Technical University of Munich
TUM School of Life Sciences

The role of the soil microbiome as a driver for soil structure development during post-mining reclamation

Miljenka Vuko

Complete reprint of the dissertation approved by the TUM School of Life Sciences of the Technical University of Munich for the award of the

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Chair: Prof. Dr. Mohsen Zare

Examiners:

1. Prof. Dr. Michael Schloter
2. Prof. Dr. Dr. h.c. Ingrid Kögel-Knabner

The dissertation was submitted to the Technical University of Munich on 04.07.2023 and accepted by the TUM School of Life Sciences on 16.10.2023.

Miljenka Vuko:

The role of the soil microbiome as a driver for soil structure development during post-mining reclamation

Commander Sisko: “…the game wouldn’t be worth playing if we knew what was going to happen.

Prophet: “You value your ignorance of what is to come?!”

Commander Sisko: “That may be the most important thing to understand about humans: it is the unknown that defines our existence. We are constantly searching, not just for answers to our questions, but for new questions. We are explorers. We explore our lives day by day, and we explore the galaxy trying to expand the boundaries of our knowledge. And that is why I am here. Not to conquer you with weapons, or with ideas. But to co-exist…and to learn.”

“Emissary”. *Star Trek: Deep Space Nine,* S01E01/02, 1993

Table of Contents

List of figures

List of tables

List of abbreviations

ABC	ATP-binding cassette
ANOVA	analysis of variance
bp	base pairs
BCE	Before the Common Era
BSA	bovine serum albumin
CEC	cation exchange capacity
C_{mic}	microbial carbon
DC	dissolved carbon
DOC	dissolved organic carbon
DN	dissolved nitrogen
DON	dissolved organic nitrogen
EPS	extracellular polymeric substance
exo-PS	exopolysaccharide
HMM	Hidden Markov Models
ITS	internal transcribed spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
LPS	lipopolysaccharide
NCBI	National Center for Biotechnology Information
NMDS	non-metric multidimensional scaling
N_{mic}	microbial nitrogen
ORF	open reading frame
PCA	principal component analysis

qPCR	quantitative polymerase chain reaction
RA	reclamation age
SOC	soil organic carbon
OPX	outer membrane auxiliary proteins
PCP	polysaccharide co-polymerase proteins
TN	total nitrogen
TPR	tetratricopeptide repeat proteins
VAM	vesicular arbuscular mycorrhiza

Publications and contributions

Vuko, M., Cania, B., Vogel, C., Kublik, S., Schloter, M., and Schulz, S. (2020) Shifts in reclamation management strategies shape the role of exopolysaccharide and lipopolysaccharide-producing bacteria during soil formation. *Microbial Biotechnology* **13**: 584–598. (**Publication I, first author, published**)

Pihlap, E., **Vuko, M**., Lucas, M., Steffens, M., Schloter, M., Vetterlein, D., Endenich, M., Kögel-Knabner, I. (2019) Initial soil formation in an agriculturally reclaimed open-cast mining area - the role of management and loess parent material. *Soil Tillage Research* **191**: 224-237. (**Publication II, second author, published**)

Summary

Soil is an endangered non-renewable resource that supports all terrestrial life on Earth and can take thousands of years to develop. Hence, losing huge amounts of fertile soil due to open-cast coal mining presents a big global issue since such mines usually occupy large areas. A major concern during the reclamation of mined soils is erosion. This threat can be counteracted by developing stable aggregates that can resist physical and chemical stress. The formation of stable aggregates can be induced by extracellular polymeric substances (EPSs) produced by bacteria. EPSs consist mainly of polysaccharides, among which exopolysaccharides (exo-PSs) and lipopolysaccharides (LPSs) show adhesive properties that can glue soil particles together. Little is known about these compounds during soil formation, particularly in a reclamation setting where they could enhance and accelerate soil structure stabilization. Therefore, this thesis aimed to follow soil formation during 24 years of reclamation in a chronosequence of post-mining soils, with a special emphasis on soil structure and bacteria that can improve aggregation using metagenomics. To this end, an analysis of genes involved in the biosynthesis and export of exo-PSs and LPSs was performed in combination with direct measurements of EPSs from the soils.

Results showed that at the beginning of reclamation soil structure development was dominated by calcium carbonate, making it particularly sensitive to increased moisture and prone to erosion. The low abundant microbial community consisted primarily of stochastically distributed bacteria capable of withstanding stressful environmental conditions, such as *Sphingomonas* and *Ramlibacter.* When the first plant (leguminous *Medicago sativa* (alfalfa)) was introduced, organic matter content, abundance of microbes, as well as the concentration of extracellular polymeric substances (EPSs), substantially increased. This coincided with an increase in the relative abundance of exo-PS and LPS genes assigned to Rhizobiales, bacterial order known to use these compounds to initiate symbiosis with legumes. Once the reclamation practice changed to agricultural management, there was a further increase in organic matter and EPS concentration, but the abundance of bacteria and fungi in the topsoil decreased. At the same time, there was a peak in the relative abundances of *wza* (polysaccharide export protein Wza), *kpsE* (capsular polysaccharide export permease KpsE), *wcaF* (colanic acid biosynthesis acetyltransferase WcaF), *sacB* (levansucrase SacB), as well as *wzt* (LPS ATP-binding protein Wzt) and *lptF* and *G* (LptBFGC LPS permease LptF and G) genes, showcasing polysaccharide production as a defense strategy against disturbances such as deep tillage.

After 24 years of reclamation, bacterial abundance and diversity further increased, exhibiting a continuous shift to a community of low-abundant species typically found in agricultural soils where they perform different ecosystem services. Of note, members from the *Nitrospiraceae* family, which are important nitrifiers, increased in relative abundance and replaced *Sphingomonadaceae* as the most abundant exo-PS producer. Even though EPS concentrations increased likely as a result of accumulation over time, the relative abundances of most of the analyzed exo- and lipopolysaccharide genes decreased, pointing to a lower need for polysaccharide production and bacterial adaptation to regular agricultural practices. Importantly, organic C and N accumulation was not sustained after 12 and 24 years of reclamation, confirming bacterial EPSs to be an important source of organic matter stabilizing soil structure in the long term.

Zusammenfassung

Böden sind eine gefährdete, nicht erneuerbare Ressource. Sie sind Grundlage allen terrestrischen Lebens, während ihre Entwicklung Tausende von Jahren dauern kann. Der Verlust großer Mengen fruchtbarer Böden durch den Braunkohletagebau stellt somit ein großes globales Problem dar, zumal solche Tagebaue in der Regel große Flächen einnehmen. Ein Hauptproblem bei der Rekultivierung dieser Flächen ist dabei die Erosion. Dieser Bedrohung kann durch die Entwicklung stabiler Bodenaggregate entgegengewirkt werden, die physikalischen und chemischen Belastungen widerstehen können. Die Bildung stabiler Aggregate kann durch extrazelluläre polymere Substanzen (EPS) angeregt werden, die von Bakterien produziert werden. EPS bestehen hauptsächlich aus Polysacchariden, von denen Exopolysaccharide (Exo-PS) und Lipopolysaccharide (LPS) adhäsive Eigenschaften aufweisen, durch welche Bodenpartikel verklebt werden können. Es ist nur wenig über diese Verbindungen im Verlaufe der Bodenbildung bekannt, insbesondere bei der Rekultivierung, wo sie die Stabilisierung der Bodenstruktur verbessern und beschleunigen könnten. Ziel meiner Dissertation war es daher, die Bodenbildung während einer 24-jährigen Rekultivierungsphase anhand einer Chronosequenz von Tagebauböden unter Verwendung von Metagenomik zu verfolgen. Ein besonderes Augenmerk lag dabei auf der Bodenstruktur und den Bakterien, die die Aggregation verbessern können. Zu diesem Zweck wurde eine Analyse derjenigen Gene, die an der Biosynthese und dem Export von Exo-PS und LPS beteiligt sind, in Kombination mit direkten Messungen von EPS aus den Böden durchgeführt.

Die Ergebnisse zeigten, dass zu Beginn der Rekultivierung die Entwicklung der Bodenstruktur von Kalziumkarbonat dominiert wurde, was den Boden besonders empfindlich gegenüber erhöhter Feuchtigkeit und anfällig für Erosion machte. Die spärliche mikrobielle Gemeinschaft bestand in erster Linie aus stochastisch verteilten Bakterien, die in der Lage sind, widrigen Umweltbedingungen zu widerstehen, wie z. B. *Sphingomonas* und *Ramlibacter*. Nachdem mit der Kultivierung der ersten Pflanze – die Leguminose *Medicago sativa* (alfalfa) – begonnen wurde, nahmen der Gehalt an organischer Substanz, die Fülle an Mikroben sowie die Konzentration extrazellulärer polymerer Substanzen (EPS) erheblich zu. Dies fiel mit einer Zunahme der relativen Häufigkeit von Exo-PS- und LPS-Genen zusammen, die den Rhizobiales zugeordnet werden. Von dieser bakteriellen Ordnung ist bekannt, dass sie diese Verbindungen nutzt, um eine Symbiose mit Leguminosen einzugehen. Nach der Umstellung der

Rekultivierungspraxis auf eine landwirtschaftliche Bewirtschaftung nahmen die organische Substanz und die EPS-Konzentration weiter zu, während die Abundanz von Bakterien und Pilzen im Oberboden abnahm. Gleichzeitig stieg die relative Häufigkeit von *wza* (Polysaccharid-Exportprotein Wza), *kpsE* (Kapselpolysaccharid-Exportpermease KpsE), *wcaF* (Colansäure-Biosynthese-Acetyltransferase WcaF), *sacB* (Levansucrase SacB) sowie *wzt* (LPS ATP-bindendes Protein Wzt) und *lptF* und *G* (LptBFGC LPS-Permease LptF und G), was zeigt, dass die Polysaccharidproduktion eine Verteidigungsstrategie gegen Störungen wie tiefe Bodenbearbeitung darstellt.

Nach 24 Jahren der Rekultivierung nahmen Häufigkeit und Vielfalt der Bakterien weiter zu, was eine kontinuierliche Verschiebung hin zu einer Gemeinschaft von Arten mit geringer Häufigkeit zeigt. Viele dieser Arten sind typisch für landwirtschaftliche Böden und erbringen dort verschiedene Ökosystemleistungen. Auffällig war dabei eine relative Zunahme von Mitgliedern der Familie *Nitrospiraceae*, die wichtige Nitrifikanten beinhaltet, und die damit die *Sphingomonadaceae* als häufigste Exo-PS-Produzenten ablösten. Obwohl die EPS-Konzentrationen wahrscheinlich infolge der Akkumulation im Laufe der Zeit zunahmen, nahmen die relativen Häufigkeiten der meisten analysierten Exo- und Lipopolysaccharid-Gene ab. Dies deutet auf einen geringeren Bedarf an Polysaccharid-Produktion und eine bakterielle Anpassung an regelmäßige landwirtschaftliche Praktiken hin. Bemerkenswerterweise wurde die Anreicherung von organischem Kohlenstoff und Stickstoff nach 12 und 24 Jahren der Rekultivierung nicht mehr aufrechterhalten. Dadurch zeigt sich, dass bakterielle EPS eine wichtige Quelle organischer Masse sind, die die Bodenstruktur langfristig stabilisiert.

1 Introduction

1.1 Soil formation

In a recent book chapter, Hartemink (2016) has listed 81 selected definitions of the soil found across two centuries of research material, demonstrating that it depends on the perspective and context in which it is contemplated. Broadly, it can be said that the soil is a multiphase system consisting of mineral material, organic matter at various stages of decay, water, gases, and plant roots. It also provides a medium for an astonishing diversity of organisms to live in (Bardgett, 2005a). Throughout history, however, the soil has been considered in many different ways: as mere inert support, a tool for food production, a source of raw materials and minerals, an element of tax assessment, a base for physical, chemical, and biological processes, an organized natural body, etc. (Buol, 1973; Dazzi and Lo Papa, 2022). For much of history, humans were mostly concerned with the agricultural aspect of soil, starting from the Neolithic Era (10,000-4,500 BCE) when human groups started transitioning from hunter-gatherer societies to sedentary agricultural civilizations. These civilizations often worshipped deities they connected to soil production, revealing the intimate relationship humans have always had with soil. In fact, ancient civilizations have risen and fallen based on the fertility and depletion or erosion of soils (e.g. the fall of Babylon and Nineveh after the highly productive lands in Mesopotamia degraded and lost rich topsoil leaving a barren saline landscape) (McNeill and Winiwarter, 2004; Churchman, 2010). Knowledge about soil has greatly evolved from the first soil classifications found in ancient Chinese, Indian, Middle Eastern, Roman, and Greek texts and literature dating back up to 2000 BCE, all the way to the dawn of modern soil science at the beginning of the 19th century (McNeill and Winiwarter, 2004).

In 1883, Vasilli V. Dokuchaev, the founder of pedology (discipline of soil science focused on soil formation and evolution), established a fundamental principle by which the soil was seen as an independent natural-historical body resulting from the influence of subsoils, flora and fauna, geological age, and relief of the locality (Churchman, 2010). Dokuchaev's work was the basis upon which Jenny (1941) constituted the five factors of soil formation: climate, parent material, organisms, topography, and time (Brevik and Hartemink, 2010). These factors are highly interactive and can be practically infinitely variable during the process of soil formation (Bardgett, 2005a). On the one hand, according to the traditional soil-formation theory (i.e. developmental theory), a 'young' soil forms or develops progressively under these factors and stops once the 'mature' soil is in equilibrium with

prevailing environmental conditions. On the other hand, according to a more recent, evolutionary theory of pedogenesis, inconstancy of environmental conditions that change dynamically and non-linearly is key to soil evolution (Huggett, 1998). Turnover of materials (additions, losses, transfers, transformations) in such dynamic systems can be multidirectional and lead to a variety of evolutionary pathways, indicating that it is improbable for a developmental soil sequence to ever run its full course (Huggett, 1998).

Soil is a dynamic system constantly being shaped by soil-forming processes (Jenny, 1941). Its formation starts with the alteration of geologic material, which gradually develops into a soil profile that consists of physically, chemically, and biologically distinct horizontal layers. This process can also be restarted by the removal or covering of the soil profile, either by natural or artificial events, such as erosion, flood deposits, or volcanic eruptions (Harrison and Strahm, 2008). The organization of soil into distinct depth profiles is the main attribute that sets it apart from the geological parent material that it originates from. Horizons that are very close to the parent material are usually only slightly changed by soil-forming processes. Subsequent horizons are gradually more developed towards the soil surface, which is often a distinct layer of organic litter derived from plants and animal remains (Figure 1) (Harrison and Strahm, 2008). A well-developed soil is also characterized by stable structure, pH, water-holding capacity, water infiltration rate, low erodibility, high organic matter content, nutrient content and cycling, and biological diversity. These properties develop and change at different rates and under different driving forces. Some can change rapidly under external environmental changes (e.g. pH), while some change very slowly or not at all (e.g. soil texture) (Chadwick and Chorover, 2001).

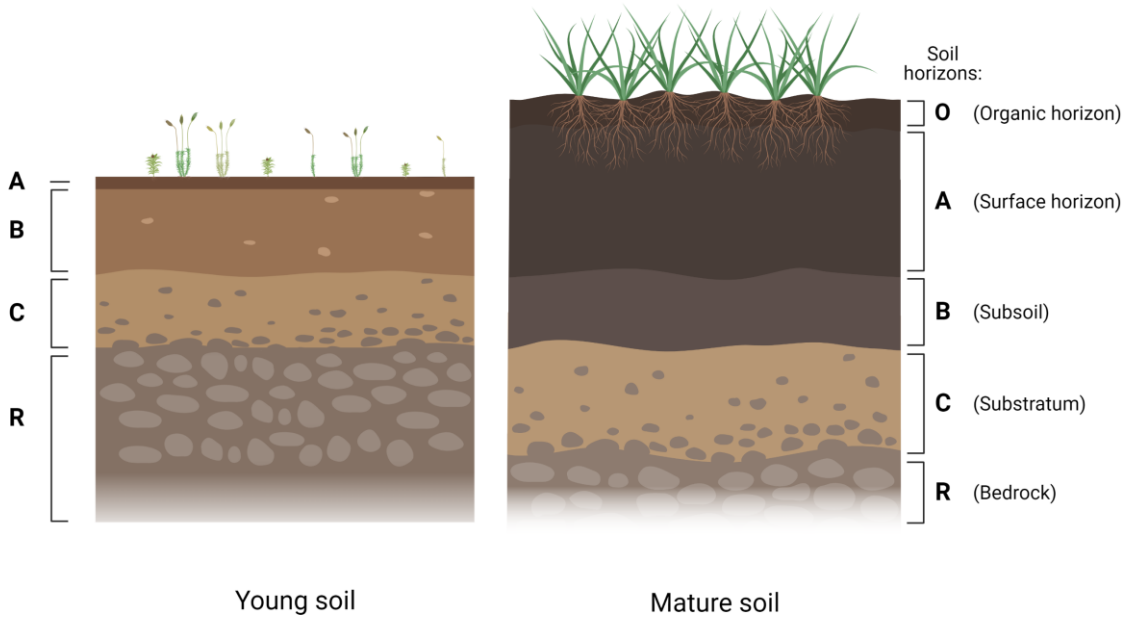


Figure 1. Soil profiles of a young and mature soil
In the young soil, horizons are less differentiated, devoid of organic matter, and covered by pioneering species, such as mosses. In the mature soil, horizons are well differentiated, with a distinct layer of organic litter derived from plants and animals. Created with BioRender.com.

Since soil formation is a process that can last for thousands of years, knowledge about it was gathered inferentially through estimates, such as the differentiation of soil horizons (Stevens and Walker, 1970). Another approach to studying soil genesis is to investigate soils that are presumably under the equal influence of four out of five soil-forming factors, but differ in time since they started developing (i.e. soils of different developmental ages). This approach is possible with the use of soil chronosequences.

1.2 Soil structure

One of the key features of developed soils is soil structure. Soil structure influences many important soil functions, such as soil water movement, erosion, nutrient cycling, crop yield, groundwater pollution, carbon sequestration, and CO_2 emissions (Bronick and Lal, 2005). Soil structure refers to the size, shape, and spatial arrangement of solids and voids across different spatial scales and their capacity to hold and transfer fluids, air, and organic and inorganic material (Lal, 1991; Pihlap et al., 2019). Consequently, soil structure can be viewed from the solid phase, as the assembly of soil particles into aggregates, or from the void phase, as the connectivity of different types of pores (Chenu and Cosentino, 2011).

Aggregates are considered to be the basic units or building blocks of soil structure, and their stability is commonly used as an indicator of soil structure (Lynch and Bragg, 1985; Six et al., 2000). As defined by Martin et al. (1955), an aggregate is a ‘group of primary soil particles that adhere to each other more strongly than to other surrounding particles’ (Nimmo, 2005). There have been four major milestones in the research history of soil aggregation. First, Emerson (1959) developed a model of a "soil crumb", which consisted of oriented clay and quartz particles stabilized by soil organic matter. Second, Edwards and Bremner (1967) proposed the formation of microaggregates by the bonding of clay particles, polyvalent metals, and organic matter complexes. Third, Tisdall and Oades (1982) established the generally accepted hierarchical theory of aggregate formation in which different binding agents (temporary vs persistent) act at different hierarchical stages of aggregation. They divided aggregates into two major classes: microaggregates (< 250 µm), i.e. particles bound by persistent binding agents, and macroaggregates (> 250 µm), i.e. microaggregates bound by temporary and transient binding agents. Microaggregates are therefore more stable and less impacted by physical stresses than macroaggregates (Six et al., 2004). Fourth, Oades (1993) later modified the hierarchical theory to add that aggregation does not necessarily follow a sequential order where particles are bound to microaggregates, which are then bound to macroaggregates, but that microaggregates can also be formed inside macroaggregates. Aggregated soil structure ultimately results in a spatially heterogeneous and complex three-dimensional architecture of soil that is built from a variety of mineral, organic, and biotic materials (Ritschel and Totsche, 2019).

Major factors influencing aggregation, and therefore soil structure development, are (1) soil fauna, (2) soil microorganisms, (3) roots, (4) inorganic binding agents, and (5) environmental conditions (Six et al., 2004). These can be further dissected and generally divided into abiotic and biotic. Abiotic factors influencing aggregation include soil mineralogy and texture, pH, water, climate, cation exchange capacity (CEC), organic (e.g. organic carbon, N, P) and inorganic (e.g. clay, carbonates, cations) binding agents (Bronick and Lal, 2005; Yang et al., 2022). Aggregation and aggregate stabilization generally increase with soil organic carbon (SOC), clay surface area, and CEC, but different factors dominate in different soil types (Bronick and Lal, 2005). SOC is often thought to have a dominant influence on soil aggregate stability by creating regions of heterogeneity leading to “hot spots” of aggregation (Bronick and Lal, 2005; Jensen et al., 2019). In soils that are low in SOC and clay content, aggregation is usually dominated by cations, e.g. Al^{3+} and Fe^{3+} oxides in Oxisols (Bronick and Lal, 2005; Zotarelli et al., 2005). Soil texture plays a

significant role in aggregation; in coarse-textured soils, SOC has a great influence on structure, while in soils with a high clay content the type of clay is an important factor (Kay, 1997). In calcareous soils, calcium carbonate is usually the main stabilizing agent (Mataix-Solera et al., 2011), where it has a cementing role (Totsche et al., 2018).

Climatic conditions that likely have the strongest effect on soil aggregation are freeze-thaw and dry-wet cycles. Frost acts as a physical force that can mechanically displace, mix, stress, and geometrically reorient aggregates (Totsche et al., 2018) as it causes the expansion of ice crystals in pores, which break particle-to-particle bonds (Six et al., 2004). These disruptive effects of frost are increased with increasing water content in the soil (Six et al., 2004), while at low water contents, ice crystals are formed rapidly before they can create significant disruption in the pores (Bullock et al., 1988). The destruction of aggregates by freezing is somewhat reversed by the subsequent thawing and drying processes that result in cohesion and reformation of bonds between adjacent microaggregates (Bullock et al., 1988; Six et al., 2004). Soils are in a continuous cycle of wetting and drying through the effect of rain, snow, and wind, and these processes result in decreases and increases in aggregation (Six et al., 2004). Repeated wet cycles with sufficient drying times have been shown to increase aggregate stabilization as the disintegration of aggregates leads to a better settlement and packaging of particles, which results in a greater bonding with the next drying event (Kemper and Rosenau, 1984; Denef et al., 2001). Generally, aggregate stabilization of soils with higher organic matter content benefit from wet-dry cycles due to the hydrophobic characteristics of the organic matter, while soils with a lower organic matter content get rewetted faster and are under a stronger disruptive influence of wetting than the stabilizing effect of drying (Six et al., 2004).

1.3 Biotic factors influencing soil structure development

1.3.1 Plants and animals

Biotic factors influencing aggregation include different soil-dwelling animals, plant roots, and microorganisms. The best-known animals that influence soil characteristics are earthworms, whose importance in the formation of soil structure has already been recognized by Charles Darwin in his book from 1881 (Darwin, 1881), where he describes fine soil particles (“animal mold”) that have passed many times through the intestinal canals of earthworms. By ingesting organic matter and inorganic soil material, earthworms pass the mixture through their gut where it is encrusted with mucus and excrete it in the form of biogenic organo-mineral aggregates, i.e. casts (Six et al., 2004; Lemtiri et al., 2014; Vidal

et al., 2019). In general, higher structural stability and favorable conditions for root growth have been found in earthworm-treated soils in comparison to control soils (Bossuyt et al., 2004; Barré et al., 2009). Additionally, earthworms can improve soil structure by burrowing and thus mechanically perturbating soil particles and leaving organic mucilage on the surrounding soil, which can form a stable structure (Six et al., 2004). Of note, among other invertebrates known to influence soil structure are termites, especially in tropical soils (Jouquet et al., 2016). Some termite species ingest organic matter and build galleries and nests using fecal matter as the gluing agent, while fungus-growing species use saliva as the binding agent, leading to an improved soil structure (Garnier-Sillam and Harry, 1995; Jouquet et al., 2016).

The roles of plants in the development of soil structure include root penetration, modifying soil water regime, root exudates, decomposition of dead roots, and root entanglement (Six et al., 2004). Root growth and penetration happen in axial and radial directions, i.e. elongation and diameter increase, respectively (Kolb et al., 2012). This growth exerts pressure on the surrounding soil and causes compaction and an increased bulk density in the rhizosphere soil (Young, 1998). It is also often associated with the fragmentation of soil because it induces soil loosening and aggregate formation (Angers and Caron, 1998). Plants with coarser roots create pores that can mainly be characterized as macropores larger than 30 µm, by which water and other fluids flow and thus avoid the soil matrix (Gibbs and Reid, 1988; Angers and Caron, 1998). On the other hand, plants with fine and dense root systems induce a more heterogeneous pore size distribution with numerous pores of <15 µm in diameter (Bodner et al., 2014). Pore networks created by roots are especially important in undisturbed soils, as tillage usually breaks the continuity of long pores (Angers and Caron, 1998). Plant roots can also stimulate soil structure formation through different indirect effects. For example, they can modify the soil water regime by up-taking water and causing a localized drying and shrinkage of the soil (Kay, 1990), or increase soil stability and resilience to mechanical stresses (Hallett et al., 2009). In addition, roots are a major source of organic carbon through root exudates and mucilage, which drive the activity of soil microorganisms and fauna, but are also known to directly affect soil structure formation (Oleghe et al., 2017).

1.3.2 Microorganisms

The stabilization of aggregates has been largely attributed to biological processes mainly induced by microorganisms (Denef and Six, 2005). In fact, soil structure is often referred

to as soil “architecture” (Vogel et al., 2022) and since microorganisms are major factors in its formation, stabilization, and destruction (Chenu and Cosentino, 2011), they can be referred to as soil architects. However, the contribution of microbial communities depends on the microbial species, available substrates, and soil management (Beare et al., 1994).

1.3.2.1 Fungi

Fungi are known to promote soil structure by anchoring soil particles with hyphae and thus creating macroaggregates (> 250 µm), while bacteria are particularly important for the stabilization of microaggregates (< 250 µm) (Six et al., 2004; Costa et al., 2018). In a global meta-analysis of 279 soil biota species from 183 studies, Lehmann et al. (2017) found bacteria and fungi to be more important for soil aggregation compared to soil animals, with fungi strongly affecting macro-aggregation and bacteria affecting both micro- and macro-aggregation. The degree of aggregation by fungal hyphae depends on the strength of adhesion to soil particles, the tensile strength of the hyphae, and on the architecture of the mycelial network in the soil (Chenu and Cosentino, 2011). Soil particles can be bound and enmeshed into aggregates by vesicular arbuscular mycorrhiza (VAM) and saprophytic fungi. VAM form hyphal networks in association with roots and thus act together in the physical bonding of soil particles, while saprophytic fungi colonize dead roots and grow hyphal networks along them, using the carbon and nutrients from decaying roots (Tisdall and Oades, 1979; Degens, 1997). The location of fungal hyphae within the soil matrix is also of importance; the effect they exert on the formation of macroaggregates is likely to be greater when they grow in large pores between microaggregates (Degens, 1997). In an experiment where different fungal traits that affect soil macro-aggregation were compared between 31 fungal saprophytic strains, hyphal density was found to be the highest contributor (Lehmann et al., 2020). It should be noted that saprophytic fungi can also have a negative effect on soil aggregation via enzymes they release to degrade organic matter, e.g. leucine aminopeptidase (Baldrian et al., 2011; Lehmann et al., 2020).

1.3.2.2 Bacteria

Bacteria influence soil aggregation mainly by exuding biopolymers that bind soil particles together (Chenu, 1995), however, the data on bacteria responsible for soil aggregation is scarce and mostly performed on isolated strains and in laboratory conditions (Caesar-TonThat et al., 2007; Caesar-Tonthat et al., 2014; Costa et al., 2018). This introduces a bias towards cultivable species and ignores the vast majority of species still considered uncultivable (Vartoukian et al., 2010; Chaudhary et al., 2019). Hence, cultivation-

independent and community-based metagenomic studies are necessary to access the microbial genetic pool not reachable by classical cultivation methods in order to explore the compounds produced by bacteria in natural environments (Costa et al., 2020).

Bacteria can bind and stabilize soil aggregates in two ways: (i) via releasing organic substances with ionic charges and a slimy texture holding mineral particles together, and (ii) via adhering to mineral particles with their outer cell membrane proteins, thus cementing them into aggregates (Chenu, 1995). In general, a major factor that facilitates microbial adhesion to soil particles is the production of extracellular polymeric substances (EPS) (Chenu, 1995), which are complex three-dimensional matrices composed of polysaccharides, lipids, proteins or peptides, and nucleic acids (Costa et al., 2018). EPS are ubiquitous in the environment and their properties, as well as functions, depend on the environmental conditions and vary greatly. There are thus many examples of the functions bacterial EPS perform in ecology, but also in industrial use.

One of the most studied examples is the attachment of nitrogen-fixing bacteria to plant roots. Rhizobia are soil bacteria that form symbiotic relationships with leguminous plants, where rhizobia fix nitrogen for the plant and the plant provides bacteria with carbohydrates as a source of carbon and energy (Skorupska et al., 2006). Production of EPS represents an advantageous strategy for rhizobia to survive nutrient starvation and desiccation while living in the bulk soil, and it aids in the complex interaction with the host plant, increasing the inoculum of bacteria on plant roots (Abdian and Zorreguieta, 2016). Other types of bacteria also use EPS for creating interactions with plants: some beneficial, e.g. certain *Pseudomonas* strains that can improve plant growth and protect against pathogens (Lugtenberg et al., 2001; Espinosa-Urgel et al., 2002), and some pathogenic, e.g. *Xanthomonas campestris* pv. *campestris* or *Pantoea stewartii* subsp. *stewartii*, which causes black rot on cruciferous plants (Dow et al., 2003) and Stewart's wilt disease in maize (von Bodman et al., 1998), respectively. Production of EPS is used by some bacteria as means of protection against being ingested by predators, e.g. the human pathogen *Yersinia pseudotuberculosis*, which forms biofilms on the surface of the nematode worm, *C. elegans* (Atkinson et al., 2011), or *Pseudomonas aeruginosa*, which reduces the motility of *C. elegans* by entrapping it within the matrix (Chan et al., 2021). Since EPS also serve as biosorbing mediums by accumulating nutrients from the environment, they also play a crucial role in the adsorption of heavy metals, immobilizing them within the matrix (Pal and

Paul, 2008). It is this property of EPS that makes them vital in the bioremediation of contaminated sites, e.g. wastewater systems (Pal and Paul, 2008).

Secreting EPSs promotes the initial colonization of abiotic surfaces by planktonic cells, enabling the formation of complex network structures such as biofilms, microbial mats, and soil crusts (Velmourougane et al., 2017). EPSs can be secreted by bacteria, fungi, cyanobacteria, yeasts, microalgae, and protists (Costa et al., 2018) and can protect them against environmental stresses (Flemming et al., 2016; Costa et al., 2018) by creating favorable and sheltered niches (Velmourougane et al., 2017). EPS matrix in these niches can be metaphorically called the “house of the biofilm cells” (Flemming et al., 2007). What is important in the soil environment is the ability of EPS to entrap nutrients and soil particles (Costerton et al., 1995; Costa et al., 2018), thus contributing to overall soil aggregation and structure stabilization. Major fractions of EPS that determine its adhesive properties are exopolysaccharides and lipopolysaccharides (Velmourougane et al., 2017).

1.3.2.2.1 Bacterial exopolysaccharides (exo-PS)

Exopolysaccharides (exo-PSs) excreted by bacteria are long chains of high-molecular-weight sugars with great variation in composition, structure, and function (Sutherland, 2001), and a molecular mass of 0.5 x 10^6 to 2 x 10^6 daltons (Flemming and Wingender, 2010). They are high in carbon content, thus their production is energetically expensive and requires a sufficient supply of carbon (Flemming and Wingender, 2010; Costa et al., 2018). Some exopolysaccharides are homopolysaccharides, such as cellulose produced by e.g. *Acetobacter xylinum*, *Agrobacterium tumefaciens*, or *Rhizobium* spp. (Ross et al., 1991), but most are heteropolysaccharides that consist of a mixture of sugar residues (Flemming and Wingender, 2010). Despite their large diversity, the biosynthesis and export of exo-PSs are grouped into four distinct mechanisms: (i) Wzx and Wzy-dependent, (ii) ATP-binding cassette (ABC) transporter-dependent, (iii) synthase-dependent pathway (Figure 2), and (iv) extracellular synthesis using a single sucrase protein (Whitney and Howell, 2013; Schmid et al., 2015).

The majority of cell-surface and outer polysaccharides (e.g. alginate, capsular polysaccharides, O-antigens) in both Gram-positive and negative bacteria, are synthesized via the Wzx/Wzy pathway (Islam and Lam, 2014). The main components in this pathway are the Wzx flippase and Wzy polymerase (Islam and Lam, 2014). Subsequent components of the pathway differ depending on the bacterial species, but all involve polysaccharide co-polymerase (PCP) proteins (Morona et al., 2009). In the first step of the

Wzx/Wzy-dependent pathway, the individual repeating sugar units are assembled by several glycosyltransferases at the inner leaflet of the inner membrane and translocated across the cytoplasmic membrane by a Wzx protein, also called flippase (Schmid et al., 2015). Then, they are polymerized at the periplasmic space by the Wzy protein with the help of auxiliary PCP proteins, such as Wzc in Gram-negative bacteria (Cuthbertson et al., 2009; Schmid et al., 2015). Wzc plays a critical role in the final exportation of some exopolysaccharides by interacting with the outer membrane auxiliary proteins encoded by the *wza* gene. Wza and Wzc span the cell envelope and are required in the final stage of polysaccharide export (Cuthbertson et al., 2009). Polysaccharides assembled by the Wzx/Wzy pathway are heteropolymers commonly composed of up to five types of sugars within their chemical structures, e.g. xanthan, sphingan, succinoglycan, and colanic acid (Schmid et al., 2015).

