



## Biological Soil Crusts from Different Soil Substrates Harbor Distinct Bacterial Groups with the Potential to Produce Exopolysaccharides and Lipopolysaccharides

Cania, Barbara; Vestergaard, Gisle Alberg; Kublik, Susanne; Köhne, John Maximilian; Fischer, Thomas; Albert, Andreas; Winkler, Barbro; Schloter, Michael; Schulz, Stefanie

*Published in:*  
Microbial Ecology

*Link to article, DOI:*  
[10.1007/s00248-019-01415-6](https://doi.org/10.1007/s00248-019-01415-6)

*Publication date:*  
2020

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Cania, B., Vestergaard, G. A., Kublik, S., Köhne, J. M., Fischer, T., Albert, A., Winkler, B., Schloter, M., & Schulz, S. (2020). Biological Soil Crusts from Different Soil Substrates Harbor Distinct Bacterial Groups with the Potential to Produce Exopolysaccharides and Lipopolysaccharides. *Microbial Ecology*, 79, 326–341. <https://doi.org/10.1007/s00248-019-01415-6>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

[Click here to view linked References](#)

1 Biological soil crusts from different soil substrates harbor distinct bacterial groups with the potential to  
2 produce exopolysaccharides and lipopolysaccharides

3

4 revised version June 19, 2019

5

6 Barbara Cania<sup>1</sup>, Gisle Vestergaard<sup>1,2</sup>, Susanne Kublik<sup>1</sup>, John Maximilian Köhne<sup>3</sup>, Thomas Fischer<sup>4</sup>,  
7 Andreas Albert<sup>5</sup>, Barbro Winkler<sup>5</sup>, Michael Schloter<sup>1,6</sup>, Stefanie Schulz<sup>1</sup>

8

9 <sup>1</sup>Research Unit Comparative Microbiome Analysis, Helmholtz Zentrum München Research Center for  
10 Environmental Health (GmbH), Ingolstädter Landstraße 1, 85764 Neuherberg

11 <sup>2</sup>Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, 2800  
12 Lyngby, Denmark

13 <sup>3</sup>Department of Soil System Science, Helmholtz Centre for Environmental Research (UFZ), Theodor-  
14 Lieser-Straße 4, 06120 Halle, Germany

15 <sup>4</sup>Central Analytical Laboratory, Brandenburg Technical University, Konrad-Wachsmann-Allee 6, 03046  
16 Cottbus, Germany

17 <sup>5</sup>Research Unit Environmental Simulation, Helmholtz Zentrum München Research Center for  
18 Environmental Health (GmbH), Ingolstädter Landstraße 1, 85764 Neuherberg

19 <sup>6</sup>Chair for Soil Science, Technical University of Munich, Emil-Ramann-Straße 2; 85354 Freising, Germany

20

21 Corresponding author: Stefanie Schulz, stefanie.schulz@helmholtz-muenchen.de, +49 89 3187-3054

22

## 23 **Acknowledgements**

24 The authors wish to thank Gudrun Hufnagel for measuring the biochemical parameters, Christoph Schmidt  
25 and Abilash Chakravarthy Durai Raj for bioinformatical advice, as well as Viviane Radl and Antonios Michas  
26 for constructive feedback on the previous version of the manuscript. This study was performed as part of  
27 the Transregional Collaborative Research Centre 38 (SFB/TRR 38), which is financially supported by the  
28 Deutsche Forschungsgemeinschaft (DFG, Bonn) and the Brandenburg Ministry of Science, Research and  
29 Culture (MWFK, Potsdam), and the project “The influence of agricultural management practices on  
30 microbial functions and networks in biological soil crusts” funded by the DFG in frame of the DFG-  
31 Nachwuchsakademie “Agrarökosystemforschung: Bodenressourcen und Pflanzenproduktion”. The authors  
32 also gratefully acknowledge the funding provided by the German Federal Office for Agriculture and Food  
33 (BLE).

