Bacterial community structure and function of *Anthelia* biological soil crust

Rúna Björk Smáradóttir

Faculty of Life and Environmental Sciences
University of Iceland
2020

Bakteríusamsetning og virkni í hélumosalífskurn

Rúna Björk Smáradóttir

60 eininga ritgerð sem er hluti af
Magister Scientiarum gráðu í líffræði

Leiðbeinandi
Dr. Ólafur S. Andrésson

Meistaranefnd
Dr. Ólafur S. Andrésson
Dr. Ingibjörg Svala Jónsdóttir

Prófdómari
Dr. Viggó Þór Marteinsson

Líf- og umhverfisvísindadeild
Verkfræði- og náttúruvísindasvið
Háskóli Íslands
Reykjavík, janúar 2020

Bacterial community structure and function of *Anthelia* biological soil crust

Rúna Björk Smáradóttir

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Biology

Advisor
Dr. Ólafur S. Andrésson

MS Committee
Dr. Ólafur S. Andrésson
Dr. Ingibjörg Svala Jónsdóttir

Master's Examiner
Dr. Viggó Þór Marteinsson

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, January 2020

Bacterial community structure and function of *Anthelia* biological soil crust
Bacteria in *Anthelia* biocrust
60 ECTS thesis submitted in partial fulfillment of a *Magister Scientiarum* degree in Biology

Faculty of Life and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugötu 7
101, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Rúna Björk Smáradóttir, 2020, *Bacterial community structure and function of* Anthelia *biological soil crust*, Master's thesis, Faculty of Life and Environmental Sciences, University of Iceland, pp. 69.

Printing: Háskólaprent
Reykjavik, Iceland, January 2020

Abstract

Biological soil crust (biocrust) dominated by the liverwort *Anthelia juratzkana* is widespread in the Icelandic highlands. In this study the bacterial community structure and function of the biocrust in various habitats within four areas in the highlands was analysed using high-throughput metagenomic sequencing. A clear difference was found between the biocrust and underlying soil strata both in taxonomic analysis and functional gene analysis.

The most abundant phyla of the biocrust were Acidobacteria, Actinobacteria and Proteobacteria. The phyla Acidobacteria, Armatimonadetes, Bacteroidetes, Chloroflexi and Cyanobacteria were enriched in the biocrust compared to subsurface soil. The most abundant genera were *Ktedonobacter* of the phylum Chloroflexi, *Bradyrhizobium* of the phylum Proteobacteria and the Acidobacterial genus Candidatus *Solibacter*.

No statistical difference in the bacterial composition was found between different habitat types, sample areas or seasons.

Genes encoding various functional pathways were enriched in the biocrust compared to lower soil strata including carbohydrate metabolism, which was among the most abundant functional systems, photosynthesis, motility and chemotaxis, and potassium and sulfur metabolisms.

Útdráttur

Lífskurn sem einkennist af soppmosanum hélumosa (*Anthelia juratzkana*) er útbreidd á hálendi Íslands. Í þessari rannsókn var bakteríusamsetning og virkni lífskurnarinnar í ýmsum vistgerðum á fjórum svæðum á hálendi Íslands skoðuð með notkun á háhraða DNA raðgreiningu á víðerfðamengi skurnarinnar. Greinilegur munur sást á milli lífskurnarinnar og undirliggjandi jarðvegslags bæði m.t.t. flokkunarfræði og starfrænna þátta.

Algengustu fylkingar lífskurnarinnar voru Acidobacteria, Actinobacteria og Proteobacteria. Fylkingarnar Acidobacteria, Armatimonadetes, Bacteroidetes, Chloroflexi og Cyanobacteria voru marktækt algengari í lífskurninni en í neðra jarðvegslagi. Algengustu ættkvíslir lífskurnarinnar voru *Ktedonobacter* sem tilheyrir fylkingunni Chloroflexi, *Bradyrhizobium* sem telst til Proteobacteria og Candidatus *Solibacter* sem tilheyrir fylkingunni Acidobacteria.

Í rannsókninni fannst enginn munur á bakteríusamsetningu milli ólíkra vistgerða, sýnatökustaða og árstíða.

Gen sem ákvarða ýmis lífvirk ferli voru algengari í lífskurninni en í undirliggjandi jarðvegslagi, meðal annars sykruefnaskipti sem voru meðal algengustu ferla, ljóstillífun, hreyfanleiki og efnasækni, og efnaskipti kalíums og brennisteins.

Table of Contents

List of Figures ..... **vi**

List of Tables ..... **viii**

Abbreviations ..... **ix**

Acknowledgements ..... **xi**

1 Introduction **1**
1.1 Characteristics and ecological functions of biological soil crusts 1
1.2 Liverwort dominated biocrust in Iceland 4
1.3 Soil bacteria 7
1.4 Bacterial communities of biocrusts 9
1.5 Objectives 11

2 Methods **13**
2.1 Sample collection 13
2.2 DNA isolation 15
2.3 DNA shotgun sequencing 15
2.4 Taxonomic analysis 15
2.5 Functional gene analysis 16
2.6 Statistical analysis 16

3 Results **17**
3.1 Taxonomic analysis 17
3.1.1 Overview of taxonomic analysis 17
3.1.2 Bacterial composition of the biocrust and comparison to lower soil strata 19
3.1.3 Comparison between seasons, habitat types and sampling areas 26
3.2 Functional gene analysis 31
3.2.1 Overview of functional gene analysis 31
3.2.2 Functional traits of the biocrust and comparison to lower soil strata 32

4 Discussion **39**
4.1 Taxonomic analysis 40
4.1.1 Methodology 40
4.1.2 Overview of taxonomic analysis 40
4.1.3 Most abundant taxa 41
4.1.4 Seasons, habitat types and sample areas 45
4.2 Functional traits 45

5 Conclusions **49**

References **51**

Appendix **57**

List of Figures

Figure 1.1: Sketches of typical biocrust types 2

Figure 1.2: Biocrust types 3

Figure 1.3: *Anthelia juratzkana* at 20x magnification 4

Figure 1.4: *Anthelia* biocrust in different habitat types 6

Figure 1.5: Distribution of EUNIS habitat types with high biocrust coverage 7

Figure 2.1: Sample areas 13

Figure 2.2: Sample of biocrust showing subsample locations 15

Figure 3.1: DNA recovery from topcrust and subsoil 18

Figure 3.2: PCA of taxonomic analysis on bacterial genus level 19

Figure 3.3: Percentage and relative abundance of dominant and other genera in topcrust 20

Figure 3.4: Relative abundance of total bacterial reads and dominant bacteria classified at phylum/class level 21

Figure 3.5: Relative abundance of total bacterial reads and dominant bacteria and comparison between topcrust and subsoil 22

Figure 3.6: Relative abundance of Acidobacteria, Alphaproteobacteria and Actinobacteria in topcrust classified to orders 23

Figure 3.7: Relative abundance of the most abundant bacterial genera 25

Figure 3.8: Relative abundance of most abundant Cyanobacterial genera and total Cyanobacterial reads in topcrust classified to orders 26

Figure 3.9: Relative abundance of the most abundant bacterial genera in spring and fall at Gagnheiði 27

Figure 3.10: PCA analysis on the genus level emphasizing different habitat types and sample areas 28

Figure 3.11: Relative abundance of the most abundant bacterial genera of different habitat types 29

Figure 3.12:Relative abundance of the most abundant bacterial genera of different sample areas 30

Figure 3.13: PCA of functional gene analysis on Level 3 SEED subsystems 32

Figure 3.14: Relative abundance of Level 1 SEED subsystems 34

Figure 3.15: Relative abundance of carbohydrates Level 2 SEED subsystems 35

Figure 3.16: Relative abundance of cofactors, vitamins, prosthetic groups and pigments Level 2 SEED subsystems 35

Figure 3.17: Relative abundance of motility and chemotaxis Level 2 SEED subsystems 36

Figure 3.18: Relative abundance of photosynthesis Level 2 SEED subsystems..... 37

Figure 3.19: Relative abundance of potassium metabolism Level 2 SEED subsystems..... 37

Figure 3.20: Relative abundance of sulfur metabolism Level 2 SEED subsystems 38

List of Tables

Table 1.1: EUNIS habitat types with high biocrust coverage 5

Table 2.1: Sampling sites 14

Table 3.1: Overview of DNA recovery and taxonomic analysis of topcrust 17

Table 3.2: Overview of DNA recovery and taxonomic analysis of subsoil 18

Table 3.3: Relative abundance of total bacterial reads classified at phylum/class level and comparison between topcrust and subsoil 20

Table 3.4: Relative abundance of dominant bacteria classified at phylum/class level and comparison between topcrust and subsoil 21

Table 3.5: Relative abundance of the most abundant bacterial genera of topcrust and comparison between topcrust and subsoil 24

Table 3.6: Overview of functional trait analysis of topcrust 31

Table 3.7: Overview of functional trait analysis of subsoil 31

Table 3.8: Relative abundance of Level 1 SEED subsystems and comparison between topcrust and subsoil 33

Abbreviations

ANOVA	analysis of variance
CO_2	carbon dioxide
DNA	deoxyribonucleic acid
EPS	extracellular polymeric substances
EUNIS	European nature information system
OTU	operational taxonomic unit
PC	principal component
PCA	principal component analysis
USA	United States of America

Acknowledgements

I want to thank my supervisor, Dr. Ólafur S. Andrésson, for giving me the opportunity to go back to school and for all the adventures and travels both in Iceland and overseas. Thank you for your guidance, support and patience and for always being ready to assist me on my educational journey.

I want to thank Dr. Ingibjörg Svala Jónsdóttir for the guidance and conversations and also for giving me the opportunity to travel to Svalbard.

My fellow students and everyone else in Askja I want to thank for a great time.

This project was supported by the University of Iceland Research Fund and The Icelandic Centre for Research (RANNÍS).

Finally I want to thank my family and friends for their support and patience and for always having faith in me.

1 Introduction

1.1 Characteristics and ecological functions of biological soil crusts

Biological soil crust (biocrust) is the community of lichens, bryophytes, non-lichenized fungi, algae, Cyanobacteria and other prokaryotes that live within or immediately on top of the uppermost millimeters of soil, typically in vascular plant interspaces (Weber *et al*., 2016). Soil particles are aggregated through the microbial production of extracellular polymeric substances (EPS) that glue soil particles together (Rossi *et al*., 2018) and the biocrust is further stabilized by lichen rhizines, bryophyte rhizoids and fungal hyphae. The majority of the biomass is within rather than on top of the soil (Weber *et al*., 2016).

Biocrusts provide various ecological functions and services which have value for human society (Weber *et al*., 2016; Rodríguez-Caballero *et al*., 2018). Biocrusts aggregate soil particles and provide protection against wind and water erosion, they improve ecosystem hydrology and reduce runoff. Biocrusts also influence biogeochemical cycles by fixing carbon and nitrogen and are the main primary producers in extreme and disturbed habitats where vascular plant cover is low and during times when vascular plants are dormant (Weber *et al*., 2016; Rodríguez-Caballero *et al*., 2018).

Biocrusts are the first colonizers on disturbed soil and an effective force in soil stabilization and recovery. They interact with vascular plants and in some cases enhance their propagation and growth through the establishment of seedlings and increased fertility of soil. Also, biocrusts are a unique habitat for a broad range of microfauna (Weber *et al*., 2016).

Biocrusts are often classified by their dominant photoautotrophic group; cyanobacterial crust, algal crust, bryophyte crust and lichen crust. In addition to this, biocrusts can be classified according to their morphology into smooth, rugose, rolling or pinnacled biocrust (Figure 1.1) (Weber *et al*., 2016).

The light Cyanobacterial biocrust (Figure 1.1A) is generally regarded as an early successional stage of biocrust development. The biocrust then shifts to thicker, dark Cyanobacterial crust with more complex assemblages (Figure 1.1B). Bryophyte dominated biocrusts (Figure 1.1D and F) are generally thought to be a late stage in biocrust succession (Weber *et al*., 2016). Bryophytes are very tolerant of environmental extremes, such as desiccation and variations in temperature. Bryophytes in biocrusts stabilize the soil, increase fertility and facilitate the microbial community growth (Weber *et al*., 2016).

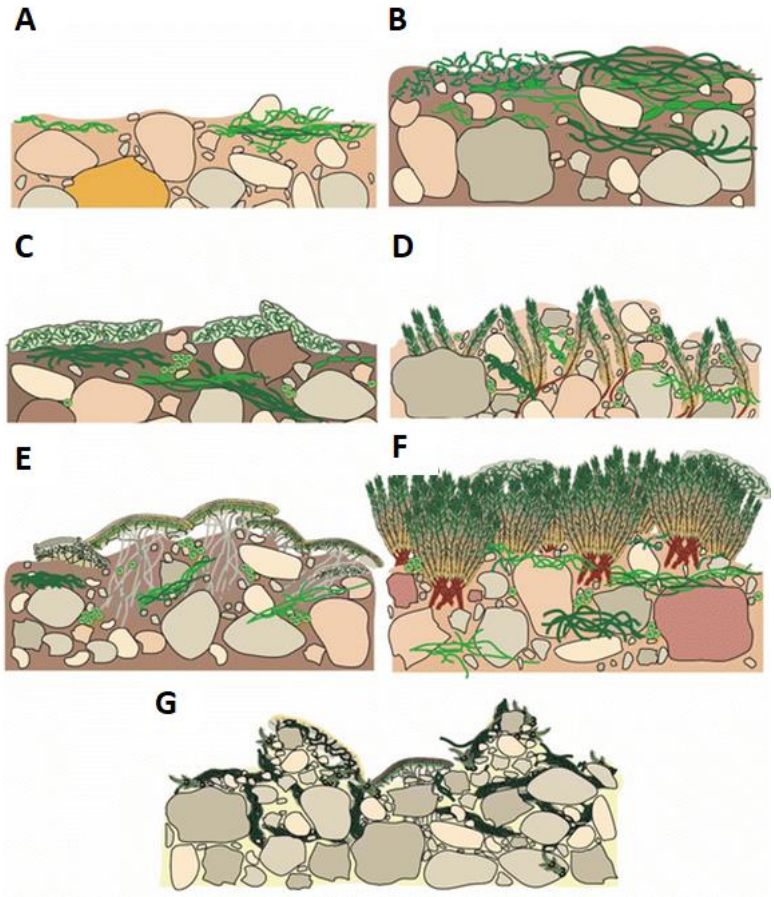

Figure 1.1: Sketches of typical biocrust types. A) Light cyanobacterial biocrust characterized by Microcoleus *sp. B) Dark cyanobacterial biocrust characterized by a diverse community of Cyanobacteria. C) Crustose cyanolichen biocrust characterized by cyanolichens and free living Cyanobacteria and algae. D) Rugose moss biocrust with moss stems embedded in the soil and only the uppermost leaves or fruiting bodies rising over the soil surface. E) Rolling chlorolichen biocrust characterized by crustose and squamulose chlorolichens along with free living cyanobacteria and green algae. F) Rolling "thick" moss biocrust with up to 5 mm thick moss carpets and free living cyanobacteria and green algae living on top or in between the stems. G) Pinnacled biocrust with irregular elevated structures where organisms prevent soil erosion. (Colesie* et al.*, in Weber* et al.*, 2016)*

Figure 1.2 shows some examples of different biocrust types. Dark pinnacled biocrust (Figure 1.2A) is common in the southwestern USA, e.g. the Mojave desert (Mogul *et al*., 2017) and the Colorado Plateau (Steven *et al*., 2015). This type of biocrust is generally dominated by Cyanobacteria but can have substantial lichen and moss cover. Lichen dominated biocrust (Figure 1.2B) is found in various locations, e.g. in Tabernas, Spain (Maier *et al*., 2014) and the Arabic peninsula (Abed *et al*., 2019). Light cyanobacterial biocrust (Figure 1.2C) has been reported from around the globe, e.g. in the Arabic peninsula (Abed *et al*., 2019) and the Kalahari desert in Africa (Elliott *et al*., 2014). Biocrusts dominated by thallous liverworts (Figure 1.2D), e.g. *Riccia* and *Asterella*, are common in Australia (Eldridge & Delgado-Baquerizo, 2019) and South Africa (Büdel *et al.*, 2009). In these types of biocrust the liverwort coverage can be very high and in some cases the biocrust can be composed entirely

of liverwort thalli (Weber *et al*., 2016). Liverworts are usually found in areas of high humidity or heavy rainfalls (Raven *et al*., 1992) and in southeast Australia, Eldridge & Delgado-Baquerizo (2019) found that increased precipitation resulted in increased cover of thallous liverworts.

Figure 1.2: Biocrust types. A) Pinnacled biocrust from Canyonlands National Park, Utah. Picture taken 28.09.2016 on a field trip at the Biocrust3 conference. B) Lichen dominated biocrust from Tabernas, Spain (Maier et al.*, 2014). C) Light Cyanobacterial biocrust from the Kalahari desert (Elliott* et al.*, 2014). D) Liverwort (*Riccia *sp.) dominated biocrust from Namibia (Weber* et al.*, 2016).*

Extracellular polymeric substances (EPS) are a collection of microbially produced substances that form a matrix in biofilms in which the cells are embedded. The main components of EPS are various polysaccharides, proteins, lipids and extracellular DNA (Flemming *et al*., 2016). The polysaccharides are often referred to as exopolysaccharides and the matrix is referred to as extracellular polymeric matrix (Rossi *et al*., 2018). EPS of biocrusts have the same properties as in other microbial associations, such as biofilms, i.e. retaining moisture, accumulating nutrients and protecting the microbial communities from harmful biological and physical agents as well as aggregating soil particles and gluing together the biotic and abiotic components (Rossi *et al*., 2018).

Many microorganisms of biocrusts produce EPS but Cyanobacteria and algae have been considered the main producers along with some microfungi and members of Proteobacteria and Actinobacteria (Flemming *et al*., 2016; Rossi *et al*., 2018). Recently EPS production has also been reported among members of Acidobacteria (Kielak *et al*., 2017).

1.2 Liverwort dominated biocrust in Iceland

In Iceland biocrust dominated by the liverwort *Anthelia juratzkana* is widespread in the highlands. *A. juratzkana* (Figure 1.3) is a monoicous leafy liverwort with very short (2-7 mm long) silver or blue-gray shoots (Bjarnason, 2018). The biocrust is typically found in areas with late melting snow or in snowbeds at elevations above 400 m but can also be found at lower elevations (Ottosson *et al*., 2016).

Figure 1.3: Anthelia juratzkana *at 20x magnification. Captured with a stereo microscope.*

The *Anthelia* biocrust has a high coverage in several EUNIS (European nature information system) habitat types in Iceland (Table 1.1) (Ottosson *et al*., 2016). A habitat type is an ecological unit with certain characteristics in terms of vegetation, wildlife, soil and climate. Areas of the same habitat type have similar assemblages of plants and animals (Ottosson *et al*., 2016). The pan-European EUNIS habitat type classification is a hierarchical system used

to facilitate the synchronized description and collection of data across Europe (https://www.eea.europa.eu/data-and-maps/data/eunis-habitat-classification#tab-based-on-data).

Table 1.1: EUNIS habitat types with high biocrust coverage (Ottosson et al.*, 2016).*

EUNIS habitat type	**Estimated total area**	**Crust cover average (%)**	**Soil C (%)**	**Soil pH**
E4.26 Icelandic *Racomitrium ericoides* heath	2,300 km^2	33.5	1.09	6.59
E4.115 Boreal moss snowbed communities	1,600 km^2	28.2	1.03	6.81
E4.241 Icelandic lava field lichen heaths	650 km^2	23.9	1.63	6.30
H5.2 Glacial moraines with very sparse or no vegetation.	2,700 km^2	12.0	1.4	6.6

The habitat types with the highest *Anthelia* biocrust cover are Icelandic *Racomitrium ericoides* heaths (EUNIS E4.26) (Figure 1.4A), Boreal moss snowbed communities (EUNIS E4.115) (Figure 1.4B) and Icelandic lava field lichen heaths (EUNIS E4.241) (Figure 1.4C) with 33.5%, 28.2% and 23.9% crust cover respectively. The soil carbon in these habitat types is low, ranging from 1.0-1.6% and the pH is acidoneutral, ranging from 6.3-6.8. In all these habitat types the most common moss is *Racomitrium ericoides*, the most common vascular plant is *Salix herbacea* and the most common lichens are various *Stereocaulon* species (Ottosson *et al*., 2016).

Figure 1.5 shows the distribution and estimated total area of the EUNIS habitat types with high biocrust cover highlighting how widespread the *Anthelia* biocrust is in Iceland. The estimated total coverage of *Anthelia* biocrust in these four habitat types is approximately 1,700 km^2.

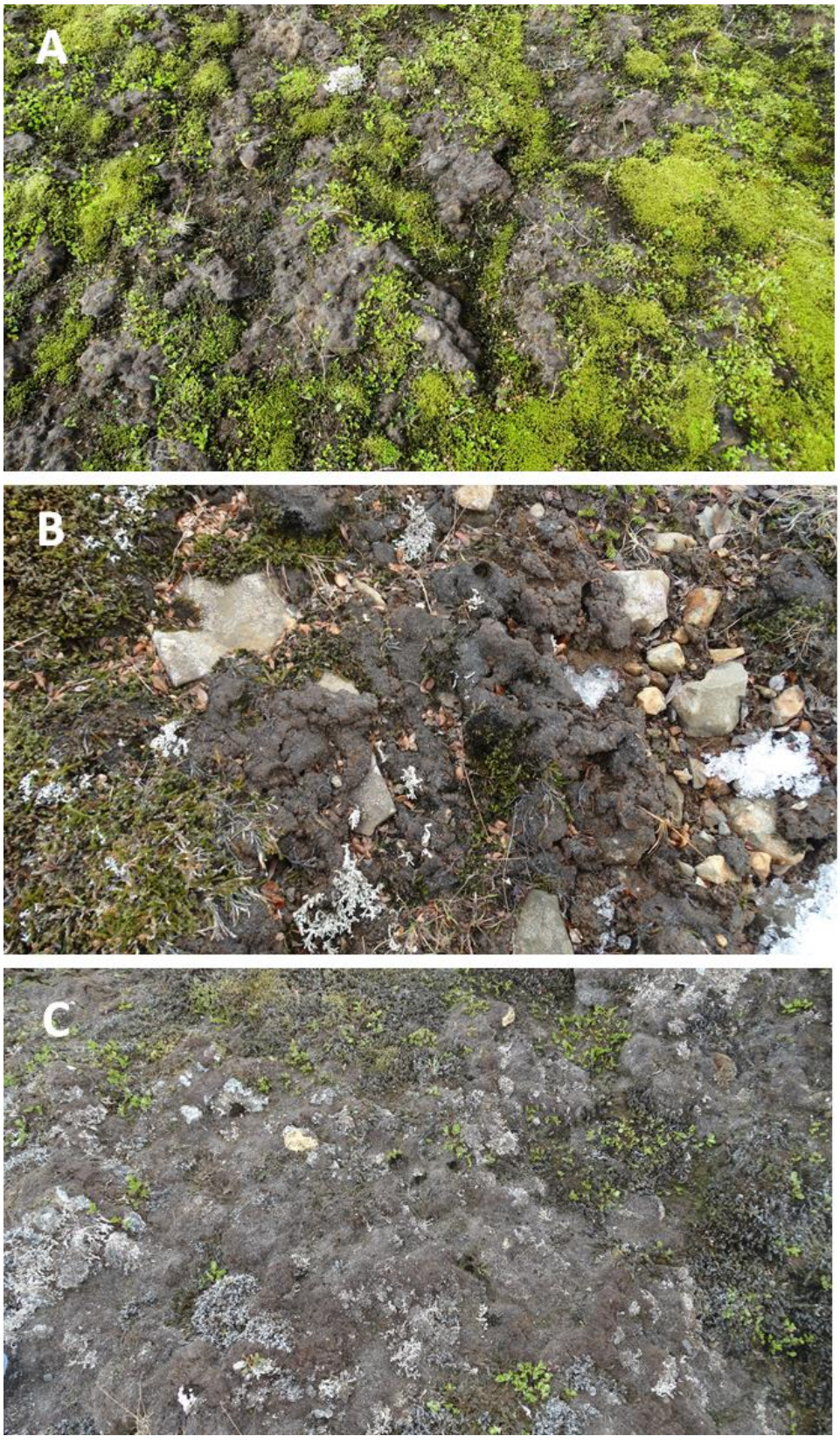

Figure 1.4: Anthelia *biocrust in A) Icelandic* Racomitrium ericoides *heath (EUNIS E4.26) from Skaftártunga, B) Boreal moss snowbed communities (EUNIS E4.115) from Gagnheiði and C) Icelandic lava field lichen heaths (EUNIS E4.241) from Laki. Pictures taken during field collection in summer and fall 2016.*

