkane hydroxylases. Predicted open reading frames (ORFs) of the annotated metagenomic samples were used as queries in the BLAST search, and the results were filtered by sequence identity (>50%) and minimum alignment length (> 50 amino acids). Sequences that matched a given protein family were associated to an experimentally validated catabolic enzyme performing an aromatic degradation reaction, identified by a substrate and a degradation product. The putative substrate and product were then connected to reconstruct the catabolic networks. The number of genes assigned to the same function (or degradation reaction) determined the thickness of the line in the figure. The relative abundance of each function and the complete list of the substrate is shown in the appendix. Scripts and command allowing the network reconstruction in a graphical format and the comparison of the abundance levels of genes encoding catabolic enzymes have previously been reported (Bargiela et al., 2015b). In order to establish the taxonomic affiliation of the identified catabolic genes, the amino acid sequences were used as query in BLASTP search against NCBI non-redundant database. The complete taxonomy of the first ten homologs was retrieved and used for the taxonomic lineage definition.

3.3. Results

3.3.1.   DNA extraction and sequencing

After 20 days from the germination, 3 g of technical Cashew Nut Shell Liquid (tCNSL) were added to the soil and, after sixty days, a further enrichment of the rhizosphere was carried out in shaking flasks with the addition of tCNSL. DNA extraction from samples collected at timepoint zero and after 1, 3 and 5 weeks of tCNSL microcosm enrichment (samples IDs = 0W, 1W, 3W, 5W) was performed and two replicates were subjected to Illumina *de novo* sequencing. The output of the assembling and annotation is summarised in Table 3.1. Due to sequencing error, only one replicate was sequenced for sample 0W. Due to differences in size in terms of number of sequences, the outcome of the further analyses was normalised on the total number of protein coding genes.

Table 3.1. Summary of the metagenomes assembly data and gene annotation. Average values of two sequenced replicates is shown.

ID	N. sequences	N. bases	Protein coding genes			
				- with COG	- with Pfam	- with Enzyme
0W	27966	14574824	38669	21308	18124	10274
1W	111702	91688760	182293	89478	84398	38144
3W	145541	105634269	232009	121399	110141	54662
5W	262012	176082657	389525	196226	183768	86769

3.3.2.   Analysis of bacterial community dynamics

A comparative analysis of the samples was carried out in order to highlight possible modifications in the microbial community composition. Taxonomic classification was based on the distribution of the best BLAST hits of the annotated protein-coding genes. Phylotypes abundance was analysed at the phylum, class, order and family level. At the phylum level (Fig. 3.2), the most predominant microorganisms belonged to *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Firmicutes*. Significant changes in the relative abundances of these groups became apparent. At the timepoint zero (0W) the bacterial community was composed mainly by *Actinobacteria* (38.26% of the annotated sequences), *Proteobacteria* (32.46%), *Gemmatimonadetes* (5.90%) *Firmicutes* (1.91%), *Cyanobacteria* (1.45%) and *Bacteroidetes* (1.42%). Following addition of tCNSL, *Proteobacteria* increased in abundance to become the dominant phylum (1W: 79.74%; 3W: 53.06%; 5W: 58.35%),

whereas a considerable drop was observed in the relative abundance of *Actinobacteria* (1W: 15.6%; 3W: 14.7%; 5W: 19.4%). The third most-represented phylum after enrichment was *Bacteroidetes* (1W: 0.76%; 3W: 8.87%; 5W: 12.74%). The relative abundance of *Firmicutes* remained quite constant, accounting for approximately 1.3-2.8% of the annotated sequences.
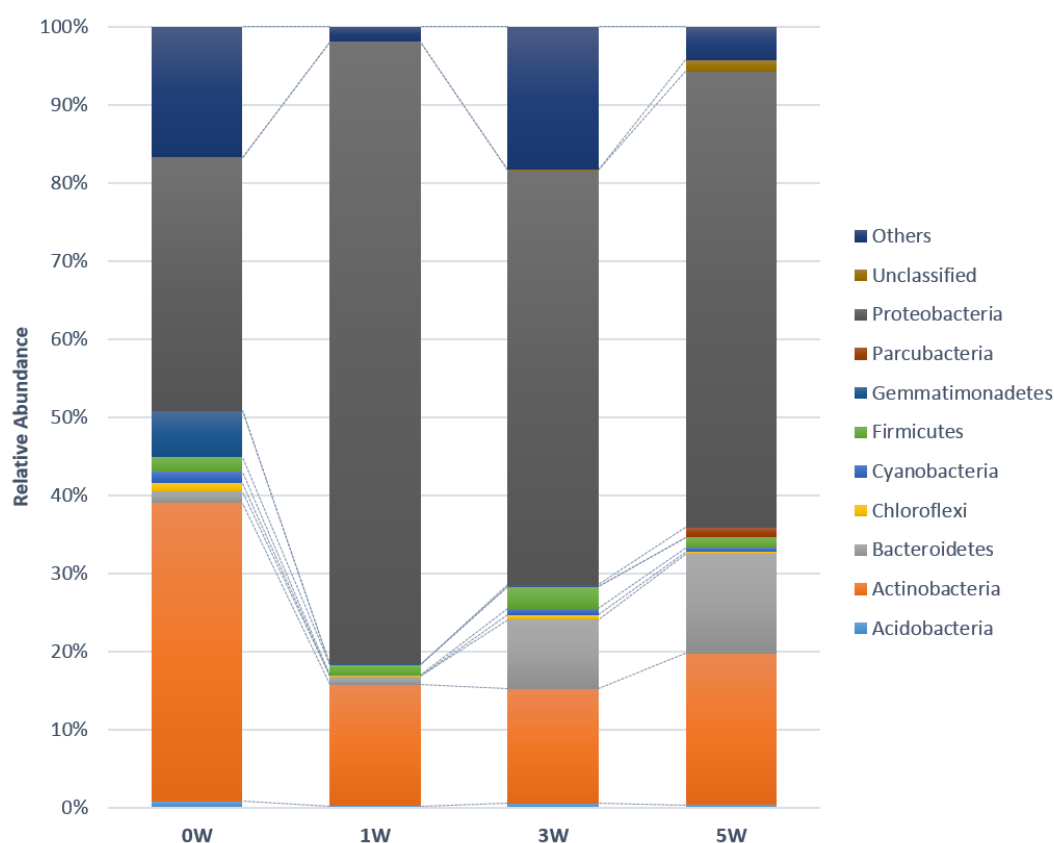
Fig. 3.2. Microbial diversity in the tCNSL enriched samples and ratio of the most abundant taxonomic groups at the phylum level. The phylogenetic composition was based on the distribution of best BLAST hits of protein-coding genes in the dataset. The stacked bars represent the composition of each sample estimated from the result of BLASTP analysis with a 30% threshold identity.

Interesting changes in the community structure occurred at the class level (Table 3.2). The class profile of *Proteobacteria* showed a pronounced alteration within the enrichment period. *Alphaproteobacteria* was the dominant class at timepoint zero (26.02% of total annotated sequences), while *Beta-* and *Gammaproteobacteria* accounted for only 1.90%. After 1 week, *Betaproteobacteria* and *Gammaproteobacteria* showed 28 and 10-fold increase respectively (53.10% and 19.10%); after 5 weeks *Betaproteobacteria* represented 25% of the microbial population, *Alphaproteobacteria* 21%, and *Gammaproteobacteria* 12%. The increment in *Bacteroidetes* was ascribed to a small enrichment

in *Chitinophagia* (0W: 0.13%; 5W: 4.30%), *Sphingobacteriia* (0W: 0.11%; 5W: 3.73%), *Cytophagia* (0W: 0.61%; 5W: 2.99%), and *Flavobacteriia* (0W: 0.15%; 5W: 1.15%).

Table 3.2 Relative abundance (%) of the taxonomic groups at class level of the top 4 most abundant phyla based on the assignment of protein-coding genes after BLASTP analysis.

Phylum	Class	0W	1W	3W	5W
Proteobacteria	Acidithiobacillia	0.07	0.02	0.06	0.03
	Alphaproteobacteria	26.02	7.31	27.19	21.15
	Betaproteobacteria	1.89	53.10	14.80	25.01
	Deltaproteobacteria	2.58	0.21	0.85	0.55
	Epsilonproteobacteria	0.01	0.01	0.06	0.02
	Gammaproteobacteria	1.90	19.10	10.08	11.59
	Zetaproteobacteria	0.01	0.00	0.04	0.01
Actinobacteria	Acidimicrobiia	0.04	0.22	0.10	0.28
	Actinobacteria	32.68	15.35	14.35	18.97
	Coriobacteriia	0.03	0.00	0.12	0.03
	Nitriliruptoria	0.05	0.00	0.00	0.02
	Rubrobacteria	0.12	0.00	0.01	0.01
	Thermoleophilia	5.34	0.01	0.12	0.10
Bacteroidetes	Bacteroidia	0.06	0.03	0.26	0.32
	Chitinophagia	0.13	0.30	3.27	4.30
	Cytophagia	0.61	0.14	1.26	2.99
	Flavobacteriia	0.15	0.09	0.95	1.15
	Saprospiria	0.04	0.02	0.18	0.14
	Sphingobacteriia	0.11	0.49	2.79	3.73
	Unclassified	0.32	0.03	0.17	0.13
Firmicutes	Bacilli	0.87	1.26	1.26	0.73
	Clostridia	0.86	0.10	1.28	0.47
	Erysipelotrichia	0.01	0.00	0.01	0.00
	Negativicutes	0.15	0.01	0.17	0.07
	Tissierellia	0.02	0.00	0.05	0.02
	Unclassified	0.00	0.00	0.00	0.00

The intra-class diversity of the dominant groups *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* was investigated and changes in the relative abundance of the included members were observed (Fig. 3.3 and 3.4). At the timepoint zero (blue pie charts), the class *Actinobacteria* was composed mainly by microorganisms of the order *Micrococcales* (45.6%) and *Propionibacteriales* (37.2%), with *Intrasporangiaceae* and *Nocardioidaceae* as the most abundant families. After tCNSL addition, *Micrococcales* percentage dropped to 3.0% and showed at

the same time a more diversified composition at family level including *Microbacteriaceae*, *Cellulomonadaceae* and *Micrococcaceae* bacteria. The relative abundance of *Propionibacteriales* decreased to 11.76% after 1 week of enrichment (orange pie chart), to increase again to 28.22% and 44.26% after 3 and 5 weeks of enrichment (grey and yellow pie charts), becoming the second most represented order. *Corynebacteriales* became the dominant order after the enrichment (0W: 2.45%; 1W: 3.44%; 3W: 26.05%; 5W: 44.45%), mainly represented by the genus *Mycobacterium.* Curiously, the abundance of *Streptomycetales* fluctuated between 2%-71% (0W: 2.52%; 1W: 71.56%; 3W: 30.21%; 5W: 2.90%), settling on a negligible value after 5 weeks.
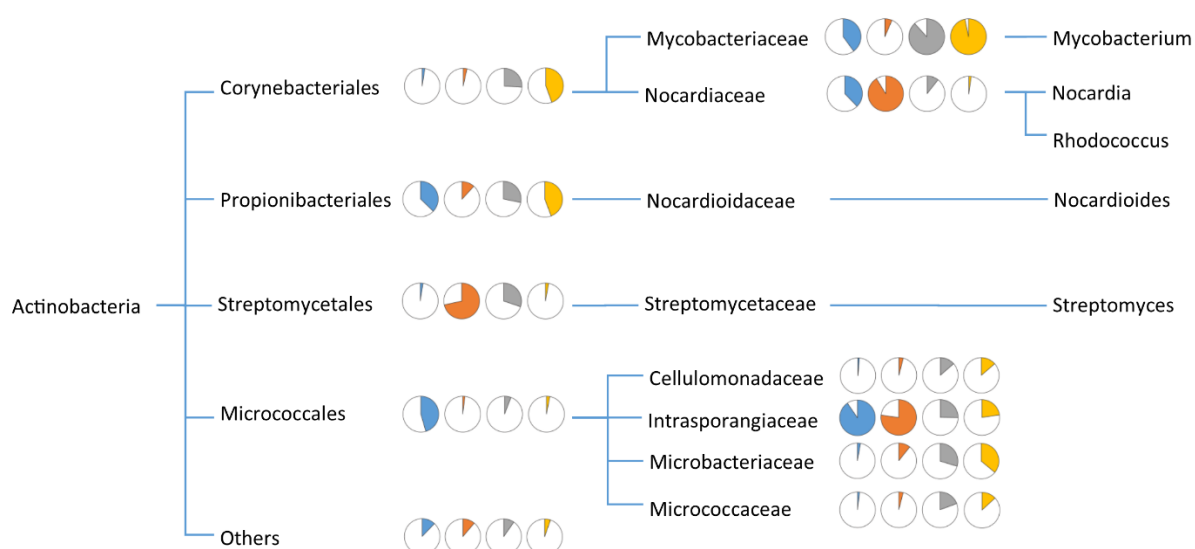
Fig 3.3. Taxonomic composition of the bacterial class *Actinobacteria* based on the assignment of protein-coding genes. Distribution of the included orders and families at different timepoints is shown as pie charts with the following colour code: blue for 0W, orange for 1W, grey for 3W, and yellow for 5W.

The class *Alphaproteobacteria* consisted essentially of *Rhizobiales* and *Sphingomonadales* (Fig. 3.4). *Rhizobiales* was the dominant order in 0W (78.11%), represented for the 88% by microorganisms of the family *Methylobacteriaceae.* In course of enrichment the abundance of *Rhizobiales* lowered (1W: 23.59%; 3W: 43.23%; 5W: 47.44%) and a more diversified composition was observed with the increase of microorganisms belonging to *Brazyrhizobiaceae* (0W: 3.03%; 1W: 34.84%; 3W: 40.68%; 5W: 29.69%), *Phyllobacteriaceae* (0W: 1.58%; 1W: 20.32%; 3W: 36.79%; 5W: 51.18%) and *Rhizobiaceae* (0W: 3.55%; 1W: 27.31%; 3W: 7.57%; 5W: 8.34%). *Sphingomonadales* bacteria appeared enriched from the initial 16.23% to a final 43.9%, with *Sphingomonas* being the most represented genus. In the class *Betaproteobacteria* the initial predominance of *Burkholderiales* was confirmed and favoured after enrichment (0W: 79.59%; 1W: 99.05%; 3W: 97.90%; 5W: 99.02%),

with the predominant affiliation to the families *Burkholderiaceae* and *Oxalobacteraceae* (Fig. 3.4). Similarly, before CNSL amendment *Gammaproteobacteria* comprised a wide number of species including *Xanthomonadales, Chromatiales, Enterobacterales, Methylococcales* and *Pseudomonadales*. The enrichment process supported the growth of the only *Xanthomonadales* (0W: 22.59%; 1W: 86.54%; 3W: 85.32%; 5W: 90.23%); the main members of the order were *Rhodonobacteriaceae*, including bacteria of the genera *Frateuria*, *Rhodonobacter*, *Luteibacter* and *Dyella* (Fig. 3.4).
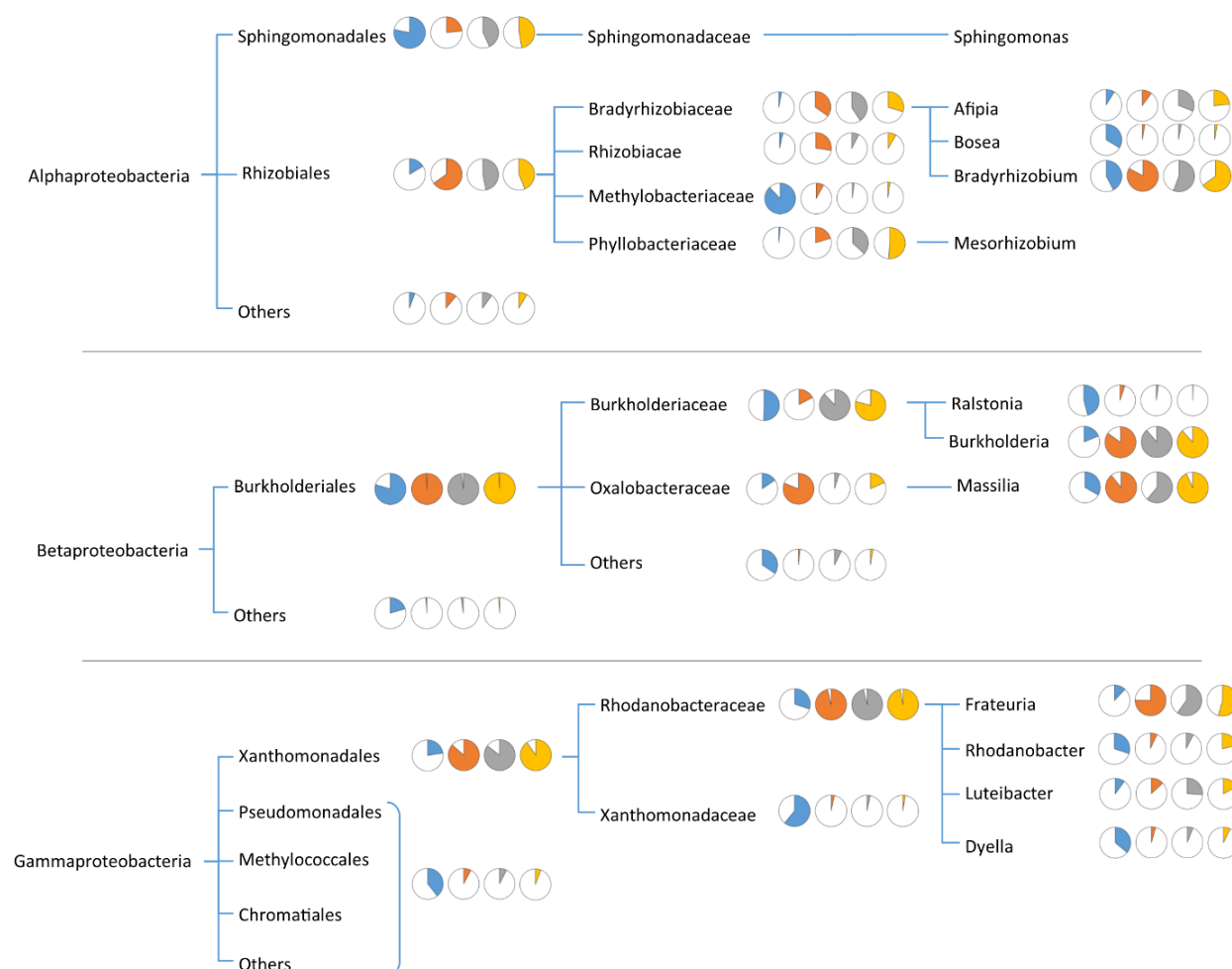
Fig. 3.4 Phylogenetic composition of bacterial classes *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* based on the assignment of protein-coding genes. Distribution of the included orders and families at different timepoints is shown as pie charts with the following colour code: blue for 0W, orange for 1W, grey for 3W, and yellow for 5W.

