TECHNISCHE UNIVERSITÄT MÜNCHEN

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Mikrobiologie

Biodegradation of aromatic hydrocarbons in oxygen-limited groundwater

Lauren Bradford

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzende(r):	Hon.-Prof. Dr. Michael Schloter
Prüfer der Dissertation:	1. Prof. Dr. Tillmann Lüders
	2. Prof. Dr. Wolfgang Liebl

Die Dissertation wurde am 27.01.2020 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 08.06.2020 angenommen.

Abstract

Hydrocarbon contamination of groundwater is a global problem, negatively impacting human health and ecosystem functioning. One of the main ways to mitigate this issue is bioremediation, the use of microorganisms to sequester or break down pollutants. Bacteria capable of metabolizing hydrocarbons are widespread in the environment, including groundwater. They rely on a variety of metabolic and respiratory strategies, distinguished primarily by the availability of oxygen. Oxygen can act as a terminal electron acceptor (TEA) during respiration, but also as a metabolic co-substrate during the breakdown of stable compounds like aromatic hydrocarbons. Anaerobic degradation relies on alternative TEAs and mechanisms of aromatic destabilization. Strictly aerobic and anaerobic hydrocarbon degradation have been well studied in recent decades, but hot spots of degradation occur at the fringes of contaminant plumes, where mixing of oxygen and other TEAs create dynamic, TEA-limited conditions. Microaerobes and other microbes with specialized strategies that could offer advantages in such conditions have been poorly investigated, despite their probable contribution to pollutant degradation in environmentally-relevant situations. In this thesis, I focus on two such strategies: oxygenesis via nitric oxide dismutation, and oxygen-activation, nitrate-respiration whereby an organism respires an alternative TEA and reserves limited oxygen for aromatic destabilization. To unambiguously connect the identities and functional expression patterns of these degraders, I improved upon the method of RNA stable isotope probing (RNA-SIP) by developing semi-quantitative total-RNA-SIP.

In the first part of this doctoral work, I developed PCR assays for the putative nitric oxide dismutase gene (*nod*) and probed for its abundance in contaminated aquifers and other aquatic systems. Dismutation of nitric oxide into N_2 and O_2 has been linked to oxygen-dependent metabolism of methane and alkanes in anoxic enrichment cultures. The *nod* gene is believed to code for the dismutase enzyme catalyzing this process. We hypothesize that it could also provide oxygen for aromatic hydrocarbon degradation, even in anoxic and low-oxygen conditions. Following the design and testing of nod-targeted degenerate primers, I found evidence of *nod* genes in contaminated aquifers in Siklós, Hungary and Düsseldorf, Germany, as well as in experimental wastewater treatment systems. This shows that oxygenesis may be more widespread and relevant to bioremediation and the global nitrogen cycle than previously thought, although the respective organisms still resist cultivation and more detailed characterisation.

I followed this with the development and a proof-of-principle experiment for total-RNA-SIP. RNA-SIP is a well-established technique used to identify microbes in a complex community actually metabolizing a substrate of interest by means of ^{13}C-labelling. Most RNA-SIP studies to date have focused solely on identification via sequencing of PCR-amplified "heavy" small subunit (SSU) rRNA. By instead combining SIP with sequencing of total heavy RNA, I was able to analyze the identity (via SSU rRNA) and function (via mRNA) of degraders from SIP gradient fractions. In combination with enrichment factor calculations, this is a powerful tool for identifying not only process-relevant microbes, but also the biochemical pathways used. I performed a proof-of-principle experiment on microcosms containing sediment from a hydrocarbon-contaminated microoxic aquifer and provided with ^{13}C-labelled toluene. Members of the Rhodocyclaceae family were overwhelmingly the most abundant and highly labelled members of the sediment community, consistent with previous work on this site. The functional transcript data showed interesting patterns that would have been missed using other methods, such as the importance of cell motility, prevalence of catechol-2,3-dioxygenase over catechol-1,2-dioxygenase for aromatic ring opening, and use of phenol hydroxylase for aromatic activation.

Moreover, I applied this new method to a series of microcosms designed to unravel the catabolic and respiratory adaptations of toluene degraders to simultaneous but limited availability of oxygen and nitrate as TEAs. I successfully resolved the involvement of distinct microaerobic vs. aerobic and anaerobic populations in pollutant degradation. In particular, differential labelling of catabolic and respiratory transcripts including phenol-hydroxylase (*dmpN*) revealed the adaptation of degraders within the Commamonadaceae under microoxic conditions with nitrate amendment. This demonstrates, for the first time, the previously hypothesized oxygen-activating, nitrate-respiring degradation strategy for a complex aquifer microbiota, explaining how degraders can increase catabolic efficiency under limited electron acceptor supply.

Overall, this thesis provides important insights into the largely unrecognized contributions to hydrocarbon degradation made by microbes employing unconventional metabolic and respiratory adaptations. This has implications for future bioremediation efforts, for example by encouraging treatment with low oxygen concentrations combined with nitrate, rather than more expensive oxygen sparging. It also represents a novel and valuable method for overcoming the notorious limitations of environmental non-target transcriptomics, by unambiguously linking substrate metabolism to the identity and gene expression of process-relevant microbes.

Zusammenfassung

Die Verunreinigung von Grundwasser mit Kohlenwasserstoffen ist ein globales Problem, das sich negativ auf die menschliche Gesundheit und auf Ökosysteme auswirkt. Ein grundlegender Ansatz zur Lösung dieses Problems ist die Bioremediation, d.h. der Abbau von Schadstoffen durch Mikroorganismen. Bakterien, die in der Lage sind, Kohlenwasserstoffe zu metabolisieren, sind in der Umwelt, einschließlich des Grundwassers, weit verbreitet. Sie sind auf eine Vielzahl von Stoffwechsel- und Atmungsaktivitäten spezialisiert, die sich vor allem durch die Verwendung von Sauerstoff unterscheiden. Sauerstoff kann als terminaler Elektronenakzeptor (TEA) bei der Atmung, aber auch als metabolisches Co-Substrat im Abbau von chemisch stabilen Schadstoffen wie z.B. aromatischen Kohlenwasserstoffen wirken. Der anaerobe Abbau basiert auf alternativen Atmungsprozessen und Aktivierungsmechanismen zur Destabilisierung der Aromaten. Der sowohl rein aerobe oder anaerobe Abbau von Kohlenwasserstoffen wurde in den letzten Jahrzehnten bereits intensiv untersucht. Am Rand von Schadstofffahnen gibt es in natürlichen Grundwasserleitern allerdings wichtige «Hot-Spots» des Abbaus, in denen die Durchmischung von Sauerstoff und anderen TEAs dynamische und teils auch limitierende Bedingungen für verschiedene Atmungen schafft. Mikroaerobier und andere Mikroben mit spezifischen Strategien, die unter solchen Bedingungen physiologische Vorteile bieten können, sind trotz ihres möglichen wichtigen Beitrags zum Schadstoffabbau in umweltrelevanten Situationen bisher nur unzureichend untersucht worden. In dieser Dissertation konzentriere ich mich auf zwei Strategien: Oxygenese durch Stickoxid-Disproportionierung und Sauerstoff-Aktivierung bei Nitrat-Atmung, in welcher ein Organismus Nitrat veratmet, während limitierender Sauerstoff nur zur Aktivierung aromatischer Substrate verwendet wird. Um die Identitäten und prozessrelevante Genexpression solcher Abbauer gezielt zu untersuchen, habe ich die Methodik des RNA «Stable Isotope Probings» (RNA-SIP) einen neuen, semi-quantitativen Ansatzes für Transkriptom-SIP wesentlich erweitert.

Im ersten Teil dieser Doktorarbeit habe ich PCR-Assays für das Gen der postulierten Stickoxid-Dismutase (nod) entwickelt, um dessen Häufigkeit in kontaminierten Aquiferen und anderen aquatischen Systemen zu untersuchen. Die Disproportionierung von Stickstoffmonoxid zu N_2 und O_2 wurde bereits mit der Oxidation von Methan und Alkanen in anoxischen Anreicherungskulturen in Verbindung gebracht. Es wird angenommen, dass das Nod-Gen für ein Dismutase-Enzym kodiert, welches diesen Prozess katalysiert. Wir vermuten, dass es auch Sauerstoff für den Abbau von aromatischen Kohlenwasserstoffen bereitstellen könnte, sowohl

unter anoxischen oder sauerstoffarmen Bedingungen. Nach der Entwicklung geeigneter PCR Primer, fand ich Hinweise auf eine unerwartete Diversität und Abundanz von Nod-Genen in kontaminierten Grundwasserleitern und in Kläranlagen. Dies zeigt, dass die Oxygenese möglicherweise weiter verbreitet und relevanter ist, als bisher angenommen, sowohl in der Bioremediation als auch im globalen Stickstoffkreislauf. Die entsprechenden Mikroorganismen entziehen sich jedoch weiterhin einer Kultivierung und genaueren Charakterisierung.

Das Projekt wurde mit der Entwicklung und einer ersten Demonstration des Transkriptom-SIP weiterentwickelt. RNA-SIP ist eine bewährte Methode, um diejenigen Mikroben in einer komplexen Gemeinschaft gezielt zu identifizieren, die ein bestimmtes Substrat abbauen. Dies wird mittels einer ^{13}C-Markierung erreicht. Die meisten RNA-SIP Studien haben sich bisher ausschließlich auf die Identifizierung von PCR-amplifizierter, ^{13}C-markierter 16S rRNA fokussiert. Durch die Kombination von RNA-SIP mit der Sequenzierung der gesamten ^{13}C-RNA (mRNA und rRNA) wurde so der Nachweis der Identität und Genexpression der Schadstoffabbauer erstmals möglich. Die Berechnung von ^{13}C-Anreicherungsfaktoren für bestimmte Transkriptgruppen war dabei ein leistungsfähiges Werkzeug, um nicht nur prozessrelevante Mikroben, sondern auch deren Physiologie zu identifizieren. Ein Proof-of-Principle-Experiment wurde mit Sedimenten aus einem kontaminierten, mikro-oxischen Aquifer und mit ^{13}C-Toluol durchgeführt. Mitglieder der *Rhodocyclaceae* waren, in Übereinstimmung mit früheren Untersuchungen des Standorts, die am stärksten ^{13}C-markierten Mikroben im Sediment. Die markierten mRNA Transkripte zeigten interessante Aktivitätsmuster, die mit anderen Methoden übersehen worden wären, wie z.B. die Rolle der Motilität, die Bedeutung der Brenzcatechin-2,3-Dioxygenase gegenüber der Brenzcatechin-1,2-Dioxygenase für die Ringöffnung, und die Verwendung der Phenol-Hydroxylase für die Aktivierung des aromatischen Substrats.

Abschließend habe ich ein Experiment mit Mikrokosmen durchgeführt, um die katabolische und respiratorische Anpassung von Toluol-Abbauern an die gleichzeitige und doch limitierende Verfügbarkeit von Sauerstoff und Nitrat als Elektronenakzeptoren zu untersuchen. Mit Transkriptom-SIP habe die Beteiligung unterschiedlicher aerober, mikroaerober und anaerober Populationen am Schadstoffabbau nachgewiesen. Insbesondere die differentielle Markierung katabolischer und respirativer Transkripte, einschließlich der Phenol-Hydroxylase (*dmpN*), zeigte die Anpassung von Abbauern innerhalb der *Commamonadaceae* an mikro-oxische Bedingungen und Nitrat. Erstmals gelang es dadurch, die zuvor nur postulierte

Strategie der Sauerstoff-Aktivierung bei Nitrat-Atmung für den Abbau von Aromaten in einer natürlichen mikrobiellen Gemeinschaft nachzuweisen. Dies zeigt auf, wie Schadstoffabbauer ihre katabolische Effizienz unter Elektronenakzeptor-Limitierung erhöhen können.

Zusammenfassend liefert diese Arbeit wichtige Einblicke in einige noch wenig verstandene und unkonventionelle Anpassungen von Stoffwechsel- und Atemwegen, die im Abbau von Schadstoffen unter limitierender Elektronenakzeptor-Verfügbarkeit relevant sind. Möglicherweise können daraus neue Strategien zur Bioremediation entwickelt werden, indem z.B. eine niedrige und deshalb kostengünstige Sauerstoffzufuhr mit einer gleichzeitigen Zugabe von Nitrat kombiniert wird. Gleichzeitig demonstriert diese Arbeit erstmals den neuen Ansatz des Transkriptom-SIPs, in dem der Umsatz eines bestimmten Substrates gezielt mit der Identität und Genexpression prozessrelevanter Mikroben verknüpft werden kann. Dadurch können einige der fundamentalen Limitierungen der Umwelt-Transkriptomik überwunden werden.

Table of Contents

List of Figures

List of Tables

Abbreviations

$^{13}C_7$ toluene	Toluene in which all carbon atoms are the heavy stable isotope
BTEX	Benzene, toluene, ethylbenzene, and xylenes (*m*-, *p*-, and *o*-)
COG	Clusters of orthologous groups of protein
CsTFA	Cesium trifluroacetate
DGGE	Denaturing gradient gel electrophoresis
DO	Dissolved oxygen
EB	Elution buffer
EF	Enrichment factor
GC-MS	Gas chromatography mass spectrometry
HMGU	Helmholtz Zentrum München
HMM	Hidden markov model
IPTG	Isopropyl β- d-1-thiogalactopyranoside
IVT	In vitro transcription
KEGG	Kyoto encyclopedia of genes and genomes
LB	Luria broth; rich growth medium
MEGAN	MEtaGenome ANalyzer
nanoSIMS	Nanoscale secondary ion mass spectrometry
NGS	Next generation sequencing
OANR	Oxygen-activating, nitrate-respiring
PAH	Polyaromatic hydrocarbon
PCI	Phenol-chloroform-isoamyl alcohol
PLFA	Phospholipid fatty acids
QIIME	Quantitative Insights Into Microbial Ecology
R2A medium	Reasoner's 2A medium
ROX	Passive reference dye used in qPCR
SCC	Sodium chloride-citrate buffer
SDS	Sodium dodecyl sulfate
SIP	Stable isotope probing
T7 polymerase	RNA polymerase from the T7 bacteriophage
Taq polymerase	Thermostable DNA polymerase from the thermophilic bacterium *T. aquaticus*
TEA	Terminal electron acceptor
TPB	Transcripts per kilobase billion
TPM	Transcripts per kilobase million
TRLFP	Terminal restriction fragment length polymorphism analysis
WWTP	Wastewater treatment plant
DNA	Dioxyribonucleic acid
cDNA	Complementary DNA
dsDNA	Double-stranded DNA
ssDNA	Single-stranded DNA
dNTP	Deoxynucleoside triphosphate
ATP	Adenosine triphosphate
CTP	Cytidine triphosphate
GTP	Guanosine triphosphate
UTP	Uridine triphosphate
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RT-PCR	Reverse transcriptase PCR
RT-qPCR	Reverse transcriptase quantitative PCR
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SSU rRNA	Small subunit ribosomal RNA
16S	SSU rRNA found in prokaryotes
18S	SSU rRNA found in eukaryotes
mRNA	Messenger RNA

1. Introduction

1.1 Groundwater contamination

Leaks, spills, and improper disposal of hydrocarbons lead to widespread contamination of environmental systems, including groundwater. The removal of contaminants can happen by physical processes (sorption, diffusion, advection, and abiotic transformation) or by degradation. Of these, biodegradation is the only process which actually transforms the contaminants and the "only sustainable component of natural attenuation" (Griebler & Lueders 2009). Biodegradation occurs because of microbes that use the contaminants as carbon and energy sources, and in the case of chlorinated compounds, as electron acceptors in respiration (Meckenstock et al. 2015; Azubuike et al. 2016; Das & Chandran 2011). Groundwater resources are vital to human health and to many industrial processes, and since these environments are dominated by microbes, it is important to understand microbial processes that occur at polluted sites in order to develop best practices for clean-up and attenuation. This understanding goes from the ecosystem level (which organisms are present and how do they interact?) to the biochemical level (how do individual organisms degrade hydrocarbons?).

Hydrocarbons, as the name suggests, are compounds composed only of carbon and hydrogen. They are formed by natural processes, either as metabolites in living cells or by decomposition of biomass over geologic time. Though these compounds are naturally-occurring, anthropogenic extraction, refining, transportation, and storage of petroleum resources has greatly changed their distribution in the environment (Fayemiwo et al. 2017; Ramesh et al. 2011). Interaction with hydrocarbons can have profound effects on human health (Schwarzenbach et al. 2010; Alharbi et al. 2018) and ecosystem functioning (Fleeger et al. 2003; Dew et al. 2015).

Hydrocarbons can be divided into three classes based on their structure (McNaught et al. 1997). (i) Saturated hydrocarbons contain only single bonds between carbon atoms and are saturated with hydrogen. (ii) Unsaturated hydrocarbons have double bonds (alkenes) or triple bonds (alkynes). Both saturated and unsaturated hydrocarbons can exist as straight chains, branched chains, or rings. (iii) Aromatics are distinguished by a ring-type bonding pattern that creates a system of delocalized π-electrons, resulting in highly stable molecules that are difficult to break down and are therefore persistent in contaminated environments (Fuchs et al. 2011). Aromatics are further divided into monoaromatics (which have a single ring) and polyaromatics (PAHs, which have multiple rings). Of particular interest within the monoaromatics are benzene,

toluene, ethylbenzene and xylenes – collectively called BTEX – which are common constituents of petroleum and spread easily through water systems because of their relatively high solubility as compared to PAHs (Njobuenwu et al. 2005). Despite their aromatic-conferred stability and toxicity, many bacteria have developed ways to use hydrocarbons as carbon sources and electron donors.

1.2 Bacterial hydrocarbon catabolism

1.2.1 Aerobic degradation

The delocalized π-electron system of the aromatic ring makes these hydrocarbons extremely stable, so specialized catabolic pathways are required to destabilize these bonds. The mechanisms used to activate aromatics as the first step in degradation differ based on oxygen availability (Fig. 1). When oxygen is available, it acts as a co-substrate in the initial attack performed by mono- or di-oxygenase enzymes, which add hydroxyl groups to destabilize the ring and form catechol, methylcatechol, or protocatechuate central intermediates. The catabolic steps leading to central intermediate formation are collectively called "peripheral pathways" or "upper pathways", as opposed to the "central catabolic pathways" entered by the central intermediate. Multiple mono- and di-oxygenases exist, and genes coding for these enzymes are used as biomarkers for aerobic aromatic metabolism (for review, see Gülensoy & Alvarez (1999), Parales et al. (2008), Fuchs et al. (2011), and Di Gennaro et al. (2011)). Genes encoding toluene degradation pathways are known to be located on mobile genetic elements that can be shared by horizontal gene transfer (Sentchilo et al. 2000; Khomenkov et al. 2008; Bouhajja et al. 2017), so establishing phylogenetic connections is difficult.

The catechol-based intermediates must undergo another oxygen-dependent step for ring opening by either *ortho*- or *meta*-cleavage, catalyzed by ring-cleaving dioxygenases. Intradiol dixoygenases cleave the ring between the hydroxyl substituents (*ortho*- cleavage), whereas extradiol dioxygenases cleave the bond next to the substituents (*meta*- cleavage). These two families are exemplified by intradiol catechol-1,2-dioxygenase (C12O) (Hayaishi & Hashimoto 1950) and extradiol catechol-2,3-dioxygenase (C23O) (Kojima et al. 1961), and though both groups function by ring cleavage they belong to evolutionarily distinct classes, with completely different structures and catalytic mechanisms. Intradiol dioxygenases all belong to a single lineage, while extradiol dioxygenases belong to three: types I, II, and III (Vaillancourt et al. 2006). C23O enzymes are members of Type I and can be further subdivided

into three subfamilies: I.2.A contains mostly C23O from *Pseudomonadaceae* and I.2.B mostly from *Sphingomonadaceae*, while I.2.C represents C23O from a wide range of bacterial genera (Eltis & Bolin 1996). Kukor and Olsen (1996) found that C23O in subfamily 1.2.C efficiently used low amounts of dissolved oxygen, and likely evolved in environments with low oxygen and substrate concentrations. Subfamily I.2.C C23O is frequently found at contaminated sites and is used as a functional marker for aerobic, especially microaerobic, aromatic degradation (Táncsics et al. 2010; Alfreider & Vogt 2007; Hendrickx, Junca, et al. 2006; Benedek et al. 2016). After ring opening, products of ring-opening undergo a series of further steps to produce intermediates fed into the citric acid cycle.

1.2.2 Anaerobic degradation

While aerobic degradation of hydrocarbons has long been known, it was only in the 1980's that mechanisms for anaerobic activation began to come to light. The first-discovered and best-understood mechanism is fumarate addition, in which the methyl group of the aromatic is added to fumarate (Biegert et al. 1996), destabilizing the aromatic ring and producing benzylsuccinate, which is further metabolized to the central intermediate benzoyl-CoA. For many BTEX compounds, this first step is catalyzed by benzylsuccinate synthase (Bss). The gene *bssA* codes for the alpha subunit of this enzyme and is commonly used as a biomarker to investigate anaerobic aromatic degrader communities in the field (Winderl et al. 2007; Winderl et al. 2010; Winderl et al. 2008; von Netzer et al. 2013; Yagi et al. 2010; Tan et al. 2013; Larentis et al. 2013; Espinola et al. 2018). More recently, carboxylation has been suggested as an activation mechanism for unsubstituted aromatics benzene and naphthalene, producing benzoate that can be converted to benzoyl-CoA (Abu Laban et al. 2010). Hydroxylation of benzene to phenol may present a third mechanism for activation of unsubstituted aromatics (Zhang et al. 2013). All anaerobic activation mechanisms known to date funnel monoaromatic compounds into the central intermediate benzoyl-CoA.

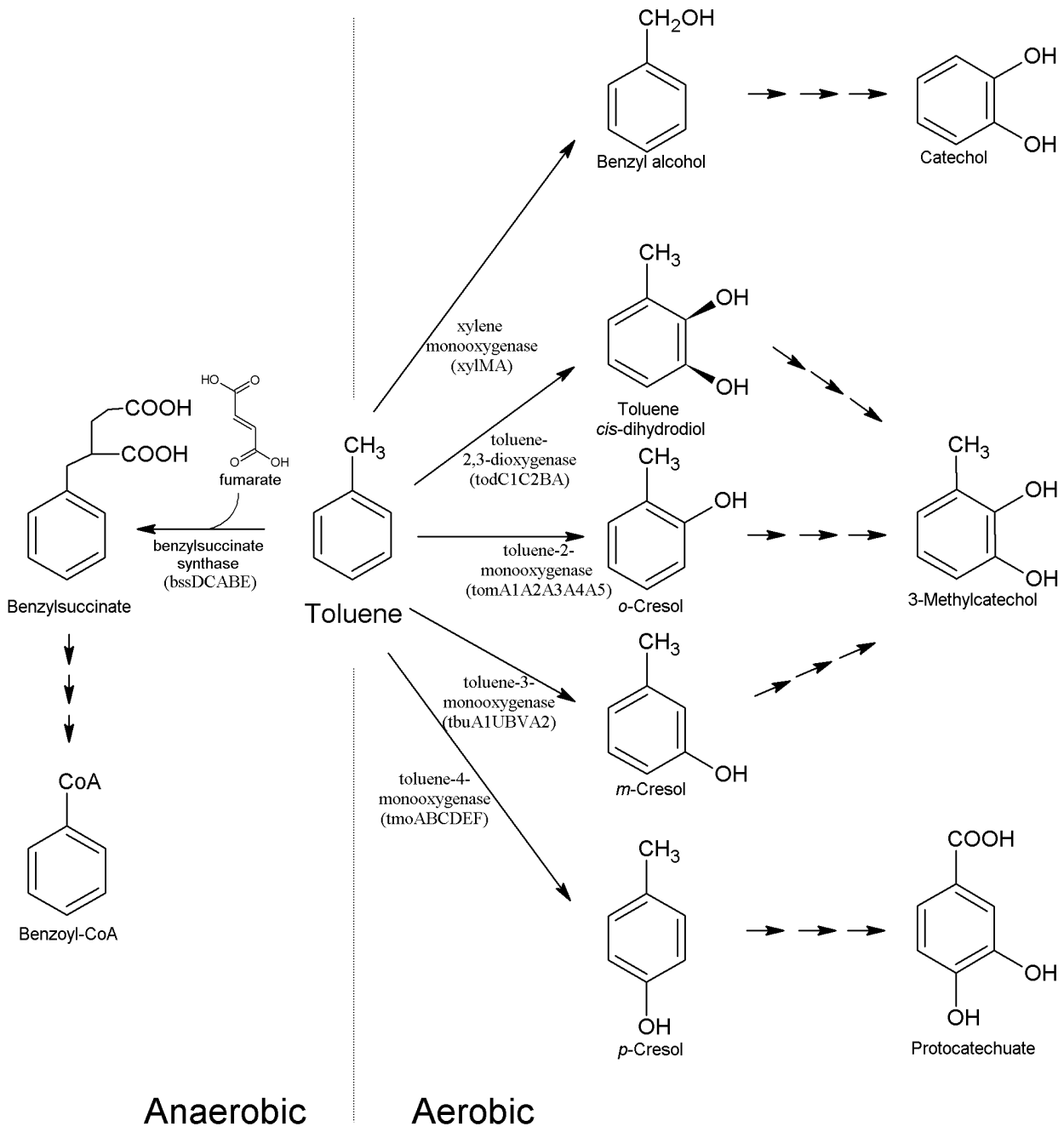

Figure 1: Peripheral (upper) pathways of aerobic and anaerobic toluene activation, with central intermediates. Activating enzymes are shown below reaction arrows, with genes in brackets. Modified from Foght (2008), Gülensoy & Alvarez (1999) and Parales et al. (2008).

Benzoyl-CoA is then dearomatized by benzoyl-CoA reductases (BCR), which are oxygen-sensitive (Boll et al. 2002). Class 1 BCR catalyze an ATP-dependent reaction and are used by facultative anaerobes, while Class 2 BCR are ATP-independent and are used by strict anaerobes. Genes coding for BCR are sometimes used as catabolic markers for hydrocarbon degradation (Sperfeld et al. 2018; Fahrenfeld et al. 2014; Sun et al. 2014) but results must be interpreted with caution since benzoyl-CoA is also a central intermediate in degradation of

common natural aromatics like humic acids, amino acids, and lignins (Carmona et al. 2009). Following dearomatization, the ring is opened by a hydrolase, for which *bamA* is a common marker gene (Porter & Young 2013), and goes through a series of ß-oxidation-like steps to produce CO_2 and three acetyl-CoA molecules.

1.3 Bacterial respiration

Respiration requires a terminal electron acceptor (TEA) at the end of an electron transport chain. Microbes which use oxygen as a TEA are defined as aerobes, while those using alternate TEAs are anaerobes. The energy yield from metabolism of a given carbon source depends on the redox regime (Tiehm & Schulze 2003), and the availability of electron acceptors is a key limiting factor for the activity of degrader populations (Callaghan 2013; Meckenstock et al. 2015). Respiring oxygen produces the most energy per molecule of substrate oxidized (Table 1), so contaminant degradation generally proceeds fastest under oxic conditions. Anaerobic respiration is more diverse in energy yield (Table 1) and therefore in kinetics.

Hydrocarbon contamination of groundwater introduces an abundance of carbon sources that stimulate microbial metabolism, rapidly depleting oxygen. As a contaminant plume spreads, a physical distribution of redox environments forms with the most reduced environments closest to the contamination source and the most oxidized at the distal end and outer edges (Christensen et al. 2000). Contrary to this classical model of reverse thermodynamic zonation, redox divisions are not absolute, and several processes can take place in the same zone. Oxygen and other terminal electron acceptors are continuously replenished by groundwater flow and mixing at the plume fringes (Bauer et al. 2008). This mixing of electron donors and acceptors creates hot spots of biodegradation (Meckenstock et al. 2015) where specific degrader lineages are enriched and overall diversity is decreased (Winderl et al. 2008; Ning et al. 2018). It is generally thought that aerobes can be active only at the capillary fringe of a contaminant plume, where dissolved oxygen is available because of diffusion and mixing, while anaerobes typically thrive towards the core of the contaminant plume, where oxygen has been depleted but other TEAs are available.

Table 1: Thermodynamics of toluene degradation with various aerobic (white), anaerobic (grey), and syntrophic (blue) conditions. From Lueders (2017).

Electron acceptor	Stoichiometry	ΔG°' (kJ per mol toluene)
O_2 / H_2O	$C_7H_8 + 9\ O_2 + 3\ H_2O \rightarrow HCO_3^- + 7\ H^+$	-3790
NO_3- / N_2	$5\ C_7H_8 + 36\ NO_3^- + H+ \rightarrow 35\ HCO_3^- + 18\ N_2 + 3\ H_2O$	-3555
$Fe(OH)_3$ / $FeCO_3$	$C_7H_8 + 36\ Fe(OH)_3 + 29\ HCO_3^- + 29\ H^+ \rightarrow 36\ FeCO_3 + 87\ H_2O$	-1497
SO_4^{2-} / HS^- (complete)	$2\ C_7H_8 + 9\ SO_4^{2-} + 6\ H_2O \rightarrow 14\ HCO_3^- + 9\ HS^- + 5\ H^+$	-203 (-45 kJ/mol SO_4^{2-})
SO_4^{2-} / HS^- (incomplete)	$2\ C_7H_8 + 3\ SO_4^{2-} + 6\ H_2O \rightarrow 6\ CH_3COO^- + 2\ HCO_3^- + 3\ HS^- + 5\ H^+$	-61 (-41 kJ/mol SO_4^{2-})
CO_2 / CH_4 (sum)	$2\ C_7H_8 + 15\ H_2O \rightarrow 9\ CH_4 + 5\ HCO_3^- + 5\ H^+$	-130

Obligate anaerobes cannot be active or sometimes even survive in the presence of oxygen (Fig. 2). In some cases, metabolic enzymes like Bss, BCRs, or nitrogenases are inactivated by oxygen. Other obligate anaerobes lack enzymes that can detoxify oxygen-free radicals or peroxide (Holland 2013). Aerotolerant anaerobes can survive oxygen exposure but do not use oxygen for respiration or catabolism. Facultative anaerobes are more adaptable and can use various methods of respiration depending on environmental conditions, taking advantage of the high energy yield of aerobic respiration when oxygen is available but relying on alternate TEAs to remain active in its absence.

Aerobes thrive at atmospheric or equivalent oxygen levels (~20 % gaseous O_2 or ~8 mg/l dissolved O), while microaerobes are typically adapted to much lower levels (Fig. 2). There is no standard definition for microoxic conditions; literature definitions range from 2 mg/L dissolved oxygen (DO) (Kim & Jaffé 2008; Chiang et al. 1989) to 10 µM (0.32 mg/L) (Unden 1999) and even down to a few nanograms per litre (Zakem & Follows 2017). Microaerobes survive at low oxygen concentration by expressing high-affinity terminal oxidases for respiration (Morris & Schmidt 2013) and other high-affinity metabolic enzymes like certain types of oxygenases as discussed above (Kukor & Olsen 1996; Leahy & Olsen 1997).

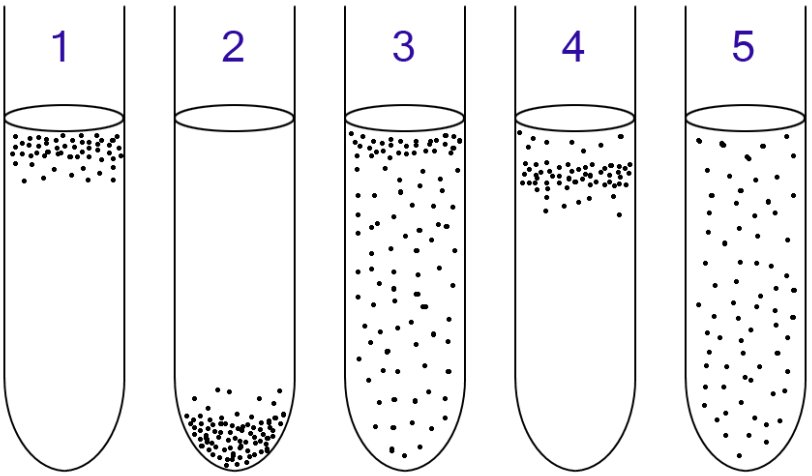

Figure 2: Conceptual demonstration of the behaviour of bacteria grown in liquid culture: Obligate aerobes (1) gather at the surface to access oxygen; obligate anaerobes (2) gather at the bottom to avoid toxic oxygen; facultative anaerobes (3) spread throughout but concentrate at the top where oxygen is available; microaerobes (4) gather in an area with low oxygen concentration; aerotolerant bacteria (5) spread evenly. From Wikimedia: user:Pixie / Wikimedia Commons / Public Domain.

Strictly aerobic and anaerobic hydrocarbon degradation have been intensively studied, but much less attention has been given to the "grey area" of degraders adapted to oxygen-limited conditions or dynamic changes in TEA availability. Considering the degradation hot-spots found at contaminant plume fringes, such organisms could make substantial but largely unrecognized contributions to bioremediation efforts. Furthermore, these organisms could be making use of a wider array of metabolic or respiratory strategies than generally considered at present.

1.4 Alternative strategies for hydrocarbon metabolism under limited electron acceptor availability

1.4.1 Oxygen-activation, nitrate-respiration

Some specialized microorganisms in hydrocarbon-contaminated environments seem to have adapted to oxygen-limited conditions by employing oxygen only to activate aromatics while using nitrate as an alternative TEA, reducing the oxygen requirements for metabolism and perhaps offering a competitive advantage in fluctuating environments. By definition these microbes are anaerobes, but they must be aerotolerant and could even function as facultative anaerobes if oxygen levels rise high enough. Concurrent use of oxygen and nitrate in this manner was hypothesized to occur in contaminated aquifer systems nearly 30 years ago

(Britton 1989) and has since been supported in a few studies of pure cultures (Leahy & Olsen 1997), environmental samples (Durant et al. 1999), and engineered systems (Ma & Love 2001; Liu et al. 2018). This oxygen-activating, nitrate-respiring (OANR) strategy may overlap with "aerobic denitrification", which is understood as the simultaneous use of both O_2 and NO_3^- as TEAs in individual microbes. Aerobic denitrification is known to occur in at least 37 isolated bacterial species, with potential utility in wastewater treatment and nitrogen removal from groundwater (Ji et al. 2015). A few pure cultures that seem to be capable of using the OANR strategy have been isolated and studied (Leahy & Olsen 1997; Britton 1989), but overall very little is known about which organisms employ this strategy, what mechanisms they use, or how common it is in environmental systems.

1.4.2 Oxygenic/intra-aerobic pathways

Recent work by Larentis et al. (2013) challenges the idea that aerobes and anaerobes are confined to distinct areas of a contaminant plume. This study investigated a fine-scale microbial community profile in the Flingern aquifer in Düsseldorf, Germany, a tar-oil contaminated site where toluene is the dominant contaminant (Anneser et al. 2010). The study analyzed 16S rRNA, dioxygenases (*todA* and *todC*), monooxygenase (*tmoA*), and benzylsuccinate synthase (*bssA*) genes. The expectation was that there would be a succession of aerobic to anaerobic toluene degraders from the upper plume fringe to the plume core. Instead, *bssA* was found across the oxic/anoxic redox gradient and, very surprisingly, *tmoA* genes were present right into the centre of the toluene plume, with the maximum abundance in the strongly reduced plume core (Fig. 3). The *tmoA* genes were cloned and sequenced, and found to be most closely related to other environmental clones from BTEX contaminated sites. Toluene monooxygenase gene sequences have also been recovered from other contaminated aquifers in zones of low or no dissolved oxygen (Hendrickx, Dejonghe, et al. 2006; Hendrickx et al. 2005; Cavalca et al. 2004; Ogram et al. 1995; Nebe et al. 2009).

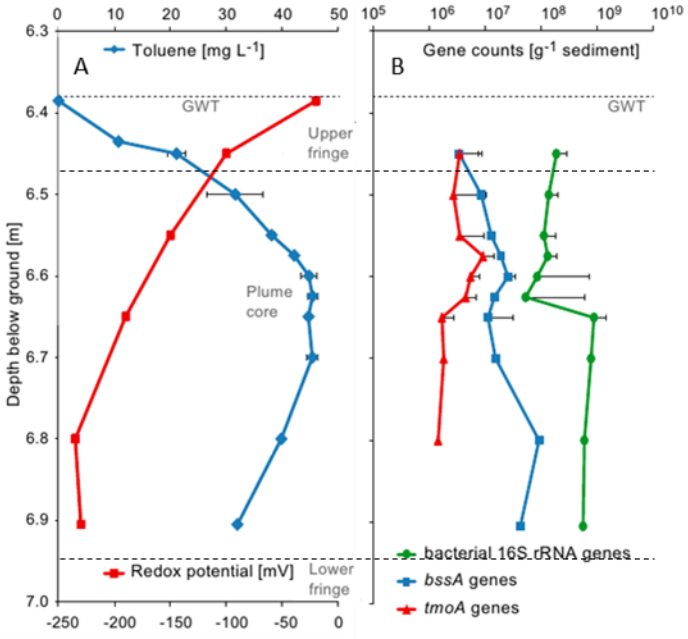

Figure 3: Depth profile of the contaminated Flinger aquifer showing (A) toluene concentration, redox potential, and plume zonation over depth (modified from Anneser et al. (2010)) and (B) Abundance of 16S rRNA, bssA and tmoA genes as measured by qPCR. Modified from Larentis et al. (2013).

Since toluene-4-monooxygenase, the enzyme coded for by *tmo* genes, is oxygen-dependent, the presence of its coding DNA (Larentis et al. 2013) or mRNA (Ogram et al. 1995) in anoxic environments is currently a mystery. Larentis et al. (2013) suggests that the *tmoA* sequences may be "memory effects": the genes remain present because oxygen was previously available when the water table was in a different position, but are no longer active. There could also be micro-niches with sufficient oxygen for aerobic metabolism. The mRNA found by Ogram et al. (1995) must have been actively transcribed in their anoxic subsurface site, though the authors suggest that it may be an artifact of sample handling if there was accidental exposure to oxygen which induced transcription, or if cross-hybridization occurred with their gene probe. But recent findings in anaerobic methane and alkane metabolizers present another possibility: an intra-aerobic pathway dependent on anaerobic oxygenesis using nitrogen-oxygen species.

Research by Ettwig et al. (2010) on enrichments from nitrate-rich freshwater sediments identified the bacterium "*Candidatus* Methylomirabilis oxyfera" and assembled its genome. These enrichment cultures grow anaerobically on methane, with complete reduction of nitrate to N_2. However, the genome of "*Ca.* Methylomirabilis oxyfera" is missing the *nosZDFY* gene set that codes for nitrous oxide reductase, required for reduction of N_2O to N_2. However, it has the entire complement of genes necessary for aerobic methane oxidation. Transcriptomics and

proteomics found that the necessary sequences are not only present in DNA, but are transcribed and translated in anoxic conditions. The enrichment culture was unable to metabolize methane with only nitrate or N_2O available as electron acceptors, but became active upon addition of nitrite. Particulate methane monooxygenase (pMMO) carries out the initiation step of aerobic methane oxidation by using oxygen to hydroxylate methane, yielding methanol and water. Assays of pMMO activity found higher activity rates with nitrite present than with oxygen, and no activity with nitrate or N_2O. They also found that N_2 was produced, despite the lack of genes for reduction of N_2O to N_2. Based on these results, Ettwig et al. (2010) proposed the existence of an intra-aerobic pathway taking place in "*Ca.* Methylomirabilis oxyfera", wherein NO is split into N_2 and O_2 by an unknown dismutase enzyme, and the oxygen used by pMMO to activate methane (Fig. 4B).

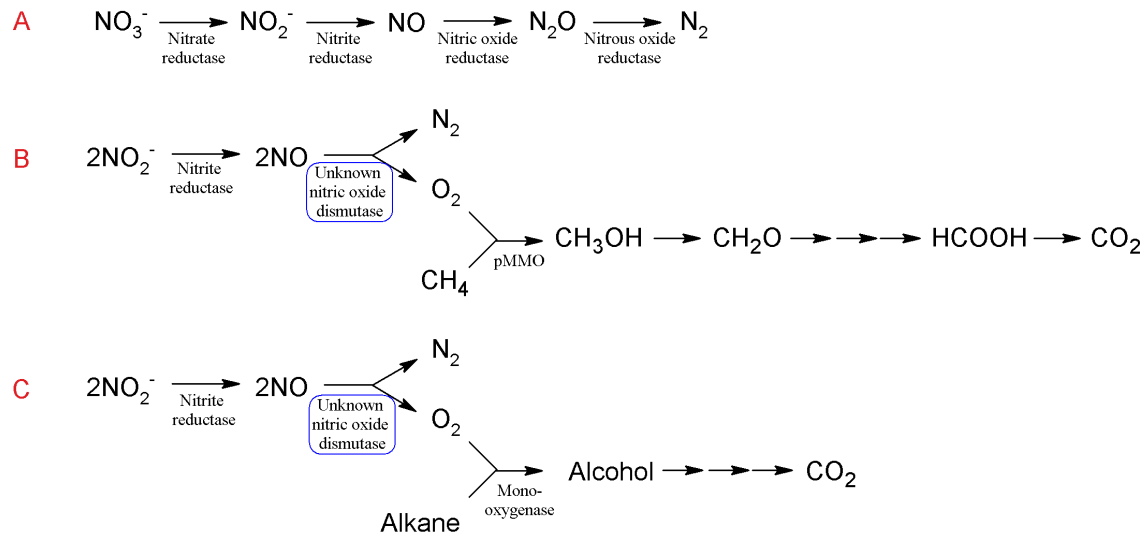

Figure 4. (A) The complete canonical denitrification pathway; proposed oxygen production by dismutation connected to (B) methane oxidation (proposed by Ettwig et al. (2010)) and (C) alkane degradation (proposed by Zedelius et al. (2011)). Adapted from Ettwig et al. (2012).

Zedelius et al. (2011) suggested an analogous mechanism in the denitrifying alkane-degrading *Gammaproteobacteria* strain HdN1. Aerobic activation of alkanes is performed by monooxygenases, and anaerobic activation by addition to a fumarate molecule by 1-(methylalkyl)succinate synthase or alkylsuccinate synthase. The genome of HdN1 does not contain the *mas* or *ass* genes for these anaerobic activation enzymes, but does contain genes for two, possibly three, monooxygenases. HdN1 does not produce metabolites expected from addition to fumarate. Unlike "*Ca.* methylmirabilis oxyfera", the HdN1 genome has the entire set of genes necessary for complete denitrification. Tests showed that it was able to grow when

nitrate or nitrite was provided, but not with N_2O, despite having the necessary genes for N_2O reduction. Addition of NO was toxic to the culture, so this study was unable to ascertain if nitrite, NO, or some unknown product is the essential ingredient for activation.