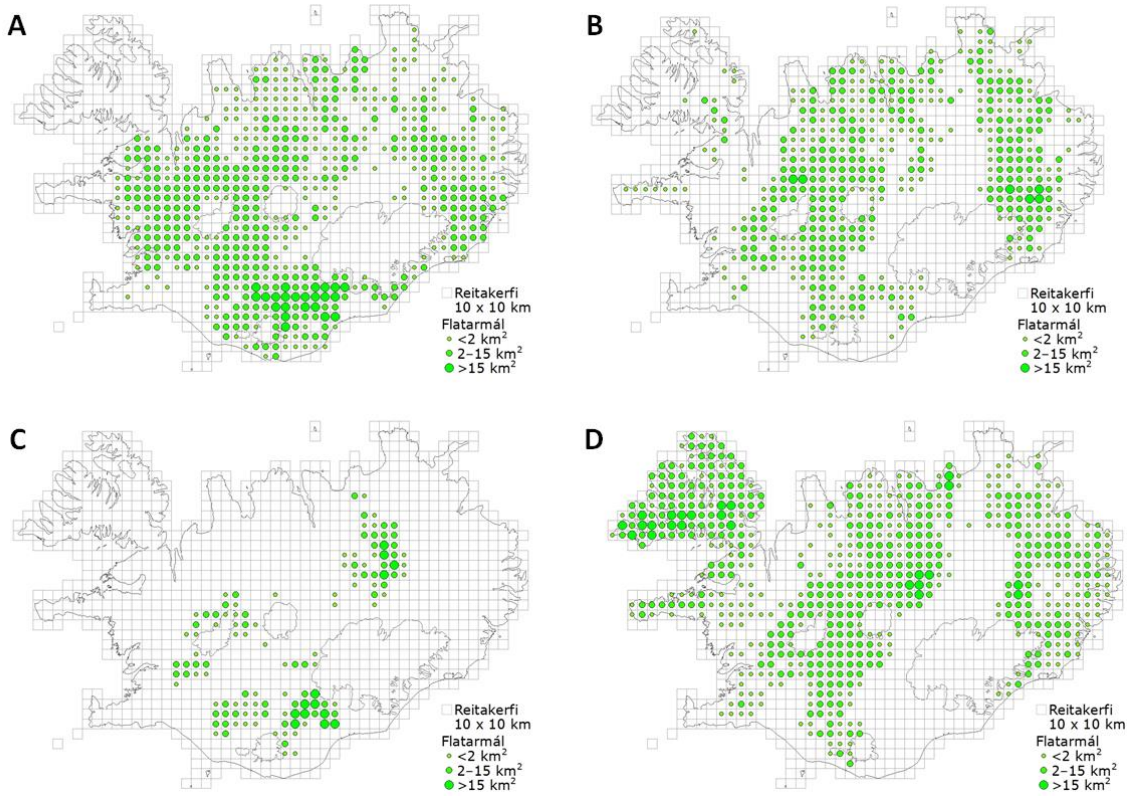

Figure 1.5: Distribution of EUNIS habitat types with high biocrust coverage. A) Icelandic Racomitrium ericoides *heath (EUNIS E4.26) B) Boreal moss snowbed communities (EUNIS E4.115) C) Icelandic lava field lichen heaths (EUNIS E4.241) D) Glacial moraines with very sparse or no vegetation (EUNIS H5.2) (Ottosson* et al.*, 2016).*

1.3 Soil bacteria

Soil bacteria are a very diverse and abundant group that contribute to many ecological processes, such as nitrogen, phosphorus and carbon cycling as well as interactions with plant and animal communities (Bardgett & Van Der Putten, 2014; Delgado-Baquerizo et al., 2018). Many soil bacteria are slow growing and difficult to cultivate in laboratory settings but in recent years new methods in molecular environmental microbiology have advanced our knowledge in this field (Delgado-Baquerizo et al., 2018; Fierer et al., 2012; Kielak et al., 2016; Malard et al., 2019).

Delgado-Baquerizo *et al*. (2018) analysed soil samples from across the globe and found that only a small fraction (2%) of phylotypes accounted for 41% of the soil bacteria, that is, only a few taxa dominate the soil bacteria communities. These phylotypes were classified as dominant based on being highly abundant and ubiquitous, i.e. found in more than half of the samples. Over 35% of these dominant phylotypes belong to the Proteobacteria, mainly Alphaproteobacteria (nearly 30%), about 30% belong to Actinobacteria, about 13% to Acidobacteria and about 7% to Planctomycetes with other phyla representing less than 5% (Delgado-Baquerizo *et al*., 2018). The major environmental factors driving habitat

preferences for the dominant phylotypes were grouped into five ecological clusters; high pH, low pH, drylands, low plant productivity and dry forest environments but only half of the phylotypes could be separated into these clusters whereas the other half was considered a core microbiome with no identifiable habitat preferences (Delgado-Baquerizo *et al*., 2018).

Fierer *et al*. (2012) analysed soil communities in various biomes (hot and cold deserts, tropical, temperate and boreal forests, prairie and tundra) and found that all biomes were dominated by the bacterial phyla Acidobacteria, Actinobacteria and Proteobacteria. Actinobacteria, Bacteroidetes and Cyanobacteria were on average more abundant in desert than non-desert biomes but vice versa was found for Acidobacteria and Verrucomicrobia (Fierer *et al*., 2012).

In the Arctic Malard *et al*. (2019) found that soil pH is a major factor driving bacterial diversity. A clear distinction was found in bacterial community structure between acidic (pH < 5), acidoneutral (pH 5-7) and alkaline (pH > 7) soil samples. Samples from the northern part of Iceland included in the study were in the acidoneutral group. In the study only a small fraction (0.3%) of all bacterial taxa were considered dominant based on having abundance over 0.1% in all samples. In acidoneutral soil the most abundant phyla of the total bacterial community were Acidobacteria, Proteobacteria and Verrucomicrobia but in the dominant group the relative abundance of Acidobacteria and Verrucomicrobia was reduced while the abundance of Planctomycetes, Chloroflexi and Bacteroidetes was increased. The relative abundance of Proteobacteria was about the same in the total and dominant groups (Malard *et al*., 2019). Unique and generalist taxa for each pH category were also identified. The most abundant taxa found exclusively in the acidoneutral group could be assigned to Actinobacteria, Verrucomicrobia and Acidobacteria, each with abundance of about 20%. Indicator species for each pH category were also identified based on abundant OTUs that were specifically associated with the different pH ranges. For the acidoneutral group six taxa were identified, three belonging to Acidobacteria (family Blastocatellaceae), two to Verrucomicrobia (order Chthoniobacterales) and one to Gemmatimonadetes (family Gemmatimonadaceae) (Malard *et al*., 2019).

The Proteobacteria is the largest and most diverse bacterial phylum (Madigan *et al*., 2006). The phylum is divided into several subdivisions or classes, Alpha-, Beta-, Gamma-, Delta- and Epsilonproteobacteria. Members of the Proteobacteria are gram negative and very diverse in both form and metabolic activity. Many Proteobacterial species are photoautotrophs that carry out anoxygenic photosynthesis and many are involved in the nitrogen cycle and can fix nitrogen, either free living or in symbiotic relationships with plants (Madigan *et al.*, 2006).

The Actinobacteria is a gram positive, spore forming phylum that forms branched hyphae (Barka *et al*., 2016). Most Actinobacterial species are saprophytic organisms that live in acidoneutral soil, both in the surface layers and below ground. They can utilize a wide spectrum of complex carbohydrates but survive as spores under low nutrient conditions. In soil ecosystems *Streptomyces* is usually the dominant Actinobacterial genus. Many

Actinobacteria produce a wide array of secondary metabolites, such as antibiotics, and some Actinobacteria are nitrogen fixing, e.g. the symbiotic genus *Frankia* (Barka *et al*., 2016).

The Acidobacteria are very abundant and widespread in soil ecosystems but they are difficult to cultivate and much of our knowledge about their ecology comes from environmental molecular microbiology studies (Kielak *et al*., 2016). Many members of the phylum are slow growing and require low concentrations of nutrients, non-conventional carbon sources and a raised CO_2 concentration for growth. The ability to utilize complex polysaccharides is common among the Acidobacteria and most of them are able to use many types of oligosaccharides (Kielak *et al*., 2016). Most Acidobacteria are tolerant of low pH and various pollutants in soil (Kielak *et al*., 2016). The ability to produce EPS in culture is found among many Acidobacteria (Kielak *et al*., 2016, 2017; Ward *et al*., 2009).

1.4 Bacterial communities of biocrusts

Biocrusts are complex ecosystems with high biodiversity and metabolic activities which are different from barren soil (Weber *et al*., 2016 ; Moreira-Grez *et al*., 2019). Their microbial composition also differs from underlying soil strata (Steven *et al*., 2013a; Elliott *et al*., 2014; Maier *et al*., 2014; Mogul *et al*., 2017; Moreira-Grez *et al*., 2019).

Many factors influence the microbial composition of biocrusts, e.g. their dominant photoautotrophic group (Moreira-Grez *et al*., 2019), soil properties (Steven *et al*., 2013a; Abed *et al*., 2019), morphology (Chilton *et al*., 2018), climate (Blay *et al*., 2017; Steven *et al*., 2015), crust cover (Mogul *et al*., 2017), physical disturbance (Steven *et al*., 2015) and vegetation zone (Elliott *et al*., 2014).

Successional stages can also influence the bacterial composition of biocrusts. In Cyanobacteria dominated biocrust in Oman a decrease was observed in Actinobacterial proportions and an increase in Cyanobacterial proportions with crust development (Abed *et al*., 2019). Chilton *et al*. (2018) found that late successional stages of a biocrust in southeastern Australia were dominated by Alphaproteobacteria but Cyanobacteria dominated early and mid stages. In late stages the biocrust was dominated by lichen and mosses which was not the case in earlier stages (Chilton *et al*., 2018). In an early stage biocrust on a grazed grassland in the Kalahari desert, Proteobacteria and Actinobacteria were the most abundant phyla, together representing 63% of the reads (Elliott *et al*., 2014).

Historically Cyanobacteria have been regarded as the most important bacterial phylum of biocrusts and it is often the most abundant (Weber *et al*., 2016). Cyanobacteria are generally considered pioneers in biocrust formation, especially the "ecosystem engineer" *Microcoleus vaginatus* (Couradeau *et al*., 2019). Cyanobacteria are known to produce EPS which bind soil and provide a stable and hydrated environment enhancing biocrust formation (Rossi & De Philippis, 2015). Furthermore, the filamentous Cyanobacteria, *Microcoleus*, *Schizothrix* and *Hydrocoleum*, are encased in exopolysaccharidic sheaths which are considered the first

accumulation of EPS contributing to the biocrust structure (Rajeev *et al*., 2013). These sheaths then remain in the soil, binding the particles together, after filament migration or death (Rossi *et al*., 2018).

Many Cyanobacteria are nitrogen fixers, e.g. the heterocystous Nostocales which includes the genera *Nostoc*, *Scytonema* and *Tolypothrix* that are found in biocrusts worldwide (Weber *et al*., 2016). Even though the pioneer *M. vaginatus* does not fix nitrogen, it has a so called cyanosphere associated with the filamentous sheaths, a Cyanobacterial community that is enriched in nitrogen fixing genera (Couradeau *et al*., 2019). It has been noted in later successional stages of biocrust that *M. vaginatus* is replaced by other Cyanobacteria, e.g. *Nostoc* and *Scytonema* (Belnap, 2002).

In a Cyanobacteria dominated biocrust in the Arabic peninsula Cyanobacteria accounted for 29% of bacterial taxa and 20% in lichen dominated biocrust. Other abundant taxa were Actinobacteria, Alphaproteobacteria and Bacteroidetes (Abed *et al*., 2019).

Cyanobacteria was the most dominant phylum in Cyanobacteria dominated biocrust from western Australia but in lichen dominated crusts Acidobacteria and Proteobacteria (mainly Alphaproteobacteria) were most abundant (Moreira-Grez *et al*., 2019).

In a pinnacled biocrust from the Mojave desert, USA, Cyanobacteria was the most dominant phylum representing 33% of the bacterial community with Proteobacteria and Chloroflexi accounting for 26% and 12% of the reads respectively (Mogul *et al*., 2017). The study also found that Chloroflexi and Cyanobacteria abundance and diversity increased with increasing biocrust cover. In an earlier study in the Mojave desert Cyanobacteria dominated the biocrust representing 42% of the bacterial community while Proteobacteria accounted for over 20% and Actinobacteria about 15% (Steven *et al*., 2014).

In a biocrust from the Colorado Plateau, USA, Cyanobacteria, Proteobacteria and Actinobacteria were the most abundant phyla but varied between different soil material (Steven *et al*., 2013a).

In lichen dominated biocrust from Spain the most abundant phyla were Actinobacteria and Proteobacteria (mainly Alphaproteobacteria) each representing over 25% of bacterial reads. Bacteroidetes represented about 12%, Cyanobacteria about 7% and other phyla less than 5% of the reads (Maier *et al*., 2014).

In rolling biocrust from Idaho, USA, Actinobacteria were the dominant phylum representing 36-51% of the bacterial community and the study also found that the abundance of Actinobacteria and Firmicutes increased with higher elevation while the abundance of Cyanobacteria, Proteobacteria and Chloroflexi decreased. Cyanobacterial abundance was highest in lower elevations (about 11%) but was below 1% in the colder, wetter climate of higher elevations (Blay *et al*., 2017).

Another example of low Cyanobacterial abundance in biocrust comes from the high Arctic in Svalbard where Proteobacteria were the dominant phylum accounting for 50-70% of the bacterial community while the remaining community was represented by Bacteroidetes, Acidobacteria and Actinobacteria. Other phyla were found in low abundance, including Cyanobacteria which represented less than 1% of the community. In spite of this the crust had an abundance of chlorophyll a, a signature of phototrophic microorganisms. The study also found that Proteobacteria were the main producers of the EPS of the crust (Mugnai *et al.*, 2015). In another study from the high Arctic (Ellismere Island, Canada), Cyanobacteria represented 10-35% of the bacterial community, Acidobacteria accounted for 10-25%, Planctomycetes 5-20% and Proteobacteria and Verrucomicrobia about 15% each (Steven *et al.*, 2013b).

In a long term warming and wetting experiment Steven *et al*. (2015) found that 2°C soil warming had little effect on the bacterial community while wetting altered the bacterial composition and increased Cyanobacterial biomass. Warming and wetting combined however had a dramatic effect on bacterial composition and decreased Cyanobacterial abundance from about 60% in the control to 4% with an increase in Chloroflexi, Actinobacteria and Acidobacteria (Steven *et al.*, 2015).

1.5 Objectives

The objectives of this study are:

I. To analyse the bacterial composition of the *Anthelia* biological soil crust (biocrust).
II. To compare the bacterial composition between the biocrust (top 5 mm) and lower soil strata (subsoil).
III. To compare the bacterial composition of the biocrust between different seasons, habitat types and sample areas.
IV. To analyse the main functional traits of the microbial community of the biocrust via functional gene analysis.

2 Methods

2.1 Sample collection

Samples of biocrust were collected in four sample areas in Iceland (Figure 2.1 and Table 2.1) chosen in consultation with the Iceland Institute of Natural History as EUNIS habitat types with high biocrust coverage (Table 1.1). The soil is andosol (volcanic mineral soil) with low carbon content (1-2%) and the pH ranges from about 6.3 to 6.8 (Ottosson *et al.*, 2016). At Gagnheiði samples were collected both in early spring and late fall for comparison between seasons. Sample containers were thrown randomly for sampling. Three 15x15 cm samples (Figure 2.2) were collected at each sample area (total of 12 samples). Samples were then stored at -20°C until further analysis. For balanced sampling, samples collected in spring (24.05.2016) at Gagnheiði were excluded from general taxonomic and functional analyses and only used to compare bacterial composition between seasons.

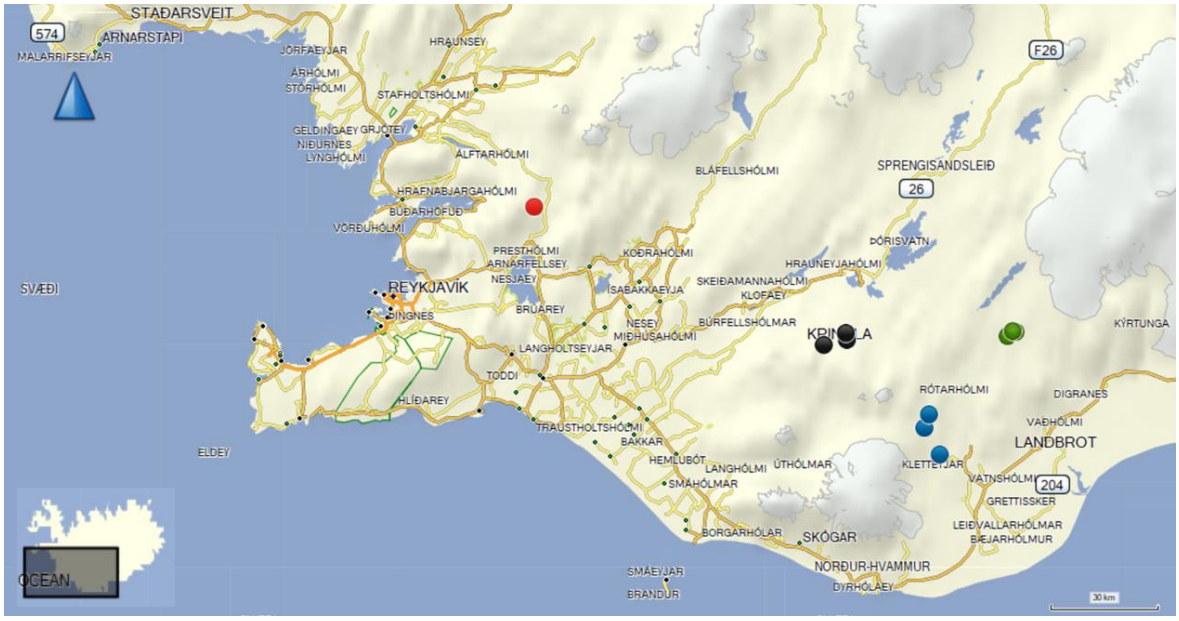

Figure 2.1: Sample areas. Red: Gagnheiði close to Þingvellir. Black: Fjallabak. Blue: Skaftártunga. Green: Laki.

Table 2.1: Sampling sites.

Sampling area	Sample	GPS location (elevation)	Habitat type [a]	Time of sampling	MG-RAST ID (subsoil)
Gagnheiði	G1	N64° 22.053' W21° 03.768' (520 m)	Boreal moss snowbed communities (EUNIS E4.115)	24.05.2016 and 14.09.2016 [b]	4746930.3 (4746914.3)
	G2	N64° 22.044' W21° 03.703' (511 m)	Boreal moss snowbed communities (EUNIS E4.115)	24.05.2016 and 14.09.2016 [b]	4746927.3 (4746913.3)
	G3	N64° 22.081' W21° 03.623' (493 m)	Boreal moss snowbed communities (EUNIS E4.115)	24.05.2016 and 14.09.2016 [b]	4746925.3
Fjallabak	F1	N64°01.458' W19°21.357' (773 m)	Glacial moraines with very sparse or no vegetation (EUNIS H5.2)	24.08.2016	4746911.3 (4746915.3)
	F2	N64°02.220' W19°13.191' (597 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	24.08.2016	4746924.3
	F3	N64°03.230' W19°13.545' (598 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	24.08.2016	4746929.3
Skaftár-tunga	S1	N63° 45.107' W18° 40.340' (380 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	10.07.2016	4746917.3
	S2	N63° 49.078' W18° 45.765' (500 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	10.07.2016	4746923.3
	S3	N63° 51.095' W18° 44.137' (560 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	10.07.2016	4746916.3
Laki	L1	N64° 02.780' W18° 16.358' (560 m)	Icelandic lava field lichen heaths (EUNIS E4.241)	09.07.2016	4746919.3 (4746920.3)
	L2	N64° 03.399' W18° 13.504' (573 m)	Icelandic lava field lichen heaths (EUNIS E4.241)	09.07.2016	4746926.3
	L3	N64° 03.511' W18° 14.532' (587 m)	Icelandic *Racomitrium ericoides* heath (EUNIS E4.26)	09.07.2016	4746912.3

(a) Ottosson *et al*., 2016. (b) Samples collected in spring (24.05.2016) and fall (14.09.2016) at Gagnheiði sampling area. Spring samples were excluded from general taxonomic and functional analyses and only used to compare bacterial composition between seasons.

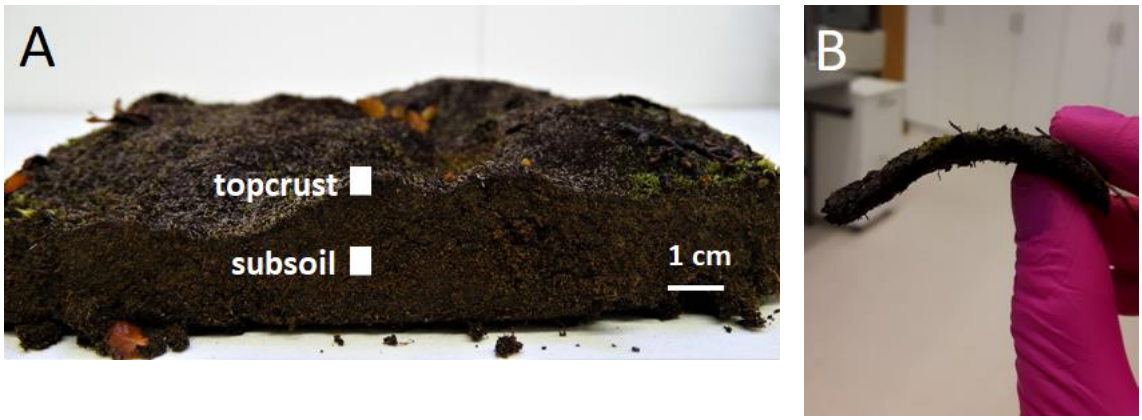

Figure 2.2: A) Sample collected in the field showing subsample locations in the top 5 mm (topcrust) and at 15 mm depth (subsoil). B) Sample of topcrust.

2.2 DNA isolation

Random 5 g subsamples from the top 5 mm (topcrust) were cut out and dried at room temperature and then manually homogenized using mortar and pestle for DNA isolation. For four samples (G1, G2, F1 and L1) random 5 g subsamples were also cut out at 15 mm depth (subsoil) (Figure 2.2). For deciding the vertical location of the subsamples the topcrust (top 5 mm) easily peeled off and subsoil samples were collected low enough to be adequately far away from the biocrust but not low enough for anaerobic conditions.

DNA isolation was performed using the DNeasy PowerLyzer® PowerSoil® DNA Isolation Kit (QiaGen) according to the manufacturer's protocol using 0.2 g of dried soil sample.

2.3 DNA shotgun sequencing

DNA sequencing libraries were prepared using the Illumina Nextera® XT DNA kit according to the manufacturer's protocol. Samples were multiplexed using the Nextera XT Index Kit v2 and six samples run each time on the Illumina MiSeq v.2 platform using a 2x150 base sequencing kit. Samples concentrations were adjusted and rerun until a minimum of ~3 million reads (900 Mb of sequence) was reached (Table 3.1).

2.4 Taxonomic analysis

For taxonomic analysis unassembled reads were uploaded to the Kaiju web server (Menzel *et al*., 2016) using the default „greedy" mode with minimum match length 11, minimum match score 75 and 5 allowed mismatches. The NCBI BLAST non redundant protein database of bacteria, archaea, viruses, fungi and microbial eukaryotes was chosen as a

reference database (http://kaiju.binf.ku.dk/). The resulting classification data was downloaded and summaries of bacterial community produced.

2.5 Functional gene analysis

For functional gene analysis unassembled reads were uploaded to the MG-RAST web server (Meyer *et al*., 2008) (http://www.mg-rast.org/) and analysed using SEED subsystems (http://pubseed.theseed.org). In subsystems classification, genes are grouped into collections of functional roles, which together implement a specific biological process or structural complex. Proteins within a protein family (e.g. share a common domain structure) may exhibit the same or multiple functional roles (Overbeek *et al*., 2005). The resulting classification data was downloaded and summaries of functional genes produced.

All sequence data sets are publicly available in MG-RAST (http://www.mg-rast.org/) under the project name "Anthelia_Biocrust", see MG-RAST identification numbers in Table 2.1.

