UiT The Arctic University of Norway

Department of Arctic and Marine Biology

Analysis of Bacterial Communities associated with the Tidal Zone Grass *Puccinellia* in Response to Diesel Spills

Judith Freyer

Master's thesis in Biology, BIO-3950, November 2024

Abstract

Contamination from petroleum products and hydrocarbons is one of the major sources of pollution worldwide. The use of plants or microbes to reduce pollution is called bioremediation and is a cheap and environmentally friendly alternative to mechanical or chemical clean-up. A good understanding of native plants and microbes is required to implement bioremediation. In this study, the common and widespread tidal grass Puccinellia sp. and its associated root microbiome were subjected to diesel contamination to assess their bioremediation potential. The experiment lasted 13 weeks and included three treatments: control, low diesel concentration and high diesel concentration. Rhizosphere samples were taken at four time points during this period and the bacterial community was analysed using 16 S sequencing. Prior to sequencing, several changes had to be made to the protocol to overcome difficulties. However, the method used did not yield sufficient DNA sequences and the analysis was limited to a descriptive study. The results show that there was a clear difference between rhizosphere and bulk soil bacterial communities. Rhizosphere communities have higher diversity and relatively stable community composition over time and treatment. In contrast, bulk soil communities have lower diversity and change over time and treatment. All results indicate a strong influence of plant on bacterial community composition, more than diesel. All plants survived the diesel contamination, and all treatments contained up to 60 % of bioremediation-relevant bacteria, indicating the bioremediation potential of Puccinellia sp. and its rhizosphere.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACE	Abundance-based Coverage Estimator
BC	Barcode
BTEX	Benzene, Toluene, Ethylbenzene, Xylene
CI	chloroform:isoamyl alcohol
EPS	Extracellular polymeric substances
EtBr	Ethidium bromide
GEP	Gel Electrophoresis
GM	Genetically Modified
HC	Hydrocarbon
HS	High Sensitivity (Qubit)
LCA	Last Common Ancestor
ND	NanoDrop
ONT	Oxford Nanopore Technologies
OTU	Operational Taxonomic Unit
PAH	Polycyclic Aromatic Hydrocarbon
PCI	Phenol:chloroform:isoamyl alcohol
PCR	Polymerase Chain Reaction
PGPR	Plant Growth Promoting Rhizobacteria
Q Score	Quality Score
RAP	Rapid Adapter
SARA	Saturates, Aromatics, Resins, Aspahltenes
SD	Standard Deviation
TAE	Tris-Acetate-EDTA buffer
TBE	Tris-Borate-EDTA buffer
TNS	Tris-NaCl-SDS buffer

1 Introduction

1.1 Petroleum Pollution

Petroleum, also known as crude oil, is a naturally occurring liquid that originates from decomposed organic material. It is typically found in porous rock formations and is commonly extracted for its industrial use. the main application is as a source of fossil fuels, but it is also used as an industrial intermediate to produce plastics, elastomers or pharmaceuticals (Baker, 1970; Gad, 2005a, 2005b; Sullivan, 2005). The exploration of petroleum creates large volumes of contaminated waste and is the quantitative largest source for environmental pollutants. Moreover, significant environmental contamination occurs throughout the production, refining, storage, transportation and utilisation of petroleum products (D´Surney & Smith, 2005; Gad, 2005a; Sullivan, 2005). The environmental impact of accidental oil spills, such as the Exxon Valdez or Deepwater Horizon incidents, is significant and receives considerable media attention. However, chronic low-level exposure is often more damaging to organisms (Bragg et al., 1994; D´Surney & Smith, 2005).

1.1.1 Chemical Composition and Environmental Fate

Hydrocarbons (HC) constitute between 50 and 98 % of petroleum products (Gad, 2005b). The remaining components are sulphur, nitrogen, and oxygen compounds, as well as trace metals (Gad, 2005b; Sullivan, 2005). Petrogenic HC can be divided into four main chemical fractions, saturates, aromatics, resins, and asphaltenes (SARA), according to their polarity (Prendergast & Gschwend, 2014). For simplicity, the last two fractions are combined (Figure 1). The main component are saturated aliphatic HCs, such as branched, linear (n) or cyclic alkanes, which contain only single bonds. The second fraction is aromatic HC, which are compounds with at least one benzene ring. Compounds with at least two benzene rings, such as naphthalene, are defined as polycyclic aromatic hydrocarbons (PAHs). The final fraction are asphaltenes and resins, which are complex compounds containing multiple fused aromatic rings and alkane side chains. These can make up to 10% of the crude oil (Baker, 1970; Fakher et al., 2020; Koshlaf & Ball, 2017; Prendergast & Gschwend, 2014; Ward et al., 2009).

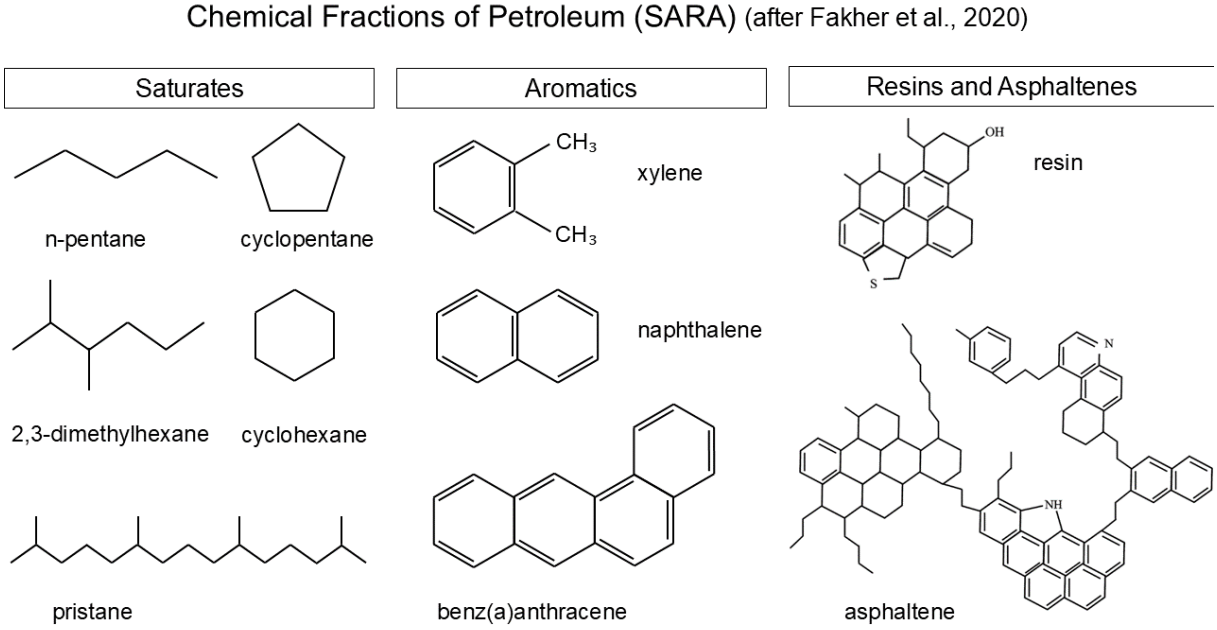

***Figure 1**: Examples compounds for the chemical fractions of petroleum products.*

The ratio and composition of the fractions determine the chemical and physical properties of the resulting petroleum product, such as volatility, solubility, boiling point, and viscosity. This allows for petroleum to be distilled into fractions with desired properties, a process known as refining (Gad, 2005a, 2005b; Sullivan, 2005). Diesel fuel, for example, can be distilled in the boiling range of approximately 160 to 360 °C and contains predominantly branched and cyclic alkanes (Figure 1) ranging from C9 to C20, but can also contain monoaromatics (Gad, 2005a). It is he most produced petroleum product, with over 200 million tonnes in the European Union alone (Eurostat, 2022).

The molecular weight of a HC increases with complexity (Figure 1), which also increases viscosity and hydrophobicity, and decreases volatility. The heavier and thicker petroleum products contain a greater proportion of asphaltenes and resins, whereas the lighter products contain a greater proportion of saturates (Prendergast & Gschwend, 2014). Moreover, this also determines the environmental fate of petroleum products is also determined, as biotic and abiotic weathering processes can separate and transform the different fractions. Compounds with a small and light structure, such as short alkanes and small aromatics, evaporate readily and are photodegraded in the atmosphere. Alternatively, they can be dissolved in water where they can be biodegraded by microbes. More complex compounds exhibit greater hydrophobicity and a high affinity for organic matter and soil. Such compounds are recalcitrant and often resistant to degradation. As the oil becomes more weathered, it will exhibit increased viscosity and adsorption in the soil (Gad, 2005b; Koshlaf & Ball, 2017; Phillips, 2005). However, weathering depends on many environmental factors, as well as the composition of the petroleum product. For example, increasing temperature decreases viscosity but increases solubility (Koshlaf & Ball, 2017), and waves and wind can disperse the oil over large areas (Hershner & Lake, 1980). However, its spread across an environment can cause severe damage.

1.1.2 Toxicity

The formation of a thick, oily slick following an oil spill can have a number of adverse biological effects. The blockage of light and water can hinder gas exchange and limit the availability of nutrients. In addition, animals can become covered in oil, which can have further detrimental effect. Moreover, HC themselves are toxic for a number of organisms, with the toxicity varying depending on the species, the type of oil as well as the duration and level of exposure (Koshlaf & Ball, 2017). In general, smaller compounds are more readily absorbed by tissues, for example through interaction with cell membranes, and cause rapid damage or irritation (Baker, 1970; Phillips, 2005). However, HC does not typically bioaccumulate in organisms (Phillips, 2005; Sullivan, 2005).

The most prevalent toxic effect on animals and humans is irritation and damage on the organ of contact, which is typically reversible but can become chronic with continuous exposure. The development of cancer has been observed to occur at higher concentrations of PAHs in the petroleum product (Gad, 2005a, 2005b; Phillips, 2005; Sullivan, 2005). In humans, exposure can result in narcosis and central nervous system depression. In particular, products with low viscosity present a significant risk of aspiration. Even minimal quantities can result in the development of chemical pneumonitis and subsequent lung damage. Dermal exposure results in chemical irritation (Gad, 2005b; Sullivan, 2005).

As reported by Baker (1970), the majority of shoreline plants show consistently reduced photosynthesis and transpiration, but varying effects on respiration rates. While germination is often inhibited, it is not a permanent phenomenon. Leaves, particularly when in contact with petroleum products, frequently exhibit symptoms such as wilting or discolouration. However, the degree of resistance exhibited by different species varies, and in some cases plants simply die. Where plants secure the shoreline, for example in salt marshes, the death of one species can cause secondary damage, such as erosion (Cowell, 1969; Hershner & Lake, 1980). Macroalgae exposed to crude oil exhibited a reduction in photosynthetic activity accompanied by an increase in respiration rates. Additionally, they often displayed discolouration following short-term exposure. Following chronic exposure, the overall growth rate was found to be reduced (Stepaniyan, 2008).

A number of microbial species are capable of degrading a wide range of petrogenic HCs. However, increased mutation rates might occur if PAHs are present and the community will shift in compositions to respond to the contamination (Ribeiro et al., 2013; Sullivan, 2005).

1.2 Clean-up of Shoreline Spills

The primary objective of the clean-up of petroleum spills at sea is to contain the spill and prevent it from reaching the shoreline (Dave & Ghaly, 2011). Moreover, given the significant biological damage that oil slicks can cause (Koshlaf & Ball, 2017), the minimisation of the effects on marine wildlife is a crucial aspect of the clean-up process. In the event of an oil spill reaching the shore, the response is conducted in three stages: 1) the emergency phase, to remove floating bulk oil, stage 2) the project phase, to remove stranded oil and oiled material, and stage 3) the polishing phase, to remove stains and leftover oil (ITOPF, 2014). The applied methods

for clean-up can often be used for more than one stage of response, and are either physical, chemical or biological (ITOPF, 2014; Prendergast & Gschwend, 2014).

If the contamination is excessive, for example thick oily mats, the first layer of vegetation or sediment may need to be removed completely. This method is an effective means of removing oil but does not ensure good recovery of the vegetation and should only be considered if there is significant damage to other wildlife, such as shorebirds. In some cases, replanting of native vegetation may be necessary. During the winter, such thick oily mats can be burned in situ as the damage to the vegetation is less severe and recovery in the following spring is improved (Baker et al., 1993).

The removal of floating and accumulated soil on water surfaces can be achieved through the use of skimmers, pumps, or vacuum trucks. These heavy machines are very effective, especially when time is of the essence. However, their operation requires road access and can be very damaging to the shoreline (ITOPF, 2014). If specialised machinery is not available, civil engineering machinery such as excavators can be used. However, these can only be applied to open coastlines such as beaches. In addition, the removal process is very imprecise and produces large amounts of waste (ITOPF, 2014). On very sensitive or inaccessible shorelines, oiled material might have to be collected manually. This method overall ensures good recovery of the habitat and very selective collection of material, however, it is very labour intensive (ITOPF, 2014).

Flushing can be applied in instances where the oil has permeated the sediment or was buried during the initial collection. It is also suitable for use on sensitive coastlines, such as mangroves or salt marshes. Large volumes of seawater are pumped into the sediment at low pressure. The water flow causes the oil to float and rise to the surface where it can be collected. On some coastlines, this can be achieved by using waves to transport moderately contaminated material below the waterline. This method is most effective on sandy or muddy sediments and less suitable on rocky shores. However, this approach can result in erosion of the shoreline and damage to surrounding vegetation (Baker et al., 1993; ITOPF, 2014).

The previous methods can be combined and repeated, depending on the residual oil and the geography. In particular, flushing will often need to be carried out over several tidal cycles. In addition, oil can be soaked up with highly sorbent materials. However, this method dramatically increases the volume of waste produced but is highly effective for surface oil (Baker et al., 1993; ITOPF, 2014).

Once the significant amount of bulk oil has been removed, the focus shifts to 'polishing' the environment (stage 3). At this point, the removal of heavy, thick oil slicks has been completed and what remains are either stains, oily films on substrates and lighter oil that has been absorbed into the sediment. In the case of a hard substrate or an industrial coastline, such as dams, hot or cold pressure washing can be applied. This method is often costly, time consuming and can cause significant damage to the surface. Hard substrate organisms such as algae or barnacles will never survive (Baker et al., 1993; ITOPF, 2014). On pebble, cobble or gravel shorelines, neither flushing, pressure washing nor in-situ burning is very effective, and rocks often need to

be washed in tanks or in converted cement trucks. This usually requires the addition of a solvent or detergent and is labour intensive and costly (ITOPF, 2014).

Chemical agents, such as surfactants, dispersant, emulsifier or solidifiers can be applied throughout various cleaning steps to enhance breakdown and removal of petroleum products. However, several studies have reported higher mortality of shoreline vegetation and poorer recovery following the use of surfactants and similar detergents. It is therefore recommended to minimise the use of chemicals (Baker et al., 1993; Cowell, 1969; ITOPF, 2014).

The most expensive and time-consuming part is the treatment or disposal of the collected oily material. Removed and contaminated material can be incinerated in special facilities, for example, or biodegraded in controlled bioreactors. In addition, it is often not possible to remove all of the oil completely, and significant amounts of oil are left to natural degradation and weathering (Chaudhry et al., 2005; Das & Chandran, 2011; ITOPF, 2014).

1.3 Bioremediation

HCs are ubiquitous in the environment and originate not only from anthropogenic pollution but also from natural sources such as deep-sea oil seeps, combustion and shale. Furthermore, some metabolic processes in plants result in the synthesis of compounds that are similar in structure to complex HC. This makes HC a continuously available substrate for a wide range of organisms (Gerhardt et al., 2009; Ron & Rosenberg, 2014; Ward et al., 2009). The purposeful use of organisms, such as bacteria or plants, to enhance the natural degradation of environmental contaminants, is called bioremediation (Shugart, 2005; Ward et al., 2009). In contrast to expensive and time-consuming cleaning methods, bioremediation is considered to be easy to implement, inexpensive, environmentally friendly, gentle on the coastline, applicable to large areas and resulting in complete mineralisation of contaminants (Das & Chandran, 2011; Gerhardt et al., 2009; Koshlaf & Ball, 2017). Bioremediation potential has been shown primarily in heterotrophic bacteria and fungi, but also in cyanobacteria, yeast, and even algae for a diversity of different habitats (Chaudhry et al., 2005; Das & Chandran, 2011; Ward et al., 2009).

1.3.1 Microbial Bioremediation

High adaptability and mutation rate, coupled with metabolic diversity, and fast dispersal over vast areas (Konopka, 2009), makes microbes particularly well-suited for biodegradation. The capacity of microbes to degrade petroleum was first demonstrated during the cleanup of the Exxon Valdez oil spill in 1989 (Bragg et al., 1994). Subsequent to this, it has become a standard approach especially for terrestrial ecosystems (Das & Chandran, 2011).

Microbes use HC not only for energy, but also as catabolic products in other metabolic pathways, or for direct incorporation into their cell structure, for example n-alkanes in cell membranes (Ward et al., 2009). In some cases transformations are merely done to reduce toxicity (Chaudhry et al., 2005). This may cause an energy deficit, however, contaminated soils are often nutrient depleted (Chaudhry et al., 2005; Gerhardt et al., 2009; Koshlaf & Ball, 2017), and especially nitrogen and phosphorus become limiting factors (Das & Chandran, 2011; Ron

& Rosenberg, 2014). Under aerobic conditions, HC degradation is initiated by oxygenases, which use molecular oxygen as an electron acceptor while oxidising HC as an electron donor. Monooxygenases incorporate only one atom of molecular oxygen, whereas dioxygenases incorporate two. Under anaerobic conditions, substrate activation is usually by hydroxylation with water, methylation or fumarate addition. However, these are energetically less favourable than oxygen, so anaerobic HC degradation is usually slower and only carried out by specialised organisms (Table 1) (Koshlaf & Ball, 2017; Ward et al., 2009).

The chemical composition of the petroleum product determines the rate and effectiveness of bioremediation, as the susceptibility of a HC to degradation depends on its complexity. Susceptibility can be ranked in decreasing order: linear alkanes, branched alkanes, monoaromatics, cyclic alkanes, PAHs and asphaltenes (Das & Chandran, 2011; Ward et al., 2009). In addition, more complex compounds have a greater affinity for organic matter and may absorb so much that their susceptibility is reduced even further (Koshlaf & Ball, 2017).

To increase bioavailability, microbes have developed two main strategies: splitting metabolic pathways and increasing uptake. A single genus rarely contains the entire metabolic pathway for the degradation of complex HCs. Instead, several genera form a complete pathway (Table 1), a process called co-metabolism (Gerhardt et al., 2009). High diversity means higher functional redundancy, and such communities are better equipped to degrade a wide range of hydrocarbons. However, this means that the bioremediation potential is limited by the diversity of the community (Konopka, 2009; Ron & Rosenberg, 2014). Some bacteria produce biosurfactants or emulsifiers (Table 1), a diverse group of amphipathic compounds that reduce the surface tension of oil, but have many other uses such as cell signalling or biofilm formation. In addition, micelles can be produced to encapsulate microdroplets, increasing the surface area of the oil (Das & Chandran, 2011; Ron & Rosenberg, 2014; Ward et al., 2009). An alternative approach is to enhance the hydrophobicity of cell membranes, thereby facilitating the uptake of HC (Ward et al., 2009).

Temperature has a strong influence on several aspects of petroleum HC degradation. The viscosity (and volatility) of the petroleum product decreases with temperature, reducing solubility and therefore bioavailability (Das & Chandran, 2011). Moreover, metabolic rates and chemical reactions decrease with temperature (Ratkowsky et al., 1982). As temperature decreases, the stiffness of biological membranes increases, thereby limiting membrane transport and reducing the uptake of HC and nutrients (Konopka, 2009).

***Table 1**: Table of bacterial families and their orders, relevant for this study. Listed are the most relevant (not all) substrates used by the family, as well as key ecological features that relate to HC degradation or plant growth promotion (PGPR).*

Order	Family	Compound, Substrate	Ecology	Source
Bacillales	Bacillaceae	Crude oil, n-alkanes, Aromatics (pyrene), Soil organic matter	Ubiquity and endospore formation De-emulsification and PGPR, Aerobic or facultative anaerobic, Chemoorganotrophy	(Das & Chandran, 2011; Koshlaf & Ball, 2017; Mandic-Mulec et al., 2015; Ward et al., 2009)
	Planococcaceae	Sodium salicylate, Aromatics Phenol, Lignocellulose,	Marine ubiquity, Emulsification and PGPR, Antibiotics and metal resistance, Aerobic or facultative anaerobic	(Ron & Rosenberg, 2014; Shivaji et al., 2014; Ward et al., 2009)
Burkholderiales	Alcaligenaceae	Crude oil, Aromatics (carbazole), n-alkanes (- C33)		(Das & Chandran, 2011; Koshlaf & Ball, 2017)
	Burkholderiaceae	Aromatics, n-alkanes		(Chaudhry et al., 2005; Das & Chandran, 2011; Koshlaf & Ball, 2017)
	Comamonadaceae	Naphthalene	Free-living in soil or water, Plant associated or plant pathogen, Animal associated or clinical samples, Mesophily, Aerobic organotrophy or anaerobic denitrification, Fe(III) reduction and H_2 oxidation	(Ward et al., 2009; Willems, 2014)
	Methylophilaceae	Methanol, Methylamine, Dichloromethane, Formaldehyde	Plant symbiosis in phyllosphere and rhizosphere, Aerobic obligate methylotrophy	(Doronina et al., 2014)
	Rhodocyclaceae	Aromatics, Recalcitrant compounds, Organic matter	Anaerobic photoheterotrophy, Chemoautotrophy or methylotrophy, Fermentation, Denitrification and nitrogen fixation in plant association	(Oren, 2014a; Ward et al., 2009)
Corynebacteriales	Corynebacteriaceae	Crude oil	De-emulsification	(Das & Chandran, 2011; Koshlaf & Ball, 2017; Ward et al., 2009)
	Mycobacteriaceae	Aromatics, Alkanes		(Koshlaf & Ball, 2017; Ward et al., 2009)
	Nocardiaceae	Aromatics, n-alkanes,	Ubiquity, Biosurfactant production,	(Das & Chandran, 2011; Koshlaf & Ball, 2017; Ward et al., 2009)

		Crude oil, volatile HC	Emulsification, de-emulsification, dispersion Aerobic	
Desulfuromonadales	Desulfuromonadaceae	Sulphur, Metal, Organic matter, Chlorinated solvents and BTEX compounds	Freshwater and marine, sediment and mud, Syntrophy with bacteria, Bioremediation of heavy metals, Mesophily, Obligate anaerobic or aerotolerant, Chemoorganotrophy, Respiration or fermentation,	(Greene, 2014)
Enterobacterales	Enterobacteriaceae		Ubiquity and endosymbiosis, PGPR and biosurfactant production, Facultative anaerobic	(Octavia & Lan, 2014; Ward et al., 2009)
Flavobacteriales	Flavobacteriaceae	Crude oil, Macromolecules (polysaccharides, proteins)	Ubiquity, Marine or freshwater, soil, plants or animals Capnophily, Aerobic or facultative anaerobic, Chemoorganotrophy	(Chaudhry et al., 2005; Das & Chandran, 2011; Koshlaf & Ball, 2017; McBride, 2014)
Geobacterales	Geobacteraceae	Mono-aromatics	Soils and sediments, High abundance in HC contaminated soils, Reduction of iron and heavy metals, Anaerobic	(Röling, 2014; Ward et al., 2009)
Granulosicoccales	Granulosicoccaceae		Marine or endophyte, Obligate aerobic chemoorganoheterotrophy	(Ivanova & Webb, 2014)
Hyphomicrobiales	Phyllobacteriaceae	n-alkanes, Aromatics (naphthalene)	aerobic	(Koshlaf & Ball, 2017; Ward et al., 2009)
Micrococcales	Brevibacteriaceae	Aromatics (phenanthrene), n-alkanes, Crude oil	Halotolerance (aquatic and terrestrial), Main application in dairy production	(Das & Chandran, 2011; Forquin-Gomez et al., 2014; Koshlaf & Ball, 2017)
	Cellulomonadaceae	Crude oil, n-alkanes, Aromatics		(Koshlaf & Ball, 2017)

	Micrococcaceae	Crude oil and kerosene, Aromatics (xylene), n-alkanes (C10 - C40), Alkenes	Soil, rhizosphere, marine, mammalian skin, De-emulsification Aerobic or facultative anaerobic,	(Chaudhry et al., 2005; Das & Chandran, 2011; Dastager et al., 2014; Koshlaf & Ball, 2017; Ward et al., 2009)
Mycobacteriales	Gordoniaceae	Crude oil, n-alkanes, Aromatics		(Das & Chandran, 2011; Koshlaf & Ball, 2017)
	Dietziaceae	Crude oil, n-alkanes, Aromatics	Soil and marine sediment, Petroleum contaminated sites	(Das & Chandran, 2011; Koshlaf & Ball, 2017; Pukall, 2014)
Oceanospirillales	Oceanospirillaceae		Marine ubiquity	(Ron & Rosenberg, 2014; Ward et al., 2009)
	Oleiphilaceae	n-alkanes	Marine ubiquity, Aerobic metabolism	(Cappello & Yakimov, 2010; Ron & Rosenberg, 2014; Ward et al., 2009)
Propionibacteriales	Nocardioidaceae	Aromatics, Haloalkanes, Polymeric polyesters	Ubiquity, Polluted soils and wastewater, Rhizosphere or endophyte, Chemoorganotrophy and respiration	(Das & Chandran, 2011; Ma et al., 2023; Tóth & Borsodi, 2014)
	Propionibacteriaceae	Aromatics	Activated sludge, soil and water, dairy, human epidermis, Aerobic or facultative anaerobic	(Stackebrandt, 2014)
Pseudomonadales	Alcanivoracaceae	Crude oil, n-alkanes (C10 - C20), Aromatics	Marine ubiquity, High abundance after marine oil spills, Biosurfactant production and biofilm formation Aerobic	(Koshlaf & Ball, 2017; Ron & Rosenberg, 2014; Ward et al., 2009)
	Halomonadaceae	Crude oil, n-alkanes Aromatics	Saline environment (marine or terrestrial),	(de la Haba et al., 2014; Koshlaf & Ball, 2017)
	Moraxellaceae	Crude oil, n-alkanes (C10 - C40), Volatile HC	Emulsification and dispersion	(Das & Chandran, 2011; Klein et al., 2007; Ward et al., 2009)
	Pseudomonadaceae	Crude oil, Alkyl aromatics, Volatile HC	Ubiquity and rhizosphere PGPR and biosurfactant production, Emulsification and dispersion, Aerobic or anaerobic chemoorganotrophy	(Chaudhry et al., 2005; Das & Chandran, 2011; Dazzo & Ganter, 2009; Garrity et al., 2015; Koshlaf & Ball, 2017; Ward et al., 2009)
Rhizobiales	Beijerinckiaceae	Crude oil, EDTA	Soil and water habitats, Phyllosphere and PGPR, Aerobic,	(Marín & Arahal, 2014)

			Facultative methylotrophy, Chemoorganoheterotrophy,	
	Rhizobiaceae		PGPR	(Chaudhry et al., 2005; Ward et al., 2009)
	Xanthobacteraceae	Chlorinated or brominated Alkanes, Alkenes, Aromatics, Terpenes, thiophenes, (and more)	Ubiquity in wet soil and sediment, Plant associated, Aerobic chemoheterotrophy or facultative chemolithotrophy, Nitrogen fixation	(Oren, 2014b)
Rhodobacterales	Rhodobacteraceae		Aquatic and marine, Sulphur and carbon biogeochemical cycling, Aerobic photo or chemoheterotrophy	(Pujalte et al., 2014)
Sphingomonadales	Sphingomonadaceae	Aromatics, Volatile HC	Ubiquity, Freshwater or marine, soil, rhizosphere and phyllosphere, Emulsification and dispersion, Aerobic or anaerobic chemoorganotrophy	(Das & Chandran, 2011; Glaeser & Kämpfer, 2014; Koshlaf & Ball, 2017; Ward et al., 2009)
Thiotrichales	Piscirickettsiaceae	Aromatics	Marine oil spill	(Koshlaf & Ball, 2017; Ron & Rosenberg, 2014)

1.3.1.1 Strategies to Implement Microbial Bioremediation

The simplest method of microbial bioremediation is known as natural attenuation. The contamination is simply allowed to degrade naturally in the environment by native microbes, with regular monitoring of the process. However, this process is only suitable for low to moderate contamination and can take up to years. Environmental conditions such as temperature, nutrients or soil organic matter content have a significant impact on the effectiveness of the process. Tilling or harrowing can be used to improve soil aeration and subsequently aerobic HC degradation (Table 1). However, the presence of indigenous and metabolically active microbes, together with the necessary metabolic pathways, is essential for effective degradation of petrogenic HCs (Koshlaf & Ball, 2017).