Anaerobic oxygenesis is known to be performed in some perchlorate and chlorate respiring bacteria (Van Ginkel et al. 1996; Rikken et al. 1996) via chlorite dismutation by means of well-characterized chlorite dismutase enzymes (Schaffner et al. 2015). In most cases, this oxygen is used as the terminal electron acceptor for aerobic respiration, but it may be used in monooxygenase-catalyzed alkane activation by *Pseudomonas chloritidismutans* (Mehboob et al. 2009; Mehboob et al. 2016) or aromatic activation by *Alicycliphilus denitrificans* (Weelink et al. 2008). Ettwig et al. (2012) elaborated on the possible identity of the proposed nitric oxide dismutase (NOD) that would be necessary for these intra-aerobic pathways. Bacterial nitric oxide reductases (NORs) reduce toxic NO to harmless N_2O. The "*Ca.* Methylomirabilis oxyfera" and HdN1 genomes share two paralogs of quinone-using NORs, which contain "both resemblances and significant differences with respect to canonical qNORs" (Ettwig et al. 2012). In particular, substitutions to some amino acids in the active site inactivate quinol-binding. The gene is highly transcribed in "*Ca.* Methylomirabilis oxyfera", so it can be assumed that the product has an important biological function. The authors propose that these NOR-relatives are the putative NODs necessary for oxygenesis in nitrate-reducing conditions, and call for purification and characterization of the enzymes to explore this hypothesis.

Primers targeting "*Ca.* Methylomirabilis oxyfera"-related putative *nod* genes were recently developed in a study of a methane-oxidizing, nitrite-reducing laboratory reactor inoculated with river sediments (Bhattacharjee et al. 2016), but finding evidence of more diverse *nod* genes that may exist in the environment requires development and testing of a suite of suitable primers on a range of environmental samples.

Considering the low oxygen requirements by OANR strategists, and the apparent ability of at least a few organisms to produce oxygen for hydrocarbon activation via nitrite dismutation, an intriguing possibility exists for overlap between these groups. Understanding the prevalence of these processes in the environment and their role in degradation of compounds beyond methane and alkanes could open the way to new strategies for bioremediation of hydrocarbon contaminated sites as well as elimination of nitrogen species from groundwater environments. The responsible organisms and their optimal conditions remain a mystery, so cultivation is extremely challenging and insights must instead rely on molecular biology methods.

1.5 Molecular biology methods to investigate pollutant degraders in complex microbiota

1.5.1 Stable isotope probing

Stable isotope probing (SIP) is a well-established method in molecular ecology which allows the identification of organisms that metabolize a specific substrate of interest. A substrate artificially labelled with rare stable isotopes (most commonly ^{13}C, but also ^{15}N or ^{18}O) is provided, and any organisms that metabolize it incorporate the heavy isotope into their biomolecules. Various methods exist to analyze these biomolecules, which include phospholipid fatty acids (PLFA), DNA, RNA, and proteins (Gutierrez-Zamora & Manefield 2010; Lueders et al. 2016). The first SIP studies were performed in 1998 using PLFA, and the technique was quickly adapted for use with DNA and RNA (2000 and 2002, respectively) (Uhlik et al. 2013). An overview of the associated methods, advantages, and disadvantages of each biomolecule are shown in Table 2.

Table 2: Analysis method, strengths, and limitations of major biomarkers used in SIP. From Lueders et al. (2016).

	Analysis Method	**Strength**	**Limitation**
DNA	Isopycnic centrifugation, various downstream analysis options (ex. fingerprinting, sequencing)	- Community composition from SSU rRNA - Metabolic potential via functional genes - Potential for genome assembly	-Active state unknown -Labelling depends on replication
rRNA	Isopycnic centrifugation, various downstream analysis options (ex. fingerprinting, sequencing)	- Community composition from SSU rRNA - Active organisms only - Resolution of rapid changes - Rapid labelling, independent of replication	-No functional gene data -Less stable than DNA; risk of degradation
mRNA	Isopycnic centrifugation, various downstream analysis options (ex. fingerprinting, sequencing)	- Direct ties to community function - Actively transcribed genes only - Differential expression under varying conditions - Resolution of rapid changes - Rapid labelling, independent of replication	-No taxonomic data -Unstable; high risk of degradation -Low quantities (<5% of total RNA); enrichment or amplification required
PLFA	Gas chromatography – (isotope ratio) mass spectrometry	- High sensitivity - Can calculate bacterial abundance	-Very low taxonomic resolution
Protein	Mass spectrometry	- Direct ties to metabolic activity - Most sensitive SIP - Short incubation time required	-Labour intensive -Requires metagenome sequences for interpretation

RNA is more sensitive to labelling than DNA and responds more dynamically to changes in environmental conditions (Dumont et al. 2011). This is because RNA is produced continuously, while DNA is only produced (and therefore labelled) during cellular replication. RNA-SIP was first developed to investigate degradation of hydrocarbons in a water treatment system (Manefield et al. 2002), and this remains a popular application. The typical workflow for an RNA-SIP experiment is shown in Fig. 5, while the wide applicability of the method along with selected publications are summarized in Fig. 6.

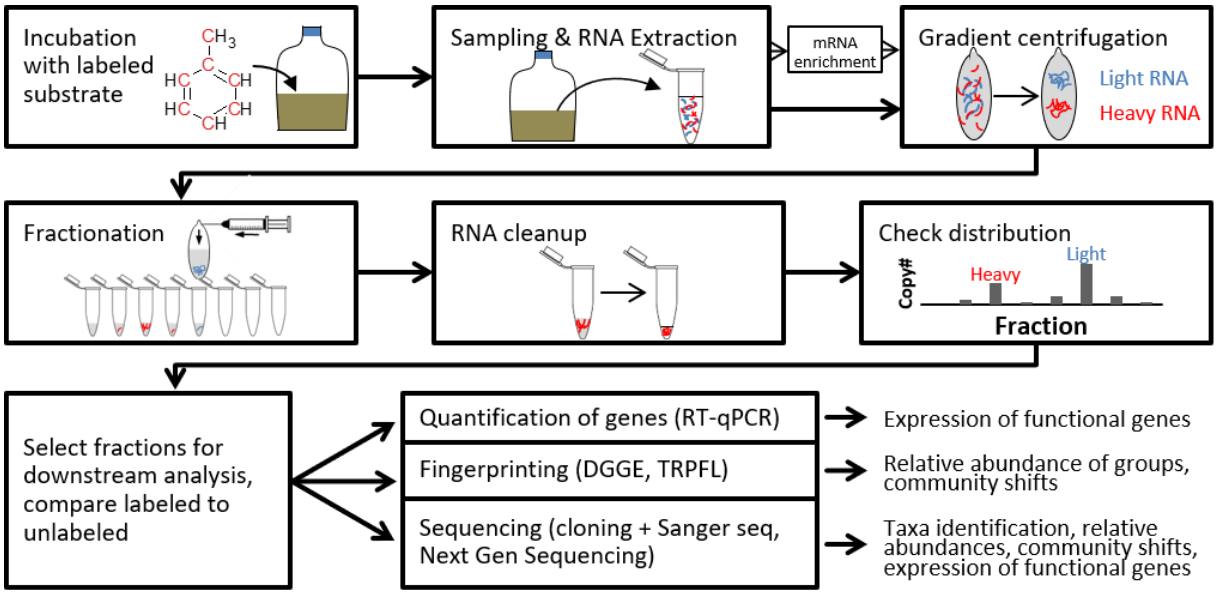

Figure 5: General workflow of an RNA-SIP experiment with suggestions for downstream analyses. From Lueders et al. (2016).

In the context of this thesis, SIP – specifically the novel method of total-RNA-SIP developed as part of this thesis and discussed in section 1.5.1.1 - was employed to target relevant hydrocarbon (specifically toluene) degraders in a variety of relevant redox conditions. This is valuable compared to more common methods like enrichment and untargeted sequencing of the whole community for multiple reasons. Enrichment relies on pre-knowledge of optimal growth conditions for the organism(s) of interest, which is especially challenging for putative nitrite-dismutating and/or oxygen-activating nitrate-respiring microbes whose identity and preferred habitat conditions are currently still unclear. Even with such knowledge, many microbes defy isolation because they rely on syntrophic metabolic partners. Enrichment can also be extremely slow, especially for anaerobic cultivation, where doubling times can sometimes be measured in weeks or months. Sequencing of the entire community, whether

directly from an environmental sample or during cultivation, can reveal community constituents or expressed functions, but cannot always distinguish between active degraders and those who metabolize other electron donors.

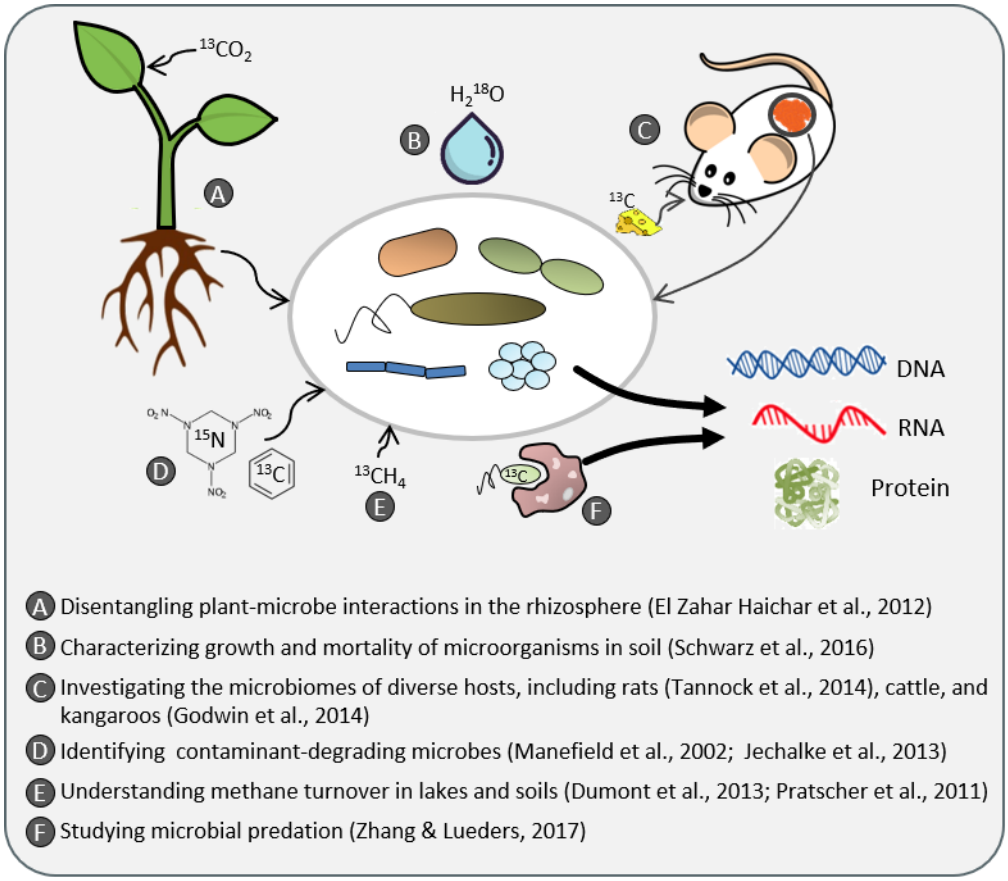

Figure 6: Overview of RNA-SIP applications, with selected example publications.

Various methodological improvements have increased the effectiveness of nucleic-acid SIP since its inception, notably the replacement of ethidium-bromide visualization and band selection with gradient fractionation (Lueders et al. 2004). More recently, quantitative SIP (qSIP), wherein all fractions recovered from SIP gradients are subjected to PCR amplification and sequencing, has enabled calculation of the isotope incorporation for each individual taxa (Hungate et al. 2015). This also alleviates the problem of GC-content based density differences in genes or transcripts, which complicate interpretation in more traditional SIP approaches wherein only selected "heavy" and "light" fractions are analyzed (Youngblut & Buckley 2014). The more common and considerably less expensive method for SIP, dubbed "Heavy-SIP" by a recent publication (Youngblut et al. 2018) is to sequence fractions from specific "heavy" and

"light" density fractions. In this case, it is essential to include controls provided with un labelled substrate in order to account for GC-based density shifts during calculations. Kramer et al. (2016) introduced an enrichment factor (EF) calculation using abundance data from the heavy and light fractions of isotopically labelled and un labelled samples. Although EFs are not strictly quantitative, they can be used to rank the degree of labelling between taxonomic or functional groups of interest, giving a much more detailed view of each group's contribution to substrate metabolism than qualitative analyses.

Nucleic acid SIP is generally followed by PCR amplification of the small subunit ribosomal RNA (SSU rRNA; 16S in bacteria or archaea and 18S in eukaryotes) from gradient fractions. SSU rRNA codes for essential cellular machinery and is present in all cells in varying copy numbers. Amplification and sequencing of SSU rRNA reveals the taxonomic identity of the organisms in a system, and when combined with SIP this can answer the question "who in this system metabolizes my substrate-of-interest?" However, primers considered "universal" for prokaryotes are known to miss some organisms (Brooks et al. 2015; van Dijk et al. 2014; Tremblay et al. 2015). Resulting biases from primers and from preferential amplification in PCR, which propagates exponentially, can influence apparent community composition (Poretsky et al. 2014; Laursen et al. 2017). For this reason, and as sequencing technologies continuously improve, more researchers in environmental studies are turning to PCR-free shotgun metagenomics and metatranscriptomics.

However, use of PCR-free methods can complicate the calculation of EFs, because the miniscule amount of RNA available in heavy fractions from unlabelled controls, or light fractions from labelled microcosms, can be insufficient for sequencing. This problem, along with an improved EF calculation to overcome it, is further discussed in Chapter 2.8.3 (Methods) and Appendix 1.

1.5.1.1 mRNA-SIP

The key limitation of SSU rRNA SIP is that it provides only taxonomic, not functional, information about process-relevant microbes. Isopycnic gradients generally have the capacity to resolve both labelled rRNA and mRNA (Rickwood 1992). A small number of studies in recent years have advanced mRNA-SIP, in most cases interrogating the resulting fractions with reverse-transcription PCR and/or DGGE fingerprinting (Huang et al. 2009; el Zahar Haichar et al. 2012; Pratscher et al. 2011). Huang et al. (2009), combined rRNA-SIP and mRNA-SIP

with single-cell Raman fluorescence in situ hybridization (FISH) and DGGE to study aerobic naphthalene degraders in groundwater. Dumont et al. (2011) compared DNA-, rRNA, and mRNA-SIP to study aerobic methanotrophs in lake sediment, using reverse transcription PCR (RT-PCR) and terminal restriction fragment length polymorphism analysis (TRLFP) to investigate SIP-labelled transcripts of *pmoA*, a methane monooxygenase gene. This study confirmed that labelling of mRNA was much faster than labelling of DNA. el Zahar Haichar et al. (2012) used DNA- and mRNA-SIP to study nitrate-reducing bacteria in a rhizosphere system. They used RT-PCR to investigate denitrification genes, then performed DGGE and sequenced DGGE-excised bands. These studies proved that functional transcripts can be explored using SIP, but could only provide insight into the labelling of specific transcripts of interest. Such approaches can be limited by lack of suitable primers, and the workload to investigate a high number of genes at the same time is untenable. Only three studies to date have combined mRNA-SIP with next-generation metatranscriptome sequencing (Fortunato & Huber 2016; Dumont et al. 2013; Trembath-Reichert et al. 2019). In their pioneering 2013 study, the Dumont group used mRNA-SIP to look at aerobic methanotrophs in lake sediment, this time in combination with next-generation sequencing (Dumont et al. 2013).They were able to trace expressed pathways of methane oxidation to CO_2, and discovered highly abundant labelled motility transcripts that may allow methanotrophs to precisely position themselves within redox gradients. Since mRNA typically makes up a very small proportion of total RNA (<5 %) (Lodish et al. 2000), their workflow involved the enrichment of mRNA by rRNA depletion prior to isopycnic centrifugation, a step commonly used in transcriptome studies to increase mRNA sequencing depth.. This removed all taxonomic data present in the form of 16S rRNA reads, requiring additional preparation of PCR-amplified 16S rRNA libraries for a comprehensive interpretation of taxonomic and functional labelling. They also performed linear amplification of mRNA fractions recovered from gradients to produce sufficient material for 454 pyrosequencing, which may have introduced biases to expression profiles.

The majority of transcriptome studies to date focus on mRNA by enriching mRNA or depleting rRNA, which makes up the vast majority of total RNA (Petrova et al. 2017; Lodish et al. 2000), to open up sequencing space for functional transcripts. In eukaryotes, mRNA has a polyadenylated tail which can be exploited by bead hybridization to select only mRNA from extracted total RNA. Polyadenylation in prokaryotic mRNA, on the other hand, signals for exonuclease degradation (Belasco 2010; Deutscher 2006), so that poly-A tails in prokaryotic mRNA are rare and enrichment impossible. Instead, the focus is often on rRNA depletion, with

the two most popular methods being subtractive hybridization and exonuclease treatment (Carvalhais et al. 2012), both available as commercial kits. However, this removes most 16S-based taxonomic information, which is of great interest in research on communities in environmental samples. For this reason, a few studies have employed total-RNA sequencing (aka the "double-RNA" approach) as pioneered by Urich et al. (2008). This approach provides both taxonomic and functional information while avoiding possible biases introduced by primers and exponential amplification inherent in PCR amplicon studies. It also avoids the additional processing step of rRNA depletion, which increases the risk of RNA degradation and introduces varying amounts of bias to the results (He et al. 2010).

Of the few mRNA-SIP studies done so far (outside of this thesis), just two have performed total-RNA-SIP (i.e. without rRNA depletion). Fortunato & Huber (2016) collected deep-sea hydrothermal vent fluid samples and provided them with ^{13}C-labelled bicarbonate under hydrogen-rich conditions to explore chemolithoautotrophic metabolic potential at varying incubation temperatures. They found a high degree of labelling and that different phylogenetic groups dominated at different temperature regimes, with hydrogen oxidation coupled to oxygen or nitrate reduction as the primary metabolism. A later study by the same group (Trembath-Reichert et al. 2019) again used total-RNA-SIP, now combined with nanoSIMS analysis, to study similar deep-sea hydrothermal vent communities. This time, they provided the first microbial community characterization of hydrothermal vent fluids from the Marian back-arc basin using ^{13}C-labelled bicarbonate SIP incubations. However, neither of these projects made use of EFs to aid interpretation, and the difficulties of deep-sea sampling and shipboard laboratory work prevented adequate microcosm replication for quantitative analyses. Fortunato & Huber (2016) acknowledged that density fractions with very low RNA concentrations are largely comprised of sequences known to be a common contaminant in extraction kits. This highlights a major limitation of working with heavy-unlabelled or light-labelled fractions, which is also required for EF calculation. As part of this thesis work, I developed an updated EF calculation (Appendix 1) that does not require sequence data from heavy-unlabelled fractions, bypassing the problem of library production from extremely-low RNA amounts.

1.5.1.2 Linear amplification

Linear amplification of RNA is one way to effectively work with the small amounts of RNA that can be retrieved from environmental samples or SIP gradients. The process involves polyadenylating all RNA strands and using a T7 primer to prime for complementary DNA (cDNA) synthesis. Double-stranded cDNA is then used as a template for in vitro transcription, which results in multiple copies of each original RNA template strand and hundreds to thousands of times more RNA output than input (Fig. 7). Cao et al. (2010) hypothesized that mRNA should be enriched during polyadenylation-derived amplification because the structured conformations and lack of canonical polyadenylation of rRNA and tRNA will reduce the efficacy of the first step in the procedure for these groups of RNA. This could result in a second advantage of linear amplification by increasing the proportion of mRNA-based reads in a total-RNA sequencing project.

1.5.2 RNA-seq

DNA sequencing was pioneered in the 1970's and was originally based on using dideoxynucleotide analogues as specific chain-terminating inhibitors of DNA polymerase (Sanger et al. 1977). This method was later combined with capillary electrophoresis and fluorescent detection technology to increase throughput, and "Sanger sequencing" is still in use today for projects which require long read lengths and high accuracy but low-throughput, including the clone library method used during primer development in this thesis. Development of Next Generation Sequencing (NGS) followed, driven largely by the Human Genome Project, with a few companies developing their own high-throughput sequencing technologies. Illumina currently holds the greatest market share (Terry 2019), with a range of machines that perform short-read, massively high-throughput sequencing. Recent developments by Pacific Biosciences and Oxford Nanopore have introduced long-read sequencing - for a review, see van Dijk et al. (2018). Nanopore has recently explored direct sequencing of single RNA molecules (Garalde et al. 2018), but to date this intriguing method has been used in only one proof-of-principle study and thus is not suitable for routine work. Since these long-read technologies currently cannot provide the quantitative read counts necessary to interpret transcript expression in my projects, Illumina sequencing was used for sequencing in conjunction with SIP work in this thesis.

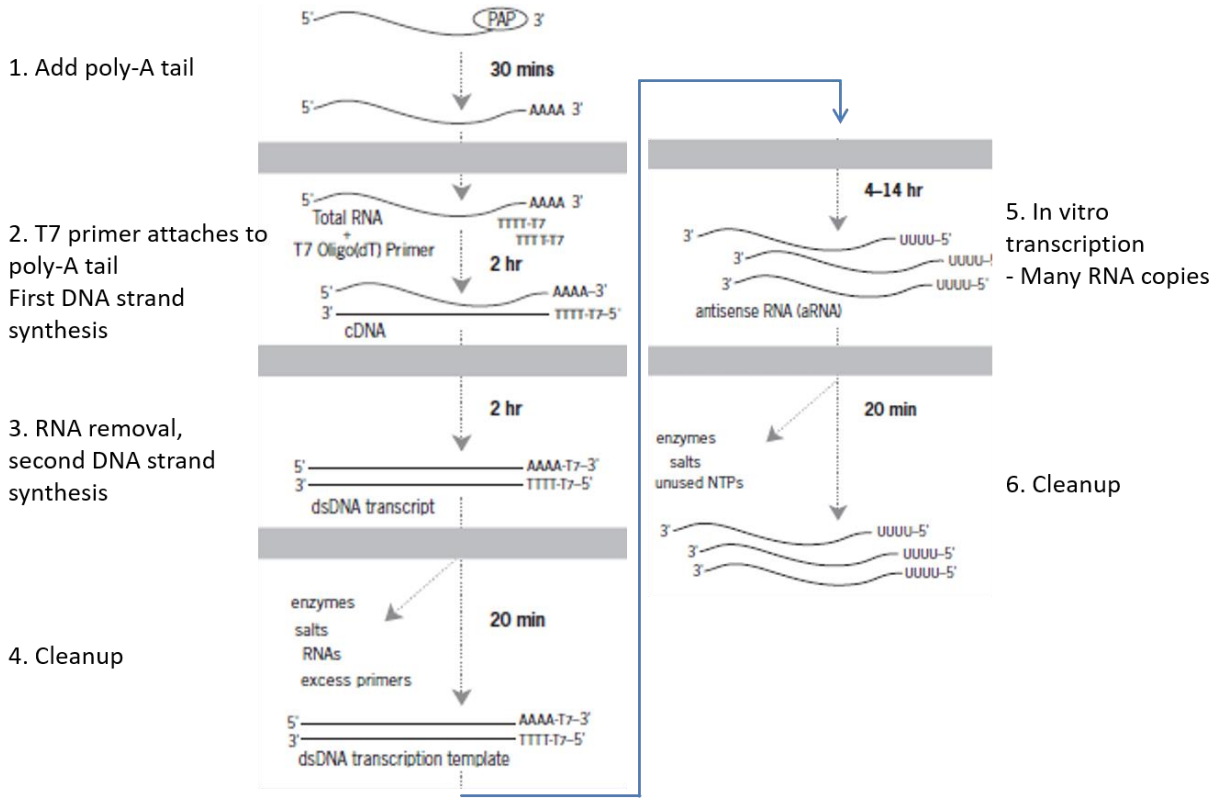

Figure 7: Overview of linear RNA amplification. Adapted from the MessageAmp II Bacteria protocol (Thermofisher).

Illumina sequencing technology is based on bridge amplification of template strands which are attached to a glass flow-cell, then on detection of fluorescently- labelled nucleotides as they are added one-by-one to match the template strand (Fig. 8B). Only double-stranded DNA can be ligated to adapters, so RNA library preparation relies on reverse transcription of RNA into double-stranded complementary DNA (cDNA) before adapter ligation. Alternately, cDNA synthesis can be performed prior to library preparation to protect samples from RNAse degradation during storage and further processing (Fig. 8A).

Products of library preparation are bound to adapter-complementary sequences scattered across the flow-cell, and the second strand is cleaved away, leaving a single-stranded product which can undergo bridge amplification (Fig. 8B). Bridge amplification makes a cluster of forward and reverse sequences which would obscure signals, so one type is cleaved away, leaving a cluster of identical sequences in a patch. This is necessary to produce a strong enough signal during nucleotide addition for the machine to actually read DNA. Now that cluster generation is complete, sequencing-by-synthesis can proceed by step-wise addition of fluorescently labelled nucleotides. The first generation of Illumina machines could only produce sequence

lengths of 25 nt, but the current chemistry gives single reads of up to 300 nt. The effective length can be extended by paired-end sequencing, wherein the template strand is sequenced in one direction, then flipped and sequenced from the primer at the other end, giving forward and reverse reads which overlap in the middle and can be collapsed into a single sequence read during data processing.

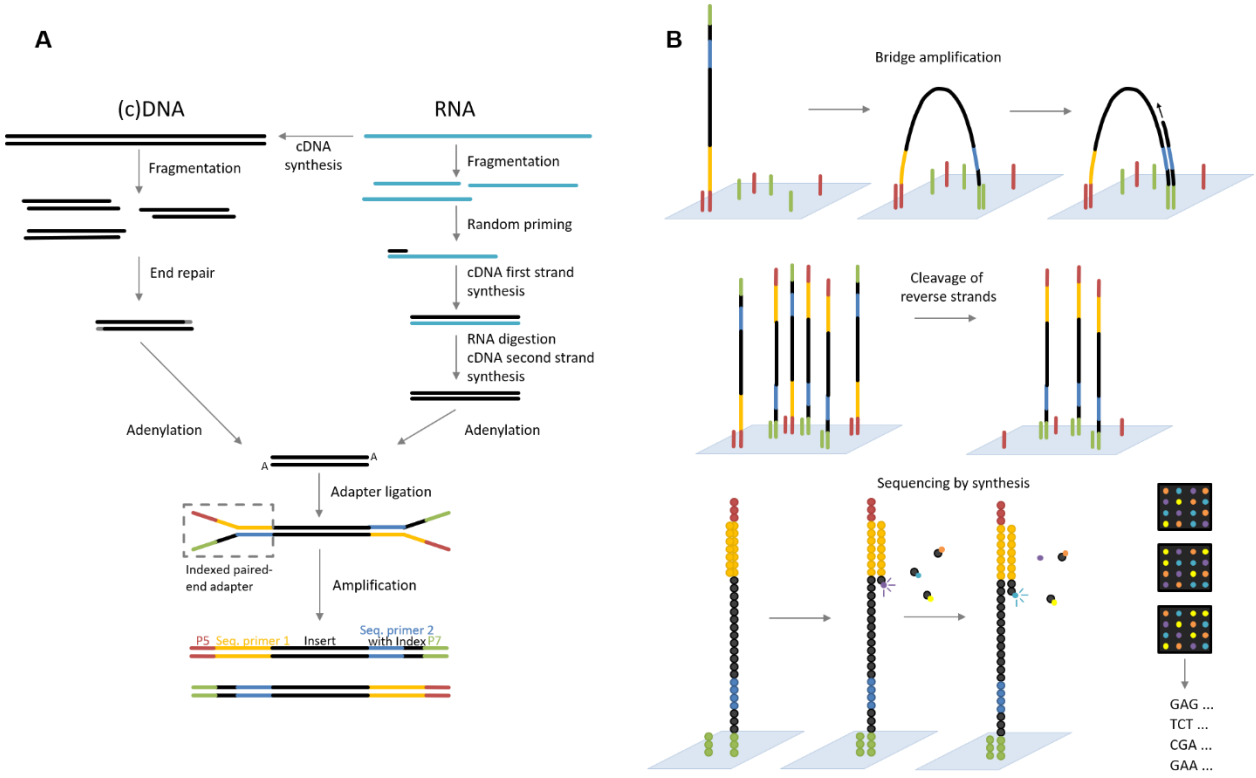

Figure 8: Conceptual overview of (A) Illumina library preparation from DNA or RNA; (B) steps in Illumina sequencing

Sequencing output can be processed for quality and each read annotated by comparison to databases of known taxonomic assignment or functional gene identity. Combined with SIP, RNA-seq of total-RNA (or cDNA derived from total-RNA) made it possible to explore the taxonomy and functional gene expression of hydrocarbon degraders in microoxic and fluctuating redox microcosms derived from contaminated sediment.

1.6 Project outlines and goals of the thesis

While strictly aerobic and anaerobic hydrocarbon degradation have been well studied in recent decades, the hotspots of degradation in groundwater contaminant plumes occur at the fringes, where availability of oxygen and other TEAs that drive microbial metabolism may be in constant fluctuation. Accordingly, we posit that the importance of microaerobes, degraders that reserve limited oxygen supply for activating aromatics while respiring alternative TEAs, or possibly even "intra-aerobic" degraders that produce their own oxygen, may have been largely overlooked. Moreover, some enrichments and pure cultures are known to employ an oxygen-activating, nitrate-respiring (OANR) strategy to metabolize aromatic hydrocarbons during oxygen limitation (Leahy & Olsen 1997; Liu et al. 2018). Uncovering the identity and gene expression of OANR strategists will bring us closer to identifying and isolating degraders that may be much more relevant in complex natural settings than currently perceived.

This results in three overarching questions addressed in this thesis:

1. How widespread is the putative nitric oxide dismustase (*nod*) gene in contaminated aquifers and other aquatic systems? Putative Nod is believed to produce oxygen that is used for aerobic activation of hydrocarbons in anoxic environments. Based on the discovery of dismutating oxygenesis in methane and alkane metabolism (Ettwig et al. 2010; Zedelius et al. 2011) and the presence of oxygen-dependent aromatic metabolism genes in an anoxic contaminant plume (Larentis et al. 2013), I hypothesized that oxygenic metabolisms are widespread in the environment, including contaminated groundwater. To investigate this, I iteratively designed and tested PCR primers targeting *nod*, and used the best primers to examine the prevalence of this gene in hydrocarbon-contaminated aquifers and other aquatic systems. The results are published in the journal Applied and Environmental Microbiology (Zhu et al. 2017).

2. How can the development of semi-quantitative total-RNA stable isotope probing (SIP) improve upon the functionality of SIP compared to more common amplification-based and/or non-quantitative methods? I performed a proof-of-principle total-RNA-SIP experiment on a set of microcosms that were based on sediment from a BTEX-contaminated microoxic aquifer in Siklós, Hungary and incubated with ^{13}C-labelled toluene. To optimize data interpretation, I established a bioinformatics workflow for SSU rRNA and mRNA sequencing reads and used the enrichment factor (EF) calculation first published by Kramer et al. (2016) to semi-quantitatively rank degrees of labelling. A DNA-SIP experiment on these same microcosms provided a direct comparison for results during method development. To further extend the

utility of total-RNA-SIP, I performed independent experiments testing the effectiveness and biases of two methods that could improve sequencing space for mRNA: linear amplification and targeted 23S rRNA depletion. Results are published in the journal Frontiers in Microbiology (Bradford et al. 2018).

3. Are oxygenic or oxygen-activating, nitrate-respiring (OANR) microbes involved in hydrocarbon degradation under limited terminal electron acceptor (TEA) conditions? I hypothesized that OANR strategists have a competitive advantage under oxygen limitation when nitrate is available, and that the necessary oxygen could be produced by oxygenesis. I used sediment from a BTEX-contaminated aquifer in Dusseldorf, Germany, for a series of microcosms incubated under various redox conditions, including microoxic conditions with and without nitrate. Using ^{13}C-labelled toluene and the total-RNA-SIP method developed in the previous project, I compared the identities and functional transcripts of hydrocarbon degraders. I employed the bioinformatics workflow developed in the proof-of-concept project and searched for *nod* and transcripts involved in toluene catabolism by a variety of methods. To overcome the reliance on sequencing library preparation from the extremely low amounts of RNA retrieved from certain SIP gradient fractions, I developed an improved EF equation that requires data from fewer fractions. The manuscript resulting from this project is in preparation for submission to the ISME journal.

2. Methods

2.1 Study sites and sampling

2.1.1 Siklós study site

The Siklós aquifer in southwest Hungary is contaminated by BTEX leaked from underground storage tanks at a former petrol station. Xylene and ethylbenzene are the most abundant compounds and have approximately equal plume sizes. Further details of the Siklós aquifer, including well placement, can be found in Táncsics et al. (2012).

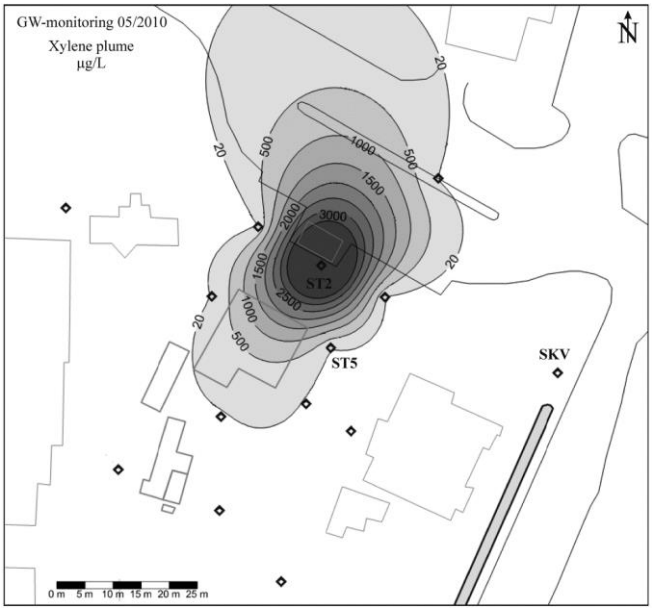

Figure 9: Map of Siklós aquifer. Sampling wells are indicated with diamonds. Xylene concentrations (µg/L) measured in 2010 are indicated by shading. From Táncsics et al. (2012).

Sediment sampling was performed by by Dr. András Táncsics (Szent István University, Godollo, Hungary), Milán Farkas (SIU), and Anna Szalay (IGOE, Helmholtz Zentrum München) in April 2015. Sediment was retrieved from the bottom of monitoring well ST-2 (Fig. 9) at 6 m depth by vacuum pumping into a clean 10 L plastic jerrycan. Sediment was allowed to settle for approximately 20 minutes before being transferred to a sterile 1 L glass bottle along with *in situ* groundwater. Bottles were completely filled to minimize atmospheric exposure and were kept cool during transport to the laboratory.

2.1.2 Flingern study site

The aquifer in the Flingern district of Düsseldorf, Germany, sits beneath a former gasworks site that was in operation from 1893 until 1967. Production of benzene, coke, gas, and tar oil led to the formation of an anoxic hydrocarbon plume composed primarily of BTEX along with some PAHs. Remediation efforts in the mid-1990's removed much of the tar oil, reduced the plume size, and brought BTEX concentrations from > 100 mg/L to an average of 20 mg/L (Anneser et al. 2008). This site has been extensively studied in recent decades, particularly in regard to various aspects of the microbial communities active in hydrocarbon remediation (Anneser et al. 2008; Winderl et al. 2008; Anneser et al. 2010; Winderl et al. 2010; Larentis et al. 2013; von Netzer et al. 2013; Grösbacher et al. 2016; Müller 2017; Bouhajja et al. 2017; Pilloni et al. 2019).

Sediment from two sampling campaigns at the Flingern site was used in the course of this thesis. Development of *nod* assays used sediment collected by drilling in September 2013. Samples were stored in a cool box in the field until they could be transferred to -20 °C upon return to the lab.

Microcosms were set up using sediment collected from the Flingern aquifer in November 2016. Sediment was retrieved by drill coring and preserved by being scooped into bottles full of sterilized anoxic deionized water. Bottles were sealed and stored at ambient temperature for transport, then stored at 12°C until use. Sediment used for microcosms was from a depth of ~7 m below ground surface, corresponding to the lower fringe of the contaminant plume (Pilloni et al. 2019).

2.2 Nucleic acid extraction

2.2.1 DNA

DNA from Siklós sediment and Flingern sediment collected in 2013 was extracted using a phenol-chloroform method as previously described (Pilloni et al. 2012). Between 0.9 – 1 g of sediment were added to 2 ml screw-top tubes containing a 1:1 mix of 0.1 and 0.7 mm diameter zirconia-silica beads (Roth, Karlsruhe, Germany), suspended in 650 µL PTN buffer (120 mM Na_2HPO_4, 125 mM Tris, 0.25 mM NaCl, pH 8) and incubated with 10 µL proteinase K (10 mg/ml) and 40 µL lysozyme (50 mg/ml) for 15 min at 37 °C. Then 150 µL 20 % (v/w) sodium dodecyl sulfate was added, followed by 15 min of incubation at 65 °C with 500 rpm shaking, then addition of 100 µL 25:24:1 phenol:chloroform:isoamyl alcohol (PCI; Carl Roth,

Karlsruhe, Germany). Tubes were bead-beaten for 45 s at 6.5 m/s in a FastPrep-24 (MP Biomedicals, Solon, Ohio, United States), then centrifuged at 20800 g for 5 min at 4 °C. Aqueous supernatants were transferred to Phase Lock Gel tubes (Quantabio, Beverly, MA, USA) and extracted with one volume PCI, centrifuged, transferred, and likewise extracted with 1 vol of 24:1 chloroform-isoamyl alcohol. Purified supernatant was mixed with 2 vol of 30 % polyethylene glycol, incubated at 4 °C overnight, and precipitated by centrifugation at 20,000 g for 30 min at 20°C, followed by a wash with ice-cold fresh 70 % ethanol. DNA was resuspended in 30 µL EB (QIAgen, Hilden, Germany). Extract quality was checked by Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, USA).

2.2.2 RNA

RNA from Siklós microcosms, used for the Siklós RNA-SIP project, was extracted by Dr. András Táncsics and Milán Farkas using the RNA PowerSoil® Total RNA Isolation Kit (QIAgen). Extracted RNA was transferred to Helmholtz München on dry ice.

RNA from the Flingern sediment-based microcosms was extracted using a bead-beating phenol-chloroform protocol modified from Schmitt et al. (1990). Approximately 0.4 g of sediment was added to sterile screw-cap tubes containing a 1:1 mix of 0.1 mm and 0.7 mm zirconia/silica beads. Each tube received 750 µL $NaPO_4$ buffer (pH 5.6), 250 µL TNS buffer (500 mM Tris pH 8, 100 mM NaCl, 10% w/v sodium dodecyl sulfate), and 250 µL acidic phenol/chloroform/isoamyl alcohol (PCI; 25:24:1; pH 4-5 from Carl Roth). Tubes were shaken on a TissueLyzer II (QIAgen) at 30 Hz for 1 min, then centrifuged at 20800 g for 5 min at 4 °C. Aqueous supernatants were transferred to Phase Lock Gel tubes (Quantabio) and extracted with one volume acidic PCI (Carl Roth) and spun again. Aqueous supernatants were again transferred to Phase Lock tubes, extracted with 1 vol chloroform:isoamyl alcohol (24:1; Carl-Roth), and spun. Nucleic acids were precipitated by mixing with 1 vol isopropanol (Carl-Roth) and 0.5 vol 7.5 M ammonium acetate, incubating at room temperature for 10 min, and centrifuging for ≥ 30 min at 20800 g and 4 °C. Supernatant was discarded and the pellet washed with 500 µL cold 70 % ethanol, dried, and resuspended in RNAse-free water. DNA was removed by treatment with TURBO DNAse (Thermo Fisher Scientific) according to manufacturer instruction, and the remaining RNA purified with the Zymo RNA Clean & Concentrator-5 kit (Zymo, Irvine, USA) and eluted twice into a total of 21 µL RNAse-free water. Triplicate extractions were performed and were pooled at the purification step to produce highly concentrated RNA extracts.

2.3 Putative nitric oxide dismutase assays

2.3.1 Primer design

Primers for the putative *nod* gene were designed iteratively (Fig. 10). The first round of primers (Table 3) were produced based on *nod* sequences from "*Candidatus* Methylomirabilis oxyfera" DAMO_2434 (CBE69496) and DAMO_2437 (CBE69502), HdN1 HDN1F_02620 (CBL43845), and the assembled metagenome of a NC10-AAA enrichment culture (KX364454 and KX364455), which were aligned with selected *qnor* and *cnorB* sequences in MEGA6 using the ClustalW algorithm. The second round (Table 3) were optimized based on results from round 1 (Fig. 10).

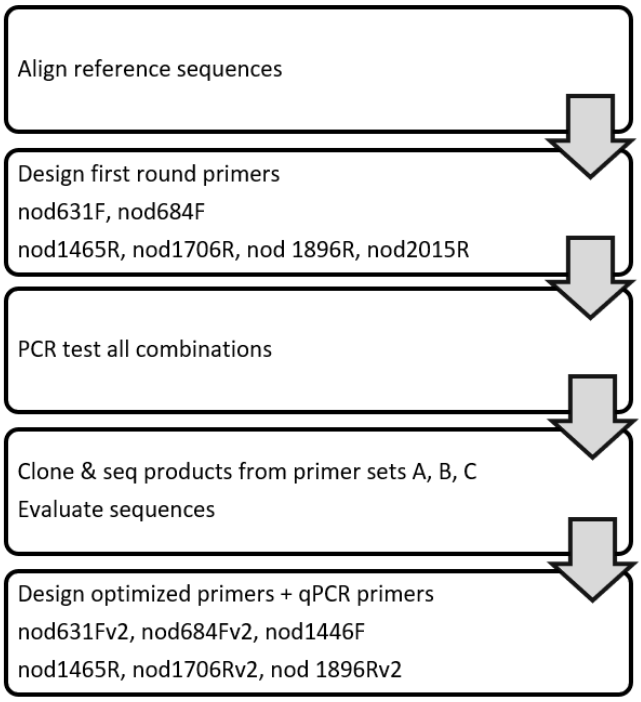

Figure 10: Schematic of the *nod* primer design and testing process.

2.3.2 PCR

Extracted DNA was diluted by 10 or 100 times for use as PCR template to overcome inhibition from co-extracted materials. Each PCR reaction mix had a volume of 25 µL containing nuclease-free water, 1.5 mM $MgCL_2$, 0.1 mM dNTP mix, 1x PCR buffer, 0.5 U Taq polymerase (Fermentas GmbH, Basel, Switzerland), 5 ug bovine serum albumin (Roche Diagnostics GmbH, Basel, Switzerland), 0.5 uM of each primer (Table 3), and 1 µL template DNA. Gradient PCR between 50 and 65 °C, depending on primer properties, was used to determine optimal annealing temperatures. PCR conditions were: a 3 min initial dissociation

at 96 °C, followed by 35 cycles of amplification (95 °C for 45 s, 52-62 °C for 60 s, 72 °C for 90 s), then a final 72 °C extension for 5 min. PCR products were visualized by gel electrophoresis with the Generuler 1 kb DNA ladder (Thermo Fisher Scientific).