34

## 35 **Abstract**

36 Biological soil crusts (BSCs) play an important role in improving soil stability and resistance to erosion by  
37 promoting aggregation of soil particles. During initial development, BSCs are dominated by bacteria. Some  
38 bacterial members of the crusts can contribute to the formation of soil aggregates by producing  
39 exopolysaccharides and lipopolysaccharides that act as “glue” for soil particles. However, little is known  
40 about the dynamics of “soil glue” producers during the initial development of BSCs. We hypothesized that  
41 different types of initial BSCs harbor distinct producers of adhesive polysaccharides. To investigate this,  
42 we performed a microcosm experiment, cultivating BSCs on two soil substrates. High-throughput shotgun

43 sequencing was used to obtain metagenomic information on microbiomes of bulk soils from the beginning  
44 of the experiment, and BSCs sampled after four and ten months of incubation. We discovered that the  
45 relative abundance of genes involved in the biosynthesis of exopolysaccharides and lipopolysaccharides  
46 increased in BSCs compared to bulk soils. At the same time, communities of potential “soil glue” producers  
47 that were highly similar in bulk soils underwent differentiation once BSCs started to develop. In the bulk  
48 soils, the investigated genes were harbored mainly by *Betaproteobacteria*, whereas in the BSCs, the major  
49 potential producers of adhesive polysaccharides were, aside from *Alphaproteobacteria*, either  
50 *Cyanobacteria* or *Chloroflexi* and *Acidobacteria*. Overall, our results indicate that the potential to form  
51 exopolysaccharides and lipopolysaccharides is an important bacterial trait for initial BSCs, and is  
52 maintained despite the shifts in bacterial community composition during BSC development.

53

#### 54 **Keywords**

55 Biological soil crusts, exopolysaccharides, lipopolysaccharides, microbiome, metagenomics

56

#### 57 **Introduction**

58 Biological soil crusts (BSCs) are important biotic components of many terrestrial ecosystems [1, 2]. They  
59 consist of highly specialized and complex communities of algae, mosses, lichens, fungi, cyanobacteria and  
60 other prokaryotes [3]. These organisms live in a close association with soil particles, forming a coherent  
61 layer within the uppermost few millimeters of the topsoil, or directly on the soil surface [1]. An important  
62 structural element of BSCs is the extracellular polymeric matrix (EPM) which is composed mostly of  
63 polysaccharides, and connects organisms and soil particles [4]. EPM ensures BSC integrity, provides  
64 protection from external harmful agents, and alters moisture content as well as nutrient availability [4].

65 EPM also fosters the stabilization of soil aggregates, and protects soils from erosion by wind or water [5-  
66 8]. Among organisms forming BSCs, the best-studied producers of polysaccharides are cyanobacteria and  
67 algae [1]. However, although not as thoroughly studied, also non-photosynthetic microbial members of  
68 BSCs, including fungi, proteobacteria and actinobacteria, are prominent producers of these compounds  
69 [9-11].

70 The composition and chemical properties of polysaccharides in EPM strongly depend on the community  
71 of organisms forming BSCs. For example, it has been demonstrated that non-photosynthetic bacteria  
72 primarily produce simple polysaccharides, composed mainly of mannose, galactose and glucose [12],  
73 while cyanobacteria, algae and fungi produce more complex polysaccharides, which may contain high  
74 amounts of non-neutral sugars [13-15]. As it was shown that even slight differences in the sugar  
75 composition can result in completely different physical traits of the polysaccharide [16], the properties of  
76 EPM could be influenced by any factor that changes the structure of polysaccharide-producing  
77 communities. It is known that the composition of organisms forming BSCs changes depending on (i) the  
78 developmental stage of BSCs [17-19], (ii) environmental factors like radiation, humidity, elevation,  
79 temperature [17-24], and (iii) edaphic factors like soil pH, texture and nutrient content [1, 17, 20, 21].  
80 However, not all members of BSCs have the ability to produce polysaccharides, and little is known about  
81 the dynamics of polysaccharide-producing organisms during the development of different types of BSCs.