Overall, before the enrichment with tCNSL, the most representative taxa were *Rhizobiales*, followed by *Micrococcales*, *Propionibacteriales*, *Gemmatimonadales*, *Solirubrobacterales* and *Sphingomonadales* (Fig. 3.5). The composition drastically changed after 1 week of enrichment, going

through a transitional stage that found its final structure in the 3W and 5W communities, in which the most prevalent orders were *Burkholderiales*, *Xanthomonadales*, Rhizobiales, *Sphingomonadales, Streptomycetales*, *Corynebacteriales* and *Propionibacteriales.*
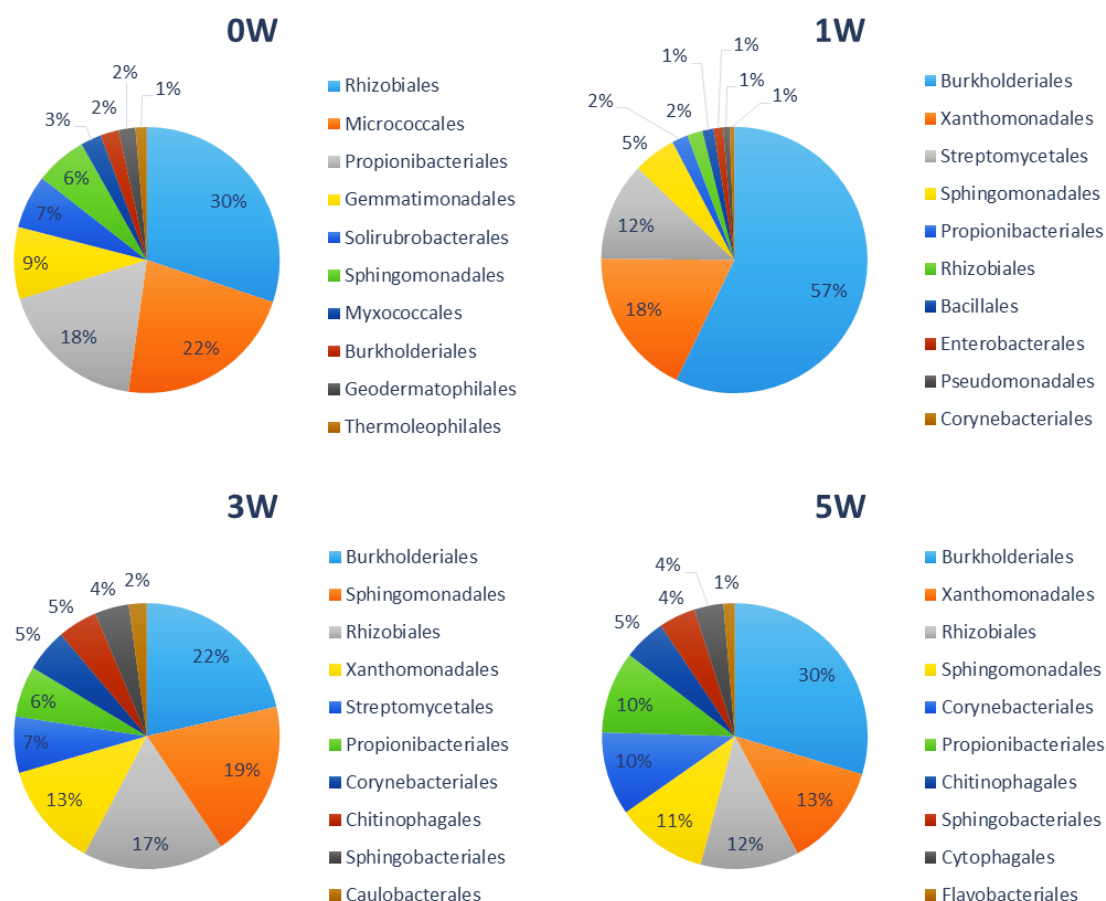
Fig. 3.5 Relative abundance of the top ten most represented orders at different timepoints.

3.3.3. Functional assignment of predicted genes

Functional assignment of the predicted protein-coding genes was performed based on the comparison against pre-defined groups of protein families, as represented by different types of functional classifications such as COGs, KO terms and SEED subsystems. About 51% of the predicted genes were assigned to 4,242 COGs. The hierarchical clustering analysis based on similar functional profiles according to the COG Categories is displayed in a tree format in Fig. 3.6A. The placement in the tree reflects the relative degree of similarity of samples to each other. Sample 0W and 1W showed

distinctive profiles, whereas 3W and 5W clustered together suggesting a similar pattern in functional characterization.

The comparative analysis of COG-assigned predicted genes indicated a quite unvaried distribution throughout the enrichment for general functions associated with routine microbial activity, which resulted to be also the most abundant categories, such as carbohydrate metabolism, amino acids and protein metabolism. Over-representation of COGs within the categories "cell wall/membrane biogenesis", "transcription", "intracellular trafficking/secretion and vesicular transport", "cell motility", and with unknown function were observed in the metagenomic samples after 1, 3 and 5 weeks of enrichment (Fig. 3.6B).
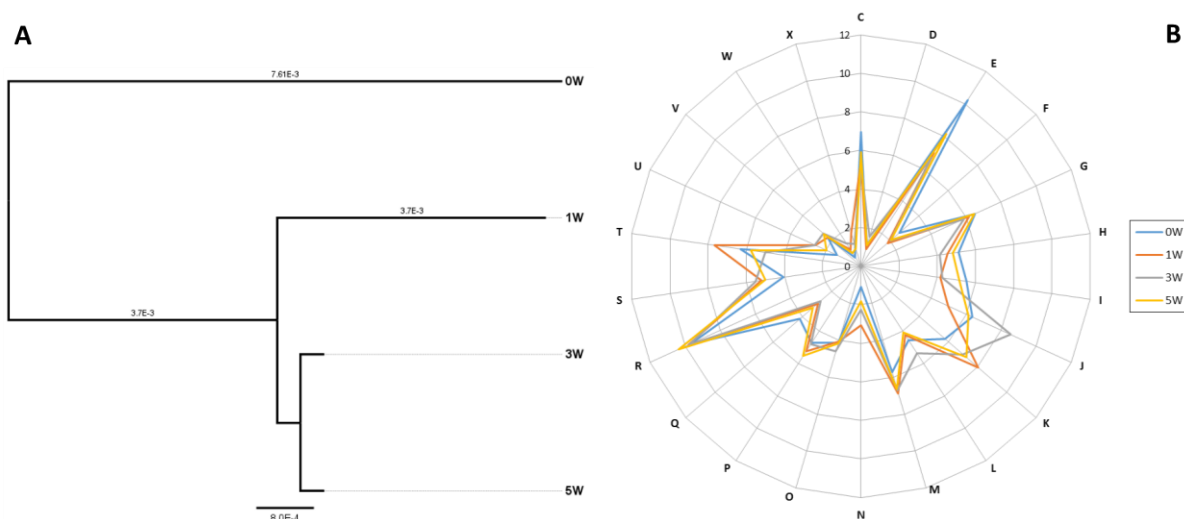
Fig. 3.6. A) Hierarchical clustering by COG function categories and B) distribution pattern of COG-assigned predicted genes. The estimated gene copies values were considered and the analysis was performed using the Function Category Comparison tool on the IMG web interface. C: Energy production and conversion; D: Cell cycle control, mitosis, and meiosis; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane biogenesis; N: Cell motility; O: Post-translational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking and secretion; V: Defense mechanisms; W: Extracellular structures; X: Mobilome prophages, transposons. The genes not assignable to any COGs are not shown in this figure.

The analysis based upon SEED database performed on the MG-RAST platform revealed similar patterns (Fig. 3.7). Metagenomic sequences were annotated against SEED level 1 subsystems. Among the metabolic categories, genes assigned to the metabolism of carbohydrates, amino acids and protein resulted the most abundant as it often happens in soil metagenomes (Uroz et al., 2013; Urich

et al., 2008), since they are related to the housekeeping functions of all living organisms. However, the comparative analysis of the metabolic categories relative distribution highlighted an enrichment in genes annotated into the groups "cell wall and capsule", "membrane transport", "virulence and defense" and "metabolism of aromatic compounds" in samples 1W, 3W and 5W.
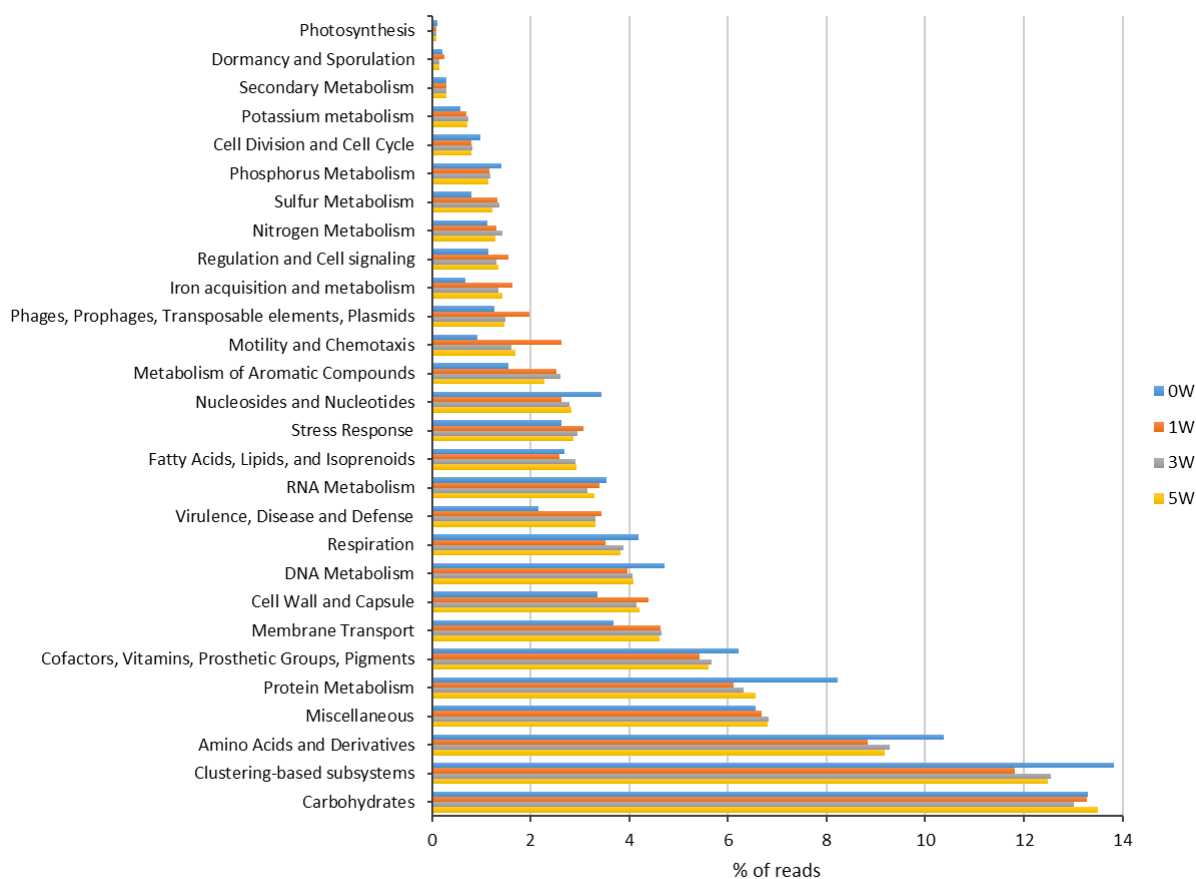
Fig. 3.7. Relative distribution of reads in major level 1 subsystems. Metagenomic data were annotated against SEED subsystems in MG-RAST based on a minimum percentage identity cut-off of 60%, E-value of 1e-5 and length of 15 bp.

Proteins classified in the metabolism of aromatic compounds were assigned to eight subsystems at level 2, among which peripheral pathway for catabolism of aromatic compounds was the most abundant, followed by metabolism of central aromatic intermediates (Fig. 3.8). In particular, the subcategory of the peripheral pathway catabolism, which comprise proteins involved in the degradation of phenol, phenylpropanoid, benzoate, naphthalene and phenyl-alkanoic acids derivatives, showed an increment in the enriched samples. The metabolism of central aromatic intermediates was mainly represented by the homogentisate pathway and by proteins involved in the N-heterocyclic aromatic compound degradation.

Metabolism of aromatic compounds
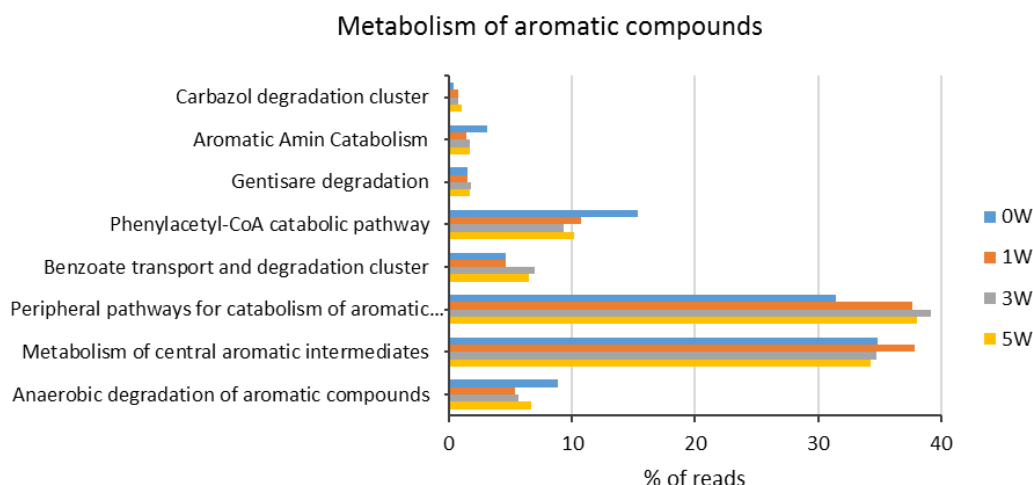
Fig 3.8. Distribution of sequences associated to subcategories of the aromatic compound metabolism.

The sequences classified in "Virulence, disease and defense" accounted for 2.16%, 3.44%, 3.32% and 3.32% of the total reads annotated by SEED subsystems in 0W, 1W, 3W and 5W respectively. Resistance to antibiotics and toxic compounds, important features for microbial survival and adaptation in response to the presence of toxic compounds, was the most abundant subcategory in each sample. The "Cell wall and capsule" category (0W: 3.35%, 1W: 4.40%, 3W: 4.16%, 5W: 4.22%) showed an increment in "bacterial peptidoglycan (*murein*) *hydrolases2"* and "Gram-negative cell wall components" level 2 subsystems in tCNSL amended samples, suggesting a possible enrichment in Gram-negative microorganisms which are known for their aromatics degradation capabilities. Furthermore, an increase in the abundance of the category "Ton and Tol transport systems" that includes TonB-dependent transporter proteins, which enable in Gram-negative bacteria the transport of trace elements, siderophores, heme and vitamin B12 was also observed (Noinaj et al., 2010). Carbohydrates, amino acids, organic acids and aromatic compounds are also TonB-dependent substrates, as observed for *Xanthomonas* and *Sphingomonas* bacteria (Miller et al., 2010; Tang et al., 2012; Blanvillain et al., 2007). In addition, the KEGG pathway for degradation of aromatic compounds was specifically analysed. Before enrichment, this specific pathway comprised only 30 KEGG orthology (KO) identifiers, whereas this number increased to 98 KO after 5 weeks of CNSL enrichment. The functional diversity of samples with respect to the degradation of aromatic compounds pathway was compared in order to estimate richness and evenness of the included KO groups; the Shannon diversity indexes 3.21 (0W), 3.78 (1W), 3.75 (3W), 3.81 (5W) suggested some enrichment in functional diversity of proteins potentially involved in degradation of a wide-range of aromatic compounds.

Altogether, the exposure to CNSL constituents caused significant changes in the taxonomic composition of microbial community and its genetic potential, with effects on the polyaromatics and monoaromatics catabolic.

3.3.4. Biodegradation network reconstruction

The metagenome analysis was also applied to pinpoint possible modifications in the genetic potential of degradation of aromatic compounds in the rhizosphere microbial community in response to the exposure to tCNSL. The search of enzymes putatively involved in the aerobic bacterial degradation of aromatics was carried out with the utilization of the AromaDeg database. This manually curated database contains sequences of key catabolic functions with validated biochemical functions that were used as query sequences in the BLAST search. Results that matched to one of the protein families within the database were filtered by amino acid sequence identity (>50%) and minimum alignment length (>50 amino acids) and associated to a putative catabolic function. Corresponding substrates and degradation products were then linked and the partial reconstruction of the theoretical aerobic aromatic catabolic routes was achieved. The complete list of substrates and definition of the abbreviations assigned to the catabolic functions can be found in the appendix (Table S3.1). The analysis allowed also the visualization and comparison of the relative abundance of the functions identified in the community at different timepoint of the enrichment. However, due to the limited sequence coverage, the reconstructed pathways were partially incomplete. Only one hit was retrieved from the community at timepoint zero (0W). This low number may be linked to the broad spectrum of carbon sources that are present in soil, as well as to the low number of assembled sequences due to the high microbial diversity resulting in low sequence coverage of metagenome (Gans et al., 2005; Howe et al., 2014). Moreover, this resulted in short and fragmented ORFs, difficult to use as queries for an unambiguous functional affiliation.