2.6 Statistical analysis

Statistical analysis was performed using R-3.4.3 and RStudio 1.1.423 (R-Core-Team, 2017; RStudio, 2018). Differences between topcrust and subsoil were evaluated using Welch two sample t-test both for taxonomic and functional gene analyses. Differences between habitat types and sample areas were evaluated using analysis of variance (ANOVA) and Tukey's honest significance test was conducted to confirm where the differences occurred between groups where statistical difference was observed. To evaluate the difference between seasons paired t-test was used for each site on the genus level. Additionally Welch two sample t-test was used for evaluation of differences between seasons for the most abundant genera. Principal component analysis (PCA) was conducted to analyze patterns of clustering.

Tables with detailed information of t-test and ANOVA output and PCA taxa and gene loading scores can be found in Appendix.

3 Results

3.1 Taxonomic analysis

3.1.1 Overview of taxonomic analysis

Table 3.1 shows DNA recovery of samples and overview of the taxonomic analysis of bacteria, fungi and archaea in topcrust and Table 3.2 shows the same information for subsoil. DNA recovery from topcrust was very variable between samples but on average higher than from subsoil (*p*-value 0.024) (Figure 3.1).

Table 3.1: Overview of DNA recovery and taxonomic analysis of topcrust.

Sample	DNA (ng/µl)	read count	identified reads (%)[a]	Fungi (%)	Archaea (%)	Bacteria (%)	genus (%)[b]
G1_september	24.0	7,782,856	44.5	10.1	0.4	89.0	68.5
G2_september	25.5	7,590,373	44.6	6.2	0.6	92.7	68.7
G3_september	38.6	6,169,792	44.4	7.9	0.5	91.3	68.2
G1_may	50.0	8,056,774	45.2	6.7	0.5	92.5	68.6
G2_may	26.0	8,921,188	49.1	7.4	0.5	91.8	68.0
G3_may	47.0	8,545,534	49.8	5.8	0.5	93.4	66.5
F1	36.2	6,541,857	45.9	8.8	0.4	90.1	71.0
F2	30.7	5,777,778	45.1	6.6	0.3	92.4	72.5
F3	37.0	7,995,364	45.1	8.1	0.3	91.1	70.0
S1	34.3	3,655,245	45.9	8.6	0.4	90.2	69.3
S2	10.0	4,296,852	39.7	5.5	0.4	93.6	71.4
S3	27.0	3,859,272	45.2	11.8	0.4	87.3	69.1
L1	34.8	2,785,906	48.1	4.8	0.5	94.0	67.5
L2	23.9	3,983,033	36.5	13.3	0.4	85.5	68.5
L3	17.9	4,242,066	46.9	2.3	0.5	97.0	67.7
Average	**30.9**	**6,013,593**	**45.1**	**7.6**	**0.4**	**91.5**	**69.0**
St. dev	**10.2**	**1,996,196**	**3.2**	**2.7**	**0.1**	**2.7**	**1.5**

(a) Proportion of reads that could be identified. (b) Proportion of total bacterial reads that could be identified to the genus level.

Table 3.2: Overview of DNA recovery and taxonomic analysis of subsoil.

Sample	DNA (ng/µl)	read count	identified reads (%)[a]	Fungi (%)	Archaea (%)	Bacteria (%)	identified to genus (%)[b]
G1_subsoil [c]	22.8	5,438,637	47.5	0.3	0.5	98.8	56.4
G2_subsoil [c]	20.2	6,739,137	46.0	0.8	0.6	98.4	60.3
L1_subsoil	20.1	8,705,862	46.5	1.3	0.5	97.8	62.8
F1_subsoil	10.8	7,884,510	43.9	0.6	0.6	98.6	60.3
Average	**18.5**	**7,192,037**	**46.0**	**0.8**	**0.5**	**98.4**	**59.9**
St.dev	**4.6**	**1,229,910**	**1.3**	**0.4**	**0.0**	**0.4**	**2.3**

(a) Proportion of reads that could be identified. (b) Proportion of total bacterial reads that could be identified to the genus level. (c) Subsoil samples from Gagnheiði collected in fall.

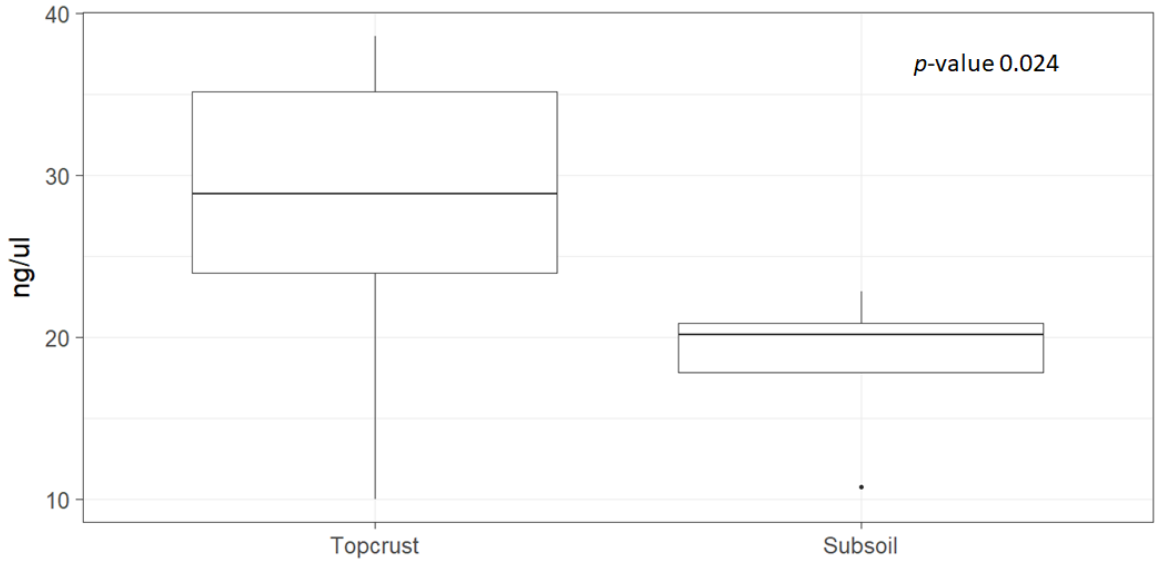

Figure 3.1: DNA recovery from topcrust and subsoil. N *topcrust=12;* N *subsoil=4.*

Total read count in topcrust samples was on average about 6 million reads. Identified reads were on average 45.1% of the total reads. Bacteria accounted for 91.5% of the identified reads on average. About 7.6% of the reads could be assigned to fungi and 0.4% to Archaea.

In subsoil total read count was on average about 7.2 million reads. Identified reads were on average 46% of the total reads. 98.4% of the identified reads could be assigned to Bacteria, 0.8% to Fungi and 0.5% to Archaea.

The most abundant fungal phylum was Ascomycota, representing over 90% of the fungal reads in topcrust and about 70% in subsoil. The most abundant Archaea phylum was Euryarchaeota accounting for about 70% of the Archaea reads in topcrust and about 65% in subsoil.

3.1.2 Bacterial composition of the biocrust and comparison to lower soil strata

In topcrust 69% and in subsoil 59.9% of identified bacterial reads could be assigned to the genus level (Table 3.1 and Table 3.2). In total over 1,900 bacterial genera were found in each sample. Figure 3.2 shows PCA analysis on the genus level. The analysis shows a separation between topcrust and subsoil samples.

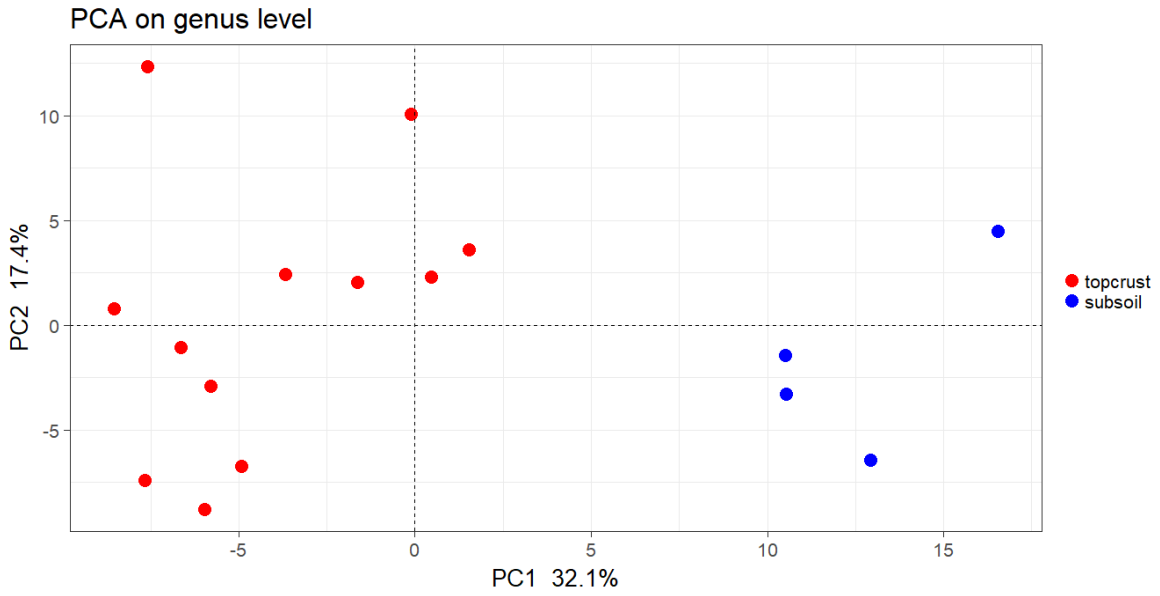

Figure 3.2: PCA of taxonomic analysis on bacterial genus level. N *topcrust=12;* N *subsoil=4. Genera whith the highest loading scores for PC1 and PC2 are shown in Appendix (tables 1 and 2).*

In topcrust 86 genera have abundance over 0.1% of bacterial reads across all topcrust samples, representing 43.2% of total bacterial reads (Figure 3.3). In subsoil 89 genera have abundance over 0.1% of bacterial reads across all subsoil samples, representing 35.7% of total bacterial reads. These genera account for about 4.5% of all bacterial genera found in the community and will henceforth be referred to as dominant.

Figure 3.4 shows the relative abundance of bacterial phyla and Proteobacterial classes both for total bacterial reads (Table 3.3) and the dominant bacteria (Table 3.4). The most abundant groups are Acidobacteria, Actinobacteria and Alphaproteobacteria. Unclassified bacterial reads are about 2.5% of total reads in topcrust and 5.3% in subsoil.

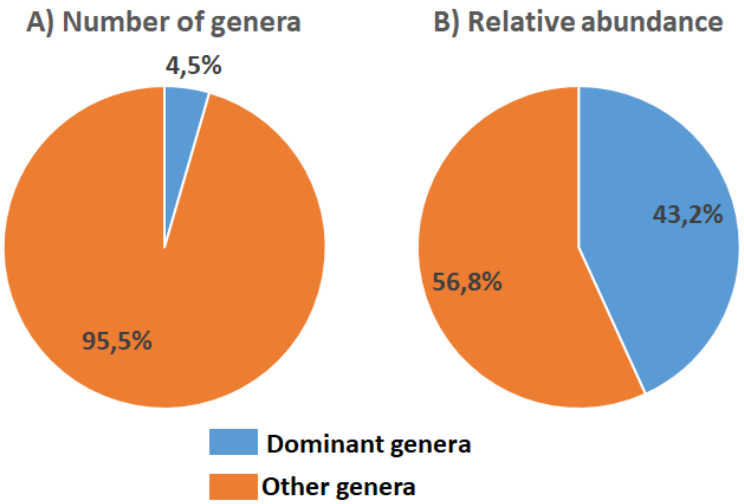

*Figure 3.3: A) Percentage and B) Relative abundance of dominant and other genera in topcrust. Dominant = genera with over 0.1% of total bacterial reads in all topcrust samples (*N*=12).*

*Table 3.3: Relative abundance (mean %) of total bacterial reads classified at phylum/class level and comparison between topcrust and subsoil (*p*-values).* N *topcrust=12;* N *subsoil=4.*

Phylum/class	**Topcrust**	**St.dev**	**Subsoil**	**St. dev**	***p*-values [(a)]**
Acidobacteria	16.2	2.6	14.9	1.6	0.31
Actinobacteria	18.4	3.1	21.4	3.0	0.19
Armatimonadetes	0.7	0.1	0.4	0.0	<0.001*** [(b)]
Bacteroidetes	4.1	0.9	3.4	0.6	0.15
Chloroflexi	7.5	1.8	5.5	0.2	0.0032** [(b)]
Cyanobacteria	4.1	1.0	1.6	0.1	<0.001*** [(b)]
Deinococcus-Thermus	0.3	0.0	0.3	0.0	0.08
Firmicutes	3.6	0.5	3.2	0.1	0.03* [(b)]
Gemmatimonadetes	0.5	0.1	1.1	0.2	0.0063** [(c)]
Nitrospirae	0.3	0.0	0.6	0.2	0.096
Planctomycetes	4.9	1.2	5.5	0.4	0.17
Alphaproteobacteria	18.9	1.8	17.1	1.6	0.15
Betaproteobacteria	5.1	0.3	6.3	0.4	0.01* [(c)]
Gammaproteobacteria	4.3	0.4	4.1	0.3	0.3
Deltaproteobacteria	2.9	0.4	3.9	0.4	0.0098** [(c)]
Verrucomicrobia	2.1	0.5	2.2	0.1	0.58
Other	3.6	0.4	3.3	0.1	<0.001*** [(b)]
Unclassified Bacteria	2.5	0.2	5.3	0.5	0.008** [(c)]

(a) Statistical difference: $p < 0{,}001$*** ; $p < 0{,}01$** ; $p < 0{,}05$* obtained by Welch two sample t-test. Detailed output is shown in Appendix (table 6). (b) Enriched in topcrust. (c) Enriched in subsoil.

*Table 3.4: Relative abundance (mean %) of dominant bacteria classified at phylum/class level and comparison between topcrust and subsoil (*p*-values). Dominant = Genera with over 0.1% of total bacterial reads in all topcrust or all subsoil samples.* N *topcrust=12;* N *subsoil = 4.*

Phylum/class	**Topcrust**	**St. dev**	**Subsoil**	**St. dev**	***p*-values [(a)]**
Acidobacteria	19.2	3.7	13.3	2.5	0.014* [(b)]
Actinobacteria	24.5	4.0	30.6	4.1	0.07
Armatimonadetes	0.6	0.2	0.0	0.0	<0.001*** [(b)]
Bacteroidetes	2.2	0.6	1.3	0.2	0.0013** [(b)]
Chloroflexi	11.1	3.1	4.9	1.2	<0.001*** [(b)]
Cyanobacteria	2.8	0.6	0.3	0.0	<0.001*** [(b)]
Deinococcus-Thermus	0.4	0.1	0.4	0.0	0.7
Firmicutes	2.5	0.4	2.5	0.2	0.99
Gemmatimonadetes	0.0	0.0	1.4	0.3	0.0025** [(c)]
Nitrospirae	0.0	0.0	0.6	0.3	0.04* [(c)]
Planctomycetes	6.5	1.7	9.8	1.1	0.004** [(c)]
Alphaproteobacteria	23.2	2.9	24.0	2.5	0.67
Betaproteobacteria	3.2	0.3	4.0	0.7	0.15
Gammaproteobacteria	1.4	0.7	1.3	0.1	0.77
Deltaproteobacteria	0.8	0.1	2.8	0.2	<0.001*** [(c)]
Verrucomicrobia	1.5	0.7	2.2	0.2	0.008** [(c)]
Other	0.0	0.0	0.5	0.1	0.008** [(c)]

(a) Statistical difference: $p < 0{,}001$*** ; $p < 0{,}01$** ; $p < 0{,}05$* obtained by Welch two sample t-test. Detailed output is shown in Appendix (table 7). (b) Enriched in topcrust. (c) Enriched in subsoil.

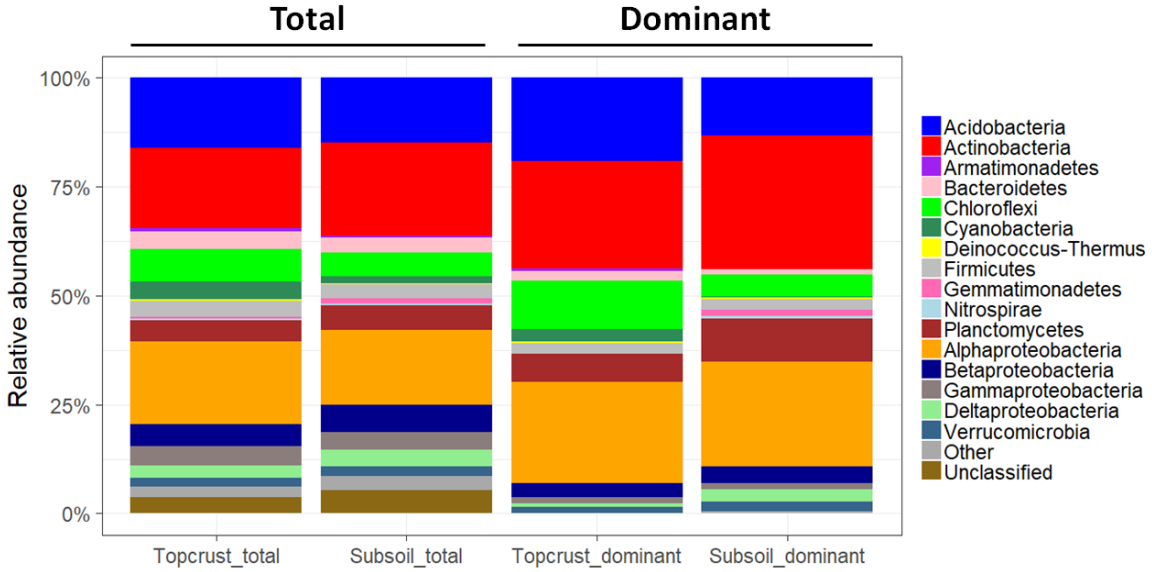

Figure 3.4: Relative abundance of total bacterial reads and dominant bacteria classified at phylum/class level. Dominant = Genera with over 0.1% of total bacterial reads in all topcrust or all subsoil samples. N *topcrust=12 ;* N *subsoil=4.*

Comparison between topcrust and subsoil on the phylum/class level can be seen in Figure 3.5, Table 3.3 and Table 3.4. In the total community Armatimonadetes, Chloroflexi, Cyanobacteria and Firmicutes are significantly more abundant in topcrust than in subsoil but Gemmatimonadetes, Betaproteobacteria and Deltaproteobacteria are enriched in subsoil. No statistical difference between topcrust and subsoil is found for other groups. Among the dominant bacteria Acidobacteria, Armatimonadetes, Bacteroidetes, Chloroflexi and Cyanobacteria are enriched in topcrust but Gemmatimonadetes, Nitrospirae, Planctomycetes, Deltaproteobacteria and Verrucomicrobia are more abundant in subsoil.

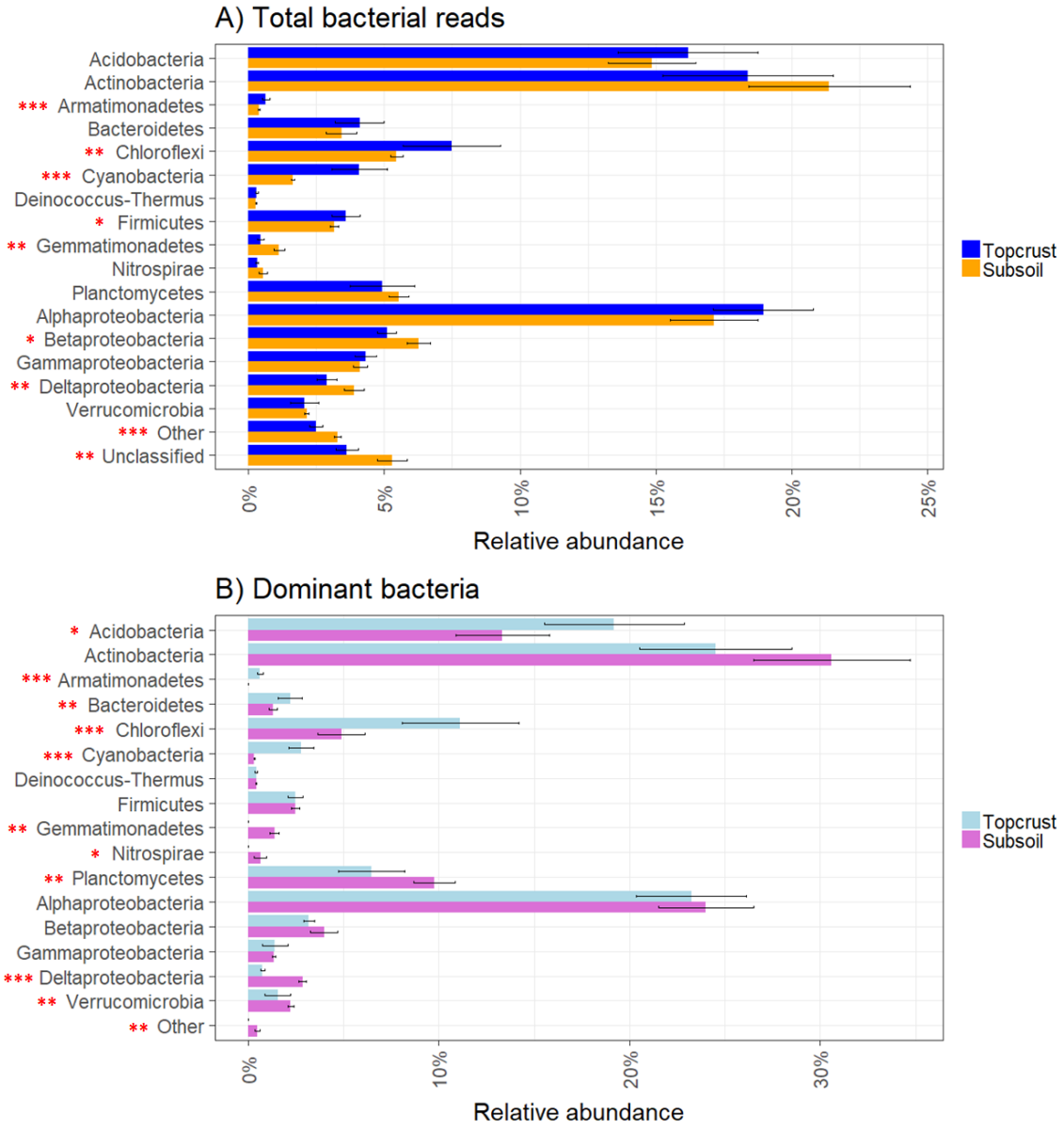

Figure 3.5: Relative abundance (mean) of A) Total bacterial reads and B) Dominant bacteria classified on the phylum/class level and comparison between topcrust and subsoil. Error bars show standard deviation. Dominant = Genera with over 0.1% of total bacterial reads in all topcrust or all subsoil samples. Statistical difference: $p < 0.001$*** ; $p < 0.01$**; $p < 0.05$*. N *topcrust=12 ;* N *subsoil=4.*

The phylum Actinobacteria is very abundant in topcrust representing 18.4% of the total bacterial community and 24.5% of dominant bacteria. In the dominant group of topcrust the 22 Actinobacterial genera are divided between 11 orders. Members of the orders Corynebacteriales and Streptomycetales are most abundant representing 3% and 2% of both total and dominant bacterial reads respectively (Figure 3.6).