82 In this respect, bacterial polysaccharides, specifically exopolysaccharides (EPSs) and lipopolysaccharides  
83 (LPSs), are of great interest, as cyanobacteria and non-phototrophic bacteria form BSCs in the initial stage  
84 of BSC development [25]. EPSs are either synthesized intracellularly, and excreted by one of three  
85 different pathways: the Wzy-dependent pathway, the ABC transporter-dependent pathway and the  
86 synthase-dependent pathway, or synthesized extracellularly [26, 27]. In contrast, parts of LPSs are initially  
87 synthesized inside a cell, then ligated together at the inner membrane and transported to the cell surface  
88 as mature molecules [28, 29]. While LPSs are present in most Gram-negative bacteria [30], EPSs are

89 exuded by a wide range of taxa [16]. Among the most-recognized producers of EPSs are cyanobacterial  
90 members of *Oscillatoria*, *Nostoc*, *Lyngbya* and *Schizothrix*, as well as bacterial members of  
91 *Microbacterium*, *Pseudomonas*, *Bacillus*, *Paenibacillus* and *Streptomyces* [31]. These microorganisms are  
92 the first colonizers of bare soils, and their EPSs as well as LPSs are considered as essential for the initial  
93 consolidation of soil particles and the preparation of conditions for the establishment of cryptogamic  
94 surface cover in the later stages of BSC development [32]. Thus, a better understanding of the dynamics  
95 of polysaccharide-producing organisms during the initial development of BSCs requires more in-depth  
96 knowledge on cyanobacteria and other bacteria that initialize BSC establishment [4].

97 Many researchers studied polysaccharide-producing bacterial strains that were isolated from BSCs at  
98 different stages of development [33-36]. However, data on the community dynamics of bacterial EPS and  
99 LPS producers under natural conditions is missing. Thus, our aim was to investigate polysaccharide-  
100 producing bacterial communities during the initial stage of BSC development. We assumed that the  
101 relative abundance of genes related to EPS and LPS formation would increase once BSC development  
102 starts. Moreover, we hypothesized that different types of initial BSCs would harbor different communities  
103 of potential EPS and LPS producers. To test our hypotheses, we performed a microcosm experiment  
104 cultivating BSCs on two different soil substrates. As the soil substrates came from sites with different types  
105 of naturally occurring BSCs, we expected that the BSCs cultivated in the microcosm experiment would also  
106 consist of distinct microbial communities. To address our research questions, we used a high-throughput  
107 shotgun sequencing of DNA extracted from bulk soils from the beginning of the experiment, as well as  
108 initial BSCs sampled after four and ten months of incubation. We employed a bioinformatics pipeline  
109 described by Cania et al. [37] targeting genes specific for EPS and LPS production to obtain information on  
110 bacteria potentially involved in the production of adhesive polysaccharides.

## 112 **Materials and Methods**

### 113 *Sites description*

114 Soils for the incubation experiment were collected in 2011 from two sites at the initial stages of ecosystem  
115 development. One was the artificial catchment Chicken Creek (51°36'18" N, 14°15'58" E) and the other  
116 was an initial moving sand dune close to Lieberose (51°55'49" N, 14°22'22" E). Both sites are located in  
117 the state of Brandenburg in eastern Germany, approximately 37 km apart. The climate of the region is  
118 temperate continental with a mean air temperature of 8.9 °C and mean rainfall of 569 mm a<sup>-1</sup> [19].

119 The Chicken Creek catchment was constructed in 2005 in an opencast mine near Cottbus by dumping and  
120 contouring sand and loamy sand material originating from Pleistocene sediments. Details on the  
121 construction works and site conditions are given by Gerwin et al. [38] and Russell et al. [39]. After  
122 construction, no restoration was undertaken and the area was allowed to undergo natural succession.  
123 BSC development at the site was heterogeneous depending on the appearance of vascular vegetation,  
124 which still was dynamic at the time of sampling [40]. For the Lusatian post-mining sites, the cyanobacterial  
125 species *Microcoleus vaginatus*, *Nostoc* spec., *Phormidium* spec., *Schizothrix* spec., *Tolypothrix* spec., the  
126 green algal species *Bracteacoccus minor*, *Chlorococcum* spec., *Cylindrocystis* spec., *Elliptochloris* spec.,  
127 *Gloeocysis* spec., *Klebsormidium*, *Chlorella* spec., *Zygogonium* spec., *Ulothrix* spec., *Haematococcus* spec.,  
128 the lichens *Placynthiella oligotropha* and *Cladonia subulata*, as well as the mosses *Polytrichum piliferum*  
129 and *Ceratodon purpureus* were reported [18, 41, 42]. The Chicken Creek site heterogeneity was also  
130 reflected by high variability of moss coverages, which were recorded on 107 vegetation monitoring plots  
131 each having a size of 5 x 5 m<sup>2</sup>, which ranged from 0.1 to 95 % with a median coverage of 30 ±25 % [40].  
132 Terminal successional stages of cryptogamic surface cover development could not be identified, mainly  
133 due to BSC extinction caused by vascular plant overgrowth.