Differences could be observed in the relative abundance of genes for the main catabolic enzymes families (Fig. 3.9). Type I extradiol dioxygenase (EXDO I) superfamily comprises enzymes that catalyse the ring cleavage of mono- and polyaromatics such as catechol, benzene, toluene, biphenyl, naphthalene; the bicyclic substrates subfamily increased 2.4-fold and variations in the monocyclic and miscellaneous families were also observed. The extradiol dioxygenases (EXDO) of the cupin superfamily, comprising enzymes such as gentisate dioxygenases, increased 2-fold between week 1 and 3. The abundance of EXDO LigB superfamily, which includes protocatechuate and gallate dioxygenases, and the proteobacterial homoprotocatechuate 2,3-dioxygenases, varied in the range of

13-17% of total catabolic genes. Remarkable was the 3.8-fold increase registered in alkane hydroxylases.
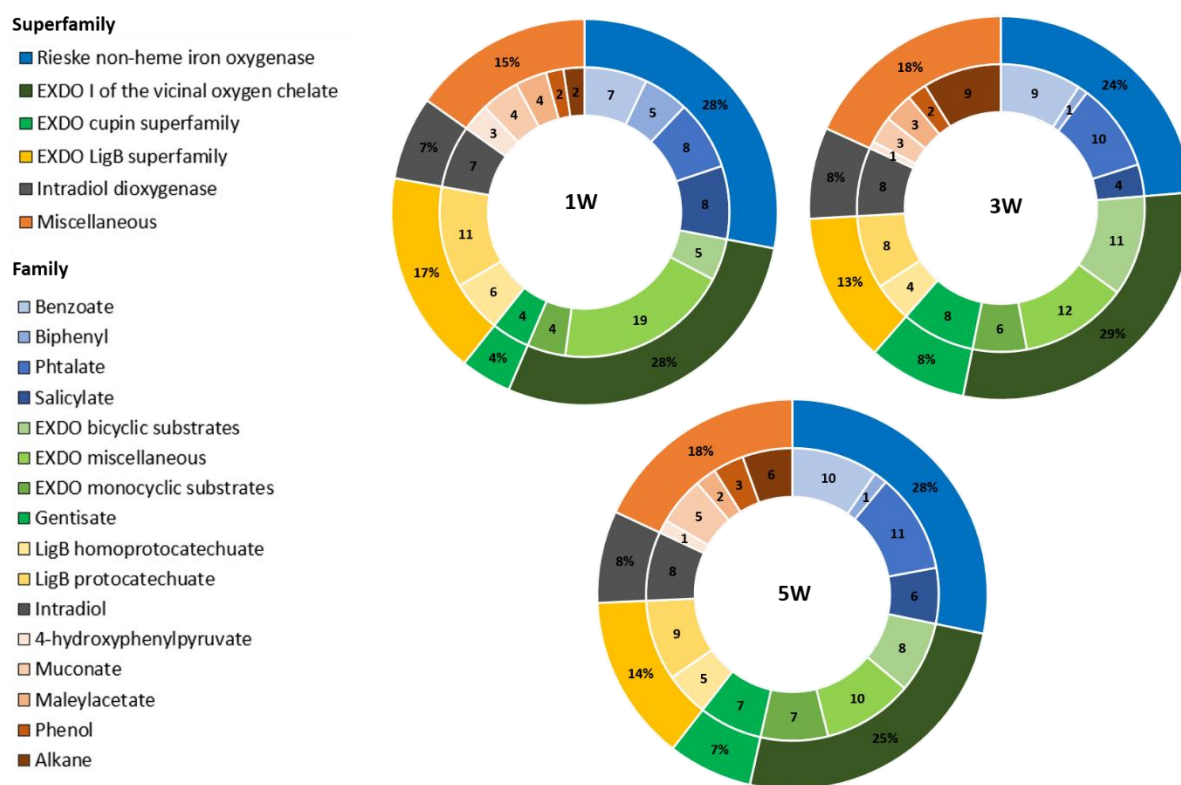
Fig. 3.9. Relative abundance and diversity of key catabolic enzyme groups involved in the degradation of aromatic compounds as annotated via comparison with the AromaDeg database.

Two distinct catabolic networks were built illustrating the degradation capabilities on a whole-community genomic scale of polyaromatic (Fig 3.10A) and monoaromatic compounds (Fig 3.10B). Catabolic function could be assigned to approximately 0.08% of the total predicted protein-coding sequences of 1W, 3W and 5W (257 protein-coding genes in 1W, 363 in 3W, and 471 in 5W) that matched with the validated database, covering 57 unique functions. Modifications in the relative abundance of functions associated with the catabolism of at least 18 substrates were found. However, no clear substrate specificity could be assigned to 26-30% of the retrieved enzymes (313 sequences in total), therefore not considered for the purpose of network reconstruction.
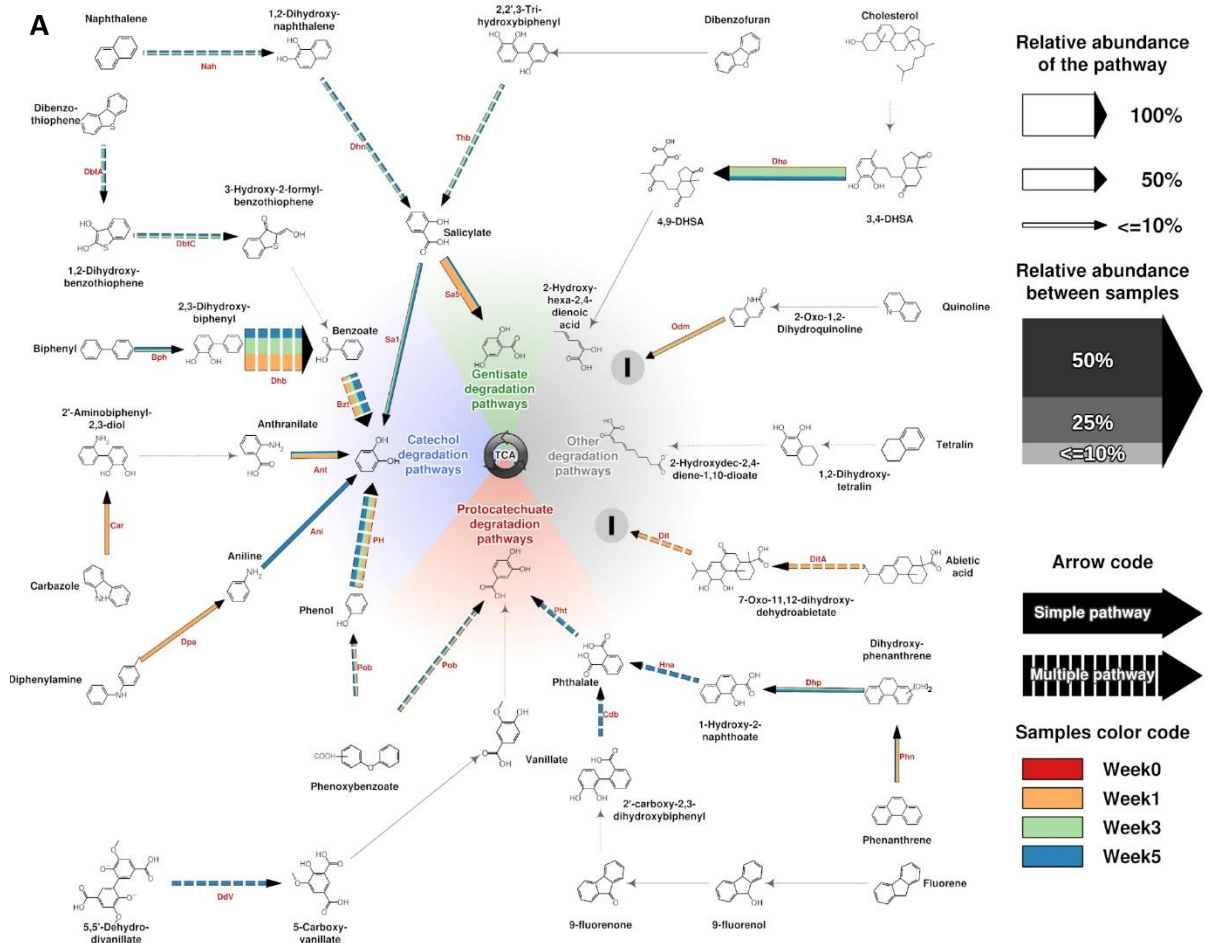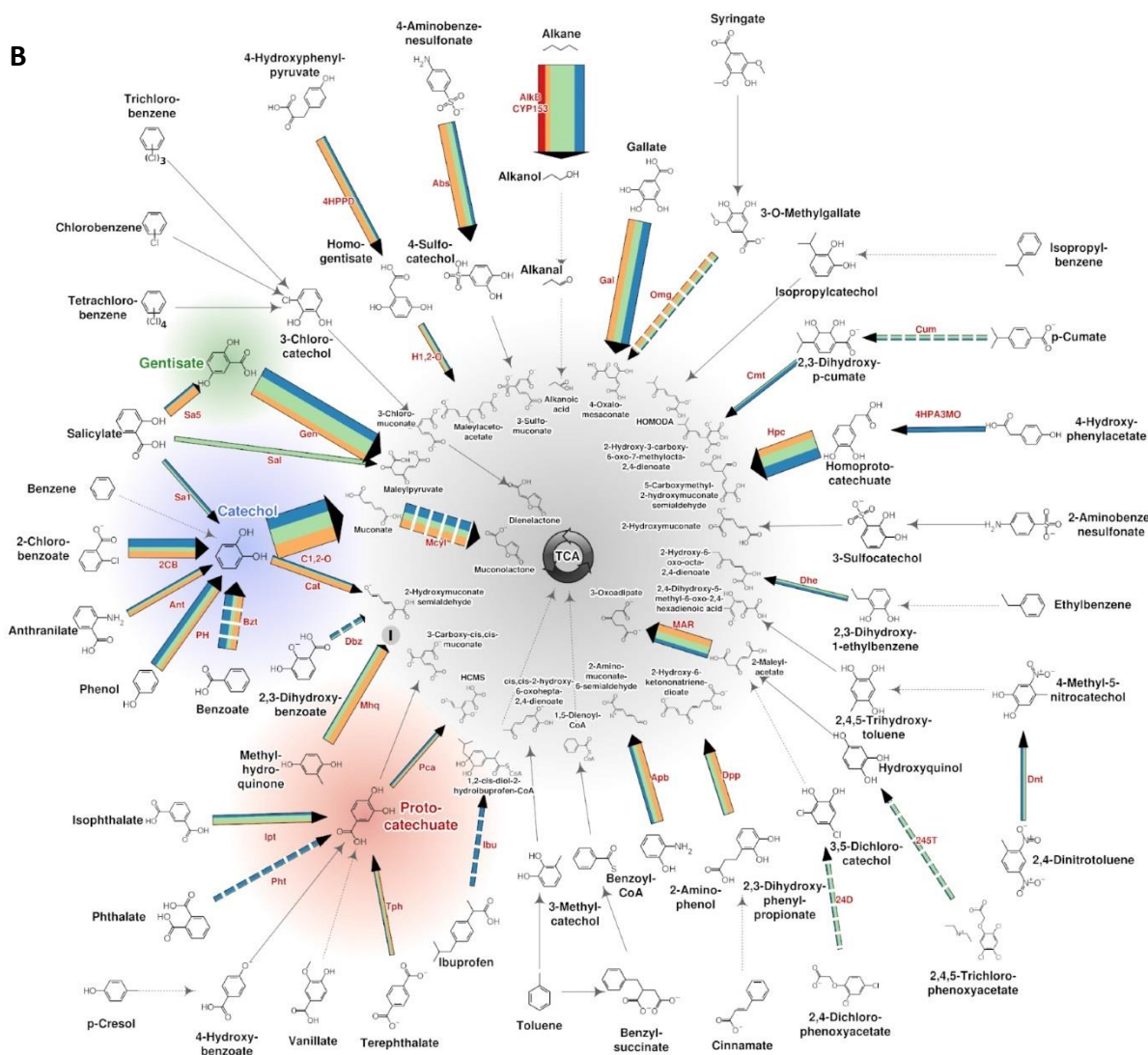
Fig. 3.10. Catabolic networks of (A) polyaromatics, (B) monoaromatics and alkenes in the CNSL-enriched sorghum rhizosphere microcosms sampled at week 0, week 1, week 3, week 5 (see colour code). Putative substrate and product of each degrading function were connected and an abbreviation was assigned to the catalysed reaction. The number of catabolic genes assigned to a degrading reaction is visualised by the thickness of the arrow. The values were normalised on the total number of protein coding genes in the given sample to avoid bias due to differences in sample size.

The catabolic network of polyaromatic compounds (Fig. 3.10A) emphasised benzoate, salicilate and phthalate as main intermediates of degradation. The majority of the annotated functions were coded by genes only detected or enriched in week 3 and 5. Benzoate appeared to be the product of the activity of biphenyl (Bph) and 2,3-dihydroxybiphenyl dioxygenases (Dhb); Dhb were encoded by the highest number of genes, equally abundant after one and three weeks of enrichment. Genomic evidence for enzymes associated with the potential degradation of naphthalene, by naphthalene and dihydroxynaphthalene dioxygenase (Nah, Dhn), and of 2,3',2-trihydroxybiphenyl (Thb) was found in

week 3 and week 5. Activities included in the phenanthrene and 2'-carboxy-2,3-dihydroxybiphenyl pathway (Phn, Dhp, Hna, Cdb) were responsible for the phthalate accumulation, which then enters in the protocatechuate degradation pathway through the action of phthalate 4,5-dioxygenases (Pht). A 11-fold increase from week 1 was detected in genes encoding catabolic enzymes associated with the putative degradation of 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione propionate (3,4-DHSA by Dha). Finally, a number of sequences of sample 1W matched with functions involved in the hydroxylation of diverse polyaromatic molecules (Car, Dpa, DitA, Dit, Odm), but no evidence of these genes was found in 3W and 5W.

The monoaromatic and alkene compounds network (Fig. 3.10B) showed a more diversified web of potential catabolic reactions funnelled into the Krebs cycle. As first observation, the degradation potential of the majority of the pollutants and their intermediates was equally detected in week 1, week 3 and week 5 microbial community. Nevertheless, a difference in the relative abundance of genes encoding for enzymes involved in particular degradation steps was observed, showing a positive correlation with the duration of enrichment. Catechol confirmed its role as a central intermediate in the degradation of cyclic hydrocarbons (Ladino-Orjuela et al., 2016). Genes encoding for benzoate (Bzt) and 2-chlorobenzoate dioxygenases (2CB), phenol hydroxylase (PH), and salicylate 1-hydroxylase (Sa1) showed positive correlation with the exposure time to tCNSL, contributing to the formation of catechol as common catabolic intermediate. Catechol and its derivatives are then cleaved by dioxygenases, either between the hydroxyl groups (*ortho*-cleavage) or adjacent to one of the hydroxyl group (*meta*-cleavage). The *ortho*-cleavage of catechol to muconate by catechol-1,2-intradiol dioxygenases (C1,2-O) appeared to be the preferential pathway, and C1,2-O resulted to be the most abundant gene class in the network. Evidence of the further transformation of muconate to muconolactore by carboxymuconate cycloisomerase (Mcyl) were also obtained. Genes encoding alkane monooxygenase/hydroxylase enzymes AlkB and CYP153 composed the second most abundant group, with the highest number of hits recorded after 3 weeks of enrichment (5-fold increase compared to 1W), likely suggesting that fatty moieties are more susceptible to degradation and mineralisation than other more recalcitrant aromatic components. Other genes related with biodegradation were accumulated in weeks 3 and 5 and were relevant to the conversion of other intermediates of aromatic degradation pathways: gentisate, by gentisate dioxygenases (Gen) with 2.3-fold change from 1W to 3W; gallate transformed by gallate dioxygenases (Gal) to 4-oxalomesaconate, and homoprotocatechuate metabolised by homoprotocatechuate 3,4-dioxygenases (Hpc).

On the contrary, a negative correlation with the time of incubation was found for functions associated with the metabolism of 4-hydroxyphenylpyruvate (4HPPD), 4-aminobenzenesulfonate (Abs), 3-O-methylgallate (Omg), catechol (by the *meta*-cleavage of catechol 2,3-dioxygenases Cat),

and with the functions involved in the degradation of 2,3-dihydroxyphenylpropionate and 2-aminophenol (Dpp and Apb). Genomic indications exclusive for samples 3W and 5W were found for activities related to the metabolism of phthalate (Pht), in the p-cumate pathway (Cum and Cmt), 2,3-dihydroxybenzoate (Dbz), and di- and trichlorophenoxyacetate (2,4-D and 2,4,5-T).