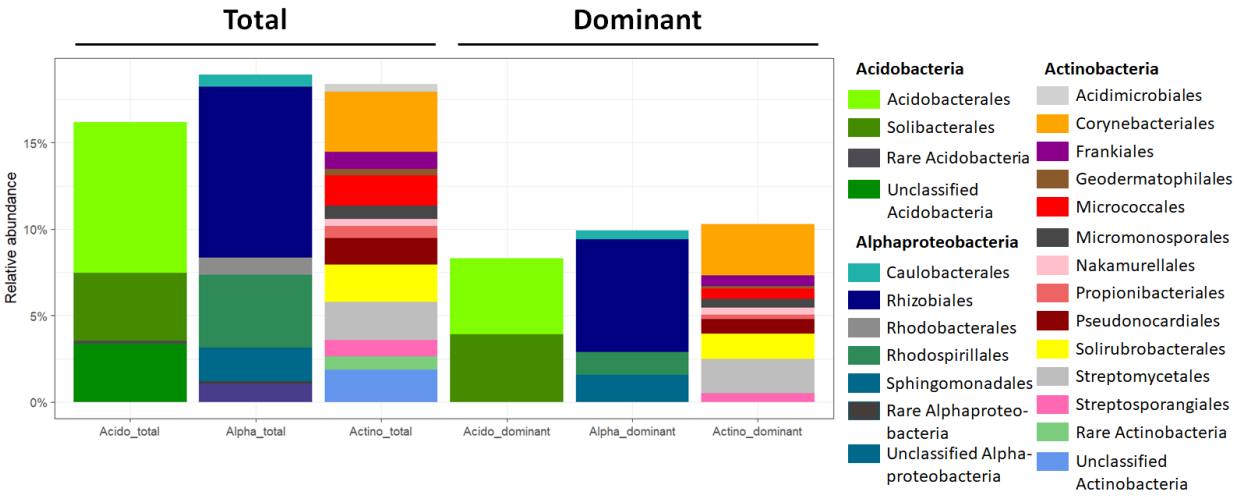

Figure 3.6: Relative abundance of Acidobacteria, Alphaproteobacteria and Actinobacteria in topcrust classified to orders. The total community is shown on the left and the dominant genera on the right. Proportions of total bacterial reads. Dominant = Genera with over 0.1% of total bacterial reads in all topcrust samples. N=*12*

The Alphaproteobacteria are very abundant in topcrust, representing 18.9% of the total bacterial community and 23.2% of dominant bacteria. Over 50% of the Alphaproteobacteria belong to the order Rhizobiales, both in the total community and dominant bacteria, representing 9.9% and 6.6% of bacterial reads respectively (Figure 3.6).

Acidobacteria account for 19.2% of the dominant bacteria and are significantly more abundant in topcrust than in subsoil where they account for 13.3% (*p*-value 0.014). About half of the dominant Acidobacterial genera belong to the order Acidobacterales and half to Solibacterales, accounting for 4.4% and 3.9% of total bacterial reads respectively (Figure 3.6). In the total community Acidobacterales account for 8.7% and Solibacterales 3.9% of total reads indicating that many Acidobacterales genera are in low abundance. A large proportion of Acidobacterial reads is unclassified (Figure 3.6).

Figure 3.7 and Table 3.5 show the relative abundance of the most abundant bacterial genera of topcrust and comparison to subsoil. The most abundant genera in topcrust are *Ktedonobacter*, *Bradyrhizobium* and Candidatus *Solibacter* each accounting for 3-4% of total bacterial reads.

*Table 3.5: Relative abundance (mean %) of the most abundant bacterial genera of topcrust and comparison between topcrust and subsoil (*p*-values). Proportions of total bacterial reads.* N *topcrust=12;* N *subsoil=4.*

Genus (Phylum)	**Topcrust**	**St.dev**	**Subsoil**	**St.dev**	***p*-values [(a)]**
Ktedonobacter (Chloroflexi)	4.0	1.2	1.5	0.4	<0.001*** [(b)]
Bradyrhizobium (Proteobacteria)	3.9	0.9	4.2	0.7	0.54
Candidatus *Solibacter* (Acidobacteria)	3.5	1.2	2.9	0.9	0.41
Mycobacterium (Actinobacteria)	2.3	0.6	1.5	0.2	0.0023** [(b)]
Streptomyces (Actinobacteria)	1.9	0.3	2.0	0.3	0.45
Granulicella (Acidobacteria)	1.2	0.5	0.2	0.1	<0.001*** [(b)]
Singulisphaera (Planctomycetes)	1.2	0.3	1.2	0.2	0.74
Sphingomonas (Proteobacteria)	0.9	0.1	0.4	0,0	<0.001*** [(b)]
Thermogemmatispora (Chloroflexi)	0.8	0.2	0.3	0.1	<0.001*** [(b)]
Conexibacter (Actinobacteria)	0.7	0.2	0.7	0.2	0.82
Frankia (Actinobacteria)	0.7	0.1	0.6	0.1	0.89
Solirubrobacter (Actinobacteria)	0.7	0.2	0.7	0.2	0.46
Terriglobus (Acidobacteria)	0.6	0.2	0.2	0.0	<0.001*** [(b)]
Gemmata (Planctomycetes)	0.6	0.2	0.7	0.1	0.22
Pseudomonas (Proteobacteria)	0.6	0.3	0.5	0.1	0.21
Acidobacterium (Acidobacteria)	0.6	0.1	0.2	0.1	<0.001*** [(b)]
Pseudonocardia (Actinobacteria)	0.5	0.2	0.7	0.2	0.28
Chthoniobacter (Verrucomicrobia)	0.5	0.2	0.3	0.1	0.048* [(b)]
Methylobacterium (Proteobacteria)	0.5	0.1	0.3	0.0	<0.001*** [(b)]
Silvibacterium (Acidobacteria)	0.5	0.1	0.2	0.0	<0.001*** [(b)]

(a) Statistical difference: $p < 0{,}001$*** ; $p < 0{,}01$** ; $p < 0{,}05$* obtained by Welch two sample t-test. Detailed output is shown in Appendix (table 8). (b) Enriched in topcrust.

The Chloroflexi genera *Ktedonobacter* and *Thermogemmatispora* are significantly more abundant in topcrust than in subsoil. This also applies to the Acidobacterial genera *Granulicella*, *Terriglobus*, *Acidobacterium* and *Silvibacterium*. The Acidobacterial genus Candidatus *Solibacter* is very variable between samples and the difference between topcrust and subsoil is not statistically significant (*p*-value 0.41). The Alphaproteobacterial genera *Sphingomonas* and *Mehylobacterium*, the Actinobacterial genus *Mycobacterium* and the Verrucomicrobial genus *Chthoniobacter* are also significantly more abundant in topcrust than subsoil.

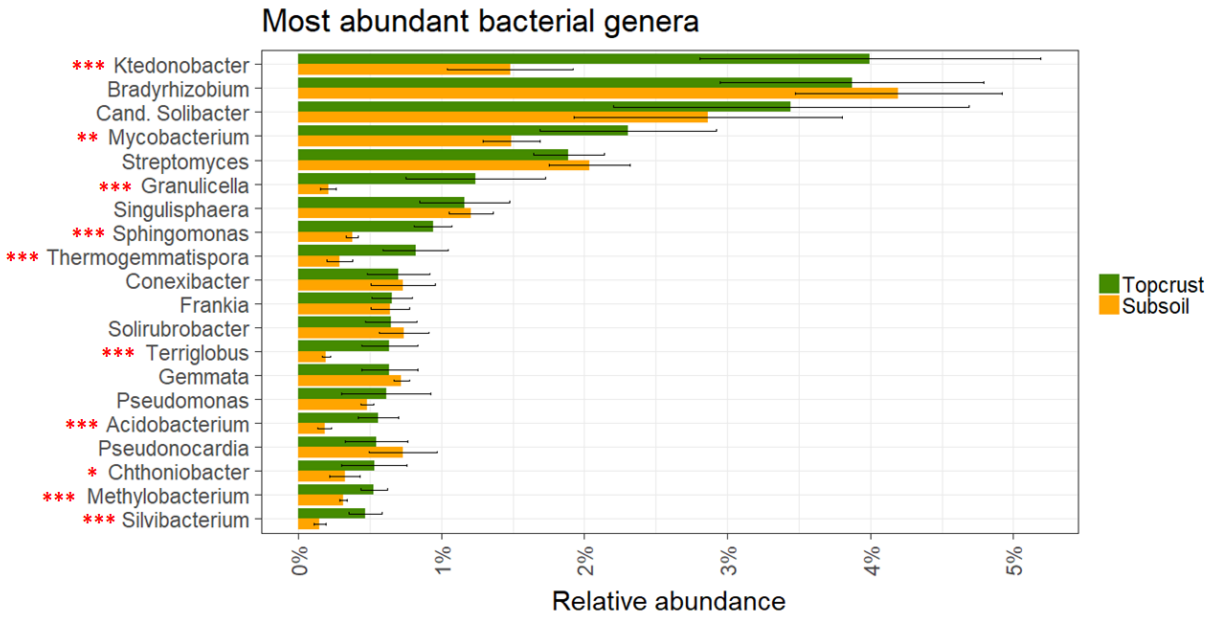

Figure 3.7: Relative abundance (mean) of the most abundant bacterial genera of topcrust and comparison to subsoil. Error bars show standard deviation. Proportions of total bacterial reads. N *topcrust=12 ;* N *subsoil=4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**

Cyanobacteria represent 4.1% of total bacterial reads in topcrust (Table 3.3) and 2.8% of dominant bacteria (Table 3.4). This difference indicates that most Cyanobacterial genera are in very low abundance. About 80 Cyanobacterial genera were found in each sample but only 17 had an abundance of over 0.1% of bacterial reads in at least one sample, further highlighting the low abundance of most Cyanobacterial genera. These 17 genera (Figure 3.8A) represent 53.5% of total Cyanobacterial reads. *Nostoc* is the most abundant Cyanobacterial genus representing 0.33% of total bacterial reads. Figure 3.8B shows the total Cyanobacterial community of the biocrust classified to orders. The most abundant order is Nostocales which accounts for 1.8% of total bacterial reads. The genera *Nostoc*, *Scytonema*, *Tolypotrix*, *Calothix*, *Fischerella*, *Anabaena* and *Hassallia* belong to the Nostocales. The Oscillatoriales genus *Microcoleus* is in low abundance and very variable between samples.

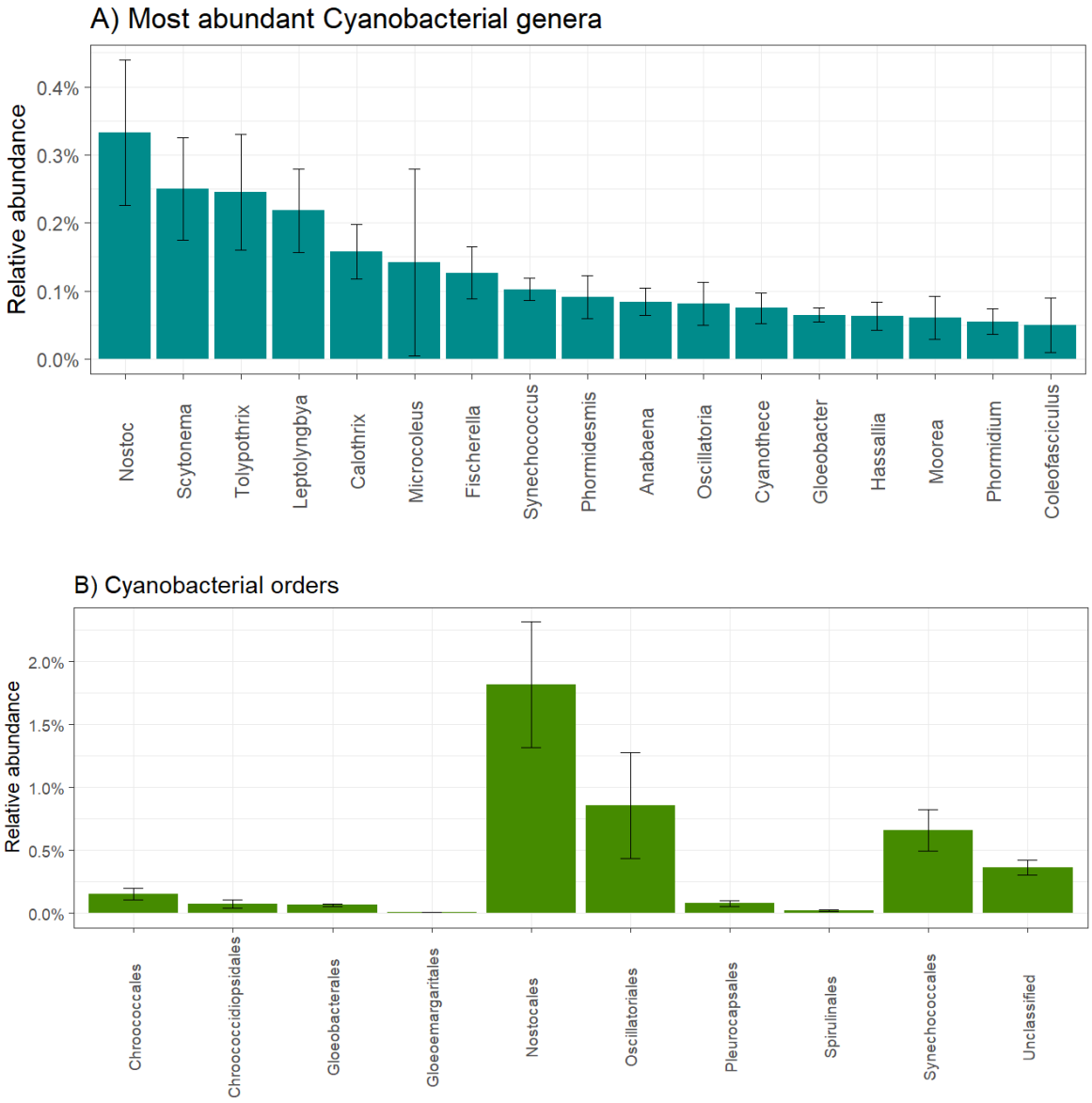

Figure 3.8: Relative abundance (mean) of A) Most abundant Cyanobacterial genera in topcrust. B) Total Cyanobacterial reads in topcrust classified to orders. Error bars show standard deviation. Proportions of total bacterial reads. N=*12*

3.1.3 Comparison between seasons, habitat types and sampling areas

In one sampling area, Gagnheiði, three paired samples were collected in early spring (24.05.2016) and late fall (14.09.2016) to compare the bacterial composition of the biocrust between seasons. The spring samples were collected very soon after the snow had melted. Paired t-tests were performed on the genus level for each site and no statistical difference was found. Additionally Welch two sample t-tests were conducted for the most abundant

genera shown in Figure 3.9. Statistical difference between seasons was only found for one of the genera, the Alphaproteobacteria *Sphingomonas* which was more abundant in the fall (*p*-value 0.04).

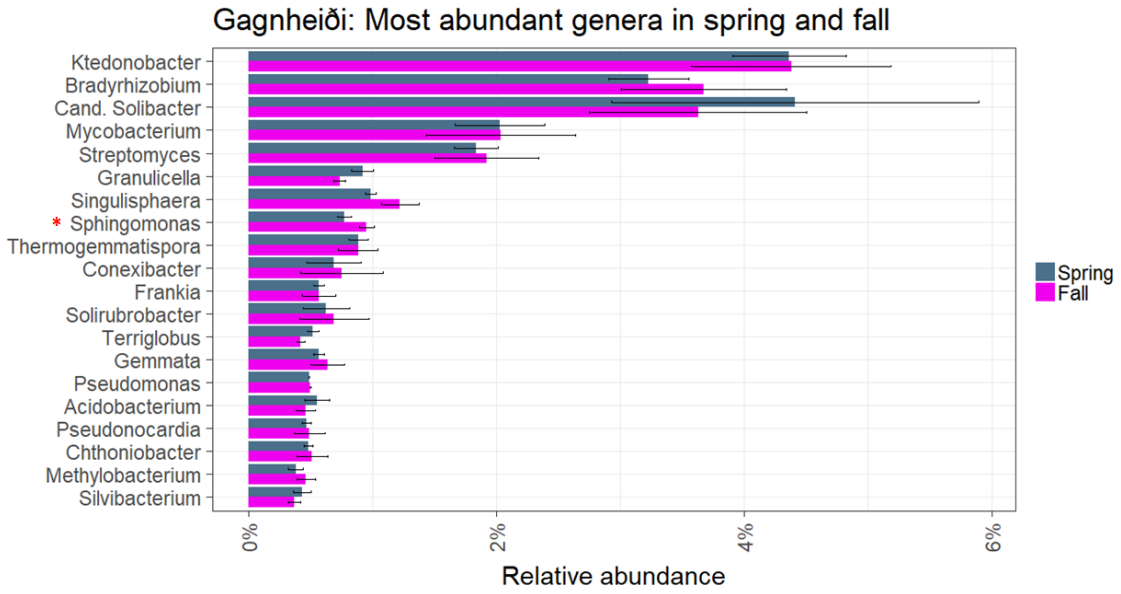

Figure 3.9: Relative abundance (mean) of the most abundant bacterial genera in spring and fall at Gagnheiði. Error bars show standard deviation. Proportions of total bacterial reads. N*=3. Statistical difference:* p *< 0.05*. Detailed output of tests is shown in Appendix (tables 9 and 19).*

To compare the bacterial composition of the biocrust between different habitats samples were collected in four habitat types (Table 1.1 and Table 2.1). Three samples were collected in Boreal moss snowbed communities (EUNIS E4.115) (all from Gagnheiði), six in Icelandic *Racomitrium ericoides* heaths (EUNIS E4.26) (two from Fjallabak, three from Skaftártunga and one from Laki), two in Icelandic lava field lichen heaths (EUNIS E4.241) (both from Laki) and one in Glacial moraines with very sparse or no vegetation (EUNIS H5.2) (from Fjallabak). Figure 3.10A shows PCA analysis of the samples emphasizing the habitat types. No habitat type pattern can be seen. ANOVA on the most abundant genera found a significant difference between habitat types for two of the genera, the Planctomycetes genus *Gemmata* (*p*-value=0.04) and the Verrucomicrobial genus *Chthoniobacter* (*p*-value <0.001), both significantly more abundant at Lava field lichen heaths compared to the other habitat types (Figure 3.11). EUNIS H5.2 (Glacial moraines with very sparse or no vegetation) was excluded from the analysis because there is only one sample representing that habitat type.

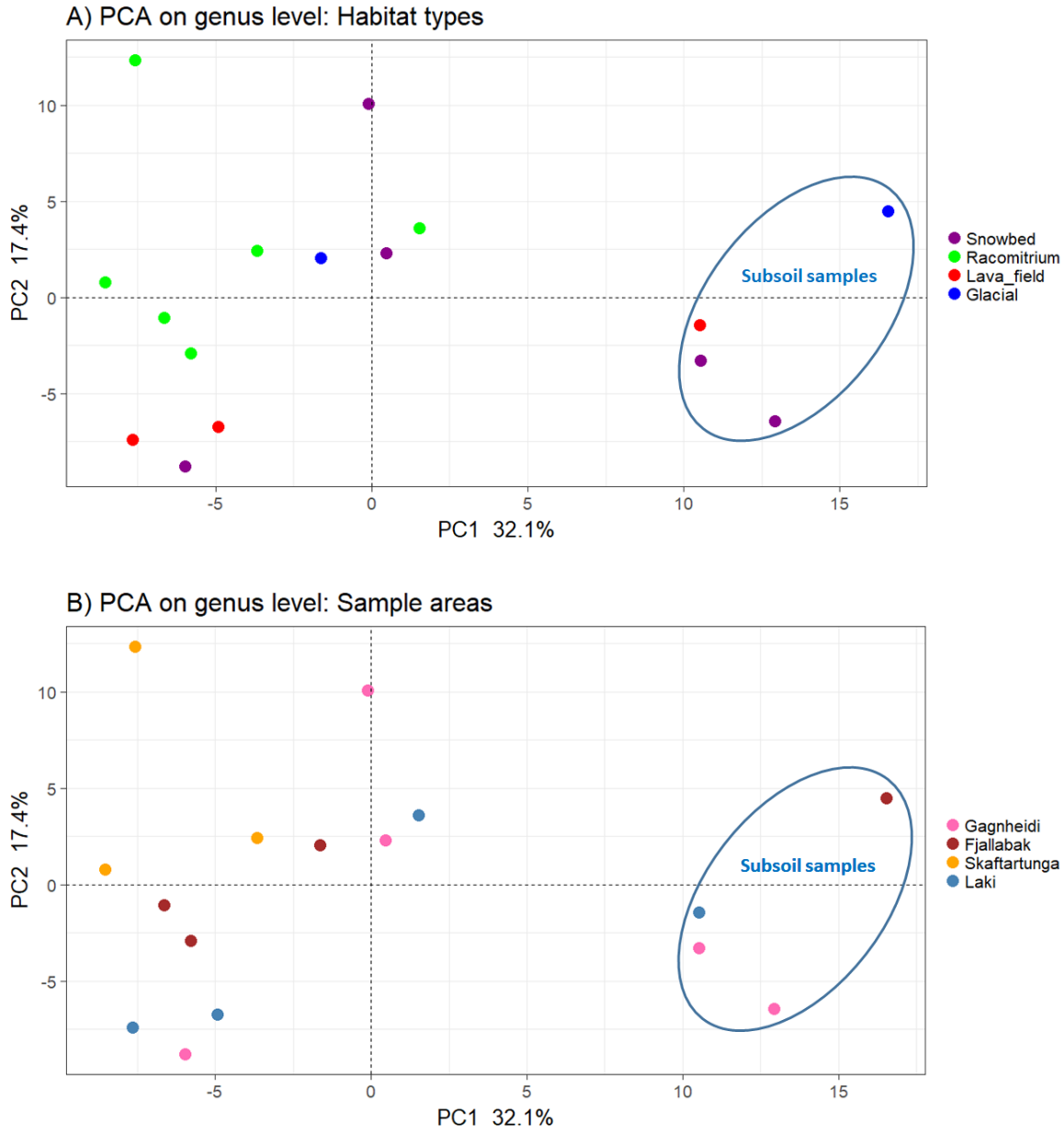

*Figure 3.10: A) PCA analysis on the genus level emphasizing different habitat types. Snowbed = Boreal moss snowbed communities (EUNIS E4.115) (*N=*3) ; Racomitrium = Icelandic* Racomitrium ericoides *heaths (EUNIS E4.26) (*N=*6) ; Lava_field = Icelandic lava field lichen heaths (EUNIS E4.241) (*N=*2) ; Glacial = Glacial moraines with very sparse or no vegetation (EUNIS H5.2) (*N=*1). B) PCA analysis on the genus level emphasizing the sample areas.* N *Gagnheiði=3;* N *Fjallabak=3;* N *Skaftártunga=3;* N *Laki=3.*

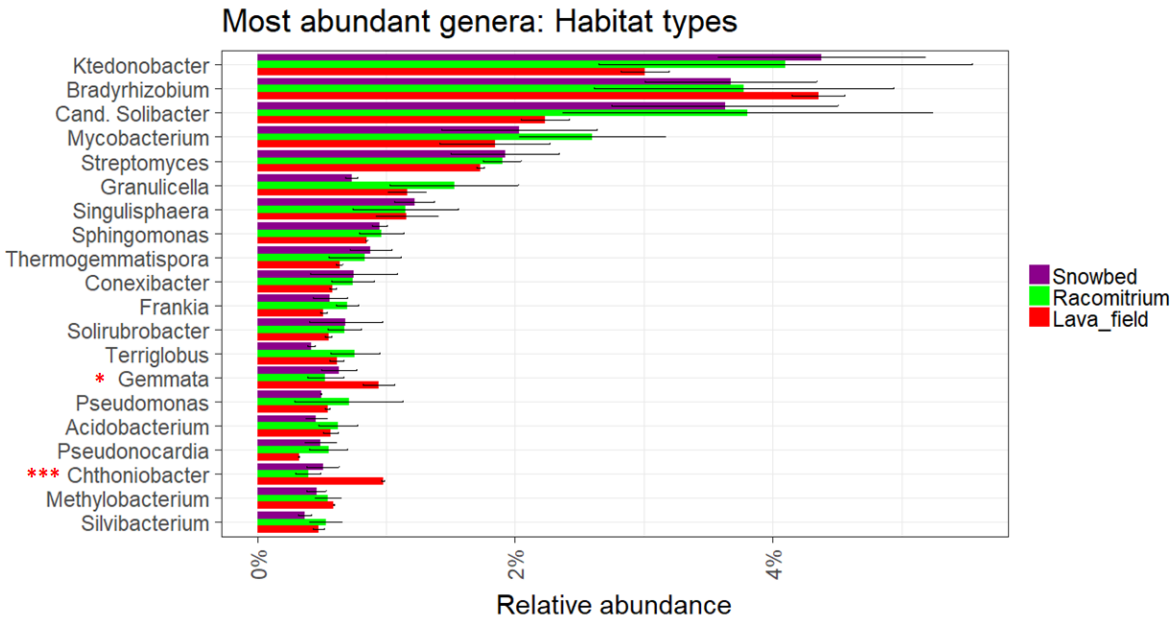

Figure 3.11: Relative abundance (mean) of the most abundant bacterial genera of different habitat types. Error bars show standard deviation. Proportions of total bacterial reads. Snowbed = Boreal moss snowbed communities (EUNIS E4.115) (N=3) ; Racomitrium = Icelandic Racomitrium ericoides *heaths (EUNIS E4.26) (N=6) ; Lava_field = Icelandic lava field lichen heaths (EUNIS E4.241) (N=2). Statistical difference:* $p < 0.001$*** ;* $p < 0.01$**; $p < 0.05$*. *Detailed output of tests is shown in Appendix (tables 17, 20 and 21).*

To compare the bacterial composition of the biocrust between different geographical areas, samples were collected in four sample areas (Table 2.1). Three samples were collected from each area; Gagnheiði, Fjallabak, Laki and Skaftártunga. Figure 3.10B shows PCA analysis on the genus level emphasizing the sample areas. No distinction between sample areas can be seen. ANOVA on the most abundant genera revealed significant differences between sample areas for two Acidobacterial genera, *Granulicella* (*p*-value=0.014) and *Terriglobus* (*p*-value=0.02), both significantly less abundant at Gagnheiði compared to Skaftártunga (Figure 3.12).