The process of introducing bacterial strains with the capacity to degrade HC (Table 1) to improve bioremediation in situ is referred to as bioaugmentation (Das & Chandran, 2011). The introduced strains are either isolates from other contaminated sites or genetically modified (GM) organisms. However, the strong influence of environmental factors remains and studies have shown mixed results and rarely successful field applications (Chaudhry et al., 2005; Koshlaf & Ball, 2017; Kuiper et al., 2004; Viebahn et al., 2009). The majority of GM strains show a rapid decline in survival after introduction and exhibit low metabolic activity. This may be because the expression of the inserted genes in GM bacteria may require more energy, or because native bacteria adapted to the environment simply outcompete the GM bacteria (Gerhardt et al., 2009; Viebahn et al., 2009).

One method of enhancing the environmental conditions on the site and supporting the native microbial population is the addition of nutrients and electron acceptors. This process is referred to as biostimulation or landfarming (Das & Chandran, 2011; Ward et al., 2009). Fertilizers can be natural, such as manure or compost (Koshlaf & Ball, 2017) or artificial, such as the nitrogen-rich fertilizers Inipol EAP 22 and Customblem (Bragg et al., 1994; Ron & Rosenberg, 2014). The efficacy of both fertilisers and other treatments has been demonstrated on contaminated shorelines affected by the Exxon Valdez oil spill. However, the success of these treatments has been observed only in the aerobic layer, and the outcome is influenced by a range of environmental parameters (Baker et al., 1993; Bragg et al., 1994; Das & Chandran, 2011; Koshlaf & Ball, 2017; Kuiper et al., 2004). Furthermore, it is not always evident which nutrient is the limiting factor (Kuiper et al., 2004). In contrast, excess nutrients, particularly nitrogen, can impede the degradation process (Das & Chandran, 2011; Koshlaf & Ball, 2017). However, it is not always advisable to use this method for oil spill clean-up management, as it requires close monitoring and a high level of familiarity with the specifics of the site and its indigenous microbes (ITOPF, 2014).

Knowing the bacterial composition is important knowledge for any application of microbial bioremediation. However, identifying and analysing the microbes involved in HC degradation is not always straightforward. Only a small percentage of bacteria can be cultured, and it is particularly difficult to isolate endosymbionts and other plant-associated bacteria (Brenner & Farmer III, 2015). In addition, studying bacteria in culture can be misleading because they adapt so quickly to culture conditions. Instead, a common approach is DNA sequencing.

1.3.2 Phytoremediation

A less understood, yet promising approach is the use of plants for bioremediation. Several processes are relevant for phytoremediation, which can occur simultaneously. Phytotransformation describes the ability of a plant to break down a compound, either by metabolic processes or the production of enzymes (Gerhardt et al., 2009). Some plants able to degrade even complex compounds such as polychlorinated biphenyls, nitroaromatics or trichloroethylene (Meagher, 2000). Physically restricting the movement of a contaminant is called phytostabilisation. This process can be particularly important on coastlines to prevent offshore oil spills from reaching the mainland (Gerhardt et al., 2009; Ward et al., 2009). The storage of a contaminant is referred to as phytoextraction, while its transport through the plant is referred to as phytovolatilization. However, both processes leave the contaminant unaltered (Gerhardt et al., 2009; Ward et al., 2009). However, contamination hinders plant growth, and most plants will die at high concentrations (Baker, 1970; Gerhardt et al., 2009; Ward et al., 2009). This depends very much on the plant species and the type of contamination and exposure. For example, many grasses wilt and die when exposed to petroleum through the leaves, but survive when only the roots are exposed (Baker, 1970). In general, the lipophilicity of petroleum products can disrupt cell membranes, allowing oil to enter cells and leak cell contents. It can also block stomata and intercellular spaces, reducing photosynthesis, transpiration and translocation rates (Baker, 1970). Therefore, one of the most important mechanisms provided by plants, is phytostimulation, which refers to the enhancement of the associated rhizosphere microbiome. The rhizosphere displays the highest microbial activity and is the main site for bioremediation (Dazzo & Ganter, 2009).

1.3.3 Rhizoremediation

The rhizosphere is the soil directly affected by the plant's root system. That includes the endorhiza (interior root) as well as the rhizoplane (root surface). In contrast to this is the soil beyond the influence of the roots (Figure 2), which is also referred to as the bulk soil (Dazzo & Ganter, 2009). The main role of plants is to provide optimal conditions for microbial activity in the rhizosphere. The root system provides additional structural support and habitat for the microbes, while aerating the soil. Most importantly, living roots secrete water-soluble exudates such as sugars, amino acids, organic acids and vitamins, known as rhizodeposition. In addition, as the plant grows, dead cells are shed from the epidermis or root cap and root mucilage can be deposited. Carbon dioxide from respiration or other volatile compounds may also be exuded from the roots. As a result, the abundance and activity of microbes in the rhizosphere is greater than in the bulk soil, a phenomenon called rhizosphere effect (Figure 2) (Chaudhry et al., 2005; Dazzo & Ganter, 2009; Kuiper et al., 2004; Ward et al., 2009).

Rhizoremediation is also referred to as microbe-assisted phytoremediation. It is the most effective bioremediation method, as the interactions between plant and root microbiome (Figure 2) can overcome the limitations of microbial bioremediation and phytoremediation. The microbes involved are typically classified as either plant growth promoting rhizobacteria (PGPR) or active contaminant degraders (Table 1) (Gerhardt et al., 2009). The main limitation to microbial biodegradation is the availability of nutrients. Root exudates provide a readily available nutrient source for microbes. In addition, the formation of chelates between root

exudates and toxic heavy metals can occur, reducing the overall toxicity of the environment (Chaudhry et al., 2005). Bacterial degradation of petrogenic HC is most effective under aerobic conditions. As the root system develops, it creates channels in the soil and loosens the soil structure, allowing additional oxygen to reach the bacteria (Chaudhry et al., 2005; Gerhardt et al., 2009). For microbes to take up a substrate, contact with the outer membrane is required. Physical retention of nutrients and contaminants by roots (phytostabilisation) can therefore enhance the degradation of HC (Ron & Rosenberg, 2014; Ward et al., 2009). In particular, the degradation of recalcitrant compounds, such as PAHs, often requires the involvement of multiple metabolic pathways. The high abundance and diversity of microorganisms in the rhizosphere allows for the development of diverse and complementary communities that engage in co-metabolism (Chaudhry et al., 2005). The uptake of compounds with a slightly hydrophobic character by plants is optimal, and strongly soil-absorbed nutrients or HC are inaccessible. Plants are also less tolerant of contaminants than microbes (Baker, 1970). Microbial oxidation of petrogenic HCs reduces their toxicity while increasing their availability to plants. Microbes, and especially bacteria, also increase the solubility of soil-bound or otherwise unavailable forms of nutrients, such as phosphorus. Alternatively, they may be capable of nitrogen fixation (Chaudhry et al., 2005; Dazzo & Ganter, 2009; Gerhardt et al., 2009; Koshlaf & Ball, 2017). Bacteria, in particular, can also act as biological control agents for plant pathogens, either by stimulating the plant's natural defences or by targeting microbial plant pathogens (hyperparasitism). However, plant-microbe interactions can also be detrimental to the plant. This can happen when the bacterium becomes a plant pathogen or when the microbe competes with the plant for nutrients (Dazzo & Ganter, 2009; Roquigny et al., 2017).

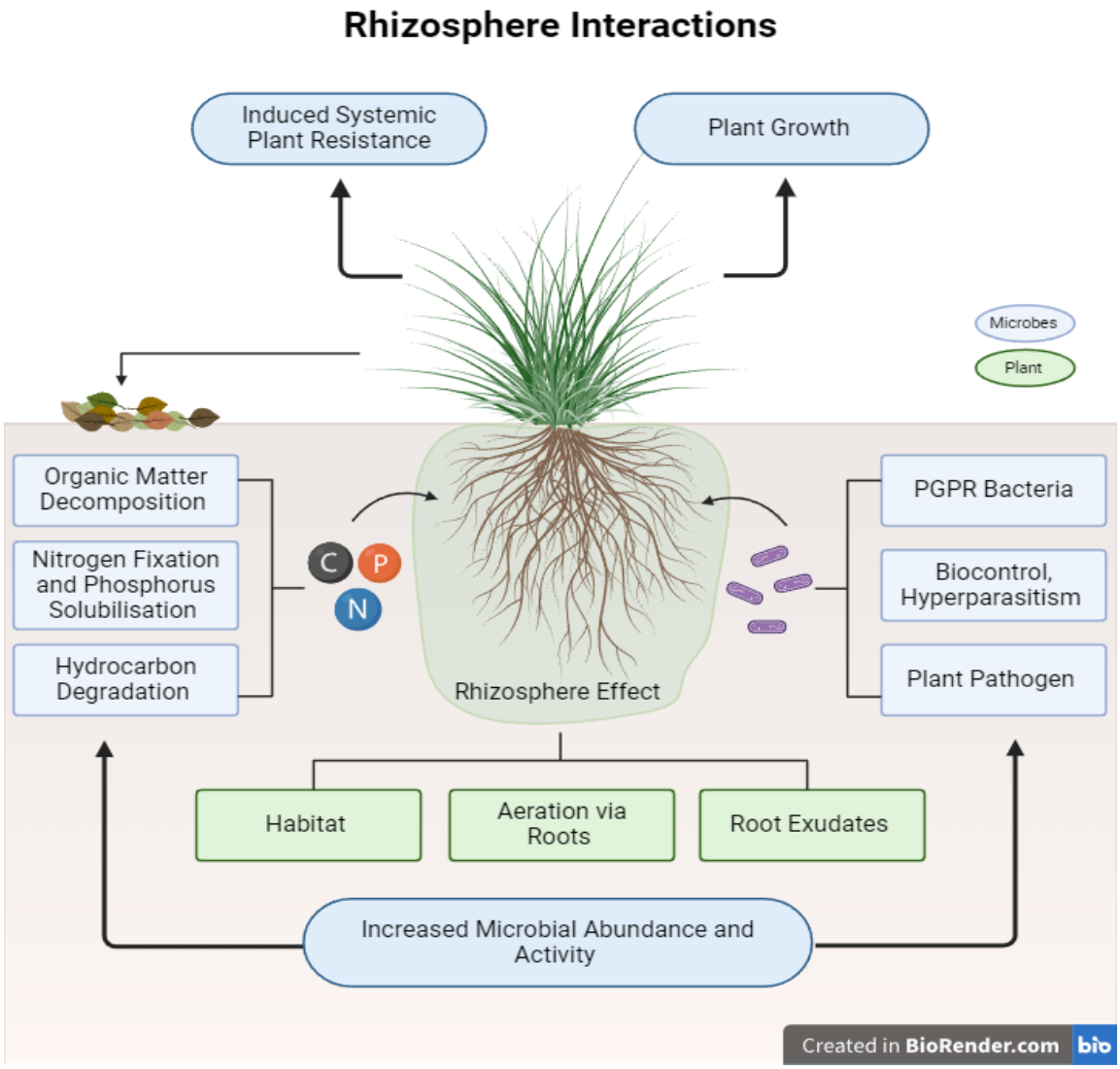

***Figure 2**: Schematic illustration of the most important plant-microbe interactions in the rhizosphere*

While the introduction of petrogenic HC can alter the microbial community, the plant itself is the most significant factor in forming a rhizosphere community, exceeding the influence of nutrients or soil structure (Ribeiro et al., 2013). The release of root exudates has the potential to alter the pH of the soil, the rate of water flux, and the availability of oxygen in the rhizosphere, and is a direct response to environmental stressors. Furthermore, the composition of exudates can be actively modified by the plant to create an optimal environment for specific microbes. For example, this can lead to an increase in the abundance and activity of HC degrading bacteria. This process is called rhizosphere breeding (Chaudhry et al., 2005). It is possible that these complementary interactions, facilitated by the plant, represent the plant's response to stress (Chaudhry et al., 2005). Nevertheless, these co-evolutionary processes remain poorly understood (Konopka, 2009). Therefore, for optimal rhizoremediation, both the plant and the bacteria need to be considered (Kuiper et al., 2004). Possible pairs are plants with PGPR bacteria, or plants with HC degrading bacteria (Table 1).

1.4 Role of Coastal Marshes

The vegetation of coastal marshes has important ecological functions, especially in protecting the shoreline. Plants increase the friction of incoming waves, reducing their energy and leading to wave attenuation. The roots provide structure and prevent erosion, while the upper parts of the plant increase the sedimentation of particles in the water, stabilising the whole shoreline. Plants can retain water, while their roots provide better drainage, leading to better flood attenuation (Shepard et al., 2011).

In addition, in case of an oil spill reaching the shoreline, the local vegetation can trap the oil (phytostabilisation) and stop it from reaching the inland. This means that shoreline vegetation can act as a first line of defence as well as an important bioremediation site. Plants in the intertidal zone are often halophytic and more robust than other plants, and many have been shown to have bioremediation potential (Couto et al., 2011). A common plant of intertidal zones and shorelines are grasses, such as the temperate grass *Puccinellia* sp. However, not all halophytic plant species are suitable for bioremediation, as their resistance to oil contamination is species-specific (Cowell, 1969; Hershner & Lake, 1980). Nevertheless, grasses have great potential for bioremediation. They tend to have many fine roots that can spread over a large area and provide many habitats for bacteria (Chaudhry et al., 2005; Das & Chandran, 2011; Kuiper et al., 2004). The size of the root system is proportional to the size of the rhizosphere effect (Chaudhry et al., 2005; Dazzo & Ganter, 2009). For instance, the combination of diverse grasses and Pseudomonas strains has demonstrated the capacity to mitigate environmental stressors associated with contamination (Gerhardt et al., 2009). However, grasses can take up oil via their crowns, which can make them wither or die, while their roots are more resistant (Baker, 1970; Cowell, 1969). To date, bioremediation, especially the addition of microbes or plants, is rarely used on shorelines because of the lack of control due to high variability. Instead, after an initial clean-up with mechanical or chemical methods, shorelines are often left to natural attenuation (ITOPF, 2014). To fill this knowledge gap, more knowledge is needed about the intertidal and plant-associated microbiome.

1.5 Study Aim

Bioremediation, in particular rhizoremediation, is a cost-effective and versatile method for reducing petroleum contamination in the environment. It requires knowledge of native and resilient plants and their associated microbiome. Especially accidental oil spills pose a great threat to shorelines, but bioremediation is not commonly applied.

The aim of this study is to assess the bioremediation potential for petroleum HC of the common intertidal grass *Puccinellia* and its rhizobiome. The bacterial community will be identified using in-house DNA sequencing of the 16 S gene. The effectiveness and appropriate modifications of the method will be evaluated. The influence of the experimental design on community composition will also be assessed. Changes in diversity and community composition over time and treatment will be described. As part of this, the occurrence and development of the bacterial rhizoremediation consortium over time and treatment will be described. It is expected that 1) the community composition will change in response to diesel pollution, and 2) that the community composition displays a difference between rhizosphere and bulk soil.

2 Methods

2.1 Study Site and Sampling Design

Sampling took place in spring 2021 on the Tisnes peninsula (69°36'12.7"N 18°49'55.6"E) southwest of Tromsø, northern Norway, and the sampling site was chosen for its remote location (Figure 3A). It is a protected area (Norwegian: Naturreservat), which means that human disturbance is prohibited. Therefore, pre-exposure to anthropogenic HCs is expected to be negligible, making the area a good 'pristine' reference compared to the experiment. Nine plants of the halophytic grass Puccinellia sp. including their surrounding soil, and three buckets of soil were collected. After transportation to the Climate Laboratory Holt, Tromsø, plants were acclimated and maintained in buckets with holes. The buckets were placed in larger trays containing salt water and holes at a certain height to keep the water level at a certain threshold (Figure 3B). Whenever the water level became low, the trays were refilled with a solution of 35 g of salt per litre of tap water of unknown temperature. This corresponds to a 3.5% seawater solution, or 'standard seawater' (Millero et al., 2008).

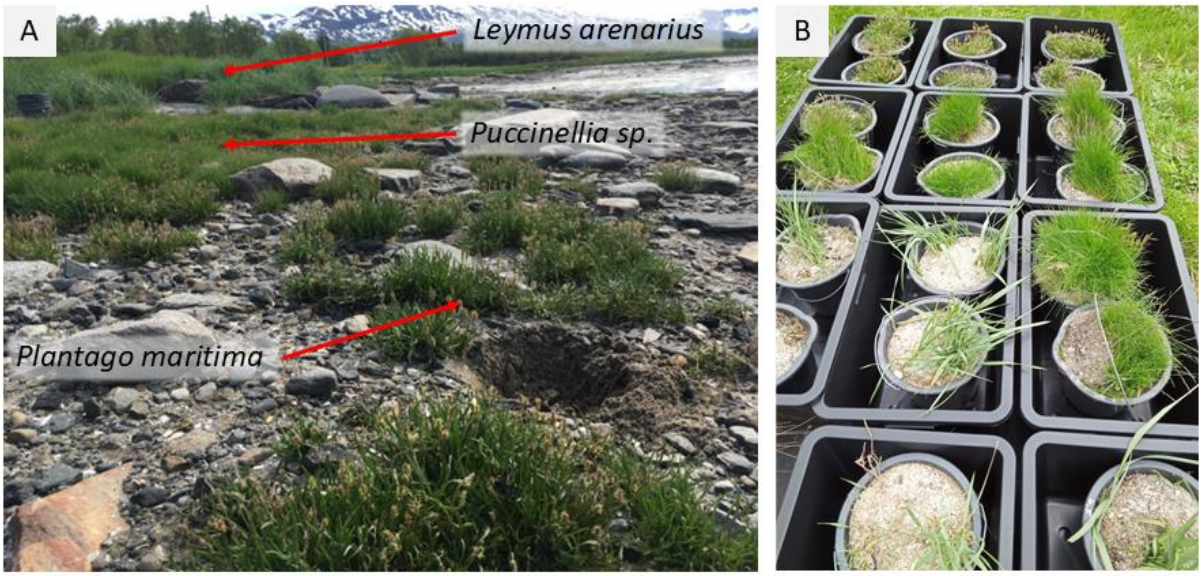

***Figure 3**: **A** – Picture of the sampling site on Tisnes showing the zonation of the original three study plants. **B** – Experimental setup in Tromsø, seen are buckets for all plants.*

The plants and soil buckets were then divided into three diesel contamination treatments: Control, Low (0.5% w/v) and High (1% w/v) and each treatment contained three plants and one soil bucket. The diesel was poured directly onto the soil without touching the leaves. Small cores were taken from the rhizosphere samples at four time points, week 0 (immediately after diesel addition), week 1, week 7 and week 13. In addition, three extra rhizosphere samples and one soil sample were taken from the original sampling site at Tisnes, at week 1 and again at

week 13, giving a total of 56 samples (Figure 4). Samples were collected in 50ml Falcon tubes and stored at -20°C until laboratory work commenced in spring 2023.

***Figure 4**: Scheme of the experiment setup.*

2.2 DNA Extraction

2.2.1 Soil Homogenisation

Before extracting DNA, the soil was homogenised by placing a few teaspoons of frozen soil into a tissue grinder (10 ml stainless steel Grinding Jar Set with grinding balls, Qiagen™). To ensure the sample will stay frozen during homogenisation, the jar was submerged in liquid nitrogen before putting it in the tissue lyser (TissueLyser II, Qiagen™) for at least two minutes at 22.0 Hz. The homogenised soil was then placed in a 15 ml Eppendorf tube and kept frozen. Between samples, all the used equipment was washed with water, sprayed down with 70 % ethanol, and then dried and sterilized under UV light (UVC 500 UV crosslinker, Hoefer). Approximately 0.2 g of homogenised soil was weighed (Sartorius CP225D Semi-Micro Balance, DWS) into a lysis tube (Matrix E lysis tube, MP Biomedical™) and stored at - 20 °C.

2.2.2 DNA Extraction

The centrifuge (Centrifuge 5415 R, Eppendorf™) was pre-cooled to 4 °C and always used at 13,000 g. In Appendix 1, one will find recipes for buffers that were not purchased premade.

First, 500 µl of phenol:chloroform:isoamylalcohol (PCI) in a 25:24:1 solution (Qiagen™) and autoclaved TRIS-NaCl-SDS (TNS) buffer were added to the lysis tube, and homogenised using a bead beater (FastPrep-24® Classic, MP Biomedicals™) for 30 seconds at 5.0 m/ s. The samples were centrifuged for 10 minutes, and the supernatant transferred to a fresh 2 ml tube (RNase/DNase free, Eppendorf™). This step was repeated two more times with changed volumes for PCI and TNS. For the second time 300 µl were used, for the third 200 µl, and the supernatant of each step pooled together. The solution was extracted with 750 µl of chloroform:isoamylalcohol (CI) in a 24:1 solution (Qiagen™) and centrifuged for five minutes. Then, 350 µl of the resulting supernatant was transferred into a fresh 2 ml tube (RNase/DNase

free, Eppendorf™). Next, 700 µl of PEG-6000 buffer and 5 µl of linear acrylamide (Ambion™) were added, the samples vortexed and centrifuged for 60 minutes. All fluids were carefully decanted without disturbing the pellet. Then, 1 ml of ice-cold 70 % ethanol was added and centrifuged for 10 minutes. The ethanol was decanted, and the step repeated. Leftover ethanol was pipetted out and the samples dried (Thermomixer comfort, Eppendorf™) at 50 °C for a few minutes until the ethanol had evaporated. The pellet was eluted in 50 µl Milli-Q® (Millipore®) water by gentle up and down pipetting. Finally, 50 µl RNase stock solution containing 5 µl Recombinant RNase A (Ambion™) with 2 ml Milli-Q® water was added. Samples were vortexed, centrifuged briefly and stored at -20°C until further analysis.

2.2.3 DNA Verification

To verify the amount of extracted DNA, NanoDrop® (ND-1000 Spectrophotometer) was used. The thawed samples were vortexed and briefly centrifuged (SPROUT™ Mini Centrifuge, Heathrow Scientific), then stored on ice. Milli-Q® water was used to clean the contacts and to take an initial blank measurement. Each sample was measured twice using 2 µl, and the contacts were cleaned with lens paper after every measurement. At regular intervals, the NanoDrop® (ND) was checked for correct measurements using a water (blank) sample. The DNA amount (ng/ µl) and the absorbance ratio at 260/ 280 nm were recorded.