3.3.5. Correlation between community composition and phylogenetic identities of catabolic genes

We intended to project taxonomy information over the genes encoding catabolic enzymes used for network reconstruction in order to detect the main taxonomic groups involved in the pollutant degradation. Approximately 80% of genomic fragments could be assigned to a taxonomic qualifier at class level and about 70% at order level (Fig. 3.11); approximately 28-31% of sequences remained unclassified at the order level. This can be attributed to generally high diversity of microorganisms in the soil before enrichment resulting in a low sequence coverage and short lengths of both assemblies and predicted protein sequences, and also to the fact that the majority of soil microorganisms still remains uncharacterised (Fierer, 2017).
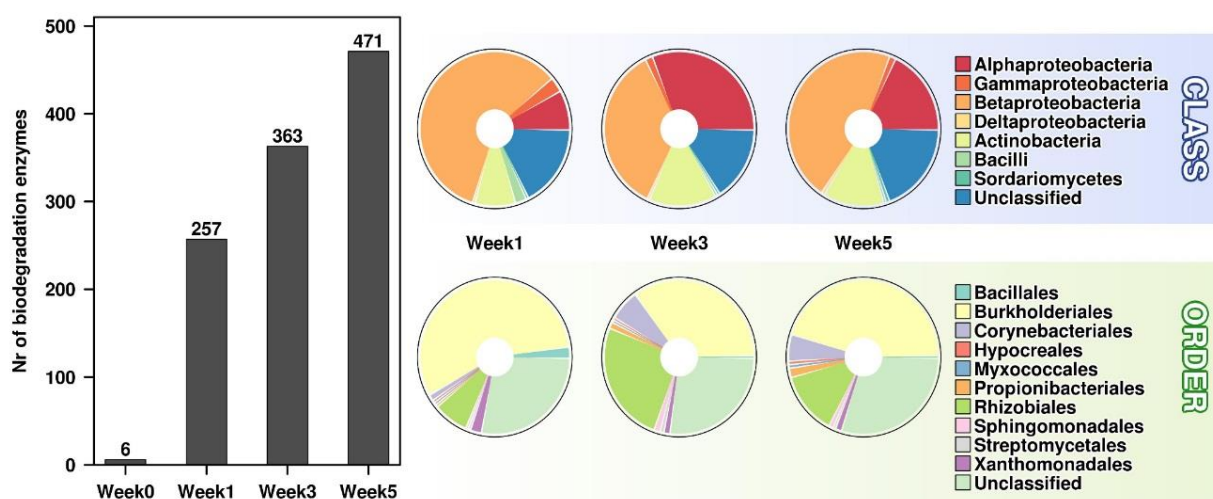
Fig. 3.11. Relative contribution of catabolic genes as per microbial lineage after enrichment and incubation for 1, 3 and 5 weeks. Left panel corresponds to the total number of orthologs matching AromaDeg database and meeting cut-off criteria of 50% amino acid sequence identity.

Key taxonomic groups included microorganisms closely related to members of *Alphaproteobacteria* (1W: 8%; 3W: 32%; 5W: 18%), *Betaproteobacteria* (1W: 61%; 3W: 37%; 5W:

48%), *Actinobacteria* (1W: 9%; 3W: 15%; 5W: 13%) and *Gammaproteobacteria* (1W: 3%; 3W: 1%; 5W: 1%). Members of *Burkholderiales* were the major contributors to the network (1W: 59%; 3W: 36%; 5W: 47%), followed by *Rhizobiales* (1W: 7%; 3W: 27%; 5W: 13%) and *Corynebacteriales* (1W: 1%; 3W: 6%; 5W: 5%), altogether covering the abilities to potentially transform the majority of the substrates. Little contribution was provided by *Xanthomonadales*, *Bacillales*, *Sphingomonadales*, *Propionibacteriales*, and *Streptomycetales*.

The contribution of the microbial species to each of the catabolic gene classes was studied (Fig. 3.12). Many enzymes were encoded simultaneously by different microbial groups suggesting the co-participation of them in the degradation of hydrocarbons, although the difference in relative abundance highlighted the predominance and important role of *Burkholderiales*. The taxonomic affiliation of proteins for 17 functions out of 62 was marked as unclassified and those were not included in the heatmap. The majority of catabolic functions appeared to be carried through enzymes encoded by betaproteobacterial *Burkholderiales* genomes. Alphaproteobacteria of the order *Rhizobiales* encoded enzymes for degradation of a number of substrates including homoprotocatechuate (by Hpc), alkene (by AlkB), catechol (by C1,2-O), 4-hydroxyphenylpyruvate (by AHPPD), 4-aminobenzenesulfonate (by Abs), maleylacetate (by MAR), and phenol (by PH). The degradation of gentisate (by Gen) appeared to be supported exclusively by *Burkholderiales* in 1-week microcosm, but the additional participation of *Rhizobiales* appeared after 3 weeks. *Corynebacteriales* and *Propionibacteriales* appeared to supply the genetic potential for transforming polyaromatics 2,3-dihydroxybiphenyl and 3,4-DHSA by Dhb and Dha after 3 and 5 weeks. As mentioned before, the catabolic function of a significant number of sequences (26-30%) was ambiguous and genes were not included in the network reconstruction. However, the phylogenetic assignment of these sequences was possible and indicated their similarity to the most represented *Burkholderiales*, *Rhizobiales* and *Corynebacteriales* (data not shown).

Interestingly, some functions were exclusively associated with specific microbial taxa. The enzymes responsible for the catabolism of the polyaromatics phenanthrene and naphthalene (by Phn, Nah), 2-chlorobenzoate (by 2CB), anthranilate (by Ant), biphenyl to catechol (by Bph, Bzt), dibenzothiophene (by DbtA), dinitrotoluene (by Dnt), 2,3-dihydroxyphenylpropionate (by Dpp), gallate and 3-O-methylgallate (by Gal and Omg), 3-carboxy-cis,cis-muconate (by Mcyl), methylhydroquinone (by Mhq), phthalate (Pht), and salicylate (Sa5) were attributed only to members of *Burkholderiales*. The catabolic pathway of *p*-cumate by Cum and Cmt, the degradation of 2,3-dihydroxybenzoate (by Dbz), 2,4,5-trichlorophenoxyacetate (by 245T) and 2,4-dichlorophenoxyacetate (by 24D) were found in members of *Rhizobiales*. Gammaproteobacterial *Xanthomonadales* exhibited the potential for degradation of 2-aminophenol (by Apa). These facts

indicate the necessity of microbial networking to enable the effective biodegradation of a multitude of natural and anthropogenic compounds in the soil and maintain its health and resilience.
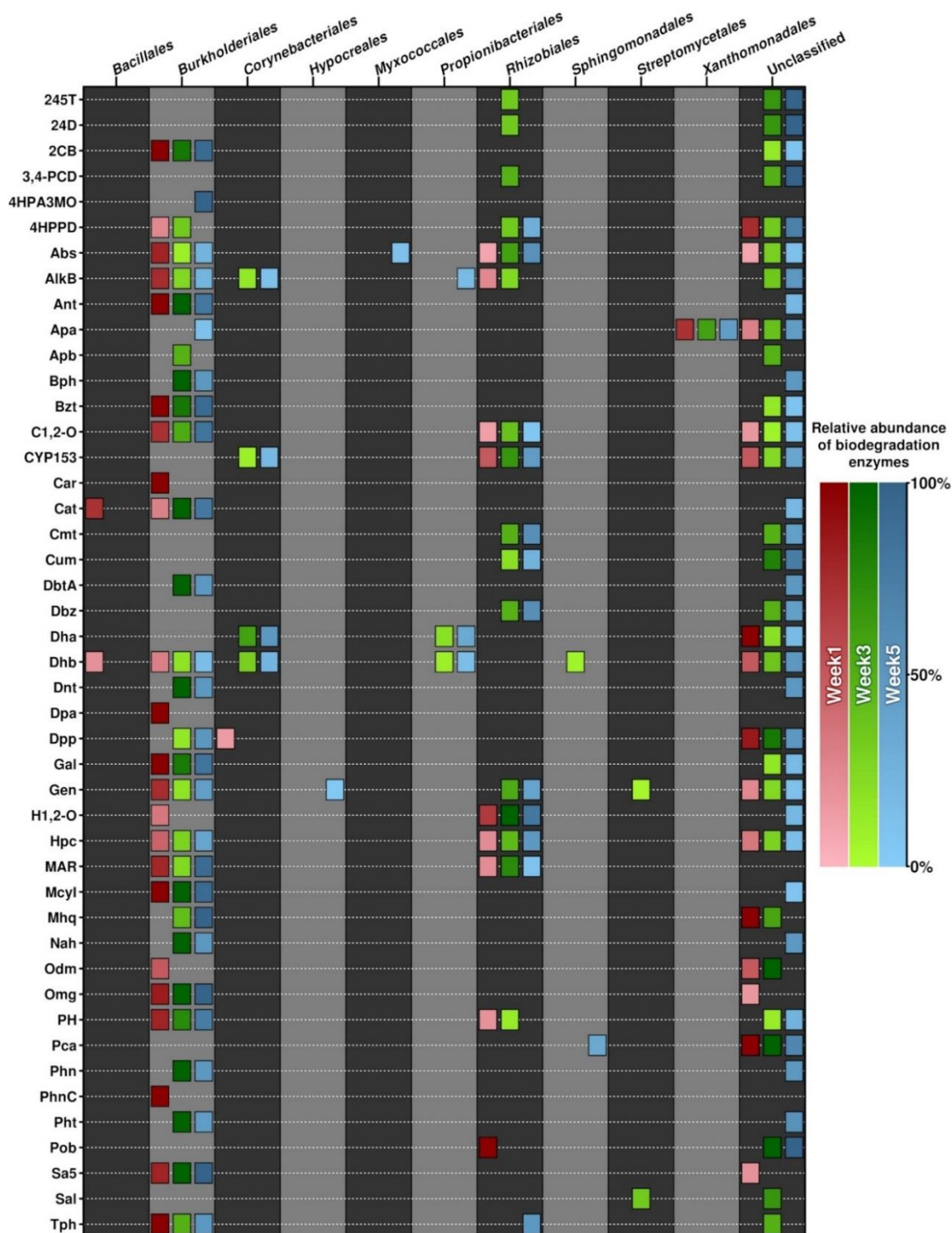
Fig. 3.12. Heat map showing the taxonomy affiliation at order level of the sequences used for the degradation network reconstruction differentiated per each catabolic function. Catabolic classes with ambiguous taxonomic assignation (only unclassified) were not included in the figure. Relative abundance of biodegrading enzymes was calculated on each function per week of enrichment.

3.4. Discussion

The rhizosphere is a unique environment harbouring complex microbial community that quickly respond to the changes in physical-chemical conditions (Philippot et al., 2013). The presence of anthropogenic pollutants in soil, such as heavy metals and hydrocarbons, has proven to impose a selective pressure to soil microbiome that favours the growth of metabolically and physiologically more adapted species. Bacteria are known to have evolved to metabolize diverse aromatic compounds including environmental pollutants, using these compounds as carbon and energy source (Nešvera et al., 2015). We postulated that the addition of CSNL rich in phenolic lipids would shape the Sorghum rhizosphere bacterial community favouring the growth of aromatic-degrading microbes. The final goal was to enrich in functions that interact and modify the main components of the cashew oil, in order to clone and express the corresponding enzymes which may be of interest for establishing new biotechnological processes.

After CNSL amendment in soil and a further microcosm enrichment performed in shaking flasks for 5 weeks, *de novo* sequencing of the microbial community allowed us to explore taxonomic and functional changes happening at specific interval of time. In the first part of this work, phylogenetic composition of the samples was delineated based on the distribution of the best BLAST hits of the annotated protein-coding genes. In the second part of the study, the modifications in response to the addition of cashew oil were investigated from a functional point of view, focusing in particular on the catabolic ability of the microbial community, with respect to polyaromatic and aromatic compounds. The results of the metagenomics analyses allowed us to make some considerations that are in agreement with the findings of previous metagenomics studies.

3.4.1. Phylogenetic analysis of microbial community composition

As a first observation, the analysis at phylum level identified *Proteobacteria, Actinobacteria* and *Bacteroidetes* as most represented groups, which has been identified as common inhabitants in the rhizosphere soil of crop plants (Li et al., 2014a), rice (Edwards et al., 2015), and sorghum plants (Oberholster et al., 2018). However, the comparison of samples at different timepoints of the enrichment highlighted the modification in the relative abundance of the taxonomic groups. The abundance of *Proteobacteria* remarkably increased in course of enrichment; this group comprises Gram-negative bacteria that has been extensively reported to participate in oil polluted environments to the degradation of petroleum hydrocarbons (Jurelevicius et al., 2013; Zhang et al., 2017).

Actinobacteria was the second most abundant phylum, although its abundance decreased 2-fold after 5 weeks compared to timepoint zero. *Actinobacteria* are well studied for their aromatic catabolic pathways (Pérez-Pantoja et al., 2010), playing an important role in bioremediation (Chen et al., 2015; Zhang et al., 2012) and plant growth promotion (Palaniyandi et al., 2013). Members of *Bacteroidetes* are common in both marine and terrestrial ecosystems and they are specialized in the degradation of complex organic matter, including petroleum hydrocarbons (Thomas et al., 2011; Drury et al., 2013). In the proteobacterial community, *Beta-* and *Gammaproteobacteria* were dominated by members of *Burkholderiales* and *Xanthomonadales* respectively, and were favoured over *Alphaproteobacteria* group that was instead equally represented by *Sphingomonadales* and *Rhizobiales* (genera *Bradyrhizobium* and *Mesorhizobium*). Although *Actinobacteria* became less represented after CNSL enrichment, few changes within this phylum were observed, with the rise of *Mycobacterium* and *Nocardioides* to become the most abundant genera. The phylum *Bacteroidetes* had a significant increase in the abundance of sequences associated to orders *Chitinophagales*, *Cytophagales* and *Sphingobacteriales.* This large group of bacteria, formerly classified under the group "Cytophaga-Flavobacterium-Bacteroides", is widespread in the biosphere and it has been found in forest soil, rhizosphere soil and root endophytes (Madhaiyan et al., 2015). Overall, these findings appear to be consistent with other studies conducted on rhizosphere samples. Genera such as *Rhizobium*, *Sphingomonas*, *Burkholderia* and *Pseudomonas* are known to contain strains with plant growth-promoting abilities (Mendes et al., 2013), and a similar predominance of bacteria was also discovered in petroleum-contaminated sites.

Members of *Burkholderiales* are frequently occurring in the rhizosphere of crop plants, beneficially associated with them, as well as in a variety of polluted sites, which correlates with their potential for biodegradation of recalcitrant organic compounds. A number of *Burkholderia* species are known to possess versatile catabolic traits that enable them to degrade root exudates, recalcitrant and xenobiotic pollutants (O'Sullivan & Mahenthiralingam, 2005; Suárez-Moreno et al., 2012) including phenanthrene and naphthalene (Seo et al., 2006; Kang et al., 2003; Kim et al., 2003). The *Burkholderia cepacia* complex (BCC) is a group of nine closely related bacterial species that have useful properties in the natural environment as plant pest antagonists, plant growth promoters, and degradative agents of toxic substrates (Chiarini et al., 2006; Revathy et al., 2015). Nearly all BCC species have been found in the rhizosphere of important plant crops (Ramette et al., 2005), and reported to degrade a variety of environmental contaminants, with degradation processes influenced by pH, temperature, concentration and inoculum size. Catabolic abilities of *Burkholderia fungorum* strain FLU100 were recently demonstrated with respect to mineralization of toluene and any mixture of benzene, chlorobenzene, bromobenzene, fluorobenzene, and iodobenzene. The aromatic ring is

initially metabolised by a dioxygenase and a dehydrogenase to 3-substituted catechols; complete degradation occurred via a ortho-cleavage pathway yielding 2-methyl-2-enelactone and further metabolites (Dobslaw & Engesser, 2015). *Burkholderia vietnamiensis* strain G4 is known to mediate crop protection and growth promotion (Gillis et al., 1995), as well as to degrade trichloroethylene, an industrial solvent and degreasing agent contaminating soil and underground water, along with many other aromatic hydrocarbons (Nelson et al., 1987). *B. vietnamiensis* G4 biofilm was successfully used for trichloroethylene treatment in bioreactors (Lee et al., 2003). Its catabolic plasmid pTOM carries all the genes required to degrade phenol, toluene, o-cresol, m-cresol and benzene (O'Sullivan et al., 2007). *Burkholderia xenovorans* LB400 is considered the model organism for aerobic polychlorinated biphenyl (PCB) biodegradation (Goris et al., 2004). Firstly isolated from a PCB-contaminated soil from a landfill New York State and identified as *Pseudomonas sp.* LB400 (Bopp, 1986), the organisms has been recently reclassified as *Parabulkholderia xenovorans* LB400 (Sawana et al., 2014). This strain co-metabolizes many PCB congeners (Gibson et al., 1993), as well as isoflavonoids (Seeger et al., 2003), diterpenoids (Smith et al., 2007) and sulfonates (Ruff et al., 2003). Eleven central aromatic catabolic pathways and twenty peripheral aromatic catabolic pathways have been identified by in silico analysis of the large 9.73 Mbp LB400 genome (Chain et al., 2006). Evidence of the functionality of the homogentisate and homoprotocatechuate pathways were found for the growth of LB400 on 3-hydroxyphenylacetate and 4-hydroxyphenylacetate as sole carbon source (Méndez et al., 2011). *Burkholderia* sp. FDS-1 was isolated from soils contaminated by fenitrothion, a nitrophenolic pesticide, and described as an effective degrader of this contaminant (Zhonghui et al., 2005). Furthermore, microcosm experiments with fenitrothion-spiked soil highlighted the good performance in the pesticide biodegradation in a range of pollutant concentrations, soil pH and temperature, with results showing a complete degradation of fenitrothion by FDS-1 up to 50 mg kg^{-1} soil (Hong et al., 2007).

The second most abundant order after 5 weeks of enrichment was the gammaproteobacterial *Xanthomonadales*, which comprises important hydrocarbon-degrading bacteria frequently found in terrestrial and marine environments contaminated with petrochemicals. Interestingly, some members of this order have been described as obligate hydrocarbonoclastic bacteria for their ability to utilise exclusively hydrocarbons as carbon source (Gutierrez, 2017). After microcosm enrichment, this group was dominated by members of the families *Rhodanobacteraceae* (genera *Frateuria*, *Rhodanocabter*, *Luteibacter* and *Dyella*) and *Xanthomonadaceae* (genera *Lysobacter* and *Xanthomonas*). A stable isotope probing study with ^{13}C-labeled biphenyl, benzoate and naphthalene (Uhlik et al., 2012) reported the bacterial population of a long-term hydrocarbon contaminated soil. *Proteobacteria* were the principle actors in mediating biotransformation of the aromatic compounds, with members of *Rhodanobacteraceae* and *Burkholderia* as the main consumers of biphenyl and benzoate. The strain

Dyella ginsengisoli LA-4 isolated from activated sludge was reported to efficiently degrade biphenyl up to concentration of 500 mg L^{-1} within 40 h (Li et al., 2009). Phenanthrene degradation was also observed by *Dyella* sp., *Luteibacter* sp., and *Burkholderia* sp. isolated from mangrove sediments (Muangchinda et al., 2013).

