Analysis of temporal and spatial variation of bacterial diversity in temperate soils

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades
eines Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Carlo Robert Marzini
aus Würzburg

1. Referent: Professor Dr. Jörg Overmann
2. Referent: Professor Dr. Michael Steinert

eingereicht am: 04.07.2023

mündliche Prüfung (Disputation) am: 08.12.2023

Druckjahr 2024

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Posterbeiträge

Marzini, C., Sikorski, J., Overmann, J.: Profund 2: Prokaryotic functional diversity under different land use. (Poster) 11. Biodiversity Exploratories Assembly (2018)

Marzini, C., Vieira, S., Sikorski, J., Overmann, J.: Prokaryotic functional diversity under different land use. (Poster) 12. Biodiversity Exploratories Assembly (2019)

Marzini, C., Overmann, J.: Characteristics between free-living and sessile bacteria in soil. (Poster) 13. Biodiversity Exploratories Assembly (2020)

Contents

Abstract **V**

Zusammenfassung **VII**

1 Introduction **1**

1.1 Next generation sequencing of soil bacterial communities 1
1.2 Determination of bacterial activity in soil 2
1.3 Temporal variation and climate change of bacterial communities in soil . . . 3
1.4 Spatial variation of soil bacteria 4
1.5 Aim of this work 5

2 Does climate change affect soil bacterial communities already? Results from a German large scale study **6**

2.1 Starting point of the investigation 6
2.2 Material and Methods 11
2.2.1 Study design 11
2.2.2 Nucleic acid extraction and library preparation 11
2.2.3 Bioinformatic processing of sequencing data 13
2.2.4 Metadata collection and gap filling 13
2.2.5 Environmental predictor engineering 14
2.2.6 Predictor selection and data structure 15
2.2.7 Sequencing data normalization 16
2.2.8 Alpha diversity analysis 16
2.2.9 Beta diversity analysis 19
2.2.10 Statistical niche modelling 21
2.3 Results 23
2.3.1 Data structure 23
2.3.2 Sequencing results and community data structure 26
2.3.3 Drivers of alpha diversity 32
2.3.4 Generalized dissimilarity modelling of beta diversity for grassland and forest sites 37
2.3.5 GradientForest analysis of beta diversity for grassland and forest sites 41
2.3.6 Ecological niche modelling 49
2.4 Discussion 53
2.4.1 The abundance of rare and abundant soil bacteria declined over time 53
2.4.2 Climatic controls of alpha diversity 55
2.4.3 Edaphic predictors and alpha diversity 57

2.4.4 Community turnover is resistant to climate and climate change . . . 58
2.4.5 Statistical niche modelling based on gradientForest analysis 60
2.5 Conclusion 62

3 Bacterial life-style not fertilization shapes soil bacteria communities from a long-term agricultural fertilization experiment 63
3.1 Starting point of the study 63
3.2 Material and Methods 66
3.2.1 Study design 66
3.2.2 Soil sampling 67
3.2.3 Separation of free-living and surface-associated bacteria by Nycodenz extraction 68
3.2.4 Direct cell counting 70
3.2.5 Nucleic acid extraction and library preparation 71
3.2.6 Bioinformatics 73
3.2.7 Sequencing data normalization 74
3.2.8 Calculation of alpha and beta diversity 74
3.2.9 Determination of metabolic active bacteria 74
3.2.10 Statistical analysis 75
3.3 Results 76
3.3.1 Cell extraction from soil differs by bacterial life-style and not by fertilization 76
3.3.2 Sequencing results and contamination removal 81
3.3.3 Alpha diversity is shaped by bacterial life-style but not soil fertilization 83
3.3.4 Beta diversity of total and active communities is shaped by bacterial life-style and fertilization 84
3.3.5 Taxonomic composition of the total community of free-living and surface-associated bacteria 90
3.3.6 Taxonomic composition of the active free-living and surface-associated communities 96
3.4 Discussion 104
3.4.1 Evaluation of cell extraction of free-living and surface-associated bacteria from soil 104
3.4.2 Nitrogen fertilization and a free-living bacterial life-style reduce bacterial alpha diversity 105
3.4.3 Bacterial life-style dominates beta diversity of the total community . 106
3.4.4 Beta diversity of active bacteria is tighter bound to bacterial life-style and fertilization than the total communities 108

3.4.5 Determination of active bacteria 109
3.5 Conclusions 111

4 Synthesis and future perspectives **112**

References **114**

Abstract

Soils contain up to 10^{10} bacteria g^{-1} soil. Due to this high species diversity and the difficult distinction between free bacterial DNA, inactive and active bacteria by next generation sequencing, soils harbour one of the most complex bacterial communities. Moreover, spatial heterogeneity and temporal trends influence soil bacterial interaction with the environment and between cells. Spatial effects influence bacterial interaction by high spatial heterogeneity due to soil structure, pore distribution nutrient and water availability having a patchy distribution in soil in unsaturated water conditions. In addition, soil bacterial communities are influenced by the soil environment and climate on larger spatial scales. On a global scale, recent studies showed that the influence of climatic conditions was exceeding the impact of established predictors such as pH. Furthermore, environmental predictors exert lagged effects on bacterial communities, called legacy effects, representing the delayed influence of past environmental or climatic conditions on the current microbial community. It is challenging to distinct all the above mentioned effects in microbial community analysis.

In this thesis, I present two studies examining the environmental, spatial and temporal effects in bacterial communities of arable, grassland and forest soils under temperate climate conditions. First, I examined how temporal variation, climatic predictors and climate change affect bacterial communities of grassland and forest soils, taking the effects of edaphic predictors into account. This project was part of the Biodiversity Exploratories, a large-scale, long-term monitoring project, focussing on the interaction between land use, land use intensity and multitrophic biodiversity across three regions in Germany.
I could show that bacterial alpha diversity is declining in German grassland and forest soils from 2011 to 2017. Soil warming could in part explain this decline. Furthermore, soil moisture and soil temperature contained legacy effects, showing delayed reactions of soil bacteria to soil temperature and soil moisture changes. This is the first observational study showing that changes in climate decline soil bacterial diversity in Germany, without experimental manipulation introducing strong drought or soil warming treatments. In contrast to alpha diversity, beta diversity was more stable over time. Soil bacterial communities are considered robust and stable over time. Hence, this emphasizes the importance of the results reported here.

In the second study, bacteria from an agricultural long-term fertilization experiment were extracted from soil and separated by bacterial life-style. Thereby the free-living bacteria, living in the aqueous pore space, and the surface-associated bacteria, living attached to soil particles, were compared. Furthermore, the potential activity of the free-living and surface-associated bacteria were determined. This experiment examined the long-term influence

of different nitrogen fertilization treatments on bacterial life-style and activity in an arable soil. Theory suggests that biofilm formation and a surface-associated life-style support bacterial activity. However little evidence is available how the distinction in bacterial life-style affects community composition and bacterial activity in soil. In contrast to literature, I could show that active bacteria dominated the free-living community. Moreover, the induced changes in community composition by nitrogen fertilization were larger for the active free-living and surface-associated communities compared to the total communities, containing active and non-active bacteria. This leads to the conclusion that bacterial activity is determined by other parameters than life-style in bulk soil.

Zusammenfassung

Böden enthalten bis zu 10^{10} Bakterien g^{-1} Boden. Neben dieser hohen Abundanz und der schwierigen methodischen Trennung zwischen freier bakterieller DNS, inaktiven und aktiven Bakterien mittels Sequenzierung, enthalten Böden eine der komplexesten bakteriellen Gemeinschaften. Zudem beeinflussen die gegebene räumliche Heterogenität in Böden und die zeitliche Variabilität bakterielle Interaktionen mit der Bodenumgebung und zwischen einzelnen Bakterien. Weiterhin beeinflussen das Klima und die Bodenumgebung die bakterielle Gemeinschaft. Derzeitiger Stand des Wissens ist, dass, auf globaler Ebene, Klimafaktoren wichtiger für die Zusammensetzung der bakteriellen Gemeinschaft im Boden sind als etablierte Bodenparameter wie der pH. Es stellt eine große Herausforderung für die Analyse der mikrobiellen Gemeinschaft dar, all diese Effekte voneinander zu trennen. Räumliche Effekte beeinflussen bakterielle Interaktionen durch hohe räumliche Variabilität, Bodenstruktur und Porenverteilung. Außerdem zeigt sich unter ungesättigten Bedingungen, eine ungleichmäßige Verteilung der Wasser- und Nährstoffverfügbarkeit. Darüber hinaus können Umweltfaktoren wie Klima zeitlich verzögert Einfluss auf die bakterielle Gemeinschaft im Boden ausüben, wobei vergangene Klimabedingungen die derzeitige Bakteriengemeinschaft stärker beeinflussen können als die derzeitigen Bedingungen.

In dieser Arbeit präsentiere ich zwei Einzelstudien, die die räumlichen und zeitlichen Effekte auf die bakterielle Gemeinschaft in landwirtschaftlich genutzen, sowie Grünland- und Waldböden in Deutschland untersuchen. In der ersten Studie untersuche ich, wie sich die Bodenbakterien in Grünland- und Waldböden unter Berücksichtigung der bodenbürtigen Umweltfaktoren, im Hinblick auf den Zeitverlauf, die klimatischen Bedingungen und Klimawandel verhalten. Dieses Projekt war Teil der Biodiversitätsexploratorien, einem Langzeitprojekt, das die Erforschung von multitrophischer Diversität verteilt über drei Regionen in Deutschland in Bezug auf Landnutzung und Landnutzungsintensität untersucht.
Ich konnte zeigen, dass die bakterielle Alphadiversität in Grünland- und Waldböden von 2011 bis 2017 signifikant abnahm. Ein Anstieg der Bodentemperaturen ist zum Teil für diesen Effekt verantwortlich. Außerdem reagiereten die Bodenbakterien verzögert auf Änderungen der Bodentemperatur und -feuchte. Dies ist die erste Monitoringstudie in Deutschland die eine Abnahme der bakteriellen Alphadiverisität unter den derzeitigen Klimabedingungen ohne experimentelle Manipulation nachweist.
Im Gegensatz dazu war die bakterielle Betadiversität stabiler. Es traten zwar signifikante zeitliche Effekte auf, allerdings waren die Auswirkungen auf die Änderung der Zusam-

mensetzung der bakteriellen Gemeinschaft geringer als für die Alphadiversität. Aufgrund der erwarteten zeitlichen Stabilität von Bakteriengemeinschaften in Böden sind diese Ergebnisse von Bedeutung für die kommende Entwicklung unter sich zunehmend ändernden Klimabedingungen.

In der zweiten Studie wurden die Bodenbakterien aus einem landwirtschaftlichen Langzeitdüngeexperiment nach Lebensstil getrennt und aus dem Boden extrahiert. Dieses Experiment untersucht die Langzeiteffekte unterschiedlicher Stickstoffdüngeintensität auf den Lebensstil und die Aktivität von Bakterien in einem landwirtschaftlich genutzten Boden. Dabei wurde die freilebende bakterielle Gemeinschaft, die sich im Porenraum des Bodens befindet, und die oberflächenassoziierte Gemeinschaft, die angeheftet an Bodenpartikel lebt, verglichen. Darüber hinaus wurden die Anteile und die Zusammensetzung an aktiven Bakterien in den jeweiligen Gemeinschaften untersucht.
Die Theorie besagt bisher, dass die Ausbildung von Biofilmen, unabhängig vom Lebensraum, bakterielle Aktivität in einem oberflächenassoziierten Lebensstil unterstützt. Allerdings gibt es kaum experimentelle Ergebnisse, die diese These für Böden stützen. Im Gegensatz zu den Erwartungen konnte ich zeigen, dass aktive Bakterien die freilebende bakterielle Gemeinschaft im Boden dominierten. Außerdem waren die Änderungen der Artzusammensetzung durch die Unterschiede in der Stickstoffdüngung größer für die jeweilige aktive Gemeinschaft als die gesamte Gemeinschaft, die sowohl potentiell aktive als auch inaktive Bakterien enthält. Diese Ergebnisse legen nahe, dass für die bakterielle Aktivität in Oberböden landwirtschaftlich genutzter Flächen andere Einflussfaktoren als der Lebensstil von Bedeutung sind.

1 Introduction

Soils are diverse multitrophic environments. Especially microorganisms are highly abundant in soil. More than 10^{10} bacteria can live in a single gram of soil (Torsvik et al., 1989) harbouring up to 10^6 bacterial phylotypes (Thompson et al., 2017; Bahram et al., 2018). However, this immense diversity was not know for a long time, since cultivation provided the only access to soil bacterial diversity. Another rough estimate of bacterial diversity in soil is the analysis of phospholipid fatty acid composition, denaturating gradient electrophoresis and terminal restriction fragment length polymorphism. These methods share the property that they are not able to capture the vast bacterial diversity in soil, e.g. whole phyla escaped cultivation and hence recognition in microbial ecology. A great proportion of the progress gained in the last decades in microbial ecology was enabled by next generation sequencing. The in depth analysis of marker genes such as variable regions of the 16S ribosomal RNA gene and transcripts unravelled the sheer magnitude of bacterial diversity in soils (Thompson et al., 2017). However, the majority of bacterial phylotypes remain uncultivated to this day (Overmann et al., 2017). Inferring the physiological properties of soil bacteria by targeted cultivation and ecological functions from sequencing data by marker gene analysis, metagenomics and transcriptomics is a major aim of soil microbial ecology.

1.1 Next generation sequencing of soil bacterial communities

The culture independent approach of next generation sequencing technology offers deep insights in bacterial diversity in soil. In addition, it enables in situ analysis of experimental or observational studies of soil bacterial diversity.

Initially, analysis of sequencing data was performed after assigning sequences to operational taxonomic units (OTUs). However, over the last years, lower error rates of sequencing platforms and reference-free sequencing error removal calls for the change from operational taxonomic unit clustering to amplicon sequence variants (SV). Those can resolve bacterial phylotypes down to single nucleotide differences. Furthermore, SV are comparable and reproducible across studies, since they are independent of a reference database (Callahan et al., 2017).

The achieved progress, however, could not resolve intrinsic limitations of next generation sequencing data. SV are usually mapped using a taxonomic reference data base to infer the taxonomic community composition and estimate traits of community members. The underlying hypothesis is that phenotypic traits are assumed to be more similar for related

bacteria (Allison, 2019). Morrissey et al. (2019) could show that evolutionary history and taxonomic groupings were more predictive for soil bacteria than the ecosystem type. However, the conservation of traits depends on the complexity, with traits involving more genes being more deeply conserved on higher taxonomic levels than traits including less traits (Allison, 2019). Another approach is the prediction of functional profiles from metagenomic data bases by matching the sequenced regions of the 16S rRNA gene (Aßhauer et al., 2015). Both approaches can only serve as a very rough estimate of potential properties from bacterial communities. A more precise functional profile of microbial communities can be gained by metagenomics or metatranscriptomics allowing for the analysis of the community's genetic and functional potential or the transcribed genes, respectively.

1.2 Determination of bacterial activity in soil

Sequencing analysis of marker genes cannot differentiate active from non-active bacteria. This is important, since the vast majority of soil bacteria is estimated to be dormant, non-active or simply dead (Blagodatskaya and Kuzyakov, 2013), remaining as cell debris in soil, harbouring relic DNA which can still be sequenced (Carini et al., 2016). The major cause for the dominance of dormancy are unfavourable environmental conditions, such as lack of nutrients (Jones and Lennon, 2010). To avoid the bias introduced by dormant community members, different approaches were developed estimating bacterial activity. However, most methods are not capable of high-throughput and are limited to a rather small sample size or scale such as nano-scale secondary ion mass spectrometry, bromodeoxyuridine incorporation, (quantitative) stable isotope probing and cell staining methods, e.g. using 5-cyano-2,3-ditolyl tetrazolium chloride. All those methods rely on the uptake of a labelling compound into the cells or incorporation into metabolites or DNA, indicating an active cell metabolism or bacterial growth. Another early attempt accessing the active community was to sequence rRNA transcripts instead of the rRNA gene, since the content of rRNA and growth rates correlated well in pure culture studies (Kerkhof and Ward, 1993; Blazewicz et al., 2013). However, it was shown that inactive cells can maintain substantial ribosomal content (Sukenik et al., 2012; Blazewicz et al., 2013). Nonetheless, the analysis of rRNA transcripts avoids bias introduced by dead bacterial cells or relic DNA, providing a more realistic representation of the live bacteria.
Normalizing the rRNA transcripts by rRNA gene abundance is another widely used method to determine bacterial activity (Campbell et al., 2011; Bowsher et al., 2019). In a second step, bacterial activity is determined for each SV by applying a threshold to the respective rRNA:rDNA ratio. To a great advantage, this method relying on the rRNA:rDNA ratio offers high throughput at comparable low costs (Bowsher et al., 2019). In this thesis, I use a new dynamic threshold to determine bacteria activity in soil.

1.3 Temporal variation and climate change of bacterial communities in soil

The majority of studies dealing with temporal variation of soil bacteria measured the reaction of bacterial biomass (Fierer et al., 2009), enzyme activities or soil carbon mineralization to climate change, but did not access bacterial diversity in depth (Lipson et al., 2002; Lipson, 2007; Fierer et al., 2010; Melillo et al., 2017). The reaction of those parameters examined is quite consistent, showing sensitivity towards soil warming (Romero-Olivares et al., 2017; Walker et al., 2018). Carbon mineralization in response to experimental warming over more than 20 years revealed a multistage process (Melillo et al., 2017), starting with a short term increase in mineralization (Davidson et al., 2012), followed by a reduction in carbon mineralisation back to control levels (Melillo et al., 2002, 2017) and reduced biomass. In contrast, soil bacteria community composition is considered to be one of the most stable over time (Lauber et al., 2013; Shade et al., 2013). In studies using next generation sequencing, this might be biased by relic DNA obscuring temporal pattern (Carini et al., 2016). With respect to climate change, which could induce changes over time, the reaction of soil bacterial diversity is inconsistent. No significant change in soil bacterial community composition was found after 20 years of experimental warming (DeAngelis et al., 2015). Others confirmed no effect on soil bacteria after up to 5 years of warming (Andrade-Linares et al., 2021; Sünnemann et al., 2021). In contrast, if the trends are of sufficient duration and strength, community composition is reactive to climate (Cruz-Martínez et al., 2009; Zhou et al., 2020). Furthermore, Guo et al. (2018) found a significant effect of warming on beta diversity. This difference in results might be a consequence of several reasons: either differences in experimental design, local climate conditions, duration and strength of the applied treatments or fundamental differences in the examined ecosystem types, since the response of microbial communities to temporal variation is not similar between ecosystem types (Blankinship et al., 2011; Pold et al., 2015; Kivlin and Hawkes, 2020). For example, Hermans et al. (2020) showed that temporal variation was higher for grassland than forest communities. Furthermore, local climate conditions might affect the response of bacterial diversity to climate change. It was shown that warming decreases the abundance of soil biota in regions with lower mean annual temperature and mean annual precipitation (Blankinship et al., 2011).

In conclusion, temporal variation of bacterial diversity in soil is still understudied (Hermans et al., 2020; Carini et al., 2020), since detailed community analyses tracking bacterial diversity over several years are still rare. If available, the effect of climate change on bacterial diversity is assessed using manipulation experiments, applying rather strong climate gradients or treatments (DeAngelis et al., 2015; Nottingham et al., 2022). Those experimental studies project future climate conditions diverging strongly from current climate. However,

observational studies are lacking exploring the impact of current climatic development in long-term experiments to determine if present climate change is already affecting bacterial diversity in soil. In addition, tracking the temporal variation cannot be considered in isolation from spatial variation (Hermans et al., 2020). Representative sampling of sampling sites is pivotal to detect temporal variation. If considered, accounting for spatial variation even makes short term variation detectable (Carini et al., 2020).

1.4 Spatial variation of soil bacteria

Soils are by no means homogeneous environments. Aggregation of soil organic matter and mineral particles produces a large variety of habitats differing in aeration, pore size distribution, water availability and retention, nutrient availability and quality, considering soil organic matter. Therefore bacterial life in soil is heterogeneously distributed. Bacterial communities differ on the pore scale with different bacterial communities preferring different aggregate size classes (Neumann et al., 2013; Hemkemeyer et al., 2018; Szoboszlay and Tebbe, 2021). Moreover, soil bacteria live in patches and the cell density declines with distance to soil pores (Nunan et al., 2003). The distances between bacterial cells range from 0.5 to 12.5 µm (Raynaud and Nunan, 2014; Probandt et al., 2018). Therefore, cell interactions are limited due to the limited number of cells close-by (Raynaud and Nunan, 2014). Furthermore, the pore size distribution and water content is influencing the fragmentation of aqueous phases in soil, limiting bacterial interaction (Wang and Or, 2013; Bickel and Or, 2020) as well as dispersal, which is rapidly limited under reduced soil moisture conditions <-2 kPa (Dechesne et al., 2010; Ebrahimi and Or, 2014).

Most studies neglect these differentiations among soil bacterial communities on the pore scale. However, the high spatial differentiation between bacteria on the pore scale is impossible to implement in large scale studies due to the extensive workload during sample processing and sequencing. Homogenized and well replicated samples of bulk soil are conveniently considered as good representations of soil bacterial communities, representative for a sampling site. Thereby the microscale effect of soil structure is averaged out by sieving the bulk soil material. A good sample replication is inevitable to provide a good representation of the sampling site and avoid effects based on spatial small scale variation (Armstrong et al., 2016; Hermans et al., 2020).

Another aspect of spatial heterogeneity neglected so far for soils is life-style separating soil bacterial communities. In marine environments, the existence of planktonic free-living and particle-associated bacteria is widely recognized (Simon et al., 2014; Denef et al., 2016; Mestre et al., 2017). Those distinct communities differ in genome size (Swan et al., 2013). Furthermore, free-living bacteria are more adapted to low nutrient environments, providing lower metabolic potential (Dupont et al., 2012). In contrast particle-associated bacteria

show higher metabolic potential (Simon et al., 2014).
For soils, this differentiation in life-style between free-living bacteria living in the aqueous phase of the soil and the surface-associated community, attached to particles, is unanswered. So far only a single study differentiated soil bacterial communities by life-style, showing that the free-living and surface-associated communities in soil are separable. Thereby most bacteria occurred in both communities, showing no definite separation between communities (Bystrianský et al., 2019). However, no evidence is available if the activity of bacteria differs between free-living and surface-associated communities in soil.

1.5 Aim of this work

The analysis of spatial and temporal pattern is heavily researched in soil microbial ecology. However, specific aspects causing spatial and temporal variation in soil bacterial communities are still neglected. Therefore, I investigate two specific aspects of soil bacterial diversity:

First, I will investigate the first 6 years of a long-term and large-scale study in Germany, spanning 300 grassland and forest sites. The analysed data are not the result of experimental manipulation of the sampling sites, but of observational data tracking the development of multitrophic diversity over time. More specifically, I want to investigate how temporal variation is affecting bacterial diversity and examine if effects of climate over time, i.e. climate change, influence the soil bacterial diversity of forest and grassland communities in Germany.

Second, spatial heterogeneity in soil is not considered in most work analysing soil bacterial diversity. It is unknown for soils, how a differentiation by life-style between the free-living and surface-associated communities affect the respective bacterial activity. Also, I investigate how bacterial life-style is reacting to nutrient availability, examining the free-living and surface-associated community along a gradient of soil nitrogen content in a German long-term fertilization experiment.

2 Does climate change affect soil bacterial communities already? Results from a German large scale study

2.1 Starting point of the investigation

Climate change can already be measured on a global scale. The mean land surface air temperature has risen by 1.53 °C since the pre-industrial period. This climate change includes an increase in frequency and intensity of extremes (Pörtner et al., 2022). This rise in temperature and the occurrence of accelerated heatwave trends is recently shown to be three to four times faster for Europe than the northern midlatitudes (Rousi et al., 2022). As a result, climate change is starting to affect cropland and grassland productivity in Germany. The increase in maximum summer temperature is already threatening wheat production in Germany (Pinke et al., 2022). Similarly, maize production is predicted to decline for the next 30 years (Peichl et al., 2019). In grassland systems, yield is coupled to aridity. Induced by an observed rise in temperature and decrease in precipitation, this is threatening grassland productivity in northern Germany (Emadodin et al., 2021). More importantly, temperate grasslands might have lost their CO_2 sink function already over the last 100 years due to climate change (Baca Cabrera et al., 2021). Furthermore, a loss of plant diversity in Europe is anticipated (Bellard et al., 2012). However, a potential reduction in plant diversity does not affect belowground diversity (Delgado-Baquerizo et al., 2019).

So far, no information is available if climate change already affects soil bacterial communities in Germany. To this point, recent examination of next generation sequencing data has identified a comprehensive set of predictors shaping soil bacterial communities. On a global scale, climate, pH, soil carbon and other nutrients, C/N ratio, land use type as well as soil texture are confirmed as major drivers of bulk soil bacterial communities (Fierer and Jackson, 2006; Chau et al., 2011; Delgado-Baquerizo et al., 2016; Bahram et al., 2018; Ladau et al., 2018; Bickel et al., 2019; Delgado-Baquerizo and Eldridge, 2019). It has to be considered that climate predictors are often more important than soil texture (Delgado-Baquerizo and Eldridge, 2019) or even pH (Delgado-Baquerizo et al., 2016).
A closer look at climate-induced change in bacterial diversity shows that different aspects of climate affect soil bacterial diversity. Diurnal temperature range, mean annual temperature, aridity, mean annual precipitation and climatic water content influence bacterial diversity (Delgado-Baquerizo et al., 2016). Most notably, predictors representing soil moisture are more important than temperature related predictors (Bickel et al., 2019).
These global studies revealed that bacterial diversity is peaking in temperate grasslands (Bahram et al., 2018). Tropic, boreal and dessert biomes or temperate forests harbour less

bacterial species (Delgado-Baquerizo and Eldridge, 2019). Hence, under climate change scenarios, with increasing temperature and reduced precipitation, the maximum bacterial diversity observed for temperate grasslands will face a potential decline.

This is of major importance since soil biodiversity loss is threatening ecosystem multifunctionality (Wagg et al., 2014) and influencing aboveground diversity (Bardgett and van der Putten, 2014). Recently it was shown that long-term warming in subarctic grasslands leads to a reduction in soil C and nutrient concentrations (Söllinger et al., 2022). However, transcriptomic data suggested that the level of CO_2 emission was maintained, although ribosomal content and protein biosynthesis were reduced. Moreover, taxonomic composition and diversity did not differ significantly, although a trend towards a reduction in richness was observed. This is supported by DeAngelis et al. (2015), who found no effect of warming on bacterial diversity in bulk soil after 20 years. Other experiments showed that bacterial diversity declined under experimental warming (Nottingham et al., 2022; Wu et al., 2022). This effect is most likely influenced by a concomitant reduction in soil moisture (Guo et al., 2018). Apparently the soil temperature causing a reduction in richness is depending on the climatic legacy of the soil and was between 27-30°C for alpine soils (Pietikäinen et al., 2005). This is in line with results from Ladau et al. (2018) showing that bacterial communities were more closely adapted to historic than recent climate conditions. The fact that conditions preceding the actual sampling period still influence current soil bacterial communities is well documented and called legacy effects (Hariharan et al., 2017; Ladau et al., 2018; Schmid et al., 2021).

In contrast to temperature, the effects of changes in precipitation pattern and subsequently soil moisture on bacterial diversity are even less clear. Literature suggests that short-term precipitation manipulation has little effects on bacterial diversity (Stres et al., 2008; Cruz-Martínez et al., 2009; Engelhardt et al., 2018) or bacterial abundance (Waring and Hawkes, 2018). This might be a consequence of bacterial resistance to short-term drought (Lennon et al., 2012; Meisner et al., 2017). Furthermore, responses on ecosystem scale are difficult to estimate due to small-scale controls of soil moisture (Jansson and Hofmockel, 2020). Similar to temperature, legacy effects play an important role for bacterial responses to changes in soil moisture (Banerjee et al., 2016; Jansson and Hofmockel, 2020). However, Zhang et al. (2013) could show the soil moisture effects were higher than temperature effects for grassland steppe soils in Inner Mongolia. Besides, *Acidobacteria* were the least resistant phylum to changes in soil moisture (Castro et al., 2010; Zhang et al., 2013). Moreover, Huber et al. (2022) showed that particularly dessication decreased the fraction of active *Acidobacteria*. In conclusion, most evidence for the effects of warming and changes in soil moisture over time on bacterial diversity is either coming from extreme environments, short-term experiments or the effects on bacterial diversity were not analysed in depth, since the focus was on C cycling and fluxes.

Studies analysing the effects of climate on bacterial diversity revealed contrasting results and no clear conclusion is reached how rare and abundant bacteria are affected by climate. For example, bacterial richness was reported to decrease with increasing temperature and decreasing soil moisture (Liang et al., 2015; Prober et al., 2015; Wu et al., 2022). In soils, most bacteria are rare in abundance (Curtis et al., 2002; Delgado-Baquerizo et al., 2018). Therefore, bacterial richness, the total number of species in a sample, is dominated by rare bacteria. Furthermore, rare species were reported to be more sensitive to climate than abundant species (Bickel et al., 2019). Albeit, a higher sensitivity of rare species to climate change must not lead to a reduction in richness (Zhang et al., 2013; Zhou et al., 2020). In addition, abundant but not rare bacteria were found to be sensitive to climate as well (Liu et al., 2019).

Moreover, the analysis of effects of climate change requires longitudinal data. This involves temporal variability and stochasticity, although soil bacteria are amongst the most stable communities over time (Ladau et al., 2018; Shade et al., 2013). Thereby, soil bacterial communities showed a high resistance to climate change over a 10 month period (Waring and Hawkes, 2018). However, soil bacteria communities decreased with change in temperature over the course of several years (Liang et al., 2015). This emphasizes that the duration of climate change is important, with longer duration implying stronger reactions (Seaton et al., 2021). Nevertheless, there is still a lack of knowledge, how temporal variation affect bacterial diversity (Hermans et al., 2020). Furthermore, a common property of climate predictors is their correlation with other climatic and edaphic predictors. This property is called multicollinearity (Dormann et al., 2013). Although constituting independent biological drivers, correlated predictors represent the same signal from a statistical point of view (Ellis et al., 2012). Therefore, it is difficult to dissect effects of correlated predictors in modelling. Another obstacle is that a generalization of global results or smaller scale studies is difficult due to a high variability of predictors on regional scales and spatial correlation, which can influences soil bacteria more than temporal variation (Mod et al., 2021). On the plot scale, bacterial communities are quite stable over time and space, with spatial variation explaining more variance than time (Richter-Heitmann et al., 2020). Furthermore, effective global predictors like mean annual precipitation and mean annual temperature are ineffective on a regional scale, providing only minor variability.

In conclusion, literature suggests that soil warming will most likely reduce bacterial richness, but we lack information whether rare and abundant bacteria react similarly. Moreover, we lack in depth information which aspects of climate affect soil bacteria. Evidence from global studies cannot be transferred to more regional scales, since the main predictors used, e.g. mean annual precipitation and mean annual temperature, do not differ sufficiently between sampling sites to induce relevant changes in bacterial diversity. Identifying the relevant aspects of climate is essential to predict community shifts under climate

change scenarios and gain more insights on how soil bacteria react to climate change. So far, no observational long-term study from Germany is available, examining whether and how soil bacterial communities change over time.

In this study, I present first results from the Biodiversity Exploratories, a large scale and long term experimental setup in Germany. The focus of this interdisciplinary network is the interaction between land use type and intensity, biodiversity and ecosystem functioning on multiple trophic levels (Fischer et al., 2010). In 3 regions across Germany, 300 experimental plots were established and soil bacterial communities are monitored every 3 years. The regions are located in the south-west (Schwäbische Alb), center (Hainich) and north-east (Schorfheide) of Germany. Each region contains 50 grassland and 50 forest sites, respectively. A preceding study in the Biodiversity Exploratories network, analysing bacterial diversity confirmed that pH, soil texture, soil water content, C/N ratio, soil type and the main tree species affected bacterial community composition on grassland and forest sites in 2011 (Kaiser et al., 2016). The effects of land use type and intensity are not consistent between studies. None (Nacke et al., 2011) or little effects of land use type are reported (Kaiser et al., 2016). Besides, land use intensity (LUI) of grassland sites, estimated as the joint influence of N-fertilization, grazing and mowing, showed neglectable influence on soil bacteria (Gossner et al., 2016; Kaiser et al., 2016). These effects are apparently higher for rhizosphere communities than bulk soil communities (Schöps et al., 2018). Interestingly, the effect of LUI was found to have a delayed effect on bacterial communities, with the LUI of the past year being stronger correlated to bacterial diversity than the LUI of the sampling year (Boeddinghaus et al., 2019).

For the above mentioned reasons, the aim of this work is to identify the effects of climate, edaphic predictors and time on soil bacterial alpha and beta diversity taking spatial correlation structures into account. Alpha diversity is commonly understood as measure of average single-location diversity, i.e the diversity in a sample. Beta diversity is a measure of the relative change in species composition between locations (Jost, 2007).

Moreover, different aspects of alpha and beta diversity were analysed. As mentioned above, rare and abundant bacterial species might react differently to environmental predictors and time. Therefore richness and inverse Simpson indices were chosen to represent rare and abundant bacteria (Hill, 1973; Jost, 2006; Alberdi and Gilbert, 2019). Likewise, two independent approaches were used to analyse beta diversity. First, modelling Bray-Curtis dissimilarity in a distance based approach to track community turnover along gradients (Ferrier et al., 2007) and second, modelling each bacterial species separately to allow for an in depth analysis of specific groups or parts of a community (Ellis et al., 2012). This species specific modelling approach might give deeper insight into niche differentiation between soil bacteria based on the predictors retrieved as influential on bacterial community turnover along predictor gradients.

Bacterial communities were analysed from bulk soil of the Biodiversity Exploratories from May 2011, 2014 and 2017 using next generation sequencing of the V3 region of 16S rRNA. The analysis of RNA based communities avoids potentially obscuring pattern from relic DNA in soil (Carini et al., 2016), especially when analysing spatial and temporal pattern (Carini et al., 2020).

With this approach, I want to tackle the following questions: Which climatic aspect of soil moisture, soil temperature and precipitation affect alpha and beta diversity in grassland and forest soils in Germany? Does soil bacterial diversity change over time? If so, is climate change causing changes over time in bacterial diversity? Furthermore, I want to examine potential legacy effects of climate and LUI on bacterial diversity. Furthermore, which edaphic predictors shape bacterial diversity for this data set? Can niches of soil bacteria be identified by relevant predictors emerging from gradientForest modelling of beta diversity?

2.2 Material and Methods

2.2.1 Study design

The Biodiversity Exploratories comprise of 300 sites, located in 3 regions across Germany (Fig. 1). Each region in the Schwäbische Alb, Hainich and Schorfheide harbours 50 forest and grassland sites, respectively.

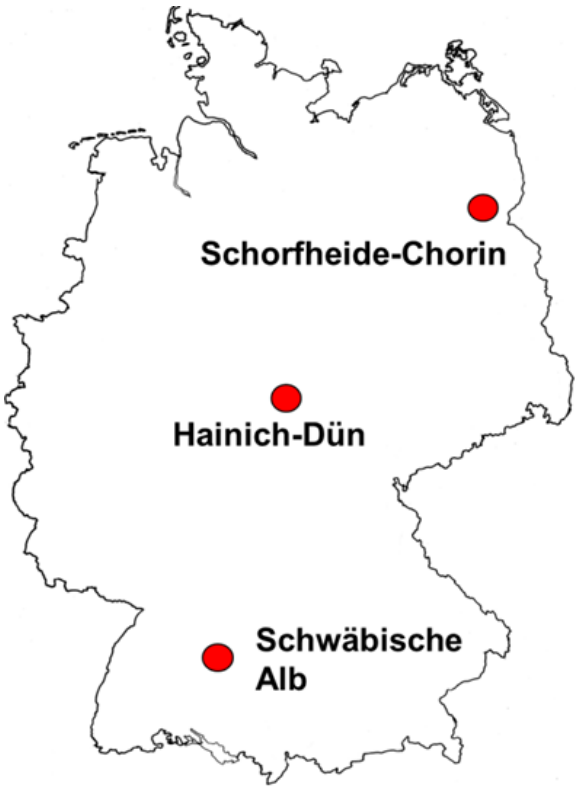

Figure 1: Distribution of the Biodiversity Exploratory regions across Germany.

Bulk soil of all 300 sites was sampled in May 2011, 2014 and 2017 during a joint sampling campaign from members of different institutes. At each site, 14 sub-samples of top soil samples of 0-10 cm depth were mixed and homogenized through 2 mm mesh sieve. Plant roots and stones were removed and samples immediately flash frozen in liquid nitrogen to preserve soil bacterial RNA communities. Until sample processing, soil samples were kept in liquid nitrogen.

2.2.2 Nucleic acid extraction and library preparation

RNA and DNA were co-extracted from soil following the protocol by Lueders et al. (2003) implementing changes according to Wüst et al. (2016).
Cell lysis for 0.6 g of soil was performed in 750 µl of 120 mM Na_2HPO_4 (pH 8) and 250 µl

Tris-HCl (pH 8) using 0.7 g zirconium beads (0.1 mm diameter) for 45 s at 6.5 m/s on a bead beater. After centrifugation at 20000 xg for 20 min, the supernatant containing nucleic acids was mixed with 500 µl Phenol/Chloroform/Isoamylalcohol (25:24:1, v/v) of pH 8 and centrifuged at max. speed for 5 min. The resulting supernatant was mixed with Chlorofom/Isoamyalcohol (24:1, v/v) and centrifuged for 5 min at 20000 xg. The supernatant was precipitated with 1 ml 60 % PEG and centrifuged for 90 min at 20000 xg. The nucleic acid pellet was washed in 500 µl ice cold 70 % ethanol and centrifuged again for 30 min at 20000 xg. All supernatant was removed and DNA and RNA were resuspended in 50 µl 100 mM Tris-HCl (pH 8) and stored at -80 °C. DNA was digested using DNAse I from the Turbo DNA-free kit (ThermoFischer) according to the manufacturer protocol. Recovered RNA was precipitated with 0.1 volume 3 M sodium acetate (pH 5.2) and 2 volumes isopropanol for 1 h and centrifuged for 5 min at 20000 xg at room temperature. The RNA pellet was resuspended in 50 µL RNAse-free water. Reverse transcription of RNA to cDNA was accomplished using the GoScript reverse transcription kit (Promega) according to the manufacturer protocol. RNA was reverse transcribed using 4 µl of RNA extracts. Next, 1 µl of Random Hexamers were added and heated on a PCR-cycler for 5 min at 70 °C. The mixture was cooled on ice for 5 min. Afterwards, 15 µL of PCR-buffer (1x GoScript reaction buffer, 3 mM $MgCl_2$, 0.5 mM nucleotide mix, 0.05 U μl^{-1} GoScript reverse transcriptase) was added. The reverse transcription continued for 5 min at 25 °C, 1 h at 42 °C, 5 min at 70 °C and finished with a cooldown to 4 °C.

The library preparation was executed as described in Bartram et al. (2011) on the V3 region of the 16 S rRNA. For each sample, 3 technical replicates were processed. 5 µl of reverse transcription product were added to 45 µl PCR-buffer (1x HF Phusion buffer, 0.2 mM nucleotide mix, 3.5 mM $MgCl_2$, 5 % DMSO, 0.4 mg/ml BSA, 0.2 pmol/µl 341f forward (5'- CCTACGGGWGGCWGCAG-3') and 515r reverse (5'-CCGCGGCTGCTGGCAC-3') primers, and 0.04 U/µl Phusion HotStart II High Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA). Reverse primers contained a variable index to facilitate sample demultiplexing after sequencing. After initial denaturation for 30 s at 98 °C, denaturation for 10 s at 98 °C, annealing for 10 s at 59 °C and elongation for 45 s at 72 °C were repeated for 15 cycles, followed by a final elongation for 7 min at 72 °C and a cooldown to 4 °C. All amplification products were gel-purified to remove dimers on a 2 % agarose gel (Lonza group, Basel, Switzerland) in 1x TAE buffer for 2.5 h at 120 V at 4 °C. The PCR product of 150 base pair length was cut out and purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany) according to the instructions of the manufacturer.

In preparation of the sequencing run, sample libraries were quantified with the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, US) and 4nM of each sample were pooled. All runs were spiked with 15 % PhiX. Samples of 2011 were sequenced on

a HighSeq 2500 platform (Illumina, San Diego, CA, USA). For 2014 and 2017 samples, the NextSeq 500 platform (Illumina, San Diego, CA, USA) was used with a 150 base pair paired-end-run.

2.2.3 Bioinformatic processing of sequencing data

After sequencing, sample libraries were demultiplexed according to the varying indices in each sequencing run. The demultiplexed sequences were processed using the Python based QIIME2 2019.1 bioinformatic pipeline (Caporaso et al., 2010; Bolyen et al., 2019) run in a Linux environment on a high performance cluster with default settings if not stated otherwise. All pipeline code can be found in detail in Supplementary Materials (page 141). Bacteria were distinguished at the level of single nucleotide differences as exact sequence variants (SV), defined as a specific nucleotide sequence of the V3 16S rRNA gene region (Callahan et al., 2017).
In short, the demultiplexed paired-end reads were joined using plugin *vsearch*. Joined reads were processed using plugin *quality-filter* (Bokulich et al., 2013). Reads having a quality score below 25 over at least 75 % of the read length were discarded. Plugin *deblur* (Amir et al., 2017) was used for denoising the joined sequences and checking for chimera.
First, all sequences were trimmed to a length of 165 base pairs. Second, all sequences were discarded occurring less than 10 times in all samples and less than 2 times in each sample. For sequence alignment MAFFT (Katoh and Standley, 2013) was used. The phylogenetic tree was calculated using the FastTree plugin (Price et al., 2010). The taxonomy was annotated with the plugin *feature-classifier*, using a Naïve Bayes classifier trained on the SILVA 132 database (Quast et al., 2013; Yilmaz et al., 2014) for the V3 region. In a final step, sequences of mitochondria and chloroplast were removed and the abundance data, taxonomy table and phylogenetic tree - in Newick format - were retrieved and exported to R ({R Core Team}, 2021). Bacteria were differentiated on the level of sequencing variants (SV). However, the concept of species is the bottom line of diversity analysis. Bacterial species are defined as a threshold of 70 % DNA-DNA hybridization (Wayne et al., 1987). Hence, the term sequence variant has to be used instead of "species".

2.2.4 Metadata collection and gap filling

All downstream data preparation and analysis was performed in R, version 4.1.0. All sequencing informations were imported into R as a phyloseq-object using package `phyloseq`. Package `tidyverse` was used for data handling and packages `ggplot2` and `gridExtra` for visualization. Environmental predictors measured on the plot scale and climatic data of the study region were obtained from BExiS, the central data base of the Biodiversity Exploratories (*https://www.bexis.uni-jena.de/*). The used metadata can be retrieved via

the following accession numbers: 19346, 10580, 14686, 16586, 26207, 20626, 19346, 19009, 14446, , 14447, 14448, 17086, 17166, 20040, 20248, 20246, 20250, 19807, 20626, 18386, 18787, 19067, 19366, 20045, 20247, 20249, 20251, 20266, 23906, 23486, 22246, 23846, 24346, 25186, 25408, 25586, 26147.
Generally, predictors containing $> 20\,\%$ missing values were discarded. Gap filling for edaphic predictors was based on distance dependent k-nearest neighbour imputation, using 5 neighbour sites with function `VIM::kNN`.
The raw data of plot specific climate data contained high proportions of missing values especially at the beginning of the records (Fig. 2). Therefore the longest possible timeframe to be analysed for legacy effects was limited to 1067 days before sampling. In case of small gaps, chained random forests with function `missRanger::missRanger` were used to impute missing values, since k-nearest neighbour imputation was not capable of imputing the cycling nature of climatic predictors. Gap filling is a common practice in meteorology and time series analysis.

2.2.5 Environmental predictor engineering

Predictor engineering is a potent tool to extract valuable information and aspects from predictors for modelling (Kuhn and Johnson, 2019). This is true especially for longitudinal data such as climate data, in case methods of time series analysis cannot be applied. Hence, changes in the data over time have to be aggregated and mapped as discrete predictors. Generally, calculating any summary statistics, such as the mean, median and identification of extreme values, is a form of predictor engineering if used as a predictor for statistical modelling. Soil temperature in 5 cm depth, soil moisture in 10 cm depth estimated as volumetric water content via frequency domain reflectometry sondes and radar precipitation data were aggregated as daily means. Soil moisture data from 5 cm depth was available as well. However, more data gaps were present. Therefore soil moisture data from 10 cm was selected for analysis due to better raw data quality. The mean daily records of climatic data were summarised over 14, 30, 90, 180, 365, 730 and 1056 days before sampling, deciphering potential legacy effects, i.e. delayed effects, of climate on bacterial communities. For these different temporal aggregations, the median, maximum, minimum and range of climatic predictors were calculated. Threshold data of soil temperature and soil moisture represent the number of days in a specific timeframe (see above) in which the target variable remained above a certain threshold (Frindte et al., 2019). Threshold data were calculated in increments of 1°C for soil temperature or 1 % for volumetric water content over a range of -36-29 °C and 1-63 % respectively. Specific for precipitation data, the sum of rain days was calculated, representing the number of days with precipitation in a specific timeframe. Moreover, land use intensity (LUI) (Blüthgen et al., 2012) and

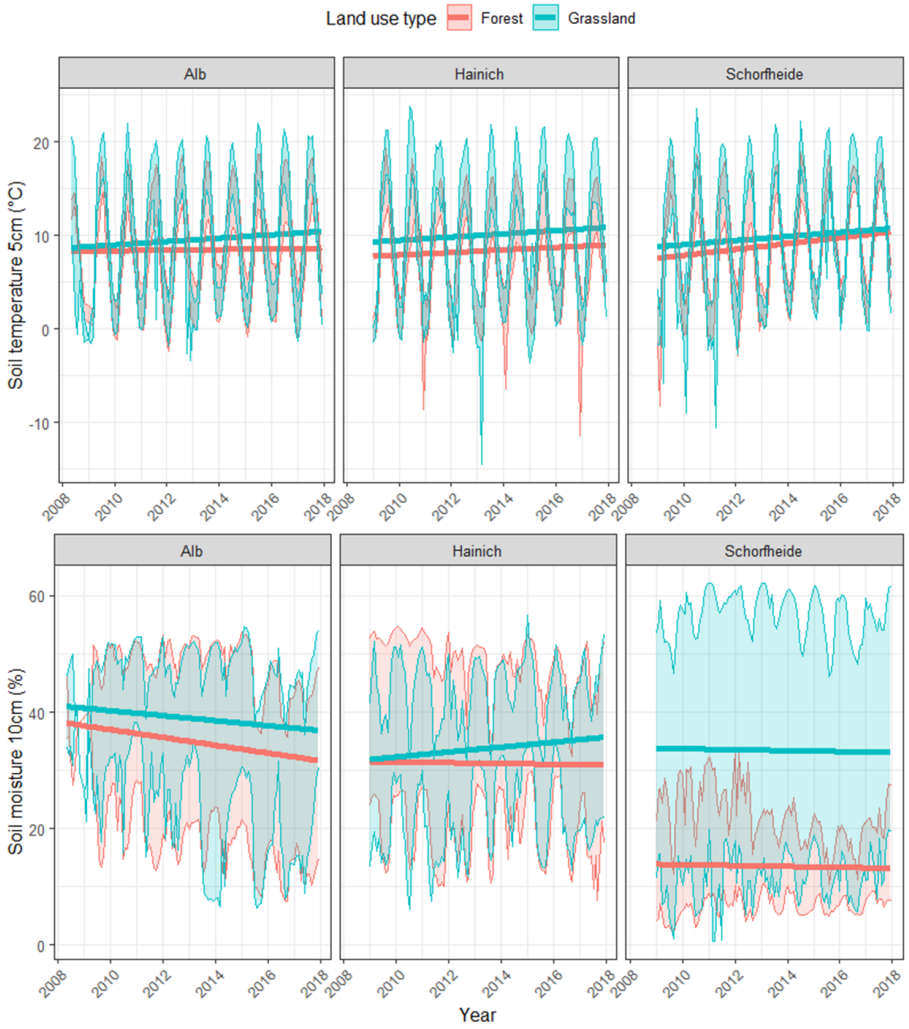

Figure 2: Ranges of soil temperature and soil moisture data separated by land use type and Exploratory region. Trends in climate data are displayed as linear models.

the respective raw data of grazing, mowing and nitrogen fertilization of grassland sites, available as annual values, were calculated as means over 1-5 years preliminary to soil sampling. Furthermore, trends in time series of LUI, soil moisture and temperature were extracted by fitting a linear model with function `lm`. The respective model slopes were extracted and used as predictors, indicating trends in climate data and LUI. An overview about all engineered predictors is provided in Supplementary materials (Tab. 7, p. 145).

2.2.6 Predictor selection and data structure

In total 296 of 300 forest and grassland sites were examined for 2011, 2014 and 2017. Four experimental plots were discontinued or replaced and therefore had to be excluded

from the analysis. Many engineered predictors derived from soil moisture (SM) and soil temperature (Ts) showed multicollinearity, meaning that some predictors were not independent and correlated among each other (Fig. 49, 50). In addition, edaphic predictors showed multicollinearity as well. Soil nutrients (organic C, total N, total C, organic S, extractable organic C and N), soil pore content and water retention (Fig. 51-53) were correlated. Statistically, collinear predictors represent identical signals towards a response variable, although the predictors may represent different causal biological drivers. Hence, all predictors and models were checked for multicollinearity. Moreover, predictors showing inhomogeneous distributions over the predictor range, not covering the predictor gradient entirely, were excluded from the analysis. Predictor distributions and correlation were assessed by plotting and visual inspection All available edaphic predictors are presented in Table 8.

Statistical testing for significant differences between groups for environmental predictors relied on the `vegan` package. Differences between groups were tested for significance with omnibus ANOVA testing and post-hoc TukeyHSD using functions `aov` and `TukeyHSD` respectively.

2.2.7 Sequencing data normalization

For the analysis of alpha and beta diversity, coverage-based rarefaction was performed using the function `metagMisc::phyloseq_coverage_raref`. Coverage-based rarefaction normalizes samples to identical sample completeness rather than identical number of reads (Chao and Jost, 2012). This approach avoids data loss and consequently loss in precision, resulting from discarding the amount of reads higher than minimum sequencing depth (Chao et al., 2020).

2.2.8 Alpha diversity analysis

Alpha diversity is commonly defined as the diversity of species in a sample (Sepkoski, 1988) and is usually estimated as indices related to species counts, their abundance and phylogeny or combinations thereof (Whittaker, 1972; Chao et al., 2010).
In this stud,y species richness, Shannon and inverse Simpson diversity indices were used. In more detail, alpha diversity was calculated as Hill numbers using function `hill_div` in package `hilldiv`. The concept of Hill numbers was originally developed by Hill (1973), creating a unifying framework of alpha diversity indices with identical mathematical properties (Jost, 2006). It was shown that many popular alpha diversity indices are special cases of

$$^{q}D \equiv \left(\sum_{i=1}^{S} p_i^q \right)^{\frac{1}{(1-q)}} \tag{1}$$

where D represents the diversity measure, S is the number of species, p_i the relative abundance of the ith species and the order of q.
Therefore, estimating the diversity in a sample is dependent only on the order of q and the species abundance (Jost, 2006). The advantages of Hill numbers are that for $q = 0, 1, 2$ equation 1 equals richness, the exponential of Shannon entropy and the inverse Simpson concentration respectively (Chao et al., 2010), integrating different alpha diversity indices in a unifying framework. In case of $q = 1$, referring to Shannon entropy, equation 1 is undefined. Therefore its limit is used as

$$^{1}D = \exp\left(- \sum_{i=1}^{S} p_i \ln p_i \right) \tag{2}$$

Unlike their original counterparts (richness, Shannon and inverse Simpson indices representing a species count, an entropy and a concentration respectively), Hill numbers of order (q) represent an "effective number of species" similar to the alpha diversity level in a sample (MacArthur, 1965). Hence, the effective number of species between different orders of q are directly comparable. This is a property no other framework or alpha diversity index can offer. Besides, this framework obeys the replication principle. The replication principle "for species-neutral diversity states that if we have N equally large, equally diverse groups with no species in common, the diversity of the pooled groups must be N times the diversity of a single group" (Chao et al., 2010). Recently, a consensus has grown that Hill numbers are the method of choice for diversity analysis (Chao et al., 2020).
Richness ($q = 0$) represents the total number of species in a sample, ignoring relative abundance. Richness of natural soil bacterial communities is therefore dominated by rare species, since the majority of SV are rare (Bickel and Or, 2021). In contrast, the Shannon ($q = 1$) and Simpson ($q = 2$) indices take species relative abundance into account (see equation 1), with Shannon describing the common species (Chao et al., 2010) and the Simpson index representing highly abundant species due to an exponent of $q = 2$.

In this work, alpha diversity was analysed for rare and highly abundant species independently, represented by richness and Simpson indices respectively. The Shannon index was excluded from the analysis since this index was highly correlated with both richness and Simpson indices (Tab. 1). A separate modelling of either index in combination with Shannon diversity is a form of pseudo-replication. Moreover, the analysis of the Shannon index is not advised in diversity analysis, due to a high influence of evenness on species counts

(Strong, 2016).