2.3.3 Cloning

PCR products were purified with PCRextract spin columns (5Prime, Hamburg, Germany) and quantified by Nanodrop Spectrophotometer (Thermo Fisher Scientific). Products were ligated into pGEM-T Easy vectors (Promega, Madison, USA) according to manufacturer protocol with overnight incubation at 4 °C. Ligation products were chemically transformed into JM109 High Efficiency Competent Cells (Promega) according to manufacturer instructions and plated onto LB-ampicillin plates treated with IPTG and X-gal for blue/white screening. Plates were incubated overnight at 37 °C. White colonies were selected and re-grown on LB plates marked with grids. Plasmids were extracted from clonal colonies by boiling prep and used as template for PCR with M13 primers (M13f: GTAAAACGACGCCAGT; M13r: CAGGAAACAGCTATGAC). Each 50 µL PCR reaction contained 39.3 µL water, 1.5 mM $MgCL_2$, 0.1 mM dNTP mix, 1x PCR buffer, 1 U Taq polymerase (Fermentas GmbH), 0.5 µM of each primer, and 1 µL boiling prep product. PCR conditions were: a 3 min initial dissociation at 94 °C, followed by 24 cycles of amplification (94 °C for 30 s, 55 °C for 30 s, 70 °C for 90 s), and a final 70 °C extension step for 5 min. Products were visualized by gel electrophoresis and cleaned with the PCRextract kit as above, quantified by Nanodrop Spectrophotometer (Thermo Fisher Scientific), and used as template for cycle sequencing with the BigDye v3.1 kit, according to manufacturer instruction. Products of the BigDye reaction were purified with the DyeEx 2.0 spin kit (QIAgen) and sequenced on an ABI 3730 DNA Analyzer (Applied Biosystems, Carlsbad, CA).

2.3.4 Data analysis

Sequences from clone libraries were analyzed with Seqman software (DNAstar, Madison, USA). High-quality sequences were translated to amino acids using MEGA6 and aligned with selected cNOR and qNOR reference sequences obtained from the NCBI-nr database using ClustalW on default settings. A neighbour-joining tree was constructed from the amino acid alignment in MEGA6. Tree topology robustness was checked by bootstrap analysis (1000 replicates).

Table 3: *nod*-targeted primers, and their position on the putative *nod* gene of "Candidatus Methylomirabilis oxyfera" nod DAMO_2437. From Zhu, et al. (2017).

Primer	Sequence (5' – 3')	Positions	Design round
nod631F	TTCTTCTGGGGHGGYTGGG	631–649	1
nod684F	CTAYACHCACAACTGGCC	684–701	1
nod1465R	CGAAGAACAGGAACAGMACCATG	1465–1443	1
nod1706R	GGCTTGGCRATCCAGTAGAAG	1706–1686	1
nod1896R	GATGTTCCAGAAGTTRACGSC	1896–1876	1
nod2015R	ATGTTACCYTTKACACCGAAC	2015–1995	1
nod684Fv2	STAYACHCAYAACTGGCC	684–701	2
nod1706Rv2	GGCTTSGCRATCCAGTAGAAG	1706–1686	2
nod1896Rv2	GATRTTCCAGAAGTTRACGSC	1896–1876	2
nod1446F	GGTGBYBTTCCTGTTCTTYRG	1446–1466	2

2.4 Microcosm design

2.4.1 Microoxic microcosms for Siklós sediment

Each microcosm consisted of a 100 ml serum bottle with 5 g (wet weight) sediment material and 50 ml artificial groundwater medium (Winderl et al. 2010) with a N_2/CO_2 (80/20, v/v) headspace. This amount of sediment was chosen to prevent redox stratification during incubation. Five µl of unlabelled (^{12}C; Carl Roth) or fully labelled ($^{13}C_7$; Sigma-Aldrich, Munich, Germany) toluene were added by injection with a glass syringe after headspace exchange. Abiotic control microcosms were set up in the same manner, but autoclaved 3 x before addition of non-labelled toluene. These microcosms controlled for possible abiotic toluene loss during incubation. Microcosms were incubated at 16 °C with 145 rpm rotary shaking. Dissolved oxygen levels were measured with planar oxygen sensor spots and a Fibox 3 Oxygen Meter (PreSens Precision Sensing GmbH, Regensburg, Germany) and maintained between 0 and 0.5 mg/L by daily replenishment of air injected through a 0.2 µm filter.

RNA was extracted from duplicate labelled and unlabelled microcosm sediments after seven days of incubation, as described above, stored at -20 °C, and transported to Munich on dry ice.

2.4.2 Microcosms from Flingern sediment under various redox conditions

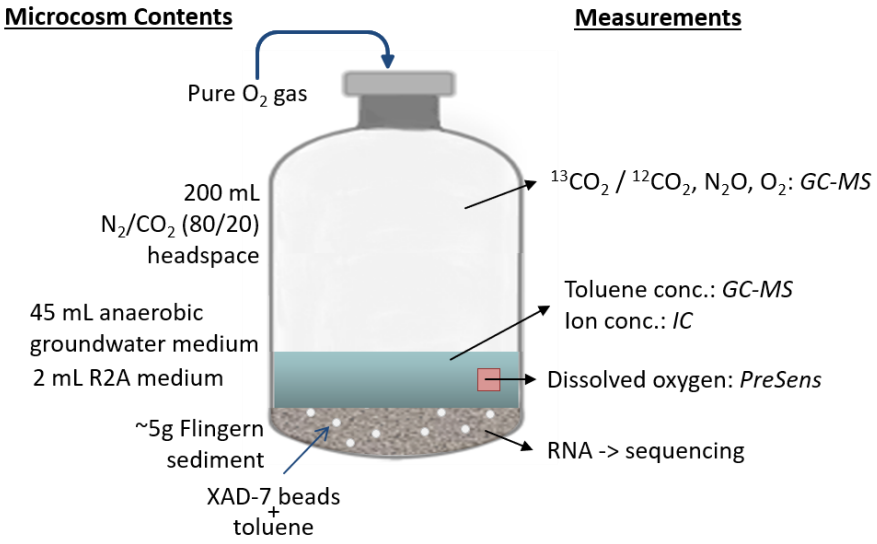

Figure 11: Overview of Flingern sediment microcosm setup and measurements

Microcosms were set up in 250 ml serum bottles. Amberlite XAD7 beads (Sigma-Aldrich) were prepared by washing 3x with absolute ethanol (Sigma-Aldrich) then ≥ 3x with ultrapure water and dried at 90 °C for 2 days. Each bottle received 0.3 g of XAD7 beads and was sterilized by autoclaving. A small square of oxygen sensor foil SF-RPSu4 (PreSens Precision Sensing GmbH) was glued to the inside of each bottle and allowed to set for >12 hours, then 45 ml of 10x diluted autoclaved artificial groundwater medium (Widdel 1980) was added. A glass syringe rinsed with acetone was used to add ~5 µL of either fully ^{13}C-labelled ($^{13}C_7$; Sigma-Aldrich) or unlabelled (^{12}C; Carl Roth) toluene to each bottle, resulting in a total toluene concentration of ~1 mM before adsorption to XAD7 beads. Bottles were sealed with rubber butyl stoppers to prevent toluene loss and incubated at 16 °C in the dark with gentle shaking (50 rpm) for 3 days to allow toluene adsorption to beads, after which measurable toluene concentrations in the media dropped to 55 ± 13 µM. Each microcosm bottle then received ~5 g sediment from the November 2016 campaign to Flingern, as described above, and 2 ml R2A medium (DSMZ recipe 830; sterilized by autoclaving). R2A is a medium frequently used to study bacteria in water systems; in this case, it provided carbon and nutrients to facilitate growth of the entire sediment community. Nitrate-amended microoxic and anoxic bottles each received 360 µL 1M filter-sterilized $NaNO_3$ solution for a concentration of ~7.6 mM. Sediment, R2A medium, and nitrate solution were added in an anoxic tent with an N_2/H_2 (90/10) atmosphere. Bottles were removed from the anoxic hood and the headspace exchanged with an

80/20 N_2/CO_2 gas mix filtered through a sterile 0.22 µm syringe filter (Merk-Millipore, Darmstadt, Germany) until dissolved oxygen was below the limit of detection of PreSens measurements. Bottles were incubated at 16 °C with 50 rpm rotary shaking.

Pure gaseous oxygen filtered through a sterile 0.22 µm syringe filter (Merk-Millipore) was provided to oxic and microoxic microcosms using a syringe pushed through the butyl stopper, with an additional needle in the stopper to allow outflow of gas and prevent overpressure. Oxic bottles initially received 40 ml and microoxic bottles 1.5 ml pure oxygen. Between 1.5 and 2.5 ml of oxygen were added to microoxic bottles periodically (Table 4) to maintain low, fluctuating oxygen availability, based on sensor spot oxygen measurements. No replenishment was performed for fully oxic bottles.

Table 4: Volume in millilitres of pure oxygen gas injected into headspaces. Millimole calculations assume 1 atm pressure and 16 °C temperature.

	Days of incubation						
	0	2	4	6	7	Total ml	Millimoles
Oxic	40	0	0	0	0	40	1.7
Microoxic+NO_3^-	1.5	2	2	2	0	7.5	0.32
Microoxic	1.5	2	2.5	2	2	10	0.42

2.5 Microcosm monitoring

2.5.1 Dissolved oxygen

Dissolved oxygen was measured at least once daily using a Fibox 3 Oxygen Meter (PreSens Precision Sensing GmbH) with immersed sensor spots cut from Oxygen Sensor Foil SF-RPSu4 that were affixed inside the serum bottle with latex glue.

2.5.2 Toluene concentration

Toluene concentration in the Siklós microoxic microcosms was measured at initial setup and after 3 days and seven days of incubation. Headspace analysis was performed by Dr. András Táncsics and Milán Farkas using an ISQ Single Quadropole GC-MS (Thermo Fischer) with an SLB-5ms fused silica capillary column (Sigma-Aldrich) on the following temperature program: 40 °C for 3 min, ramp 20 °C/min to 190◦C, hold for 1 min. The mass spectrometer (MS) was operated at 250 °C in full scan mode.

Toluene concentration in the Flingern sediment-based microcosms was measured by transferring 1 ml of medium by syringe to a crimp-cap headspace vial and spiking with 20 µL of 20 mM ethylbenzene internal standard dissolved in acetone. Vials were sealed by crimping and agitated at 70 °C for 17 minutes prior to injection of 250 µL headspace to the GC-MS. Ratios of toluene:ethylbenzene in the headspace vial were measured on a Finnigan TraceGC ultra with a DB-5 column (Agilent, Santa Clara, USA) coupled to a Finnigan Trace DSQ MS (Thermo Fisher Scientific) with the following temperature program: hold at 40 °C for 1 min, ramp 15 °C/min to 200 °C, ramp 25 °C/min to 300 °C and hold for 1.33 min. The DSQ was operated at 250 °C in selected ion mode and measured m/z 51, 77, 78, 91, 92, 97, 98, 105, and 106. The carrier gas was helium with a 1 ml/min flow rate. Concentrations were calculated using a standard curve produced from toluene stock and the same ethylbenzene internal standard.

2.5.3 Headspace gases

For the Flingern sediment based microcosms, headspace gases were monitored by GC-MS (Agilent 7890A GC coupled to a 5975C inert XL EI/CI MSD; Agilent). Forty µL of headspace was withdrawn from the microcosm with an airtight Hamilton syringe and injected. Gases were separated on a GS-Q column (Agilent) at 80 °C for 9 minutes with helium carrier gas at a flow rate of 1.6 ml/min. Masses 28 Da (N_2), 32 Da (O_2), 44 Da ($^{12}CO_2$ and N_2O, with retention times of 8.2 and 8.5 min, respectively), and 45 Da ($^{13}CO_2$) were measured by the MSD in selected ion mode (SIM). $^{13}CO_2$ atom percent was calculated as $^{13}CO_2$ / ($^{13}CO_2$ + $^{12}CO_2$). Headspace oxygen was calculated as (Oxygen signal) / (total signal all measured ions). Headspace N_2O was calculated in the same way. Quantification of N_2 produced from denitrification was not possible because of the N_2/CO_2 headspace.

2.5.4 Dissolved ions

Nitrate and nitrite were monitored qualitatively during the course of the experiment using Meckoquant test strips (Merck, Darmstadt, Germany). Liquid samples were taken by syringe concurrent with sampling for toluene and frozen until further processing. Samples were diluted with ultrapure H_2O and filtered through 0.22 µm syringe filters, then analyzed with two ICS 1100 Ion Chromatographers equipped with a 4mm diameter, 250mm length Ionpac AS4A analytical column for anions and a 4mm diameter, 250mm length Ionpac CS12A analytical column for cations (Thermo Fisher Scientific).

2.6 Density-based RNA separation

2.6.1 Ultracentrifugation

Isopycnic density gradient centrifugation was carried out according to Lueders (2015). Gradients were prepared with 5 ml of 2 g/ml cesium trifluroacetate (CsTFA; GE Healthcare, Munich, Germany), 185 µL formamide, and 1 ml of gradient buffer (0.1 M Tris-HCl, 0.1 M KCl, 1 mM EDTA) containing up to 1 µg of RNA. Gradients were loaded into 5.1 ml Quick Seal centrifuge tubes (Beckman Coulter, Indianapolis, USA), sealed, and spun on a VT1 65.2 rotor in an Optima XE90 ultracentrifuge (Beckman Coulter) at 125 000 g for ~65 hours. Fractions from each tube were collected according to Lueders (2015). For the Siklós RNA-SIP project, seven fractions were collected and density was measured by weighing known volumes. For the Flingern sediment microcosms project, twelve fractions were collected and density was measured by refractometer (Krüss, Hamburg, Germany). RNA was recovered by precipitation with 1 vol. of isopropanol, washed with cold 70 % EtOH, and resuspended in either 25 µL EB (QIAgen) for Siklós microcosms, or 20 µL RNAse-free water for Flingern microcosms.

2.6.2 RNA Quantification

rRNA standards for RT-qPCR were generated using the Riboprobe *in vitro* transcription (IVT) kit (Promega) from plasmids containing a cloned 16S rRNA gene from *Methylobacterium sp*.. 16S rRNA inserts, and surrounding M13 primer sequences, were amplified by PCR with M13 primers as described in section 2.3.3, visualized with gel electrophoresis, and purified with the PCRextract kit (5Prime). Each IVT reaction contained 20 µL of 5x Riboprobe reaction buffer, 10 µL x 100 mM DTT, 2.5 µL x 40 U/µL RNAsin, 5 µL each 2.5 mM ATP, CTP, GTP, and UTP, 2 µL x 20 U/µL T7 polymerase, ~2 µg amplicon template, and enough nuclease-free water to bring the total volume to 100 µL. Reactions were incubated at 37 °C for two hours, followed by addition of 2 µL of RQ1 DNAse (Promega) and an additional 15 min incubation. DNA-free RNA was initially purified by phenol-chloroform extraction as described in section 2.2.2; the Zymo RNA Clean and Concentrator-5 kit (Zymo) was used during later rounds of standard generation to reduce RNA loss. RNA standards were quantified with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific) on an Mx3000p (Agilent) and single-use aliquots were stored at -80 °C.

RNA recovered from fractions in the Siklós project was quantified with RT-qPCR using the AccessQuick RT-PCR kit (Promega) and Ba519f and Ba907r universal 16S primers on an Mx3000p machine (Agilent) as described by Glaubitz et al. (2009). Each 40 µL RT-qPCR reaction contained: 20 µL 2 x AccessQuick Master Mix, 0.4 µL 20 mg/ml bovine serum

albumin, 0.2 µL 1/500 diluted SYBRgreen (Thermo Fisher Scientific), 0.6 µL 1/500 diluted ROX, 10 µM each primer, and 2 µL of standard or unknown RNA. The temperature program was: 45 °C reverse transcription for 30 min, initial denaturation of 95 °C for 3 min, then 35 amplification cycles (30s at 95 °C, 30s at 52 °C, 30s at 68 °C), with a final extension step at 68 °C for 5 min. After the run, a dissociation curve from 55 °C to 95 °C was recorded to discriminate products from unspecific amplification products. RNA recovered from fractions in the Flingern sediment microcosms project was quantified either by RT-qPCR, as above, or with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific) on an Mx3000p (Agilent).

2.7 Post-gradient RNA treatment and sequencing

2.7.1 Microoxic microcosms from Siklós sediment

Based on quantitative rRNA profiles across gradients, selected RNA fractions from duplicate microcosms and gradients were pooled to obtain sufficient RNA amounts for linear amplification and sequencing (Fig. 17). These pools were designated unlabelled light (^{12}C-light, density 1.778-1.779 g/ml), unlabelled heavy (^{12}C-heavy, 1.795-1.798 g/ml), labelled light (^{13}C-light, 1.776-1.794 g/ml) and labelled heavy (^{13}C-heavy, 1.824-1.828 g/ml). RNA amounts present in the heavy fractions of the ^{12}C gradients were too low for transcriptome sequencing. Instead, the "heaviest" fractions from which sufficient RNA could be obtained from the ^{12}C gradients were chosen, in order to control for possible shifts caused by physical characteristics of RNA such as GC content (Fig. 17). Pooled RNA was concentrated via precipitation with isopropanol, washed with 70 % EtOH, and resuspended in 15 µL EB (QIAgen).

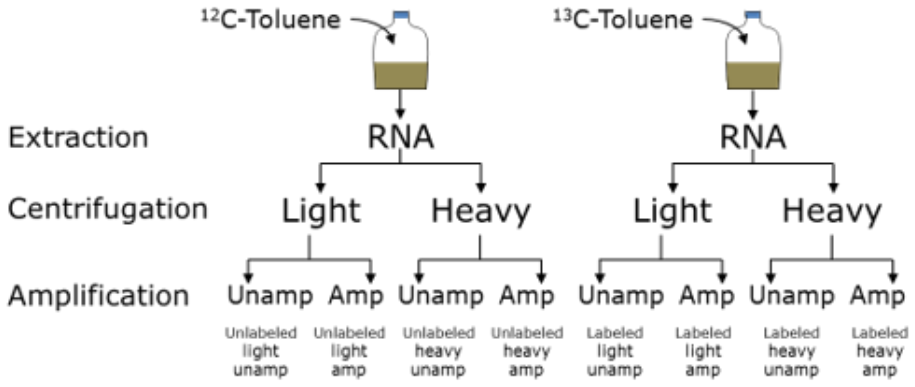

Figure 12: Conceptual overview of the Siklós total-RNA-SIP experimental design.

2.7.1.1 Linear amplification

Approximately 100 ng of RNA from each of the four pools (^{12}C-light, ^{12}C-heavy, ^{13}C-light, ^{13}C-heavy) was subjected to linear amplification, while remaining RNA was set aside and sequenced without amplification, resulting in eight RNA-sequencing libraries (Fig. 12). Linear amplification was performed using the MessageAmp II Bacteria Kit (Thermo Fisher Scientific) according to manufacturer instructions. Briefly, a poly-A tail is added to the 3' end of each RNA strand in a reaction with E. coli Poly(A) Polymerase (PAP), providing an attachment site for an oligo(dT) primer bearing a T7 promoter. ArrayScript reverse transcriptase (Thermo Fisher Scientific) produces first-strand cDNA, which is converted to double-stranded DNA (dsDNA) by DNA Polymerase, with RNase H simultaneously degrading the original RNA template. This dsDNA is the template for in vitro transcription by T7 RNA Polymerase, resulting in hundreds to thousands of antisense RNA copies of each RNA strand in the original sample.

2.7.1.2 Sequencing

RNA was quantified and purity-checked using a NanoDrop ND-1000 spectral photometer (VWR, Ismaning, Germany) and a 2100 Bioanalyzer (Agilent) with the RNA 6000 Nano LabChip. Library preparation with input templates ≥20 ng of total RNA was performed with the TruSeq® Stranded mRNA HT technology High Sample kit according to manufacturer protocols. The generated libraries were quality checked with the 2100 Bioanalyzer (Agilent) using a DNA 1000 LabChip, and quantified with the Qubit®dsDNA HS Assay Kit on a Qubit machine (Thermo Fisher Scientific). A 1 % PhiX v3 control library spike-in (Illumina, San Diego, USA) was added to the library before cluster generation. Library preparation, cluster generation, and sequencing were performed on the NextSeq500 sequencing system (Illumina) using a high-output paired-end 150 cycle (2 x 150 PE) by IMGM Laboratories GmbH (Munich, Germany). All sequencing raw data generated in this study as well as sequences representing important 16S OTUs and functional transcripts have been submitted to the NCBI sequence read archive (SRA) via the Gene Expression Omnibus under the accession number GSE119644.

2.7.2 Varying redox microcosms from Flingern sediment

2.7.2.1 cDNA synthesis

RNA recovered from SIP fractions was immediately converted to single-stranded complementary DNA (ss-cDNA) after quantification using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) with random hexamer primers. Single-stranded products were converted to double-stranded cDNA (ds-cDNA) according to instructions from

the SuperScript Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific). Each ds-cDNA reaction contained the 20 µL ss-cDNA template, 91 µL RNAse-free water, 30µL of 5x 2nd strand buffer, 0.2 mM dNTPs, 10 U *E. coli* DNA ligase, 40 U *E. coli* DNA polymerase, and 2 U *E. coli* RNAse H (all from Thermo Fisher Scientific). Reactions were incubated in a thermocycler for 2 hours at 16 °C, after which 10 U T4 DNA polymerase (Thermo Fisher Scientific) were added and the incubation continued for 5 additional minutes at 16 °C. ds-cDNA was purified with the Minelute PCR Purification Kit (QIAgen) and eluted in 12 µL of provided elution buffer (10 mM Tris·Cl, pH 8.5). Integrity and quantity of ds-cDNA were checked with the High Sensitivity DNA kit on an Agilent 2100 Bioanalyzer (Agilent).

2.7.2.2 Sequencing

Sequencing of cDNA was performed by CeGaT GmbH (Tuebingen, Germany). Library preparation was performed using the Nextera XT kit (Illumina) according to manufacturer protocols. Cluster generation and sequencing were performed on a NovaSeq 6000 to produce paired-end 150 bp reads. Demultiplexing was performed with Illumina bcl2fastq (v2.19) and adapters were trimmed with Skewer v.0.2.2 (Jiang et al. 2014).

2.8 Data handling

2.8.1 Bioinformatics

An in-house bioinformatics pipeline was developed for transcriptome-SIP data handling with the help of Dr. Gisle Vestergaard (a member of the Comparative Microbiome Analysis unit at Helmholtz München at the time of these experiments; now at the Technical University of Denmark). This pipeline was used for data handling from both microcosm projects. Commands and a guide to the bioinformatics process can be found in Appendix 2. Removal of sequencing adapters and merging of paired-end sequences was performed with Adapterremoval (Lindgreen 2012) with default settings except: trim ambiguous bases (N) at 5'/3' termini , trim bases at 5'/3' termini with quality scores ≤ 15 and minimum read length of 50. Deconseq (Schmieder & Edwards 2011) was used to remove PhiX sequences from Siklós samples; spike-in PhiX sequences were removed from data from Flingern samples by the sequencing company before data transmission. RNA was sorted into 16S, 23S, and non-rRNA using SortmeRNA v2 (Kopylova et al. 2012) with comparisons against subsets of the Silva v123 database (Quast et al. 2013). The Biopieces framework (Hansen et al. 2008) was used for command line processes. Non-rRNA tags were compared against the NCBI-nr (downloaded February 2016), COG

(Galperin et al. 2014), and KEGG v58.1 (Kanehisa et al. 2015) databases using DIAMOND (Buchfink et al. 2015). Non-rRNA analysis was performed using MEGAN 5 (Huson et al. 2007), which uses a lowest common ancestor (LCA) algorithm to assign reads to phylogenetic or functional groups, which places transcripts at the most specific level possible (Huson et al. 2007).

16S OTU assignment was performed using QIIME v1.9.1 (Caporaso et al. 2010). Preprocessing of reads identified as 16S by SortmeRNA was done with the split_libraries_fastq.py (quality=2, p=0.01), truncate_reverse_primer.py, identify_chimeric_seqs.py (method: usearch61), and filter_fasta.py scripts. Minimum Phred score was 2 (Bokulich et al. 2013) . OTU calling was performed with a subsampled open-reference based method using RDP classifier (Wang et al. 2007) using the Silva database release 123 (Quast et al. 2013). For this, the scripts pick_open_reference_otus.py (97 % similarity) and filter_otus_from_otu_table.py were used. Further analysis, ordination plotting, and statistical tests were performed in R (R Core Team 2013) using the phyloseq package (McMurdie & Holmes 2013). Significance tests were performed in Microsoft Excel 2010 using the Student's t-test (paired, two-tailed).

2.8.2 Nod transcript detection

Using MEGAN 5, all non-rRNA reads assigned to "*Candidatus* Methylomirabilis oxyfera" (between 0 and 62 reads per library) were extracted and matched to their functional annotation according to comparison with the KEGG v58.1 database. All transcripts annotated as canonical nitric oxide reductases belonging to "*Ca.* Methylomirabilis oxyfera" were compared to the NCBI-nr database via nucleotide BLAST.

Multiple methods were used to search for transcripts of putative nitric oxide dismutase in the sequences from the Flingern microcosms. Three hidden markov models (HMM) were created from alignments of environmental putative *nod* sequences (Zhu et al. 2017) using the hmmbuild command from HMMER v3.1 (Eddy 1998). All reads sorted as non-RNA were translated using Prodigal v2.6.2 (Hyatt et al. 2010) and searched with these HMMs using the hmmsearch command, and the top hits ($e < 10^{-10}$) were compared to the NCBI-nr database by blastp (Altschul et al. 1990). Having found misannotation of putative *nod* as *qNOR* in the microoxic Siklós sediment-based microcosm experiment, all reads identified as NO-reductase by comparison to KEGG were exported, clustered at 97 % identity using pick_de_novo_otus.py from Qiime v1.9.1 (Caporaso et al. 2010), and representative reads from

each “OTU” were compared to the NCBI-nr database. Finally, all translated non-rRNA was searched for short motifs in the catalytic site that distinguish putative nitric oxide dismutase from NO-reductases (Zhu et al. 2017).

2.8.3 Transcript enrichment calculation

Enrichment factors in the Siklós microcosm experiment were calculated as previously done in rRNA-SIP (Kramer et al. 2016; Zhang & Lueders 2017) using enrichment factor equation 1 (EF_1):

$$EF_1 = \frac{^{13}C_{heavy}}{^{13}C_{light}} - \frac{^{12}C_{heavy}}{^{12}C_{light}}$$

where “$^{13}C_{heavy}$” and “$^{13}C_{light}$” were the relative abundance of a specific rRNA taxon or mRNA transcript – assigned to the same KEGG/COG category – in heavy and light fractions of the ^{13}C-toluene microcosms, respectively, and “$^{12}C_{heavy}$” and “$^{12}C_{light}$” the same for respective ^{12}C-control microcosms. Positive values indicate isotopic labelling of a transcript; negative enrichment indicates more frequent detection in the light fraction of ^{13}C-RNA.

The near-absence of RNA in heavy fractions from ^{12}C microcosms in the Flingern sediment-based microcosm experiment prevented sequencing, necessitating development of an updated enrichment factor calculation (EF_2).

$$EF_2 = \frac{^{13}C_{heavy}}{^{13}C_{light}} - \frac{(^{13}C_{heavy} + {}^{13}C_{light})/2}{^{12}C_{light}}$$

The development of this equation is explained in Appendix 1. To account for differences in labelling due to inconsistent toluene degradation between microcosms at time of sacrifice, EFs were adjusted by dividing by headspace $^{13}CO_2$ atom % (Fig. 26).

2.8.4 Transcript abundance calculation

Transcript expression levels, beyond just isotopic enrichment, were of interest in the various redox conditions investigated in the Flingern sediment microcosm project. A modified Transcripts per Kilobase Million (TPM) method (Wagner et al. 2012) was used to compare expression of genes of interest across samples. Counts were divided by gene length from a representative organism (Table 5), then divided by the total sample reads and multiplied by one

billion to produce intelligible numbers ("transcripts per billion", TPB). This calculation was only performed on specific genes of interest; when considering groups of genes in categories or pathways, copy numbers were relativized by MEGAN 5 and gene lengths were not considered.

Table 5: Functional genes of interest. Genes in the KEGG denitrification module are highlighted in grey.

Gene	KEGG ID	Length (amino acids)	Length (nucleotides)	Representative organism	Uniprot entry
ccoN	K00404	474	1422	Pseudomonas stutzeri	D9IA43
ctaD	K02274	616	1848	Bacillus sp. (strain PS3)	P98005
cydA	K00425	522	1566	Escherichia coli (strain K12)	P0ABJ9
cyoB	K02298	663	1989	Escherichia coli (strain K12)	P0ABI8
dmpN	K16242	517	1551	Pseudomonas sp. (strain CF600)	P19732
tmoA	K15760	500	1500	Pseudomonas mendocina	Q00456
dmpB	K00446	307	921	Pseudomonas sp. (strain CF600)	P17262
napA	K02567	828	2484	Escherichia coli (strain K12)	P33937
narG	K00370	1247	3741	Escherichia coli (strain K12)	P09152
narH	K00371	512	1536	Escherichia coli (strain K12)	P11349
narI	K00374	225	675	Escherichia coli (strain K12)	P11350
napB	K02568	149	447	Escherichia coli (strain K12)	Q53177
nirK	K00368	376	1128	Alcaligenes faecalis	P38501
nirS	K15864	568	1704	Pseudomonas aeruginosa	P24474
norB	K04561	474	1422	Pseudomonas stutzeri	P98008
norC	K02305	146	438	Pseudomonas stutzeri	Q52527
nosZ	K00376	638	1914	Pseudomonas stutzeri	P19573

2.9 Linear amplification testing

A direct test of the effect of linear amplification on the proportion of mRNA in the total RNA was conducted in a SIP-independent manner using total RNA from *P. aeruginosa* and RT-qPCR of housekeeping genes. *P. aeruginosa* was chosen as a model organism because of the availability of well-established primers for several housekeeping genes. *P. aeruginosa* was grown in LB broth at 37°C with shaking for 70 hours. Cells were pelleted at 1700 x g for 3 minutes and RNA was extracted from the the cell mass using a bead-beating phenol-chloroform protocol modified from (Schmitt et al. 1990), as described above. Nucleic acids were precipitated with 2 volumes of 30% polyethylene glycol, 1.6 mM NaCl solution rather than isopropyl alcohol. DNA was removed from extracts by treatment with RQ1 DNAse (Promega).

Housekeeping genes *rpoD, rpoS,* and *gyrA* were amplified for cloning by RT-PCR with PCR primers from Savli et al. (2003) or Qin et al. (2003) using the AccessQuick RT-PCR kit (Promega). The temperature program was: 45 °C reverse transcription for 45 min, initial denaturation of 95 °C for 5 min, then 35 amplification cycles (15s at 95 °C, 10s at 60 °C, 15s at 72 °C), with a final extension step at 72 °C for 7 min. Products were cloned using the pGEM-Easy plasmid kit (Promega) and *E. coli* JM109 competent cells as described in the *nod* assays section above and sequenced via Sanger sequencing to ensure correct primer targeting. Clones were also used to generate RT-qPCR standards via in vitro transcription, as described in the RNA Quantification section above. Fresh RNA was then extracted and either amplified with the MessageAmp II Bacteria kit (Ambion) or kept untreated (Fig. 13).

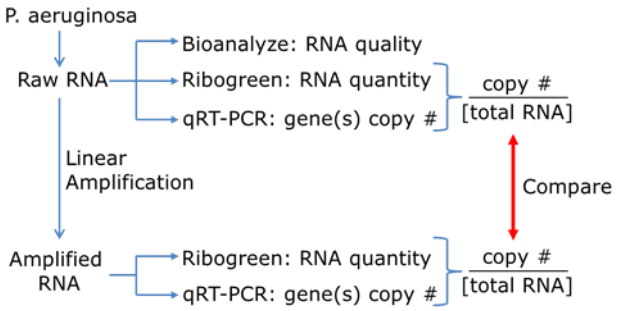

Figure 13: Workflow of sequencing-free linear amplification experiment.

Total RNA in amplified or unamplified subsamples was quantified with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific) on an Mx3000p (Agilent). Copy numbers of housekeeping genes were quantified by RT-qPCR on the Mx3000p using the AccessQuick kit and SYBR green dye, with a final primer concentration of 100pM. Each 40 µL RT-qPCR reaction contained: 20 µL 2 x AccessQuick Master Mix, 0.4 µL 20 mg/ml bovine serum albumin, 0.2 µL 1/500 diluted SYBRgreen (Thermo Fisher Scientific), 0.6 µL 1/500 diluted ROX, 100 pM each primer, and 2 µL of RT-qPCR standard or freshly extracted RNA. The temperature program was: 45 °C reverse transcription for 25 min, initial denaturation of 95 °C for 5 min, then 45 amplification cycles (15s at 95 °C, 10s at 60 °C, 15s at 72 °C), with a dissociation curve from 55 °C to 95 °C recorded after the run. Ratios of housekeeping gene copy number to total RNA were calculated for comparison.

2.10 23S Depletion Testing

A 23S rRNA depletion method was developed based on the probe-based rRNA depletion method developed by Stewart et al. (2010). In the first round of testing, DNA and RNA were simultaneously extracted from two pre-experiment microcosm sediments (4A and 4B) from Work Package 2 using a bead-beating phenol-chloroform protocol modified from (Schmitt et al. 1990) as described above. Extractions were performed in duplicate and extracts were not pooled. In the later round, E. coli JM109 were grown in LB broth overnight, and nucleic acids were extracted from cell pellets as above. To produce DNA-free RNA extracts, a subset of each extract was treated with Turbo DNAse (Thermo Fisher Scientific) and purified with the Zymo RNA Clean and Concentrator-5 kit (Zymo) according to manufacturer instructions. Purified RNA was quantified with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific) on an Mx3000p (Agilent). The remaining subset was left untreated to provide a DNA template for probe generation.

2.10.1 PCR and cloning

The 23S gene was amplified from DNA templates by PCR using forward primer Eub23S_189F (GAASTGAAACATCTHAGTA) and reverse primer Eub23S_2490R_T7 (GCCAGTGAATTGTAATACGACTCACTATAGGGCGACATCGAGGTGCCAAAC) (Stewart et al. 2010). The reverse primer contains a T7 promotor region (underlined). Each PCR reaction mix had a volume of 50 µL containing nuclease-free water, 1.5 mM $MgCL_2$, 0.1 mM dNTP mix, 1x PCR buffer, 2.5 U Taq polymerase (Fermentas GmbH), 5 ug bovine serum albumin (Roche Diagnostics GmbH), 2.5 uM of each primer, and 2 µL of ¼ diluted template DNA. PCR conditions were: hot start at 95 °C, a 2 min initial dissociation at 95 °C, followed by 35 cycles of amplification (95 °C for 20 s, 39 °C for 20 s, 72 °C for 2 min), then a final 72 °C extension for 3 min. PCR products were visualized by gel electrophoresis with the Generuler 1 kb DNA ladder (Thermo Fisher Scientific). Products were purified with the QIAgen MinElute PCR purification kit (QIAgen) according to manufacturer instruction.

Products of the PCR reactions targeting 23S rRNA were cloned and sequenced via Sanger sequencing, as described in section 2.3.3, to ensure that no spurious PCR products were formed as this could result in improperly targeted probes.

2.10.2 Probe generation

In vitro transcription (IVT) was performed using the Riboprobe In Vitro Transcription kit (Promega) using 23S PCR products as template. Each IVT reaction contained 4 µL x 5x Riboprobe reaction buffer, 4 µL x 100 mM DTT, 0.5 µL x 40 U/µL RNAsin, 0.5 mM each ATP and GTP, 0.375 mM each CTP and UTP, 0.125 mM biotin-labelled CTP and biotin-labelled UTP (Sigma-Aldrich), 1 µL x 20 U/µL T7 polymerase, ~400 ng template, and enough nuclease-free water to bring the total volume to 20 µL. All components except biotin-labelled nucleotides were from the Riboprobe kit (Promega). Reactions were incubated at 37 °C for 1.5 hours, followed by addition of 2 U Turbo DNAse (Thermo Fisher Scientific) and further incubation at 37 °C for 30 min to remove DNA. Biotin-labelled RNA probes were purified with the MegaClear Transcription kit (Thermo Fisher Scientific) according to manufacturer instruction and eluted in 50 µL nuclease-free water. Probes were quality-checked with a Nanodrop Spectrophotometer (Thermo Fisher Scientific) and quantified with with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher Scientific) on an Mx3000p (Agilent).

2.10.3 Hybridization and bead-binding

Hybridization reactions were performed according to Stewart et al. (2010), with a probe-to-RNA ratio of 2:1. Each reaction contained: 1 µL x 2500 U RNAsin (Thermo Fisher Scientific), 2.5 µL x 20X RNAse-free Sodium chloride-citrate (SCC) buffer, 10 µL 100 % formamide, RNA template and IVT-generated, biotin-labelled probes in a 2:1 ratio, and RNAse-free water to bring the volume to a total of 50 µL. Reactions were performed in 0.2 ml PCR tubes in an Eppendorf thermocycler with the following program: 70 °C for 5 min, ramp down 5 °C/min to 25 °C, hold at 25 °C.

Streptavidin-coated magnetic beads (New England Biosciences, Ipswich, MA, USA) were rinsed prior to bead binding. The stock of tubes was vortexed to mix, then 100 µL per hybridization reaction was transferred to a single 1.5 ml RNAse-free Eppendorf tube. Beads were bound to a Dynamag magnetic separation rack (Thermo Fisher Scientific) and the supernatant discarded, then beads were resuspended in an equal volume of 0.1 N NaOH to deactivate RNAses. Binding, discarding, and resuspension was then repeated three times using 1X sodium chloride-citrate (SCC) buffer. On the third time, 100 µL aliquots were made for each hybridization reaction, the beads were bound, and supernatant (SCC) was discarded.

In the initial test, hybridization reactions were diluted with 40 µL of 1X SCC and 10 µL of 100% formamide to a total volume of 100 µL, maintaining a 20 % concentration of formamide

as outlined by Stewart et al. (2010). The 100 µL hybridization reaction was added to the dried beads and incubated at room temperature for 10 minutes, with occasional gentle mixing. Beads, which should be bound to probes and 23S rRNA fragments, were captured by magnet as above and the 23S-depleted supernatant transferred to fresh tube. Beads were washed in an additional 100 µL 1x SCC, which was added to the tube of previous supernatant. RNA was purified from the supernatant with the RNeasy MinElute Cleanup kit (QIAgen) and eluted in 12 µL nuclease-free water. Purified RNA was run on a 5200 Fragment Analyzer System (Agilent) using the DNF-471-33 (Total RNA) kit to visualize and quantify 16S and 23S rRNA peaks.

2.10.4 Additional probe removal attempts

Since probe-23S complexes remained in the purified RNA (Fig. 25B), a series of additional removal attempts were performed. A 10 µL aliquot of purified RNA was added to 10 µL of rinsed beads, incubated for 10 minutes as above, then pipetted off. This supernatant was run on the Fragment analyzer (Fig. 25C). The supernatant was then brought to 100 µL total volume with 20% formamide (10 µL supernatant, 20 µL formamide, and 70 µL 1X SCC) and cleaned with beads, then purified with the RNeasy Minelute Cleanup kit (QIAgen) as in the protocol, and the resulting double-purified supernatant run on the Fragment Analyzer.

The protocol from beginning until hybridization (sections 2.10.1-2.10.3) was then repeated with nucleic acids from E. coli JM109 in order to test the effect of varied formamide concentration on probe-bead binding. Binding reactions were set up with 10 %, 15 %, and 20 % formamide. Finally, the Zymo RNA Clean & Concentrator-5 kit (Zymo) was tried as an alternative method for probe-23S complex removal, since it had previously removed probes when mistakenly used for purification after probe generation.

3. Results

3.1 Detection assays for putative nitric oxide dismustase genes

3.1.1 Primer design

Both "*Ca.* Methylomirabilis oxyfera" and gammaproteobacterium HdN1 have shown evidence of aerobic hydrocarbon metabolism under anoxic conditions (Ettwig et al. 2010; Zedelius et al. 2011), and both contain a gene variant of canonical quinol-dependent nitric oxide reductase (*qnor*) that is hypothesized to act as the oxygen-generating nitric oxide dismutase (Ettwig et al. 2012). I designed primers targeting putative nitric oxide dismutase (*nod*) genes in order to explore the potential for NO-driven oxygenesis in hydrocarbon contaminated aquifers and other environments. Initial primer designs were based on the few available reference *nod* sequences from these two organisms, and sequences generated in this first round informed a second round of primer design.

In the first round, two forward (nod631F and nod684F) and four reverse (nod1465R, nod1706R, nod1896R and nod2015R) primers were designed based on available reference sequences. Five pairs of forward/reverse primers were selected based on compatible annealing temperatures (Fig. 33). When tested on template DNA from the "*Candidatus* Methylomirabilis oxyfera" enrichment culture, PCR with three of these pairs produced amplicons of the correct size, suggesting that primers were correctly targeted. Gradient PCR using template DNA from sediment sampled from Siklós wells ST2 and ST5 (Fig. 9) and from the Flingern aquifer also produced the expected size bands with no spurious products (Fig. 14). Amplicons from the aquifer sediments with selected primer pairs (nod631F/nod1465R, nod684F/nod1465R, and nod684F/nod2015R) and PCR conditions were cloned and sequenced. These sequences were aligned against reference *nod* from "*Ca.* Methylomirabilis oxyfera" and the gammaproteobacterium HdN1, and the new alignment informed the design of a second round of primers (Table 3).

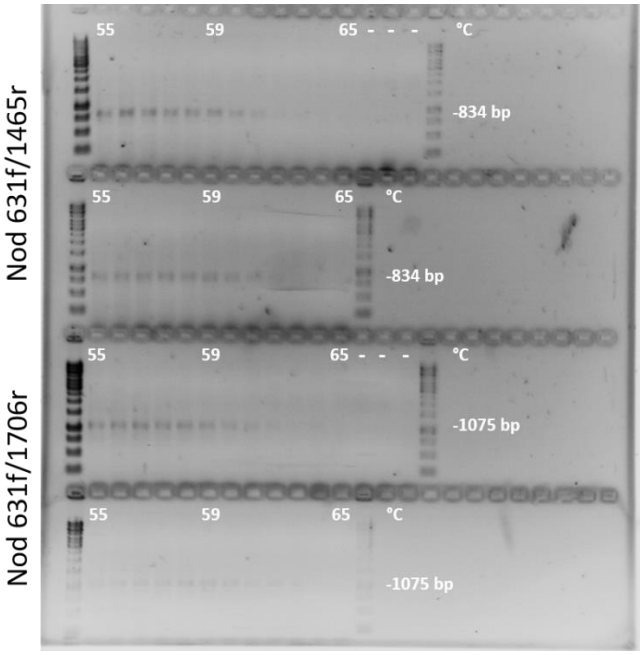

Figure 14: Example of results from gradient PCR to determine optimal annealing temperatures for two sets of *nod* primers. Template DNA was extracted from sediments from the Siklós aquifer. Minimum, median, and maxium annealing temperatures are shown at the top of each row. Minus signs denote PCR negative control reactions. Expected product sizes are labelled on the right. Marker: Generuler 1 kb DNA ladder (Thermo Fisher Scientific).