3.4.2. Functional annotation and taxonomic affiliation of catabolic genes

Data analysis using COG, KEGG and SEED database showed that the majority of genes were involved in routine microbial activities, such as metabolism of carbohydrates, amino acids, cofactors and vitamins. A small portion of genes was involved in the adaptation of bacteria to hydrocarbon compounds and in the degradation of them; this functional category displayed an increase of relative abundance in samples with longer exposure to tCNSL. First steps of the aerobic degradation of aromatic compounds (as part of the so-called upper pathway) involve the activation of the aromatic ring by oxidation catalysed by mono- or dioxygenases (Ladino-Orjuela et al., 2016), typically belonging to Rieske non-heme iron oxygenases superfamily (Gibson & Parales, 2000), flavoprotein monooxygenases (van Berkel et al., 2006) or soluble diiron multicomponent oxygenases (Leahy et al., 2003). A huge variety of hydrocarbon compounds is consequently transformed to central di- or trihydroxylated intermediates, which can be catechols or hydroxyl-substituted aromatic carboxylic acids. Further aerobic degradation of the intermediates (lower pathway) requires ring-cleavage catalysed by intradiol and extradiol dioxygenases, and the products of degradation are then directed into the tricarboxylic acid cycle. Enzymes catalysing the extradiol ring cleavage can be categorised in three superfamilies: type I extradiol dioxygenases of the vicinal oxygen chelate superfamily (Gerlt & Babbitt, 2001), type II or LigB superfamily (Sugimoto et al., 1999), and the type III cupin superfamily (Dunwell et al., 2000).

The annotated protein-coding genes from the sequenced metagenomes were compared to the manually curated AromaDeg database and presumptive functions were assigned to a number of sequences that showed similarity to key enzymes known for their role in the aerobic metabolism of aromatic compounds. The taxonomic affiliation of the catabolic genes was also performed in the attempt to highlight the contribution of particular set of microbes to the reconstructed catabolic network. This strategy was previously applied in metagenomic and metaproteomic studies that investigated the biodegradation potential of chronically crude oil-contaminated marine sediments (Bargiela et al., 2015b, 2015a; Guazzaroni et al., 2013). The analysis suggested that the microbial population after CNSL enrichment have the presumptive capacity to degrade more than 60 substrates, and the catabolic networks were built accordingly. A convergence of degradation reactions towards

the formation of central metabolites including catechol, gentisate, protocatechuate and homoprotocatechuate was observed. In particular, catechol confirmed its role as key intermediate in the biphenyl/benzoate, 2-chlorobenzoate, phenol, anthranilate and salicylate pathways. Subsequently, the intradiol cleavage of catechol was favoured over the *meta*-cleavage. Intradiol dioxygenases are almost exclusively found in *Actinobacteria* and *Proteobacteria*, and all *Burkholderia* genomes showed presence of catechol pathway (Pérez-Pantoja et al., 2010). Alkane hydroxylating enzymes and CYP153 cytochrome P450 monooxygenases, which are responsible for the oxidation of n-alkanes to their corresponding alkanols, were abundant and enriched in the 3W sample

The result of the taxonomic affiliation of the catabolic genes agreed, to some extents, with the taxonomic reconstruction based on BLASTP affiliation of the protein-coding genes. Similarly, the microorganisms that contributed the most to the degradation network were among the most abundant members in the established microcosms after 3-5 weeks of CNSL enrichment. *Burkholderiales* was appointed as a keystone taxon within the network, contributing for the majority of the catabolic enzymes, followed by members of *Rhizobiales*. Interestingly, *Rhizobiales* catabolic potential appeared to be mostly associated with functions of the lower pathway. Well-known for their nitrogen-fixation role as well as for provision of nutrients, phytohormones and precursors for plant metabolites, members of *Rhizobiales* have been also reported as capable to degrade aromatic compounds (Carvalho et al., 2006; Huang et al., 2013; Pagé et al., 2015; Li et al., 2014b). In studies conducted on ryegrass, bird's foot trefoil and willow rhizosphere members of the order *Rhizobiales* (e.g. *Bradyrhizobium*) proved to play important role in the alkane 1-monooxygenase- and cytochrome P450 monooxygenase-driven rhizoremediation of alkanes (Yousaf et al., 2010; Pagé et al., 2015). However, groups that resulted enriched after microcosm experiment, such as *Xanthomonadales* (e.g. *Frateuria*), *Spinghomonadales* (e.g. *Sphingomonas*) and *Corynebacteriales* (e.g. *Mycobacterium*), seemed not to contribute directly to the catabolic functions reviewed in the network. Considering that approximately 31% of the catabolic genes remained unclassified, it is important to underline that the effective utilization and catabolism of contaminants has to be considered the result of the interaction of multiple organisms in a specified environment. The complexity of polycyclic aromatic hydrocarbons, for example, imposes the cooperation of whole communities of bacteria and eukaryotes for an effective breakdown of these molecules. As previously reported, the cooperative interactions enhances the degradation rates by mechanisms of co-metabolism, synthesis of bio-surfactants, and adhesion via biofilm (Zhang et al., 2013; Yang et al., 2016; Kanaly & Harayama, 2000). Furthermore, the beneficial effect of microbe co-operation has been extensively described in plant rhizosphere, where the degradation of environmental contaminants is generally enhanced (Pizarro-Tobías et al., 2015). Congruently with the result of our investigation, previous experimental data demonstrated the

co-occurrence and cooperation of *Sphingomonas* and *Burkholderia* in degrading consortia (Zhang et al., 2013). The biodegradation of diesel fuel constituents by soil microbial community occurred only through the cooperation of members of *Burkholderia, Sphingomonas, Mycobacterium and Rhodanobacter,* with the latter (*Rhodanobacter* sp. strain BPC1) identified as the most active organism involved in the solubilisation and mineralization of benzo[α]pyrene (Kanaly et al., 2002). The gammaproteobacterial *Lysobacter*, as part of a bacterial guild comprising different genera, was found to be positively correlated with petroleum hydrocarbons removal thanks to the ability to produce biosurfactants that increase the bioavailability of lipophilic and oily compounds enhancing their degradation (Hayward et al., 2010; Susilaningsih et al., 2013). The beneficial impact of endophytes and plant growth-promoting rhizobacteria (PGPR) in the transformation of contaminants was demonstrated for bacteria such as *Pseudomonas*, *Burkholderia*, *Bacillus*, *Enterobacter*, *Rhizobium* (Badri et al., 2009; Wenzel, 2009; Wang et al., 2010). The inoculation of PGPR boosted petroleum removal efficiency during phytoremediation in an aged petroleum contaminated soil (Hou et al., 2015).

It should be noted that in our study no analysis of metabolites (intermediates) was conducted to confirm these *in silico*, therefore the reconstructed network has to be considered as the assessment of genomic potential for biodegradation. It was furthermore proven to be useful for pinpointing genes for functional enzymes that increased in numbers upon incubation with CNSL, which have been successfully cloned, expressed and characterised (see Chapter IV and V). The combination of the results of the conducted analyses on the established metagenomes, allowed us to make some considerations that are in line with the findings of previous metagenomic studies.

3.5. Conclusions

Metagenomics approach applied to the soil slurry enrichment experiment, integrated with the analysis of the large dataset generated, offered the possibility to unravel the soil microbial community complexity and its response to the perturbation caused by cashew oil addition. Key bacterial members inhabiting sorghum rhizosphere were identified and the differences prior to and after exposure to the phenolic lipids contained in cashew oil were highlighted; the modification of the indigenous composition was observed and the establishment of typical aromatic hydrocarbon degraders was promoted. Along with changes in community structure, the metagenomics-based catabolic network reconstruction allowed us to predict the degradation capacities of the community and the role of individual taxa in the utilization of aromatic substrates.

The most abundant microorganisms were classified predominantly into Gram-negative *Proteobacteria* (*Sphingomonas*, *Burkholderia*, *Bradyrhizobium*, *Mesorhizobium*, *Rhodobacteraceae*) and *Actinobacteria* (*Mycobacterium, Nocardioides*), known to typically colonise rhizosphere habitats and to be involved in the degradation of polluting compounds. The increment of these bacteria correlated with the increase in the relative abundance of functions for important biodegradation pathways. Genes encoding enzymes involved in the peripheral pathways of aerobic catabolism of various phenolic compounds in bacteria (e.g. gentisate, benzoate, phenol, homoprotocatechuate) showed a positive correlation with time of enrichment. The degradation reactions appeared to be carried out mainly by bacteria of *Burkholderiales* and *Rhizobiales*, although 30% of the sequences remained unclassified. Results that suggest CNSL to be a promoting agent for bacteria with the capability of degrading aromatic compounds, as well as for bacteria beneficial for mobilizing nutrients and increasing availability of organic substrates.

In addition, the *in silico* analysis of the metagenome datasets supported further significant outcomes described in the next chapters. The annotation of protein-coding genes with insights into function and conserved domains allowed to shortlist enzymes putatively capable of modifying the constituents of CNSL for the fulfilment of the sorgoleone synthesis pathway (Chapter IV). The genomic material was also used for the construction of fosmid libraries utilised for activity-based screening of industrially relevant enzymes (Chapter V).

In conclusion, we demonstrated for the first time the effects of cashew oil contamination on the indigenous bacterial community in rhizosphere environment through metagenomics approach. In this context, it would be interesting to conduct experimental validation assays to prove the extent of agreement with the metagenomics-based predictions.

3.6. References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 25 (17). p.pp. 3389–402.

Anderson, T.A., Guthrie, E.A. & Walton, B.T. (1993). Bioremediation in the rhizosphere. *Environmental Science & Technology*. 27 (13). p.pp. 2630–2636.

Badri, D. V., Weir, T.L., van der Lelie, D. & Vivanco, J.M. (2009). Rhizosphere chemical dialogues: plant-microbe interactions. *Current Opinion in Biotechnology*. 20 (6) p.pp. 642–650.

Bargiela, R., Gertler, C., Magagnini, M., Mapelli, F., Chen, J., Daffonchio, D., Golyshin, P.N. & Ferrer, M. (2015a). Degradation Network Reconstruction in Uric Acid and Ammonium Amendments in Oil-Degrading Marine Microcosms Guided by Metagenomic Data. *Frontiers in microbiology*. 6. p.p. 1270.

Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., Richter, M., Del Pozo, M. V, Cappello, S., Gertler, C., Genovese, M., Denaro, R., Martínez-Martínez, M., Fodelianakis, S., Amer, R.A., Bigazzi, D., Han, X., Chen, J., Chernikova, T.N., Golyshina, O. V, Mahjoubi, M., Jaouanil, A., Benzha, F., Magagnini, M., Hussein, E., Al-Horani, F., Cherif, A., Blaghen, M., Abdel-Fattah, Y.R., Kalogerakis, N., Barbas, C., Malkawi, H.I., Golyshin, P.N., Yakimov, M.M., Daffonchio, D. & Ferrer, M. (2015b). Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Scientific reports*. 5. p.p. 11651.

Bauer, J.E. & Capone, D.G. (1988). Effects of co-occurring aromatic hydrocarbons on degradation of individual polycyclic aromatic hydrocarbons in marine sediment slurries. *Applied and environmental microbiology*. 54 (7). p.pp. 1649–55.

van Berkel, W.J.H., Kamerbeek, N.M. & Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *Journal of Biotechnology*. 124 (4). p.pp. 670–689.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., Vasse, J., Lauber, E. & Arlat, M. (2007). Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PloS one*. 2 (2). p.p. e224.

Bopp, L.H. (1986). Degradation of highly chlorinated PCBs byPseudomonas strain LB400. *Journal of Industrial Microbiology*. 1 (1). p.pp. 23–29.

Bossio, D.A. & Scow, K.M. (1998). Impacts of Carbon and Flooding on Soil Microbial Communities: Phospholipid Fatty Acid Profiles and Substrate Utilization Patterns. *Microbial Ecology*. 35 (3). p.pp. 265–278.

de Brito, A.M., Gagne, S. & Antoun, H. (1995). Effect of compost on rhizosphere microflora of the tomato and on the incidence of plant growth-promoting rhizobacteria. *Applied and environmental microbiology*. 61 (1). p.pp. 194–9.

Brooijmans, R.J.W., Pastink, M.I. & Siezen, R.J. (2009). Hydrocarbon-degrading bacteria: the oil-spill clean-up crew. *Microbial biotechnology*. 2 (6). p.pp. 587–94.

Carvalho, M.F., Ferreira, M.I.M., Moreira, I.S., Castro, P.M.L. & Janssen, D.B. (2006). Degradation of fluorobenzene by Rhizobiales strain F11 via ortho cleavage of 4-fluorocatechol and catechol. *Applied and Environmental Microbiology*. 72 (11). p.pp. 7413–7417.

Chain, P.S.G., Denef, V.J., Konstantinidis, K.T., Vergez, L.M., Agulló, L., Reyes, V.L., Hauser, L., Córdova, M., Gómez, L., González, M., Land, M., Lao, V., Larimer, F., LiPuma, J.J., Mahenthiralingam, E.,

Malfatti, S.A., Marx, C.J., Parnell, J.J., Ramette, A., Richardson, P., Seeger, M., Smith, D., Spilker, T., Sul, W.J., Tsoi, T. V, Ulrich, L.E., Zhulin, I.B. & Tiedje, J.M. (2006). Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (42). p.pp. 15280–7.

Chen, I.M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., Varghese, N., Hadjithomas, M., Tennessen, K., Nielsen, T., Ivanova, N.N. & Kyrpides, N.C. (2017). IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Research*. 45 (D1). p.pp. D507–D516.

Chen, M., Xu, P., Zeng, G., Yang, C., Huang, D. & Zhang, J. (2015). Bioremediation of soils contaminated with polycyclic aromatic hydrocarbons, petroleum, pesticides, chlorophenols and heavy metals by composting: Applications, microbes and future research needs. *Biotechnology Advances*. 33 (6) p.pp. 745–755.

Chen, S.H. & Aitken, M.D. (1999). Salicylate stimulates the degradation of high-molecular weight polycyclic aromatic hydrocarbons by Pseudomonas saccharophila P15. *Environmental Science and Technology*. 33 (3). p.pp. 435–439.

Chiarini, L., Bevivino, A., Dalmastri, C., Tabacchioni, S. & Visca, P. (2006). Burkholderia cepacia complex species: health hazards and biotechnological potential. *Trends in microbiology*. 14 (6). p.pp. 277–86.

Copley, S.D. (2003). Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Current Opinion in Chemical Biology*. 7 (2) p.pp. 265–272.

Copley, S.D. (2009). Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nature chemical biology*. 5 (8). p.pp. 559–66.

Dalton, H., Stirling, D.I. & Quayle, J.R. (1982). Co-Metabolism [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 297 (1088). p.pp. 481–496.

Dayan, F.E., Kagan, I. a & Rimando, A.M. (2003). Elucidation of the biosynthetic pathway of the allelochemical sorgoleone using retrobiosynthetic NMR analysis. *The Journal of biological chemistry*. 278 (31). p.pp. 28607–11.

Dayan, F.E., Rimando, A.M., Pan, Z., Baerson, S.R., Gimsing, A.L. & Duke, S.O. (2010). Sorgoleone. *Phytochemistry*. 71 (10). p.pp. 1032–9.

Dobslaw, D. & Engesser, K.-H. (2015). Degradation of toluene by ortho cleavage enzymes in Burkholderia fungorum FLU100. *Microbial Biotechnology*. 8 (1). p.pp. 143–154.

Drury, B., Rosi-Marshall, E. & Kelly, J.J. (2013). Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Applied and Environmental Microbiology*. 79 (6). p.pp. 1897–1905.

Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H. & Pieper, D.H. (2014). AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database : the journal of biological databases and curation*. 2014. p.p. bau118.

Dunwell, J.M., Khuri, S. & Gane, P.J. (2000). Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiology and molecular biology reviews : MMBR*. 64 (1). p.pp. 153–79.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S., Eisen, J.A. & Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proceedings of the National Academy of Sciences*. 112 (8).

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*. 15 (10). p.pp. 579–590.

Fierer, N. & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*. 103 (3).

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*. 44 (D1). p.pp. D279–D285.

Gans, J., Wolinsky, M. & Dunbar, J. (2005). Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science*. 309 (5739).

George, J. & Kuttan, R. (1997). Mutagenic, carcinogenic and cocarcinogenic activity of cashewnut shell liquid. *Cancer Letters*. 112 (1). p.pp. 11–16.

Gerlt, J.A. & Babbitt, P.C. (2001). Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annual Review of Biochemistry*. 70 (1). p.pp. 209–246.

Gibson, D.T., Cruden, D.L., Haddock, J.D., Zylstra, G.J. & Brand, J.M. (1993). Oxidation of polychlorinated biphenyls by Pseudomonas sp. strain LB400 and Pseudomonas pseudoalcaligenes KF707. *Journal of bacteriology*. 175 (14). p.pp. 4561–4.

Gibson, D.T. & Parales, R.E. (2000). Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Current Opinion in Biotechnology*. 11 (3) p.pp. 236–243.

Gillis, M., VAN VAN, T., BARDIN, R., GOOR, M., HEBBAR, P., WILLEMS, A., SEGERS, P., KERSTERS, K., HEULIN, T. & FERNANDEZ, M.P. (1995). Polyphasic Taxonomy in the Genus Burkholderia Leading to an Emended Description of the Genus and Proposition of Burkholderia vietnamiensis sp. nov. for N2-Fixing Isolates from Rice in Vietnam. *International Journal of Systematic Bacteriology*. 45 (2). p.pp. 274–289.

Glick, B.R. (2010). Using soil bacteria to facilitate phytoremediation. *Biotechnology Advances*. 28 (3) p.pp. 367–374.

Goldfarb, K.C., Karaoz, U., Hanson, C.A., Santee, C.A., Bradford, M.A., Treseder, K.K., Wallenstein, M.D. & Brodie, E.L. (2011). Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical Recalcitrance. *Frontiers in Microbiology*. 2. p.p. 94.