To verify the quality of the extracted DNA, gel electrophoresis (GEP) was conducted. The agar was prepared using 0.8 % or 1 % agarose (A9539, Sigma-Aldrich) in 1X Tris-Acetate-EDTA (TAE) buffer (T9650, Sigma-Aldrich) according to the tray size and 0.5 µl of GelRed® Nucleic Acid Gel Stain (Biotium) for every 10 ml of buffer. The amount of ladder (GeneRuler 1 kb DNA Ladder, ThermoFischer Scientific™) and loading dye (6X) were calculated according to the previously with ND measured amount of DNA in the sample and topped off with Milli-Q® water to the desired amount. The gel ran at 60 V for approximately one hour. Samples that did not show a proper band at ~1500 bp were re-extracted. For those samples that were unsuccessful after several DNA extractions, three extractions were pooled together. However, it was still impossible to extract DNA from some samples.

2.3 Library Preparation

A library for sequencing is the DNA from the sample that has been converted through several steps into molecules suitable for the sequencing method. The 16 S Barcoding Kit 1-24 (SQK-16S024) from Oxford Nanopore Technologies (ONT) was used for sequencing. In this protocol (Appendix 2), a polymerase chain reaction (PCR) was required to amplify the 16 S gene and simultaneously attach barcodes (BC) for multiplex sequencing. The primers and barcodes were delivered as one (BC primers) and the PCR was purified using magnetic beads (AMPure XP beads, Beckman Coulter). The amount of DNA in each sample was measured using Qubit® (2.0 Fluorometer, Invitrogen) and ultimately the samples were pooled into a final sequencing library according to their barcodes (1 to 24) and their amount of DNA.

However, after the first PCR using barcode primers (BC PCR) and DNA purification, the amount of DNA in the samples was not sufficient to proceed to sequencing. To improve the DNA yield, some changes were made to the protocol.

2.3.1 Protocol Optimisation

The problems encountered, their presumed causes, and the measures attempted during protocol optimisation are summarised (Table 2). In some cases, an iterative approach was necessary to achieve adequate results. Based on the observations from the applied measures during the protocol optimisation, changes to the original protocol (Appendix 2) were made.

***Table 2**: Issues during the library preparation, the possible reasons and the applied measures to find a solution.*

Problem	Possible Reason	Measures and Solutions
Low or undetectable DNA after purification	Low detectivity from Qubit	Using the high sensitivity (HS) kit
	Inadequate purification protocol	Increasing pelleting time
		Increasing incubation time and temperature
		Increasing ethanol percentage during DNA washing
	Unsuccessful amplification	Verifying PCR with GEP before purification
No or faint GEP bands	Inadequate sample amount	Testing different sample dilutions
	Detection limit of GEP	Using TBE buffer and EtBr stain
	Unsuccessful amplification	*See below*
Only GEP smears	Inadequate DNA amount	Testing different sample dilutions
	Insufficient GEP conditions	Making fresh buffer
	Contaminated DNA (salt, protein)	Testing with different lab replicate
	Fractured DNA	Increasing agarose percentage
	Unsuccessful amplification	*See below*
Unsuccessful BC PCR amplification	Expired/ non-working reagents	Using a fresh BC kit and Master Mix
	Non-amplifiable DNA	Retesting samples with 16 S primers (fD1, rD1) and separate reagents
	Inadequate primer amount	Testing different primer concentration
	Contaminated DNA, obstructed reaction	Testing different DNA dilutions
		Adding magnesium chloride ($MgCl_2$)
	Too specific annealing temperature	Lowering annealing temperature

2.3.2 Improved Protocol

The reagents for the barcoded PCR were mixed according to the following formula (Table 3).

***Table 3**: Formula for the PCR with barcode primers.*

Reagent	Volume
Nuclease free water	11 µl
LongAmp Hot Start Taq 2X Master Mix (New England BioLabs)	25 µl

$MgCl_2$ (25 mM stock solution, Roche Diagnostics GmbH)	3 µl
Input DNA/ sample (1:10 dilution)	1 µl
16S Barcode Primer (SQK-16S024, Oxford Nanopore Technologies)	10 µl
TOTAL	50 µl

Each of the reagents was thawed and well mixed before use. For each sample, the PCR reaction was prepared in a separate 0.25 ml thin-walled PCR tubes (Fisherbrand). To access the BC primers, each foil covering the well was carefully pierced with a fresh pipette tip and carefully mixed by pipetting. Then, 10 µl of BC primers were transferred into the respective sample containing PCR tube and mixed by pipetting. Amplification was done in a thermocycler (T Professional Thermocycler, Biometra or Mastercycler® ep, Eppendorf) using the following cycling conditions (Table 4).

***Table 4**: Cycling conditions during BC PCR amplification.*

Cycle step	Temperature (°C)	Time (min)	Cycles
Initial denaturation	95	1:00	1
Denaturation	95	0:20	25
Annealing	55 / 50	0:30	25
Extension	65	2:00	25
Final extension	65	5:00	1
Hold	4	∞	

Then, 5 µl of PCR product were used to perform a GEP, using 1 % agarose, TAE buffer and GelRed®, as described above. Preferred, samples with a clear band at ~ 1500 bp were used for sequencing.

The remaining 45 µl of BC PCR product were transferred into a fresh 2 ml tube (Eppendorf) before adding 27 µl of re-suspended AMPure XP beads to the sample. After mixing by pipetting, the mixture was incubated at room temperature for around 5 minutes on a Hula mixer (Rotator, Helmut Saur Laborbedarf). The amount of beads was lowered from the original protocol (Appendix 2) to accommodate the reduced amount of PCR reaction. The samples were quickly spun down (SPROUT™ minicentrifuge, Heathrow Scientific) and kept on a magnetic stand (Magna-Sep™, Invitrogen) until a stable pellet had formed. The supernatant was removed using a pipette while keeping the tube on the magnetic stand. Each tube was then washed for up to a minute using 200 µl of freshly prepared 75 % ethanol (puriss. p.a., absolute, ≥ 99.8 % Ethanol, Sigma-Aldrich with nuclease-free water). During this process the pellet was not disturbed. The ethanol was pipetted out and the step repeated. After a short spin with the mini centrifuge, the residual ethanol was pipetted out, and the sample was allowed to dry for up to a minute but without letting the pellet crack. The pellet was then resuspended in 10 µl of 10 mM Tris-HCL pH 8.0 with 50 mM NaCl by gentle flicking the tube. The samples were then incubated for around 10 minutes at 30 °C. After that, the sample was pelleted on the magnet

until the eluate was clear and colourless. At last, 10 µl of the eluate were transferred into a fresh 2 ml tube (Eppendorf).

The amount of DNA in each purified sample was quantified with using the Qubit™ dsDNA High Sensitivity (HS) Assay Kit. Based on the measurements, samples were pooled together to approximately equal ratios in 10 mM Tris-HCl. From that pooled library, 10 µl were transferred into a fresh tube before adding 1 µl of rapid adapter (RAP) from the kit. To minimise DNA absorption into the walls, a 1.5 ml Eppendorf DNA LoBind tube was used for pooling the libraries. Each final library contained barcodes 1 to 24, but if some 'high quality' samples had the same barcode, they were used in the next library and the free spot was filled with a lower quality sample, i.e. low or undetectable DNA according to Qubit®.

2.4 Sequencing and Alignment

For sequencing, the MinION™ device (Mk1B) in combination with the Flongle Adapter (ADP-FLG001) from ONT was used. The flow cell was loaded and prepared without changes to their protocol (Appendix 2). The sequence run was started from the MinKNOW software, and standard settings were used (Table 5).

***Table 5**: MinKNOW settings for each sequencing run.*

Run options	
Run limit	24 hours
Minimum read length	200 bp
Adaptive sampling	off
Analysis	
Basecalling	On (high-accuracy, 450 bps)
Barcoding	On
Alignment	Off
Output	
Basecalled reads	On (.FASTQ, Every 10 minutes, Split files by Barcode, GZip)
Raw reads	On (.FAST5)
Read filtering	Qscore: 9 | Read length: Unfiltered

An initial basecalling was performed in real-time during sequencing (Table 5). However, to improve accuracy it was re-done using Guppy for cpu (version 6.5.7) from ONT. Used were the raw reads, with the configuration file ‘dna_r4.9.1_450bps_hac.cfg’. The samples were demultiplexed and, the barcodes, the adapters and the primers trimmed.

Sequence alignment was done with EPI2ME software (version 5.1.14) by ONT. The workflow was ‘wf-16S’. Sequences were aligned using the Minimap2 classifier to the SILVA (version 138.1) database at genus level. Threshold for identity was 95 % and 90 % for coverage. If a sequence could not be matched to a genus, the lowest common ancestor (LCA) was used. Results were reported as abundance table.

2.5 Data Analysis

All samples were used for analysis, including biological and laboratory replicates. However, they were pooled together to increase the number of reads per sample. Analysis was done on family level, and unidentifiable (unknown) bacteria were excluded for the analysis. Data analysis was carried out using R version 4.4.1 (R Core Team, 2024) with RStudio version 4.3.3 (Posit Team, 2024). Data cleaning and transformation were done using the 'tidyverse' package version 2.0.0 (Wickham et al., 2019). Community analyses, such as rarefaction curves and diversity indices, were carried out using the 'vegan' package version 2.6-6.1 (Oksanen et al., 2024). The script, as well as the other supportive packages are shown in Appendix 5. As the resulting number of reads were not sufficient enough for multivariate community analysis (Kuczynski et al., 2010), analysis of the community composition will be primarily descriptive. Moreover, accurate quantification of abundance is not attempted.

3 Results

3.1 DNA Extraction

According to ND, the mean amount of DNA is 8.9 ± 11.89 ng/ µl and the ratio of 260/ 280 nm is 1.53 ± 2.82 for all treatments together. Table 6 shows the mean amount of DNA and the mean 260/280 nm ratio for each treatment. For calculation, only the successfully sequenced samples (n = 68) were used. The full table including every DNA measurement is in Appendix 3.

***Table 6**: Nanodrop results for the mean amount of DNA and its standard deviation (SD) after extraction and the associated 260/ 280 nm ratio and its standard deviation (SD).*

	Plant Tisnes	**Plant Control**	**Plant Low**	**Plant High**
DNA (ng/ µl)	11.91 ± 10.97	12.15 ± 11.26	12.91 ± 18.15	12.32 ± 11.14
260/ 280 nm	1.17 ± 1.57	1.16 ± 2.84	0.67 ± 2.90	1.68 ± 1.18
	Soil Tisnes	**Soil Control**	**Soil Low**	**Soil High**
DNA (ng/ µl)	2.18 ± 1.05	2.12 ± 1.05	2.06 ± 0.83	2.15 ± 1.03
260/ 280 nm	-0.04 ± 3.17	2.07 ± 1.58	2.52 ± 2.54	3.40 ± 4.96

3.2 Optimisation of Library Preparation

3.2.1 Barcode PCR

The expiration date for the ONT barcoding kits is around six months (Appendix 2). The kit as well as the Master Mix that was used for the first (unsuccessful) samples, was considerably past that but was kept at -20 °C continuously. Switching to fresh reagents, did not noticeable improve the amplification, according to GEP results (figure not shown).

The annealing temperature during a PCR is primer specific. The annealing temperature in the protocol (Appendix 2) was 55 °C. However, a lower temperature of 50 °C was tested. The PCR reaction was improved (figure not shown), and the less specific temperature kept for all following reactions.

Too much primer can inhibit the PCR reaction, so different BC primer concentrations were tested. None of the lower concentrations yielded a successful reaction (figure not shown), and the original concentration according to the protocol (Appendix 2) kept. However, many samples still did not amplify, therefore all samples were tested using the general 16 S primer fD1 and rD1 with the same Master Mix. From there, only successful were used for the BC PCR. The majority of samples were amplifiable with the general primers, but not all with the BC primers.

The low 260/ 280 nm ratio (Table 6) indicates contamination in the sample, therefore, different sample dilutions and the addition of $MgCl_2$ were tested. Figures 5 and 6 show the GEP scan of three different sample dilutions (1:10, 1:25, and 1:50), with and without $MgCl_2$ addition, and for general 16 S primers fD1 and rD1 (right) in comparison to the BC primer (left), for two different samples. Additional $MgCl_2$ improved the amplification and the best sample dilutions was 1:10, which corresponds to 1 µl of sample with 9 µl of nuclease-free water. The BC primer amplified not as well as the general 16 S primer.

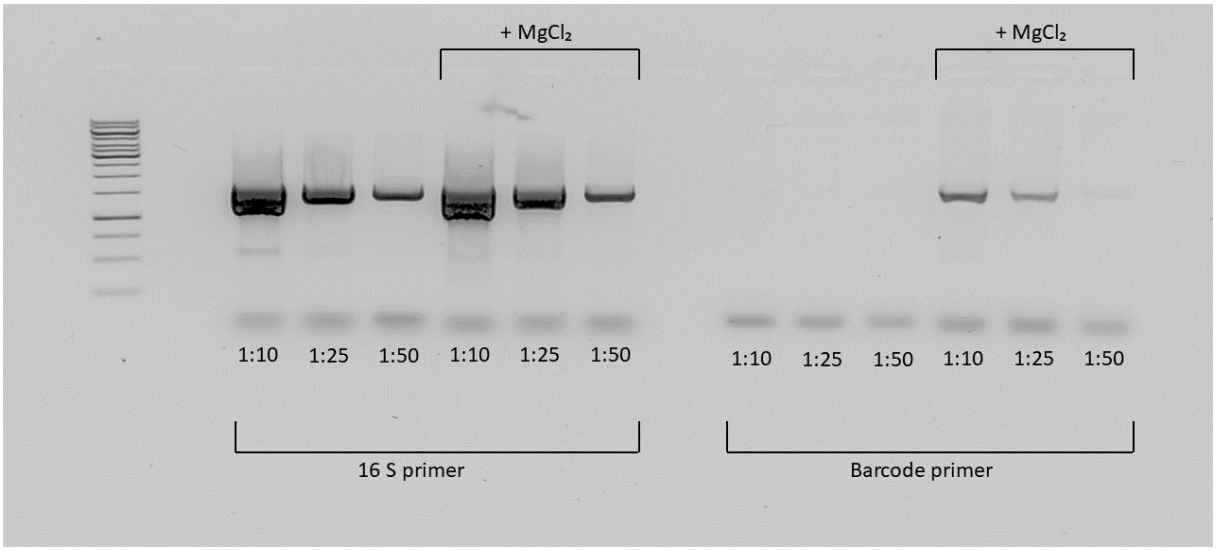

Figure 5: *GEP scan for the sample 3_PC_3a, using a 1kb DNA ladder. Testing the effect of $MgCl_2$ addition and different sample dilutions on the effectiveness of PCR reaction. Compared are the results between the general 16 S primer fD1 and rD1 (left) and the 16 S primer with barcode (right).*

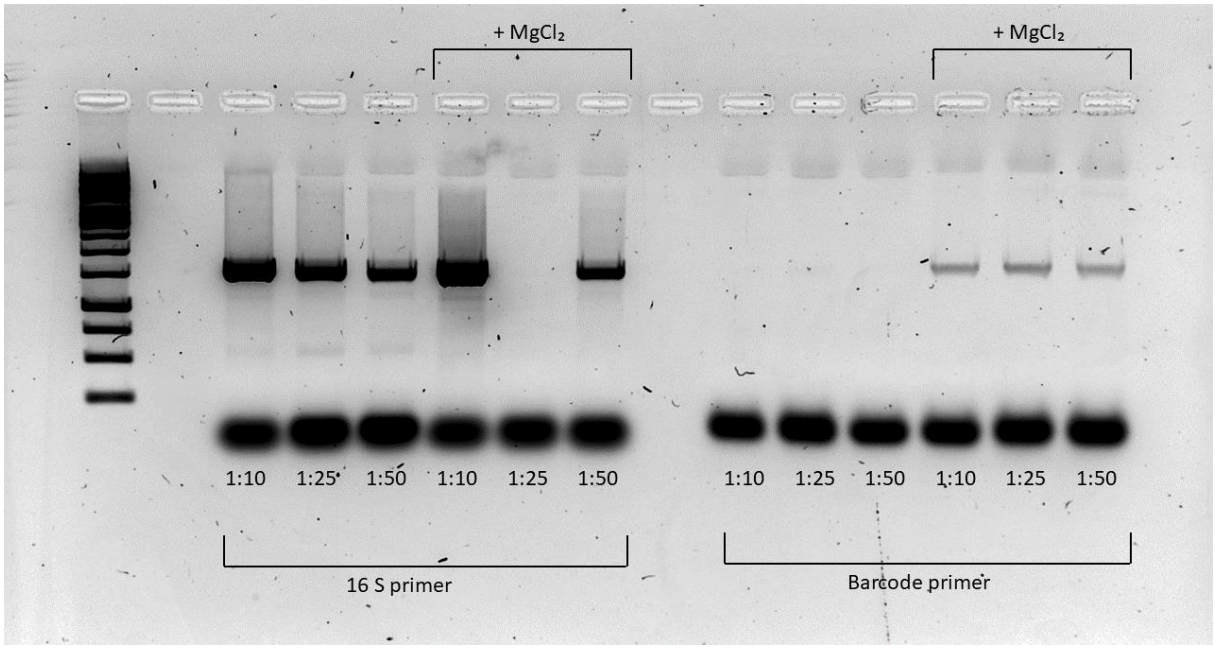

***Figure 6**: GEP scan for the sample 2_PH_3b, using a 1kb DNA ladder. Testing the effect of $MgCl_2$ addition and different sample dilutions on the effectiveness of PCR reaction. Compared are the results between the general 16 S primer fD1 and rD1 (left) and the 16 S primer with barcode (right).*

3.2.2 Gel Electrophoresis

To check if the BC PCR was successful, a GEP was performed with a small amount of the PCR product. As there were often no visible products, tests were carried out to make sure that the GEP was working correctly. The TAE buffer was freshly made and also replaced in the tray. There was no noticeable improvement. Using only 1 µl of PCR product often did not produce any bands on the gel (figure not shown), so 5 µl were used instead. As Tris-Borate-EDTA (TBE) buffer (T4415, Sigma-Aldrich) and ethidium bromine (EtBr) stain have a higher sensitivity this was also tested. However, it did not significantly improve resolution and was not continued (figure not shown). Therefore, failure to see PCR products using GEP is caused by the PCR product and not the GEP.

3.2.3 DNA Purification

Increasing the pelleting time during DNA purification, allowed the pellet to form and settle more stable, and was not as easily disturbed. The pellet was then washed with 75 % ethanol, instead of the recommended 70 % from the protocol. After resuspending the pellet in Tris-HCl, it was incubated for around 10 minutes at 30 °C, instead of two minutes at room temperature (Appendix 2). All changes were made at the same time and the DNA yield improved immediately; however, it is unclear which change was the most significant.

Moreover, using the high sensitivity (HS) kit from Qubit instead of the broad range (BR) kit, improved the accuracy for measuring DNA after purification, as DNA yield after purification was lower than expected.

3.3 Sequence Alignment

Figure 7 displays the number of reads and the percentage of identified families for all un-pooled replicates. The first number of sample ID is the week the sample was taken (0, 1, 7, or 13), the second number represents the biological replicate (1, 2, or 3), and the letter represent the laboratory replicate (a, b, or c). If a replicate was done twice, '.2' was added to the end. The bars are divided into identified (beige) and unknown (grey) families. The number above the bar represents the percentage of successfully classified families. There is no trend across treatment or time, the only difference is between plant and soil treatments. Plant treatments (top row) have more replicates and higher number of reads per sample ID than soil treatments (bottom row). In addition, a higher percentage of sequences were identified for plant treatments. On average, 70.27 ± 17.55 % of all reads were identifiable and assigned to a taxonomic level.

After sequencing, basecalling, and alignment, in total 68 samples passed the quality scores (Table 5 and Figure 7). Of these samples, the number of reads per sample varied strongly, between 6 and 1135 reads. Therefore replicates (biological and laboratory replicates) were pooled together to increase comparability in all subsequent in all other analyses.

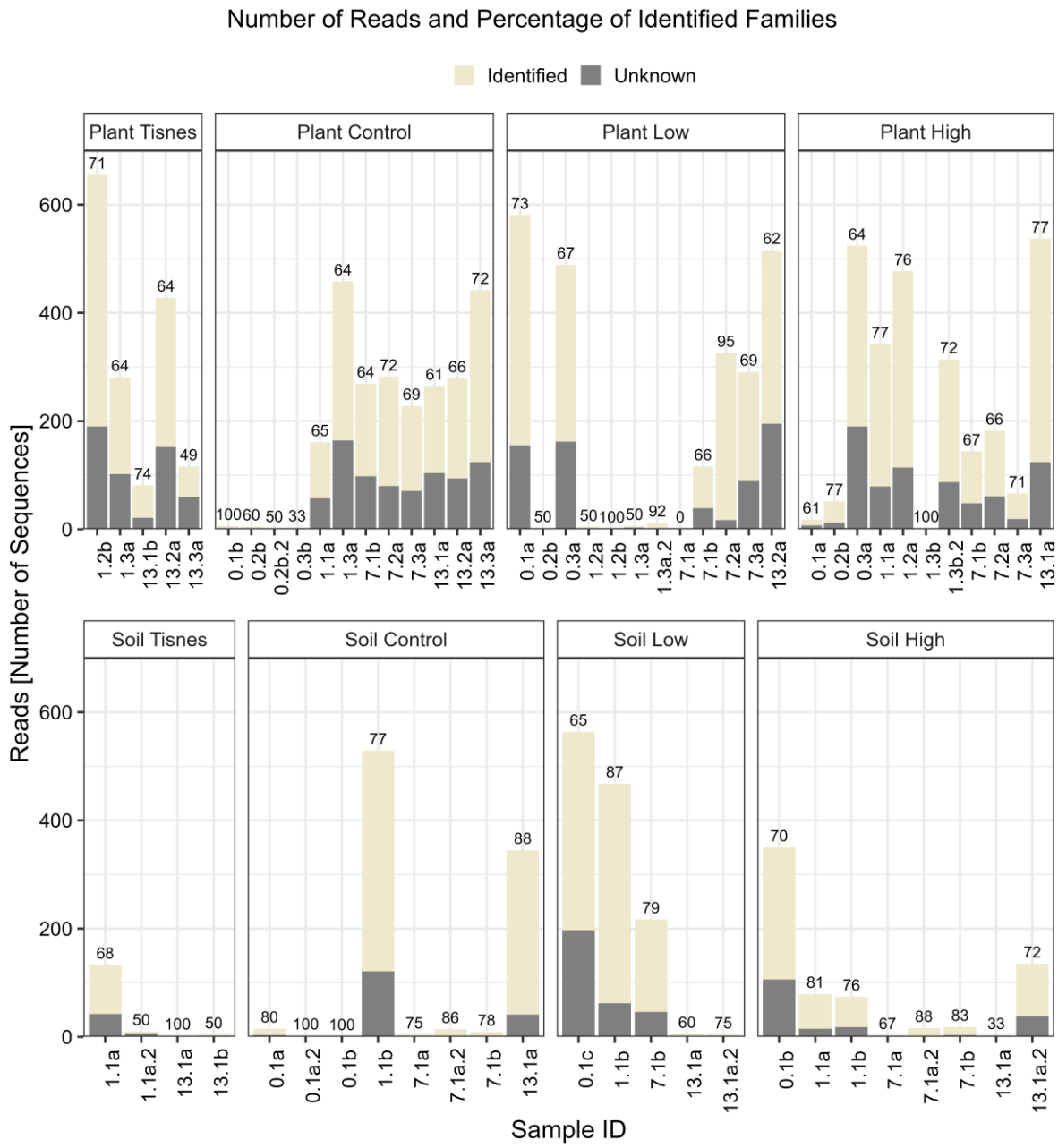

***Figure 7**: Number of reads for every replicate (bars) that passed sequencing (n = 68) for all treatments. The grey area represents the percentage of families that were not identified during alignment. The beige area, as well as the number above each bar, represents the percentage of identified families from alignment.*

Explanation of sample ID: first number identifies the timestep (0, 1, 7, 13), second number identifies the biological replicate (1, 2, 3), the letter identifies the laboratory replicate (a, b, c), and the last number identifies a duplicate.

3.4 Diversity Indices

3.4.1 Family Richness

Figure 8 displays the rarefaction curves for each timepoint (indicated by the number) in a treatment. The curve represents the relationship between the number of reads and the number of identified families for a pooled sample. A flattened-out curve means that more reads will not yield more families, e.g., the true number of families is reached. All except one curve (Soil Control, Week 13), are very steep and are not flattening out. Rarefaction curves of plant treatments (top row) reach higher number of families than curves in the soil treatments (bottom row). No trend is seen over time.