Table 1: Pearson correlation coefficient between alpha diversity indices.

	Richness	Shannon	Simpson
Richness	1		
Shannon	0.75	1	
Simpson	0.52	0.92	1

As an initial analysis, differences in alpha diversity between land use types, sampling years and Exploratory region were calculated using analysis of variance (ANOVA) with function `aov`. Post-hoc pairwise comparisons were tested for significance using Tukey honest significant difference (HSD) post-hoc test as in function `TukeyHSD`. Correction for multiple testing controlling the false discovery rate is accounted for by using intervals based on the studentized range statistic as implemented in Tukey HSD (Tukey, 1991).
Richness and Simpson indices were modelled separately for grassland and forest sites. Additionally, models of richness and Simpson indices were fit on all samples to take advantage of the high sample count. Due to their large number of predictors, influential climatic predictors were preselected by generalized linear models with elastic net regularization using function `caret::train` and method "glmnet". Generalized linear models with elastic net regularization can handle co-correlated predictors even in scenarios, where the number of predictors exceeds the number of samples (Friedman et al., 2010). Elastic net regression is a mixture of ridge and lasso regression. Thereby linear models were fitted identifying predictors which affect richness and Simpson indices based on variable importance with function `varImp`. Two hyperparameters, λ and α , were tuned automatically using package `caret::train`. For richness and Simpson index λ and α corresponded to 31.84 and 1 as well as 12.25 and 0.1, respectively. Before model fitting, all predictors were centred and scaled. Predictors with near zero variation were removed. Models were trained on subset of data comprising 2/3 of samples and cross-validated 10 times on 1/3 of the data to check for model overfitting, comparing the root mean square error. Predictors with a variable importance >0 were further used for model selection of alpha diversity. However, remaining predictors were again checked by visual inspection and dropped in case of high correlation or bad distributions.
Generalised additive models (GAM) were used to fit alpha diversity. GAM models can account for non-linear fits of predictor variables for richness and Simpson index using function `bam` in package `mgcv`. A negative binomial distribution with log-link function was applied to control for overfitting, using fast restricted maximum likelihood. Furthermore, model parameters gamma, scale and rho were set to 1.4, 1 and 0.6 respectively. A sample site identifier was used in all GAMs as random intercept to account for systematic variation

between plots not covered by fixed effects. Model selection was done via forward selection, adding the predictor resulting in the maximum drop in AIC ≥ 5. Continuous predictors were fit as non-parametric thin plate regression splines. Factor variables were fit as parametric terms. The effect of time on alpha diversity was modelled as the sampling year. Although continuous, sampling year was fit as an ordinal scaled predictor, since only three time points were available. Spatial autocorrelation in model residuals were accounted for as the interaction of latitude and longitude of sample site locations in all models. For model fitting, sample sites were split randomly in a training set for model selection and a test dataset for cross-validation, comprising 2/3 and 1/3 of data, respectively. Concurvity, the non-linear extentsion of collinearity, between predictors selected in a GAM was checked by running function `mgcv::concurvity`. 10-fold cross-validation with random subsampling of 80 % of sites from train and test data was used to check for model overfitting. Therefore root mean square errors were compared. After successful cross-validation, the final model was applied to all sites. The performance of the model was evaluated using the proportion of deviance explained for the final model. Deviance is a goodness of fit statistic and generalization of sum of squares statistics adopted for models fitted using maximum likelihood.

2.2.9 Beta diversity analysis

Beta diversity is referred to as the diversity between samples (Sepkoski, 1988). In general, most bacterial species in soils are rare (Nemergut et al., 2013; Lynch and Neufeld, 2015). Consequently, abundance data of soil bacteria generated by next generation sequencing are sparse and document the absence of most species in the majority of samples. Hence, it is difficult to infer the behaviour of rare species to environmental gradients in a statistically sound way in models based on bacterial abundance data. Therefore, only SV with a relative abundance of ≥ 0.1 % in at least one sample were selected for beta diversity modelling following the approach in Szoboszlay and Tebbe (2021). Ordination of principal coordinate analysis (PCoA) and procrustes rotation in package `vegan` with function `cmdscale` and `procrustes` were calculated, exploring beta diversity and testing for similarity between ordinations using function `protest`. Function `procrustes` rotates an ordination to maximum similarity with another ordination using scaled Y axis and symmetric Procrustes statistic. Function `protest` retrieves the Pearson correlation between ordinations rotated with `procrustes` and tests for the non-randomness of similarity between ordinations using 999 permutations (Peres-Neto and Jackson, 2001).
In this study, generalised dissimilarity modelling (GDM) (Ferrier et al., 2007) and gradientForest modelling (GF) (Ellis et al., 2012) were applied to analyse changes in beta diversity and render the results more robust as recommended in Ellis et al. (2012).

Both methods share the same goal modelling community turnover along environmental gradients. Thereby, two fundamentally different approaches are compared. GDM is based on Bray-Curtis dissimilarity (Bray and Curtis, 1957). This distance measure between samples is widely used to model beta diversity, taking species abundance but not phylogeny into account (Lozupone and Knight, 2005). Furthermore, a major advantage of GDM is the capability to incorporate and estimate the influence of spatial pattern, a property which is not available for GF. Hence, it is possible to estimate community turnover over spatial distances. GDM is as an extension of matrix regression fitting an I-spline GAM to Bray-Curtis dissimilarities (Ferrier et al., 2007). This allows for the modelling of non-linear pattern, similar to the GAMs used for alpha diversity analysis.

On the other hand, GF fits the abundance of each SV individually to environmental predictors. Turnover can be derived for every modelled SV for each predictor. Community turnover caused by a specific predictor is derived by averaging the predictor importance of all SV along the predictor gradient. Generally, both GDM and GF models are non-parametric models. In more detail, GDMs from function `gdm::gdm` were used to model community turnover along environmental gradients. Unfortunately, nominal predictors cannot be fit and were incorporated as dummy variables. Spatial autocorrelation was considered by including the pairwise distances between sites. GDM model fitting was performed on training data, comprising 2/3 of the data using random subsampling. Model selection was performed via backward selection until only significant predictors ($p<0.05$) remained, as implemented in function `gdm.varImp`. Finally, 10-fold cross-validation was performed on the 1/3 test data, testing for model overfitting, via `gdm.crossvalidation`. The final models were fit on all data.

Complementary to GDM, GF random forest models in function `gradientForest::gf` were used to analyse contributions of environmental predictors to community turnover (Ellis et al., 2012). Cross-validation is implemented by excluding observation from modelling, called out of bag samples (OOB), and testing the prediction accuracy upon OOB. Furthermore, GF is robust towards skewed predictor distributions, due to their structure rooted in binary decision trees (Breiman, 2001). Moreover, GF can deal with multicollinearity of environmental predictors by conditional permutation (Ellis et al., 2012). Thereby, a predictor is randomly permuted, breaking the relation to the response variable. In case a strong association existed between the unconditioned predictor and the response variable, the prediction accuracy for the OOB values would drop. Thus, variable importance obtained by conditional permutation are not biased by spurious correlation (Strobl et al., 2008). Conditional permutation was applied if predictors were correlated >0.5 for a maximum number of 64 partitions. Therefore, all initial predictors used for GDM modelling were fitted simultaneously. 500 trees were calculated for each model. In contrast to GDM, GF can handle nominal predictors without dummy transformation.

Moreover, grassland and forest sites of the Biodiversity Exploratory regions were analysed in separate models for GDM and GF to incorporate predictors being exclusively available for grassland or forest sites.

2.2.10 Statistical niche modelling

The niche is an abstract n-dimensional space of environmental predictors describing the presence of a species (Holt, 2009). However, only an educated guess based on literature is possible which environmental predictors are relevant defining the niches of soil bacteria in temperate forests and grasslands. Moreover, even if the relevant predictors with biological meaning to a soil bacterial community can be identified, it is not know how important each predictor is for each bacterial species.
Here, relevant predictors for mulitdimensional niche space were derived from GF models for all SV having $R^2 > 0.5$ in GF modelling of grassland and forest communities. Hence, only SV were analysed for which a minimum threshold of predictive power was provided. Initially, trajectory clustering in function `kmlShape::kml` was used to group the fits of predictor importance for all SV to the most important predictors. However, the retrieved groups of soil bacteria could only be described by qualitative predictor properties, e.g. low or high pH optimum in conjunction with low to high importance of pH. Therefore, optimal niche values were estimated along environmental predictor gradients for SV where gF modelling identified predictive power, retrieving quantitative optimal niche values for each predictor for each SV. Optimal niche values represent the maximum relative abundance of a bacterium along a predictor gradient and were determined by fitting hierarchical logistic regression as in package `eHof` (Jansen and Oksanen, 2013). In more detail, optimal niche values were determined with function `HOF` using a poisson error distribution and 50 bootstrap samplings to check model robustness. This approach is similar to the method used in Sikorski et al. (2022). All predictors were scaled between 0 and 1 for direct comparability between predictor gradients. Binomial model types were excluded, allowing only for model types fitting explicit optimal niche values. After bootstrapping the best model fit is retrieved.
In addition to optimal niche values, SV specific variable importances were derived from GF models, estimating the relative importance of each predictor for each SV. Hereby I assume that not all predictors are equally important for each soil bacterium. I imply a hierarchy of environmental predictor importance, allowing for a more flexible estimation of bacterial coexistence in soil, weighted by SV-specific variable importance of optimal niche values.
Data matrices of optimal niche values and predictor importance for each SV were transformed into two separated Euclidean distance matrices representing the position of each SV in a multidimensional space of optimal niche values and predictor importance. The ef-

fects of variable importance and optimal niche value were joined by addition of the square root of squared Euclidean distances for each SV. Like that, the position of each SV in multidimensional predictor space, defined by its pattern in optimal niche values, is reshaped by the pattern in variable importance of each SV for each predictor. These joint distance matrices were clustered to identify potentially coexisting bacteria sharing regions of similar predictor space using Density-based spatial clustering of applications with noise (DBSCAN), available in package `dbScan` with function `dbScan`. DBSCAN estimates the density of other SV around each SV according to a specified *eps* value and applies a specified minimum number of points *(minPts)* to separate different core and noise points (Ester et al., 1996). SV were joined into a cluster if there is a chain of SV where one falls inside the *eps*-neighbourhood of another SV. SV identified as border points were defined as noise. Ranges of parameters *eps* and *minPts* of 0.2-0.3 and 5-15 were scanned respectively for optimal clustering results. For niche modelling of grassland and forest bacteria, *eps* and *minPts* were 0.25 and 14 and 0.25 and 12, respectively.

2.3 Results

2.3.1 Data structure

The experimental sites of the Biodiversity Exploratories are organized in three regions across Germany. Edaphic properties clustered accordingly, e.g. high clay contents in soil were clustered in the Schwäbische Alb and decreased for Hainich and especially Schorfheide (Fig. 3). This was already shown by Kaiser et al. (2016). Furthermore, climatic predictors, e.g. sum of rain days over 1056 days and Ts over 4°C in 1056 days, were clustered by region as well, with Alb being cooler and having more days with precipitation than Hainich. Schorfheide was even more dry and warmer than Hainich and Alb. Furthermore, climatic showed a longitudinal changes as well (Fig. 4). The sum rain days over 1056 d, i.e. the sum of rain days summed up over a period of 1056 d before sampling in 2011, 2014 and 2017 respectively, decreased on average by 6.3 % between 2011 and 2017. Moreover, the Ts over 5°C in 1056 d, i.e the number of days with soil temperature over 5°C over a period of 1056 d before sampling, increased by 8.7 % between 2011 and 2017. Likewise, SM over 22 % in 365 d decreased on average by 12.17 % from 2011 to 2017- Considering the respective sample size of 296 sites per time point, those differences were statistically significant ($p<0.05$).

In consequence, these examples demonstrate that regions, climate and soil texture properties showed multicollinearity in this experimental setup, although representing independent potential drivers for soil bacterial diversity in the Biodiversity Exploratories. This poses a challenge to statistical modelling, since an identification of the most relevant drivers of bacterial diversity might be biased.

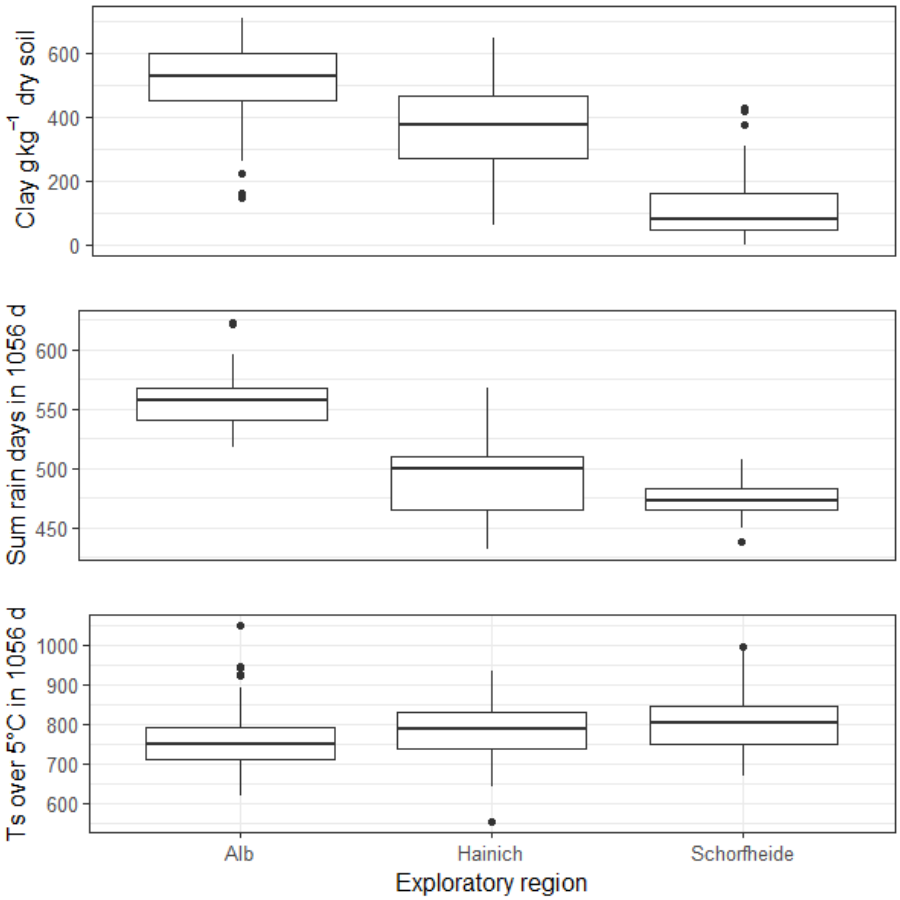

Figure 3: Differences of representative environmental predictors between regions of the Biodiversity Exploratories. Clay = soil clay content, sum rain days in 1056 d = the number of days with precipitations over 1056 d before sampling, Ts over 4°C in 1056 d = the number of days with mean soil temperature over 4°C in 1056 d before sampling.

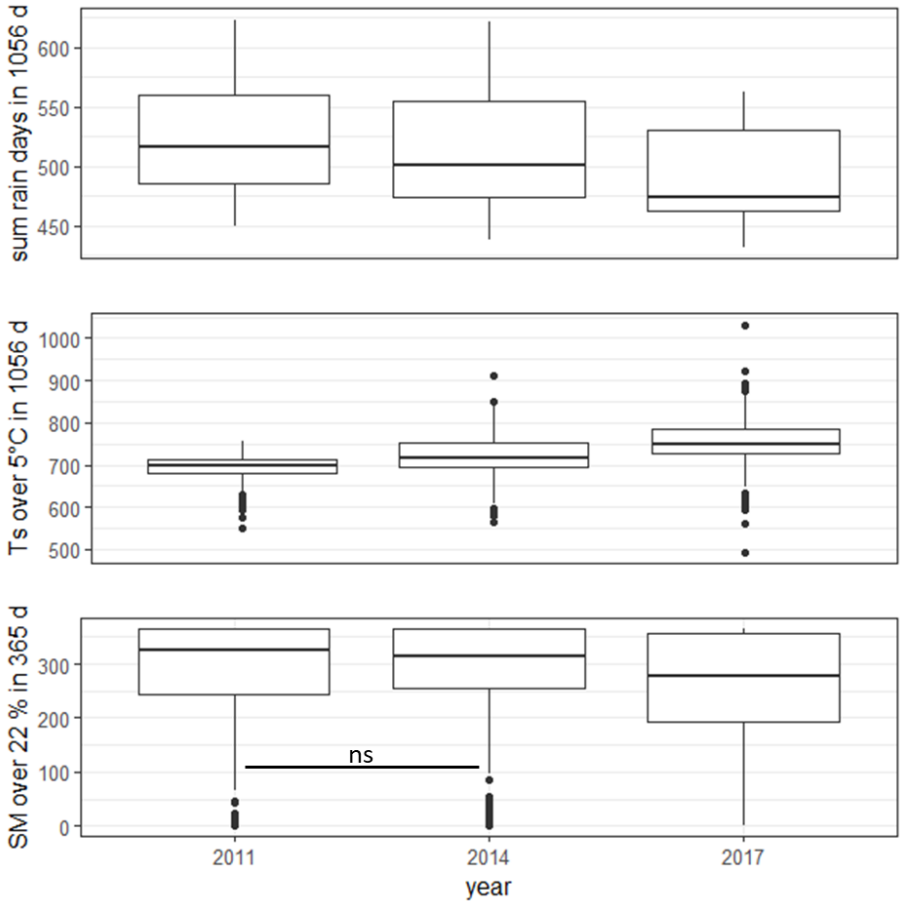

Figure 4: Differences of representative environmental predictors between sampling years. Sum rain days in 1056 d = the number of days with precipitations over 1056 d before sampling, Ts over 4°C in 1056 d = the number of days with mean soil temperature over 4°C in 1056 d before sampling, SM over 22 % in 365 d = number of days with mean soil moisture over 22 % in 365 d before sampling. Pairwise comparisons between groups were determined by Tukey HSD post-hoc testing following ANOVA. Non-significant pairwise comparisons were marked by "ns". All other pairwise comparisons were significant ($p<0.05$).

2.3.2 Sequencing results and community data structure

In total 296 forest and grassland sites were examined for 2011, 2014 and 2017. The experimental sites of the Biodiversity Exploratories are organized in three regions across Germany. This structure applied also to edaphic predictor properties like soil texture and climate, differing on average between regions (Fig. 3). This was already shown elsewhere (Kaiser et al., 2016).
The sequencing of the V3 region of the 16S cDNA resulted in 230,659 SV detected in 888 samples. All samples ranged in sequencing depth from 50,801 to 1,850,110 reads. There was no systematic difference in sequencing depth between sampling years (Fig. 5).

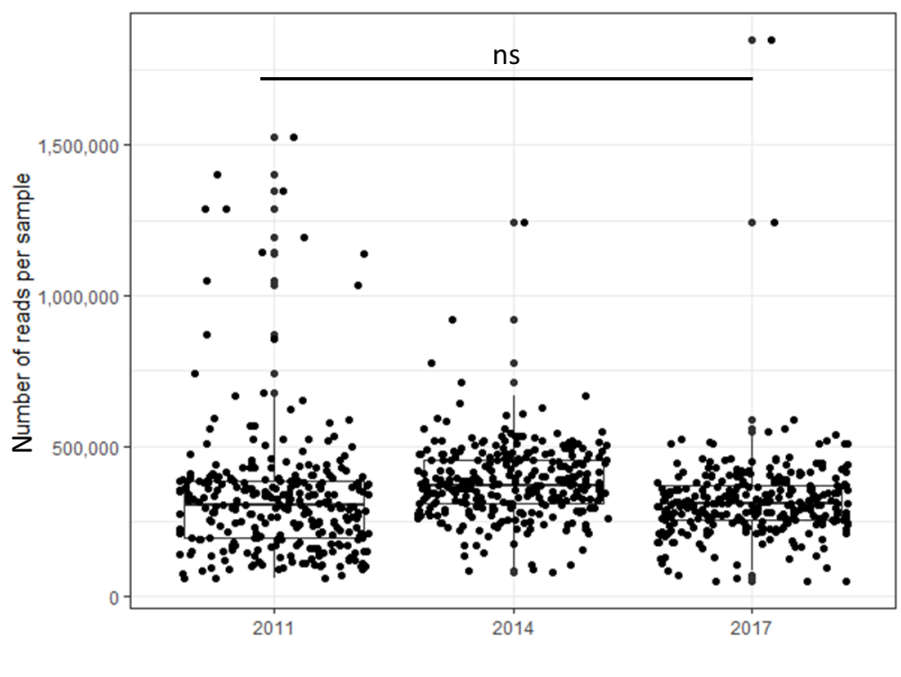

Figure 5: Raw sequencing depth distribution between years. Samples of 2011 were sequenced on HiSeq 2500. Samples of 2014 and 2017 were sequenced on NextSeq 500. Non-significant differences ($p>0.05$) in sequencing depth between sampling years are represented by black bars labelled "ns" for non-significant. Pairwise comparisons between groups were determined by Tukey HSD post-hoc testing following ANOVA. Black dots represent independent samples.

All samples were coverage rarefied (Chao and Jost, 2012) to a minimum sample coverage of 98.2 %. After sample normalization sequencing depth ranged from 19,197 to 304,705 reads.

Alpha diversity describes the diversity contained in a sample. Here, richness and Simpson indices were used to represent the total number and highly abundant SV in a sample. The mean richness over all sites was 8,236 and ranged from 1,049 to 19,161. In comparison,

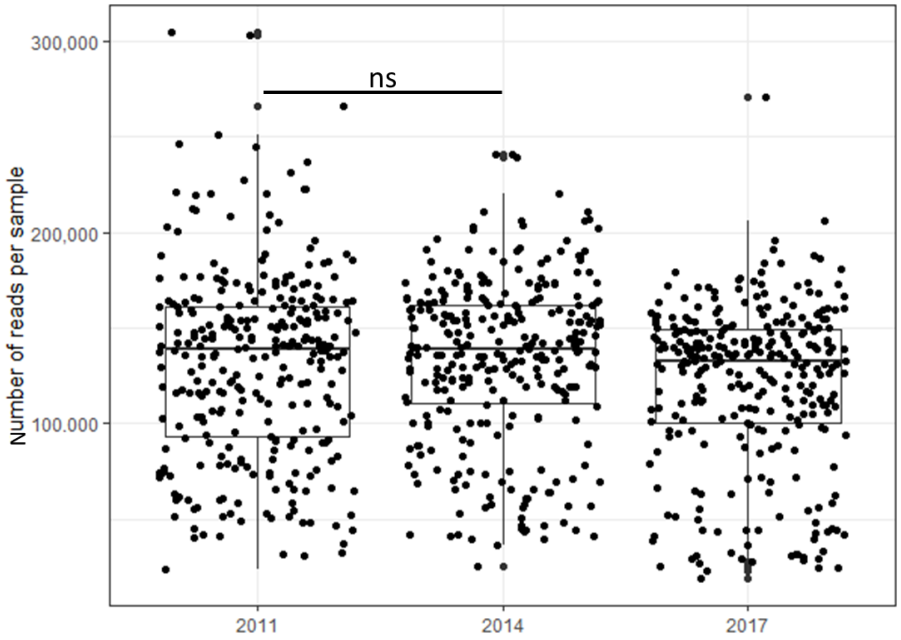

Figure 6: Sequencing depth of coverage rarefied samples. Samples of 2011 were sequenced on HiSeq 2500. Samples of 2014 and 2017 were sequenced on NextSeq 500. Non-significant differences ($p>0.05$) in sequencing depth between sampling years are represented by black bars labelled "ns" for non-significant. Pairwise comparisons between groups were determined by Tukey HSD post-hoc testing following ANOVA. Black dots represent independent samples.

mean Simpson diversity index was 365 SV and ranged from 10 to 1,431 SV.
Moreover, richness and Simpson indices differed between regions. Alb and Hainich were more similar than Schorfheide. For richness no significant differences between Alb and Hainich and different land use types were detected. In contrast, Schorfheide had significantly higher richness for grassland and significantly lower richness for forest sites with averages 10,744 and 3,126 SV, respectively (Fig. 7A).
A systematic difference between grassland and forest was detected for the Simpson index. Forest sites of all Exploratories contained significantly less highly abundant species than grassland sites. Besides, the Simpson index was highest for Alb. Hainch and Schorfheide contained lower Simpson index values, with Schorfheide harbouring the least number of abundant species.

Similar to this geographic effect, alpha diversity varied over time (Fig. 7B). Richness differed significantly between 2011 and 2017. Grassland and forest sites dropped by 1041 and 1500 SV, representing a decline of 10.3 % and 19.5 % respectively.
Again, this pattern was even more pronounced for the Simpson index, declining signifi-

cantly from 2011 to 2014 and from 2011 to 2017. Though, differences between 2014 and 2017 were not significant. On average, the difference in abundant species between 2011 and 2017 for grassland and forest sites was 95 and 188 SV, corresponding to a reduction of 30.2 % and 34.2 % respectively.

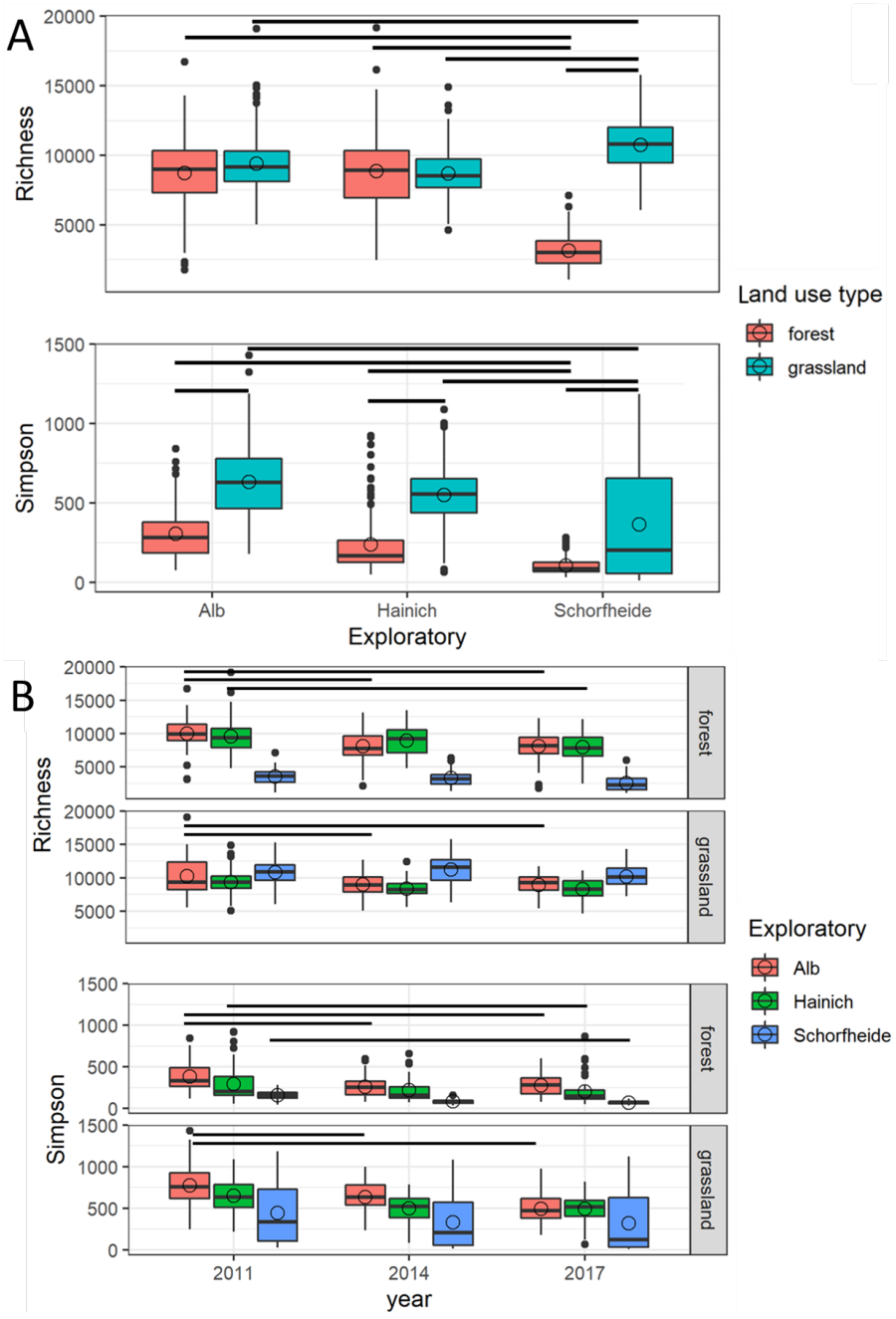

Figure 7: Alpha diversity of Biodiversity Exploratory regions separated by land use type (A) and alpha diversity in Exploratory regions over time separated by land use type (B). Richness represents all detectable species in a sample irrespective of abundance. The Simpson index represents highly abundant species. Transparent circles represent the mean next to the median displayed as horizontal bars. Vertical axis represents number of species. Differences of alpha diversity between land use type in an Exploratory and between Explorataries of an land use type were tested by Tukey HSD post-hoc testing following ANOVA. Significant differences ($p<0.05$) between groups are indicated by by black bars.

In addition to alpha diversity, the aim of this study was to deduce the influence of environmental predictors on community composition and turnover. Principal coordinate analysis of beta diversity, the diversity between sites, represented by Bray-Curtis dissimilarities revealed no systematic trajectory of bacterial communities over time (Fig. 8A). Observed shifts between sampling time points were small as displayed for each site. Together, axis 1 and 2 accounted for 51.8 % of data variation. Pairwise comparisons of beta diversity ordinations between sampling years revealed that the similarity of ordinations were not significant ($p>0.05$) with a Pearson correlation of 0.61 and 0.49 between 2011-2014 and 2014-2017 using procrustes rotation. However, the similarity between 2011-2017 was significant with a Pearson correlation of 0.85 and p-value of 0.001. Albeit, an arch-effect of sample projection was present in the 2-dimensional ordination space. This pattern is common in bacterial community analysis and indicates niche differentiation and community turnover along predictor gradients (Morton et al., 2017). Here, the overlay of soil pH onto beta diversity ordination revealed a clear gradient (Fig. 8B).

Most species in soil bacterial communities are rare and only sporadically abundant (Bickel and Or, 2021). For that reason, it is statistically challenging to differentiate between biological signal and random noise for rare SV. Therefore, the analysis of beta diversity was restricted to SV being $\geq 0.1\,\%$ relative abundance in any sample. 6,192 SV were recovered, providing more reliable abundance distributions for statistical analysis (Szoboszlay et al., 2017). Also, sparsity could be reduced by 20 %. Coverage rarefied sequences and cropped coverage rarefied sequences were compared using Procrustes rotation and significance testing, measuring the similarity between ordinations, to see how cropping of SV affected beta diversity (Fig. 8C). The structure between ordinations was preserved and significantly ($p<0.001$) correlated (Pearson = 0.92), but cropped data were drawn closer to the center for all samples.

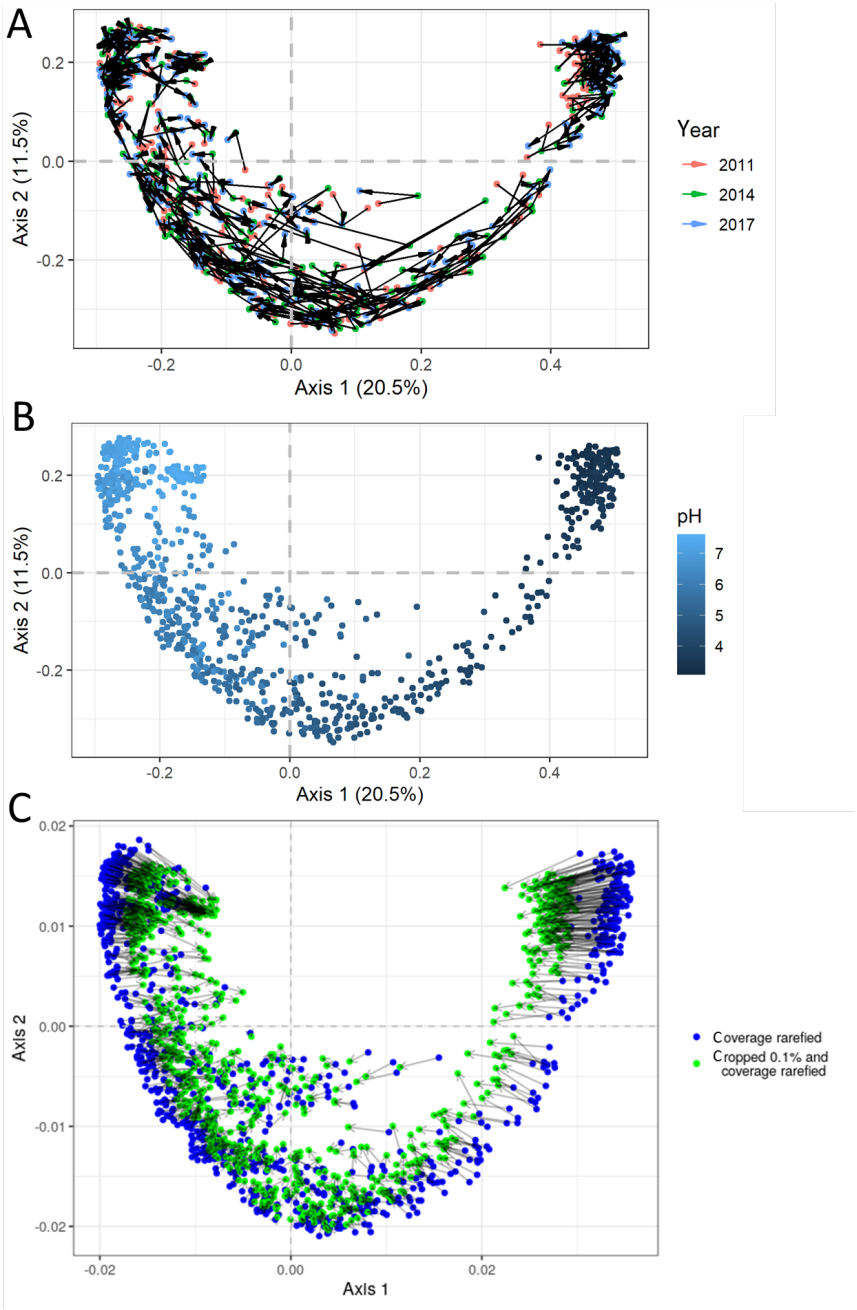

Figure 8: Principal coordinate analysis of Bray-Curtis dissimilarities on coverage rarefied sequences from 888 samples. Arrows represent the trajectories of each site from sampling timepoints 2011 to 2014 and 2017 (A). Blue colour gradient depicts soil pH (B). Procrustes rotation of coverage rarefied sequences (blue) and coverage rarefied sequences cropped to SV with $\geq$0.1 % relative abundance in any sample (C). The similarity between ordinations was significant ($p<0.001$) and Pearson correlation = 0.92.

2.3.3 Drivers of alpha diversity

Generalized additive modelling was used to identify environmental drivers for richness and Simpson index of forest and grassland sites of the Biodiversity Exploratories. It is not know yet for the present data whether climatic predictors influence alpha diversity and whether climate change is already affecting soil bacteria. Furthermore, temporal and legacy effects were examined as well.

GAMs fitted on grassland and forest sites separately did not identify any predictors exclusively available to either data sets (Fig. 54-55 in Supp. Material). Therefore, GAMs were fitted on all sites independent of land use type for richness (Fig. 9). In case a difference in alpha diversity is present between grassland and forest sites, land use type was included as categorical predictor in the model fitting process.

The GAMs of alpha diversity showed independent effects of significant predictors, displayed as partial effect plots, depicting the model fit for a specific predictor, keeping all other predictors constant at their median. Randomly selected site 27 of of Hainich grassland was chosen as reference to account for the random intercept effect. Each fit must be imagined as a transect through a multi-dimensional sample space, representing a single predictor fit. Therefore, the obvious misfit of raw data and partial predictor effect is a result of partial effect plotting. Although spatial pattern were not the scope of this study, the interaction spline of latitude and longitude of sampling sites were fit in the model to account for spatial autocorrelation between sites in all models. However, effects of spatial autocorrelation were not analysed further.

Overall, this model accounted for 87.5 % deviance explained. Deviance is a generalization of sum of squares for models fitted using maximum likelihood. Richness of the Biodiversity Exploratories was dominated by temporal variation and pH. Richness declined significantly, indicating a trend over time. Moreover, pH showed a non-linear fit with a maximum richness around 6.5. Soil texture reduced richness for high sand contents. Below >75 %, no effect of sand content on richness was visible. However, soil texture was only significant to a level ($p<0.1$). Soil moisture (SM), represented in the model as SM >14 % in 1056 d and Range of SM in 1056 d, increased richness linearly. Soil temperature (Ts), as Ts >15°C in 365 d reduced richness linearly, however the slope was narrow, since maximally 142 d with Ts >15°C were documented. Similarly, richness dropped for sites having >850 d Ts >4°C in 1056 d. Dithionite extractable Fe lead to an increase in richness up to 5 $g\,kg^{-1}$.

In addition to richness, the alpha diversity of highly abundant bacterial species was analysed as Simpson diversity index (Fig. 10). Similar to richness, separate models for grassland and forest sites did not identify any predictor exclusively available to either grassland or forest data sets (Fig. 57 - 56). Therefore, a GAM for the Simpson index was fitted on all sites to increase statistical power, by doubling in sample size. Year, bulk density, pH,

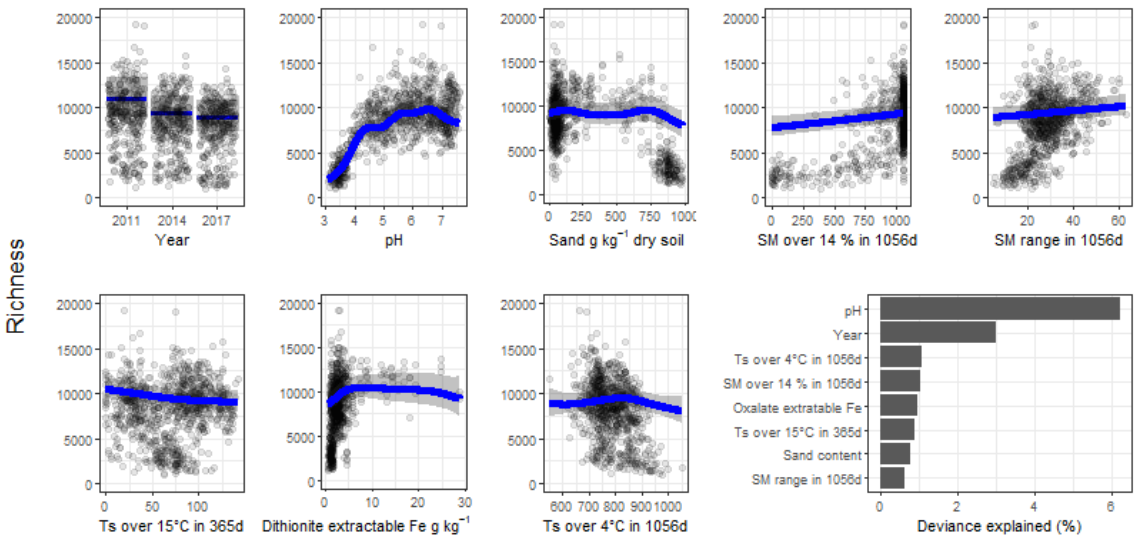

Figure 9: GAM model of bacterial richness fit on all forest and grassland sites. Subplots show the partial fit of all predictors as black lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

range of SM in 365 d, clay content and SM >42 % in 1056 d were selected as significant predictors. In total 88.8 % of deviance explained was covered by the model. Solely the soil temperature effect of forest sites was not recovered for all sites. Similar to richness, a decline of highly abundant species over time was detected. Also, pH and temporal variation were the most important predictors for the Simpson index. In contrast to richness, the Simpson index increased linearly with pH. Bulk density reduced the Simpson index non-linearly. This effect was largest for low density soils. SM led to an increase in Simpson index, indicated by the range or number of days >42 % in 1056 d. Soil texture represented by clay content was least important, increasing the abundance of highly abundant species with increasing clay content.

Deviance explained by climatic predictors was larger than soil texture in all models, emphasizing the impact of climate on bacterial alpha diversity in soil. For richness those predictors were Ts >4 °C in 1056 d, Ts >15 °C in 365 d and SM over 14 % in 1056 d. For the Simpson index, SM >42 % in 1056 d and the range of SM over 1056d were more important than soil texture. Those climatic predictors retained by GAMs, were only moderately correlated with each other or did not occur in the same model (Fig. 11). Hence, no multicollinearity was retained in the final alpha diversity models. This is important to mention since multicollinearity between climate predictors occurred (Fig. 49 in Supp. material) and could obscure model results. As an interesting property, Ts threshold data showed

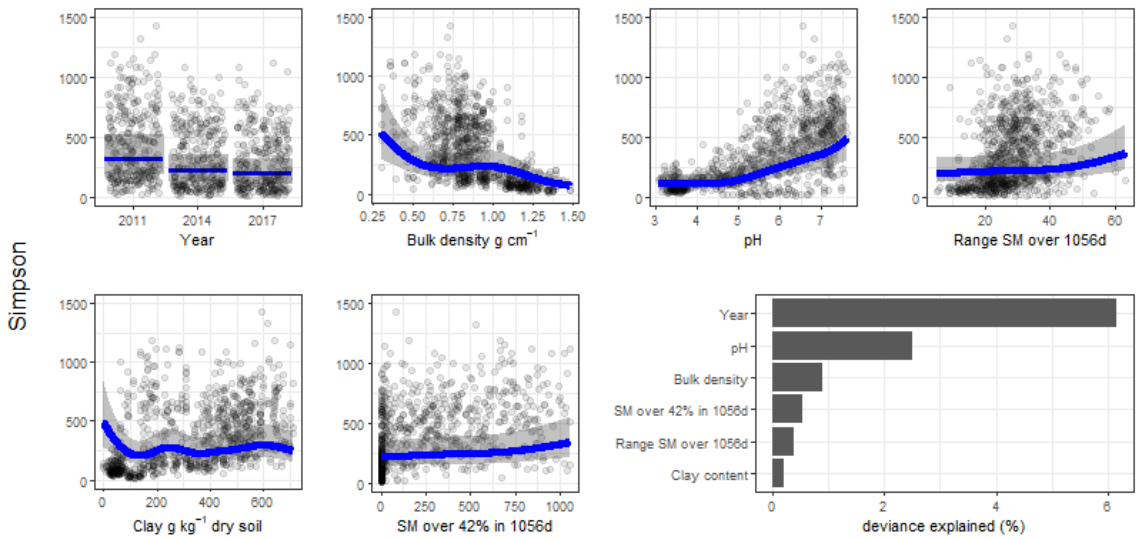

Figure 10: GAM fitted to Simpson diversity index of all sites. Subplots show the partial fit of all significant predictors as black lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

changes over time as highlighted in Fig. 11 and 12. ANOVA and Tukey HDS post-hoc testing revealed a significant increase in the number of days with soil temperature >4 °C and >15 °C over the years. Moreover, during initial model fitting, the predictor "sum rain days" was always selected during model fitting. This predictor declined significantly over time as well, which indicates a consequence of climate change (Fig. 12). However, this promising predictor had to be excluded from models due to collinearity with soil texture data, caused by the grouping of similar precipitation pattern as well as the soil texture composition in the same Exploratory regions (data not shown). A significant decline in soil moisture over time could only be observed for Alb, indicated by the number of days with SM >42 %. Although significant, the differences between SM >14 % were neglectable. Range of SM declined significantly for Hainich and Schorfheide between 2011 and 2017. Alb however increased between 2011 and 2017. There was no indication in the underlying data if either minimum or maximum SM influenced SM range (data not shown).

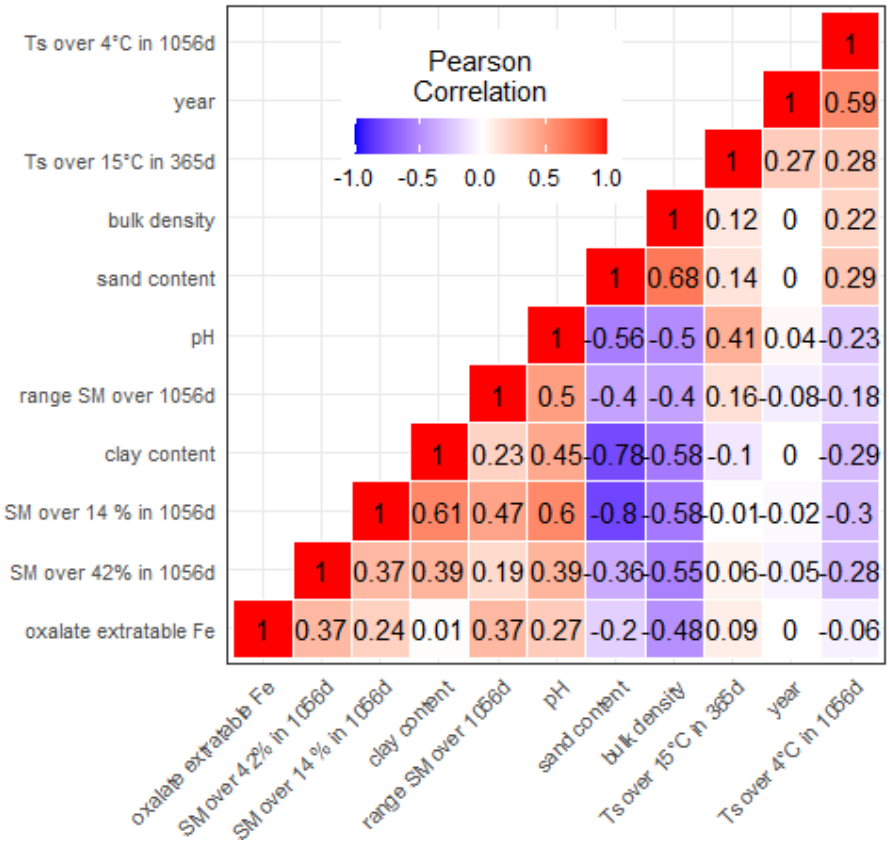

Figure 11: Pearson correlation coefficients of significant predictors retained by GAM forward selection in alpha diversity modelling.

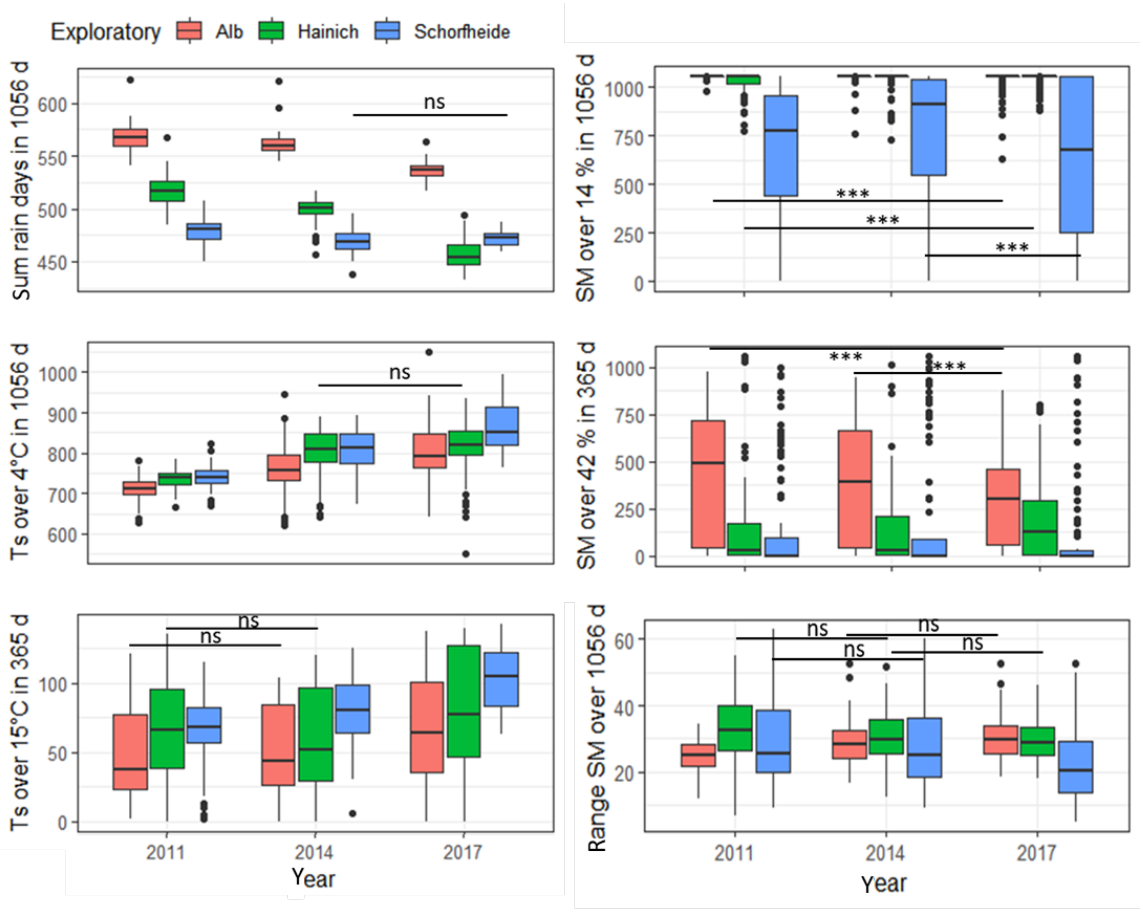

Figure 12: Significance testing of climate predictors between years retrieved by alpha diversity modelling. Pairwise comparisons between years were tested for each Exploratory separately. For better readability, black bars represent either significant ($p<0.05$) differences indicated by $***$ or non-significant differences indicated by "ns".

2.3.4 Generalized dissimilarity modelling of beta diversity for grassland and forest sites

Beta diversity of the Biodiversity Exploratories for 2011, 2014 and 2017 were analysed separately for grassland and forest sites due to large number of predictors solely available for either group using generalized dissimilarity modelling (GDM).

In GDM, beta diversity was modelled as a distance matrix on Bray-Curtis dissimilarity. 9 predictors affected beta diversity of grassland sites significantly (Fig. 13) and explained 65.79 % of model deviance. pH was by far the most important predictor and shaped community turnover constantly over the whole gradient. Similarly, the fungal to bacteria ratio of phospholipid fatty acid (PLFA) data linearly influenced the bacterial community data, however retrieved only half the importance of variable pH. Porosity gained importance for beta diversity especially for soils with high pore space. Also, soil moisture (SM) as median SM in 1067 days before sampling was of similar importance as porosity, with high average SM having more impact on community turnover, whereas low SM was of no effect. Dithionite extractable Al shaped beta diversity mostly for concentrations $< 5\,\mathrm{g\,kg^{-1}}$. Geographic distance between sites influenced community composition up to a distance of 200 km. Again, geographic distance was included to account for spatial autocorrelation. Soil texture in form of clay content shaped bacterial community up to 40 %, higher contents were of no effect for beta diversity. Community turnover over time from 2011 to 2017 implied a constant bacterial turnover rate. However, variable importance was small. GDM cannot model categorical predictors. Therefore soil type was dummy transformed. Consequently, GDM could only identify that not sharing the same soil type influences community turnover.

The GDM analysis of forest sites explained 81.6 % of model deviance. Again, 9 significant predictors were found and pH the most important predictor (Fig. 14). However, the drop in variable importance between pH and the remaining predictors was even more pronounced as for grassland sites. Community turnover was higher for low pH sites and reduced towards higher pH, but community turnover was present all along pH gradient. Soil clay content influenced beta diversity only up to 30 % clay content. Beyond, clay content did not affect community composition. Dithionite extractable Al affected beta diversity only $> 10\,\mathrm{g\,kg^{-1}}$. Maximum SM < 40 % in 365 days before sampling affected community change. The same applied to the soil stable phosphorus (P) pool with minor influence for low P contents of up to 200 $\mathrm{mg\,kg^{-1}}$. In contrast, soil temperature as the number of days over 15 °C in 365 days before sampling implied a constant rate of community turnover. The C content in O_e horizon $> 40\,\%$ contributed to biological turnover. The effects of the categorical predictors main tree species and soil type allowed only for a rudimentary analysis, showing that sites not sharing the same soil type or main tree species affected community turnover.

However, the effect of soil type was neglectable, as already seen for grassland sites. Comparing grassland and forest sites, the effect of pH as most important predictor was confirmed for both land use types. Moreover, the significant effects of clay, extractable Al, soil moisture and a minor effect of soil type were also found in both models.