Golyshina, O. V., Toshchakov, S. V., Makarova, K.S., Gavrilov, S.N., Korzhenkov, A.A., La Cono, V., Arcadi, E., Nechitaylo, T.Y., Ferrer, M., Kublanov, I. V., Wolf, Y.I., Yakimov, M.M. & Golyshin, P.N. (2017). 'ARMAN' archaea depend on association with euryarchaeal host in culture and in situ. *Nature Communications*. 8 (1). p.p. 60.

Goris, J., Vos, P. De, Caballero-Mellado, J., Park, J., Falsen, E., III, J.F.Q., Tiedje, J.M. & Vandamme, P. (2004). Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as Burkholderia xenovorans sp. nov. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*. 54 (5). p.pp. 1677–1681.

Guazzaroni, M.-E., Herbst, F.-A., Lores, I., Tamames, J., Peláez, A.I., López-Cortés, N., Alcaide, M., Del Pozo, M. V, Vieites, J.M., von Bergen, M., Gallego, J.L.R., Bargiela, R., López-López, A., Pieper, D.H., Rosselló-Móra, R., Sánchez, J., Seifert, J. & Ferrer, M. (2013). Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. *The ISME journal*. 7 (1). p.pp. 122–36.

Gutierrez, T. (2017). Aerobic Hydrocarbon-Degrading Gammaproteobacteria: Xanthomonadales. In: *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*. Cham: Springer International Publishing, pp. 1–15.

Hayward, A.C., Fegan, N., Fegan, M. & Stirling, G.R. (2010). Stenotrophomonas and Lysobacter :

ubiquitous plant-associated gamma- proteobacteria of developing significance in applied microbiology. *Journal of Applied Microbiology*. 108 (3). p.pp. 756–770.

Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L., Stringfellow, W.T., Bill, M., Conrad, M.E., Tom, L.M., Chavarria, K.L., Alusi, T.R., Lamendella, R., Joyner, D.C., Spier, C., Baelum, J., Auer, M., Zemla, M.L., Chakraborty, R., Sonnenthal, E.L., D'haeseleer, P., Holman, H.-Y.N., Osman, S., Lu, Z., Van Nostrand, J.D., Deng, Y., Zhou, J. & Mason, O.U. (2010). Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science*. 330 (6001).

Heitkamp, M.A. & Cerniglia, C.E. (1988). Mineralization of polycyclic aromatic hydrocarbons by a bacterium isolated from sediment below an oil field. *Applied and environmental microbiology*. 54 (6). p.pp. 1612–4.

Van Herwijnen, R., Wattiau, P., Bastiaens, L., Daal, L., Jonker, L., Springael, D., Govers, H.A.J. & Parsons, J.R. (2003). Elucidation of the metabolic pathway of fluorene and cometabolic pathways of phenanthrene, fluoranthene, anthracene and dibenzothiophene by Sphingomonas sp. LB126. *Research in Microbiology*. 154 (3). p.pp. 199–206.

Hong, Q., Zhang, Z., Hong, Y. & Li, S. (2007). A microcosm study on bioremediation of fenitrothion-contaminated soil using Burkholderia sp. FDS-1. *International Biodeterioration & Biodegradation*. 59 (1). p.pp. 55–61.

Hou, J., Liu, W., Wang, B., Wang, Q., Luo, Y. & Franks, A.E. (2015). PGPR enhanced phytoremediation of petroleum contaminated soil and rhizosphere microbial community response. *Chemosphere*. 138. p.pp. 592–598.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M. & Brown, C.T. (2014). Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences*. 111 (13). p.p. 4904 LP-4909.

Huang, Y., Zeng, Y., Yu, Z., Zhang, J., Feng, H. & Lin, X. (2013). In silico and experimental methods revealed highly diverse bacteria with quorum sensing and aromatics biodegradation systems - A potential broad application on bioremediation. *Bioresource Technology*. 148. p.pp. 311–316.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Tennessen, K., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.-M.A., Pati, A., Nielsen, T., Markowitz, V.M. & Kyrpides, N.C. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Standards in Genomic Sciences*. 11 (1). p.p. 17.

Husain, S. (2008). Literature overview: Microbial metabolism of high molecular weight polycyclic aromatic hydrocarbons. *Remediation Journal*. 18 (2). p.pp. 131–161.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Sanderlin, J.S., Reeves, J.H., Jenkins, M.B., Endale, D.M., Coleman, D.C. & Whitman, W.B. (2008). Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biology and Biochemistry*. 40 (11). p.pp. 2843–2853.

Jung, J., Philippot, L. & Park, W. (2016). Metagenomic and functional analyses of the consequences of reduction of bacterial diversity on soil functions and bioremediation in diesel-contaminated microcosms. *Scientific Reports*. 6. p.p. 23012.

Jurelevicius, D., Alvarez, V.M., Marques, J.M., de Sousa Lima, L.R.F., Dias, F. de A. & Seldin, L. (2013). Bacterial community response to petroleum hydrocarbon amendments in freshwater, marine, and hypersaline water-containing microcosms. *Applied and environmental microbiology*. 79 (19). p.pp. 5927–35.

Kamath, R., Schnoor, J.L. & Alvarez, P.J.J. (2004). Effect of Root-Derived Substrates on the Expression

of nah-lux Genes in Pseudomonas fluorescens HK44: Implications for PAH Biodegradation in the Rhiziosphere. *Environmental Science and Technology*. 38 (6). p.pp. 1740–1745.

Kanaly, R.A. & Harayama, S. (2000). Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *Journal of bacteriology*. 182 (8). p.pp. 2059–67.

Kanaly, R.A., Harayama, S. & Watanabe, K. (2002). Rhodanobacter sp. strain BPC1 in a benzo[a]pyrene-mineralizing bacterial consortium. *Applied and environmental microbiology*. 68 (12). p.pp. 5826–33.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*. 44 (D1). p.pp. D457–D462.

Kang, H., Hwang, S.Y., Kim, Y.M., Kim, E., Kim, Y.S., Kim, S.K., Kim, S.W., Cerniglia, C.E., Shuttleworth, K.L. & Zylstra, G.J. (2003). Degradation of phenanthrene and naphthalene by a Burkholderia species strain. *Canadian journal of microbiology*. 49 (2). p.pp. 139–44.

Kim, T.J., Lee, E.Y., Kim, Y.J., Cho, K. & Ryu, H.W. (2003). Degradation of polycyclic aromatic hydrocarbons by Burkholderia cepacia 2A-12. *World J Microbiol Biotechnol*. 19 (Vogel 1996). p.pp. 411–417.

Kivisaar, M. (2009). Degradation of nitroaromatic compounds: a model to study evolution of metabolic pathways. *Molecular Microbiology*. 74 (4). p.pp. 777–781.

Ladino-Orjuela, G., Gomes, E., da Silva, R., Salt, C. & Parsons, J.R. (2016). Metabolic Pathways for Degradation of Aromatic Hydrocarbons by Bacteria. In: *Reviews of environmental contamination and toxicology*. pp. 105–121.

Lauber, C.L., Hamady, M., Knight, R. & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and environmental microbiology*. 75 (15). p.pp. 5111–20.

Leahy, J.G., Batchelor, P.J. & Morcomb, S.M. (2003). Evolution of the soluble diiron monooxygenases. *FEMS microbiology reviews*. 27 (4). p.pp. 449–79.

Lee, E.Y., Ye, B.D. & Park, S. (2003). Development and operation of a trickling biofilter system for continuous treatment of gas-phase trichloroethylene. *Biotechnology Letters*. 25 (20). p.pp. 1757–1761.

Li, A., Qu, Y., Zhou, J. & Ma, F. (2009). Characterization of a newly isolated biphenyl-degrading bacterium, Dyella ginsengisoli LA-4. *Applied biochemistry and biotechnology*. 159 (3). p.pp. 687–95.

Li, H.-Y., Wei, D.-Q., Shen, M. & Zhou, Z.-P. (2012). Endophytes and their role in phytoremediation. *Fungal Diversity*. 54 (1). p.pp. 11–18.

Li, X., Rui, J., Mao, Y., Yannarell, A. & Mackie, R. (2014a). Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. *Soil Biology and Biochemistry*. 68. p.pp. 392–401.

Li, X., Rui, J., Xiong, J., Li, J., He, Z., Zhou, J., Yannarell, A.C. & Mackie, R.I. (2014b). Functional Potential of Soil Microbial Communities in the Maize Rhizosphere S. Wang (ed.). *PLoS ONE*. 9 (11). p.p. e112609.

Lu, Z., Deng, Y., Van Nostrand, J.D., He, Z., Voordeckers, J., Zhou, A., Lee, Y.-J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D'haeseleer, P., Hazen, T.C. & Zhou, J. (2012). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The ISME journal*. 6 (2). p.pp. 451–60.

Madhaiyan, M., Poonguzhali, S., Senthilkumar, M., Pragatheswari, D., Lee, J.-S. & Lee, K.-C. (2015). Arachidicoccus rhizosphaerae gen. nov., sp. nov., a plant-growth-promoting bacterium in the family Chitinophagaceae isolated from rhizosphere soil. *INTERNATIONAL JOURNAL OF*

SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY. 65 (Pt 2). p.pp. 578–586.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y. & Bryant, S.H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. 45 (D1). p.pp. D200–D203.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N.N. & Kyrpides, N.C. (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic acids research*. 42 (Database issue). p.pp. D560-7.

Martin, B.C., George, S.J., Price, C.A., Ryan, M.H. & Tibbett, M. (2014). The role of root exuded low molecular weight organic anions in facilitating petroleum hydrocarbon degradation: Current knowledge and future directions. *Science of the Total Environment*. 472 p.pp. 642–653.

Mendes, R., Garbeva, P. & Raaijmakers, J.M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*. 37 (5). p.pp. 634–663.

Méndez, V., Agulló, L., González, M., Seeger, M., Oliver, A. & Seeger, M. (2011). The Homogentisate and Homoprotocatechuate Central Pathways Are Involved in 3- and 4-Hydroxyphenylacetate Degradation by Burkholderia xenovorans LB400 A. Driessen (ed.). *PLoS ONE*. 6 (3). p.p. e17583.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R., Venter, J., Remington, K., Heidelberg, J., Halpern, A., Rusch, D., Eisen, J., Wu, D., Paulsen, I., Nelson, K., Nelson, W., Tyson, G., Chapman, J., Hugenholtz, P., Allen, E., Ram, R., Richardson, P., Solovyev, V., Rubin, E., Rokhsar, D., Banfield, J., Huse, S., Huber, J., Morrison, H., Sogin, M., Welch, D., Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Diaz, N., Chuang, H.-Y., Cohoon, M., Crécy-Lagard, V. de, Disz, T., Edwards, R., McNeil, L., Reich, C., Aziz, R., Bartels, D., Cohoon, M., Disz, T., Edwards, R., Gerdes, S., Hwang, K., Kubal, M., Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., Berka, J., Braverman, M., Chen, Y., Chen, Z., Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Field, D., Morrison, N., Selengut, J., Sterk, P., Altschul, Sf., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D., Jarvie, T., DeSantis, T., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E., Keller, K., Huber, T., Dalevil, D., Hu, P., Andersen, G., Cole, J., Chai, B., Farris, R., Wang, Q., Wuyts, J., Peer, Yv. de, Winkelmans, T., Wachter, R. De, Leplae, R., Hebrant, A., Wodak, S., Toussaint, A., Meyer, F., Overbeek, R., Rodriquez, A., Tringe, S., Mering, C. von, Kobayashi, A., Salamov, A., Chen, K., Chang, H., Podar, M., Short, J., Mathur, E., Detter, J., Rodriguez-Brito, B., Rohwer, F., Edwards, R., McHardy, A., Martin, H., Tsirigos, A., Hugenholtz, P., Rigoutsos, I., Edwards, R., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D., Saar, M., Alexander, S., Alexander, E., Rohwer, F., Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R., Felts, B., Rayhawk, S., Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., Rohwer, F., Mou, X., Edwards, R., Hodson, R., Moran, M., Dinsdale, E., Edwards, R., Hall, D., Angly, F., Breitbart, M., Brulc, J., Furlan, M., Desnues, C., Haynes, M., Li, L., Krause, L., Diaz, N., Bartels, D., Edwards, R., Puhler, A., Rohwer, F., Meyer, F., Stoye, J., Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S., Quackenbush, J. & Rohwer, F. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9 (1). p.pp. 386.

Miller, T.R., Delcher, A.L., Salzberg, S.L., Saunders, E., Detter, J.C. & Halden, R.U. (2010). Genome

sequence of the dioxin-mineralizing bacterium Sphingomonas wittichii RW1. *Journal of bacteriology*. 192 (22). p.pp. 6101–2.

Mittal, A. & Singh, P. (2009). Isolation of hydrocarbon degrading bacteria from soils contaminated with crude oil spills. *Indian journal of experimental biology*. 47 (9). p.pp. 760–5.

Miya, R.K. & Firestone, M.K. (2001). Enhanced Phenanthrene Biodegradation in Soil by Slender Oat Root Exudates and Root Debris. *Journal of Environment Quality*. 30 (6). p.p. 1911.

Miya, R.K. & Firestone, M.K. (2000). Phenanthrene-Degrader Community Dynamics in Rhizosphere Soil from a Common Annual Grass. *Journal of Environment Quality*. 29 (2). p.p. 584.

Mohammed, D., Ramsubhag, A. & Beckles, D.M. (2007). An assessment of the biodegradation of petroleum hydrocarbons in contaminated soil using non-indigenous, commercial microbes. *Water, Air, and Soil Pollution*. 182 (1–4). p.pp. 349–356.

Muangchinda, C., Pansri, R., Wongwongsee, W. & Pinyakong, O. (2013). Assessment of polycyclic aromatic hydrocarbon biodegradation potential in mangrove sediment from Don Hoi Lot, Samut Songkram Province, Thailand. *Journal of Applied Microbiology*. 114 (5). p.pp. 1311–1324.

Nannipieri, P., Ascher-Jenull, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G. & Valori, F. (2008). *Effects of Root Exudates in Microbial Diversity and Activity in Rhizosphere Soils*.

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G. & Renella, G. (2017). Microbial diversity and soil functions. *European Journal of Soil Science*. 68 (1). p.pp. 12–26.

Nelson, M.J., Montgomery, S.O., Mahaffey, W.R. & Pritchard, P.H. (1987). Biodegradation of trichloroethylene and involvement of an aromatic biodegradative pathway. *Applied and environmental microbiology*. 53 (5). p.pp. 949–54.

Nešvera, J., Rucká, L. & Pátek, M. (2015). Catabolism of Phenol and Its Derivatives in Bacteria: Genes, Their Regulation, and Use in the Biodegradation of Toxic Pollutants. In: *Advances in Applied Microbiology*. Academic Press, pp. 107–160.

Noinaj, N., Guillier, M., Barnard, T.J. & Buchanan, S.K. (2010). TonB-dependent transporters: regulation, structure, and function. *Annual review of microbiology*. 64. p.pp. 43–60.

Nzila, A. (2013). Update on the cometabolism of organic pollutants by bacteria. *Environmental Pollution*. 178. p.pp. 474–482.

O'Sullivan, L.A. & Mahenthiralingam, E. (2005). Biotechnological potential within the genus Burkholderia. *Letters in Applied Microbiology*. 41 (1). p.pp. 8–11.

O'Sullivan, L.A., Weightman, A.J., Jones, T.H., Marchbank, A.M., Tiedje, J.M. & Mahenthiralingam, E. (2007). Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium Burkholderia vietnamiensis. *Environmental Microbiology*. 9 (4). p.pp. 1017–1034.

Oberholster, T., Vikram, S., Cowan, D. & Valverde, A. (2018). Key microbial taxa in the rhizosphere of sorghum and sunflower grown in crop rotation. *Science of the Total Environment*. 624. p.pp. 530–539.

Oh, S., Suzuki, Y., Hayashi, S., Suzuki, Y., Koike, S. & Kobayashi, Y. (2017). Potency of cashew nut shell liquid in rumen modulation under different dietary conditions and indication of its surfactant action against rumen bacteria. *Journal of animal science and technology*. 59. p.p. 27.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F. & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*. 42 (Database issue). p.pp. D206-14.

Pagé, A.P., Yergeau, É. & Greer, C.W. (2015). Salix purpurea stimulates the expression of specific bacterial xenobiotic degradation genes in a soil contaminated with hydrocarbons A. C. Singer

(ed.). *PLoS ONE*. 10 (7). p.p. e0132062.

Palaniyandi, S.A., Yang, S.H., Zhang, L. & Suh, J.W. (2013). Effects of actinobacteria on plant disease suppression and growth promotion. *Applied Microbiology and Biotechnology*. 97 (22) p.pp. 9621–9636.

Pérez-Pantoja, D., Donoso, R., Junca, H., González, B. & Pieper, D.H. (2010). Phylogenomics of Aerobic Bacterial Degradation of Aromatics. In: *Handbook of Hydrocarbon and Lipid Microbiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1355–1397.

Philippot, L., Raaijmakers, J.M., Lemanceau, P. & van der Putten, W.H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*. 11 (11). p.pp. 789–799.

Pizarro-Tobías, P., Niqui, J.L., Roca, A., Solano, J., Fernández, M., Bastida, F., García, C. & Ramos, J.L. (2015). Field trial on removal of petroleum-hydrocarbon pollutants using a microbial consortium for bioremediation and rhizoremediation. *Environmental Microbiology Reports*. 7 (1). p.pp. 85–94.

Polasa, K. & Rukmini, C. (1987). Mutagenicity tests of cashewnut shell liquid, rice-bran oil and other vegetable oils using the Salmonella typhimurium/microsome system. *Food and Chemical Toxicology*. 25 (10). p.pp. 763–766.

Ramette, A., LiPuma, J.J. & Tiedje, J.M. (2005). Species abundance and diversity of Burkholderia cepacia complex in the environment. *Applied and environmental microbiology*. 71 (3). p.pp. 1193–201.

Reilley, K.A., Banks, M.K. & Schwab, A.P. (1996). Dissipation of polycyclic aromatic hydrocarbons in the rhizosphere. *Journal of Environment Quality*. 25 (2). p.p. 212.