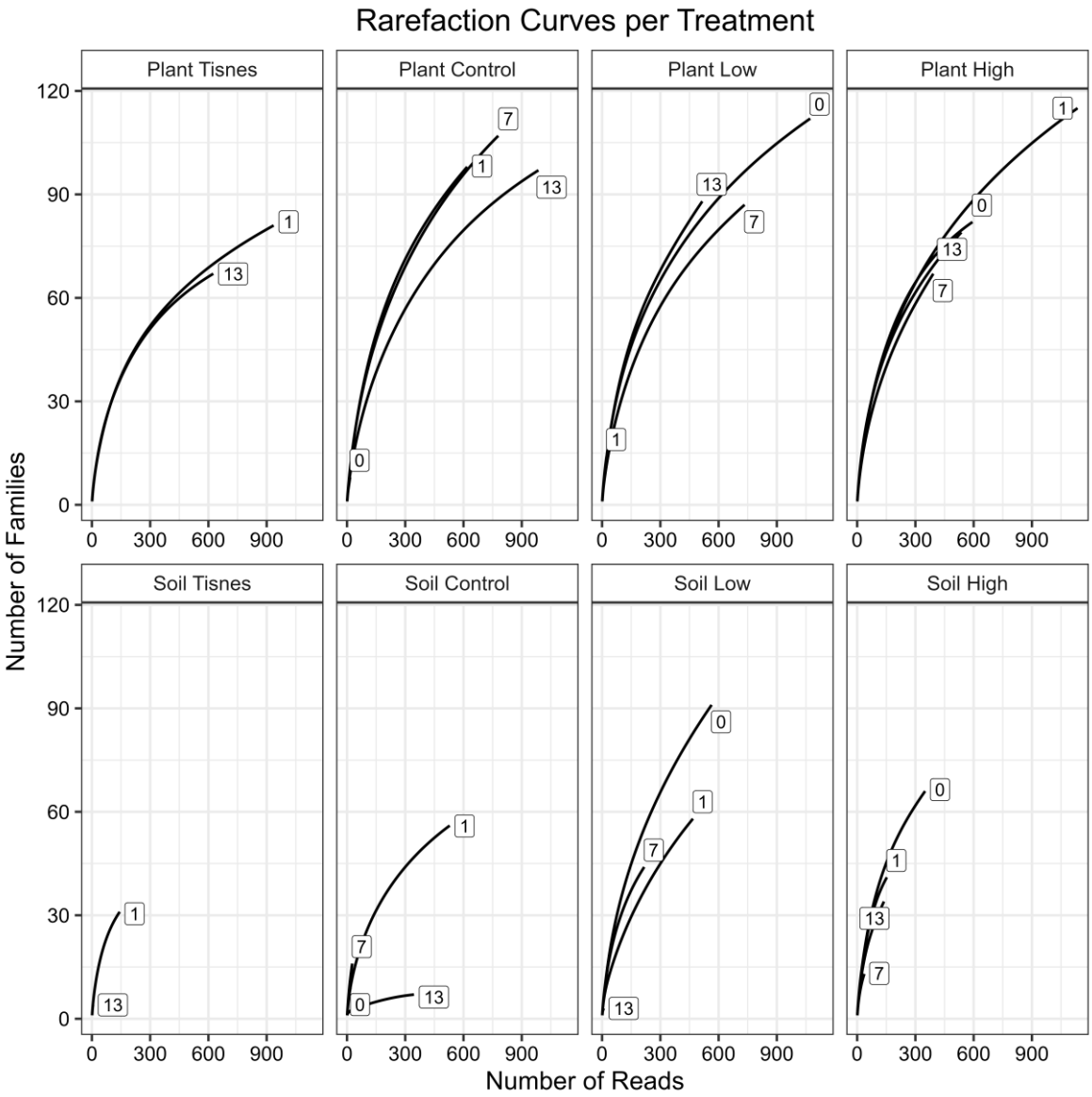

Figure 8: *Rarefaction curves for the pooled replicates in every treatment. Curves represent the theoretical sample size (= number of reads) necessary to reach a certain number of species (= bacterial families). Boxed numbers at the end of a curve represent its timestep.*

Rarefaction curves show if the true family richness is achieved, but not how high it is. Therefore, an estimate for family richness was calculated.

Figure 9 shows the observed (counted) family richness (green) and the estimated family richness (beige), as calculated by the Abundance-based Coverage Estimator (ACE) index for the pooled replicates, over time and treatment. The standard deviation (SD) for the ACE index is displayed as error bars. The estimated family richness is always higher than the observed family richness. Plant treatments (top row) show a generally higher family richness than soil treatments (bottom row). However, there is no trend over time, for neither plant nor soil treatments. Instead, family richness correlates with the number of reads per sample (Figure 7).

The mean family richness (count) across all treatments is 56.25 ± 37.10. In contrast, the mean ACE index across all treatments is 87.52 ± 55.49.

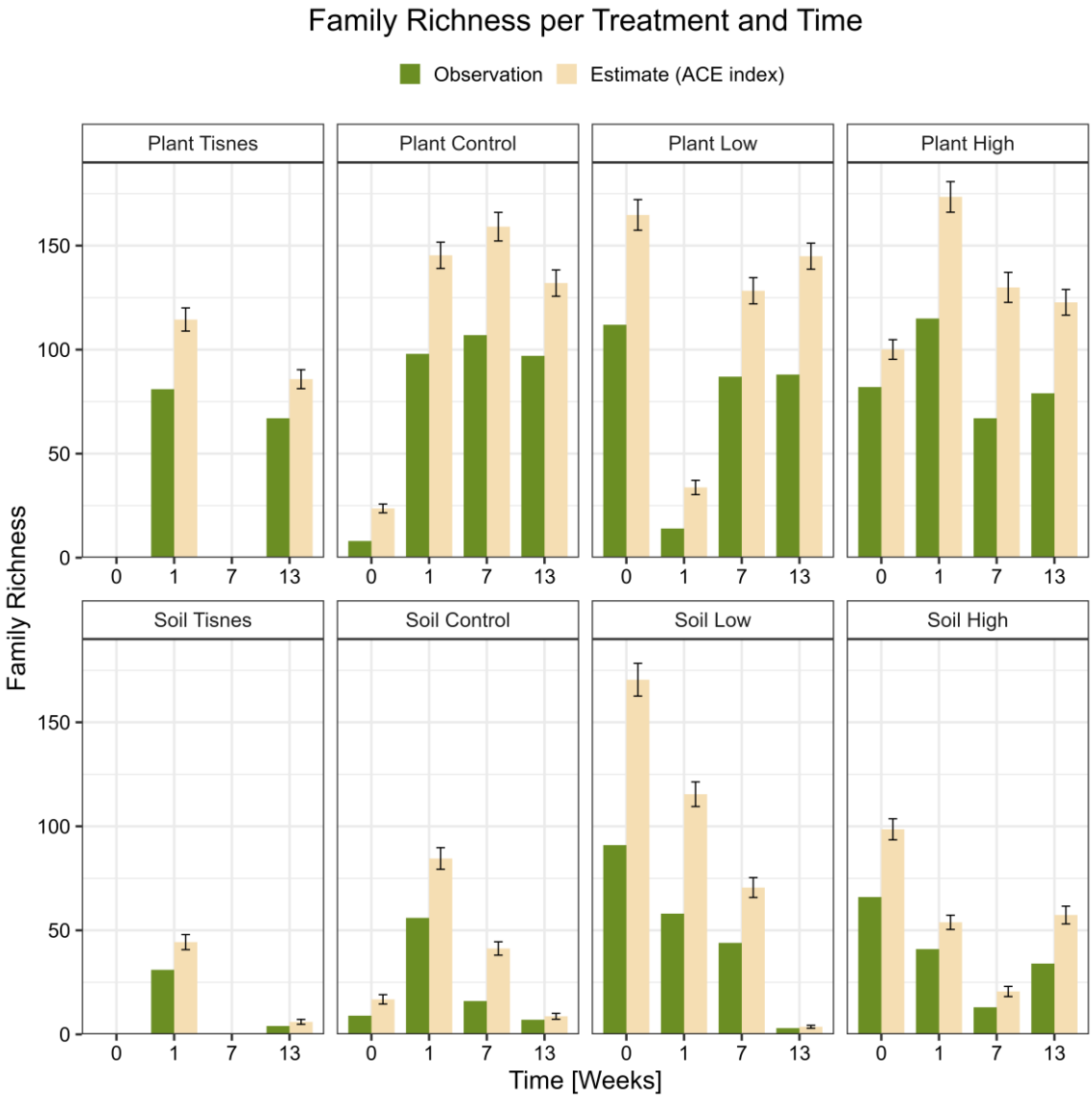

***Figure 9**: Family Richness of the pooled replicates over time and treatment. Green bars represent the observed richness (= number of families). Beige bars represent the estimated richness according to the ACE index. Error bars indicate the standard deviation (SD).*

3.4.2 Shannon Index

The Shannon-Wiener diversity index is a measurement for diversity, that considers the richness (count) as well as the evenness (distribution) of families across treatments. The index is unitless, the higher the index, the higher the diversity.

Figure 10 displays the Shannon index for every timepoint in a treatment (dots). For plant treatments (top row), the index is roughly between 2.5 and 4, with one exception (Plant Control, week 0). In contrast, the index for soil treatments (bottom row) is roughly between 0.5 and 3.8. The mean Shannon index across all treatments is 2.64 ± 0.74. Moreover, diversity increases over treatment (from Tisnes to High), however, that trend is more pronounced for soil treatments (bottom row).

The linear regression line gives an estimate of changes in diversity over time (Figure 10). The R^2 is calculated for every regression line to indicate the fit of the regression. Diversity in soil treatments (bottom row) always decreases over time, but no such trend is clear in plant treatments (top row). Moreover, the R^2 in plant treatments shows a poor fit.

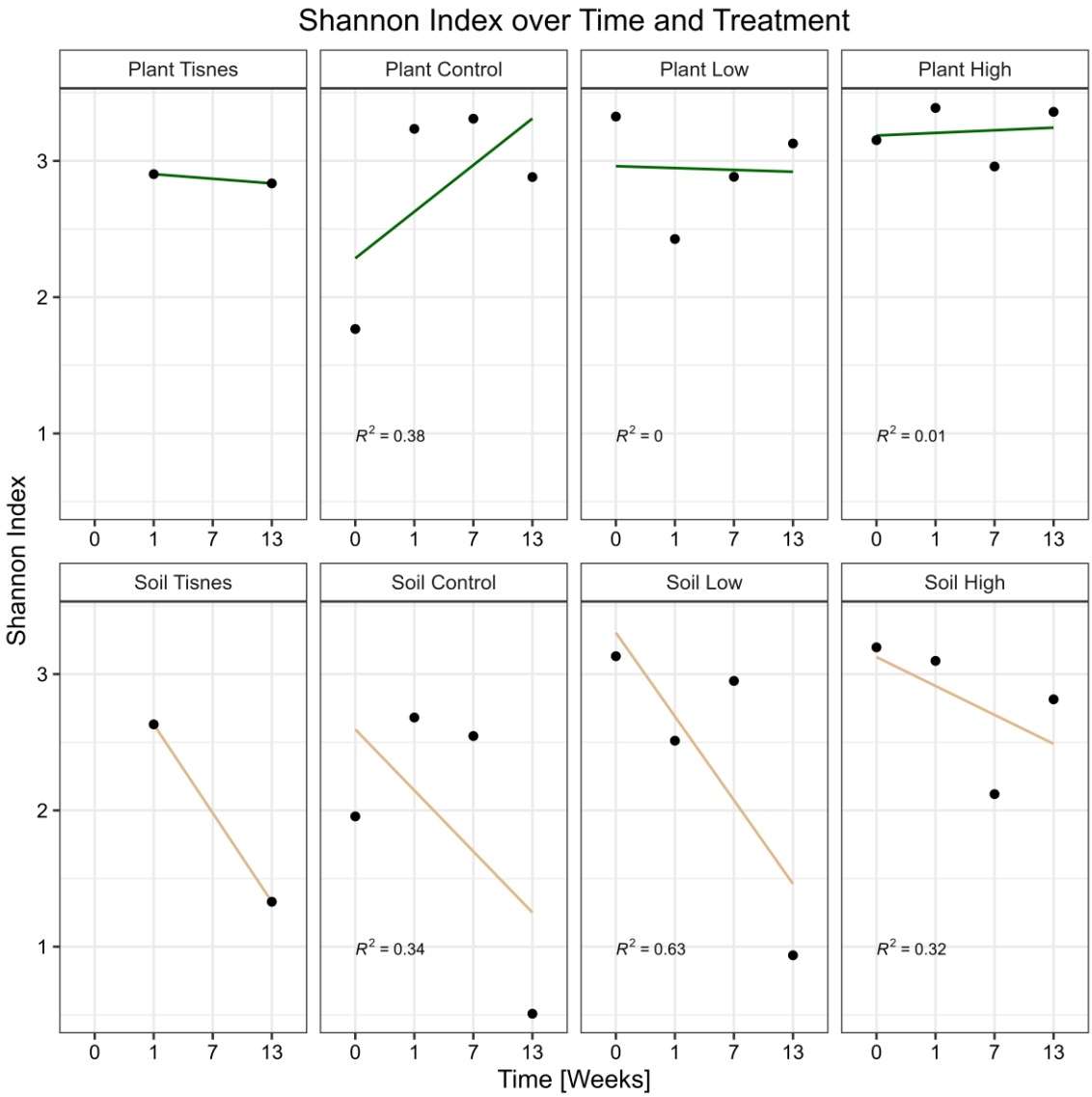

***Figure 10**: Shannon index over time and treatment. Dots represent the index for every timestep. Regression line indicates a development of diversity over time, R^2 indicates how well the regression fits the data. Green represents the plant treatments (top), light brown represents soil treatments (bottom).*

3.5 Family Composition

In total, 591 different genera, 260 families, and 131 order were identified. In descending order, the ten most abundant families in all treatments and according to their total number are Bacillaceae, Granulosicoccaceae, Rhodobacteraceae, Clostridiaceae, Pirellulaceae, Planococcaceae, Streptococcaceae, Comamonadaceae, Propionibacteriaceae and Sphingomonadaceae.

3.5.1 Bucket Effect

As the plants were grown under artificial conditions, the question arises as to whether the conditions themselves influence the family composition. The samples from Tisnes provide a

reference for how the original community developed over time. Comparison with the control treatments can show whether changes are natural or caused by the experimental design.

Figure 11 displays the relative family abundance for the Tisnes samples and the corresponding control samples at the same time. Abundance is displayed as percentage, the total number of reads for each treatment is displayed below each bar. Recorded are the five most abundant families (based on number of reads), all other families are summarized as 'Other'.

All plant treatments (top row) contain Rhodobacteraceae and Pirellulaceae. Moreover, Granulosicoccaceae are also in all plant treatments, except Plant Control 13. Changes over time in the Plant Tisnes community is the shift from Pseudomonadaceae and Microtrichaceae to Actinomarinales and Clostridiaceae. However, Microtrichaceae and the uncultured Actinomarinales, are both from the class Acidimicrobiia. Changes over time in the Plant Control treatments is the shift from Comamonadaceae and the family SC-I-84 (order Burkholderiales) to Clostridiaceae, the cyanobacteria Nostocaceae, and Lachnospiraceae. Plant Control after 13 weeks is the most dissimilar in terms of family composition and relative abundances of all plant treatments.

In contrast, soil treatments (bottom row) are dissimilar in their community composition. However, the Soil Tisnes community in the first week also contains Rhodobacteraceae, Pirellulaceae, and Granulosicoccaceae which makes it similar to the plant treatments. Moreover, it contains Comamonadaceae and Flavobacteriaceae. Soil Tisnes after 13 weeks shows Rhodobacteraceae, Rhodanobacteraceae, and Rhizobiaceae. However, reads are so low, it can virtually not be compared. In contrast, Soil Control after one week contains Bacillaceae, Planococcaceae, Clostridiaceae as well as Actinomarinales and Peptostreptococcaceae. After 13 weeks, that completely shifted to around 97 % Streptococcaceae. The remaining families are Propionibacteriaceae, Family XI (order Peptostreptococcales-Tissierales), and Staphylococcaceae. However, together they make less than 2 %.

Plant Tisnes and Soil Tisnes after one week share three out of five of the most abundant families. After 13 weeks it is only one family. In contrast, Plant Control and Soil Control at any time point do not share families as their most abundant. In summary, the most similar communities are Plant Tisnes, Plant Control after one week, and Soil Tisnes after one week. Soil treatments are dissimilar between treatments but also in comparison to plant treatments.

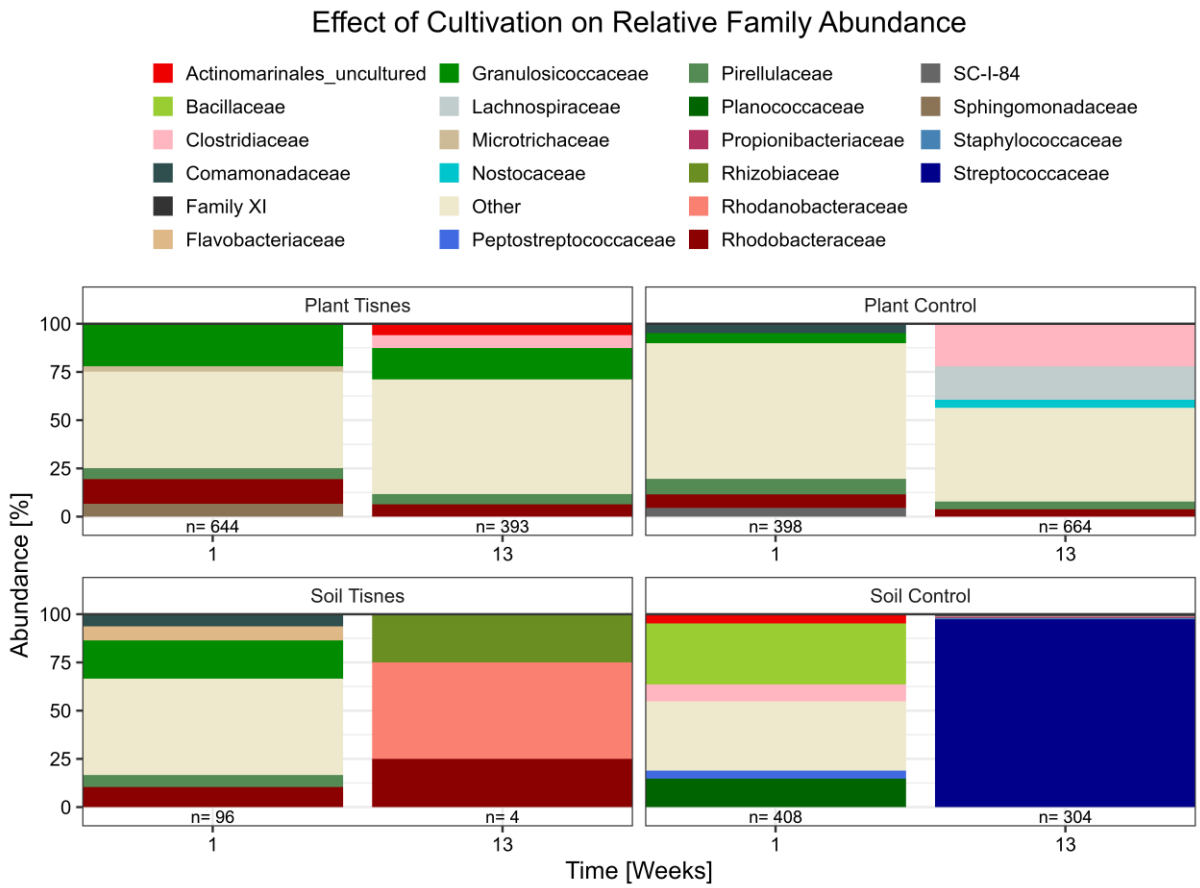

***Figure 11**: The influence of the experimental setup on the composition of bacterial families (= bucket effect). Different families are indicated by their colour (legend), shown are the five most abundant families for every sample. All other families are summarised as 'Other' (beige). The abundance is reported as percentage of total number of reads (n). Compared are the natural communities (Plant and Soil Tisnes) to the experimental communities (Plant and Soil Control) from the same timestep.*

3.5.2 Relative Abundance

Figure 12 displays the relative family abundance of the five most abundant families across all treatments and time, all other families are summarised as 'Other'. The total number of sequences is displayed below each bar. As Tisnes treatments are already displayed in figure 10, they are excluded here to not inflate the legend.

There is no obvious pattern over all plant treatments (top row), however, the families Comamonadaceae, Pirellulaceae, Bacillaceae, and Granulosicoccaceae appear most frequently. The community on Plant Control takes a subtle shift over time. However, Plant Control zero has very low reads and can virtually not be compared. Pirellulaceae and Rhodobacteraceae occur from week one until week 13. Granulosicoccaceae appear after one and seven weeks, but not anymore at the end of the experiment. Clostridia appear after seven and 13 weeks. Comamonadaceae in Plant Low appear over the entire time, except after one week. However, that can be masked from low number of reads. Clostridia appear only until week one, at that point, Propionibacteriaceae appear until week seven. Pirellulaceae are found at the beginning and the end of the experiment. The families Staphylococcaceae, Family XI (order Peptostreptococcales-Tissierales), Blastocatellaceae, and Nostocaceae do not appear at several

time points. The community in Plant High also displays a subtle shift. Bacillaceae occur during the entire time, while Granulosicoccaceae appear until week seven. Rhodobacteraceae and Pirellulaceae appear until week one. Weeks one and seven show Planococcaceae. Comamonadaceae appear in the beginning and in the end. At week seven Nostocaceae and Sphingomonadaceae, and at week 13 Pseudomonadaceae and Actinomarinales appear in the five most abundant.

There is no obvious pattern over soil treatments (bottom row), but the families Bacillaceae and Planococcaceae appear most frequently. The community in Soil Control is difficult to analyse, as the number of reads are low and Streptococcaceae are overly abundant at week 13. However, a subtle shift is notable. Until week seven Bacillaceae show high abundances, but from week seven Streptococcaceae become dominant. Actinomarinales appear in week one and seven. In the Soil Low treatment, Bacillaceae appear until week seven, but are least abundant at week zero. Planococcaceae are seen in week one and seven. Week 13 has very low reads and is dominated by Clostridiaceae and Nostocaceae. In Soil High, no shift is visible. Bacillaceae only appear in week seven and 13. Comamonadaceae and Planococcaceae are shown at week one and 13. Rhodobacteraceae occur until week one, and Pseudomonadaceae occur in week one and seven. Overall, it is difficult to see a pattern over time for Soil communities, in part due to varying and low number of reads.

An interesting observation is that the occurrence of some families is restricted to treatments. The families Fusobacteriaceae, Deinococcaceae, Nitrospiraceae, Nitrosomonadaceae, Vicinamibacteraceae, Methylophilaceae, Desulfuromonadaceae, Caulobacteraceae, and Flavobacteriaceae occur only in soil treatments. In contrast, the families Enterobacteriaceae, Lachnospiraceae, the Family XI (order Peptostreptococcales-Tissierales), and Blastocatellaceae occur only in plant treatments. Overall, the community differences between plant and soil are more pronounced than those between treatments or over time.

***Figure 12**: Bacterial family's composition over time and treatment. Reported are the five most abundant families per timestep, differentiated by their colour (legend), all other families are summarised as 'Other' (beige). Abundance is reported as percentage of total number of reads (n).*

3.5.3 Target Families

Figure 13 shows the relative abundance of the rhizoremediation families that were found in this study. Abundance is reported as percentage from the full sample and the total number of reads is shown below each bar. Because those target families are just a subset, reads are even lower. In descending order, the most abundant families across all treatments are Bacillaceae, Granulosicoccaceae, Rhodobacteraceae, Planococcaceae, Comamonadaceae,

Sphingomonadaceae, Pseudomonadaceae, Rhizobiaceae, Flavobacteriaceae, Rhodocyclaceae, and Xanthobacteraceae. In total, 24 different target families occur in the samples.

All plant treatments (top row) contain the most abundant target families, except Plant Control 0 and Plant Low 1, due to very low read numbers. There is no trend over treatment or time. In fact, all treatments display similar compositions of target families, even treatments without diesel addition (Plant Tisnes and Plant Control). Target families in plant treatments make up roughly 18 % to more than 60 % of the entire community.

In the soil treatments (bottom row) only Bacillaceae, Granulosicoccaceae and Pseudomonadaceae are the most abundant families. However, there is a difference between the treatments. The treatments without diesel (Soil Tisnes and Soil Control) are dominated by the three most abundant families. In contrast, the treatments with diesel addition (Soil Low and Soil High) have more families, especially Comamonadaceae, Rhodobacteraceae and Sphingomonadaceae, and are more similar to the plant communities. However, number of reads are even lower, and Soil Control and Soil Low after 13 weeks show (almost) no target families. Similar to plant treatments, target families make up around 20 % to more than 60 % of the bacterial community.

In summary, soil treatments have fewer families, but these are more abundant. In contrast, the plant treatments have a higher number of families, but those in lower abundance. In addition, the composition of the plant communities is quite uniform between treatments, whereas the soil communities differ between treatments.