3.1.2 nod from environmental samples

The sequences obtained from cloning amplicons made in the first round of the design process were included in a first screening for the presence of putative *nod* genes in a range of environmental samples. *Nod* genes were successfully found in sediment from both BTEX-contaminated aquifers investigated: the Flingern aquifer, located under a former gasworks site in Düsseldorf, Germany, and the Siklós aquifer, located under a gas station with leaking storage tanks in the Siklós region of Hungary. Five clone sequences from Flingern samples and 38 clone sequences from Siklós were used for further analysis, including construction of a phylogenetic tree and the second round of primer design. Primers from the second round were used by my colleague Dr. Baoli Zhu to investigate material from wastewater treatment plants (WWTP) and related bioreactors (Table 6).

Nucleotide sequences generated by Sanger sequencing of clone libraries were then translated to amino acids, aligned, and used to construct a neighbor-joining phylogenetic tree with selected nitric oxide reductase (NOR) reference sequences and the original putative *nod* reference sequences (Fig. 15). Almost all sequences obtained from the designed primers, along with putative *nod* from "*Ca.* Methylomirabilis oxyfera" and HdN1, formed a distinct cluster

from the reference *nor* sequences. One of the Siklós sequences (Sik2DC09) and a WWTP sequence (WWTP-Kempten clone R2-7) clustered with a set of Nor-related sequences from microbes in the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) superphylum. The five Flingern sequences were nearly identical to each other, with ≥ 99 % similarity, but were not very similar to "*Ca.* Methylomirabilis oxyfera" or HdN1 reference sequences, with approximately 70 % similarity to each on the nucleotide level. Clones from the Siklós aquifer, on the other hand, split into two groups with high similarity (between 84 and 99 %) to either "*Ca.* Methylomirabilis oxyfera" DAMO_2437 (18 clones) or DAMO_2434 *nod* (17 clones).

Table 6: Environmental samples investigated for *nod* genes. Grey boxes indicate samples investigated in the course of this thesis work. Adapted from Zhu et al. (2017).

Description	Designation	Redox/Process	Primer pair	No. of clones
BTEX-contaminated Siklós aquifer sediment	Siklós	Microoxic	nod631F/nod1465R	16
			nod631F/nod1706R	13
			nod684F/nod2015R	9
BTEX-contaminated Flingern aquifer, 8m depth	Flingern	Anoxic	nod631F/nod1465R	5
Swing-redox deammonification reactor	Swing-redox	Nitration and annamox	nod684Fv2/nod1706Rv2	13
Two-stage sequencing batch deammonification reactor cascade	2-stage-nitritation	Nitritation	nod684Fv2/nod1706Rv2	16
	2-stage-AMX	Annamox	nod684Fv2/nod1706Rv2	14
CANDO reactor system	CANDO-nitritation	Nitritation	nod684Fv2/nod1706Rv2	7
	CANDO-N2O	Nitrous denitritation to N2O	nod684Fv2/nod1706Rv2	5
Suspension deammonification reactor	Suspension	Nitritation and anammox	nod684Fv2/nod1706Rv2	16
Wastewater treatment plant, Garching	WWTP-TUM	Nitritation and anammox	nod684Fv2/nod1706Rv2	25
Wastewater treatment plant, Kempten	WWTP-Kempten	Nitritation and anammox	nod684Fv2/nod1706Rv2	6
			nod684Fv2/nod1896Rv2	4

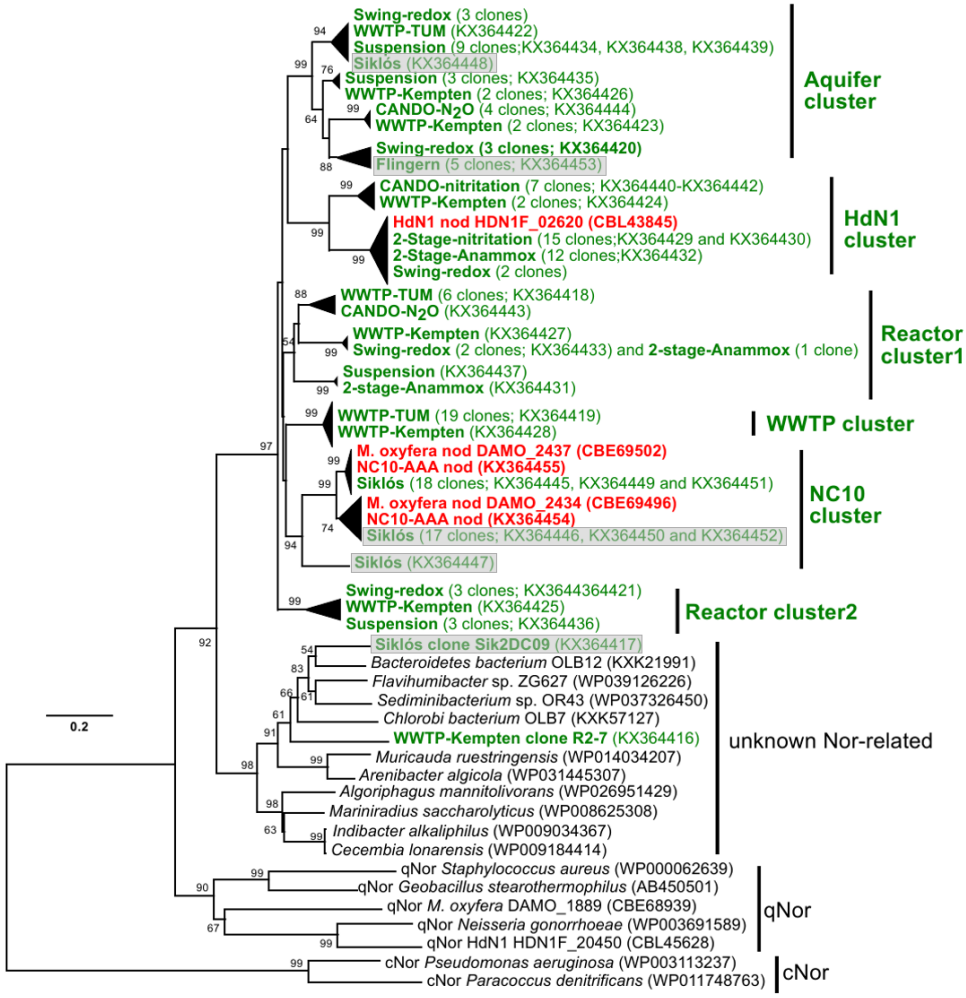

Figure 15: Neighbour-joining phyogenetic tree of putative *nod* from environmental (green) and reference (red) sequences, along with reference *qnor* and *cnorB* sequences (black). Siklós and Flingern sequences are highlighted with grey boxes. Accession numbers are shown in parentheses. The six sub-clusters of *nod* identified in this study are labelled. Values at nodes indicate bootstrap support (1000 replicates). The scale bar indicates 20% amino acid divergence. From Zhu et al. (2017).

With the exception of Siklós Sik2DC09 and WWTP-Kempten clone R2-7, all the *nod* clones and reference sequences formed a well-defined phylogenetic cluster, distinct from known cNor and qNor. Within the *nod* grouping, we classified six subclusters and named them for the organism or habitat in which they were first discovered.

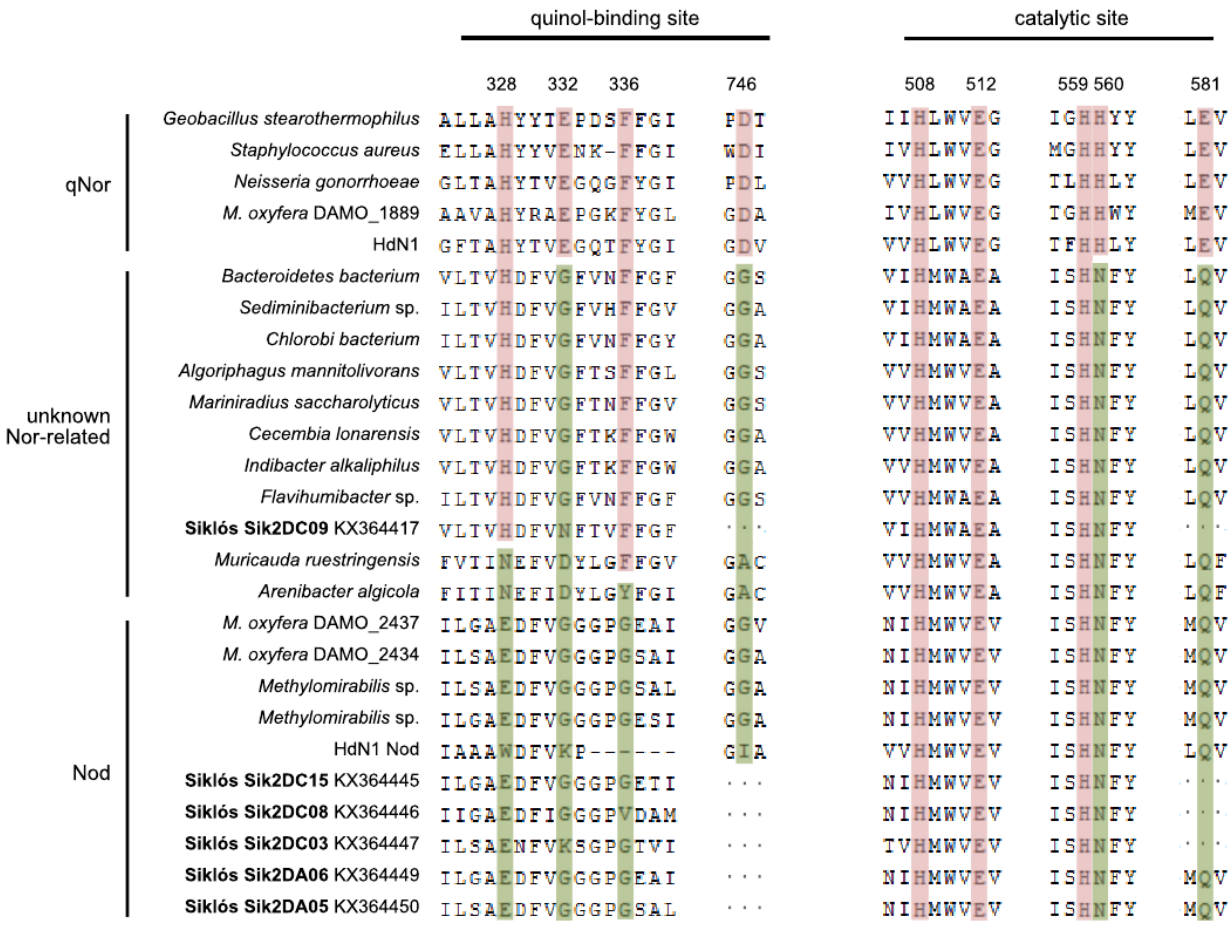

Figure 16: Multiple sequence alignment of selected environmental putative *nod* (in bold) and qNor reference sequences surrounding the quinol-binding and catalytic sites of qNor. The conserved residues for quinol-binding and catalytic functioning in qNor are highlighted in red, and substitutions at these sites in putative Nod and putative unknown Nor-related are shown in green. Positions numbers are based on the "*Ca.* methylmirabilis oxyfera*"* DAMO_2437 *nod* sequence. Accession numbers are the same as in Fig. 15. From Zhu et al. (2017).

The alignment of translated environmental sequences with reference amino acid sequences revealed that putative *nod* differs from canonical qNOR (quinol-dependent nitric oxide reductase) by several amino acid substitutions in positions essential for qNOR functioning (Fig. 16). Specifically, the reference putative *nod* sequences used for primer design as well as all but one of the environmental sequences (in which this region was covered by primers) differ from canonical qNORs at His328, Glu332, and Phe336 (Geobacillus stearothermophilus qNor numbering) in the quinol binding region. These substitutions likely render the putative nod protein incapable of binding quinol, so they cannot function as NO reductases. Additionally, nonheme metal-coordinating His560 was substituted for an asparagine in Nod, possibly altering the active site configuration. Glu581, which may act as a terminal proton donor in qNOR, was substituted for glutamine. Unfortunately, only reverse primer nod2015R extended

far enough to cover these regions, so there is no data on substitutions in many of the clones investigated. The site Asp746 was not covered by any primer set, but reference Nod and unknown Nor-related sequences show glycine substitution. Clone Siklós Sik2DC09 and the unknown Nor-related reference sequences with which it clustered (Fig. 15) had the Nod-type substitutions at Glu332, Asp746, His560, and Glu581, but matched reference qNor amino acids at His328 and Phe336 (Fig. 16).

The success of these assays show that *nod* is present in hydrocarbon contaminated groundwater as well as other aquatic systems that are vital to maintaining water quality and controlling nitrogen cycling. *Nod* may be widespread in a variety of other environmental or constructed systems; the PCR primers presented here offer a means to find out.

3.2 Development of a total-RNA-SIP pipeline and application to microaerobic toluene degraders

This microcosm experiment provided a proof-of-concept for total-RNA-SIP. Microcosms set up with sediment from the BTEX-contaminated, microoxic Siklós aquifer were incubated under microoxic conditions with ^{13}C-labelled toluene. Instead of using PCR amplification to target only SSU rRNA, or depleting rRNA to focus on mRNA, I sequenced total RNA recovered from selected gradient fractions after isopycnic centrifugation.

3.2.1 Total RNA centrifugation

Total RNA extracted after 7 d of microcosm incubation and active degradation of toluene was subjected to isopycnic centrifugation. A clear shift in rRNA distribution over density gradients was observed between microcosms provided with ^{13}C-labelled vs. unlabelled toluene (Fig. 17).

As was previously observed for DNA from these same microcosms (Táncsics et al. 2018), the small amounts of RNA remaining in light fractions of ^{13}C gradients suggested a community highly enriched in active toluene-degraders. Likewise, very little rRNA was present from unlabelled microcosms at a density >1.81 g/ml. This complicated the selection of appropriate control fractions for sequencing, as this study, unlike most previous RNA-SIP studies, could not rely on post-gradient PCR amplification. Thus, I selected total RNA from the “heavier” and “lighter” slopes of unlabelled RNA (<1.78; >1.79 g/ml, respectively) for sequencing of control fractions (Fig. 17), in order to identify possible intrinsic density shifts of transcript pools caused by physical properties of RNA such as GC content.

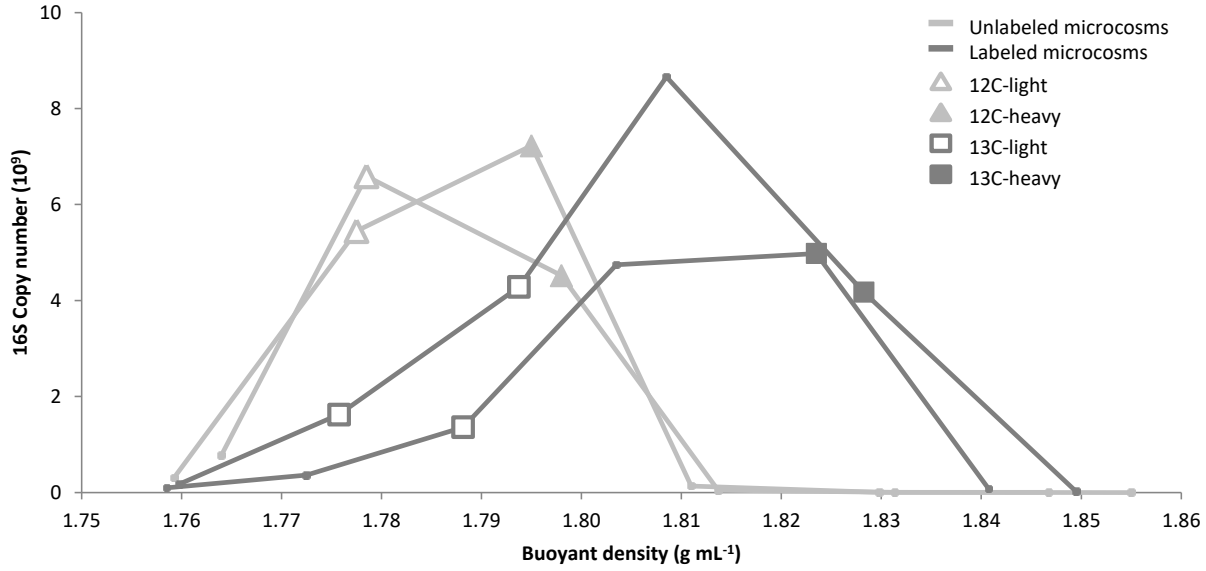

Figure 17: Quantitative distribution of rRNA in SIP gradients after isopycnic centrifugation, as measured by RT-qPCR of bacterial SSU rRNA across buoyant density fractions. RNA fractions pooled from duplicate gradients for direct sequencing are given with respective symbols.

After pooling from replicate centrifugation runs, as indicated in Fig. 17, subsets were made of each of the four libraries (^{12}C-light, ^{12}C-heavy, ^{13}C-light, and ^{13}C-heavy). One subset was sequenced without further processing, while the other was subjected to linear amplification. This resulted in a total of eight sequencing libraries (Fig. 12).

3.2.2 Composition of total RNA-seq libraries

A total of 279 million paired-end reads were obtained by RNA-seq across the eight libraries (4 unamplified, 4 which underwent linear amplification) with an average of ~30 million per library. After QC, a total of about 1 million reads could be identified as mRNA using the KEGG or COG databases (Table 7). Total SSU (small subunit) and LSU (large subunit) rRNA reads accounted for 35 ± 3 and 64 ± 3 % of sequencing reads, respectively. The ratio of non-rRNA reads was slightly but significantly increased in linearly amplified sequencing libraries ($p = 0.023$).

Table 7: Number of RNA sequencing reads and proportion of reads identified as rRNA and non-rRNA in RNA-seq libraries of density-resolved total RNA in the Siklós total-RNA-SIP project.

	Starting reads (10^6)	SSU rRNA reads	LSU rRNA	non-rRNA	KEGG mRNA tags (10^3)	COG mRNA tags (10^3)
^{12}C Light	42.5	31 %	68 %	0.6 %	101	110
^{12}C Heavy	30.3	37 %	62 %	0.8 %	93	101
^{13}C Light	37.2	37 %	60 %	1.8 %	150	171
^{13}C Heavy	46.5	33 %	66 %	0.6 %	102	136
^{12}C Light Amp	29.9	40 %	58 %	1.4 %	151	162
^{12}C Heavy Amp	53.6	31 %	67 %	1.0 %	192	211
^{13}C Light Amp	15.7	34 %	62 %	2.2 %	88	101
^{13}C Heavy Amp	23.4	34 %	65 %	1.2 %	122	131

GC-content of SSU rRNA reads appeared unchanged across gradient densities, and was also not significantly affected by linear amplification (Fig. 18). However, a slight increase in average GC-content of non-rRNA reads (including mRNA) was observed with buoyant density (~58% GC at ~1.78 g/ml; 59.5% GC at ~1.83 g/ml, respectively). Moreover, linear amplification significantly reduced the GC-content of non-rRNA reads by an average of 2.1 ± 0.9 % compared to unamplified libraries ($p = 0.025$).

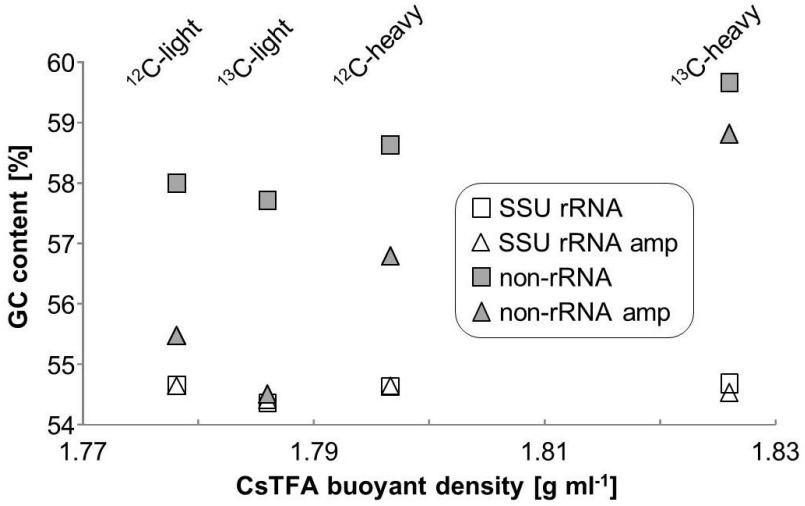

Figure 18: GC content of unamplified and amplified RNA-seq libraries of RNA-SIP buoyant density fractions.

3.2.3 Reads enriched in ^{13}C-labelled rRNA

Nearly all SSU rRNA sequences in the libraries were bacterial, with a maximum of 0.09 % of reads in libraries matching to eukaryotic rRNA and no reads matching archaeal rRNA. The community was overwhelmingly comprised of Proteobacteria (Table A1), averaging 98 % of SSU reads, with only small contributions of Bacteroidetes (avg. 1.8 %) and other phyla (avg. 0.6 %). Most of the transcripts within the Proteobacteria were affiliated to the Betaproteobacteria; of these the majority belonged to members of the *Rhodocyclaceae* (Fig. 19). Five of the six most abundant genera in total rRNA (unidentified *Rhodocyclaceae, Dechloromonas, Pseudomonas, Quatrionicoccus, Zoogloea*, and *Azonexus*) belonged to the *Rhodocyclaceae*[1]. All genera within the *Rhodocyclaceae* were strongly enriched in ^{13}C-labelled RNA, with the exception of reads affiliated to *Azoarcus* spp. (Fig. 19); no labelling appeared in members outside of this family. About 21 % of the total SSU rRNA reads were members of the *Rhodocyclaceae* whose more precise affiliation remained unclear. rRNA of *Pseudomonas* spp. comprised an average 10 % of the communities and appeared unlabelled.

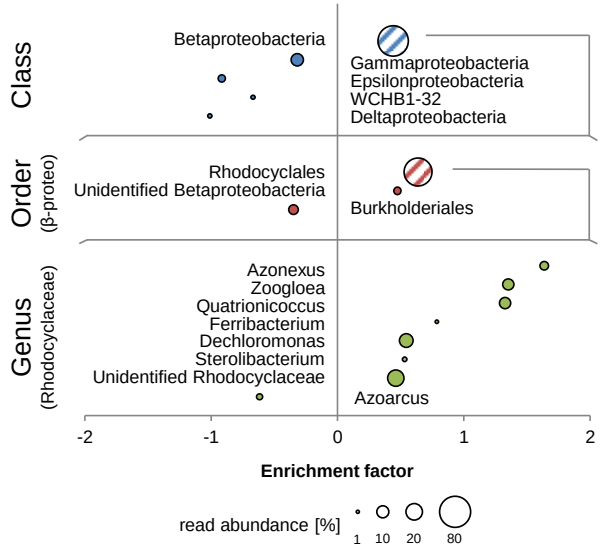

Figure 19: ^{13}C-enriched bacterial taxa identified by SIP in unamplified RNA of toluene-degrading microcosms. Rankings of SSU rRNA enrichment factors (EFs) are resolved at the class-level (A), for orders within Betaproteobacteria (B), and for genera within Rhodocyclaceae (C). EFs are shown in combination with relative read abundances averaged across all RNA-seq libraries. Groups shown are those with > 1% average read abundance at each level.

[1] It should be noted that the taxonomic status of the Rhodocyclaceae family underwent a major reclassification since the release of the Silva v123 database used in this study. The previously monofamiliar order Rhodocyclales now contains three families: Rhodocyclaceae (genera *Rhodocyclus*, *Azospira*, and *Propionivibrio*), Azonexaceae (genera *Azonexus*, *Dechloromonas*, *Ferribacterium*, and *Quatrionicoccus*), and Zoogloeaeceae (genera *Zoogloea*, *Thauera*, *Uliginosbacterium*, and *Azoarcus*) (Boden et al., 2017). As shown in Fig. 19, members of the newly-definied Azonexaceae were the most highly enriched in this experiment.

Considerable differences in taxon-level abundances were found between unamplified and amplified libraries (Table A1). For example, *Dechloromonas* spp. averaged at 11.4 % of the unamplified community but at 18.3 % after amplification, while *Zoogloea* rRNA was detected at 13.4 % in the unamplified community and at 6 % in the amplified.

3.2.4 Functional transcripts enriched in ^{13}C-labelled RNA

The number of non-rRNA reads identified as mRNA was between 88 and 192 x 10^3 reads per library (Table 7). The transcript ratios discussed in this section are given as a percentage of those reads, not of total non-rRNA, which also included tRNA and other RNA types. The most abundant category of functional transcripts could not be further identified using KEGG and COG databases (avg. 30 % of reads using KEGG and 26 % of reads using COG). Among those reads which could be further identified, transcripts linked to genes involved in carbohydrate, amino acid, and energy metabolism were most abundant and were all enriched in ^{13}C-RNA, as calculated via enrichment factor equation 1 (EF_1) for the different gene categories, pathways and transcripts (Fig. 20).

Cell motility was the transcript category most highly enriched in ^{13}C-RNA (Fig. 20A). Many transcripts linked to motility-associated genes were highly abundant in mRNA (e.g. *fliC*, coding for flagellin, at ~2.65 % of all transcripts in KEGG annotation), and almost all were enriched in ^{13}C-labelled RNA. Table A2 shows the phylogenetic affiliation of transcripts of the *fliC* gene based on MEGAN 5's lowest common ancestor algorithm. Between 35-40 % of these transcripts could only be assigned at the phylum level, at which nearly all (96-100 %) belonged to the *Proteobacteria*. Some could be specifically assigned at the species level, and 10-20 % of all transcripts linked to *fliC* were assigned to *Azovibrio restrictus*.

The second most clearly labelled transcript category was linked to genes coding for biosynthesis of secondary metabolites (Fig. 20A). The main ^{13}C-enriched transcripts in this group were identified as genes coding for a catalase-peroxidase (K03782), which averaged over 1000 copies per library. Labelling observed for the most abundant COG categories was generally consistent with KEGG results (Fig. 21), but biodegradation pathways are not resolved as a separate category in this database. The COG category "Secondary metabolite biosynthesis" includes similar pathways as the respective KEGG category, but also includes transcripts linked to genes involved in BTEX degradation, such as phenol hydroxylase/phenol-2-monooxygenase, which were highly enriched in ^{13}C-RNA. Signal transduction was the third

most labelled transcript category (Fig. 20A). This is primarily a result of the high labelling of *fliC*, which made up approximately 20 % of transcripts assigned to signal transduction, and is part of the two-component signalling system in addition to its role in motility.

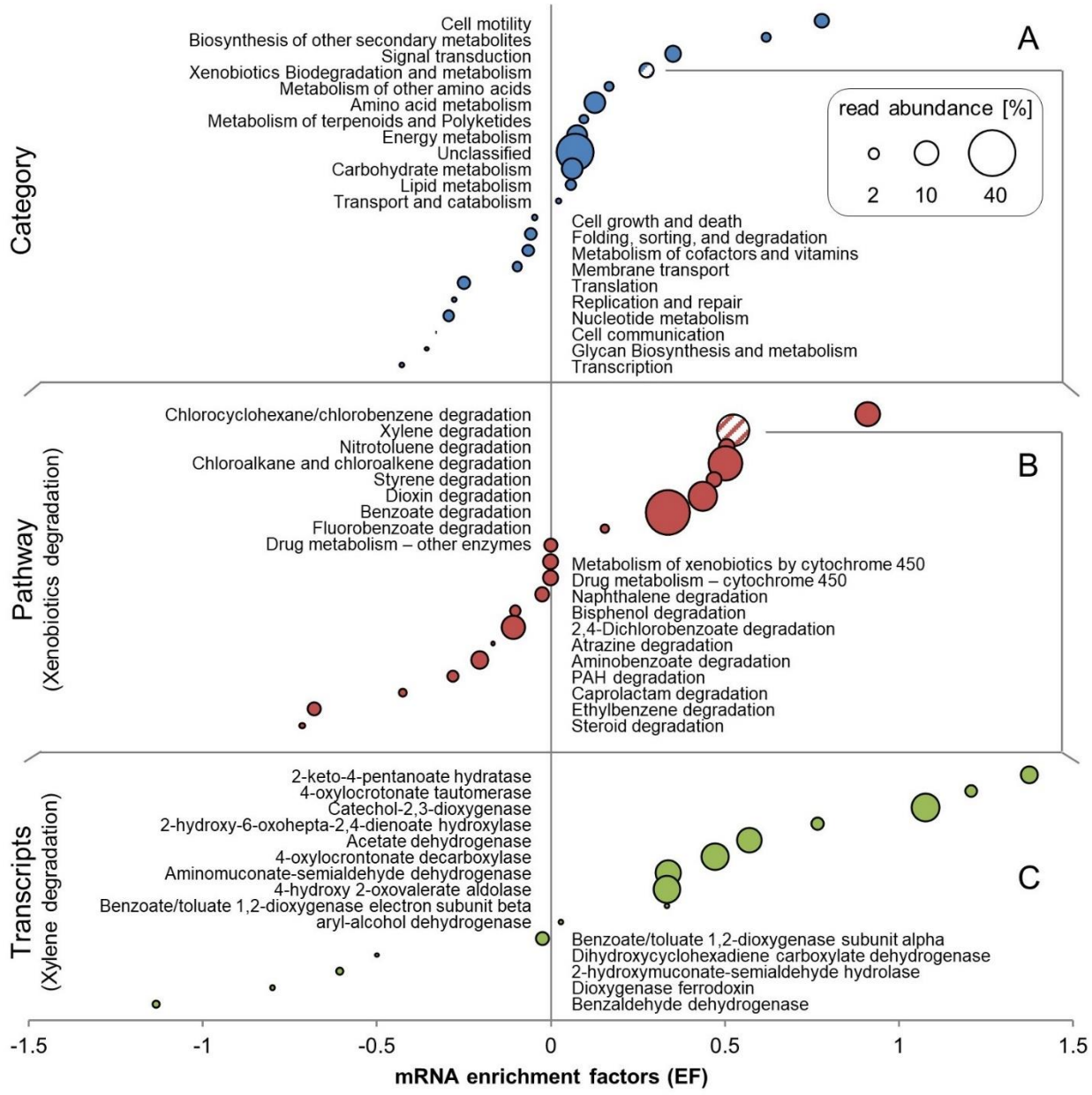

Figure 20: ^{13}C-enriched mRNA transcripts identified by SIP in unamplified RNA of toluene-degrading microcosms. Rankings of mRNA enrichment factors (EFs) are resolved at the level of (A) categories (KEGG level 1), (B) pathways within the xenobiotics biodegradation and metabolism category (KEGG level 2), and (C) individual transcripts within the xylene degradation pathway (KEGG level 4). EFs are shown in combination with relative read abundances averaged across all RNA-seq libraries. Individual transcripts shown are those with ≥ 20 total reads.

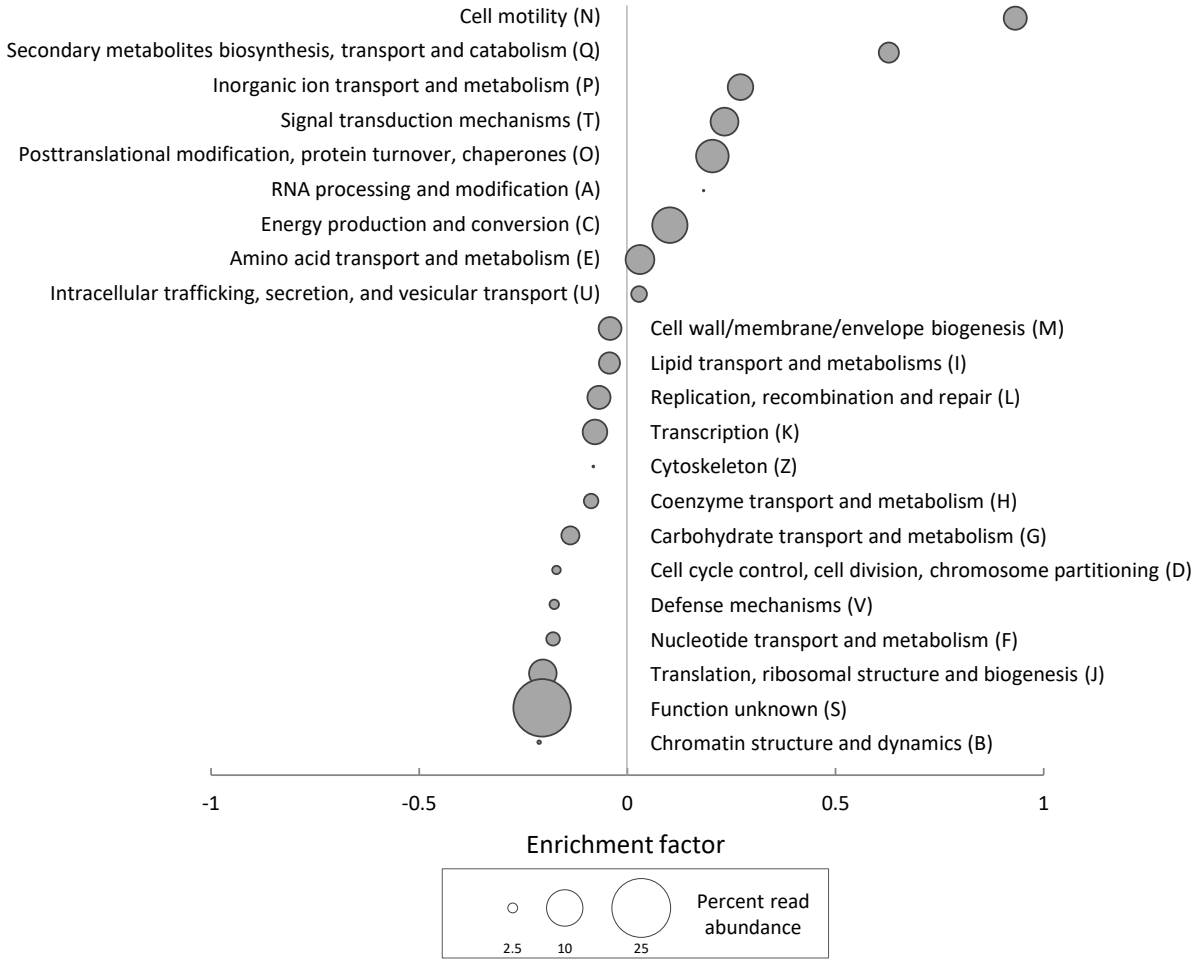

Figure 21: ^{13}C-labelled mRNA transcripts identified by RNA-SIP in toluene-degrading microcosms. Rankings of mRNA enrichment factors (EFs) are resolved at the level of COG categories. EFs are shown in combination with relative read abundances averaged across all eight RNA-seq libraries. Individual transcripts shown are those with ≥ 20 total reads.

Functional transcript annotations using the COG database generally confirm the results obtained via the KEGG database. The most abundant group was "Unknown function" (26 %), equivalent to the "Unknown" category that made up 30 % of KEGG annotations. While the "unclassified" KEGG category showed a slight positive enrichment, "Unknown Function" in COG was unlabelled, reflecting some difference between which transcripts could be annotated by the two databases. For the most part, similar categories between the two databases showed similar patterns of enrichment (Figs. 20-21). Cell motility was the most highly enriched category, followed by secondary metabolite biosynthesis. While COG does not have a category devoted to xenobiotic degradation, many of the transcripts involved in these metabolisms, for example phenol-2-monooxygenase, belong to the secondary metabolite biosynthesis category.

"Energy metabolism" (KEGG) and "Energy production and conversion" (COG) were similarly abundant and labelled, likewise "Amino acid metabolism" and "Amino acid transport and metabolism". Categories related to replication, cell growth/cycling, transcription, and translation were unlabelled in both databases.

3.2.4.1 Xenobiotic degradation transcripts

"Xenobiotics degradation" was the fourth most labelled transcript category in KEGG (Fig. 20B), though about half of the pathways in that category were themselves not enriched in ^{13}C-RNA (Fig. 20C). It should be noted that many of the pathways expressed and enriched in ^{13}C shared enzymes – for example, phenol hydroxylase (K16249), which can be involved in chlorocyclohexane and chlorobenzene degradation, benzoate degradation, toluene degradation, and in the degradation of aromatic compounds in general. Transcripts of the benzoate degradation pathway were the most abundant, though not the most highly labelled (Fig. 20B).

Transcripts of catechol-2,3-dioxygenase (C23O), an enzyme previously associated with central pollutant metabolism at the investigated site (Táncsics et al. 2012; Táncsics et al. 2013; Táncsics et al. 2010), were identified as very abundant and enriched in ^{13}C in KEGG (Fig. 20C). Taxonomic association of these C23O transcripts was done using MEGAN 5 (Table A3). Of the transcripts that could be assigned at the family level, the highest percent belonged to the *Rhodocyclaceae* (avg. 7.8 %), followed closely by *Pseudomonadaceae* (avg. 6.5 %). However, most of these transcripts could not be assigned more specifically than at the phylum level. Transcripts of the catechol-1,2-dioxygenase (K03381), on the other hand, were barely present, appearing only in very low read numbers (~1) per library.

Amongst initial hydrocarbon degradation mechanisms, transcripts of the phenol hydroxylase gene (*dmpL*; K16243) were highly abundant and labelled, with transcript numbers of up to 160 per library. In contrast, transcripts of genes coding for toluene monooxygenase subunits (K15760-K15764, which may code for toluene-2,3, or 4-monooxygenases) were present in low read numbers (~15), with varying amounts of labelling. Transcripts linked to genes coding for benzylsuccinate synthase (*bssA*; K07540) were present in all treatments but not very abundant (~20 reads) and not labelled.

3.2.4.2 Transcripts of respiratory pathways

Transcripts of genes required for prokaryotic oxidative phosphorylation were present in the microoxic microcosms, but the majority were unlabelled. The full operon of *nuoA-N*, coding for Type I NADH dehydrogenase, was expressed, while no transcripts of the *ndh* operon for Type II NADH dehydrogenase were detected. Complete pathways of both denitrification and dissimilatory nitrate reduction to ammonium (DNRA) were transcribed, but the only transcript enriched in ^{13}C-RNA found in these pathways codes for NapA (K02567). I was also interested in whether the putative NO-dismutase genes previously detected at the site would be expressed during microaerobic toluene degradation. For this, I compared all mRNA transcripts assigned to "*Candidatus* Methylomirabilis oxyfera" (between 0 and 62 reads per library) to the NCBI-nr database via nucleotide BLAST. All transcripts annotated as canonical quinol-dependent nitric oxide reductase (*qnor*) belonging to "*Ca.* Methylomirabilis oxyfera" actually matched more closely to the putative nitric oxide dismutase (*nod*). Up to 21 % of transcripts from all organisms that were annotated as nitric oxide reductase subunit B (*norB*; K04561) were putative *nod* transcripts. These *nod* transcripts appeared in read numbers of ten or fewer, so enrichment factor calculations were not robust.

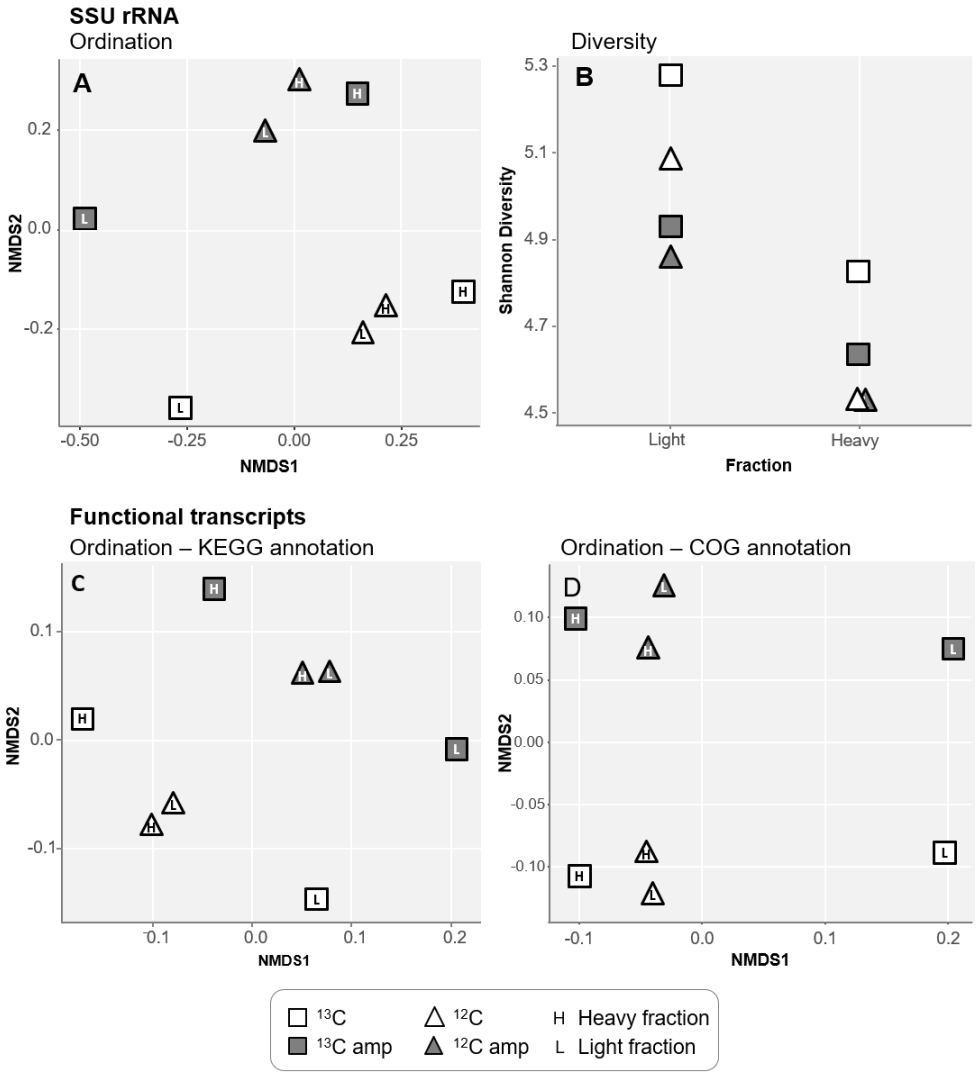

Figure 22: NMDS ordinations of SSU rRNA profiles (A), Shannon diversity of SSU rRNA (B), and NMDS ordinations based on functional profiles as annotated with the KEGG databse (C) or COG database (D).

3.2.5 Strategies to enhance the sensitivity of total-RNA-SIP

One of the challenges of sequencing total RNA to gain information about both identity and expression of functional transcripts is that mRNA constitutes a very small percentage of total RNA, typically less than 5 % (Lodish et al. 2000). This means that deeper sequencing is required for adequate expression data compared to methods that employ rRNA depletion or focus on only specific genes of interest. I explored two techniques to enhance the proportion of mRNA in total RNA sequencing libraries: linear amplification, which has been hypothesized to preferentially amplify mRNA (Cao et al. 2010), and selective depletion of 23S rRNA.