Rentz, J.A., Alvarez, P.J.J. & Schnoor, J.L. (2005). Benzo[a]pyrene co-metabolism in the presence of plant root extracts and exudates: Implications for phytoremediation. *Environmental Pollution*. 136 (3). p.pp. 477–484.

Revathy, T., Jayasri, M.A. & Suthindhiran, K. (2015). Biodegradation of PAHs by Burkholderia sp. VITRSB1 Isolated from Marine Sediments. *Scientifica*. 2015. p.p. 867586.

Ruff, J., Denger, K. & Cook, A.M. (2003). Sulphoacetaldehyde acetyltransferase yields acetyl phosphate: purification from Alcaligenes defragrans and gene clusters in taurine degradation. *The Biochemical journal*. 369 (Pt 2). p.pp. 275–85.

Sabna Prabha, S., Sisu Pramod, S. & Potty, V. (2011). Isolation, Identification and Methanogenesis of CNSL Degrading Bacteria and Immobilized Bioremediation Techniques of CNSL and its Contaminated Water from Cashew Industry. *Middle-East Journal of Scientific Research*. 10 (3). p.pp. 327–331.

Samanta, S.K., Singh, O. V. & Jain, R.K. (2002). Polycyclic aromatic hydrocarbons: Environmental pollution and bioremediation. *Trends in Biotechnology*. 20 (6) p.pp. 243–248.

Sawana, A., Adeolu, M. & Gupta, R.S. (2014). Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov. harboring environmental species. *Frontiers in genetics*. 5. p.p. 429.

Schloss, P.D. & Handelsman, J. (2006). Toward a census of bacteria in soil. *PLoS computational biology*. 2 (7). p.p. e92.

Schutter, M.E., Sandeno, J.M. & Dick, R.P. (2001). Seasonal, soil type, and alternative management influences on microbial communities of vegetable cropping systems. *Biology and Fertility of Soils*. 34 (6). p.pp. 397–410.

Seeger, M., González, M., Cámara, B., Muñoz, L., Ponce, E., Mejías, L., Mascayano, C., Vásquez, Y. &

Sepúlveda-Boza, S. (2003). Biotransformation of natural and synthetic isoflavonoids by two recombinant microbial enzymes. *Applied and environmental microbiology*. 69 (9). p.pp. 5045–50.

Seo, J.-S., Keum, Y.-S., Hu, Y., Lee, S.-E. & Li, Q.X. (2006). Degradation of phenanthrene by Burkholderia sp. C3: initial 1,2- and 3,4-dioxygenation and meta- and ortho-cleavage of naphthalene-1,2-diol. *Biodegradation*. 18 (1). p.pp. 123–131.

Shuttleworth, K.L. & Cerniglia, E. (1995). Environmental aspects of PAH biodegradation. *Applied Biochemistry and Biotechnology*. 54 (1–3). p.pp. 291–302.

Siciliano, S.D., Fortin, N., Mihoc, A., Wisse, G., Labelle, S., Beaumier, D., Ouellette, D., Roy, R., Whyte, L.G., Banks, M.K., Schwab, P., Lee, K. & Greer, C.W. (2001). Selection of Specific Endophytic Bacterial Genotypes by Plants in Response to Soil Contamination. *Applied and Environmental Microbiology*. 67 (6). p.pp. 2469–2475.

Sindhu, M.R. & Potty, V.P. (2015). Biodegradation of Cashew Nut Shell Liquid by Delftia Acidovorans and Pseudomonas Aeruginosa Isolated from Marine Environment. *International Journal of Agriculture, Environment and Biotechnology*. 8 (4). p.p. 837.

Sindhu, M.R. & Potty, V.P. (2016). Isolation of Novel Fungal Strains from Marine Environment capable of Cashew Nut Shell Liquid Degradation. *International Journal of Biotechnology and Biochemistry ISSN*. 12 (1). p.pp. 55–61.

Smith, D.J., Park, J., Tiedje, J.M. & Mohn, W.W. (2007). A Large Gene Cluster in Burkholderia xenovorans Encoding Abietane Diterpenoid Catabolism. *Journal of Bacteriology*. 189 (17). p.pp. 6195–6204.

Suárez-Moreno, Z.R., Caballero-Mellado, J., Coutinho, B.G., Mendonça-Previato, L., James, E.K. & Venturi, V. (2012). Common Features of Environmental and Potentially Beneficial Plant-Associated Burkholderia. *Microbial Ecology*. 63 (2) p.pp. 249–266.

Sugimoto, K., Senda, T., Aoshima, H., Masai, E., Fukuda, M. & Mitsui, Y. (1999). Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. *Structure*. 7 (8). p.pp. 953–965.

Susilaningsih, D., Okazaki, F., Yopi, Y., Widyastuti, Y. & Harayama, S. (2013). Isolation and Screening of Surfactant-producing Bacteria from Indonesian Marine Environments and Its Application on Bioremediation. *ANNALES BOGORIENSES*. 17 (2). p.pp. 43–53.

Tang, K., Jiao, N., Liu, K., Zhang, Y. & Li, S. (2012). Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. *PLoS one*. 7 (7). p.p. e41204.

Tatusov, R.L., Koonin, E. V & Lipman, D.J. (1997). A genomic perspective on protein families. *Science (New York, N.Y.)*. 278 (5338). p.pp. 631–7.

Teklehaimanot, Z., Jones, M. & Sinclair, F.L. (2002). Tree and livestock productivity in relation to tree planting configuration in a silvopastoral system in North Wales, UK. *Agroforestry Systems*. 56 (1). p.pp. 47–55.

Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. (2011). Environmental and Gut Bacteroidetes: The Food Connection. *Frontiers in Microbiology*. 2. p.p. 93.

Uhlik, O., Wald, J., Strejcek, M., Musilova, L., Ridl, J., Hroudova, M., Vlcek, C., Cardenas, E., Mackova, M. & Macek, T. (2012). Identification of Bacteria Utilizing Biphenyl, Benzoate, and Naphthalene in Long-Term Contaminated Soil J. A. Gilbert (ed.). *PLoS ONE*. 7 (7). p.p. e40653.

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C. & Schuster, S.C. (2008). Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome

N. Ward (ed.). *PLoS ONE*. 3 (6). p.p. e2527.

Uroz, S., Ioannidis, P., Lengelle, J., Cébron, A., Morin, E., Buée, M. & Martin, F. (2013). Functional Assays and Metagenomic Analyses Reveals Differences between the Microbial Communities Inhabiting the Soil Horizons of a Norway Spruce Plantation J. A. Gilbert (ed.). *PLoS ONE*. 8 (2). p.p. e55929.

Wackett, L.P. (2004). Evolution of enzymes for the metabolism of new chemical inputs into the environment. *The Journal of biological chemistry*. 279 (40). p.pp. 41259–62.

Wang, Y., Li, H., Zhao, W., He, X., Chen, J., Geng, X. & Xiao, M. (2010). Induction of toluene degradation and growth promotion in corn and wheat by horizontal gene transfer within endophytic bacteria. *Soil Biology and Biochemistry*. 42 (7). p.pp. 1051–1057.

Watanabe, Y., Suzuki, R., Koike, S., Nagashima, K., Mochizuki, M., Forster, R.J. & Kobayashi, Y. (2010). In vitro evaluation of cashew nut shell liquid as a methane-inhibiting and propionate-enhancing agent for ruminants. *Journal of Dairy Science*. 93 (11). p.pp. 5258–5267.

Wenzel, W.W. (2009). Rhizosphere processes and management in plant-assisted bioremediation (phytoremediation) of soils. *Plant and Soil*. 321 (1–2). p.pp. 385–408.

Xue, W. & Warshawsky, D. (2005). Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review. *Toxicology and Applied Pharmacology*. 206 (1). p.pp. 73–93.

Yakimov, M.M., Timmis, K.N. & Golyshin, P.N. (2007). Obligate oil-degrading marine bacteria. *Current Opinion in Biotechnology*. 18 (3) p.pp. 257–266.

Yan, S., Wang, Q., Qu, L. & Li, C. (2013). Characterization of Oil-Degrading Bacteria from Oil-Contaminated Soil and Activity of their Enzymes. *Biotechnology & Biotechnological Equipment*. 27 (4). p.pp. 3932–3938.

Yang, S., Wen, X., Shi, Y., Liebner, S., Jin, H. & Perfumo, A. (2016). Hydrocarbon degraders establish at the costs of microbial richness, abundance and keystone taxa after crude oil contamination in permafrost environments. *Scientific reports*. 6. p.p. 37473.

Ye, D., Siddiqi, M.A., Maccubbin, A.E., Kumar, S. & Sikka, H.C. (1996). Degradation of Polynuclear Aromatic Hydrocarbons by Sphingomonas paucimobilis. *Environmental Science & Technology*. 30 (1). p.pp. 136–142.

Yousaf, S., Andria, V., Reichenauer, T.G., Smalla, K. & Sessitsch, A. (2010). Phylogenetic and functional diversity of alkane degrading bacteria associated with Italian ryegrass (Lolium multiflorum) and Birdsfoot trefoil (Lotus corniculatus) in a petroleum oil-contaminated environment. *Journal of Hazardous Materials*. 184 (1–3). p.pp. 523–532.

Zhang, D.C., Mörtelmaier, C. & Margesin, R. (2012). Characterization of the bacterial archaeal diversity in hydrocarbon-contaminated soil. *Science of the Total Environment*. 421–422. p.pp. 184–196.

Zhang, K., Sun, Y., Cui, Z., Yu, D., Zheng, L., Liu, P. & Lv, Z. (2017). Periodically spilled-oil input as a trigger to stimulate the development of hydrocarbon-degrading consortia in a beach ecosystem. *Scientific Reports*. 7 (1). p.p. 12446.

Zhang, L., Wang, X., Jiao, Y., Chen, X., Zhou, L., Guo, K., Ge, F. & Wu, J. (2013). Biodegradation of 4-chloronitrobenzene by biochemical cooperation between Sphingomonas sp. strain CNB3 and Burkholderia sp. strain CAN6 isolated from activated sludge. *Chemosphere*. 91 (9). p.pp. 1243–1249.

Zhonghui, Z., Qing, H., Guoshun, Z., Jianhong, X. & Shunpeng, L. (2005). Studies on isolation identification of fenitrothion degradation bacteria FDS--1 and its degradation characteristics. *Zhongguo Huanjing Kexue*. 25 (1) p.pp. 52–56.

3.7. Appendix III

Table S3.1 List of catabolic functions retrieved after functional annotation of protein-coding genes via analysis using AromaDeg and in-house validate database. Putative functions, substrate utilised and abbreviation are listed in the table. Normalized values performed on the total number of protein-coding genes as annotated by IMG-M system are indicated.

Enzyme	Substrate	Abbreviation	0W	1W	3W	5W
2,4-D and 2,4,5-T oxygenases	2,4,5-trichlorophenoxyacetate	245T	0.000	0.000	0.065	0.026
2,4-D and 2,4,5-T oxygenases	2,4-Dichlorophenoxyacetate	24D	0.000	0.000	0.065	0.026
Benzoate and 2-Chlorobenzoate dioxygenases	2-Chlorobenzoate	2CB	0.000	0.219	0.151	0.295
4-Aminobenzenesulfonate 3,4-dioxygenase and related enzymes	4-Aminobenzenesulfonate	Abs	0.000	0.274	0.216	0.116
Aniline dioxygenases	Aniline	Ani	0.000	0.000	0.000	0.013
Anthranilate dioxygenases of Burkholderia and some other organisms	Anthranilate	Ant	0.000	0.137	0.065	0.064
2-Aminophenol 1,6-dioxygenase beta subunit	2-Aminophenol	Apb	0.000	0.192	0.151	0.116
Naphthalene dioxygenases of Proteobacteria and related enzymes	Biphenyl or polycyclic aromatic hydrocarbons	Bph	0.000	0.000	0.022	0.026
Benzoate and 2-Chlorobenzoate dioxygenases	Benzoate	Bzt	0.000	0.219	0.151	0.295
Carbazol and diphenylamin dioxygenases	Carbazol	Car	0.000	0.027	0.000	0.000
Enzymes related to 2,3-dihydroxybiphenyl dioxygenases of Bacillus sp. JF8 - Proteobacterial catechol 2,3-dioxygenases (EXDO 1.2 A-C as defined by Eltis and Bolin, 1996, EXDO 1.2D as defined by Eltis and Bolin, 1996)	Catechol	Cat	0.000	0.192	0.043	0.064
Extradiol dioxygenases involved in metabolism of 2'-substituted 2,3-dihydroxibiphenyls	2'-Carboxy-2,3-dihydroxybiphenyl	Cdb	0.000	0.000	0.000	0.013
2,3-Dihydroxy-p-cumate dioxygenases and 2,3-dihydroxy benzoate dioxygenases	2,3-Dihydroxy-p-cumate	Cmt	0.000	0.000	0.043	0.103
p-Cumate dioxygenases	p-Cumate	Cum	0.000	0.000	0.108	0.051
Naphthalene dioxygenases of Proteobacteria and related enzymes	Dibenzothiophene	DbtA	0.000	0.000	0.022	0.026

Dihydroxynaphthalene dioxygenases and related enzymes in Proteobacteria	dihydroxydibenzothiophene	DbtC	0.000	0.000	0.022	0.026
2,3-Dihydroxy-p-cumate dioxygenases and 2,3-dihydroxy benzoate dioxygenases	2,3-dihydroxybenzoate	Dbz	0.000	0.000	0.043	0.090
Enzymes related to 5,5'-dehydrodivanillate demethylase	5,5'-Dehydrodivanillate	DdV	0.000	0.000	0.000	0.013
Enzymes involved in steroid metabolism by Proteobacteria and related enzymes - Enzymes most probably involved in cholesterol / steroid metabolism. Some enzymes with dihydroxybiphenyl dioxygenase activity	3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dionepropionate (3,4-DHSA	Dha	0.000	0.027	0.323	0.154
Enzymes most probably involved in cholesterol / steroid metabolism, some enzymes with dihydroxybiphenyl dioxygenase activity - 2,3-dihydroxybiphenyl dioxygenases of Firmicutes and Actinobacteria - Extradiol dioxygenases of unknown function, some enzymes with 2,3-dihydroxybiphenyl cleavage activity - 2,3-Dihydroxybiphenyl dioxygenases and enzymes of methylhydroquinone metabolism - Extradiol dioxygenases related to BphC5 of Rhodococcus erythropolis TA421 - Extradiol dioxygenases related to Swit 1826 DHB dioxygenases of Sphingomonas wittichii RW1 - Enzymes involved in steroid metabolism by Proteobacteria and related enzymes- Protobacterial 2,3-dihydroxybiphenyl dioxygenases	2,3-Dihydroxybiphenyl	Dhb	0.000	0.658	0.647	0.359
Dihydroxynaphthalene dioxygenases and related enzymes in Proteobacteria	2,3-dihydroxy-1-ethylbenzene	Dhe	0.000	0.000	0.022	0.026
Dihydroxynaphthalene dioxygenases and related enzymes in Proteobacteria	Dihydroxynaphthalene	Dhn	0.000	0.000	0.022	0.026
Dihydroxynaphthalene dioxygenases and related enzymes in Proteobacteria	Dihydroxyphenanthrene	Dhp	0.000	0.000	0.022	0.026
Extradiol dioxygenases of diterpenoid degradation	7-Oxo-11,12-dihydroxydehydroabietate	Dit	0.000	0.110	0.000	0.000
Diterpenoid dioxygenases and related enzymes	Abietane diterpenoids	DitA	0.000	0.165	0.000	0.000
Naphthalene dioxygenases of Proteobacteria and related enzymes	Dinitrotoluene	Dnt	0.000	0.000	0.022	0.026
Carbazol and diphenylamin dioxygenases	Diphenylamine	Dpa	0.000	0.027	0.000	0.000
2,3-Dihydroxyphenylpropionate dioxygenases	2,3-Dihydroxyphenylpropionate	Dpp	0.000	0.192	0.151	0.026
Gallate dioxygenases	Gallate	Gal	0.000	0.247	0.280	0.295
Gentisate dioxygenases of Actinobacteria and Proteobacteria	Gentisate	Gen	0.000	0.219	0.517	0.321

1-Hydroxy-2-naphthoate dioxygenases of	1-Hydroxy-2-naphthoate	Hna	0.000	0.000	0.000	0.013
Homoprotocatechuate 3,4-dioxygenases (except of β- and γ-Proteobacteria) - Homoprotocatechuate 2,3-dioxygenases	Homoprotocatechuate	Hpc	0.000	0.247	0.388	0.372
Ibuprofen-CoA dioxygenases and related enzymes	Ibuprofen-CoA	Ibu	0.000	0.000	0.000	0.013
Isophthalate dioxygenases and related enzymes	Isophthalate	Ipt	0.000	0.055	0.194	0.090
2,3-Dihydroxybiphenyl dioxygenases and enzymes of methylhydroquinone metabolism	Methylhydroquinone	Mhq	0.000	0.247	0.151	0.013
Naphthalene dioxygenases of Proteobacteria and related enzymes	Naphthalene	Nah	0.000	0.000	0.022	0.026
1-Oxo-2,3-dihydroquinoline monooxygenases and related enzymes	1-Oxo-2,3-Dihydroquinoline	Odm	0.000	0.055	0.022	0.000
3-O-methylgallate 3,4-dioxygenase DesZ	3-O-methylgallate	Omg	0.000	0.165	0.108	0.051
Protocatechuate dioxygenases	Protocatechuate	Pca	0.000	0.027	0.022	0.039
Naphthalene dioxygenases of Proteobacteria and related enzymes	Phenanthrene	Phn	0.000	0.055	0.022	0.026
Phthalate 4,5-dioxygenases	Phthalate	Pht	0.000	0.000	0.022	0.116
Enzymes related to phenoxybenzoate, function often unknown unknown	Phenoxybenzoate	Pob	0.000	0.027	0.022	0.039
Salicylate 1-hydroxylases	Salicylate	Sa1	0.000	0.000	0.043	0.051
Salicylate 5-hydroxylases	Salicylate	Sa5	0.000	0.274	0.022	0.077
Gentisate dioxygenases of Actinobacteria and Proteobacteria	Salicylate	Sal	0.000	0.000	0.065	0.000
Enzymes typically involved in 2,3,2'-trihydroxybiphenyl degradation (e.g. formed from dibenzofuran)	2,3,2'-Trihydroxybiphenyl	Thb	0.000	0.000	0.022	0.013
Dihydroxynaphthalene dioxygenases and related enzymes in Proteobacteria	1,2-dihydroxy-5,6,7,8-tetrahydronaphthalene	Thn	0.000	0.000	0.022	0.026
Terephthalate dioxygenases	Terephthalate	Tph	0.000	0.055	0.043	0.026
Carboxymuconate cycloisomerase	3-Carboxy-cis,cis-muconate	McyI	0.000	0.302	0.237	0.321
4-hydroxyphenylpyruvate dioxygenase - 4-Hydroxyphenylpyruvate dioxygenase	4-Hydroxyphenylpyruvate	4HPPD	0.000	0.219	0.065	0.090
Alkane hydroxylase	Alkane	AlkB CYP153	0.259	0.165	0.884	0.372
Catechol 1,2-dioxygenase	Catechol	C1,2-O	0.000	0.494	0.560	0.436
4-Hydroxyphenylacetate dioxygenase	Homogentisate	4HPA3MO	0.000	0.000	0.000	0.013