134 The moving sand dune occurs near Lieberose as a result of extensive disturbances of the land surface by  
135 former military activities (until approximately 1992). The dune is composed of Pleistocene aeolian sand.  
136 A detailed description of the site is provided by Dümig et al. [43], and Fischer and Veste [19]. Depending  
137 on their position downslope an inland dune catena, three stages of BSC development could be identified.  
138 In microdepressions and at the lee side of tussocks consisting of *Corynephorus canescens* located in the  
139 center of the dune slope, dominating sand grains were physically stabilized in their contact zones by  
140 accumulated organic matter and by few filamentous algae (BSC stage 1, surface coverage 20 % [44]). At  
141 surface patches of the lower dune slope, filamentous algae enmeshed the sand grains and partially filled  
142 in the soil pores (BSC stage 2, surface coverage 40 % [44]). BSC stage 3 was characterized by full cover  
143 with filamentous and coccoid algae, and by few mosses, the latter covering less than 5 % of the surface.  
144 The dominating green algal and moss species were *Zygonium ericetorum* and *Polytrichum piliferum*,  
145 respectively [45]. Cyanobacteria were a minor component within the *Zygonium* crust, which did not  
146 form individual patches, whereas lichens could not be observed at the sampling site. The terminal  
147 successional stage of cryptogamic surface cover development, which was found in the vicinity of a less  
148 disturbed neighboring Scots pine forest (distance to the sampling site of around 500 m), was characterized  
149 by co-appearance of *Cladonia* spec. and *P. piliferum*, which formed dense surface covers.

150

#### 151 *Sampling and incubation experiment*

152 Bulk soil from the Chicken Creek catchment and the Lieberose sand dune was used to establish a  
153 microcosm experiment. A total amount of 100 kg of soil was taken per site from the top 20 cm. At  
154 Lieberose, the soil was collected from five spots on the top of the dune, where no plants were growing,  
155 while at Chicken Creek, five plant-free spots were used for the soil sampling. The soil was transported and  
156 afterwards stored in the dark at room temperature for approximately six months before the incubation

157 experiment. During that time, pre-experiments to adjust the incubation conditions for BSC growth were  
158 performed.

159 The soil was mixed and passed through a 2 mm sieve, then packed into plastic pots (10 cm x 10 cm x 10  
160 cm) and compacted to the natural soil density of approximately  $1.6 \text{ g cm}^{-3}$  [43]. In total, the microcosm  
161 experiment consisted of 18 pots (9 per site). The water content was set to 50 % of the maximum water  
162 holding capacity of the soil samples, and adjusted weekly from the bottom, which ensured very low  
163 disturbance for BSC development. Realistic climatic and light conditions were simulated in the sun  
164 simulator facility of the Helmholtz Zentrum München (Neuherberg, Germany) by generating the entire  
165 spectrum from the ultraviolet (UV, 280-400 nm; UV-B, 280-315 nm; UV-A, 315-400 nm) to the near  
166 infrared (NIR) light with a combination of four types of lamps: metal halide lamps (Osram Powerstar HQI-  
167 TS 400W/D), quartz halogen lamps (Osram Haloline 400), blue fluorescent tubes (Philips TL-D 36W/BLUE)  
168 and UV-B fluorescent tubes (Philips TL 40W/12). The lamps were arranged in several groups to obtain the  
169 natural diurnal variations of solar irradiance by switching appropriate groups of lamps on and off. The  
170 short-wave cut-off was achieved by selected borosilicate and soda-lime glass filters as previously  
171 described [46, 47]. The pots were exposed to radiation for 16 h per day. Maximum radiation was reached  
172 in the middle of the day for 8 h at PAR (photosynthetic active radiation, 400-700 nm) of  $940 \mu\text{mol m}^{-2} \text{ s}^{-1}$ ,  
173 UV-A of  $17.7 \text{ W m}^{-2}$  and UV-B of  $0.37 \text{ W m}^{-2}$ . The climatic conditions were adjusted to a night-day cycle  
174 from  $18 \text{ }^\circ\text{C}$  to  $25 \text{ }^\circ\text{C}$ , and a relative air humidity of 95 – 90 %, respectively.