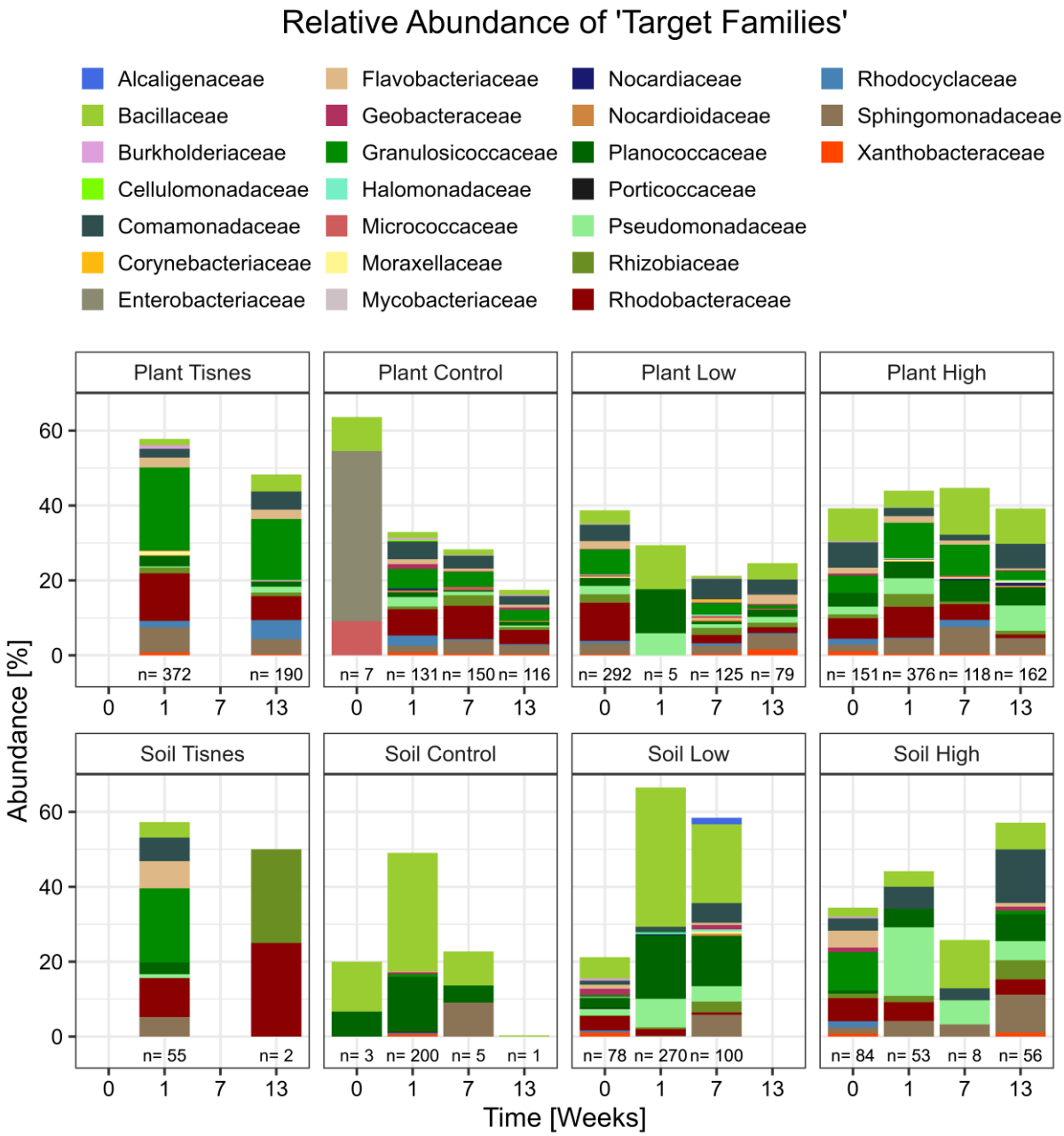

***Figure 13**: Family composition of bacteria associated with rhizoremediation (= target families) over time and treatment. Different families are indicated by their colour (legend). Abundance is reported as percentage of total number of reads per timestep. N is the number of reads assigned to target families, not the total number of reads.*

4 Discussion

4.1 Sampling Design

The original experiment included two other plants, *Leymus arenarius* and *Plantago maritima*, at different heights on the coast (Figure 3A) to analyse the bioremediation potential in the event of an oil spill reaching the coast. However, neither of the other two plants has yet been analysed. As *Puccinellia* sp. is located directly in the tidal zone and all plants survived the experiment, it was a logical choice to analyse it first. In addition, none of the *Plantago* plants survived the experiment and *Leymus* is outside the tidal zone and is unlikely to be affected by an oil spill.

The experimental setup was the best that could be done with the available resources. Moreover, it resembles similar experiments that studied bioremediation of Saltmarsh plants (Couto et al., 2011). However, it was almost impossible to control the environmental conditions in the buckets. Tidal cycles frequently introduce new nutrients, oxygen, organic matter, and additional microbes. However, this input is missing in the experimental setup. It is likely that the buckets became more anoxic and nutrient depleted than on Tisnes, particularly in deeper soil layers. This may have shifted the bacterial community towards more oligotrophic and anaerobic families. Most importantly, the water temperature in the fjords never rises above 10 °C, even in summer (Eilertsen & Skarðhamar, 2006). Therefore, intertidal sediments and plant roots are constantly cool. In addition, the buckets are made of black plastic, which means that heat absorption is high. As a result, the soil in the buckets is expected to be warmer than the soil in Tisnes. As each bacterial family has a specific preferred temperature range for optimal growth, an increase in temperature can shift the community towards more mesophilic or less psychrophilic families. However, no environmental parameters were measured during sampling. As this is a significant knowledge gap, improvements to the sampling design would include additional measurements such as temperature and oxygen saturation. In addition, due to limited resources, the actual HC concentration was not measured, and natural attenuation cannot be assessed. Another unexplored factor is the soil disturbance from sampling, which might have increased the heterogeneity of the soil. Soil characteristics have a strong effect on the microbial community (Ribeiro et al., 2013). Some disturbances are also expected on Tisnes, however, the effect is not quantifiable.

To account for such variables and uncertainties, samples were taken in triplicates. However, only the plant treatments were in triplicates while there was only one bucket for soil. Moreover, due to difficulties during the laboratory work, all replicates were pooled together by their timestep. This means information on natural variability within a treatment and timepoint was omitted.

To avoid the aforementioned effects of experimental conditions, e.g. the bucket effect, field experiments are more appropriate for studying a natural community. However, this is not feasible or permitted for petroleum products due to their toxicity. Some data on petroleum pollution in the natural environment exist anyways, mostly as observational studies following accidental oil spills, such as Exxon Valdez or Deepwater Horizon (Bragg et al., 1994; Hester et al., 2016). These studies highlighted the damage associated with oil spills and their clean-up,

but also demonstrated the effectiveness of bioremediation. Another type of study comes from a time before strict environmental legislation. For example, Hershner and Lake (1980) studied the effects of chronic diesel pollution on a salt marsh by polluting a previously pristine environment with almost 900 litres of diesel fuel. Although informative, the system suffered long-term damage. Therefore, despite their limitations, experimental setups are the only option.

4.2 DNA Amount after Extraction

Plant treatments have a mean DNA amount of 12 ng/µl, and soil treatments have a mean DNA amount of 2 ng/ µl, both with a very high SD (Table 6). A higher amount of DNA in plant treatments could reflect a higher microbial abundance. However, the DNA extraction method is not selective for bacteria alone. The DNA could have been from archaea, fungi, the plant itself or even microscopic animals. To increase the low DNA amount of some samples, several extractions could be pooled together, as it was done for some unsuccessful samples.

The mean 260/ 280 nm ratio for the plant treatments is below 1.8, for soil treatments usually above 1.8, and with high SD for both (Table 6). A ratio of around 1.8 indicates pure DNA, and a ratio lower than that indicates contamination such as salts or phenol. In addition, a DNA amount below 10 ng/µl is considered very low, and 2 ng/ µl are the detection limit of ND, which may influence the ratio (Matlock, 2015). Therefore, a 'good' ratio does not automatically mean a good sample and vice versa. Although each sample was measured twice (Appendix 6), the SD was still very high (Table 6). This could be due to an inadequate blank measurement, e.g. contaminated water. However, this is unlikely as Milli-Q® water was used for the blank. It could be the instrument itself, as the ND used is relatively old. To further reduce the variation, an additional cleaning or washing step could be added immediately after extraction.

4.3 Optimisation of Library Protocol

The annealing temperature for the BC PCR has been reduced from 55°C to 50°C. At lower temperatures, even a partial match between primer and DNA template is stable enough for successful amplification. This may be the reason why the lower temperature was more successful (Singh et al., 2014).

To test whether the samples or the BC primer were the reason for unsuccessful amplification, all samples were amplified using the general 16 S primers fD1 and rD1 (Appendix 3 and 4). Almost all samples are amplifiable with these general primers, but not all of those successful samples were amplifiable using the ONT BC primers. The 16 S primers used in the kit are 27F and 1492R. However, the primer sequence is complemented by the sequence for the barcodes and a flanking sequence (https://nanoporetech.com/document/chemistry-technical-document). However, there is no mention of whether this longer than usual sequence affected the efficiency of the amplification. In addition, full sequences are not reported due to 'proprietary information'.

The addition 3 µl of $MgCl_2$ to the 50 µl PCR reaction, immediately improved the PCR yield (Figure 5 and 6). $MgCl_2$ is a cofactor for the polymerase activity and helps counteracting possible inhibitors in the reaction, such as salts or phenol (Singh et al., 2014). A low sample dilution (1:10) gave the best results for PCR amplification (Figures 5 and 6). This may also

indicate the presence of inhibitors, as a higher dilution of inhibitors was required for successful amplification. Since the ND results were too variable (Table 6), 1 µl of sample was used regardless of DNA amount. In figure 6, one reaction was not successful, and is most likely a pipetting mistake.

During DNA purification, incubation time and temperature was increased to around 10 minutes at around 30 °C. In contrast, the protocol (Appendix 2) states an incubation time for around two minutes at room temperature. It is curious, that the protocol states these shorter times and temperatures, even though it does not seem to be effective enough. For the final purification, 75 % ethanol was used, instead of 70 % as stated in the protocol (Appendix 2). The higher concentration might have been necessary to remove all salts from the sample. As the samples are from a tidal environment, higher salt concentrations are expected. All changes were made immediately after the recommendations were received from the ONT support and Alena Didriksen (personal communication, November 2023). It is not clear whether any one change had the greatest impact, or whether they were all necessary.

4.4 Sequencing and Alignment

Basecalling is the process of translating the electrical signals from sequencing into nucleotide sequences. Here, the ONT basecaller Guppy was used to simultaneously trim primers, adapters and BC from the 16 S sequence to produce a strictly biological sequence. The accuracy of basecalling is highly dependent on the algorithm used. Guppy is an older software and is no longer updated or improved. Therefore, re-running the raw sequences with a better or more updated algorithm may improve the results. The alignment of the nucleotide sequences to a reference database and the identification of a bacterial family were carried out using the ONT programme EPI2ME. This software offers different options for sequencing, such as the database (SILVA or NCBI), the algorithm (kraken2 or minimap2) or the taxonomic level. The alignment is fully automated, based on the chosen parameters. The user has almost no influence on any of the processes, and little can be adjusted. For example, instead of an Operational Taxonomic Unit (OTU) table, the output is a simple abundance table (Appendix 7), which is not always convenient to analyse. Furthermore, the exact sequence assigned to a genus is not reported. Nevertheless, EPI2ME provides a convenient and user-friendly platform for sequence alignment that requires little prior knowledge. Other pipelines exist, of course, but would have been beyond the scope of this project. Furthermore, it was one of the original aims of the study to experience the entire ONT pipeline

Approximately 70% of all sequences are successfully identified (Figure 7). The remaining 30% of unidentified sequences could be due to a number of reasons. There could be no matching references in the library, but this is very unlikely as at least a higher taxonomic level should be identifiable. Instead, it is more likely that these sequences were somehow damaged or too short to be assigned. For example, the DNA strands could have been fragmented during the laboratory work, or nucleotides could have been misidentified during basecalling.

Several options were tried, and the use of the SILVA database (version 138.1) with the minimap2 classifiers (sequence alignment) gave the best results. Bacteria were identified to

genus level. In addition, the SILVA database is more commonly used in environmental microbiome studies, making comparisons easier. Only about 30% of all sequences could be identified using the NCBI database, although it is the basis for SILVA. However, SILVA also reports the last common ancestor (LCA) and uses slightly different taxa to the NCBI database, which could explain the better performance.

In total, 68 samples were successfully aligned by EPI2ME (Figure 7), but the original sampling design only included 56 samples (Figure 4). The reason for this discrepancy is the number of replicates required to obtain good results. This was because the barcodes were always in sets of 1 to 24 but could never be on the same flow cell twice, and this did not always match the samples from the BC PCR. So 'slots' on a flow cell were not wasted, some low DNA samples were sequenced anyway if they had the correct barcode.

The number of reads varied from 6 to 1135 for pooled samples. Now, the Flongle flow cell generates up to 2.8 Gb of data, which is the lowest compared to other flow cell options, such as MinION or PromethION (https://nanoporetech.com/platform/technology/flow-cells-and-nanopores). But even for that, the number of reads is still extremely low (Kuczynski et al., 2010). The reason for such low read numbers is most likely an accumulation of the difficulties during DNA extraction, PCR amplification and purification. Plant treatments had more reads and more replicates than soil treatments (Figure 7), which could reflect more successful DNA extraction and higher bacterial abundance within the plant treatments. However, the number of reads does not reflect true abundance (Kerkhof, 2021). To reveal strong ecological patterns, as few as 100 sequences per sample may be sufficient. However, for less obvious patterns, more than 1000 sequences per sample are required, especially for clustering (Kuczynski et al., 2010). As almost all samples have less than 1000 reads and many less than 100 reads (Figure 7), no multivariate or ordination-based community analysis was performed.

4.5 Diversity Indices

None of the rarefaction curves flatten out (Figure 8), meaning that the true family richness is not achieved. Some curves in the plant treatment (top row) start flatten out at around 1000 reads, which corresponds to the maximum number of reads (1153). However, it also shows that more than 1000 reads per sample are required to achieve adequate community representation. Figure 9 shows that the estimated family richness (ACE) is always higher than the observed family richness (count). However, by reducing the family composition to a count, it can introduce a bias, as information about their spatial distribution is lost. In a very heterogeneous environment such as soil, sub-communities in different niches may be very different from each other. As the ACE index is calculated on the basis of the observed richness, the same bias is introduced. Nevertheless, it confirms the rarefaction curves (Figure 8) by showing that the true number of families is much higher than observed.

However, neither shows a trend across treatments or time, the only difference being between the plant and soil treatments. This would suggest that the plant is more important in shaping bacterial diversity than the diesel addition.

Just like family richness, the Shannon index (Figure 10) differs between plant and soil treatments, with higher diversity in plant treatments. Over time, diversity in soil treatments decreases. Over treatment, diversity increases for both plant and soil treatments. This trend is more pronounced for soil treatments. Moreover, treatments without diesel have lower diversity than treatments with diesel addition (Figure 10). Biologically, that could reflect the high bacterial diversity and abundance in the rhizosphere in comparison to bulk soil (Chaudhry et al., 2005; Kuiper et al., 2004). Moreover, for the plant associated community that would mean that there are more families at the end of the experiment. However, there is no discernible difference between treatments without diesel addition (Plant Tisnes and Plant Control) and the treatments with diesel addition (Plant Low and Plant High). This would indicate that the plant has a stronger influence on community than the diesel addition. In fact, studies have shown that the plant shapes its rhizosphere community in response to contaminants or other environmental stressors (Chaudhry et al., 2005). For the soil communities, it would indicate that the soil itself cannot support a high number of families and diversity decreases over time. It does represent the difference between bulk soil and rhizosphere. As bulk soils lack the stabilising effect of a plant, diesel addition might become the strongest influence in shaping the community (Ribeiro et al., 2013).

4.6 Family Composition

Relative abundance was calculated to make family composition more comparable across treatments and time steps, and to compensate for different numbers of reads. However, it also introduces a bias into the analysis. Highly abundant families in treatments with very low numbers of reads may not be as abundant in the environment. There are simply not enough sequences to represent the natural environment, as shown by the rarefaction curves (Figure 8). Furthermore, it is difficult to detect changes in family composition over time and across treatments due to the overall variable or simply low number of reads. Therefore, it is more appropriate to simply state who is present rather than trying to quantify their abundance.

For the analysis, the five most abundant families per sample are reported. Five was chosen to reduce the number of different families in the legend. On the other hand, showing only the ten most abundant across all samples would have introduced bias. High read treatments with very abundant families would have biased the results for the lower read treatments. However, the ten most abundant families per sample are shown in Appendix 7.

4.6.1 Bucket Effect

Figure 11 compares the composition between the natural Tisnes community and the Control community from the experiment at the same time. The most similar communities are Plant Tisnes after week one and 13, Plant Control after one week, and Soil Tisnes after one week. Soil treatments are dissimilar between across themselves but also in comparison to plant treatments.

All plant treatments, except Plant Control, share three out of five of the most abundant families, which are Granulosicoccaceae, Pirellulaceae, and Rhodobacteraceae. These three families share some characteristics, most importantly that all are aerobic chemo organoheterotrophs (Ivanova

& Webb, 2014; Lage et al., 2022; Pujalte et al., 2014). The most important shift for the plant Tisnes community over time, is the shift from aerobic Sphingomonadaceae to the anaerobic Clostridiaceae (Glaeser & Kämpfer, 2014; Wiegel et al., 2006), which could indicate that the soil became a slightly anoxic over the summer. However, soil is very heterogenic and could contain oxic as well as anoxic compartments. The Plant Control community changes from all aerobic families to be dominated by the obligate anaerobes Clostridiaceae and Lachnospiraceae (Meehan & Beiko, 2014; Wiegel et al., 2006), as well as the cyanobacteria Nostocaceae, known for their resilience to adverse conditions (Garcia-Pichel, 2009). This may indicate that the soil has become more anoxic.

In contrast, communities in soil treatment are quite different from each other (Figure 11). All families occurring in Soil Tisnes are aerobic, which suggests that the soil stayed oxic during the summer. However, number of reads are too low, especially for Soil Tisnes after 13 weeks, to draw any conclusions about their community. The community of Soil Control after one week, contains mostly aerobic families, but also the anaerobic Clostridiaceae and Peptostreptococcaceae. Moreover, especially Bacillaceae but also Clostridiaceae are known to form spores (Mandic-Mulec et al., 2015; Slobodkin, 2014; Wiegel et al., 2006). The presence of so many anaerobes in combination with spore formers could be an indication that the soil has already become more anoxic and less habitable for bacteria. In contrast, Soil Control after 13 weeks contains around 97 % Streptococcaceae. As these are all of the genus Streptococcus (Appendix 7), it is most likely that contamination occurred during sampling or laboratory work, as this genus is usually a human pathogen (Lory, 2014).

In summary, the community differences of Plant Control over time, as well as the difference between Soil Tisnes and Soil Control after one week, indicate the presence of a bucket effect on the bacterial community. However, the similarities between Plant Tisnes, and Plant Control after one week indicate a stabilising effect of the plant on the family composition. The difference between Plant Control and Soil Control after one week, is more pronounced than the difference between Plant and Sol Tisnes after one week. This indicates a greater difference between rhizosphere and bulk soil in the experiment than in the natural environment. This is most likely due to the effect of environmental parameters during the experiment. This means that the experimental design significantly influenced the bacterial composition.

4.6.2 Family Ecology

Figure 12 shows the relative abundance of families across all treatments and timesteps. There is no trend over time nor treatment, however, some families occur more frequent than others. The ten most abundant families, in descending order, are Bacillaceae, Granulosicoccaceae, Rhodobacteraceae, Clostridiaceae, Pirellulaceae, Planococcaceae, Streptococcaceae, Comamonadaceae, Propionibacteriaceae, and Sphingomonadaceae. All these families generally occur in a wide range of habitats and have been found in intertidal or Arctic coastal systems (Gorrasi et al., 2019; Wang et al., 2012). It is therefore not surprising that they are the most common.

The families Fusobacteriaceae, Deinococcaceae, Nitrospiraceae, Nitrosomonadaceae, Vicinamibacteraceae (including the uncultured Vicinamibacterales), Methylophilaceae, Desulfuromonadaceae, Caulobacteraceae, and Flavobacteriaceae occur only in the soil treatments (Figure 12).

Nitrosomonadaceae and Nitrospiraceae are important families for nitrogen cycling, and even complement each other. Nitrosomonadaceae oxidise ammonia into nitrite, and are the main nitrifiers in soil. They can be inhibited by too high acidity or too much nitrite and are often rate limiting for nitrification (Prosser et al., 2014). Nitrospiraceae contains three genera that have very different forms of metabolism. However, the most diverse and widespread in the environment is the type genus *Nitrospira*, which is the main oxidiser of nitrite into nitrate in soil (Daims, 2014). Additionally, they are also common in intertidal sediment (Wang et al., 2012). However, both have relatively slow growth rates and are easily outcompeted by heterotrophic bacteria. That could explain, why they only occur in Soil Low at the start of the experiment. As they are the main functional group they occur most likely also in every other treatment, but simply not abundant enough to show up. In addition, especially ammonia oxidation is also performed by archaea, especially Thaumarcheaota. They compete with Nitrosomonadaceae and can further decrease their abundance (Prosser et al., 2014). However, archaea were not part of this study, so their presence and role can only be assumed.

Another important ecological niche is occupied by the family Methylophilaceae, which occur in Soil Low after seven weeks, as well as Soil High after one week (Figure 12). This family contains aerobic methylotrophs, which oxidise exclusively single carbon compounds as energy source. Such compounds are methanol or methylamine, but never methane. They are often closely associated with plants, as plants are the man methanol source in the biosphere. During plant growth the pectin in the cell wall is demethylated, which produces methanol. In return, Methylophilaceae produce for example phytohormones or vitamins, and are therefore often PGPR bacteria. Unsurprising, they are frequently found in the phyllosphere as well as the rhizosphere (Doronina et al., 2014). It is curious that they occur in the soil treatments, given their symbiosis with plants. It is possible that they also occur in plant treatments, but are simply not reported due to low abundance.

A common characteristic amongst the remaining families that only appear in soil treatments, is the ability to degrade complex substrates. Vicinamibacteraceae also degrade simple sugar, however, but prefer complex proteinaceous compounds (Huber & Overmann, 2019). Caulobacteraceae are well adapted to oligotrophic conditions. Their unique cell cycle includes a sessile, nonreproductive stage that is triggered by carbon depletion. This stage is characterised by increased nutrient uptake and degradation of complex compounds. They are also able to withstand toxic compounds and phosphorus deficiency (Abraham et al., 2014). Desulfuromonadaceae are important reducers of sulphur and metals and can completely oxidise organic compounds, including petrogenic HC. As such, they are often found in the degrading consortium of benzene, toluene, ethylbenzene, xylene (BTEX) compounds and other syntrophic communities. They can be found in a wide range of habitats, including Arctic intertidal sediments (Greene, 2014; Wang et al., 2012). The family Deinococcaceae can survive many

environmental challenges but is most known for its ability to withstand radiation. As such they can be found in many polluted or extreme habitats, but usually requires complex substrates (Rosenberg, 2014). Flavobacteriaceae are known to degrade a range of macromolecules, such as proteins, polysaccharides, and even PAHs. They can be found in wide range of habitat and are considered cosmopolitan (McBride, 2014). Fusobacteriaceae are most commonly found in the mucous membranes of mammals and can be associated with many infections in their hosts. In the environment they are often found in anoxic sediments and their metabolism is always fermentative. They are able to use a wide range of sugars as energy sources and occupy a range of ecological niches (Olsen, 2014).

The families Enterobacteriaceae, Lachnospiraceae, the Family XI (order Peptostreptococcales-Tissierellales), and Blastocatellaceae occur only in plant treatments (Figure 12). Enterobacteriaceae are a large family of facultative anaerobes. Almost all are enteric, and many are endosymbionts or associated with the rhizosphere (Octavia & Lan, 2014). Blastocatellaceae have mostly been isolated from sand and soil and can tolerate a wide range of temperatures and pH. Their preferred substrate are complex proteinaceous compounds, and other complex carbon compounds (Huber et al., 2017). Lachnospiraceae are strictly anaerobic bacteria, that most commonly found in human and cow guts, and subsequent in sewage water. However, some members also occur in a wide range of environmental habitats (Meehan & Beiko, 2014).

In summary, the families that occur only in soil occupying either an important ecological niche, such as nitrogen cycling, or are able to degrade complex molecules. In contrast, families occurring only in plant treatments are less specialised. Due to the release of readily available root exudates, soil aeration by roots, and the additional habitat on and in the roots (Figure 2), microbial abundance and activity in the rhizosphere is higher (Chaudhry et al., 2005; Dazzo & Ganter, 2009; Kuiper et al., 2004). Moreover, high input of organic nutrients, increases the competition between bacteria, and leads to leads to rapid bacterial growth. Without those nutrients, growth rate rapidly decreases (Dazzo & Ganter, 2009). Moreover, by changing the type and composition of root exudates, the plant can actively select for a specific rhizosphere consortium, a process that is called rhizosphere breeding (Chaudhry et al., 2005). However, such easy energy sources are scarce in the bulk soil and quickly taken up, and bacteria need to evade to more complex substrates. As a result, bulk soil communities often display lower diversity and activity, but high abundance of a few key members (Dazzo & Ganter, 2009; Kuiper et al., 2004). Therefore, soil treatments, contain more specialist and bacteria adapted to degrade complex substrates.

4.6.3 Target Families

There is a wide variety of microbes associated with the bioremediation of petroleum HCs, either through direct degradation or by promoting phytoremediation. The most common are discussed in more detail here, but less common or absent families are listed in Table 1. However, not all members of a family or even a genus contain the necessary metabolic pathways. In this study, bacteria were identified to genus level but analysed at family level. This was a compromise between taxonomic resolution and the ability to generalise. If only a few genera in a family are relevant to rhizoremediation, but the entire family is listed, this could lead to bias. Target

families may appear more important even though it is not certain that the one relevant genus is present in the samples. However, many studies report their relevant bacteria as genus or strain (Anekwe & Isa, 2024; Freyria et al., 2024; Koshlaf & Ball, 2017). Analysis at a higher taxonomic level would have made the analysis easier, but would have excluded some important details.

Figure 13 shows the relative abundance of target families across treatments and timepoints. Because target families represent only a subset of the full community, the number of available sequences per treatment and timestep is even lower. Moreover, the last week of Soil Low does not contain any target families.

In descending order, the most abundant families across all treatments are Bacillaceae, Granulosicoccaceae, Rhodobacteraceae, Planococcaceae, Comamonadaceae, Sphingomonadaceae, Pseudomonadaceae, Rhizobiaceae, Flavobacteriaceae, Rhodocyclaceae, and Xanthobacteraceae. In total, 24 different target families occur in the samples.