Homogentisate 1,2-dioxygenase	Homogentisate	H1,2-O	0.000	0.082	0.086	0.064
Maleylacetate reductase	Maleylacetate	MAR	0.000	0.247	0.259	0.141
Phenol hydroxylase	Phenol	PH	0.000	0.137	0.172	0.205
Protocatechuate 3,4-dioxygenase	Protocatechuate	3,4-PCD	0.000	0.000	0.043	0.026

CHAPTER VI

General Discussion and Conclusions

The overall aim of this work, which was part of the BBSRC and Innovate UK-funded project "Biotechnology for anti-weeds" carried out in collaboration with the BioComposites Centre (Bangor University, UK), and with Almac Group (Craigavon, UK) and Hockley International (Manchester, UK) as industrial partners, was to identify and characterise novel biocatalysts of industrial interest that can sustain enzyme-based "greener" processes. Many chemical synthetic processes require harsh and hazardous conditions, such as high temperatures and pressures, acidic or alkaline reaction conditions, which result in high-energy costs as well as in a negative impact on the environment. Furthermore, the low specificity of these reactions may produce unwanted side products that may be costly to remove, harmful for the environment and can lead to a lower overall product yield. The implementation of biocatalysts in the form of isolated enzymes or whole cells proved to be an effective strategy to mitigate these issues and promoted the development of a wide-range of biotechnological applications (Lorenz & Eck, 2005). The characteristic properties of the enzymes, including the substrate and stereo-specificity, the versatility in the catalysed reactions, and the mild operating conditions, justify the exceptional growth of the enzyme market and the constant demand in new enzymes.

The first objective of the project was to develop an enzymatic route for the production of the novel bio-herbicide sorgoleone. Sorgoleone, the natural component of sorghum root exudate, has demonstrated in a number of studies its potent activity as a broad range anti-weed compound (Dayan et al., 2010). Its synthesis was hypothesized through enzymatic conversion of a widely available substrate, the phenolic lipid cardol triene (5-pentadecatrienyl resorcinol), the component of the Cashew Nut Shell Liquid (CNSL). The identification and characterization of microbial activities necessary for the fulfilment of the enzymatic process and the optimization of the reaction conditions were the key stages of this investigation. In addition, the microbial enzymatic diversity was further explored with the intention of identifying novel biocatalysts with potential applications in industry. Metagenomic approaches were applied in order to accomplish the objectives of this study. Metagenomics is a powerful method to unravel the uncultured microbial diversity and its biochemical potential; it has successfully been applied in several studies to a variety of environmental samples (e.g. soil, water, sediment, rumen, human gut) characterised by the most diverse physical, biological and chemical conditions. Thanks to metagenomics studies, the microbial composition of unexplored

niches was enlightened and novel candidate catalysts required by the market were identified (Ferrer et al., 2016; Lam et al., 2015).

The rhizobiome is strongly influenced by root activity where plant secretion is a key determinant of its structure (De-la-Pena & Loyola-Vargas, 2014). Therefore, the rhizosphere of *Sorghum bicolor* plants was chosen as starting material of our investigation. In fact, it is in the rhizosphere of sorghum plants that the exudation of the unique herbicidal metabolite is localised during the germination and development of the plant. After the germination of sorghum seedlings, an enrichment step was carried out through the addition of CNSL directly in soil and again in flasks in mineral medium. In this way, we aimed to stimulate the preferential growth of microorganisms that possess the enzymatic machinery for utilisation or conversion of aromatic lipophilic constituents of the cashew seed oil. Samples of the enriched soil slurry were collected systematically at different timepoints, and the metagenomic DNA was isolated and used for the construction of fosmid clone libraries and *de novo* sequencing by Illumina platform.

Firstly, *de novo* sequencing of the environmental DNA allowed the definition of the taxonomic configuration of the sample throughout the enrichment process. Modifications in the relative abundance of the included taxonomic groups was observed, suggesting a possible adaptation of the community to the presence of tCNSL. The culture-independent characterization revealed a clear predominance of *Proteobacteria,* as previously found in contaminated soil samples (Greer et al., 2010). *Proteobacteria* were represented by members of the orders *Burkholderiales*, *Xhantomonadales*, *Rhizobiales* and *Sphingomonadales*, similarly to those in previously studied oil-contaminated sites where these bacteria proved their ability to degrade aromatic and polyaromatic hydrocarbons (Gutierrez, 2017). Among *Actinobacteria*, genera *Mycobacterium* and *Nocardioides* were strongly enriched. The phylum *Bacteroidetes* is known to comprise microorganisms involved in the degradation of organic compounds (Thomas et al., 2011), and representatives of this phylum showed an increase in their abundance after five weeks from the beginning of the enrichment. Further evidence of the community adaptation was provided by the interpretation of the metagenomic data from a functional point of view, which highlighted an increase in protein-coding genes associated with the degradation of aromatic compounds. The pathways converging around the formation of central intermediates, catechol and protocatechuate, were particularly enriched, followed by dioxygenases catalysing the breakdown of gentisate, gallate and homoprotocatechuate. Interestingly, the majority of the identified catabolic genes were affiliated to *Burkholderiales* and *Rhizobiales*, that are known to be potentially the most versatile and efficient degraders.

Successively, the annotated protein-coding genes and the fosmid libraries generated from the same genomic material were used for the mining of the biocatalysts of interest. The first step of the

hypothesised sorgoleone synthetic pathway requires the methylation of one hydroxyl group on the aromatic moiety of cardol triene. Therefore, efforts were directed to the screening of *O*-methyltransferases (OMTs) capable of accommodating cardol triene as substrate. Initially, the methyltransferase SbOMT3 (Baerson et al., 2010) encoded by a *E. coli* codon usage adapted synthetic version of the sorghum gene was tested; in a previous study, this methyltransferase was isolated from sorghum root hairs and described as putatively involved in the biosynthetic pathway of sorgoleone (Baerson et al., 2008). No activity was observed against cardol triene under the conditions applied. Following the inconclusive result of the functional screening of the fosmid libraries, the homology screening strategy was attempted. The comparison with protein databases allowed for the selection of a set of putative SAM dependent *O*-methyltransferases. The recombinant OMTs were tested using *in vitro* enzymatic assays and one active enzyme (OMT33) towards our substrate was identified. This methyltransferase from *Mycobacterium* sp.*,* one of the microorganisms favoured by tCNSL microcosm enrichment with an approximate 18-fold increase in relative abundance compared to not enriched rhizosphere community, was annotated in the nr database as a hydroxyneurosporene methyltransferase with no biochemical information; biochemical characterization of the protein allowed to define the optimal conditions for maximisation of substrate conversion, that was estimated at approximately 75%. Furthermore, OMT33 revealed substrate promiscuity towards structurally different aromatic compounds, such us vanillin, orcinol and resorcinol, which may open the possibility for other biotechnological applications. In a further effort of minimising the cost of the enzymatic process, the necessary cofactor SAM was enzymatically produced through the recombinant bacterial MetK SAM synthetase; the simultaneous production of the cofactor and utilization in cardol methylation in the same reaction system produced promising results. Finally, the identification of OMT33 shed a new light to the properties of numerous homologous proteins ambiguously annotated in public databases and represented a decisive step towards setting up the process for the industrial production of the new bio-herbicide sorgoleone.

The second step of the sorgoleone enzymatic production involves the dihydroxylation of the aromatic moiety of the methylated intermediate (3-methyl-5-pentadecatrienyl resorcinol) to produce dihydrosorgoleone, direct precursor of sorgoleone. This reaction is likely to be catalysed *in vivo* by one or multiple cytochrome P450 monooxigenases (Pan et al., 2018). As with the functional screening performed for methyltransferase activities, the fosmid libraries were mined in the attempt of isolating active bacterial monooxygenases. In addition, a panel of plant P450 genes was synthesised with *E.coli* codon usage, comprising of sorghum P450s hypothetically involved in the sorgoleone biosynthesis and other proteins with reasonable identity. No activity was observed against the designated substrate during this screening. Due to the inconclusive result, the description of the work conducted on this

matter was not included in this thesis. Nevertheless, a sequence-based screening approach was not tested yet and the possibility of finding a suitable biocatalyst within the annotated metagenome is still valid. However, researches conducted by our partners at the BioComposites Centre identified a simple and selective method to perform the second hydroxylation step through the oxidation reaction with hydrogen peroxide. A semi-synthetic production route was therefore proposed, and sorgoleone was produced in amounts sufficient for toxicity and herbicidal activity tests that are currently being evaluated by Hockley International Ltd.

The established fosmid libraries were also screened for other common targets, including lipases/esterases, glycosyl hydrolases and proteases. Functional screening protocols for hydrolytic activities are well-established and are usually performed on agar plates incorporating specific substrates (Placido et al., 2015). A number of positive hits were visualised by the appearance of clear zone of lipolysis or by the development of particular colour. The majority of the sequences were coming from alphaproteobacterial microorganisms, such as *Sphingomonadales* and *Rhizobiales*, particularly enriched in our soil samples. Four esterases that appeared extremely active from *in vivo* lipolytic assay were biochemically characterized. The novel metagenomic esterases, which did not have any previous records on functionalities, displayed organic solvent stability and prominent substrate promiscuity, both being attractive properties for possible industrial applications. Substrate promiscuity, together with catalytic promiscuity, are interesting features observed in some enzymes and they represent a diversion from the universally accepted enzyme specificity model referred to as "lock and key". Possible mechanisms and practical implications in the development of biocatalysts for organic synthesis have been recently reviewed (Gupta, 2016; Arora et al., 2014). Promiscuous enzymes are being frequently exploited in applied enzymology, furthermore, the application of rational protein engineering on a promiscuous enzyme can be the starting point for the creation of a new enzyme activity (Schmidt et al., 2003; Aharoni et al., 2005). The discovery of these new esterases fits in the recent trend of the metagenomic investigations for novel biocatalysts focused especially on lipases and esterases, which are extensively used for many industrial applications (López-López et al., 2014). A recent survey analysed and compared the substrate promiscuity of more than one hundred metagenomic esterases (Martínez-Martínez et al., 2018), including model industrial catalysts such as Lipase B (CALB) from *Candida antarctica*, a well-studied enzyme catalysing a surprising diversity of reactions including many different regio- and enantio-selective syntheses. Unexpectedly, the esterases isolated from the sorghum rhizosphere were capable of hydrolysing a number of substrates higher of than the majority of the enzymes in that collection. Crystal structures of the esterases were resolved: these data will add a considerable knowledge on the mechanisms of enzyme promiscuity and support the improvement of tools for the prediction of esterases substrate promiscuity (Martínez-

Martínez et al., 2018), besides opening the way for the design of new efficient catalysts for synthetic reactions.

In conclusion, the outcomes of this PhD work pinpointed the enormous potential of enrichment protocols combined with metagenomics analyses for the identification of novel powerful biocatalysts and for the disclosure of the microbial diversity within complex environmental samples. The results have also highlighted several key areas for further research:

- Sequence-based screening of the sorghum rhizosphere enriched metagenome for CYP450 monooxygenases activities performing the conversion of 3-methyl-5-pentadecatrienyl resorcinol into dihydrosorgoleone. In this way, the complete enzymatic route for the production of sorgoleone would be possible.

- Preparation of mutants of the methyltransferase OMT33 using random and site-directed mutagenesis in order to enhance enzyme properties, such as activity, thermostability, and substrate specificity.

- Setting up of a cost-effective scalable process for the industrial production of sorgoleone. The process development would involve optimization of factors such as substrate concentration, enzyme formulation, cofactor recycling, use of solvents, product extraction and purification.

- Site-directed mutagenesis of the metagenomic esterases in order to further expand the substrate promiscuity and enhance temperature stability and solvent stability.

6.1. References

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C. & Tawfik, D.S. (2005). The 'evolvability' of promiscuous protein functions. *Nature Genetics*. 37 (1). p.pp. 73–76.

Arora, B., Mukherjee, J. & Gupta, M.N. (2014). Enzyme promiscuity: using the dark side of enzyme specificity in white biotechnology. *Sustainable Chemical Processes*. 2 (1). p.p. 25.

Baerson, S.R., Dayan, F.E., Rimando, A.M., Nanayakkara, N.P.D., Liu, C.-J., Schröder, J., Fishbein, M., Pan, Z., Kagan, I. a, Pratt, L.H., Cordonnier-Pratt, M.-M. & Duke, S.O. (2008). A functional genomics investigation of allelochemical biosynthesis in Sorghum bicolor root hairs. *The Journal of biological chemistry*. 283 (6). p.pp. 3231–47.

Baerson, S.R., Rimando, A.M., Dayan, F.E. & Pan, Z. (2010). *O-methyltransferase gene from sorghum cloning, expression, transformation and characterization*.

Dayan, F.E., Rimando, A.M., Pan, Z., Baerson, S.R., Gimsing, A.L. & Duke, S.O. (2010). Sorgoleone. *Phytochemistry*. 71 (10). p.pp. 1032–9.

De-la-Pena, C. & Loyola-Vargas, V.M. (2014). Biotic Interactions in the Rhizosphere: A Diverse Cooperative Enterprise for Plant Productivity. *PLANT PHYSIOLOGY*. 166 (2). p.pp. 701–719.

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W.R., Golyshina, O. V. & Golyshin, P.N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology*. 9 (1). p.pp. 22–34.

Greer, C.W., Whyte, L.G. & Niederberger, T.D. (2010). Microbial Communities in Hydrocarbon-Contaminated Temperate, Tropical, Alpine, and Polar Soils. In: *Handbook of Hydrocarbon and Lipid Microbiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 2313–2328.

Gupta, R.D. (2016). Recent advances in enzyme promiscuity. *Sustainable Chemical Processes*. 4 (1). p.p. 2.

Gutierrez, T. (2017). Aerobic Hydrocarbon-Degrading Gammaproteobacteria: Xanthomonadales. In: *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*. Cham: Springer International Publishing, pp. 1–15.

Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D. & Charles, T.C. (2015). Current and future resources for functional metagenomics. *Frontiers in microbiology*. 6. p.p. 1196.

López-López, O., Cerdán, M.E. & González Siso, M.I. (2014). New extremophilic lipases and esterases from metagenomics. *Current protein & peptide science*. 15 (5). p.pp. 445–55.

Lorenz, P. & Eck, J. (2005). Metagenomics and industrial applications. *Nature*. 3 (June). p.pp. 510–516.

Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P.J., Bargiela, R., Gertler, C., Navarro-Fernández, J., Bollinger, A., Thies, S., Méndez-García, C., Popovic, A., Brown, G., Chernikova, T.N., García-Moyano, A., Bjerga, G.E.K., Pérez-García, P., Hai, T., Del Pozo, M. V., Stokke, R., Steen, I.H., Cui, H., Xu, X., Nocek, B.P., Alcaide, M., Distaso, M., Mesa, V., Peláez, A.I., Sánchez, J., Buchholz, P.C.F., Pleiss, J., Fernández-Guerra, A., Glöckner, F.O., Golyshina, O. V., Yakimov, M.M., Savchenko, A., Jaeger, K.-E., Yakunin, A.F., Streit, W.R., Golyshin, P.N., Guallar,

V., Ferrer, M. & The INMARE Consortium (2018). Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chemical Biology*. 13 (1). p.pp. 225–234.

Pan, Z., Baerson, S.R., Wang, M., Bajsa-Hirschel, J., Rimando, A.M., Wang, X., Nanayakkara, N.P.D., Noonan, B.P., Fromm, M.E., Dayan, F.E., Khan, I.A. & Duke, S.O. (2018). A cytochrome P450 CYP71 enzyme expressed in *Sorghum bicolor* root hair cells participates in the biosynthesis of the benzoquinone allelochemical sorgoleone. *New Phytologist*. 218 (2). p.pp. 616–629.

Placido, A., Hai, T., Ferrer, M., Chernikova, T.N., Distaso, M., Armstrong, D., Yakunin, A.F., Toshchakov, S. V., Yakimov, M.M., Kublanov, I. V., Golyshina, O. V., Pesole, G., Ceci, L.R. & Golyshin, P.N. (2015). Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Applied Microbiology and Biotechnology*. 99 (23). p.pp. 10031–10046.

Schmidt, D.M.Z., Mundorff, E.C., Dojka, M., Bermudez, E., Ness, J.E., Govindarajan, S., Babbitt, P.C., Minshull, J. & Gerlt, J.A. (2003). Evolutionary Potential of (β/α) 8 -Barrels: Functional Promiscuity Produced by Single Substitutions in the Enolase Superfamily †. *Biochemistry*. 42 (28). p.pp. 8387–8393.

Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. (2011). Environmental and Gut Bacteroidetes: The Food Connection. *Frontiers in Microbiology*. 2. p.pp. 93.