The ABC transporter-dependent pathway is mainly used for the biosynthesis of capsular polysaccharides, which are linked to the cell surface (Whitney and Howell, 2013; Schmid et al., 2015). Broadly, ABC transporters import and export molecules across the inner membrane of Gram-negative bacteria or the single membrane of Gram-positive bacteria (Zeng and Charkowski, 2021). Unlike the Wzx/Wzy pathway, in the ABC-dependent pathway, the entire polysaccharide is assembled on the inner leaflet of the inner membrane and then transported across by an ABC transporter (Whitney and Howell, 2013). The common feature of all ABC transporters is that they have at least two cytoplasmic nucleotide-binding domains, which have an ATPase activity that provides the energy for translocation, and at least two transmembrane domains, which are embedded in the membrane bilayer (Davidson et al., 2008; Maruyama et al., 2015). The export of the polymer through the periplasm and outer membrane of Gram-negative bacteria also requires the PCP protein, KpsE, and the outer membrane auxiliary protein, KpsD, which are analogous to Wza and Wzc proteins of the Wzx/Wzy pathway (Cuthbertson et al., 2010; Pereira et al., 2013).

In the synthase-dependent pathway, complete polymer strands are secreted across the membrane and the cell wall, independently of flippase proteins (Schmid et al., 2015). Here, both the polymerization and translocation processes are performed by a single synthase protein, which can be a subunit of a cross-membrane multiprotein complex (Rehm, 2010). In Gram-negative bacteria, this complex involves periplasmic tetratricopeptide proteins and an outer-membrane β-barrel porin protein, which exports it across the outer membrane

(Whitney and Howell, 2013). Examples of exopolysaccharides synthesized via the synthase-dependent pathway include cellulose, curdlan, alginate, and hyaluronic acid (Schmid et al., 2015).

In the extracellular synthesis of exo-PSs, the glycosyltransferases that biosynthesize the polysaccharides are secreted and covalently linked to the cell surface (Schmid et al., 2015). Levan is one example of extracellularly synthesized exo-PSs, and it is produced by some bacterial species that excrete levansucrase when extracellular sucrose is available (Limoli Dominique et al., 2015). Other examples include dextran, produced by dextransucrases, and inulin, produced by inulinosucrases (Schmid et al., 2015).

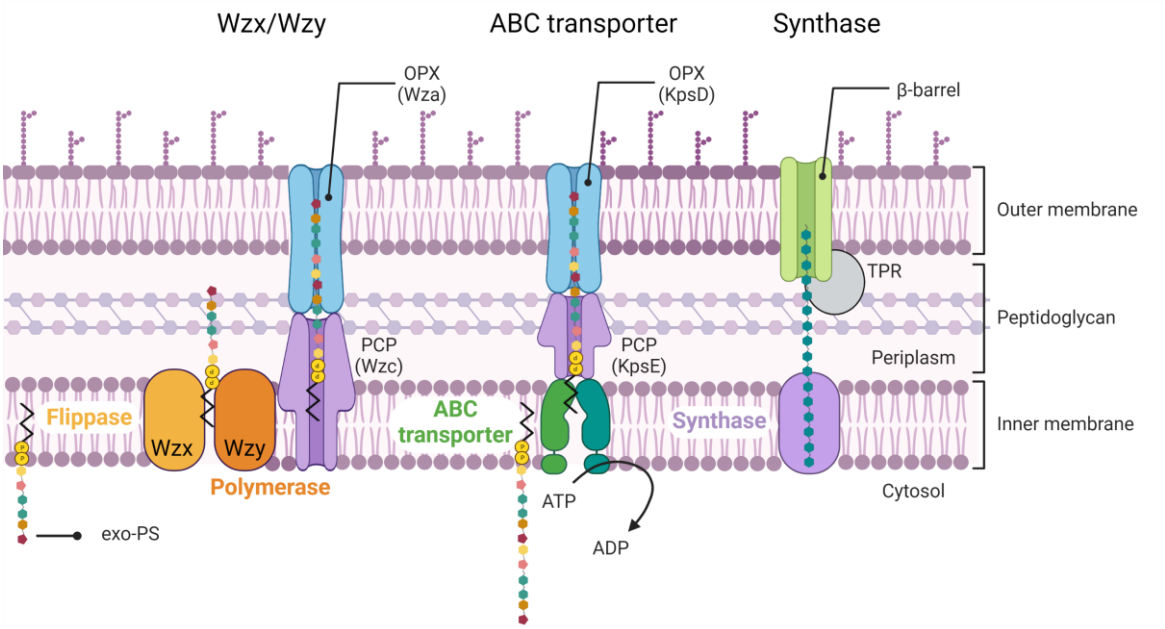


Figure 2. Main mechanisms of exopolysaccharide export through the membrane of Gram-negative bacteria
OPX: outer membrane auxiliary proteins, PCP: polysaccharide co-polymerase proteins, TPR: tetratricopeptide repeat proteins. Adapted from Cuthbertson et al., 2009; Whitney and Howell, 2013; Willis and Whitfield, 2013; Islam and Lam, 2014; Pereira et al., 2015; Schmid et al., 2015. Created with BioRender.com.

1.3.2.2.2 Bacterial lipopolysaccharides (LPS)

Lipopolysaccharides (LPSs) are complex glycolipids attached to the outer membrane of most Gram-negative bacteria and they make up the outer leaflet of their outer membranes (Jacques, 1996; Islam and Lam, 2014). They provide cell integrity and are in many cases the component of the bacterial cell that interacts with surfaces in proximity (Esteban et al.,

2014; Whitfield and Trent, 2014). LPSs are built of two or three components: lipid A, a core oligosaccharide, and, in some bacteria, a long-chain O-antigenic polysaccharide (Figure 3) (Putker et al., 2015). Lipid A serves as an outer membrane anchor, the core oligosaccharide contains important phosphate motifs, and the O-antigenic polysaccharides are heteropolymer sugar repeat units of different compositions and lengths (Islam and Lam, 2014). The structure of lipid A is highly conserved in different bacteria, while core oligosaccharides show little variation, and their conservation is presumably driven by their crucial roles in the integrity of the outer membrane (Whitfield and Trent, 2014). Conversely, O-antigens exhibit considerable diversity, also within the same bacterial species, likely resulting from selective environmental pressures (Wang and Quinn, 2010; Whitfield and Trent, 2014).

Lipid A is synthesized on the inner leaflet of the inner membrane, via nine enzymes in a highly conserved Raetz pathway (Whitfield and Trent, 2014). There are, however, exceptions, such as the *Sphingomonas* species where the Raetz pathway is absent and the outer membrane contains glycosphingolipids instead of LPSs (Kawasaki et al., 1994). After it is synthesized, lipid A is translocated across the membrane by ABC transporters. Core oligosaccharides and O-antigens can be synthesized by the Wzx/Wzy-dependent, ABC transporter-dependent, or synthase-dependent pathways (Whitfield and Trent, 2014). Depending on the pathway, the components are ligated on the inner leaflet of the inner membrane or in the periplasmatic space, resulting in a mature LPS (Wang and Quinn, 2010; Whitfield and Trent, 2014). Completed LPSs are then exported via the Lpt (LPS transport) proteins (Chng et al., 2010). The components of the Lpt export system are an ABC exporter (LptBFGC), a chaperone in the periplasm LptA, and outer membrane proteins LptD and LptE (Zeng and Charkowski, 2021).

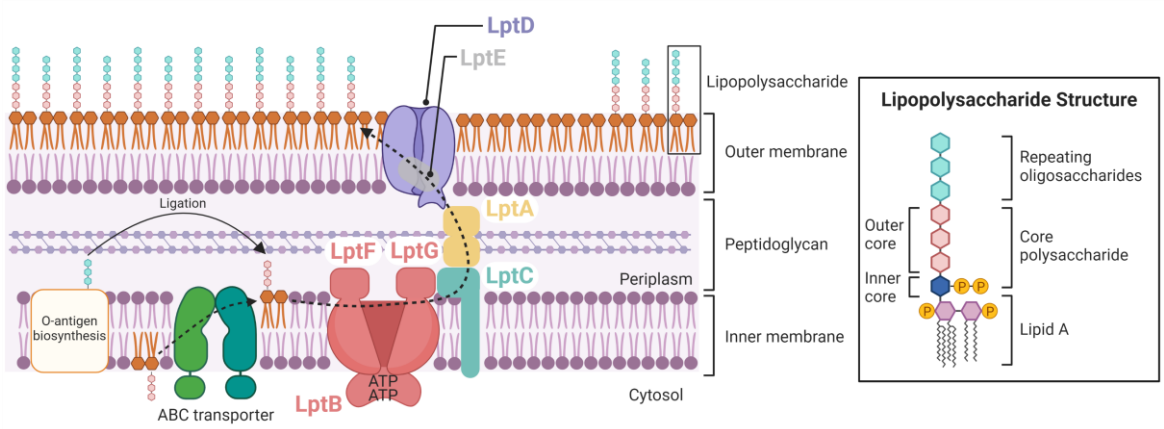


Figure 3. Export of LPSs through the membrane of Gram-negative bacteria and their structure
Lipid A is synthesized in the cytoplasm and flipped through the inner membrane by an ABC transporter. Core polysaccharides and O-antigens can be synthesized biosynthesized and exported through the inner membrane by ABC transporters or Wzy/Wzx pathways. Core polysaccharides can be ligated to the lipid A in the cytoplasm or the periplasm, and O-antigens are ligated in the periplasm. Nascent LPSs are then extracted from the periplasm leaflet of the inner membrane and exported to the outer leaflet of the outer membrane by the Lpt protein machinery. Adapted from Davidson et al., 2008; Chng et al., 2010; Wang and Quinn, 2010; Whitfield and Trent, 2014; Putker et al., 2015; Zeng and Charkowski, 2021. Created with BioRender.com.

1.4 Study area

Investigation of soil properties and functional microbial communities during soil development provides valuable insight into the important and complex process of soil formation, without which life on Earth would not be sustainable. In the present thesis, I followed this process in a chronosequence of reclaimed soils, with a special focus on soil structure and bacteria that have the potential to aid in its formation and stabilization. A chronosequence is a set of soils originating from the same parent material or substrate, which differ in the time since they were formed. Chronosequences have long been used as space-for-time substitutions, usually for studying long-term effects that cannot be observed directly (Walker et al., 2010) and are therefore valuable tools for investigating rates and directions of soil evolution (Huggett, 1998). The main assumption in research based on soil chronosequences is that four out of the five soil formation factors remain relatively constant, with time being the main factor that changes and affects soil evolution.

This study was conducted in an open-cast lignite mining area of the ‘Garzweiler’ mine (51°50 N, 6°280 E) located in western Germany, near the city of Cologne. In this area, the

mining company (RWE Power AG, Essen, Germany) extracts lignite and loess parent material using large bucket excavator machines. Loess is the natural parent material found in this area where it was deposited 12,000 – 100,000 years ago during the Weichselian glaciation period and can be found in thicknesses of up to 20 m (Pihlap et al., 2021). For the reclamation of mined sites, the company mixes approximately two meters of developed Luvisol topsoil and 18 m of loess parent material (ratio of approximately 1:10) and stores it in stockpiles for about three months. Finally, this material is deposited on top of the overburden material to a thickness of two meters, with the end goal of creating fertile agricultural soils.

The company manages the newly-formed sites during the first seven years of reclamation after which the land is returned to farmers. During the first three years of reclamation, described as the initial phase, *Medicago sativa* (alfalfa) is cultivated for its deep rooting system and nitrogen-fixing potential. In the fourth year, there is a transition from the initial phase to the agricultural management phase. During this time, soils are tilled to a depth of 30 cm and fertilized with 30 t ha^{-1} of plant-based organic compost. In years four to seven, up to 1.5 t ha^{-1} of mineral fertilizers in different forms of nitrogen, phosphorus, boron, manganese, potassium, and calcium are applied. Organic fertilizer is added again in the seventh year in the form of manure, after which only mineral fertilizers are applied. During this time, a crop rotation of wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.) is cultivated, followed by a cropping sequence of winter and summer wheat, winter barley (*Hordeum vulgare* L.), rapeseed (*Brassica napus* L.), and maize (*Zea mays* L.), which are typical crops for the region. The mean annual temperature in the area is 9.5°C and annual precipitation is 720 mm. The typical soil type is Haplic Luvisol with a silty clay loam texture, and a pH of 7.4 to 7.6 (**Publication II**).

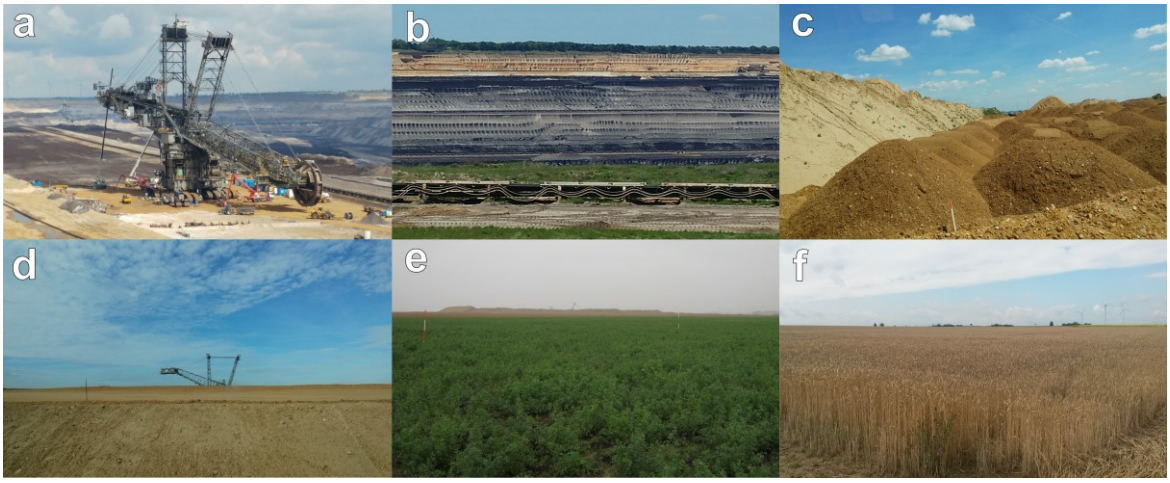


Figure 4. Mining and reclamation in the Garzweiler mine
a: bucket excavator machine used for mining coal. b: layers excavated during mining; brown Luvisol topsoil and dark lignite layers underneath. c: stockpiles of the Luvisol topsoil and loess parent material used for reclamation. d: stockpiles deposited on top of the overburden post-mining material and leveled to create a field. e: reclaimed field after one year of Medicago sativa growth. f: wheat field after 12 years of reclamation.

1.5 Aims and hypotheses

The basis of the work presented in this thesis is a comprehensive comparison of soils in different stages of development with a review of their chemical and biological characteristics. The main focus is the interplay between biological factors, particularly bacteria, and the formation of soil structure through the investigation of nutrient composition, microbial biomass, bacterial community composition, and bacteria harboring genes involved in promoting the formation of soil aggregates. Most previous studies of developing soils focused on nutrient quantification, microbial taxonomic characterization, or functional potential, yet there were no studies investigating the bacterial role in soil structure development through the production of extracellular polymeric substances along a post-mining chronosequence. Such a study, where I combined findings from metagenomic shotgun sequencing data with a direct measurement of bacterial products that act as a gluing agent for soil particles was first published in *Microbial Biotechnology* (**Publication I**) and is described in this dissertation. Therefore, in the frame of this dissertation, it was my aim to:

(1) describe nutrient and biomass dynamics in a chronosequence of developing soils as influenced by time, management, soil depth, and season. This was performed through an investigation of available C and N, nutrient quantity and composition, and microbial biomass.
(2) investigate the consequences of soil management on microbial community composition and functional potential in an agricultural chronosequence of soils. The investigation was particularly focused on bacterial exopolysaccharide and lipopolysaccharide producers combined with the quantification of extracellular polymeric substances (EPS) along the chronosequence.

In this respect, the following hypotheses were tested on a 24-year-old agricultural chronosequence in a post-mining reclamation site:

Hypothesis 1a – microbial and nutrient dynamics along the chronosequence

We expect all measured soil parameters to vary between soils of different ages and show distinct patterns between the initial reclamation phase and the agricultural management phase. Specifically, newly-formed soils should be low in macronutrients, organic matter, microbial biomass and diversity, and these parameters should increase with time and decrease with depth. After the creation of new reclaimed fields, we expect the first boost

to follow the introduction of plants to the system in the initial reclamation phase and have conditions especially favorable for fungal growth. After the transition to the agricultural management phase, we expect fungal networks to be destroyed and the abundance of bacteria to peak due to the addition of easily-accessible nutrients.

Hypothesis 1b – seasonal dynamics

We expect seasonal changes to inconsistently vary between the soils in different development stages. In the initial reclamation phase, we expect all measured parameters to increase during the growing season while the first legume plants grow without perturbation. As reclamation is then converted to the agricultural management phase, we expect to find the highest amounts of organic matter, nutrients, and microbial biomass in samples taken after wheat seeding in spring, and a decrease thereafter.

Hypothesis 2a – bacterial diversity and community composition

We expect the newly-formed soils to be populated mainly by typical colonizers of oligotrophic environments, especially nitrogen fixers, and for the early microbial communities to have a stochastic distribution influenced by local conditions in niches formed in the initial mixture of parent material and previous agricultural soil. With the introduction of the first legume plant, we expect an increase in symbiotic nitrogen-fixing bacteria. In later stages of the agricultural management phase, we expect an increase in bacterial groups typically found in agricultural soils where they perform important ecosystem services, such as nutrient and water cycling (e.g. *Nitrospiraceae, Nitrosomonadaceae, Streptomycetaceae, Pseudomonadaceae*).

Hypothesis 2b – exopolysaccharide and lipopolysaccharide-forming bacteria

In the initial reclamation phase, symbiotic nitrogen-fixing bacteria should be the main carriers of exopolysaccharide and lipopolysaccharide genes, as these bacterial groups are known to use extracellular polymeric substances in the initial colonization of legume roots. After the system is converted to agricultural management by tillage, fertilization, and crop rotation, we expect the bacterial community to use extracellular polysaccharides to trap nutrients and protect themselves from the physical stress induced by tillage. We hypothesize the bacterial groups typically found in agricultural soils dominate as polysaccharide producers in this phase. However, with time and constant agricultural management, we expect the bacterial community to adapt and thus have a lower need to

produce extracellular polysaccharides. We hypothesize that there will be a peak in the amount of extracellular polymeric substances and bacteria capable of producing them at the points of the biggest disturbances. Specifically, when alfalfa is first planted, and especially when the system is converted to agricultural cropping associated with tillage and organic input.

To test the hypotheses, a series of experiments were performed in a post-mining agricultural reclamation site, with samples taken from a chronosequence of soils that originate from the same substrate but are in different stages of development. Details of the experiments and sampling schemes are described in chapter 2.1 Study Design.

2 Materials and methods

2.1 Study design

To simulate a chronosequence that spans 24 years of soil development, fields in different stages of the reclamation process were selected and sampled for use in two experiments. The first experiment investigates the development of available nutrients and microbial abundances along the chronosequence and: (a) through the soil depth profile, (b) throughout a growing season. The second experiment investigates the development of bacterial potential to produce exopolysaccharides and lipopolysaccharides and improve soil structure development. An illustration of the study design, samples used in each experiment, and methods applied is shown in Figure 5.

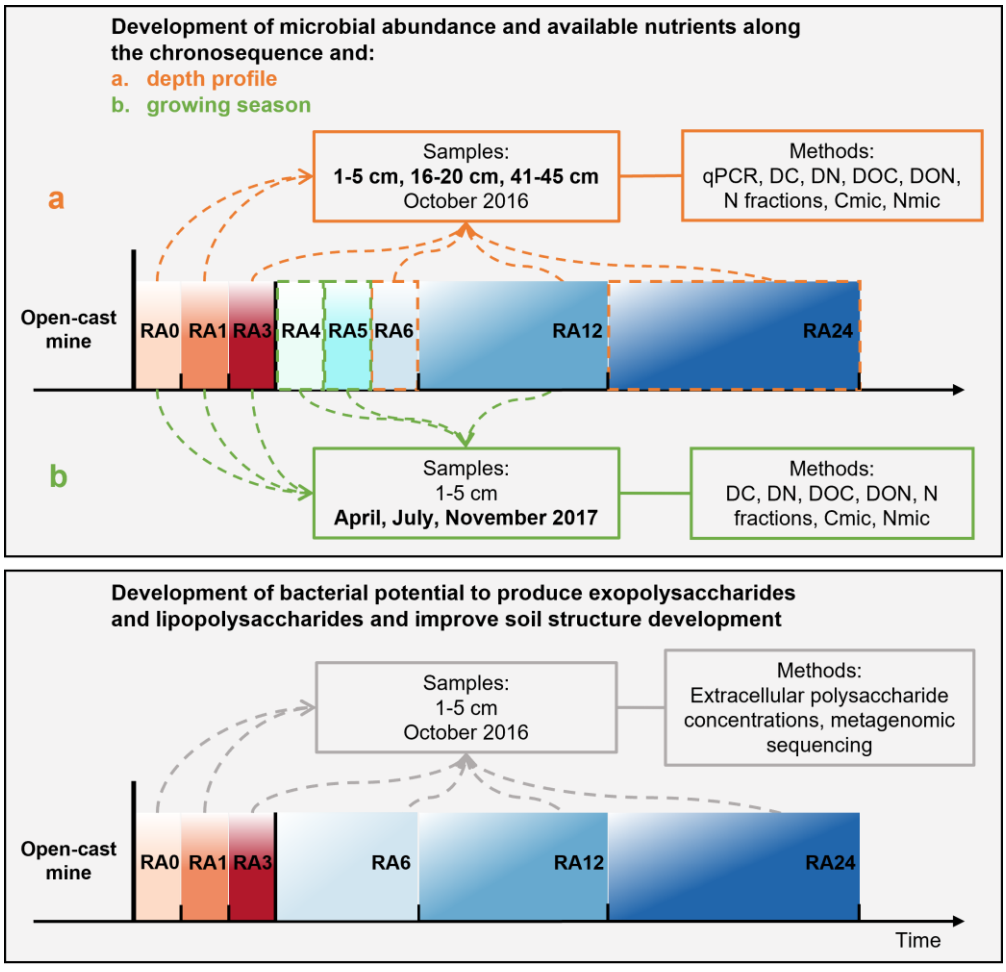


Figure 5. Study design and samples collected for two experiments
Colored boxes represent reclaimed fields, which originate from the same loess parent material and undergo the same management, but differ in reclamation age (RA) and soil development stage.

With permission from the RWE mining company, I collected soil samples in October 2016 and in April, July, and November 2017. For the first experiment, I collected samples from three soil depths in October 2016 and from topsoil during the growing season of 2017. For the second experiment, I used topsoil samples collected in October 2016.

The first three sites, aged 0, 1, and 3 years, represent the initial reclamation phase in which alfalfa is planted. Older sites, aged 4, 5, 6, 12, and 24 years, represent the agricultural management phase characterized by a crop rotation. The sites are referred to as RA0, RA1, RA3, RA4, RA5, RA6, RA12, and RA24, respectively. Field sizes were in the range of 2 to 35 ha, and samples were taken from three plots of approximately 200 m^2 in each field for biological replication. In sites sampled in October 2016, the distance between replication subplots was 4, 7, 6, 34, 10, and 51 m in subsequent reclamation ages, due to different field sizes. In sites sampled during 2017, the approximate distance between replication subplots was 38 m. A map of sampled sites is shown in Figure 6.

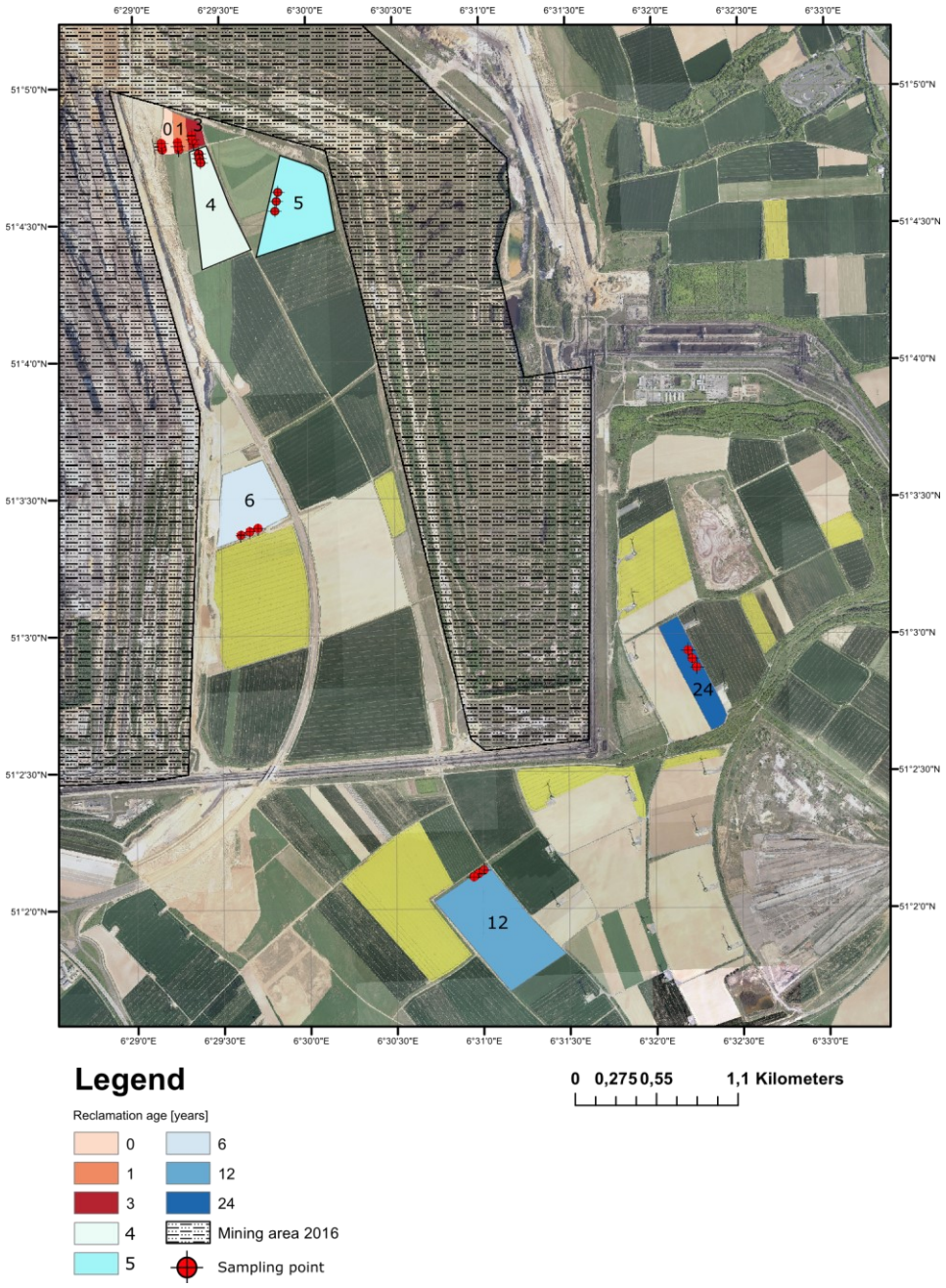


Figure 6. Map of the sampling site showing the chosen fields
Fields from the initial reclamation phase are marked with red colors and fields from the agricultural management phase are marked with blue colors.

A description of each site at the time of sampling is summarized in Table 1. RA0 was a bare field in October 2016 and April 2017 and was planted with alfalfa by July 2017. RA1 and RA3 were planted with alfalfa in October 2016 and the plants were cut down in April and November 2017. Fields from the agricultural management phase were planted with

wheat and maize in October 2016. By April 2017, maize was replaced with wheat in RA12, which was replaced with barley by November 2017. RA4 and RA5 were seeded with wheat in April 2017, which has been growing also in July. By November, the wheat was harvested and fields were seeded with new wheat plants.

Table 1. Description of fields sampled in 2016 and 2017.

			Initial phase			Agricultural management phase		
Sampling time point	Depth sampling		RA0	RA1	RA3	RA6	RA12	RA24
		October 2016	No alfalfa	Alfalfa	Alfalfa	Wheat	Maize	Wheat
			RA0	**RA1**	**RA3**	**RA4**	**RA5**	**RA12**
	Seasonal sampling	**April 2017**	No alfalfa	Alfalfa, cut down around five days before sampling	Alfalfa, cut down around five days before sampling	Wheat, seeded around 20 days before sampling	Wheat, seeded around 20 days before sampling	Wheat, seeded around 20 days before sampling
		July 2017	Alfalfa	Alfalfa	Alfalfa	Wheat	Wheat	Wheat
		November 2017	Alfalfa	Alfalfa, cut down four weeks before sampling	Alfalfa, cut down four weeks before sampling	Wheat, seeded around 10 days before sampling	Barley, seeded mid-September	Wheat, seeded around 10 days before sampling

2.2 Development of microbial abundance and available nutrients along the chronosequence

For the depth profiling in the first experiment, samples were taken from three depths: 1-5 cm, 16-20 cm, and 41-45 cm (subsoil layer) in RA0, RA1, RA3, RA6, RA12, and RA24. In the first two depths, samples were taken in four corners of each plot and pooled, while the subsoil samples were taken once per plot. In each plot, samples were homogenized and immediately cooled in the field with dry ice, and subsequently stored at 4 °C for chemical analyses and at -80 °C for DNA extraction. In total, this sampling design resulted in 54 samples (6 fields x 3 plots x 3 depths).

To study changes in the chronosequence during a growing season, samples were taken from the topsoil (1-5 cm) in the same way in sites RA0, RA1, RA3, RA4, RA5, RA6, and

RA12, in April, July, and November. In total, this sampling design resulted in another 54 samples (6 fields x 3 plots x 3 months).

2.2.1 Dissolved carbon and nitrogen

Dissolved organic carbon and nitrogen (DC and DN, respectively), as well as different nitrogen forms (i.e. ammonium, nitrate, nitrite), were extracted using the $CaCl_2$ solution. Five grams of each soil sample were extracted in 20 mL of 0.01 M $CaCl_2$ solution after 45 min of horizontal shaking. After filtering (Whatman 595 ½, Merck KGaA, Germany), samples were stored at -20 °C until further measurements.

To determine the concentration of organic dissolved carbon and nitrogen (DOC and DON, respectively), a drop of 32% HCl was added to each sample before measuring due to a high concentration of carbonate in the soils. DC, DN, DOC, and DON concentrations were measured on DIMATOC 2000 (Dimatec Analysistechnik GmBH, Germany). Total N (TN), ammonium, nitrate, and nitrite concentrations were measured from the $CaCl_2$ extracts on Skalar Continuous Flow Analyzer SA5100 (Skalar Analytical B.V., Netherlands) before the addition of HCl[1].

2.2.2 Microbial carbon and nitrogen

Microbial carbon and nitrogen (C_{mic} and N_{mic}, respectively) were determined by the chloroform fumigation method after Vance et al. (1987), and modified after Joergensen (1996). The soil was fumigated with chloroform for 24 h, after which it was extracted in $CaCl_2$ and measured on DIMATOC 2000 (Dimatec Analysistechnik GmBH, Germany), as described above. C_{mic} and N_{mic} values were calculated as the difference between DOC and DON in fumigated and non-fumigated samples. C_{mic} was calculated with the k_{ec} value of 0.45 (Joergensen, 1996), and N_{mic} with the k_{en} value of 0.54 (Joergensen and Mueller, 1996).

2.2.3 DNA extraction

DNA extraction was performed using the NucleoSpin Soil kit (Macherey-Nagel, Germany) from approximately 300 mg of soil. Sample material was suspended in lysis buffer SL1 with a 15 µl addition of the Enhancer SX. Blank extractions without sample material were performed as the negative control.

[1] DIMATOC 2000 and Skalar Continuous Flow Analyzer measurements were performed by Gudrun Hufnagel (Research Unit Comparative Microbiome Analysis, Helmholtz Zentrum München).

DNA quality and the presence of residual contaminants were assessed with a NanoDrop 1000 Spectrophotometer (PeqLab, Erlangen, Germany). DNA concentration was quantified on a SpectraMax Gemini EM Spectofluorometer (Molecular Devices GmbH, Germany) with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, USA). Extracted DNA was stored at -20 °C until further processing.

2.2.4 Quantification of bacterial, archaeal, and fungal marker genes

To measure the absolute abundance of bacteria, archaea, and fungi in the three sampled depths, quantitative PCR (qPCR) assays were performed to quantify bacterial and archaeal 16S rRNA genes and the fungal ITS region. The quantification was performed on a 7300 Real-Time PCR System (Applied Biosystems, Germany) with SYBR Green (Applied Biosystems, Germany) used for detection. The volume of each PCR reaction was 25 µl and consisted of 12.5 µl of PowerSYBR Green, 0.5 µl of primers, 3% BSA in the case of 16S rRNA genes, 1 µl of primer, and no 3% BSA in the case of the ITS region. Thermal profiles, primers, and sources of standard plasmids are shown in

Table 2.

Table 2. Thermal profiles, primer references, and standard plasmid sources of bacterial and archaeal 16S rRNA genes and fungal ITS region qPCR reactions.

Target genes	Initial denaturation time at 95°C (min)	Touch-down	Cycles	Thermal profile	Cycles	Primers and references	Standard plasmid sources
16S rRNA bacteria	10	-	-	95 °C (45s) 58 °C (45s) 72 °C (45s)	40	**FP, RP** (Bach et al., 2002)	*Pseudomonas putida*
16S rRNA archaea	10	95 °C (20s) 55 °C (60s)† 72 °C (60s)	5	95 °C (15s) 50 °C (30s) 72 °C (30s)	40	**rSAf(i)** (Nicol Graeme et al., 2005), **958r** (Bano et al., 2004)	*Methanobacterium* sp.
ITS	10	-	-	94 °C (30s) 50 °C (30s) 72 °C (30s)	40	**ITS1, ITS4** (White et al., 1990)	*Trichoderma reesei*

† -1 °C/cycle

To minimize the effect of PCR inhibitors, pre-tests with different DNA dilutions were performed, and a dilution of 1:20 was chosen as it showed no PCR inhibition. Plasmids used as standards for each gene were diluted in a range of 10^1 to 10^7 gene copies per

microliter, and all were analyzed in triplicates. Negative controls were included in each PCR run. The efficiency of amplification was calculated with the equation $Eff=[10^{(-1/slope)} – 1]$ and resulted in 92%, 84%, and 83% for 16S bacterial and archaeal genes and fungal ITS region, respectively. Melting curves were analyzed to confirm the specificity of amplicons and to ensure no primer dimers were formed.

2.3 Development of bacterial potential to produce exopolysaccharides and lipopolysaccharides and improve soil structure development

Samples used for this experiment were topsoil samples (1-5 cm) collected in October 2016 from RA0, RA1, RA3, RA6, RA12, and RA24. In total, 18 samples (6 fields x 3 plots) were analysed.