3.2.5.1 Effects of linear amplification

Linear amplification increased RNA amounts by at least 300-fold. NMDS plots of functional transcript pools and of SSU rRNA profiles (Fig. 22) in the 8 treatments were used to delineate the impact of linear amplification on total RNA-seq libraries, and on overall labelling patterns. Within the unamplified or amplified data sets, the separation between profiles was consistent with expectations for a typical SIP experiment. *I.e.*, heavy and light profiles for the control (^{12}C) microcosm were much more similar, while those of the labelled (^{13}C) microcosm were clearly separated. This separation was observed for both unamplified and amplified libraries. However, linear amplification itself seemed to introduce a further discriminant vector to overall transcript pools, showing that this treatment changed the community profile. This separation was apparent in both SSU rRNA and functional transcript profiles (Fig. 22).

Not unexpectedly, overall transcript diversity was always lower in heavy than in light fractions (Fig. 22B). Diversity of transcript libraries after linear amplification was consistently though not significantly (p= 0.073) decreased, suggesting that the amplification process could leave behind some specific transcripts or over-amplify others. Rare transcripts with copy numbers less than ten appeared with equal frequency in amplified and unamplified libraries (Fig. 23). Additionally, predicted Chao1 richness perfectly matched the richness observed (not shown), suggesting an adequate sequencing depth for sufficient community coverage.

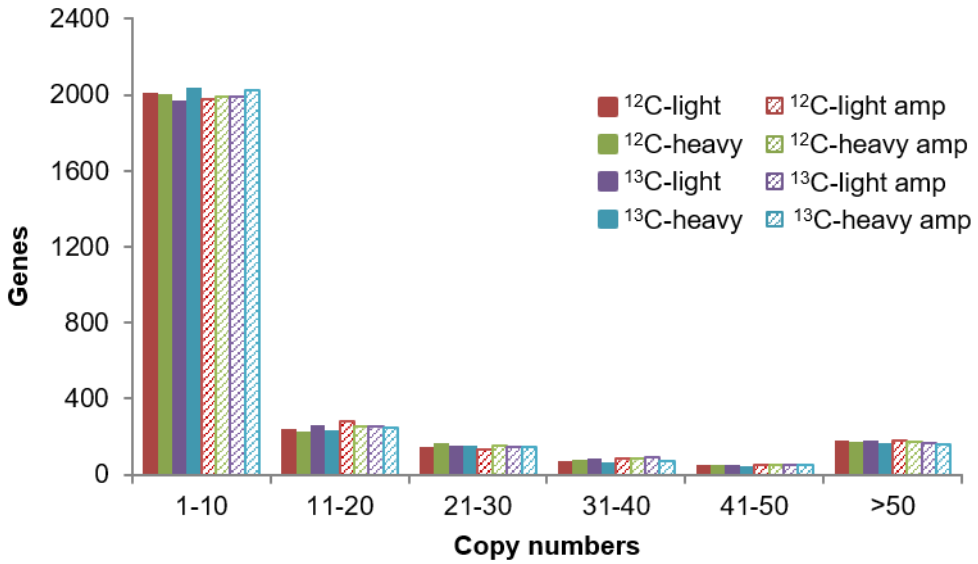

Figure 23: Histogram showing abundance distribution of the quantitative assignment of sequencing reads to specific functional transcripts in unamplified and amplified RNA-seq libraries of RNA-SIP fractions using KEGG.

To more directly test whether linear amplification increases the proportion of mRNA in total RNA, as hypothesized by Cao et al. (2010), I devised a SIP- and sequencing-independent lab experiment with pure culture RNA of *P. aeruginosa*.

RT-qPCR results of housekeeping gene transcripts did not indicate that linear amplification systematically increased mRNA ratios (Fig. 24). In fact, two of the three housekeeping genes tested had lower transcript/total RNA ratios after linear amplification than before, while one remained unchanged.

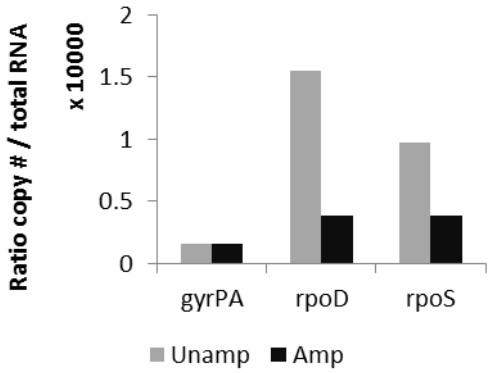

Figure 24: Ratios of gene abundance to total RNA in three P. aeruginosa housekeeping genes with and without linear amplification.

3.2.5.1 Targeted 23S rRNA depletion

I devised and tested a method for 23S rRNA depletion based on binding of 23S sequences to custom-made biotin-labelled probes, which should then be removed by streptavidin-coated magnetic beads. Amplicons of 23S rRNA were generated from sediment samples, and tested by cloning and sequencing before being used as templates for probe generation. Ninety-six clones were sequenced for each sample, resulting in 77 high quality sequences for 4A and 85 high-quality sequences for 4B. All high-quality sequences matched 100% with the 23S rRNA gene.

Probe-23S hybridization was successful, as shown by the greatly reduced 23S peak in Fragment Analyzer traces, and by the increased retention time of the probe-23S complex (Fig. 25B) compared to probes alone (Fig. 25D). However, the probe-23S complex could not be removed by streptavidin bead binding and was clearly still visible in Fragment Analyzer traces of purified, supposedly 23S-depleted RNA (Fig. 25B,C). Binding and removal were not successful when attempted with various formamide concentrations. The complex was also not removed using the Zymo RNA Clean & Concentrator-5 kit as an alternative strategy to

magnetic bead binding, although probes alone were retained by the columns when tested (data not shown). Although some variation of this method may prove viable with continued testing, the experiment was halted due to time constraints.

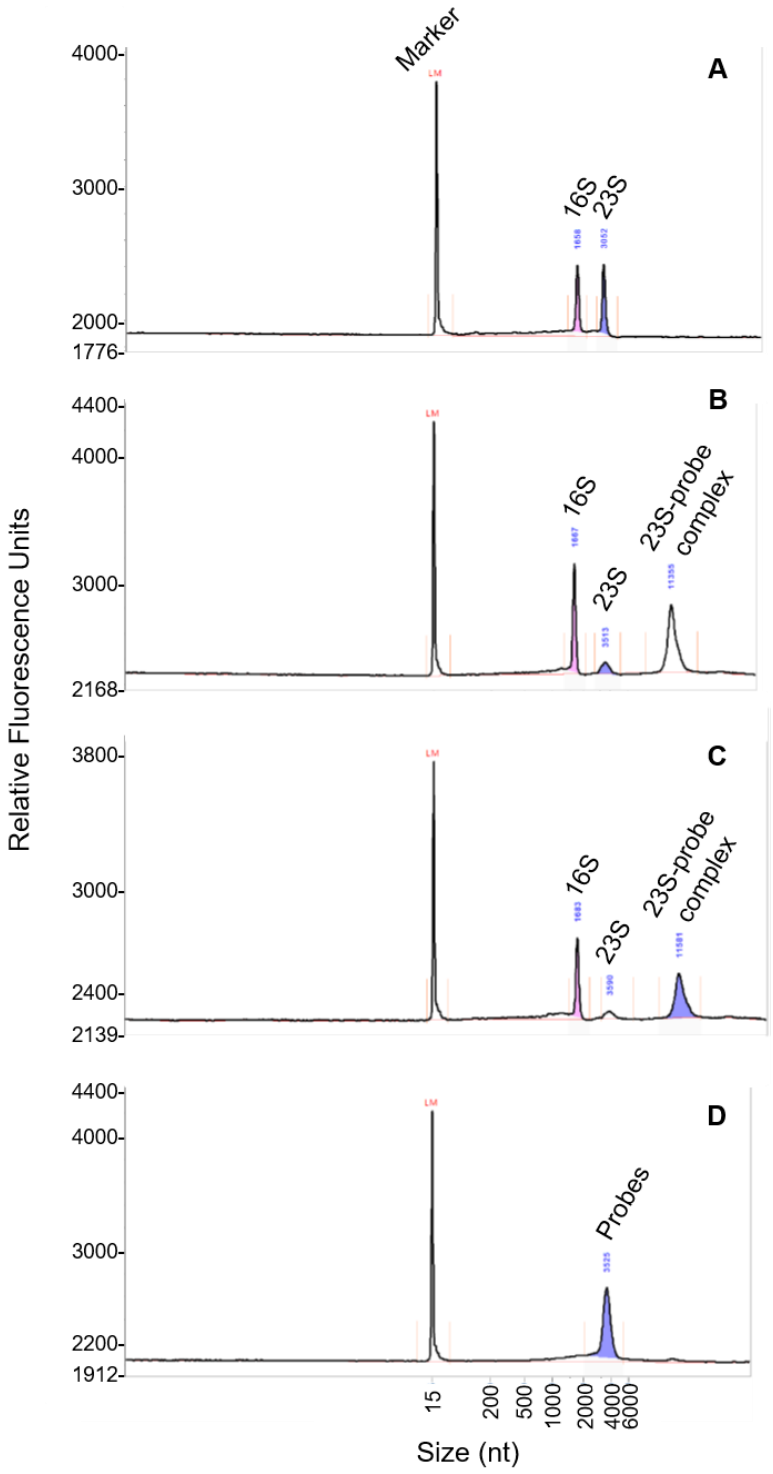

Figure 25: Fragment Analyzer traces of purified RNA during 23S depletion testing. (A) Un-depleted RNA aliquot (B) RNA after first bead-binding and cleanup (C) RNA after second bead-binding attempt, without cleanup (D) Aliquot of biotinylated 23S probes.

3.3 Degrader adaptation to limited dual electron acceptor availability

3.3.1 Microcosm incubations

Following the successful development of total-RNA-SIP in the previous microcosm experiment, I implemented this method to investigate microbial adaptations to limited and dual TEA availability. Microcosms were set up with sediment from the Flingern aquifer, provided with ^{13}C-labelled toluene, and incubated under 4 different redox conditions: oxic, microoxic, nitrate-amended microoxic, and anoxic. Changes in TEA concentration were monitored over the course of incubation, as was release of $^{13}CO_2$ as a proxy for toluene degradation. Microcosms were sacrificed during active toluene degradation, and extracted RNA was separated into heavy and light fractions by isopycnic centrifugation, then sequenced without PCR amplification or rRNA depletion.

The release of $^{13}CO_2$ (representing toluene degradation) was most rapid in oxic conditions, with onset approximately 2 days after setup and complete disappearance of toluene by day 6 (toluene concentrations not shown). Both microoxic treatments (with and without nitrate) had a longer lag phase until degradation began as compared to fully oxic conditions, after which degradation proceeded more rapidly in the presence of nitrate. Neither toluene degradation nor production of $^{13}CO_2$ were detected in fully anoxic, denitrifying microcosms, even after 45 days of incubation (Fig. 26B).

Reduction of nitrate to nitrite was rapid in the first three days of incubation, occurring at a slightly faster rate in anoxic vs. nitrate-amended microoxic conditions (Fig. 26C). After six days of incubation (the sacrifice time point for nitrate-amended microoxic microcosms), an average of ~25 % of initially amended nitrate remained in the nitrate-amended microoxic microcosms, while ~52 % had accumulated as nitrite and ~22% had been further reduced. In anoxic microcosms, 18 % of nitrate remained, 49 % accumulated as nitrite, and 33 % was further reduced after six days. Reduction of nitrate in the anoxic microcosms then nearly stopped between days six and 44, although additional nitrate was added on day 17 in an attempt to stimulate biodegradation activity in these microcosms (Fig. 26C). N_2O appeared in detectable amounts (> ~0.5 % of headspace gases) in nitrate-amended microoxic microcosms on day 7 and in anoxic microcosms on day 10 (Fig. 27). Production of N_2 gas was not quantifiable because of the N_2/CO_2 headspace of the microcosms. Ammonium concentrations were between 2.5 and 3 mg/L as measured by ion chromatography and did not change in any treatment over the course of incubation.

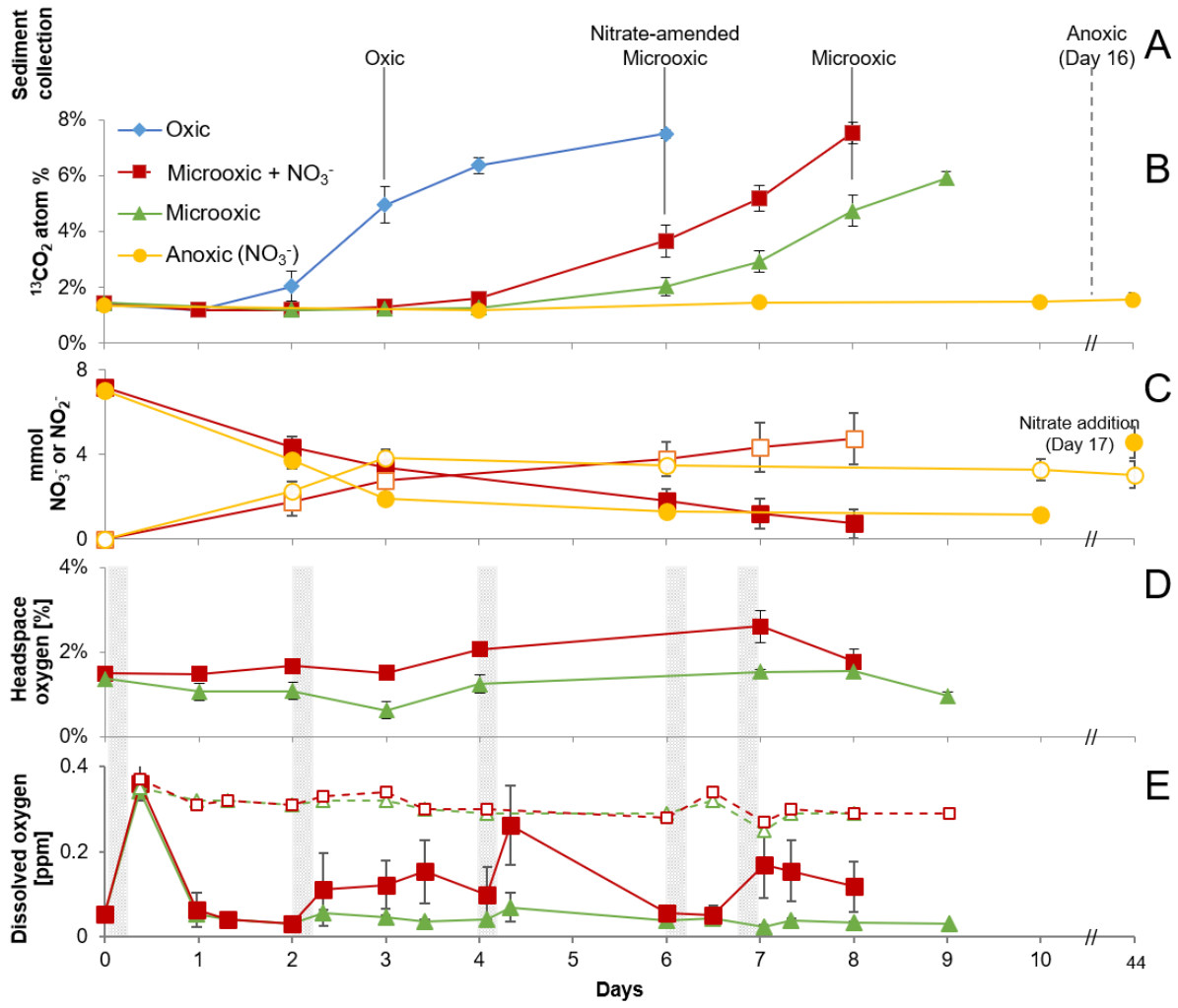

Figure 26: (A) Dates of first-round sacrifice during which sediment was collected for extraction and RNA sequencing. (B) Headspace $^{13}CO_2$ as a proxy for ^{13}C-toluene degradation. Natural abundance ~1.2 %. Complete degradation of toluene resulted in 7.5-8.0 %. (C) Nitrate (filled shapes) and nitrite (open shapes) concentrations in medium.(D) Oxygen in headspace and (E) Oxygen in medium of microoxic microcosms with or without nitrate. Vertical bars: oxygen replenishment events. Dashed lines: sterilized controls (only received first oxygen addition). Error bars are one standard deviation based on biological replicates.

Dissolved oxygen (DO) was rapidly depleted in both microoxic treatments after an initial lag period of ~12 h (Fig. 26E), but periodic re-spiking (Table 4) ensured that oxygen was never fully removed from the headspace (Fig. 26D). O_2 addition to headspace caused increased DO concentrations in microcosms only when nitrate was available at the same time. Although oxygen was not replenished in the oxic microcosms, headspace oxygen remained above 10 % and DO above 4 ppm (data not shown).

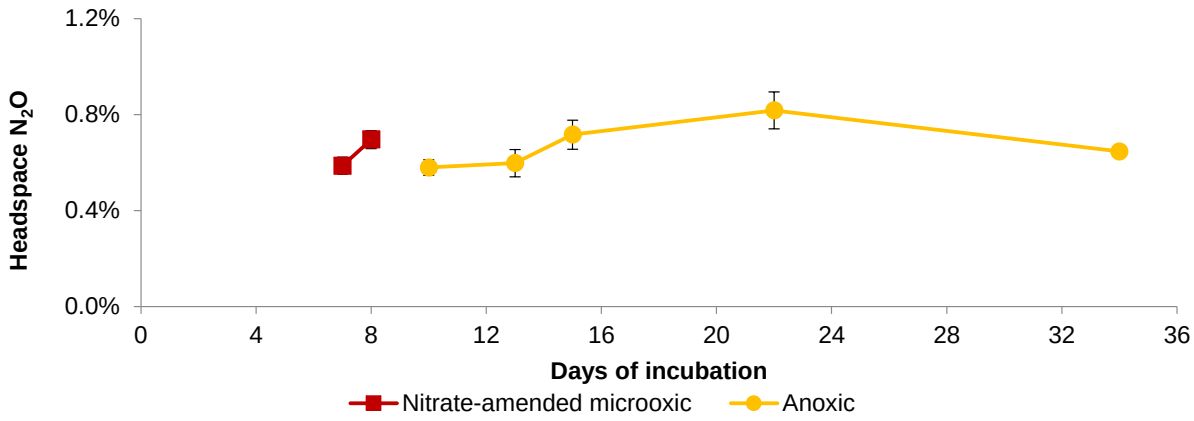

Figure 27: Headspace N_2O in microcosms amended with nitrate. Limit of detection appears to be ~ 0.5 % of total headspace signal. Error bars are standard deviation based on biological replicates.

The timing of microcosm sacrifices to collect sediment for RNA extraction and sequencing was aimed to occur during active toluene degradation, i.e. before toluene availability was exhausted. Pre-experiments indicated that complete toluene degradation resulted in 7.5 to 8 headspace atom % $^{13}CO_2$, so sacrifices for RNA extraction and sequencing were done between 3 and 5 atom % $^{13}CO_2$. Oxic microcosms were sacrificed after three days of incubation, nitrate-amended microoxic after six days, and microoxic after eight days (Fig. 26A). For anoxic, denitrifying microcosms, where toluene degradation was not detectable, sacrifices were made periodically and the earliest sacrifice point of sixteen days was chosen for sequencing to still allow for a reasonable reference to the microbes in the other microcosms.

3.3.2 Comparative rRNA labelling

RNA from light (~1.78 g/ml) and heavy (~1.81 g/ml) gradient fractions from labelled (^{13}C-toluene) microcosms was selected for sequencing, as was RNA from the light fractions of unlabelled (^{12}C-toluene) control microcosms. Triplicate labelled microcosms and duplicate unlabelled microcosms were sacrificed concurrently in each of the oxic, microoxic, and nitrate-amended microoxic conditions. Because no toluene degradation occurred in the anoxic treatments, a single microcosm was sacrificed and its RNA sequenced. cDNA sequencing of total RNA from gradient fractions revealed that almost all reads (>99 %) were bacterial. Proteobacteria constituted the majority of SSU rRNA reads in all microcosms (>86 %). Gammaproteobacteria were both the most abundant and the most enriched class in SSU rRNA under oxic conditions, while they were still most abundant but not enriched in microoxic

treatments. Though *Gammaproteobacteria* made up about one third of SSU rRNA in the light fractions of nitrate-amended microoxic conditions, they were only 5 % in the heavy fractions. Conversely, Betaproteobacteria comprised about half the SSU rRNA in the light fractions of nitrate-amended microoxic conditions but were highly enriched at 93 % of the heavy fractions (Fig. 28C). Although less abundant than Gammaproteobacteria, Betaproteobacteria were highly enriched under microoxic conditions (Fig. 28A).

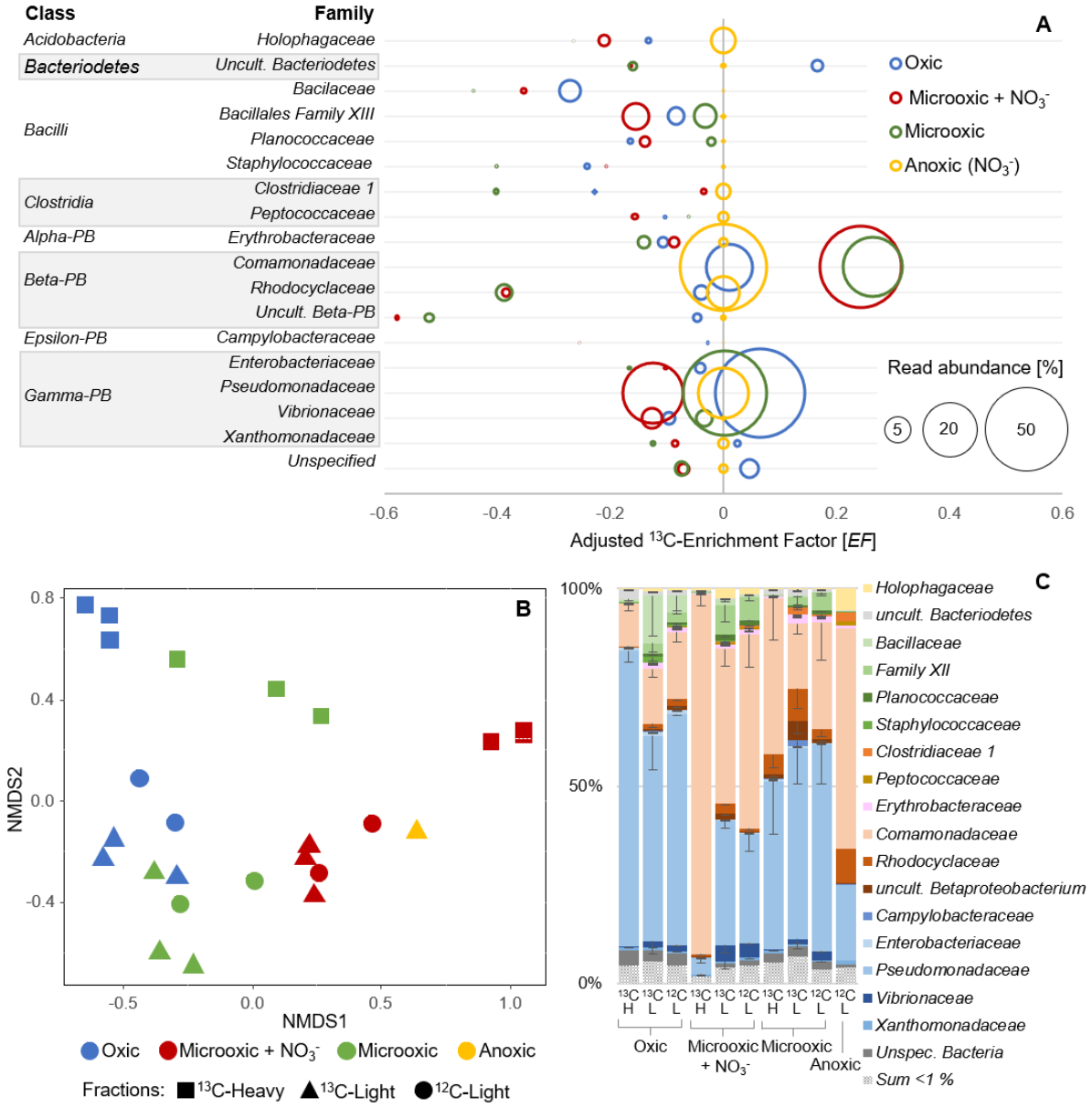

Figure 28: (A): ^{13}C-enriched bacterial families identified by SIP in the four redox conditions. EFs were calculated according to equation EF2 using averages from biological replicates. EFs for the anoxic microcosm sample were set to zero. Bubble size indicates replicate-averaged abundances in unlabelled controls. Families shown are those with at least 1 % abundance in at least one condition. (B) NMDS plot of 16S rRNA profiles at the OTU level. (C) Stacked bar blots of 16S profiles at the family level (averages of biological replicates). ^{13}C and ^{12}C indicate microcosms provided with labelled or unlabelled toluene, respectively; H and L indicate heavy and light fractions.

The genus *Pseudomonas* comprised almost all gammaproteobacterial SSU rRNA in oxic, microoxic, and anoxic conditions (≥91 % of Gammaproteobacteria) but a lower percent in nitrate-amended microoxic (≥82 %), where *Vibrio* made up most of the difference (4-11 %).

At the family level, *Comamonadaceae* dominated betaproteobacterial SSU rRNA (51 – 98 %; Fig. 28C), but there was far more variation at the genus level than observed for *Gammaproteobacteria*. *Variovorax* (19 – 30 %), *Simplicispira* (7 – 20 %), and *Acidovorax* and *Chlorochromatium* spp. (2 – 9 % each) were amongst the most abundant lineages in betaproteobacterial SSU rRNA. Uncultured *Comamonadaceae* comprised 10 – 33 % of *Betaproteobacteria*. No particular genus within the *Comamonadaceae* showed patterns of enrichment in any of the redox conditions (Fig. 29). Genera within the *Rhodocyclaceae* were <4 % each of *Betaproteobacteria*, except in the anoxic condition where *Azoarcus* was 7 % of total SSU rRNA reads, and in the ^{13}C-light fraction of microoxic conditions, where *Zoogloea* accounted for 15 % of betaproteobacterial SSU rRNA.

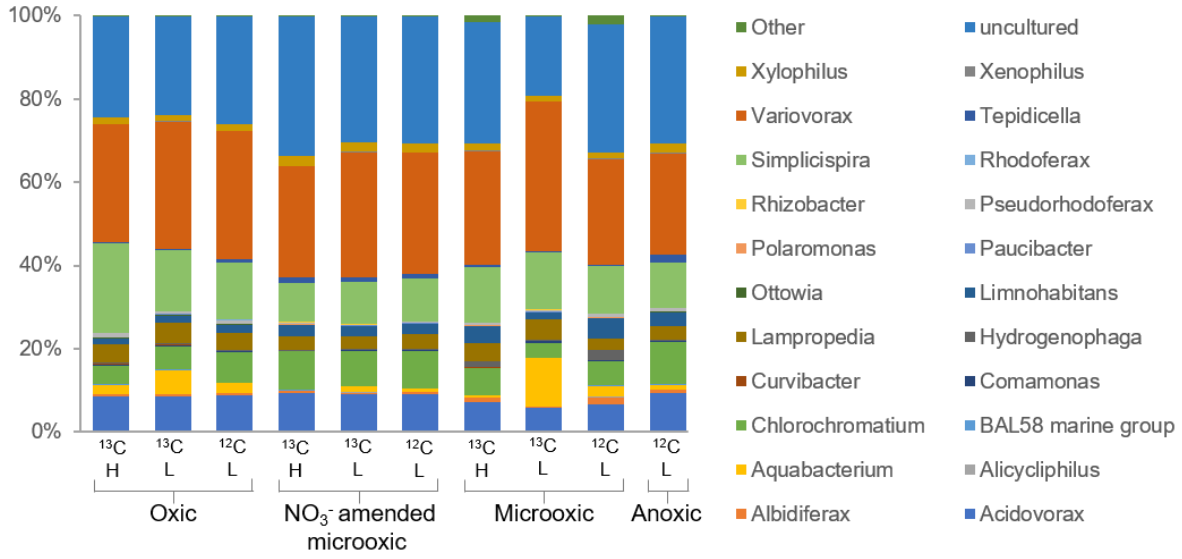

Figure 29: Genus-level rRNA composition of the Comamonadaceae family in each fraction. ^{13}C and ^{12}C indicate microcosms provided with labelled or unlabelled toluene, respectively; H and L indicate heavy and light fractions.

The NMDS ordination in Fig. 28B visualizes the similarity between density-resolved rRNA pools based on multivariate data. The analysis (Fig. 28B) showed that ^{13}C-light and ^{12}C-light fractions from all conditions clustered in the lower left corner of ordination space, while the ^{13}C-heavy fractions spread away in separate directions based on redox conditions. In a more

detailed view, within the light fractions, oxic and microoxic libraries clustered near each other, while fractions from the nitrate-amended microoxic microcosms were closer to the sole anoxic sample (^{12}C-light). Microoxic ^{13}C-heavy fractions ordinate in between ^{13}C-heavy fractions from oxic and anoxic microcosms. The tight clustering of biological replicates indicated a high reproducibility of transcriptome-SIP sequencing libraries between gradients. The larger spread between replicate microoxic ^{13}C-heavy fractions indicates more variability between these replicates than in any other group.

3.3.3 Labelling of functional gene transcripts

The highest-quality sequencing replicate(s) for each treatment and fraction were selected and interrogated for functional gene abundance and enrichment (Table A4). An average of 1.1 % ± 0.5 % of total RNA tags were non-rRNA, consistent with our previous total-RNA-SIP study on microoxic microcosms from the Siklós aquifer site (Bradford et al. 2018).

Xenobiotic degradation was the only level 2 KEGG category of functional gene transcripts appreciably enriched in ^{13}C-heavy RNA (Fig. 30). The highest enrichment was under oxic conditions, even when controlling for the amount of toluene degraded (measured via $^{13}CO_2$ release) at the time of sacrifice. Transcripts of many of the pathways (KEGG level 3) within the xenobiotic degradation category were enriched in ^{13}C, particularly those connected to degradation of aromatics, which share key enzymes for activation (e.g. phenol hydroxylase) or ring-opening (e.g. catechol-2,3-dioxygenase). Although phenol hydroxylase can activate toluene (Leahy et al. 2003), it is not included in the toluene or xylene degradation pathways in this KEGG release. Labelling was consistently highest under oxic conditions, while labelling in the two microoxic conditions (with or without nitrate amendment) was mostly similar. The higher labelling of the toluene degradation pathway without nitrate amendment was largely due to a higher abundance and labelling of transcripts involved in transformation of 2-hydroxy-5-methyl-cis,cis-muconic semialdehyde to 2-hydroxy-cis-hex-2,4-dienoate, a step of downstream aerobic toluene catabolism.

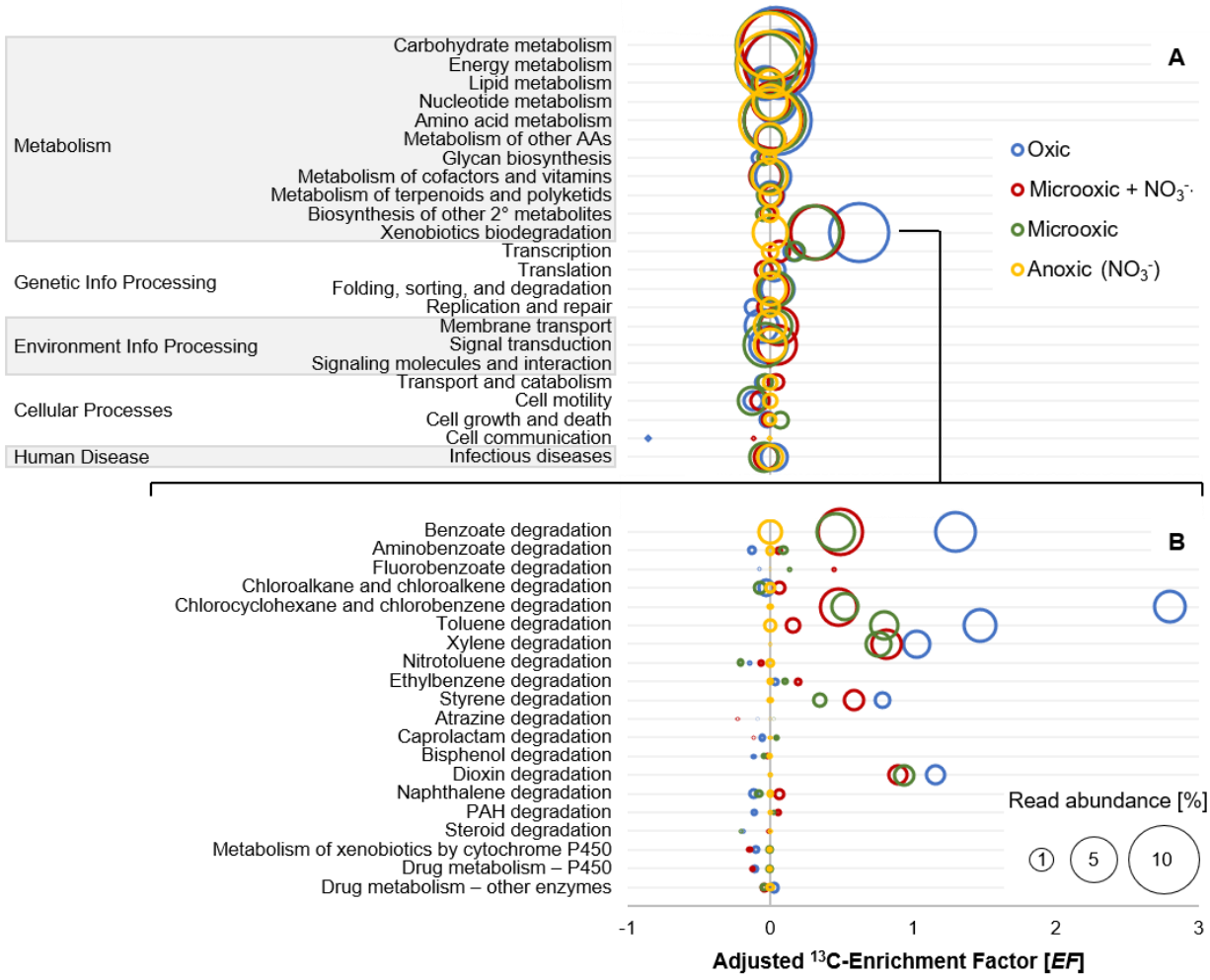

Figure 30: ^{13}C-enrichment and abundance of groups of functional genes at the level of KEGG categories (A) and pathways within xenobiotic degradation (B). Bubble size indicates abundances in unlabelled controls relative to total non-rRNA reads.

Respiration pathways (level 3) are part of the Energy Metabolism category (level 1) in KEGG. Transcripts of enzymes involved in oxidative phosphorylation or denitrification were not enriched in ^{13}C when considering contributions from all taxonomic groups together. However, when transcripts in the denitrification module as whole were separated by taxonomic assignment, enrichment was apparent in transcripts from the *Comamonadaceae* family, and higher taxonomic groups to which it belongs, in nitrate-amended microoxic conditions (Fig. 31).

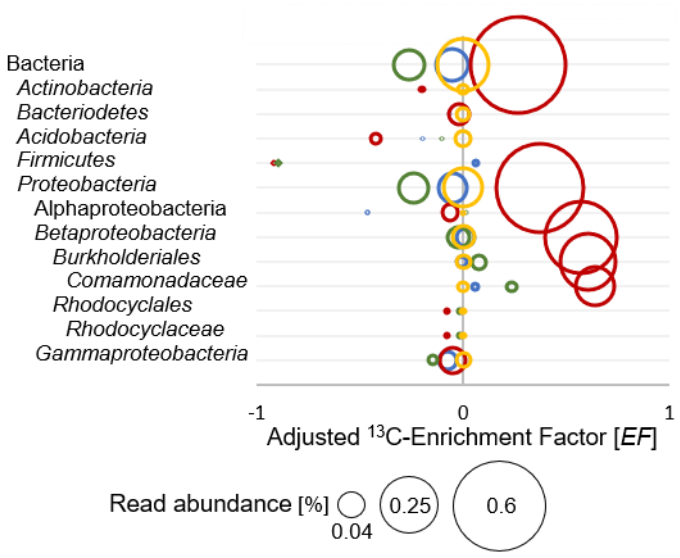

Figure 31: Bubble plot of enrichment and abundance of denitrification transcripts from selected taxonomic groups at various levels. EF from the anoxic condition are set to zero. Bubble size indicates abundances in unlabelled controls relative to total non-rRNA reads. Colours are the same as in Figure 30.

As in the Siklós-sediment based microoxic microcosms, transcripts of C23O were highly abundant and ^{13}C-enriched in all treatments with oxygen, while only a few copies of C12O were transcribed. Neither of these ring-opening oxygenases were present in the anoxic treatment.

3.3.4 Abundance of functional gene transcripts

Although the inferred EFs were highly informative for SSU rRNA and categories or pathways of functional transcripts, they may not always be reliable at the single transcript level. Here, small changes in low read numbers in different libraries could cause considerable changes in apparent EFs. Therefore, I also analysed the abundance of individual functional transcripts of interest using a Transcripts per Billion count adapted from the more common "Transcripts per Million" approach (Wagner et al. 2012).

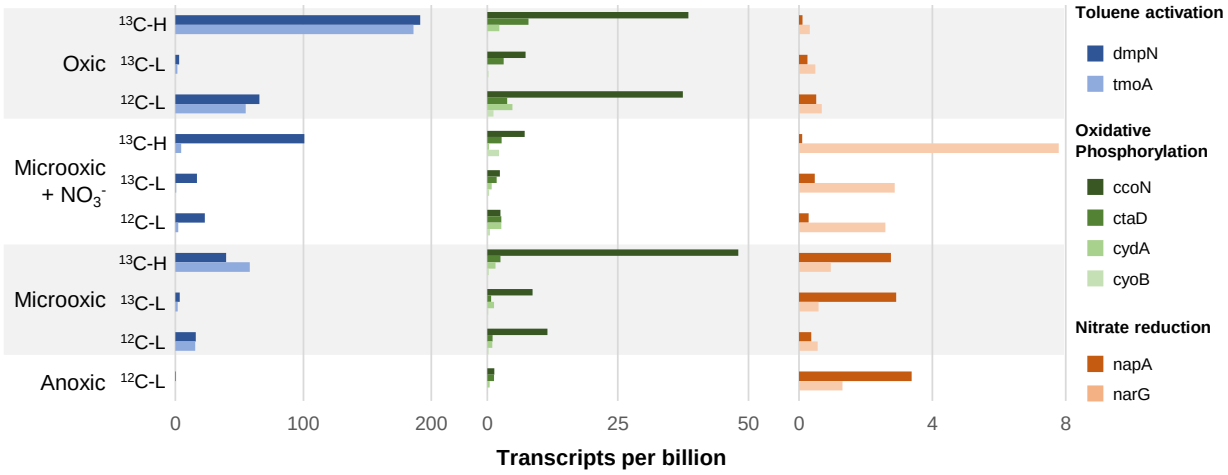

Figure 32: Bar plots of transcripts from marker genes for three process of interest: aerobic aromatic activation (left), terminal oxidases in oxidative phosphorylation (centre), and reduction of nitrate to nitrite (right). ^{13}C and ^{12}C indicate microcosms provided with labelled or unlabelled toluene, respectively; H and L indicate heavy and light fractions.

Genes *dmpN* and *tmoA* code for subunits of phenol hydroxylase (also known as phenol-2-monooxygenase) and toluene monooxygenase, respectively, and are marker genes for two different peripheral aromatic degradation pathways. These genes were transcribed in similar numbers in all fractions of oxic and microoxic microcosms (Fig. 32). In nitrate-amended microoxic microcosms, *dmpN* was much more prevalent than *tmoA*. All *tmoA* transcripts were proteobacterial, as were the majority (79 to 98 %) of *dmpN*. Only between 7 and 23 % of *tmoA* reads per fraction could be annotated more specifically than at the phylum level; of these, nearly all came from *Pseudomonas*. Between 61 and 70 % of *dmpN* reads could be annotated at the class level in nitrate-amended microoxic fractions, and nearly all of these were betaproteobacterial. Fewer (17 to 40 %) could be annotated at the class level in fractions from other conditions; these were mostly gammaproteobacterial in oxic fractions and mostly betaproteobacterial in microoxic fractions. Neither *dmpN* nor *tmoA* was expressed in anoxic microcosms (Fig. 32). Transcripts of *bssA*, a marker for anaerobic aromatic activation, were present only in very low copy numbers (< 0.1 transcripts per billion), even in anoxic microcosms (data not shown). Of the 103 unique benzylsuccinate synthase-containing genera currently known in the NCBI protein database, only two were present in the 16S rRNA from the anoxic microcosm: *Azoarcus* (6.8 %) and *Desulfosporosinus* (1.2 %).

Four variants of aerobic respiration terminal oxidases were investigated targeting marker genes described by Morris & Schmidt (2013).Transcripts of the gene *ccoN*, which codes for a subunit of high-affinity C-type oxidase, were the most highly transcribed of the four in all samples. In fractions from oxic and microoxic conditions, *ccoN* expression was between 2 and 11 times

higher than the other terminal oxidase markers combined. *ccoN* expression was lower in fractions from nitrate-amended microoxic conditions (Fig. 32). Transcripts involved in denitrification were more highly expressed in microcosms with nitrate than those without.

Transcripts of the *napA* gene, which codes for the catalytic subunit of periplasmic nitrate reductase, were more highly expressed in fractions from microoxic microcosms than in fractions from nitrate-amended microoxic conditions. The converse was true of *narG* transcripts, which code for a subunit of membrane-bound nitrate reductase (Fig. 32). Nearly all *napA* transcripts were proteobacterial, with *Epsilonproteobacteria* accounting for up to 89 % of *napA* reads in the heavy ^{13}C fraction of the nitrate-amended microoxic treatment. Betaproteobacteria reads accounted for up to 11 % of *napA* in nitrate-amended microoxic fractions, but between 10 and 45 % in microoxic fractions. Most *narG* transcripts were also affiliated to *Proteobacteria* (68-99 %) but *Geothrix fermentans* (*Acidobacteria*) also accounted for notable read counts (7.5 - 16 % of *narG* transcripts) in the nitrate-amended microoxic or anoxic microcosm. Transcripts of genes involved in either assimilatory or complete dissimilatory nitrate reduction (*nirA,nirB, nirD, nrfA, nrfH, nit-6*) were detected only in extremely low read numbers. Transcripts associated with the putative nitric oxide dismutase (*nod*) were also not found in any fraction.

In summary, TEA availability had clear impacts on both the composition of the microbial community in these microcosms, and on the metabolic and respiratory strategies used by those community members capable of degrading toluene. Although denitrification-driven toluene degradation did not occur in the absence of oxygen, nitrate amendment combined with microoxic conditions had a clear effect on the degrader community. The combination of functional and taxonomic data provided by total-RNA-SIP was necessary to determine that *Comamonoadaceae* respired nitrate while using oxygen to activate hydrocarbons.