Bacillaceae occur in all treatments and have especially high abundances in soil treatments (Figure 13). This is not surprising as they are truly ubiquitous and their primary habitat is soil. Their main contribution to rhizoremediation is as PGPRs. Functions include, but are not limited to, organic matter decomposition, nitrogen fixation, phosphorus solubilisation, phytohormone production, induction of systemic resistance, metal ion assimilation and biocontrol. However, a key ecological feature is the production of highly resistant endospores, whose production is triggered by, for example, nutrient depletion or changes in pH (Mandic-Mulec et al., 2015). However, DNA extraction and sequencing does not distinguish between DNA from active and sporulated bacteria. This could introduce bias by overestimating their abundance and importance.

Sphingomonadaceae occur most frequently in plant treatments (Figure 13), and this family could be one of the most important families for bioremediation. They occur in a wide range of habitats, including soils, marine and freshwater, in association with plants, or in contaminated or oligotrophic habitats. When in association with plants, they can be in the phyllosphere, as endophytes or as part of the rhizosphere community. They are able to suppress plant pathogens and are considered PGPR bacteria. Moreover, many genera are able to degrade complex and recalcitrant compounds, such as PAHs. Two important adaptations are responsible for this. Firstly, Sphingomonadaceae contain sphingoglycolipids instead of polysaccharides and a high carotenoid content in the cell wall. In particular, the degradation of heterocyclic compounds (carbon rings with at least one atom of another element) produces reactive oxygen species that are harmful to the organism (Glaeser & Kämpfer, 2014). Carotenoids are able to reduce the concentration of reactive oxygen species in unsaturated lipids (Widomska et al., 2019). In one sense, the cell wall of Sphingomonadaceae contains an integrated antioxidant system. Secondly, many Sphingomonadaceae have large plasmids that carry a large number of genes for different degradation pathways. These include genes for HC degradation, oxidation and metabolism (Glaeser & Kämpfer, 2014).

Planococcaceae can be found in almost any habitat, from ice to soil to plants. Many members are resistant to metals and antibiotics. They also contribute to rhizoremediation through a number of functions. Firstly, they produce the plant hormone indoleacetic acid and can also dissolve phosphorus and fix nitrogen, making them PGPR bacteria. Secondly, they can degrade a wide range of complex or toxic compounds, such as phenolic compounds (benzene ring with hydroxyl group), and can use phenol as their sole carbon source. They can also detoxify metals such as arsenic, mercury or chromium. Finally, some genera produce extracellular polymeric substances (EPS), some of which can emulsify petroleum compounds such as xylene or hexane. (Shivaji et al., 2014).

Pseudomonadaceae occur in all plant treatments, and in soil treatments with diesel addition (Figure 13). These bacteria are aerobic and are one of the most important PGPRs, either directly or indirectly. Their most important direct effect is the production of phytohormones such as auxins, cytokinins or gibberellins. They also increase the availability of nutrients to the plant by solubilising phosphorus or producing siderophores. Some Pseudomonadaceae exert biocontrol by acting against plant pathogens, especially fungi. Others suppress plant diseases, by competing with plant pathogens for nutrients and space, by inducing systemic resistance within the plant, or by producing antibiotic compounds that reduce pathogens (Roquigny et al., 2017). In addition, many strains have been found to also directly degrade a variety of petrogenic HCs in contaminated soils (Anekwe & Isa, 2024; Darsa & Thatheyus, 2014; Obayori et al., 2009).

Xanthobacteraceae occur in all plant treatments in small abundances, and in some soil treatments (Figure 13), and they are ubiquitous in wet soils and sediments. A common characteristic is nitrogen fixation, and they are often found in symbiosis with legumes. They can also grow on a variety of different complex compounds such as halogenated, chlorinated or brominated alkanes, alkenes and aromatics. In addition, many members can use hydrogen or sulphur as an alternative energy source, making them important organisms for bioremediation (Oren, 2014b).

Rhodocyclaceae are predominantly found in plant treatments (Figure 13). This family exhibits a wide range of metabolic pathways. As nitrogen fixers they are found as endo- or epiphytes on plant roots. Members are also capable of degrading a wide range of aromatic and recalcitrant compounds, particularly under anaerobic conditions. These compounds include benzene, toluene, catechol, ethylbenzene, benzoates, pyridine, menthol, cholesterol or testosterone. In soils, the degradation of PAHs is also widespread, making them valuable for bioremediation (Oren, 2014a).

Comamonadaceae occur in all plant treatments, and soil treatments with diesel addition (Figure 13). They are found in a wide range of habitats, including soil, water, and extreme or polluted environments. They exhibit a very high metabolic diversity, including photoautotrophy and photoheterotrophy, organotrophy, fermentation, iron reduction, hydrogen oxidation and more. However, heterotrophic genera in particular are capable of degrading a wide variety of complex compounds, for example the ability to accumulate natural polymers, such as polyhydroxyalkanoates in their cells. In addition, many genera carry genes for several aromatic

degradation pathways. Finally, many members can use nitrate as an alternative electron acceptor under anoxic conditions. This adaptation makes them very common in wastewater or activated sludge (Willems, 2014).

Flavobacteriaceae occur in almost all plant treatments but only in few soil treatments (figure 13). Their most notable ecological feature is the use of macromolecules, such as starch, as energy source. Moreover, some genera, are able to degrade PAHs (McBride, 2014).

Granulosicoccaceae occur predominantly in plant treatments (Figure 13). This very small family is capable of using various carbohydrates and is often found in marine environments or as endophytes (Ivanova & Webb, 2014). Moreover, it is every common family in the tidal zone and are known to degrade HC (Freyria et al., 2024; Rizzo et al., 2019).

In general, plant treatments have more families (and higher diversity) associated with rhizoremediation than soil treatment. Furthermore, no plant treatment stands out in terms of family composition, with the exception of Plant Control at week zero, due to the low number of reads. Otherwise, almost all families occur at all time steps and treatments. That could confirm that the presence of the plant is more important in shaping the bacterial community than the addition of diesel (Ribeiro et al., 2013). Moreover, the high diversity and family richness especially among rhizoremediation communities could be an indication of the rhizosphere effect (Chaudhry et al., 2005; Dazzo & Ganter, 2009; Kuiper et al., 2004). For example, plants are able to actively shape their rhizosphere community by altering root exudates, in response to environmental stressors (Chaudhry et al., 2005). This is indicated by the similar communities between all plant treatments. Moreover, soil treatments display less families, but higher dominance of a few families (for example Bacillaceae). Curiously, soil treatments also display a few families that are usually only associated with the rhizosphere, such as Rhizobiaceae. Moreover, the communities without any diesel addition (Tisnes and Control) for plant and soil treatments also display many target families. Reason for that could be, that the analysis on family level is simply too broad. Or that diesel does in fact not have a strong effect on the bacterial community.

4.7 Bioremediation Potential

Diesel is a not very complex petroleum product, consisting mainly of branched chain or cycloalkanes and a low percentage of aromatics. Despite its moderate complexity (compared to crude oil), it can be quite toxic to plants and very irritating to animals and humans (Gad, 2005a). The diesel concentration in the experiment was 50 g/ 10 l (0.5% w/v) for the ‘Low’ treatment and 100 g/ 10 l (1% w/v) for the ‘High’ treatment. These concentrations are well above the safe exposure threshold for diesel, which is around 1 mg/ 10 l (Phillips, 2005). Nevertheless, all plants of the grass *Puccinellia* survived the diesel contamination and almost all treatments contained a consortium of HC degrading bacteria (Figure 13). The use of HC and petroleum products as a substrate for microbes is common and not very specialised. However, the more complex a compound is, the more difficult its degradation pathway becomes and the less common its degradation pathway is (Gerhardt et al., 2009; Ron & Rosenberg, 2014; Ward et al., 2009). It can therefore be assumed that the diesel was actually used and at least partially

degraded during the experiment. However, this is not verifiable for this study. Nevertheless, *Puccinellia* and its rhizosphere microbiome show HC degradation potential and may be a good candidate for bioremediation.

However, there are still some uncertainties. The diesel, although in high concentration, was poured directly onto the soil, avoiding contact with the leaves. Grass roots are more resistant to oil contamination than their leaves (Baker, 1970). However, in the event of an oil spill, the entire grass is usually covered. Furthermore, most accidental oil spills involve crude oil, and *Puccinellia* is not known to survive this (Cowell, 1969). It is therefore uncertain whether *Puccinellia* would be able to effectively intercept (phytostabilisation) and degrade an oil spill on the coast. Additional cleaning may be required. However, *Puccinellia* is a very common and widespread grass, especially along shorelines or salt marshes in Arctic and temperate regions. Even if the grass and other vegetation die shortly after a spill, it can be replanted over large areas and used to restore the shoreline. Using seeds to regrow the native vegetation also allows soil microbes to effectively colonise the root system and form an effective rhizosphere. This method, called seed-planting, has proven to be a very effective bioremediation method for soil restoration (Gerhardt et al., 2009). *Puccinellia* may be a good plant to apply this method to intertidal zones. At the very least, it shows potential for natural attenuation and shoreline restoration.

Nevertheless, before Puccinellia can be implemented in bioremediation strategies, more research is required. Further investigation is required to establish the tolerance of *Puccinellia* to different concentrations and different petroleum products. The rate of HC degradation and the compounds susceptible to degradation in this environment should also be assessed. A more comprehensive examination of the microbiome is required to identify degradation pathways on a molecular level. In addition, a detailed understanding of the composition and structure of the community would make predicting its response more accurate. With these questions answered, *Puccinellia* and its rhizobiome can be incorporated into bioremediation strategies for the intertidal zone of temperate and Arctic marshes.

5 Conclusion

The grass *Puccinellia* and bulk soil from the intertidal zone were subjected to a diesel contamination trial. Rhizosphere samples were collected and analysed using 16 S sequencing to identify and analyse the community and assess its bioremediation potential.

The amount of DNA after extraction was low overall, with higher DNA in the plant treatments and also indicating some contamination. Several changes had to be made to the library preparation protocol. The most important change to the PCR reaction was the addition of $MgCl_2$, which indicates the presence of some inhibitors. In addition, tests showed that the supplied BC primers were less effective than standard 16 S primers. The main changes for the DNA purification protocol were the alterations for incubation and the use of a higher percentage of ethanol. It is not clear whether improvements were necessary because the protocol was inadequate or because inhibitors were carried over from DNA extraction. Despite optimisation,

the output of the sequencing was insufficient. The number of reads per sample was too low for in-depth analysis, so the analysis performed was mostly descriptive.

The community composition is influenced by the experimental design, despite the use of an appropriate set-up. However, the effect is minimised in the plant treatments. Consistent with the DNA extraction results, family richness and diversity are higher in the plant treatments. However, family richness is not an accurate representation of the environment but reflecting the insufficient number of reads. In contrast to plant treatments, diversity in soil treatments also changes over time and treatment. There is no such trend for the family composition, the most pronounced difference being between plant and soil treatments. However, when looking specifically at the rhizoremediation consortium, the composition in the plant treatments does not change across treatments or time, while the soil treatments show a difference between diesel and control treatments.

All results indicate that the plant itself has the strongest effect on community composition, more than diesel, and the effect of diesel on community composition is only noticeable in the soil treatments. The community composition changes but does not show a trend over treatment or time. Moreover, the grass *Puccinellia* did not die during exposure and the rhizoremediation consortium constituted up to 60 % of the bacterial community. Therefore, *Puccinellia* and its rhizosphere show biodegradation potential. However, future studies need to focus on more comprehensive experiments and a more in-depth analysis of the community composition and structure.

Works cited

Abraham, W.-R., Rohde, M., & Bennasar, A. (2014). The Family Caulobacteraceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 179-205). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_259

Anekwe, I. M. S., & Isa, Y. M. (2024). Sustainable remediation of acid mine drainage (AMD)- and crude oil-contaminated sites through wastewater-enhanced Pseudomonas aeruginosa ATCC 15442 treatment. *Biotechnology for the Environment*, *1*(1), 12. https://doi.org/10.1186/s44314-024-00012-3

Baker, J. M. (1970). The effects of oils on plants. *Environmental Pollution (1970)*, *1*(1), 27-44. https://doi.org/10.1016/0013-9327(70)90004-2

Baker, J. M., Little, D. I., & Owens, E. H. (1993). A Review of Experimental Shoreline Oil Spills. *International Oil Spill Conference Proceedings*, *1993*(1), 583-590. https://doi.org/10.7901/2169-3358-1993-1-583

Bragg, J. R., Prince, R. C., Harner, E. J., & Atlas, R. M. (1994). Effectiveness of bioremediation for the Exxon Valdez oil spill. *Nature*, *368*(6470), 413-418. https://doi.org/10.1038/368413a0

Brenner, D. J., & Farmer III, J. J. (2015). Enterobacteriaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-24). https://doi.org/https://doi.org/10.1002/9781118960608.fbm00222

Cappello, S., & Yakimov, M. M. (2010). Oleiophilus. In K. N. Timmis (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 1749-1754). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77587-4_124

Chaudhry, Q., Blom-Zandstra, M., Gupta, S., & Joner, E. J. (2005). Utilising the synergy between plants and rhizosphere microorganisms to enhance breakdown of organic pollutants in the environment. *Environ Sci Pollut Res Int*, *12*(1), 34-48. https://doi.org/10.1065/espr2004.08.213

Couto, M. N., Basto, M. C., & Vasconcelos, M. T. (2011). Suitability of different salt marsh plants for petroleum hydrocarbons remediation. *Chemosphere*, *84*(8), 1052-1057. https://doi.org/10.1016/j.chemosphere.2011.04.069

Cowell, E. B. (1969). The Effects of Oil Pollution on Salt-Marsh Communities in Pembrokeshire and Cornwall. *The Journal of Applied Ecology*, *6*(2). https://doi.org/10.2307/2401533

D´Surney, S. J., & Smith, M. D. (2005). Chemicals of Environmental Concern. In *Encyclopedia of Toxicology* (Second Edition ed., pp. 526-530).

Daims, H. (2014). The Family Nitrospiraceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (pp. 733-749). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_126

Darsa, K. V., & Thatheyus, A. J. (2014). Biodegradation of Petroleum Compound Using Pseudomonas aeruginosa. *OALib*, *01*(05), 1-9. https://doi.org/10.4236/oalib.1100735

Das, N., & Chandran, P. (2011). Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int*, *2011*, 941810. https://doi.org/10.4061/2011/941810

Dastager, S. G., Krishnamurthi, S., Rameshkumar, N., & Dharne, M. (2014). The Family Micrococcaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 455-498). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_168

Dave, D., & Ghaly, A. (2011). Remediation Technologies for Marine Oil Spills: A Critical Review and Comparative Analysis. *American Journal of Environmental Sciences*, *7*(5), 423-440. https://doi.org/10.3844/ajessp.2011.423.440

Dazzo, F. B., & Ganter, S. (2009). Rhizosphere. In *Encyclopedia of Microbiology* (Third Edition ed., pp. 335-349).

de la Haba, R. R., Arahal, D. R., Sánchez-Porro, C., & Ventosa, A. (2014). The Family Halomonadaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Gammaproteobacteria* (pp. 325-360). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38922-1_235

Doronina, N., Kaparullina, E., & Trotsenko, Y. (2014). The Family Methylophilaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 869-880). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_243

Eilertsen, H. C., & Skarðhamar, J. (2006). Temperatures of north Norwegian fjords and coastal waters: Variability, significance of local processes and air–sea heat exchange. *Estuarine, Coastal and Shelf Science*, *67*(3), 530-538. https://doi.org/https://doi.org/10.1016/j.ecss.2005.12.006

Eurostat. (2022, 24.05.2024). *Complete Energy balances*. https://ec.europa.eu/eurostat/databrowser/view/NRG_BAL_C__custom_2063970/bookmark/table?lang=en&bookmarkId=f5136eaa-2469-4b07-b87c-a579145cd3df&page=time:2020

Fakher, S., Ahdaya, M., Elturki, M., & Imqam, A. (2020). Critical review of asphaltene properties and factors impacting its stability in crude oil. *Journal of Petroleum Exploration and Production Technology*, *10*(3), 1183-1200. https://doi.org/10.1007/s13202-019-00811-5

Forquin-Gomez, M.-P., Weimer, B. C., Sorieul, L., Kalinowski, J., & Vallaeys, T. (2014). The Family Brevibacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 141-153). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_169

Freyria, N. J., Góngora, E., Greer, C. W., & Whyte, L. G. (2024). High Arctic seawater and coastal soil microbiome co-occurrence and composition structure and their potential hydrocarbon biodegradation. *ISME Communications*, *4*(1), ycae100. https://doi.org/10.1093/ismeco/ycae100

Gad, S. C. (2005a). Diesel Fuel In *Encyclopedia of Toxicology* (Second Edition ed., pp. 19-22).

Gad, S. C. (2005b). Petroleum Hydrocarbons. In *Encyclopedia of Toxicology* (Second Edition ed., pp. 377-379).

Garcia-Pichel, F. (2009). Cyanobacteria. In M. Schaechter (Ed.), *Encyclopedia of Microbiology (Third Edition)* (pp. 107-124). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012373944-5.00250-9

Garrity, G. M., Bell, J. A., & Lilburn, T. (2015). Pseudomonadaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-1). https://doi.org/https://doi.org/10.1002/9781118960608.fbm00232

Gerhardt, K. E., Huang, X. D., Glick, B. R., & Greenberg, B. M. (2009). Phytoremediation and rhizoremediation of organic soil contaminants: Potential and challenges. *Plant Science*, *176*(1), 20-30. https://doi.org/10.1016/j.plantsci.2008.09.014

Glaeser, S. P., & Kämpfer, P. (2014). The Family Sphingomonadaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 641-707). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_302

Gorrasi, S., Pesciaroli, C., Barghini, P., Pasqualetti, M., & Fenice, M. (2019). Structure and diversity of the bacterial community of an Arctic estuarine system (Kandalaksha Bay) subject to intense tidal currents. *Journal of Marine Systems*, *196*, 77-85. https://doi.org/https://doi.org/10.1016/j.jmarsys.2019.04.004

Greene, A. C. (2014). The Family Desulfuromonadaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria* (pp. 143-155). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39044-9_380

Hershner, C., & Lake, J. (1980). Effects of chronic oil pollution on a salt-marsh grass community. *Marine Biology*, *56*(2), 163-173. https://doi.org/10.1007/bf00397133

Hester, M. W., Willis, J. M., Rouhani, S., Steinhoff, M. A., & Baker, M. C. (2016). Impacts of the Deepwater Horizon oil spill on the salt marsh vegetation of Louisiana. *Environ Pollut*, *216*, 361-370. https://doi.org/https://doi.org/10.1016/j.envpol.2016.05.065

Huber, K. J., & Overmann, J. (2019). Vicinamibacteraceae. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-3). https://doi.org/https://doi.org/10.1002/9781118960608.fbm00319

Huber, K. J., Pascual, J., Foesel, B. U., & Overmann, J. (2017). Blastocatellaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-4). https://doi.org/https://doi.org/10.1002/9781118960608.fbm00290

ITOPF. (2014). *Clean-Up of Oils from Shorelines* (Technical Information Paper, Issue 7). https://www.amn.pt/DCPM/Documents/TIP%207%20Clean-up%20of%20Oil%20From%20Shorelines.pdf

Ivanova, E. P., & Webb, H. K. (2014). The Family Granulosicoccaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Gammaproteobacteria* (pp. 315-317). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38922-1_247

Kerkhof, L. J. (2021). Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol Ecol*, *97*(3). https://doi.org/10.1093/femsec/fiab001

Klein, A. N., Frigon, D., & Raskin, L. (2007). Populations related to Alkanindiges, a novel genus containing obligate alkane degraders, are implicated in biological foaming in activated sludge systems. *Environ Microbiol*, *9*(8), 1898-1912. https://doi.org/10.1111/j.1462-2920.2007.01307.x

Konopka, A. (2009). Ecology, Microbial In *Encyclopedia of Microbiology* (Third Edition ed., pp. 91-106).

Koshlaf, E., & Ball, A. S. (2017). Soil bioremediation approaches for petroleum hydrocarbon polluted environments. *AIMS Microbiol*, *3*(1), 25-49. https://doi.org/10.3934/microbiol.2017.1.25

Kuczynski, J., Liu, Z., Lozupone, C., McDonald, D., Fierer, N., & Knight, R. (2010). Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods*, *7*(10), 813-819. https://doi.org/10.1038/nmeth.1499

Kuiper, I., Lagendijk, E. L., Bloemberg, G. V., & Lugtenberg, B. J. (2004). Rhizoremediation: a beneficial plant-microbe interaction. *Mol Plant Microbe Interact*, *17*(1), 6-15. https://doi.org/10.1094/MPMI.2004.17.1.6

Lage, O. M., Calisto, R., Vitorino, I., & Øvreås, L. (2022). Pirellulaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-7). https://doi.org/https://doi.org/10.1002/9781118960608.fbm00402

Lory, S. (2014). The Family Streptococcaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Firmicutes and Tenericutes* (pp. 367-370). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_345

Ma, Y., Wang, J., Liu, Y., Wang, X., Zhang, B., Zhang, W., Chen, T., Liu, G., Xue, L., & Cui, X. (2023). Nocardioides: "Specialists" for Hard-to-Degrade Pollutants in the Environment. *Molecules*, *28*(21). https://doi.org/10.3390/molecules28217433

Mandic-Mulec, I., Stefanic, P., & van Elsas, J. D. (2015). Ecology of Bacillaceae. *Microbiol Spectr*, *3*(2), TBS-0017-2013. https://doi.org/10.1128/microbiolspec.TBS-0017-2013

Marín, I., & Arahal, D. R. (2014). The Family Beijerinckiaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 115-133). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_255

Matlock, B. (2015). Assessment of nucleic acid purity. *Technical Note*, *52646*, 1-2.

McBride, M. J. (2014). The Family Flavobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (pp. 643-676). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_130

Meagher, R. B. (2000). Phytoremediation of toxic elemental and organic pollutants. *Curr Opin Plant Biol*, *3*(2), 153-162. https://doi.org/10.1016/s1369-5266(99)00054-0

Meehan, C. J., & Beiko, R. G. (2014). A Phylogenomic View of Ecological Specialization in the Lachnospiraceae, a Family of Digestive Tract-Associated Bacteria. *Genome Biology and Evolution*, *6*(3), 703-713. https://doi.org/10.1093/gbe/evu050

Millero, F. J., Feistel, R., Wright, D. G., & McDougall, T. J. (2008). The composition of Standard Seawater and the definition of the Reference-Composition Salinity Scale. *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(1), 50-72. https://doi.org/https://doi.org/10.1016/j.dsr.2007.10.001

Obayori, O. S., Adebusoye, S. A., Adewale, A. O., Oyetibo, G. O., Oluyemi, O. O., Amokun, R. A., & Ilori, M. O. (2009). Differential degradation of crude oil (Bonny Light) by four Pseudomonas strains. *Journal of Environmental Sciences*, *21*(2), 243-248. https://doi.org/https://doi.org/10.1016/S1001-0742(08)62258-5

Octavia, S., & Lan, R. (2014). The Family Enterobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Gammaproteobacteria* (pp. 225-286). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38922-1_167

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S.,…Weedon, J. (2024). vegan: Community Ecology Package. https://doi.org/10.32614/CRAN.package.vegan

Olsen, I. (2014). The Family Fusobacteriaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Firmicutes and Tenericutes* (pp. 109-132). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_213

Oren, A. (2014a). The Family Rhodocyclaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 975-998). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_292

Oren, A. (2014b). The Family Xanthobacteraceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 709-726). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_258

Phillips, R. D. (2005). Fuel Oils. In *Encyclopedia of Toxicology* (Second Edition ed., pp. 385-387).

Posit Team. (2024). *RStudio: Integrated Development Environment for R* In Posit Software, PBC. http://www.posit.co/

Prendergast, D. P., & Gschwend, P. M. (2014). Assessing the performance and cost of oil spill remediation technologies. *Journal of Cleaner Production*, *78*, 233-242. https://doi.org/10.1016/j.jclepro.2014.04.054

Prosser, J. I., Head, I. M., & Stein, L. Y. (2014). The Family Nitrosomonadaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 901-918). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_372

Pujalte, M. J., Lucena, T., Ruvira, M. A., Arahal, D. R., & Macián, M. C. (2014). The Family Rhodobacteraceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 439-512). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_377

Pukall, R. (2014). The Family Dietziaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 327-338). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_188

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org

Ratkowsky, D. A., Olley, J., McMeekin, T. A., & Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *J Bacteriol*, *149*(1), 1-5. https://doi.org/10.1128/jb.149.1.1-5.1982

Ribeiro, H., Mucha, A. P., Almeida, C. M., & Bordalo, A. A. (2013). Bacterial community response to petroleum contamination and nutrient addition in sediments from a temperate salt marsh. *Sci Total Environ*, *458-460*, 568-576. https://doi.org/10.1016/j.scitotenv.2013.04.015

Rizzo, C., Malavenda, R., Gerce, B., Papale, M., Syldatk, C., Hausmann, R., Bruni, V., Michaud, L., Lo Giudice, A., & Amalfitano, S. (2019). Effects of a Simulated Acute Oil Spillage on Bacterial Communities from Arctic and Antarctic Marine Sediments. *Microorganisms*, *7*(12). https://doi.org/10.3390/microorganisms7120632

Röling, W. F. M. (2014). The Family Geobacteraceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria* (pp. 157-172). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39044-9_381

Ron, E. Z., & Rosenberg, E. (2014). Enhanced bioremediation of oil spills in the sea. *Curr Opin Biotechnol*, *27*, 191-194. https://doi.org/10.1016/j.copbio.2014.02.004

Roquigny, R., Novinscak, A., Biessy, A., & Filion, M. (2017). Pseudomonadaceae: From Biocontrol to Plant Growth Promotion. In S. Mehnaz (Ed.), *Rhizotrophs: Plant Growth Promotion to Bioremediation* (pp. 39-68). Springer Singapore. https://doi.org/10.1007/978-981-10-4862-3_3

Rosenberg, E. (2014). The Family Deinococcaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (pp. 613-615). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_127

Shepard, C. C., Crain, C. M., & Beck, M. W. (2011). The protective role of coastal marshes: a systematic review and meta-analysis. *PLoS One*, *6*(11), e27374. https://doi.org/10.1371/journal.pone.0027374

Shivaji, S., Srinivas, T. N. R., & Reddy, G. S. N. (2014). The Family Planococcaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Firmicutes and Tenericutes* (pp. 303-351). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_351

Shugart, L. R. (2005). Bioremediation. In *Encyclopedia of Toxicology* (Second Edition ed., pp. 297-298).