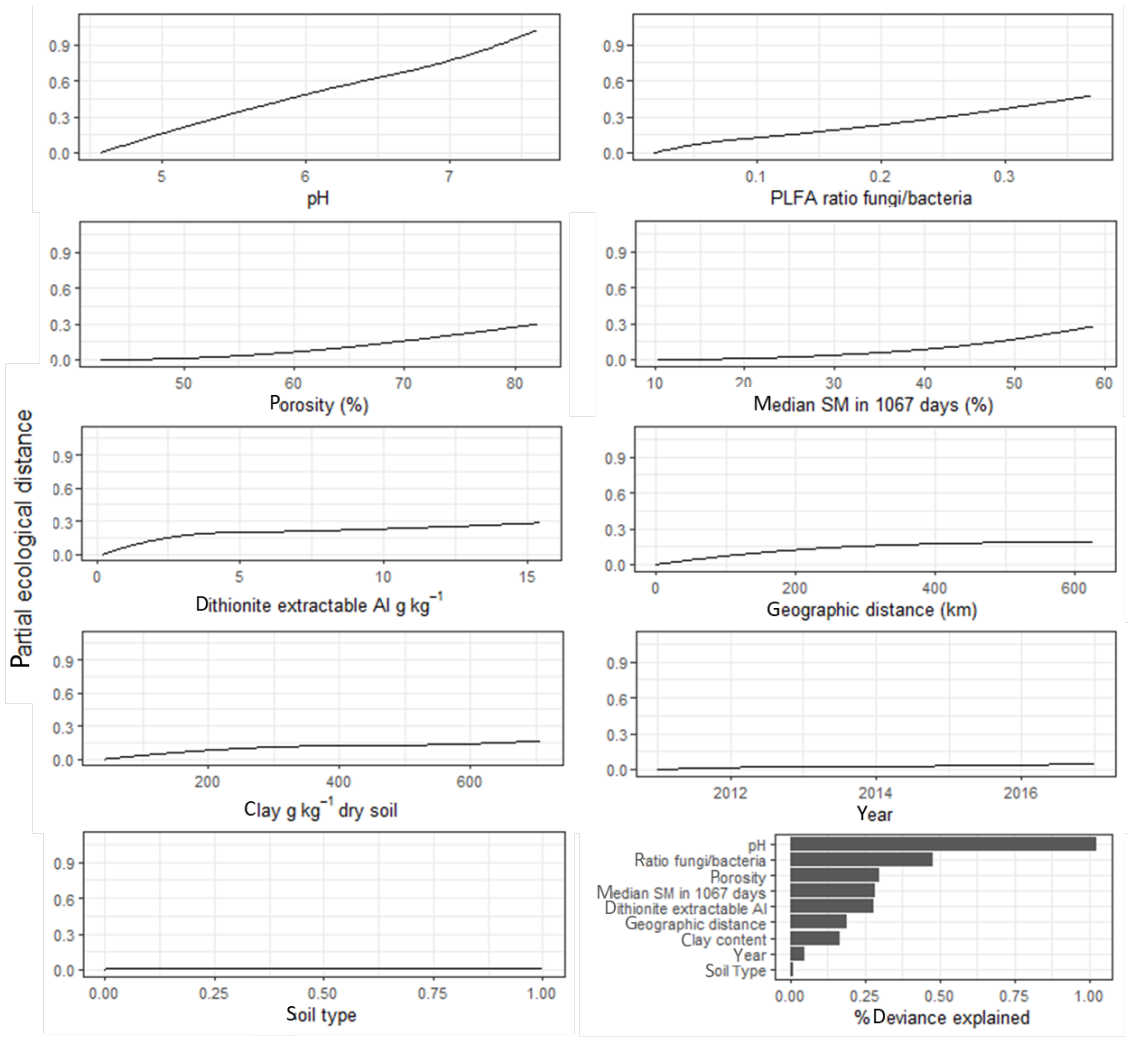

Figure 13: Significant ($p < 0.05$) predictors affecting beta diversity of grassland sites analysed by GDM of Bray-Curtis dissimilarities. Plot 1-9 display the predictor gradient on the x-axis and its effect on community turnover indicated by an increase in partial ecological distance on the y-axis. The inclination of each graph represents community turnover along each predictor gradient. The steeper the slope, the more turnover is caused by the respective value of predictor gradient on the x-axis. The final barplot displays the partial variable importance of significant predictors as deviance explained.

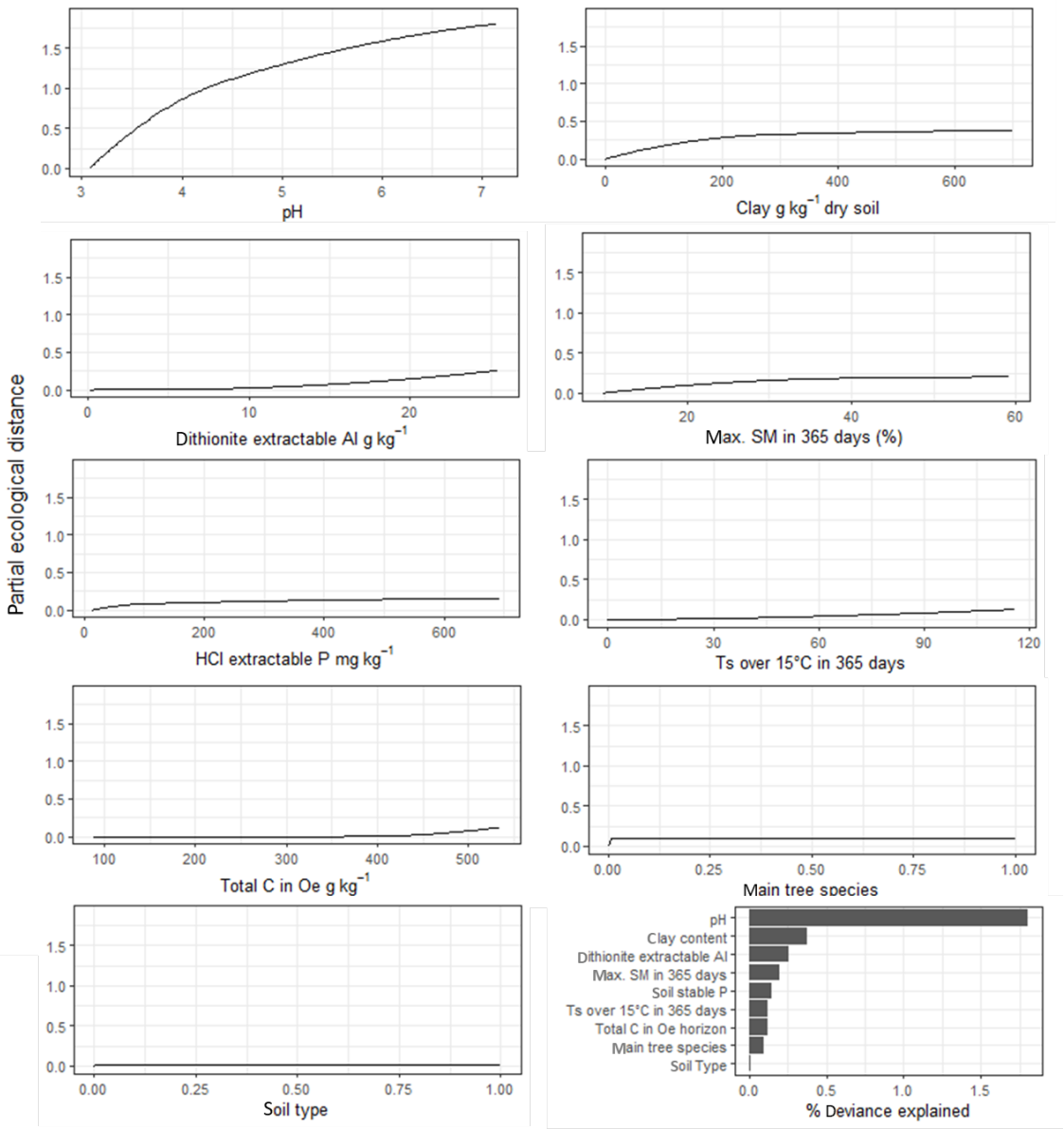

Figure 14: Significant ($p < 0.05$) predictors affecting beta diversity of forest sites analysed by GDM of Bray-Curtis dissimilarities. Plot 1-9 display the predictor gradient and its effect on community turnover indicated by an increase in partial ecological distance. The inclination of each graph represents community turnover along each predictor gradient. The steeper the slope, the more turnover is caused by the respective value of predictor gradient on the x-axis. The final barplot displays the partial variable importance of significant predictors as deviance explained.

2.3.5 GradientForest analysis of beta diversity for grassland and forest sites

GradientForest (gF) analysis is complementing GDM analysis by modelling betaG diversity directly as abundance turnover of each SV in the bacterial community. In total 6192 SV were analysed. For the grassland sites, 3848 SV retrieved $R^2 > 0$ and 740 SV had $R^2 > 0.5$, which means that more than 50 % of their variance in abundance could be explained by GF. Variable importance identified pH as the most important predictor for grassland sites, followed by soil Fe content as oxalate and dithionite extractable Fe, soil pores distribution represented by fine coarse pores, wide coarse pores and fine pores, nutrient availability in the form of organic C content and C/N ratio. Soil texture in the form of clay and sand content was selected as well (Fig. 15). It has to be considered that organic C is a proxy for soil fertility and nutrient availability due to its high correlation to other soil nutrients, such as soil nitrogen and soil sulphur contents (Fig. 51).

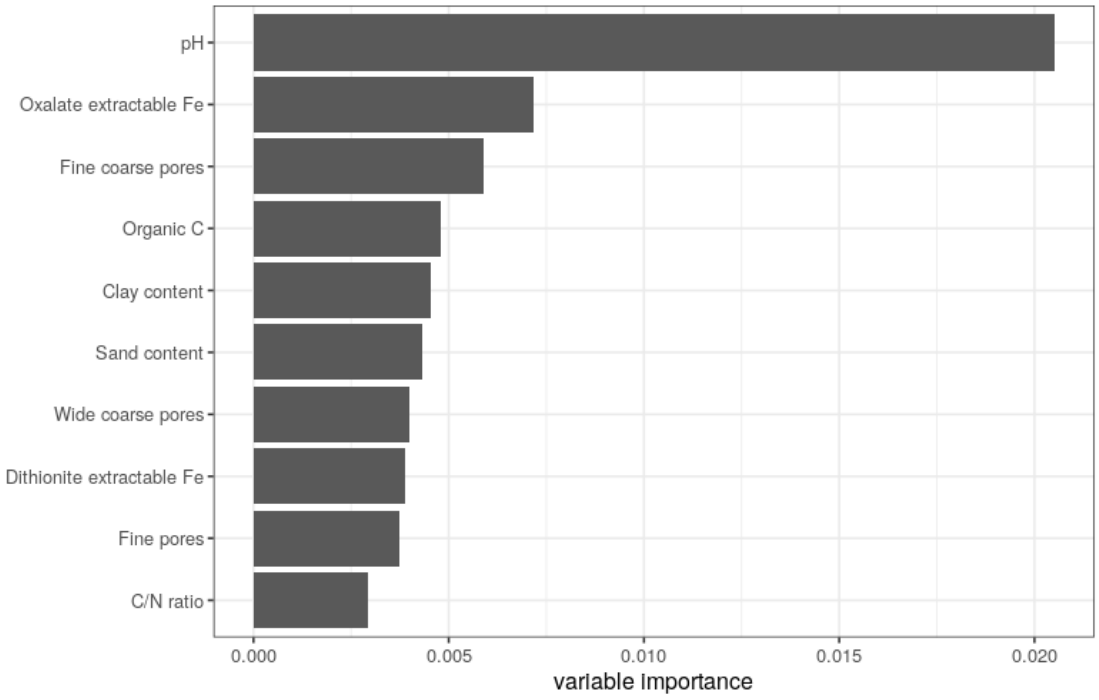

Figure 15: Variable importance of the 10 most important predictors of gradientForest analysis of grassland sites displayed as average proportion of variance explained for all species.

Community turnover for GF is depicted as the mean turnover of all modelled bacteria with a predictive power >0 in R^2 (Fig. 16). Similar to GDM modelling, species turnover occurred along the whole gradient of pH. For oxalate extractable Fe, community turnover occurred around 10 and >25 $g\,kg^{-1}$. The content of fine coarse pores in soil resulted in community turnover <4 %. Higher community turnover for high contents of organic C

occurred $> 20\,\mathrm{g\,kg^{-1}}$, corresponding to >20 %. Turnover induced by clay content was low but appeared along the whole gradient. This pattern was similar for sand content and wide coarse pores. Major community turnover was similar for dithionite and oxalate extractable Fe for minimum and maximum Fe contents. Fine pores showed a response solely <15 %. The C/N ratio of the soil induced higher turnover >12.

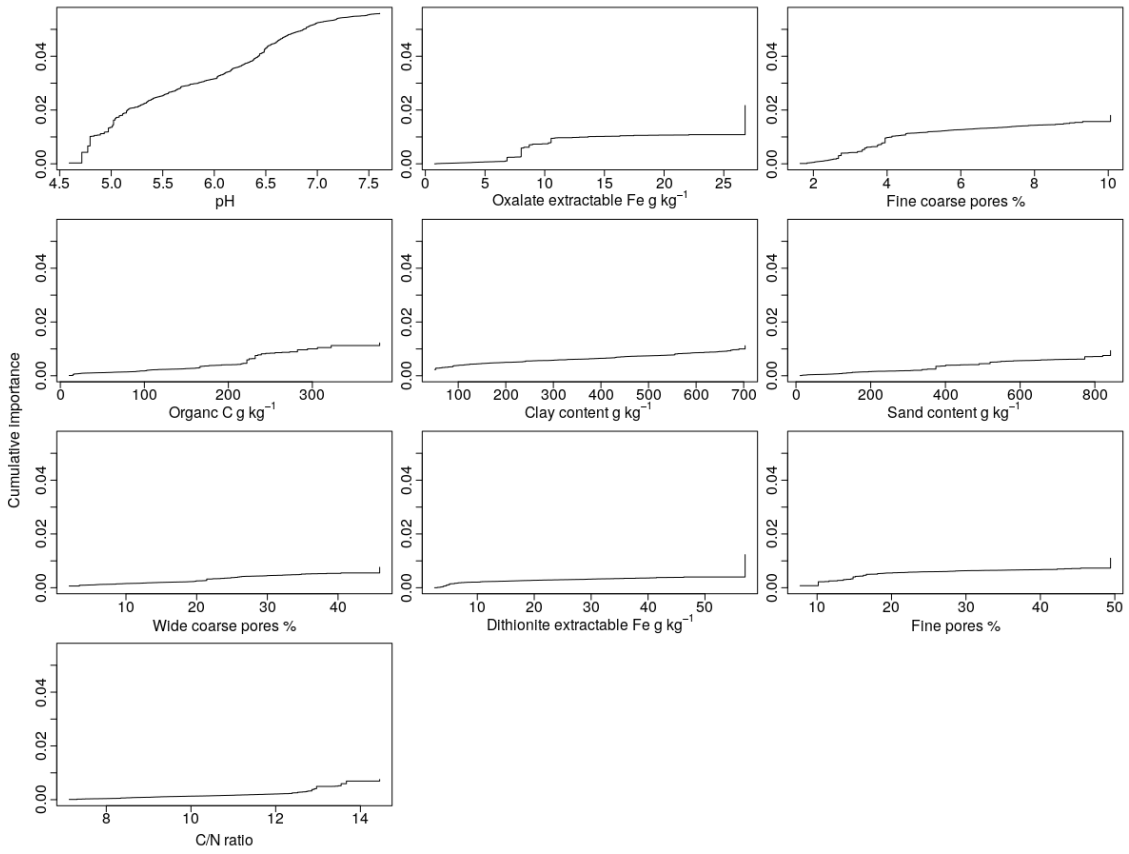

Figure 16: Community turnover along the 10 most important predictor gradients of gradientForest for grassland sites. The y-axis is normalized to integrate variable importance, displayed as the mean turnover of all modelled 6192 SV. The inclination of each graph represents community turnover along each predictor gradient. The steeper the slope, the more turnover is caused by the respective value of predictor gradient on the x-axis.

Due to the resilience of GF analysis to multicollinearity between predictors, all climatic predictor were fit. However, no climatic predictors were amongst the 10 most important predictors relevant for 6192 SV abundant bacteria >0.1 % relative abundance. Bearing the comparatively low variable importance in mind, SM threshold data clearly showed that the number of days >26 % SM in 1056 d was the most important predictor for SM threshold data (Fig. 17). However, SM threshold data > 26 % in 1056 days were almost as important. In contrast, SM threshold data aggregated over 365 or 730 days identified a

different trend. Here, variable importance increased constantly with SM.

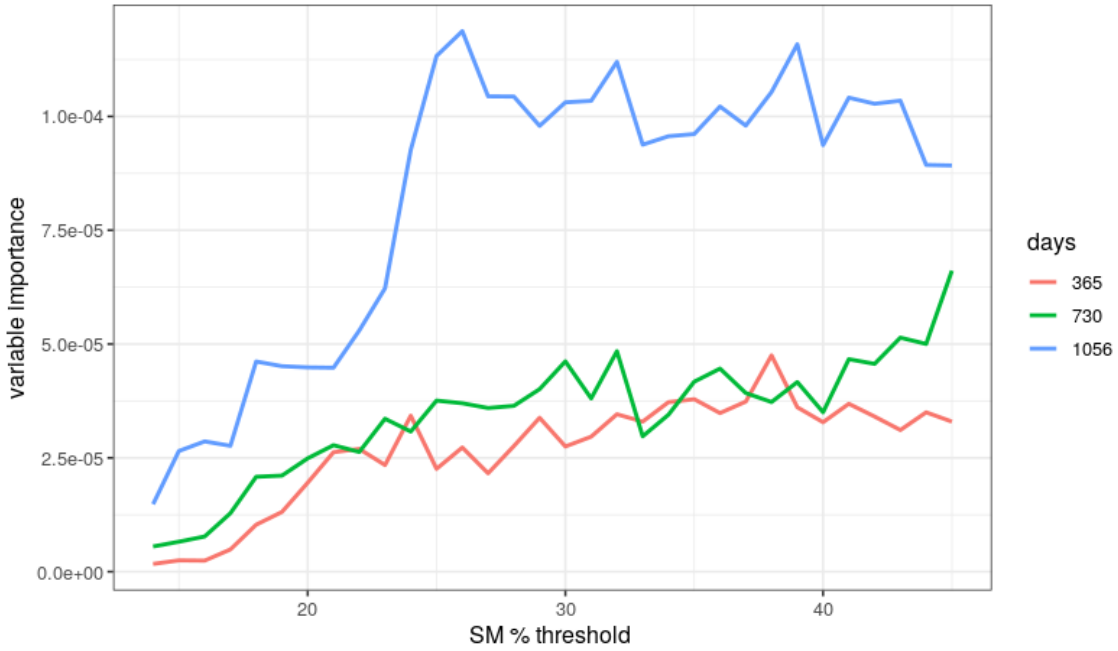

Figure 17: Variable importance of soil moisture (SM) threshold data of grassland sites as retrieved by gradientForest analysis. Line colors code the temporal aggregation of SM threshold predictors, aggregated over a specific number of days before sampling.

In contrast, soil temperature threshold data presented a different relationship to community turnover. First of all, no differences occurred between different temporal aggregations of 365, 730 and 1056 d, showing similar variable importance pattern for each temporal aggregation (Fig. 18). Second, Ts >1-2 °C was most important, representing the number of frost free days. This insights show the potential of GF to gain deeper insight in the dependence of bacteria communities on climate conditions in soil.

For the forest sites, 6179 SV were modelled, 3494 of which retrieved $R^2 > 0$. 550 SV had R^2 >0.5. Similar to the grassland model, community composition of forest sites was mainly influenced by pH (Fig. 19). Other drivers were of minor importance.
Forest sites in the Biodiversity Exploratories harboured more acidic soils <4.5 than grassland sites (Fig. 20). In contrast to grassland, community turnover based on pH was highest around 5.5 and slowed down for pH >6. Labile soil P induced community turnover up to 15 mg kg^{-1}, above that the effect was neglectable. Similar to the grassland sites, soil clay content on forest sites induced minor turnover over the whole gradient. This applied also to dithionite extractable Fe, where turnover happened along the whole gradient. Total C in O_e horizon affected community turnover below 20 % and above >45 %. The impor-

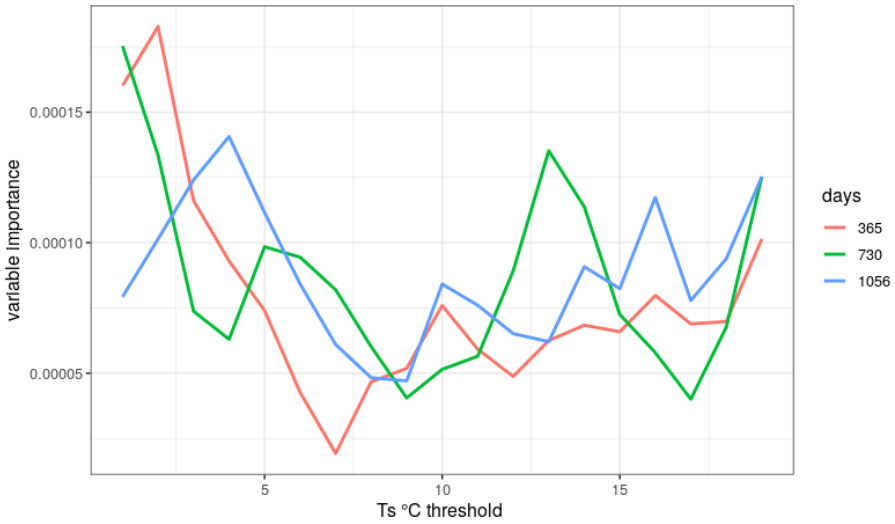

Figure 18: Variable importance of soil temperature (Ts) threshold data of grassland sites as retrieved by gradientForest analysis. Line colors code the temporal aggregation of Ts threshold predictors, aggregated over a specific number of days before sampling.

tance of organic C was minor in comparison to bacterial grassland communities. Moreover, no increased turnover > 20 % was observed. Oxalate extractable extractable Fe showed similar pattern to dithionite extractable Fe, being fundamentally different to the shift in beta diversity induced for the grassland communities. However, the range of soil Fe content relevant for grassland community turnover was not contained for forest soils under investigation. Bulk density, soil stable P content and C/N ratio induced little community turnover over their respective gradient. The C/N ratio reached much wider values than grassland sites, indicating worse substrate qualities than the grassland sites.

As already observed for the grassland sites, no climatic predictors were selected as important predictor next to pH. However, the variable importance of SM threshold data climaxed around 25 % for the data summarizing 1056 days (Fig. 21), implying that this soil moisture level is in comparison most important for community turnover. This is backed by the grassland data also identifying SM >26 % in 1056 days as the most important SM threshold predictor. As for the SM threshold data of forest, no obvious difference between different temporal aggregations was observed. Soil temperature threshold data showed no differentiation between temporal aggregations as well (Fig. 22). In contrast to the grassland sites, an increase of variable importance until days >16 °C in 1056 days for the Ts threshold data was observed.

SM >26 % summarized over 1056 days was identified as the most important threshold value independent of land use type. However, the high variable importance in grassland was influenced by a high response of 3 SVs with variable importance of >0.01, occurring

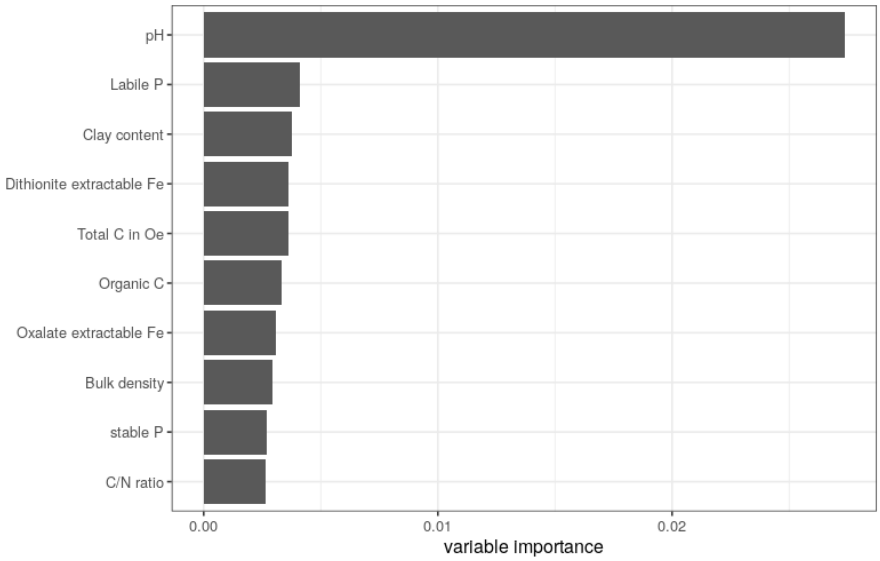

Figure 19: Variable importance of the 10 most important predictors of gradientForest analysis of forest sites displayed as average proportion of variance explained for all species.

predominantly under dry soil conditions < 26 % soil moisture (Fig. 23). They belonged to family *Xanthobacteraceae* in order *Rhizobiales*, genus *Aquisphaera* in order *Isosphaerales* and order *Gaiellales*. Most other SV showed comparatively low variable importance for SM > 26 % (Fig. 23).

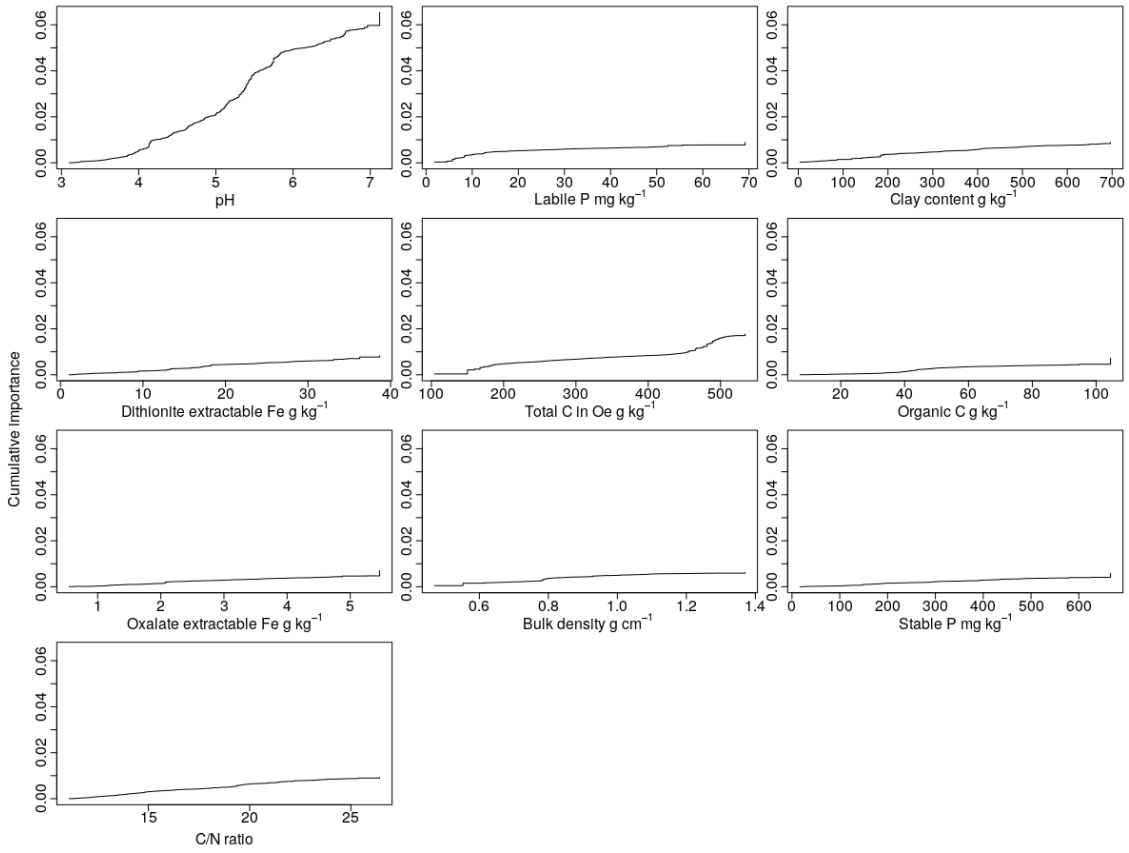

Figure 20: Community turnover along the 10 most important predictor gradients of gradientForest for forest sites. The y-axis is normalized to integrate variable importance, displayed as the mean turnover of all modelled 6192 SV. The inclination of each graph represents community turnover along each predictor gradient. The steeper the slope, the more turnover is caused by the respective value of predictor gradient on the x-axis.

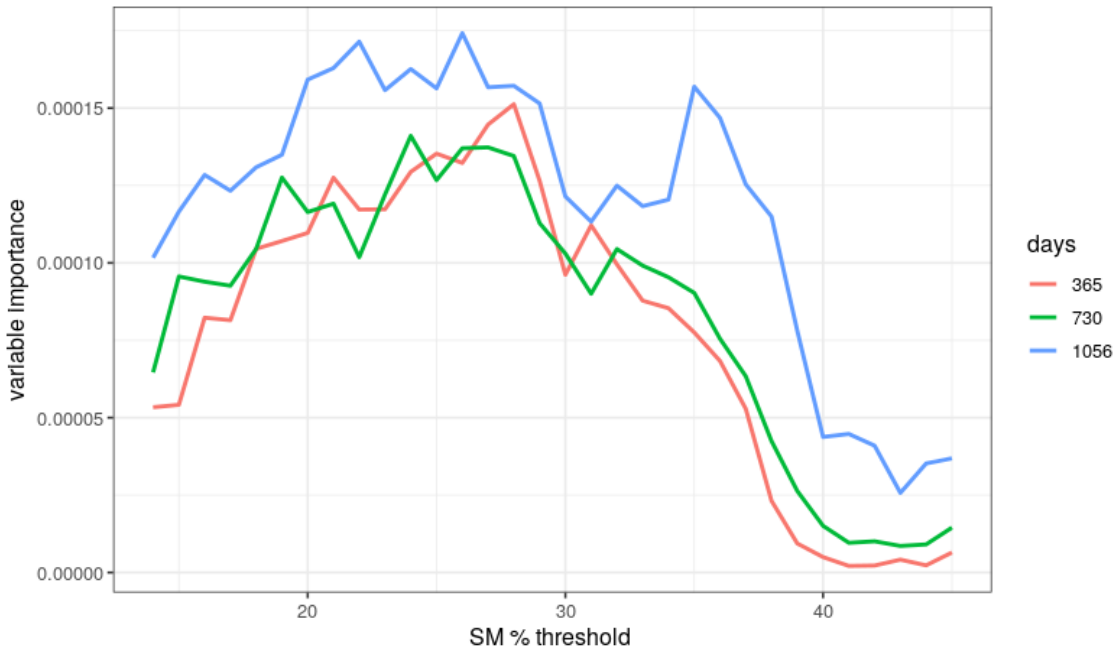

Figure 21: Variable importance of gradientForest analysis for soil moisture threshold data for forest sites. Line colors code the aggregated number of days before sampling.

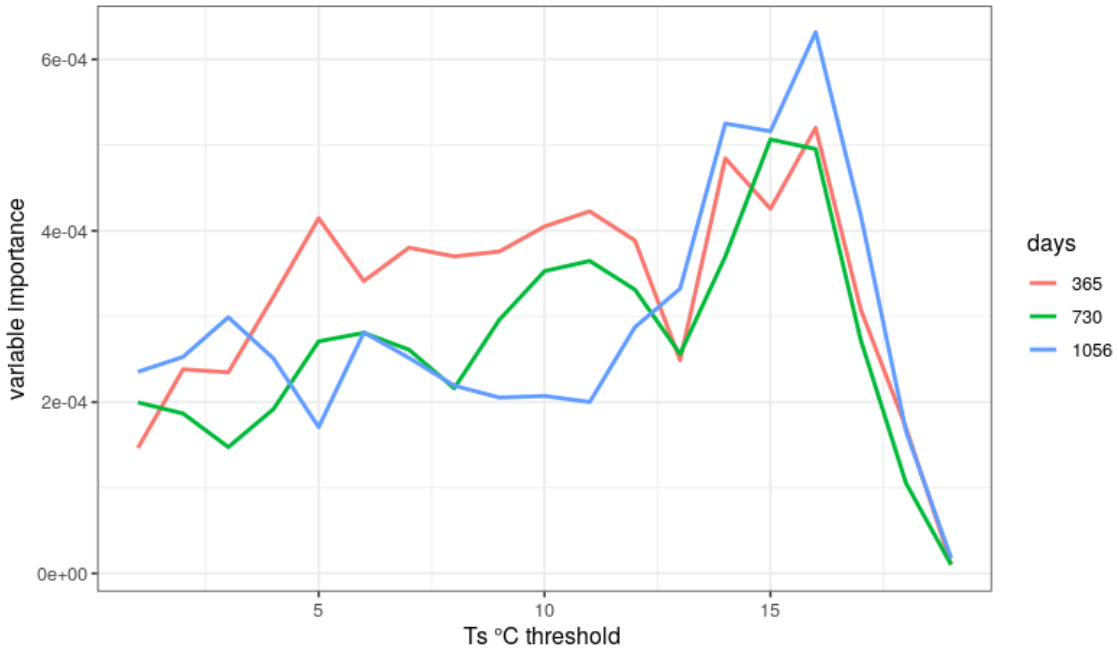

Figure 22: Variable importance of gradientForest analysis for soil temperature threshold data for forest sites. Line colors code the aggregated number of days before sampling.

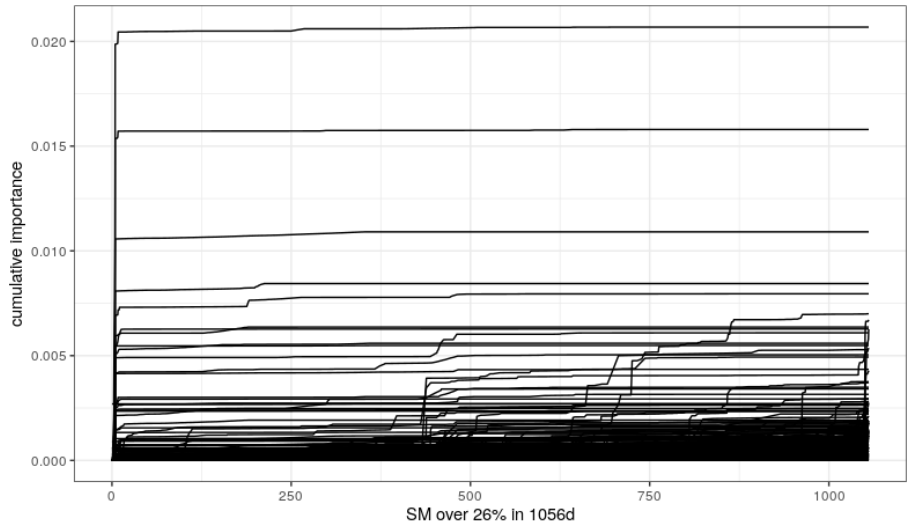

Figure 23: Species specific cumulative variable importance of all 740 SV having a $R^2 > 0.5$ from the Gradient Forest model of grassland sites of SM >26 %.

2.3.6 Ecological niche modelling

Community turnover in the GF models of soil bacteria was evaluated as the averaged turnover of all species with predictive power. However, for grassland and forest communities, only 740 and 550 SV with $R^2 > 0.5$ were retrieved respectively. For those SV, the individual turnover function for each predictor gradient clearly showed different clusters of bacteria which did not only react specifically to different regions on the predictor gradients, but differed in variable importance as well (compare Fig. 58 and 59). In a first attempt, trajectory clustering was applied to group those SV in different clusters for each predictor. The successful clustering implied that distinct groups of soil bacteria existed, sharing similar species turnover along predictor gradients and their respective profile in predictor importance. However, no quantitative niche clustering can be performed based on trajectory clustering, as only categorical groups could be retrieved for each predictor based on the clustering along the predictor gradient and cumulative variable importance.

Other modelling approaches for soil bacteria use optimal niche values to derive niches for soil bacteria (Sikorski et al., 2022). However, this approach lacks the consideration of bacterial tolerance, i.e. predictor importance, to a deviation from optimal niches values. Here, the combination of optimal niche value estimation and SV-specific predictor importance could lead to more realistic estimation of established soil bacterial niche space.
Hence, optimal niche values were estimated for the 10 most important predictors of each SV with $R^2 > 0.5$ from GF modelling of beta diversity analysis of grassland and forest communities. For 728 out of 740 SV for grassland and 550 SV for forest sites optimal niche values were retrieved. For 12 SV from grassland communities, no model to determine optimal niche values could be fitted. These SV were excluded from further analysis. Optimal niche values were determined fitting models of hierarchical logistic regression (Jansen and Oksanen, 2013). The optimal niche value of each predictor for each SV was added to the respective variable importance value from GF models. Together, clustering of optimal niche values and variable importance can be used to identify groups of SV sharing similar regions of multidimensional predictor space.

Multidimensional predictor space was clustered by density based spatial clustering with applications of noise (DBSCAN). PCoA ordination of grassland bacteria shows that bacteria harboured 6 groups (Fig. 24A). 143 SV could not be matched to any cluster. Fig. 24B shows the mean optimum niche value for all SV sharing the same cluster. Niches A and B differ mainly in wide coarse pores content. In comparison to the other niches, A and B contain clay rich sites, along with low coarse pore content. Niches D and E are quite similar, but differ mainly in C/N ratio. Niche F contained the lowest mean pH together with niche C, but both differ in organic C and oxalate extractable Fe content.
PCoA on euclidean distances between bacteria of forest sites revealed 3 niches (Fig. 25A).

18 SV could not be ascribed to a niche. Obviously, niche A was separated from niches B and C. The spatial overlap between B and C is valid, since 2-dimensional PCoA ordination was only used for visualization purposes and underlying niche clustering was performed on the full multidimensional distance matrix. As indicated by ordination, niches B and C were closely related in mean optimal niche value composition, differing only in carbon content of the O_e horizon (Fig. 25B). Considering the taxonomic composition of bacterial niches of grassland and forest soils, no taxonomic pattern occurred. Bacterial phyla were present in all groups and no taxonomic group below phylum level was exclusively bound to a certain niche.

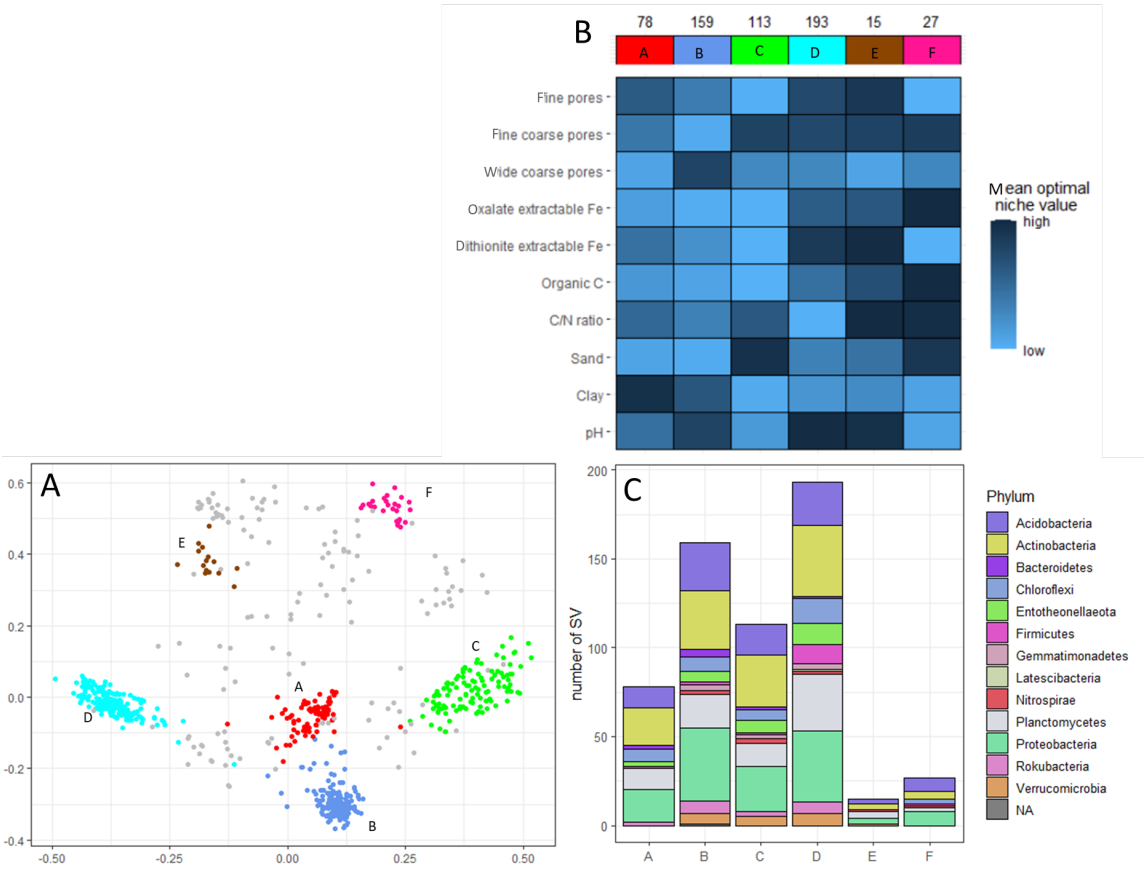

Figure 24: Niche modelling of grassland bacteria. (A) PCoA on euclidean distance matrix consisting of the added euclidean distances of optimal niche values and variable importance for the 10 most important predictors as retrieved by Gradient Forest. Each dot represents one of 728 SV. Different groups of soil bacteria identified by spatial clustering are coloured and labelled with capital letters. Gray points depict SV which could not be assigned to a specific cluster. (B) Scaled mean optimal niche values for each cluster of soil bacteria as identified by DBSCAN. The blue colour gradient in each tile depicts mean optimal niche value from all SV in a cluster. All predictor gradients are normalized to 0-1. Capital letters on top of coloured tiles identify the cluster of soil bacteria. Numbers on top represent the number of SV in each cluster. (C) Taxonomic composition of each cluster on phylum level. Y-axis depict the number of species in each taxonomic group.

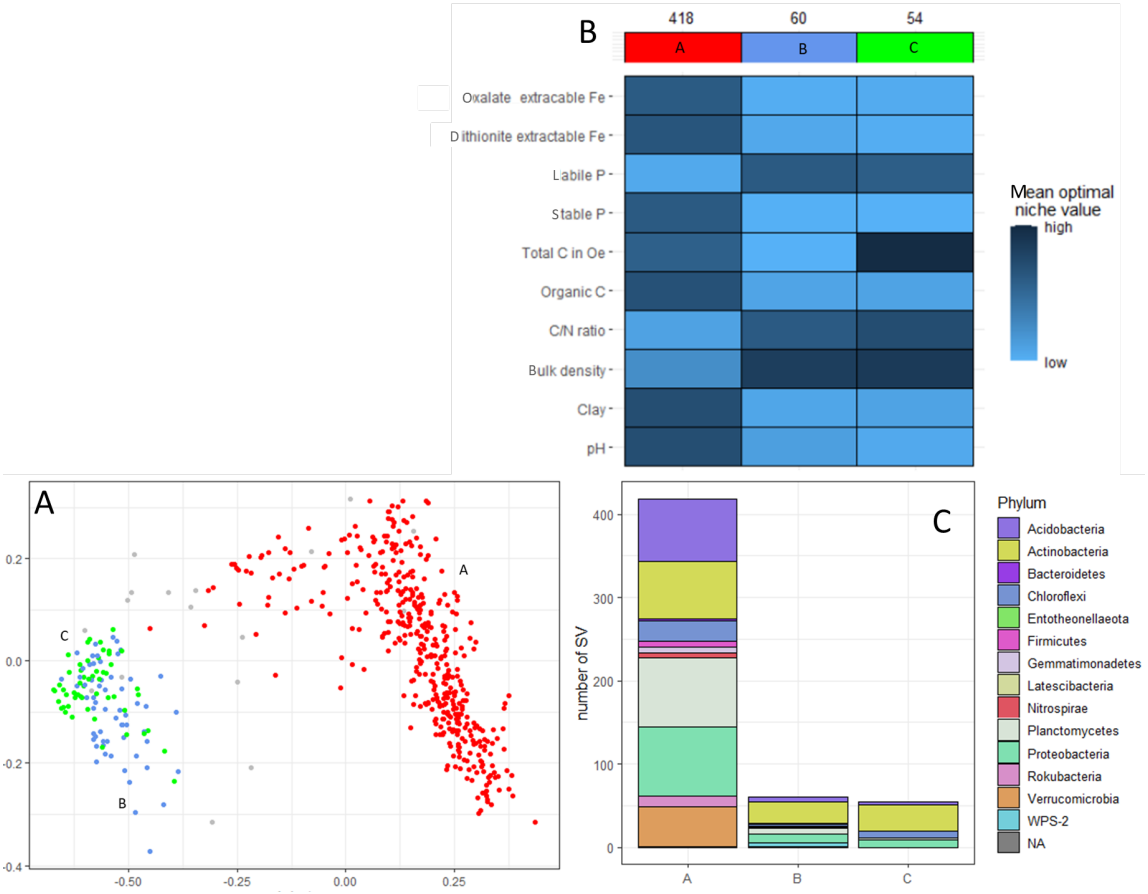

Figure 25: Niche modelling of forest bacteria. (A) PCoA on euclidean distance matrix consisting of the added euclidean distances of optimal niche values and variable importance for the 10 most important predictors as retrieved by Gradient Forest analysis. Each dot represents one of 550 SV. Different groups of soil bacteria identified by spatial clustering are coloured and labelled with capital letters. Gray points depict SV which could not be assigned to a specific cluster. (B) Scaled mean optimal niche values for each cluster of soil bacteria as identified by DBSCAN. The blue colour gradient in each tile depicts mean optimal niche value from all SV in a cluster. All predictor gradients are normalized to 0-1. Capital letters on top of coloured tiles identify the cluster of soil bacteria. Numbers on top represent the number of SV in each cluster. (C) Taxonomic composition of each cluster on phylum level. Y-axis depict the number of species in each taxonomic group.

2.4 Discussion

Climate predictors are amongst the most important predictors of soil bacterial diversity on a global scale (Bickel et al., 2019; Delgado-Baquerizo and Eldridge, 2019; Zheng et al., 2020). However, insights from large scale studies are difficult to translate to regional scales (Mod et al., 2021). Furthermore, little information is available how bulk soil bacterial communities react to climate and its longitudinal effects over time. This is important if considered that climate is already changing in Germany and ecosystem functions become affected by summer drought.

In more detail, we are lacking knowledge how and whether alpha diversity, the abundance of rare and abundant soil bacteria, react to climate and how diversity develops over time under climate change. Also, it is not known how potential changes translate to shifts in beta diversity, the diversity between samples, represented by community turnover. To shed more light on the situation in Germany, high throughput sequencing data based on 16S rRNA transcripts were examined over 3 time points of bulk soil bacterial communities spanning 6 years across 296 forest and grassland sites, distributed across 3 regions in Germany,

2.4.1 The abundance of rare and abundant soil bacteria declined over time

It is well known from literature that alpha diversity is highly sensitive to temporal variation (Lauber et al., 2013). Moreover, 16S rRNA transcript based communities, as used in this study, were found to be sensitive to the sampling timepoint (Herzog et al., 2015), suggesting a vulnerability of bacterial alpha diversity to short-term changes and temporal autocorrelation. Therefore the major question in this study was whether and how bacterial alpha diversity changed over time. Alpha diversity analysis revealed that rare and abundant soil bacteria, represented by richness and Simpson diversity indices declined significantly over time (Fig. 9 - 10). For Simpson this effect was even more important than pH. However, pH represented the most important predictor for richness.

The observed decline was possibly a consequence of stochasticity and short-term variation. However, the consistent decline in alpha diversity also found for highly abundant species is of special interest, since mostly rare bacteria were shown to account for the majority of temporal variability in soils (Shade and Gilbert, 2015; Zhou et al., 2020). Besides, the importance of sampling year as predictor for the Simpson index was twice as high compared to richness and the importance of pH. This shows that the observed decline in alpha diversity is affecting all soil bacteria independent of abundance. Certainly, a possible caveat could be that intra annual changes influence this result, since results from a grassland site of the Biodiversity Exploratories revealed changes in alpha diversity on the plot scale over the course of a year (Richter-Heitmann et al., 2020). However, I consider the decline in

alpha diversity to be robust and no consequence of short-term variation. This is supported by the similar sampling time points in May for all years, the large sample size of 888 samples per model, the distribution of sites in three regions across Germany, the random effect structure accounting for plot specific effects and the fact that available predictors describing short-term soil moisture and temperature conditions during sampling were not influential.

Considering the comparability of data between years, another point to discuss is the change in sequencing platform from Hiseq 2500 to NextSeq 500 from 2011 to 2014 and 2017. This means a change from 4-colour to 2-colour chemistry which might affect base calling (DeKayne et al., 2021) and consequently affect species detection. However, sample ordination did not reveal a systematic bias (Fig. 8). More importantly, it is well known that alpha diversity and especially richness estimation is primarily affected by sequencing depth (Kleine Bardenhorst et al., 2022; Sanchez-Cid et al., 2022). Here, no systematic difference was observed in sequencing depth between years (Fig. 5 - 6). Besides, sequencing data normalization by sample completeness rather than size was applied to make alpha diversity comparisons between samples and sampling years more reliable and preventing loss of sequencing information (McMurdie, 2018; Chao and Jost, 2012; Weiss et al., 2017). Hence, it can be ruled out that the decline in richness and Simpson indesx, is a consequence of sequencing platform or systematic differences in sequencing depth.

Last but not least, spatial variation may confound temporal variation (Hermans et al., 2020). Therefore, spatial autocorrelation between sites was accounted for in GAM model fitting. Furthermore, site specific spatial variability was accounted for sample homogenization, mixing multiple sub-samples per site. Notably, it was observed during model fitting that the correction for spatial autocorrelation, reduced variable importance and affected the fit of environmental predictors (data not shown). Again, this was regarded as an effect of regional/spatial clustering and multicollinearity of environmental predictors in this study (Fig. 3, 4, 11).

In conclusion, the observed decline in alpha diversity over time is considered as a reliable signal inherent to this dataset affecting all, rare and abundant, soil bacteria. However, a continuation of this analysis including data from upcoming sampling years has to investigate whether the observed decline is still an active process and part of long-term variations >6 years of alpha diversity in temperate grassland and forest ecosystems in Germany.

2.4.2 Climatic controls of alpha diversity

Bacterial alpha diversity declined over time in Germany. Time itself, except for stochastic effects, is no causal driver for this temporal pattern. I hypothesized that climate might explain parts of this pattern, in case those predictors are temporally correlated. This would represent the first evidence of climate change already affecting soil bacterial diversity in Germany. Furthermore, I expected legacy effects of climate predictors influencing bacterial alpha diversity, describing a delay effect of climate on bacterial diversity.

Climate predictors engineered from soil moisture and soil temperature data, measured continuously since 2008, were more important than the effect of soil texture on rare and abundant soil bacteria, confirming their significance for bacterial alpha diversity. This corresponds to earlier results of Delgado-Baquerizo and Eldridge (2019), where climatic predictors were more important than texture in temperate grasslands globally.

The selection of climatic threshold data during model fitting over actual SM and Ts conditions measured in soil while sampling emphasized the delayed impact of climate on alpha diversity. This confirmed the anticipated legacy effects of climate on alpha diversity. Climatic predictors summarized over 1056 d before the sampling date resulted in higher reductions in AIC than aggregations over 1 or 2 years, emphasizing the rather long time lag of climate effects needed for alpha diversity to react. This confirmed results for soil bacteria from alpine soils using similar temperature threshold data (Frindte et al., 2019). Legacy effects of soil bacteria to climate were confirmed frequently (Waring and Hawkes, 2018; Banerjee et al., 2016; Ladau et al., 2018). Unfortunately, it was not possible to aggregate climate over longer preceding time periods, due to the prevalence of discontinuous data and missing values in climate data before 2008. Although it was shown that historic climate outperformed the influence of more recent climate data on soil bacteria (Ladau et al., 2018).

Generally, the bigger effects of climate predictors aggregated over 1-3 years chosen over shorter aggregations from weeks to months, depicted a rather temporally stable bacterial alpha diversity, unaffected by short-term changes in soil moisture and soil temperature. Besides, extreme climatic event represented by maxima or minima of soil temperature and moisture data were not selected at all during model fitting. This is interesting, since bacterial alpha diversity and especially richness are considered as rather variable over time and prone to short-term changes induced by extreme events (Richter-Heitmann et al., 2020).