2.3.1 Extracellular polymeric substance (EPS) extraction[2]

To compare the quantity and quality of extracellular polymeric substances (EPS) in soils of different reclamation stages, extraction was performed from 2.5 g dry weight of topsoil from each reclamation age after (Redmile-Gordon et al., 2014). The extraction protocol is based on a technique that uses cation exchange resin according to (Frølund et al., 1996). The protocol was tested and confirmed as the most selective method with the lowest cell-damaging effects (Redmile-Gordon et al., 2014; Redmile-Gordon et al., 2015). Exopolysaccharides and proteins are the two EPS components measured as described by DuBois et al. (1956) and Redmile-Gordon et al. (2013), using the Lowry assay.

2.3.2 Metagenome library preparation and sequencing

Shotgun metagenomics sequencing was performed for DNA extracted from the topsoil layer (1-5 cm) as described in 2.2.3. Since the DNA yield for samples from RA0 was too low for sequencing, the extraction for all RAs was performed three times and the extracts were pooled to ensure enough material could be retrieved from all samples. First, 100 ng of DNA per sample was sheared in the E220 Focused-ultrasonicator (Covaris, USA) to the target size of 500 bp, according to the manufacturer's protocol. Conditions of shearing were 175 W peak incident power, 5% duty factor, 200 cycles per burst, the temperature of 7 °C, and 45 s of treatment time. After shearing, the concentration and quality of DNA were measured with the High Sensitivity kit on the Fragment Analyzer (Advanced Analytical Technologies Inc., USA). DNA was further diluted to match the desired 100 ng in each

[2] EPS extraction was performed by Dr. Cordula Vogel, a cooperation partner at the Institute of Soil Science and Site Ecology, Technical University of Dresden.

sample. In the next step, DNA was used for end repair, adaptor ligation, size selection, and library amplification with the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs, UK). According to the manufacturer´s protocol for DNA input of 100 ng or less, the NEBNext Adaptor was diluted 10-fold. The size selection of Adaptor-ligated DNA was performed for the approximate insert size of 400-500 bp in the clean-up step. Afterward, Adaptor-ligated DNA was amplified with NEBNext Multiple Oligos for Illumina (New England Biolabs, UK) in eight PCR cycles. PCR products were purified twice with AMPure XP beads, using a 1:0.6 DNA-to-bead ratio. Length distributions of the resulting libraries were estimated using the Standard Sensitivity kit on the Fragment Analyzer (Advanced Analytical Technologies Inc., USA).

Libraries were diluted to final concentrations of 4 nM per sample, pooled equimolar, and spiked with 1% PhiX as sequencing control. Paired-end sequencing was performed on a MiSeq instrument (Illumina Inc., San Diego, CA, USA) using the Miseq Reagent Kit v3 for 600 cycles[3].

2.3.3 Bioinformatic analysis

After the sequencing, raw sequences were processed as described in Vestergaard et al. (2017) to ensure good quality controls. First, the sequencing adapters were removed and reads were trimmed with AdapterRemoval v2.1 (Schubert et al., 2016), which was set to Phred quality of at least 15 and read length of at least 50. Next, PhiX contaminations, used as the sequencing control, were removed with Deconseq v0.4 (Schmieder and Edwards, 2011)[4].

Estimation of the average coverage of sequenced metagenomes was performed with Nonpareil v2.4 (Rodriguez and Konstantinidis, 2014). In addition, the Nonpareil diversity index was calculated and used as a proxy for describing the complexity of microbial communities in the samples.

2.3.4 Microbial community structure

Filtered sequencing reads were blasted against the NCBI non-redundant protein sequences database for their taxonomic classifications. Reads were blasted with Kaiju v1.4 (Menzel et al., 2016) in Greedy mode with five mismatches allowed. In addition, reads

[3] MiSeq sequencing was performed by Susanne Kublik (Research Unit Comparative Microbiome Analysis, Helmholtz Zentrum München).
[4] Pipelines for shotgun metagenomic sequence annotations were developed by Dr. Gisle Vestergaard (Department of Biology, University of Copenhagen, Denmark).

containing the V4 region of the 16S rRNA genes were extracted with RiboTagger v0.8 (Xie et al., 2016) for taxonomic profiling. Subsequently, they were annotated to the SILVA database (release 119).

2.3.5 Exopolysaccharide and lipopolysaccharide genes

As described in Cania et al. (2019), protein sequences involved in the production and excretion of exopolysaccharides and lipopolysaccharides were downloaded from the online Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology database (in October 2016). Next, they were screened for the presence of function-specific conserved domains with NCBI's CD-search (Marchler-Bauer et al., 2015). KEGG Orthology (KO) numbers that contained those domains were used to construct an exopolysaccharide- and lipopolysaccharide-specific database with a total of 14 genes (Table 3). Then, Hidden Markov Models (HMM) corresponding to those sequences were downloaded from the TIGRFAM v15 (Haft et al., 2013) and Pfam v30 (Finn et al., 2016) databases.

Open reading frames (ORFs) in filtered sequences from soil samples were predicted with FragGeneScan v1.19 (Rho et al., 2010). Those were subsequently scanned with HMMER v3 (hmmer.org). Reads that matched HMMs downloaded from TIGRFAM and Pfam with an E-value threshold of 10^{-5} were blasted against the constructed exopolysaccharide- and lipopolysaccharide-specific KEGG Orthology databases. The criterium by which the reads were assigned with a KO number was the consistency of the top 25 blast results. Blasting was performed with Diamond v0.8 (Buchfink et al., 2014)[5].

[5] Pipelines for exopolysaccharide and lipopolysaccharide gene annotations were developed by Barbara Cania (Research Unit Comparative Microbiome Analysis, Helmholtz Zentrum München).

Table 3. Analyzed genes related to exo- and lipopolysaccharide production, the proteins they code for, and their respective KO numbers and HMM IDs.

	Gene	Protein	KO number	HMM ID
Exopolysaccharide	*wza*	Polysaccharide export outer membrane protein - Wza	K01991	PF02563
	algE	Alginate export outer membrane protein - AlgE	K16081	PF13372
	algJ	Alginate biosynthesis acetyltransferase - AlgJ	K19295	PF16822
	wcaB	Colanic acid biosynthesis acetyltransferase - WcaB	K03819	TIGR04016
	wcaF	Colanic acid biosynthesis acetyltransferase - WcaF	K03818	TIGR04008
	wcaK/amsJ	Colanic acid/amylovoran biosynthesis pyruvyl transferase - WcaK/AmsJ	K16710	TIGR04006
	kpsE	Capsular polysaccharide export system permease - KpsE	K10107	TIGR01010
	epsG	Exopolysaccharide biosynthesis transmembrane protein - EpsG	K19419	PF14897
	epsA	Exopolysaccharide biosynthesis tyrosine kinase modulator - EpsA	K19420	TIGR01006
	sacB	Levansucrase - SacB	K00692	PF02435
Lipopolysaccharide	*wzt*	Lipopolysaccharide transport system ATP-binding protein - Wzt	K09691	PF14524
	lptF	LptBFGC lipopolysaccharide export complex permease - LptF	K07091	TIGR04407
	lptG	LptBFGC lipopolysaccharide export complex permease - LptG	K11720	TIGR04408 PF03739
	lptC	LptBFGC lipopolysaccharide export complex inner membrane protein - LptC	K11719	TIGR04409 PF06835

2.4 Statistical analysis

Statistical analyses and data visualizations were performed with the R statistical software v4 (R Core Team, 2018). Data processing and visualization were mainly performed using the *tidyverse* v1.3 family of packages (Wickham et al., 2019). The analysis of variance contained in all measured chemical and biological parameters was performed with a principal component analysis (PCA) method using the *prcomp* function from base R and functions from the *factoextra* package v1.0 (Kassambara and Mundt, 2017) for additional calculations and visualization.

Statistical hypotheses were tested using the WRS2 package v1.1 (Mair and Wilcox, 2018) which contains robust statistical methods that do not rely on parametric assumptions. The effect of reclamation age on all measured parameters was tested with a robust one-way

ANOVA with trimmed means using the *t1way* function and differences between reclamation ages were tested with a post hoc *lincon* function. For parameters measured in samples taken from different depths or seasons, the effect of reclamation age and depth or season was tested with a robust two-way ANOVA with trimmed means. Results after all post hoc tests were corrected with the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) and were considered significant when the *p*-value was lower than 0.05.

All statistical analyses of sequencing results were based on relative read abundances due to different amounts of sequencing reads obtained per sample. Data processing for phylogenetic analysis was mainly performed using the *phyloseq* package v1.32 (McMurdie and Holmes, 2013). Non-metric multidimensional scaling (NMDS) analysis of bacterial reads based on the Bray-Curtis dissimilarity indices was performed using the *ordinate* function from the *phyloseq* package, and additional calculations (stress plot) were performed using the *vegan* v.2.5 (Oksanen et al., 2018) package. Differences in particular taxa and functional genes between reclamation ages were tested using the robust one-way ANOVA with trimmed means described above.

3 Results

3.1 Development of microbial abundance and available nutrients along the chronosequence

3.1.1 Depth profile

3.1.1.1 Distribution of nutrients with soil development and depth

Dissolved carbon (DC) and nitrogen (DN) were measured from $CaCl_2$ extracts of samples taken in October 2016 in three depths of six reclamation ages (Figure 7).

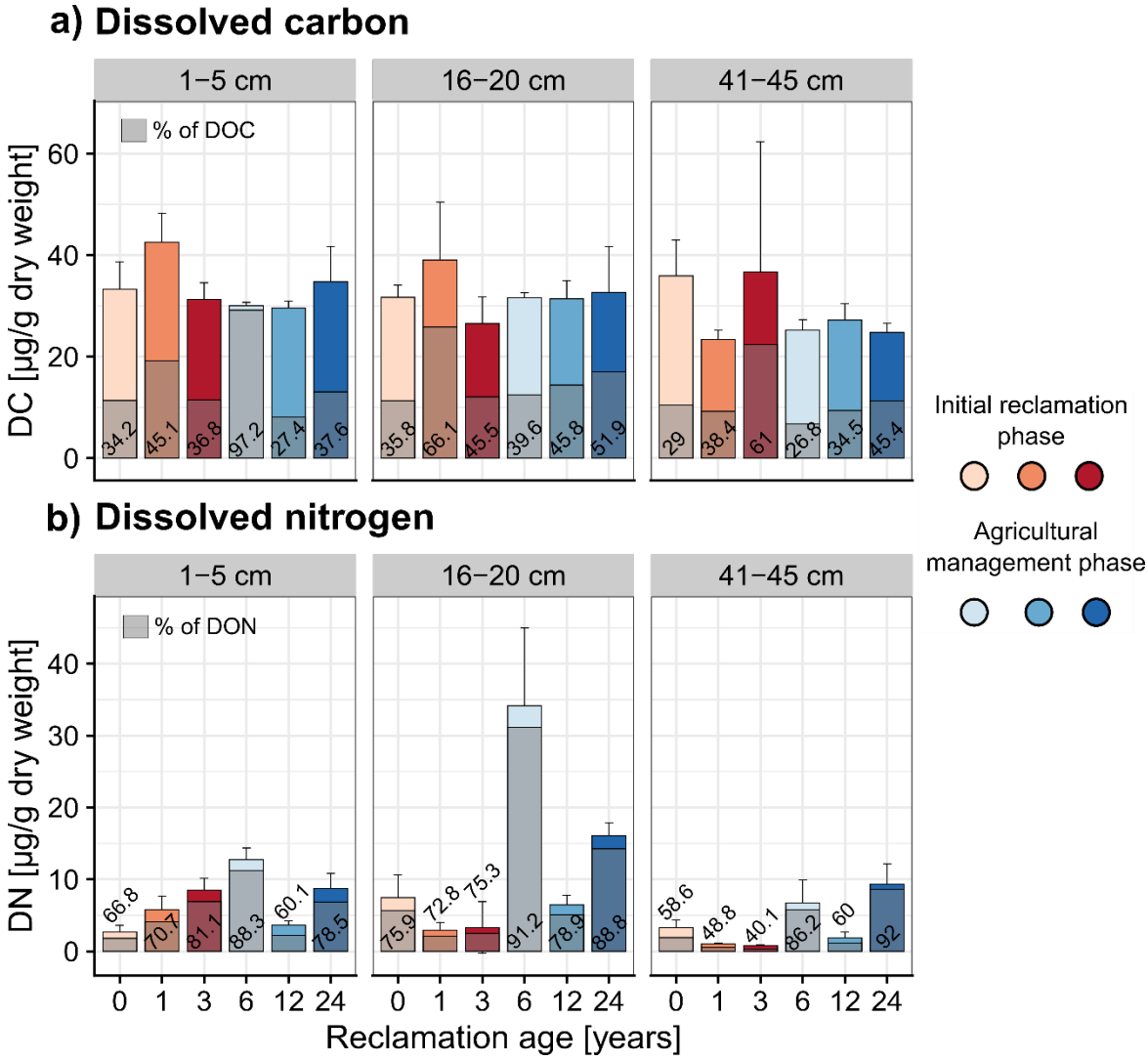


Figure 7. Concentrations of dissolved C and N in three depths
Bar plots showing the concentrations of dissolved carbon (a) and nitrogen (b) and the percentage of organic carbon and nitrogen as measured after adding HCl to the $CaCl_2$ extracts. Bars are colored by the reclamation age and represent means of triplicates, and the greyed-out parts show the organic fraction within each sample.

Average concentrations of DC (Figure 7a) ranged between 25 in the third depth of RA24 and 42.6 µg/g dw in the first depth of RA1. No significant difference was detected between reclamation ages and the three depths. On the other hand, there was a significant effect of reclamation age ($p = 0.008$) and the interaction of reclamation age and depth ($p = 0.001$) on DOC concentrations. On average, DOC constituted between 26.8 and 66.1% of DC, with the exception of RA6 depth 1-5 cm where it constituted 97.2%. This was significantly different from RA0, RA3, and RA12 in the same depth.

Average concentrations of DN (Figure 7b) ranged between 0.8 in the third depth of RA3 to 34.2 µg/g dw in the second depth of RA6. Reclamation age, depth, and their interaction had a significant effect on DN concentrations ($p = 0.001$). In the depth of 1-5 cm the concentration in RA6 was significantly higher compared to RAs 0, 1, and 12. The concentration in RA3 was also significantly higher than in RA0. Concentrations in the depth of 16-20 cm were significantly different between RA1 and RA24, as well as RA12 and RA24.

On average, DON constituted 40.1 to 91.2% of DN. Reclamation age, depth, and their interaction had a significant effect on DON concentrations ($p = 0.001$). In the depth of 1-5 cm, DON concentration was highest in RA6, where it was significantly higher compared to RA0, 1, 3, and 12. Similar to DN, the concentration of DON was significantly higher in RA3 compared to RA0. DON concentration was highest in RA6 in the second (16-20 cm), as well as the first depth.

In the second and third depths, DON concentrations were generally higher in samples from the agricultural management phase compared to the samples from the initial reclamation phase.

Concentrations of ammonium, nitrite, and nitrate were also measured from $CaCl_2$ extracts for all samples (Table 4).

Table 4. Ammonium and nitrate concentrations
Results are shown as means of triplicates and standard deviations. a,b,c next to the results show significant differences between samples after performing a post hoc test in each soil depth.

Reclamation age [years]	Depth [cm]	Ammonium [µg/g dw]	Nitrate [µg/g dw]
0	1 - 5	1.54 ± 0.68[a]	0.52 ± 0.17
1		3.29 ± 1.31[ab]	0.4 ± 0.26
3		5.65 ± 0.86[b]	0.18 ± 0.19
6		9.62 ± 1.29[c]	0.38 ± 0.03
12		2.07 ± 0.45[a]	0.25 ± 0.14
24		5.21 ± 2.56[abc]	0.28 ± 0.26
0	16 - 20	5.04 ± 2.72	0.47 ± 0.07
1		1.65 ± 0.58[a]	0.4 ± 0.02
3		2.15 ± 3.05	0.36 ± 0.09
6		**24.68 ± 7.24**	0.35 ± 0.02
12		3.81 ± 0.39[b]	0.35 ± 0.03
24		11.46 ± 1.5[c]	0.36 ± 0.05
0	41 - 45	2.46 ± 2.15	0.29 ± 0.07
1		0	0.35 ± 0.01
3		0	0.32 ± 0.11
6		5.07 ± 2.5	0.3 ± 0.03
12		0.49 ± 0.84	0.32 ± 0.02
24		6.58 ± 2	0.36 ± 0.02

Nitrite concentrations were below the detection limit, and nitrate concentrations were below one µg/g dw in all samples. Reclamation age, depth, and their interaction had a significant effect on ammonium ($p = 0.001$). Depth had a significant effect on nitrate concentrations ($p = 0.02$) and in the first depth, nitrate was highest in RA0. Ammonium concentrations had a similar trend as DN and DON, with the highest concentration found in the second depth of RA6. In the first depth, ammonium concentration was significantly higher in RA6 compared to RAs 0, 1, 3, and 12.

3.1.1.2 Distribution of microbial abundance with soil development and depth

Microbial carbon (C_{mic}) and nitrogen (N_{mic}) were measured from fumigated $CaCl_2$ extracts of samples taken in October 2016 in three depths of six reclamation ages (Figure 8).

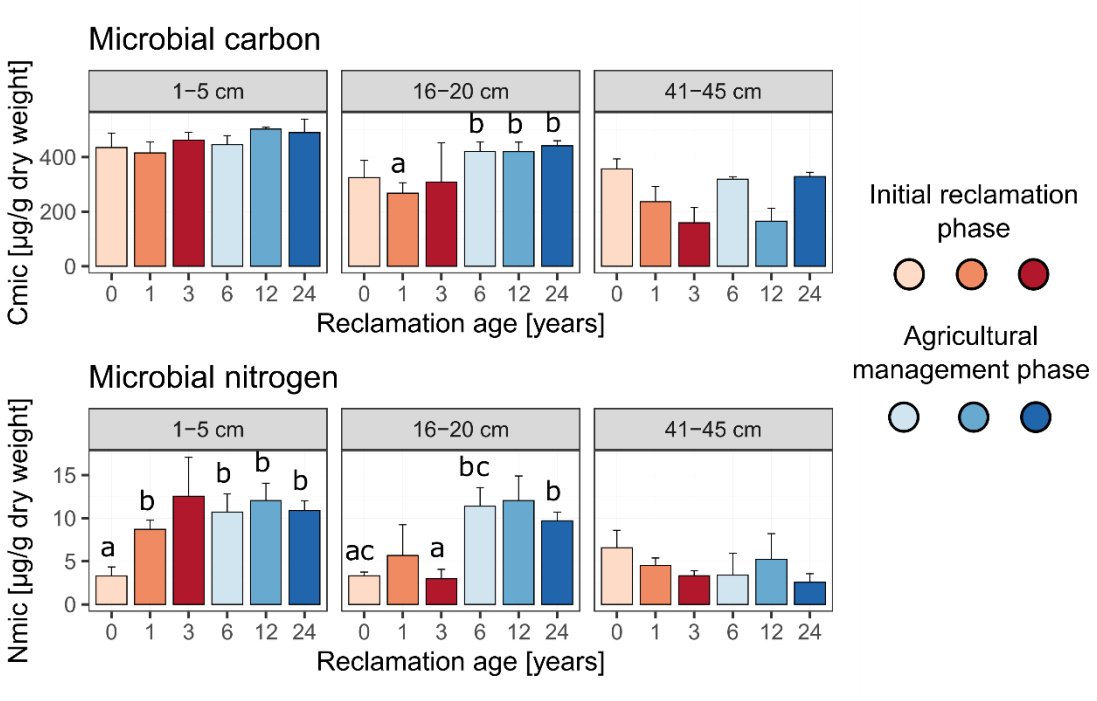


Figure 8. Concentrations of microbial C and N in three depths
Bar plots showing the concentrations of microbial carbon and nitrogen as measured after chloroform fumigation of $CaCl_2$ extracts. Bars are colored by the reclamation age and represent means of triplicates and error bars represent standard deviations. The letters above bars indicate the significant differences between samples after performing a post hoc test in each soil depth.

Reclamation age, depth, and their interaction had a significant effect on both C_{mic} and N_{mic} ($p = 0.001$). For C_{mic}, there were no significant differences between samples in the first and the third depth after performing a post hoc test. In the first depth, C_{mic} increased with time; from 434.9 ± 52.7 in RA0 to 489.8 ± 48.9 µg/g dw in RA24. In the second depth, C_{mic} was higher in the agricultural management depth, with significant differences between RA1 and RA6, 12, and 24.

The concentration of N_{mic} was increasing with time and there was a significant difference between RA0 and RA1, as well as all samples from the agricultural management phase. During the initial reclamation phase, the concentration increased from 3.3 ± 1.1 in RA0 to 12.5 ± 4.5 µg/g dw in RA3. In the agricultural management phase, the concentrations remained comparable to those measured in RA3.

In the second depth, C_{mic} and N_{mic} concentrations were higher in the agricultural management phase. For N_{mic}, the concentration was significantly higher in RA24 compared

to RA0 and RA3. In the third depth, there were no clear trends between reclamation ages. For RA0 there was no clear depth stratification in both C_{mic} and N_{mic}.

The investigation of microbial abundances continued with the extraction of DNA from each sample and using it to quantify selected genes. Differences between amounts of DNA obtained from each sample are shown in Figure 9.

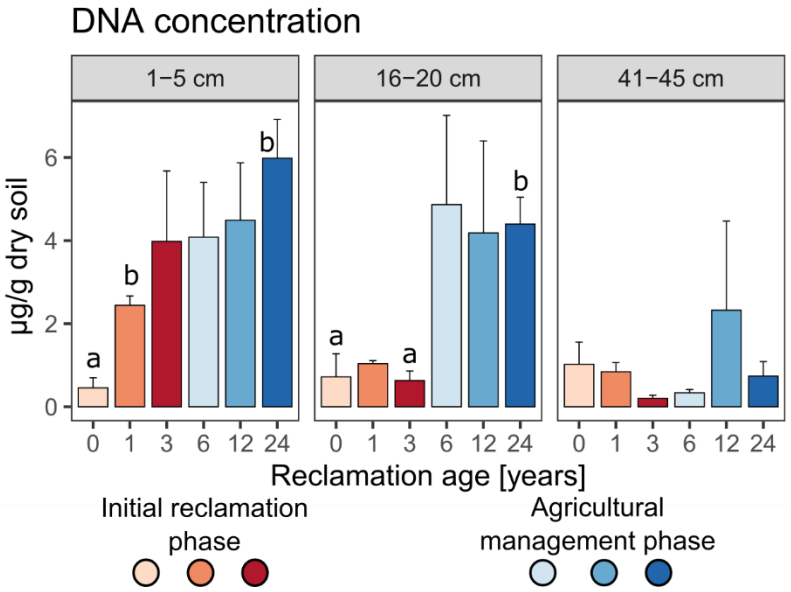


Figure 9. Extracted DNA concentrations
Bar plots showing the concentration of DNA extracted from each sample. Bars are colored by the reclamation age and represent means of triplicates and error bars represent standard deviations. The letters above bars indicate the significant differences between samples after performing a post hoc test in each soil depth.

Similar to C_{mic} and N_{mic}, all factors had a significant effect on DNA yield ($p = 0.001$). In the first depth, the amount of DNA extracted was increasing with reclamation ages. By RA1, the yield of DNA was five times higher than RA0 and by RA24 it was 13 times higher. In the second depth, the amount was higher in the agricultural management phase than in the initial reclamation phase, while no clear trends between reclamation ages were observed in the third depth.

Abundances of bacteria and archaea were calculated as 16S rRNA gene copy numbers per gram of dry soil, and the abundance of fungi was calculated as ITS region copy number per gram of dry soil using qPCR. The results are shown in Figure 10.

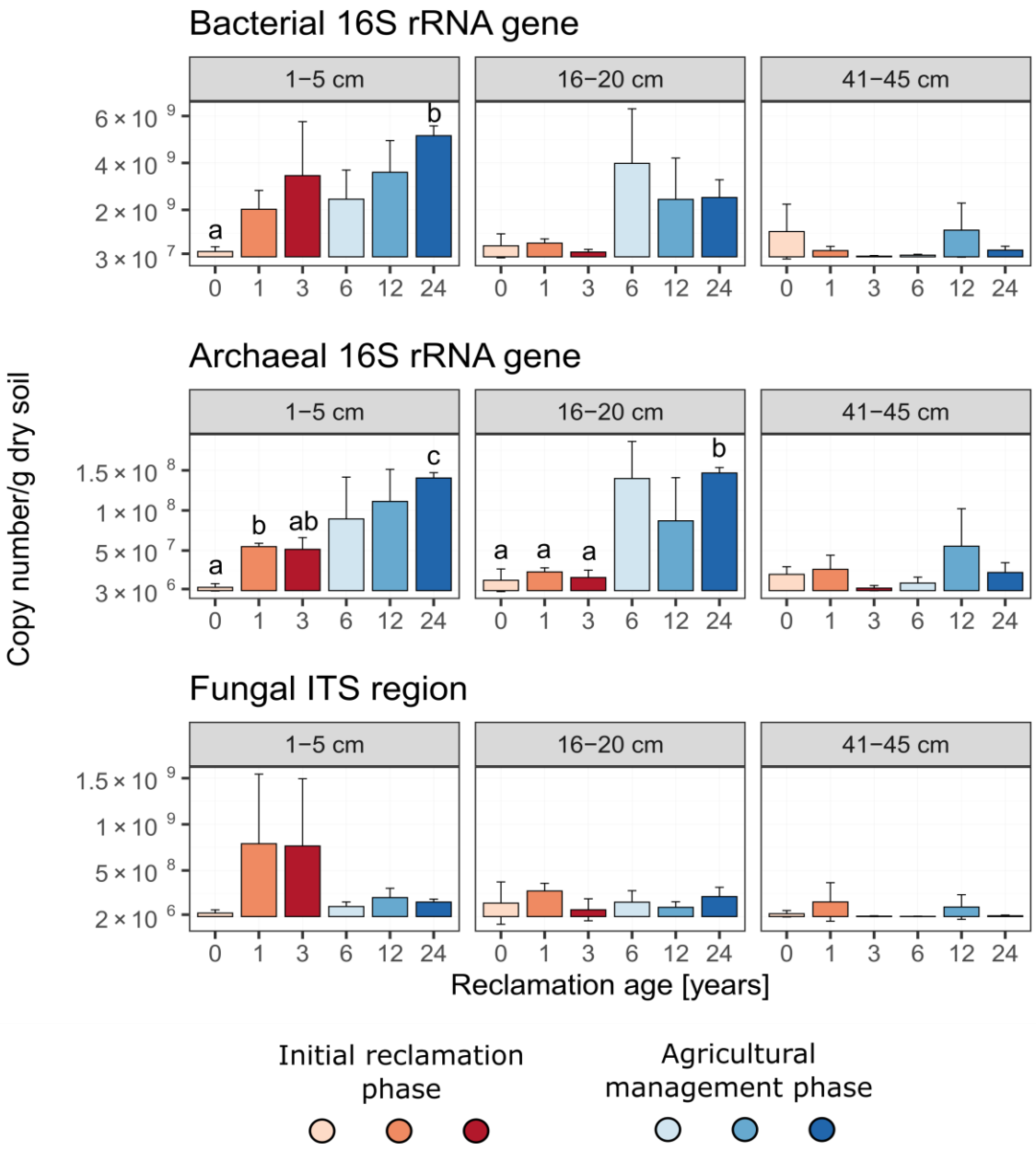


Figure 10. Microbial abundance as measured by qPCR
Bar plots showing 16S rRNA and ITS region copy numbers per g of dry soil. Bars are colored by the reclamation age and represent means of triplicates and error bars represent standard deviations. The letters above bars indicate the significant differences between samples after performing a post hoc test in each soil depth.

Bacteria dominated over the other two domains in all samples. Reclamation age, depth, and their interaction had a significant effect on bacterial abundance ($p = 0.001$). In the first depth, bacterial abundance increased during the initial reclamation stage from $2.26 \times 10^8 \pm 2 \times 10^8$ in RA0 to $3.46 \times 10^9 \pm 2.3 \times 10^9$ copies g^{-1} dry soil in RA3. After the transition to

the agricultural phase, the abundances decreased to 2.46 x 10^9 ± 1.24 x 10^9 in RA6 and reached the highest abundance of 5.16 x 10^9 ± 4.18 x 10^8 copies g^{-1} dry soil in RA24.

In the second depth, bacterial abundance was higher in the agricultural management phase, especially in RA6 where it reached 3.98 x 10^9 ± 2.32 x 10^9 copies g^{-1} dry soil. In the third depth, bacterial abundance was increased only in RA0 and decreased in all other RAs compared to the first and second depth.

The number of archaeal 16S rRNA genes was higher in the agricultural management phase, especially in the first two depths, and reached the highest abundance in the second depth of RA24 (1.47 x 10^8 ± 6.73 x 10^6 copies g^{-1} dry soil). Similar to the bacterial abundance, reclamation age, depth, and their interaction had a significant effect on archaeal abundance ($p = 0.001$).

In contrast, only depth had a significant effect on the fungal abundance ($p = 0.015$). The highest abundance of fungi was found in the soil samples of the first depth of the initial management phase, during the alfalfa growth: 7.9 x 10^8 ± 7.54 x 10^8 and 7.66 x 10^8 ± 7.29 x 10^8 copies g^{-1} dry soil at RA1 and RA3, respectively. Thereafter the abundance of fungi decreased.

3.1.1.3 Multivariate analysis of measured parameters

Principal component analysis (PCA) was performed to describe the variance contained in the chemical and biological parameters measured in the three depths of the six reclamation ages. The first two principal components of the resulting PCA explain 62.3% of the variance contained in the data (Figure 11a). Different nitrogen fractions (DN, DON, TN, and ammonium) contributed most to the variance explained by the first two principal components (Figure 11 b).

A factor map of the variables (Figure 11 c) shows the variable's quality of representation, calculated as squared coordinates (cos^2). A high cos^2 indicates a good representation of the variable on the principal component, and these variables are dark in color and positioned close to the circumference of the correlation circle. A low cos^2 indicates that the variable is not represented by the principal components, and these variables are light in color and located closer to the center of the circle. Different nitrogen fractions explained the variance in the data more than different carbon fractions (DC and DOC), which together

explained less than 1%. Microbial nitrogen and carbon, bacterial abundance, and DNA concentration are grouped together and have a good representation on the principal components. These variables are also negatively correlated to the nitrate concentration and the dry weight of samples.

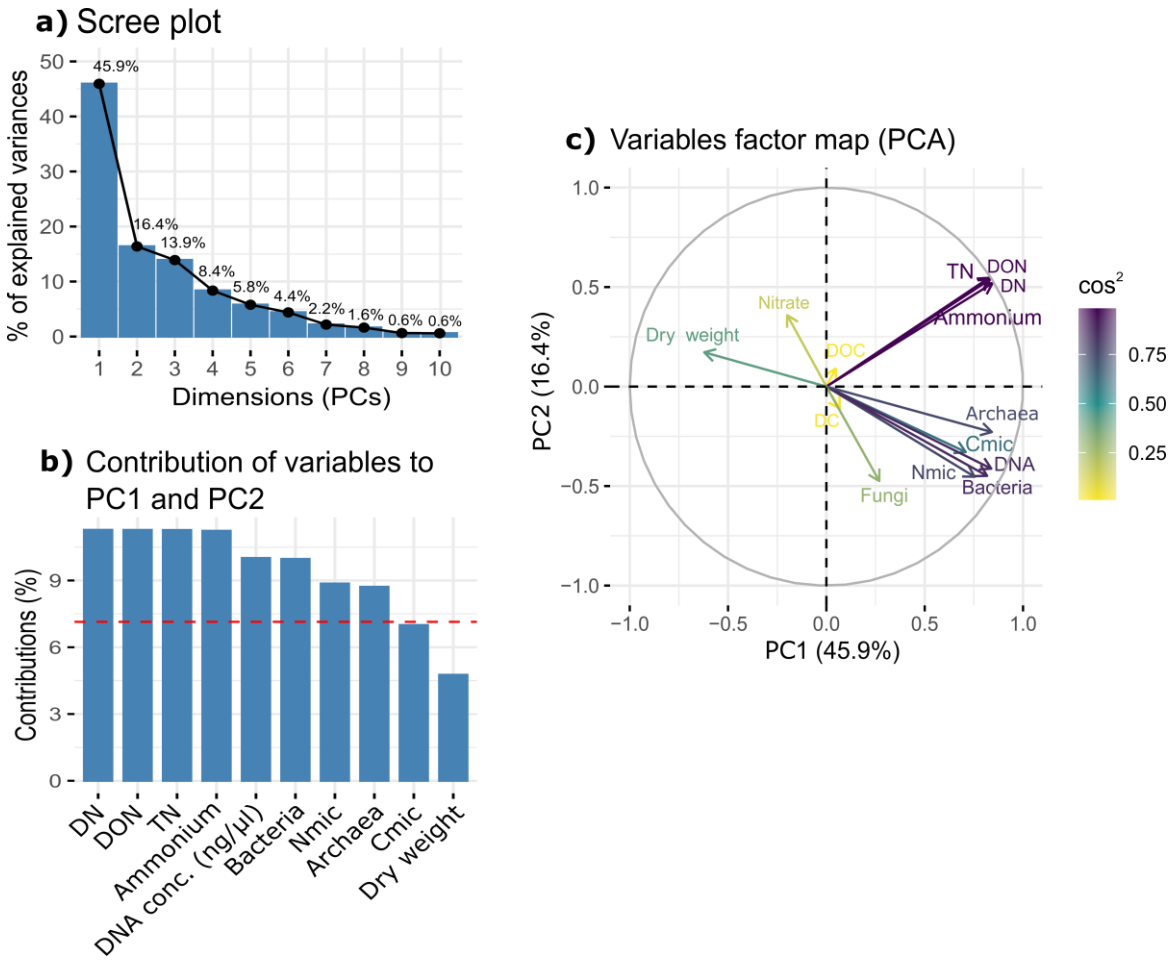


Figure 11. Principal components explaining the variance of chemical and biological parameters
a) PCA scree plot showing how much variation each PC captures from the data, b) bar plot showing how much each variable contributes to the variability contained in PC1 and PC2; the red line represents the expected average contribution, i.e. the expected contribution if contributions of all variables were uniform, c) PCA factor map with variables colored by the quality of their representation (cos^2, i.e. squared coordinates).