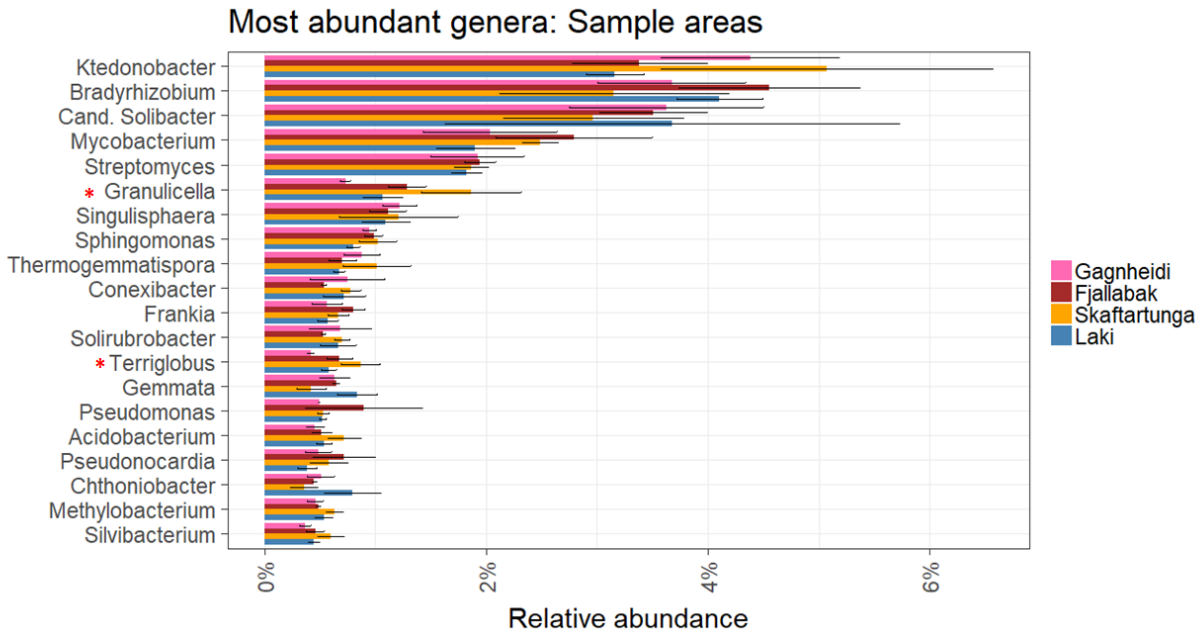

Figure 3.12: Relative abundance (mean) of the most abundant bacterial genera of different sample areas. Error bars show standard deviation. Proportions of total bacterial reads. N *Gagnheiði=3;* N *Fjallabak=3;* N *Skaftártunga=3;* N *Laki=3. Statistical difference:* p*<0.05*. Detailed output of tests is shown in Appendix (tables 18, 22 and 23).*

3.2 Functional gene analysis

3.2.1 Overview of functional gene analysis

Table 3.6 shows an overview of functional gene analysis in topcrust obtained from the MG-RAST server and Table 3.7 shows the same information for subsoil. Almost 90% of the reads could be assigned to proteins but only 33.1% of reads in topcrust and 33.3% in subsoil could be assigned to annotated proteins.

Table 3.6: Overview of functional trait analysis of topcrust. N=12

Sample	Read count	Unknown feature (%)	Unknown protein (%)	Annotated protein (%)	rDNA (%)
G1_september	9,887,406	9.5	57.4	33.0	0.1
G2_september	9,614,832	9.5	54.9	35.4	0.1
G3_september	7,131,144	11.4	58.1	30.3	0.1
F1	8,109,306	12.1	55.5	32.2	0.1
F2	6,527,781	12.1	55.3	32.5	0.1
F3	9,286,463	11.1	54.7	34.1	0.1
S1	5,158,583	9.5	57.9	32.5	0.1
S2	5,024.511	11.0	54.8	34.0	0.2
S3	5,277,008	8.5	56.8	34.6	0.1
L1	4,476,239	7.8	59.7	32.4	0.1
L2	4,727,296	11.8	58.7	29.3	0.1
L3	5,319,461	6.6	56.0	37.2	0.1
Average	**6,711,669**	**10.1**	**56.6**	**33.1**	**0.1**
St.dev	**1,947,142**	**1.7**	**1.6**	**2.0**	**0.0**

Table 3.7: Overview of functional trait analysis of subsoil. N=4.

Sample	Read count	Unknown feature (%)	Unknown protein (%)	Annotated protein (%)	rDNA (%)
G1_subsoil [(a)]	6,319,991	9.7	57.5	32.7	0.1
G2_subsoil [(a)]	7,911,981	9.8	56.2	34.0	0.1
L1_subsoil	9,792,575	12.3	54.1	33.5	0.1
F1_subsoil	8,819,386	10.4	56.4	33.1	0.1
Average	**8,210,983**	**10.6**	**56.1**	**33.3**	**0.1**
St.dev	**1,278,362**	**1.0**	**1.2**	**0.5**	**0.0**

(a) Subsoil samples from Gagnheiði collected in fall.

3.2.2 Functional traits of the biocrust and comparison to lower soil strata

Figure 3.13 shows PCA analysis of functional genes on level 3 SEED subsystems. The clustering patterns are similar as those seen for taxonomic analysis (Figure 3.2) and the same distinction between topcrust and subsoil can be seen.

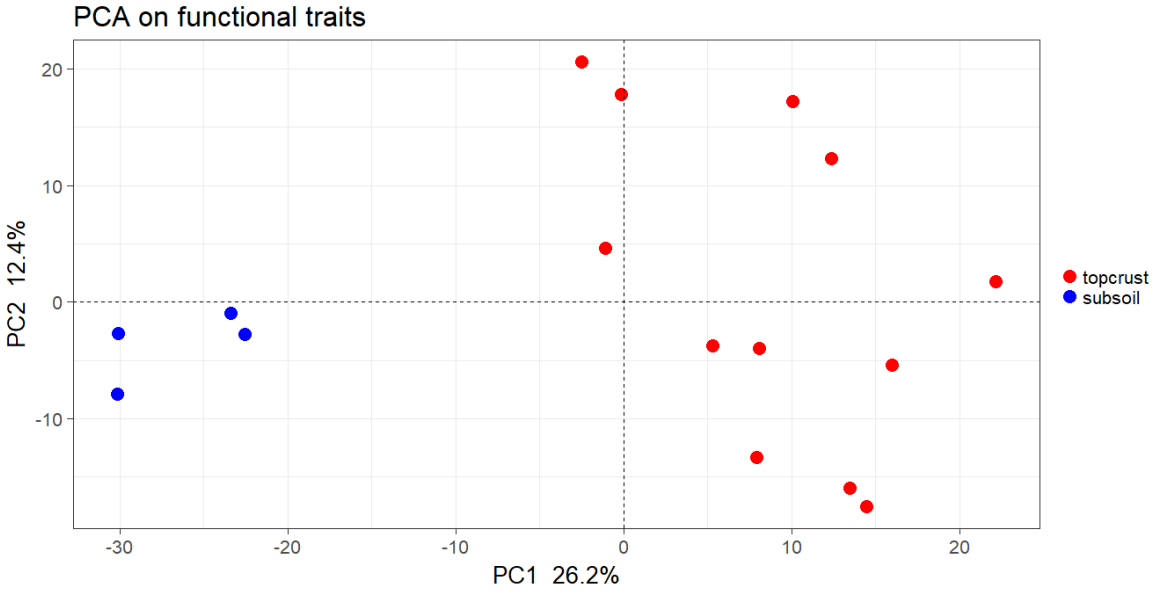

Figure 3.13: PCA of functional gene analysis on Level 3 SEED subsystems. N *topcrust=12;* N *subsoil = 4. Subsystems whith the highest loading scores for PC1 and PC2 are shown in Appendix (tables 3 and 4).*

Relative abundance of Level 1 SEED subsystems is shown in Table 3.8 and Figure 3.14. The most abundant subsystems are those for amino acids and derivatives, carbohydrates, clustering-based subsystems and protein metabolism.

*Table 3.8: Relative abundance (mean %) of Level 1 SEED subsystems and comparison between topcrust and subsoil (*p*-values). Proportions of Level 1 annotations.* N *topcrust=12;* N *subsoil=4.*

Level 1 SEED subsystems	**Topcrust**	**St.dev**	**Subsoil**	**St.dev**	***p*-values [a]**
Amino acids and derivatives	9.6	0.1	10.0	0.1	<0.001*** [c]
Carbohydrates	14.9	0.2	14.2	0.1	<0.001*** [b]
Cell division and cell cycle	1.0	0.0	0.9	0.0	0.056
Cell wall and capsule	3.5	0.1	3.5	0.0	0.78
Clustering-based subsystems	12.2	0.1	12.5	0.1	<0.001*** [c]
Cofactors, vitamins, prosthetic groups and pigments	5.4	0.1	5.2	0.0	<0.001*** [b]
DNA metabolism	4.4	0.1	4.4	0.1	0.28
Dormancy and sporulation	0.1	0.0	0.1	0.0	0.39
Fatty acids, lipids and isoprenoids	2.7	0.0	2.7	0.0	0.099
Iron acquisition and metabolism	0.6	0.0	0.5	0.0	0.24
Membrane transport	3.8	0.1	4.1	0.1	0.013* [c]
Metabolism of aromatic compounds	1.7	0.1	1.8	0.1	0.051
Miscellaneous	6.2	0.1	6.2	0.1	0.37
Motility and chemotaxis	1.1	0.1	0.9	0.0	0.0016** [b]
Nitrogen metabolism	1.2	0.1	1.3	0.0	0.073
Nucleosides and nucleotides	3.0	0.0	3.1	0.0	0.006** [c]
Phages, prophages, transposable elements, plasmids	1.3	0.0	1.4	0.0	0.83
Phosphorus metabolism	1.3	0.0	1.3	0.0	0.48
Photosynthesis	0.2	0.0	0.1	0.0	<0.001*** [b]
Potassium metabolism	0.9	0.0	0.8	0.0	0.021* [b]
Protein metabolism	8.0	0.2	8.2	0.1	0.04* [c]
RNA metabolism	3.5	0.1	3.5	0.1	0.26
Regulation and cell signaling	1.1	0.1	1.0	0.0	0.063
Respiration	4.7	0.1	4.7	0.0	0.91
Secondary metabolism	0.31	0.0	0.28	0.0	0.0013** [b]
Stress response	2.6	0.1	2.6	0.0	0.23
Sulfur metabolism	1.1	0.1	1.0	0.0	0.031* [b]
Virulence, disease and defense	3.6	0.2	3.5	0.1	0.73

(a) Statistical difference: $p < 0{,}001$*** ; $p < 0{,}01$** ; $p < 0{,}05$* obtained by Welch two sample t-test. Detailed output is shown in Appendix (table 10). (b) Enriched in topcrust. (c) Enriched in subsoil.

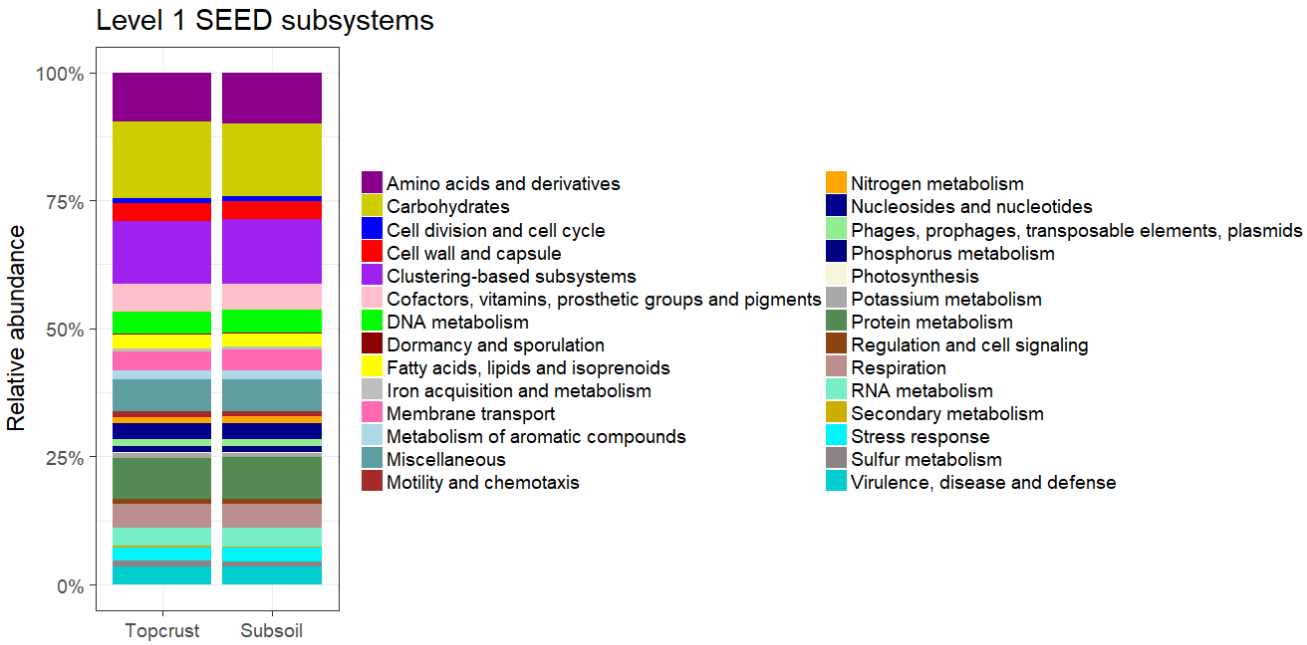

Figure 3.14: Relative abundance of Level 1 SEED subsystems. Proportions of Level 1 annotations. N *topcrust=12;* N *subsoil=4.*

Significant differences between topcrust and subsoil were found in 12 of 28 Level 1 subsystems (Table 3.8). Seven of these subsystems are more abundant in topcrust than subsoil; "carbohydrates" (*p*-value <0.001), "cofactors, vitamins, prosthetic groups and pigments" (*p*-value <0.001), „motility and chemotaxis" (*p*-value=0.0016), "photosynthesis" (*p*-value <0.001), "potassium metabolism" (*p*-value=0.021), "secondary metabolism" (*p*-value=0.0013) and "sulfur metabolism" (*p*-value=0.031). Reads from these subsystems were assigned to deeper (level 2) subsystems in the SEED hierarchy. The subsystem "secondary metabolism" was excluded from the analysis because reads from that subsystem were mainly related to plants and metazoans. The subsystems "amino acids and derivatives", "clustering-based subsystems", "membrane transport", "nucleosides and nucleotides" and "protein metabolism" were enriched in subsoil and not examined further.

In the carbohydrates subsystem the difference between topcrust and subsoil is driven by di- and oligosaccharide synthesis and utilization, monosaccharide metabolism, polysaccharide metabolism, CO_2 fixation, aminosugar synthesis and utilization, and various other carbohydrate metabolic traits (Figure 3.15).

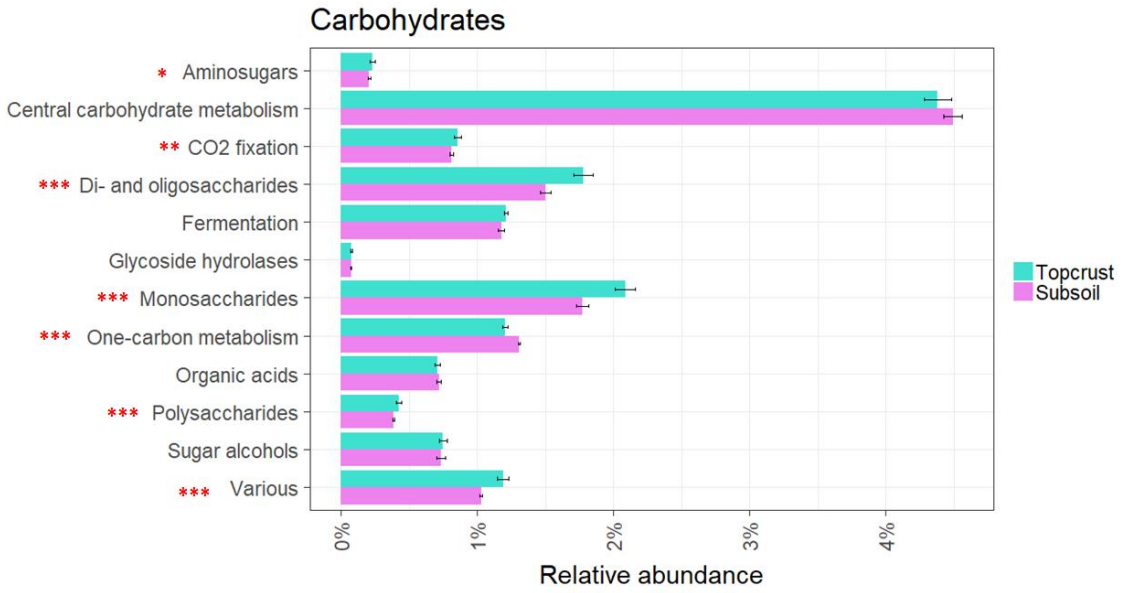

Figure 3.15: Relative abundance (mean) of carbohydrates Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12;* N *subsoil=4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**. *Detailed output of tests is shown in Appendix (table 11).*

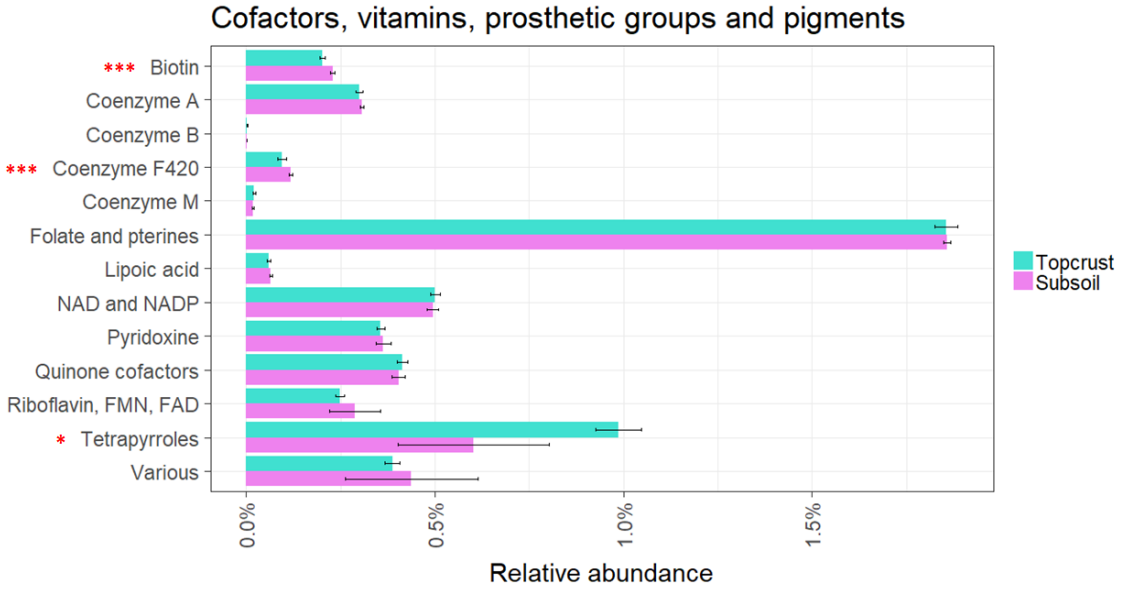

Figure 3.16: Relative abundance (mean) of cofactors, vitamins, prosthetic groups and pigments Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12 ;* N *subsoil=4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**. *Detailed output of tests is shown in Appendix (table 12).*

In the subsystem “cofactors, vitamins, prosthetic groups and pigments“ the difference in abundance between topcrust and subsoil is driven by tetrapyrroles while both biotin and coenzyme F420 are more abundant in subsoil (Figure 3.16). The tetrapyrroles subsystem contains mostly genes for vitamin B12 and chlorophyll biosynthesis.

In the motility and chemotaxis subsystem both bacterial chemotaxis and flagellar motility are enriched in topcrust (Figure 3.17).

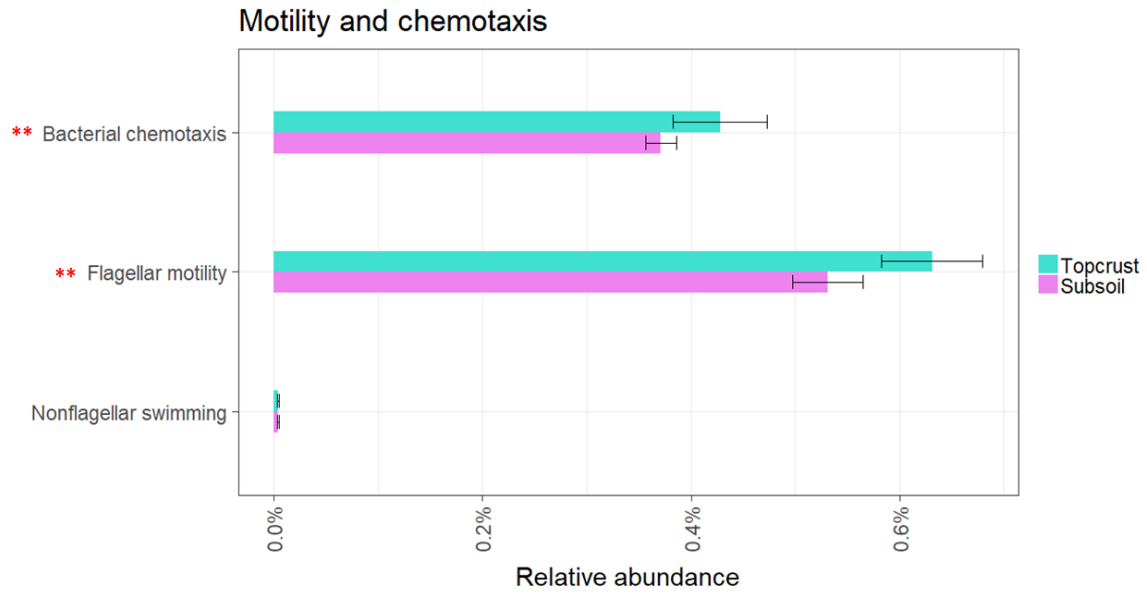

Figure 3.17: Relative abundance (mean) of motility and chemotaxis Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12;* N *subsoil=4. Statistical difference:* p < *0.001*** *;* p < *0.01*** *;* p < *0.05**. *Detailed output of tests is shown in Appendix (table 13).*

In the photosynthesis subsystem all level 2 subsystems are significantly more abundant in topcrust, especially electron transport and light harvesting subsystems which are in very low abundance in subsoil (Figure 3.18).

In the potassium metabolism subsystem the difference between topcrust and subsoil is driven by potassium homeostasis (Figure 3.19).

For sulfur metabolism the difference in abundance between topcrust and subsoil can be explained by galactosylceramide and sulfatide metabolism although sulfur oxidation is more abundant in subsoil (Figure 3.20).