According to variable importance, Ts derived predictors were equally important to SM for richness. However, for Simpson only effects of SM were selected. Initially selected climatic predictors in separate GAMs for grassland or forest sites were recovered in the models fit on all sites as well (Fig. 54, 55). This emphasized the validity of the final GAM models for alpha diversity fitted on all sites, independent of land use type. For Simpson, no Ts derived

predictors were selected. This is in line with Richter-Heitmann et al. (2020) identifying SM as being more important for abundant species and implies more resistance of abundant bacteria to trends in soil temperature.
Apart from the effect on alpha diversity, the engineered climate predictors revealed relevant Ts and SM thresholds, affecting bacteria in temperate soils. Interestingly, for both alpha diversity indices, the effect of soil moisture range over 1056 d could be confirmed. Furthermore, high soil moisture >42 % over 1056 d was more decisive for Simpson compared to SM >14 % over 1056 d for richness.
Considering the soil temperature data, richness declined with an increasing number of days > 4 °C and > 15 °C, representing an increase in soil temperature. This decline was no indirect effect of SM, since Ts threshold data were not well correlated with SM (Fig. 49). Possibly, the diversity in a sample is reduced with increasing Ts, as rare SV might be outperformed by abundant species being more resilient to rising soil temperatures, since Simpson did not respond significantly to Ts. From literature, the effects of climate and climate change on bacterial diversity are not consistent. On the one hand, it was shown recently that experimental warming decreased richness in grassland over 7 years (Wu et al., 2022). On the other hand, little to no effects were found of short and long-term manipulations of climate in forest and grassland communities (Cruz-Martínez et al., 2009; DeAngelis et al., 2015; Engelhardt et al., 2018; Sünnemann et al., 2021; Andrade-Linares et al., 2021). Comparing the effects of warming and moisture regimes, soil moisture is more important for alpha diversity than warming (Zhang et al., 2013; Wu et al., 2022).

Moreover, data exploration and correlation revealed temporal correlation for Ts and to lesser extend for SM derived predictors (Fig. 4, 11). However, time itself is no causal driver for bacterial diversity. Deviance explained increased for the selected climate predictors if year was manually excluded from the GAM models, indicating data variation caused by change over time can be explained by climatic predictors. Hence, climate change causing higher soil temperature and reduced soil moisture over time is already shaping bacterial alpha diversity in forest and grassland soils in Germany. It was recently shown that soil warming is an ongoing process in Germany (Dorau et al., 2022). Therefore, it is of highest importance to pursue the presented time series to see how climate change continues to affect soil bacterial communities in Germany. This is of outstanding importance, since other studies found no effects of soil warming on soil bacteria in Germany (Andrade-Linares et al., 2021; Sünnemann et al., 2021).
Moreover, the multicollinearity generally observed for climate data with region and time led to the exclusion of the promising predictor group "sum of rain days" from the analysis due to severe correlation with soil texture. Sum rain days over 1056 days was initially selected during model selection for alpha diversity. However, the incorporation of sum rain

days together with soil texture by GAM forward selection, led to the occurrence of Simpson paradox during model fitting (Appleton et al., 1996; Armitage and Jones, 2019), reversing the fit/effect of sum rain days over 1056 d (data not shown), compared to a model fitted excluding soil texture.

2.4.3 Edaphic predictors and alpha diversity

In addition to the temporal decline of alpha diversity caused by climate, pH remained the most important predictor for richness and the second most important predictor for Simpson. However, the link of richness and Simpson to pH differed. Rare species showed an unimodal pattern with an optimum around pH 6.5. This pattern is described extensively in literature (Fierer and Jackson, 2006; Terrat et al., 2017; Bickel et al., 2019; Delgado-Baquerizo and Eldridge, 2019). In contrast, highly abundant species showed a linear increase from pH 4.5 to pH 7. This results in communities where abundant bacteria gain in importance for soil with pH > 6.5.
Moreover, finer soil texture led to an increase in alpha diversity for rare and abundant species. This is widely confirmed in literature for richness (Kandeler et al., 2000; Zhang et al., 2007; Hemkemeyer et al., 2018) and Simpson (Zheng et al., 2021). However, also contradicting results are available (Obayomi et al., 2021). Besides, Simpson decreased with increasing bulk density. Together these effects hinted towards the importance of soil structure and aggregation for alpha diversity, since bacteria were shown to have preferences for particle size classes (Hemkemeyer et al., 2018). Overall, the effects of pH and soil texture are well known to shape bacterial alpha diversity in soil (Plassart et al., 2019; Bickel et al., 2019).
A factor not frequently encountered in soil biodiversity research is the effect of soil Fe content. Here, richness was susceptible to low soil Fe contents and declined $< 5\,\mathrm{g\,kg^{-1}}$. Possibly, low Fe contents affected rare species in soils by low solubility of Fe-oxides and hydroxides and a potential association with micronutrients, since Simpson did not react to low Fe contents. It might be that some rare species are less competitive for Fe or micronutrients and therefore superseded.

Land use type and intensity, as well as the derivatives mowing, grazing and fertilization, did not significantly affect alpha diversity. Preliminary testing revealed that land use type did not affect richness except for Schorfheide. For Simpson diversity, significant effects occurred in all regions (Fig. 7). However, during GAM modelling, these effects were not selected, which led to the conclusion that land use effects could be allocated to edaphic and climatic predictors. Moreover, land use intensity and its components did not contribute significantly to bacterial alpha diversity in this study. This is in line with literature showing

weak or neutral effects of LUI on bacterial diversity (Kaiser et al., 2016; Gossner et al., 2016; Boeddinghaus et al., 2019).

2.4.4 Community turnover is resistant to climate and climate change

Beta diversity represented by community turnover was analysed to see how strong environmental predictors shape bacterial diversity. Therefore, community turnover was modelled separately for grassland and forest communities by generalized dissimilarity modelling (GDM) and gradientForest (GF), because exclusively available predictors for either land use type were among the selected predictors. In contrast to alpha diversity analysis, temporal effects were significant but only of minor importance for grassland communities. Forest communities were unaffected by time. This supports the finding that beta diversity is generally less susceptible to temporal variation than alpha diversity (Lauber et al., 2013; Shade and Gilbert, 2015; Hermans et al., 2020). Moreover, this confirms that bacterial diversity is more stable in forest soils, compared to grassland (Hermans et al., 2020).
Furthermore, significant climate predictors were only confirmed for beta diversity of forest communities revealed by GDM, selecting max. SM over 365 d and Ts >15 °C over 365 d. Due to its non-parametric nature and general robustness towards co-correlation (Ellis et al., 2012), GF is predestined for the analysis of climatic threshold data. In statistical modelling, a priori exclusion of predictors with unknown behaviour towards beta diversity is difficult to justify, as it is unknown which predictor is likely affecting community turnover. Moreover, skewed predictor distribution and outliers do not affect tree based models. Therefore, all climatic predictors were fitted in GF models.
Although of low relative variable importance, clear pattern for SM and Ts threshold data were retrieved. This facilitated the identification of relevant aspects of SM and Ts to soil bacterial community turnover. The independent identification of soil temperatures around 4 and 15 °C by independent modelling approaches for alpha and beta diversity supports their biological relevance for soil bacterial diversity. Although the effects of climate predictors on beta diversity were low compared to alpha diversity, beta diversity is reactive to climate change if the trends are of sufficient duration and strength (Cruz-Martínez et al., 2009; Zhou et al., 2020), emphasizing robustness of beta diversity towards climate. However, Guo et al. (2018) confirmed a significant effect of warming on beta diversity.
It can be speculated that the temporal effect found for grassland communities and the climatic effects confirmed for the forest communities, represented the same inherent biological signal, since Ts threshold data contained temporal correlation. However, the relative importance of temporal and climatic effects were low for beta diversity compared to alpha diversity, suggesting that beta diversity is more stable over time (Armstrong et al., 2016). The lack of stronger temporal and climatic effects as seen for alpha diversity analysis might

be a consequence of the applied sparsity reduction method, since rare species were not included in beta diversity analysis. However, a comparison of beta diversity ordination of the total community containing all SV versus the filtered community excluding SV <0.1 % relative abundance in any sample showed no significant differences by procrustes rotation (Fig. 8C). Generally, the removal of rare species is a common practice in soil bacterial community analysis (Szoboszlay and Tebbe, 2021). In addition, it was shown that the removal of conditionally rare taxa does not alter beta diversity pattern using Bray-Curtis dissimilarity (Kaminsky and Morales, 2018).

The analysis of beta diversity by GDM and GF modelling confirmed pH as the most important predictor. Besides, soil texture, soil nutrient concentration, sesquioxide concentration and soil structure caused community turnover in grassland and forest communities. Extractable Al and Fe represented sesquioxide concentration, i.e. amorph Fe and Al oxides. Al is toxic to plants and bacteria, especially for pH < 4.5 (Wood, 1995) and affects beta diversity (Pan et al., 2014). Al and Fe were selected by GDM and GF respectively.

Regarding soil mineral P content, community turnover of forest sites was affected by low P content $< 200\,\mathrm{mg\,kg^{-1}}$. Below that threshold soil P stocks and P solubilization are possibly getting limited. For the grassland communities, the ratio of fungal/bacterial PLFA was selected as the second most important predictor causing community turnover. Unfortunately, fungal PLFAs were not available for forest sites, although the impact of fungal communities on soil microbial communities is well known, especially at low soil pH (de Vries et al., 2006; Rousk et al., 2010) and more recalcitrant C sources (de Boer et al., 2005).

Moreover, GF models complemented GDM analysis by adding the effect of organic C on community turnover for grassland and forest models. The effect of organic C is not restricted to carbon as this predictor was highly correlated with soil N and S contents and therefore representing the effect of soil nutrient concentration on community turnover. For forest models, the influence of soil nutrient supply was augmented by total C in organic litter horizon, identified in both GDM and GF. Since bacterial communities of aboveground horizons were not sampled, the effects found on community turnover especially for high C contents > 400 g $\mathrm{kg^{-1}}$ hint towards an effect of dissolved organic carbon on bacterial communities (Judd et al., 2006; Zhang et al., 2022). Soil organic carbon plays an important role for bacterial diversity (Delgado-Baquerizo et al., 2016), even more if carbon quality is considered (Neumann et al., 2013; Szoboszlay and Tebbe, 2021).

Turning now to soil structure, bulk density, soil pore distribution and porosity were selected for forest and grassland models of GDM and GF representing the soil pore distribution. These different predictors can be summarized representing a common driver, since porosity and pore distribution data were not available for forest sites, but were highly correlated with bulk density (Fig. 51).

Moreover, the substrate quality of soil organic matter as C/N ratio was identified causing community turnover, although its variable importance was low. The effect of substrate quality on beta diversity was already confirmed for the Biodiversity Exploratories (Kaiser et al., 2016).
In summary, community turnover was more stable over time than alpha diversity analysis. However, the selection of almost identical temperature thresholds for alpha and beta diversity for the forest communities supported a consistent effect of these predictors. It has to be considered though, that these effects had only little impact on community turnover.

2.4.5 Statistical niche modelling based on gradientForest analysis

Modelling the specific turnover of each SV by GF revealed that clusters of SV occurred along predictor gradients, showing similar pattern of species turnover (Fig- 58-59). Therefore, a modelling approach combining optimal niche values with SV specific predictor importance was performed to see if potential niches for soil bacteria can be derived from this species clustering approach.
For several reasons, this approach has the potential to be more precise than niche modelling approaches lacking SV specific modelling, especially if a large number of predictors is available. First, based on the GF results, only SV were selected for which predictive power was provided. Second, the availability of relative variable importance incorporated the predictor importance profile for each SV in the modelling. However, an individual weighting of each predictor by variable importance is not possible. Hence, the applied method represents a niche modelling based on the optimal niche value profile and variable importance profile rather than a niche modelling based on optimal niche values weighted by variable importance.
Generally, the addition of SV specific variable importance to SV specific optimal niche values led to a less clear separation between clusters than grouping solely based on optimal niche values (Fig. 60). Including variable importance affected the differentiation between clusters of bacteria more in forest communities. Grassland bacteria maintained their cluster structure based on optimal niche values, although the effect of variable importance was included. The variable importance structure represented mainly the influence of pH for most SV in form of a one-dimensional stretch, as seen in Fig. 60 (panels B and E in supplementary materials). The 6 grassland cluster differed clearly in their average optimal niche value composition. However, cluster D and E differed only in C/N ratio. This implied quite specific adaptations of SV to substrate quality. Forest clusters revealed 2 groups with contrasting conditions across all predictors, separating A from B and C. Furthermore, niche B and C differed only in total C in O_e. In contrast to the optimal niche values, taxonomy revealed no systematic enrichment of a taxonomic group to a certain niche. All phyla and

taxonomic groups below phylum level were evenly distributed between bacterial niches of grassland and forest soils. This suggests, that adaptations to specific soil conditions defined by different regions on gradients of environmental predictors do not lead to phylogenetic conserved adaptations. The lack of deeper niche differentiation is supported by GDM and GF beta diversity modelling which revealed the outstanding dominance of pH shaping community turnover. Moreover, the importance of pH was already visible as an arch effect in beta diversity ordination (Fig. 8). The lack of a better differentiation in taxonomy and consequently phylogeny between niches might be caused by the negligence of spatial constraints in soil. Hemkemeyer et al. (2018) and Szoboszlay and Tebbe (2021) showed that communities differ by particle size fractions in soil. Niche modelling on the pore scale differentiating between communities and soil conditions might reveal more precise aspects of bacterial coexistence in soil. However, differentiating a large number of communities on the pore scale is unfeasible for a large scale study. Even more, the differentiated acquisition of soil moisture data on the pore scale is complicated although a joint analysis would be of great interest to estimate how changes in soil moisture regimes due to particle sizes and water tension shape bacterial communities in conjunction with other edaphic predictors.

2.5 Conclusion

The analysis of alpha and beta diversity revealed independently similar drivers of soil temperature shaping soil bacteria. Apparently, soil temperature > 4°C and > 15°C were most important for richness and community turnover of soil bacteria of grassland and forest communities. However, alpha diversity was predominantly shaped by temporal decline over time and pH. Rare as well as abundant soil bacteria were found to decline significantly over time. For richness this can be attributed to an increase in soil temperature over time demonstrating that climate change already influences soil bacterial communities across Germany. The decline of highly abundant bacterial species is surprising, since abundant species are considered to be more stable over time. Therefore, the relevance of this decline over time is supported even further.

In contrast, community turnover was rather stable over time, although grassland communities were affected significantly. The dominance of pH on bacterial community turnover was more important as revealed by GDM and GF analysis. Generally, GF is a potent method providing comparable results to GDM analysis while allowing for SV specific analysis of species turnover along environmental gradients. This revealed niche differentiation of soil bacterial communities in German grassland and forest soils.

Furthermore, a higher variable importance of climatic predictors summarized over 1-3 years compared to short-term temporal aggregations from several days to weeks confirmed legacy effects of soil temperature and soil moisture on bacterial diversity and emphasized the temporal stability of bacterial alpha and beta diversity in soil. However, legacy effects of land use intensity were not observed for alpha and beta diversity.

Summing up, the presented results from the Biodiversity Exploratories revealed insight in climate change already affecting soil bacterial communities in Germany. It is therefore of great importance to further pursue this development of soil bacterial communities in response to climate change in future studies.

3 Bacterial life-style not fertilization shapes soil bacteria communities from a long-term agricultural fertilization experiment

3.1 Starting point of the study

Bacteria in the bioshpere live predominantly surface-associated (Costerton et al., 1987; Tuson and Weibel, 2013). This is true especially for soils (Flemming and Wuertz, 2019). Soils offer a high surface area, but only a small proportion of it is covered by bacteria (Nunan et al., 2003). The distance between these bacteria is in the range of tens of µm (Raynaud and Nunan, 2014). However, bacteria are not homogeneously distributed in the soil matrix and occur predominantly in patches (Nunan et al., 2003).

The attachment of bacteria to soil particles or any other surface is driven by electrostatic interactions (Tuson and Weibel, 2013). Generally, the attachment process is described as a two-stage model with a reversible and rapid adhesion based on hydrophobic and electrostatic interaction (Huysman and Verstraete, 1993b,a; Schwarz et al., 2017), followed by an irreversible stage which takes several hours and is based on van der Waals interactions between the hydrophobic region of the outer cell wall and the surface (Tuson and Weibel, 2013). Moreover, by exuding extracellular polysaccharides (EPS) , bacteria are capable of preconditioning a surface for cell attachment (Kögel-Knabner et al., 2008). EPS are chemically diverse and contain a large variety of carbohydrates as well as proteins or DNA (Whitfield, 1988). A vast community of bacteria can be involved which live on any solid-liquid interface in a matrix of EPS. Thereby, also bacteria are involved which are not attached to the surface directly, but are contained in the EPS matrix. A diverse community of bacteria in a matrix of EPS attached to a surface is summarized as a biofilm (Flemming and Wingender, 2010). The EPS matrix makes up the majority of biofilm drymass (Cai et al., 2019). Biofilms colonize preferably carbon-rich material in soil (Nunan, 2017). Moreover, a biofilm is not static, but follows a lifecycle consisting of attachment, colonization, development, maturation and cell dispersal (Cai et al., 2019). This indicates that biofilm formation depends on environmental conditions and the source and composition of the available surface.

Flemming and Wuertz (2019) claimed that biofilms harbour most active bacteria. This is supported by Wu et al. (2019), who detected 23 times more active bacteria in soil biofilms by substrate-induced respiration. Seneviratne and Jayasinghearachchi (2005) found biofilms to improve soil nutrient mineralization and higher enzyme activity. Although most bacteria in soil are described as dormant with only a small proportion being metabolically active (Blagodatskaya and Kuzyakov, 2013). This is in line with the hypothesis that the surface-associated lifestyle in a biofilm is a mechanism for low nutrient environments and enables

growth at much lower nutrient concentration than in liquid culture (Kögel-Knabner et al., 2008; Flemming and Wuertz, 2019). This was confirmed experimentally several times (Andersson and Nilsson, 2001; Huang et al., 2015). However, also contradicting results were published. Wu et al. (2019) found that high-nutrient input induced biofilm formation and low nutrients favoured a free-living lifestyle.

At the time of writing, there is no conclusion reached, if substrate availability favours a free-living or surface-associated bacterial lifestyle. These inconsistent results might be a consequence of soil types or texture and its chemical composition, since soil surfaces can provide decomposable substrates and nutrients or inhibit biological degradation by spatial and physical effects (Chenu et al., 2002). For example, soil-pores < 0.8 µm were found not being colonized by bacteria at all (Ranjard and Richaume, 2001), which could protect substrates spatially from microbial decomposition. Moreover, it was demonstrated that soil communities differ on the pore scale (Ruamps et al., 2011) with different phylogenetic groups having preferences for different particle size fractions (Hemkemeyer et al., 2018; Ranjard and Richaume, 2001). In contrast to a surface-associated lifestyle, a free-living or planktonic lifestyle can either be passive, enforced by biofilm dispersal and the consequent release of cells into the soil pore space (Flemming and Wingender, 2010) or by any form of motility providing the capabilities for targeted movement towards higher nutrient concentrations in the soil water or for dispersal.

Generally, there is no information available how and if fertilization of arable soils affects soil bacterial life-style. So far information is exclusively available for bulk soil and rhizosphere communities. It was shown that organic fertilizations increases alpha diversity in arable soils (Wei et al., 2008; Hartmann et al., 2015; Francioli et al., 2016; van der Bom et al., 2018). Moreover, mineral fertilization is reducing alpha diversity indirectly by soil acidification (Zeng et al., 2016; Li et al., 2016; Dai et al., 2018), however this effect was attributed to secondary soil acidification. Moreover, mineral N was found to reduce microbial biomass (Wallenstein et al., 2006). However, also neutral (Ge et al., 2008; Ogilvie et al., 2008; Eo and Park, 2016; Duan et al., 2022) and concentration dependent effects were reported (Sun et al., 2019). Hartmann et al. (2015) assumes that these inconsistencies in results are dedicated to differences in experimental systems and management.

The physiological mechanisms leading to the formation of biofilms in soils are well studied, but little is known about the bacterial communities living predominantly in biofilms or having a surface-associated lifestyle. Even less is know about bacterial species living temporarily or exclusively in the soil pore space. So far, only one study is available comparing surface-associated and free-living communities in soil. Bystrianský et al. (2019) found species being significantly enriched in either of the communities, with the genera *Pseudomonas*, *Rhizobium*, *Arthrobacter* and *Bosea* being significantly enriched in the surface-associated community.

In conclusion, a coherent analysis of the effects of nutrient availability on bacterial lifestyle and activity is still missing. In this study, bacterial lifestyle is defined as the contemplation of surface-associated versus a free-living bacteria in the soil pore space.

Hence, the aim of this study was to analyse how the bacterial community composition, being either free-living or surface-associated is influenced by nutrient availability, potential metabolic activity and life-style. I hypothesise, first, that for each fertilization treatment the free-living community is less diverse than the surface-associated community. However, I don't expect a significant effect of fertilization treatment on alpha diversity.

Second, the free-living and surface-associated communities differ in community composition for each fertilization treatment.

Third, I expect the differences in community composition to be more prominent for the potentially active communities for different fertilization treatments and bacterial life-style.

Fourth, the proportion of active bacteria should increase with fertilization and be higher in the respective surface-associated community due to the better nutrient supply. In contrast, the free-living community should be depleted in active bacteria due to starvation.

3.2 Material and Methods

3.2.1 Study design

Initially, it was planned to compare different soil types from the Biodiversity Exploratory network (Fischer et al., 2010), spanning a wide range of edaphic properties. Unfortunately, due to the Corona restrictions in summer 2020, the soil sampling campaign was cancelled. As a replacement, soil bacterial communities of different fertilization treatments originating from the same soil type of an agricultural long-term fertilization experiment were compared to analyse the influence of soil nutrient availability on the free-living and surface-associated soil bacterial community composition. In the laboratory, a cell extraction experiment from soil was developed to separate soil bacteria of different life-style. The extracted soil bacterial communities were analysed using Next Generation Sequencing of the variable V3/V4-region of the bacterial 16S rRNA gene copies and transcripts. It is unknown from literature to which extend the free-living and surface associated bacteria are shaped by metabolic active bacteria. To further distinguish between potentially active and non-active bacteria, a novel approach based on the 16S rRNA:rDNA ratio was employed. A conceptual overview of the experimental design is depicted in figure 26.

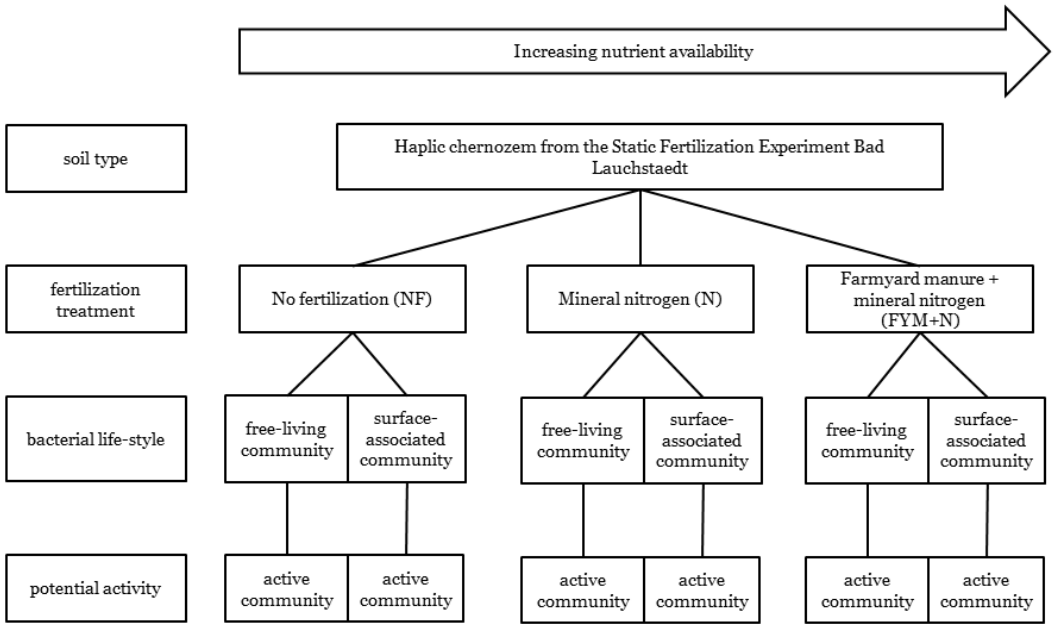

Figure 26: Experimental design of the bacterial community extraction experiment. The active community is defined as the rRNA:rDNA ratio for each bacterium, passing a dynamic threshold.

3.2.2 Soil sampling

Soil samples were taken on 30th of June 2020 from the Static Fertilization Experiment in Bad Lauchstaedt. This long-term fertilization experiment was initiated in 1902 (Köppen and Eich, 1991) on a haplic chernozem established on loess (Francioli et al., 2016), which is one of the most fertile soil types worldwide. The soil texture consists of 21.0 % clay, 67.8 % silt and 11.2 % sand. A comprehensive description of the site and the experimental design of the fertilization experiment is described in (Blair et al., 2006). Three fertilization treatments were compared, spanning the maximum available gradient in carbon and nitrogen content. The zero control plot, receiving no fertilization (NF) except liming since 1902, was compared to treatments solely fertilized with mineral nitrogen (N) and farmyard manure + mineral nitrogen (FYM+N). The respective soil nutrient concentrations are given in Table 2. FYM+N was fertilized annually with mineral fertilizer (80 kg N ha^{-1}) and every second year with organic fertilizer (30 t ha^{-1}). The last organic fertilization was applied in

2019. Treatment N received 100 kg N ha^{-1} annually. All investigated experimental plots had similar pH, avoiding a pH gradient between treatments, since pH is well known to have a major influence on soil bacterial community composition (Thompson et al., 2017; Szoboszlay et al., 2017; Kaiser et al., 2016).

Table 2: Soil properties of the investigated fertilizations treatments. NF stands for unfertilized control, N for mineral nitrogen fertilization and FYM+N for farmyard manure + mineral nitrogen. Organic carbon content (C_{org}), total nitrogen content N_t(%), organic carbon:total nitrogen ratio (C/N), available soil phosphorus (av. P), available soil potassium (av. K), pH and soil moisture (%, w/w). Data provided by Dr. Thomas Reitz (UFZ Leipzig, personal communication).

	C_{org}(%)	N_t(%)	C/N	av. P (mg/kg)	av. K (mg/kg)	pH	soil moisture
NF	1.56	0.13	11.6	42.93	44.57	7.54	5.22
N	1.74	0.16	11.26	19.37	57.36	7.46	4.64
FYM+N	2.35	0.21	11.0	239.9	238.9	7.40	4.35

For each fertilization treatment, 7 subsamples from 2-10 cm depth were randomly taken between rows of winter wheat with a soil corer. The soil from 0-2 cm could not be recovered and was lost during drilling with the soil corer, due to low soil water content. All 7 subsamples were merged and homogenized by 2 mm sieving and roots and stones were removed in the field. From the homogenized samples, 3 technical replicates of 3 g of soil per fertilization treatment were collected in 50 ml falcon tubes and fixed immediately in 4.5 ml of a mixture of 50 % (v/v) ethanol (EtOH) and phosphate buffered saline (PBS) of pH 7.3 and stored at 4 °C in the dark until cell extraction. 50 % EtOH was used as a fixative to stop any metabolic activity, since it is precipitating and not cross-linking proteins. Although fixation with 1 % paraformaldehyde, a cross-linking fixation agent, improved extraction efficiency of bacteria from soil (Eichorst et al., 2015), cross-linking fixation reagents are know to interfere with sequencing (Clingenpeel et al., 2014).
Remaining soil was stored at 4 °C in the dark in polyethylen bags for cell counting, determination of gravimetric soil water content and determination of cell extraction efficiency. The processing time per treatment from soil sampling to storage of sample replicates took 1 h for each fertilization treatment.

3.2.3 Separation of free-living and surface-associated bacteria by Nycodenz extraction

Before extraction, soil samples had been stored at 4 °C in 4.5 ml 50 % (v/v) EtOH and PBS solution in 50 ml falcon tubes. To separate the free-living bacteria in the soil suspension from surface-associated bacteria and soil particles, extractions were performed in 50 % (v/v)- EtOH/PBS solution followed by a Nycodenz density centrifugation. Evaluating the

cell extraction efficiency for the free-living community, harvesting a sufficient amount of bacteria for nucleic acid extraction and library preparation, 1 versus 5 consecutive extractions were compared

Subsequently, the surface-associated bacteria were separated from soil particles by means of cell extraction in a detergent solution containing 50 % (v/v) EtOH, 50 mM sodiumpyrophoshate, 10 mM EDTA and 0.1 % Tween 80 dissolved in PBS of pH 7.3.

Besides the extraction solution, three 8 mm and three 4 mm of heat-sterilized glass beads were added to each sample and vortexed vertically for 1 min at full force on a VortexGenie2000. Next, the tubes were vortexed for 15 min in horizontal position at half force followed by a centrifugation at 200 xg for 2 min to settle soil particles. The supernatant containing the extracted bacteria was collected with a serological pipette and transferred to a clean 50 ml tube. The remaining soil pellet was extracted 4 more times for each sample. Thereto, the soil pellet wasresuspended in 4.5 ml of 50 % (v/v) EtOH and PBS and vortexed briefly to detach the soil pellet from the tube bottom and the extraction process was repeated.

The 5th and final extraction of free-living bacteria was done similarly, however, the centrifugation step to collect the supernatant was skipped and the soil slurry was layered on top of a cushion of 10 ml Nycodenz solution of 1.3 g/ml using a serological pipette in a 35 ml ultracentrifuge tube forming a sharp interface between the soil slurry and Nycodenz. The collected supernatant containing the extracted bacteria from the preceding extraction steps was added as well, resulting in one single ultracentrifugation step for all five extraction steps. The samples were centrifuged at 10143 xg for 30 min at 4 °C in a swing-out rotor in an ultracentrifuge to separate the free-living bacteria from co-extracted soil particles and debris. After centrifugation, the interphase between extraction medium and Nycodenz containing the cells was collected using a serological pipette. To extract the surface-associated bacteria, the remaining soil pellet in the ultracentrifuge tubes was resuspended in 4.5 ml of detergent solution. Again, three 8 mm and three 4 mm glass beads were added and vortexed briefly to loosen the pellet. The soil slurry was transferred to a 50 ml screw cap tube. The tube was vortexed vertically for 1 min at full force and horizontally for 15 min at half force. The soil slurry produced was again layered on top of a 10 ml Nycodenz cushion in an ultracentrifuge tube and processed at identical conditions a the extraction in of the free-living bacteria described before.

After the ultracentrifugation, all samples were centrifuged at 17000 xg for 10 min. The supernatant was discarded and the cell pellet was resuspended in 1 ml 50 % PBS/EtOH solution by sonication three times for 5 min in a sonication bath at room temperature. In order to remove still remaining soil particles and purify the extracted cells even further a second Nycodenz extraction was conducted. Therefore, the resuspended cells were layered on top of 1 ml Nycodenz of 1.3 g/ml in 2 ml tube and centrifuged at 10000 xg for 30 min

at 4 °C in a fixed angel rotor in a Beckmann 4230 tabletop centrifuge to clean the cells even further and separate remaining soil particles and debris. After centrifugation, all supernatant above the Nycodenz cushion was collected, transferred to a new 2 ml tube and centrifuged at 20000 xg. The supernatant was discarded and the final cell pellets were resuspended in 750 µl of RNA Shield and stored at -20 °C until nucleic acid extraction.

3.2.4 Direct cell counting

Determination of the total cell number in soil samples and and the cell number of non-extractable bacteria, remaining in the soil samples after cell extraction, were determined by addition of 900 µl 10 mM HEPES buffer of pH 7 containing 2 % (v/v) Glutardialdehyde to 0.1-0.25 g of soil. The samples were vortexed briefly and fixed for 10 min. 100 µl of soil slurry were transferred to a new 2 ml tube with a 1 ml pipette using tips of which the tip has been cut before autoclaving to allow the uptake of coarse soil particles. The samples were further diluted by addition of 900 µl 10 mM HEPES. After sample homogenization by vortexing, 50 µl were transferred into a new 2 ml tube and mixed with 1 ml of 10 mM HEPES and 450 µl methanol. The samples were homogenized in a sonicator bath for 15 min at 35 °C. 500 µl were transferred to a 15 ml screw cap tube containing 9.5 ml 2mM MOPS buffer. Cells were stained by addition of 2 µl SybrGreen I stock solution (stock concentration: 10000x in DMSO) for 10 min in the dark on a rotary shaker. Afterwards, cells were transferred onto black polycarbonate filters of 2 µm pore size and 21 mm in diameter. The filters dried briefly at room temperature and half a filter was mounted onto a microscopic slide with 10 µl of anti-fading reagent (250 mg 1,4 Diazabicyclo(2,2,2)octan dissolved in 1 ml of 3x PBS and diluted with 9 ml glycerol), covered with a cover slip and sealed with nail polish. Cells of free-living and surface-associated bacteria in cell extracts were counted using 75 µl of cell extracts stored in RNA/DNA shield. 900 µl 10 mM HEPES buffer of pH 7 containing 2 % Glutardialdehyde was added and stained with 0.2 µl SybrGreen I stock solution for 10 min in the dark on a rotary shaker. The stained cells were added to 9 ml 2 mM MOPS in a 15 ml scew cap tube and transferred onto filters as described above. 10 fields of view (FOV) were counted for each sample using a Zeiss AxioVision Epifluorescence microscope counting 50-200 cells per FOV. Total cell numbers were calculated using the formula below:

$$\text{Total cell number g}^{-1}\,\text{dry soil} = \frac{\text{total cell counts} \times \text{filter factor} \times \text{dilution factor}}{\text{g dry soil}} \tag{3}$$

with:

$$\text{Filter factor} = \frac{\text{filter area } \mu\text{m}^2}{\text{FOV area } \mu\text{m}^2} \tag{4}$$

which results in:

$$\text{Filter factor} = \frac{346360590.1\,\mu m^2}{15625\,\mu m^2} = 22167.08 \tag{5}$$

The dilution factor used was 600, 600 and 10 for soil bacteria, non-extractable bacteria and extractable bacteria respectively.
In contrast to direct counting, most probable number statistics (MPN) was used to enumerate the bacteria on particles co-extracted from soil during cell extraction. The MPN was estimated following the method in (Roser et al., 1987) using table VIII2 in (Toutenburg, 1971). Simulating dilution factor of a = 2, 4 different areas of defined size, each of them half in size of the preceding one, were checked for the presence of bacteria on particles. The largest area corresponded to the area of one FOV as defined above. For each size class, n = 10 random replicates were examined on each filter. The mean number of positive fields x was calculated by summing up the number of positive fields and dividing it by the number of replicated fields n examined for each size class. For each x, K was selected from table VIII2 in (Toutenburg, 1971). Finally the cell number $\lambda\,\text{FOV}^{-1}$ was estimated using:

$$\log_{10}\lambda = x\log_{10}a - K \tag{6}$$

Finally, the total number of bacteria on particles was calculated by:

$$\text{total cell number}\,g^{-1}\,\text{dry soil} = \frac{\lambda \times \text{filter factor} \times \text{dilution factor}}{g\,\text{dry soil}} \tag{7}$$

The dry weight of soil samples used for bacterial cell counting was calculated by determining the gravimetric water content for reference samples of soil and soil after cell extraction. The samples were dried at 105 °C until reaching constant weight.

3.2.5 Nucleic acid extraction and library preparation

RNA and DNA were extracted from cell extracts using Zymo Research (ZR) Miniprep Kit (R2002). This kit was chosen, since it is specifically designed for hard to lyse cells from environmental samples. Samples stored in 750 µl ZR RNA shield were transferred to ZR BashingBead Lysis Tubes containing sterile 0.1 and 0.5 mm zirconium beads. Nucleotide-free water was used instead of sample material for negative controls. Bead beating was performed on a FastPrep24 bead beater (MP Biomedicals) at 6.5 ms^{-1} for 60 s. Samples

were centrifuged at 16000 xg for 30 s and 500 µl of supernatant were transferred into a new 2 ml tube. 1 ml of ZR Lysis Buffer was added and mixed well. The lysed cells were transferred into a ZR SpinAway Filter tube and centrifuged at 16000 xg for 30 s. For RNA extraction, 1.5 ml of flow-through were mixed with 1.5 ml of 100 % ethanol and centrifuged using a ZR ZymoSpin IIICG column at 16000 xg for 30 s. The RNA, bound to the column matrix, was washed with 400 µl of ZR Wash Buffer and centrifuged at 16000 xg for 30s. Next, the RNA was treated with 80 µl of DNAse I at a concentration of 0.625 U μl^{-1} at 30 °C for 30 min to remove any traces of genomic DNA. Both columns, containing the RNA and DNA bound to the column matrix were washed with 400 µl ZR Prep Buffer and centrifuged at 16000 xg for 30s. 700 µl of ZR Wash Buffer were added to the column and centrifuged at 16000 xg for 30 s. Again, 400 µl of ZR Wash Buffer were added to the column and centrifuged at 16000 xg for 2 min. The columns were transferred to a new 2 ml tube and incubated with 50 µl of RNAse-free water for 5 min at room temperature. After the incubation, the columns were centrifuged at 16000 xg for 30 s. To remove potential contained PCR inhibitors, 600 µl ZR HRC Prep Solution was added to ZR III-HRC Filter columns and centrifuged at 8000 xg for 3 min. The eluted RNA and DNA samples were added to the prepared ZR III-HRC Filter columns and centrifuged at 16000 xg for 3 min. RNA samples were checked for DNA contamination by adding 15 µl of nucleotide-free water to 5µl of RNA extract or 20 µl of nucleotide-free water for the negative control, and performing PCR on the V3/V4 region of the 16S rRNA gene using forward and rerverse primers 415f and 830r in combination with the Phusion HotStart II High-Fidelity Polymerase (Thermo Scientific). After initial denaturation for 30 s at 98 °C, a denaturation for 10 s at 98 °C, annealing for 10 s at 55 °C and elongation for 45 s at 72 °C were repeated for 15 cycles. After the amplification, a end-elongation followed for 2 min at 72 °C and a cooldown to 4 °C. PCR products were checked for amplification products by gel-electrophoresis. All tested RNA samples still contained genomic DNA compared to the negative control. Therefore, a second DNAse I treatment was performed using the Invitrogen Turbo DNA-free kit according to the manufacturer protocol. The RNA samples were subsequently resuspended in 20 µl RNAse-free water and the purified RNA and DNA samples were stored at -80 °C.

Library preparation for RNA samples started with reverse transcription of RNA in DNA. 4 µl of RNA extract were added to 1 µl of Random Hexamers and heated on a PCR-cycler for 5 min at 70 °C. The mixture was cooled on ice for 5 min and 15 µL of PCR-buffer (1x GoScript reaction buffer, 3 mM $MgCl_2$, 0.5 mM nucleotide mix, 0.05 U μl^{-1} GoScript reverse transcriptase) were added. The reverse transcription continued with 5 min at 25 °C, 1 h at 42 °C, 5 min at 70 °C and finished with a cooldown to 4 °C.
A two-step multiplex PCR was performed by adding 20 µl of reverse transcription product

to 30 µl PCR-buffer (1x HF Phusion buffer, 0.8 mM nucleotide mix, 3.5 mM $MgCl_2$, 3 % DMSO, 0.4 mg/ml BSA, 0.25 pmol/µl primers and 0.03 U/µl Phusion HotStart II High Fidelity DNA Polymerase (Thermo Scientific). After initial denaturation for 30 s at 98 °C, denturation for 10 s at 98 °C, annealing for 10 s at 55 °C and elongation for 45 s at 72 °C were repeated for 15 cycles. A end-elongation followed for 2 min at 72 °C and a cooldown to 4 °C. In a second step, 4 µl of PCR-product were mixed with 48 µl PCR-buffer (1x HF Phusion buffer, 0.8 mM nucleotide mix, 3.5 mM $MgCl_2$, 3 % DMSO, 0.4 mg/ml BSA, 0.25 pmol/µl primers and 0.03 U/µl Phusion HotStart II High Fidelity DNA Polymerase (Thermo Scientific)) and denaturated for 30 s at 98 °C. After initial denaturation, 15 cycles of denaturation for 10 s at 98 °C, annealing for 10 s at 55 °C and elongation for 45 s at 72 °C followed. Finally, after End-elongation for 2 min at 72 °C, the samples were cooled to 4 °C.
4nM of each sample were pooled, spiked with 15 % PhiX and sequenced on a MiSeq platform (Illumina, San Diego, CA, USA) using a 350 base pair paired-end-run kit.

3.2.6 Bioinformatics

Sequences were retrieved from the Illumina MiSeq platform already demultiplexed, according to the barcodes used for each sample. The demultiplexed sequences were processed using the Python based QIIME2 2019.1 bioinformatic pipeline (Caporaso et al., 2010; Bolyen et al., 2019) run in a Linux environment on a high performance cluster with default settings if not stated otherwise.
Bacteria were distinguished at the level of single nucleotide differences as exact exact sequence variants (SV), defined as a specific nucleotide sequence of the V3/V4 16S rRNA gene region (Callahan et al., 2017). For this project, the same pipeline infrastructure as in chapter 2.2.3 was used. However, changes to the existing protocol from chapter 2.2.3 are listed below. In short, the demultiplexed paired-end reads were joined using plugin *vsearch*. Joined reads were processed using plugin *quality-filter* (Bokulich et al., 2013). Reads having a quality score below 25 over at least 75 % of the read length were discarded. Plugin *deblur* (Amir et al., 2017) was used for denoising the joined sequences and checking for chimera. First, all sequences were trimmed to a length of 350 base pairs. Second, all sequences were discarded occurring less than 10 times in all samples and less than 2 time in each sample. For sequence alignment MAFFT (Katoh and Standley, 2013) was used. The phylogenetic tree was calculated using the FastTree plugin (Price et al., 2010). The taxonomy was annotated using the plugin *feature-classifier*. A Naïve Bayes classifier trained with the SILVA 132 database (Quast et al., 2013; Yilmaz et al., 2014) for the V3/V4 region. In case the taxonomic affiliation can not be determined, SV were kept as "uncultured bacterium" (Overmann et al., 2017). In a final step, sequences of mitochondria and

chloroplasts were removed and the abundance data, taxonomy table and phylogenetic tree - in Newick format - were retrieved and exported to R ({R Core Team}, 2021).

3.2.7 Sequencing data normalization

All downstream analysis, data normalization and statistics were done entirely in R 4.0.1. All sequencing informations were imported into R as a phyloseq-object using package `phyloseq`. Package `tidyverse` was used for data handling and packages `ggplot2`, `metacoder` and `gridExtra` for visualization. Sequencing data normalization, pivotal for alpha and beta diversity analysis, was performed as coverage-based rarefaction using function `metagMisc::phyloseq_coverage_raref`. Coverage-based rarefaction normalizes samples to identical sample completeness rather than identical number of reads (Chao and Jost, 2012). This approach avoids precision loss caused by discarding the amount of reads higher than minimum sequencing depth (Chao et al., 2020).

3.2.8 Calculation of alpha and beta diversity

Hill numbers were calculated as alpha diversity measures to calculate richness, Shannon and Simpson indices representing the total number of bacteria in a sample, the number of abundant and the very abundant bacteria in a sample, respectively. For the calculation of alpha diversity, please refer to chapter 2.2.8.
To account for data compositionality, the bacterial abundance/count data were transformed by addition of a pseudo-count of 1 and calculating the natural logarithm (Quinn et al., 2018). Sequencing data are compositional, meaning that derived abundance are not true abundances and depend on the capacity of the sequencing platform. (Gloor et al., 2017). Consequently, beta diversity was calculated on transformed read counts as weighted Unifrac distances, which incorporate abundance and phylogeny (Lozupone and Knight, 2005).

3.2.9 Determination of metabolic active bacteria

RNA, in contrast to DNA, is rapidly decomposed by RNAse in the environment. Hence, the extraction of bacterial 16S rRNA in soil, implies the existence of intact cells and ribosomes, allowing for protein synthesis. The number of ribosomes in a cell was found to correlate with growth rates of pure cultures (Kerkhof and Ward, 1993; Kemp et al., 1993). This approach is also used as an indicator of protein synthesis potential for bacteria in natural communities (Campbell et al., 2011; Denef et al., 2016; Bowsher et al., 2019). Hence, the protein synthesis potential of each SV in each sample was determined as the rRNA:rDNA ratio on a read count basis using a novel dynamic threshold approach. To do so, all reads for each SV

in a sample were added up in all 3 technical replicates. Differences in library sizes between rRNA and rDNA communities of each sample were normalized to minimum sequencing depth using function `rarefy.even.depth` implemented in package `phyloseq`. Based on a dynamic threshold, sensitive to DNA abundance, each SV was classified as active or inactive. SV having 0 reads only in the DNA community were replaced with a pseudocount of 1 to be able to calculate the rRNA:rDNA ratio. This derivation of this dynamic threshold is described in Noviana (2021). Briefly, the dynamic threshold was developed since mostly rare taxa were classified as active in literature (Jones and Lennon, 2010; Campbell et al., 2011; Klein et al., 2016). However, this finding is based on an artefact inherent to amplicon sequencing data generation. Rare SV, represented by a low number of reads, have a high variation of up to 80 % between technical replicates, accounting for variation introduced during library preparation (nucleic acid extraction, PCR and bioinformatics). On the contrary, abundant SV show a lower variation of 5 %. As a result, to avoid a inflated false discovery rate, a dynamic threshold considering SV abundance in the DNA community was applied. This measure will prevent rare SV from being identified as potentially active caused by stochastic variation. Like that, false discovery rate could be reduced to 0.4 %.

3.2.10 Statistical analysis

As stated before, pH between fertilization treatments was constant. However, differences between major soil nutrients were highly correlated between treatments (Fig. 27). Hence, C_{org} was chosen to represent soil nutrients av. P, av. K, N_t and C_{org}.
Statistical testing for alpha and beta diversity analysis relied mainly on the `vegan` package. Differences between groups were tested for significance with ANOVA and TukeyHSD post-hoc testing using functions `aov` and `TukeyHSD` respectively. Beta diversity ordination was based on weighted Unifrac distances (Lozupone and Knight, 2005). Ordination was performed using distance-based redundancy analysis (RDA) with function `capscale` and the first two dimensions were displayed. The 95 % confidence intervals of nominal scaled predictors were calculated using function `veganCovEllipse`.
Subsequent variance partitioning of environmental parameters in beta diversity of total and potentially active rRNA communities were estimated using function `varpart`. The respective p-values were estimated by PERMANOVA via function `adonis2`.
Moreover, variable importance and SV specific modelling of beta diversity and turnover was performed using gradientForest analysis (Ellis et al., 2012). For detailed information see remarks in section 2.2.9.

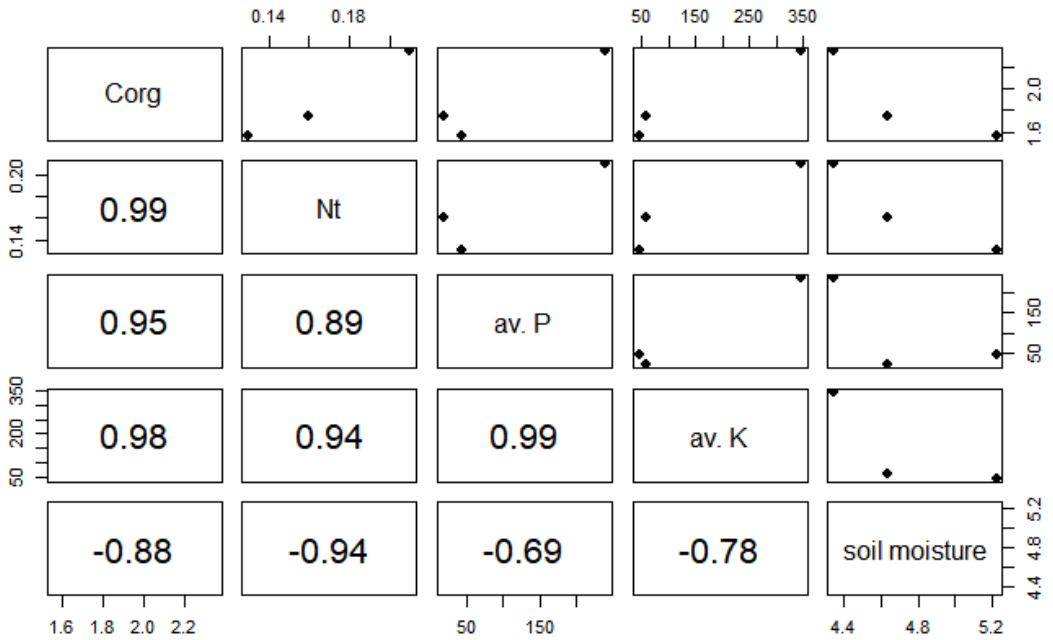

Figure 27: Pearson correlation between soil nutrients and soil moisture. Corg = organic carbon, Nt = total N, av.P = available P, av.K = available K, soil moisture = gravimetrical water content.

3.3 Results

3.3.1 Cell extraction from soil differs by bacterial life-style and not by fertilization

Bacteria were extracted from the haplic chernozem of the Static Fertilization Experiment from Bad Lauchstaedt. Shaking the soil with PBS, in combination with soil aggregate disruption using glass beads was applied to retrieve the free-living community, living in the soil pore space. Afterwards, surface-associated bacteria were extracted similarly with detergents in the extraction solution. Bacterial cells were enumerated by direct counting using an epi-fluorescence microscope, to compare the extracted number of cells used for library preparation. This was necessary to characterize and provide a reliable cell extraction method for downstream sequencing of bacterial communities.

A single extraction with PBS followed by a single extraction with detergents was compared to 5 consecutive extractions followed by a single extraction with detergents in order to analyse the effect of single or multiple PBS extractions on cell counts of free-living or surface associated communities. A median number of 2.3×10^5, 1.2×10^6, 2.2×10^6 and 1.0×10^7 cells g^{-1} dry soil were extracted respectively (Fig. 28). The results show that 5 consecutive

PBS extractions yield indeed 5 times more cells. Moreover, this increase in PBS extraction steps also increased the amount of cells extracted with the subsequent detergent extraction by factor 5. Hence, 5 consecutive extractions instead of a single extraction with PBS were chosen to represent the free-living community. Moreover, library preparation for bacterial communities failed for rRNA samples using only a single extraction step for the free-living community.

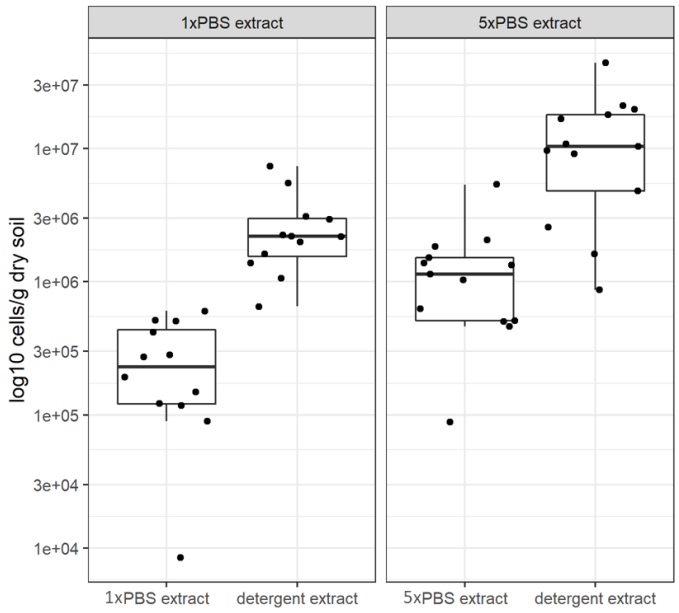

Figure 28: Cell counts per gram dry soil on a $\log_{10}$ scale for different cell extraction methods. The free-living community is represented by a single extraction (1x PBS extract) and 5 consecutive extractions with PBS (5xPBS extract). The surface associated community is represented by a single extraction with detergents (detergent extract). Dots represent independent experiments.

Extracted bacteria from soil were separated from soil particles and debris via Nycodenz density gradient centrifugation. However, the cell extracts were not free of bacteria still attached to soil particles (Fig. 29). Bacteria on particles were enumerated using most probable number (MPN), to evaluate if bacteria on particles might bias the free-living or surface-associated community. The extraction of free-living and surface-associated bacteria co-extracted equals amount of bacteria on particles. The median number of bacteria on particles represented 0.7 % and 0.07 % of the median number of extracted bacteria of the free-living and surface-associated community respectively.

Comparing the number of cells extracted from soil between different fertilization treatments, the free-living community extracted on average 10^6 cells per gram of soil for all fertilization treatments (Fig. 30). In contrast, the surface-associated community extracted

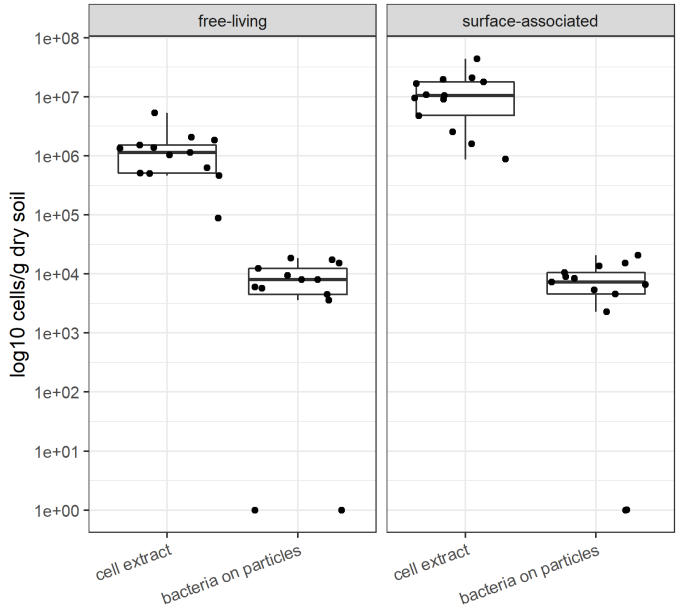

Figure 29: Number of cells extracted per gram of dry soil with either 5 consecutive PBS extractions (free-living) or a single extraction with detergents (surface-associated) compared to the number of cells adhered to particles coextracted from soil. Cell extracts were estimated by direct counting. Bacteria on particles were estimated via MPN. Dots represent independent experiments.

on average 10^7 cells for all fertilization experiments.