Figure 12 a shows the principal component analysis with the distribution of samples according to soil depth. Samples from the depth of 41-45 cm are clustered together on the left side of the ordination. Most samples from the first depth (1-5 cm) are clustered in the

lower quadrant on the right side. Samples from the second depth (16-20 cm) are scattered in all quadrants; mostly together with the samples from the third depth, and as a separate cluster in the upper right quadrant.

Figure 12b shows the principal component analysis with the distribution of samples according to reclamation age. All samples from RA0 are clustered on the left side of the ordination, alongside samples from the second and third depth of older reclamation ages. Samples from the initial reclamation stage are mostly clustered on the left side, while samples from later stages are mostly clustered on the right side, with the exception of the third-depth samples. Those samples closely resemble samples from the initial reclamation stage. Samples that are most separated from the rest, especially on the first principal component, are from the second depth of RA6. These samples are located in the upper right quadrant and are characterized by high nitrogen concentrations (Figure 11c).

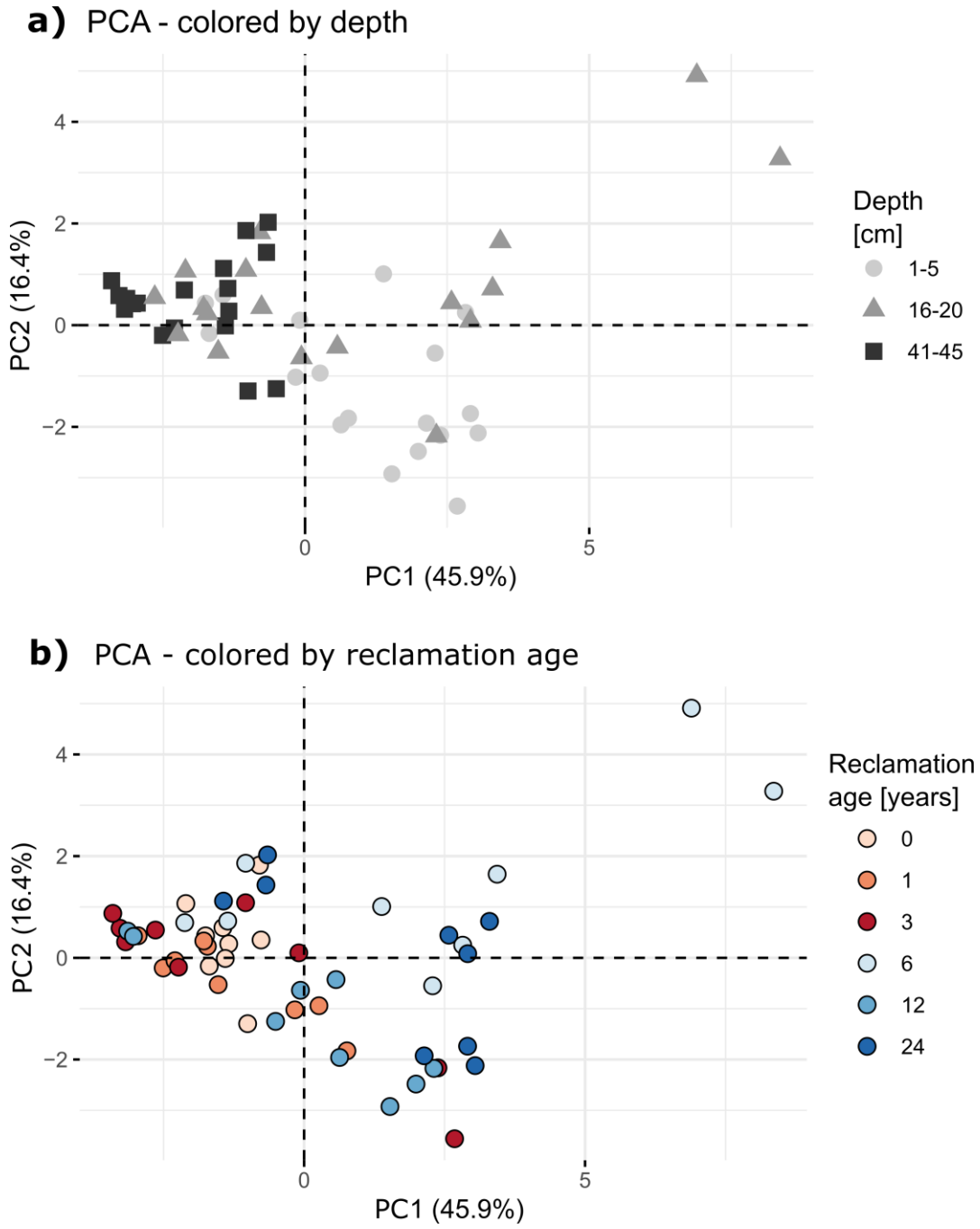


Figure 12. PCA of chemical and biological parameters
Principal component analysis with the distribution of samples according to soil depth (a) and reclamation age (b).

3.1.2 Growing season

3.1.2.1 Distribution of nutrients and microbial biomass with soil development and growing season

Dissolved carbon (DC) and nitrogen (DN) were measured in $CaCl_2$ extracts for all samples taken from six reclamation ages throughout a growing season in 2017.

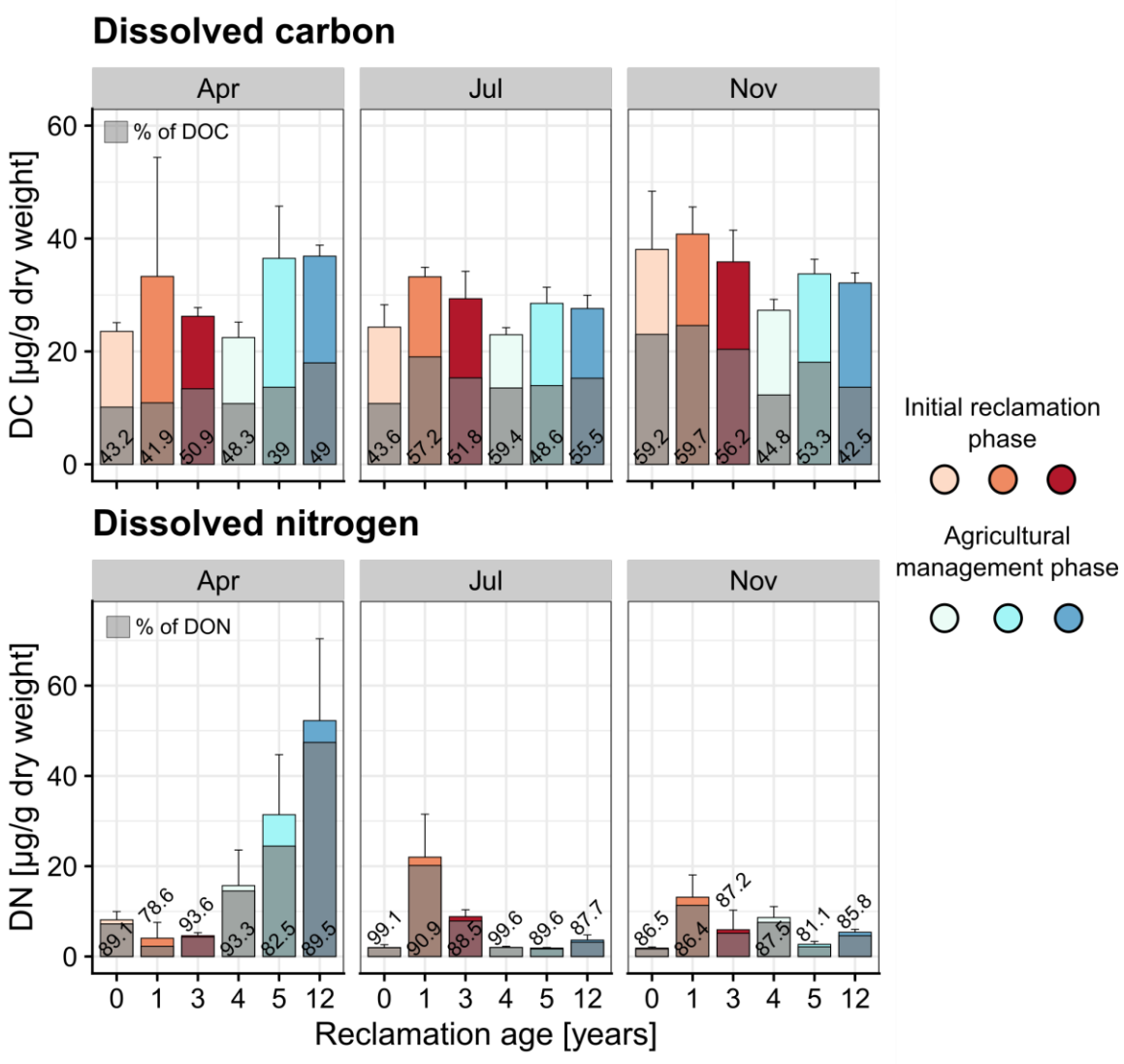


Figure 13. Concentrations of dissolved C and N during a growing season
Bar plots showing the concentrations of dissolved carbon (a) and nitrogen (b) and the percentage of organic carbon and nitrogen as measured after adding HCl to the $CaCl_2$ extracts. Bars are colored by the reclamation age and represent means of triplicates, and the greyed-out parts show the organic fraction within each sample.

Reclamation age and the interaction between reclamation age and month had a significant effect on DOC concentrations ($p = 0.01$ and $p = 0.009$, respectively), while the month alone

did not ($p = 0.057$). Average concentrations of DC and DOC (Figure 13) increased from April to November in samples from the initial reclamation phase. For these samples, the average amount of DOC was 45% of DC in April, 50% in July, and 58% in November. In samples from the agricultural management phase, DC and DOC remained relatively constant during the growing season. The average amount of DOC was 45, 54, and 46% in the three months, respectively.

Conversely, concentrations of DON were significantly affected by all three factors: reclamation age ($p = 0.001$), month ($p = 0.002$), and the interaction of the two ($p = 0.001$). In April, DN and DON concentrations were increasing from RA4 to RA12 and peaked in samples from RA12. In July, the concentrations were increasing only in samples from RA1 and RA3, and in November, they were highest in RA1. The average amount of DON was 87, 92, and 86% of DN in the initial reclamation phase in April, July, and November, respectively. Similar results were found in samples from the agricultural management phase with 88, 92, and 84% DON in the three months, respectively. However, the total concentrations of DON in the agricultural management phase were highest in April and decreased thereafter, while in the initial reclamation phase, DON concentrations were relatively stable. To further characterize the composition of nitrogen fractions, ammonium, nitrate, nitrite, and organic N concentrations were also measured from $CaCl_2$ extracts. A comparison of proportions of measured N fractions is illustrated in

Figure 14.

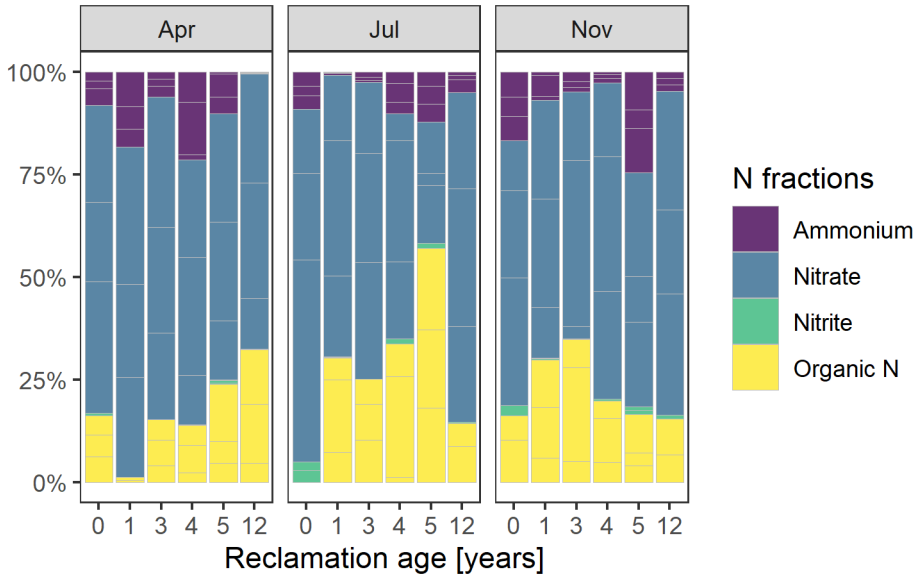


Figure 14. N fractions during a growing season
Proportions of nitrogen fractions (ammonium, nitrate, nitrite, and organic N) measured in April, July, and November. Shown are stacked bar plots with replicates visible in each stack.

Nitrate made the largest part of N in all samples except for RA5 in July where organic N is dominating. In April, there was a proportional increase of nitrate and organic N, and a decrease of ammonium in samples from the agricultural management phase (RA4, RA5, and RA12). In RA1 and RA3, there was an increase in organic N and a decrease in ammonium from April to November. Nitrite made a small proportion in all samples, the largest being in RA0 in July.

The absolute results of measured N fractions are shown in Table 5. Concentrations of ammonium were significantly affected by month ($p = 0.03$) and were generally lowest in July. The highest concentrations were found in April in RA4 and RA5. Nitrate was significantly affected by all three factors: reclamation age ($p = 0.01$), month ($p = 0.003$), and the interaction of the two ($p = 0.003$). In the fields of the agricultural management phase (RA4, RA5, RA12) nitrate was increasing in April and sharply decreased by July. In initial reclamation fields where alfalfa was growing (RA1 and RA3), nitrate was higher in July. In contrast, in RA0 nitrate was highest in April, before alfalfa was seeded. Nitrite concentrations were highest in April in RA4 and RA5, but generally low in all samples. Organic N concentrations were significantly affected by reclamation age and the interaction between age and month ($p = 0.003$ and $p = 0.001$, respectively). In RA1 and RA3 organic N was higher in July than in April, while the opposite was found in the agricultural

management phase. In November, organic N concentration remained highest in RA1 and RA3.

Table 5. Concentrations of ammonium, nitrate, nitrite, and organic N shown as means of triplicates and standard deviations.

Reclamation age [years]	Month	Ammonium [µg/g dw]	Nitrate [µg/g dw]	Nitrite [µg/g dw]	Organic N [µg/g dw]
0	April	0.63 ± 0.27	5.84 ± 1.5	0.05 ± 0.08	1.26 ± 0.18
1		0.27 ± 0.1	1.21 ± 0.26	0 ± 0	0.02 ± 0.02
3		0.25 ± 0.07	3.25 ± 0.66	0 ± 0	0.63 ± 0.13
4		3.35 ± 2.75	10.09 ± 4.07	0.5 ± 0.08	2.16 ± 1.04
5		3.02 ± 2.37	19.09 ± 5.54	0.32 ± 0.24	7.01 ± 4.5
12		0.27 ± 0.02	39.98 ± 15.42	0.05 ± 0.06	19.26 ± 9.54
0	July	0.12 ± 0.02	1.16 ± 0.72	0.07 ± 0.06	0 ± 0
1		0.2 ± 0.13	17.04 ± 6.71	0.09 ± 0.03	7.51 ± 4.93
3		0.23 ± 0.1	6.41 ± 1.58	0 ± 0	2.22 ± 0.58
4		0.14 ± 0.04	0.73 ± 0.46	0.02 ± 0.03	0.45 ± 0.48
5		0.15 ± 0.02	0.36 ± 0.22	0.01 ± 0.02	0.68 ± 0.03
12		0.17 ± 0.12	2.72 ± 0.6	0.01 ± 0.02	0.48 ± 0.45
0	November	0.19 ± 0.03	0.71 ± 0.31	0.03 ± 0.05	0.18 ± 0.17
1		1.08 ± 1.13	9.77 ± 3.49	0.06± 0.07	4.63 ± 1.65
3		0.32 ± 0.13	3.98 ± 3.77	0.01 ± 0.02	2.3 ± 1.94
4		0.25 ± 0.08	6.91 ± 2.01	0.03 ± 0.05	1.78 ± 0.97
5		0.48 ± 0.19	1.1 ± 0.42	0.04 ± 0.03	0.32 ± 0.2
12		0.24 ± 0	4.08 ± 0.79	0.05 ± 0.08	0.8 ± 0.71

The development of microbial biomass in the chronosequence and during the growing season was measured as microbial carbon and nitrogen and results are shown in Table 6.

Table 6. Dissolved organic and microbial carbon and nitrogen concentrations shown as means of triplicates and standard deviations.
Values with an asterisk mark samples that have been assessed from two replicates, due to an error during the measurement of one replicate.

Reclamation age [years]	Month	DOC [µg/g dw]	DON [µg/g dw]	C_{mic} [µg/g dw]	N_{mic} [µg/g dw]
0	April	10.12 ± 1.18	7.22 ± 1.47	228.7 ± 5.2	0.5 ± 0.8
	July	10.77 ± 3.15	1.95 ± 0.66	210.9 ± 50.7	2.1 ± 0.6
	November	19.25 ± 7.62*	1.57 ± 0.4*	294.83 ± 20.02*	3.82 ± 1.63*
1	April	10.89 ± 1.72	2.26 ± 0.33	400.2 ± 42.3	3.2 ± 1.1
	July	19.03 ± 2.22	20.16 ± 9.25	256.7 ± 29.2	14.8 ± 3.3
	November	24.59 ± 5.93	11.33 ± 4.15	281.8 ± 127.7	5.9 ± 9.3
3	April	13.38 ± 1.81	4.32 ± 0.64	386.8 ± 22.1	5.8 ± 1.8
	July	15.31 ± 3.49	7.88 ± 1.82	188.7 ± 28.1	6.8 ± 1.6
	November	20.37 ± 5.64	5.15 ± 3.72	274.4 ± 28.1	4.5 ± 0.5
4	April	10.76 ± 1.2	14.52 ± 7.12	285.9 ± 53.4	-2.1 ± 2.6
	July	13.52 ± 2.77	2.01 ± 0.66	168.5 ± 22.2	3.7 ± 0.5
	November	12.29 ± 2.32	7.56 ± 2.19	274.3 ± 40.5	2.5 ± 0.7
5	April	13.66 ± 2.92	24.46 ± 5.98	186.5 ± 38.3	1.5 ± 1.9
	July	13.94 ± 2.67	1.67 ± 0.12	135.9 ± 28.3	3 ± 0.2
	November	14.6 ± 2.28*	2.02 ± 0.28*	333.5 ± 43.22*	4.62 ± 0.44*
12	April	17.97 ± 1.6	47.41 ± 18.44	401.4 ± 79.7	-0.2 ± 7
	July	15.25 ± 0.12	3.19 ± 0.94	246.1 ± 92.1	10.4 ± 2.4
	November	13.65 ± 1.09	4.62 ± 0.55	332.8 ± 120.7	7.3 ± 3

High concentrations of C_{mic} were measured in April for RA1 and RA3; the two fields sampled not long after the alfalfa plant was cut and the cuttings were left on the fields (Table 1). The highest C_{mic} (401.4 ± 79.7) was found in RA12 in the same month, despite RA12 having been seeded with wheat at the same time as RA4 and RA5 fields. C_{mic} concentrations for all samples were lowest in July. C_{mic} was significantly affected by month ($p = 0.001$) and the interaction of month and reclamation age ($p = 0.001$).

Similar to DON, N_{mic} concentrations were significantly affected by all three factors: reclamation age ($p = 0.002$), month ($p = 0.001$), and the interaction of the two ($p = 0.015$). N_{mic} also increased from April to July in RA1 and RA3 and decreased by November. The highest concentration was found in RA1 in July. In RA0, the concentration increased from 0.5 ± 0.8 in April to 2.1 ± 0.6 in July, after the first alfalfa plant was seeded. As for the

samples from the agricultural management phase fields, the highest concentrations were found in July in RA12. Two of the samples from the same phase in April (RA4 and RA12) gave negative results, meaning that less nitrogen was found after the fumigation process. This was likely caused by an error during the fumigation or extraction, rendering these results untrustworthy.

3.1.2.2 Comparison of chronosequence and seasonal variation

Variations of all measured parameters along the chronosequence within each month are shown in Figure 15. The clustering heatmap showed differences between reclamation ages in each month.

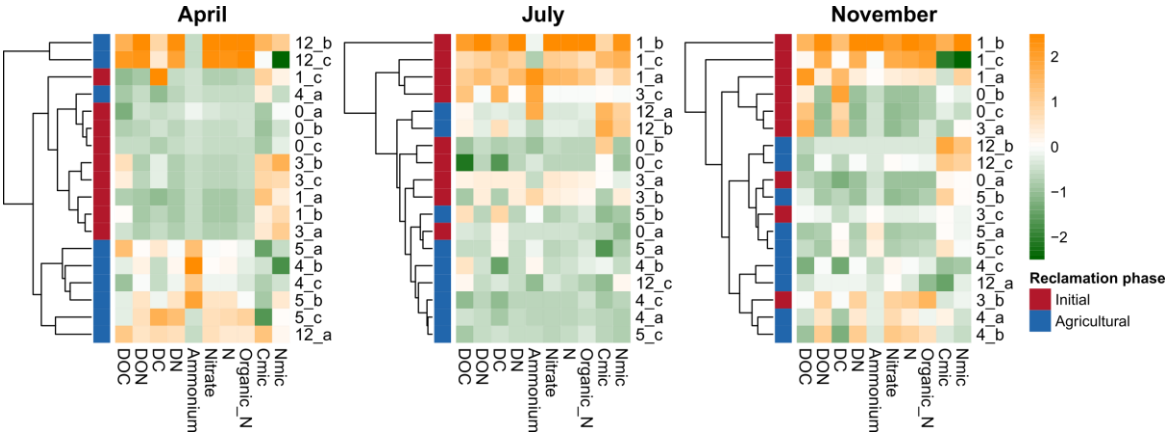


Figure 15. Comparison of measured parameters between reclamation ages within each month
Heatmaps showing all measured parameters normalized by the Z score transformation for comparison between reclamation ages within April, July, and November. Measured parameters are shown in columns and concentrations of each parameter are expressed as units of standard deviation from the mean in all samples. The mean of all samples is normalized to zero (white). When a sample is above the mean, it has a positive Z score (orange), and when it is below the mean, it has a negative Z score (green). The heatmaps include hierarchical clustering of rows, which shows similar samples in close positions.

In April, two samples from RA12 characterized by highest concentrations of DOC and majority of nitrogen fractions clustered separately from other samples. Majority of other samples from the agricultural management phase clustered at the bottom of the April heatmap. Samples from RA0 had low concentrations of all measured parameters, while majority of other samples from the initial reclamation phase clustered together and were characterized by high microbial biomass.

In July, majority of all measured parameters were highest in samples from RA1, followed by RA3. Samples from RA0 resembled samples from the agricultural management phase,

which were depleted of almost all nutrients compared to RA1 and RA3. Only in some samples of RA12 microbial biomass remained relatively higher compared to the rest.

Similar trends were observed in November, with RA1 again clustering at the top of the heatmap. However, samples from RA3 showed a higher variation and clustered separately. In samples from RA0, only carbon concentrations were high, with one sample showing microbial biomass similar to some samples from RA3 and RA5. Clustering at the bottom of the heatmap were samples from RA3 and RA4, characterized by high nitrogen concentrations. Similar to results observed in July, microbial biomass was relatively high in two samples from RA12.

Variations of all measured parameters between months within each reclamation age are shown in Figure 16.

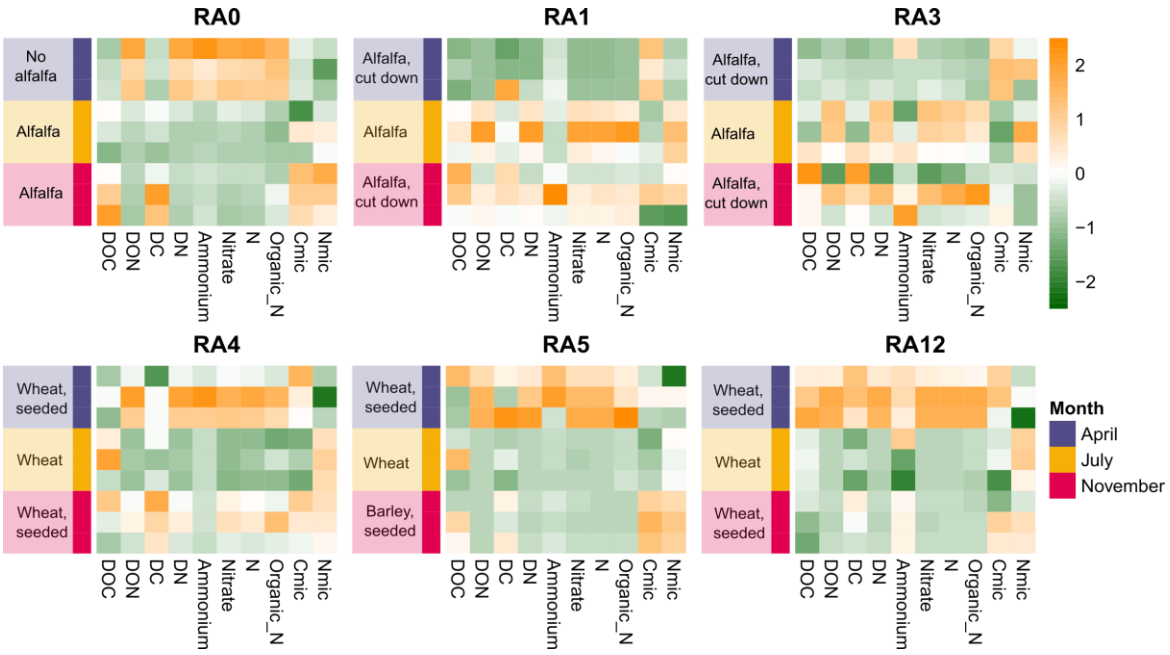


Figure 16. Comparison of measured parameters between months within each reclamation age
Heatmaps showing all measured parameters normalized by the Z score transformation for comparison between April, July, and November within each reclamation age. Measured parameters are shown in columns and concentrations of each parameter are expressed as units of standard deviation from the mean in all samples. The mean of all samples is normalized to zero (white). When a sample is above the mean, it has a positive Z score (orange), and when it is below the mean, it has a negative Z score (green).

In RA0, concentrations of nitrogen fractions were highest in April, before alfalfa was first seeded. In contrast, concentrations of carbon and microbial biomass were highest in

November. A similar trend was observed with samples from RA5, however, carbon concentrations were high in some of the samples of RA5 already in April, when wheat was freshly seeded. In contrast to these, in samples from RA1 and RA3 almost all measured carbon and nitrogen concentrations were lowest in April and increased in July and November. In RA1, C_{mic} was high in April and N_{mic} was high in July. Similar was observed with two out of three samples from RA3 and RA4, where C_{mic} was lowest in July.

In RA12, almost all measured parameters were highest in April, with N_{mic} remaining relatively high also in July and November and C_{mic} remaining relatively high in two of the samples in November.

3.2 Development of bacterial potential to produce exopolysaccharides and lipopolysaccharides and improve soil structure development

3.2.1 Extracellular polysaccharide concentrations

Extracellular polymeric substances (EPS) were extracted from the topsoil samples of six reclamation ages. The exopolysaccharide component of the EPS was measured using glucose as a standard and expressed as µg of glucose equivalent per g of dry soil. The protein component was measured using bovine serum and expressed as µg of bovine serum equivalent per g of dry soil. The results are shown in Figure 17.

Reclamation age had a significant effect on both EPS components ($p < 0.001$). The concentrations of both fractions increased with the start of reclamation in RA1 and were generally higher in the agricultural management phase compared to the initial reclamation phase.

For exopolysaccharides, the increase from RA0 to RA1 was almost twofold; from 120.21 ± 8.9 to 216.9 ± 28.1 µg g-1 dw soil, respectively. A slight decrease to 189.18 ± 18.7 µg g-1 dw soil was observed in RA3. Slight increases were visible during the agricultural management phase and the highest concentration was measured in soil samples from RA24, where it was significantly higher compared to other RAs, and reached 290.54 ± 11.2 µg g-1 dw soil.

A similar trend was observed with the concentrations of the proteinaceous component of EPS. The concentration doubled during the first year of the reclamation from 44.18 ± 5.7

in RA0 to 93.8 ± 3 µg g-1 dw soil in RA1. After 24 years of reclamation, the concentration was almost three times as high (125.54 ± 21.7 µg g-1 dw soil).

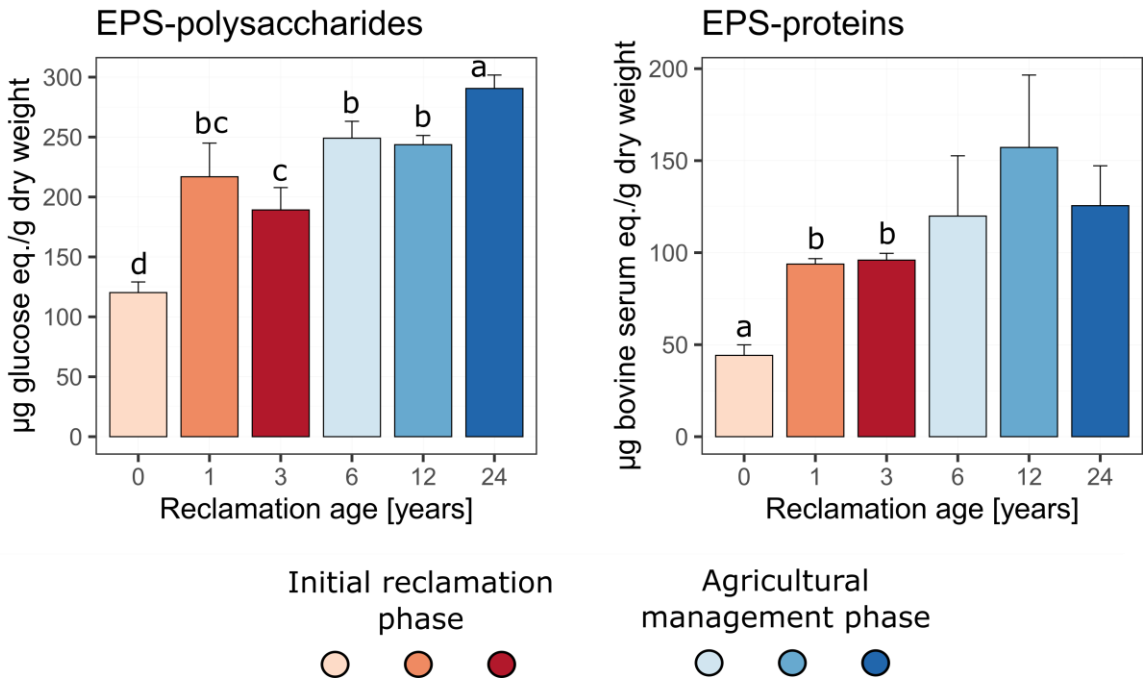


Figure 17. Concentrations of EPS-polysaccharides and EPS-proteins
Bar plots showing the exopolysaccharide and proteinaceous components of extracted EPS per g of dw soil. Bars are colored by the reclamation age and represent means of triplicates and error bars represent standard deviations. The letters above bars indicate the significant differences between samples after performing a post hoc test. EPS extraction was performed by Dr. Cordula Vogel, a cooperation partner at the Institute of Soil Science and Site Ecology, Technical University of Dresden.

3.2.2 Analysis of shotgun metagenomic data

The structure and potential function of microbiomes from the topsoil of six reclamation ages were studied in more detail by shotgun metagenomic sequencing. In total, the number of filtered sequencing reads obtained was 45.8 million. In the non-template control, only 212 reads were obtained, indicating very low contamination during DNA extraction and metagenomic library preparation (Figure 18). Those reads were excluded from further analysis. The number of reads per sample ranged from 1.6 to 3.7 million and the read length ranged from 292.23 to 296.96 bp. Taxonomic annotation per sample ranged from 69 to 79.3%.

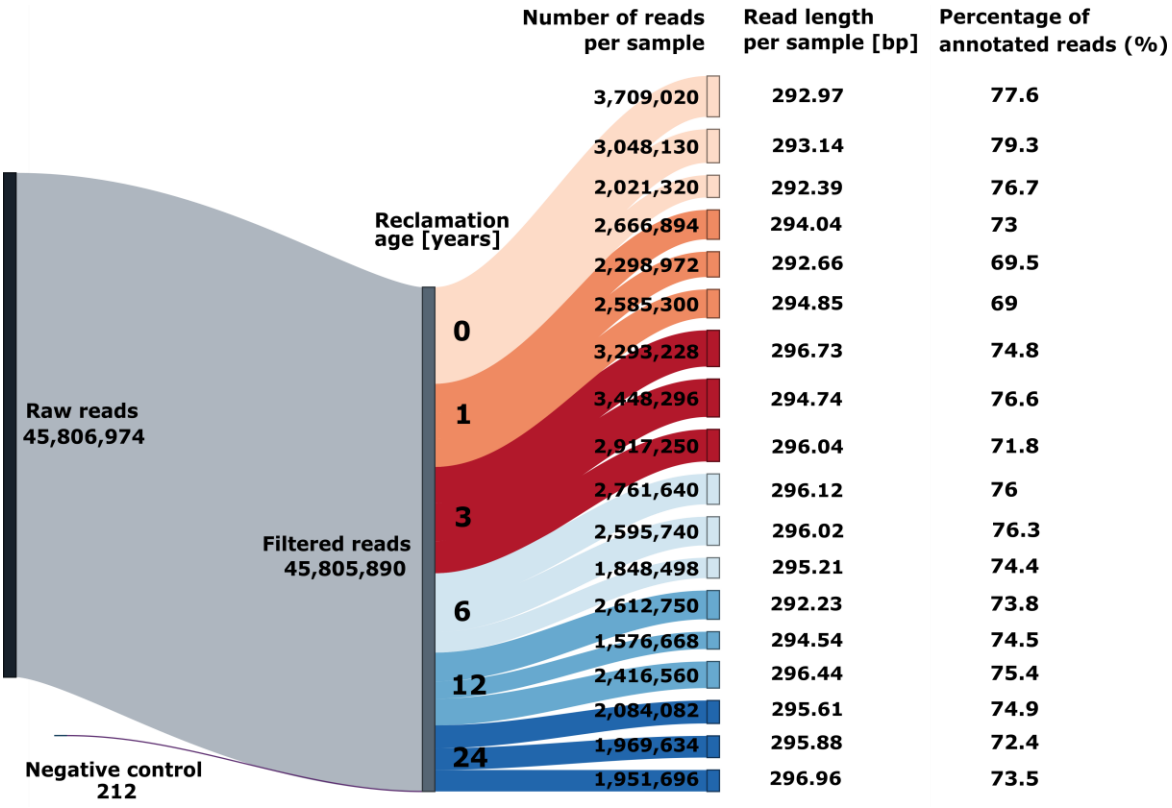
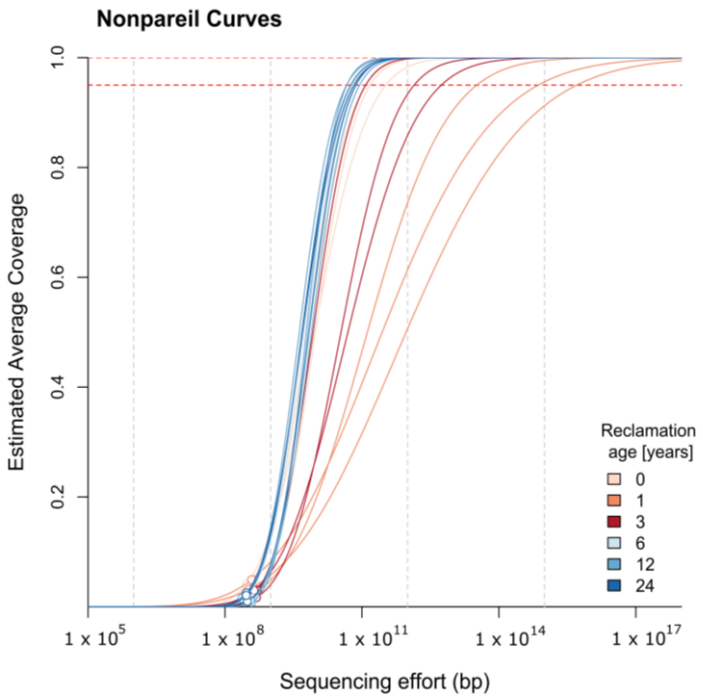


Figure 18. Sequencing summary
Average number of raw and filtered reads of all metagenomes, the number of filtered reads per sample, their length, and percentage of reads annotated to the NCBI non-redundant database. Samples from the initial management phase are colored in shades of red and samples from the agricultural management phase are colored in shades of blue.