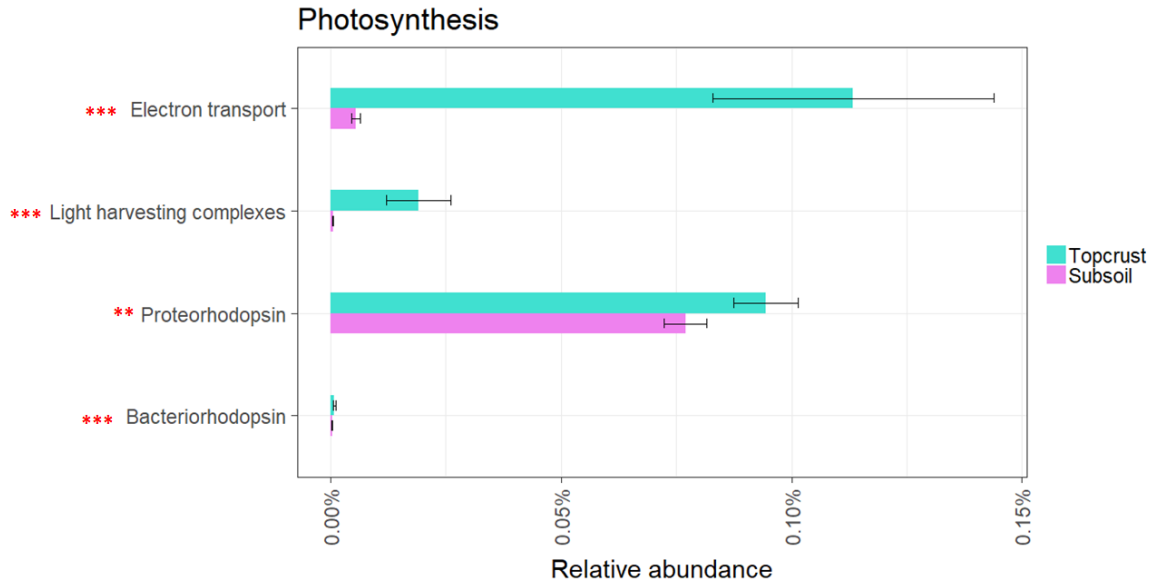

Figure 3.18: Relative abundance (mean) of photosynthesis Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12;* N *subsoil=4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**. *Detailed output of tests is shown in Appendix (table 14).*

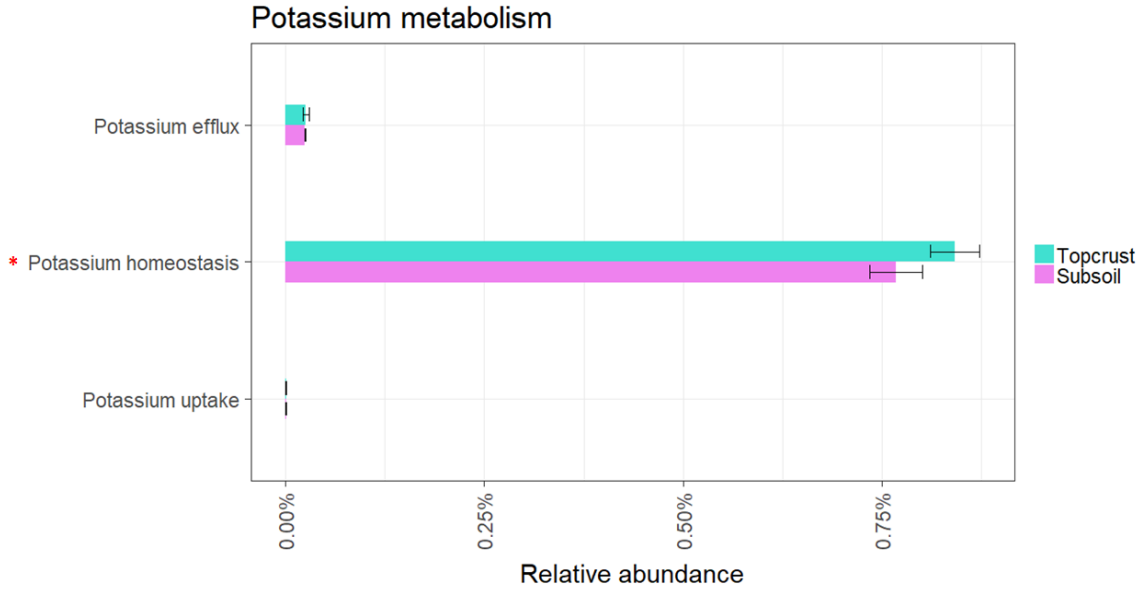

Figure 3.19: Relative abundance (mean) of potassium metabolism Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12;* N *subsoil=4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**. *Detailed output of tests is shown in Appendix (table 15).*

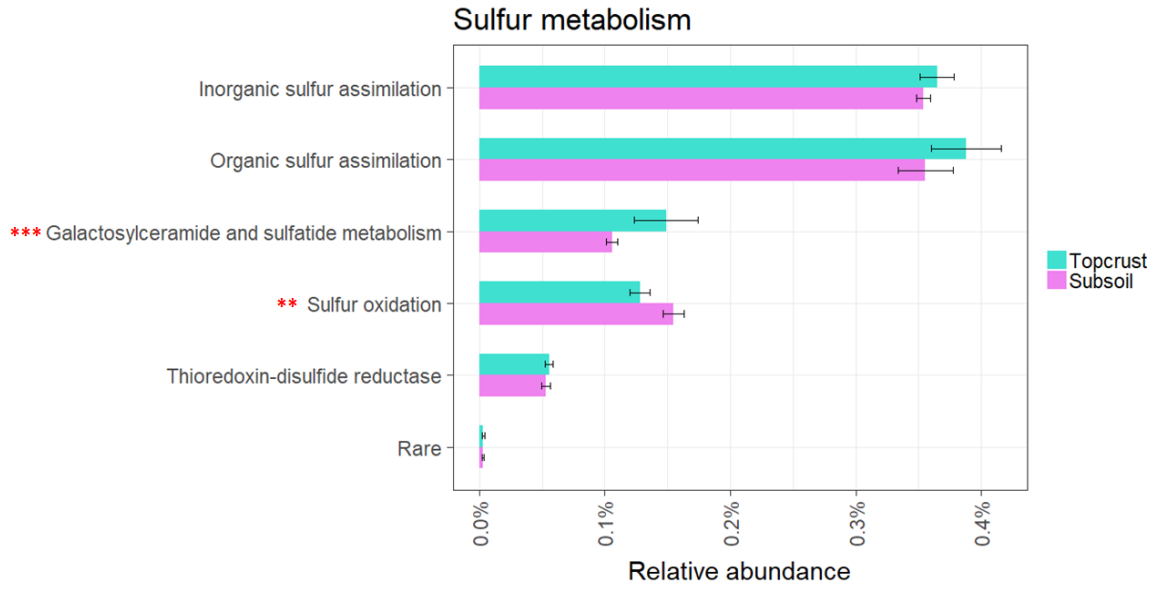

Figure 3.20: Relative abundance (mean) of sulfur metabolism Level 2 SEED subsystems. Error bars show standard deviation. Proportions of Level 2 annotations. N *topcrust=12;* N *subsoil = 4. Statistical difference:* p < *0.001*** ; p < *0.01*** ; p < *0.05**. *Detailed output of tests is shown in Appendix (table 16).*

4 Discussion

DNA recovery was very variable in topcrust but on average higher than in subsoil (*p*-value 0.024) (Figure 3.1) and could indicate higher microbial biomass of the biocrust compared to lower soil strata. This is consistent with findings from the Colorado Plateau where DNA concentration was up to four fold higher in biocrust than in subsurface soils (Steven *et al*., 2013a). In the Mojave desert Mogul *et al*. (2017) found 16S rDNA copy abundance to be six fold higher in biocrust than subsoil in sites with low biocrust coverage but found no statistical differences between soil strata for sites with high and intermediate biocrust cover. In an earlier study in the Mojave desert Steven *et al*. (2014) found no difference between DNA recovery from biocrust and shrub root zones, however bacterial rDNA copies were enriched in the biocrust.

In this study whole shotgun metagenome sequencing was applied. Most environmental microbiology studies have been conducted using 16S rDNA amplicon sequencing (e.g. Abed *et al*., 2019; Chilton *et al*., 2018; Delgado-Baquerizo *et al*., 2018; Malard *et al*., 2019; Moreira-Grez *et al*., 2019). The metagenome approach has the advantage that functional genes can be assessed in addition to taxonomy and the method overcomes some of the weaknesses of amplicon sequencing such as primer bias (Breitwieser *et al*., 2019). The main weakness of whole shotgun metagenome sequencing is the requirement of reference genomes for classification (Breitwieser *et al*., 2019). While some taxa, e.g. Cyanobacteria are underrepresented in the databases (Casaburi *et al*., 2016), others are overrepresented, e.g. human pathogens and microbiome taxa (Menzel *et al*., 2016). Species which are difficult to culture in laboratory settings are also underrepresented in databases and are not covered by the standard nomenclature (Breitwieser *et al*., 2019).

The two different approaches to analysing microbial composition of environmental samples give comparable results. When characterizing stromatolite microbiomes from the Bahamas Casaburi *et al*. (2016) used both targeted 16S rDNA amplicon sequencing and whole metagenome shotgun sequencing and found strong correlation between the two approaches although the metagenomic approach allowed higher taxonomic resolution with more complex community structure. The study also found strong correlation between phylogeny and function. Fierer *et al*. (2012) also found strong correlation between taxonomic analysis based on 16S rRNA genes obtained from amplicon sequencing and from whole shotgun sequencing.

4.1 Taxonomic analysis

4.1.1 Methodology

For taxonomic analysis the raw sequence data was uploaded to the Kaiju web server (Menzel *et al*., 2016). The program translates the sequence reads into all six possible reading frames and searches for amino acid sequence matches in a database of annotated proteins from microbial reference genomes. The setting "greedy" which allows some mismatches using the BLOSUM62 substitution matrix was chosen with a minimum match score of 75. The advantage of using protein level sequence comparison is better classification of taxa which are underrepresented in the databases although the disadvantage is that it is not possible to classify reads from non-protein coding regions (Menzel *et al*., 2016)

4.1.2 Overview of taxonomic analysis

Only about 45% of the sequence reads could be assigned to taxa (Table 3.1 and Table 3.2) using the settings on Kaiju web server described above. This can be explained by the underrepresentation of many soil taxa in the databases due to the difficulties in laboratory culturing. Menzel *et al*. (2016) could classify about 25% of the sequence reads of desert soil using similar settings on the Kaiju server.

Bacteria are dominant in the microbial communities both in topcrust and subsoil, representing 91.5% and 98.4% of identifiable reads respectively (Table 3.1 and Table 3.2). The difference can be explained by a higher abundance of fungi in the topcrust than in lower soil strata. Using metagenome data from this study Guðmundsdottir & Andresson (2019) analysed the fungal composition of the biocrust and found differences between the biocrust and lower soil strata. A high proportion of bacterial reads can be assigned to genera, 69% in topcrust and 59.9% in subsoil.

There is a clear distinction between the bacterial composition of the biocrust (topcrust) and underlying soil strata (subsoil) (Figure 3.2), which could indicate functional differences. In this study the topcrust subsamples were from the uppermost 5 mm and the subsoil subsamples were cut out at 15 mm depth. In a biocrust from the Kalahari desert Elliott *et al*. (2014) found significant differences between the biocrust (0-1 cm) and subsurface soil (1-2 cm), in particular the abundance of Cyanobacteria and Bacteroidetes was higher in the biocrust, but in subsurface soil Acidobacteria, Actinobacteria, Chloroflexi, and Firmicutes were more abundant (Elliott *et al*., 2014). In the Colorado Plateau Steven *et al*. (2013a) found Cyanobacterial and Proteobacterial abundance to be significantly higher in biocrust samples collected at 0-1 cm depth than in below-crust soils collected at 2-5 cm depth. In the subsurface soils Archaea and Chloroflexi were significantly enriched (Steven *et al*., 2013a). In a study in the Mojave desert Steven *et al*. (2014) found differences between biocrust and shrub root zones both in the bacterial community structure and functional gene analysis. Cyanobacteria were dominant in the biocrust while Actinobacteria and Proteobacteria were enriched in the root zones (Steven *et al*., 2014). In another study in the Mojave desert Mogul

et al. (2017) studied the bacterial composition across a gradient of biocrust coverage. At high coverage sites Cyanobacterial abundance in topsoil (0-1 cm) was three fold higher than in subsurface soil (1-2 cm) but when biocrust coverage was low the Cyanobacterial abundance in topsoil and subsurface soil was similar (Mogul *et al*., 2017). Maier *et al*. (2014) also found differences between biocrust (0-1 cm) and below-crust soil (1-3 cm) in the Tabernas desert, Spain. Cyanobacteria were significantly more abundant in biocrust and in below-crust soil Acidobacteria, Verrucomicrobia, Gemmatimonadetes, Planctomycetes, Armatimonadetes and Archaea were more abundant (Maier *et al*., 2014).

In the *Anthelia* biocrust 86 genera have abundance over 0.1% of bacterial reads across all topcrust samples, representing 43.2% of total bacterial reads and are considered dominant (Figure 3.3). In subsoil 89 dominant genera have abundance over 0.1% of bacterial reads across all subsoil samples, representing 35.7% of total bacterial reads. These dominant genera only account for about 4.5% of all genera highlighting the dominance of a few genera in the community. Malard *et al*. (2019) used similar criteria when defining dominant genera of Arctic soil and found that only 0.3% of all taxa accounted for 32% of total bacterial reads. In a global survey of soil bacterial communities dominant phylotypes were defined based on being both highly abundant and found in more than half of the samples collected worldwide. Using these criteria only 2% of all phylotypes representing 41% of total reads were considered dominant (Delgado-Baquerizo *et al*., 2018).

4.1.3 Most abundant taxa

In topcrust the most abundant phyla are the Actinobacteria, Proteobacteria (mainly Alphaproteobacteria) and Acidobacteria, both in the total community and amongst dominant bacteria (Table 3.3, Table 3.4, Figure 3.4 and Figure 3.5).

The Actinobacteria are more abundant in subsoil than topcrust although the difference is not statistically significant (Figure 3.5). Six Actinobacterial genera are among the most abundant genera of topcrust but only *Mycobacterium* which accounts for 2.3% of topcrust's reads is more abundant in topcrust than subsoil (*p*-value 0.0023). The other genera, *Frankia*, *Pseudonocardia*, *Streptomyces*, *Conexibacter* and *Solirubrobacter*, are of similar abundance in topcrust and subsoil (Table 3.5 and Figure 3.7). *Mycobacterium* species are generally free living saprophytes found in soil although many species of this genus can cause human diseases (Barka *et al*., 2016). Species of the genus *Frankia* are nitrogen fixers which can form symbiotic relationships with angiosperms via root nodules (Barka *et al*., 2016). Some species of the genus *Pseudonocardia* have also been shown to fix nitrogen (Mahendra & Alvarez-Cohen, 2005). *Streptomyces* species are ubiquitous and abundant in soil ecosystems where they are important saprophytes recycling insoluble carbon from plants and fungi. *Streptomyces* produce a variety of hydrolytic enzymes and secondary metabolites such as antibiotics (Barka *et al*., 2016). The genera *Conexibacter* and *Solirubrobacter* both belong to the order Solirubrobacterales (Figure 3.6). All reads matching the genus *Conexibacter* could be assigned to *C. woesei*, the type species of the genus. *C. woesei* is a slow growing aerobe which can reduce nitrate to nitrite, thus participating in the nitrogen cycle of soils

(Monciardini *et al*., 2003). About half of the *Solirubrobacter* reads could be assigned to *S. soli* which is a non-motile, non-spore forming, aerobic rod first isolated from a ginseng field soil (Kim *et al*., 2007).

Species of the Actinobacterial genera *Mycobacterum*, *Frankia*, *Pseudonocardia*, *Streptomyces*, *Conexibacter* and *Solirubrobacter* are found on the list of the most dominant phylotypes of soil bacteria worldwide (Delgado-Baquerizo *et al*., 2018).

In the total bacterial community Alphaproteobacteria are more abundant in topcrust than in subsoil but vice versa for the dominant bacteria although the difference is not statistically significant in either case (Table 3.3, Table 3.4 and Figure 3.5). The order Rhizobiales is dominating, accounting for over 50% of the Alphaproteobacterial reads of topcrust, both in the total community and dominant bacteria (Figure 3.6). Members of the order Rhizobiales are nitrogen fixers, best known for their symbiotic relationship with legumes (Madigan *et al*., 2006) although some taxa are free living (Ludwig, 1984). Rhizobiales members are also found in the rhizosphere of non-legume plants (Fischer *et al*., 2012) and in lichen endosymbioses (Erlacher *et al*., 2015).

Two Rhizobiales genera are among the most abundant bacterial genera of topcrust, *Bradyrhizobium* and *Methylobacterium* (Table 3.5 and Figure 3.7). *Bradyrhizobium* accounts for about 4% of bacterial reads and is of similar abundance in topcrust and subsoil. *Bradyrhizobium* is a very diverse genus with species that are well known for their nitrogen fixing properties, both in symbiosis and free living, and some *Bradyrhizobium* species are phototrophs (Avontuur *et al*., 2019). *Methylobacterium* accounts for 0.5% of the bacterial community in topcrust and is significantly more abundant than in subsoil (p-value <0.001). Elliott *et al*. (2014) also found *Methylobacterium* among the most abundant genera of a biocrust in the Kalahari desert and significantly enriched in biocrust compared to subsurface soil. The genus *Sphingomonas* of the order Sphingomonadales is also significantly more abundant in topcrust (p-value <0.001), representing 0.9% of bacterial reads. *Sphingomonas* species are found in numerous different habitats, including soil, and some have been shown to produce various EPS substances (White *et al*., 1996). Members of both *Methylobacterium* and *Sphingomonas* are aerobic anoxygenic phototrophs and several phototrophic isolates of these genera have been reported in biocrusts (Csotonyi *et al*., 2010).

Species of the Alphaproteobacterial genera *Bradyrhizobium*, *Methylobacterium* and *Sphingomonas* are found on the list of the most dominant phylotypes of soil bacteria worldwide (Delgado-Baquerizo *et al*., 2018).

The Acidobacteria are more abundant in topcrust than subsoil (Table 3.3, Table 3.4 and Figure 3.5) although the difference is only statistically significant for the dominant bacteria (p-value 0.014). Members of the dominant Acidobacteria are almost evenly distributed between the orders Acidobacterales and Solibacterales but in the total community the Acidobacterales account for over 50% of Acidobacterial reads and a large proportion is unclassified (Figure 3.6).

Genera of the phylum Acidobacteria are in high abundance in topcrust (Table 3.5 and Figure 3.7). The most abundant Acidobacterial genus is Candidatus *Solibacter* (mainly Candidatus *S. usitatus*) which accounts for 3.5% of total bacterial reads in topcrust. Candidatus *S. usitatus* was originally isolated from a pasture of ryegrass and clover in Virginia, Australia (Joseph *et al*., 2003). It is a slow growing, aerobic heterotroph, only culturable in low nutrient media (David *et al*., 2005; Ward *et al*., 2009). Ward *et al*. (2009) sequenced the genome of Candidatus *S. usitatus* and found that it has a very large genome encoding for 8.097 proteins. It can utilize various simple sugars in addition to complex carbohydrates such as chitin, pectin, starch, xylan and cellulose and has a large number of glycoside hydrolases. The genome also contains various genes that are advantageous in low nutrient environments such as high affinity iron transporters, carbon monoxide dehydrogenase and a large number of high affinity ABC transporters. The genome also has a very high number of PEP-CTERM proteins which are associated with EPS biosynthesis and in culture it produces an abundance of EPS (Ward *et al*., 2009). Elliott *et al*. (2014) also found Candidatus *Solibacter* among the most abundant genera of biocrust in the Kalahari desert.

The Acidobacterial genera *Granulicella, Terriglobus, Acidobacterium* and *Silvibacterium* are among the most abundant genera of topcrust. All these genera are significantly enriched in topcrust (*p*-values <0.001 in all cases) (Table 3.5 and Figure 3.7). Production of EPS has been reported for *Granulicella* (Kielak *et al*., 2017), *Terriglobus* (Whang *et al*., 2014) and *Acidobacterium* (Ward *et al*., 2009).

Candidatus *Solibacter* is found on the list of the most dominant phylotypes of soil bacteria worldwide but not the other abundant Acidobacteria. However the list contains many unclassified Acidobacteria (Delgado-Baquerizo *et al*., 2018).

The phylum Chloroflexi is abundant in topcrust and significantly more abundant than in subsoil both in the total bacterial community (*p*-value 0.003) and in dominant bacteria (*p*-value <0.001) (Table 3.3, Table 3.4 and Figure 3.5). Two Chloroflexi genera, *Ktedonobacter* and *Thermogemmatispora*, both of the class Ktedonobacteria, are in high abundance in topcrust and significantly more abundant than in subsoil (*p*-values<0.001 in both cases) (Table 3.5 and Figure 3.7). Most species of the class Ktedonobacteria, including *Ktedonobacter racemifer*, can oxidize carbon monoxide which is useful in low nutrient environments (King & King, 2014).

The genus *Ktedonobacter* (mostly *K. racemifer*) is the most abundant genus of topcrust representing 4% of total bacterial reads. *Ktedonobacter* is the type genus of the family Ktedonobacteraceae, a deeply branching lineage of the phylum Chloroflexi (Cavaletti *et al*., 2006). *K. racemifer*, which was originally isolated from Italian soil, is a mesophilic, mildly acidophilic, aerobic gram positive species. The bacterium is filamentous, produces a branched mycelium and forms spores in grape like clusters. It is capable of growing under microaerophilic conditions but not anaerobically (Cavaletti *et al*., 2006). Whole genome sequencing of *K. racemifer* revealed a very large genome (over 13 Mbp) with over 11

thousand protein coding genes and a high number of transposon associated genes (Chang *et al*., 2011).

The Chloroflexi genus *Thermogemmatispora* accounts for 0.8% of total bacterial reads in topcrust. Species of this genus are mildly acidophilic, thermophilic heterotrophs which have been isolated from geothermal soils, biofilms and compost (King & King, 2014; Yabe *et al*., 2011; Zheng *et al*., 2019). *Thermogemmatispora* species are spore forming and filamentous and produce branched mycelia (King & King, 2014; Yabe *et al*., 2011; Zheng *et al*., 2019).

Ktedonobacter and *Thermogemmatispora* are not found on the list of the most dominant phylotypes of soil bacteria worldwide. However the list contains many unclassified Chloroflexi phylotypes (Delgado-Baquerizo *et al*., 2018).

One genus of the class Gammaproteobacteria is among the most abundant genera of topcrust, *Pseudomonas* (Table 3.5 and Figure 3.7). Many *Pseudomonas* species are well known for their production of EPS and formation of biofilms (Ude *et. al*., 2006; Flemming *et al*., 2016) and may therefore be important in the formation of the biocrust. *Pseudomonas* is found on the list of the most dominant phylotypes of soil bacteria worldwide (Delgado-Baquerizo *et al*., 2018).

Compared to biocrusts worldwide, e.g. from USA (Mogul *et al*., 2017), the Arabic peninsula (Abed *et al*., 2019) and Australia (Moreira-Grez *et al*., 2019), Cyanobacteria are in low abundance in the *Anthelia* biocrust representing only 4.1% of the total bacterial community. However, their contribution to the biocrust may be of particular importance because of their well-known photosynthetic and nitrogen fixing properties in addition to their production of EPS (Weber *et al*., 2016). The most abundant Cyanobacterial order in the biocrust is Nostocales which includes the three most abundant Cyanobacterial genera, *Nostoc*, *Scytonema* and *Tolypotrix* (Figure 3.8). The Nostocales are filamentous nitrogen fixers found in various habitats, both free living and in symbiosis, e.g. with lichens (Whitton, 2013). The non-nitrogen fixing biocrust pioneer *Microcoleus* is in very low abundance and very variable between samples (Figure 3.8). The dominance of the Nostocales in the Cyanobacterial community could indicate that the biocrust is in a late successional stage (Belnap, 2002). The observed low abundance of Cyanobacteria compared to biocrust worldwide may be related to the colder Icelandic climate considering the very low Cyanobacterial abundance in biocrust from Svalbard (Mugnai *et al*., 2015) and from high elevations in Idaho, USA, where the climate is cold and wet (Blay *et al*., 2017). However, in a study in the high Arctic Ellismere Island, Canada, Cyanobacterial abundance was high (Steven *et al*., 2013b).

No Cyanobacteria are found on the list of the most dominant phylotypes of soil bacteria worldwide (Delgado-Baquerizo *et al*., 2018).

The main functions that Cyanobacteria contribute to biocrusts, i.e. photosynthesis, nitrogen fixation and EPS production, can be carried out by other biocrust organisms. Other photosynthetic organisms are in abundance in the biocrust, e.g. the liverwort *Anthelia*

juratzkana, which characterizes the biocrust and Alphaproteobacterial anoxygenic phototrophs. Other nitrogen fixers are also in abundance, e.g. the Rhizobiales and the Actinobacterial genus *Frankia*. Many other members of the biocrust can produce EPS, e.g. the Acidobacteria and many Proteobacterial genera. Also, Mugnai *et al*. (2015) found an abundance of EPS in the Svalbard biocrust in spite of low Cyanobacterial abundance and could by further analysis link the production to Proteobacteria.