In relation to the total number of bacteria in soil, the cell extraction efficiency from soil represented 0.02 %, and 0.2 % of soil cell counts for the free-living and surface-associated community (Fig. 31). Non-extractable cells, remaining in the soil cell extraction accounted for 4 % of the total cell counts. Obviously, a difference in cell counts existed between the number of extracted cells and non-extractable cells on the one hand and cell counts in soil on the other hand. The proportion of cells recovered in extraction account on average only for 4.22 % of cells in soil. I assume that this difference was caused by cell lysis during extraction.

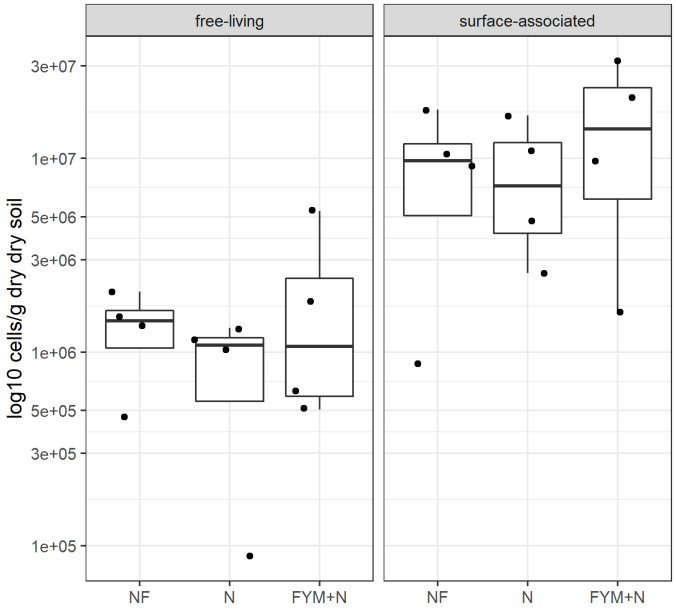

Figure 30: Direct cell counts for the free-living community extracted with 5 consecutive PBS extractions and the surface-associated community extracted with detergents displayed for 3 distinct fertilization treatments. NF = no fertilization, N = mineral nitrogen, FYM+N = farmyard manure + mineral nitrogen. Dots represent independent cell extractions from soil.

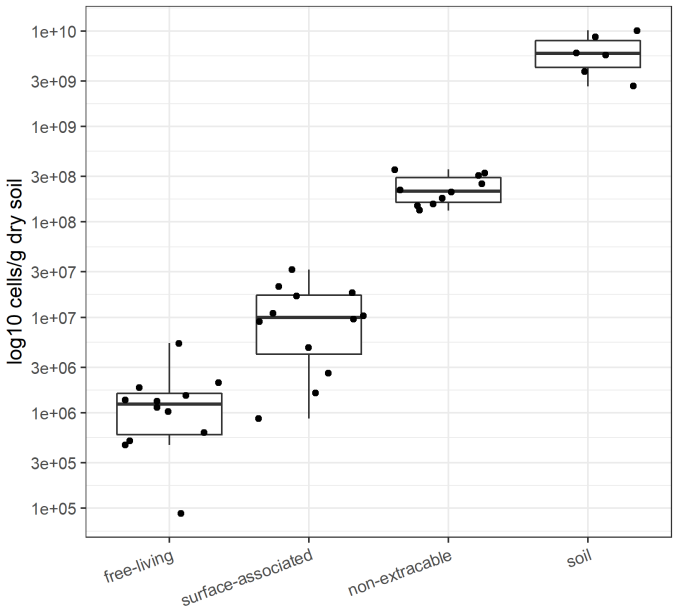

Figure 31: Cell counts for the free-living bacteria extracted with 5 consecutive PBS extractions, the surface-associated community extracted with a single detergent extraction, non-extractable bacteria remaining in the soil after extraction and the total number of cells in soil. Sample size between plots differs due to a varying number of experiments between cell extractions and cell counting in soil.

3.3.2 Sequencing results and contamination removal

The sequenced samples were comprised of low biomass samples, extracting only up to $3x10^6$ bacteria. This is challenging since standard extraction kits require normally more biomass to work properly. In addition, RNA and DNA had to be separated and sequenced, reducing the available amount of nucleic acid even further. For that reason, negative controls were sequenced to check for the reliability of library preparation and to remove contaminating sequences in the samples. Sequence variants (SV) in negative controls having $\geq$ 0.1 % relative abundance were defined as contaminating sequences and removed from the free-living and surface-associated community samples prior to analysis, following the method of Szoboszlay and Tebbe (2021). 109 SV in negative controls of DNA and 25 SV in negative controls of RNA were identified as contaminants. Contaminants made up 99.6 % and 99.7 % relative abundance in negative controls of RNA and DNA respectively (Fig. 32). RNA communities of the free-living and surface-associated bacteria showed higher levels of contamination than DNA based communities. RNA based communities of the free-living and surface-associated bacteria were comprised of 56 % and 18.9 % of contaminants. In contrast, DNA based communities of the free-living and surface-associated bacteria contained 3.02 % and 4.04 %, respectively. This high level of contamination for the RNA samples of free-living bacteria was presumably a result of low template amount for the library preparation, caused by RNA degradation due to longer sample preparation because of exhaustive DNA digestion, since DNA samples of the free-living community showed much less contamination.

The sequencing depth per sample is highly variable ranging from 9339 to 197949 reads (Fig. 33). However, all rarefaction curves were in saturation, meaning that the sequencing effort was capable to represent the diversity present in the samples. Generally, RNA-based samples contained less reads than DNA-based samples, but the sequencing depth was similar between fertilization treatments. Moreover, the number of SV in the surface-associated samples was higher than in the free-living samples for both RNA and DNA. Due to the high variation in sequencing depth, rarefaction to minimum sequencing depth (McMurdie and Holmes, 2014) for all samples would discard the majority of reads for most samples and lead to loss of precision. As a consequence, for the determination of active SV, samples were grouped by extraction type and fertilization treatment and normalized to minimum sequencing depth, minimizing the loss of reads for each group. As a result, absolute read-based comparisons of active SV between fertilization treatments are not meaningful and hence only relative comparisons based on relative abundances were analysed.

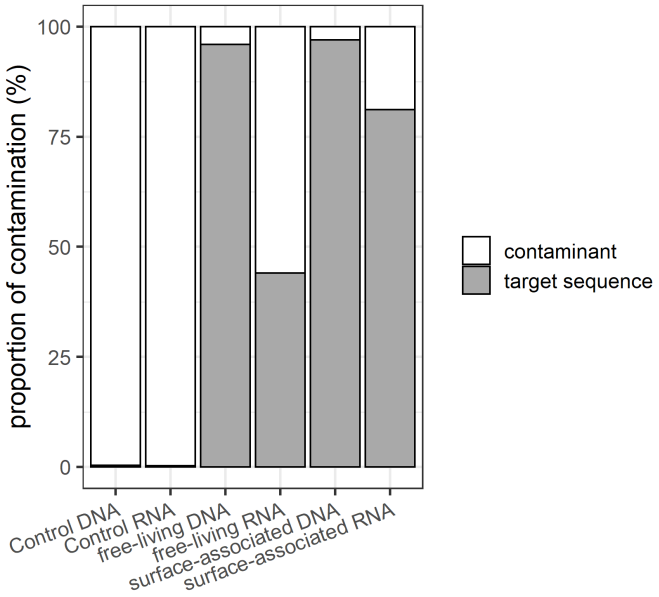

Figure 32: Average proportion of contaminating sequence variants (%) in negative controls, free-living and surface-associated communities for DNA and RNA samples respectively.

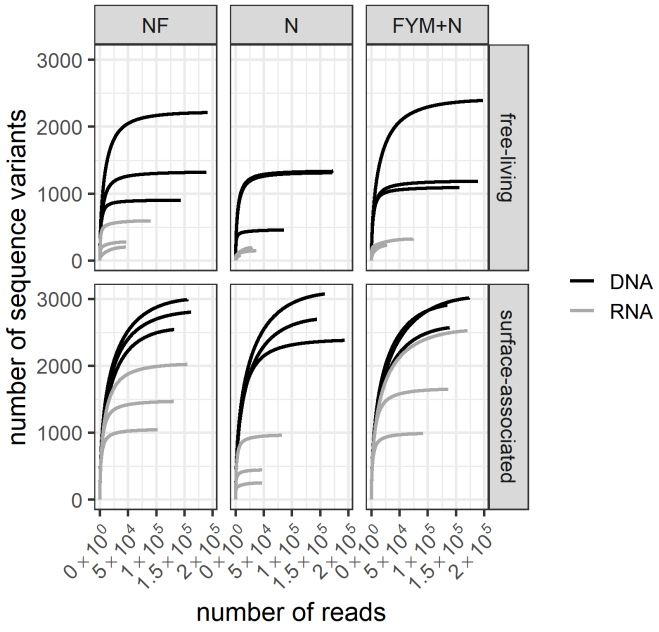

Figure 33: Rarefaction curves of un-normalized read counts displayed for the respective RNA and DNA sample. Rarefaction curves are displayed separately for each fertilization treatment and the respective cell extraction method. No fertilization = NF, mineral nitrogen = N, farmyard manure in combination with mineral nitrogen = FYM+N. Rarefaction curves of identical color represent technical replicates of respective RNA or DNA samples.

3.3.3 Alpha diversity is shaped by bacterial life-style but not soil fertilization

Alpha diversity was calculated to deduce how bacterial life-style, long-term fertilization and nucleic acid type affected the abundance of rare and abundant species in extracted soil communities.

The surface-associated DNA communities contained on average the most species (Fig. 34). In contrast, the free-living RNA communities contained the least amount of species. All, soil bacterial communities were structured similarly with much more rare than abundant species except for the free-living RNA communities which contained proportionally more abundant species. Distinct fertilization treatments contained similar species counts for each community type except for the N treatment having a tendency to lower alpha diversity for all communities except Simpson in the DNA communities (Fig. 34). Moreover, surface-associated communities showed higher diversity than free-living communities for the same nucleic acid type. However, the surface-associated community of FYM+N from DNA samples was lower for Simpson than the respective free-living community. Also, DNA samples were in general more diverse than the respective RNA samples. Testing these observed differences for statistical significance using multiway ANOVA, the effects of nucleic acid type and bacterial life-style were statistically significant (Tab. 3). However, the effect of fertilization treatments was not significant for alpha diversity. The effect of water content on alpha diversity could not be estimated.

Table 3: 3-way ANOVA testing for differences in means of alpha diversity. Df = degrees of freedom, SS = sum of square. Technical replicates of alpha diversity were averaged before calculation of ANOVA.

	Df	SS	F-value	p-value
nucleic acid type	1	2252353	9.713	0.004
life-style	1	2369491	10.218	0.003
fertilization	2	121107	0.261	0.771
residuals	29	6724849		

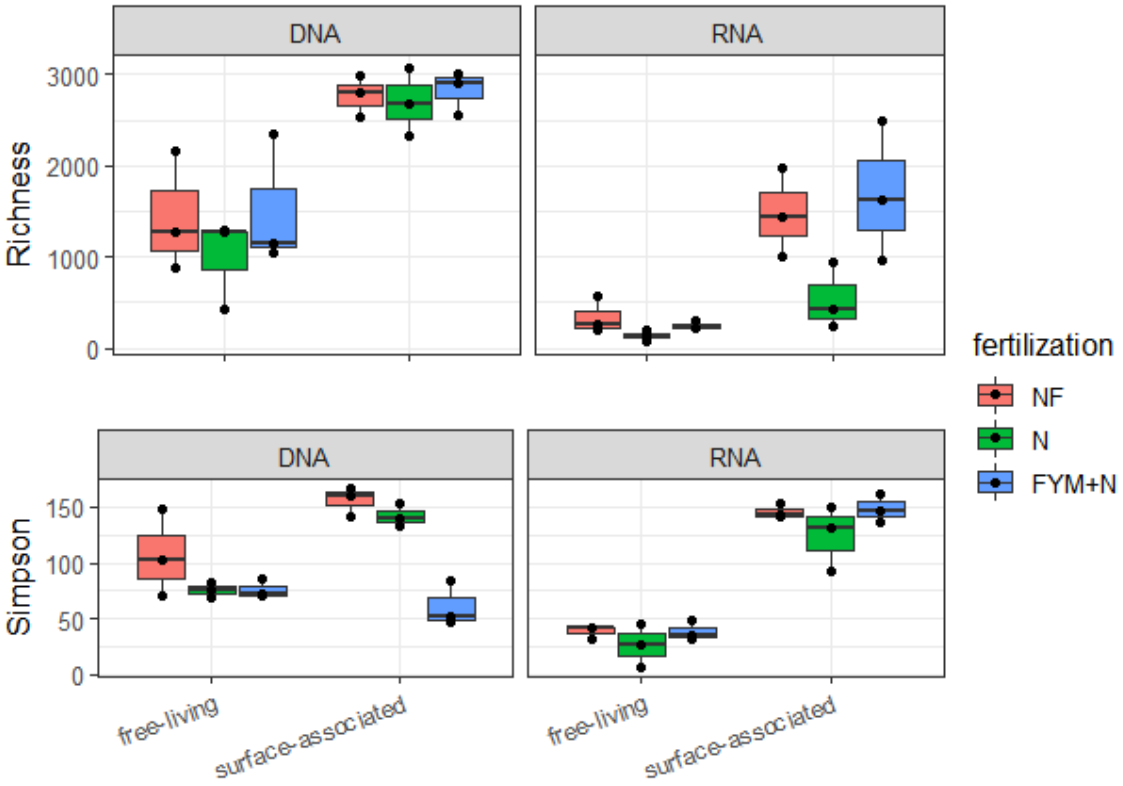

Figure 34: Alpha diversity for the DNA and RNA samples of the free-living and surface-associated bacteria, displayed as representative number of species. Richness is ignoring species abundance (q = 0), Shannon emphasizing abundant sequence variants (q = 1) and Simpson is emphasizing highly abundant sequence variants (q = 2). Each boxplots represents results for 3 technical replicates.

3.3.4 Beta diversity of total and active communities is shaped by bacterial life-style and fertilization

Beta diversity describes the diversity between samples and was based on weighted unifrac distances in combination with redundancy analysis (RDA) to explore which predictors affected bacterial community composition. I anticipated that the community compositions differ by life-style and fertilization treatment and those differences to increase even further for the active communities.

PERMANOVA revealed significant effects ($p < 0.05$) of C_{org} representing fertilization, bacterial life-style and nucleic acid type (Tab. 4). Due to high multicollinearity between soil nutrient concentrations (C_{org}, N_t, av. P, av. K), C_{org} was selected to represent the fertilization effect represented by soil nutrients (Tab. 2), since organic C before N and P was shown to drive bacterial diversity on the examined site (Francioli et al., 2016). Moreover,

the effect of gravimetric water content was not significant. The variance explained by nucleic acid type and bacterial life-style was higher than the effects of C_{org} and water content. Combined, the variance explained for all constraints accounted for 50 % of data variation.

However, RDA revealed that the influence of fertilization treatment and water content, displayed as blue arrows, separated FYM+N from NF and N samples (Fig. 35). NF and N were more similar in sample spread, showing no clear separation from each other. The arrow length displayed, indicates the fit and the direction of the arrow hints towards the direction of maximum change in community composition.

Table 4: Variance partitioning of environmental metadata and the respective p-values estimated by redundancy analysis and permanova on beta diversity calculated on weighted unifrac distances.

factor	p-value	variance explained (%)
nucleic acid type	0.001	23
life-style	0.001	16
C_{org}	0.014	6
water content	0.086	5

Complementary to RDA, a non-parametric gradientForest model, based on Random Forest (Breiman, 2001) was fitted to resolve the effects of soil nutrients on bacterial abundance on SV level. Bacterial life-style, water content and fertilization were most important with nucleic acid type playing a minor role in variable importance (Fig. 36). 65 SV showed a goodness of fit of $R^2 > 0.5$, representing 0.66 % of all SV. Fig. 37 depicts the partial variable importance of each predictor for each SV with goodness of fit > 0.5. Although negatively correlated, variable importance of C_{org} and soil water content differed for some SV. Concentrating on C_{org} representing fertilization, 4 SV had a partial importance > 0.3. Those SV were affiliated to genera *Luteimonas, Mesorhizobium, Pedomicrobium* and uncultured *Gaillela*. SV with a partial importance < 0.3 for bacterial life-style were members of *Actinobacteria*: *Solirubrobacterales, Acidimicrobiia, Microtrichales* and *Alphaproteobacteria Sphingomonadales* and *Rhizobiales* as well as a member of *Firmicutes* and *Nitrospira*. Generally, bacterial life-style affected most SV. This is in line with the overall variable importance (Fig. 36).

Examining the reproducibility of the cell extraction procedure, the variation between technical replicates of identical life-style and nucleic acid type is small (Fig. 38). Samples of identical origin, always group closely together. The grouping effect was significant ($p = 0.01$), proving that the variation within groups of samples was smaller than the variation between groups and tested on weighted unifrac distance matrix using multivariate ANOVA.

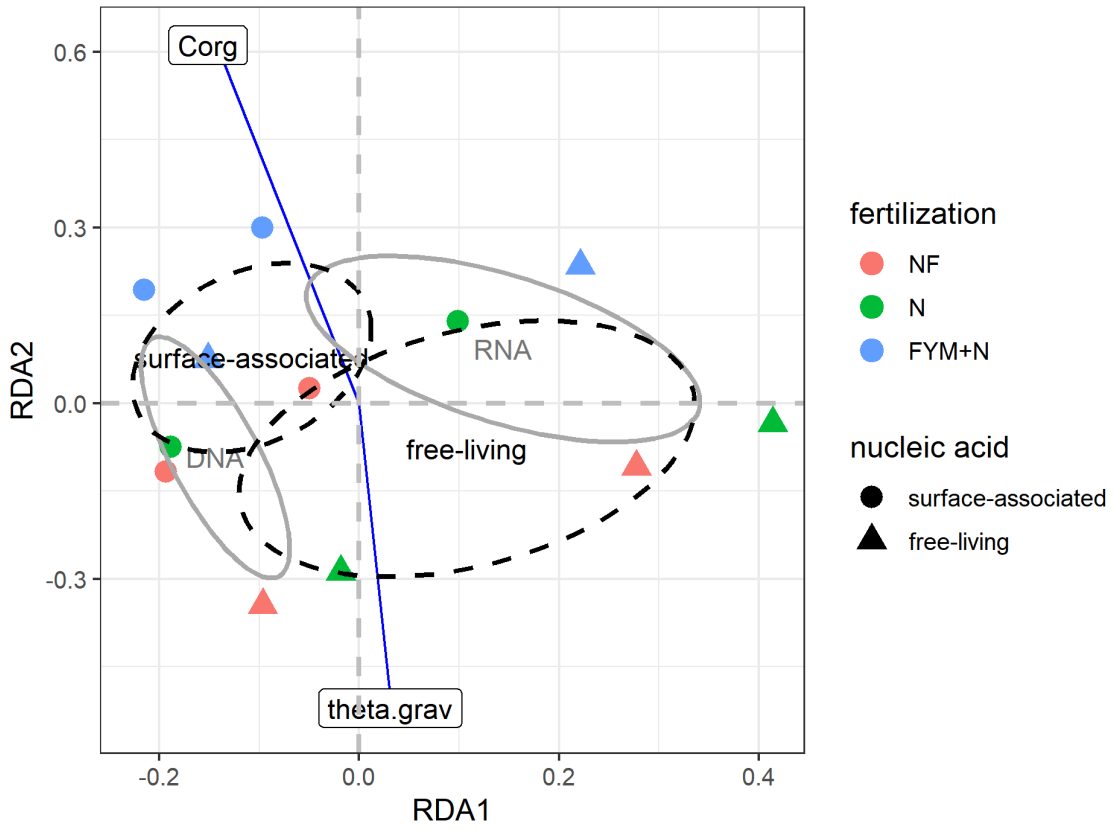

Figure 35: Redundancy analysis of complete communities, calculated on weighted unifrac distances. Technical replicates were averaged before ordination calculation. C_{org} and theta.grav represent fertilization and gravimetric water content displayed as blue arrows. Arrow length indicates goodness of fit and the direction represents the direction of maximum change. Nucleic acid type (RNA, DNA) and bacterial lifestyle (free-living, surface-associated) are factor constraints, displayed as centroids with the respective 95 % confidence interval.

The analysis of beta diversity of active communities revealed that the differences between fertilization treatments changed (Fig. 39). Similar to the alpha diversity of RNA samples, N was more distinct from NF and FYM+N in community composition. Additionally, the clear separation of communities by bacterial life-style was still prominent. This was approved by variance partitioning (Tab. 5), confirming life-style as the only significant predictor shaping the active communities. Also the variance explained increased by 7 % for bacterial life-style. Although non-significant, variance explained for C_{org}, representing fertilization, and water content increased by 7 and 11 %, respectively. In contrast to the total communities, the active communities contained no SV with an overall goodness of fit $R^2 > 0.5$.

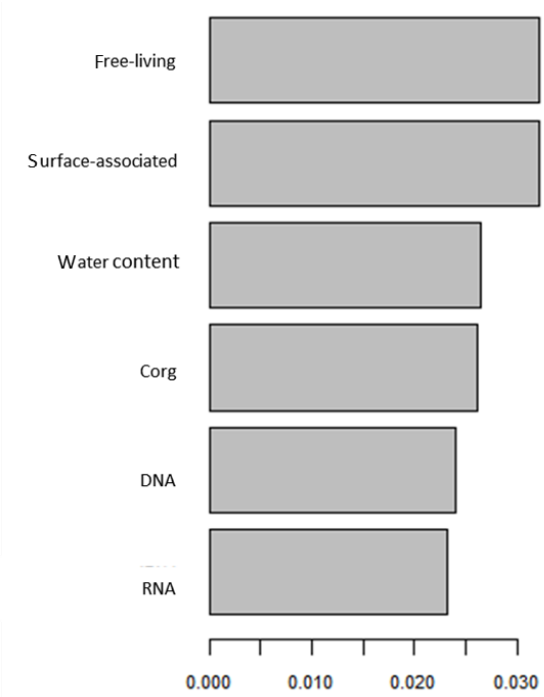

Figure 36: Variable importance weighted by species R^2 from gradientForest analysis of all species.

Table 5: Variance explained in active rRNA communities by environmental predictors estimated by redundancy analysis and their respective p-values by PERMANOVA on beta diversity calculated on weighted unifrac distances.

factor	p-value	variance explained (%)
life-style	0.03	30
C_{org}	0.58	13
water content	0.14	16

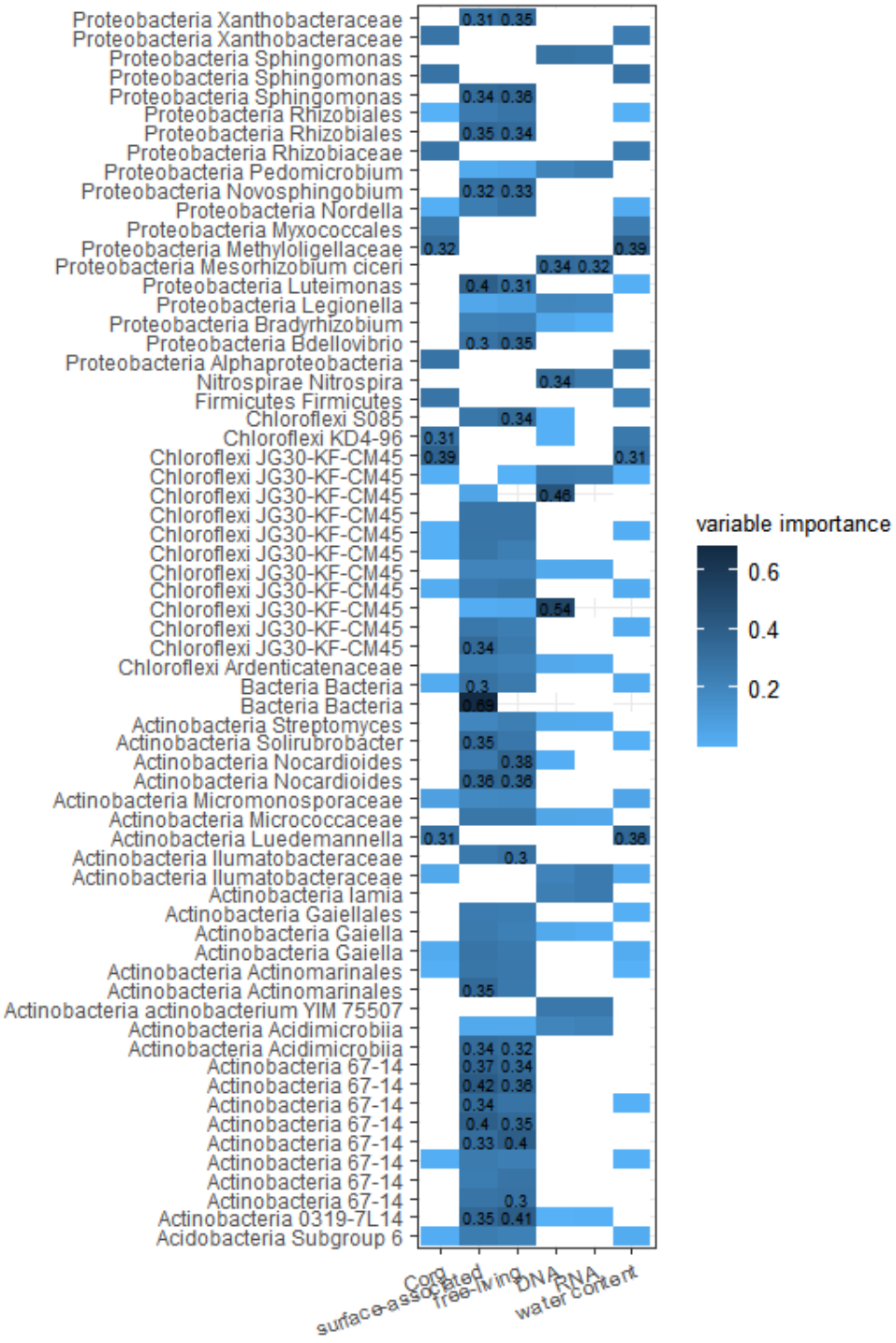

Figure 37: Variable importance of SV having a goodness of fit $R^2 > 0.5$ from gradientForest analysis. Unique SV are named by a combination of their phylum in combination with their highest meaningful taxonomic level. Variable importance >0.3 are printed in each tile. Tiles for which the fit could not be determined were left blank.

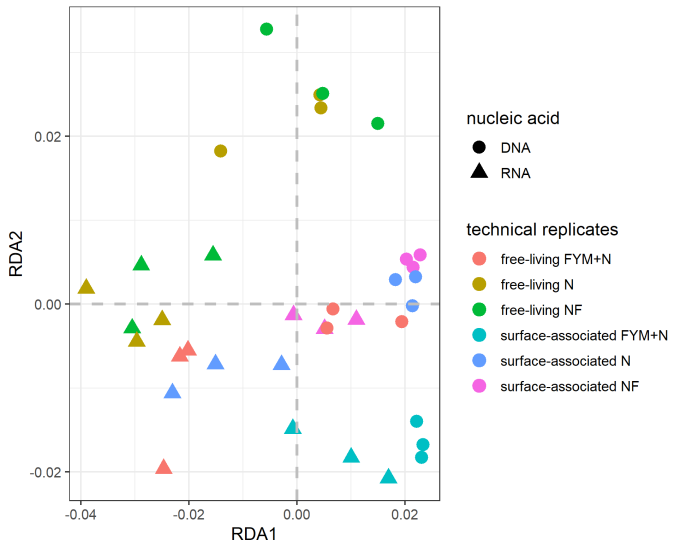

Figure 38: Redundancy analysis based on weighted unifrac distances for all samples. Technical replicates of bacterial communities are coloured by their respective combination of fertilization treatment and life-style.

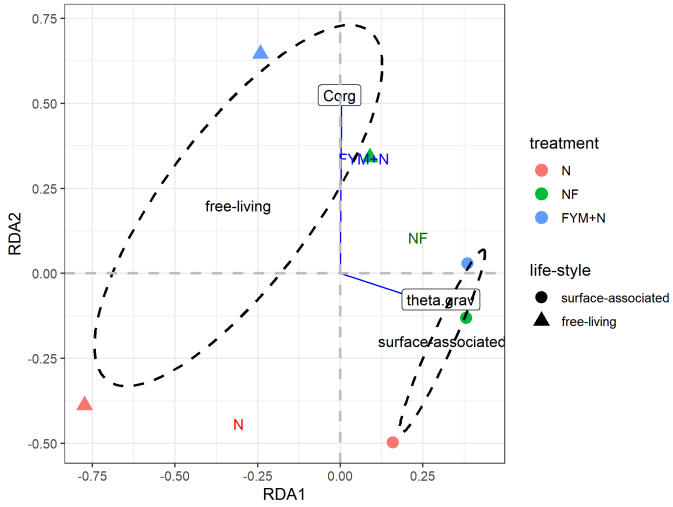

Figure 39: Redundancy analysis based on weighted unifrac distances for active communities. Replicates represent technical replicates. organic Carbon (C_{org}) and gravimetric water content (theta.grav) are displayed as blue arrows. Arrow length indicates goodness of fit and the pointing direction represents the direction of maximum change. Bacterial lifestyle (free-living, surface-associated) and fertilization are factor constraints, displayed as centroids with the respective 95 % confidence interval.

3.3.5 Taxonomic composition of the total community of free-living and surface-associated bacteria

The taxonomic compositions of RNA communities were examined to reveal pattern, enrichment or depletions of taxonomic groups in certain communities associated with bacterial life-style and fertilization, since DNA data are prone to shifts in community composition caused by DNA from soil originating from dead or inactive cells, thus biasing the soil bacterial community (Blazewicz et al., 2013; Blagodatskaya and Kuzyakov, 2013).
Members of the phyla *Actinobacteria* and *Proteobacteria* dominated the free-living and surface-associated communities (Fig. 40). The free-living communities contained 28.9-47.9 % and 41.6-61.5 % of relative abundance for *Actinobacteria* and *Proteobacteria*. For the surface-associated communities 43.6-49.2 % and 25.2-29.7 % were contained in *Actinobacteria* and *Proteobacteria*. This emphasizes the depletion in *Proteobacteria* for the surface-associated communities across all fertilization treatments compared to the free-living community. However, this depletion was not caused by a decrease in *Alphaproteobacteria*, which were the largest group for the free-living and surface-associated communities, but *Gammaproteobacteria* and *Deltaproteobacteria*, reduced by more than half of their relative abundance respectively. Moreover, the surface-associated communities were enriched in phylum *Actinobacteria* by more than 10 %, driven by increased abundances of *Acidimicrobiia* and *Thermoleophilia*. Besides, abundant members of the phylum *Chloroflexi*, represented by class *Chloroflexia*, increased in relative abundance for all fertilization treatments. Generally, the distribution of taxonomic groups was similar between fertilization treatments for the surface-associated community.
As mentioned before, the free-living communities were depleted in *Actinobacteria*. In contrast to the surface-associated communities, *Actinobacteria* were dominated by members of the class *Actinobacteria* with 18.7-25 % relative abundance across all fertilization treatments. The NF free-living community was specifically enriched in in phylum *Actinobacteria* by more than 10 % represented by class *Actinobacteria*, *Acidimicrobiia* and *Thermoleophilia*. For all free-living *Proteobacteria*, *Alphaproteobacteria* provided the majority of reads with 30.7-34.4 % relative abundance. Moreover, *Gammaproteobacteria* were enriched in the free-living communities of N and FYM+N.

In more detail, examining the taxonomic distribution of *Actinobacteria* of the rRNA communities down to family level, *Ilumatobacteraceae*, *Gaiellaceae* and *Pseudonocardiaceae* accounted for the observed enrichment of phylum *Actinobacteria* in NF for the free-living community compared to N and FYM+N (Fig. 41). Similarly, the free-living community of NF appeared to be of similar composition than the respective surface-associated community. Furthermore, N was enriched in *Streptomycetales*. Generally, *Streptomycetaceae*

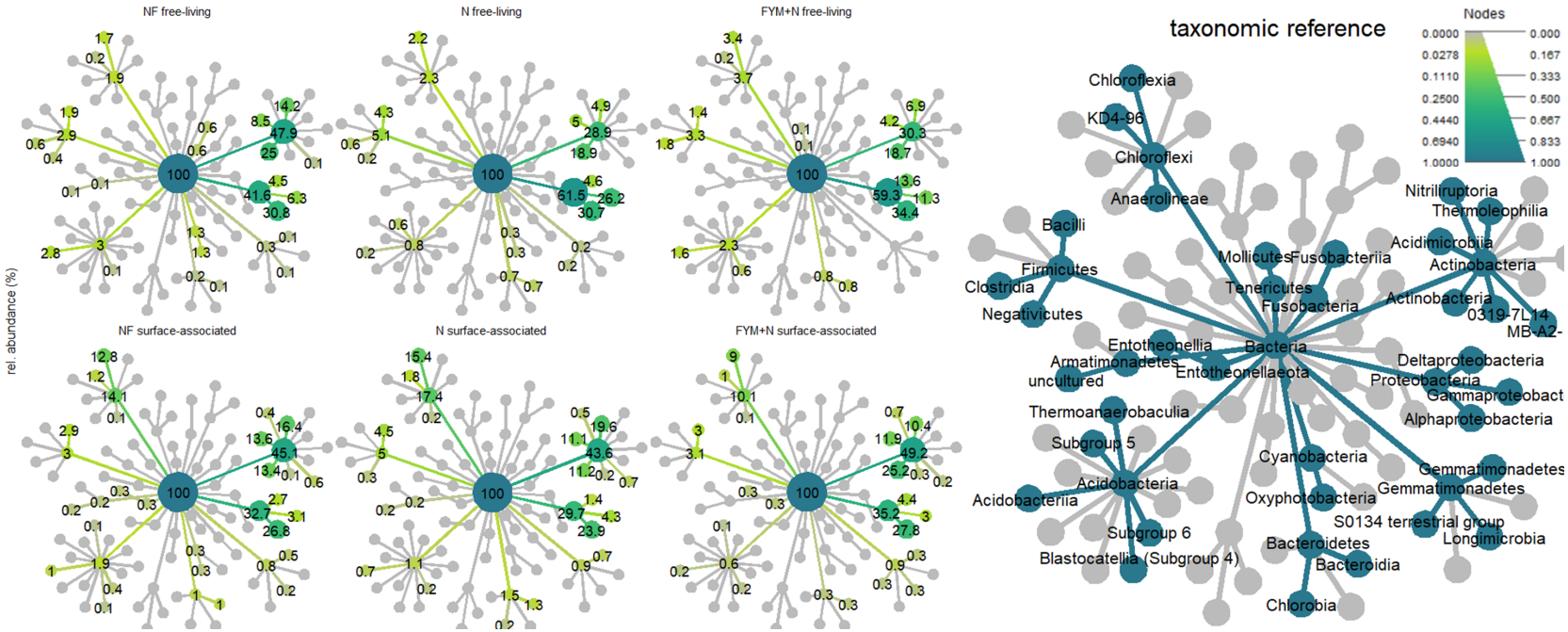

Figure 40: Taxonomic trees showing the relative abundance (%) of rRNA based community on phylum and class levels. The upper row of subplots represents the free-living bacterial community for each fertilization treatment. The lower row of subplots displays the surface-associated community for each fertilization treatment. NF represents no fertilization, N stands for mineral nitrogen and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of each taxonomic group.

made up the largest proportion of class *Actinobacteria* for the free-living community.
As already pointed out, the surface-associated communities extracted higher fractions of *Actinobacteria* from soil. Moreover, *Ilumatobacteraceae* of class *Acidimicrobiia* were enriched for all fertilization treatments as well as *Solirubrobacteraceae*. Moreover, FYM+N was enriched in *Streptosporangiales* and *Streptomycetales*.
A closer look at the *Proteobacteria* showed that for the free-living and surface-associated communities, *Rhizobiaceae* and *Xanthobacteraceae* of the class *Alphaproteobacteria* were the largest taxonomic group (Fig. 42). However *Rhizobiales* were reduced by 5.8-11.5 % for the surface-associated community. Also, *Gammaproteobacteria* were depleted for the surface-associated communities. However, due to the homogeneous abundance distribution no taxonomic group could be identified being responsible for that shift since *Gammaproteobacteria* were comprised of a large variety of families of low abundance. N was enriched of *Burkholderiaceae* in the free-living community. *Myxococcales* were enriched for FYM+N. Besides, the observed increased proportion of *Chloroflexi* for the surface-associated community was attributed to a preferential extraction of *Thermomicrobiales* (*JG30-KF-CM45*) (Fig. 43). Also class *KD4-96* was enriched with an entirely uncultured family.

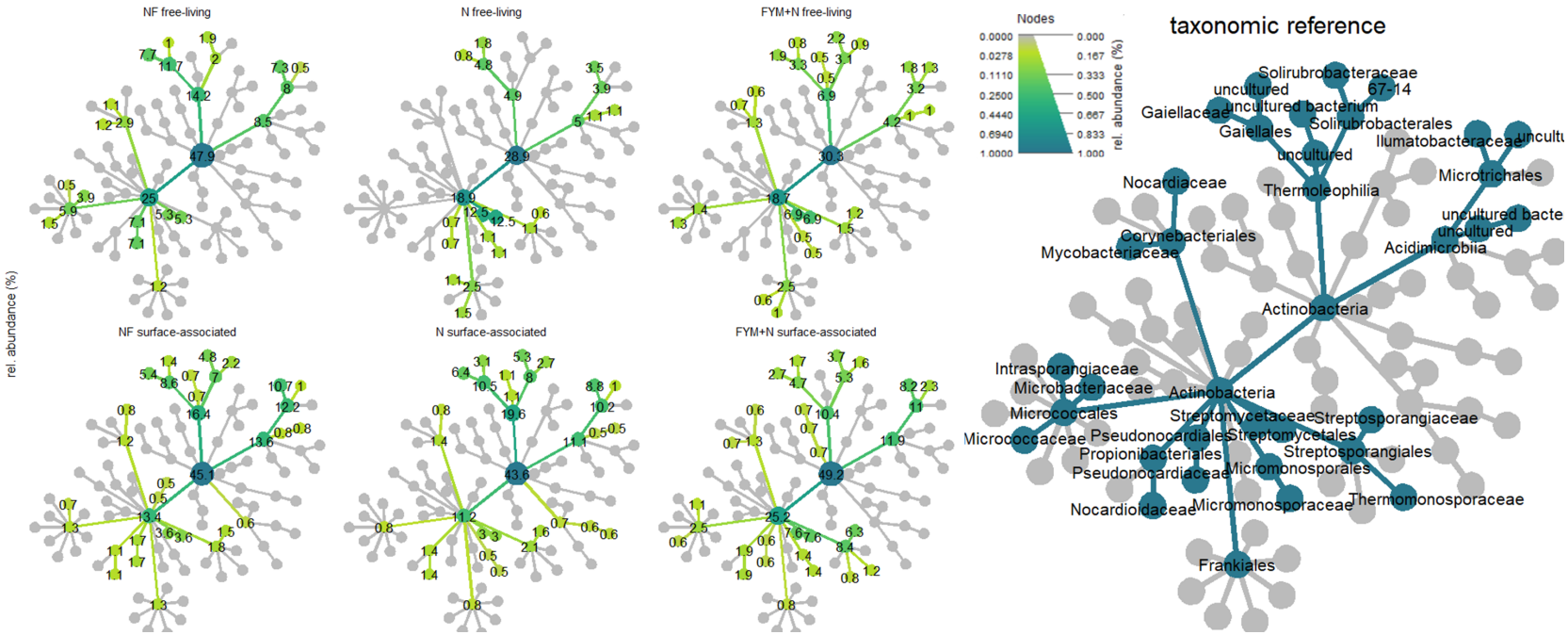

Figure 41: Taxonomic trees showing the relative abundance (%) of rRNA based data showing class, order and family level of phylum *Actinobacteria*. The upper row of subplots represents the free-living communities. The lower row of subplots displays the surface-associated communities. NF, N and FYM+N stands for different fertilization treatments of no fertilization, mineral N fertilization and farmyard manuare in combination with mineral N.The relative abundance of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed.

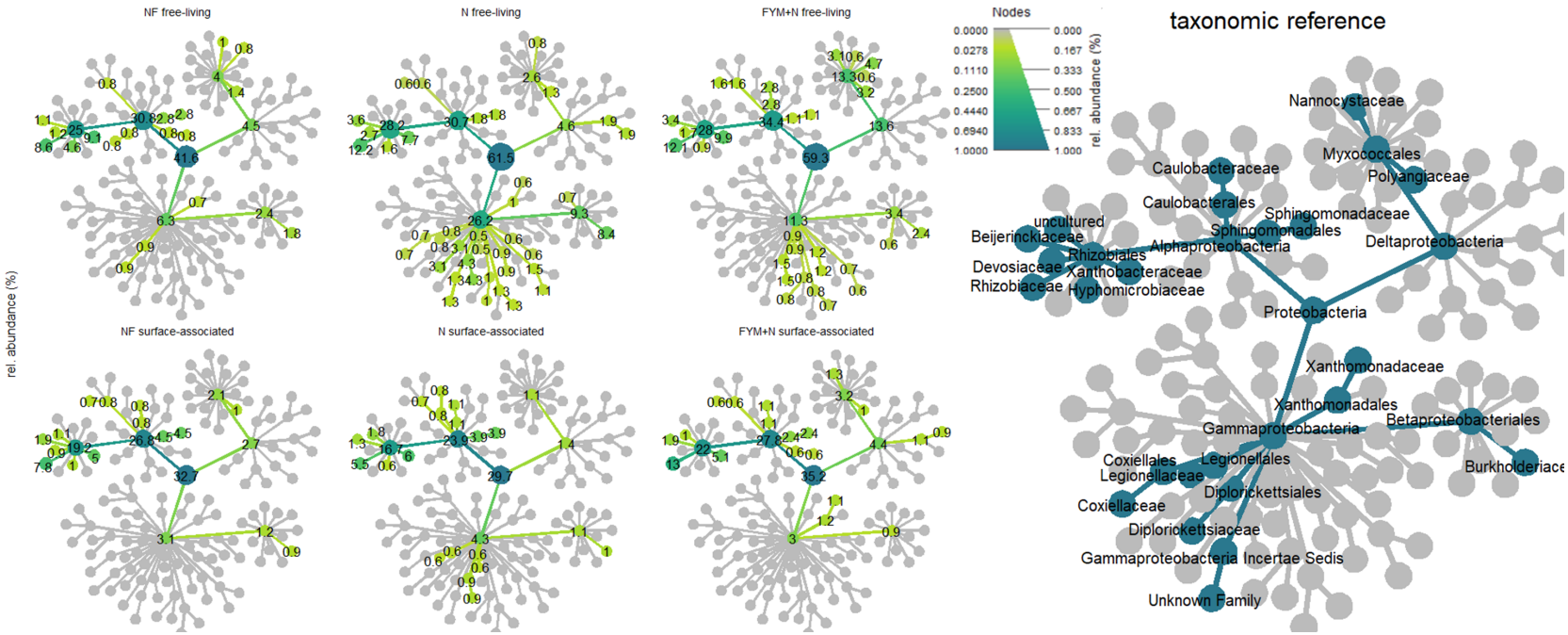

Figure 42: Taxonomic trees showing the relative abundance (%) of rRNA based data showing class, order and family level of phylum *Proteobacteria*. The upper row of subplots represents the free-living communities. The lower row of subplots displays the surface-associated communities. NF, N and FYM+N stands for different fertilization treatments of no fertilization, mineral N fertilization and farmyard manuare in combination with mineral N.The relative abundance of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed.

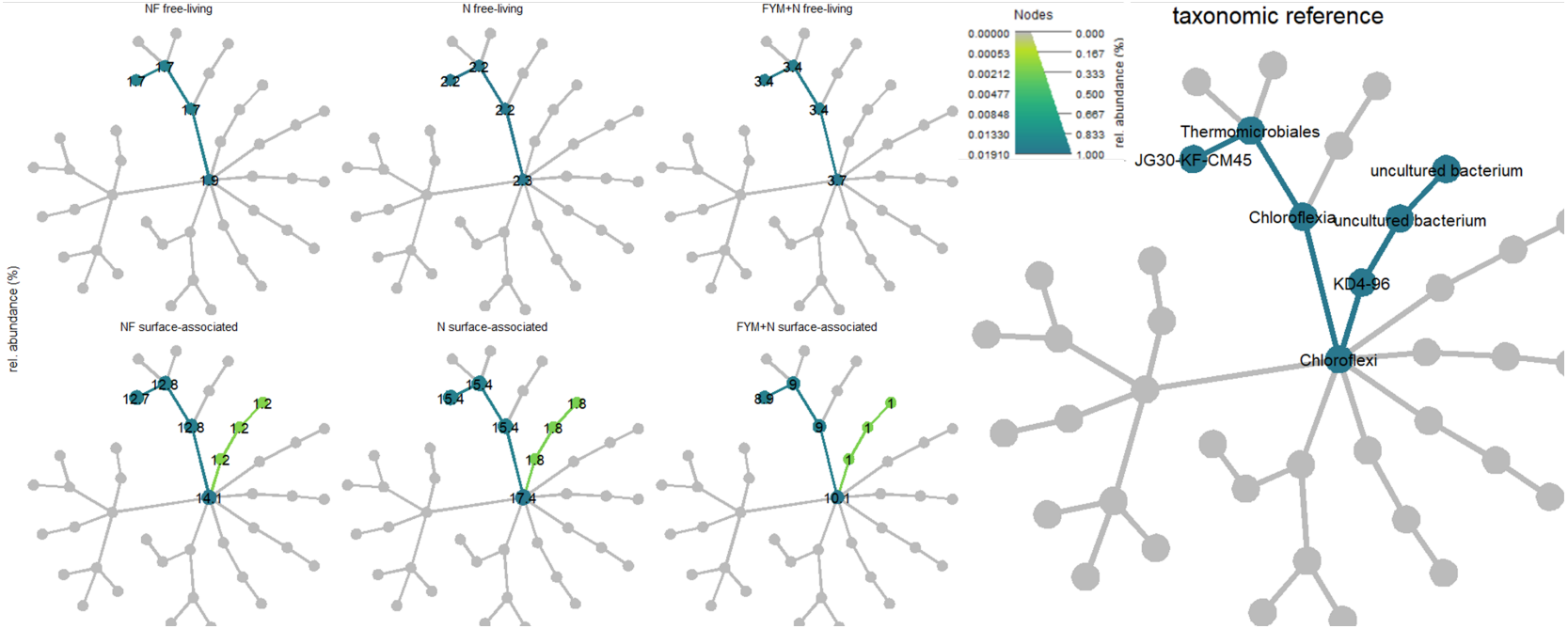

Figure 43: Taxonomic trees showing the relative abundance (%) of rRNA based data showing class, order and family level of phylum *Chloroflexi*. The upper row of subplots represents the free-living communities. The lower row of subplots displays the surface-associated communities. NF, N and FYM+N stands for different fertilization treatments of no fertilization, mineral N fertilization and farmyard manuare in combination with mineral N.The relative abundance of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.\,\%$ are not displayed.

3.3.6 Taxonomic composition of the active free-living and surface-associated communities

In this study, the status of potential metabolic activity for each SV was determined through the rRNA:rDNA ratio.. The relative abundances of active SV was based on the rRNA community data, since some active SV occurred only in rRNA communities. Furthermore, DNA based communities are biased by free DNA outlasting in soil.
I hypothesised that the proportion of active bacteria increases with fertilization and to be higher in the respective surface-associated communities. Comparing the proportion of active SV to the total number of SV and the relative abundance of active SV to all bacteria, the free-living communities contained higher proportions of active SV (Tab. 6). Focussing first on the relative number of active SV, the proportion was generally small, ranging from 1.5 to 7.3 %. NF had the highest proportion of active SV in the free-living and surface-associated community. Second, the relative abundance of active SV was quite similar between fertilization treatments and did not follow the trend indicated by the proportion of active SV to the total number of species. The treatment fertilized with mineral N had the highest relative abundance for all treatments for both communities.

Table 6: Proportions of active SV counts in the free-living and surface-associated RNA communities across fertilization treatments. The left column describes the proportion of the number of active SV to the total number of SV in a sample. The right column shows the relative abundance of active SV based on the number of sequence reads.

		prop. active SV (%)	rel. abund. of active SV (%)
free-living	NF	7.3	63.3
	N	4.1	73.3
	FYM+N	4.9	65.4
surface-associated	NF	3.7	29.9
	N	1.5	44.7
	FYM+N	2.3	23.8

Looking at the taxonomic composition of active SV, phylum *Actinobacteria* and *Proteobacteria* dominated the communities (Fig. 44). This pattern was already observed in the previous section for the total communities comprised of all SV independent of activity. In the free-living community, *Proteobacteria* was the most abundant phylum. *Alphaproteobacteria* were most abundant followed by *Gamma-* and *Deltaproteobacteria*. However, the proportion of *Proteobacteria* differed between fertilization treatments. N contained with 71.9 % the highest proportion of *Proteobacteria*, equally distributed between *Alpha-* and *Gammaproteobacteria*, followed by FYM+N with 58.8 % which was enriched in *Deltaproteobacteria* compared to the other treatments. NF contained only 35.6 % *Proteobacteria*

of mainly *Alphaproteobacteria* and in consequence was depleted in *Gammaproteobacteria*. However, NF was enriched in active *Actinobacteria*. *Actinobacteria* represented the predominant class with *Acidimicrobiia* and *Thermoleophilia* showing less relative abundance. N and FYM+N contained less *Actinobacteria*, with class *Actinobacteria* being the most abundant. Furthermore, N was enriched in *Firmicutes* of class *Bacili*.

As expected from the previous results, the active surface-associated communities consisted predominantly of *Actinobacteria* and *Proteobacteria* for all fertilization treatments. *Actinobacteria* contained mostly members of the class *Actinobacteria*, *Acidimicrobiia* and *Thermoleophlia*, which were enriched compared to the free-living communities. Moreover, the relative abundance of *Proteobacteria*, especially *Gammaproteobacteria*, were depleted for all fertilization treatments. Nonetheless, *Alphaproteobacteria* were the predominant class and comprised similar abundances for all fertilization treatments compared to the free-living community. Unlike the free-living communities, the taxonomic composition was more similarly distributed between fertilization treatments. Only FYM+N differed, being enriched in *Alphaproteobacteria* and class *Actinobacteria* by roughly 10 % compared to NF and N. In contrast, *Thermoleophilia* were reduced for FYM+N. Generally, the observed enrichment of *Chloroflexia* for the surface-associated communities was confirmed. Moreover, active *Firmicutes* of the class *Bacilli* were enriched for NF in the surface-associated community and for N in the free-living and surface-associated community.

Having a closer look at lower taxonomic levels, the free-living active RNA community composition for phylum *Actinobacteria* (Fig. 45), the fertilization treatments differed in taxonomic composition. For NF, *Microbacteriaceae*, *Pseudonocardiaceae*, *Gaiellaceae* and *Ilumatobacteraceae* were enriched in the free-living community. N was generally depleted in *Actinobacteria*.

In the surface-associated community, *Solirubrobacteraceae* were enriched for all fertilization treatments. For class *Acidimicrobiia*, *Microtrichales* were predominant. Comparing the different fertilization treatments with each other, FYM+N was enriched in *Streptomycetaceae*. NF was again enriched in *Pseudonocardiaceae* as seen for the free-living community.

In phylum *Proteobacteria* (Fig. 46), all communities showed that the high relative abundance of *Alphaproteobacteria* was represented by order *Rhizobiales* and more specifically *Rhizobiaceae*, especially for N fertilization. Moreover, active *Gammaproteobacteria* and *Burkholderiaceae* were enriched in the free-living communities of N and FYM+N. Besides, FYM+N was enriched in *Myxococcales* of class *Deltaproteobacteria*.

Even though the surface-associated community contained lower proportions of *Proteobacteria*, *Alphaproteobacteria* and *Rhizobiales* remained the most abundant group and *Rhizobiaceae* were enriched in relative abundance for NF and FYM+N.