The coverage of sequenced metagenomes was calculated by the Nonpareil algorithm and was based on the read redundancy values. The average coverage is presented as a function of sequencing effort in the form of Nonpareil curves (Figure 19). Coverage ranged from 0.8 to 5%. More diverse communities require larger sequencing efforts to achieve higher coverage. The highest diversity indices were found in samples from RA1 and RA3, which would require higher sequencing efforts. The highest coverage was obtained for samples from RA0 and it was generally higher in samples from the initial phase than agricultural management phase.



Reclamation age [years]	Nonpareil coverage [%]	Diversity index
0	3.3	22.9
0	5	22.51
0	4	22.81
1	2.3	25.35
1	3	27
1	4.9	25.86
3	1.7	24.15
3	2.9	22.79
3	3	24.42
6	1.5	22.59
6	1.5	22.54
6	1.6	22.35
12	0.8	22.49
12	1.4	22.11
12	1.2	22.72
24	1	22.6
24	2.5	22.27
24	2	22.25

Figure 19. Sequencing coverage and estimated diversity
Comparison of coverage estimations and diversity between samples using Nonpareil curves and calculated indices. Reclamation age had a significant effect on both the coverage and diversity indices with p values of 0.04 and 0.01, respectively.

3.2.3 Phylogenetic annotation

Filtered sequencing reads were annotated by blasting against the NCBI-nr database. Annotation resulted in 69.1 to 79.3% of reads annotated at the superkingdom level. The results are shown in Table 7.

Most reads, 63 to 75%, corresponded to Bacteria. Two to four percent corresponded to Eukaryotes, their percentage being higher in the initial reclamation phase than in the agricultural management phase. In contrast, archaeal reads were higher in the agricultural management phase. However, they were under one percent in all samples. Reads corresponding to Viruses did not exceed 0.15% in any of the samples. As this study is focused on bacteria, subsequent analyses were performed for bacterial reads only.

Table 7. Percentages of metagenomic reads assigned to Bacteria, Archaea, and Viruses, and unassigned reads in each sample.

Reclamation age [years]	Bacteria [%]	Archaea [%]	Eukaryotes [%]	Viruses [%]	Unassigned reads [%]
0	73.56	0.66	3.3	0.12	22.36
0	75.3	0.44	3.46	0.12	20.67
0	72.05	0.43	4.13	0.14	23.25
1	68.52	0.68	3.65	0.15	27
1	64.78	0.63	3.91	0.14	30.54
1	63.88	0.81	4.27	0.14	30.9
3	70.75	0.51	3.42	0.14	25.2
3	72.57	0.58	3.37	0.15	23.33
3	67.66	0.53	3.5	0.13	28.17
6	72.24	0.81	2.9	0.12	23.97
6	72.2	0.73	3.25	0.13	23.7
6	70.45	0.95	2.87	0.13	25.59
12	69.84	0.89	2.92	0.1	26.23
12	70.66	0.87	2.88	0.09	25.49
12	71.44	0.84	3	0.1	24.61
24	71.01	0.93	2.88	0.12	25.1
24	68.62	0.84	2.8	0.1	27.63
24	69.53	0.91	2.95	0.12	26.49

An NMDS analysis was performed based on the bacterial reads (Figure 20) to compare community structures in all reclamation ages. Samples from different reclamation ages were dominantly separated along the x-axis. Samples from RA0 clustered separately on the left side of the plot, with the largest distance between replicates. Samples from other RA's were ordinated on the right side of the plot, with RA24 clustering furthest away from RA0. Replicates from RA6 were positioned closest to each other.

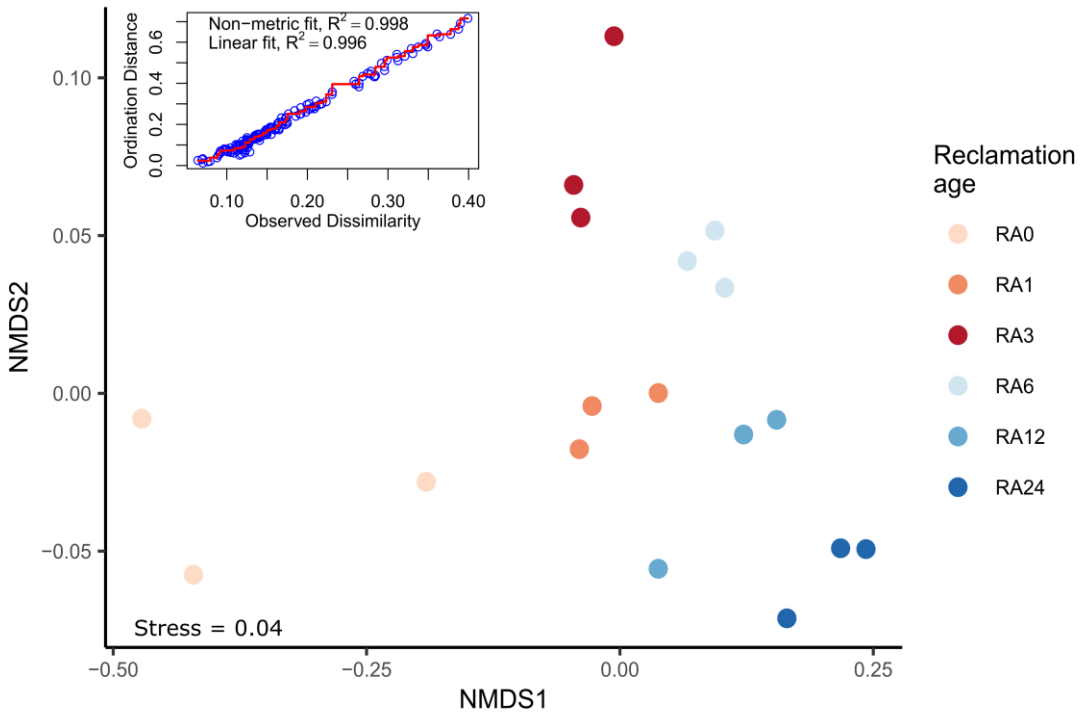


Figure 20. NMDS analysis of bacterial reads based on the Bray-Curtis dissimilarity indices
Shepard (stress) plot in the upper left corner shows goodness-of-fit statistics for the calculated NMDS, where Bray-Curtis dissimilarity indices are plotted against the NMDS ordination distances. Stress is calculated from the residuals (blue circles) around the regression (red) line, and ranges from zero to one. Stress that is close to zero indicates a high goodness-of-fit for the NMDS.

Proteobacteria was the most abundant prokaryotic phylum in all samples (Figure 21), but became less abundant during the reclamation process, as the number of low-abundant phyla increased. The same was observed with Actinobacteria and Bacteroidetes. In contrast, Acidobacteria increased in abundance. A confirmation of this taxonomic assignment of all reads was performed by comparing it with the assignment of extracted 16S rRNA gene using Ribotagger (Xie et al., 2016) and the SILVA database. The relative abundance of extracted 16S rRNA genes was below 0.5% of all filtered reads. Nevertheless, a phylogenetic analysis of reads assigned to 16S rRNA showed similar trends for most of the dominant phyla compared with results obtained with a phylogenetic analysis of all assigned reads (Figure 12b).

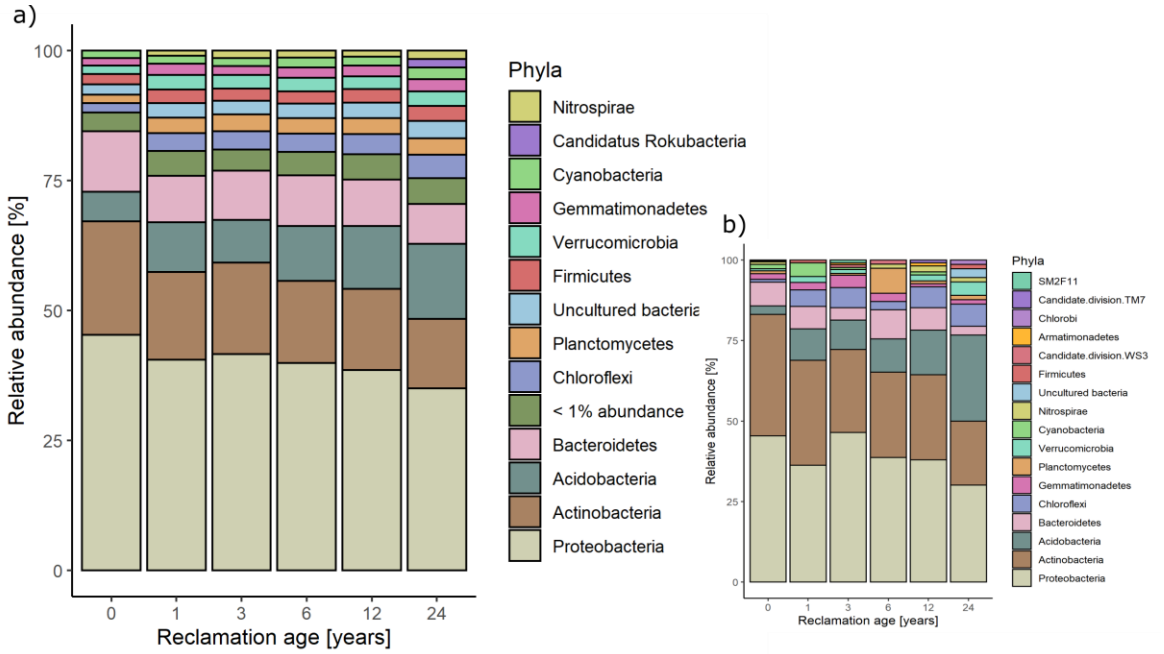


Figure 21. Taxonomic profiles of bacterial communities on the Phylum level
a) Relative abundance of top 10 most abundant Phyla found in each reclamation age; b) Relative abundance of reads assigned to Phyla using Ribotagger assignment of extracted 16S rRNA genes.

Within Proteobacteria, the most abundant families were *Sphingomonadaceae*, *Comamonadaceae*, and *Bradyrhizobiaceae (*Figure 22*)*. *Sphingomonadaceae* had a significantly higher relative abundance in RA0 compared to all other RAs, and it represented 7.5% of all bacterial reads in this RA. That number decreased to 3.9% by RA1, and to 1.9% by RA24. A similar, but less pronounced trend was observed with several other families, such as *Comamonadaceae*, *Xanthomonadaceae*, and *Oxalobacteraceae*.

In contrast, families such as *Bradyrhizobiaceae, Rhizobiaceae, Burkholderiaceae,* and *Hyphomicrobiaceae* increased in relative abundance after RA0. The highest increase was observed with *Bradyrhizobiaceae* and *Burkholderiaceae*, which had a significant difference between RA0 and RA1 (2.15% to 2.86% and 1.21% to 1.35%, respectively). Relative abundances of these families, with the exception of *Bukrholderiaceae*, decreased by RA24.

Within Actinobacteria*,* the most abundant families were *Streptomycetaceae* and *Micrococcaceae.* The relative abundance of *Streptomycetaceae* remained stable throughout the reclamation, while *Micrococcaeceae* sharply decreased from 2.72% in RA0 to 0.82% in RA1.

Among Bacteroidetes, the most abundant families decreased in abundance after RA0 (*Chitinophagaceae*, *Sphingobacteriaceae*, *Flavobacteriaceae).* This was most evident for *Sphingobacteriaceae,* which followed a trend similar to *Micrococcaceae* and decreased from 2.23% in RA0 to 0.83% in RA1.

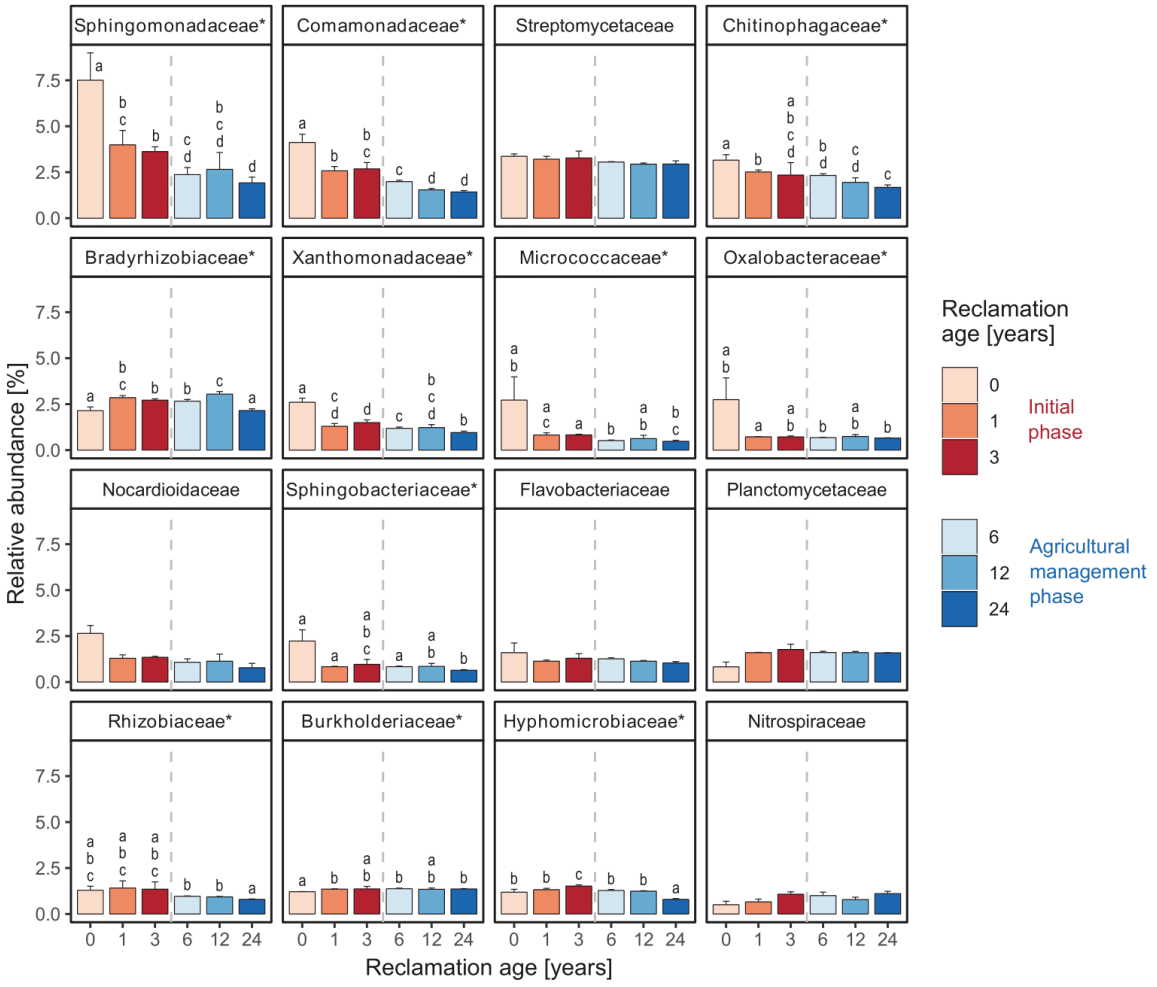


Figure 22. Top 12 most abundant bacterial families found in the initial and agricultural management phase
Abundances of families marked with an asterisk were significantly different between reclamation ages, and letters above bars indicate post-hoc results.

Figure 23 shows the relative abundances of bacteria that were not assigned to the family level. In RA0, 23.7% of bacterial reads were not classified on the family level, while in RA24 that number increased to 41.9%. The difference between RA0 and RA24 was particularly visible in the low abundant bacterial reads (< 1%) and in the abundance of bacterial reads assigned to Acidobacteria.

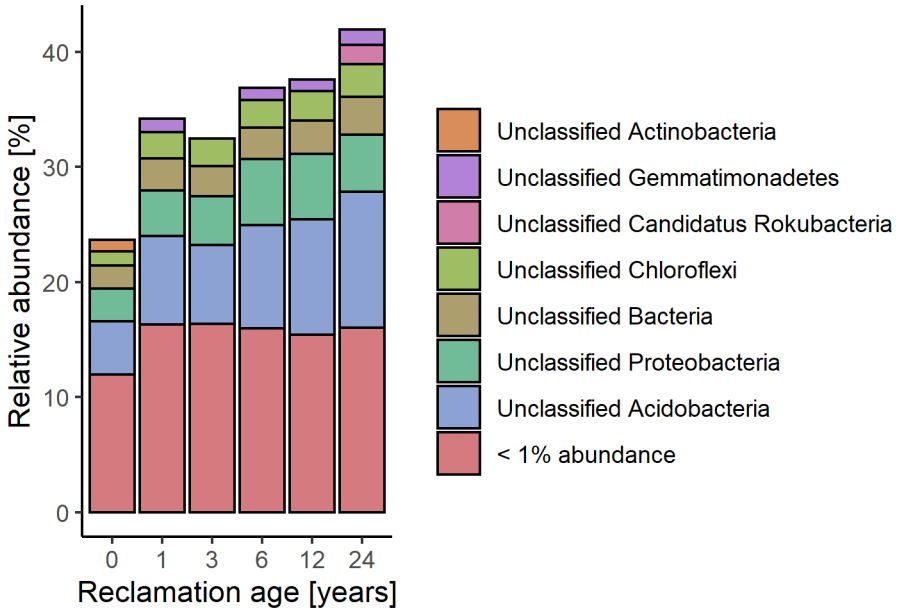


Figure 23. Relative abundances of bacterial reads that were not classified on the family level

The genera composition of the most abundant families is shown in Figure 24. In most cases, reads were largely assigned to one to three genera per family.

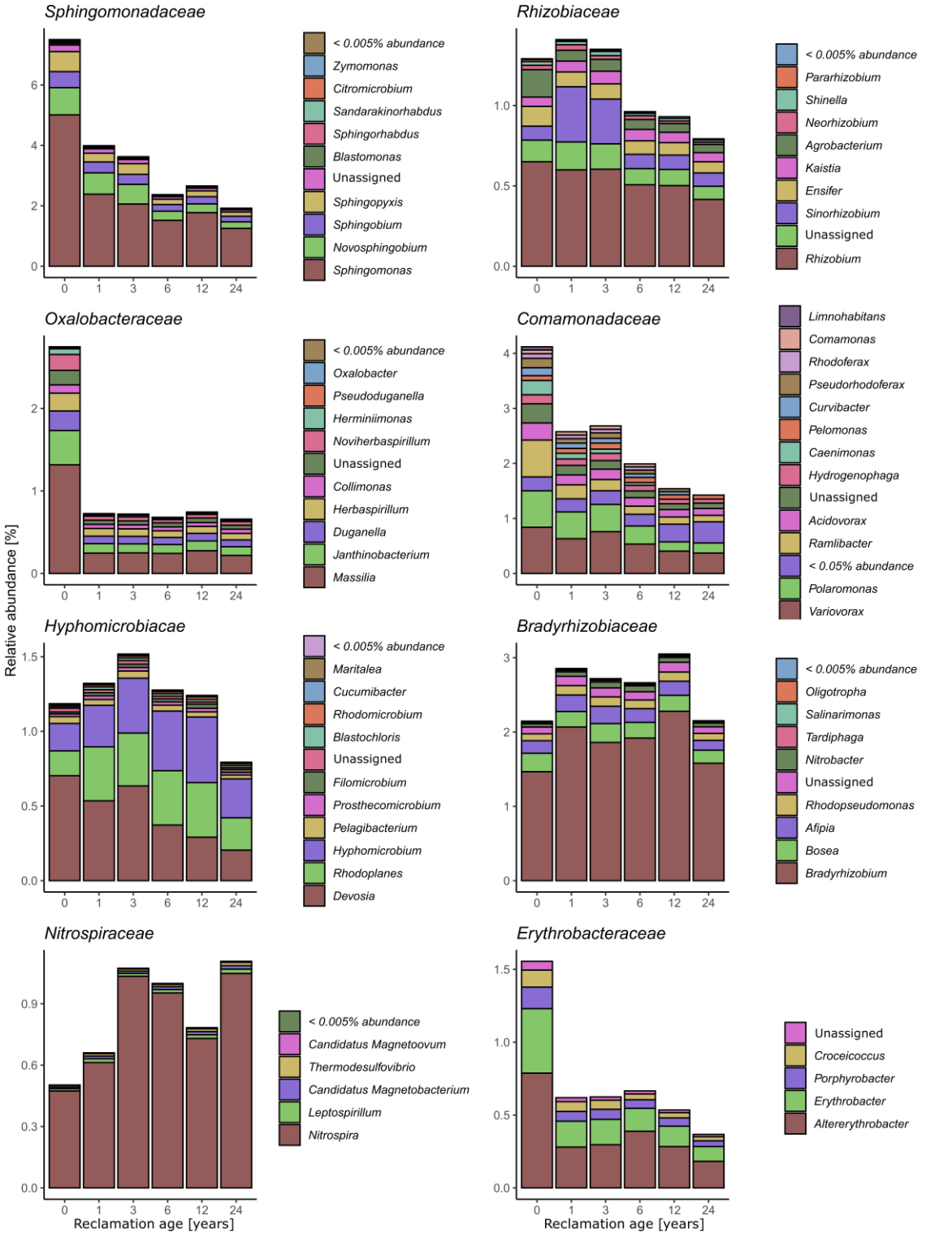


Figure 24. Genera composition of reads assigned to most abundant families

3.2.4 Functional annotation

Functional annotation focused on exopolysaccharide and lipopolysaccharide genes was performed on reads assigned to Bacteria. Genes specific for exopolysaccharide and lipopolysaccharide production and excretion were targeted using hidden Markov models (HMMs) based on data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Table 3). Further analysis was performed on 10 out of the 14 genes, as reads assigned to *algE*, *espG*, *espA*, and *wcaK/amsJ* had very low abundances of less than 10 reads per sample.

Relative abundances of all 10 analysed genes ranged from 0.032% in RA24 to 0.047% in RA6. The two most abundant genes were *lptF* and *lptG*, encoding for parts of the permease that exports lipopolysaccharides (Figure 25). The third most abundant gene was *wza*, encoding for the polysaccharide export outer membrane protein. Between reclamation ages, the highest abundance of these three genes was observed in RA6, followed by RA1. In RA6, their mean relative abundances were as follows: 0.013% for *lptF*, 0.012% for *lptG*, and 0.009% for *wza*. The peak in abundance in RA6 was also observed for *wcaF* and *kpsE.* In fact, seven out of the ten analysed genes showed the highest abundance in RA6. Among the genes already mentioned, these also include *sacB* and *wzt*. Additionally, relative abundances of all mentioned genes showed an increase from RA0 to RA1.

With the exception of *wza*, all genes had the lowest relative abundance in RA24. Moreover, all analysed lipopolysaccharide genes and two exopolysaccharide genes (*kpsE* and *wcaF*) showed lowest abundances in RA0 and RA24.

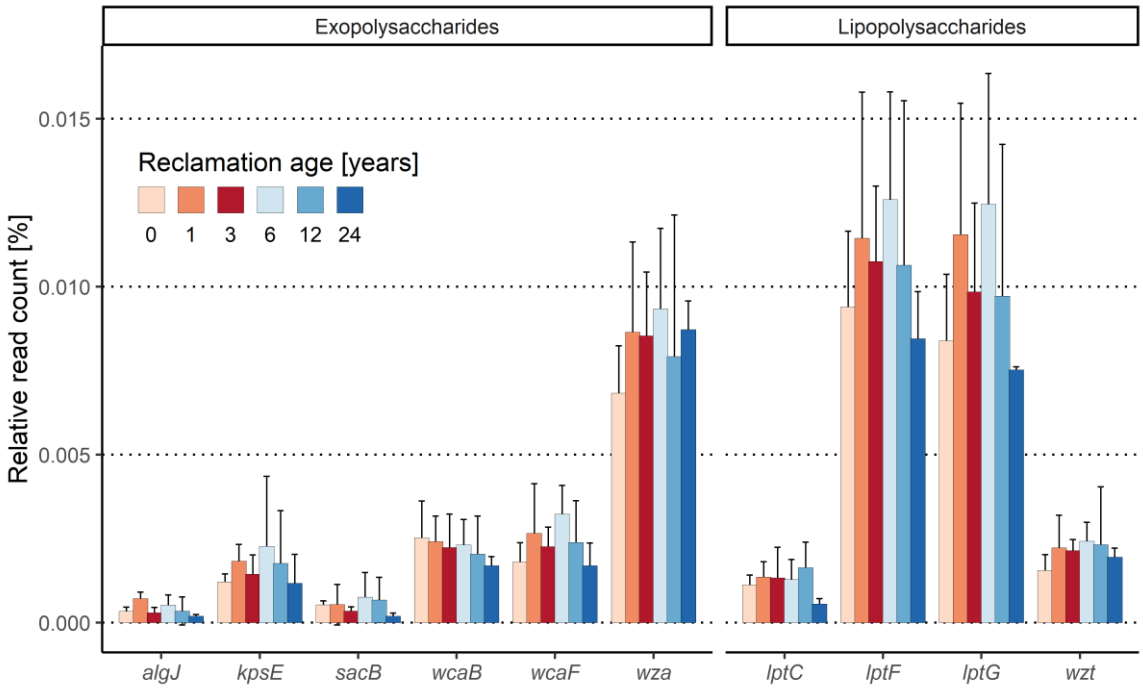


Figure 25. Relative read count of the ten analyzed genes involved in exopolysaccharide and lipopolysaccharide biosynthesis
Bars are colored by the reclamation age and represent means of triplicates and error bars represent standard deviations. Differences between reclamation ages were tested with the robust ANOVA with trimmed means, which showed no significant differences.

To test whether the same trend could be observed with actual amounts of genes in reclamation soil samples, we roughly estimated their absolute abundances using the 16S rRNA gene copy numbers obtained with qPCR. Results and a description of this approach are shown in Figure 26.

In contrast to gene relative abundances, estimates of their absolute abundances followed the trend observed with 16S rRNA gene copy results (Figure 10): the most abundant genes, namely *wza*, *lptF*, *lptG*, *wzt*, *wcaB*, and *wcaF* showed the strongest increase in soils from RA1, a slight decrease in RA6 and increases in RA12 and especially RA24.

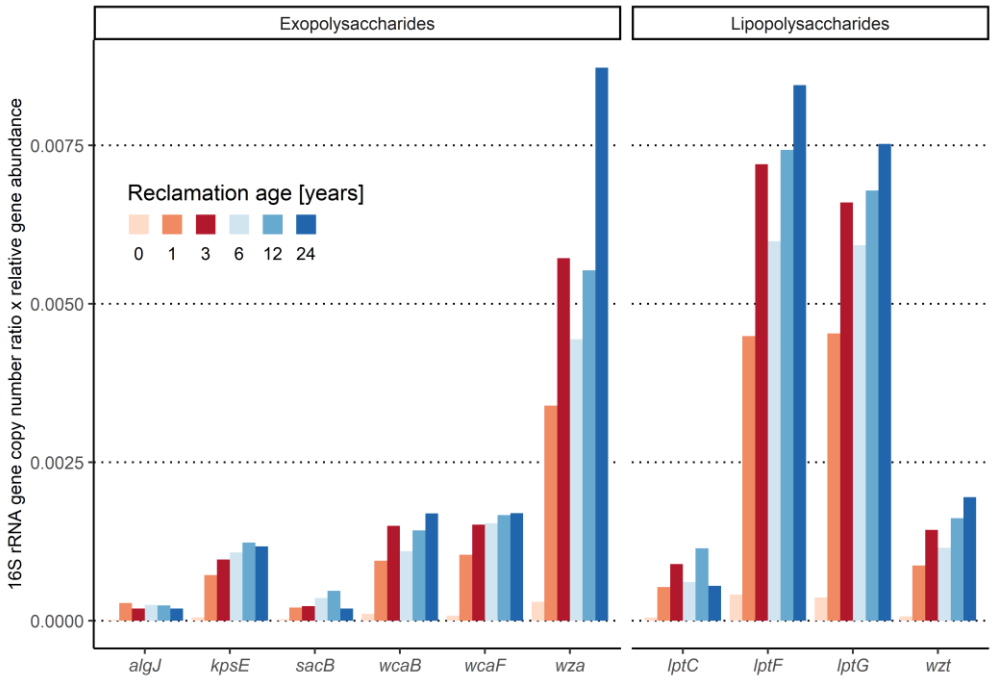


Figure 26. Estimation of genes' absolute abundances
First, the mean values of 16S rRNA gene copy numbers (qPCR) in each reclamation age were divided by the highest mean value, which was found in RA24. Then, the resulting ratios were multiplied by the mean relative abundance of each exo- and lipopolysaccharide gene analyzed from the metagenome. This resulted in a rough estimation of the absolute gene abundance in each reclamation age compared to RA24.

3.2.4.1 Connection of functions to taxonomy

Reads assigned to analyzed exopolysaccharide and lipopolysaccharide genes were blasted against the NCBI-nr database to obtain their phylogenetic annotation. In total, genes were assigned to 245 different bacterial families. A more detailed analysis was performed on the most abundant exopolysaccharide and lipopolysaccharide carriers in each reclamation age. To obtain these, 10 families carrying the most genes in each RA were first filtered. When put together, these amounted to 26 distinct families harboring the most genes in all reclamation ages (Figure 27).

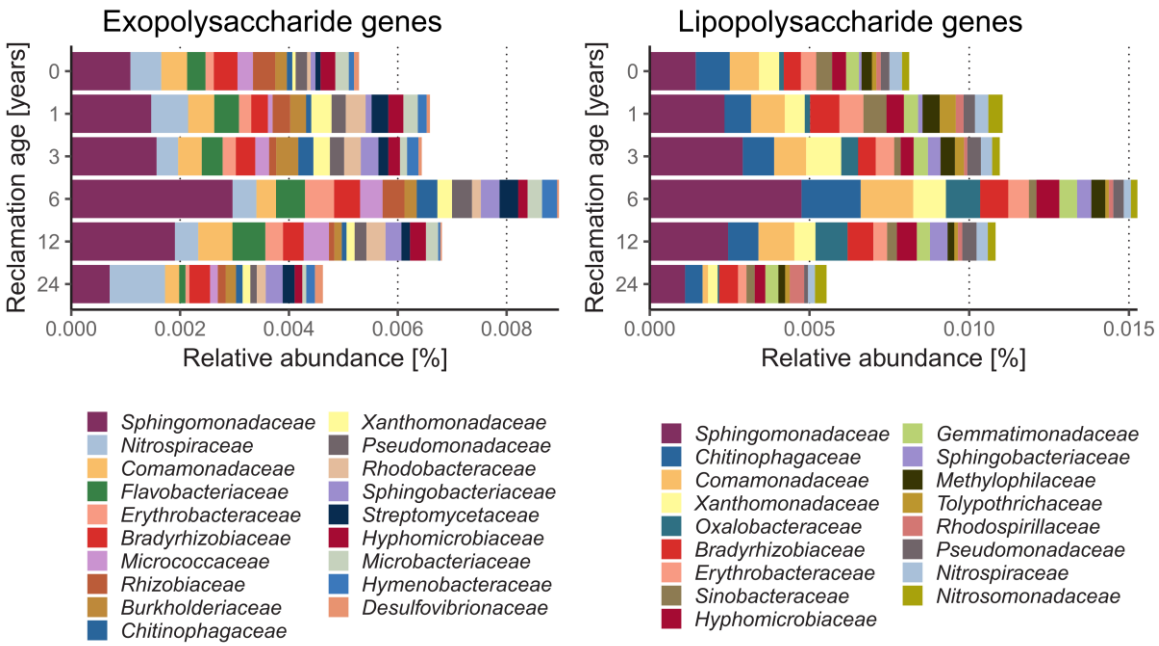


Figure 27. Phylogenetic annotation of potential exo- and lipopolysaccharide producers
Shown are distributions of gene copies among 10 bacterial families that harbored most genes in each reclamation age. Relative abundances were obtained by summing up mean relative abundances of exo- and lipopolysaccharide genes and dividing them by the number of genes: six for exo- and four for lipopolysaccharides.