The bacterial communities of liverwort based biocrusts have not been investigated to date according to the author's knowledge. In addition there is lack of standardization in the literature regarding bacterial composition of biocrusts and it is therefore difficult to compare the *Anthelia* biocrust to biocrusts worldwide.

4.1.4 Seasons, habitat types and sample areas

Sampling in two different seasons (spring and fall) at Gagnheiði showed no significant difference (Figure 3.9) suggesting stability in the bacterial community composition of the biocrust. Statistical difference was only found for one of the most abundant genera, *Sphingomonas*, which was more abundant in the fall.

Samples were collected at four different EUNIS habitat types and in four different sample areas (Table 2.1). The habitat types all have similar properties (Table 1.1) and the appearance of the biocrust was uniform in all habitat types and sample sites (Figure 1.4). PCA analysis of the samples did not indicate any separation between habitat types or sample areas (Figure 3.10). Two of the most abundant genera were significantly more abundant in Lava field lichen heath compared to the other habitat types (Figure 3.11) and two Acidobacterial genera were significantly less abundant at Gagnheiði compared to Skaftártunga (Figure 3.12). This further highlights the stability and homogeneity of the bacterial community of the biocrust.

4.2 Functional traits

In this study sequence reads were uploaded to the MG-RAST web server (Meyer *et al*., 2008) for functional gene analysis and the hierarchical SEED subsystems used as a reference database. A subsystem is a collection of functional roles which account for a biological process or structural complex and may be thought of as generalization of the term pathway. Protein families, i.e. proteins that share a common domain structure, may conduct the same or multiple function and non-homologous proteins that implement a single function are grouped together (Overbeek *et al*., 2005).

Table 3.6 and Table 3.7 show an overview of functional gene analysis. Only a third of the reads could be assigned to known proteins indicating that much of the data represents poorly characterized genes. This may be explained by the underrepresentation of soil bacteria genomes in the databases and possibly the lack of proper annotation of the existing genomes. PCA analysis of the samples on level 3 in the SEED subsystem hierarchy (Figure 3.13)

shows the same clear distinction between topcrust and subsoil as seen in the taxonomic analysis (Figure 3.2). This confirms the difference between the two soil strata and reflects a structural and functional uniqueness of the biocrust.

The most abundant level 1 SEED subsystems are amino acids and derivatives, protein metabolism, and clustering-based subsystems which are enriched in subsoil, and carbohydrates, which is enriched in topcrust (Table 3.8 and Figure 3.14). The amino acid and derivatives subsystem includes various metabolic pathways of amino acids. Over 60% of the protein metabolism subsystem is protein biosynthesis which includes ribosomal proteins, tRNA aminoacylation enzymes, translation initiation factors among others. The subsystem also includes protein degradation, protein folding and various other protein processing subsystems. The clustering-based subsystem includes miscellaneous undefined functional roles of hypothetical or putative proteins, further indicating poor annotation of reference genomes for soil microbiota in the databases.

The carbohydrate subsystem is significantly more abundant in topcrust than in subsoil (p-value <0.001). The difference in abundance can be explained by many factors, including polysaccharide metabolism (p-value<0.001) (Figure 3.15). This may be of importance in biocrust formation. Polysaccharides are crucial components of the EPS of biofilms and biocrusts (Flemming *et al.*, 2016; Rossi *et al*., 2018). As an example, bacterial cellulose is an important factor of the exopolymeric matrix in biofilms (Ude *et al*., 2006). Functions related to monosaccharides and di- and oligosaccharides are also enriched in topcrust (p-value<0.001 in both cases). The CO_2 fixation subsystem, which includes genes related to photosynthesis, is also more abundant in topcrust than subsoil (p-value 0.002).

Five other level 1 subsystems are enriched in topcrust compared to subsoil (Table 3.8). In the subsystem cofactors, vitamins, prosthetic groups and pigments, the difference is driven by tetrapyrroles (p-value 0.043) (Figure 3.16). This subsystem includes heme and siroheme biosynthesis, coenzyme B12 biosynthesis and chlorophyll biosynthesis.

In the subsystem motility and chemotaxis the difference in abundance can be explained by bacterial chemotaxis (p-value 0.004) and flagellar motility (p-value 0.005) (Figure 3.17). Bacterial chemotaxis is the movement towards environments that contain higher concentrations of beneficial, or lower concentrations of toxic, chemicals and is often driven by flagella (Wadhams & Armitage, 2004). Chemotaxis is important in biofilm formation and development (Hölscher *et al*., 2015; Wadhams & Armitage, 2004) and surface attached bacteria respond strongly to chemoattractants (Oliveira *et al*., 2016).

Some of the above mentioned subsystems enriched in topcrust are related to photosynthesis, i.e. CO_2 fixation and chlorophyll biosynthesis. All level 2 photosynthesis subsystems are also significantly enriched in topcrust (Figure 3.18), especially electron transport and light harvesting complexes (p-values<0.001) which are in very low abundance in subsoil. This is consistent with taxonomic analysis in which photosynthetic taxa, such as Cyanobacteria are more abundant in topcrust than subsoil.

Potassium metabolism is enriched in topcrust (*p*-value 0.02) and the difference is driven by potassium homeostasis (*p*-value 0.022) (Figure 3.19). The subsystem includes potassium uptake, efflux and transport systems. Sulfur metabolism is also enriched in topcrust (*p*-value 0.03) and the difference can be explained by the subsystem galactosylceramide and sulfatide metabolism (*p*-value<0.001). Sulfur oxidation is however enriched in subsoil (*p*-value 0.006) (Figure 3.20). The difference in abundance for both potassium and sulfur metabolism could be explained by potentially higher nutrient cycling rates closer to the soil surface as the surface receives more plant litter than the subsurface. In a biocrust in the Mojave desert Steven *et al*. (2014) also found potassium metabolism to be more abundant in biocrust than in root zones. In the Mojave desert potassium concentrations of root zones were higher than in biocrust areas but no potassium data is available for this study. Fierer *et al*. (2012) found higher abundances of genes associated with both potassium and sulfur metabolism in nondesert soils compared to desert soils and explained the difference by lower nutrient cycling in the desert soils.

5 Conclusions

The *Anthelia* biocrust is widespread in Iceland, especially in the highlands, and the biology and ecological functions of the biocrust are presumably important for the ecosystems of the highlands. This study shows a clear difference between the biocrust and subsoil for both taxonomic and functional gene analyses indicating functional differences between the two soil strata. This is most evident for photosynthetic taxa, e.g. Cyanobacteria, and EPS producing taxa, e.g. Acidobacteria which are enriched in topcrust as well as for various photosynthetic pathways and related genes.

This study found the microbial community of the biocrust to be stable across seasons and uniform between EUNIS habitat types and sample areas.

Despite the presumed importance of the biocrust there have been few studies that focus on this micro ecosystem. This study is the first to assess the bacterial composition and function of *Anthelia* biocrust and can serve as a reference and baseline of the biocrust's bacterial community and function for further studies to build on.

References

Abed, R. M. M., Tamm, A., Hassenrück, C., Al-Rawahi, A. N., Rodríguez-Caballero, E., Fiedler, S., … Weber, B. (2019). Habitat-dependent composition of bacterial and fungal communities in biological soil crusts from Oman. *Scientific Reports*, 9(6468). doi:10.1038/s41598-019-42911-6

Avontuur, J. R., Palmer, M., Beukes, C. W., Chan, W. Y., Coetzee, M. P. A., Blom, J., … Steenkamp, E. T. (2019). Genome-informed *Bradyrhizobium* taxonomy: where to from here? *Systematic and Applied Microbiology*, 42(4), 427–439. doi:10.1016/j.syapm.2019.03.006

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

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., … van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and Molecular Biology Reviews*, 80(4), iii–iii. doi:10.1128/mmbr.00044-16

Belnap, J. (2002). Nitrogen fixation in biological soil crusts from southeast Utah, USA. *Biology and Fertility of Soils*, 35(2), 128–135. doi:10.1007/s00374-002-0452-x

Bjarnason, Á.H. (2018). *Mosar á Íslandi*. Ágúst H. Bjarnason.

Blay, E. S., Schwabedissen, S. G., Magnuson, T. S., Aho, K. A., Sheridan, P. P., & Lohse, K. A. (2017). Variation in Biological Soil Crust Bacterial Abundance and Diversity as a Function of Climate in Cold Steppe Ecosystems in the Intermountain West, USA. *Microbial Ecology*, 74(3), 691–700. doi:10.1007/s00248-017-0981-3

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4), 1125–1139. doi:10.1093/bib/bbx120

Büdel, B., Darienko, T., Deutschewitz, K., Dojani, S., Friedl, T., Mohr, K. I., … Weber, B. (2009). Southern african biological soil crusts are ubiquitous and highly diverse in drylands, being restricted by rainfall frequency. *Microbial Ecology*, 57(2), 229–247. doi:10.1007/s00248-008-9449-9

Casaburi, G., Duscher, A. A., Reid, R. P., & Foster, J. S. (2016). Characterization of the stromatolite microbiome from Little Darby Island, The Bahamas using predictive and whole shotgun metagenomic analysis. *Environmental Microbiology*, 18(5), 1452–1469. doi:10.1111/1462-2920.13094

Cavaletti, L., Monciardini, P., Bamonte, R., Schumann, P., Ronde, M., Sosio, M., & Donadio, S. (2006). New lineage of filamentous, spore-forming, gram-positive bacteria from soil. *Applied and Environmental Microbiology*, 72(6), 4360–4369. doi:10.1128/AEM.00132-06

Chang, Y. juan, Land, M., Hauser, L., Chertkov, O., del Rio, T. G., Nolan, M., … Lapidus, A. (2011). Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21 T). *Standards in Genomic Sciences*, 5(1), 97–111. doi:10.4056/sigs.2114901

Chilton, A. M., Neilan, B. A., & Eldridge, D. J. (2018). Biocrust morphology is linked to marked differences in microbial community composition. *Plant and Soil*, 429(1–2), 65–75. doi:10.1007/s11104-017-3442-3

Couradeau, E., Giraldo-Silva, A., De Martini, F., & Garcia-Pichel, F. (2019). Spatial segregation of the biological soil crust microbiome around its foundational cyanobacterium, *Microcoleus vaginatus*, and the formation of a nitrogen-fixing cyanosphere. *Microbiome*, 7(1), 1–12. doi:10.1186/s40168-019-0661-2

Csotonyi, J. T., Swiderski, J., Stackebrandt, E., & Yurkov, V. (2010). A new environment for aerobic anoxygenic phototrophic bacteria: biological soil crusts. *Environmental Microbiology Reports*, 2(5), 651–656. doi:10.1111/j.1758-2229.2010.00151.x

Davis, K. E. R., Joseph, S. J., & Janssen, Peter, H. (2005). Effects of Growth Medium, Inoculum Size, and Incubation Time on Culturability and Isolation of Soil Bacteria. *Applied and Environmental Microbiology*, 71(2), 826–834. doi:10.1128/AEM.71.2.826

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-gonzález, A., Eldridge, D. J., Bardgett, R. D., … Fierer, N. (2018). A Global Atlas of the Dominant Bacteria Found in Soil. *Science*. doi:10.1126/science.aap9516

Eldridge, D. J., & Delgado-Baquerizo, M. (2019). The influence of climatic legacies on the distribution of dryland biocrust communities. *Global Change Biology*, 25(1), 327–336. doi:10.1111/gcb.14506

Elliott, D. R., Thomas, A. D., Hoon, S. R., & Sen, R. (2014). Niche partitioning of bacterial communities in biological crusts and soils under grasses, shrubs and trees in the Kalahari. *Biodiversity and Conservation*, 23(7), 1709–1733. doi:10.1007/s10531-014-0684-8

Erlacher, A., Cernava, T., Cardinale, M., Soh, J., Sensen, C. W., Grube, M., & Berg, G. (2015). Rhizobiales as functional and endosymbiontic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Frontiers in Microbiology*, 6(2), 1–9. doi:10.3389/fmicb.2015.00053

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., … Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(52), 21390–21395. doi:10.1073/pnas.1215210110

Fischer, D., Pfitzner, B., Schmid, M., Simões-Araújo, J. L., Reis, V. M., Pereira, W., … Hartmann, A. (2012). Molecular characterisation of the diazotrophic bacterial community in uninoculated and inoculated field-grown sugarcane (*Saccharum* sp.). *Plant and Soil*, 356(1–2), 83–99. doi:10.1007/s11104-011-0812-0

Flemming, H. C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., & Kjelleberg, S. (2016). Biofilms: An emergent form of bacterial life. *Nature Reviews Microbiology*, 14(9), 563–575. doi:10.1038/nrmicro.2016.94

Guðmundsdottir, P. L. and Andresson, O. S. (2019). Fungi in liverwort-based biocrust. *Icel. Agric. Sci.*, 32, 43-60.

Hölscher, T., Bartels, B., Yu-Cheng, L., Gallegos-Monterrosa, R., Price-Whelan, A., Kolter, R., … Kovács, Á. T. (2015). Motility , Chemotaxis and Aerotaxis Contribute to Competitiveness during Bacterial Pellicle Biofilm Development. *Journal of Molecular Biology*, 427(23), 3695–3708. doi:10.1016/j.jmb.2015.06.014

Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A., & Janssen, P. H. (2003). Laboratory Cultivation of Widespread and Previously Uncultured Soil Bacteria. *Applied and Environmental Microbiology*, 69(12), 7210–7215. doi:10.1128/AEM.69.12.7210-7215.2003

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., & Kuramae, E. E. (2016). The ecology of Acidobacteria: Moving beyond genes and genomes. *Frontiers in Microbiology*, 7(5), 1–16. doi:10.3389/fmicb.2016.00744

Kielak, A. M., Castellane, T. C. L., Campanharo, J. C., Colnago, L. A., Costa, O. Y. A., Da Silva, M. L. C., … Kuramae, E. E. (2017). Characterization of novel Acidobacteria exopolysaccharides with potential industrial and ecological applications. *Scientific Reports*, 7, 1–11. doi:10.1038/srep41193

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

King, C. E., & King, G. M. (2014). Description of *Thermogemmatispora carboxidivorans* sp. nov., a carbon-monoxideoxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 64(4), 1244–1251. doi:10.1099/ijs.0.059675-0

Ludwig, R. A. (1984). *Rhizobium* free-living nitrogen fixation occurs in specialized nongrowing cells. *Isotopenpraxis*, 20(1), 1566–1569. doi:10.1073/pnas.81.5.1566

Madigan, M. T., Martinko, J. M., Bender, K. S., Buckley, D. H., & Stahl, D. A. (2006). *Brock Biology of Microorganisms* (Eleventh edition.) Pearson, Upper Saddle River. New Jersey.

Mahendra, S., & Alvarez-Cohen, L. (2005). *Pseudonocardia dioxanivorans* sp. nov., a novel actinomycete that grows on 1,4-dioxane. *International Journal of Systematic and Evolutionary Microbiology*, 55(2), 593–598. doi:10.1099/ijs.0.63085-0

Maier, S., Schmidt, T. S. B., Zheng, L., Peer, T., Wagner, V., & Grube, M. (2014). Analyses of dryland biological soil crusts highlight lichens as an important regulator of microbial communities. *Biodiversity and Conservation*, 23(7), 1735–1755. doi:10.1007/s10531-014-0719-1

Malard, L. A., Anwar, M. Z., Jacobsen, C. S., & Pearce, D. A. (2019). Biogeographical patterns in soil bacterial communities across the Arctic region. *FEMS Microbiology Ecology*, 95, 1-13. doi:10.1101/655431

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7, 1–9. doi:10.1038/ncomms11257

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., … Edwards, R. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9(1), 386. doi:10.1186/1471-2105-9-386

Mogul, R., Vaishampayan, P., Bashir, M., McKay, C. P., Schubert, K., Bornaccorsi, R., … Wilhelm, M. B. (2017). Microbial community and biochemical dynamics of biological soil crusts across a gradient of surface coverage in the central Mojave Desert. *Frontiers in Microbiology*, 8(10), 1–19. doi:10.3389/fmicb.2017.01974

Monciardini, P., Cavaletti, L., Schumann, P., Rohde, M., & Donadio, S. (2003). *Conexibacter woesei* gen. nov., sp. nov., a novel representative of a deep evolutionary line of descent within the class Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 53(2), 569–576. doi:10.1099/ijs.0.02400-0

Moreira-Grez, B., Tam, K., Cross, A. T., Yong, J. W. H., Kumaresan, D., Nevill, P., … Whiteley, A. S. (2019). The Bacterial Microbiome Associated With Arid Biocrusts and the Biogeochemical Influence of Biocrusts Upon the Underlying Soil. *Frontiers in Microbiology*, 10(9), 1–13. doi:10.3389/fmicb.2019.02143

Mugnai, G., Ventura, S., Mascalchi, C., Rossi, F., Adessi, A., De Philippis, R. (2015). Biological soil crusts from Arctic environments: characterization of the prokaryotic community and exopolysaccharidic matrix analysis. *Geophysical Research Abstracts*, Vol. 17. EGU2015-6389, European Geosciences Union, Vienna, 12-17 April 2015.

Oliveira, N. M., Foster, K. R., & Durham, W. M. (2016). Single-cell twitching chemotaxis in developing biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6532–6537. doi:10.1073/pnas.1600760113

Ottosson, J. G., Sveinsdottir, A., & Hardardottir, M. (Eds.) (2016). *Vistgerðir á Íslandi (Habitats in Iceland)*. In: Vol. 54: Icelandic Institure of Natural History.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., … Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17), 5691–5702. doi:10.1093/nar/gki866

R-Core-Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rajeev, L., Da Rocha, U. N., Klitgord, N., Luning, E. G., Fortney, J., Axen, S. D., … Mukhopadhyay, A. (2013). Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *ISME Journal*, 7(11), 2178–2191. doi:10.1038/ismej.2013.83

Raven, P.H., Evert, R.F. and Eichhorn, S.E. (1992). *Biology of Plants* (Fifth edition). Worth publishers. New York.

Rodríguez-Caballero, E., Castro, A. J., Chamizo, S., Quintas-Soriano, C., Garcia-Llorente, M., Cantón, Y., & Weber, B. (2018). Ecosystem services provided by biocrusts: From ecosystem functions to social values. *Journal of Arid Environments*, 159, 45–53. doi:10.1016/j.jaridenv.2017.09.005

Rossi, F., & De Philippis, R. (2015). Role of cyanobacterial exopolysaccharides in phototrophic biofilms and in complex microbial mats. *Life*, 5(2), 1218–1238. doi:10.3390/life5021218

Rossi, F., Mugnai, G., & De Philippis, R. (2018). Complex role of the polymeric matrix in biological soil crusts. *Plant and Soil*, 429(1–2), 19–34. doi:10.1007/s11104-017-3441-4

RStudio. (2018). RStudio: Integrated development environment for R (Version 1.1.423). Retrieved from http://www.rstudio.org/

Steven, B., Gallegos-Graves, L. V., Belnap, J., & Kuske, C. R. (2013a). Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiology Ecology*, 86(1), 101–113. doi:10.1111/1574-6941.12143

Steven, B., Lionard, M., Kuske, C. R., & Vincent, W. F. (2013b). High Bacterial Diversity of Biological Soil Crusts in Water Tracks over Permafrost in the High Arctic Polar Desert. PLoS ONE, 8(8). doi:10.1371/journal.pone.0071489

Steven, B., Gallegos-Graves, L. V., Yeager, C., Belnap, J., & Kuske, C. R. (2014). Common and distinguishing features of the bacterial and fungal communities in biological soil crusts and shrub root zone soils. *Soil Biology and Biochemistry*, 69, 302–312. doi:10.1016/j.soilbio.2013.11.008

Steven, B., Kuske, C. R., Gallegos-graves, L. V., Reed, S. C., & Belnap, J. (2015). Climate change and physical disturbance manipulations result in distinct biological soil crust communities. *Appl. Environ. Microbiol*, 81(21), 7448–7459. doi:10.1128/AEM.01443-15

Ude, S., Arnold, D. L., Moon, C. D., Timms-Wilson, T., & Spiers, A. J. (2006). Biofilm formation and cellulose expression among diverse environmental *Pseudomonas* isolates. *Environmental Microbiology*, 8(11), 1997–2011. doi:10.1111/j.1462-2920.2006.01080.x

Wadhams, G. H., & Armitage, J. P. (2004). Making sense of it all: Bacterial chemotaxis. *Nat Rev Mol Cell Biol*, 5(12), 1024–1037. doi:10.1038/nrm1524

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., … Kuske, C. R. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology*, 75(7), 2046–2056. doi:10.1128/AEM.02294-08

Weber, B., Budel, B. & Belnap, J. (Eds.). (2016). *Biological soil crusts: an organizing principle in drylands*. New York, NY: Springer Berlin Heidelberg.

Whang, K. S., Lee, J. C., Lee, H. R., Han, S. I., & Chung, S. H. (2014). *Terriglobus tenax* sp. nov., an exopolysaccharide producing acidobacterium isolated from rhizosphere soil of a medicinal plant. *International Journal of Systematic and Evolutionary Microbiology*, 64(2), 431–437. doi:10.1099/ijs.0.053769-0

White, D. C., Suttont, S. D., & Ringelberg, D. B. (1996). The genus *Sphingomonas*: Physiology and ecology. *Current Opinion in Biotechnology*, 7(3), 301–306. doi:10.1016/S0958-1669(96)80034-6

Whitton, B.A. (Eds.). (2013). *Ecology of Cyanobacteria II. Their Diversity in Space and Time*. Springer.

Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M., & Yokota, A. (2011). *Thermogemmatispora onikobensis* gen. nov., sp. nov. and *Thermogemmatispora foliorum* sp. nov., isolated from fallen leaves on geothermal soils, and description of Thermogemmatisporaceae fam. nov. and Thermogemmatisporales ord. nov. within the class Ktedonobacteria. *International Journal of Systematic and Evolutionary Microbiology*, 61(4), 903–910. doi:10.1099/ijs.0.024877-0

Zheng, Y., Wang, C. M., Sakai, Y., Abe, K., Yokota, A., & Yabe, S. (2019). *Thermogemmatispora aurantia* sp. nov. and *Thermogemmatispora argillosa* sp. nov., within the class Ktedonobacteria, and emended description of the genus *Thermogemmatispora*. *International Journal of Systematic and Evolutionary Microbiology*, 69(6), 1744–1750. doi:10.1099/ijsem.0.003388

Appendix

PCA loading scores

Table 1: Genera with the highest loading scores for PC1.

Genus (Phylum)	**PC1 loading score**
Brevundimonas (Alphaproteobacteria)	-0.1161737
Sphingobium (Alphaproteobacteria)	-0.1146062
Acetobacter (Alphaproteobacteria)	-0.1143166
Candidatus *Microthrix* (Actinobacteria)	0.1139766
Gaiella (Actinobacteria)	0.1134140
Caulobacter (Alphaproteobacteria)	-0.1131592
Methyloceanibacter (Alpharoteobacteria)	0.1124244
Acidisphaera (Alphaproteobacteria)	-0.1124165
Granulibacter (Alphaproteobacteria)	-0.1119385
Anabaena (Cyanobacteria)	-0.1107701
Novosphingobium (Alphaproteobacteria)	-0.1107012
Acidocella (Alphaproteobacteria)	-0.1102146
Azospirillum (Alphaproteobacteria)	-0.1098087
Silvibacterium (Acidobacteria)	-0.1095176
Acidiphilium (Alphaproteobacteria)	-0.1090209

Table 2: Genera with the highest loading scores for PC2.

Genus (Phylum)	**PC2 loading score**
Corynebacterium (Actinobacteria)	0.064822011
Microbacterium (Actinobacteria)	0.055438841
Nitrobacter (Alphaproteobacteria)	0.054112866
Streptomyces (Actinobacteria)	0.052644736
Blastococcus (Actinobacteria)	0.052626134
Geodermatophilus (Actinobacteria)	0.045449646
Afipia (Alphaprotebacteria)	0.043203336
Nocardia (Actinobacteria)	0.041782012
Opitutus (Verrucomicrobia)	-0.029655182
Gemmata (Planctomycetes)	0.024397886
Gordonia (Actinobacteria)	0.023974844
Rhodopseudomonas (Alphaproteobacteria)	0.020148931
Schlesneria (Planctomycetes)	0.016401950
Streptacidiphilus (Actinobacteria)	0.016086529
Bradyrhizobium (Alphaproteobacteria)	0.005489254

Table 3: Level 3 SEED subsystems with the highest loading scores for PC1.