Active *Chloroflexi* were represented by *Thermomicrobiales* of family *JG30-KF-CM45* for

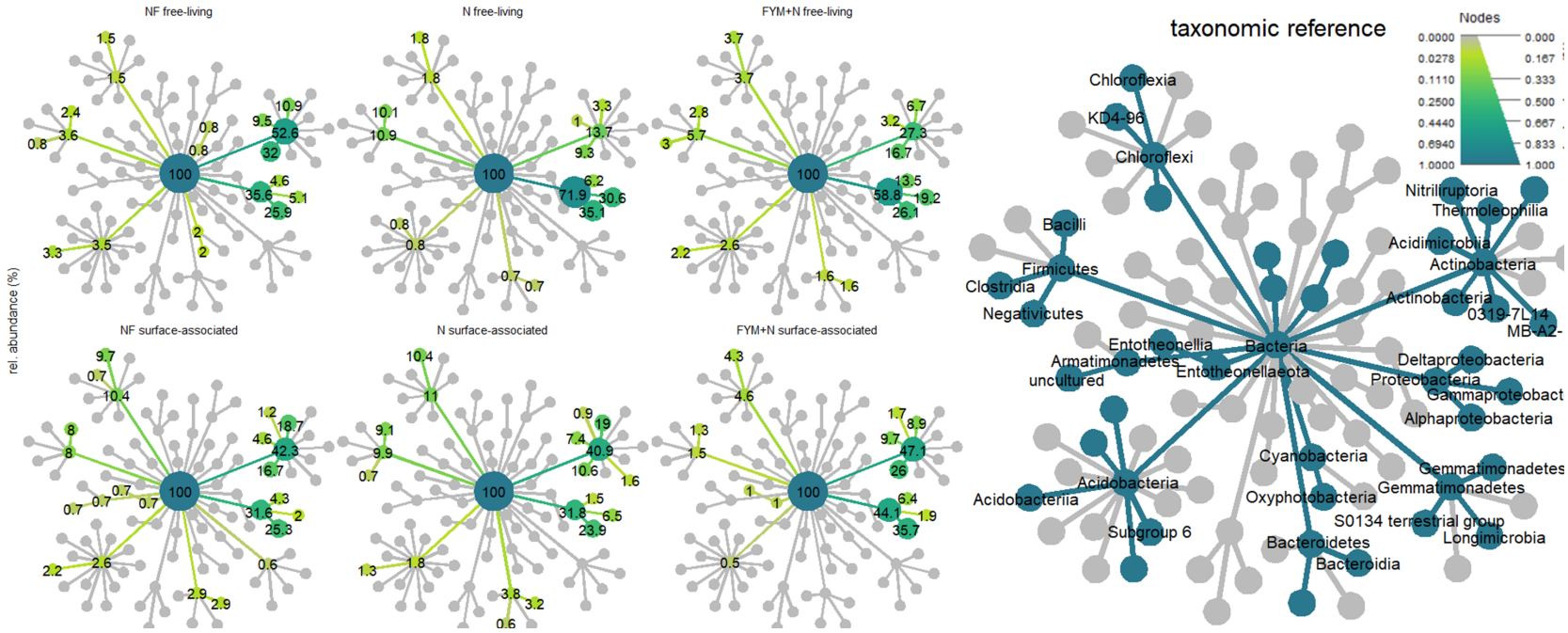

Figure 44: Taxonomic trees showing the relative abundance (%) of active SV in the rRNA based community on phylum and class levels. The upper row of subplots depicts the free-living community for each fertilization treatment. The lower row of subplots showss the surface-associated community for each fertilization treatment. NF represents no fertilization, N is mineral nitrogen and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance (%) of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of taxonomic groups.

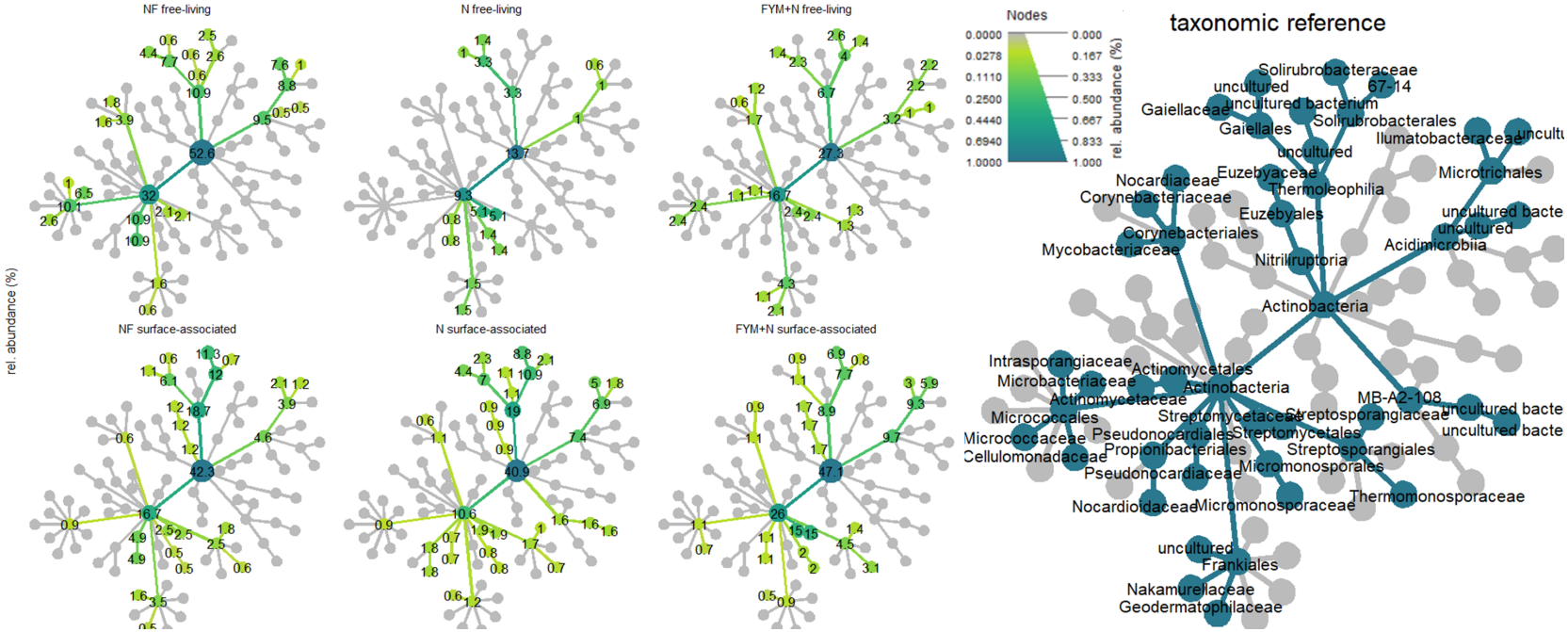

Figure 45: Taxonomic trees showing the relative abundance (%) of active SV in the rRNA based community on class, order and family level for phylum *Actinobacteria*. The upper row of subplots represents the free-living community for the different fertilization treatments. The lower row of subplots displays the surface-associated community. NF represents no fertilization, N depicts mineral nitrogen fertilization and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance (%) of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of taxonomic groups.

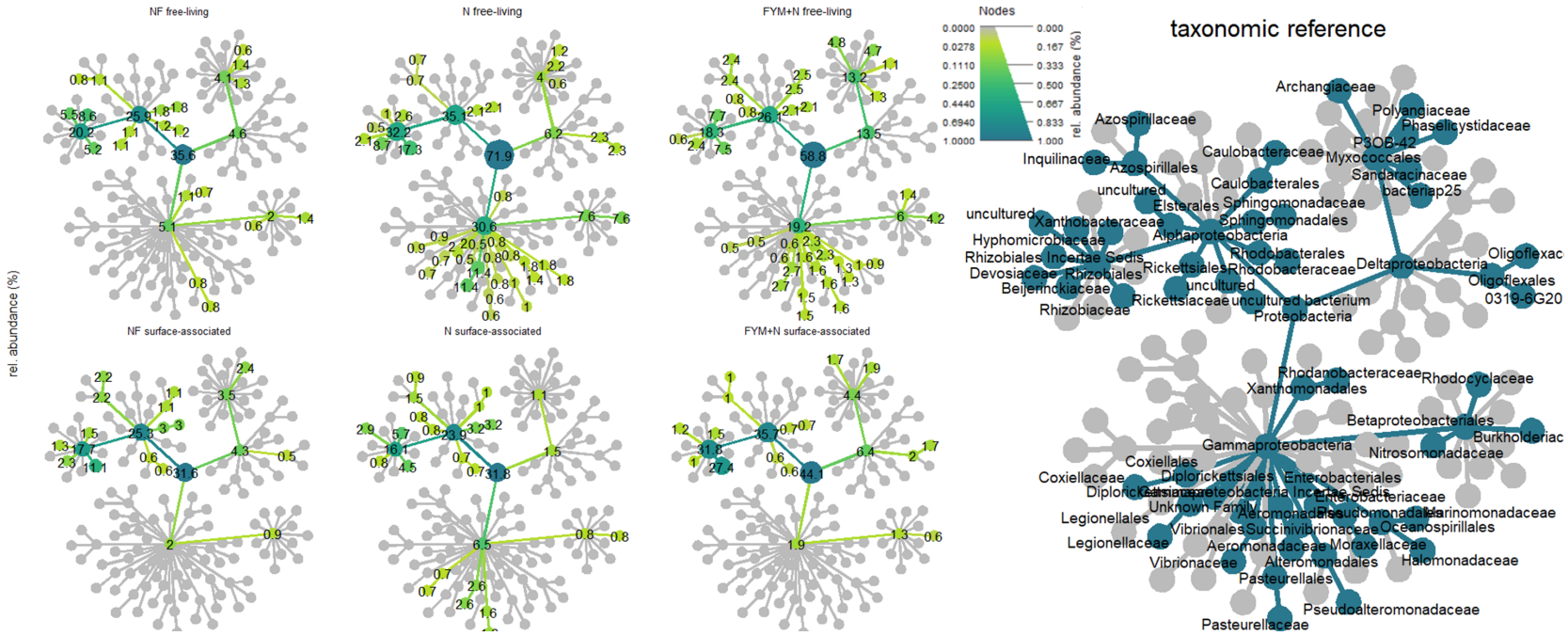

Figure 46: Taxonomic trees showing the relative abundance (%) of active SV in the rRNA based community on class, order and family level of phylum *Proteobacteria*. The upper row of subplots represents the free-living community for the different fertilization treatments. The lower row of subplots displays the surface-associated community. NF represents no fertilization, N depicts mineral nitrogen fertilization and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance (%) of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of taxonomic groups.

all communities and enriched in the surface-associated communities for NF and N. Lastly, active *Firmicutes* were enriched in *Bacillales* in the free-living community of N and the surface-associated community of NF and N.

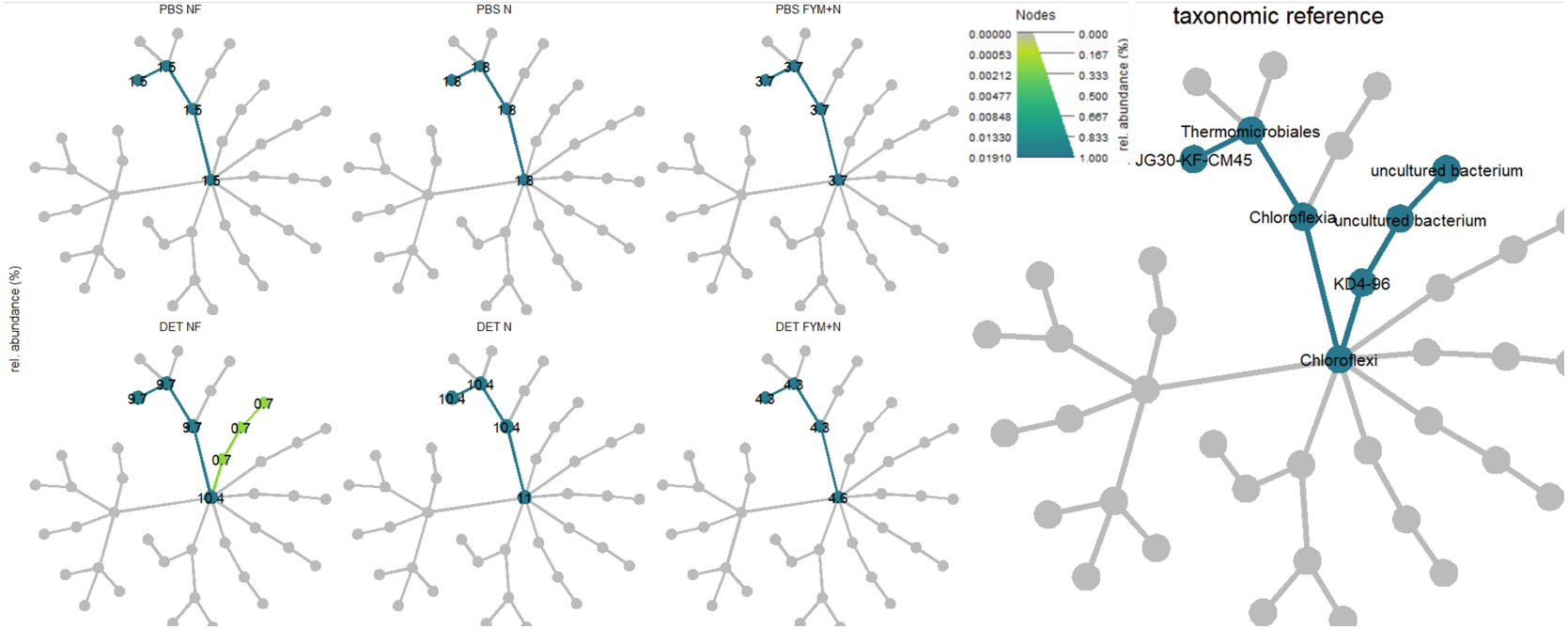

Figure 47: Taxonomic trees showing the relative abundance (%) of active SV in the rRNA based community on class, order and family level of phylum *Chloroflexi*. The upper row of subplots represents the free-living community for the different fertilization treatments. The lower row of subplots displays the surface-associated community. NF represents no fertilization, N depicts mineral nitrogen fertilization and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance (%) of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of taxonomic groups.

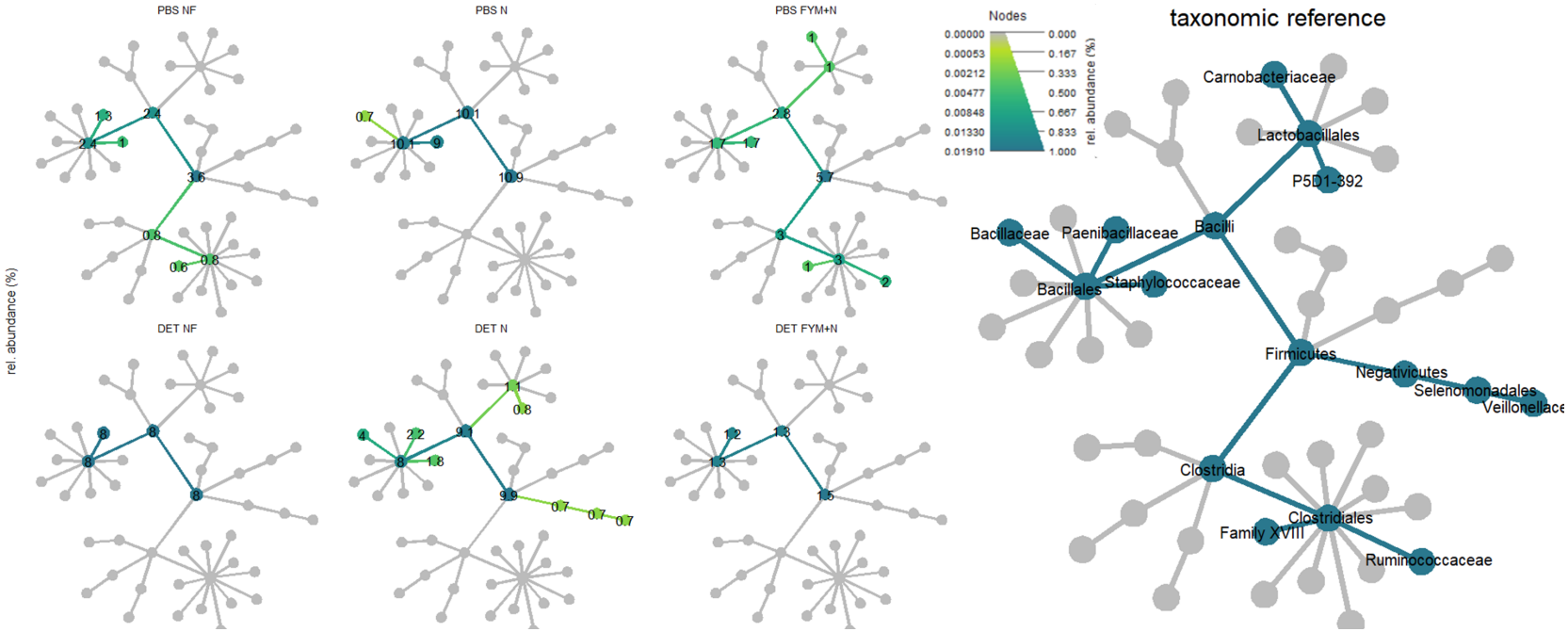

Figure 48: Taxonomic trees showing the relative abundance (%) of active SV in the rRNA based community on class, order and family level of phylum *Firmicutes*. The upper row of subplots represents the free-living community for the different fertilization treatments. The lower row of subplots displays the surface-associated community. NF represents no fertilization, N depicts mineral nitrogen fertilization and FYM+N is farmyard manure in combination with mineral nitrogen. The relative abundance (%) of each group is displayed by size and colour. Taxonomic groups having a relative abundance $< 0.1\,\%$ are not displayed. The reference tree displays the names of taxonomic groups.

3.4 Discussion

3.4.1 Evaluation of cell extraction of free-living and surface-associated bacteria from soil

In this study, a cell extraction method was established to separate free-living and surface-associated bacteria from soil using Nycodenz gradient centrifugation. This method was evaluated, considering the amount of cells extracted, the contamination level with bacteria on particles remaining in the extracted cells, the similarity between replicates and the removal of contaminating sequences in the final bacterial communities.

As expected, PBS extracted less cells than detergent extraction, since no detergents except ethanol, used as fixation agent, were used. Moreover, free-living bacteria in the soil pore space are assumed to be lower in number by orders of magnitude (Flemming and Wuertz, 2019). However, 5 consecutive extractions were necessary for the free-living community to extract a sufficient number of cells for a successful library preparation. The observation, that multiple extractions with PBS, also increased the number of extracted cells with detergents, is probably related to a better aggregate dispersion due to a longer aggregate disruption time while shaking.

Furthermore, around 10^4 bacteria still attached to particles were co-extracted with the PBS and detergent extractions, contaminating the extractable free-living and surface-associated communities (Fig. 29). However, the proportion of bacteria on particles represented only 0.7 % and 0.07 % of extracted cells for the free-living and surface-associated community, respectively. Therefore, the anticipated bias introduced by bacteria on particles was considered insignificant. Also, no differences in cell extraction efficiency between fertilization treatments was observed (Fig. 30).

In general, the applied cell extraction method produced good reproducibility between technical replicates. Beta diversity ordination showed a narrow grouping for technical replicates of each sample type (Fig. 38). Moreover, ANOVA confirmed that the variance between technical replicates was significantly lower than the variation between groups. This was supported by alpha diversity analysis, showing low variability between replicates for abundant SV. The larger variation for richness between replicates was expected, as any SV being present in a sample represented by even a single read influences richness similar to any abundant SV. This behaviour is well described in literature (Bickel and Or, 2021).

In conclusion, the applied cell extraction method together with the library preparation was capable of reproducing and differentiating between different bacterial communities. Nevertheless, the achieved cell extraction efficiency of 0.02-0.02 % was low in comparison to other published Nycodenz cell extractions (Maron et al., 2006; Kallmeyer et al., 2008; Lindahl, 1996; Courtois et al., 2001). However, a direct comparison to literature is difficult, since extraction protocols have to be adapted for each soil (Ranjard and Richaume,

2001) and quantitative and qualitative cell recovery are not necessarily related (Maron et al., 2006). Besides, most literature data used old-fashioned dyes to count bacteria in soil as a reference for cell extraction efficiency, achieving comparably low cell counts in soil (Lindahl and Bakken, 1995; Kallmeyer et al., 2008) and therefore indirectly increasing the extraction efficiency.

3.4.2 Nitrogen fertilization and a free-living bacterial life-style reduce bacterial alpha diversity

Bacterial diversity pattern within samples were estimated by richness and the Simpson index, representing rare and highly abundant bacteria. The free-living communities contained much lower alpha diversity in richness and Simpson for rRNA and rDNA samples than the respective surface-associated communities. This was expected, since the majority of bacteria in soils is attached to mineral/organic surfaces or living in biofilms (Flemming and Wuertz, 2019). Moreover, the alpha diversity of free-living and surface-associated communities was composed of a small number of high abundant SV, estimated by Simpson and a high number of rare SV estimated by richness. This abundance distribution pattern, with the vast majority of SV being rare and only a minor number of species being abundant, is commonly reported for bacteria in soil (Bickel and Or, 2021; Magurran and Henderson, 2003; Delgado-Baquerizo et al., 2018). As a consequence, the applied cell extraction method was capable of reproducing this pattern without bias introduction or over-representation of abundant SV.

RNA samples showed a lower richness for both the free-living and the surface-associated communities compared to DNA samples. This pattern in well documented (Li et al., 2021) and most likely affected by sample processing after cell extraction. RNA is prone to degradation by RNAses (Alm and Stahl, 2000). Especially digestion of genomic DNA at elevated temperatures during sample preparation posed a risk to RNA degradation. This is critical especially for diverse, but low biomass samples as used in this study, since even small losses of rRNA might change the examined diversity of the communities considerably. On the other hand, next generation sequencing of rDNA based samples overestimate the diversity in soil caused by co-extraction of free DNA from dead bacteria (Nielsen et al., 2007) and representation of mostly dormant/inactive bacteria (Blagodatskaya and Kuzyakov, 2013). Therefore, bacterial communities based on 16S rRNA provide less diverse communities, but give a more stringent picture of the potentially protein synthesizing soil bacteria (Blazewicz et al., 2013).

The comparison of alpha diversity between fertilization treatments showed that N fertilization contained in tendency lower alpha diversity. However, this effect was not significant.

Many recent studies showed that long-term mineral N fertilization reduced alpha diversity (Zeng et al., 2016; Li et al., 2016; Dai et al., 2018) and microbial biomass (Wallenstein et al., 2006). This effect, however, was attributed to soil acidification induced by mineral fertilization. In contrast, some reported an increase in bacterial alpha diversity (Wei et al., 2008; Xu et al., 2020) and others found no effect (Ge et al., 2008; Ogilvie et al., 2008; Leff et al., 2015; Duan et al., 2022; Feng et al., 2022). Furthermore, it has to be considered that the NF treatment received up to 60 kg N ha^{-1} annually as anthropogenic wet deposition (Merbach and Körschens, 2002), biasing the fertilization estimation, especially for the NF control. In general, the high fertility of the chernozem soil in Bad Lauchstädt most likely diminished fertilization effects. Nonetheless, the observed decline in alpha diversity for the mineral N fertilization treatment in this study is no indirect effect of soil pH, since all plots had constant pH, therefore supporting the observation that mineral N fertilization decreases alpha diversity.

In contrast to mineral N fertilization, organic fertilization is reported to increase richness in soils (Hartmann et al., 2015; Francioli et al., 2016; van der Bom et al., 2018). This pattern was not observed comparing the free-living and surface-associated communities of different fertilization treatments. However, the positive effect of organic fertilization in literature is again compromised by alterations in soil pH (e.g. Wei et al., 2008; Hartmann et al., 2015). Moreover, literature data refer to bulk soil communities and do not differentiate sol bacteria by life-style, impeding a direct comparison of results to literature data.

3.4.3 Bacterial life-style dominates beta diversity of the total community

The analysis of beta diversity for the total community, containing all SV irrespective of activity status, confirmed significant effects of nucleic acid type, bacterial life-style and fertilization on bacterial communities. The free-living communities were enriched in *Proteobacteria* and depleted in *Actinobacteria* and *Chloroflexi* compared to the surface-associated communities. In more detail, *Proteobacteria* in the free-living communities were enriched in a variety of *Burkholderiaceae*. Many type strains summarised in *Burkholderiaceae* are motile (Coenye, 2014). Hence, the reduction in the surface-associated community might result from motile members of this groups preferably extracted in the free-living community.

The surface-associated communities of *Actinobacteria* showed increased relative abundances of *Solirubrobacteraceae*. Described type strains of *Solirubrobacter* comprise only non-motile generalist (Singleton et al., 2003; Kim et al., 2007; Zhang et al., 2014). Moreover, gradientForest analysis revealed that 5 SV of family *67-10* of order *Solirubrobacterales* were associated with bacterial life-style. Together with the enrichment in the surface-associated communities, this indicated a systematic association to a surface-associated

life-style. The dominance of phylum *Actinobacteria* for the surface-associated community was confirmed by Francioli et al. (2016). In the latter publication bulk soil communities were dominated by *Actinobacteria*, emphasizing that surface associated communities are more closely related to the bulk soil community.

Furthermore, *Choloroflexi* of the family *JG30-KF-CM45* were enriched in the surface-associated communities. Unfortunately, no cultivated representatives with known lifestyle and biology are available (Lladó Fernández et al., 2019). However, the superordinate class of *Thermomicrobiales* were found in other arables soils and are presumably involved in nitrogen cycling (Rummel et al., 2020). This was further supported by 3 SV of class *JG30-KF-CM45* which were closely related to a surface-associated life-style by gradientForest analysis. Furthermore, *KD4-96* was enriched in the surface-associated community and was found to correlate with iron (Fe) and aluminium (Al) concentration in soil (Wegner and Liesack, 2017). Since Fe and Al concentrations are generally low in the soil solution, this reported correlation is in accordance with a surface-associated lifestyle and a presumed role in Fe cycling in soil.

RDA ordination separated FYM+N from the other fertilization treatments, along the fertilization gradient emphasizing the effect of organic fertilization on community composition. In general, organic and N fertilization are well known for affecting beta diversity (Ramirez et al., 2012; Leff et al., 2015; Hartmann et al., 2015; Zheng et al., 2021). However, the variance explained by fertilization was small in this study (Tab. 4). Hence, differences in taxonomic composition were small between fertilization treatments. This was confirmed by Francioli et al. (2016) who found only little differences between bulk soil community composition of fertilization treatments. *Streptosporangiales* and *Streptomycetales* from phylum *Actinobacteria* and *Myxococcales* from phylum *Proteobacteria* were enriched for FYM+N. *Streptomycetales* were shown to be involved in the initial decomposition of organic material (Chater et al., 2010). This is supported by the decrease in relative abundance in the free-living communities.

Overall, taxonomic groups were shared between communities. This is in line with the results of Bystrianský et al. (2019) and Hujslová et al. (2020) who identified genera being significantly enriched, but shared between the free-living or surface-associated community in soil. In conclusion, bacterial life-style induced systematic changes in taxonomic composition of the total communities across distinct fertilization treatments, confirming the anticipated effect of bacterial life-style on community composition. Also, the effect of fertilization was significant, however, the effect on community composition was small and no systematic shifts picturing the N gradient between fertilization treatments was observed.

3.4.4 Beta diversity of active bacteria is tighter bound to bacterial life-style and fertilization than the total communities

Activity of soil bacteria was determined by applying a dynamic threshold based on rRNA to rDNA ratio to bacteria of the free-living and surface-associated community. Therefore, active bacteria represent a subset of rRNA communities. I hypothesized the taxonomic differences between different fertilization treatments to increase for active bacteria. Furthermore, I expected the proportion of active SV to increase with fertilization and a surface-associated life-style.

It turned out, that active bacteria represented mostly abundant SV. Especially the free-living community was dominated by active SV in all fertilization treatments. This is in contrast to the anticipated higher activity for the surface-associated community as hypothesised by Flemming and Wuertz (2019). It can be ruled out that the increased activity of free-living bacteria is a consequence of bacterial growth or activation during sample processing since soil bacteria were fixed with ethanol immediately while sampling. Furthermore, the anticipated increase in activity with increasing fertilization, i.e. nutrient availability, was not confirmed. Surprisingly, N fertilization had the highest relative abundance of active bacteria in both the free-living and surface-associated community compared to NF and FYM+N. A potential explanation is the observed lack of available P for the N treatment (Tab. 2) which induced a delayed ripening of wheat (Dr. Thomas Reitz, personal communication). This lead to longer plant activity for the N treatment in comparison to NF and FYM+N, where the crops were already ripe and dry on sampling date.

The beta diversity of active communities differed significantly only by bacterial life-style as revealed by variance partitioning (Tab 5) and RDA (Fig. 39). However, the variance explained increased for life-style, fertilization and water content in comparison to the total community, indicating a more stringent coupling of active bacteria to environmental predictors. Besides, the lack of significant effects for fertilization and soil water content for the active community was most likely dedicated to a reduced sample size for the active communities having only 6 samples to compare instead of 12 samples for the total communities. Additionally, the RDA of active communities showed a clear separation between fertilization treatments. N samples were distinct from NF and FYM+N. This finding is supported by alpha diversity analysis, which indicated a lower alpha diversity for N samples compared to NF or FYM+N.

However, taxonomic differences between N and other fertilization treatments were small. Members of the order *Bacillales* in the phylum *Firmicutes* were increased in relative abundances for the free-living and surface-associated community for N. This in in line with literature (Ramirez et al., 2012; Francioli et al., 2016; Dai et al., 2018). Unfortunately, this enrichment was not seen for the FYM+N treatment but NF, indicating a negative

connection to organic fertilization. In contrast, *Firmicutes* were enriched in the bulk soil community in Bad Lauchstädt (Francioli et al., 2016). However, this report might be biased by pH, since plots with varying soil pH were compared.

The lack of substantially diverging bacterial communities mapping the gradient in soil nutrient availability between the fertilization treatments for the free-living and surface-associated communities was not expected, since it is well documented that community composition change and microbial biomass and enzyme activities increase with fertilization (Tab. 2) (e.g. Klimanek, 2000; Kautz et al., 2004; Chakraborty et al., 2011). Apparently, this does not imply systematic shifts in taxonomic composition of bacterial communities. Neumann et al. (2013) demonstrated that in Bad Lauchstädt 70 % of the bacterial population lived in the clay fraction, which was not affected by the long-term fertilization gradient after more than 100 years. However, significant changes in the bulk soil community were reported for the non-limed version of the fertilization experiment (Francioli et al., 2016; Böhme et al., 2005). This is affected by a pH gradient between treatment plots, since pH is a major driver shaping soil bacterial communities (Jiao and Lu, 2020; Karimi et al., 2018; Kaiser et al., 2016). Albeit, this does not apply to the experimental setup in this study, since plots of constant pH were compared. Also, the co-extraction of bacteria on particles together with the free-living and surface-associated communities is not considered responsible for the similarities in community composition since these bacteria represented only between 0.7-0.07 % of extracted cells.

Possibly, the differences in edaphic properties between treatment plots were too small to generate fundamental shifts in the extractable community using a Nycodenz-based cell extraction method. More than 100 years of different fertilization regimes resulted on the limed plots in a difference of 0.79 % and 0.8 % for C_{org} and N_t respectively (Tab. 2).

3.4.5 Determination of active bacteria

In this study, activity of bacteria is deduced by calculating the ratio of rRNA:rDNA reads per SV per sample and using a dynamic abundance threshold separating active from non-active SV. This concept is based on the assumption that high cellular amounts of ribosomes and consequently rRNA indicates a protein synthesis potential. This approach is well established and used for a long time already (e.g. Bowsher et al., 2019; Loeppmann et al., 2018; Kerkhof and Ward, 1993; Denef et al., 2016). It was also found to be consistent with staining methods inducing activity (Bowsher et al., 2019). However, some limitations have to be discussed. It was shown that rRNA content and growth rate do not necessarily scale linear (Mandelstam and Halvorson, 1960; Kerkhof and Ward, 1993; Worden and Binder, 2003) and dormant cells can have increased amounts of ribosomes as well (Sukenik

et al., 2012). Furthermore, some studies use a constant threshold of 1 for rRNA:rDNA ratio data to define active species (Franklin et al., 2013), following the work of Jones and Lennon (2010). To reduce bias, we further developed this concept by applying a dynamic threshold based on statistical permutation to reduce the false discovery rate of active SV. This was necessary, since mostly rare SV were described as active in literature (Jones and Lennon, 2010; Campbell et al., 2011; Klein et al., 2016). However, this finding is based on an artefact inherent to data generation by amplicon sequencing. Rare SV, represented by a low number of reads, have a high variation in abundance of up to 80 % between technical replicates, accounting for variation introduced during library preparation (i.e. nucleic acid extraction, PCR and bioinformatics). On the contrary, abundant SV show a lower variation of 5 % (Noviana, 2021). Therefore, rare SV might be identified as active, based on a higher random variation. Applying the dynamic threshold, the false discovery rate could be reduced to 0.4 % in comparison to a Null-model, representing the stochastic variation in the data.

However, only a minor proportion of bacteria in soil are considered metabolically active (Blagodatskaya and Kuzyakov, 2013). The majority is potentially active or dormant. Potentially active bacteria can become active in only a few minutes. Dormant cells take longer to change their status (Blagodatskaya and Kuzyakov, 2013). Moreover, the response of bacteria to substrate addition, activating soil bacteria, was found to be soil texture dependent (Loeppmann et al., 2018). Therefore, sequencing a soil community yields a large number of SV which are not necessarily active, but dormant. Additionally, in case of rDNA based libraries, those communities harbour rDNA from dead cells persisting in soil (Carini et al., 2016). In contrast, sequencing of rRNA based communities removes dead cells biasing the community, since extracellular rRNA is labile and prone to degradation. Hence, rRNA based communities can be more reliably used to describe soil bacterial communities (Loeppmann et al., 2018). Nonetheless, rRNA-based communities do not equal active communities, because it was shown that metabolic activity and cellular rRNA content do not necessarily scale linear (Blazewicz et al., 2013). Moreover, the amount of rRNA in a cell is dependent on cell size which biases activity based on a rRNA:rDNA ratio in favour of large cells (Denef et al., 2016). Blazewicz et al. (2013) concluded that rRNA based communities can be considered as indicators of protein synthesis potential. Therefore, deducing activity of soil bacteria using a dynamic threshold of rRNA:rDNA ratios, is currently the only cost-efficient high throughput method available.

3.5 Conclusions

Little evidence is available how bacterial communities differ in soils between the soil pore space and surface associated habitats and how this difference in bacterial life-style is affected by nutrient availability. In this study, bulk soil bacterial communities from a long-term fertilization experiment were separated based on bacterial life-style by means of cell extraction from soil. This cell extraction method involved soil washing with PBS and detergents to extract free-living bacteria and surface-associated bacteria respectively. The applied cell extraction method based on Nycodenz gradient centrifugation managed to extract representative communities from soil. Furthermore, the influence of life-style and fertilization on bacterial activity was examined. Bacterial life-style was the most important factor shaping bacterial communities. Generally, free-living communities were enriched in *Proteobacteria*. The surface-associated community consisted of a more diverse community, enriched in phylum *Actinobacteria* and *Chloroflexi* of order *Thermomicrobiales*. In contrast, *Gammaproteobacteria* were depleted.

Cell counting and alpha diversity analysis revealed that less bacteria inhabited the soil pore space. However, in contrast to my expectations, the free-living community was dominated by active bacteria, which is opposed to theory, where most active bacteria inhabit solid surfaces and biofilms (Flemming and Wuertz, 2019). The effect of fertilization affected community composition, but induced higher variation in taxonomic composition of bacteria communities between fertilization treatments for the potentially active community. However, no consistent pattern in community composition was found mapping the increase in nutrient availability between fertilization treatments (NF<N<FYM+N). In tendency mineral nitrogen fertilization was reduced in alpha diversity, supporting the reported negative effect of mineral nitrogen fertilization on bacterial diversity, independent of soil pH. Many bacteria which differed between fertilization treatments were uncultivated. Hence, the results presented in this study provide the data basis for targeted analysis of active but yet uncultured bacteria and their role in soil nutrient cycling.

4 Synthesis and future perspectives

The work of my doctoral thesis presented here examined spatial and temporal variation in soil bacterial communities. I could show that soil bacterial alpha diversity is declining over time in grassland and forest ecosystems in Germany under recent climate conditions. In contrast, beta diversity was less affected by temporal variation and hence more stable over time. I could further show that climate change in the form of soil warming can in part explain this decline. This is the first evidence that ongoing climate change in Germany is already affecting bacterial diversity in soil. However, it was not possible to check if the observed decline is a consequence of past climate conditions >3 years preceding the soil sampling, restricted by climate data availability. The Biodiversity Exploratory network is a long-term project and future work should trace this time series further. Upcoming analysis has to reveal if the decline of soil bacterial alpha diversity is an ongoing trend or reaching a new steady-state. Hence, great care has to be taken to make future data analysis comparable to the existing dataset considering sequencing depth.

The analysis of the free-living and surface-associated community along a nitrogen gradient in a long-term fertilization experiment revealed the dominance of active bacteria in the free-living fraction. This is opposed to the expectations of free-living and particle-associated communities from marine environments or as expected from theory regarding biofilm forming communities in soil (Flemming and Wingender, 2010; Flemming and Wuertz, 2019). A more diverse sampling of soil and land use types, texture and nutrient availability, was cancelled in 2020 due to the restrictions of the Corona pandemic. Therefore, prospective follow-up experiments should examine if the obtained results in this study can be verified for bacterial grassland and forest communities in addition to the results obtained here for an arable soil. In the light of the restrictions of soil bacterial habitat connectivity due to the fragmentation of the soil aqueous phase (Wang and Or, 2010, 2013), the effect of soil water content on community composition and activity of the free-living and surface-associated community composition should be in focus.

Acknowledgements

First of all, I want to thank Prof. Dr. Jörg Overmann for the possibility to write my thesis as a PhD student at DMSZ in his departments of microbial ecology and diversity research. Moreover, I want to thank Johannes Sikorski for general help, support and introduction to bacterial diversity analysis in R. Selma Gomez-Vieira for the soil sampling in 2011, 2014 and 2017 in the Biodiversity Exploratories, valuable discussion and funny times in the office and troubleshooting in R. Furthermore, I want to thank our three lab queens Alicia Geppert, Anika Methner and Franziska Burkhard for the soil sampling campaign in the Biodiversity Exploratories, support in nucleic acid extraction and library preparation and a very pleasant working atmosphere during my stay at DSMZ.

I want to thank Boyke Bunk and his team for proving and maintaining the server environment, allowing me to handle the sophisticated models and data processing. I want to thank Prof. Dr. Christoph Tebbe and Prof. Dr. Cornelia Smalla for their participation in my thesis committee.

Furthermore, I want to thank Martinique Frentrump and all my friends trying to support me my during this time. I really enjoyed spending my time with all of you.

In the framework of the Biodiversity Exploratories, I want to thank the managers of the three Exploratories, Kirsten Reichel-Jung, Juliane Vogt, Miriam Teuscher and all former managers for their work in maintaining the plot and project infrastructure; Christiane Fischer for giving support through the central office, Andreas Ostrowski for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. We thank the administration of the Hainich national park, the UNESCO Biosphere Reserve Swabian Alb and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all land owners for the excellent collaboration. The central soil sampling campaign of the Biodiversity Exploratories was supported by Ingo Schöning, Prof. Dr. Marion Schrumpf, Steffan Ferber, Beatrix Schnabel, Runa Boeddinghaus, Sven Marhan, and Prof. Dr. Ellen Kandeler. I want to thank Dr. Thomas Reitz for uncomplicated help and permission to sample soil from the Static Fertilization Experiment in Bad Lauchstädt.

References

Alberdi, A. and Gilbert, M. T. P., 2019, A guide to the application of Hill numbers to DNA-based diversity analyses. *Molecular Ecology Resources*, 19(4), 804–817, doi: 10.1111/1755-0998.13014.

Allison, S. D., 2019, Traits track taxonomy. *Nature Ecology & Evolution*, doi: 10.1038/s41559-019-0937-8.

Alm, E. W. and Stahl, D. A., 2000, Critical factors influencing the recovery and integrity of rRNA extracted from environmental samples: Use of an optimized protocol to measure depth-related biomass distribution in freshwater sediments. *Journal of microbiological methods*, 40(2), 153–162.

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A. et al., 2017, Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*, 2(2), e00191–16, doi:10.1128/mSystems.00191-16.

Andersson, S. and Nilsson, S. I., 2001, Influence of pH and temperature on microbial activity, substrate availability of soil-solution bacteria and leaching of dissolved organic carbon in a mor humus. *Soil Biology and Biochemistry*, 33(9), 1181–1191, doi:10.1016/S0038-0717(01)00022-0.

Andrade-Linares, D. R., Zistl-Schlingmann, M., Foesel, B., Dannenmann, M., Schulz, S. and Schloter, M., 2021, Short term effects of climate change and intensification of management on the abundance of microbes driving nitrogen turnover in montane grassland soils. *Science of The Total Environment*, 780, 146672, doi: 10.1016/j.scitotenv.2021.146672.

Appleton, D., French, J. and Vanderpump, M., 1996, Ignoring a Covariate: An Example of Simpson's Paradox. *American Statistician - AMER STATIST*, 50, 340–341, doi:10.1080/00031305.1996.10473563.

Armitage, D. W. and Jones, S. E., 2019, How sample heterogeneity can obscure the signal of microbial interactions. *The ISME Journal*, 13(11), 2639–2646, doi: 10.1038/s41396-019-0463-3.

Armstrong, A., Valverde, A., Ramond, J.-B., Makhalanyane, T. P., Jansson, J. K., Hopkins, D. W., Aspray, T. J., Seely, M., Trindade, M. I. and Cowan, D. A., 2016, Temporal dynamics of hot desert microbial communities reveal structural and functional responses to water input. *Scientific Reports*, 6(1), 34434, doi:10.1038/srep34434.

Aßhauer, K. P., Wemheuer, B., Daniel, R. and Meinicke, P., 2015, Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA. *Bioinformatics*, 31(17), 2882–2884, doi:10.1093/bioinformatics/btv287.

Baca Cabrera, J. C., Hirl, R. T., Schäufele, R., Macdonald, A. and Schnyder, H., 2021, Stomatal conductance limited the CO_2 response of grassland in the last century. *BMC Biology*, 19(1), 50, doi:10.1186/s12915-021-00988-4.

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H. et al., 2018, Structure and function of the global topsoil microbiome. *Nature*, 560(7717), 233–237, doi:10.1038/s41586-018-0386-6.

Banerjee, S., Helgason, B., Wang, L., Winsley, T., Ferrari, B. C. and Siciliano, S. D., 2016, Legacy effects of soil moisture on microbial community structure and N2O emissions. *Soil Biology and Biochemistry*, 95, 40–50, doi:10.1016/j.soilbio.2015.12.004.

Bardgett, R. D. and van der Putten, W. H., 2014, Belowground biodiversity and ecosystem functioning. *Nature*, 515(7528), 505–511, doi:10.1038/nature13855.

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. and Neufeld, J. D., 2011, Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Applied and Environmental Microbiology*, 77(11), 3846–3852, doi:10.1128/AEM.02772-10.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. and Courchamp, F., 2012, Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15(4), 365–377, doi:10.1111/j.1461-0248.2011.01736.x.

Bickel, S., Chen, X., Papritz, A. and Or, D., 2019, A hierarchy of environmental covariates control the global biogeography of soil bacterial richness. *Scientific Reports*, 9(1), 12129, doi:10.1038/s41598-019-48571-w.

Bickel, S. and Or, D., 2020, Soil bacterial diversity mediated by microscale aqueous-phase processes across biomes. *Nature Communications*, 11(1), 116, doi:10.1038/s41467-019-13966-w.

Bickel, S. and Or, D., 2021, The chosen few—variations in common and rare soil bacteria across biomes. *The ISME Journal*, pp. 1–11, doi:10.1038/s41396-021-00981-3.

Blagodatskaya, E. and Kuzyakov, Y., 2013, Active microorganisms in soil: Critical review of estimation criteria and approaches. *Soil Biology and Biochemistry*, 67, 192–211, doi:10.1016/j.soilbio.2013.08.024.

Blair, N., Faulkner, R. D., Till, A. R., Korschens, M. and Schulz, E., 2006, Long-term management impacts on soil C, N and physical fertility: Part II: Bad Lauchstadt static and extreme FYM experiments. *Soil and Tillage Research*, 91(1), 39–47, doi:10.1016/j.still.2005.11.001.

Blankinship, J. C., Niklaus, P. A. and Hungate, B. A., 2011, A meta-analysis of responses of soil biota to global change. *Oecologia*, 165(3), 553–565, doi:10.1007/s00442-011-1909-0.

Blazewicz, S. J., Barnard, R. L., Daly, R. A. and Firestone, M. K., 2013, Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses. *The ISME journal*, 7(11), 2061.

Blüthgen, N., Dormann, C. F., Prati, D., Klaus, V. H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A. et al., 2012, A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. *Basic and Applied Ecology*, 13(3), 207–220, doi:10.1016/j.baae.2012.04.001.

Boeddinghaus, R. S., Marhan, S., Berner, D., Boch, S., Fischer, M., Hölzel, N., Kattge, J., Klaus, V. H., Kleinebecker, T., Oelmann, Y. et al., 2019, Plant functional trait shifts explain concurrent changes in the structure and function of grassland soil microbial communities. *Journal of Ecology*, 107(5), 2197–2210, doi:10.1111/1365-2745.13182.

Böhme, L., Langer, U. and Böhme, F., 2005, Microbial biomass, enzyme activities and microbial community structure in two European long-term field experiments. *Agriculture, Ecosystems & Environment*, 109(1), 141–152, doi:10.1016/j.agee.2005.01.017.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A. and Caporaso, J. G., 2013, Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10(1), 57–59, doi:10.1038/nmeth.2276.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F. et al., 2019, Reproducible, interactive, scalable and extensible microbiome data

science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857, doi:10.1038/s41587-019-0209-9.

Bowsher, A. W., Kearns, P. J. and Shade, A., 2019, 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. *mSystems*, 4(2), e00003–19, doi:10.1128/mSystems.00003-19.

Bray, J. R. and Curtis, J. T., 1957, An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325–349, doi:10.2307/1942268.

Breiman, L., 2001, Random Forests. *Machine Learning*, 45(1), 5–32, doi:10.1023/A:1010933404324.

Bystrianský, L., Hujslová, M., Hršelová, H., Řezáčová, V., Němcová, L., Šimsová, J., Gryndlerová, H., Kofroňová, O., Benada, O. and Gryndler, M., 2019, Observations on two microbial life strategies in soil: Planktonic and biofilm-forming microorganisms are separable. *Soil Biology and Biochemistry*, 136, 107535, doi:10.1016/j.soilbio.2019.107535.

Cai, P., Sun, X., Wu, Y., Gao, C., Mortimer, M., Holden, P. A., Redmile-Gordon, M. and Huang, Q., 2019, Soil biofilms: Microbial interactions, challenges, and advanced techniques for ex-situ characterization. *Soil Ecology Letters*, 1(3-4), 85–93, doi:10.1007/s42832-019-0017-7.

Callahan, B. J., McMurdie, P. J. and Holmes, S. P., 2017, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11(12), 2639–2643, doi:10.1038/ismej.2017.119.

Campbell, B. J., Yu, L., Heidelberg, J. F. and Kirchman, D. L., 2011, Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences*, 108(31), 12776–12781, doi:10.1073/pnas.1101405108.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I. et al., 2010, QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335–336, doi:10.1038/nmeth.f.303.

Carini, P., Delgado-Baquerizo, M., Hinckley, E.-L. S., Holland-Moritz, H., Brewer, T. E., Rue, G., Vanderburgh, C., McKnight, D. and Fierer, N., 2020, Effects of spatial variability and relic DNA removal on the detection of temporal dynamics in soil microbial communities. *mBio*, 11(1), e02776–19, /mbio/11/1/mBio.02776–19.atom, doi:10.1128/mBio.02776-19.

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S. and Fierer, N., 2016, Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology*, 2(3), 1–6, doi:10.1038/nmicrobiol.2016.242.

Castro, H. F., Classen, A. T., Austin, E. E., Norby, R. J. and Schadt, C. W., 2010, Soil microbial community responses to multiple experimental climate change drivers. *Applied and Environmental Microbiology*, 76(4), 999–1007, doi:10.1128/AEM.02874-09.

Chakraborty, A., Chakrabarti, K., Chakraborty, A. and Ghosh, S., 2011, Effect of long-term fertilizers and manure application on microbial biomass and microbial activity of a tropical agricultural soil. *Biology and Fertility of Soils*, 47(2), 227–233, doi:10.1007/s00374-010-0509-1.

Chao, A., Chiu, C.-H. and Jost, L., 2010, Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3599–3609, doi:10.1098/rstb.2010.0272.

Chao, A. and Jost, L., 2012, Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, 93(12), 2533–2547, doi:10.1890/11-1952.1.

Chao, A., Kubota, Y., Zelený, D., Chiu, C.-H., Li, C.-F., Kusumoto, B., Yasuhara, M., Thorn, S., Wei, C.-L., Costello, M. J. et al., 2020, Quantifying sample completeness and comparing diversities among assemblages. *ECOLOGICAL RESEARCH*, p. 39.

Chater, K. F., Biró, S., Lee, K. J., Palmer, T. and Schrempf, H., 2010, The complex extracellular biology of Streptomyces. *FEMS Microbiology Reviews*, 34(2), 171–198, doi:10.1111/j.1574-6976.2009.00206.x.

Chau, J. F., Bagtzoglou, A. C. and Willig, M. R., 2011, The effect of soil texture on richness and diversity of bacterial communities. *Environmental Forensics*, 12(4), 333–341, doi:10.1080/15275922.2011.622348.

Chenu, C., Stotzky, G., Huang, P. and Bollag, J., 2002, Interactions between microorganisms and soil particles: An overview. *Interactions between soil particles and microorganisms: Impact on the terrestrial ecosystem*, 1, 1–40.

Clingenpeel, S., Schwientek, P., Hugenholtz, P. and Woyke, T., 2014, Effects of sample treatments on genome recovery via single-cell genomics. *The ISME Journal*, 8(12), 2546–2549, doi:10.1038/ismej.2014.92.

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

Costerton, J. W., Cheng, K. J., Geesey, G. G., Ladd, T. I., Nickel, J. C., Dasgupta, M. and Marrie, T. J., 1987, Bacterial Biofilms in Nature and Disease. *Annual Review of Microbiology*, 41(1), 435–464, doi: 10.1146/annurev.mi.41.100187.002251.

Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P. and Simonet, P., 2001, Quantification of bacterial subgroups in soil: Comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environmental Microbiology*, 3(7), 431–439, doi:10.1046/j.1462-2920.2001.00208.x.

Cruz-Martínez, K., Suttle, K. B., Brodie, E. L., Power, M. E., Andersen, G. L. and Banfield, J. F., 2009, Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *The ISME Journal*, 3(6), 738–744, doi:10.1038/ismej.2009.16.

Curtis, T. P., Sloan, W. T. and Scannell, J. W., 2002, Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences*, 99(16), 10494–10499, doi:10.1073/pnas.142680199.

Dai, Z., Su, W., Chen, H., Barberán, A., Zhao, H., Yu, M., Yu, L., Brookes, P. C., Schadt, C. W., Chang, S. X. et al., 2018, Long-term nitrogen fertilization decreases bacterial diversity and favors the growth of Actinobacteria and Proteobacteria in agro-ecosystems across the globe. *Global Change Biology*, 24(8), 3452–3461, doi:10.1111/gcb.14163.

Davidson, E. A., Samanta, S., Caramori, S. S. and Savage, K., 2012, The Dual Arrhenius and Michaelis–Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, 18(1), 371–384, doi:10.1111/j.1365-2486.2011.02546.x.

de Boer, W., Folman, L. B., Summerbell, R. C. and Boddy, L., 2005, Living in a fungal world: Impact of fungi on soil bacterial niche development. *FEMS Microbiology Reviews*, 29(4), 795–811, doi:10.1016/j.femsre.2004.11.005.

De-Kayne, R., Frei, D., Greenway, R., Mendes, S. L., Retel, C. and Feulner, P. G. D., 2021, Sequencing platform shifts provide opportunities but pose challenges

for combining genomic data sets. *Molecular Ecology Resources*, 21(3), 653–660, doi:10.1111/1755-0998.13309.

de Vries, F. T., Hoffland, E., van Eekeren, N., Brussaard, L. and Bloem, J., 2006, Fungal/bacterial ratios in grasslands with contrasting nitrogen management. *Soil Biology and Biochemistry*, 38(8), 2092–2103, doi:10.1016/j.soilbio.2006.01.008.

DeAngelis, K. M., Pold, G., Topçuoğlu, B. D., van Diepen, L. T. A., Varney, R. M., Blanchard, J. L., Melillo, J. and Frey, S. D., 2015, Long-term forest soil warming alters microbial communities in temperate forest soils. *Frontiers in Microbiology*, 6, doi:10.3389/fmicb.2015.00104.

Dechesne, A., Wang, G., Gülez, G., Or, D. and Smets, B. F., 2010, Hydration-controlled bacterial motility and dispersal on surfaces. *Proceedings of the National Academy of Sciences*, 107(32), 14369–14372, doi:10.1073/pnas.1008392107.

Delgado-Baquerizo, M., Bardgett, R. D., Vitousek, P. M., Maestre, F. T., Williams, M. A., Eldridge, D. J., Lambers, H., Neuhauser, S., Gallardo, A., García-Velázquez, L. et al., 2019, Changes in belowground biodiversity during ecosystem development. *Proceedings of the National Academy of Sciences*, 116(14), 6891–6896, doi:10.1073/pnas.1818400116.