Out of the 26 families, 16 were involved in both exo- and lipopolysaccharide processes. Families that were found to carry only exopolysaccharide genes are *Flavobacteraceae*, *Micrococcaceae*, *Rhodobacteraceae*, *Desulfovibrionaceae*, *Microbacteraceae*, *Streptomycetaceae*, and *Hymenobacteraceae*. On the other hand, families that were found to carry only lipopolysaccharide genes are *Gemmatimonadaceae*, *Tolypothrichaceae*, *Sinobacteraceae*, *Rhodospirillaceae*, *Nitrosomonadaceae*, and *Methylophilaceae*.

Sphingomonadaceae was found to be the most prominent carrier of both analyzed gene categories in almost all reclamation ages. Only in RA24, that role was taken by *Nitrospiraceae* for exopolysaccharide genes. *Nitrospiraceae* were among the most abundant potential exopolysaccharide producers in other RAs as well, together with *Comamonadaceae* and *Flavobacteraceae*.

For lipopolysaccharide genes, the second and third families carrying the most genes differed between reclamation ages. Second most abundant family for these genes in RA0, RA6 and RA24 was *Chitinophagaceae*. In RA1 and RA12 it was *Comamonadaceae*, and in RA3 *Xantomonadaceae*. *Nitrospiraceae* was among the families carrying a low

abundance of lipopolysaccharide genes. For families with a low abundance of both exo- and lipopolysaccharide genes, differences between reclamation ages can be observed as well.

A detailed analysis of the eight most abundant genes distributed across 10 prominent bacterial families is shown in Figure 28.

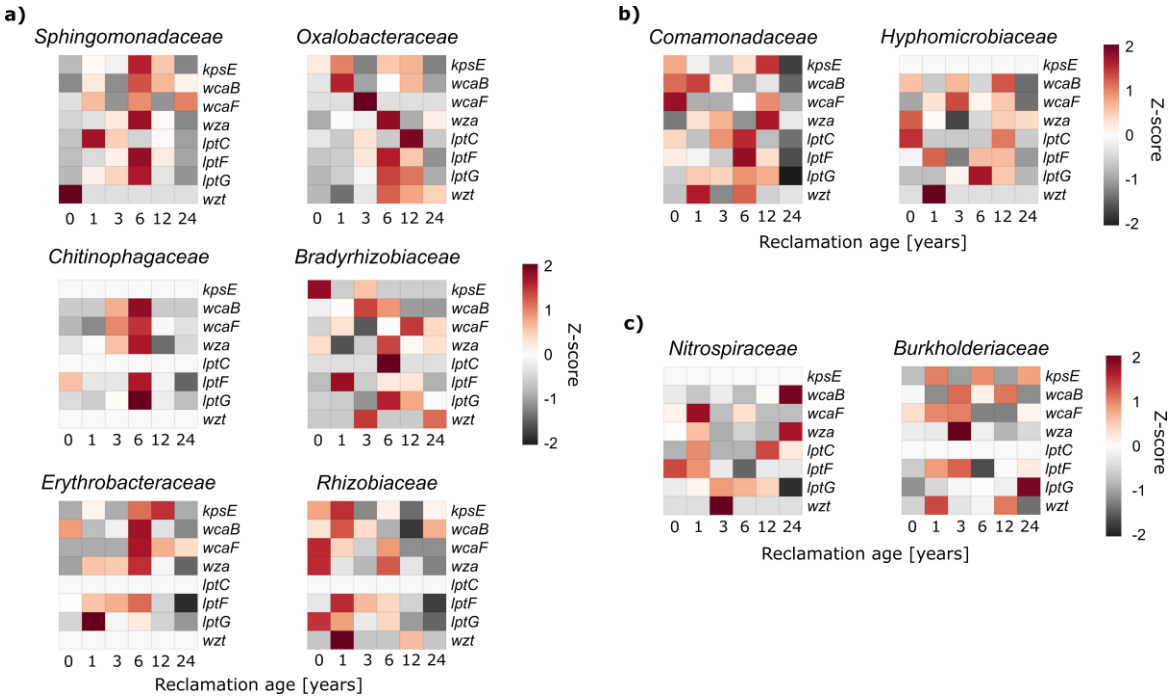


Figure 28. Relative gene abundances of eight genes with the highest abundance harbored by particular bacterial families
Genes are presented in rows and the amount of gene per sample is expressed as a unit of standard deviation from the mean of gene abundance in all samples (ie. normalized by the Z score transformation). The mean of all samples is normalized to zero (white). When a sample is above the mean, it has a positive Z score (red), and when it is below the mean, it has a negative Z score (gray). a: Families with most exopolysaccharide/lipopolysaccharide genes assigned in RAs 1 and/or 6. b: Families with a decrease of exopolysaccharide, but an increase of lipopolysaccharide genes assigned in RA6. c: Families with a decrease of exo- and lipopolysaccharide genes assigned in RA6.

For most of the families, the highest relative abundance of genes driving exopolysaccharide and lipopolysaccharide synthesis was found in RA6, namely *Sphingomonadaceae*, *Oxalobacteraceae*, *Chitinophagacaeae*, *Bradyrhizobiacae*, and *Erythrobacteracae* (Figure 28a). For *Sphingomonadaceae*, this is evident in three of the exopolysaccharide (capsular polysaccharide export system permease – *kpsE*, colanic acid biosynthesis acetyltransferase – *wcaB,* and polysaccharide export outer membrane protein -

wza) and two of the lipopolysaccharide (parts of the LptBFGC lipopolysaccharide export complex permease - *lpF* and *lptG*) genes. Similar trends were observed for the other four families. However, no genes coding for another part of the LptBFGC permease – *lptC,* as well as for the lipopolysaccharide transport system ATP-binding protein - *wzt* were assigned to *Chitinophagacaeae* and *Erythrobacteracae*, and the latter family had a peak of *kpsE* genes in RA12. *Oxalobacteraceae* harbored some of the analyzed genes in samples from the initial reclamation phase, while *wza*, *lptF*, *lptG*, and *wzt* all peaked in RA6. For *Rhizobiaceae*, seven out of the ten genes showed the highest relative abundance in RAs 0 and 1 (*kpsE*, *wcaB*, *wcaF*, *wza*, *lptF*, *lptG*, and *wzt*).

A decrease in relative abundances of exopolysaccharide and an increase of lipopolysaccharide genes following the transition to the agricultural management phase in RA6 was observed for *Comamonadaceae* and *Hyphomicrobiaceae* (Figure 27). Upon closer inspection, abundances of *lptC*, *lptF*, and *lptG* peaked in *Comamonadaceae* in RA6, while for *Hyphomicrobiaceae* this increase was mostly attributed to *lpG* (Figure 28b). For both of these families, there was a decrease in the abundance of analyzed genes in the later stage of the reclamation process at RA24.

For some families, such as *Nitrospiraceae* and *Burkoholderiaceae*, a decrease of gene relative abundances was observed in RA6 (Figure 28c). For *Nitrospiraceae*, *wcaB* and *wza* genes peaked at RA24, while for *Burkoholderiaceae* reads assigned to these genes, along with *wcaF* and *lptF,* peaked at RA3.

3.2.4.2 Comparison of potential exo- and lipopolysaccharide producers with general taxonomy

A comparison of ten highly abundant bacterial families between their general abundance and the abundance of analyzed exopolysaccharide and lipopolysaccharide genes they harbored is shown in Figure 29. The relative abundances of many of the highly abundant families were decreasing in the later stages of the reclamation process. At the same time, the amount of exopolysaccharide and lipopolysaccharide genes they carry peaked at RA6. This trend is evident for *Sphingomonadaceae*, *Comamonadaceae, Oxalobacteraceae*, *Chitinophagacaeae*, and *Erythrobacteracae.*

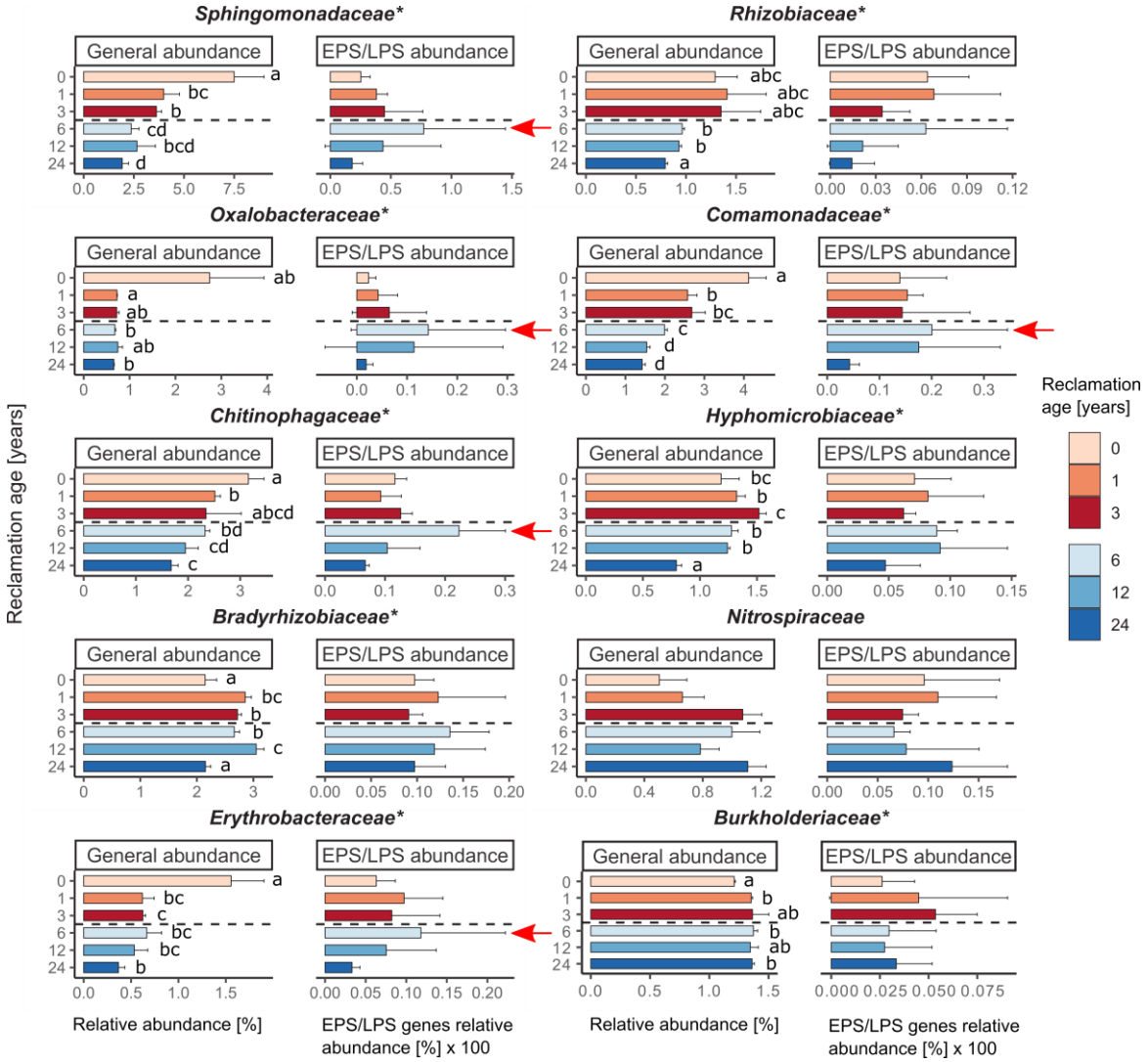


Figure 29. Comparison of the relative abundance of ten highly abundant families with their potential to produce exopolysaccharides and lipopolysaccharides
Asterisks indicate families that were found to have significantly different relative abundances between reclamation ages and letters next to the bars represent pairwise comparisons after the Benjamini-Hochberg correction. Relative abundances of all analyzed exopolysaccharide and lipopolysaccharide genes are shown as percentages multiplied by 100. Red arrows point to families with a decrease in general abundance and an increase in gene abundance in RA6.

The abundance of *Sphingomonadaceae* decreased by a factor of almost four from RA0 to RA24, while for *Comamonadaceae* and *Chitinophagacaeae* the decrease between the same RAs was by a factor of almost three and two, respectively. For *Oxalobacteraceae* and *Erythrobacteraceae*, a sharp decrease was observed already at RA1.

Bradyrhizobiaceae and *Rhizobiaceae* carried the most exopolysaccharide/lipopolysaccharide genes at RA1 and RA6, while the overall relative

abundance of *Bradyrhizobiaceae* was highest in RA12 and of *Rhizobiaceae* in RA1. During the agricultural management phase, the overall abundance of *Rhizobiaceae* decreased, similar to *Hyphomicrobiaceae*. The overall abundance of *Burkholderiaceae* increased and stabilized after RA0, while the exopolysaccharide and lipopolysaccharide genes assigned to this family peaked at RA3 and decreased in the later stages. Among the 10 families, *Nitrospiraceae* was the only one that increased both in the overall abundance and in the amount of exopolysaccharide/lipopolysaccharide genes it carried at RA24.

Despite the trends of decrease described for most of the highly abundant families, the absolute abundance of bacteria (Figure 10) and the amount of EPS-polysaccharides (Figure 17) measured in the soil samples showed an increase after RA0 and were highest in RA24.

4 Discussion

4.1 Use of a chronosequence as a substitute for time

Despite the advantages of using chronosequences in research of long-lasting processes and the useful information they provide about soil formation, their value has been questioned in the literature (Johnson and Miyanishi, 2008; Walker et al., 2010). For one, chronosequences represent only a partial historical record since not all events are recorded within the soil profile, which may lead to overemphasis on resistant soil components and processes as opposed to rapidly changing properties (Huggett, 1998). In addition, some of the basic principles of chronosequences are often difficult to meet or determine, e.g. whether soils differ only with respect to time while other soil-forming factors are indeed constant (Walker et al., 2010), whether all sites are following the same biodiversity trajectory (De Palma et al., 2018), or whether characteristics of the parent material they originate from are constant (Zwanzig et al., 2021).

In the present study, the type of soils in each reclamation age was very similar in soil texture and $CaCO_3$ concentrations, as well as pH, indicating that the parent material used for reclamation was the same in all time points, fulfilling an important requirement for the chronosequence approach (**Publication II**). In addition, unlike hundreds-of-years-old (or older) natural chronosequences, the history of the chosen reclamation site is well known and does not need to be assumed, since it is managed by the mining company and farmers. However, microbial communities in soil are sensitive to numerous factors, such as biogeochemical and vegetation changes (e.g. root exudates), which are usually highly dynamic and operate over short timescales (Bardgett et al., 2005b). Thus, a more granular approach with samples from each year of reclamation sampled multiple times throughout a year, or even followed for several years, would be beneficial to precisely distinguish between the effects of reclamation age and management versus other possible confounding factors. Nevertheless, since the same established reclamation strategy was applied to all sites in this study and since the sites are located in a relatively small geographical region, we believe that age and management practices outweigh other possible factors. In addition, studies performed in an independent chronosequence from another open-cast coal mine in West Germany showed similar patterns of succession of bacterial and fungal communities following changes in management (Roy et al., 2017; Schmid et al., 2020).

4.2 Microbial and nutrient dynamics along the chronosequence

4.2.1 Mineral phase driving soil structure formation in the organic matter-poor initial soil substrate

Initial soil substrate (RA0) was characterized by low nutrient content and microbial biomass, and no horizontal stratification. In that sense, the topsoil of RA0 resembled the lower depth of older and more mature soils of the chronosequence, as seen in the distribution of samples on the PCA (Figure 12). Of note, the difference between RA0 and older sites was evident in the content of nitrogen and its fractions more so than in the content of dissolved carbon, which did not vary throughout the chronosequence (Figure 7). This is likely due to the nature of the parent material used for the reclamation, which is a calcareous loess substrate, and the loess belt in Western Europe comprises of soils that usually contain a relatively high amount of calcium carbonate (10 to 15 percent) (Finke and Hutson, 2008). Findings from **Publication II**, where calcium carbonate ($CaCO_3$) was measured in the same chronosequence, showed no significant change in the concentration of $CaCO_3$ between reclamation ages or depth. This constant presence of $CaCO_3$ likely also explains the relatively constant pH measured in the samples, which were slightly alkaline and ranged from 7.4 to 7.6. In contrast, no $CaCO_3$ was found in an agricultural site in the vicinity of the mine that was chosen as a reference site for comparison. Results from the same publication also showed that a substantial part of soil's organic carbon content at the beginning of the reclamation is the "dead" carbon inherited from parent loess material. Interestingly, the concentration of total carbon measured by dry combustion showed a strong difference between reclamation ages; a three times increase by RA24 in the 1-5 cm depth, and a two times increase in the 16-20 cm depth compared to RA0. Low concentrations of total carbon and dissolved organic carbon and high $CaCO_3$ in RA0 showed that $CaCO_3$ accounts for a substantial part of soil's total carbon at the beginning of the reclamation.

Nitrogen fractions as major contributors explaining the variance between reclamation ages and depths in the PCA (Figure 12) was not a surprising result, as nitrogen is often found to limit primary production in terrestrial ecosystems (Vitousek et al., 1989) and it is a long-held theory in models of long-term ecosystem development that early successional ecosystems are nitrogen-limited since most initial substrates do not contain N (Menge et al., 2012; Darcy et al., 2018). Supporting the N-limitation theory are also data showing very low microbial abundances in RA0, as seen through the measured abundances of bacteria,

archaea, and fungi. These showed abundances lower by up to two orders of magnitude compared to other RAs, exhibiting a trend similar to those observed in the early stages of different initial ecosystems, such as glacier forefields (Esperschütz et al., 2011) and salt marshes (Dini-Andreote et al., 2014). Interestingly, C_{mic} concentrations showed no differences between reclamation ages in the topsoil, while N_{mic} had significantly lower concentrations in RA0 compared to later RAs (Figure 8). Furthermore, amounts of DNA extracted from different soils followed the same trend as concentrations of N_{mic}. These results showed that the system established at the beginning of the reclamation is highly dependent on external N sources, reinforcing the introduction of a leguminous plant as the key step to biological restoration.

Low organic content in RA0 is also mirrored in the low concentration of extracellular polymeric substances (EPS) extracted from the topsoil (Figure 17). Both polysaccharide and proteinaceous components of EPS were significantly lower in RA0 compared to other reclamation ages, and their concentrations were comparable to results obtained from a bare fallow site where the extraction and measurement were performed using the same method (Redmile-Gordon et al., 2014) as the one applied in this study. Given that aggregates, the main constituents of soil structure, are mineral particles connected by various biotic and abiotic binding agents, the lack of organic material in the calcareous carbonate-rich soil in RA0 suggests that carbonates are the main cementing materials at the beginning of the reclamation. It is interesting to note that during a test for a greenhouse experiment that would compare plant growth and microbial communities in soils from RA0 and RA6, the RA0 soil could not hold water and did not survive transport from the laboratory to a nearby greenhouse, destroying the experiment. This happened likely because loess substrates rich in calcium carbonate tend to form soft rock-like structures under dry environmental conditions that are dissipated and lost when moisture content increases (Li et al., 2016; Yates et al., 2018). In fact, a study of soils from the same chronosequence where macroaggregates were submerged in water showed complete disintegration of the aggregate structure, suggesting severe water instability and low plasticity, as well as the sensitivity of the soils to erosion (Pihlap et al., 2021).

4.2.2 Biological activation of the soil following the growth of *Medicago sativa*

A drastic change in soil characteristics in RA1 of the studied chronosequence was initially described in **Publication II**, where different types of data were combined to describe the overall initial soil formation as influenced by management and loess parent material. The

main improvement in soil quality with the introduction of alfalfa was seen through the accumulation of organic matter in the topsoil, mostly due to plant litter and rhizodeposition. Alfalfa is a nitrogen-fixing legume plant with the ability to act as a source of the primary limiting nutrient in most initial ecosystems (Brankatschk et al., 2011), making its introduction an essential step for the formation of a fully functional soil. However, soil physical parameters in the initial reclamation phase (RA0-RA3) were still characterized mainly by low macroporosity and high bulk density, which is typical for post-mined soils (Mukhopadhyay et al., 2016). Despite the high bulk density, which usually negatively impacts root growth (Nosalewicz and Lipiec, 2014), we observed a deep rooting system of alfalfa of more than 1 m already in RA1.

Biological activation of the soil was particularly evident through an increase of microbial biomass observed in RA1 after the introduction of alfalfa (Figure 8, Figure 9, Figure 10). A sharp increase in microbial N portrays the legume's ability to fix nitrogen through its symbiotic interactions with rhizobial bacteria (Fonouni-Farde et al., 2016). Microbial N is the biologically active N pool in soil and an important participant in soil N cycling (Deng et al., 2000; Xing et al., 2022), making it one of the crucial elements for development of the young reclaimed soils. This increase was corroborated by the significantly higher amount of DNA extracted from RA1 compared to RA0 samples, which was also reflected in the quantified copy numbers of bacterial, archaeal, and fungal genes, all of which showed substantial increases. These findings are in line with a study that focused on bacterial communities in a chronosequence of reclaimed soils from the same mining region in West Germany, where a sharp increase in bacterial abundance was observed following the cultivation of alfalfa (Schmid et al., 2020). In addition, we found an almost 2-fold increase in the concentrations of EPS-polysaccharides and proteins (Figure 17), which were most likely produced by the enriched microbial community fueled by the carbon from fresh organic matter input by alfalfa. However, the concentrations found in RA1 and RA3 were still considerably lower than those found in mature grassland soil (Redmile-Gordon et al., 2014).

4.2.3 Transition to the agricultural management phase

Following the transition to agricultural management phase of the reclamation, several changes were observed in the soil ecosystem in RA6 samples: increase in dissolved organic N and ammonium in the first two depths (Table 4, Figure 7), decrease in N_{mic} and bacterial 16S rRNA gene abundance in the first depth (1-5 cm) but an increase in the

second depth (16-20 cm), and decrease in fungal ITS region copy numbers (Figure 8, Figure 10). The second depth of RA6 showed a particular increase in N concentrations as well as microbial C and N, which can also be seen in the clustering of those samples in a separate quadrant of the PCA analysis (Figure 12). This was accompanied by a decrease in bulk density and an increase in macroporosity in the plough layer of RA6 (**Publication II**). These results are likely owing to the effects of deep tillage, which incorporates surface residues to deeper horizons and increases oxidation, leading to quicker decomposition of organic residues and release of nutrients (Johnson and Hoyt, 1999). Moreover, it is in line with studies comparing conservational and conventional tillage systems, where results show that under conservational tillage microbes tend to concentrate at the surface (accompanying the accumulation of organic carbon) and decrease at lower depths, whereas under conventional tillage this trend is reversed and equal or greater amounts of microbes exist below the surface (Doran, 1987; Johnson and Hoyt, 1999). Results observed in the plough layer of RA6 showed that with the transition to the agricultural management phase, nutrients and microbial biomass accumulated during the initial reclamation phase in the topsoil are diluted by nutrient-low subsoil, leading to decreased microbial abundances in the first depth and increased nutrients and biomass in the second depth.

The strong effect of tillage and the change from alfalfa to wheat on the established microbial community was also observed in a steep decrease in fungal abundances as measured by copy numbers of the fungal ITS region. Detrimental effects of mechanical soil mixing on the disruption of fungal hyphal networks are well established in the literature (Evans and Miller, 1990; Frey et al., 1999; Lu et al., 2018). In addition, legumes are generally more mycotrophic (i.e. receiving carbon, water, or nutrients through symbiotic associations with fungi) than grasses, among which e.g. maize was shown to carry more mycorrhizal fungi than wheat (Plenchette et al., 2005). Therefore, the sharp decrease in fungal abundance following the deep tillage event and transition to wheat in the agricultural management phase was not surprising.

4.2.4 Soil structure after 24 years of development and management

The amount of dissolved carbon did not increase throughout the agricultural management phase and different depths, while dissolved nitrogen concentrations increased from RA12 to RA24 in all depths. Still, they were lower than RA6 in the first two depths (Figure 7). The same was observed for ammonium concentrations (Table 4). A clear depth stratification in

RA24 was observed in the PCA analysis where only samples from the third depth clustered with samples from the initial reclamation phase on the left side, while samples from the first two depths clustered on the right side (Figure 12). Interestingly, the only variable that explains variance in the data in the upper left quadrant of the PCA is nitrate (Figure 11c). Subsoils have the ability to store nutrients for plant growth (Kautz et al., 2013), especially those prone to leaching below the topsoil following the addition of mineral fertilizers or mineralization of organic matter, such as nitrate (Dunbabin et al., 2003; Schneider and Don, 2019).

Despite the addition of organic fertilizers in the agricultural management phase, a loss of organic C stock from RA6 to RA24 and far lower concentrations of organic C in RA24 were found compared to a nearby mature soil unaffected by mining (**Publication II**). Using radiocarbon ^{14}C activity, these losses were confirmed to be affecting recent organic carbon, since the proportion of “dead” organic carbon increased with continuous tillage in RA12. These results may be a reflection of inadequate organic C inputs since the last addition of organic fertilizers happens in the seventh year of the reclamation followed only by mineral fertilizers. Another possible reason may be higher organic matter mineralization rates by microbes, since we observed the highest abundance of bacteria and archaea in RA24, at least in the first depth (Figure 10).

Nevertheless, some improvements in soil structure by RA24 were observed through a decrease in bulk density and an increase in macroporosity in the plough layer (**Publication II**). Beneath the plough layer, alfalfa roots created a system of biopores that could be used by following crops in the agricultural management phase to explore the subsoil (Lucas et al., 2019). However, the bulk density in the first depth of RA24 was still considerably higher (1.55 ± 0.04 g cm^{-3}) compared to a nearby mature soil (1.28 ± 0.04 g cm^{-3}) (**Publication II**), which is in line with other studies that compared bulk densities in reclaimed and natural soils (Shrestha and Lal, 2011; Liu et al., 2017). Reclaimed mining soils are generally characterized by high bulk density (Ussiri et al., 2006), which can cause poor aeration, limited nutrients, decreased water retention, and impede root growth (Stirzaker et al., 1996). In fact, it is often a limiting factor for the growth of native plants in reclaimed mining soils (Sinnett et al., 2008), since bulk densities higher than 1.5 g cm^{-3} can restrict root growth of many plants (Kozlowski, 1999). This was also confirmed by a lab experiment that used soils from the same chronosequence compacted to different bulk densities, where results showed that at a density of 1.45 g cm^{-3} roots likely cannot elongate into the

compacted rhizosphere (Lucas et al., 2019). The fact that the bulk density measured in topsoil of RA24 was still higher than 1.5 g cm^{-3} even after 24 years of reclamation is indicative of an incomplete soil structure development.

4.3 Seasonal dynamics

4.3.1 Rapid N turnover in the initial reclamation phase

Revitalizing effects of alfalfa rhizodeposition and above-ground litter on the initial reclamation substrate can directly be observed in samples taken throughout a growing season. In the RA0 field, almost all parameters measured in April samples (before the introduction of alfalfa) were lower compared to older RAs. Interestingly, when looking at relative changes during the season within RA0 (Figure 16), nitrogen fraction concentrations were highest and microbial biomass was lowest in April, before alfalfa planting. In contrast, in samples taken in July, after alfalfa was planted, nitrogen fractions were depleted and microbial biomass increased. Low concentrations of nitrogen fractions could be due to sampling taking place before the proper establishment of root nodules that could fill the N pool by fixation. Additionally, it could be the result of competition between the young alfalfa plants and microbes. Another likely competitor is weed plants, which are known to invade new alfalfa fields during the seeding year (Yang et al., 2022). In November, C_{mic} and N_{mic} had the highest concentrations, along with dissolved carbon but not dissolved nitrogen, suggesting that the six-month period of alfalfa growth in a new, nutrient-poor ecosystem is not enough to build up reserves of nitrogen.

In RA1 and RA3 fields, the aboveground plant parts were freshly cut down before the sampling in April. Compared to other RAs at the same time, C_{mic} and N_{mic} had the highest concentrations in RA1 and RA3, but the concentrations of all measured nitrogen fractions were low. These results suggest that all N introduced in this young ecosystem is quickly taken up by the microbial community, or that the sampling was performed before the decomposition of plant litter started. There could be multiple factors causing decreased rates of decomposition in the initial reclamation phase: placement of the litter (incorporated litter is decomposed more rapidly than litter left on the soil surface), high bulk density leading to low oxygen concentrations that decelerate ammonification and disable nitrification processes (Stefanakis et al., 2014), or limited microbial activity due to low nutrient content and their sub-optimal ratios (Hoyos-Santillan et al., 2018). In July, there were increases in dissolved and organic nitrogen concentrations, as well as nitrate and N_{mic}, especially in RA1. These increases, together with the positive correlation of N

concentrations with microbial N, demonstrate a flush of mineralized nutrients from the plant litter and suggest an increased microbial activity in response to the added nitrogen, alleviating the N limitation in the system (Wei et al., 2019). In November, four weeks after another cutting of alfalfa, all measured parameters were higher in RA1 than RA3, possibly indicating that the input of fresh C and N during the initial reclamation phase is not sufficient to build up reserves of organic matter and nutrients in the young reclaimed soil. In degraded lands, legumes can meet up to 90% of their N needs via biological N fixation, which is why their inclusion is a well-known strategy in e.g. managing pastures for intensifying forage production (Mosier et al., 2021). However, in the case of post-mining reclamation of loess soils, where newly created fields likely resemble primary succession more than a degraded land (e.g. an abandoned agricultural field), including legumes might not be enough to create a fully productive and healthy ecosystem.

4.3.2 Effect of agricultural management practices on N_{mic}

The transformation from the initial reclamation phase to the agricultural management phase in the fourth year of reclamation is made by deep tilling (30 cm), organic matter fertilization with compost, and subsequent crop rotation. To investigate the effects of this transition during a growing season, samples were taken from fields RA4 and RA5 to cover time points close to this transition. In samples taken in April, after the first seeding of wheat, there was a marked increase in nitrogen concentrations (Table 5, Figure 13) compared to the initial reclamation phase, but a decrease thereafter; a trend opposite to the one observed during alfalfa growth.

Surprisingly, the N increase in April was not accompanied by an increase in N_{mic} (Table 6). A meta-analysis of 96 studies that compared the effect of agricultural practices on microbial biomass found that conservation agriculture practices increased N_{mic} by up to 64% compared to conventional tillage practices (Li et al., 2018). This result is in line with the low concentrations of N_{mic} found in April samples taken from RA4, as well as RA5 and RA12, confirming the high sensitivity of N_{mic} to changes in soil ecosystem and agricultural management practices and thus its potential to be used as an early indicator of changes in soil N stability (Powlson et al., 1987; Brookes, 1995; McDaniel et al., 2014). Conversely, C_{mic} concentrations showed higher stability between both reclamation ages and seasons (Figure 13). This might be due to the constant presence of $CaCO_3$, which can adsorb organic compounds and stabilize soil organic C (Rowley et al., 2018).

4.4 Bacterial diversity and community composition

4.4.1 Low diversity of the bacterial community at the beginning of reclamation

In the initial substrate of RA0, bacterial community structure showed a strong dissimilarity compared to older RAs, as seen from the clustering in the NMDS analysis of bacterial reads (Figure 20). In addition, the distance between RA0 replicates was larger even than the distances between e.g. RA1 and RA6. This result suggests a high heterogeneity of plots and a stochastic distribution of microbes, which is in line with studies indicating that the assembly of bacterial communities in early successional soils is largely governed by stochastic processes (Dini-Andreote et al., 2015). Generally, the assembly of bacterial communities is influenced by deterministic and stochastic ecological processes. In deterministic processes, the composition and relative abundance of species is selected through the influence of external factors, whereas stochastic processes involve events that are equivalent to those produced by random chance alone (Chase and Myers, 2011; Stegen et al., 2012; Tripathi et al., 2018). In RA0, bacterial community assembly may be influenced by local conditions in niches formed in the initial mixture of parent material and previous agricultural soil, leading to an increased relative influence of stochastic over deterministic processes.

Compared to more mature soil, the diversity of the microbial community in RA0 was low: only up to 23% of metagenomic reads were left unassigned on the family level and there were relatively few bacterial families, especially those of low abundance (Figure 23). The community was primarily composed of a handful of families, with *Sphingomonadacaeae* and *Comamonadaceae* comprising already more than 10% of the total community. The genus *Sphingomonas* comprised almost 5% of the *Sphingomonadaceae* family found in RA0, which was also recognized as the main potential producer of exo- and lipopolysaccharides (Figure 27). Members from this genus are known to promote plant growth during stress conditions, such as drought, heavy metal contamination, and salinity, but they are also noted for their ability to degrade organometallic compounds (Asaf et al., 2020). Metallic compounds are often found in open-cast coal mining areas, since the mining process includes a phase when rocks of different compositions (e.g. iron sulfide minerals) are removed and exposed to atmospheric conditions, triggering their accelerated weathering (Bhuiyan et al., 2010). Furthermore, some genera from the *Comamonadaceae* family are photoautotrophs (Willems, 2014), which is a trait useful for survival in low nutrient

and organic matter conditions, like those found in RA0. *Variovorax, Polaromonas,* and *Ramlibacter* were the most abundant *Comamonadaceae* genera in RA0, and the sharpest decrease by RA1 was observed for *Ramlibacter*. This genus was first isolated from weathered fragments of meteorite minerals from sub desert soil in Tunisia and described as a nanometric bacterium able to produce cysts (i.e. cells encased by thick capsules), which provide long-term resistance to desiccation (Heulin et al., 2003). These results show that in the rare bacterial community in RA0, the main microbe inhabitants are bacteria capable of withstanding extreme and stressful conditions.

The structure of the enriched bacterial community in RA1 and RA3 was more similar to samples from the agricultural management phase than to RA0 (Figure 20). Furthermore, RA3 was separating from RA1 and other RAs along the y-axis in the NMDS analysis, suggesting higher dissimilarity of RA3. After RA0, the relative abundance of the most abundant families decreased (Figure 22), while the abundance of reads not assigned to the family level increased (Figure 23). In addition, the diversity index calculated with the Nonpareil algorithm peaked at RA1 (Figure 19) and the percentage of metagenomic reads unassigned to the kingdom level increased (Table 7). Normally, soils are one of the most complex and phylogenetically diverse microbial habitats with an impressive genetic diversity, so it is not unusual to find 40-60% of sequenced genes to be of unknown function since these analyses rely on reference databases that largely originate from human health and biotechnology research efforts (Choi et al., 2017; Bernard et al., 2018; Chen et al., 2020). Initial findings from the metagenome analysis in this study reflect the development and biological diversification of reclaimed soil following the introduction of alfalfa.