4. Discussion

4.1 Putative nitric oxide dismutase in the environment

Nitric oxide dismutation to O_2 and N_2 has been hypothesized to be a new pathway in the nitrogen cycle, bypassing N_2O production and potentially providing the means for oxygen-dependent processes to occur under anoxic conditions. Evidence of this process was first reported in the methanotroph "Ca. Methylomirabilis oxyfera" (Ettwig et al. 2010), a member of the NC10 candidate phylum, followed closely by alkane-oxidizing gammaproteobacterium HdN1 (Zedelius et al. 2011). A follow-up paper by Ettwig et al. (2012) hypothesized that the key enzyme in this reaction, putative nitric oxide dismutase (*nod*), was a variant of quinol-dependent nitric oxide reductases (qNORs). These widespread enzymes are members of the heme-copper oxidase superfamily that perform the essential task of reducing toxic NO to non-toxic N_2O using electrons accepted from quinol (Matsumoto et al. 2012). Despite high sequence similarity in most sections, amino acid substitutions in the quinol-binding site compromise this function in putative *nod*. Furthermore, substitutions in the catalytic centre cause an altered active site configuration. These changes and the high expression levels of these qNOR-related gene products in "Ca. Methylomirabilis oxyfera" under anoxic conditions (Ettwig et al. 2010) led to the suggestion that these enzymes are functional in NO dismutation (Ettwig et al. 2012). The aim of the work presented here was to (i) develop workable primers targeting *nod* and (ii) use these primers to explore the prevalence and diversity of *nod* in hydrocarbon-contaminated aquifers and other environments.

4.1.1 Primer development

Prior to this work, the prevalence of *nod* in environmental systems was largely unknown. Primer design was hampered by the existence of only two reference sequences: putative *nod* genes in the genomes of "Ca. Methylomirabilis oxyfera" and HdN1. Bhattacharjee et al. (2016) developed a set of primers based on an alignment of these reference sequences, and found putative *nod* in a nitrite-reducing methane-oxidizing laboratory reactor inoculated with river sediments. However, these PCR products were very short at 329 – 426 bp, making phylogenetic analysis and consideration of amino acid substitutions difficult.

The current study aimed to iteratively design and test targeted primers for *nod* that would produce longer PCR products with higher sequence diversity. The designs were initially based on the few available reference sequences and then expanded iteratively using the sequences

retrieved from environmental samples. We used a “classical” cloning and Sanger sequencing technique for expedience and cost-effectiveness, especially considering the iterative rounds of sequencing needed for primer design. In addition, Sanger sequencing typically provides longer reads than widely-available high-throughput sequencing techniques like Illumina sequencing, which was beneficial when dealing with PCR products of > 1000 bp. In the end, we developed four forward primers and five reverse primers, in six primer pair combinations (Fig. 33). Primer pair D (nod684Fv2/nod1706Rv2) incorporates environmental *nod* sequences and produces a ~1000 bp amplicon covering several sites that distinguish *nod* from canonical *qnor*. It successfully recovered a high diversity of lineages from WWTP samples. For these reasons, I recommend primer pair D for testing of environmental samples. Nested PCR by pair E, followed by pair D, could also be useful for difficult samples. One pair, combination F, makes a comparatively short product of suitable length for quantitative PCR, which was employed in this thesis to investigate the abundance of *nod* compared to 16S in WWTP samples.

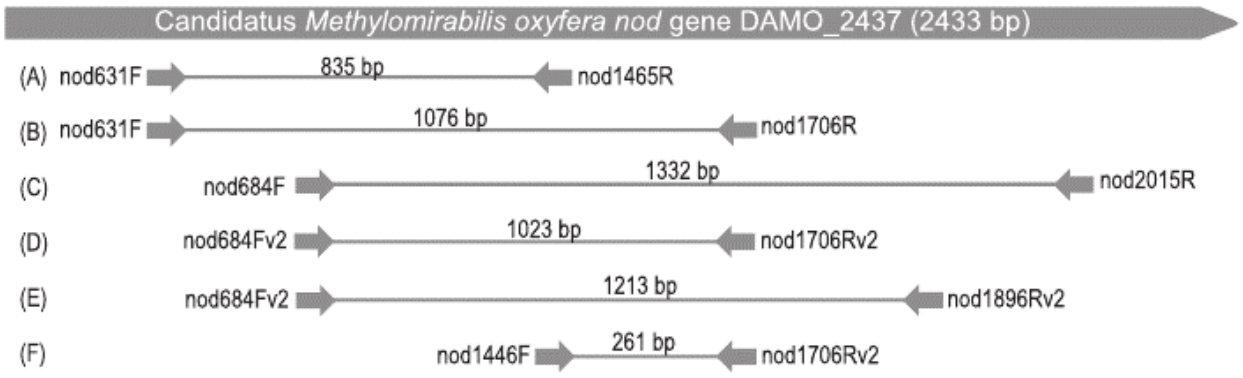

Figure 33: *nod*-targetted primer sets developed in this study. Positioning numbers are based on the "Ca. methylmirabilis oxyfera" DAMO_2437 nod gene. Expected amplicon sizes are indicated. Lengths are not drawn to scale. From Zhu (2016).

4.1.2 Diversity and abundance of environmental nod gene lineages

Despite the initial use of such a limited number of reference sequences, the environmental *nod* genes from contaminated aquifers were clearly distinct from those found in “Ca. Methylomirabilis oxyfera” and HdN1. The gene pools in Siklós and Flingern aquifers were also strikingly different from each other, possibly reflecting the distinct contamination and redox conditions at the two sites. The five Flingern clones shared >99 % sequence identity and were more closely related to the HdN1 sequence than to “Ca. Methylomirabilis oxyfera”. Conversely, the majority of clones from the Siklós aquifer were closely related to one of the

two nod paralogs in “Ca. Methylomirabilis oxyfera”, while one clone (Sik2DC09) seemed to be neither true qNor nor Nod.

This study was expanded by my colleague Dr. Baoli Zhu, who used the second round of primers to investigate the diversity and abundance of *nod* in a series of engineered water systems. The recovered genes were more diverse than those from aquifer systems (Fig. 15). Using primer set F for qPCR, he found comparable ratios of *nod* to 16S rRNA (~ 2 %) in the two systems, but a higher absolute abundance of *nod* in the Siklós aquifer. This was probably due to the collection of high-organic sump material collected from Siklós compared to the more oligotrophic mineral sediments from the Flingern aquifer. The abundance of *nod* was up to 10 % of 16S counts in reactors designed for partial nitrification and anammox, where a high abundance of anammox organisms can be assumed.

Since publication of our study results in 2017, our *nod*-targeted primers have been used to search for evidence of nitric oxide dismutation in diverse environments by other research groups. Abercron et al. (2017) found no evidence of *nod* in their oil-contaminated aquifer sediments or enrichments provided with naphthalene in anoxic, denitrifying conditions. Zhang et al. (2018) investigated soil from a methanogenic alpine wetland in Tibet which was known to harbour NC10 bacteria. Using qPCR, they found a higher abundance of *nod* genes than NC10-specific 16S, suggesting either multiple copies of *nod* per genome, or the presence of *nod* genes in microbes outside the NC10 group. When translated, ~ 75 % of their *nod* sequences showed the amino acid substitutions characteristic of *nod*, while ~ 25 % had only some of the substitutions and clustered with unknown Nor-related sequences. Hu et al. (2019) found diverse *nod* genes in samples from WWTPs, hydrocarbon-impacted wastewater, hydrocarbon contaminated soil, and marsh soil. They also produced more degenerate *nod*-targeting primer sets, which they recommend for future use despite the failure of these new primers to amplify *nod* genes from environmental samples. Padilla et al. (2016) also found transcripts annotated as *nod* in transcriptome libraries from marine oxygen minimum zones. Further work by members of my research group compared the phylogeny of *nod*, *pmoA*, and 16S rRNA from four oxygenic NC10 enrichment cultures and one Siklós aquifer sediment sample (Zhu et al. 2019). Based on the congruence of these three phylogenies, we concluded that *nod* is suitable for future use as a functional and phylogenetic of methane-driven oxygenic denitrifiers in the NC10 group.

4.1.3 Clues to functionality from amino acid sequence

All environmental *nod* found in our study had amino acid substitutions characteristic of "Ca. Methylomirabilis oxyfera" *nod* (Fig. 16), supporting our assumption that the primers specifically target true *nod* without amplifying closely-related *qnor* genes. One each of the Siklós and engineered water system clones, as well as about one quarter of the sequences later retrieved by Zhang et al. (2018), formed an interspersed cluster on the phylogenetic tree between *nod* and *nor*s (Fig. 15). This cluster also contained reference sequences from members of the Fibrobacteres-Chlorobi-Bacteroidetes (FCB) superphylum which were annotated as q*nor* in the NCBI database. Translated sequences in this cluster contained only some of the *nod*-type amino acid substitutions, which may render them unable to use quinol for electron binding. As yet, no biochemical evidence exists for the function of this intermediate cluster, but it may be a clue to the evolution of *nod* via mutation of *qnor* genes.

The success of these primers in retrieving *nod* sequences from a variety of environments proves that *nod* is widespread, and its distribution and abundance in many other environments remains to be explored. However, the connection between the protein product of the *nod* gene and active dismutation has not yet been definitely shown. Isolation of transmembrane proteins is very technically challenging (Carpenter et al. 2008), so the putative Nod protein has not been isolated and tested. Until the biochemical activity and mechanism of nitric oxide dismutation can be confirmed, the activity of Nod enzymes remains putative.

4.2 Total-RNA-SIP and BTEX degraders in a microoxic aquifer

The primary goal of this work package was to provide a proof-of-concept for the feasibility of total-RNA-SIP (also termed transcriptome-SIP), showing the advantages in providing targeted insights into microbial activities at taxonomic and functional levels from the same sequencing library. I did this using RNA extracted from microoxic, pollutant-degrading aquifer sediment microcosms, so the results not only show the utility of this novel method, but also the identity and activity of poorly understood microaerobic toluene degraders. The DNA-SIP study undertaken from these same microcosms (Táncsics et al. 2018) offered a valuable point of comparison for the sequencing results and showcased the additional depth of information given by total-RNA-SIP vs. amplicon-based DNA-SIP.

4.2.1 Methodological considerations

The level of detail in functional transcript data clearly depended on the number of mRNA reads obtained from gradient fractions. The amount of non-rRNA reads was between 0.6 and 2.2 % in each library (Table 7), but this small proportion was not unexpected for total RNA-seq and was consistent with the percentage of non-RNA or putative mRNA reported for other total-RNA-seq studies (Table A5). Most transcriptomic studies use mRNA enrichment (eukaryotes) or rRNA depletion (prokaryotes) to increase the depth of mRNA sequencing. However, my interest in SSU-based microbial community composition and desire to avoid potential skewing effects motivated a depletion-free, total-RNA approach. This made deep sequencing necessary for adequate non-rRNA coverage, preventing the sequencing of true biological replicates for RNA fractions. However, the inclusion of both amplified and unamplified libraries, in which overall labelling patterns were generally consistent (Fig. 22, Table A1), provided a certain level of confidence to this end. A further cautioning statement needs to be added with reference to the comparably late time point in microcosm incubation (7 d) chosen for transcriptome SIP analysis. Certain patterns of transcript labelling during initial phases of toluene catabolism may have been missed due to the more rapid turnover of distinct mRNA species. The late time point was motivated by the need for abundant and highly ^{13}C-labelled RNA in this primary proof-of-concept study.

The interpretation of SIP data strongly relies on comparing sequencing results between ^{12}C and ^{13}C-treatments, and across similar ranges of buoyant densities (Whiteley et al. 2007). This was not fully possible in the Siklós total-RNA-SIP study, mainly because post-gradient PCR amplification of ^{12}C-RNA recovered from heavy fractions, and vice versa, could not be relied upon. Direct sequencing of total RNA and parallel linear amplification necessitated a certain minimum amount (~120 ng) of total RNA recovered after fractionation. To increase yields for low-RNA fractions, several fractions of similar density across duplicate centrifugation gradients were pooled (Fig. 17). In essence, RNA from all meaningful fractions of the density-resolved total RNA obtained were sequenced from both ^{12}C and ^{13}C gradients, which was the best that could be accomplished considering these methodological limitations. While the lack of sequencing replicates precludes statistical calculations or differential expression analysis that have been previously used in amplicon-based SIP studies (Pepe-Ranney et al. 2016), the enrichment factors based on differences in labelled and control microcosms offer a semi-quantitative way to interpret RNA labelling (Kramer et al. 2016; Zhang & Lueders 2017).

4.2.2 Taxonomic insights via rRNA of labelled degraders

Since contaminated aquifers are well-known to harbour not only bacteria, but also archaea and microeukaryotes (Griebler & Lueders 2009), the lack of respective rRNA in these libraries came as a surprise. Possibly, the limited availability of oxygen in the microoxic microcosms as well as the substantial toluene loads (~1 mM; Táncsics et al. (2018)) contributed to selection of specific bacterial degraders during SIP incubation. Betaproteobacteria have been reported previously to be prevalent in oxygen-limited, aromatic hydrocarbon-contaminated groundwater environments (Fahy et al. 2006). *Rhodocyclaceae* in particular are known to be important at the investigated site, both from previous samplings (Táncsics et al. 2012; Táncsics et al. 2013) and from DNA-SIP of the same microcosms as used in this study (Táncsics et al. 2018). Members of the *Rhodocyclaceae* can degrade aromatics both aerobically and anaerobically (Song et al. 2001; Chakraborty et al. 2005). *Dechloromonas aromatica*, for example, can degrade several aromatic compounds using oxygen, nitrate, perchlorate, or chlorate as electron acceptors (Chakraborty et al. 2005). *Dechloromonas* rRNA was very abundant and clearly enriched in ^{13}C-rRNA, despite being represented by less than 1% of the DNA amplicons sequenced in DNA-SIP of the same microcosms (Táncsics et al. 2018). Since labelling of DNA requires cellular replication, while RNA labelling can occur independent of cellular growth (Manefield et al. 2002), this could indicate that *Dechloromonas* was metabolically active despite slow growth during microoxic SIP incubation. Alternately, this differential representation could be a result of primer bias during the production of PCR amplicons for sequencing. Primer bias was even more clearly apparent for the genus *Azonexus*, which was abundant (~6%; Table A1) and highly enriched in ^{13}C-rRNA, while not detected at all in DNA amplicons (Táncsics et al. 2018).

Quatrionicoccus spp. (a genus within the *Rhodocyclaceae* – originally called *Quadricoccus* but reclassified as *Quatrionicoccus* (Tindall & Euzéby 2006)) – were first isolated from activated sludge biomass in an enhanced biological phosphorus removal reactor with cycling aerobic/anaerobic conditions (Maszenan et al. 2002). Members of this genus are considered to be strict aerobes. This genus comprised an average 10 % of the rRNA reads and was highly labelled, indicating effective toluene metabolism under microoxic conditions. This was consistent with labelling results from DNA-SIP of the same microcosms (Táncsics et al. 2018). The genus is also highly abundant in Siklós groundwater itself (Farkas et al. 2016). Before this work, members of the genus have not been described as hydrocarbon degraders. This contrasts other members of the *Rhodocyclacaeae* such as *Zoogloea* spp. which were clearly identified as

labelled in both rRNA and mRNA in the present study, and which have been frequently reported as aerobic degraders of aromatic hydrocarbons (Farkas et al. 2015; Jechalke et al. 2013).

In contrast, other microbes of the BTEX contaminated sediment were abundant in RNA but without any indication of labelling, including members of the *Pseudomonadaceae* and *Comamonadaceae*. Both are well-known to degrade hydrocarbons (Wald et al. 2015; Fahy et al. 2006), under either aerobic (Martinez-Lavanchy et al. 2010) or nitrate-reducing conditions (Sun & Cupples 2012). These organisms must have been active and producing RNA transcripts, potentially thriving on electron donors introduced with the original sediment inoculum, but not receiving labelled carbon from the added toluene.

4.2.3 Functional insights via labelled mRNA

The combination of stable isotope probing with metatranscriptomic sequencing provided information on a whole range of functional transcripts produced by active degraders during incubation. This depth of information is both the great advantage and disadvantage of metatranscriptomic sequencing compared to targeted methods like RT-qPCR; specific patterns of interest must be selected from a wealth of possible insights. Here, I discuss transcript labelling patterns which appeared most relevant in understanding degrader activity under microoxic conditions.

Transcripts linked to genes coding for proteins involved in cell motility were the most highly labelled group in both KEGG and COG analyses (Figs. 20-21). The importance of chemotaxis of degraders towards aromatics, including BTEX, is well-known (Krell et al. 2013). In highly contaminated systems, motile cells have advantages not only in moving towards contaminant sources, but also in escaping toxically high local concentrations (Parales et al. 2000). Still, the apparent importance of motility amongst degraders in agitated SIP microcosms came as a surprise, and could potentially be connected to a lower availability of toluene or oxygen during the late phase of degradation (Táncsics et al. 2018), triggering a motility response (Lacal 2017).

Transcripts of genes coding for flagellin, the principle component of bacterial flagella, were the most abundant of the transcripts linked to motility and were highly ^{13}C-enriched. Between 10 and 20 % of all transcripts of the *fliC g*ene were assigned to *Azovibrio restrictus* (Table A2), a N_2-fixing microaerophile within the *Rhodocyclaceae* (Reinhold-Hurek & Hurek 2000), while only ~0.1 % of 16S rRNA reads were affiliated to *Azovibrio* sp. (Table A1). It is highly unlikely

that an organism so rare within the community could have contributed so much of one of the most abundant and highly labelled transcripts in these microcosms. Either lateral gene transfer, or the lack of more appropriate reference genome data for other members of the *Rhodocyclaceae*, such as *Azonexus* spp., could potentially explain this finding. No other single species contributed more than 1.6 % of transcripts of the *fliC* within each sample, even known motile toluene degraders such as *Dechloromonas* (Chakraborty et al. 2005) or *Zoogloea spp.* (Farkas et al. 2015), which were much more abundant in rRNA reads than *Azovibrio*.

4.2.3.1 Xenobiotic degradation transcripts

The relatively high copy numbers and labelling of transcripts for the phenol hydroxylase gene and the low numbers or complete absence of other possible oxygen-dependent toluene-activating gene transcripts indicated that phenol hydroxylase was the primary driver of toluene activation in these microcosms. Phenol hydroxylase has a broad substrate specificity for monoaromatics and its marker gene *dmpN* has been previously found in microoxic BTEX-contaminated groundwater (da Silva & Corseuil 2012) and used as a biomarker for aerobic aromatic degradation (Nebe et al. 2009). However, it is also possible that transcripts of other genes in upper pathways of toluene catabolism were transcribed at earlier time points of microcosm incubation, but may have been subject to more rapid turnover and decay.

Catechol-2,3-dioxygenase (C23O) is an extradiol dioxygenase which has long been used as a conserved functional marker gene for aerobic aromatic degraders (Táncsics et al. 2012; Táncsics et al. 2010; Táncsics et al. 2013). Kukor & Olsen (1996) found that C23O in the subfamily 1.2.C of extradiol dioxygenases efficiently used low amounts of dissolved oxygen, and likely evolved in environments with low oxygen and substrate concentrations. Considering its frequent presence in contaminated environments (Táncsics et al. 2012; Hendrickx, Dejonghe, et al. 2006) and the previously seen high diversity at this site (Táncsics et al. 2012; Táncsics et al. 2013; Táncsics et al. 2010), it is not surprising that C23O was highly expressed and the related transcripts highly ^{13}C-labelled in SIP microcosms. The majority of these transcripts could only be identified to the phylum level (Table A3). This may be because gene sequences are so similar across different lineages that affiliation by MEGAN's lowest-common-ancestor algorithm stopped at a higher level. A more likely explanation is the lack of more specifically assigned reference sequences in the KEGG database, since most C23Os could not be linked to any cultured bacterium.

Within the transcripts that could be assigned more specifically, *Rhodocyclaceae* were most prominent. This included matches to *Azovibrio restrictus*, previously found in hydrocarbon-contaminated soil (Yang et al. 2014) and to *Zoogloea oleivorans*, a facultative aerobic hydrocarbon-degrader (Farkas et al. 2015). Transcripts of the C23O gene affiliated to the *Pseudomonadaceae* were nearly as frequent as that of the *Rhodocyclaceae* (avg. 6.5 %), although *Pseudomonadaceae* were clearly not isotopically enriched (Table A3). Members of this family are well-known as aerobic degraders of BTEX compounds (Martinez-Lavanchy et al. 2010) and have long been known to harbour C23O (Kojima et al. 1961). The results of this labelling experiment suggest that they were not competitive in toluene degradation in the microcosms, or assimilated C from alternative substrates for biomass buildup.

Microoxic sedimentary systems may well contain anoxic micro-niches where anaerobic metabolism may take place. Furthermore, the daily replenishment of oxygen during microcosm incubation (Táncsics et al. 2018) resulted in periods of anoxia. For these reasons, I also searched for transcripts of marker genes of anaerobic aromatic degradation, *i.e.* benzylsuccinate synthase (Bss) (von Netzer et al. 2016). While *bssA* was indeed transcribed in all microcosms, transcripts were not labelled, indicating that anaerobic degraders were not competitive in the microcosms.

4.2.3.2 Transcripts of respiratory pathways

Genes involved in aerobic respiratory pathways were transcribed in low to moderate copy numbers. However, only transcripts of genes coding for a Type I NADH dehydrogenase were present, not Type II, although the majority of *Betaproteobacteria*, which dominated the system, are known to harbour genes for Type II as well (Marreiros et al. 2016). This may indicate an adaptation to microoxic conditions, as some obligate microaerophiles harbor only Type I (Smith et al. 2000), and expression of Type II in *E. coli* occurs only under fully oxic conditions (Spiro et al. 1989).

Despite not having added nitrate to the microcosms, transcripts of genes coding for all steps of canonical denitrification and dissimilatory nitrate reduction to ammonium (DNRA) were detected. All related transcripts were unlabelled except for *napA*, which codes for a nitrate reductase adapted to low-nitrate conditions (Potter et al. 1999) and suggested to be involved in aerobic denitrification (Ji et al. 2015), indictating that such physiologies could have been relevant during microcosm incubation.

4.2.4 Methods to enhance total-RNA-SIP

Linear amplification is one potential strategy to alleviate the technical limitations involved in total RNA-seq following SIP gradients. I compared this to direct sequencing in the Siklós microcosm total-RNA-SIP study. Linear amplification increased RNA amounts, but also decreased GC content in non-rRNA (Fig. 18), decreased library diversity, and caused profound changes to community profiles for SSU rRNA and functional transcripts (Fig. 22). Linearly amplified libraries contained a much higher percentage of small (<50 bp) read fragments which were removed early in the bioinformatic pipeline. Previous studies have also found changes in transcript ratios when using various linear amplification protocols (Degrelle et al. 2008; Nygaard et al. 2003), which may be due to inefficiency or bias in the initial reverse transcription step (Levesque-Sergerie et al. 2007). The equal numbers of rare transcripts in amplified and unamplified libraries (Fig. 23) indicate that the amplification process did not simply miss less common transcripts, but must have selected based on another characteristic, perhaps GC-content. While the proportion of non-rRNA was somewhat increased by linear amplification (Table 7), this increase was too small to provide tangible benefits in the face of the negative effects caused by the treatment.

Our independent test based on RT-qPCR targeting *P. aeruginosa* housekeeping genes did not indicate an increase in mRNA ratios as a result of linear amplification (Fig. 24). Thus I cannot confirm the idea that linear amplification could increase the proportion of mRNA to total RNA reads in sequencing libraries in a meaningful manner as hypothesized by Cao et al. (2010). Because of this and the marked skewing effects observed in overall library composition (Fig. 22), I recommend that linear amplification should be used only if total RNA yield is too low for direct analysis, and samples studied using linear amplification should only be compared to samples treated likewise.

Many transcriptome studies remove rRNA before library preparation in order to focus only on mRNA. This is most commonly achieved with a combination of SSU and LSU rRNA-targetted biotin-labelled probes and streptavidin-coated beads. This allows a much higher amount of mRNA data a given sequencing depth, but of course erases the taxonomic data in SSU rRNA, and may introduce some bias to the mRNA data remaining. To circumvent this, I aimed to adapt this depletion method to remove only LSU (in this case, 23S) rRNA, retaining the valuable SSU taxonomic data, along with mRNA, while opening up sequencing space to reduce the necessary sequencing depth of total-RNA sequencing. Had the depletions worked as planned, I intended to deep-sequence a series of RNA samples with and without 23S depletion

for comparison in order to quantify induced biases. Unfortunately, probe removal by magnetic bead binding or spin columns was unsuccessful, despite a number of attempts with varied conditions. Thus, 23S depletion could not be established as a viable strategy to increase mRNA read abundance in total RNA-seq libraries, and the experiment was discontinued due to time limitations.

4.3 Degrader adaptation to limited dual electron acceptor availability

Under fully oxidized conditions, oxygen can act as both an activator of aromatic hydrocarbons and as TEA in aerobic respiration. Under oxygen limitation, oxygen has been postulated to instead be used only for substrate activation, while respiration can proceed using alternative TEAs such as nitrate (Liu et al. 2018; Chayabutra & Ju 2000; Durant et al. 1999; Nestler et al. 2007). Adaptation and competition amongst different aerobic, microaerobic, and denitrifying degrader lineages potentially involved, however, remains poorly investigated. Using our recently developed total-RNA-SIP approach (Bradford et al. 2018), I aimed to contribute to a better understanding of the composition, catabolism, and respiration of degrader communities active under various redox conditions.

My first total-RNA-SIP study offered a proof-of-principle for the feasibility of this novel method, but lacked biological replication. In the present study, I demonstrate that SSU rRNA profiles were highly reproducible between replicated gradient fractions (Fig. 28). In principle, comparable reproducibility should also be feasible for non-rRNA reads, but >95 % of total RNA in microbial cells is typically rRNA (Lodish et al. 2000). Technical problems encountered by the sequencing company while preparing some cDNA libraries resulted in lower-quality mRNA data for a few replicates, so I chose to select the highest-quality dataset from each condition for functional gene analysis rather than using all replicates. In some treatments (nitrate-amended microoxic ^{13}C-heavy fractions, microoxic ^{12}C-light fractions), two equally high-quality datasets were available. While this lack of replication means that absolute counts of functional genes may be approached with some skepticism, interesting patterns of transcription under the different conditions are still readily apparent. Additionally, confidence can be gained by grouping transcripts into pathways or overall categories and considering abundance and enrichment at those levels.

4.3.1 Biodegradation processes under distinct electron acceptor availability

Addition of nitrate to microoxic contaminated sediments or wastewater has previously been shown to stimulate degradation of contaminants while reducing oxygen demand (Leahy & Olsen 1997; Ma & Love 2001; Liu et al. 2018; Durant et al. 1999). This was also apparent in the pattern of $^{13}CO_2$ release in this experiment. Nitrate enhanced the rate of toluene degradation in microoxic conditions (Fig. 26B). The initial rapid reduction of nitrate in both nitrate-amended microoxic and anoxic microcosms was probably due to metabolism of electron donors either from the R2A spike or intrinsic to the sediment. Reduced rates of nitrate reduction after 4 to 6 days, then, probably signified exhaustion of these alternative substrates. Nitrate reduction occurred at relatively constant rates regardless of oxygen availability in the nitrate-amended microoxic microcosms (Fig. 26C). However, a reduced demand for oxygen in microoxic conditions when nitrate was available, despite faster toluene degradation, was also apparent.

4.3.2 Degrader communities under distinct redox conditions

TEA availability clearly influenced microbial community composition, especially that of active toluene degraders (Fig. 28B). The main distinction between microoxic and nitrate-amended microoxic conditions appeared to be the higher abundance of *Comamonadaceae* in heavy labelled fractions under nitrate amendment (Fig. 28C). Previous studies have reported a high abundance of various Proteobacteria throughout the contaminant plume *in situ*, with *Pseudomonas* spp. primarily detectable at the upper fringe, and *Comamonadaceae* throughout the anoxic plume core (Larentis et al. 2013; Pilloni et al. 2019). In the current study, *Pseudomonas* spp. consistently dominated in fully oxic microcosms, but appeared not to be competitive for toluene degradation under microoxic conditions (Fig. 28A), especially when nitrate was present. This brings into question its relevance for biodegradation under TEA limitation *in situ*. At the same time, several *Pseudomonas* spp. have been widely considered as suitable model organisms for investigating aromatic hydrocarbon degradation under both aerobic and denitrifying conditions (Lalucat et al. 2006; Domingues et al. 2017; Philippot 2002), and even under microoxic conditions (Tribelli et al. 2018).

Instead, *Comamonadaceae* were identified as the major toluene degraders in microoxic microcosms, whether or not nitrate was amended (Fig. 28A). Within the *Comamonadaceae*, *Variovorax* and *Acidovorax* spp. are well known aromatic degraders (Posman et al. 2017; Rooney-Varga et al. 1999; Fahy et al. 2006), thus their importance in microoxic microcosms

was no surprise. Within the *Comamonadaceae*, transcripts of the entire KEGG denitrification module were enriched in ^{13}C in nitrate-amended microoxic conditions (Fig. 31). Considering that anaerobic toluene degradation was not observed, and the high ^{13}C enrichment of both *Comamonadaceae* SSU rRNA and oxygen-dependent aromatic activation transcripts, it seems that members of this family must have activated toluene with oxygen while respiring nitrate in my microcosms. The ability to degrade toluene during periods of oxygen limitation, reserving available oxygen for activation but not respiration, could convey a competitive advantage over other degraders (Wilson & Bouwer 1997) and might also explain the higher toluene degradation rate compared to microoxic microcosms without nitrate.

Lui et. al (Liu et al. (2018)) established a membrane biofilm reactor with just enough oxygen for aerobic benzene activation , while nitrate was available for respiration. Although it was inoculated with *Pseudomonas putida* F1, the final community was dominated primarily by *Comamonadaceae*. They predicted that phenol hydroxylase should be the primary means of oxygen-dependent aromatic activation based on PICrust extrapolation (Langille et al. 2013), but direct evidence was not provided. These present findings clearly demonstrate that phenol hydroxylase (*dmpN*) was much more active than toluene monooxygenase (*tmoA*) under nitrate-amended microoxic conditions (Fig. 32). Since transcription levels of *tmoA* and *dmpN* were approximately equal in microoxic treatments without nitrate, it is also possible that an inhibitory effect of nitrate, rather than low oxygen availability, could have triggered these expression patterns. These results fully support the conclusion that members of the *Comamonadaceae* employed an oxygen-activating, nitrate-respiring strategy for degradation of toluene in the investigated aquifer sediments. The similar composition of *Comamonadaceae* genera in all fractions (Fig. 29) suggests that this capacity may be conserved across different genera of the family, rather than being specific to one genus or species.

4.3.3 Lack of toluene degradation in anoxic treatments

Comamonadaceae also accounted for more than half of the rRNA in anoxic microcosms, in which toluene degradation was not active. Several genera known to harbor the benzylsuccinate synthase pathway (von Netzer et al. 2016) were present in SSU rRNA from the anoxic microcosm, notably *Azoarcus* spp., but only a negligible number of *bssA* transcripts were detected after 16 d of incubation. Denitrifying toluene degradation usually occurs readily in aquifer sediments (Lueders 2017). In the present study, despite the rapid onset of nitrate

reduction and the presence of denitrifying bacteria, the apparent lack of toluene degradation remains unexplained.

4.3.4 Patterns of ^{13}C-enrichment and abundance in catabolic and respiratory transcripts

Many of the enriched xenobiotic degradation pathways (Fig. 30) share enzymes relevant for toluene degradation. Although phenol hydroxylase can activate toluene (along with other aromatics) and was clearly detected in these microcosms (Fig. 32), it is not included in the toluene degradation or xylene degradation pathways of the KEGG release used in this study. This explains why toluene and xylene degradation were neither the most abundant nor most highly labelled pathways.

Aerobic denitrification is distinct from the oxygen-activating, nitrate-respiring strategy discussed above. Rather, it involves the simultaneous use of both oxygen and nitrate as TEAs within individual organisms in environments with low (Li et al. 2017) or fluctuating oxygen availability (Ji et al. 2015). The *napA* gene, which codes for the catalytic subunit of periplasmic nitrate reductase, has been used as a biomarker for aerobic denitrification (Marchant et al. 2017). I expected to find evidence of this metabolism in the nitrate-amended microoxic microcosms. However, transcripts of *narG*, coding for the catalytic subunit of membrane-bound nitrate reductase, were far more abundant than *napA* in all fractions from this condition (Fig. 32). Conversely, *napA* transcripts were more abundant in microoxic microcosms without nitrate and under fully anoxic conditions. Studies of nitrate reductase in *E. coli* strain K12 found that *napA* expression was high under nitrate limitation, but inhibited at nitrate concentrations above 2 mM (Potter et al. 1999; Wang et al. 1999). Since the microcosms were initially amended with 7.6 mM nitrate, this may explain the observed patterns.

The comparatively high abundance of transcripts for *ccoN*, which codes for a high-affinity cbb_3-type terminal oxidase (Pitcher & Watmough 2004), was unsurprising under microoxic conditions but unexpected in oxic microcosms, where low-affinity oxidases should be sufficient to sustain aerobic respiration. However, Morris & Schmidt (2013) have reported high affinity oxidases in two-thirds of the ~1000 bacterial genomes and most of the environmental metagenomes they investigated, suggesting that high-affinity terminal oxidases are far more prevalent than previously thought. Transcription of *ccoN* was much lower under nitrate-amended microoxic conditions than in oxic or nitrate-free microoxic conditions (Fig. 32).

Again, this supports the conclusion that nitrate reduction was used as a main mode of respiration in these treatments, reserving available oxygen for aromatic substrate activation.

4.3.5 Absence of nitric oxide dismutase transcripts

The *nod* gene was previously found in sediment samples from the Flingern aquifer site during initial *nod* assays in Work Package 1. The unexpected abundance of *tmoA* genes previously found in the anoxic plume core *in situ* (Larentis et al. 2013) also suggested that oxygenesis could take place there. I hypothesized that oxygenic bacteria would have an additional advantage in fluctuating oxygen conditions, beyond that shown by others operating under the oxygen-activation, nitrate-respiration strategy. By creating their own oxygen via dismutation, these bacteria could continue to use oxygen activation during periods of anoxia, without needing to change their metabolic machinery. This hypothesis proved untrue, at least in the incubation conditions used in the current study, since *nod* transcripts were absent in all fractions from all conditions. These results also call into question our hypothesis that oxygenesis contributes to hydrocarbon degradation in the anoxic or limited-TEA conditions common to contaminated sites. Very little is known about the mechanism of dismutation and whether it is employed by any groups outside the NC10 clade. Exploration of more sites and sediment incubation with a wider array of substrates and TEAs is necessary before discarding or expanding upon this hypothesis.

4.4 Synthesis and outlook

The overarching goal of this thesis was to investigate the adaptations of pollutant degraders to the interphase of aerobic and anaerobic conditions. This was tackled by (i) the development and application of a functional marker gene assay for NO dismutase genes, and (ii) by developing and applying cutting-edge improvements to the stable isotope analysis (SIP) technique to oxygen-limited sediment microcosms. Along with the assumed intra-aerobic degradation via NO-dismutation, aerobic denitrification and oxygen-activating, nitrate-respiring (OANR) degradation were of specific interest to this thesis work.

Using primers developed early in my thesis, I found evidence of *nod* genes in both the Siklós and Flingern contaminated aquifer sites, which were also the sources of sediments for later microcosm experiments. Colleagues and other research groups using the primers developed in this work have found *nod* genes in wastewater treatment plants and experimental water treatment bioreactors (Zhu et al. 2017), as well as in an alpine wetland (Zhang et al. 2018).

Based on the postitive PCR results from a range of environmental samples, including contaminated aquifer sediments, I also expected to find *nod* transcripts in the microcosm-based total-RNA-SIP projects presented here. Respective mRNA was indeed present in the microoxic Siklós sediment microcosms, though not in high numbers. The majority of these transcripts, which I could successfully identify as *nod* based on the characteristic substitutions and/or by their placement in alignments with other *nod*s and closely-related nitric oxide reductases (*nor*s), were mistaken for canonical *nor* in the KEGG database. This suggests that *nod* has already been captured in a wider range of environments by metagenomic or metatranscriptomic sequencing projects, but its prescence is obscured by misleading annotation in gene databases.

NO-distmutating bacteria could in fact be a subset of OANR strategists, which should have a competitive advantage in low-oxygen, nitrate-containing environments. One aim of the Flingern sediment microcosm experiment was to provide conditions that would enrich OANR organisms, perhaps including NO-dismutators. However, although I found clear evidence of OANR and identified the bacteria performing it in conjunction with toluene degradation, there were no transcripts of *nod* in the metatranscriptomic datasets. Thus, NO-dismuation could not be substantiated as a viable adaptation of degraders to limiting O_2-availability in this work. More suitable enrichment conditions and the identities of *nod*-containing bacteria outside of the NC10 phylum thus remain to be elaborated. Furthermore, whether the Nod protein coded for by *nod* is a functional enzyme that actually performs nitric oxide dismutation has yet to be biochemically proven. Protein isolation and testing has not yet been possible, but would be a great step forward in understanding the mechanism of dismutation and its impact on pollutant degradation and the global nitrogen cycle.

The two microcosm projects clearly show the impact of electron acceptor availability on the composition and activities, both catabolic and respiratory, of complex toluene-degrading aquifer sediment microbiota. *Rhodocyclaceae* dominated microoxic microcosms when the sediment was sourced from the Siklós aquifer, but were not at all abundant or competitive against the dominant *Comamonadaceae* and *Pseudomonadaceae* in microoxic microcosms from Flingern sediment. Despite this marked distinction in key populations, the expressed catabolic and respiratory functional transcripts showed great similarity, highlighting functional redundancy in contaminated environments and the importance of considering function rather than community composition when studying bioremediation potential. Phenol hydroxylase as the aromatic-activation monooxygenase and C23O as the ring-opening hydroxylase were both shown to play important roles in toluene degradations in microoxic conditions. Cell motility,

conversely, only appeared important when toluene was limited. Rather than discovering novel groups of hydrocarbon degrading microbes in these microcosms, the combination of taxonomic and functional data highlights previously unknown or under-recognized physiological adaptations of well-known degraders in limited-TEA conditions.

Total-RNA-SIP revealed the dominant role of diverse *Comamonadaceae* in oxygen-limited sediments when nitrate was available as a simultaneous TEA. These degraders readily switched respiration strategy, reducing their oxygen demand without forfeiting the mechanistic advantage of oxygen-dependent aromatic activation. For the first time, this directly demonstrates the long-hypothesized OANR degradation strategy for aromatic hydrocarbons in polluted aquifer sediments. Considering the frequent presence of *Comamonadaceae* at contaminated sites (as reported, for example, by Fahy et al. (2006), Singleton et al. (2009), Sun & Cupples (2012), Saidi-Mehrabad et al. (2013), and Posman et al. (2017)), OANR may be much more widespread and relevant to bioremediation than previously understood.

The power of total-RNA-SIP to reveal both taxonomic and functional information of degraders was instrumental to these findings. Environmental systems are seldom static and microbiomes almost never consist of a single taxonomic group, as is often the case in laboratory enrichments and pure cultures. To understand the roles and mechanisms of microbes as they relate to real-world situations, they must be studied as members of complex microbiota, ideally in *in situ*-like conditions. The advent of metagenomics and metatranscriptomics opened up a whole new way to explore microbial ecology, but untangling process-relevant data from complex communities remains a considerable challenge. Stable isotope probing can target microbes with relevant metabolic capabilities without relying on highly artificial conditions necessary for enrichment. mRNA-SIP, and the novel method of total-RNA-SIP, go one step further by providing data on the expression of functional transcripts by relevant organisms. This thesis built on recent advances in stable isotope probing and total-RNA sequencing to produce one of the first instances of total-RNA-SIP, and the only instance to date to make use of replication and control incubations to semi-quantitatively interpret labelling via enrichment factor calculations. Previous publications from our working group introduced the Enrichment Factor (EF_1) calculation to rank the amount of isotopic labelling in specific taxonomic groups, but the amplification-free nature of total-RNA-SIP and lack of heavy RNA in control microcosms (fed with ^{12}C substrate) prevented its use in my second microcosm experiment (and likely in many future projects). I overcame this methodological limitation by developing a new calculation (EF_2), which allows me and future users of total-RNA-SIP to understand the degree of labelling

of both taxonomic groups and functional transcripts. I also made use of annotations from multiple databases to connect functional transcripts with their taxonomic origin. These annotations, and indeed any interpretation of functional data, depend on the completeness and correctness of available gene databases. The high percentage of "unknown" transcripts, as well as the observed misannotation of *nod* sequences, show that work remains to be done in this area, especially pertaining to 'omics information from environmental samples.

The total amounts of RNA recovered from gradient fractions and the net costs of transcriptome sequencing will remain a major limitation of total-RNA-SIP work in the future. I tested two possible methods to overcome these limitations. Linear amplification does greatly increase RNA yield, but introduces biases and does not reduce the necessary depth of sequencing. Targeted depletion of 23S rRNA may be beneficial in opening up sequencing space for 16S and mRNA within a given depth, but this unfortunately could not be accomplished within the course of this work. It is possible that this depletion would also introduce an unacceptable amount of bias, but this can only be revealed by successful depletion and comparative sequencing in the future.

My findings in the course of this thesis should be of significant interest to two overlapping groups of colleagues and stakeholders: those seeking to understand and improve bioremediation in the many anthropogenically contaminated sites worldwide, and those who wish to implement total-RNA-SIP to their own systems of interest, whether environmental, medical, or beyond.

References

Abercron, M.-V. et al., 2017. Naphthalene biodegradation under oxygen-limiting conditions: community dynamics and the relevance of biofilm-forming capacity. *Microb. Biotechnol.*, 10(6), pp.1781–1796.

Alfreider, A. & Vogt, C., 2007. Bacterial diversity and aerobic biodegradation potential in a BTEX-contaminated aquifer. *Water, Air, & Soil Pollution*, 183(1-4), pp.415–426.

Alharbi, O.M. et al., 2018. Health and environmental effects of persistent organic pollutants. *J. Mol. Liq*, 263, p.4420453.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215(3), pp.403–410.

Anneser, B. et al., 2010. High resolution analysis of contaminated aquifer sediments and groundwater—what can be learned in terms of natural attenuation? *Geomicrobiol. J.*, 27(2), pp.130–142.

Anneser, B. et al., 2008. High-resolution monitoring of biogeochemical gradients in a tar oil-contaminated aquifer. *Appl. Geochem.*, 23(6), pp.1715–1730.

Azubuike, C.C., Chikere, C.B. & Okpokwasili, G.C., 2016. Bioremediation techniques-classification based on site of application: principles, advantages, limitations and prospects. *World J. Microbiol. Biotechnol.*, 32(11), p.180.

Bauer, R.D. et al., 2008. Mixing-controlled biodegradation in a toluene plume—results from two-dimensional laboratory experiments. *J. Contam. Hydrol.*, 96(1-4), pp.150–168.