Delgado-Baquerizo, M. and Eldridge, D. J., 2019, Cross-biome drivers of soil bacterial alpha diversity on a worldwide scale. *Ecosystems*, 22(6), 1220–1231, doi:10.1007/s10021-018-0333-2.

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Trivedi, P., Osanai, Y., Liu, Y.-R., Hamonts, K., Jeffries, T. C. and Singh, B. K., 2016, Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecological Monographs*, 86(3), 373–390, doi:10.1002/ecm.1216.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K. and Fierer, N., 2018, A global atlas of the dominant bacteria found in soil. *Science*, 359(6373), 320–325, doi:10.1126/science.aap9516.

Denef, V. J., Fujimoto, M., Berry, M. A. and Schmidt, M. L., 2016, Seasonal succession leads to habitat-dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. *Frontiers in Microbiology*, 0, doi:10.3389/fmicb.2016.00606.

Dorau, K., Bamminger, C., Koch, D. and Mansfeldt, T., 2022, Evidences of soil warming from long-term trends (1951–2018) in North Rhine-Westphalia, Germany. *Climatic Change*, 170(1), 9, doi:10.1007/s10584-021-03293-9.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J. et al., 2013, Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46, doi:10.1111/j.1600-0587.2012.07348.x.

Duan, N., Li, L., Liang, X., Fine, A., Zhuang, J., Radosevich, M. and Schaeffer, S. M., 2022, Variation in bacterial community structure under long-term fertilization, tillage, and cover cropping in continuous cotton production. *Frontiers in Microbiology*, 13, 847005, doi:10.3389/fmicb.2022.847005.

Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Alexander Richter, R., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J. D., Haft, D. H. et al., 2012, Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal*, 6(6), 1186–1199, doi:10.1038/ismej.2011.189.

Ebrahimi, A. N. and Or, D., 2014, Microbial dispersal in unsaturated porous media: Characteristics of motile bacterial cell motions in unsaturated angular pore networks. *Water Resources Research*, 50(9), 7406–7429, doi:10.1002/2014WR015897.

Eichorst, S. A., Strasser, F., Woyke, T., Schintlmeister, A., Wagner, M. and Woebken, D., 2015, Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *FEMS Microbiology Ecology*, 91(10), fiv106, doi:10.1093/femsec/fiv106.

Ellis, N., Smith, S. J. and Pitcher, C. R., 2012, Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, 93(1), 156–168, doi:10.1890/11-0252.1.

Emadodin, I., Corral, D. E. F., Reinsch, T., Kluß, C. and Taube, F., 2021, Climate change effects on temperate grassland and Its implication for forage production: A case study from northern Germany. *Agriculture*, 11(3), 232, doi:10.3390/agriculture11030232.

Engelhardt, I. C., Welty, A., Blazewicz, S. J., Bru, D., Rouard, N., Breuil, M.-C., Gessler, A., Galiano, L., Miranda, J. C., Spor, A. et al., 2018, Depth matters: Effects of precipitation regime on soil microbial activity upon rewetting of

a plant-soil system. *The ISME Journal*, 12(4), 1061–1071, doi:10.1038/s41396-018-0079-z.

Eo, J. and Park, K.-C., 2016, Long-term effects of imbalanced fertilization on the composition and diversity of soil bacterial community. *Agriculture, Ecosystems & Environment*, 231, 176–182, doi:10.1016/j.agee.2016.06.039.

Ester, M., Kriegel, H., Sander, J. and Xu, X., 1996, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Knowledge Discovery and Data Mining.

Feng, Y., Delgado-Baquerizo, M., Zhu, Y., Han, X., Han, X., Xin, X., Li, W., Guo, Z., Dang, T., Li, C. et al., 2022, Responses of soil bacterial diversity to fertilization are driven by local environmental context across china. *Engineering*, 12, 164–170, doi:10.1016/j.eng.2021.09.012.

Ferrier, S., Manion, G., Elith, J. and Richardson, K., 2007, Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13(3), 252–264, doi:10.1111/j.1472-4642.2007.00341.x.

Fierer, N. and Jackson, R. B., 2006, The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103(3), 626–631, doi:10.1073/pnas.0507535103.

Fierer, N., Nemergut, D., Knight, R. and Craine, J. M., 2010, Changes through time: Integrating microorganisms into the study of succession. *Research in Microbiology*, 161(8), 635–642, doi:10.1016/j.resmic.2010.06.002.

Fierer, N., Strickland, M. S., Liptzin, D., Bradford, M. A. and Cleveland, C. C., 2009, Global patterns in belowground communities. *Ecology Letters*, 12(11), 1238–1249, doi:10.1111/j.1461-0248.2009.01360.x.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D. et al., 2010, Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic and Applied Ecology*, 11(6), 473–485, doi:10.1016/j.baae.2010.07.009.

Flemming, H.-C. and Wingender, J., 2010, The biofilm matrix. *Nature reviews microbiology*, 8(9), 623–633, doi:10.1038/nrmicro2415.

Flemming, H.-C. and Wuertz, S., 2019, Bacteria and archaea on earth and their abundance in biofilms. *Nature Reviews Microbiology*, 17(4), 247–260, doi:10.1038/s41579-019-0158-9.

Francioli, D., Schulz, E., Lentendu, G., Wubet, T., Buscot, F. and Reitz, T., 2016, Mineral vs. Organic Amendments: Microbial Community Structure, Activity and Abundance of Agriculturally Relevant Microbes Are Driven by Long-Term Fertilization Strategies. *Frontiers in Microbiology*, 7, doi:10.3389/fmicb.2016.01446.

Franklin, R. B., Luria, C., Ozaki, L. S. and Bukaveckas, P. A., 2013, Community composition and activity state of estuarine bacterioplankton assessed using differential staining and metagenomic analysis of 16S rDNA and rRNA. *Aquatic Microbial Ecology*, 69(3), 247–261, doi:10.3354/ame01635.

Friedman, J. H., Hastie, T. and Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33, 1–22, doi:10.18637/jss.v033.i01.

Frindte, K., Pape, R., Werner, K., Löffler, J. and Knief, C., 2019, Temperature and soil moisture control microbial community composition in an arctic–alpine ecosystem along elevational and micro-topographic gradients. *The ISME Journal*, 13(8), 2031–2043, doi:10.1038/s41396-019-0409-9.

Ge, Y., Zhang, J.-b., Zhang, L.-m., Yang, M. and He, J.-z., 2008, Long-term fertilization regimes affect bacterial community structure and diversity of an agricultural soil in northern China. *Journal of Soils and Sediments*, 8(1), 43–50, doi:10.1065/jss2008.01.270.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. and Egozcue, J. J., 2017, Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, 8, doi:10.3389/fmicb.2017.02224.

Gossner, M. M., Lewinsohn, T. M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K., Renner, S. C., Sikorski, J., Wubet, T. et al., 2016, Land-use intensification causes multitrophic homogenization of grassland communities. *Nature*, 540(7632), 266–269, doi:10.1038/nature20575.

Guo, X., Feng, J., Shi, Z., Zhou, X., Yuan, M., Tao, X., Hale, L., Yuan, T., Wang, J., Qin, Y. et al., 2018, Climate warming leads to divergent succession of grassland microbial communities. *Nature Climate Change*, 8(9), 813–818, doi:10.1038/s41558-018-0254-2.

Hariharan, J., Sengupta, A., Grewal, P. and Dick, W. A., 2017, Functional predictions of microbial communities in soil as affected by long-term tillage practices. *Agricultural & Environmental Letters*, 2(1), 170031, doi:10.2134/ael2017.09.0031.

Hartmann, M., Frey, B., Mayer, J., Mäder, P. and Widmer, F., 2015, Distinct soil microbial diversity under long-term organic and conventional farming. *The ISME Journal*, 9(5), 1177–1194, doi:10.1038/ismej.2014.210.

Hemkemeyer, M., Dohrmann, A. B., Christensen, B. T. and Tebbe, C. C., 2018, Bacterial preferences for specific soil particle size fractions revealed by community analyses. *Frontiers in Microbiology*, 9, doi:10.3389/fmicb.2018.00149.

Hermans, S. M., Buckley, H. L., Curran-Cournane, F., Taylor, M. and Lear, G., 2020, Temporal variation in soil bacterial communities can be confounded with spatial variation. *FEMS Microbiology Ecology*, 96(12), fiaa192, doi:10.1093/femsec/fiaa192.

Herzog, S., Wemheuer, F., Wemheuer, B. and Daniel, R., 2015, Effects of fertilization and sampling time on composition and diversity of entire and active bacterial communities in German grassland soils. *Plos*, 10(12), e0145575, doi:10.1371/journal.pone.0145575.

Hill, M. O., 1973, Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432, doi:10.2307/1934352.

Holt, R. D., 2009, Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences*, 106, 19659–19665, doi:10.1073/pnas.0905137106.

Huang, Q., Wu, H., Cai, P., Fein, J. B. and Chen, W., 2015, Atomic force microscopy measurements of bacterial adhesion and biofilm formation onto clay-sized particles. *Scientific Reports*, 5(1), 16857, doi:10.1038/srep16857.

Huber, K. J., Vieira, S., Sikorski, J., Wüst, P. K., Fösel, B. U., Gröngröft, A. and Overmann, J., 2022, Differential response of Acidobacteria to water content, soil type, and land use during an extended drought in African savannah soils. *Frontiers in Microbiology*, 13, 750456, doi:doi.org/10.3389/fmicb.2022.750456.

Hujslová, M., Gryndlerová, H., Bystrianský, L., Hršelová, H. and Gryndler, M., 2020, Biofilm and planktonic microbial communities in highly acidic soil (pH $<$ 3) in the Soos National Nature Reserve, Czech Republic. *Extremophiles*, 24(4), 577–591, doi:10.1007/s00792-020-01177-x.

Huysman, F. and Verstraete, W., 1993a, Effect of cell surface characteristics on the adhesion of bacteria to soil particles. *Biology and Fertility of Soils*, 16(1), 21–26, doi:10.1007/BF00336510.

Huysman, F. and Verstraete, W., 1993b, Water-facilitated transport of bacteria in unsaturated soil columns: Influence of cell surface hydrophobicity and soil properties. *Soil Biology and Biochemistry*, 25(1), 83–90, doi:10.1016/0038-0717(93)90245-7.

Jansen, F. and Oksanen, J., 2013, How to model species responses along ecological gradients – Huisman–Olff–Fresco models revisited. *Journal of Vegetation Science*, 24(6), 1108–1117, doi:10.1111/jvs.12050.

Jansson, J. K. and Hofmockel, K. S., 2020, Soil microbiomes and climate change. *Nature Reviews Microbiology*, 18(1), 35–46, doi:10.1038/s41579-019-0265-7.

Jiao, S. and Lu, Y., 2020, Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environmental Microbiology*, 22(3), 1052–1065, doi:10.1111/1462-2920.14815.

Jones, S. E. and Lennon, J. T., 2010, Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences*, 107(13), 5881–5886, doi:10.1073/pnas.0912765107.

Jost, L., 2006, Entropy and diversity. *Oikos*, 113(2), 363–375, doi:10.1111/j.2006.0030-1299.14714.x.

Jost, L., 2007, Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), 2427–2439, doi:10.1890/06-1736.1.

Judd, K. E., Crump, B. C. and Kling, G. W., 2006, Variation in dissolved organic matter controls bacterial production and community composition. *Ecology*, 87(8), 2068–2079, doi:10.1890/0012-9658(2006)87[2068:VIDOMC]2.0.CO;2.

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., Schrumpf, M. and Daniel, R., 2016, Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Scientific Reports*, 6(1), doi:10.1038/srep33696.

Kallmeyer, J., Smith, D. C., Spivack, A. J. and D'Hondt, S., 2008, New cell extraction procedure applied to deep subsurface sediments: Cell extraction of deep subsurface sediments. *Limnology and Oceanography: Methods*, 6(6), 236–245, doi:10.4319/lom.2008.6.236.

Kaminsky, R. and Morales, S. E., 2018, Conditionally rare taxa contribute but do not account for changes in soil prokaryotic community structure. *Frontiers in Microbiology*, 9, 809, doi:10.3389/fmicb.2018.00809.

Kandeler, E., Tscherko, D., Bruce, K. D., Stemmer, M., Hobbs, P. J., Bardgett, R. D. and Amelung, W., 2000, Structure and function of the soil microbial community in microhabitats of a heavy metal polluted soil. *Biology and Fertility of Soils*, 32(5), 390–400, doi:10.1007/s003740000268.

Karimi, B., Terrat, S., Dequiedt, S., Saby, N. P. A., Horrigue, W., Lelièvre, M., Nowak, V., Jolivet, C., Arrouays, D., Wincker, P. et al., 2018, Biogeography of soil bacteria and archaea across France. *Science Advances*, 4(7), eaat1808, doi:10.1126/sciadv.aat1808.

Katoh, K. and Standley, D. M., 2013, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780, doi:10.1093/molbev/mst010.

Kautz, T., Wirth, S. and Ellmer, F., 2004, Microbial activity in a sandy arable soil is governed by the fertilization regime. *European Journal of Soil Biology*, 40(2), 87–94, doi:10.1016/j.ejsobi.2004.10.001.

Kemp, P. F., Lee, S. and LaRoche, J., 1993, Estimating the growth rate of slowly growing marine bacteria from RNA content. *Applied and Environmental Microbiology*, 59(8), 2594–2601, doi:10.1128/aem.59.8.2594-2601.1993.

Kerkhof, L. and Ward, B. B., 1993, Comparison of nucleic acid hybridization and fluorometry for measurement of the relationship between RNA/DNA ratio and growth rate in a marine bacterium. *Applied and Environmental Microbiology*, 59(5), 1303–1309, doi:10.1128/aem.59.5.1303-1309.1993.

Kim, M. K., Na, J.-R., Lee, T.-H., Im, W.-T., Soung, N.-K. and Yang, D.-C. ., 2007, Solirubrobacter soli sp. nov., isolated from soil of a ginseng field. *International Journal of Systematic and Evolutionary Microbiology*, 57(7), 1453–1455, doi:10.1099/ijs.0.64715-0.

Kivlin, S. N. and Hawkes, C. V., 2020, Spatial and temporal turnover of soil microbial communities is not linked to function in a primary tropical forest. *Ecology*, 101(4), doi:10.1002/ecy.2985.

Klein, A. M., Bohannan, B. J. M., Jaffe, D. A., Levin, D. A. and Green, J. L., 2016, Molecular evidence for metabolically active bacteria in the atmosphere. *Frontiers in Microbiology*, 0, doi:10.3389/fmicb.2016.00772.

Kleine Bardenhorst, S., Vital, M., Karch, A. and Rübsamen, N., 2022, Richness estimation in microbiome data obtained from denoising pipelines. *Computational and Structural Biotechnology Journal*, 20, 508–520, doi:10.1016/j.csbj.2021.12.036.

Klimanek, E.-M., 2000, Die Wirkung von Düngungsänderungen auf die mikrobielle Aktivität von Löss-Schwarzerde nach 96 Jahren differenzierter organischer und mineralischer düngung im "Statischen Düngungsversuch"; Bad Lauchstädt. *Archives of Agronomy and Soil Science*, 45(5), 381–397, doi:10.1080/03650340009366137.

Kögel-Knabner, I., Guggenberger, G., Kleber, M., Kandeler, E., Kalbitz, K., Scheu, S., Eusterhues, K. and Leinweber, P., 2008, Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry. *Journal of Plant Nutrition and Soil Science*, 171(1), 61–82, doi:10.1002/jpln.200700048.

Köppen, D. and Eich, D., 1991, Einfluß 85jähriger differenzierter organischer und mineralischer Düngung auf Bodeneigenschaften im Statischen Versuch Bad Lauchstädt. *Zeitschrift für Pflanzenernährung und Bodenkunde*, 154(4), 245–252, doi:10.1002/jpln.19911540402.

Kuhn, M. and Johnson, K., 2019, Engineering numeric predictors. In Feature Engineering and Selection, Chapman and Hall/CRC.

Ladau, J., Shi, Y., Jing, X., He, J.-S., Chen, L., Lin, X., Fierer, N., Gilbert, J. A., Pollard, K. S. and Chu, H., 2018, Existing climate change will lead to pronounced shifts in the diversity of soil prokaryotes. *mSystems*, 3(5), e00167–18, /msystems/3/5/msys.00167–18.atom, doi:10.1128/mSystems.00167-18.

Lauber, C. L., Ramirez, K. S., Aanderud, Z., Lennon, J. and Fierer, N., 2013, Temporal variability in soil microbial communities across land-use types. *The ISME Journal*, 7(8), 1641–1650, doi:10.1038/ismej.2013.50.

Leff, J. W., Jones, S. E., Prober, S. M., Barberán, A., Borer, E. T., Firn, J. L., Harpole, W. S., Hobbie, S. E., Hofmockel, K. S., Knops, J. M. H. et al., 2015, Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proceedings of the National Academy of Sciences*, 112(35), 10967–10972, doi:10.1073/pnas.1508382112.

Lennon, J. T., Aanderud, Z. T., Lehmkuhl, B. K. and Schoolmaster Jr., D. R., 2012, Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology*, 93(8), 1867–1879, doi:10.1890/11-1745.1.

Li, H., Xu, Z., Yang, S., Li, X., Top, E., Wang, R., Zhang, Y., Cai, J., Yao, F., Han, X.-G. et al., 2016, Responses of soil bacterial communities to nitrogen deposition and precipitation increment are closely linked with aboveground community variation. *Microbial Ecology*, 71, doi:10.1007/s00248-016-0730-z.

Li, J., Wen, Y. and Yang, X., 2021, Understanding the responses of soil bacterial communities to long-term fertilization regimes using DNA and RNA sequencing. *Agronomy*, 11(12), 2425, doi:10.3390/agronomy11122425.

Liang, Y., Jiang, Y., Wang, F., Wen, C., Deng, Y., Xue, K., Qin, Y., Yang, Y., Wu, L., Zhou, J. et al., 2015, Long-term soil transplant simulating climate change with latitude significantly alters microbial temporal turnover. *The ISME Journal*, 9(12), 2561–2572, doi:10.1038/ismej.2015.78.

Lindahl, V., 1996, Improved soil dispersion procedures for total bacterial counts, extraction of indigenous bacteria and cell survival. *Journal of Microbiological Methods*, 25(3), 279–286, doi:10.1016/0167-7012(95)00102-6.

Lindahl, V. and Bakken, L. R., 1995, Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiology Ecology*, 16(2), 135–142, doi:10.1111/j.1574-6941.1995.tb00277.x.

Lipson, D., Schadt, C. and Schmidt, S., 2002, Changes in soil microbial community structure and function in an alpine dry meadow following spring snow melt. *Microbial Ecology*, 43(3), 307–314, doi:10.1007/s00248-001-1057-x.

Lipson, D. A., 2007, Relationships between temperature responses and bacterial community structure along seasonal and altitudinal gradients. *FEMS Microbiology Ecology*, 59(2), 418–427, doi:10.1111/j.1574-6941.2006.00240.x.

Liu, N., Hu, H., Ma, W., Deng, Y., Liu, Y., Hao, B., Zhang, X., Dimitrov, D., Feng, X. and Wang, Z., 2019, Contrasting biogeographic patterns of bacterial and archaeal diversity in the top- and subsoils of temperate grasslands. *mSystems*, 4(5), 18, doi:10.1128/mSystems.00566-19.

Lladó Fernández, S., Větrovský, T. and Baldrian, P., 2019, The concept of operational taxonomic units revisited: Genomes of bacteria that are regarded as closely related are often highly dissimilar. *Folia Microbiologica*, 64(1), 19–23, doi:10.1007/s12223-018-0627-y.

Loeppmann, S., Semenov, M., Kuzyakov, Y. and Blagodatskaya, E., 2018, Shift from dormancy to microbial growth revealed by RNA:DNA ratio. *Ecological Indicators*, 85, 603–612, doi:10.1016/j.ecolind.2017.11.020.

Lozupone, C. and Knight, R., 2005, UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12), 8228–8235, doi:10.1128/AEM.71.12.8228-8235.2005.

Lueders, T., Manefield, M. and Friedrich, M. W., 2003, Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients: Quantitative analysis of SIP gradients. *Environmental Microbiology*, 6(1), 73–78, doi:10.1046/j.1462-2920.2003.00536.x.

Lynch, M. D. J. and Neufeld, J. D., 2015, Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13(4), 217–229, doi:10.1038/nrmicro3400.

MacArthur, R. H., 1965, Patterns of Species Diversity. *Biological Reviews*, 40(4), 510–533, doi:10.1111/j.1469-185X.1965.tb00815.x.

Magurran, A. E. and Henderson, P. A., 2003, Explaining the excess of rare species in natural species abundance distributions. *Nature*, 422(6933), 714–716, doi:10.1038/nature01547.

Mandelstam, J. and Halvorson, H., 1960, Turnover of protein and nucleic acid in soluble and ribosome fractions of non-growing Escherichia coli. *Biochimica et Biophysica Acta*, 40, 43–49, doi:10.1016/0006-3002(60)91313-5.

Maron, P.-A., Schimann, H., Ranjard, L., Brothier, E., Domenach, A.-M., Lensi, R. and Nazaret, S., 2006, Evaluation of quantitative and qualitative recovery of bacterial communities from different soil types by density gradient centrifugation. *European Journal of Soil Biology*, 42(2), 65–73, doi:10.1016/j.ejsobi.2005.08.003.

McMurdie, P. J., 2018, Normalization of microbiome profiling fata. In Mcirobiome Analysis: Methods and Protocols, pp. 143–168, Springer Science+Business Media, LLC, part of Springer Nature.

McMurdie, P. J. and Holmes, S., 2014, Waste not, want not: Why rarefying microbiome data is inadmissible. *PLOS Computational Biology*, 10(4), e1003531, doi:10.1371/journal.pcbi.1003531.

Meisner, A., Leizeaga, A., Rousk, J. and Bååth, E., 2017, Partial drying accelerates bacterial growth recovery to rewetting. *Soil Biology and Biochemistry*, 112, 269–276, doi:10.1016/j.soilbio.2017.05.016.

Melillo, J. M., Frey, S. D., DeAngelis, K. M., Werner, W. J., Bernard, M. J., Bowles, F. P., Pold, G., Knorr, M. A. and Grandy, A. S., 2017, Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, 358(6359), 101–105, doi:10.1126/science.aan2874.

Melillo, J. M., Steudler, P. A., Aber, J. D., Newkirk, K., Lux, H., Bowles, F. P., Catricala, C., Magill, A., Ahrens, T. and Morrisseau, S., 2002, Soil warming and carbon-cycle feedbacks to the climate system. *Science*, 298(5601), 2173–2176, doi:10.1126/science.1074153.

Merbach, I. and Körschens, M., 2002, The Static Fertilization Experiment at the start and the end of the 20th century. *Archives of Agronomy and Soil Science*, 48(5), 413–422, doi:10.1080/03650340215649.

Mestre, M., Borrull, E., Sala, M. M. and Gasol, J. M., 2017, Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME Journal*, 11(4), 999–1010, doi:10.1038/ismej.2016.166.

Mod, H. K., Buri, A., Yashiro, E., Guex, N., Malard, L., Pinto-Figueroa, E., Pagni, M., Niculita-Hirzel, H., van der Meer, J. R. and Guisan, A., 2021, Predicting spatial patterns of soil bacteria under current and future environmental conditions. *The ISME Journal*, 15(9), 2547–2560, doi:10.1038/s41396-021-00947-5.

Morrissey, E. M., Mau, R. L., Hayer, M., Liu, X.-J. A., Schwartz, E., Dijkstra, P., Koch, B. J., Allen, K., Blazewicz, S. J., Hofmockel, K. et al., 2019, Evolutionary history constrains microbial traits across environmental variation. *Nature Ecology & Evolution*, 3(7), 1064–1069, doi:10.1038/s41559-019-0918-y.

Morton, J. T., Toran, L., Edlund, A., Metcalf, J. L., Lauber, C. and Knight, R., 2017, Uncovering the horseshoe effect in microbial analyses. *mSystems*, 2(1), 8.

Nacke, H., Thürmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., Schöning, I., Schrumpf, M. and Daniel, R., 2011, Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS ONE*, 6(2), e17000, doi:10.1371/journal.pone.0017000.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P. et al., 2013, Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews*, 77(3), 342–356, doi:10.1128/MMBR.00051-12.

Neumann, D., Heuer, A., Hemkemeyer, M., Martens, R. and Tebbe, C. C., 2013, Response of microbial communities to long-term fertilization depends on their microhabitat. *FEMS Microbiology Ecology*, 86(1), 71–84, doi:10.1111/1574-6941.12092.

Nielsen, K. M., Johnsen, P. J., Bensasson, D. and Daffonchio, D., 2007, Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research*, 6(1-2), 37–53, doi:10.1051/ebr:2007031.

Nottingham, A. T., Scott, J. J., Saltonstall, K., Broders, K., Montero-Sanchez, M., Püspök, J., Bååth, E. and Meir, P., 2022, Microbial diversity declines in warmed tropical soil and respiration rise exceed predictions as communities adapt. *Nature Microbiology*, 7(10), 1650–1660, doi:10.1038/s41564-022-01200-1.

Noviana, Z., 2021, Structure and composition of bacterial communities in the rhizosphere and root of Hypericum species. Ph.D. thesis, Technische Universität Braunschweig, Braunschweig.

Nunan, N., 2017, The microbial habitat in soil: Scale, heterogeneity and functional consequences. *Journal of Plant Nutrition and Soil Science*, 180(4), 425–429, doi:10.1002/jpln.201700184.

Nunan, N., Wu, K., Young, I. M., Crawford, J. W. and Ritz, K., 2003, Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. *FEMS Microbiology Ecology*, 44(2), 203–215, doi:10.1016/S0168-6496(03)00027-8.

Obayomi, O., Seyoum, M. M., Ghazaryan, L., Tebbe, C. C., Murase, J., Bernstein, N. and Gillor, O., 2021, Soil texture and properties rather than irrigation water type shape the diversity and composition of soil microbial communities. *Applied Soil Ecology*, 161, 103834, doi:10.1016/j.apsoil.2020.103834.

Ogilvie, L. A., Hirsch, P. R. and Johnston, A. W. B., 2008, Bacterial diversity of the broadbalk 'classical' winter wheat experiment in relation to long-term fertilizer inputs. *Microbial Ecology*, 56(3), 525–537, doi:10.1007/s00248-008-9372-0.

Overmann, J., Abt, B. and Sikorski, J., 2017, Present and future of culturing bacteria. *Annual Review of Microbiology*, 71(1), 711–730, doi:10.1146/annurev-micro-090816-093449.

Pan, Y., Cassman, N., de Hollander, M., Mendes, L. W., Korevaar, H., Geerts, R. H., van Veen, J. A. and Kuramae, E. E., 2014, Impact of long-term N,

P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS Microbiology Ecology*, 90(1), 195–205, doi:10.1111/1574-6941.12384.

Peichl, M., Thober, S., Samaniego, L., Hansjürgens, B. and Marx, A., 2019, Climate impacts on long-term silage maize yield in Germany. *Scientific Reports*, 9(1), 7674, doi:10.1038/s41598-019-44126-1.

Peres-Neto, P. R. and Jackson, D. A., 2001, How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, 129(2), 169–178, doi:10.1007/s004420100720.

Pietikäinen, J., Pettersson, M. and Bååth, E., 2005, Comparison of temperature effects on soil respiration and bacterial and fungal growth rates. *FEMS Microbiology Ecology*, 52(1), 49–58, doi:10.1016/j.femsec.2004.10.002.

Pinke, Z., Decsi, B., Jámbor, A., Kardos, M. K., Kern, Z., Kozma, Z. and Ács, T., 2022, Climate change and modernization drive structural realignments in European grain production. *Scientific Reports*, 12(1), 7374, doi:10.1038/s41598-022-10670-6.

Plassart, P., Prévost-Bouré, N. C., Uroz, S., Dequiedt, S., Stone, D., Creamer, R., Griffiths, R. I., Bailey, M. J., Ranjard, L. and Lemanceau, P., 2019, Soil parameters, land use, and geographical distance drive soil bacterial communities along a European transect. *Scientific Reports*, 9(1), 605, doi:10.1038/s41598-018-36867-2.

Pold, G., Melillo, J. M. and DeAngelis, K. M., 2015, Two decades of warming increases diversity of a potentially lignolytic bacterial community. *Frontiers in Microbiology*, 6.

Pörtner, H. O., Roberts, D. C., Poloczanska, E. S., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S. and Möller, V., 2022, IPCC, 2022: Summary for Policymakers. *Climate Change*.

Price, M. N., Dehal, P. S. and Arkin, A. P., 2010, FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLOS ONE*, 5(3), e9490, doi:10.1371/journal.pone.0009490.

Probandt, D., Eickhorst, T., Ellrott, A., Amann, R. and Knittel, K., 2018, Microbial life on a sand grain: From bulk sediment to single grains. *The ISME Journal*, 12(2), 623–633, doi:10.1038/ismej.2017.197.

Prober, S. M., Leff, J. W., Bates, S. T., Borer, E. T., Firn, J., Harpole, W. S., Lind, E. M., Seabloom, E. W., Adler, P. B., Bakker, J. D. et al., 2015, Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters*, 18(1), 85–95, doi:10.1111/ele.12381.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O., 2013, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596, doi:10.1093/nar/gks1219.

Quinn, T. P., Erb, I., Richardson, M. F. and Crowley, T. M., 2018, Understanding sequencing data as compositions: An outlook and review. *Bioinformatics*, 34(16), 2870–2878, doi:10.1093/bioinformatics/bty175.

{R Core Team}, 2021, R: A Language and Environment for Statistical Computing.

Ramirez, K. S., Craine, J. M. and Fierer, N., 2012, Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Global Change Biology*, 18(6), 1918–1927, doi:10.1111/j.1365-2486.2012.02639.x.

Ranjard, L. and Richaume, A., 2001, Quantitative and qualitative microscale distribution of bacteria in soil. *Research in Microbiology*, 152(8), 707–716, doi:10.1016/S0923-2508(01)01251-7.

Raynaud, X. and Nunan, N., 2014, Spatial ecology of bacteria at the microscale in soil. *PLoS ONE*, 9(1), e87217, doi:10.1371/journal.pone.0087217.

Richter-Heitmann, T., Hofner, B., Krah, F.-S., Sikorski, J., Wüst, P. K., Bunk, B., Huang, S., Regan, K. M., Berner, D., Boeddinghaus, R. S. et al., 2020, Stochastic dispersal rather than deterministic selection explains the spatio-temporal distribution of soil bacteria in a temperate grassland. *Frontiers in Microbiology*, 11.

Romero-Olivares, A., Allison, S. and Treseder, K., 2017, Soil microbes and their response to experimental warming over time: A meta-analysis of field studies. *Soil Biology and Biochemistry*, 107, 32–40, doi:10.1016/j.soilbio.2016.12.026.

Roser, D. J., Bavor, H. J. and McKERSIE, S. A., 1987, Application of Most-Probable-Number Statistics to direct enumeration of microorganisms. *APPL. ENVIRON. MICROBIOL.*, 53, 6.

Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F. and Coumou, D., 2022, Accelerated western European heatwave trends linked to more-persistent double jets

over Eurasia. *Nature Communications*, 13(1), 3851, doi:10.1038/s41467-022-31432-y.

Rousk, J., Brookes, P. C. and Bååth, E., 2010, Investigating the mechanisms for the opposing pH relationships of fungal and bacterial growth in soil. *Soil Biology and Biochemistry*, 42(6), 926–934, doi:10.1016/j.soilbio.2010.02.009.

Ruamps, L. S., Nunan, N. and Chenu, C., 2011, Microbial biogeography at the soil pore scale. *Soil Biology and Biochemistry*, 43(2), 280–286, doi:10.1016/j.soilbio.2010.10.010.

Rummel, P. S., Pfeiffer, B., Pausch, J., Well, R., Schneider, D. and Dittert, K., 2020, Maize root and shoot litter quality controls short-term CO_2 and N_2O emissions and bacterial community structure of arable soil. *Biogeosciences*, 17(4), 1181–1198, doi:10.5194/bg-17-1181-2020.

Sanchez-Cid, C., Tignat-Perrier, R., Franqueville, L., Delaurière, L., Schagat, T. and Vogel, T. M., 2022, Sequencing depth has a stronger effect than DNA extraction on soil bacterial richness discovery. *Biomolecules*, 12(3), 364, doi:10.3390/biom12030364.

Schmid, M. W., van Moorsel, S. J., Hahl, T., De Luca, E., De Deyn, G. B., Wagg, C., Niklaus, P. A. and Schmid, B., 2021, Effects of plant community history, soil legacy and plant diversity on soil microbial communities. *Journal of Ecology*, 109(8), 3007–3023, doi:10.1111/1365-2745.13714.

Schöps, R., Goldmann, K., Herz, K., Lentendu, G., Schöning, I., Bruelheide, H., Wubet, T. and Buscot, F., 2018, Land-Use intensity rather than plant functional identity shapes bacterial and fungal rhizosphere communities. *Frontiers in Microbiology*, 9, 2711, doi:10.3389/fmicb.2018.02711.

Schwarz, M., Kloß, S., Stöckel, S., Pollok, S., Holländer, A., Cialla-May, D., Weber, K. and Popp, J., 2017, Pioneering particle-based strategy for isolating viable bacteria from multipart soil samples compatible with Raman spectroscopy. *Analytical and Bioanalytical Chemistry*, 409(15), 3779–3788, doi:10.1007/s00216-017-0320-z.

Seaton, F. M., Reinsch, S., Goodall, T., White, N., Jones, D. L., Griffiths, R. I., Creer, S., Smith, A., Emmett, B. A. and Robinson, D. A., 2021, Long-term drought and warming alter soil bacterial and fungal communities in an upland heathland. *Ecosystems*, doi:10.1007/s10021-021-00715-8.

Seneviratne, G. and Jayasinghearachchi, H. S., 2005, A rhizobial biofilm with nitrogenase activity alters nutrient availability in a soil. *Soil Biology and Biochemistry*, 37(10), 1975–1978, doi:10.1016/j.soilbio.2005.02.027.

Sepkoski, J. J., 1988, Alpha, beta, or gamma: Where does all the diversity go? *Paleobiology*, 14(3), 221–234, doi:10.1017/S0094837300011969.

Shade, A. and Gilbert, J. A., 2015, Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, 23(6), 335–340, doi:10.1016/j.tim.2015.01.007.

Shade, A., Gregory Caporaso, J., Handelsman, J., Knight, R. and Fierer, N., 2013, A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME Journal*, 7(8), 1493–1506, doi:10.1038/ismej.2013.54.

Sikorski, J., Baumgartner, V., Birkhofer, K., Boeddinghaus, R. S., Bunk, B., Fischer, M., Fösel, B. U., Friedrich, M. W., Göker, M., Hölzel, N. et al., 2022, The evolution of evological diversity in Acidobacteria. *Frontiers in Microbiology*, 13, 715637, doi:10.3389/fmicb.2022.715637.

Simon, H. M., Smith, M. W. and Herfort, L., 2014, Metagenomic insights into particles and their associated microbiota in a coastal margin ecosystem. *Frontiers in Microbiology*, 5.

Singleton, D. R., Furlong, M. A., Peacock, A. D., White, D. C., Coleman, D. C. and Whitman, W. B. ., 2003, Solirubrobacter pauli gen. nov., sp. nov., a mesophilic bacterium within the Rubrobacteridae related to common soil clones. *International Journal of Systematic and Evolutionary Microbiology*, 53(2), 485–490, doi:10.1099/ijs.0.02438-0.

Söllinger, A., Séneca, J., Borg Dahl, M., Motleleng, L. L., Prommer, J., Verbruggen, E., Sigurdsson, B. D., Janssens, I., Peñuelas, J., Urich, T. et al., 2022, Down-regulation of the bacterial protein biosynthesis machinery in response to weeks, years, and decades of soil warming. *Science Advances*, 8(12), eabm3230, doi:10.1126/sciadv.abm3230.

Stres, B., Danevcic, T., Pal, L., Fuka, M. M., Resman, L., Leskovec, S., Hacin, J., Stopar, D., Mahne, I. and Mandic-Mulec, I., 2008, Influence of temperature and soil water content on bacterial, archaeal and denitrifying microbial communities in drained fen grassland soil microcosms: Influence of T and soil moisture on microbial communities. *FEMS Microbiology Ecology*, 66(1), 110–122, doi:10.1111/j.1574-6941.2008.00555.x.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. and Zeileis, A., 2008, Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307, doi:10.1186/1471-2105-9-307.

Strong, W. L., 2016, Biased richness and evenness relationships within Shannon–Wiener index values. *Ecological Indicators*, 67, 703–713, doi:10.1016/j.ecolind.2016.03.043.

Sukenik, A., Kaplan-Levy, R. N., Welch, J. M. and Post, A. F., 2012, Massive multiplication of genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria). *The ISME Journal*, 6(3), 670–679, doi:10.1038/ismej.2011.128.

Sun, R., Zhang, P., Riggins, C. W., Zabaloy, M. C., Rodríguez-Zas, S. and Villamil, M. B., 2019, Long-term N fertilization decreased diversity and altered the composition of soil bacterial and archaeal communities. *Agronomy*, 9(10), 574, doi:10.3390/agronomy9100574.

Sünnemann, M., Alt, C., Kostin, J. E., Lochner, A., Reitz, T., Siebert, J., Schädler, M. and Eisenhauer, N., 2021, Low-intensity land-use enhances soil microbial activity, biomass and fungal-to-bacterial ratio in current and future climates. *Journal of Applied Ecology*, 58(11), 2614–2625, doi:10.1111/1365-2664.14004.

Swan, B. K., Tupper, B., Sczyrba, A., Lauro, F. M., Martinez-Garcia, M., González, J. M., Luo, H., Wright, J. J., Landry, Z. C., Hanson, N. W. et al., 2013, Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences*, 110(28), 11463–11468, doi:10.1073/pnas.1304246110.

Szoboszlay, M., Dohrmann, A. B., Poeplau, C., Don, A. and Tebbe, C. C., 2017, Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. *FEMS Microbiology Ecology*, 93(12), doi:10.1093/femsec/fix146.

Szoboszlay, M. and Tebbe, C. C., 2021, Hidden heterogeneity and co-occurrence networks of soil prokaryotic communities revealed at the scale of individual soil aggregates. *MicrobiologyOpen*, 10(1), doi:10.1002/mbo3.1144.

Terrat, S., Horrigue, W., Dequietd, S., Saby, N. P. A., Lelièvre, M., Nowak, V., Tripied, J., Régnier, T., Jolivet, C., Arrouays, D. et al., 2017, Mapping and predictive variations of soil bacterial richness across France. *PLOS ONE*, 12(10), e0186766, doi:10.1371/journal.pone.0186766.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G. et al., 2017, A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681), 457–463, doi:10.1038/nature24621.

Torsvik, V., Goks, J. and Daae, F. L., 1989, High Diversity in DNA of Soil Bacteria. *APPL. ENVIRON. MICROBIOL.*, p. 6.

Toutenburg, H., 1971, Fisher, R. A., and F. Yates: Statistical Tables for Biological, Agricultural and Medical Research. 6th Ed. Oliver & Boyd, Edinburgh and London 1963. X, 146 P. Preis 42 s net. *Biometrische Zeitschrift*, 13(4), 285–285, doi:10.1002/bimj.19710130413.

Tukey, J. W., 1991, The Philosophy of Multiple Comparisons. *Statistical Science*, 6(1), 100–116.

Tuson, H. H. and Weibel, D. B., 2013, Bacteria–surface interactions. *Soft Matter*, 9(17), 4368, doi:10.1039/c3sm27705d.

van der Bom, F., Nunes, I., Raymond, N. S., Hansen, V., Bonnichsen, L., Magid, J., Nybroe, O. and Jensen, L. S., 2018, Long-term fertilisation form, level and duration affect the diversity, structure and functioning of soil microbial communities in the field. *Soil Biology and Biochemistry*, 122, 91–103, doi:10.1016/j.soilbio.2018.04.003.

Wagg, C., Bender, S. F., Widmer, F. and van der Heijden, M. G. A., 2014, Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences*, 111(14), 5266–5270, doi:10.1073/pnas.1320054111.

Walker, T. W. N., Kaiser, C., Strasser, F., Herbold, C. W., Leblans, N. I. W., Woebken, D., Janssens, I. A., Sigurdsson, B. D. and Richter, A., 2018, Microbial temperature sensitivity and biomass change explain soil carbon loss with warming. *Nature Climate Change*, 8(10), 885–889, doi:10.1038/s41558-018-0259-x.

Wallenstein, M. D., McNulty, S., Fernandez, I. J., Boggs, J. and Schlesinger, W. H., 2006, Nitrogen fertilization decreases forest soil fungal and bacterial biomass in three long-term experiments. *Forest Ecology and Management*, 222(1), 459–468, doi:10.1016/j.foreco.2005.11.002.

Wang, G. and Or, D., 2010, Aqueous films limit bacterial cell motility and colony expansion on partially saturated rough surfaces. *Environmental Microbiology*, 12(5), 1363–1373, doi:10.1111/j.1462-2920.2010.02180.x.

Wang, G. and Or, D., 2013, Hydration dynamics promote bacterial coexistence on rough surfaces. *The ISME Journal*, 7(2), 395–404, doi:10.1038/ismej.2012.115.

Waring, B. and Hawkes, C. V., 2018, Ecological mechanisms underlying soil bacterial responses to rainfall along a steep natural precipitation gradient. *FEMS Microbiology Ecology*, 94(2), doi:10.1093/femsec/fiy001.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E. et al., 1987, Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic and Evolutionary Microbiology*, 37(4), 463–464, doi:10.1099/00207713-37-4-463.

Wegner, C.-E. and Liesack, W., 2017, Unexpected dominance of elusive Acidobacteria in early industrial soft coal slags. *Frontiers in Microbiology*, 0, doi:10.3389/fmicb.2017.01023.

Wei, D., Yang, Q., Zhang, J.-Z., Wang, S., Chen, X.-L., Zhang, X.-L. and Li, W.-Q., 2008, Bacterial community structure and diversity in a black soil as affected by long-term fertilization. *Pedosphere*, 18(5), 582–592, doi:10.1016/S1002-0160(08)60052-1.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A. et al., 2017, Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), doi:10.1186/s40168-017-0237-y.

Whitfield, C., 1988, Bacterial extracellular polysaccharides. *Canadian Journal of Microbiology*, 34(4), 415–420, doi:10.1139/m88-073.

Whittaker, R. H., 1972, Evolution and measurement of species diversity. *TAXON*, 21(2-3), 213–251, doi:10.2307/1218190.

Wood, M., 1995, A mechanism of aluminium toxicity to soil bacteria and possible ecological implications. *Plant and Soil*, 171, 63–69.

Worden, A. Z. and Binder, B. J., 2003, Growth regulation of rRNA content in Prochlorococcus and Synechococcus (marine Cyanobacteria) measured by whole-cell hybridization of rRNA-targeted peptide nucleic acids. *Journal of Phycology*, 39(3), 527–534, doi:10.1046/j.1529-8817.2003.01248.x.

Wu, L., Zhang, Y., Guo, X., Ning, D., Zhou, X., Feng, J., Yuan, M. M., Liu, S., Guo, J., Gao, Z. et al., 2022, Reduction of microbial diversity in grassland

soil is driven by long-term climate warming. *Nature Microbiology*, 7(7), 1054–1062, doi:10.1038/s41564-022-01147-3.

Wu, Y., Cai, P., Jing, X., Niu, X., Ji, D., Ashry, N. M., Gao, C. and Huang, Q., 2019, Soil biofilm formation enhances microbial community diversity and metabolic activity. *Environment International*, 132, 105116, doi:10.1016/j.envint.2019.105116.

Wüst, P. K., Nacke, H., Kaiser, K., Marhan, S., Sikorski, J., Kandeler, E., Daniel, R. and Overmann, J., 2016, Estimates of soil bacterial ribosome content and diversity are significantly affected by the nucleic acid extraction method employed. *Applied and Environmental Microbiology*, 82(9), 2595–2607, doi:10.1128/AEM.00019-16.

Xu, A., Li, L., Coulter, J. A., Xie, J., Gopalakrishnan, S., Zhang, R., Luo, Z., Cai, L., Liu, C., Wang, L. et al., 2020, Long-term nitrogen fertilization impacts on soil bacteria, grain yield and nitrogen use efficiency of wheat in semiarid loess plateau, China. *Agronomy*, 10(8), 1175, doi:10.3390/agronomy10081175.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O., 2014, The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, 42(D1), D643–D648, doi:10.1093/nar/gkt1209.

Zeng, J., Liu, X., Song, L., Lin, X., Zhang, H., Shen, C. and Chu, H., 2016, Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition. *Soil Biology and Biochemistry*, 92, 41–49, doi:10.1016/j.soilbio.2015.09.018.

Zhang, L., Zhu, L., Si, M., Li, C., Zhao, L., Wei, Y. and Shen, X., 2014, Solirubrobacter taibaiensis sp. nov., isolated from a stem of Phytolacca acinosa Roxb. *Antonie van Leeuwenhoek*, 106(2), 279–285, doi:10.1007/s10482-014-0194-4.

Zhang, P., Zheng, J., Pan, G., Zhang, X., Li, L. and Rolf, T., 2007, Changes in microbial community structure and function within particle size fractions of a paddy soil under different long-term fertilization treatments from the Tai Lake region, China. *Colloids and Surfaces B: Biointerfaces*, 58(2), 264–270, doi:10.1016/j.colsurfb.2007.03.018.

Zhang, X., Zhang, G., Chen, Q. and Han, X., 2013, Soil bacterial communities respond to climate changes in a temperate steppe. *PLOS ONE*, 8(11), e78616, doi:10.1371/journal.pone.0078616.

Zhang, Y., Heal, K. V., Shi, M., Chen, W. and Zhou, C., 2022, Decreasing molecular diversity of soil dissolved organic matter related to microbial community along an alpine elevation gradient. *Science of The Total Environment*, 818, 151823, doi:10.1016/j.scitotenv.2021.151823.

Zheng, M., Zhu, P., Zheng, J., Xue, L., Zhu, Q., Cai, X., Cheng, S., Zhang, Z., Kong, F. and Zhang, J., 2021, Effects of soil texture and nitrogen fertilisation on soil bacterial community structure and nitrogen uptake in flue-cured tobacco. *Scientific Reports*, 11(1), 22643, doi:10.1038/s41598-021-01957-1.

Zheng, Y., Ji, N.-N., Wu, B.-W., Wang, J.-T., Hu, H.-W., Guo, L.-D. and He, J.-Z., 2020, Climatic factors have unexpectedly strong impacts on soil bacterial β-diversity in 12 forest ecosystems. *Soil Biology and Biochemistry*, 142, 107699, doi:10.1016/j.soilbio.2019.107699.

Zhou, Z., Wang, C. and Luo, Y., 2020, Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nature Communications*, 11(1), 3072, doi:10.1038/s41467-020-16881-7.

Supplementary Material

Biodiversity Exploratories diversity analysis

QIIME2 pipeline code

QIIME2 pipeline code used with version 2019.1 processing the Illumina sequencing data of the V3 region of 16S rRNA transcripts from bulk soil samples of the Biodiversity Exploratories.

```
#!/bin/sh
#$ -S /bin/bash

#SBATCH --mail-type=END
#SBATCH --mail-user=cam17@dsmz.de

echo "START TIME: " `date`


while getopts l:s:r:f:t: option
do
 case "${option}"
 in
 l) LENGTH=${OPTARG};;
 s) MINSIZE=${OPTARG};;
 r) MINREADS=${OPTARG};;
 f) FOLDER=${OPTARG};;
 t) THREAD=${OPTARG};;
 esac
done

cd $FOLDER

if [ ! -f joined-demux.qza ];
then
    echo "import"
qiime tools import --type 'SampleData[JoinedSequencesWithQuality]' --
    input-path pe-33-manifest-4S --output-path joined-demux.qza --input-
    format SingleEndFastqManifestPhred33
fi

 ## quality filtering
if [ ! -f demux-joined-filtered.qza ] && [ ! -f demux-joined-filter-
    stats.qza ];
```

```
then
        echo "quality filtering"
qiime quality-filter q-score-joined --i-demux joined-demux.qza --p-min-
    quality 25 --p-quality-window 3 --p-min-length-fraction 0.75 --o-
    filtered-sequences demux-joined-filtered.qza --o-filter-stats demux-
    joined-filter-stats.qza
fi

   ## sequencing scoring after quality filtering
if [ -f demux-joined-filtered.qza ];
then
echo "visualize after quality filtering"
 qiime demux summarize --i-data demux-joined-filtered.qza --o-
     visualization demux-joined-filtered.qzv
fi

        ## visualize stats
if [ -f demux-joined-filter-stats.qza ];
then
echo "visualize after quality filtering"
 qiime metadata tabulate --m-input-file demux-joined-filter-stats.qza
     --o-visualization demux-joined-filter-stats.qzv
fi

        ## deblur
if [ ! -f rep-seqs.qza ] && [ ! -f table.qza ] && [ ! -f deblur-
    stats.qza ];
then
echo "qiime deblur denoise-16S --i-demultiplexed-seqs demux-joined-
    filtered.qza --p-min-size $MINSIZE --p-min-reads $MINREADS --p-trim-
    length $LENGTH --o-representative-sequences rep-seqs.qza --o-table
    table.qza --p-sample-stats --p-jobs-to-start 10 --o-stats deblur-
    stats.qza"
 qiime deblur denoise-16S --i-demultiplexed-seqs demux-joined-filtered.
     qza --p-min-size $MINSIZE --p-min-reads $MINREADS --p-trim-length
     $LENGTH --o-representative-sequences rep-seqs.qza --o-table table.
     qza --p-sample-stats --p-jobs-to-start $THREAD --o-stats deblur-
     stats.qza
fi

 ## visualize after deblur
if [ -f rep-seqs.qza ];
then
```

```
echo "visualize after deblur rep-seqs"
 qiime feature-table tabulate-seqs --i-data rep-seqs.qza --o-
    visualization rep-seqs.qzv
fi
if [ -f table.qza ];
then
echo "visualize after deblur table"
 qiime feature-table summarize --i-table table.qza --o-visualization
    table.qzv
fi
if [ -f deblur-stats.qza ];
then
echo "visualize after deblur stats"
 qiime deblur visualize-stats --i-deblur-stats deblur-stats.qza --o-
    visualization deblur-stats.qzv
fi

if [ ! -f aligned-rep-seqs.qza ];
then
        echo "mafft"
qiime alignment mafft --i-sequences rep-seqs.qza --o-alignment aligned-
   rep-seqs.qza --p-n-threads $THREAD
fi

if [ ! -f masked-aligned-rep-seqs.qza ];
then
        echo "mask"
qiime alignment mask --i-alignment aligned-rep-seqs.qza --o-masked-
   alignment masked-aligned-rep-seqs.qza
fi

 ## Phylogeny --> can be threaded, but not recommendable, cause tree
    could be different
if [ ! -f unrooted-tree.qza ];
then
        echo "unrooted-tree"
qiime phylogeny fasttree --i-alignment masked-aligned-rep-seqs.qza --o-
   tree unrooted-tree.qza --p-n-threads $THREAD
fi

if [ ! -f rooted-tree.qza ];
then
        echo "rooted-tree"
```

```
qiime phylogeny midpoint-root --i-tree unrooted-tree.qza --o-rooted-
    tree rooted-tree.qza
fi

# used the classifier from Selma
if [ ! -f taxonomy.qza ];
then
echo "silva taxonomy"
qiime feature-classifier classify-sklearn --i-classifier "/home/cam17/
    Documents/Taxonomy_classifier/classifier_132_V3.qza" --p-n-jobs
    $THREAD --i-reads rep-seqs.qza --o-classification taxonomy.qza
fi



qiime tools export \
--input-path taxonomy.qza \
--output-path taxonomy-with-spaces
qiime metadata tabulate \
  --m-input-file taxonomy-with-spaces/taxonomy.tsv  \
  --o-visualization taxonomy-as-metadata.qzv
qiime tools export \
--input-path taxonomy-as-metadata.qzv \
--output-path taxonomy-as-metadata
qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-path taxonomy-as-metadata/metadata.tsv \
  --output-path taxonomy-without-spaces.qza



## remove mitochondria and chloroplast sequences (optional), based on
    taxonomy.qza
if [ ! -f table-no-MitChloro.qza ];
then
qiime taxa filter-table --i-table table.qza --i-taxonomy taxonomy-
    without-spaces.qza --p-exclude mitochondria,chloroplast --o-filtered
    -table table-no-MitChloro.qza
fi

# table.qza is changed to feature-table.biom, taxonomy.qza -> taxonomy.
    tsv, rooted-tree.qza -> tree.nwk
if [ -f table.qza ];
then
echo "export"
```

```
qiime tools export --input-path table-no-MitChloro.qza --output-path
    exported
fi

if [ -f taxonomy.qza ];
then
echo "export"
qiime tools export --input-path taxonomy.qza --output-path exported
fi

if [ -f rooted-tree.qza ];
then
echo "export"
qiime tools export --input-path rooted-tree.qza --output-path exported
fi

if [ -f exported ];
then
cd exported
fi

if [ -f feature-table.biom ] && [ -f taxonomy.tsv ] && [ -f tree.nwk ];
then
cp taxonomy.tsv biom-taxonomy.tsv
fi

echo "END TIME: " `date`
```

Predictor correlation

Table 7: Overview over all engineered predictors created for modelling of bacterial alpha and beta diversity.

data	temporal aggregation (days before sampling)	predictor	
soil moisture	1056 days	thresholds	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
		slope	
soil moisture	730 days	thresholds	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
soil moisture	365 days	threshold	1-63 % in 1 % steps
		minimum	
		maximum	
		range	

soil moisture	180 days	threshold	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
soil moisture	90 days	threshold	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
soil moisture	30 days	threshold	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
soil moisture	14 days	threshold	1-63 % in 1 % steps
		minimum	
		maximum	
		range	
soil temperature	1056 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
		slope	
soil temperature	730 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
soil temperature	365 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
soil temperature	180 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
soil temperature	90 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
soil temperature	30 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
soil temperature	14 days	threshold	-36-29°C in 1°C steps
		minimum	
		maximum	
		range	
precipitation	1056 days	sum rain days	
	730 days	sum rain days	
	365 days	sum rain days	
land use intensity	5 years	mean LUI	
		mean grazing	
		mean mowing	
		mean fertilization	
		slope LUI	
		slope grazing	
		slope mowing	
		slope fertilization	
land use intensity	4 years	mean LUI	
		mean grazing	
		mean mowing	
		mean fertilization	
land use intensity	3 years	mean LUI	

		mean grazing
		mean mowing
		mean fertilization
land use intensity	2 years	mean LUI
		mean grazing
		mean mowing
		mean fertilization
land use intensity	1 years	mean LUI
		mean grazing
		mean mowing
		mean fertilization

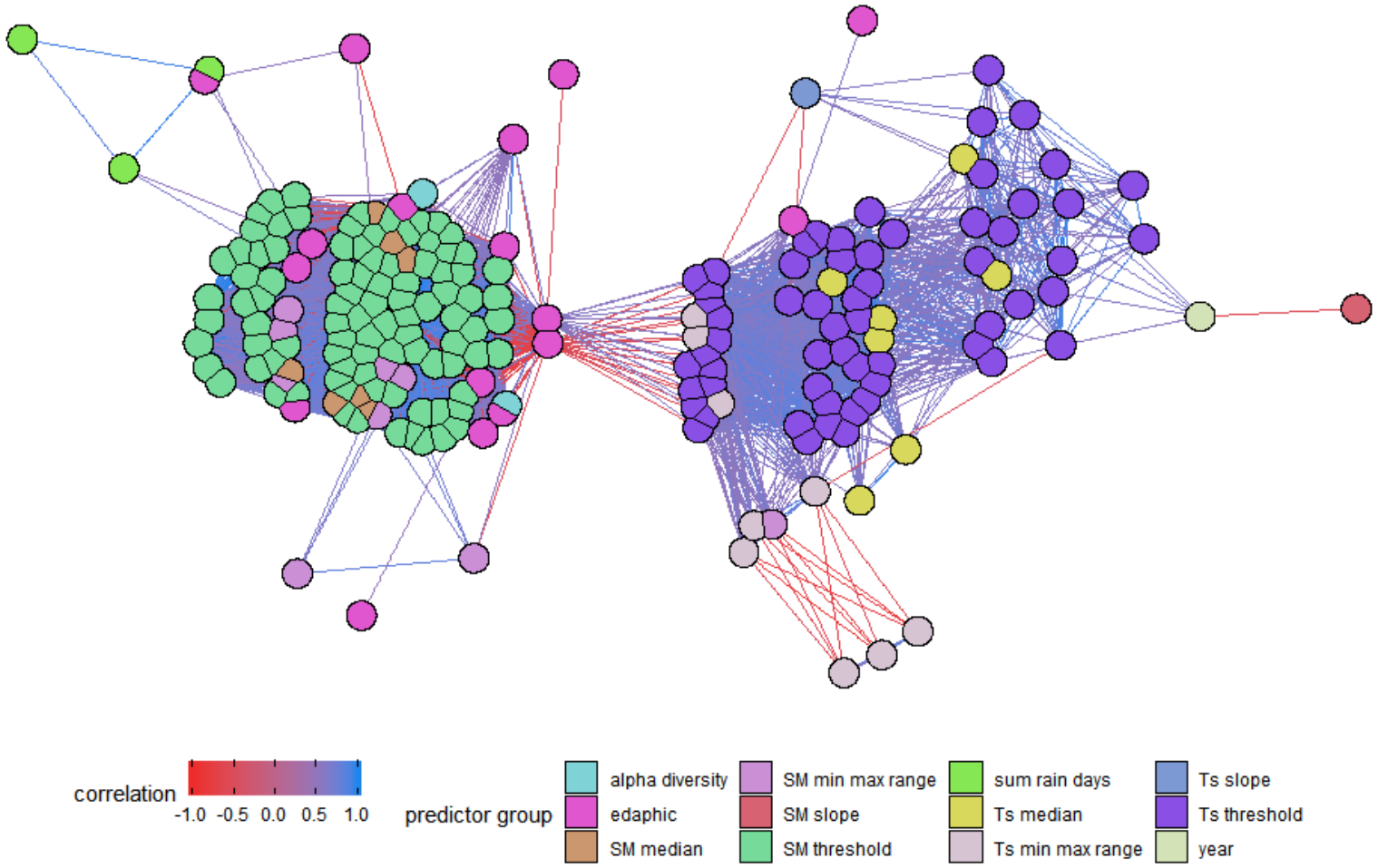

Figure 49: Pearson correlation network of all climatic predictors used in this study and sampling year. For better readability predictor groups were color coded. Only Pearson correlation coefficients >0.5 are displayed. SM = soil moisture, Ts = soil temperature.