4.4.2 Effect of the transition to the agricultural management phase on the bacterial community

The transition in reclamation stages had a strong influence on the bacterial community, as observed from the analysis of bacterial community structure (Figure 20) where samples from RA6 clustered close together, indicating homogenization following the mechanical mixing of soil. The number of bacterial reads not assigned to the family level increased, especially for unclassified Acidobacteria and Proteobacteria (Figure 23), the most common bacterial phyla found in soils (Lauber Christian et al., 2009), where they exhibit great metabolic diversity (Kielak et al., 2016). Not many genomes from Acidobacteria have been thoroughly studied despite their abundance in many different environments, but they have been found to carry a comprehensive set of genes that allow them high adaptability, e.g.

genes involved in regulating carbon, nitrogen, and sulfur cycles, polysaccharide degradation, exopolysaccharide production, as well as genes encoding for different transporters, imparting resilience to starvation and stress (Kielak et al., 2016; Kalam et al., 2020). Collectively, these results indicate a shift in the microbial community from mostly symbiotic N-fixing bacteria and increasingly abundant fungi to a diverse and relatively low abundant bacteria commonly found in soil following a drastic change in reclamation strategy in a young soil ecosystem.

4.5 Exopolysaccharide and lipopolysaccharide-forming bacteria

4.5.1 Measuring bacterial EPSs in soil

Extracellular polymeric substances are the primary material of bacterial biofilms (Sutherland, 2001) and consist primarily of exopolysaccharides and proteins (Redmile-Gordon et al., 2014). Extraction of these substances from complex materials such as soils is notoriously difficult mainly due to two reasons: (i) almost all components of EPSs can also be found inside the cells living in biofilms, which means that the extraction method needs to be delicate enough not to release intracellular material into the extract (Flemming and Wingender, 2010), (ii) EPSs are not the only source of organic matter in soil so the extraction method needs to be sensitive enough to distinguish between EPS and other organic biochemicals in bulk soil (Redmile-Gordon et al., 2014). The exopolysaccharide and protein EPS fractions measured here were extracted using a modified cation exchange resin method: first, soluble microbial products are extracted from soil using a $CaCl_2$ solution, then they are mixed with cation exchange resin and an extraction buffer, and finally the mixture is shaken and centrifuged to derive supernatant that contains EPS polysaccharides and proteins (Frølund et al., 1996; Redmile-Gordon et al., 2014). In comparison to four other methods, this one showed significant increases in polysaccharide production in a heterotrophic soil biofilm responding to desiccation stress and substrate addition, as well as the highest selectivity and lowest cell-damaging effects (Redmile-Gordon et al., 2014; Redmile-Gordon et al., 2015).

Adding to the complexity of extracting bacterial EPS exopolysaccharides and proteins from the soil is the fact that EPSs are not produced only by bacteria but also by microalgae, yeasts, fungi, protists, and plants (Gunina and Kuzyakov, 2015). Hence, it is difficult to assess a direct link between measured polysaccharide concentrations and their source of origin. To minimize the effect plants may have on these concentrations, the sampling strategy in this study excluded the sampling of any plant root materials or plant litter. In

addition, the conclusion from a meta-analysis that included results from 64 studies focusing on the origin, composition, and fate of sugars in soil was that only 20% of all sugars in soil originate from a primary source, i.e. decomposition of plant litter and rhizodeposits, and 80% originate from secondary sources, i.e. microorganisms and their residues (Gunina and Kuzyakov, 2015). This confirms that at least the majority of the measured EPS-polysaccharide concentrations are of microbial origin. Still, the drastic increase in EPSs observed in RA1 could also be attributed to an increasing fungal, as well as bacterial biomass. However, we also observed an increase in relative abundances of exo- and lipopolysaccharide bacterial genes in RA1, especially for nitrogen-fixing bacteria, suggesting that bacteria play an important role in the production of polysaccharides in the early stages of reclamation. Moreover, concentrations of both EPS exopolysaccharides and proteins increased in the agricultural management phase – the phase in which fungal biomass drastically decreased compared to the years of alfalfa growth.

Unlike exopolysaccharides, methods commonly used for lipopolysaccharide extraction are based on biological tissue assays with the aim of detecting lipopolysaccharides in human or animal tissues where they can elicit an immune response and are therefore primarily investigated for their pathological properties (Raetz and Whitfield, 2002; Stromberg et al., 2017). They can also be extracted from pure cell cultures (Reitz et al., 2000), but these methods are not well adapted for use on complex environmental samples such as soil (Tunlid and White, 1992; Rylander, 2002). Consequently, lipopolysaccharide extraction and quantification were not performed as part of this study. Hence, a direct link between increased relative lipopolysaccharide gene abundances at certain points during the reclamation and final lipopolysaccharide products cannot be made.

4.5.2 Symbiotic nitrogen fixers as important exo- and lipopolysaccharide producers

As expected, we observed a strong increase of bacteria belonging to the order Rhizobiales in RA1 and RA3. Rhizobia are a group of soil bacteria able to establish symbiosis with legume plants, but before the symbiosis is established they must survive the often harsh conditions in the environment and compete with local microbiota (Poole et al., 2018). Rhizobia are saprophytes that can adapt to oligotrophic conditions allowing them to survive in complex and competitive microbial communities (Poole et al., 2018), which is likely the case in the nitrogen-scarce environment of the initial reclamation phase. An important mechanism that allows rhizobia to live on plant surfaces, but also on soil mineral surfaces

under different environmental conditions is biofilm formation (Russo et al., 2015). Rhizobia are known to secrete exopolysaccharides in the environment both during symbiosis with plants and while free living in bulk soil (Skorupska et al., 2006). Lipopolysaccharides also play a central role in biofilm formation since they determine cells adhesiveness and cell-to-cell aggregation, and were found to be important for rhizobia as the regulation factor in biofilm formation, e.g. for *Rhizobium leguminosarum* (Russo et al., 2015). In addition, it is suggested that bacterial surface polysaccharides help in the evasion of the plant immune system (D'Haeze and Holsters, 2004). In accordance with this, there was an increase in the relative abundance of *Rhizobiaceae*, *Bradyrhizobiaceae*, and *Burkholderiaceae*, but also in the relative abundance of exo- and lipopolysaccharide genes that they carry in RA1 (Figure 29).

Within the *Rhizobiaceae* family, a particular increase was observed for *Sinorhizobium* (Figure 24), which is precisely the genus whose host legume is *Medicago sativa* (Poole et al., 2018). *Sinorhizobium meliloti* produces high- and low-molecular-weight exopolysaccharides, which promote symbiosis and are essential for cell-to-cell interactions and surface attachment (Sorroche et al., 2018). In addition, an interplay between LPSs and low-molecular-weight exopolysaccharides was found to influence the interactions of *S. meliloti* with biotic and abiotic surfaces (Sorroche et al., 2018). Similar to *Sinorhizobium* species, multiple experiments with *Rhizobium leguminosarum* in which exopolysaccharide genes or their regulators were deleted, showed that such mutants had a decreased ability to attach and form biofilms on surfaces of roots and root hairs, whereas over-expression of these genes increased attachment (Williams et al., 2008; Janczarek et al., 2010; Janczarek and Rachwał, 2013). The same was observed for LPSs, where a lack of genes encoding for enzymes involved in the synthesis of the O antigen in the LPS core produced scattered micro-colonies with poor attachment to root surfaces (Russo et al., 2015). Results presented in this thesis showed a peak in the relative abundance of exo- and lipopolysaccharide genes assigned to *Rhizobiaceace* in RA1, corroborating the importance of these molecules for successful attachment and symbiosis of bacterial members from this family with *Medicago sativa*.

Relative abundance of the *Burkholderiaceae* family increased after RA0 and remained high through all RAs, but carried most exo- and lipopolysaccharide genes at RA3. *Burkholderiaceae* are saprophytic bacteria primarily found in soil and water (Coenye, 2014), but their biofilm-forming characteristics are mostly studied in pathogenic species

from *Burkholderia*, the type genus, due to the important role of biofilms in *Burkholderia* pathogenesis processes (Sawasdidoln et al., 2010; Chen et al., 2014; Limmathurotsakul et al., 2014). However, some species (e.g. *Burkholderia phymatum*) do carry symbiosis genes on plasmids and are known as nitrogen-fixing symbionts of *Mimosa* plant species (Moulin et al., 2014). Hence, biofilm formation is likely an important characteristic for *Burkholderia* soil-dwelling species as well as for their human pathogen counterparts. *Bradyrhizobiaceae* is another rhizobial family that increased in relative abundance after RA0 (Figure 22), particularly the *Bradyrhizobium* genus (Figure 24). Similar to *Rhizobiaceae*, exopolysaccharide production and excretion is suggested to be an important factor in symbiosis formation for different *Bradyrhizobium* species, e.g. *B. japonicum* and *B. elkanii* with soybean (Quelas et al., 2010). These results, together with the increased amounts of EPS-polysaccharides and proteins (Figure 17) and the estimated absolute abundances of exo- and lipopolysaccharide genes found in RA1 (Figure 26) demonstrate that symbiotic nitrogen fixers are important players in triggering soil structure development in young reclamation soils.

By RA3, the relative abundance of most of the analyzed exo- and lipopolysaccharide genes decreased (Figure 25), as did the measured amount of EPS-polysaccharides extracted from the soils (Figure 17). This result is in concordance with experiments that report decreased exopolysaccharide production of bacterial colonies grown in conditions of sufficient N compared to N starvation (López-García et al., 2001; Sanin et al., 2003; Quelas et al., 2006), since the concentrations of dissolved nitrogen (Figure 7) and ammonia (Table 4) increased in RA3, alleviating N limitation after 3 years of alfalfa growth.

4.5.3 Stress conditions facilitate polysaccharide production

The introduction of tillage practices in a sensitive new ecosystem most likely acted as a disturbance for the microbial community established during alfalfa growth. Niches shaped by fungi and nitrogen-fixing microbial communities during the unperturbed initial reclamation phase were likely fractured into mineral particles and organic residue by mechanical soil mixing. Macroaggregates are particularly susceptible to degradation by tillage, leading to disintegration and destabilization of the previously-protected organic matter, increased rates of organic carbon decomposition, and change in nutrient availability to plants (Beare et al., 1994; Balota et al., 2003). Additionally, tillage improves soil aeration and creates physical conditions that are more favorable for microbial activity. The disintegration of macroaggregates also provides an increased surface area available for

microbial attack or attachment (Johnson and Hoyt, 1999), explaining the increased amounts of EPS-polysaccharides and proteins (Figure 17) as well as increased relative abundances of exo- and lipopolysaccharide-producing genes (Figure 25), despite a decrease in microbial biomass (Figure 8, Figure 10).

Interestingly, no significant differences were observed in aggregate sizes in different reclamation ages due to the cementing effect of carbonate, which was found in high concentration throughout the chronosequence (**Publication II**). However, there was an increase of organic carbon stored within the smallest aggregates at RA6 and this aggregate fraction is most likely to be affected by bacterial exopolysaccharide and lipopolysaccharide excretion and biofilm formation (Six et al., 2004). Add to that, seven out of ten exo- and lipopolysaccharide genes analyzed here showed the highest relative abundances at RA6 (Figure 25). By excreting exopolysaccharides and lipopolysaccharides, bacterial cells can surround themselves with clay particles (Burns and Davies, 1986), thus gluing soil particles together and improving soil structural properties (Costa et al., 2018). In addition, the formation of biofilm EPSs is a strategy of protection from environmental stressors and is well known to be utilized by microbes living in disturbed ecosystems (Rossi and De Philippis, 2015), such as the one here created by deep tillage.

At RA6, an increase in relative abundances of exopolysaccharide genes was observed mostly for two families from the *Sphingomonadales* order: *Sphingomonadaceae* and *Erythrobacteraceae*. This result indicates a change in the exo/lipopolysaccharide-harboring bacteria (Figure 27) compared to the initial reclamation phase, where the main carriers of these genes were N-fixing bacteria. Members of the *Sphingomonadaceae* family members are known to exhibit the ability to stabilize soil structure. In a study where bacteria were isolated from microaggregates of soils under different management agroecosystems, Caesar-TonThat et al. (2007) found that *Sphingobacterium*, next to *Stenotrophomonas* from the *Xantomonadaceae* family, have the greatest ability to increase aggregate strength. *Sphingomonadaceae* are known for characteristic glycosphingolipids on the outer membrane (Kawasaki et al., 1994), which have strong adhesive properties, and *Sphingomonas*, the type genus, produces exopolysaccharides named sphingans (Caesar-TonThat et al., 2007; Glaeser and Kämpfer, 2014). They also produce gellan, while *Xanthomonas* species produce xanthan: exopolysaccharides widely used in biotechnology for their adsorption properties (Balíková et al., 2022). Other notable families with most exo-

and lipopolysaccharide genes assigned at RA6 include *Oxalobacteraceae* and *Chitinophagacaeae.*

4.5.4 Bacterial adaptation to regular agricultural practices

Similar patterns in bacterial communities were observed in RA3 and RA24: increased general biomass of bacteria (Figure 10) and decreased relative abundances of exo- and lipopolysaccharide genes, with the exception of *wza* (Figure 25). These results indicate that the bacterial community increases in abundance with time and accumulation of organic matter and eventually adapts to management practices, as observed through a decreased need to produce EPS as a reaction to stress conditions, which is in accordance with other metagenome studies showing that bacterial communities remain stable under long-term treatments, at least on the DNA level (de Vries et al., 2015; Cania et al., 2019). Despite the decrease in relative gene abundances, estimated absolute abundances of exo- and lipopolysaccharide genes were highest in RA24 (Figure 26) since the total bacterial abundance, as measured by qPCR, was highest at the same time point. In addition, the concentration of EPS-polysaccharides peaked in RA24, likely as a result of an accumulation over time from an increasing biomass of bacteria that have a lower need to produce exopolysaccharides and lipopolysaccharides. Considering the losses of fresh organic C in RA12 and RA24, the peak in EPS-polysaccharide concentration at RA24 underlines the importance of soil bacteria for the stabilization of soil structure and improvement of soil quality during reclamation in the long term.

Relative abundances of most bacterial families decreased, while the number of genes not assigned to the family level, especially for unclassified Acidobacteria, peaked at RA24. These results show a continuous shift to a diverse community of low-abundant species typically found in agricultural soils. One exception was the *Nitrospiraceae* family, dominated by the *Nitrospira* genus (Figure 24), which increased in both general relative abundance and abundance of exo- and lipopolysaccharide genes (Figure 29), replacing *Sphingomonadaceae* as the most abundant exopolysaccharide producer (Figure 27). Members of the *Nitrospiraceae* family have the ability to perform an important step in nitrogen cycling; particularly the *Nitrospira* genus*,* which is the predominant known nitrite oxidizer found in nature (Daims, 2014). In fact, some *Nitrospira* members are complete nitrifiers able to perform oxidation of ammonia via nitrite to nitrate, providing them a growth advantage over sole ammonia or nitrite oxidizers (Daims et al., 2015). Hence, *Nitrospira* are readily found in agricultural soils (Xia et al., 2011). In addition, all known *Nitrospira* have

a tendency to produce large amounts of EPS and cells in microcolonies are often embedded in EPS (Daims, 2014). These results underline the advantage of bacteria involved both in important nutrient cycles as well as exo- and lipopolysaccharide production over solely exo- and lipopolysaccharide producers in agricultural soils.

5 Conclusions and Outlook

The post-mining reclamation chronosequence in Garzweiler allowed us to investigate the genetic potential of bacterial communities to produce polysaccharides during different stages of soil formation. Combined with measurements of EPS polysaccharide concentrations, this study highlights the role of bacteria and their production of EPSs for the development of soil structure by gluing soil particles together. Improvements in soil structure greatly decrease sensitivity to erosion, which is a major concern during post-mining reclamation. Our results showed this is particularly relevant at the start of reclamation when concentrations of organic matter are low and soil structure is stabilized predominately by calcium carbonate-rich minerals, making it sensitive to erosion when the moisture content is increased. The first major change in the newly-established ecosystem followed the introduction of the first leguminous plant, alfalfa. With the growing plant, the soil was biologically activated by a flush of fresh organic matter and nitrogen, which helped alleviate the initial N limitation. At this point, soil structure formation was accelerated by an increased concentration of EPS-polysaccharides, and the bacterial community able to produce exo- and lipopolysaccharides was comprised mainly of rhizobia, corroborating these polysaccharides as important factors in the initiation of legume-rhizobia symbioses. However, soil physical parameters during the initial reclamation phase still resembled typical poor-quality post-mined soils (**Publication I**) and organic matter accumulation during a growing season was not substantial.

Transition to the agricultural management phase after three years of alfalfa growth acted like a major disturbance event to this sensitive ecosystem. It was marked by a decrease in microbial abundances in the topsoil, while the concentration of EPSs, as well as the relative abundance of exo- and lipopolysaccharide-producing families, increased. These results showed that the surviving bacterial community was efficient in producing polysaccharides and shifted from mostly symbiotic N-fixing bacteria to relatively low abundant species commonly found in agricultural soils. Production of EPSs in this phase was likely triggered as a protection mechanism against stress conditions (deep tillage event) and was fueled by carbon from the organic fertilizer, which acted as a necessary energy source for their production.

After 24 years of reclamation and crop rotation, there was a clear depth stratification between the plough layers and subsoil, similar to what is expected in soils under

conventional agriculture. However, despite the inputs of fresh organic matter, concentrations of organic C were lower and bulk density was still higher compared to a nearby mature soil, indicating an incomplete soil structure development. On the other hand, bacterial biomass, as well as the concentration of EPS-polysaccharides peaked in the topsoil of RA24 as a result of accumulation over time and production by a growing biomass of bacteria. However, these bacteria had decreased relative abundances of exo- and lipopolysaccharide genes, indicating a lower need for EPS production due to adaptation to agricultural management practices (**Publication II**). Nevertheless, bacterial EPSs are likely to be a very important source of organic matter stabilizing soil structure during reclamation in the long term, considering the inadequate organic C accumulation by RA24.

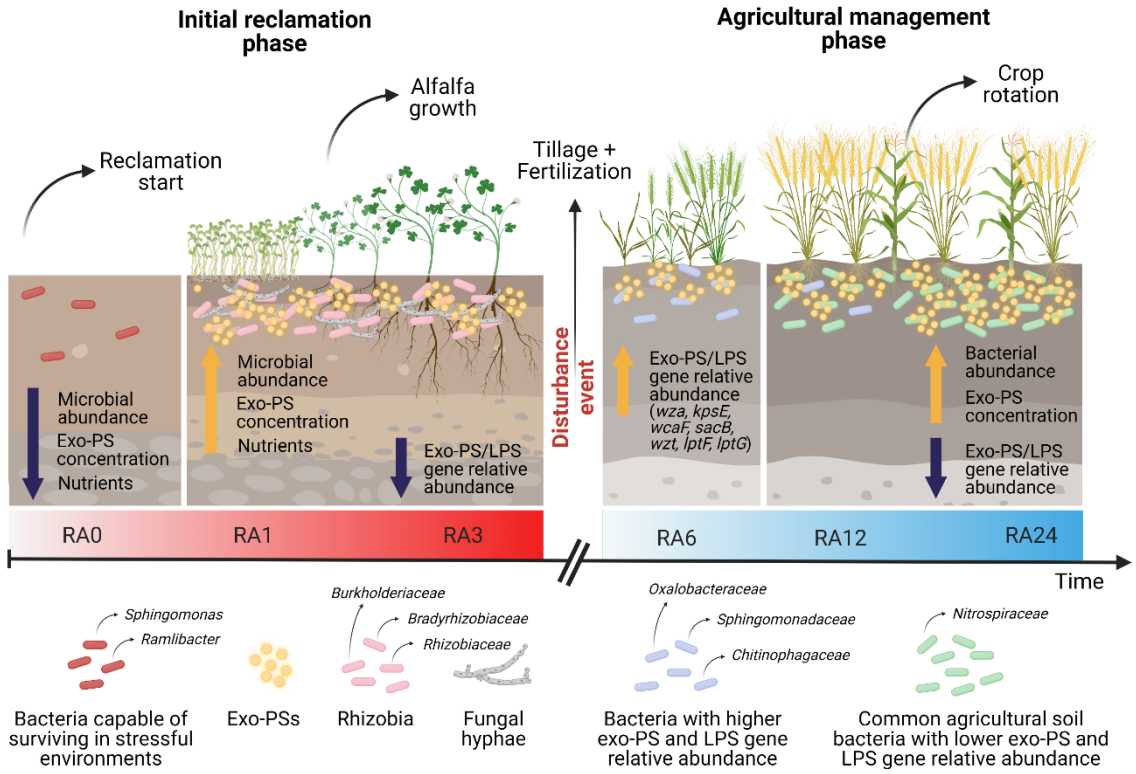


Figure 30. Infographic of the main findings
Visual depiction of the changes observed during the reclamation with notable bacterial genera and families found in each phase, as well as exo-PS and LPS genes that peaked in relative abundance at RA6. Created with BioRender.com.

Overall, this study shows that the exopolysaccharide and lipopolysaccharide-producing bacterial community is especially favored following disturbance events in the sensitive reclaimed soil. It is therefore possible that subjecting such soils to planned and carefully managed disturbance events, combined with higher organic matter input, could help build

the resistance necessary for withstanding severe future environmental conditions. However, these results are based on DNA sequencing, allowing us to examine the potential of bacterial communities. Hence, confirmation of this potential translating into the actual production of polysaccharides would require further research using techniques such as transcriptomics or proteomics. In addition, laboratory experiments in controlled conditions would allow for testing of different bacterial strains and EPSs they produce to find those most suitable for use *in situ*. Our findings provide a starting point for this type of research and add to the body of knowledge necessary for the successful restoration of destroyed ecosystems. Considering the ecological, climate change-related, and consequently humanitarian crises humankind is bound to face in the near future, successful restoration of degraded ecosystems will likely be one of the key solutions.

References

Abdian, P.L., and Zorreguieta, A. (2016) Extracellular factors involved in biofilm matrix formation by Rhizobia. In *The Perfect Slime: Microbial Extracellular Polymeric Substances (EPS)*. Flemming, H.-C., Neu, T.R., and Wingender, J. (eds): IWA Publishing, pp. 227-247.

Angers, D.A., and Caron, J. (1998) Plant-induced Changes in Soil Structure: Processes and Feedbacks. *Biogeochemistry* **42**: 55-72.

Asaf, S., Numan, M., Khan, A.L., and Al-Harrasi, A. (2020) Sphingomonas: from diversity and genomics to functional role in environmental remediation and plant growth. *Critical Reviews in Biotechnology* **40**: 138-152.

Atkinson, S., Goldstone, R.J., Joshua, G.W.P., Chang, C.-Y., Patrick, H.L., Cámara, M. et al. (2011) Biofilm Development on Caenorhabditis elegans by Yersinia Is Facilitated by Quorum Sensing-Dependent Repression of Type III Secretion. *PLOS Pathogens* **7**: e1001250.

Bach, H.J., Tomanova, J., Schloter, M., and Munch, J.C. (2002) Enumeration of total bacteria and bacteria with genes for proteolytic activity in pure cultures and in environmental samples by quantitative PCR mediated amplification. *The Journal of Microbiological Methods* **49**: 235-245.

Baldrian, P., Voříšková, J., Dobiášová, P., Merhautová, V., Lisá, L., and Valášková, V. (2011) Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi from the upper layers of forest soil. *Plant and Soil* **338**: 111-125.

Balíková, K., Farkas, B., Matúš, P., and Urík, M. (2022) Prospects of Biogenic Xanthan and Gellan in Removal of Heavy Metals from Contaminated Waters. *Polymers* **14**: 5326.

Balota, E.L., Colozzi-Filho, A., Andrade, D.S., and Dick, R.P. (2003) Microbial biomass in soils under different tillage and crop rotation systems. *Biology and Fertility of Soils* **38**: 15-20.

Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J.T. (2004) Phylogenetic Composition of Arctic Ocean Archaeal Assemblages and Comparison with Antarctic Assemblages. *Applied and Environmental Microbiology* **70**: 781-789.

Bardgett, R. (2005a) *The Biology of Soil: A community and ecosystem approach*. Oxford, UK: Oxford University Press.

Bardgett, R.D., Bowman, W.D., Kaufmann, R., and Schmidt, S.K. (2005b) A temporal approach to linking aboveground and belowground ecology. *Trends in Ecology & Evolution* **20**: 634-641.

Barré, P., McKenzie, B.M., and Hallett, P.D. (2009) Earthworms bring compacted and loose soil to a similar mechanical state. *Soil Biology and Biochemistry* **41**: 656-658.

Beare, M.H., Hendrix, P.F., Cabrera, M.L., and Coleman, D.C. (1994) Aggregate-Protected and Unprotected Organic Matter Pools in Conventional and No-Tillage Soils. *Soil Science Society of America Journal* **58**: 787-795.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **57**: 289-300.

Bernard, G., Pathmanathan, J.S., Lannes, R., Lopez, P., and Bapteste, E. (2018) Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery. *Genome Biology and Evolution* **10**: 707-715.

Bhuiyan, M.A.H., Parvez, L., Islam, M.A., Dampare, S.B., and Suzuki, S. (2010) Heavy metal pollution of coal mine-affected agricultural soils in the northern part of Bangladesh. *Journal of Hazardous Materials* **173**: 384-392.

Bodner, G., Leitner, D., and Kaul, H.P. (2014) Coarse and fine root plants affect pore size distributions differently. *Plant and Soil* **380**: 133-151.

Bossuyt, H., Six, J., and Hendrix, P.F. (2004) Rapid incorporation of carbon from fresh residues into newly formed stable microaggregates within earthworm casts. *European Journal of Soil Science* **55**: 393-399.

Brankatschk, R., Towe, S., Kleineidam, K., Schloter, M., and Zeyer, J. (2011) Abundances and potential activities of nitrogen cycling microbial communities along a chronosequence of a glacier forefield. *The ISME Journal* **5**: 1025-1037.

Brevik, E.C., and Hartemink, A.E. (2010) Early soil knowledge and the birth and development of soil science. *CATENA* **83**: 23-33.

Bronick, C.J., and Lal, R. (2005) Soil structure and management: a review. *Geoderma* **124**: 3-22.

Brookes, P.C. (1995) The use of microbial parameters in monitoring soil pollution by heavy metals. *Biology and Fertility of Soils* **19**: 269-279.

Brookes, P.C., Landman, A., Pruden, G., and Jenkinson, D.S. (1985) Chloroform fumigation and the release of soil nitrogen: A rapid direct extraction method to measure microbial biomass nitrogen in soil. *Soil Biology and Biochemistry* **17**: 837-842.

Buchfink, B., Xie, C., and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**: 59.

Bullock, M.S., Nelson, S.D., and Kemper, W.D. (1988) Soil Cohesion as Affected by Freezing, Water Content, Time and Tillage. *Soil Science Society of America Journal* **52**: 770-776.

Buol, S. (1973) *Soil Genesis and Classification*. Ames, IA: Iowa State University Press.

Burns, R.G., and Davies, J.A. (1986) The Microbiology of Soil Structure. *Biological Agriculture & Horticulture* **3**: 95-113.

Caesar-TonThat, T.C., Caesar, A.J., Gaskin, J.F., Sainju, U.M., and Busscher, W.J. (2007) Taxonomic diversity of predominant culturable bacteria associated with microaggregates from two different agroecosystems and their ability to aggregate soil in vitro. *Applied Soil Ecology* **36**: 10-21.

Caesar-Tonthat, T.C., B Stevens, W., Sainju, U., Caesar, A.J., West, M., and F Gaskin, J. (2014) Soil-Aggregating Bacterial Community as Affected by Irrigation, Tillage, and Cropping System in the Northern Great Plains. *Soil Science* **179**: 11-20.

Cania, B., Vestergaard, G., Krauss, M., Fliessbach, A., Schloter, M., and Schulz, S. (2019) A long-term field experiment demonstrates the influence of tillage on the bacterial potential to produce soil structure-stabilizing agents such as exopolysaccharides and lipopolysaccharides. *Environmental Microbiome* **14**: 1.

Chadwick, O.A., and Chorover, J. (2001) The chemistry of pedogenic thresholds. *Geoderma* **100**: 321-353.

Chan, S.Y., Liu, S.Y., Seng, Z., and Chua, S.L. (2021) Biofilm matrix disrupts nematode motility and predatory behavior. *The ISME Journal* **15**: 260-269.

Chase, J.M., and Myers, J.A. (2011) Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**: 2351-2363.

Chaudhary, D.K., Khulan, A., and Kim, J. (2019) Development of a novel cultivation technique for uncultured soil bacteria. *Scientific Reports* **9**: 6666.

Chen, L.X., Anantharaman, K., Shaiber, A., Eren, A.M., and Banfield, J.F. (2020) Accurate and complete genomes from metagenomes. *Genome Research* **30**: 315-333.

Chen, Y.S., Shieh, W.J., Goldsmith, C.S., Metcalfe, M.G., Greer, P.W., Zaki, S.R. et al. (2014) Alteration of the phenotypic and pathogenic patterns of Burkholderia pseudomallei that persist in a soil environment. *The American Journal of Tropical Medicine and Hygiene* **90**: 469-479.

Chenu, C. (1995) Extracellular Polysaccharides: An interface Between Microorganisms and Soil Constituents. In *Environmental Impacts of Soil Component Interactions: Land Quality, Natural and Anthropogenic Organics*. M., H.P., J., B., M., B.J., B., M.W., and L., P.A. (eds). Boca Raton, FL: CRC Lewis Publishers, pp. 217-233.

Chenu, C., and Cosentino, D. (2011) Microbial Regulation of Soil Structure Dynamics. In *The Architecture and Biology of Soils: Life in Inner Space*. Ritz, K., and Young, I.M. (eds): CABI International, pp. 37-70.

Chng, S.-S., Gronenberg, L.S., and Kahne, D. (2010) Proteins Required for Lipopolysaccharide Assembly in Escherichia coli Form a Transenvelope Complex. *Biochemistry* **49**: 4565-4567.

Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., Williams, R. et al. (2017) Strategies to improve reference databases for soil microbiomes. *The ISME Journal* **11**: 829-834.

Churchman, G.J. (2010) The philosophical status of soil science. *Geoderma* **157**: 214-221.

Coenye, T. (2014) The Family Burkholderiaceae. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 759-776.

Costa, O., Raaijmakers, J., and Kuramae, E. (2018a) Microbial Extracellular Polymeric Substances: Ecological Function and Impact on Soil Aggregation. *Frontiers in Microbiology* **9**: 1636.

Costa, O.Y.A., de Hollander, M., Pijl, A., Liu, B., and Kuramae, E.E. (2020) Cultivation-independent and cultivation-dependent metagenomes reveal genetic and enzymatic potential of microbial community involved in the degradation of a complex microbial polymer. *Microbiome* **8**: 76.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995) Microbial biofilms. *Annual Review of Microbiology* **49**: 711-745.

Cuthbertson, L., Kos, V., and Whitfield, C. (2010) ABC Transporters Involved in Export of Cell Surface Glycoconjugates. *Microbiology and Molecular Biology Reviews* **74**: 341.

Cuthbertson, L., Mainprize, I.L., Naismith, J.H., and Whitfield, C. (2009) Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. *Microbiology and molecular biology reviews* **73**: 155-177.

D'Haeze, W., and Holsters, M. (2004) Surface polysaccharides enable bacteria to evade plant immunity. *Trends in Microbiology* **12**: 555-561.

Daims, H. (2014) The Family Nitrospiraceae. In *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 733-749.

Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M. et al. (2015) Complete nitrification by Nitrospira bacteria. *Nature* **528**: 504-509.

Darcy, J.L., Schmidt, S.K., Knelman, J.E., Cleveland, C.C., Castle, S.C., and Nemergut, D.R. (2018) Phosphorus, not nitrogen, limits plants and microbial primary producers following glacial retreat. *Science advances* **4**: eaaq0942.

Darwin, C. (1881) *The formation of vegetable mould, through the action of worms, with observations on their habits*. London, UK: John Murray.

Davidson, A.L., Dassa, E., Orelle, C., and Chen, J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews* **72**: 317-364.

Dazzi, C., and Lo Papa, G. (2022) A new definition of soil to promote soil awareness, sustainability, security and governance. *International Soil and Water Conservation Research* **10**: 99-108.

De Palma, A., Sanchez-Ortiz, K., Martin, P.A., Chadwick, A., Gilbert, G., Bates, A.E. et al. (2018) Chapter Four - Challenges With Inferring How Land-Use Affects Terrestrial Biodiversity: Study Design, Time, Space and Synthesis. In *Advances in Ecological Research*. Bohan, D.A., Dumbrell, A.J., Woodward, G., and Jackson, M. (eds): Academic Press, pp. 163-199.

de Vries, M., Schöler, A., Ertl, J., Xu, Z., and Schloter, M. (2015) Metagenomic analyses reveal no differences in genes involved in cellulose degradation under different tillage treatments. *FEMS Microbiology Ecology* **91**: fiv069.

Degens, B.P. (1997) Macro-aggregation of soils by biological bonding and binding mechanisms and the factors affecting these: a review. *Australian Journal of Soil Research* **35**.

Denef, K., and Six, J. (2005) Clay mineralogy determines the importance of biological versus abiotic processes for macroaggregate formation and stabilization. *European Journal of Soil Science* **56**: 469-479.

Denef, K., Six, J., Bossuyt, H., Frey, S.D., Elliott, E.T., Merckx, R., and Paustian, K. (2001) Influence of dry–wet cycles on the interrelationship between aggregate, particulate organic matter, and microbial community dynamics. *Soil Biology and Biochemistry* **33**: 1599-1611.

Deng, S.P., Moore, J.M., and Tabatabai, M.A. (2000) Characterization of active nitrogen pools in soils under different cropping systems. *Biology and Fertility of Soils* **32**: 302-309.

Dini-Andreote, F., Stegen, J.C., van Elsas, J.D., and Salles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E1326-1332.

Dini-Andreote, F., de Cassia Pereira e Silva, M., Triado-Margarit, X., Casamayor, E.O., van Elsas, J.D., and Salles, J.F. (2014) Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. *The ISME Journal* **8**: 1989-2001.

Doran, J.W. (1987) Microbial biomass and mineralizable nitrogen distributions in no-tillage and plowed soils. *Biology and Fertility of Soils* **5**: 68-75.

Dow, J.M., Crossman, L., Findlay, K., He, Y.-Q., Feng, J.-X., and Tang, J.-L. (2003) Biofilm dispersal in Xanthomonas campestris is controlled by cell–cell signaling and is required for full virulence to plants. *Proceedings of the National Academy of Sciences* **100**: 10995-11000.