Level 3 SEED subsystem	PC1 loading score
Denitrification	-0.05673434
Photosystem_II-type_photosynthetic_reaction_center	0.05606829
Glutamine_Glutamate_Aspartate_and_Asparagine_Biosynthesis	-0.05525813
L-fucose_utilization_temp	0.05517832
Flavohaemoglobin	-0.05493028
Oxygen_and_light_sensor_PpaA.PpsR	0.05487880
Synechocystis_experimental	0.05481561
DMSP_breakdown	-0.05468645
Molybdopterin_cytosine_dinucleotide	0.05447546
Streptococcus_agalactiae_virulome	-0.05427298
CO_Dehydrogenase	-0.05393795
Cobalamin_synthesis	0.05392088
Carotenoids	0.05375957
Vir.like_type_4_secretion_system	0.05375224
Succinate_dehydrogenase	-0.05370159

Table 4: Level 3 SEED subsystems with the highest loading scores for PC2.

Level 3 SEED subsystem	PC2 loading score
Biotin_biosynthesis_Experimental	-0.0262091059
Glycolysis_and_Gluconeogenesis	-0.0198247623
Glutathione.regulated_potassium.efflux_system_and_associated_functions	0.0189343946
Soluble_cytochromes_and_functionally_related_electron_carriers	0.0152895668
Transport_of_Iron	-0.0140935698
Photorespiration_oxidative_C2_cycle.	-0.0130339465
GABA_and_putrescine_metabolism_from_cluters	-0.0130139128
USS.DB.6	0.0120322526
NiFe_hydrogenase_maturation	-0.0116573006
Proteolysis_in_bacteria._ATP.dependent	0.0109708887
EC49.61	0.0062489569
Biogenesis_of_cbb3.type_cytochrome_c_oxidases	-0.0061332106
Naphtalene_and_antracene_degradation	-0.0055860371
Competence_or_DNA_damage.inducible_protein_CinA	-0.0035232768
Phosphoenolpyruvate_phosphomutase	0.0008401544

Welch two sample t-tests

Table 5: Detailed Welch two sample t-test output for DNA recovery with depth as factor.

	t-value	**df**	**Topcrust mean (ng/ul)**	**Subsoil mean (ng/ul)**	**95% CI**	***p*-value**
DNA recovery	-2.71	8.8	28.3	18.5	-18.1 - 1.6	0.024

Table 6: Detailed Welch two sample t-test output for total bacterial reads classified at phylum/class level with depth as factor.

Phylum/class	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Acidobacteria	-1.09	7.62	0.162	0.149	-0.04 - 0.02	0.31
Actinobacteria	1.53	4.97	0.184	0.214	-0.02 - 0.08	0.19
Armatimonadetes	-5.73	13.98	0.007	0.004	-0.004 - -0.002	<0.001
Bacteroidetes	-1.60	7.70	0.041	0.034	-0.02 - 0.003	0.15
Chloroflexi	-3.66	12.18	0.075	0.055	-0.03 - -0.008	0.003
Cyanobacteria	-7.91	11.26	0.041	0.016	-0.03 - -0.02	<0.001
Deinococcus-Thermus	-1.92	13.0	0.003	0.003	-6.9e-04 - 3.7e-05	0.07
Firmicutes	-2.50	14.0	0.036	0.032	-0.008 - -0.0006	0.03
Gemmatimonadetes	5.65	3.66	0.005	0.011	0.003 - 0.01	0.006
Nitrospirae	2.37	3.12	0.003	0.006	-0.0007 - 0.005	0.096
Planctomycetes	1.45	13.78	0.049	0.055	-0.003 - 0.02	0.17
Alphaproteobacteria	-1.68	5.3	0.189	0.171	-0.05 - 0.009	0.15
Betaproteobacteria	4.44	4.19	0.051	0.063	0.004 - 0.02	0.01
Gammaproteobacteria	-1.11	7.4	0.043	0.041	-0.006 - 0.002	0.3
Deltaproteobacteria	4.15	4.78	0.029	0.039	0.004 - 0.02	0.0098
Verrucomicrobia	0.56	12.42	0.021	0.022	-0.003 - 0.004	0.58
Other	7.79	10.34	0.025	0.033	0.006 - 0.01	<0.001
Unclassified Bacteria	4.89	4.018	0.036	0.053	0.007 - 0.03	0.008

Table 7: Detailed Welch two sample t-test output for dominant bacteria classified at phylum/class level with depth as factor.

Phylum/class	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Acidobacteria	-3.26	7.09	0.192	0.133	-0.1 - -0.02	0.014
Actinobacteria	2.29	4.67	0.245	0.306	-0.009 - 0.1	0.07
Armatimonadetes	-12.94	11	0.006	0	-0.007 - -0.005	<0.001
Bacteroidetes	-4.05	13.39	0.022	0.013	-0.01 - -0.004	0.001
Chloroflexi	-5.35	12.18	0.111	0.049	-0.09 - -0.04	<0.001
Cyanobacteria	-12.56	11.05	0.027	0.003	-0.03 - -0.02	<0.001
Deinococcus-Thermus	-0.39	13.93	0.004	0.004	-0.0007 - 0.0005	0.7
Firmicutes	0.004	9.1	0.025	0.025	-0.004 - 0.004	0.99
Gemmatimonadetes	9.45	3	0	0.014	0.009 - 0.02	0.003
Nitrospirae	3.47	3	0	0.006	0.0005 - 0.01	0.04
Planctomycetes	4.05	7.7	0.065	0.098	0.01 - 0.05	0.004
Alphaproteobacteria	0.45	5.41	0.232	0.24	-0.03 - 0.05	0.67
Betaproteobacteria	1.88	3.25	0.032	0.04	-0.005 - 0.02	0.15
Gammaproteobacteria	-0.30	12.38	0.014	0.013	-0.005 - 0.004	0.77
Deltaproteobacteria	16.97	3.59	0.008	0.028	0.02 - 0.02	<0.001
Verrucomicrobia	3.069	13.8	0.015	0.022	0.002 - 0.01	0.008
Other	6.46	3	0	0.005	0.002 - 0.007	0.008

Table 8: Detailed Welch two sample t-test output for the most abundant genera with depth as factor.

Genus	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Acidobacterium	-7.39	13.38	0.006	0.002	-0.005 - -0.003	<0.001
Bradyrhizobium	0.65	5.94	0.039	0.042	-0.009 - 0.02	0.54
Candidatus *Solibacter*	-0.88	6.19	0.034	0.029	-0.02 - 0.01	0.41
Chthoniobacter	-2.23	10.44	0.005	0.003	-4.1e-03 - -1.4e-05	0.048
Conexibacter	0.24	4.67	0.007	0.007	-0.003 - 0.004	0.82
Frankia	-0.14	4.91	0.007	0.006	-0.002 - 0.002	0.89
Gemmata	1.28	13.99	0.006	0.007	-0.0006 - 0.002	0.22
Granulicella	-6.87	11.96	0.012	0.002	-0.01 - -0.007	<0.001
Ktedonobacter	-5.72	12.97	0.04	0.015	-0.03 - -0.02	<0.001
Methylobacterium	-6.74	13.99	0.005	0.003	-0.003 - -0.001	<0.001
Mycobacterium	-3.74	13.69	0.023	0.015	-0.01 - -0.003	0.002
Pseudomonas	-1.33	12.55	0.006	0.005	-0.003 - 0.0008	0.21
Pseudonocardia	1.22	4.43	0.005	0.007	-0.002 - 0.006	0.28
Silvibacterium	-7.64	13.34	0.005	0.002	-0.004 - -0.002	<0.001
Singulisphaera	0.34	10.04	0.012	0.012	-0.002 - 0.003	0.74
Solirubrobacter	0.8	4.96	0.006	0.007	-0.002 - 0.004	0.46
Sphingomonas	-12.4	13.9	0.009	0.004	-0.007 - -0.005	<0.001
Streptomyces	0.82	4.35	0.019	0.02	-0.003 - 0.006	0.45
Terriglobus	-7.16	12.77	0.006	0.002	-0.006 - -0.003	<0.001
Thermogemmatispora	-6.09	12.48	0.008	0.003	-0.007 - -0.003	<0.001

Table 9: Detailed Welch two sample t-test output for the most abundant genera in spring and fall at Gagnheiði with season as factor.

Genus	**t-value**	**df**	**Spring mean**	**Fall mean**	**95% CI**	***p*-value**
Acidobacterium	-1.0	3.8	0.005	0.005	-0.004 - 0.002	0.38
Bradyrhizobium	0.005	2.89	0.032	0.037	-0.01 - 0.02	0.46
Candidatus *Solibacter*	-0.64	3.24	0.044	0.036	-0.05 - 0.03	0.56
Chthoniobacter	0.35	2.28	0.005	0.005	-0.003 - 0.004	0.76
Conexibacter	0.23	3.47	0.007	0.007	-0.008 - 0.009	0.83
Frankia	-0.01	2.44	0.006	0.006	-0.004 - 0.004	0.99
Gemmata	0.73	2.38	0.006	0.006	-0.003 - 0.005	0.53
Granulicella	-2.64	3.03	0.009	0.007	-0.004 - 0.0004	0.077
Ktedonobacter	0.026	3.18	0.044	0.044	-0.02 - 0.02	0.98
Methylobacterium	1.21	3.88	0.004	0.005	-0.001 - 0.003	0.3
Mycobacterium	0.017	3.27	0.02	0.02	-0.01 - 0.02	0.99
Pseudomonas	2.15	3.32	0.005	0.005	-3.1e-05 - 1.8e-04	0.11
Pseudonocardia	0.28	2.3	0.005	0.005	-0.003 - 0.004	0.8
Silvibacterium	-0.99	3.57	0.004	0.004	-0.002 - 0.001	0.39
Singulisphaera	2.10	2.29	0.01	0.012	-0.002 - 0.007	0.15
Solirubrobacter	0.26	3.47	0.006	0.007	-0.006 - 0.008	0.81
Sphingomonas	3.08	4.0	0.008	0.01	0.0008 - 0.003	0.04
Streptomyces	0.28	2.7	0.018	0.019	-0.01 - 0.01	0.8
Terriglobus	-2.58	3.41	0.005	0.004	-0.002 - 0.0002	0.07
Thermogemmatispora	-0.02	2.9	0.009	0.009	-0.004 - 0.004	0.99

Table 10: Detailed Welch two sample t-test output for level 1 SEED subsystems with depth as factor.

Level 1 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Amino acids and derivatives	7.11	6.81	0.096	0.1	0.003 - 0.005	<0.001
Carbohydrates	-6.89	7.88	0.149	0.142	-0.009 - -0.005	<0.001
Cell division and cell cycle	-2.32	6.55	0.01	0.009	-5.2e-04 - 8.5e-06	0.06
Cell wall and capsule	0.29	6.14	0.035	0.035	-0.0007 - 0.001	0.78
Clustering based subsystems	7.04	5.32	0.122	0.125	0.002 - 0.004	<0.001
Cofactors, vitamins, prosthetic groups and pigments	-9.56	13.8	0.054	0.052	-0.003 - -0.002	<0.001
DNA metabolism	1.19	6.29	0.044	0.044	-0.0006 - 0.002	0.28
Dormancy and sporulation	-0.88	13.96	0.001	0.001	-1.0e-04 - 4.2e-05	0.39
Fatty acids, lipids and isoprenoids	-2.12	4.1	0.027	0.027	-0.001 - 0.0002	0.1
Iron acquisition and metabolism	-1.27	8.75	0.006	0.005	-0.0007 - 0.0002	0.24
Membrane transport	3.99	4.52	0.038	0.041	0.0008 - 0.004	0.01
Metabolism of aromatic compounds	2.64	4.48	0.017	0.018	-9.9e-06 - 2.1e-03	0.051
Miscellaneous	-0.99	5.22	0.062	0.062	-0.001 - 0.0005	0.37
Motility and chemotaxis	-4.22	10.61	0.011	0.009	-0.002 - -0.0007	0.002
Nitrogen metabolism	2.22	5.44	0.012	0.013	-8.5e-05 - 1.4e-03	0.07
Nucleosides and nucleotides	5.1	4.21	0.03	0.031	0.0005 - 0.0017	0.006
Phages, prophages, transposable elements, plasmids	0.23	4.1	0.013	0.014	-0.0007 - 0.0008	0.83
Phosphorus metabolism	0.76	5.08	0.013	0.013	-0.0002 - 0.0004	0.48
Photosynthesis	-11.68	12.24	0.002	0.0008	-0.002 - -0.001	<0.001
Potassium metabolism	-3.58	4.28	0.009	0.008	-0.001 - -0.0002	0.02
Protein metabolism	2.38	9.3	0.08	0.082	0.0001 - 0.004	0.04
RNA metabolism	1.33	3.65	0.035	0.035	-0.0009 - 0.002	0.26
Regulation and cell signaling	-2.03	13.0	0.011	0.01	-8.6e-04 - 2.6e-05	0.06
Respiration	-0.11	8.55	0.047	0.047	-0.0008 - 0.0007	0.91
Secondary metabolism	-4.81	8.03	0.0031	0.0028	-0.0004 - 0.0001	0.001
Stress response	1.27	14.0	0.026	0.026	-0.0002 - 0.0006	0.22
Sulfur metabolism	-2.6	8.07	0.011	0.01	-1.18e-03 - -7.5e-05	0.03
Virulence, disease and defense	-0.36	8.16	0.036	0.035	-0.002 - 0.002	0.73

Table 11: Detailed Welch two sample t-test output for carbohydrates level 2 SEED subsystems with depth as factor.

Carbohydrates level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Aminosugars	-2.9	7.49	0.0023	0.0021	-4.4e-04 - -4.8e-05	0.022
Central carbohydrate metabolism	2.28	7.0	0.0438	0.045	-4.2e-05 - 2.2e-03	0.057
CO_2 fixation	-3.9	11.33	0.0086	0.0081	-0.0007 - -0.0002	0.002
Di- and oligosaccharides	-9.03	8.94	0.0178	0.015	-0.003 - -0.002	<0.001
Fermentation	-2.43	3.59	0.0121	0.0118	-8.0e-04 - 7.2e-05	0.08
Glycoside hydrolases	-1.0	6.83	0.0008	0.0007	-1.0e-04 - 4.2e-05	0.35
Monosaccharides	-9.12	7.58	0.0209	0.0177	-0.004 - -0.002	<0.001
One carbon metabolism	16.94	13.96	0.012	0.0131	0.0009 - 0.001	<0.001
Organic acids	0.93	5.98	0.0071	0.0072	-0.0002 - 0.0003	0.39
Polysaccharides	-5.16	13.34	0.0042	0.0039	-0.0005 - -0.0002	<0.001
Sugar alcohols	-0.67	4.54	0.0075	0.0073	-0.0007 - 0.0004	0.54
Various	-12.2	13.39	0.0119	0.0103	-0.002 - -0.001	<0.001

Table 12: Detailed Welch two sample t-test output for cofactors, vitamins, prosthetic groups and pigments level 2 SEED subsystems with depth as factor.

Cofactors level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Biotin	7.3	5.52	0.002	0.0023	0.0002 - 0.0004	<0.001
Coenzyme A	1.92	8.9	0.003	0.0031	-1.4e-05 - 1.7e-04	0.09
Coenzyme B	-1.98	13.63	1.9e-05	1.3e-05	-1.5e-05 - 6.0e-07	0.07
Coenzyme F420	5.03	13.24	0.001	0.0012	0.0001 - 0.0003	<0.001
Coenzyme M	-1.66	8.28	0.0002	0.0002	-6.2e-05 - 9.9e-06	0.13
Folate and pterines	0.2	14.0	0.0186	0.0186	-0.0002 - 0.0002	0.84
Lipoic acid	2.07	6.02	0.0006	0.0007	-1.0e-05 - 1.2e-04	0.08
NAD and NADP	-0.73	4.4	0.005	0.0049	-0.0003 - 0.0002	0.5
Pyridoxine	0.6	3.43	0.004	0.0036	-0.0003 - 0.0004	0.59
Quinone cofactors	-0.95	4.02	0.0041	0.004	-0.0004 - 0.0002	0.4
Riboflavin, FMN, FAD	1.03	3.05	0.0025	0.003	-0.0008 - 0.002	0.38
Tetrapyrroles	-3.27	3.15	0.01	0.006	-0.007 - -0.0002	0.04
Various	0.5	3.02	0.0039	0.0044	-0.003 - 0.004	0.65

Table 13: Detailed Welch two sample t-test output for motility and chemotaxis level 2 SEED subsystems with depth as factor.

Motility and chemotaxis level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Flagellar motility	-4.12	6.85	0.0063	0.0053	-0.002 - -0.0004	0.005
Nonflagellar swimming	0.21	4.69	4.1e-05	4.1e-05	-8.9e-06 - 1.0e-05	0.84
Bacterial chemotaxis	-3.52	13.69	0.0043	0.0037	-0.0009 - -0.0002	0.004

Table 14: Detailed Welch two sample t-test output for photosynthesis level 2 SEED subsystems with depth as factor.

Photosynthesis level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Electron transport	-11.74	11.08	1.1e-03	5.5e-05	-0.001 - -0.0009	<0.001
Light harvesting complexes	-8.72	11.02	1.9e-04	5.3e-06	-0.0002 - -0.0001	<0.001
Bacteriorhodopsin	-4.88	13.94	7.9e-06	3.1e-06	-6.9e-06 - -2.7e-06	<0.001
Proteorhodopsin	-5.16	7.30	0.0009	0.0008	-2.5e-04 - -9.5e-05	0.001

Table 15: Detailed Welch two sample t-test output for potassium metabolism level 2 SEED subsystems with depth as factor.

Potassium level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Potassium efflux	-1.0	12.9	0.0003	0.0002	-3.8e-05 - 1.4e-05	0.33
Potassium homeostasis	-3.45	4.49	0.0084	0.0077	-0.001 - -0.0002	0.02
Potassium uptake	-0.24	5.53	4.9e-06	4.4e-06	-5.9e-06 - 4.8e-06	0.82

Table 16: Detailed Welch two sample t-test output for sulfur metabolism level 2 SEED subsystems with depth as factor.

Sulfur level 2 subsystem	**t-value**	**df**	**Topcrust mean**	**Subsoil mean**	**95% CI**	***p*-value**
Inorganic sulfur assimilation	-2.1	12.02	0.0036	0.0036	-2.2e-04 - 4.0e-06	0.06
Organic sulfur assimilation	-2.12	5.82	0.0039	0.0036	-7.0e-04 - 5.3e-05	0.08
Galactosylceramide and sulfatide metabolism	-5.36	13.19	0.0015	0.0011	-0.0006 - -0.0003	<0.001
Sulfur oxidation	4.85	4.52	0.0013	0.0015	0.0001 - 0.0004	0.006
Thioredoxin-disulfide reductase	-1.1	4.35	0.0006	0.0005	-8.6e-05 - 3.6e-05	0.33
Rare	-0.52	7.63	3.2e-05	2.9e-05	-1.5e-05 - 9.7e-06	0.62

Table 17: Detailed ANOVA output for the most abundant genera in different habitat types with habitat type as factor. The habitat type Glacial moraines with very sparse or no vegetation (EUNIS H5.2) was excluded from the analysis because only one sample represents that habitat type. Tukey's honest significance test results for Chthoniobacter *and* Gemmata *are shown in Appendix tables 20 and 21.*

Genus	**df**	**F-value**	***p*-value**
Acidobacterium	2	1.4	0.3
Bradyrhizobium	2	0.27	0.77
Candidatus *Solibacter*	2	1.03	0.4
Chthoniobacter	2	19.73	<0.001
Conexibacter	2	0.32	0.74
Frankia	2	2.79	0.12
Gemmata	2	5.11	0.04
Granulicella	2	3.32	0.09
Ktedonobacter	2	0.67	0.54
Methylobacterium	2	1.31	0.32
Mycobacterium	2	1.39	0.3
Pseudomonas	2	0.41	0.68
Pseudonocardia	2	1.86	0.22
Silvibacterium	2	2.01	0.2
Singulisphaera	2	0.03	0.97
Solirubrobacter	2	0.32	0.73
Sphingomonas	2	0.43	0.66
Streptomyces	2	0.32	0.73
Terriglobus	2	4.09	0.06
Thermogemmatispora	2	0.55	0.59

Table 18: Detailed ANOVA output for the most abundant genera in sample areas with sample area as factor. Tukey's honest significance test results for Granulicella *and* Terriglobus *are shown in Appendix tables 22 and 23.*

Genus	**df**	**F-value**	**p-value**
Acidobacterium	3	2.54	0.13
Bradyrhizobium	3	1.23	0.36
Candidatus *Solibacter*	3	0.15	0.93
Chthoniobacter	3	2.98	0.1
Conexibacter	3	0.61	0.63
Frankia	3	2.09	0.18
Gemmata	3	3.33	0.08
Granulicella	3	6.81	0.014
Ktedonobacter	3	1.9	0.2
Methylobacterium	3	2.45	0.14
Mycobacterium	3	1.33	0.33
Pseudomonas	3	1.03	0.43
Pseudonocardia	3	1.25	0.35
Silvibacterium	3	2.95	0.098
Singulisphaera	3	0.09	0.97
Solirubrobacter	3	0.43	0.74
Sphingomonas	3	1.71	0.24
Streptomyces	3	0.1	0.96
Terriglobus	3	5.9	0.02
Thermogemmatispora	3	1.48	0.29

Paired t-tests

Table 19: Detailed paired t-test output for each site in spring and fall at Gagnheiði on genus level with season as factor.

Sample site	t-value	df	Mean of the differences	95% CI	*p*-value
G1	0.075	208	1.28e-05	-0.0003 - 0.0004	0.94
G2	0.068	208	2.69e-06	-7.5e-05 - 8.1e-05	0.95
G3	-0.94	208	-6.63e-05	-2.1e-04 - 7.3e-05	0.35

Tukey's honest significance tests

Table 20: Detailed Tukey's honest significance test output for the genus Chthoniobacter *with habitat type as factor. Snowbed = Boreal moss snowbed communities (EUNIS E4.115); Racomitrium = Icelandic Racomitrium ericoides heaths (EUNIS E4.26); Lava_field = Icelandic lava field lichen heaths (EUNIS E4.241).*

	diff	95% CI	*p*-value
Racomitrium vs lava field	-0.0058	-0.008 - -0.003	0.0006
Snowbed vs lava field	-0.0047	-0.008 - -0.002	0.005
Snowbed vs Racomitrium	0.0011	-0.001 - 0.003	0.37

Table 21: Detailed Tukey's honest significance test output for the genus Gemmata *with habitat type as factor. Snowbed = Boreal moss snowbed communities (EUNIS E4.115); Racomitrium = Icelandic Racomitrium ericoides heaths (EUNIS E4.26); Lava_field = Icelandic lava field lichen heaths (EUNIS E4.241).*

	diff	95% CI	*p*-value
Racomitrium vs lava field	-0.0041	-0.008 - -0.0004	0.03
Snowbed vs lava field	-0.0031	-0.007 - 0.001	0.15
Snowbed vs Racomitrium	0.0011	-0.002 - 0.004	0.62

Table 22: Detailed Tukey's honest significance test output for the genus Granulicella *with sample area as factor.*

	diff	95% CI	*p*-value
Gagnheiði vs Fjallabak	-0.0055	-0.01 - 0.003	0.22
Laki vs Fjallabak	-0.0022	-0.01 - 0.006	0.83
Skaftártunga vs Fjallabak	0.0058	-0.002 - 0.01	0.19
Laki vs Gagnheiði	0.0034	-0.005 - 0.01	0.58
Skaftártunga vs Gagnheiði	0.011	0.003 - 0.02	0.01
Skaftártunga vs Laki	0.008	-0.0003 - 0.02	0.06

Table 23: Detailed Tukey's honest significance test output for the genus Terriglobus *with sample area as factor.*

	diff	**95% CI**	***p*-value**
Gagnheiði vs Fjallabak	-0.0026	-0.006 - 0.0009	0.16
Laki vs Fjallabak	-0.00096	-0.004 - 0.003	0.82
Skaftártunga vs Fjallabak	0.0019	-0.002 - 0.005	0.37
Laki vs Gagnheiði	0.0016	-0.002 - 0.005	0.49
Skaftártunga vs Gagnheiði	0.0045	0.001 - 0.008	0.014
Skaftártunga vs Laki	0.0029	-0.0006 - 0.006	0.11