Table 8: Edaphic predictors used in this study.

Predictor	definition	explanation
Wide coarse pore content (%)	> 50 µm	crucial for soil aeration and fast drainage
Fine coarse pore content (%)	> 10 µm	crucial for soil aeration and fast drainage
Medium sized pore content (%)	0.2-10 µm	storage of plant available water
Fine pore content (%)	< 0.2 µm	water stored in fine pores
pF 1.8		soil water content in pore size > 10 µm
pf 2.5		soil water content in pore size 0.2-10 µm
pF 4.2		soil water content in pore size < 0.2 µm
Porosity	%	proportion of potential air filled soil volume
Bbulk density	density of sol (g/cm^3)	
Sand content	g/kg	0.063-2 mm
Clay content	g/kg	<0.002 mm
Dry matter content	%	proportion of dry matter in a defined soil volume
Dithionite extractable Fe	g/kg soil	
Oxalate extractable Fe	g/kg soil	
Dithionite extractable Al	g/kg soil	
Oxalate extractable Al	g/kg soil	
Nmin	µg N g^{-1} soil DM	nitrate + ammonium content
NH_4^+	µg N g^{-1} soil DM	ammonium content
NO_3	µg N g^{-1} soil DM	nitrate content
EN		extractable nitrogen
EOC		extractable organic carbon
Total N	g/kg soil	
Total C	g/kg soil	
Organic C	g/kg soil	
C/N ratio	ratio of organic C to total N content in soil	well-established indicator for the nitrogen availability in soil
Organic S	mg/kg soil	calculated organic S concentration for extraction with KH_2PO_4
Total S	mg/kg soil	total S concentration
Inorganic S	mg/kg soil	inorganic S concentration after extraction with KH_2PO_4
N_{mic}	µg Nmic g^{-1} DM	microbial carbon (without KEN-Factor)
C_{mic}	µg Cmic g^{-1} DM	microbial carbon (without KEC-Factor)
C_{mic}/N_{mic}		
Invertebrates PLFA	nmol FAME g-1 soil DM	invertebrate marker PLFA FAME 20:4?6,9,12,15
Fungi PLFA	nmol FAME g-1 soil DM	PLFA FAME 18:26,9
Fungi/Bacteria PLFA	nmol FAME g-1 soil DM	ratio fungi to bacteria
Ts	soil temperature in 10 cm depth (°C)	
SWG	gravimetric water content (%)	
pH	hydronium ion activity in a soil suspension	
Sample year	year of sampling	2011, 2014 or 2017
Plant biomass	g/m^2	weight of dried biomass in gram per m^2
Soil type		soil type according World Reference Base

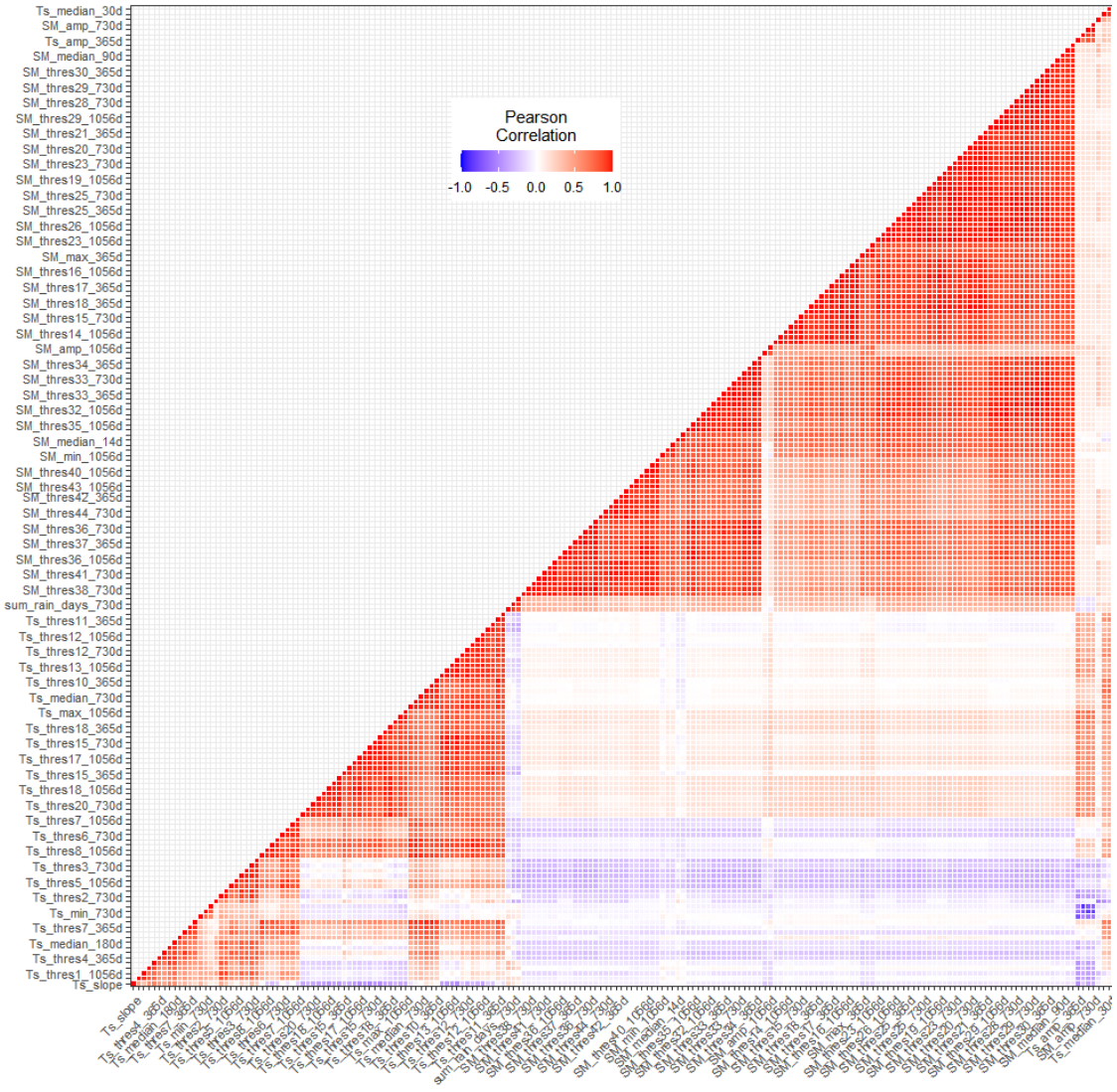

Figure 50: Pearson correlation of engineered climatic predictors available for all land use types. Climate predictors are clustered by similarity. For better readability overlapping predictor labels were dropped.

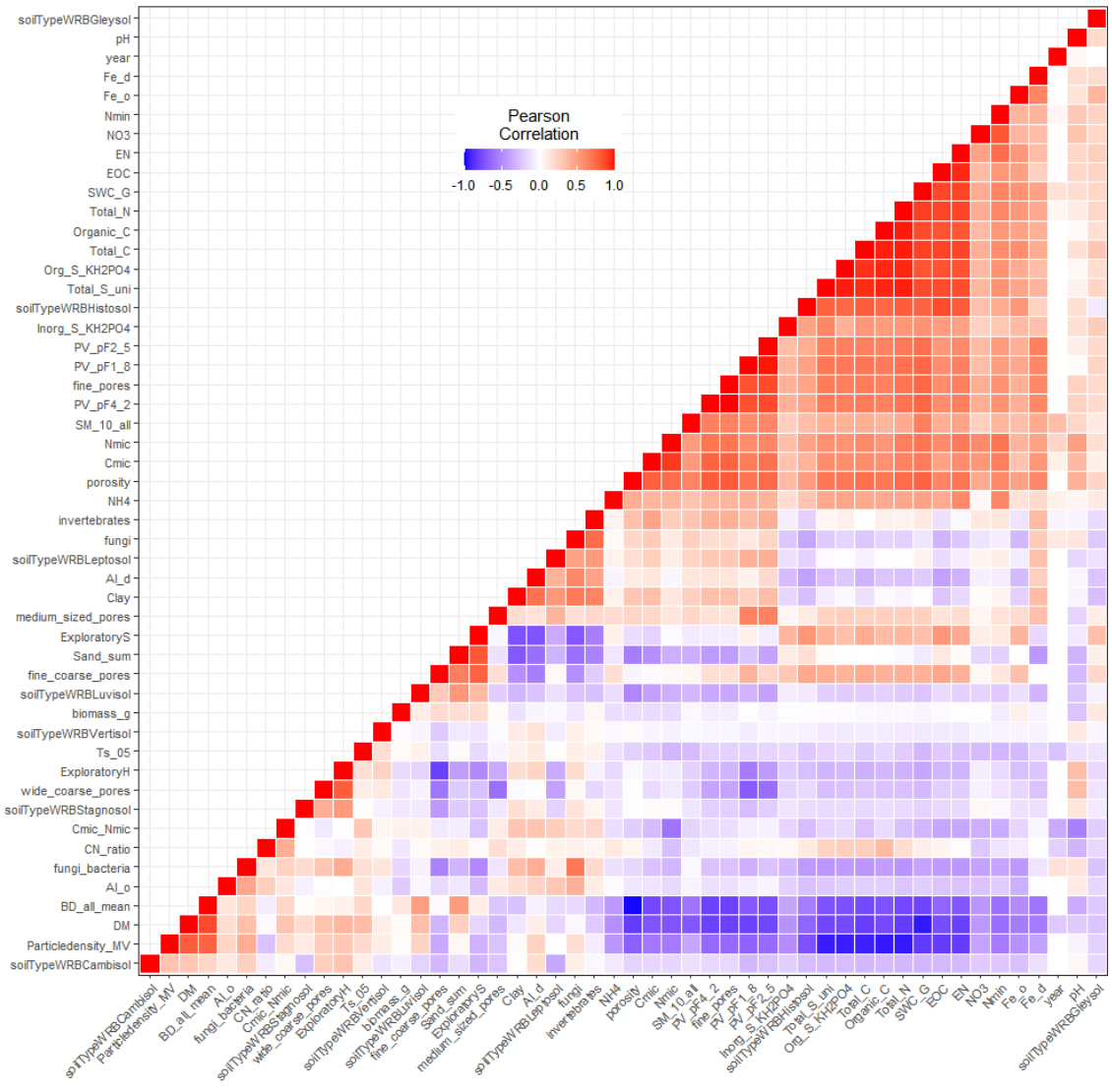

Figure 51: Pearson correlation of edaphic predictor groups for grassland sites.

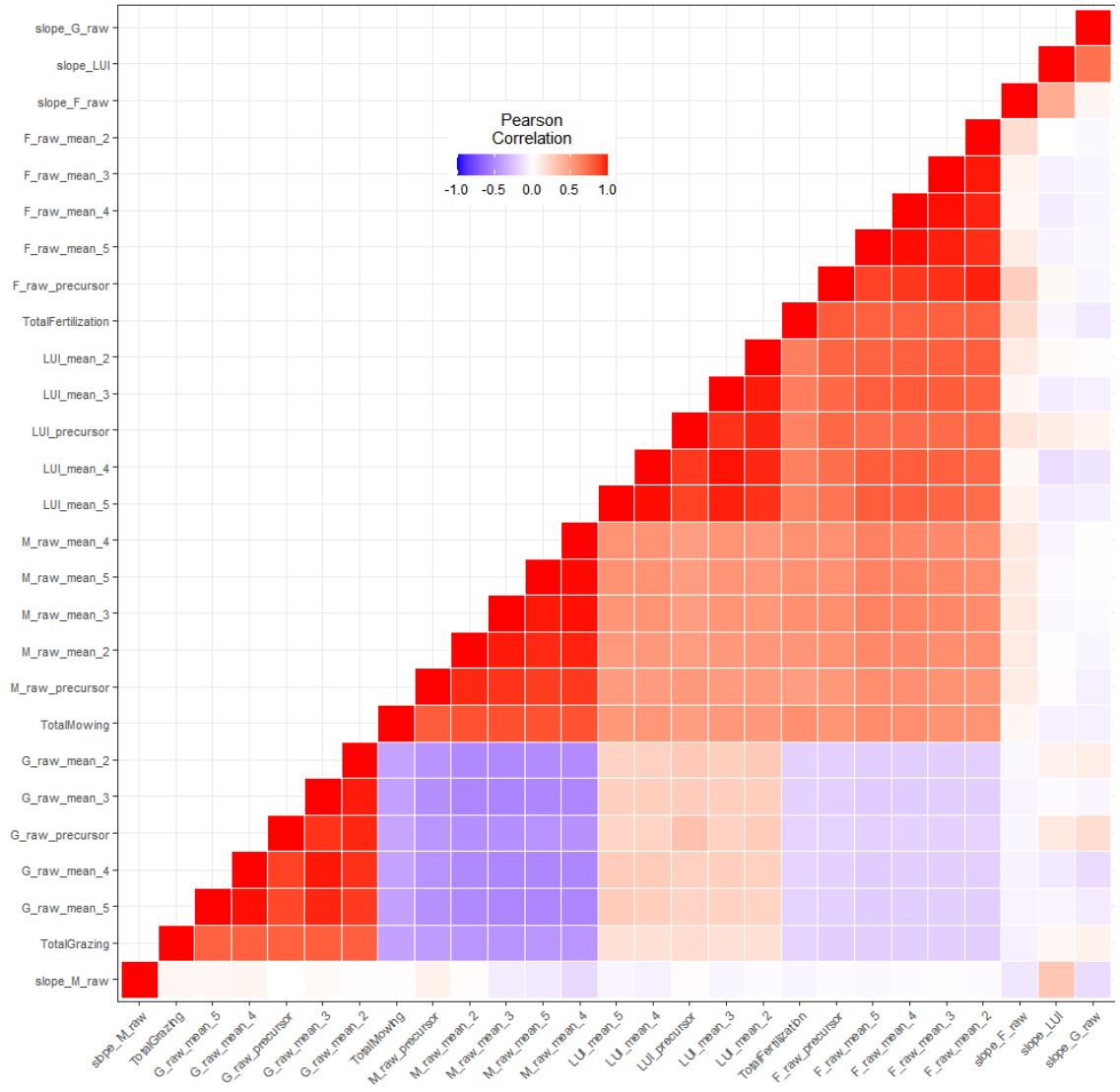

Figure 52: Pearson correlation of land use intensity and raw components grazing, moving, N-fertilization of grassland sites.

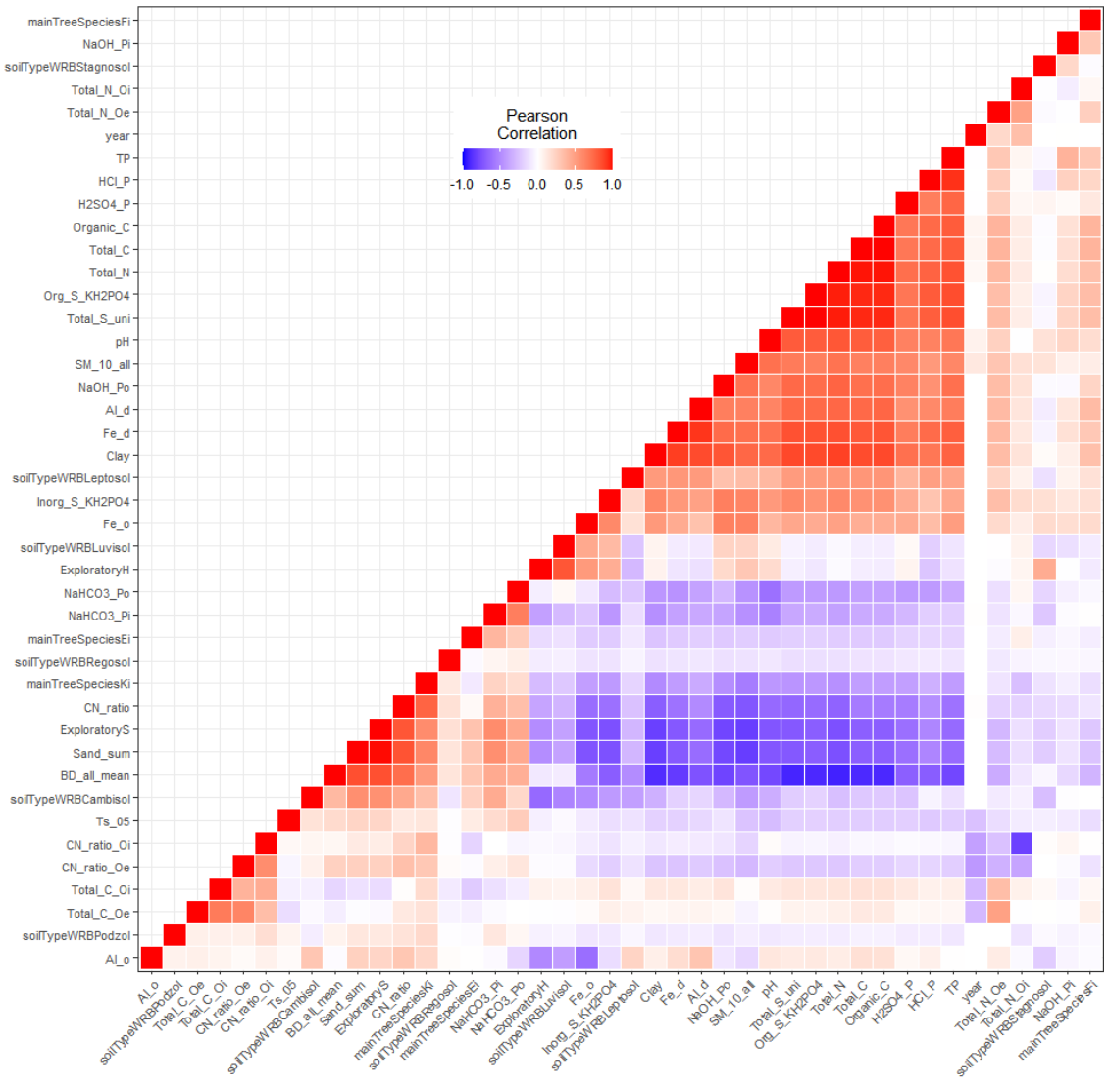

Figure 53: Pearson correlation of edaphic predictor groups for forest sites.

GAMs for alpha diversity

First, GAMs were trained on richness separately for grassland and forest sites, since only a small subset of edaphic predictors were available for both land use types. However, none of the exclusively available predictors were selected for forest or grassland models by forward selection (Fig. 54 and 55). The grassland model explained 60.8 % of deviance explained with year, pH and Ts >5°C in 730 d as significant ($p<0.05$) predictors. Relative variable importance showed temporal variation to be the most important predictor.

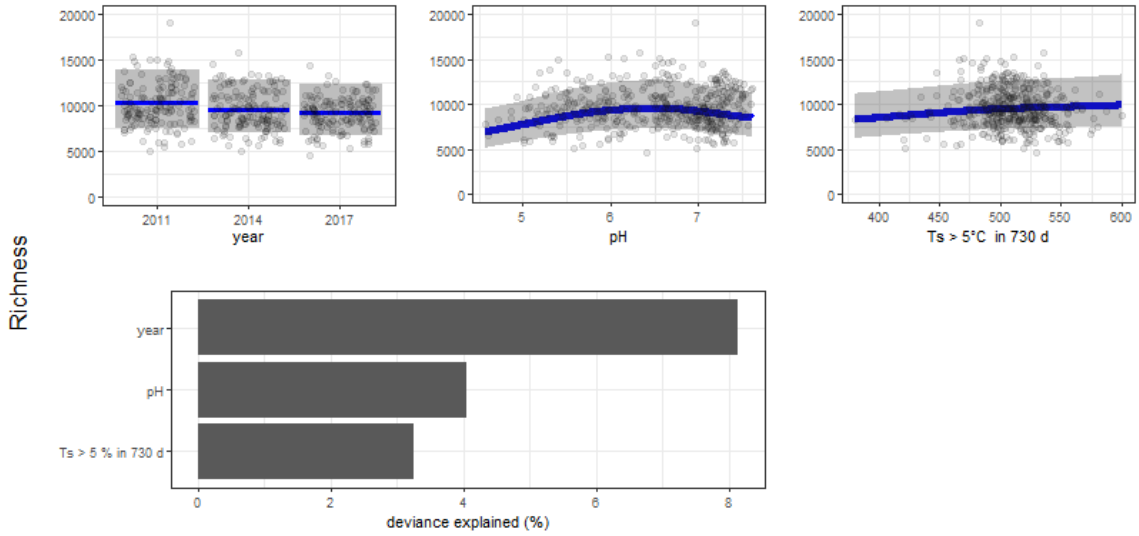

Figure 54: GAM model of bacterial richness fit on grassland sites. Subplots show the partial fit of all predictors as blue lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

The GAM fit on forest sites, identified year, pH, soil moisture as SM >21 % in 365 d and range of SM in 365 d, organic C content, soil texture as sand content and soil temperature as median Ts in 730 d. Again temporal effects and pH were the most important predictors. All other predictors except range of SM in 365 d showed a negative partial variable importance, which implied that the partial effects of those predictors could be replaced by all other predictors and have no true explanatory power for richness of forest sites. However, in total the model accounted for 90.2 % of data variation.

The fit of abundant bacteria, represented by Simpson, on grassland sites identified year, pH, sand content and SM >42 % in 1056 d as significant predictors, explaining 85.6 % of model deviance. The analysis of Simpson for forest sites identified year, pH, range of SM in 365 d and median Ts over 90 d prior to sampling as significant predictors accounting for 86.5 % of deviance explained. Unfortunately, no exclusive predictors for grassland or forest

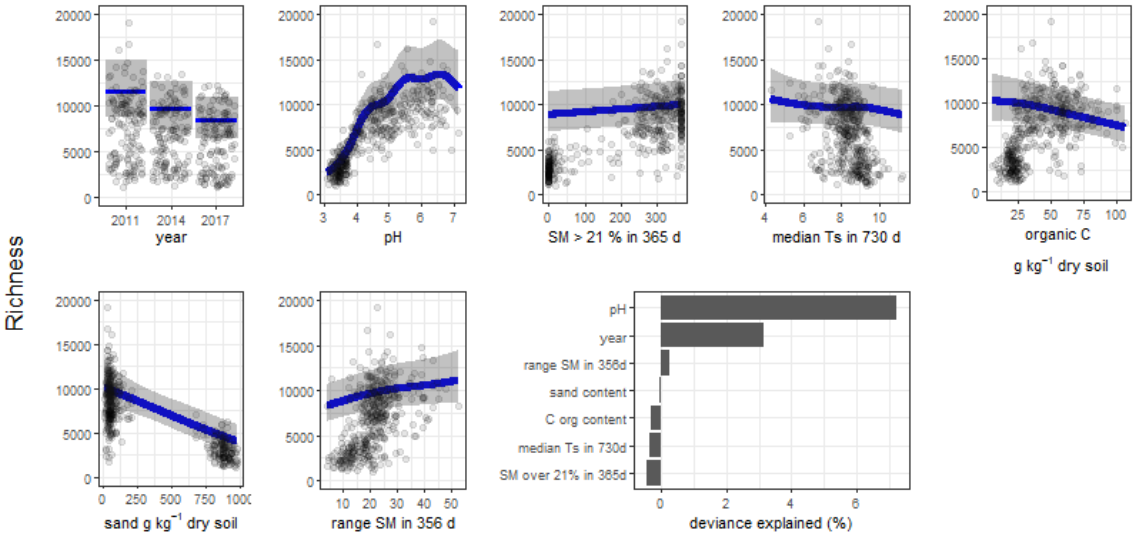

Figure 55: GAM model of bacterial richness fit on all forest sites. Subplots show the partial fit of all predictors as blue lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

sites were selected. Therefore alpha diversity in the form of richness and Simpson were modelled independent of land use, since no exclusively available predictor was selected. In case a significant effect of land use type is present, land use was included as categorical predictor during model fitting.

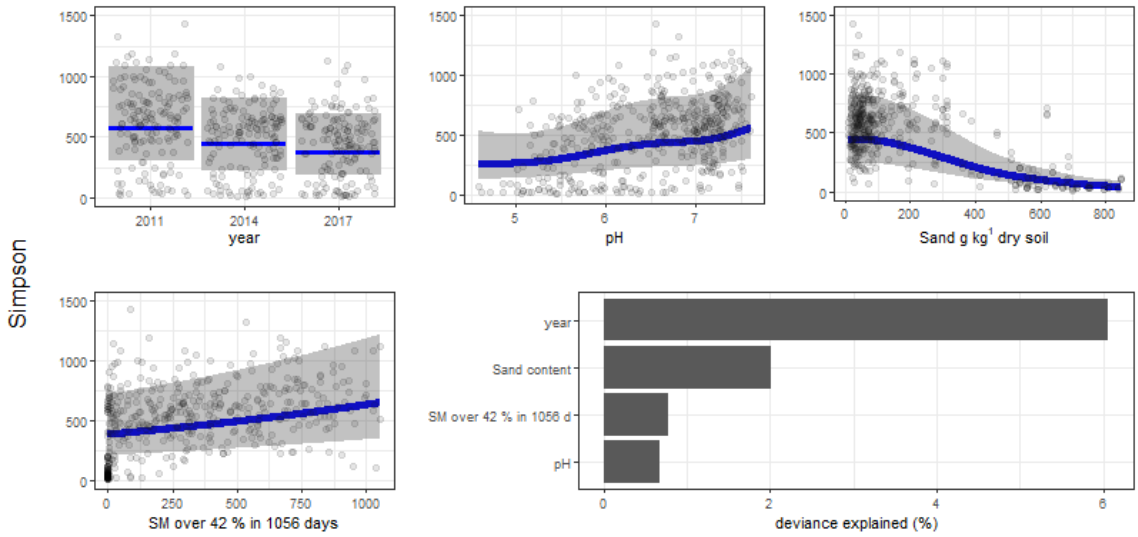

Figure 56: GAM fitted to Simpson diversity for grassland sites. Subplots show the partial fit of all significant predictors as blue lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

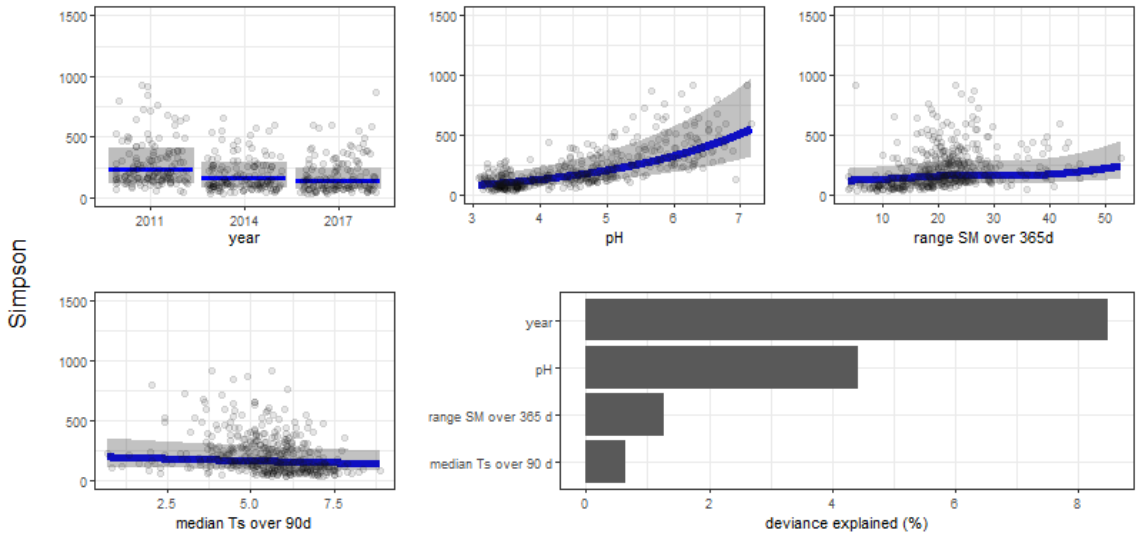

Figure 57: GAM fitted to Simpson diversity for forest sites. Subplots show the partial fit of all significant predictors as blue lines in the order of detection by forward selection. Shaded grey areas represent 95 % confidence intervals. Points display raw data for each predictor. Variable importance is displayed as partial deviance explained for each predictor.

Trajectory clustering

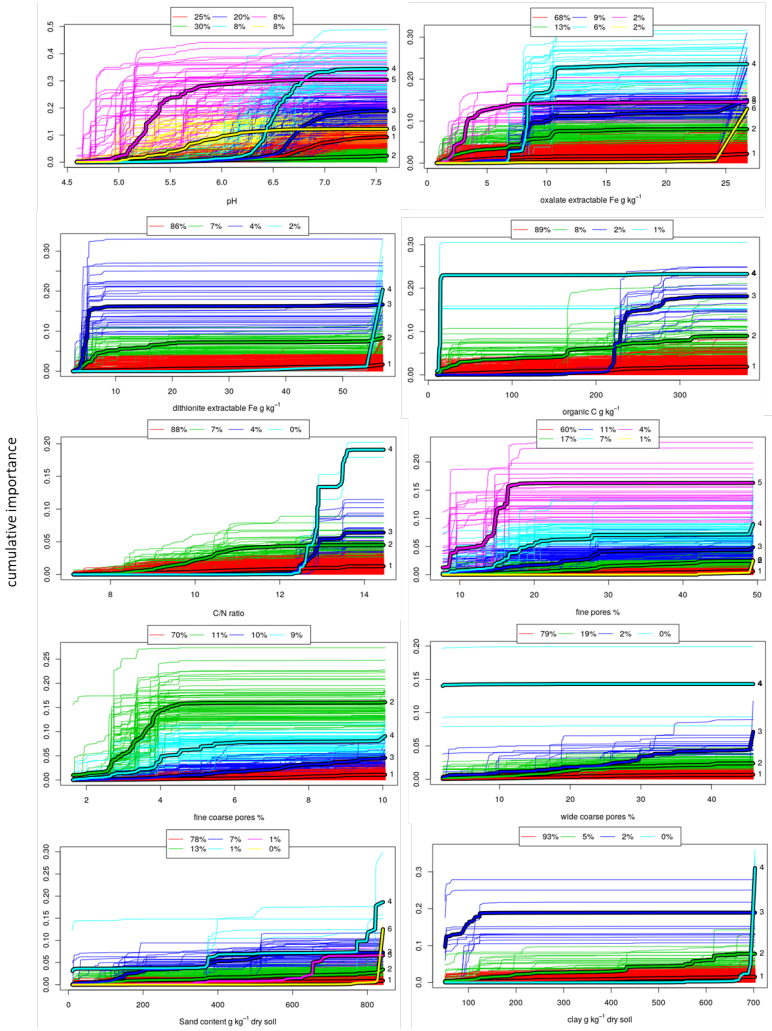

Figure 58: Trajectory clustering of cumulative variable importance for 740 SV of $R^2 > 0.5$ in gradient Forest analysis of grassland communities. The x-axis display the on 10 most important predictors. Different colour define clusters of soil bacteria being similar in their abundance turnover as identified by trajectory clustering. Boxes above each plot depict the proportion of bacteria in each cluster.

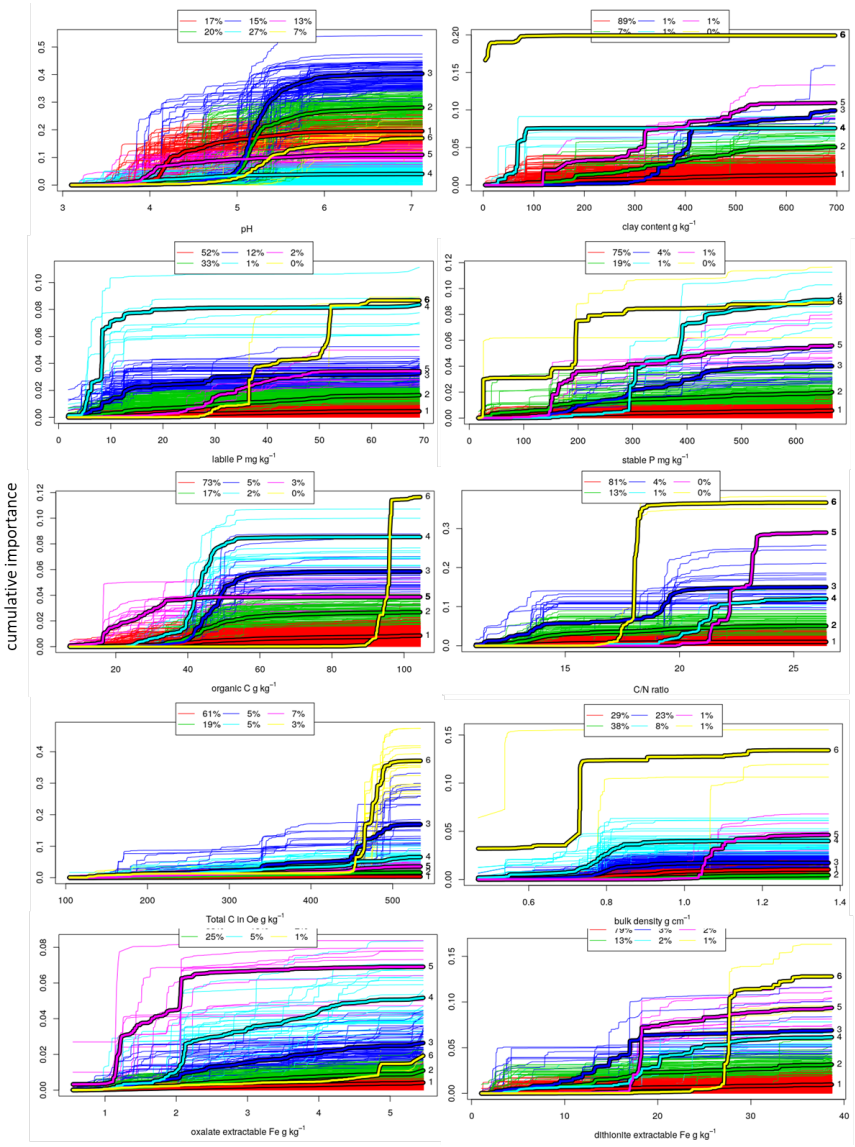

Figure 59: Trajectory clustering of cumulative variable importance for 550 SV of $R^2 > 0.5$ in gradient Forest analysis of forest communities. The x-axis display the on 10 most important predictors. Different colour define clusters of soil bacteria being similar in their abundance turnover as identified by trajectory clustering. Boxes above each plot depict the proportion of bacteria in each cluster.

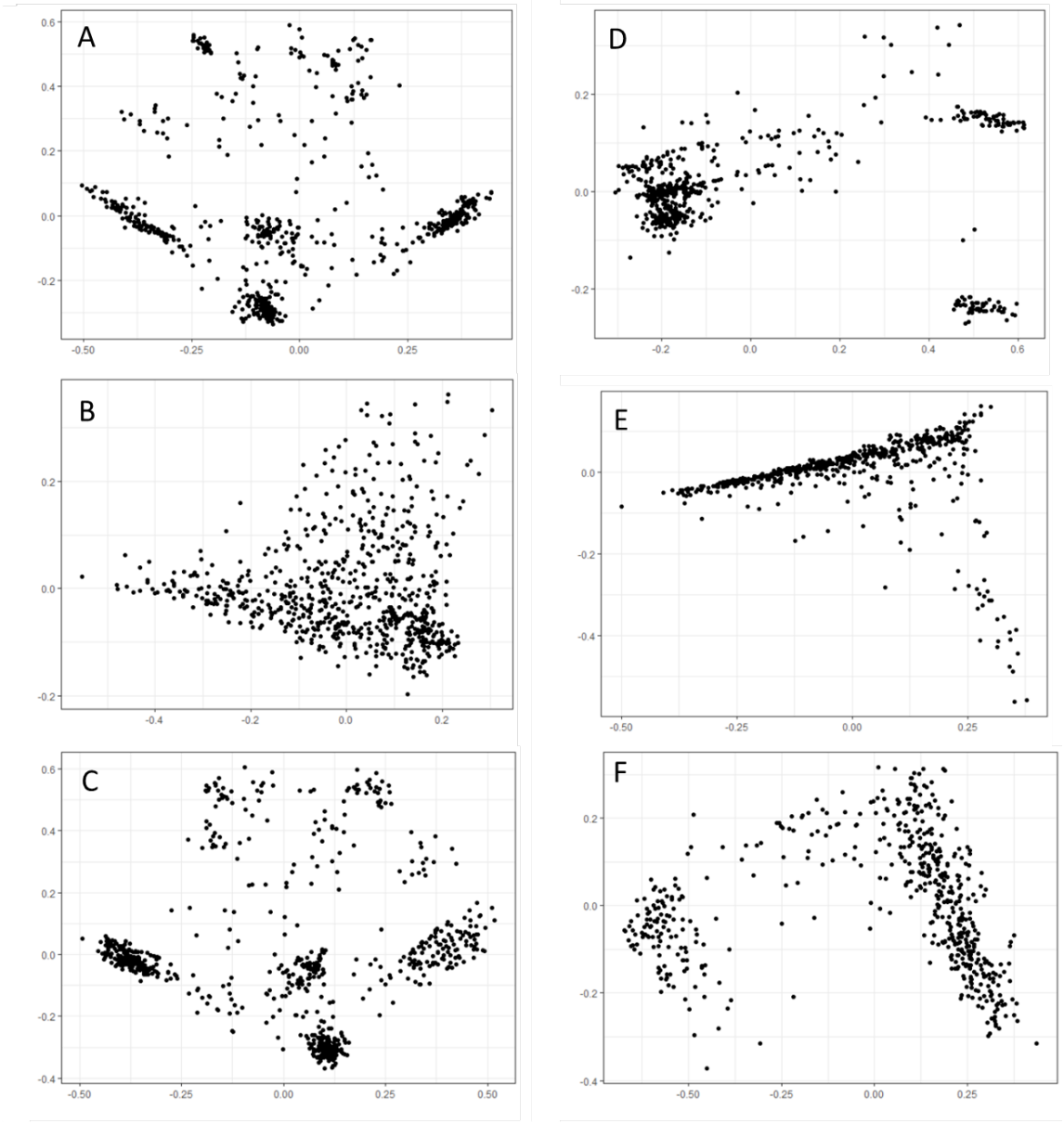

Figure 60: PCoA ordinations of euclidean of 728 SV and 550 SV from grassland (in plots A,B,C) and forest (in plots D, E, F) communities. Each black dot represents a single species. Plot A and D represent the ordination based on euclidean distances of optimal niches values from the 10 most important predictors identified by grandient Forest analysis. Plot B and E represent the ordination of euclidean distances of SV-specific variable importances for grassland and forest communities. Plots C and F contain the synthesis (addition) of euclidean distance matrices of SV-specific optimal niche values and variable importances.

Soil bacteria extraction Bad Lauchstädt

QIIME2 pipeline code

QIIME2 pipeline code used with version 2019.1 processing the Illumina MiSeq sequencing data of V3/V4 region of 16S rRNA and rRNA gene samples of soil bacteria extraction from different fertilization treatments from the static fertilization experiment in Bad Lauchstädt.

```
#!/bin/sh
#$ -S /bin/bash

#SBATCH --mail-type=END
#SBATCH --mail-user=cam17@dsmz.de

echo "START TIME: " `date`


while getopts l:s:r:f:t: option
do
 case "${option}"
 in
 l) LENGTH=${OPTARG};;
 s) MINSIZE=${OPTARG};;
 r) MINREADS=${OPTARG};;
 f) FOLDER=${OPTARG};;
 t) THREAD=${OPTARG};;
 esac
done

cd $FOLDER

if [ ! -f joined-demux.qza ];
then
    echo "import"
qiime tools import --type 'SampleData[JoinedSequencesWithQuality]' --
   input-path pe-33-manifest-4S --output-path joined-demux.qza --input-
   format SingleEndFastqManifestPhred33
fi

 ## quality filtering
if [ ! -f demux-joined-filtered.qza ] &&  [ ! -f demux-joined-filter-
   stats.qza ];
then
        echo "quality filtering"
```

```
qiime quality-filter q-score-joined --i-demux joined-demux.qza --p-min-
    quality 25 --p-quality-window 3 --p-min-length-fraction 0.75 --o-
    filtered-sequences demux-joined-filtered.qza --o-filter-stats demux-
    joined-filter-stats.qza
fi

   ## sequencing scoring after quality filtering
if [ -f demux-joined-filtered.qza ];
then
echo "visualize after quality filtering"
 qiime demux summarize --i-data demux-joined-filtered.qza --o-
     visualization demux-joined-filtered.qzv
fi

        ## visualize stats
if [ -f demux-joined-filter-stats.qza ];
then
echo "visualize after quality filtering"
 qiime metadata tabulate --m-input-file demux-joined-filter-stats.qza
     --o-visualization demux-joined-filter-stats.qzv
fi

        ## deblur
if [ ! -f rep-seqs.qza ] && [ ! -f table.qza ] && [ ! -f deblur-
    stats.qza ];
then
echo "qiime deblur denoise-16S --i-demultiplexed-seqs demux-joined-
    filtered.qza --p-min-size $MINSIZE --p-min-reads $MINREADS --p-trim-
    length $LENGTH --o-representative-sequences rep-seqs.qza --o-table
    table.qza --p-sample-stats --p-jobs-to-start 10 --o-stats deblur-
    stats.qza"
 qiime deblur denoise-16S --i-demultiplexed-seqs demux-joined-filtered.
     qza --p-min-size $MINSIZE --p-min-reads $MINREADS --p-trim-length
     $LENGTH --o-representative-sequences rep-seqs.qza --o-table table.
     qza --p-sample-stats --p-jobs-to-start $THREAD --o-stats deblur-
     stats.qza
fi

 ## visualize after deblur
if [ -f rep-seqs.qza ];
then
echo "visualize after deblur rep-seqs"
```

```
 qiime feature-table tabulate-seqs --i-data rep-seqs.qza --o-
    visualization rep-seqs.qzv
fi
if [ -f table.qza ];
then
echo "visualize after deblur table"
 qiime feature-table summarize --i-table table.qza --o-visualization
    table.qzv
fi
if [ -f deblur-stats.qza ];
then
echo "visualize after deblur stats"
 qiime deblur visualize-stats --i-deblur-stats deblur-stats.qza --o-
    visualization deblur-stats.qzv
fi

if [ ! -f aligned-rep-seqs.qza ];
then
        echo "mafft"
qiime alignment mafft --i-sequences rep-seqs.qza --o-alignment aligned-
   rep-seqs.qza --p-n-threads $THREAD
fi

if [ ! -f masked-aligned-rep-seqs.qza ];
then
        echo "mask"
qiime alignment mask --i-alignment aligned-rep-seqs.qza --o-masked-
   alignment masked-aligned-rep-seqs.qza
fi

 ## Phylogeny --> can be threaded, but not recommendable, cause tree
    could be different
if [ ! -f unrooted-tree.qza ];
then
        echo "unrooted-tree"
qiime phylogeny fasttree --i-alignment masked-aligned-rep-seqs.qza --o-
   tree unrooted-tree.qza --p-n-threads $THREAD
fi

if [ ! -f rooted-tree.qza ];
then
        echo "rooted-tree"
```

```
qiime phylogeny midpoint-root --i-tree unrooted-tree.qza --o-rooted-
    tree rooted-tree.qza
fi

# used the classifier from Selma
if [ ! -f taxonomy.qza ];
then
echo "silva taxonomy"
qiime feature-classifier classify-sklearn --i-classifier "/home/cam17/
    Documents/Taxonomy_classifier/classifier_132_V3_V4.qza" --p-n-jobs
    $THREAD --i-reads rep-seqs.qza --o-classification taxonomy.qza
fi



qiime tools export \
--input-path taxonomy.qza \
--output-path taxonomy-with-spaces
qiime metadata tabulate \
  --m-input-file taxonomy-with-spaces/taxonomy.tsv  \
  --o-visualization taxonomy-as-metadata.qzv
qiime tools export \
--input-path taxonomy-as-metadata.qzv \
--output-path taxonomy-as-metadata
qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-path taxonomy-as-metadata/metadata.tsv \
  --output-path taxonomy-without-spaces.qza



## remove mitochondria and chloroplast sequences (optional), based on
    taxonomy.qza
if [ ! -f table-no-MitChloro.qza ];
then
qiime taxa filter-table --i-table table.qza --i-taxonomy taxonomy-
    without-spaces.qza --p-exclude mitochondria,chloroplast --o-filtered
    -table table-no-MitChloro.qza
fi

# table.qza is changed to feature-table.biom, taxonomy.qza -> taxonomy.
    tsv, rooted-tree.qza -> tree.nwk
if [ -f table.qza ];
then
echo "export"
```

```
qiime tools export --input-path table-no-MitChloro.qza --output-path
    exported
fi

if [ -f taxonomy.qza ];
then
echo "export"
qiime tools export --input-path taxonomy.qza --output-path exported
fi

if [ -f rooted-tree.qza ];
then
echo "export"
qiime tools export --input-path rooted-tree.qza --output-path exported
fi

if [ -f exported ];
then
cd exported
fi

if [ -f feature-table.biom ] && [ -f taxonomy.tsv ] && [ -f tree.nwk ];
then
cp taxonomy.tsv biom-taxonomy.tsv
fi

echo "END TIME: " `date`
```

Curriculum vitae

Carlo Marzini
Born 26 January 1990 in Würzburg

Kommendestraße 12
38173 Lucklum

Working experience

Sine 05/2021	**Livestock and direct marketing manager** for Lucklumer Biogut GmbH & Co.KG
05/2017 – 12/2020	**PhD student** in molecular soil ecology studying the „Analysis of temporal and spatial variation of bacterial diversity in temperate soils“ at the Leibniz Institute German Collection of Microorganisms and cell cultures (DSMZ) GmbH in Braunschweig in the department of Microbial Ecology and Diversity.
11/2016 – 05/2017	Sampling and documentation for **R&H Umwelt** in Würzburg project work for subsoil investigation and remediation

Education and Qualifications

07/2016 – 08/2016	**Internship** at the **Scottish Rural College** in Edinburgh in the department of soil ecology
10/2013 – 01/2016	**Master of Science in Geoecology**, University of Bayreuth Specialisation in environmental chemistry Master thesis in the department of soil ecology Graduation grade **1.3**
10/2011 – 03/2014	**Bachelor of Science in Geoecology**, University of Bayreuth Bachelor thesis in soil science and -geography Graduation grade **1.9**
07/2009 – 03/2010	Basic military service
2000 – 2009	Deutschhaus-Gymnasium in Würzburg A-levels: Chemistry, History, German and social studies Graduation grade **1.8**

Curriculum vitae

05/2017 – 12/2020	Additional training courses in leadership and conversation techniques
04/2018	Conference VAAM in Wolfsburg
01/2018 – 12/2018	PhD representative at DSMZ
08/2012 – 09/2012	**R&H Umwelt** in Würzburg Duties: Sampling and documentation
Teaching	Supervision of practical courses in soil ecology Tutor in inorganic chemistry

Skills

Driver licence	**B**
Language skills	German – native speaker English – Proficient French – Basic
Computer skills	**Microsoft Office** - Proficient **ArcGIS** – Basic **R** – Proficient multivariate data analysis, linear and nonlinear modeling, machine learning and data science **Python** – **QIIME2** pipeline for sequencing data information processing
Lab methods	Fluorescence and light microscopy Fluorescence In Situ Hybridisation (FISH) Fluorescence activated cell sorting (FACS) Liquid scintillation counting (LSC) 13C/14C tracer studies Illumina 16S Amplicon sequencing for RNA and DNA MiSeq and NextSeq 550 Library preparation Mass spectrometry – basic ICP-OES – basic

Hobbies and interests

Agriculture
Sports: Calisthenics and running