Singh, J., Birbian, N., Sinha, S., & Goswami, A. (2014). A critical review on PCR, its types and applications. *Int. J. Adv. Res. Biol. Sci, 1*(7), 65-80.

Slobodkin, A. (2014). The Family Peptostreptococcaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Firmicutes and Tenericutes* (pp. 291-302). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_217

Stackebrandt, E. (2014). The Family Propionibacteriaceae: Genera other than Propionibacterium. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 725-741). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_194

Stepaniyan, O. V. (2008). Effects of crude oil on major functional characteristics of macroalgae of the Barents Sea. *Russian Journal of Marine Biology*, *34*(2), 131-134. https://doi.org/10.1134/S1063074008020077

Sullivan, M. J. (2005). Oil, Crude. In *Encyclopedia of Toxicology* (Second Edition ed., pp. 292-295).

Tóth, E. M., & Borsodi, A. K. (2014). The Family Nocardioidaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 651-694). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_193

Viebahn, M., Smit, E., Glandorf, D. C. M., Wernars, K., & Bakker, P. A. H. M. (2009). Effect of Genetically Modified Bacteria on Ecosystems and Their Potential Benefits for Bioremediation and Biocontrol of Plant Diseases – A Review. In E. Lichtfouse (Ed.), *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms: Climate change, intercropping, pest control and beneficial microorganisms* (pp. 45-69). Springer Netherlands. https://doi.org/10.1007/978-90-481-2716-0_4

Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., Tam Nora, F.-Y., & Zhou, H.-W. (2012). Comparison of the Levels of Bacterial Diversity in Freshwater, Intertidal Wetland, and Marine Sediments by Using Millions of Illumina Tags. *Applied and Environmental Microbiology*, *78*(23), 8264-8271. https://doi.org/10.1128/AEM.01821-12

Ward, O. P., Singh, A., Van Hamme, J. D., & Voordouw, G. (2009). Petroleum Microbiology. In *Encyclopedia of Microbiology* (Third Edition ed., pp. 443-456).

Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, *173*(2), 697-703. https://doi.org/10.1128/jb.173.2.697-703.1991

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V.,…Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43). https://doi.org/10.21105/joss.01686

Widomska, J., Welc, R., & Gruszecki, W. I. (2019). The effect of carotenoids on the concentration of singlet oxygen in lipid membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1861*(4), 845-851. https://doi.org/https://doi.org/10.1016/j.bbamem.2019.01.012

Wiegel, J., Tanner, R., & Rainey, F. A. (2006). An Introduction to the Family Clostridiaceae. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Eds.), *The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria* (pp. 654-678). Springer US. https://doi.org/10.1007/0-387-30744-3_20

Willems, A. (2014). The Family Comamonadaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 777-851). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_238

Tools used

DeepL Write, DeepL SE: https://www.deepl.com/en/write, used for: reformulation and grammar correction of text phrases. No new content was created.

Appendices

Appendix 1: Buffers for DNA Extraction

1. **TNS buffer (autoclaved)**

***Table 7**: Formula to make Tris-NaCl-SDS buffer.*

500 mM TRIZMA (M.W 121.14)	15.76 g
100 mM NaCl (M.W 58.44)	1.17 g
10 % Sodium Dodecyl Sulphate (SDS) (M.W 288.38)	20 g
Autoclaved Milli-Q® water	200 ml

2. **PEG-6000 (autoclaved)**

***Table 8**: Formula to make Polyethylene glycol buffer.*

Autoclaved Milli-Q® water	up to 400 ml
100 mM NaCl (M.W 58.44)	37.4 g
Polyethylene glycol (6000)	120 g

Appendix 2: Protocol for Library Preparation

MINION | GRIDION | FLONGLE

Rapid sequencing amplicons - 16S barcoding (SQK-16S024)

Overview

Before beginning, we strongly recommend that first-time users read and review the complete online protocol on the Nanopore Community which can be accessed via the QR code. The online protocol provides the full instructions, additional commentary, and considerations for various steps, as well as guidance on input amounts, DNA handling, and sequencing. **This reference card is designed for experienced users.**

1. Library preparation (~100 minutes)

1 Flongle Sequencing Expansion

There are three buffers that come into direct contact with a flow cell at the point of loading (Sequencing Buffer, Flush Buffer/Flow Cell Flush, and Loading Beads/Library Beads or Loading Solution/Library Solution). We have found that there is a very low level of contaminants seeping out of the plastic vials that contain these buffers, and this impacts the robustness of the Flongle flow cell system (MinION and PromethION are not impacted by this).

To rapidly deploy this to Flongle users, we have produced a Flongle Sequencing Expansion (EXP-FSE001 and EXP-FSE002), which can perform 12 Flongle flow cell loads in total. The buffers in the Flongle Sequencing Expansion are shipped in glass vials to avoid the contaminant seeping issue. It is important to use these buffers when using Flongle Flow Cells.

To load a library onto your Flongle flow cell, you will need to use the following components:

Flongle Sequencing Expansion (EXP-FSE001) components

- Sequencing Buffer II (SBII)
- Flush Buffer (FB)
- Loading Beads II (LBII) or Loading Solution (LS)

Sequencing or Flow Cell Priming Kit components

- Flush Tether (FLT)

Flongle Sequencing Expansion (EXP-FSE002) components for Kit 14

- Sequencing Buffer (SB)
- Flow Cell Flush (FCF)
- Library Beads (LIB)
- Library Solution (LIS)

Oxford Nanopore Technologies deem the useful life of the Flongle Flow Cell Priming Kit to be 6 months from receipt by the customer.

2 Take one 96-well plate containing 16S barcodes. Break one set of barcodes (1-24, or as desired) away from the plate and return the rest to storage.

3 The 96-well plates are designed to break in one direction only. Strips, or multiple strips, of eight wells/barcodes can be removed from the plate at any one time.

4 Thaw the desired barcodes, make sure the liquid is at the bottom of the tubes, and place on ice.

5 Thaw the LongAmp Hot Start Taq 2X Master Mix, spin down briefly, mix well by pipetting and place on ice.

6 Prepare the DNA in nuclease-free water.

- Transfer 10 ng genomic DNA into a DNA LoBind tube
- Adjust the volume to 10 µl with nuclease-free water
- Mix thoroughly by flicking the tube, to avoid unwanted shearing
- Spin down briefly in a microfuge

7 For each sample to be tested, prepare the following mixture in separate 0.2 ml thin-walled PCR tubes.

Reagent	Volume
Nuclease-free water	5 µl

Reagent	Volume
Input DNA (10 ng)	10 µl
LongAmp Hot Start Taq 2X Master Mix	25 µl
Total	**40 µl**

If the amount of input material is altered, the number of PCR cycles may need to be adjusted to produce the same yield.

8. Ensure the components are thoroughly mixed by pipetting, and spin down.

9. Using clean pipette tips, carefully pierce the foil surface of the required barcodes. Use a new tip for each barcode to avoid cross-contamination. Make a note of which barcode numbers will be run for each sample.

10. Using a multichannel pipette, mix the 16S barcodes by pipetting up and down 10 times. Transfer 10 µl of each 16S Barcode into respective sample-containing tubes.

The layout of the barcodes in the plate are as follows:

11. Mix thoroughly by pipetting up and down ten times.

12. Amplify using the following cycling conditions:

Cycle step	Temperature	Time	No. of cycles
Initial denaturation	95 °C	1 min	1
Denaturation	95 °C	20 secs	25
Annealing	55 °C	30 secs	25
Extension	65 °C	2 mins	25
Final extension	65 °C	5 mins	1
Hold	4 °C	∞	

13. Transfer each sample to a separate 1.5 ml DNA LoBind Eppendorf tube. Carry out steps 11-21 for each sample, before pooling the samples at step 22.

14. Resuspend the AMPure XP beads by vortexing.

15. Add 30 µl of resuspended AMPure XP beads to the reaction and mix by pipetting.

16. Incubate on a Hula mixer (rotator mixer) for 5 minutes at room temperature.

17. Prepare 500 µl of fresh 70% ethanol in nuclease-free water.

18. Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette off the supernatant when clear and colourless.

19. Keep the tube on the magnet and wash the beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the ethanol using a pipette and discard.

20. Repeat the previous step.

21. Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.

22. Remove the tube from the magnetic rack and resuspend pellet in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. Incubate for 2 minutes at room temperature.

23. Pellet the beads on a magnet until the eluate is clear and colourless.

24. Remove and retain 10 µl of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube.
    - Dispose of the pelleted beads

25. Quantify 1 µl of eluted sample using a Qubit fluorometer.

26. Pool all barcoded libraries in the desired ratios to a total of 50-100 fmoles in 10 µl of 10 mM Tris-HCl pH 8.0 with 50 mM NaCl. For 16S amplicons of ~1500 bp, 50-100 fmoles equates to ~50-100 ng.

27. Add 1 µl of RAP to the barcoded DNA.

28. Mix gently by flicking the tube, and spin down.

29. Incubate the reaction for 5 minutes at room temperature.

30. Following standard input recommendations, the protocol should produce enough final library (adapted DNA in EB) to load at least two Flongle flow cells. We recommend reserving enough library to load a second Flongle flow cell. We recommend loading 3-20 fmol of this final

prepared library onto the flow cell. Loading more than 20 fmol can have a detrimental effect on throughput. Dilute the library in EB or nuclease-free water up to a final volume of 5 µl.

31 Half of the prepared library is used for loading into the Flongle flow cell. Store the library on ice until ready to load. The remaining half of the library can be stored at 4° C for one or two days if required.

2. Loading the Flongle flow cell

1 Flongle Sequencing Expansion

There are three buffers that come into direct contact with a flow cell at the point of loading (Sequencing Buffer, Flush Buffer/Flow Cell Flush, and Loading Beads/Library Beads or Loading Solution/Library Solution). We have found that there is a very low level of contaminants seeping out of the plastic vials that contain these buffers, and this impacts the robustness of the Flongle flow cell system (MinION and PromethION are not impacted by this).

To rapidly deploy this to Flongle users, we have produced a Flongle Sequencing Expansion (EXP-FSE001 and EXP-FSE002), which can perform 12 Flongle flow cell loads in total. The buffers in the Flongle Sequencing Expansion are shipped in glass vials to avoid the contaminant seeping issue. It is important to use these buffers when using Flongle Flow Cells.

To load a library onto your Flongle flow cell, you will need to use the following components:

Flongle Sequencing Expansion (EXP-FSE001) components

- Sequencing Buffer II (SBII)
- Flush Buffer (FB)
- Loading Beads II (LBII) or Loading Solution (LS)

Sequencing or Flow Cell Priming Kit components

- Flush Tether (FLT)

Flongle Sequencing Expansion (EXP-FSE002) components for Kit 14

- Sequencing Buffer (SB)
- Flow Cell Flush (FCF)
- Library Beads (LIB)
- Library Solution (LIS)

Oxford Nanopore Technologies deem the useful life of the Flongle Flow Cell Priming Kit to be 6 months from receipt by the customer.

2 Please note that the Sequencing Tether (SQT) tube will NOT be used in this protocol. It is provided in the kit for potential future product compatibility.

3 Do NOT touch the reverse side of the Flongle flow cell array or the contact pads on the Flongle adapter. ALWAYS wear gloves when handling Flongle flow cells and adapters to avoid damage to the flow cell or adapter.

4 The diagram below shows the components of the Flongle flow cell:

The seal tab, air vent, waste channel, drain port and sample port are visible here. The sample port, drain port and air vent only become accessible once the seal tab is peeled back.

5 Thaw the Sequencing Buffer II (SBII), Loading Beads II (LBII) and Flush Buffer (FB) from the Flongle Sequencing Expansion and Flush Tether (FLT) from your sequencing kit at room temperature.

6 Mix the Sequencing Buffer II (SBII), Flush Buffer (FB) and Flush Tether (FLT) tubes by vortexing and spin down at room temperature.

7 Place the Flongle adapter into the MinION or one of the five GridION positions.

The adapter should sit evenly and flat on the MinION Mk1B or GridION platform. This ensures the flow cell assembly is flat during the next stage.

8 The adapter needs to be plugged into your device, and the device should be plugged in and powered on before inserting the Flongle flow cell.

9 Place the flow cell into the Flongle adapter, and press the flow cell down until you hear a click.

The flow cell should sit evenly and flat inside the adapter, to avoid any bubbles forming inside the fluidic compartments.

10 In a fresh 1.5 ml Eppendorf DNA LoBind tube, mix 117 µl of Flush Buffer (FB) with 3 µl of Flush Tether (FLT) and mix by pipetting.

11 Peel back the seal tab from the Flongle flow cell, up to a point where the sample port is exposed, as follows:

1. Lift up the seal tab:

2. Pull the seal tab to open access to the sample port:

3. Hold the seal tab open by using adhesive on the tab to stick to the MinION Mk 1B lid:

12 To prime your flow cell with the mix of Flush Buffer (FB) and Flush Tether (FLT) that was prepared earlier, ensure that there is no air gap in the sample port or the pipette tip. Place the P200 pipette tip inside the sample port and slowly dispense the priming fluid into the Flongle flow cell by slowly pipetting down. We also recommend twisting the pipette plunger down to avoid flushing the flow cell too vigorously.

13 Vortex the vial of Loading Beads II (LBII). Note that the beads settle quickly, so immediately prepare the Sequencing Mix in a fresh 1.5 ml Eppendorf DNA LoBind tube for loading the Flongle, as follows:

Reagents	Volume
Sequencing Buffer II (SBII)	15 µl
Loading Beads II (LBII) mixed immediately before use, or Loading Solution (LS), if using. LS can be used instead of LBII when preparing libraries with the Ligation Sequencing Kit (SQK-LSK110)	10 µl
DNA library	5 µl
Total	**30 µl**

14 To add the Sequencing Mix to the flow cell, ensure that there is no air gap in the sample port or the pipette tip. Place the P200 tip inside the sample port and slowly dispense the Sequencing Mix into the flow cell by slowly pipetting down. We also recommend twisting the pipette plunger down to avoid flushing the flow cell too vigorously.

15 Seal the Flongle flow cell using the adhesive on the seal tab, as follows:

1. Stick the transparent adhesive tape to the sample port.

2. Replace the top (Wheel icon section) of the seal tab to its original position.

16 Replace the sequencing platform lid.

Appendix 3: PCR Formula

***Table 9**: Formula for PCR reaction using the 16 S primer fD1 and rD1.*

Autoclaved Milli-Q® water	20.4 µl (variable with sample)
Buffer (x10)	2.5 µl
dNTP (100 mmol)	0.4 µl
forward primer fD1 (100 mmol)	0.25 µl
Reverse primer rD1 (100 mmol)	0.25 µl
Enzyme (5U/ µl)	0.2 µl
Template DNA	1 µl (variable)

Appendix 4: Primer Sequences

***Table 10**: Sequence for 16 S primer fD1 and rD1.*

Primer	Sequence (5' - 3')	Reference
fD1	AGA GTT TGA TCC TGG CTC AG	(Weisburg et al., 1991)
rD1	AAG GAG GTG ATC CAG CC	(Weisburg et al., 1991)

Appendix 5: R script for Analysis

```
# LIBRARIES + FUNCTIONS
library(tidyverse)

library(scales)

library(ggthemes)
library(ggfortify)
library(ggrepel)
library(ggpubr)

library(vegan)
library(fossil)

library(patchwork)

give.n <- function(y) {
  return(data.frame(y = y, label = paste0("n = ", length(y))))

# CUSTOM COLOR PALETTE
pal <- c("Other" = "cornsilk2", "Identified" = "cornsilk2",
"Unknown" = "gray50", "plant" = "darkgreen", "soil" = "burlywood",
"Bacillaceae" = "yellowgreen", "Granulosicoccaceae" = "green4",
"Pseudomonadaceae" = "lightgreen", "Micrococcaceae" = "indianred",
"Burkholderiaceae" = "plum", […])


##############################################################################
#  CLEANING + FORMATTING
```

```r
##############################################################################

data  <- read.delim("align_silva.tsv", sep = "\t", header = T)
align <- data %>% gather(sample_ID, abund, 2:66)
align <- align %>% separate(tax, c("domain", "kingdom","phylum", "class",
"order", "family", "genus"), sep = ";") %>%
  mutate(ident = ifelse(family != "Unknown", "Identified", "Unknown"))

align$sample_ID <- substr(align$sample_ID, 2, 10)  # removing 'X'
align$sample    <- substr(align$sample_ID, 1, 4)   # pooling replicates
align$kingdom   <- substr(align$kingdom, 1, 8)     # kingdom clean-up

matrix      <- align %>%  select(family, abund, sample) %>%
create.matrix(tax.name = "family", locality = "sample", abund.col =
"abund", abund = T ) %>% t()
# "Unknown_Family" is counted as separate family


##############################################################################
#  VARIABLES
##############################################################################

variables <- unique(align$sample) %>% as.data.frame() %>%
set_names("sample")

# treatment
variables$treatment <- substr(variables$sample, 3, 4)
variables$treatment <- factor(variables$treatment, levels = c("PT", "PC",
"PL", "PH", "ST", "SC", "SL", "SH"))
variables$treatment <- recode_factor(variables$treatment,
"PT" = "Plant Tisnes", "PC" = "Plant Control", "PL" = "Plant Low",
"PH" = "Plant High", "ST" = "Soil Tisnes", "SC" = "Soil Control",
"SL" = "Soil Low", "SH" = "Soil High")

# group
variables$group <- substr(variables$sample, 3, 3)
variables <- variables %>% mutate(group = case_when(group == "P" ~
"plant", group == "S" ~ "soil"))
variables$group <- factor(variables$group, levels = c("plant", "soil"))

# time
variables$time  <- substr(variables$sample, 1, 1)
variables       <- variables %>% mutate(time = case_when(
time %in% "1" ~ "0", time %in% "2" ~ "1",
time %in% "3" ~ "7", time %in% "4" ~ "13"))
variables$time  <- factor(variables$time, levels = c("0", "1", "7", "13"))


##############################################################################
#  TAXONOMY
##############################################################################

max_fam <- align %>% group_by(family) %>% summarise("total" = sum(abund))
%>%
```

```
  arrange(desc(total)) %>% rownames_to_column(var = "position") %>%
head(10)

family <- align %>% filter(family != "Unknown") %>%
  group_by(sample, family) %>% reframe(abund = sum(abund)) %>%
  filter(abund != 0) %>% arrange(desc(abund)) %>%
  group_by(sample) %>%
  mutate(common = ifelse(row_number() <= 5, family, "Other"),
         percentage = round(abund/(sum(abund))*100,digits = 2),
         label = paste("n=", sum(abund)))


##############################################################################
#  MERGING
##############################################################################

align <- merge(align, variables, by = "sample")
family<- merge(family, variables, by = "sample")


##############################################################################
# 1. DNA EXTRACTION
##############################################################################

DNA_full <- read.csv("DNA.csv", sep = ";", header = T, dec = ",")

DNA      <- DNA_full %>% subset(sample.ID %in% align$sample_ID) %>%
subset(select = -soil_weight) %>% pivot_longer(cols = !sample.ID, names_to
= c(".value", "run"), names_sep = "_")

DNA$sample  <- substr(DNA$sample.ID, 1, 4)
DNA         <- merge(DNA, variables, by = "sample")
means_DNA   <- DNA %>% group_by(treatment) %>% summarise(mean(DNA),
sd(DNA), mean(X260280), sd(X260280))


##############################################################################
# 2. IDENTIFIED VS UNKNOWN SPECIES
##############################################################################

pct <- align %>%
  group_by(sample_ID, sample) %>% count(ident, wt = abund) %>%
  mutate(ratio = n/sum(n) * 100,
         position = sum(n) + 1,
         label = ifelse(ident != "Identified", NA, round(ratio, digits =
0)),
         xlabel = substr(sample_ID, 6, 9),
         xlabel = str_replace(xlabel, "_2", ".2"))

pct <- merge(pct, variables, by = "sample")

# plot
a <- pct %>% subset(group == "plant") %>%
  mutate(xlabel = paste(as.character(time), xlabel, sep = ".")) %>%
```

```
  ggplot(aes(x = fct_reorder(xlabel, as.integer(time)), y=n, fill =
ident)) +
  geom_col(position = "stack") +
  geom_text(aes(label = label, y = position),  vjust = -0.5, size = 4.5) +
  facet_grid(.~ treatment, space = "free_x", scales = "free_x") +
  scale_y_continuous(expand = c(0,0), limits = c(0, 700)) +
  scale_fill_manual(values = pal) +
  labs(y = "Reads [Number of Sequences]", x = "", fill = "") +
  theme_bw(base_size = 20) +
  theme(axis.text = element_text(color = "black"),
        axis.text.x = element_text(angle = 90, hjust = 0.9),
        axis.title.x = element_blank(),
        strip.background = element_rect(fill="white"),
        plot.title = element_text(hjust = 0.5))

b <- pct %>% subset(group == "soil") %>%
  mutate(xlabel = paste(as.character(time), xlabel, sep = ".")) %>%
  ggplot(aes(x= fct_reorder(xlabel, as.integer(time)), y=n, fill=ident)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = label, y = position),  vjust = -0.5, size = 4.5) +
  facet_grid(.~ treatment, space = "free_x", scales = "free_x") +
  scale_y_continuous(expand = c(0,0), limits = c(0, 700)) +
  scale_fill_manual(values = pal) +
  labs(x = "Sample ID", y = "Reads [Number of Sequences]", fill = "") +
  theme_bw(base_size = 20) +
  theme(axis.text = element_text(color = "black"),
        axis.text.x = element_text(angle = 90, hjust = 0.9),
        strip.background = element_rect(fill="white"),
        plot.title = element_text(hjust = 0.5))

(a/b) + plot_annotation(title = 'Number of Reads and Percentage of
Identified Families',
theme = theme(plot.title = element_text(hjust = 0.5, size = 20))) +
        plot_layout(axes = "collect") +
        plot_layout(guides = "collect") & theme(legend.position = "top")


##############################################################################
#  3. DIVERSITY INDICES
##############################################################################

shannon   <- diversity(matrix, "shannon")
richness  <- estimateR(matrix)
evenness  <- shannon/ log(specnumber(matrix))

diversity <- cbind(shannon, t(richness), evenness) %>% as.data.frame() %>%
             rownames_to_column(var = "sample")
diversity <- merge(diversity, variables, by = "sample")


#### RAREFACTION CURVE ####

rarefaction <- rarecurve(matrix, tidy = T)
colnames(rarefaction) <- c("sample", "samplesize", "species")
```

```
rarefaction <- merge(rarefaction, variables, by = "sample")

# plot
rarefaction %>% group_by(sample) %>%
  mutate(time = as.character(time)) %>%
  mutate(label = if_else(samplesize == max(samplesize),
as.character(time), NA_character_)) %>%
  ggplot(aes(samplesize, species, group = sample) ) +
  geom_line(linewidth = 1) +
  geom_label_repel(aes(label = label), na.rm = T, nudge_x = 1, size = 5) +
  facet_wrap(~ treatment, nrow = 2, axes = "all_x") +
  labs(y = "Number of Families", x = "Number of Reads",
       title = "Rarefaction Curves per Treatment") +
  theme_bw(base_size = 20) +
  theme(axis.text = element_text(color = "black"),
        plot.title = element_text(hjust = 0.5),
        strip.background=element_rect(fill="white"),
        legend.position = "none")


#### SPECIES RICHNESS ####

# plot
diversity %>%
  select(sample, treatment, time, S.obs, S.ACE, se.ACE) %>%
  pivot_longer(cols = S.obs:S.ACE, names_to = "index") %>%
  mutate(se.ACE = ifelse(index == "S.obs", NA, se.ACE)) %>%
  ggplot(aes(time, value, fill = fct_relevel(index, "S.obs"))) +
  geom_col(position = "dodge", width = 0.8) +
  geom_errorbar(aes(x= time, ymax = value + se.ACE, ymin = value -se.ACE),
width = 0.3, position = position_dodge(width = 0.8)) +
  facet_wrap(~treatment, nrow = 2, axes = "all_x") +
  scale_y_continuous(expand = c(0,0), limits = c(0, 190)) +
  scale_fill_manual(values = c("olivedrab", "wheat"),
                    labels = c("Observation", "Estimate (ACE index)")) +
  theme_bw(base_size = 20) +
  theme(plot.title = element_text(hjust = 0.5),
        strip.background=element_rect(fill="white"),
        legend.position = "top",
        axis.text = element_text(color = "black")) +
  labs(title = "Family Richness per Treatment and Time", fill = "",
       x = "Time [Weeks]", y = "Family Richness")


##### SHANNON INDEX #####

# plot
ggplot(diversity, aes(time, shannon, group = treatment)) +
  geom_smooth(aes(group = treatment, colour = group), method="lm", se=F) +
  geom_point(size = 3) +
  stat_cor(label.y=1,r.accuracy=0.01, aes(label=  ..rr.label.., size=6)) +
  facet_wrap(~ treatment, nrow = 2, axes = "all_x") +
  scale_color_manual(values = pal) +
  theme_bw(base_size = 20) +
```

```
  theme(plot.title = element_text(hjust = 0.5),
        legend.position = "none",
        strip.background=element_rect(fill="white"),
        axis.text = element_text(color = "black")) +
  labs(title = "Shannon Index over Time and Treatment", y = "Shannon
Index", x = "Time [Weeks]")


##############################################################################
#  4. SPECIES COMPOSITION
##############################################################################

# NUMBER OF DIFFERENT TAXA
n_distinct(align$genus)         # 591 genera
n_distinct(align$family)        # 260 families
n_distinct(align$order)         # 131 orders
n_distinct(align$class)         #  48 classes


##### BUCKET EFFECT #####

bucket <- family %>%
  mutate(bucket = case_when(
sample %in% c("2_PT", "2_PC") ~ "Plant Week 1",
sample %in% c("4_PT", "4_PC") ~ "Plant Week 13",
sample %in% c("2_ST", "2_SC") ~ "Soil Week 1",
sample %in% c("4_ST", "4_SC") ~ "Soil Week 13")) %>% na.omit()

# plot
ggplot(bucket, aes(x = time, y = percentage, fill = common)) +
  geom_col() +
  stat_unique(geom = "text", aes(x = time, y=-5, label = label), size=5) +
  facet_wrap(~ treatment, nrow = 2, axes = "all_x") +
  scale_fill_manual(values = pal) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(-10, 100))+
  scale_x_discrete(expand = c(0, 0)) +
  labs(title = "Effect of Cultivation on Relative Family Abundance",
       y = "Abundance [%]", fill = "", x = "Time [Weeks]") +
  theme_bw(base_size = 20) +
  theme(plot.title = element_text(hjust = 0.5),
        strip.background=element_rect(fill="white"),
        legend.position = "top",
        axis.text = element_text(color = "black")) +
  guides(fill = guide_legend(nrow = 6))


##### RELATIVE FAMILY ABUNDANCE #####

# plot
family %>% filter(!treatment %in% c("Soil Tisnes", "Plant Tisnes")) %>%
  ggplot(aes(x = time, y = percentage, fill = common)) +
  geom_col() +
  stat_unique(geom = "text", aes(x = time, y=-5, label = label), size=6) +
```

```
  facet_wrap(~ treatment, nrow = 2, axes = "all_x") +
  scale_fill_manual(values = pal) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(-10, 100)) +
  scale_x_discrete(expand = c(0, 0)) +
  labs(y = "Abundance [%]", fill = "", x = "Time [Weeks]",
       title = " Relative Family Abundance over Time per Treatment") +
  theme_bw(base_size = 20) +
  theme(plot.title = element_text(hjust = 0.5),
        strip.background=element_rect(fill="white"),
        legend.position = "top",
        axis.text = element_text(color = "black")) +
  guides(fill = guide_legend(nrow = 9))


##############################################################################
#  5. TARGET FAMILIES
##############################################################################

target_family <- c("Actinomycetes", "Alcaligenaceae", "Bacillaceae",
"Brevibacteriaceae", "Burkholderiaceae", "Cellulomonadaceae",
"Corynebacteriaceae", "Dietziaceae", "Flavobacteriaceae", "Gordoniaceae",
"Halomonadaceae", "Micrococcaceae", "Moraxellaceae", "Mycobacteriaceae",
"Nocardiaceae", "Nocardioidaceae",  "Oleiphilaceae", "Phyllobacteriaceae",
"Piscirickettsiaceae", "Pseudomonadaceae", "Sphingomonadaceae",
"Oceanospirillaceae", "Planococcaceae", "Rhodocyclaceae",
"Geobacteraceae", "Comamonadaceae", "Xanthobacteraceae",
"Enterobacteriaceae", "Rhizobiaceae", "Granulosicoccaceae",
"Roseobacteraceae", "Rhodobacteraceae", "Porticoccaceae")

# plot
family %>% filter(family %in% target_family, .preserve = T) %>%
  group_by(sample) %>%
  mutate(label = paste("n=", sum(abund))) %>%
  ggplot(aes(time, percentage, fill = family)) +
  geom_col() +
  stat_unique(geom = "text", aes(time, -5, label = label), size = 4.5) +
  facet_wrap(~ treatment, nrow = 2, axes = "all_x") +
  labs(y = "Abundance [%]", fill = "", x = "Time [Weeks]",
       title = "Relative Abundance of 'Target Families'") +
  scale_fill_manual(values = pal) +
  theme_bw(base_size = 20) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text = element_text(color = "black"),
        strip.background = element_rect(fill="white"),
        legend.position = "top")+
  guides(fill = guide_legend(nrow = 7))
```