DuBois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., and Smith, F. (1956) Colorimetric Method for Determination of Sugars and Related Substances. *Analytical Chemistry* **28**: 350-356.

Dunbabin, V., Diggle, A., and Rengel, Z. (2003) Is there an optimal root architecture for nitrate capture in leaching environments? *Plant, Cell & Environment* **26**: 835-844.

Edwards, A.P., and Bremner, J.M. (1967) Microaggregates in Soil. *Journal of Soil Science* **18**: 64-73.

Emerson, W.W. (1959) The structure of soil crumbs. *Journal of Soil Science* **10**: 235-244.

Esperschütz, J., Pérez-de-Mora, A., Schreiner, K., Welzl, G., Buegger, F., Zeyer, J. et al. (2011) Microbial food web dynamics along a soil chronosequence of a glacier forefield. *Biogeosciences* **8**: 3283-3294.

Espinosa-Urgel, M., Kolter, R., and Ramos, J.-L. (2002) Root colonization by Pseudomonas putida: love at first sight. *Microbiology* **148**: 341-343.

Esteban, J., Pérez-Tanoira, R., Pérez-Jorge-Peremarch, C., and Gómez-Barrena, E. (2014) Chapter 3 - Bacterial Adherence to Biomaterials Used in Surgical Procedures. In *Microbiology for Surgical Infections*. Kon, K., and Rai, M. (eds). Amsterdam: Academic Press, pp. 41-57.

Evans, D.G., and Miller, M.H. (1990) The role of the external mycelial network in the effect of soil disturbance upon vesicular—arbuscular mycorrhizal colonization of maize. *New Phytologist* **114**: 65-71.

Finke, P.A., and Hutson, J.L. (2008) Modelling soil genesis in calcareous loess. *Geoderma* **145**: 462-479.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L. et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**: D279-285.

Flemming, H., and Wingender, J. (2010) The biofilm matrix. *Nature Reviews Microbiology* **8**: 623-633.

Flemming, H.C., Neu, T.R., and Wozniak, D.J. (2007) The EPS matrix: the "house of biofilm cells". *Journal of Bacteriology* **189**: 7945-7947.

Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., and Kjelleberg, S. (2016) Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* **14**: 563.

Fonouni-Farde, C., Tan, S., Baudin, M., Brault, M., Wen, J., Mysore, K.S. et al. (2016) DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. *Nature Communications* **7**: 12636.

Frey, S.D., Elliott, E.T., and Paustian, K. (1999) Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. *Soil Biology and Biochemistry* **31**: 573-585.

Frølund, B., Palmgren, R., Keiding, K., and Nielsen, P.H. (1996) Extraction of extracellular polymers from activated sludge using a cation exchange resin. *Water Research* **30**: 1749-1758.

Garnier-Sillam, E., and Harry, M. (1995) Distribution of humic compounds in mounds of some soil-feeding termite species of tropical rainforests: its influence on soil structure stability. *Insectes Sociaux* **42**: 167-185.

Gibbs, R.J., and Reid, J.B. (1988) A Conceptual Model of Changes in Soil Structure Under Different Cropping Systems. In *Advances in Soil Science*. Stewart, B.A. (ed). New York, NY: Springer New York, pp. 123-149.

Glaeser, S.P., and Kämpfer, P. (2014) The Family Sphingomonadaceae. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 641-707.

Gunina, A., and Kuzyakov, Y. (2015) Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. *Soil Biology and Biochemistry* **90**: 87-100.

Haft, D.H., Selengut, J.D., Richter, R.A., Harkins, D., Basu, M.K., and Beck, E. (2013) TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Research* **41**: D387-395.

Hallett, P.D., Feeney, D.S., Bengough, A.G., Rillig, M.C., Scrimgeour, C.M., and Young, I.M. (2009) Disentangling the impact of AM fungi versus roots on soil structure and water transport. *Plant and Soil* **314**: 183-196.

Harrison, R.B., and Strahm, B.D. (2008) Soil Formation. In *Encyclopedia of Ecology*. Jørgensen, S.E., and Fath, B.D. (eds). Oxford: Academic Press, pp. 3291-3295.

Hartemink, A.E. (2016) Chapter Two - The definition of soil since the early 1800s. In *Advances in Agronomy*. Sparks, D.L. (ed): Academic Press, pp. 73-126.

Heulin, T., Barakat, M., Christen, R., Lesourd, M., Sutra, L., De Luca, G., and Achouak, W. (2003) Ramlibacter tataouinensis gen. nov., sp. nov., and Ramlibacter henchirensis sp. nov., cyst-producing bacteria isolated from subdesert soil in Tunisia. *International Journal of Systematic and Evolutionary Microbiology* **53**: 589-594.

Hoyos-Santillan, J., Lomax, B.H., Turner, B.L., and Sjögersten, S. (2018) Nutrient limitation or home field advantage: Does microbial community adaptation overcome nutrient limitation of litter decomposition in a tropical peatland? *Journal of Ecology* **106**: 1558-1569.

Huggett, R.J. (1998) Soil chronosequences, soil development, and soil evolution: a critical review. *CATENA* **32**: 155-172.

Islam, S.T., and Lam, J.S. (2014) Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Canadian Journal of Microbiology* **60**: 697-716.

Jacques, M. (1996) Role of lipo-oligosaccharides and lipopolysaccharides in bacterial adherence. *Trends in Microbiology* **4**: 408-410.

Janczarek, M., and Rachwał, K. (2013) Mutation in the pssA gene involved in exopolysaccharide synthesis leads to several physiological and symbiotic defects in Rhizobium leguminosarum bv. trifolii. *International Journal of Molecular Sciences* **14**: 23711-23735.

Janczarek, M., Kutkowska, J., Piersiak, T., and Skorupska, A. (2010) Rhizobium leguminosarum bv. trifolii rosR is required for interaction with clover, biofilm formation and adaptation to the environment. *BMC Microbiology* **10**: 284.

Jenny, H. (1941) *Factors of soil formation: a system of quantitative pedology*. New York: McGraw-Hill Book Company, Inc.

Jensen, J.L., Schjønning, P., Watts, C.W., Christensen, B.T., Peltre, C., and Munkholm, L.J. (2019) Relating soil C and organic matter fractions to soil structural stability. *Geoderma* **337**: 834-843.

Joergensen, R.G. (1996) The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEC value. *Soil Biology and Biochemistry* **28**: 25-31.

Joergensen, R.G., and Mueller, T. (1996) The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEN value. *Soil Biology and Biochemistry* **28**: 33-37.

Johnson, A.M., and Hoyt, G.D. (1999) Changes to the Soil Environment under Conservation Tillage. *HortTechnology* **9**: 380-393.

Johnson, E.A., and Miyanishi, K. (2008) Testing the assumptions of chronosequences in succession. *Ecology letters* **11 5**: 419-431.

Jouquet, P., Bottinelli, N., Shanbhag, R.R., Bourguignon, T., Traoré, S., and Abbasi, S.A. (2016) Termites: The Neglected Soil Engineers of Tropical Soils. *Soil Science* **181**: 157-165.

Kalam, S., Basu, A., Ahmad, I., Sayyed, R.Z., El-Enshasy, H.A., Dailin, D.J., and Suriani, N.L. (2020) Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. *Frontiers in Microbiology* **11**.

Kassambara, A., and Mundt, F. (2017) factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7.

Kautz, T., Amelung, W., Ewert, F., Gaiser, T., Horn, R., Jahn, R. et al. (2013) Nutrient acquisition from arable subsoils in temperate climates: A review. *Soil Biology and Biochemistry* **57**: 1003-1022.

Kawasaki, S., Moriguchi, R., Sekiya, K., Nakai, T., Ono, E., Kume, K., and Kawahara, K. (1994) The cell envelope structure of the lipopolysaccharide-lacking gram-negative bacterium Sphingomonas paucimobilis. *Journal of bacteriology* **176**: 284-290.

Kay, B.D. (1990) Rates of Change of Soil Structure Under Different Cropping Systems. In *Advances in Soil Science 12: Volume 12*. Stewart, B.A. (ed). New York, NY: Springer New York, pp. 1-52.

Kay, B.D. (1997) Soil structure and organic carbon: A review. In *Soil Processes and the Carbon Cycle*. Lal, R., Kimble, J. M., Follett, R. F., Stewart, B. A. (ed). Boca Raton: CRC Press, pp. 169-197.

Kemper, W.D., and Rosenau, R.C. (1984) Soil Cohesion as Affected by Time and Water Content. *Soil Science Society of America Journal* **48**: 1001-1006.

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., and Kuramae, E.E. (2016) The Ecology of Acidobacteria: Moving beyond Genes and Genomes. *Frontiers in Microbiology* **7**: 744.

Kolb, E., Hartmann, C., and Genet, P. (2012) Radial force development during root growth measured by photoelasticity. *Plant and Soil* **360**: 19-35.

Kozlowski, T.T. (1999) Soil Compaction and Growth of Woody Plants. *Scandinavian Journal of Forest Research* **14**: 596-619.

Lal, R. (1991) Soil Structure and Sustainability. *Journal of Sustainable Agriculture* **1**: 67-92.

Lauber Christian, L., Hamady, M., Knight, R., and Fierer, N. (2009) Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Applied and Environmental Microbiology* **75**: 5111-5120.

Lehmann, A., Zheng, W., and Rillig, M.C. (2017) Soil biota contributions to soil aggregation. *Nature Ecology & Evolution* **1**: 1828-1835.

Lehmann, A., Zheng, W., Ryo, M., Soutschek, K., Roy, J., Rongstock, R. et al. (2020) Fungal Traits Important for Soil Aggregation. *Frontiers in Microbiology* **10**: 2904.

Lemtiri, A., Colinet, G., Alabi, T., Cluzeau, D., Zirbes, L., Haubruge, E., and Francis, F. (2014) Impacts of earthworms on soil components and dynamics. A review. *Biotechnologie, Agronomie, Société et Environnement* **18**: 121-133.

Li, Y., Chang, S.X., Tian, L., and Zhang, Q. (2018) Conservation agriculture practices increase soil microbial biomass carbon and nitrogen in agricultural soils: A global meta-analysis. *Soil Biology and Biochemistry* **121**: 50-58.

Li, Z., Nie, X., Chang, X., Liu, L., and Sun, L. (2016) Characteristics of Soil and Organic Carbon Loss Induced by Water Erosion on the Loess Plateau in China. *PLOS ONE* **11**: e0154591.

Limmathurotsakul, D., Paeyao, A., Wongratanacheewin, S., Saiprom, N., Takpho, N., Thaipadungpanit, J. et al. (2014) Role of Burkholderia pseudomallei biofilm formation and lipopolysaccharide in relapse of melioidosis. *Clinical Microbiology and Infection* **20**: O854-856.

Limoli Dominique, H., Jones Christopher, J., and Wozniak Daniel, J. (2015) Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiology spectrum* **3**: 223-247.

Liu, X., Bai, Z., Zhou, W., Cao, Y., and Zhang, G. (2017) Changes in soil properties in the soil profile after mining and reclamation in an opencast coal mine on the Loess Plateau, China. *Ecological Engineering* **98**: 228-239.

López-García, S.L., Vázquez, T.E., Favelukes, G., and Lodeiro, A.R. (2001) Improved soybean root association of N-starved Bradyrhizobium japonicum. *Journal of Bacteriology* **183**: 7241-7252.

Lu, X., Lu, X., and Liao, Y. (2018) Effect of Tillage Treatment on the Diversity of Soil Arbuscular Mycorrhizal Fungal and Soil Aggregate-Associated Carbon Content. *Frontiers in Microbiology* **9**.

Lucas, M., Schlüter, S., Vogel, H.-J., and Vetterlein, D. (2019) Soil structure formation along an agricultural chronosequence. *Geoderma* **350**: 61-72.

Lugtenberg, B.J.J., Dekkers, L., and Bloemberg, G.V. (2001) Molecular determinants of rhizosphere colonization by Pseudomonas. *Annual Review of Phytopathology* **39**: 461-490.

Lynch, J.M., and Bragg, E. (1985) Microorganisms and Soil Aggregate Stability. In *Advances in Soil Science*. Stewart, B.A. (ed). New York, NY: Springer New York, pp. 133-171.

Mair, P., and Wilcox, R. (2018) WRS2: Wilcox robust estimation and testing. R Package Version 1.1.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**: D222-D226.

Martin, J.P., Martin, W.P., Page, J.B., Raney, W.A., and de Ment, J.D. (1955) Soil Aggregation. In *Advances in Agronomy*. Norman, A.G. (ed): Academic Press, pp. 1-37.

Maruyama, Y., Itoh, T., Kaneko, A., Nishitani, Y., Mikami, B., Hashimoto, W., and Murata, K. (2015) Structure of a Bacterial ABC Transporter Involved in the Import of an Acidic Polysaccharide Alginate. *Structure* **23**: 1643-1654.

Mataix-Solera, J., Cerdà, A., Arcenegui, V., Jordán, A., and Zavala, L.M. (2011) Fire effects on soil aggregation: A review. *Earth-Science Reviews* **109**: 44-60.

McDaniel, M.D., Tiemann, L.K., and Grandy, A.S. (2014) Does agricultural crop diversity enhance soil microbial biomass and organic matter dynamics? A meta-analysis. *Ecological Applications* **24**: 560-570.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**: e61217.

McNeill, J.R., and Winiwarter, V. (2004) Breaking the Sod: Humankind, History, and Soil. *Science* **304**: 1629-1628.

Menge, D.N.L., Hedin, L.O., and Pacala, S.W. (2012) Nitrogen and Phosphorus Limitation over Long-Term Ecosystem Development in Terrestrial Ecosystems. *PLOS ONE* **7**: e42045.

Menzel, P., Ng, K.L., and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* **7**: 11257.

Morona, R., Purins, L., Tocilj, A., Matte, A., and Cygler, M. (2009) Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. *Trends in Biochemical Sciences* **34**: 78-84.

Mosier, S., Córdova, S.C., and Robertson, G.P. (2021) Restoring Soil Fertility on Degraded Lands to Meet Food, Fuel, and Climate Security Needs via Perennialization. *Frontiers in Sustainable Food Systems* **5**.

Moulin, L., Klonowska, A., Caroline, B., Booth, K., Vriezen, J.A.C., Melkonian, R. et al. (2014) Complete Genome sequence of Burkholderia phymatum STM815T, a broad host range and efficient nitrogen-fixing symbiont of Mimosa species. *Standards in Genomic Sciences* **9**: 763-774.

Mukhopadhyay, S., Masto, R.E., Yadav, A., George, J., Ram, L.C., and Shukla, S.P. (2016) Soil quality index for evaluation of reclaimed coal mine spoil. *Science of The Total Environment* **542**: 540-550.

Nicol Graeme, W., Tscherko, D., Embley, T.M., and Prosser James, I. (2005) Primary succession of soil Crenarchaeota across a receding glacier foreland. *Environmental Microbiology* **7**: 337-347.

Nimmo, J.R. (2005) Aggregation: Physical Aspects. In *Encyclopedia of Soils in the Environment*. Hillel, D. (ed). Oxford: Elsevier, pp. 28-35.

Nosalewicz, A., and Lipiec, J. (2014) The effect of compacted soil layers on vertical root distribution and water uptake by wheat. *Plant and Soil* **375**: 229-240.

Oades, J.M. (1993) The role of biology in the formation, stabilization and degradation of soil structure. *Geoderma* **56**: 377-400.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. (2018) vegan: Community Ecology Package. R Package Version 2.5.

Oleghe, E., Naveed, M., Baggs, E.M., and Hallett, P.D. (2017) Plant exudates improve the mechanical conditions for root penetration through compacted soils. *Plant and Soil* **421**: 19-30.

Pal, A., and Paul, A.K. (2008) Microbial extracellular polymeric substances: central elements in heavy metal bioremediation. *Indian Journal of Microbiology* **48**: 49-64.

Pereira, S., Mota, R., Santos, C., De Philippis, R., and Tamagnini, P. (2013) Chapter Seven - Assembly and Export of Extracellular Polymeric Substances (EPS) in Cyanobacteria: A Phylogenomic Approach. In *Advances in Botanical Research*. Chauvat, F., and Cassier-Chauvat, C. (eds): Academic Press, pp. 235-279.

Pereira, S.B., Mota, R., Vieira, C.P., Vieira, J., and Tamagnini, P. (2015) Phylum-wide analysis of genes/proteins related to the last steps of assembly and export of extracellular polymeric substances (EPS) in cyanobacteria. *Scientific reports* **5**: 14835-14835.

Pihlap, E., Steffens, M., and Kögel-Knabner, I. (2021) Initial soil aggregate formation and stabilisation in soils developed from calcareous loess. *Geoderma* **385**: 114854.

Pihlap, E., Vuko, M., Lucas, M., Steffens, M., Schloter, M., Vetterlein, D. et al. (2019) Initial soil formation in an agriculturally reclaimed open-cast mining area - the role of management and loess parent material. *Soil and Tillage Research* **191**: 224-237.

Plenchette, C., Clermont-Dauphin, C., Meynard, J.M., and Fortin, J.A. (2005) Managing arbuscular mycorrhizal fungi in cropping systems. *Canadian Journal of Plant Science* **85**: 31-40.

Poole, P., Ramachandran, V., and Terpolilli, J. (2018) Rhizobia: from saprophytes to endosymbionts. *Nature Reviews Microbiology* **16**: 291-303.

Powlson, D.S., Prookes, P.C., and Christensen, B.T. (1987) Measurement of soil microbial biomass provides an early indication of changes in total soil organic matter due to straw incorporation. *Soil Biology and Biochemistry* **19**: 159-164.

Putker, F., Bos, M.P., and Tommassen, J. (2015) Transport of lipopolysaccharide to the Gram-negative bacterial cell surface. *FEMS microbiology reviews* **39**: 985-1002.

Quelas, J.I., López-García, S.L., Casabuono, A., Althabegoiti, M.J., Mongiardini, E.J., Pérez-Giménez, J. et al. (2006) Effects of N-starvation and C-source on Bradyrhizobium japonicum exopolysaccharide production and composition, and bacterial infectivity to soybean roots. *Archives of Microbiology* **186**: 119-128.

Quelas, J.I., Mongiardini, E.J., Casabuono, A., López-García, S.L., Althabegoiti, M.J., Covelli, J.M. et al. (2010) Lack of galactose or galacturonic acid in Bradyrhizobium japonicum USDA 110 exopolysaccharide leads to different symbiotic responses in soybean. *Molecular Plant-Microbe Interactions* **23**: 1592-1604.

R Core Team (2018) R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Raetz, C.R., and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annual Review of Biochemistry* **71**: 635-700.

Redmile-Gordon, M., Brookes, P., Evershed, R., Goulding, K., and Hirsch, P. (2014) Measuring the soil-microbial interface: Extraction of extracellular polymeric substances (EPS) from soil biofilms. *Soil Biology and Biochemistry* **72**: 163-171.

Redmile-Gordon, M.A., Armenise, E., White, R.P., Hirsch, P.R., and Goulding, K.W.T. (2013) A comparison of two colorimetric assays, based upon Lowry and Bradford techniques, to estimate total protein in soil extracts. *Soil biology & biochemistry* **67**: 166-173.

Redmile-Gordon, M.A., Evershed, R.P., Hirsch, P.R., White, R.P., and Goulding, K.W.T. (2015) Soil organic matter and the extracellular microbial matrix show contrasting responses to C and N availability. *Soil Biology and Biochemistry* **88**: 257-267.

Rehm, B.H. (2010) Bacterial polymers: biosynthesis, modifications and applications. *Nature Reviews Microbiology* **8**: 578-592.

Reitz, M., Rudolph, K., Schröder, I., Hoffmann-Hergarten, S., Hallmann, J., and Sikora, R.A. (2000) Lipopolysaccharides of Rhizobium etliStrain G12 Act in Potato Roots as an Inducing Agent of Systemic Resistance to Infection by the Cyst Nematode Globodera pallida. *Applied and Environmental Microbiology* **66**: 3515-3518.

Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research* **38**: e191.

Ritschel, T., and Totsche, K.U. (2019) Modeling the formation of soil microaggregates. *Computers & Geosciences* **127**: 36-43.

Rodriguez, R.L., and Konstantinidis, K.T. (2014) Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**: 629-635.

Ross, P., Mayer, R., and Benziman, M. (1991) Cellulose biosynthesis and function in bacteria. *Microbiological Reviews* **55**: 35-58.

Rossi, F., and De Philippis, R. (2015) Role of Cyanobacterial Exopolysaccharides in Phototrophic Biofilms and in Complex Microbial Mats. *Life* **5**: 1218-1238.

Rowley, M.C., Grand, S., and Verrecchia, É.P. (2018) Calcium-mediated stabilisation of soil organic carbon. *Biogeochemistry* **137**: 27-49.

Roy, J., Reichel, R., Bruggemann, N., Hempel, S., and Rillig, M.C. (2017) Succession of arbuscular mycorrhizal fungi along a 52-year agricultural recultivation chronosequence. *FEMS Microbiology Ecology* **93**.

Russo, D.M., Abdian, P.L., Posadas, D.M., Williams, A., Vozza, N., Giordano, W. et al. (2015) Lipopolysaccharide O-chain core region required for cellular cohesion and compaction of in vitro and root biofilms developed by Rhizobium leguminosarum. *Applied and environmental microbiology* **81**: 1013-1023.

Rylander, R. (2002) Review: Endotoxin in the environment — exposure and effects. *Journal of Endotoxin Research* **8**: 241-252.

Sanin, S.L., Sanin, F.D., and Bryers, J.D. (2003) Effect of starvation on the adhesive properties of xenobiotic degrading bacteria. *Process Biochemistry* **38**: 909-914.

Sawasdidoln, C., Taweechaisupapong, S., Sermswan, R.W., Tattawasart, U., Tungpradabkul, S., and Wongratanacheewin, S. (2010) Growing Burkholderia pseudomallei in biofilm stimulating conditions significantly induces antimicrobial resistance. *PLoS One* **5**: e9196.

Schmid, C.A.O., Reichel, R., Schröder, P., Brüggemann, N., and Schloter, M. (2020) 52 years of ecological restoration following a major disturbance by opencast lignite mining does not reassemble microbiome structures of the original arable soils. *Science of The Total Environment* **745**: 140955.

Schmid, J., Sieber, V., and Rehm, B. (2015) Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Frontiers in Microbiology* **6**: 496.

Schmieder, R., and Edwards, R. (2011) Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS One* **6**: e17288.

Schneider, F., and Don, A. (2019) Root-restricting layers in German agricultural soils. Part I: extent and cause. *Plant and Soil* **442**: 433-451.

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

Shrestha, R.K., and Lal, R. (2011) Changes in physical and chemical properties of soil after surface mining and reclamation. *Geoderma* **161**: 168-176.

Sinnett, D., Poole, J., and Hutchings, T.R. (2008) A comparison of cultivation techniques for successful tree establishment on compacted soil. *Forestry: An International Journal of Forest Research* **81**: 663-679.

Six, J., Elliott, E.T., and Paustian, K. (2000) Soil Structure and Soil Organic Matter II. A Normalized Stability Index and the Effect of Mineralogy. *Soil Science Society of America Journal* **64**: 1042-1049.

Six, J., Bossuyt, H., Degryze, S., and Denef, K. (2004) A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research* **79**: 7-31.

Skorupska, A., Janczarek, M., Marczak, M., Mazur, A., and Król, J. (2006) Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microbial cell factories* **5**: 7-7.

Sorroche, F., Bogino, P., Russo, D.M., Zorreguieta, A., Nievas, F., Morales, G.M. et al. (2018) Cell Autoaggregation, Biofilm Formation, and Plant Attachment in a Sinorhizobium meliloti lpsB Mutant. *Molecular Plant-Microbe Interactions* **31**: 1075-1082.

Stefanakis, A., Akratos, C.S., and Tsihrintzis, V.A. (2014) Chapter 5 - Treatment Processes in VFCWs. In *Vertical Flow Constructed Wetlands*. Stefanakis, A., Akratos, C.S., and Tsihrintzis, V.A. (eds). Boston: Elsevier, pp. 57-84.

Stegen, J.C., Lin, X., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal* **6**: 1653-1664.

Stevens, P.R., and Walker, T.W. (1970) The Chronosequence Concept and Soil Formation. *The Quarterly Review of Biology* **45**: 333-350.

Stirzaker, R.J., Passioura, J.B., and Wilms, Y. (1996) Soil structure and plant growth: Impact of bulk density and biopores. *Plant and Soil* **185**: 151-162.

Stromberg, L., Mendez, H., and Mukundan, H. (2017) Detection Methods for Lipopolysaccharides: Past and Present. In *Escherichia coli - Recent Advances on Physiology, Pathogenesis and Biotechnological Applications*. Amidou, S. (ed). Rijeka: IntechOpen, pp. 141–168.

Sutherland, I.W. (2001) Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology (Reading, England)* **147**: 3-9.

Tisdall, J.M., and Oades, J.M. (1979) Stabilization of soil aggregates by the root systems of ryegrass. *Soil Research* **17**: 429-441.

Tisdall, J.M., and Oades, J.M. (1982) Organic matter and water-stable aggregates in soils. *Journal of Soil Science* **33**: 141-163.

Totsche, K.U., Amelung, W., Gerzabek, M.H., Guggenberger, G., Klumpp, E., Knief, C. et al. (2018) Microaggregates in soils. *Journal of Plant Nutrition and Soil Science* **181**: 104-136.

Tripathi, B.M., Stegen, J.C., Kim, M., Dong, K., Adams, J.M., and Lee, Y.K. (2018) Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *The ISME Journal* **12**: 1072-1083.

Tunlid, A., and White, D. (1992) Biochemical analysis of biomass, community structure, nutritional status, and metabolic activity of microbial communities in soil. In *Soil biochemistry*. Stotzky, G., and Bollag, J. (eds). New York: Marcel Dekker, pp. 229-262.

Ussiri, D.A.N., Lal, R., and Jacinthe, P.A. (2006) Soil Properties and Carbon Sequestration of Afforested Pastures in Reclaimed Minesoils of Ohio. *Soil Science Society of America Journal* **70**: 1797-1806.

Vance, E.D., Brookes, P.C., and Jenkinson, D.S. (1987) An extraction method for measuring soil microbial biomass C. *Soil Biology and Biochemistry* **19**: 703-707.

Vartoukian, S.R., Palmer, R.M., and Wade, W.G. (2010) Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters* **309**: 1-7.

Velmourougane, K., Prasanna, R., and Saxena, A.K. (2017) Agriculturally important microbial biofilms: Present status and future prospects. *Journal of Basic Microbiology* **57**: 548-573.

Vestergaard, G., Schulz, S., Schöler, A., and Schloter, M. (2017) Making big data smart—how to use metagenomics to understand soil quality. *Biology and Fertility of Soils* **53**: 479-484.

Vidal, A., Watteau, F., Remusat, L., Mueller, C.W., Nguyen Tu, T.-T., Buegger, F. et al. (2019) Earthworm Cast Formation and Development: A Shift From Plant Litter to Mineral Associated Organic Matter. *Frontiers in Environmental Science* **7**: 55.

Vitousek, P.M., Matson, P.A., and Van Cleve, K. (1989) Nitrogen availability and nitrification during succession: Primary, secondary, and old-field seres. *Plant and Soil* **115**: 229-239.

Vogel, H.-J., Balseiro-Romero, M., Kravchenko, A., Otten, W., Pot, V., Schlüter, S. et al. (2022) A holistic perspective on soil architecture is needed as a key to soil functions. *European Journal of Soil Science* **73**: e13152.

von Bodman, S.B., Majerczak, D.R., and Coplin, D.L. (1998) A negative regulator mediates quorum-sensing control of exopolysaccharide production in Pantoea stewartii subsp. stewartii. *Proceedings of the National Academy of Sciences* **95**: 7687-7692.

Vuko, M., Cania, B., Vogel, C., Kublik, S., Schloter, M., and Schulz, S. (2020) Shifts in reclamation management strategies shape the role of exopolysaccharide and lipopolysaccharide-producing bacteria during soil formation. *Microbial Biotechnology* **13**: 584–598.

Walker, L.R., Wardle, D.A., Bardgett, R.D., and Clarkson, B.D. (2010) The use of chronosequences in studies of ecological succession and soil development. *Journal of Ecology* **98**: 725-736.

Wang, X., and Quinn, P.J. (2010) Lipopolysaccharide: Biosynthetic pathway and structure modification. *Progress in Lipid Research* **49**: 97-107.

Wei, X., Reich, P.B., and Hobbie, S.E. (2019) Legumes regulate grassland soil N cycling and its response to variation in species diversity and N supply but not CO2. *Global Change Biology* **25**: 2396-2409.

White, T.J., Bruns, T., Lee, S., Taylor, J., A Innis, M., H Gelfand, D., and Sninsky, J. (1990) Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In *PCR Protocols, a Guide to Methods and Applications*. Innis, M.A., Gelfand, D.H., Shinsky, J.J., and White, T.J. (eds). San Diego: Academic Press, Inc., pp. 315-322.

Whitfield, C., and Trent, M.S. (2014) Biosynthesis and Export of Bacterial Lipopolysaccharides. *Annual review of biochemistry* **83**: 99-128.

Whitney, J.C., and Howell, P.L. (2013) Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends in microbiology* **21**: 63-72.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R. et al. (2019) Welcome to the Tidyverse. *Journal of Open Source Software* **4**: 1686.

Willems, A. (2014) The Family Comamonadaceae. In *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 777-851.

Williams, A., Wilkinson, A., Krehenbrink, M., Russo, D.M., Zorreguieta, A., and Downie, J.A. (2008) Glucomannan-mediated attachment of Rhizobium leguminosarum to pea root hairs is required for competitive nodule infection. *Journal of Bacteriology* **190**: 4706-4715.

Willis, L.M., and Whitfield, C. (2013) Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydrate Research* **378**: 35-44.

Xia, W., Zhang, C., Zeng, X., Feng, Y., Weng, J., Lin, X. et al. (2011) Autotrophic growth of nitrifying community in an agricultural soil. *The ISME Journal* **5**: 1226.

Xie, C., Goi, C.L.W., Huson, D.H., Little, P.F.R., and Williams, R.B.H. (2016) RiboTagger: fast and unbiased 16S/18S profiling using whole community shotgun metagenomic or metatranscriptome surveys. *BMC Bioinformatics* **17**: 508.

Xing, T.-t., Cai, A.-d., Lu, C.-a., Ye, H.-l., Wu, H.-l., Huai, S.-c., Wang, S.-c., Xu, M.-g., Lin, Q.-m. (2022) Increasing soil microbial biomass nitrogen in crop rotation systems by improving nitrogen resources under nitrogen application. *Journal of Integrative Agriculture* **21**: 1488-1500.

Yang, Y., Zhang, Y., Jia, G., Wang, L., and Yu, X. (2022) Contributions of biotic and abiotic factors to soil aggregation under different thinning intensities. *Ecological Indicators* **139**: 108958.

Yates, K., Fenton, C.H., and Bell, D.H. (2018) A review of the geotechnical characteristics of loess and loess-derived soils from Canterbury, South Island, New Zealand. *Engineering Geology* **236**: 11-21.

Young, I.M. (1998) Biophysical interactions at the root–soil interface: a review. *The Journal of Agricultural Science* **130**: 1-7.

Zeng, Y., and Charkowski, A.O. (2021) The Role of ATP-Binding Cassette Transporters in Bacterial Phytopathogenesis. *Phytopathology* **111**: 600-610.

Zotarelli, L., Alves, B.J.R., Urquiaga, S., Torres, E., dos Santos, H.P., Paustian, K. et al. (2005) Impact of Tillage and Crop Rotation on Aggregate-Associated Carbon in Two Oxisols. *Soil Science Society of America Journal* **69**: 482-491.

Zwanzig, L., Zwanzig, M., and Sauer, D. (2021) Outcomes of a quantitative analysis of 48 soil chronosequence studies in humid mid and high latitudes: Importance of vegetation in driving podzolization. *CATENA* **196**: 104821.

Acknowledgements

First of all, I want to express my gratitude to Prof. Dr. Michael Schloter and Dr. Stefanie Schulz for the opportunity to embark on this journey. Thank you for your trust, patience, and support while I was learning how to navigate the amazing world of science. I have truly learned a lot from you.

I also want to thank our project partners: Prof. Dr. Ingrid Kögel-Knabner, Prof. Dr. Doris Vetterlein, Dr. Markus Steffens, Dr. Evelin Pihlap, and Dr. Maik Lucas, for all the fascinating discussions and wonderful collaboration. In addition, I am grateful to Dr. Cordula Vogel, whose work on EPS extraction elevated the results presented here. Thank you also to Susanne Kublik and Gudrun Hufnagel for all your help in the lab.

I consider myself very lucky for having met all my wonderful colleagues at COMI. First thanks go to my best-office people: Marta, Jenny, Luhua, David, Roberto, (+ Antonis), for all the laughs and support. Next big thank you goes to the rest of the COMIs for the wonderful lunch and coffee breaks, which were always filled with interesting and stimulating discussions, even when they revolved around the most banal things. Thank you Matea, Urška, Sebastian, Vivi, Catarina, Christoph, Helga, Yvonne, Akane, Clara, Michi, Nicole, Yuri, Kate, and others (too many to count!). All our scientific discussions and brainstorming were an invaluable help in shaping this work and in my own scientific development. All of you have made this PhD a really special period that I will always remember fondly.

Reaching this finish line also took a lot of external support from friends and family. Dearest Anja, you accompanied me (both knowingly and unknowingly) on my numerous writing guilt trips over the last few years. And you did so with incredible patience and understanding. I don’t think I would have been able to reach this point without your support. Thank you for putting up with me (green heart emoji). I am forever grateful to my Balkan Cluster; meeting you two was a wonderful clash of lucky coincidences and it made all the difference. Dear Barb, thank you for reminding me what true friendship is. All of you have been there for me through some of my worst moments in life and kept me from collapsing in on myself, thank you.

Finally, thank you to my friends and family back home; knowing that I have a base I can always return to is a real gift. Thank you for all the fun during vacations, which allowed me to decompress from time to time while putting in the work necessary to finish this thesis.

My dear Mare, Jele, Željka, and others always made sure of that – thank you! Finally, I want to express my deepest gratitude to my loving parents who always supported me in all my endeavors.

Hvala mojim kućnim Vukićima na potpori i svim kavama, ručkovima, večerama, smiju i zabavi kad sam doma. Posebno hvala roditeljima koji su mi omogućili ovaj, i sve prethodne korake; bez vaše ljubavi i potpore nikad ne bih stigla do ovdje.