Belasco, J.G., 2010. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell Biol*, 11(7), pp.467–478.

Benedek, T. et al., 2016. Polyphasic analysis of an Azoarcus-Leptothrix-dominated bacterial biofilm developed on stainless steel surface in a gasoline-contaminated hypoxic groundwater. *Environ. Sci. Pollut. Res. Int.*, 23(9), pp.9019–9035.

Bhattacharjee, A.S. et al., 2016. Methane dependent denitrification-from ecosystem to laboratory-scale enrichment for engineering applications. *Water Res.*, 99, pp.244–252.

Biegert, T., Fuchs, G. & Heider, J., 1996. Evidence that anaerobic oxidation of toluene in the denitrifying bacterium Thauera aromatica is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur. J. Biochem.*, 238(3), pp.661–668.

Boden, R., Hutt, L. P., and Rae, A. W., 2017. Reclassification of Thiobacillus aquaesulis (Wood & Kelly, 1995) as Annwoodia aquaesulis gen. nov., comb. nov., transfer of Thiobacillus (Beijerinck, 1904) from the Hydrogenophilales to the Nitrosomonadales, proposal of Hydrogenophilalia class. nov. within the “Proteobacteria”, and four new families within the orders Nitrosomonadales and Rhodocyclales. *International Journal of Systematic and Evolutionary Microbiology* 67, pp.1191–1205.

Bokulich, N.A. et al., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods*, 10(1), p.57.

Boll, M., Fuchs, G. & Heider, J., 2002. Anaerobic oxidation of aromatic compounds and hydrocarbons. *Curr. Opin. Chem. Biol.*, 6(5), pp.604–611.

Bouhajja, E. et al., 2017. Identification of novel toluene monooxygenase genes in a hydrocarbon-polluted sediment using sequence-and function-based screening of metagenomic libraries. *Appl. Microbiol. Biotechnol.*, 101(2), pp.797–808.

Bradford, L. et al., 2018. Transcriptome-stable isotope probing provides targeted functional and taxonomic insights into microaerobic pollutant-degrading aquifer microbiota. *Front. Microbiol.*, 9, p.2696.

Britton, L.N., 1989. *Aerobic denitrification as an innovative method for in-situ biological remediation of contaminated subsurface sites*,

Brooks, J.P. et al., 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.*, 15(1), p.66.

Buchfink, B., Xie, C. & Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 12(1), pp.59–60.

Callaghan, A.V., 2013. Metabolomic investigations of anaerobic hydrocarbon-impacted environments. *Curr. Opin. Biotechnol.*, 24(3), pp.506–515.

Cao, F.-L. et al., 2010. An optimized RNA amplification method for prokaryotic expression profiling analysis. *Appl. Microbiol. Biotechnol.*, 87(1), pp.343–352.

Caporaso, J.G. et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7(5), pp.335–336.

Carmona, M. et al., 2009. Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *MMBR*, 73(1), pp.71–133.

Carpenter, E.P. et al., 2008. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.*, 18(5), pp.581–586.

Carvalhais, L.C. et al., 2012. Application of metatranscriptomics to soil environments. *J. Microbiol. Methods*, 91(2), pp.246–251.

Cavalca, L., Dell'Amico, E. & Andreoni, V., 2004. Intrinsic bioremediability of an aromatic hydrocarbon-polluted groundwater: diversity of bacterial population and toluene monoxygenase genes. *Appl. Microbiol. Biotechnol.*, 64(4), pp.576–587.

Chakraborty, R. et al., 2005. Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene compounds by Dechloromonas strain RCB. *Appl. Environ. Microbiol.*, 71(12), pp.8649–8655.

Chayabutra, C. & Ju, L.-K., 2000. Degradation of n-hexadecane and its metabolites by Pseudomonas aeruginosa under microaerobic and anaerobic denitrifying conditions. *Appl. Environ. Microbiol.*, 66(2), pp.493–498.

Chiang, C. et al., 1989. Aerobic biodegradation of benzene, toluene, and xylene in a sandy aquifer—data analysis and computer modeling. *Groundwater*, 27(6), pp.823–834.

Christensen, T.H. et al., 2000. Characterization of redox conditions in groundwater contaminant plumes. *J. Contam. Hydrol.*, 45(3), pp.165–241.

Das, N. & Chandran, P., 2011. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol. Res. Int.*, 2011.

Degrelle, S.A. et al., 2008. Amplification biases: possible differences among deviating gene expressions. *BMC Genomics*, 9(1), p.46.

Deutscher, M.P., 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.*, 34(2), pp.659–666.

Dew, W.A. et al., 2015. Biological effects and toxicity of diluted bitumen and its constituents in freshwater systems. *J. Appl. Toxicol.*, 35(11), pp.1219–1227.

Van Dijk, E.L. et al., 2018. The Third Revolution in Sequencing Technology. *Trends Genet.*

Van Dijk, E.L., Jaszczyszyn, Y. & Thermes, C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.*, 322(1), pp.12–20.

Domingues, P.M. et al., 2017. Bacterial production of biosurfactants under microaerobic and anaerobic conditions. *Rev. Environ. Sci. Bio*, 16(2), pp.239–272.

Dumont, M.G. et al., 2011. DNA-, rRNA-and mRNA-based stable isotope probing of aerobic methanotrophs in lake sediment. *Environ. Microbiol.*, 13(5), pp.1153–1167.

Dumont, M.G., Pommerenke, B. & Casper, P., 2013. Using stable isotope probing to obtain a targeted metatranscriptome of aerobic methanotrophs in lake sediment. *Environ. Microbiol. Rep.*, 5(5), pp.757–764.

Durant, L.W., D'adamo, P. & Bouwer, E., 1999. Aromatic Hydrocarbon Biodegradation with Mixtures of O2 and NO3- as Electron Acceptors. *Environ. Eng. Sci.*, 16(6), pp.487–500.

Eddy, S.R., 1998. Profile Hidden Markov Models. *Bioinformatics*, 14(9), pp.755–763.

Eltis, L.D. & Bolin, J.T., 1996. Evolutionary relationships among extradiol dioxygenases. *J. Bacteriol.*, 178(20), pp.5930–5937.

Espinola, F. et al., 2018. Metagenomic analysis of subtidal sediments from polar and subpolar coastal environments highlights the relevance of anaerobic hydrocarbon degradation processes. *Microb. Ecol.*, 75(1), pp.123–139.

Ettwig, K.F. et al., 2012. Bacterial oxygen production in the dark. *Front. Microbiol.*, 3(273), pp.1–8.

Ettwig, K.F. et al., 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature*, 464(7288), pp.543–548.

Fahrenfeld, N. et al., 2014. Insights into biodegradation through depth-resolved microbial community functional and structural profiling of a crude-oil contaminant plume. *Microbial ecology*, 68(3), pp.453–462.

Fahy, A. et al., 2006. Heterogeneous aerobic benzene-degrading communities in oxygen-depleted groundwaters. *FEMS Microbiol. Ecol.*, 58(2), pp.260–270.

Farkas, M. et al., 2016. Enrichment of dissimilatory Fe (III)-reducing bacteria from groundwater of the Siklós BTEX-contaminated site (Hungary). *Folia Microbiol.*, 62(1), pp.63–71.

Farkas, M. et al., 2015. Zoogloea oleivorans sp. nov., a floc-forming, petroleum hydrocarbon-degrading bacterium isolated from biofilm. *Int. J. Syst. Evol. Microbiol.*, 65(1), pp.274–279.

Fayemiwo, O., Daramola, M. & Moothi, K., 2017. BTEX compounds in water-future trends and directions for water treatment. *Water SA*, 43(4), pp.602–613.

Fleeger, J.W., Carman, K.R. & Nisbet, R.M., 2003. Indirect effects of contaminants in aquatic ecosystems. *Sci. Total Environ.*, 317(1-3), pp.207–233.

Fortunato, C.S. & Huber, J.A., 2016. Coupled RNA-SIP and metatranscriptomics of active chemolithoautotrophic communities at a deep-sea hydrothermal vent. *ISME J.*, 10(8), pp.1925–1938.

Fuchs, G., Boll, M. & Heider, J., 2011. Microbial degradation of aromatic compounds—from one strategy to four. *Nat. Rev. Microbiol.*, 9(11), pp.803–816.

Galperin, M.Y. et al., 2014. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.*, 43(D1), pp.D261–D269.

Garalde, D.R. et al., 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods*.

Di Gennaro, P., Bargna, A. & Sello, G., 2011. Microbial enzymes for aromatic compound hydroxylation. *Appl. Microbiol. Biotechnol.*, 90(6), pp.1817–1827.

Van Ginkel, C. et al., 1996. Purification and characterization of chlorite dismutase: a novel oxygen-generating enzyme. *Arch. Microbiol.*, 166(5), pp.321–326.

Glaubitz, S. et al., 2009. 13C-isotope analyses reveal that chemolithoautotrophic Gamma-and Epsilonproteobacteria feed a microbial food web in a pelagic redoxcline of the central Baltic Sea. *Environ. Microbiol.*, 11(2), pp.326–337.

Godwin, S. et al., 2014. Investigation of the microbial metabolism of carbon dioxide and hydrogen in the kangaroo foregut by stable isotope probing. *ISME J.*

Griebler, C. & Lueders, T., 2009. Microbial biodiversity in groundwater ecosystems. *Freshwater Biol.*, 54(4), pp.649–677.

Grösbacher, M. et al., 2016. Organic contamination versus mineral properties: competing selective forces shaping bacterial community assembly in aquifer sediments. *Aquat. Microb. Ecol.*, 76(3), pp.243–255.

Gülensoy, N. & Alvarez, P.J., 1999. Diversity and correlation of specific aromatic hydrocarbon biodegradation capabilities. *Biodegradation*, 10(5), pp.331–340.

Gutierrez-Zamora, M.-L. & Manefield, M., 2010. An appraisal of methods for linking environmental processes to specific microbial taxa. *Rev. Environ. Sci. Bio*, 9(2), pp.153–185.

Hansen, M.A. et al., 2008. Biopieces: a bioinformatics toolset and framework. In *19th International Conference on Genome Informatics*. Gold Coast, Australia.

Hayaishi, O. & Hashimoto, K., 1950. Pyrocatecase a new enzyme catalizing oxidative breakdown of pyrocatechin. *The Journal of Biochemistry*, 37(3), pp.371–374.

He, S. et al., 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods*, 7(10), pp.807–812.

Hendrickx, B., Junca, H., et al., 2006. Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *J. Microbiol. Methods*, 64(2), pp.250–265.

Hendrickx, B. et al., 2005. Dynamics of an oligotrophic bacterial aquifer community during contact with a groundwater plume contaminated with benzene, toluene, ethylbenzene, and xylenes: an in situ mesocosm study. *Appl. Environ. Microbiol.*, 71(7), pp.3815–3825.

Hendrickx, B., Dejonghe, W., et al., 2006. PCR-DGGE method to assess the diversity of BTEX mono-oxygenase genes at contaminated sites. *FEMS Microbiol. Ecol.*, 55(2), pp.262–273.

Holland, K.T., 2013. Anaerobic bacteria. In Springer Science & Business Media.

Hu, Q.-Q. et al., 2019. High microbial diversity of the nitric oxide dismutation reaction revealed by PCR amplification and analysis of the nod gene. *Int. Biodeterior. Biodegradation*, 143, p.104708.

Huang, W.E. et al., 2009. Resolving genetic functions within microbial populations: in situ analyses using rRNA and mRNA stable isotope probing coupled with single-cell Raman-fluorescence in situ hybridization. *Appl. Environ. Microbiol.*, 75(1), pp.234–241.

Hungate, B.A. et al., 2015. Quantitative microbial ecology through stable isotope probing. *Appl. Environ. Microbiol.*, 81(21), pp.7570–7581.

Huson, D.H. et al., 2007. MEGAN analysis of metagenomic data. *Genome Res.*, 17(3), pp.377–386.

Hyatt, D. et al., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), p.119.

Jechalke, S. et al., 2013. Analysis of structure, function, and activity of a benzene-degrading microbial community. *FEMS Microbiol. Ecol.*, 85(1), pp.14–26.

Ji, B. et al., 2015. Aerobic denitrification: A review of important advances of the last 30 years. *Biotechnol. Bioprocess Eng.*, 20(4), pp.643–651.

Jiang, H. et al., 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, 15(1), p.182.

Kanehisa, M. et al., 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, 44(D1), pp.D457–D462.

Khomenkov, V. et al., 2008. Organization of metabolic pathways and molecular-genetic mechanisms of xenobiotic degradation in microorganisms: A review. *Appl. Biochem. Microbiol.*, 44(2), pp.117–135.

Kim, H. & Jaffé, P.R., 2008. Degradation of toluene by a mixed population of archetypal aerobes, microaerophiles, and denitrifiers: laboratory sand column experiment and multispecies biofilm model formulation. *Biotechnol. Bioeng.*, 99(2), pp.290–301.

Kojima, Y., Itada, N. & Hayaishi, O., 1961. Metapyrocatechase: a new catechol-cleaving enzyme. *J. Biol. Chem.*, 236(8), pp.2223–2228.

Kopylova, E., Noé, L. & Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24), pp.3211–3217.

Kramer, S. et al., 2016. Resource partitioning between bacteria, fungi, and protists in the detritusphere of an agricultural soil. *Front. Microbiol.*, 7(1524), pp.1–12.

Krell, T. et al., 2013. Bioavailability of pollutants and chemotaxis. *Curr. Opin. Biotechnol.*, 24(3), pp.451–456.

Kukor, J.J. & Olsen, R.H., 1996. Catechol 2, 3-dioxygenases functional in oxygen-limited (hypoxic) environments. *Appl. Environ. Microbiol.*, 62(5), pp.1728–1740.

Abu Laban, N. et al., 2010. Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environmental microbiology*, 12(10), pp.2783–2796.

Lacal, J., 2017. The Potential of Hydrocarbon Chemotaxis to Increase Bioavailability and Biodegradation Efficiency. In Krell, T., ed. *Cellular Ecophysiology of Microbes*. Springer

International Publishing, pp. 1–14.

Lalucat, J. et al., 2006. Biology of Pseudomonas stutzeri. *Microbiol. Mol. Biol. Rev.*, 70(2), pp.510–547.

Langille, M.G. et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.*, 31(9), p.814.

Larentis, M., Hoermann, K. & Lueders, T., 2013. Fine-scale degrader community profiling over an aerobic/anaerobic redox gradient in a toluene-contaminated aquifer. *Environ. Microbiol. Rep.*, 5(2), pp.225–234.

Laursen, M.F., Dalgaard, M.D. & Bahl, M.I., 2017. Genomic GC-Content Affects the Accuracy of 16S rRNA Gene Sequencing Based Microbial Profiling due to PCR Bias. *Front. Microbiol.*, 8, p.1934.

Leahy, J.G., Batchelor, P.J. & Morcomb, S.M., 2003. Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.*, 27(4), pp.449–479.

Leahy, J.G. & Olsen, R.H., 1997. Kinetics of toluene degradation by toluene-oxidizing bacteria as a function of oxygen concentration, and the effect of nitrate. *FEMS Microbiol. Ecol.*, 23(1), pp.23–30.

Levesque-Sergerie, J.-P. et al., 2007. Detection limits of several commercial reverse transcriptase enzymes: impact on the low-and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Mol. Biol.*, 8(1), p.93.

Li, H., Wang, W. & Huang, S., 2017. Aerobic Denitrifiers: Characteristics and Recent Research. In *MSMEE 2017*.

Lindgreen, S., 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes*, 5(1), p.337.

Liu, Z. et al., 2018. Accurate O2 delivery enabled benzene biodegradation through aerobic activation followed by denitrification-coupled mineralization. *Biotechnol. Bioeng.*

Lodish, H. et al., 2000. *Molecular Cell Biology*, *4th edition*, W.H. Freeman (New York).

Lueders, T., 2015. DNA-and RNA-based stable isotope probing of hydrocarbon degraders. In *Hydrocarbon and Lipid Microbiology Protocols*. Humana Press, pp. 181–197.

Lueders, T. et al., 2016. RNA-stable isotope probing: from carbon flow within key microbiota to targeted transcriptomes. *Curr. Opin. Biotechnol.*, 41, pp.83–89.

Lueders, T., 2017. The ecology of anaerobic degraders of BTEX hydrocarbons in aquifers. *FEMS Microbiol. Ecol.*, 93(1).

Lueders, T., Manefield, M. & Friedrich, M.W., 2004. Enhanced sensitivity of DNA-and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic

centrifugation gradients. *Environ. Microbiol.*, 6(1), pp.73–78.

Ma, G. & Love, N.G., 2001. BTX biodegradation in activated sludge under multiple redox conditions. *J. Env. Eng.*, 127(6), pp.509–516.

Manefield, M. et al., 2002. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.*, 68(11), pp.5367–5373.

Marchant, H.K. et al., 2017. Denitrifying community in coastal sediments performs aerobic and anaerobic respiration simultaneously. *ISME J.*

Marreiros, B.C. et al., 2016. Type II NADH: quinone oxidoreductase family: phylogenetic distribution, structural diversity and evolutionary divergences. *Environ. Microbiol.*, 18(12), pp.4697–4709.

Martinez-Lavanchy, P.M. et al., 2010. High stability and fast recovery of expression of the TOL plasmid-carried toluene catabolism genes of Pseudomonas putida mt-2 under conditions of oxygen limitation and oscillation. *Appl. Environ. Microbiol.*, 76(20), pp.6715–6723.

Maszenan, A. et al., 2002. Quadricoccus australiensis gen. nov., sp. nov., a beta-proteobacterium from activated sludge biomass. *Int. J. Syst. Evol. Microbiol.*, 52(1), pp.223–228.

Matsumoto, Y. et al., 2012. Crystal structure of quinol-dependent nitric oxide reductase from Geobacillus stearothermophilus. *Nat. Struct. Mol. Biol.*, 19(2), p.238.

McMurdie, P.J. & Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), p.e61217.

McNaught, A., Wilkinson, A. & others, 1997. Compendium of chemical terminology. IUPAC recommendations.

Meckenstock, R.U. et al., 2015. Biodegradation: Updating the concepts of control for microbial clean-up in contaminated aquifers. *Environ. Sci. Technol.*

Mehboob, F. et al., 2016. Genome and proteome analysis of P seudomonas chloritidismutans AW-1 T that grows on n-decane with chlorate or oxygen as electron acceptor. *Environ. Microbiol.*, 18(10), pp.3247–3257.

Mehboob, F. et al., 2009. Growth of Pseudomonas chloritidismutans AW-1T on n-alkanes with chlorate as electron acceptor. *Appl. Microbiol. Biotechnol.*, 83(4), pp.739–747.

Morris, R.L. & Schmidt, T.M., 2013. Shallow breathing: bacterial life at low O 2. *Nat. Rev. Microbiol.*, 11(3), p.205.

Müller, H., 2017. *Long-distance electron transfer by cable bacteria in aquifer sediments*.

Nebe, J. et al., 2009. Quantification of aromatic oxygenase genes to evaluate enhanced bioremediation by oxygen releasing materials at a gasoline-contaminated site. *Environ.*

Sci. Technol., 43(6), pp.2029–2034.

Nestler, H. et al., 2007. Biodegradation of chlorobenzene under hypoxic and mixed hypoxic-denitrifying conditions. *Biodegradation*, 18(6), pp.755–767.

Von Netzer, F. et al., 2013. Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl. Environ. Microbiol.*, 79(2), pp.543–552.

Von Netzer, F. et al., 2016. Functional gene markers for fumarate-adding and dearomatizing key enzymes in anaerobic aromatic hydrocarbon degradation in terrestrial environments. *J. Mol. Microbiol. Biotechnol.*, 26(1-3), pp.180–194.

Ning, Z. et al., 2018. Spatial Pattern of Bacterial Community Diversity Formed in Different Groundwater Field Corresponding to Electron Donors and Acceptors Distributions at a Petroleum-Contaminated Site. *Water*, 10(7), p.842.

Njobuenwu, D.O., Amadi, S.A. & Ukpaka, P.C., 2005. Dissolution rate of BTEX contaminants in water. *Can. J. Chem. Eng.*, 83(6), pp.985–989.

Nygaard, V. et al., 2003. Effects of mRNA amplification on gene expression ratios in cDNA experiments estimated by analysis of variance. *BMC Genomics*, 4(1), p.11.

Ogram, A. et al., 1995. Isolation and characterization of RNA from low-biomass deep-subsurface sediments. *Appl. Environ. Microbiol.*, 61(2), pp.763–768.

Padilla, C.C. et al., 2016. NC10 bacteria in marine oxygen minimum zones. *ISME J.*

Parales, R. et al., 2008. Diversity of microbial toluene degradation pathways. In *Adv. Appl. Microbiol.* Elsevier, pp. 1–73.

Parales, R.E., Ditty, J.L. & Harwood, C.S., 2000. Toluene-degrading bacteria are chemotactic towards the environmental pollutants benzene, toluene, and trichloroethylene. *Appl. Environ. Microbiol.*, 66(9), pp.4098–4104.

Pepe-Ranney, C. et al., 2016. Unearthing the ecology of soil microorganisms using a high resolution DNA-SIP approach to explore cellulose and xylose metabolism in soil. *Front. Microbiol.*, 7, p.703.

Petrova, O.E. et al., 2017. Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Scientific Reports*, 7, p.41114.

Philippot, L., 2002. Denitrifying genes in bacterial and archaeal genomes. *Biochim. Biophys. Acta, Gene Struct. Expression*, 1577(3), pp.355–376.

Pilloni, G. et al., 2019. Dynamics of Hydrology and Anaerobic Hydrocarbon Degrader Communities in A Tar-Oil Contaminated Aquifer. *Microorganisms*, 7(2), p.46.

Pilloni, G. et al., 2012. Testing the limits of 454 pyrotag sequencing: reproducibility, quantitative assessment and comparison to T-RFLP fingerprinting of aquifer microbes. *PloS one*, 7(7), p.e40467.

Pitcher, R.S. & Watmough, N.J., 2004. The bacterial cytochrome cbb3 oxidases. *Biochim. Biophys. Acta, Bioenerg.*, 1655, pp.388–399.

Poretsky, R. et al., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, 9(4), p.e93827.

Porter, A.W. & Young, L.Y., 2013. The bamA gene for anaerobic ring fission is widely distributed in the environment. *Front. Microbiol.*, 4, p.302.

Posman, K.M., DeRito, C.M. & Madsen, E.L., 2017. Benzene degradation by a Variovorax species within a coal tar-contaminated groundwater microbial community. *Appl. Environ. Microbiol.*, 83(4), pp.e02658–16.

Potter, L.C. et al., 1999. Competition between Escherichia coli strains expressing either a periplasmic or a membrane-bound nitrate reductase: does Nap confer a selective advantage during nitrate-limited growth? *Biochem. J.*, 344(Pt 1), p.77.

Pratscher, J., Dumont, M.G. & Conrad, R., 2011. Assimilation of acetate by the putative atmospheric methane oxidizers belonging to the USCalpha clade. *Environ. Microbiol.*, 13(10), pp.2692–2701.

Qin, X. et al., 2003. Use of real-time PCR with multiple targets to identify Pseudomonas aeruginosa and other nonfermenting gram-negative bacilli from patients with cystic fibrosis. *J. Clin. Microbiol.*, 41(9), pp.4312–4317.

Quast, C. et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, 41(D1), pp.D590–D596.

R Core Team, 2013. *R: A Language and Environment for Statistical Computing*, Vienna, Austria. Available at: https://www.R-project.org/.

Radax, R. et al., 2012. Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community. *Environ. Microbiol.*, 14(5), pp.1308–1324.

Ramesh, A. et al., 2011. Global environmental distribution and human health effects of polycyclic aromatic hydrocarbons. *Global contamination trends of persistent organic chemicals*, pp.95–124.

Reinhold-Hurek, B. & Hurek, T., 2000. Reassessment of the taxonomic structure of the diazotrophic genus Azoarcus sensu lato and description of three new genera and new species, Azovibrio restrictus gen. nov., sp. nov., Azospira oryzae gen. nov., sp. nov. and Azonexus fungiphilus gen. nov., sp. nov. *Int. J. Syst. Evol. Microbiol.*, 50(2), pp.649–659.

Rickwood, D., 1992. Preparative Centrifugation: A Practical Approach. In Rickwood, D., ed. Oxford, UK: Oxford University Press, Oxford, UK, pp. 143–186.

Rikken, G., Kroon, A. & Van Ginkel, C., 1996. Transformation of (per) chlorate into chloride by a newly isolated bacterium: reduction and dismutation. *Appl. Microbiol. Biotechnol.*, 45(3), pp.420–426.

Rooney-Varga, J.N. et al., 1999. Microbial communities associated with anaerobic benzene degradation in a petroleum-contaminated aquifer. *Appl. Environ. Microbiol.*, 65(7), pp.3056–3063.

Saidi-Mehrabad, A. et al., 2013. Methanotrophic bacteria in oilsands tailings ponds of northern Alberta. *ISME J.*, 7(5), p.908.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *PNAS*, 74(12), pp.5463–5467.

Savli, H. et al., 2003. Expression stability of six housekeeping genes: a proposal for resistance gene quantification studies of Pseudomonas aeruginosa by real-time quantitative RT-PCR. *J. Med. Microbiol.*, 52(5), pp.403–408.

Schaffner, I. et al., 2015. Mechanism of chlorite degradation to chloride and dioxygen by the enzyme chlorite dismutase. *Arch. Biochem. Biophys.*, 574, pp.18–26.

Schmieder, R. & Edwards, R., 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One*, 6(3), p.e17288.

Schmitt, M.E., Brown, T.A. & Trumpower, B.L., 1990. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucleic Acids Res.*, 18(10), pp.3091–3092.

Schwab, C. et al., 2014. Gene expression of lactobacilli in murine forestomach biofilms. *Microb. Biotechnol.*, 7(4), pp.347–359.

Schwartz, E. et al., 2016. Stable isotope probing with 18 O-water to investigate microbial growth and death in environmental samples. *Curr. Opin. Biotechnol.*, 41, pp.14–18.

Schwarzenbach, R.P. et al., 2010. Global water pollution and human health. *Annu. Rev. Environ. Resour.*, 35, pp.109–136.

Sentchilo, V.S. et al., 2000. Molecular diversity of plasmids bearing genes that encode toluene and xylene metabolism in Pseudomonas strains isolated from different contaminated sites in Belarus. *Appl. Environ. Microbiol.*, 66(7), pp.2842–2852.

Da Silva, M.L.B. & Corseuil, H.X., 2012. Groundwater microbial analysis to assess enhanced BTEX biodegradation by nitrate injection at a gasohol-contaminated site. *Int. Biodeterior. Biodegradation*, 67, pp.21–27.

Singleton, D.R., Ramirez, L.G. & Aitken, M.D., 2009. Characterization of a polycyclic aromatic hydrocarbon degradation gene cluster in a phenanthrene-degrading Acidovorax

strain. *Appl. Environ. Microbiol.*, 75(9), pp.2613–2620.

Smith, M.A. et al., 2000. Characteristics of the aerobic respiratory chains of the microaerophiles Campylobacter jejuni and Helicobacter pylori. *Arch. Microbiol.*, 174(1-2), pp.1–10.

Song, B. et al., 2001. Characterization of halobenzoate-degrading, denitrifying Azoarcus and Thauera isolates and description of Thauera chlorobenzoica sp. nov. *Int. J. Syst. Evol. Microbiol.*, 51(2), pp.589–602.

Sperfeld, M. et al., 2018. Microbial community of a gasworks aquifer and identification of nitrate-reducing Azoarcus and Georgfuchsia as key players in BTEX degradation. *Water Res.*, 132, pp.146–157.

Spiro, S., Roberts, R. & Guest, J., 1989. FNR-dependent repression of the ndh gene of Escherichia coli and metal ion requirement for FNR-regulated gene expression. *Mol. Microbiol.*, 3(5), pp.601–608.

Stewart, F.J., Ottesen, E.A. & DeLong, E.F., 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.*, 4(7), pp.896–907.

Sun, W. & Cupples, A.M., 2012. Diversity of five anaerobic toluene-degrading microbial communities investigated using stable isotope probing. *Appl. Environ. Microbiol.*, 78(4), pp.972–980.

Sun, W., Sun, X. & Cupples, A.M., 2014. Presence, diversity and enumeration of functional genes (bssA and bamA) relating to toluene degradation across a range of redox conditions and inoculum sources. *Biodegradation*, 25(2), pp.189–203.

Tan, B. et al., 2013. Metagenomic analysis of an anaerobic alkane-degrading microbial culture: potential hydrocarbon-activating pathways and inferred roles of community members. *Genome*, 56(10), pp.599–611.

Táncsics, A. et al., 2010. Investigation of catechol 2, 3-dioxygenase and 16S rRNA gene diversity in hypoxic, petroleum hydrocarbon contaminated groundwater. *Syst. Appl. Microbiol.*, 33(7), pp.398–406.

Táncsics, A. et al., 2013. One-year monitoring of meta-cleavage dioxygenase gene expression and microbial community dynamics reveals the relevance of subfamily I. 2. C extradiol dioxygenases in hypoxic, BTEX-contaminated groundwater. *Syst. Appl. Microbiol.*, 36(5), pp.339–350.

Táncsics, A. et al., 2012. Quantification of subfamily I. 2. C catechol 2, 3-dioxygenase mRNA transcripts in groundwater samples of an oxygen-limited BTEX-contaminated site. *Environ. Sci. Technol.*, 46(1), pp.232–240.

Táncsics, A. et al., 2018. Stable isotope probing of hypoxic toluene degradation at the Siklós aquifer reveals prominent role of Rhodocyclaceae. *FEMS Microbiol. Ecol.*, 94(6),

p.fiy088.

Tannock, G.W. et al., 2014. RNA-Stable-Isotope Probing Shows Utilization of Carbon from Inulin by Specific Bacterial Populations in the Rat Large Bowel. *Appl. Environ. Microbiol.*, 80(7), pp.2240–2247.

Terry, M., 2019. Top 10 Gene Sequencing Companies by Revenue. Available at: https://www.biospace.com/article/top-10-gene-sequencing-companies-by-revenue/ [Accessed January 15, 2020].

Tiehm, A. & Schulze, S., 2003. Intrinsic aromatic hydrocarbon biodegradation for groundwater remediation. *Oil Gas Sci. Technol.*, 58(4), pp.449–462.

Tindall, B. & Euzéby, J., 2006. Proposal of Parvimonas gen. nov. and Quatrionicoccus gen. nov. as replacements for the illegitimate, prokaryotic, generic names Micromonas Murdoch and Shah 2000 and Quadricoccus Maszenan et al. 2002, respectively. *Int. J. Syst. Evol. Microbiol.*, 56(11), pp.2711–2713.

Trembath-Reichert, E., Butterfield, D.A. & Huber, J.A., 2019. Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *ISME J.*, p.1.

Tremblay, J. et al., 2015. Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.*, 6.

Tribelli, P.M. et al., 2018. Microaerophilic alkane degradation in Pseudomonas extremaustralis: a transcriptomic and physiological approach. *J. Ind. Microbiol. Biotechnol.*, 45(1), pp.15–23.

Uhlik, O. et al., 2013. Stable isotope probing in the metagenomics era: a bridge towards bioremediation. *Biotechnol. Adv.*, 31(2), pp.154–165.

Unden, G., 1999. Biology of the Prokaryotes. In Lengeler, Joseph W and Drews, Gerhart and Schlegel, Hans Günter, ed. Georg Thieme Verlag.

Urich, T. et al., 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One*, 3(6), pp.e2527–e2527.

Vaillancourt, F.H., Bolin, J.T. & Eltis, L.D., 2006. The ins and outs of ring-cleaving dioxygenases. *Crit. Rev. Biochem. Mol. Biol.*, 41(4), pp.241–267.

Wagner, G.P., Kin, K. & Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, 131(4), pp.281–285.

Wald, J. et al., 2015. Pseudomonads rule degradation of polyaromatic hydrocarbons in aerated sediment. *Front. Microbiol.*, 6, p.1268.

Wang, H., Tseng, C.-P. & Gunsalus, R.P., 1999. The napF and narG nitrate reductase operons in Escherichia coli are differentially expressed in response to submicromolar

concentrations of nitrate but not nitrite. *J. Bacteriol.*, 181(17), pp.5303–5308.

Wang, Q. et al., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.*, 73(16), pp.5261–5267.

Weelink, S.A. et al., 2008. Isolation and characterization of Alicycliphilus denitrificans strain BC, which grows on benzene with chlorate as the electron acceptor. *Appl. Environ. Microbiol.*, 74(21), pp.6672–6681.

Whiteley, A.S. et al., 2007. RNA stable-isotope probing. *Nat. Protoc.*, 2(4), pp.838–844.

Widdel, F., 1980. *Anaerober Abbau von Fettsäuren und Benzoesäure durch neu isolierte Arten sulfat-reduzierender Bakterien.*

Wilson, L. & Bouwer, E., 1997. Biodegradation of aromatic compounds under mixed oxygen/denitrifying conditions: a review. *J. Ind. Microbiol. Biotechnol.*, 18(2-3), pp.116–130.

Winderl, C. et al., 2008. Depth-resolved quantification of anaerobic toluene degraders and aquifer microbial community patterns in distinct redox zones of a tar oil contaminant plume. *Appl. Environ. Microbiol.*, 74(3), pp.792–801.

Winderl, C. et al., 2010. DNA-SIP identifies sulfate-reducing Clostridia as important toluene degraders in tar-oil-contaminated aquifer sediment. *ISME J.*, 4(10), pp.1314–1325.

Winderl, C., Schaefer, S. & Lueders, T., 2007. Detection of anaerobic toluene and hydrocarbon degraders in contaminated aquifers using benzylsuccinate synthase (bssA) genes as a functional marker. *Environ. Microbiol.*, 9(4), pp.1035–1046.

Yagi, J.M. et al., 2010. Subsurface cycling of nitrogen and anaerobic aromatic hydrocarbon biodegradation revealed by nucleic acid and metabolic biomarkers. *Appl. Environ. Microbiol.*, 76(10), pp.3124–3134.

Yang, S. et al., 2014. Crude oil treatment leads to shift of bacterial communities in soils from the deep active layer and upper permafrost along the China-Russia Crude Oil Pipeline route. *PLoS One*, 9(5), p.e96552.

Youngblut, N.D., Barnett, S.E. & Buckley, D.H., 2018. HTSSIP: An R package for analysis of high throughput sequencing data from nucleic acid stable isotope probing (SIP) experiments. *PloS one*, 13(1), p.e0189616.

Youngblut, N.D. & Buckley, D.H., 2014. Intra-genomic variation in G+ C content and its implications for DNA stable isotope probing. *Environ. Microbiol. Rep.*, 6(6), pp.767–775.

El Zahar Haichar, F., Roncato, M.-A. & Achouak, W., 2012. Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of Arabidopsis thaliana. *FEMS Microbiol. Ecol.*, 81(2), pp.291–302.

Zakem, E. & Follows, M., 2017. A theoretical basis for a nanomolar critical oxygen concentration. *Limnol. Oceanogr.*, 62(2), pp.795–805.

Zedelius, J. et al., 2011. Alkane degradation under anoxic conditions by a nitrate-reducing bacterium with possible involvement of the electron acceptor in substrate activation. *Environ. Microbiol. Rep.*, 3(1), pp.125–135.

Zhang, L. & Lueders, T., 2017. Micropredator niche differentiation between bulk soil and rhizosphere of an agricultural soil depends on bacterial prey. *FEMS Microbiol. Ecol.*, 93(9), p.fix103.

Zhang, T. et al., 2013. Anaerobic benzene oxidation via phenol in Geobacter metallireducens. *Appl. Environ. Microbiol.*, p.AEM–03134.

Zhang, Y. et al., 2018. The Occurrence of Putative Nitric Oxide Dismutase (Nod) in an Alpine Wetland with a New Dominant Subcluster and the Potential Ability for a Methane Sink. *Archaea*, 2018.

Zhu, B. et al., 2019. Nitric oxide dismutase (nod) genes as a functional marker for the diversity and phylogeny of methane-driven oxygenic denitrifiers. *Front. Microbiol.*, 10, p.1577.

Zhu, B. et al., 2017. Unexpected diversity and high abundance of putative nitric oxide dismutase (Nod) genes in contaminated aquifers and wastewater treatment systems. *Appl. Environ. Microbiol.*, 83(4), pp.e02750–16.

Appendix 1: Supporting Information

Post-submission update on taxonomy of the Rhodocyclaceae family

The Rhodocyclaceae family, as defined in the Silva v123 database used in data analysis for this thesis, was highly abundant and clearly connected to toluene degradation in sediment from the Siklós aquifer. However, this family underwent a major taxonomic reclassification since the data analysis was performed (Boden et al., 2017), which affects our interpretation of the sequencing data. Prior to reclassification, Rhodocyclales was a monofamiliar order containing only Rhodocyclaceae. It has now been split into three families: Rhodocyclaceae (containing genera *Rhodocyclus*, *Azospira*, and *Propionivibrio*), Azonexaceae (containing genera *Azonexus*, *Dechloromonas*, *Ferribacterium*, and *Quatrionicoccus*), and Zoogloeaeceae (containing genera *Zoogloea*, *Thauera*, *Uliginosbacterium*, and *Azoarcus*). This change should be noted when considering the results of the Siklós sediment-based microoxic microcosm experiment as described in sections 3.2 and 4.2.

Development of Enrichment Faction equations

Some previous SIP studies (Kramer et al. 2016; Zhang & Lueders 2017), and the Siklós total-RNA-SIP study undertaken as part of this thesis, calculated enrichment factors (EFs) by comparing gene/transcript copies or relative abundance in heavy and light fractions from microcosms fed with either isotopically labelled (^{13}C) or unlabelled (^{12}C) substrate according to equation EF_1:.

$$EF_1 = \frac{^{13}C_{heavy}}{^{13}C_{light}} - \frac{^{12}C_{heavy}}{^{12}C_{light}}$$

where "$^{13}C_{heavy}$" and "$^{13}C_{light}$" were the relative abundance of a specific rRNA taxon or mRNA transcript – assigned to the same KEGG/COG category – in heavy and light fractions of the ^{13}C-toluene microcosms, respectively, and "$^{12}C_{heavy}$" and "$^{12}C_{light}$" the same for respective ^{12}C-control microcosms.

In this case, the ($^{13}C_{heavy}$/ $^{13}C_{light}$) term defines the amount of labelling, while the ($^{12}C_{heavy}$/ $^{12}C_{light}$) term controls for shifts due to effects other than labelling, e.g. GC content. EF > 0 indicates that more of a given transcript appears in the $^{13}C_{heavy}$ fraction than can be explained by GC content or other physical effects, and therefore that isotopic labelling took place.

Equation EF_1 can be expanded to explicitly show that it relies on relative abundances. The expanded version, $EF_{1\text{-exp}}$, calculates EF for a transcript of interest (Tr. A).

$$EF_{1\text{-exp}} = \frac{\left(\frac{\text{Tr. A } {}^{13}\text{C-H}}{\text{Total tr. } {}^{13}\text{C-H}}\right)}{\left(\frac{\text{Tr. A } {}^{13}\text{C-L}}{\text{Total tr. } {}^{13}\text{C-L}}\right)} - \frac{\left(\frac{\text{Tr. A } {}^{12}\text{C-H}}{\text{Total tr. } {}^{12}\text{C-H}}\right)}{\left(\frac{\text{Tr. A } {}^{12}\text{C-L}}{\text{Total tr. } {}^{12}\text{C-L}}\right)}$$

However, the miniscule amount of RNA found in the ${}^{12}C_{heavy}$ fractions in the Flingern sediment microcosm experiment (Work Package 3) means that amplification would be necessary for reliable sequencing. This is not a problem for most RNA-SIP studies because they are based on PCR-amplified SSU rRNA. In our amplification-free method, however, the heavy ${}^{12}C$ fractions could not be sequenced and Equation EF_1 could not be used.

To develop a valid equation without heavy ${}^{12}C_{heavy}$ data, I considered whether a given transcript appears in the ${}^{12}C_{light}$ fraction in the expected amount. If the presence of a given transcript in ${}^{13}C_{heavy}$ is due to effects besides labelling, the total amount (relative to sequencing depth) across ${}^{13}C_{heavy}$ plus ${}^{13}C_{light}$ should be equal to the amount in ${}^{12}C_{light}$. This led to equation EF_2 , shown here expanded in the same manner as $EF_{1\text{-exp}}$:

$$EF_{2\text{-exp}} = \underbrace{\frac{\left(\frac{\text{Tr. A } {}^{13}C_{heavy}}{\text{Total tr. } {}^{13}C_{heavy}}\right)}{\left(\frac{\text{Tr. A } {}^{13}C_{light}}{\text{Total tr. } {}^{13}C_{light}}\right)}}_{term\ 1} - \underbrace{\frac{\left(\frac{\text{Tr. A } {}^{13}C_{heavy} + {}^{13}C_{light}}{\text{Total tr. } {}^{13}C_{heavy} + {}^{13}C_{light}}\right)}{\left(\frac{\text{Tr. A } {}^{12}C_{light}}{\text{Total tr. } {}^{12}C_{light}}\right)}}_{term\ 2}$$

Term 1 remains the same as in EF_1. If *term 2* has a value greater than *term 1*, some of transcript A that is expected to be in the ${}^{12}C_{light}$ fraction is not there and can be assumed to have migrated to a heavier fraction. Thus, labelling is not as strong as would be indicated by *term 1* alone. As in Equation EF_1 and the cited studies, $EF > 0$ indicates labelling.

The equation found in the methods section (2.8.3) is a simplified version of Equation EF_2, based on the fact that SSU rRNA values are already expressed in relative terms.

$$EF_2 = \frac{{}^{13}C_{heavy}}{{}^{13}C_{light}} - \frac{({}^{13}C_{heavy} + {}^{13}C_{light})/2}{{}^{12}C_{light}}$$

Our newly developed equation was tested on data sets from previous studies and found to give equivalent patterns of labelling as Equation EF_1.

Supplementary tables

Table A1: Phylogenetic affiliation and percent abundance of SSU rRNA reads from taxonomic units recovered in RNA-seq libraries in the Siklós total-RNA-SIP project. Isotope enrichment for taxa was calculated via equation EF1 using abundances in libraries that did not undergo linear amplification.