Appendix 6: DNA Extraction

***Table 11**: ND results after DNA extraction for all laboratory replicates.*

		1st ND measurement		2nd ND measurement	
sample ID	**Soil Weight (g)**	**DNA (ng/ µl)**	**A 260/ 280**	**DNA (ng/ µl)**	**A 260/ 280**
1_PC_1a	0.225	2	1.95	2.9	2.7
1_PC_1b	0.201	0.8	-11.06	1.7	0.8
1_PC_2a	0.209	3.9	1.6	1.8	1.28
1_PC_2b	0.204	1.7	2.54	1.3	0.76
1_PC_3a	0.206	2.9	1.41	1.4	1.14
1_PC_3b	0.196	1.7	3.32	1.8	1.15
1_PH_1a	0.209	2.6	4.95	3.6	2.27
1_PH_1b	0.201	1.5	2.07	3.1	1.45
1_PH_2a	0.215	0.8	6.1	1.8	4.85
1_PH_2b	0.2	1.3	2.39	2.1	0.98
1_PH_3a	0.203	15.2	1.63	15.5	1.88
1_PL_1a	0.208	18.4	1.74	17.3	1.59
1_PL_2a	0.198	2.7	0.96	1.3	1.74
1_PL_2b	0.199	0.2	-0.55	0.7	-12
1_PL_3a	0.215	11.4	1.74	11.8	1.73
1_SC_1a	0.203	2	1.9	1.5	2.75
1_SC_1b	0.195	0.7	1.1	1.8	0.85
1_SH_1a	0.208	0.3	-0.4	1.5	2.38
1_SH_1b	0.201	1.6	3.72	1.5	0.86
1_SL_1a	0.214	1.4	1.4	1.4	1.04
1_SL_1b	0.198	1.6	1.96	1.8	1
1_SL_1c	pooled	2.7	1.28	3	1.08
2_PC_1a	0.201	22.1	1.5	19.9	1.54
2_PC_2a	0.236	160	1.96	163.1	1.93
2_PC_3a	0.218	34.6	1.58	35	1.62
2_PH_1a	0.201	14.3	1.66	14.7	1.79
2_PH_2a	0.215	34.4	1.8	36.1	1.68
2_PH_3a	0.215	5.2	1.2	4.7	1.27
2_PH_3b	0.196	5.2	2.15	6.4	1.56
2_PL_1a	0.197	115.4	1.79	116.8	1.82
2_PL_2a	0.207	0.4	1.09	0.3	0.32
2_PL_2b	0.21	-0.1	0.23	2.4	0.7
2_PL_2c	0.205	0.4	0.53	1.2	3.41
2_PL_3a	0.208	1.8	0.87	0.9	2.48
2_PL_3b	0.214	0.2	0.22	2.6	0.99
2_PL_3b	0.205	1.6	0.96	-0.5	0.45
2_PT_1a	0.207	68.8	1.73	68.9	1.78
2_PT_2a	0.212	4.8	2.21	4	3.44
2_PT_2b	0.197	4	1.95	4.9	1.06
2_PT_3a	0.212	25.3	1.73	26	1.67
2_SC_1a	0.216	2.6	0.72	0.7	1.15

2_SC_1b	0.203	4	1.31	4.2	1.36
2_SH_1a	0.205	2.4	0.94	3.6	2.57
2_SH_1b	0.203	1	8.28	1.8	0.92
2_SH_1c	pooled	2.5	1.2	3.8	1.31
2_SL_1a	0.209	5.1	20.76	3.9	1.58
2_SL_1b	0.2	2.2	8.53	2.8	1.41
2_ST_1a	0.203	2.6	1.33	3.6	1.17
2_ST_1b	0.203	-1.3	1.33	2	0.78
3_PC_1a	0.201	4.1	1.11	4.1	1.53
3_PC_1b	0.201	3.2	4.33	3.5	0.92
3_PC_2a	0.211	8.8	1.57	8.9	1.88
3_PC_3a	0.208	12.4	1.91	12.8	1.7
3_PH_1a	0.248	8.4	1.68	9.8	1.48
3_PH_1b	0.201	0.7	-2.02	2.2	1.13
3_PH_2a	0.203	5.4	1.85	6.6	1.18
3_PH_3a	0.21	26.9	1.85	28.6	1.58
3_PL_1a	0.206	8.1	1.3	8.1	1.41
3_PL_1b	0.204	5.1	1.31	3.5	1.67
3_PL_2a	0.194	1.5	1.31	2.2	0.75
3_PL_2b	0.199	1.2	1.49	2.3	5.92
3_PL_3a	0.192	59.2	1.78	57.2	1.73
3_SC_1a	0.197	1	0.84	2.4	1
3_SC_1b	0.198	1.8	2.67	1.6	3.79
3_SH_1a	0.209	2.5	1.51	3.4	1.4
3_SH_1b	0.204	1.4	17.49	0.4	-0.52
3_SL_1a	0.208	1.8	1.48	1.9	0.8
3_SL_1b	0.197	1.6	2.78	0.9	2.55
3_SL_1c	pooled	5.2	1.62	4.2	1.25
4_PC_1a	0.206	28.1	1.71	27.8	1.49
4_PC_2a	0.206	5	1.98	5.2	1.22
4_PC_3a	0.204	16	1.52	14.9	1.44
4_PH_1a	0.205	12.6	1.8	12	1.59
4_PH_2a	0.2	139.1	1.87	136.2	1.86
4_PH_3a	0.199	110.1	1.81	109.8	1.79
4_PL_1a	0.2	86.8	1.78	83.5	1.74
4_PL_2a	0.205	37.2	1.84	36.5	1.72
4_PL_3a	0.201	52.2	1.72	51.6	1.63
4_PT_1a	0.204	3.7	1.35	3.4	1.84
4_PT_1b	0.199	2.4	1.99	0.2	-3.22
4_PT_2a	0.204	23.6	1.66	23.2	1.77
4_PT_3a	0.206	4.8	1.89	4.7	1.23
4_SC_1a	0.209	2.3	6.15	2.1	1.18
4_SC_1b	0.2	1.8	3.44	0.3	-9.42
4_SH_1a	0.203	3.4	2.52	2.8	1.09
4_SH_1b	0.196	2.2	1.04	1.7	1.55
4_SL_1a	0.195	2.4	1.94	0.9	0.58

4_SL_1b	0.204	0.9	7.8	-0.1	0.95
4_ST_1a	0.201	2.5	1.03	2.5	1.35
4_ST_1b	0.199	1.2	1.41	0.7	-6.51
4_ST_1c	pooled	2	1.13	2.9	1.09

Appendix 7: Abundance Table

***Table 12**: Abundance table for the ten most abundant genera per sample. 'n' is the number of reads for that sample, and 'total' are all reads for that genus across all treatments.*

Sample	Phylum	Class	Order	Family	Genus	n	total
1_PC	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Escherichia-Shigella	5	5
1_PC	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Jeotgalicoccus	1	1
1_PC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
1_PC	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	1	110
1_PC	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas	1	66
1_PC	Actinobacteriota	Actinobacteria	Micrococcales	Micrococcaceae	Arthrobacter	1	14
1_PC	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	1	271
1_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	20	586
1_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	18	522
1_PH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	17	253
1_PH	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae	Woeseia	16	92
1_PH	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	14	199
1_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	11	586
1_PH	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae	Cyclobacteriaceae_uncultured	9	49
1_PH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Rhodoferax	8	73
1_PH	Bacteroidota	Bacteroidia	Chitinophagales	Saprospiraceae	Saprospiraceae_uncultured	8	64
1_PH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Comamonadaceae_uncultured	8	59
1_PL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	55	263
1_PL	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	28	522
1_PL	Firmicutes	Clostridia	Eubacteriales	Alkalibacteraceae	Alkalibacter	23	30
1_PL	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	22	522
1_PL	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	18	199
1_PL	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	15	201
1_PL	Firmicutes	Desulfitobacteriia	Desulfitobacteriales	Desulfitobacteriaceae	Desulfosporosinus	15	42
1_PL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	14	586

1_PL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	13	81
1_PL	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	13	139
1_SC	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	6	271
1_SC	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Staphylococcus	2	41
1_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
1_SC	Proteobacteria	Gammaproteobacteria	Enterobacterales	Pasteurellaceae	Haemophilus	1	4
1_SC	Fusobacteriota	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	1	1
1_SC	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Veillonellaceae	Veillonella	1	1
1_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
1_SC	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Methylobacterium-Methylorubrum	1	1
1_SC	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	1	77
1_SH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	25	522
1_SH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	15	253
1_SH	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Desulfuromonadaceae	Desulfuromonas	9	72
1_SH	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae	Woeseia	8	92
1_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	8	73
1_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	6	586
1_SH	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	6	199
1_SH	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	5	90
1_SH	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacteraceae	Unclassified_Vicinamibacteraceae	5	80
1_SH	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Dokdonella	5	61
1_SH	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Maribacter	5	17
1_SL	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	26	110
1_SL	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacteraceae	Unclassified_Vicinamibacteraceae	24	80
1_SL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	19	586
1_SL	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	18	75
1_SL	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	17	253
1_SL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Nitrosomonadaceae	MND1	15	31
1_SL	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	14	73
1_SL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Nitrosomonadaceae	IS-44	11	25

1_SL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	9	263
1_SL	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae	Cyclobacteriaceae_uncultured	7	49
2_PC	Proteobacteria	Gammaproteobacteria	Burkholderiales	SC-I-84	Unclassified_SC-I-84	16	66
2_PC	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	14	522
2_PC	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	13	199
2_PC	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	10	201
2_PC	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	8	253
2_PC	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	8	90
2_PC	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Comamonadaceae_uncultured	8	59
2_PC	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	7	522
2_PC	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	7	199
2_PC	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	7	139
2_PC	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae	Woeseia	7	92
2_PC	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rhodanobacteraceae_uncultured	7	23
2_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	49	522
2_PH	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	31	139
2_PH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	25	201
2_PH	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Paenibacillus	25	105
2_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	23	522
2_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	21	586
2_PH	Proteobacteria	Gammaproteobacteria	Enterobacterales	Aeromonadaceae	Aeromonas	17	18
2_PH	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Bythopirellula	15	53
2_PH	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	14	199
2_PH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	11	253
2_PL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	2	586
2_PL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	2	263
2_PL	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Blastopirellula	1	72
2_PL	Cyanobacteria	Cyanobacteriia	Leptolyngbyales	Leptolyngbyaceae	Leptolyngbya PCC-6306	1	1
2_PL	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	1	271
2_PL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Lysinibacillus	1	119

2_PL	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	1	90
2_PL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Paenisporosarcina	1	14
2_PL	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	1	271
2_PL	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	1	201
2_PL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	1	81
2_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Nitrosomonadaceae	Ellin6067	1	22
2_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Oxalobacteraceae	Massilia	1	12
2_PL	Proteobacteria	Gammaproteobacteria	Enterobacterales	Yersiniaceae	Serratia	1	6
2_PL	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Thermobacillus	1	1
2_PT	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	114	522
2_PT	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	32	199
2_PT	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	29	522
2_PT	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	15	100
2_PT	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	13	90
2_PT	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	11	253
2_PT	Actinobacteriota	Acidimicrobiia	Microtrichales	Microtrichaceae	Sva0996 marine group	11	41
2_PT	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	10	139
2_PT	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	10	263
2_PT	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Yoonia-Loktanella	9	63
2_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	125	586
2_SC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	29	263
2_SC	Firmicutes	Bacilli	Bacillales	Planococcaceae	Lysinibacillus	25	119
2_SC	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	20	253
2_SC	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	15	77
2_SC	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	14	81
2_SC	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	11	75
2_SC	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Desulfuromonadaceae	Desulfuromonas	11	72
2_SC	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae	Cyclobacteriaceae_uncultured	9	49
2_SC	Firmicutes	Bacilli	Bacillales	Planococcaceae	Planococcaceae_uncultured	9	25
2_SH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	15	201

2_SH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	7	201
2_SH	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	4	271
2_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Methylophilaceae	Methylotenera	4	50
2_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	3	586
2_SH	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Paenibacillus	3	105
2_SH	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Desulfuromonadaceae	Desulfuromonas	3	72
2_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Methylophilaceae	Methylotenera	3	50
2_SH	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	2	110
2_SH	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	2	100
2_SH	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacteraceae	Unclassified_Vicinamibacteraceae	2	80
2_SH	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	2	77
2_SH	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Hoeflea	2	29
2_SH	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonadaceae_uncultured	2	22
2_SH	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Granulicatella	2	2
2_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	2	586
2_SH	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	2	263
2_SH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	2	253
2_SH	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	2	139
2_SH	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	2	77
2_SH	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Dokdonella	2	61
2_SH	Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosiaceae	Devosia	2	61
2_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	2	49
2_SH	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae	Rheinheimera	2	15
2_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	2	13
2_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas	2	12
2_SH	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Flavimaricola	2	8
2_SH	Proteobacteria	Gammaproteobacteria	Enterobacterales	Erwiniaceae	Erwinia	2	2
2_SL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	149	586
2_SL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Lysinibacillus	49	119
2_SL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	34	81

2_SL	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	31	201
2_SL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Planococcaceae_uncultured	10	25
2_SL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Peptostreptococcaceae_uncultured	10	13
2_SL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	6	263
2_SL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	6	166
2_SL	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Desulfuromonadaceae	Desulfuromonas	6	72
2_SL	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	5	253
2_ST	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	19	522
2_ST	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Yoonia-Loktanella	5	63
2_ST	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	4	586
2_ST	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	3	271
2_ST	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae	Woeseia	3	92
2_ST	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Rhodoferax	3	73
2_ST	Firmicutes	Desulfitobacteriia	Desulfitobacteriales	Desulfitobacteriaceae	Desulfosporosinus	3	42
2_ST	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	2	253
2_ST	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	2	100
2_ST	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	2	77
2_ST	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Blastopirellula	2	72
2_ST	Acidobacteriota	Blastocatellia	Blastocatellales	Blastocatellaceae	Blastocatella	2	63
2_ST	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonadaceae_uncultured	2	39
2_ST	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Sulfitobacter	2	21
2_ST	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Leptothrix	2	11
2_ST	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Unclassified_Flavobacteriaceae	2	11
2_ST	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Pricia	2	8
3_PC	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	13	522
3_PC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	13	263
3_PC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	10	166
3_PC	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	10	110
3_PC	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Pseudorhodobacter	10	31
3_PC	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	9	253

3_PC	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas	9	66
3_PC	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas	8	66
3_PC	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Aphanizomenon NIES81	8	30
3_PC	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	7	253
3_PC	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	7	199
3_PC	Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosiaceae	Devosia	7	61
3_PC	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	7	522
3_PC	Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosiaceae	Devosia	7	61
3_PC	Acidobacteriota	Blastocatellia	Blastocatellales	Blastocatellaceae	JGI 0001001-H03	7	57
3_PH	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia PCC-9350	20	70
3_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	19	586
3_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	10	522
3_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	9	522
3_PH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	9	253
3_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	9	586
3_PH	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Aphanizomenon NIES81	6	30
3_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	5	586
3_PH	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	5	73
3_PH	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	5	199
3_PL	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	220	271
3_PL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family XI	Anaerococcus	28	33
3_PL	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Staphylococcus	25	41
3_PL	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	14	522
3_PL	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family XI	Peptoniphilus	11	13
3_PL	Firmicutes	Clostridia	Clostridia_Incertae_sedis	Gracilibacteraceae	Lutispora	10	17
3_PL	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	7	271
3_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	7	49
3_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Comamonadaceae_uncultured	6	59
3_PL	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	5	201
3_PL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	5	166

3_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Rhizobacter	5	25
3_SC	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	5	308
3_SC	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	2	100
3_SC	Deinococcota	Deinococci	Deinococcales	Deinococcaceae	Deinococcus	2	2
3_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
3_SC	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	1	90
3_SC	Firmicutes	Bacilli	Bacillales	Planococcaceae	Sporosarcina	1	77
3_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
3_SC	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	1	271
3_SC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	1	263
3_SC	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	1	253
3_SC	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacteraceae	Unclassified_Vicinamibacteraceae	1	80
3_SC	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Staphylococcus	1	41
3_SC	Proteobacteria	Gammaproteobacteria	Burkholderiales	Oxalobacteraceae	Noviherbaspirillum	1	1
3_SC	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Peptostreptococcaceae	Romboutsia	1	81
3_SC	Proteobacteria	Gammaproteobacteria	Burkholderiales	Methylophilaceae	Methylotenera	1	50
3_SC	Proteobacteria	Gammaproteobacteria	Enterobacterales	Yersiniaceae	Serratia	1	6
3_SH	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	7	271
3_SH	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	6	271
3_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	3	586
3_SH	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Staphylococcus	2	41
3_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Neisseriaceae	Neisseria	2	2
3_SH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	2	201
3_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	2	73
3_SH	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	Terrimonas	1	28
3_SH	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae	Paraglaciecola	1	8
3_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
3_SH	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	1	308
3_SH	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family XI	Finegoldia	1	5
3_SH	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	1	100

3_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	1	49
3_SL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	32	586
3_SL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Lysinibacillus	12	119
3_SL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Methylophilaceae	Methylotenera	10	50
3_SL	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	7	201
3_SL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	6	263
3_SL	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	5	100
3_SL	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	4	110
3_SL	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	4	49
3_SL	Actinobacteriota	Acidimicrobiia	Microtrichales	Microtrichaceae	Microtrichaceae_uncultured	4	22
3_SL	Firmicutes	Bacilli	Bacillales	Planococcaceae	Caryophanon	4	16
4_PC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	92	166
4_PC	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnospiraceae_uncultured	52	76
4_PC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	25	263
4_PC	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Herbinix	21	27
4_PC	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	15	263
4_PC	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Paenibacillus	14	105
4_PC	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia PCC-9350	11	70
4_PC	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Pir4 lineage	10	139
4_PC	Firmicutes	Negativicutes	Veillonellales-Selenomonadales	Sporomusaceae	Pelosinus	10	23
4_PC	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnoclostridium	10	25
4_PH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	38	586
4_PH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	31	253
4_PH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	28	201
4_PH	Firmicutes	Bacilli	Paenibacillales	Paenibacillaceae	Paenibacillus	16	105
4_PH	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	13	166
4_PH	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	10	522
4_PH	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Sva1033	Unclassified_Sva1033	10	33
4_PH	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Lachnoclostridium	10	25
4_PH	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	8	110

4_PH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Rhodoferax	8	73
4_PH	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia PCC-9350	8	70
4_PL	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	13	586
4_PL	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia PCC-9350	11	70
4_PL	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	9	100
4_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	9	73
4_PL	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Trichormus HINDAK 2001-4	9	12
4_PL	Proteobacteria	Gammaproteobacteria	Burkholderiales	SC-I-84	Unclassified_SC-I-84	7	66
4_PL	Acidobacteriota	Blastocatellia	Blastocatellales	Blastocatellaceae	JGI 0001001-H03	7	57
4_PL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	6	263
4_PL	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacterales_uncultured	Unclassified_Vicinamibacterales_uncultured	6	110
4_PL	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter	6	90
4_PL	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	Vicinamibacteraceae	Unclassified_Vicinamibacteraceae	6	80
4_PL	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	6	75
4_PL	Desulfobacterota	Desulfuromonadia	Desulfuromonadales	Desulfuromonadaceae	Desulfuromonas	6	72
4_PL	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Aphanizomenon NIES81	6	30
4_PT	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	37	522
4_PT	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	23	263
4_PT	Proteobacteria	Gammaproteobacteria	Burkholderiales	Rhodocyclaceae	Denitromonas	18	61
4_PT	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	15	522
4_PT	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	14	586
4_PT	Proteobacteria	Gammaproteobacteria	Granulosicoccales	Granulosicoccaceae	Granulosicoccus	12	522
4_PT	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	11	253
4_PT	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	11	100
4_PT	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacteraceae_uncultured	9	199
4_PT	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Blastopirellula	9	72
4_SC	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	296	308
4_SC	Actinobacteriota	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Cutibacterium	2	271
4_SC	Firmicutes	Bacilli	Staphylococcales	Staphylococcaceae	Staphylococcus	2	41
4_SC	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Family XI	Anaerococcus	2	33

4_SC	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	1	586
4_SC	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nodularia PCC-9350	1	70
4_SH	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	6	586
4_SH	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	5	201
4_SH	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 13	4	263
4_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Methylophilaceae	Methylotenera	4	50
4_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	4	49
4_SH	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingopyxis	4	10
4_SH	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinales_uncultured	Unclassified_Actinomarinales_uncultured	3	253
4_SH	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingorhabdus	3	100
4_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	TRA3-20	Unclassified_TRA3-20	3	73
4_SH	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas	3	66
4_SH	Proteobacteria	Gammaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	3	13
4_SH	Proteobacteria	Gammaproteobacteria	Enterobacterales	Alteromonadaceae	Paraglaciecola	3	8
4_SL	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nostoc PCC-73102	3	8
4_SL	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Nostocaceae	Nostoc PCC-73102	2	8
4_SL	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium sensu stricto 1	1	166
4_ST	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Dokdonella	2	61
4_ST	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Hoeflea	1	29
4_ST	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Sulfitobacter	1	21