	Unamplified				Amplified				Enrichment (unamplified)	Average Abundance
	Unlabelled		Labelled		Unlabelled		Labelled			
	Light	Heavy	Light	Heavy	Light	Heavy	Light	Heavy		
Proteobacteria	97.9	98.3	95.4	99.9	97.7	98.1	94.7	99.8	0.04	97.7
Betaproteobacteria	81.0	84.7	62.8	93.2	80.8	83.0	66.1	90.3	0.44	80.2
Rhodocyclaceae	67.9	71.8	48.0	81.4	69.9	70.9	56.4	81.1	0.64	68.4
NA	20.0	22.1	15.9	24.8	20.9	22.6	19.3	22.0	0.46	21.0
Dechloromonas	11.3	12.4	8.3	13.7	18.9	21.3	10.8	22.1	0.55	14.8
Quatrionicoccus	12.6	12.2	6.3	14.5	9.7	4.9	5.7	14.1	1.33	10.0
Zoogloea	13.8	14.5	7.5	17.9	5.6	7.3	4.3	6.9	1.35	9.7
Azonexus	3.5	3.8	1.7	4.7	7.5	8.6	4.3	10.6	1.64	5.6
Azoarcus	2.2	1.7	4.4	0.7	3.0	1.3	8.5	1.1	-0.62	2.9
Sterolibacterium	1.6	2.1	1.2	2.2	1.4	2.4	1.0	1.7	0.53	1.7
Ferribacterium	1.0	1.0	0.6	1.2	1.1	1.1	0.6	1.1	0.79	1.0
Azovibrio	0.0	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.17	0.1
Comamonadaceae	9.1	8.2	11.1	6.1	6.4	6.3	6.8	4.3	-0.35	7.3
NA	5.9	5.5	7.7	4.1	4.4	4.4	4.7	2.9	-0.40	4.9
Acidovorax	0.8	0.7	0.9	0.5	0.6	0.6	0.6	0.4	-0.28	0.6
Polaromonas	0.5	0.4	0.6	0.2	0.2	0.2	0.3	0.1	-0.33	0.3
Gammaproteobacteria	9.7	6.7	15.4	5.8	13.9	10.3	22.6	8.7	-0.32	11.7
Pseudomonadaceae	9.2	6.3	14.1	5.7	13.5	9.9	21.6	8.6	-0.28	11.1
Pseudomonas	8.3	5.8	13.1	5.3	12.9	9.8	20.9	8.2	-0.30	10.5
NA	0.8	0.4	0.9	0.4	0.5	0.0	0.6	0.4	-0.07	0.5
Epsilonproteobacteria	5.6	5.3	13.3	0.5	1.4	2.9	2.9	0.2	-0.92	4.0
Campylobacteraceae	4.7	4.6	11.5	0.4	1.3	2.6	2.5	0.2	-0.95	3.5
Arcobacter	3.8	4.0	9.8	0.4	0.9	2.1	1.9	0.1	-1.03	2.9
Sulfurospirillum	1.0	0.6	1.7	0.1	0.4	0.6	0.6	0.1	-0.61	0.6
Others	2.1	1.7	4.6	0.1	2.3	1.9	5.3	0.2	-0.74	2.3

NA – Not affiliated below family level

Table A2: Abundance and phylogenetic assignment of *fliC* transcripts (coding for flagellin) recovered in RNA-seq libraries in the Siklós total-RNA-SIP project. Abundances are given relative to the total number of *fliC* transcripts in a given library. Only groups averaging >0.5 % of *fliC* transcripts in unamplified samples are shown. Isotope enrichment for taxa was calculated via equation EF1 using abundances in libraries that did not undergo linear amplification.

	Unamplified				Amplified				Enrichment (unamplified)	Average Abundance
	Unlabelled		Labelled		Unlabelled		Labelled			
	Light	Heavy	Light	Heavy	Light	Heavy	Light	Heavy		
Bacteria	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	0	100
Proteobacteria	99.2	99.4	95.9	99.8	99.1	99.6	97.1	99.9	0.037164	98.6
Betaproteobacteria	59.3	53.4	36.2	60.1	50.8	62.2	37.1	66.2	0.759078	52.2
Burkholderiales	6.7	4.9	5.7	3.1	3.3	3.4	3.1	2.8	-0.18418	5.1
Comamonadaceae	5.6	4.0	3.9	2.3	2.9	2.9	2.3	2.3	-0.13155	4.0
Acidovorax	0.9	1.1	1.3	0.5	0.4	0.4	0.8	0.4	-0.84297	0.9
Acidovorax sp. KKS102	0.8	0.7	1.0	0.4	0.4	0.4	0.2	0.3	-0.43885	0.7
Rhodocyclales	39.4	34.2	20.0	40.3	33.1	43.0	22.1	47.8	1.14753	33.5
Rhodocyclaceae	39.4	34.2	20.0	40.3	33.1	43.0	22.1	47.8	1.14753	33.5
Azoarcus	0.9	1.2	0.3	0.7	1.6	2.0	1.3	1.2	0.967009	0.8
Azoarcus sp. KH32C	0.6	0.8	0.3	0.6	1.5	1.9	1.3	0.9	0.227442	0.6
Azovibrio	16.1	17.5	10.4	19.8	15.1	18.3	10.0	20.9	0.81703	16.0
Azovibrio restrictus	16.1	17.5	10.4	19.8	15.1	18.3	10.0	20.9	0.81703	16.0
Deltaproteobacteria	0.6	0.8	1.6	0.3	0.1	0.1	0.8	0.1	-1.1363	0.8
Desulfuromonadales	0.6	0.7	1.3	0.1	0.1	0.1	0.4	0.1	-1.0844	0.7
Geobacteraceae	0.6	0.7	1.3	0.1	0.1	0.1	0.4	0.1	-1.0844	0.7
Geobacter	0.6	0.7	1.3	0.1	0.1	0.1	0.4	0.1	-1.0844	0.7
Geobacter metallireducens	0.3	0.6	0.9	0.1	0.0	0.1	0.2	0.0	-2.17973	0.5
Epsilonproteobacteria	3.5	4.6	17.8	0.9	12.1	7.5	30.8	2.5	-1.27913	6.7
Campylobacterales	3.4	4.6	17.6	0.9	11.9	7.5	30.4	2.5	-1.30563	6.6
Campylobacteraceae	2.8	3.7	14.2	0.9	8.9	6.5	25.2	2.2	-1.23429	5.4
Arcobacter	2.4	2.6	10.3	0.5	6.4	4.6	20.2	1.5	-1.04544	3.9
Sulfurospirillum	0.4	0.9	3.3	0.3	2.2	1.6	5.0	0.7	-2.11999	1.2
Helicobacteraceae	0.5	0.7	2.9	0.0	2.5	0.7	1.0	0.2	-1.51715	1.0
Sulfuricurvum	0.3	0.3	1.7	0.0	1.7	0.2	4.0	0.1	-0.7965	0.6
Sulfuricurvum kujiense	0.3	0.3	1.6	0.0	1.6	0.2	2.9	0.1	-0.7965	0.6
Gammaproteobacteria	1.0	0.7	1.6	0.1	0.3	0.5	2.9	0.0	-0.63152	0.9
Pseudomonadales	0.6	0.5	1.0	0.1	0.3	0.4	0.4	0.0	-0.67377	0.5
Pseudomonadaceae	0.6	0.5	1.0	0.1	0.3	0.4	0.4	0.0	-0.67377	0.5
Pseudomonas	0.6	0.5	1.0	0.1	0.3	0.4	0.4	0.0	-0.67377	0.5
Spirochaetes	0.3	0.4	1.6	0.1	0.6	0.2	0.4	0.0	-1.02376	0.6
Spirochaetia	0.3	0.4	1.6	0.1	0.6	0.2	2.1	0.0	-1.02376	0.6
Spirochaetales	0.3	0.4	1.6	0.1	0.6	0.2	2.1	0.0	-1.02376	0.6
Spirochaetaceae	0.3	0.4	1.6	0.1	0.5	0.2	2.1	0.0	-1.02376	0.6

Table A3: Abundance and phylogenetic assignment of catechol-2,3-dioxygenase transcripts recovered in RNA-seq libraries in the Siklós total-RNA-SIP project. Abundances are given relative to the total number of C23O transcripts in a given library. Only groups averaging >0.5 % of C23O transcripts in unamplified samples are shown. Isotope enrichment for taxa was calculated via equation EF1 using abundances in libraries that did not undergo linear amplification.

	Unamplified				Amplified				Average Abundance
	Unlabelled		Labelled		Unlabelled		Labelled		
	Light	Heavy	Light	Heavy	Light	Heavy	Light	Heavy	
Bacteria	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Proteobacteria	99.4	98.3	94.7	98.4	97.2	97.8	97.6	98.0	97.7
Betaproteobacteria	11.2	17.4	15.1	13.2	11.3	13.0	19.5	17.8	14.8
Burkholderiales	1.1	0.9	1.3	1.1	2.1	2.5	2.4	3.0	1.8
Burkholderiaceae	0.0	0.0	0.0	0.0	0.0	0.3	0.8	0.0	0.1
Comamonadaceae	0.6	0.9	0.7	0.0	0.7	0.0	0.0	1.5	0.5
Hydrogenophaga	0.6	0.9	0.0	0.0	0.4	0.0	0.0	1.0	0.3
T4	0.6	0.9	0.0	0.0	0.0	0.0	0.0	1.0	0.3
Methylibium	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Xenophilus	0.0	0.0	0.7	1.1	0.4	0.0	0.0	0.5	0.3
Xenophilus azovorans	0.0	0.0	0.7	1.1	0.4	0.0	0.0	0.5	0.3
Rhodocyclales	3.4	12.2	7.2	7.1	5.7	6.3	13.0	7.6	7.8
Rhodocyclaceae	3.4	12.2	7.2	7.1	5.7	6.3	13.0	7.6	7.8
Azovibrio	0.6	1.7	2.6	1.6	1.8	1.0	3.3	1.5	1.8
Azovibrio restrictus	0.6	1.7	2.6	1.6	1.8	1.0	3.3	1.5	1.8
Unclass. Rhodocyclaceae	0.0	0.9	1.3	0.0	0.7	1.3	4.9	0.5	1.2
Rhodocyclaceae strain PG1-Ca6	0.0	0.9	1.3	0.0	0.7	1.3	4.9	0.5	1.2
Zoogloea	1.7	7.0	2.0	2.2	1.4	2.2	2.4	0.5	2.4
Zoogloea oleivorans	1.7	7.0	2.0	2.2	1.4	2.2	2.4	0.5	2.4
Methyloversatilis	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.1
M. universalis	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.1
Thauera	0.6	0.0	0.0	0.0	0.4	0.6	0.0	1.0	0.3
Gammaproteobacteria	7.3	13.0	6.6	10.4	9.2	5.4	15.4	10.2	9.7
Pseudomonadales	7.3	11.3	3.3	10.4	8.9	3.2	7.3	8.1	7.5
Pseudomonadaceae	5.0	11.3	3.3	7.1	6.4	3.2	7.3	8.1	6.5
Pseudomonas	5.0	11.3	3.3	7.1	0.4	3.2	7.3	7.1	5.6
Pseudomonas taeanensis	5.0	0.9	0.0	7.1	0.4	0.0	0.0	0.0	1.7
Xanthomonadales	2.2	0.9	3.3	3.3	2.1	2.2	7.3	1.5	2.9
Xanthomonadaceae	2.2	0.9	3.3	3.3	2.1	2.2	7.3	1.5	2.9
Pseudoxanthomonas	2.2	0.9	3.3	3.3	2.1	2.2	7.3	1.5	2.9
P. spadix	2.2	0.9	3.3	3.3	2.1	2.2	7.3	1.5	2.9
Epsilonproteobacteria	0.0	0.0	0.0	0.0	0.4	0.3	1.6	0.0	0.3
Campylobacterales	0.0	0.0	0.0	0.0	0.4	0.3	1.6	0.0	0.3
Campylobacteraceae	0.0	0.0	0.0	0.0	0.4	0.3	1.6	0.0	0.3
Arcobacter	0.0	0.0	0.0	0.0	0.4	0.3	1.6	0.0	0.3

Table A4: Number of sequencing reads and proportion of reads identified as rRNA and non-rRNA in cDNA-seq libraries produced in the Flingern sediment microcosms (work package 3). Libraries analyzed for functional transcript profiles are highlighted in grey.

Library number	Condition	Isotope	Replicate	Fraction	Total reads	% SSU rRNA	% LSU rRNA	% non-rRNA
1	Oxic	^{13}C	1	Heavy	38997710	38	46	16
2	Oxic	^{12}C	1	Light	51312718	31	53	15
3	Oxic	^{13}C	2	Heavy	41654894	32	41	27
4	Oxic	^{12}C	2	Light	43123278	31	60	9
5	Oxic	^{13}C	3	Heavy	44307115	47	51	2.2
6	Oxic	^{12}C	3	Light	49077675	43	56	1.3
7	Oxic	^{12}C	1	Light	53242423	33	64	2.6
8	Oxic	^{12}C	2	Light	45748462	41	56	1.9
9	Nitrate-amended microoxic	^{13}C	1	Heavy	52649376	42	57	1.1
10	Nitrate-amended microoxic	^{12}C	1	Light	43342531	46	53	1.1
11	Nitrate-amended microoxic	^{13}C	2	Heavy	54354023	47	52	1.3
12	Nitrate-amended microoxic	^{12}C	2	Light	41914981	41	58	0.9
13	Nitrate-amended microoxic	^{13}C	3	Heavy	51779007	41	58	1
14	Nitrate-amended microoxic	^{12}C	3	Light	45185998	39	60	0.8
15	Nitrate-amended microoxic	^{12}C	1	Light	71186603	37	63	0.6
16	Nitrate-amended microoxic	^{12}C	2	Light	40823076	32	57	10.7
17	Microoxic	^{13}C	1	Heavy	41818458	35	51	14
18	Microoxic	^{12}C	1	Light	53792616	40	59	1.4
19	Microoxic	^{13}C	2	Heavy	84581274	40	54	6.6
20	Microoxic	^{12}C	2	Light	42073052	35	60	4.6
21	Microoxic	^{13}C	3	Heavy	43158416	45	52	1.6
22	Microoxic	^{12}C	3	Light	51249541	35	64	0.7
23	Microoxic	^{12}C	1	Light	63164354	48	51	0.7
24	Microoxic	^{12}C	2	Light	53848519	41	58	0.8
25	Anoxic	^{12}C	1	Light	49312228	42	59	0.7

Table A5: Proportion of reads identified as non-rRNA and as mRNA in selected references.

Reference	Percent total reads non-rRNA	Percent total reads identified as mRNA	Database	Source
Radax et al. (2012)	8 %	1.5 %	NCBI-nr	Table 1
Fortunato & Huber (2016)	1.5 - 75.5* %	0.04 – 2.5 %	KEGG Orthology	Calculated from Table S1
Schwab et al. (2014)	1.7 – 7.7 %	0.6 – 2.2 %	SEED	Calculated from Table S1

* The high percent of reads not identified as rRNA in some samples is not discussed in Fortunato and Huber (2016), but may be due to an abundance of uncultivated lineages at these sites and/or their use of an older Silva release (release 111). The percent of total reads matching mRNA sequences in the KEGG database remains < 2.5 % despite the high percent of reads considered non-rRNA.

Appendix 2: Bioinformatics Walkthrough

The majority of scripts used here are the work of Dr. Gisle Vestergaard (gisle.vestergaard@bio.ku.dk). As such, it is up to him whether to share that code in published forms (like a thesis) or on an individual basis. Here, I provide a walkthrough of the steps used, along with expected inputs/outputs and chosen parameters. For those less familiar with bioinformatic analysis, this guide should give an overview of the path taken by my data from sequencing to interpretation. Those with bioinformatics experience should be able to create a similar workflow by building their scripts around the programs used.

Overview

The state of the data when received from the sequencing company determines which steps are necessary. From CeGaT, files were fastQ format, separate forward and reverse reads, with adapters already trimmed and PhiX spike-ins already removed. I have included steps for trimming adapters and removing spike-ins if necessary for other data. From IMGM, adapters had to be trimmed.

Initial processing and sorting

1. (Trim adapters and) collapse paired-end reads together. Initial quality filtering.
   Program: Adapterremoval (v.2) (Schubert et al., 2016)
2. Remove spike-ins and other contaminants
   Program: Deconseq (v.0.4.3) (Schmieder and Edwards, 2011)
3. Sort rRNA from non-rRNA by comparison to Silva databases
   Program: SortmeRNA (v.2) (Kopylova et al., 2012)

16S

4. Sort 16S from rRNA file output of previous step by comparison to Silva databases
   Program: SortmeRNA (v.2) (Kopylova et al., 2012)
5. Preprocess 16S fastQ files for OTU calling. Second quality filtering.
   Program: Qiime (v1.9.1) (Caporaso et al., 2010)
6. Assign OTUs with open-reference method
   Program: Qiime (v1.9.1)
7. Attach taxonomy metadata to OTU file
   Program: Biom (McDonald et al., 2012)

mRNA (non-rRNA)

8. Compare non-rRNA reads against KEGG and NCBI-nr databases
   Program: DIAMOND (v0.9.22) (Buchfink et al., 2015)
   (Convert fastq to fasta format, Program: Biopieces (Hansen et al., 2008))
9. Create RMA files
   Program: MEGAN (v5.10.6) (Huson et al., 2007)
10. Create megan files to import into MEGAN
    Program: MEGAN (v5.10.6)

Tips - Biopieces

The biopieces framework (https://github.com/maasha/biopieces/wiki) has many useful commands for small but useful tasks like:

- counting the number of sequences in a file

```
read_fastq -i <input fastq> | count_records -xo
```

- taking a subset of reads

```
read_fastq -i <input fastq> -n <number of reads> | write_fastq -o <output>.fastq
```

- converting between formats

```
read_fastq -i <input fastq> | write_fasta -o <output>.fasta
```

- visualizing size distribution of reads and saving the histogram as an svg file

```
read_fasta -i <input fastq> | plot_lendist -x -t svg -o lendist_non_rRNA.svg -k SE
Q_LEN
```

And many other things. See the online manual and tutorials for more.

Initial processing

Confirm adapters

Inputs: sequences in fastQ format. File1: forward paired reads. File2: reverse paired reads
Outputs: prints identified adapter sequences to stream. No file output

```
adapterremoval --file1 <file1> --file2 <file2> --identify-adapters
```

```
Ex.
/project/genomics/Gisle/Bin/adapterremoval --file1 NG-136-0001_S15_L001_R1_001.fas
tq.gz --file2 NG-136-0001_S15_L001_R2_001.fastq.gz --identify-adapters`
```

Example output:

Processed a total of 118,453,942 reads in 8:42.6s; 226,000 reads per second on average …
Found 50801257 overlapping pairs …
Of which 14694 contained adapter sequence(s) …
Printing adapter sequences, including poly-A tails:
–adapter1:
AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGT
CTTCTGCTTG
| || |||| || || | | | ||****** ||| | | |
Consensus:
CGGTCTCTTATACACATCTCCGAGCCCACGAGACTACGCTGCATCGCGGATGGCG
GGGGGGGGGGGAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Quality:
#"#"#"#""#"#"#"##""##"#"##########"#"##"#"""""""#"""#""#"""""""#"""#"""""""##########
##
Top 5 most common 9-bp 5'-kmers:
1: CTGTCTCTT = 0.80% (94)
2: CGTGGACTA = 0.31% (36)
3: TGGTCGTCC = 0.24% (28)
4: CTGTCACTT = 0.24% (28)
5: AGTGCTCTA = 0.21% (25)
–adapter2:
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTAT
CATT
|| | | | | |||| |
Consensus:
CTGTCTCTTATACACATCTGACGCTGCCGACGACGGAGAGAGTGTAGATCTCGGG
GGGGGGGGGAGGAGGAAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Quality:
##$##"###$######"#"##########$##########"#"##"""""##"##""#""#"""""""""""""""""###
##"""""#""
Top 5 most common 9-bp 5'-kmers:
1: CTGTCTCTT = 0.58% (68)
2: ATGTCTCTT = 0.26% (31)
3: CTGTCTCTA = 0.26% (30)
4: CTGTCGCGT = 0.20% (24)
5: CTGTCACTT = 0.19% (22)

Remove adapters and collapse paired sequences

Input: FastQ files x2 (forward and reverse), adapter sequences (if applicable)
Output: 6x fastQ files: (data from adapterremoval help page)

NAME.gz.collapsed: contains overlapping mate-pairs which have been merged into a single read (PE mode) or reads for which the adapter was identified by a minimum overlap, indicating that the entire template molecule is present. This does not include which have subsequently been trimmed due to low-quality or ambiguous nucleotides.
NAME.collapsed.truncated: Collapsed reads which were trimmed due the presence of low-quality or ambiguous nucleotides
NAME.pair1.truncated: Output file containing trimmed mate1 reads
NAME.pair1.truncated: Output file containing trimmed mate2 reads
NAME.singleton.truncated: Output file to which containing paired reads for which the mate has been discarded
NAME.discarded: Contains reads discarded due to length or too many N bases

```
qsub -cwd /project/genomics/Gisle/Scripts/qsub_adapterremoval.rb -f <file1> -r <fi
le2> -o <Output_directory> -q <min quality score> -l <min length> --adapter1 <adap
ter1_seq> --adapter2 <adapter2_seq> -c
```

```
Ex.
mkdir Qsub_adapterremoval

qsub -cwd /project/genomics/Gisle/Scripts/qsub_adapterremoval.rb -f S799Nr8.1.fast
q.gz -r S799Nr8.2.fastq.gz  -o Qsub_adapterremoval -q 15 -l 50 --adapter1 AGATCGGA
AGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG --adapter2 AGATCGGAAGAGCG
TCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -c
```

Parameters

- f: input fastQ file. Forward paired reads
- r: input fastQ file. Reverse paired reads
- o: Name of output directory. Directory does not have to exist prior
- q: trim bases at 5'/3' termini with quality scores <= to set value
- l: minimum length
- adapter1, adapter2: sequences known from library prep and/or confirm adapter step
- c: collapse. When set, paired ended read alignments of --minalignmentlength or more bases are combined into a single consensus sequence, representing the complete insert, and written to either NAME.collapsed or NAME.collapsed.truncated

Combine all collapsed sequences

Concatenate both collapsed files (truncated or not). This means sequences with low-quality bases are carried forward, to be discarded on the preprocessing (pre-OTU calling) step.

```
cat <NAME>.collapsed* > <NAME>.catcoll.fastq
```

Remove spike-in sequences

Input: fastq file
Output:2x fastq files

<NAME>.catcoll_clean.fq - contains "clean sequences", everything that is not phiX
<NAME>.catcoll_cont.fq - contains any detected phiX sequences

```
qsub -cwd /project/genomics/Gisle/Scripts/qsub_deconseq.sh <NAME.catcoll.fastq> ph
ix <Output directory>
```

Sort

Sort by comparing sequences to Silva databases, which must be kept current on your server. Sortmerna v2 and above only sorts rRNA from non-rRNA, instead of giving output files of 16S, 23S, other rRNA, and non-rRNA as in previous versions. Our workaround is two steps: first sort rRNA from non-rRNA, then use a modified Sortmerna script to sort out 16S specifically.

Sort rRNA from non-rRNA

Input: "Clean" fastq file from deconseq, or (if phiX was removed by company prior to data transmission), concatenated fastq file.
Output:

<NAME>_rRNA.sam
<NAME>_rRNA.fastq
<NAME>_non-rRNA.fastq
<NAME>_rRNA.log

```
mkdir SortedRNA
```

```
qsub -cwd sortmeRNA.sh <NAME>.fq SortedRNA/
```

Sort only 16S from rRNA

Using a modified version of the sortmerna.sh script with comparisons only to Silva 16S databases

Input: fastq of all rRNA from previous step
Output:

<NAME>_rRNA_16SrRNA.sam
<NAME>_rRNA_16SrRNA.fastq

```
mkdir SortedRNA
```

```
qsub -cwd sortme16srRNA.sh <NAME>_rRNA.fq SortedRNA/
```

The difference between sortmeRNA.sh and sortme16srRNA.sh is simply in the list of databases compared against (and the name specified for output files).

- sortmeRNA compares against silva-bac-16s-id90.fasta , silva-bac-23s-id98.fasta , silva-arc-16s-id95.fasta , silva-arc-23s-id98.fasta , silva-euk-18s-id95.fasta , silva-euk-28s-id98.fasta , rfam-5s-database-id98.fasta , and rfam-5.8s-database-id98.fasta . Any sequence read that matches something in one of these databases is sorted as rRNA.
- sortme16srRNA.sh compares against only silva-bac-16s-id90.fasta , silva-arc-16s-id95.fasta , and silva-euk-18s-id95.fasta . Any sequence read that matches something in these databases is sorted as 16S. There is no output file for non-16SrRNA, though the script could be edited to provide that.
- These databases are maintained on the server by members of COMI.

16S workflow

Make a mapping file

An overview of mapping file format and function can be found on the qiime website: http://qiime.org/documentation/file_formats.html

The mapping file is essential for preprocessing files before beginning OTU assignment. It will also be very helpful for further processing like using the phyloseq package in R, and is generally a good way to keep a record of sample names and experimental conditions.

Example mapping file (first few lines) from the first microcosm project:

#SampleID	BarcodeSequence	LinkerPrimerSequence	ReversePrimer	Redox	Isotope	Fraction
AL1Heavy	AAAAAAAA	<linkerprimerseq>	<revprimer>	Oxic	13C	Heavy
AL1Light	GGGGGGGG	<linkerprimerseq>	<revprimer>	Oxic	13C	Light
AL5Heavy	CCCCCCCC	<linkerprimerseq>	<revprimer>	Oxic	13C	Heavy
AL5Light	TTTTTTTT	<linkerprimerseq>	<revprimer>	Oxic	13C	Light
AU2Light	AGGCGGGG	<linkerprimerseq>	<revprimer>	Oxic	12C	Light

Since reads are already demultiplexed, the barcodes in this case are not real, but since that column is required they must be invented. Each sample must have a unique 8-letter barcode. Primer sequences can be identical - in this case, they're a default Illumina set and are not actually used in analysis. Additional columns can be added with experimental metadata, which will be useful later on for grouping samples and generating statistics.

Check the mapping file for errors:

```
validate_mapping_file.py -m <mapping file> -o <Output directory>
```

This is a qiime script.
A message will print to the stream telling whether there are errors or not. If errors are present, check the log and corrected mapping file in the output directory.

Preprocess

(Quality filter, remove chimeras, format read names)

Input: fastq file of 16S reads, mapping file
Output: fasta file

seqs.fna - contains all 16S sequences that passed quality filtering, with sample ID appended to each read identifier.

Gisle has a script that combines the three steps using qiime scripts
split_libraries_fastq.py (quality filtering and name formatting)
truncate_reverse_primer.py (remove adapter sequences)
identify_chimeric_seqs.py (find chimeras using usearch61) and
filter_fasta.py (to remove sequences IDed as chimeric)

```
qsub -cwd /project/genomics/Gisle/Scripts/qsub_qiime_preprocess.sh <file>.fq 2 <fu
ll path to mapping file> <min size> <max size>
```

However, running this script on large fastq files very often resulted in crashes because of memory problems.Furthermore, adapter removal was already performed, and not a single chimeric read was found in the Siklós samples (for which I used Gisle's script and a work-around of splitting files, processing, recombining, and fixing name formats with the sed command. No details given because it was unnecessarily complicated).

It is therefore possible, and very simple, to use just split_libraries_fastq.py. I adapted parameters from the qsub_qiime_preprocess.sh script above.

```
split_libraries_fastq.py -i <NAME>.fastq -o <Output directory> -m <mapping file> -
-barcode_type 'not-barcoded' -q 2 -p 0.01 --phred_offset=33 --sample_id <NAME>
```

Parameters:

- barcode_type: not_barcoded (already demultiplexed)
- phred_offset: specify ascii code used. 33 is standard for modern Illumina sequencing.
- q: phred quality threshold. The maximum unacceptable score
- p: min. per read length fraction. Min number of consecutive high-quality base calls to include a read (per single end read) as a fraction of the input read length.
- sample_id: the name that will be appended to all read identifiers, and will appear as a column in the eventual OTU table. Must match the sample name in the mapping file.

Ex.

```
mkdir Preprocess


split_libraries_fastq.py -i BL5HeavyRep_16SrRNA.fastq -o Preprocess/ -m /home/igo
e/lauren.bradford/O2NO3_Mic_Experiment/micexp_mappingfile_run2.txt --barcode_type
'not-barcoded' -q 2 -p 0.01 --phred_offset=33 --sample_id BL5HeavyRep
```

OTU Calling

Input: all preprocessed 16S fasta files from whole experiment
Output: OTU table in biom format, many intermediate files

otu_table_filtered.biom - a biom formatted OTU table

The qiime_assign_silva.sh script makes use of qiime scripts:

- parallel_pick_otus_uclust_ref.py - a closed-reference OTU method, but failures are filtered out and clustered de novo in subsequent steps
- filter_fasta.py - make a fasta file of sequences that failed OTU assignment in the previous step
- pick_otus.py - pick de novo OTUs for failed sequences (subsampled with a python script)
- pick_rep_set.py - pick one sequence from each OTU
- make_otu_table.py
- filter_otus_from_otu_table.py
- merge_otu_tables.py
- parallel_align_seqs_pynast.py
- filter_alignment.py
- make_phylogeny.py

And biom scripts:

- biom add-metadata
- biom summarize-table
- biom convert

And an API script to filter out singletons

Command:

```
qsub -cwd <path>qsub_qiime_assign_silva.sh <output directory> <percent identity (u
sually 97)> <parameters file>
```

Contents of parameter file:

- pick_otus:enable_rev_strand_match True
- pick_otus:max_accepts 1
- pick_otus:max_rejects 8
- pick_otus:stepwords 8
- pick_otus:word_length 8
- pick_otus:similarity 0.97

Gisle's script is complex and automates many steps for combining OTUs from multiple samples in a pseudo-open-reference picking manner. Each sample is processed separately until the merging of OTU tables and pynast alignment.
In setting up a new workflow, this step may be more simply managed using qiime's pick_open_reference_otus.py script, though the processing time may be untenable for large datasets.
For a better explanation of this script, please contact Dr. Gisle Vestergaard (gisle.vestergaard@bio.ku.dk).

Convert biom file to txt file (OTU table)

The output files from OTU calling are in biom format (http://biom-format.org/), which can be directly imported to the phyloseq package in R. For analysis in Excel, biom files can be converted to text:

```
biom convert -i <biom file> -o <desired name of output file> --to-tsv --header-key
taxonomy
```

mRNA (non-RNA)

Compare against databases

Use DIAMOND to BLAST non-rRNA files against database fasta files for taxonomic or functional annotation. We compare by blastx against NCBI-nr, KEGG v.58.1 (the last free KEGG version), and COG (all with protein sequences). These databases must be maintained as fasta files on the server. The output file contains information needed to make RMA files with MEGAN in the next step.

Change fastq to fasta with Biopieces

Input: non-rRNA fastq file from SortmeRNA
Output: non-rRNA fasta file

```
read_fastq -i <NAME>.fq | write_fasta -o <NAME>.fa -x
```

Run DIAMOND script

Input: non-rRNA fasta file, reference database fasta (protein)
Output: tabulated text file with columns (brackets: field codes from DIAMOND manual):

Query Seq-id (qseid)
Reference ID (sseqid)
Percentage of identical matches (pident)
Alignment length (length)
Number of mismatches (mismatch)
Number of gap openings (gapopen)
Start of alignment in query (qstart)
End of alignment in query (qend)
Start of alignment in subject (sstart)
End of alignment in subject (send)
Expect value (evalue)
Bit score (bitscore)

```
qsub -pe hmp 18 /project/genomics/Gisle/Scripts/qsub_diamond.rb -i <query fasta -o
<NAME>.<database name> -c 18 -d <database>
```

Ex.

```
qsub -pe hmp 18 /project/genomics/Gisle/Scripts/qsub_diamond.rb -i AL1Heavy_non-rRN
A.fasta -o AL1Heavy.kegg -c 18 -d /project/genomics/Databases/KEGG/KEGG.fasta
```

Create RMA files with Megan script

Input: non-rRNA fasta file, DIAMOND output tab file
Output: RMA (read-match archive) file

```
qsub_megan_create_rma.sh <tabulated DIAMOND file file> <non-rRNA.fa file> <min_supp
ort value> <output directory >
```

Parameters:

- min_support value: how many reads a taxa must have to be included in the analysis. Changing this parameter can drastically affect the number of species you detect. I always use 1.

Analyze with MEGAN

Input: RMA file(s)

Open Megan - whether on the server or a personal version downloaded to your computer.
To open on the Helmholtz servers:

```
/project/genomics/Gisle/Bin/Programs/Megan_5_10_6/MEGAN
```

The following steps can be done in the MEGAN GUI:
Open the RMA files for one database (ex. KEGG) that you want to compare.
In any one of the windows (there will be one for each RMA opened), select Options -> Compare
Select all the samples you want to compare. Select “normalized counts” to normalize reads to the lowest number in one of the samples. This will open a new window with the comparison. This can be saved as a.MEGAN file to be opened again later without having to load all the RMA files again.

NCBI results can be seen in the main window. To see KEGG or COG results, click the KEGG or COG buttons near the top right.
Right-click on nodes to collapse or uncollapse subtrees and branches.
You can select sub-branches and leaves (individual transcripts) as you wish, and export read copy numbers for everything you have selected into a CSV file with File -> Export -> CSV format. Choose the format "kegg, name-counts" in the popup.

A note on transcript placement within categories and pathways:
If a transcript codes for an enzyme that is part of more than one pathway or category (which is very common!), it will be placed in ALL. For example, phenol hydroxylase plays a role in Benzoate degradation, Chlorocyclohexane and chlorobenzene degradation, Toluene degradation, and Aromatics degradation pathways.

Bioinformatics Pipeline References

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. doi:10.1038/nmeth.3176.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. doi:10.1038/nmeth.f.303.

Hansen, M. A., Oey, H., Fernandez-Valverde, S., Jung, C.-H., and Mattick, J. S. (2008). Biopieces: a bioinformatics toolset and framework. in 19th International Conference on Genome Informatics (Gold Coast, Australia.).

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. doi:10.1101/gr.5969107.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. doi:10.1093/bioinformatics/bts611.

McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D., Wilke, A., Huse, S., Hufnagle, J., Meyer, F., et al. (2012). The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience 1, 7.

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6, e17288. doi:10.1371/journal.pone.0017288.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC research notes 9, 88.

Publications and Conference Contributions

Publications

Bradford, L.M., Bartel, G., Lueders, T. (*in prep*) Transcriptome-SIP of hydrocarbon degrader adaption under dual electron acceptor availability in aquifer sediments.

Táncsics, A., Farkas, M., Horváth, B., Maróti, G., **Bradford, L.M.**, Lueders, T., and Kriszt, B. (2019). Genome analysis provides insights into microaerobic toluene-degradation pathway of Zoogloea oleivorans Buc T. *Archives of microbiology*, 1–6. doi:10.1007/s00203-019-01743-8.

Zhu, B., Wang, J., **Bradford, L.M.**, Ettwig, K., Hu, B., and Lueders, T. (2019). Nitric oxide dismutase (nod) genes as a functional marker for the diversity and phylogeny of methane-driven oxygenic denitrifiers. *Front. Microbiol.* 10, 1577. doi:10.3389/fmicb.2019.01577.

Bradford, L.M., Vestergaard, G., Táncsics, A., Zhu, B., Schloter, M., and Lueders, T. (2018). Transcriptome-stable isotope probing provides targeted functional and taxonomic insights into microaerobic pollutant-degrading aquifer microbiota. *Front. Microbiol.* 9, 2696. doi: 10.3389/fmicb.2018.02696

Zhu, B., **Bradford, L.M.**, Huang, S., Szalay, A., Leix, C., Weissbach, M., Táncsics, A., Drewes, J. E., and Lueders, T. (2017). Unexpected diversity and high abundance of putative nitric oxide dismutase (Nod) genes in contaminated aquifers and wastewater treatment systems. *Appl. Environ. Microbiol.* 83, e02750–16. doi:10.1128/AEM.02750-16.

Lueders, T., Dumont, M.G., **Bradford, L.**, and Manefield, M. (2016). RNA-stable isotope probing: from carbon flow within key microbiota to targeted transcriptomes. *Curr. Opin. Biotechnol.* 41, 83–89. doi:10.1016/j.copbio.2016.05.001.

Conference contributions

Bradford, L.M., and Lueders, T. (2018). Hydrocarbon degradation under fluctuating redox conditions. Poster at the 17th International Symposium on Microbial Ecology (ISME) in Leipzig, Germany.

Bradford, L.M., and Lueders, T. (2018). Targeted approaches towards microbial function: Smart labelling of microbiomes and transcriptomes using stable isotopes. Poster at the HMGU evaluation in Munich, Germany.

Bradford, L.M., Szalay, A.R., Vestergaard, G., Schloter, M., Zhu, B., Farkas, M., Táncsics, A., and Lueders, T. (2017). Total-RNA-SIP Reveals Identity and Function of Hydrocarbon Degraders in a Hypoxic Aquifer. Poster at the 14th Symposium on Bacterial Genetics and Ecology (BAGECO) in Aberdeen, Scotland.

Bradford, L.M., Szalay, A.R., Vestergaard, G., Zhu, B., Farkas, M., Táncsics, A., and Lueders, T. (2016). Metatranscriptome-SIP of a hypoxic toluene-degrading aquifer microbiome. Poster at the 16th International Symposium on Microbial Ecology (ISME) in Montreal, Canada.

Bradford, L.M., Zhu, B., Szalay, A.R., Vestergaard, G., Táncsics, A., Farkas, M., and Lueders, T. (2016). Stable Isotope Probing – Metatranscriptomics of Hydrocarbon Degrading Cultures. Poster at the Conference of the Association for General and Applied Microbiology (VAAM) in Jena, Germany.

Authorship Clarifications

Articles featured in this thesis

1. Bradford, Bartel, and Lueders, 2019. The experiment was designed by Lauren Bradford and Tillmann Lueders. Sediment sampling at the Flingern aquifer site was performed in November 2016 by Lauren Bradford and Tillmann Lueders, along with members of the Lueders, Griebler, and Deng working groups (all members of the Institute for Groundwater Ecology). Microcosms were set up by Lauren Bradford with the aid of Gabrielle Bartel, who also performed one round of ds-cDNA synthesis prior to sequencing library preparation. Fumihito Ogawa also assisted with microcosm preparation and oxygen measurements. Sequencing was performed by CeGaT gmbH (Tuebingen). All other laboratory work was performed by Lauren Bradford, as were bioinformatic and data analyses. The manuscript was written by Lauren Bradford, with revisions by Tillmann Lueders. Parts of this manuscript appear in the methods, results, and discussion of this thesis (Chapters 2.1.2, 2.2, 2.4.2, 2.5, 2.6, 2.7, 2.8, 3.3, and 4.3).

2. Bradford *et al.*, 2018. The experiment was designed by Lauren Bradford, Tillmann Lueders, and András Táncsics (Szent István University, Godollo, Hungary). Sampling of well material was performed by András Táncsics, Milán Farkas (also from SIU), Anna Szalay, and Tillmann Lueders (IGOE). Microcosm setup, monitoring, and sacrifice were done by András Táncsics and Milán Farkas and Tillmann Lueders, as was RNA extraction. Lauren Bradford, Tillmann Lueders, András Táncsics, and Milán Farkas executed isopycnic centrifugation and associated steps together. Linear amplification and submission for sequencing were done by Lauren Bradford. Bioinformatics and data analysis were undertaken by Lauren Bradford based on a workflow designed in collaboration with Dr. Gisle Vestergaard (previously of COMI at HMGU, now at the Technical University of Denmark). The sequencing-free linear amplification test was conceived and carried out by Lauren Bradford, with support from Tillmann Lueders. The manuscript was written by Lauren Bradford, with major revisions by Tillmann Lueders and input from other authors. Parts of this manuscript appear in the methods, results, and discussion of this thesis (Chapters 2.1.1, 2.2, 2.4.1, 2.5, 2.6, 2.7.1, 2.8, 2.9, 3.2, and 4.2)

3. Zhu *et al.*, 2017. This project was conceived and designed by Baoli Zhu, Tillmann Lueders, and Lauren Bradford. Baoli Zhu and Lauren Bradford designed the primers based on available *nod* reference sequences. Lauren Bradford tested all combinations of forward and reverse primers with compatible melting temperatures, then used rounds of gradient PCR to determine

optimal annealing temperatures for successful primer pairs. Lauren Bradford performed the PCR screening and cloning of *nod* from contaminated aquifer sediment samples and aided Baoli Zhu during some WWTP sampling events. Baoli Zhu and Sichao Huang performed cloning sequencing of WWTP samples. They also performed qPCR assays for 16S rRNA *nod* abundance, with occasional help from Lauren Bradford. The manuscript was written by Baoli Zhu with major contributions by Lauren Bradford and Tillmann Lueders. Parts of this manuscript appear in the methods, results, and discussion of this thesis (Chapters 2.2, 2.3, 3.1, and 4.1).

Contributions to other articles

4. Táncsics *et al.*, 2019. The authors compared transcriptome sequencing data generated for Bradford *et al.* (2018) against the genome of Zoogloea oleivorans Buc T to better understand the role of this microbe in hydrocarbon degradation.

5. Zhu *et al.*, 2019. Lauren Bradford produced and sequenced the 16S rRNA clone libraries from "*Ca.* methylmirabilis oxyfera" enrichments.

6. Lueders *et al.*, 2016. Lauren Bradford wrote the mRNA-SIP paragraphs and gave suggestions on sections by other authors. She also created Figure 1 and Table 1, which appear in the introductory chapter of this thesis (Fig. 5 and Table 2).

Acknowledgements

My first thanks, of course, goes to Dr. Tillmann Lueders for his guidance, encouragement, and occasional suggestions to go home, drink a beer, & worry about Plan B (or C or D…) on Monday. That wisdom will certainly serve me well going forward. Thank you, Tillmann, for offering me such a cool project but also welcoming my suggestions and forays into independent experiments. My fellow members of the Lueders group were always quick to offer assistance and encouragement, in the lab or out of it, and I'm so happy to have worked with them. I am especially grateful to Baoli for his mentorship and to Gabi for always finding time to help me despite her perpetual busyness. Thanks also to Gisle Vestergaard, whose lessons in bioinformatics helped me discover a new love for the data science side of biology research.

Thank you to Dr. Schloter and Dr. Orsi for their input throughout these projects as members of my thesis committee, and to Dr. Schloter and Dr. Liebl for serving as my defense committee. Our discussions have made me a better scientist and your advice has resulted in a thesis that I am proud to share with the world.

Moving across the world in pursuit of a PhD was not easy, and I was so lucky to land in a community of researchers who were not only brilliant but kind, welcoming, and a whole lot of fun. IGOE was a wonderful place to work and its closure is a loss for Helmholtz and the whole environmental research community. I am grateful to have been part of it. Thanks especially to my officemates (Anna, Lu, Kathrin, & Lucas) for many welcome distractions, and to the lunch group (Anna, Sviatlana, He, Feng-chao, Clemens, Armin, & Luzie) for so many delicious meals and great conversations. Thanks also to Judith and Sviatlana for being great neighbours, by which I mean chauffeuring me around and bringing supplies when I was sick.

These scientific achievements wouldn't have been possible without all the support I found outside the lab as well. Playing in the English-Speaking Music Ensembles (ESME) orchestra kept me sane and introduced fantastic opportunities for a firmly mediocre violinist. Uscha Halla shared her lovely home with me and our many, many pets. My family has supported me on every step on this path, even though it took me so very far away. Thank you, Mom, for your endless encouragement; Dad, for hassling me into finally finishing this thesis; and Dan for occasionally saying you're proud of me (which is a lot to ask of a big brother).

Finally, thanks to David, without whom I would have long ago given up and run away to live in a forest. All my love and appreciation – for infinite patience, kindness, and food in adversity.