175 BSCs were sampled after four (T1) and ten (T2) months of incubation from three independent pots per  
176 soil substrate and sampling time point. Only the upper 2 mm were considered as BSC. In addition, samples  
177 of bulk soil without BSC development were taken at the beginning of the experiment (T0). In total, 18  
178 samples were collected (3 sampling times x 2 sites x 3 replicates). Samples for DNA analyses were directly  
179 frozen at  $-80 \text{ }^\circ\text{C}$ , while samples for biochemical analyses were stored at  $4 \text{ }^\circ\text{C}$  until further processing. For

180 the determination of water repellency and the computed tomography analysis, undisturbed samples from  
181 the end of the experiment were taken using Petri dishes and stored at 4 °C until further analysis.

182

### 183 *Physicochemical measurements*

184 For the analysis of dissolved organic carbon (DOC) and nitrogen (DON), bulk soils and BSC samples were  
185 suspended with a 0.01 M CaCl<sub>2</sub> solution in a 1:3 ratio (w/v), and shaken horizontally for 45 min. After  
186 passing through a Millex-HV 0.45 µm filter (Merck Millipore, USA), extracts were analyzed for DOC by  
187 means of a DIMA-TOC 100 analyzer (Dimatec Analysentechnik GmbH, DE), and for DON – using a Skalar  
188 Continuous Flow Analyzer SA5100 (Skalar Analytical B.V., NL) [48]. Soil pH of bulk soil samples was  
189 measured in 0.01 M CaCl<sub>2</sub> solution with a soil:solution ration of 1:5 (w/v) after 3 hours of incubation time.  
190 Water repellency of BSCs was measured as a dimensionless „repellency index“ using the ethanol/water  
191 microinfiltrometric sorptivity method according to Fischer et al. [49], where a theoretical value of 1  
192 characterizes totally non-repellent soils [50], and may exceed 50 for highly repellent soils [51].

193 Pre-experiments indicated that only BSCs from T2 grown on substrate from Chicken Creek developed a  
194 thickness sufficient for visualization by computed tomography (CT). Thus, only these samples were used  
195 to determine connectivity of the three-dimensional pore system of the BSCs and the underlying soil as  
196 described previously [52]. The structure of the undisturbed samples was analyzed using a micro-computed  
197 tomography scanner (X-Tek HMX 225, Nikon Metrology, BE) equipped with a fine-focus X-ray tube (spot  
198 size of 5 µm) and a digital flat panel detector with a resolution of 512 by 512 pixels (width by height). The  
199 resulting X-ray computed microtomography (XCMT) images were used to calculate Euler characteristics  
200 for 26 nearest neighbors of each voxel. So defined Euler numbers were computed as a function of pore  
201 size in the range between 15 and 291 µm [53].

202

203 *DNA extraction, library preparation and sequencing*

204 DNA was extracted from bulk soil and BSC samples using the 'Genomic DNA from soil' NucleoSpin Soil Kit  
205 (Macherey-Nagel, DE) according to the manufacturer's manual. Based on a pretest performance (data not  
206 shown), Buffer SL1 was chosen for sample lysis. DNA purity was verified by means of a NanoDrop 1000  
207 spectrophotometer (Thermo Fisher Scientific, USA). The quantity was also measured using a SpectraMax  
208 Gemini EM microplate reader (Molecular Devices, USA) together with a Quant-iT PicoGreen dsDNA Assay  
209 Kit (Life Technologies, USA), and is presented in Table 2. DNA was sheared using an E220 Focused-  
210 ultrasonicator (Covaris, USA) with the following conditions: peak incident power = 175 W, duty factor =  
211 10 %, cycles per burst = 200, treatment time = 100 s, temperature = 7 °C, water level = 6, sample volume  
212 = 50 µl, intensifier = yes. Library preparation was performed using the NEBNext Ultra DNA Library Prep Kit  
213 for Illumina and the NEBNext Multiplex Oligos for Illumina (both New England Biolabs, UK) as described  
214 in the protocol of the producer. Due to lower DNA concentrations (Table 2), samples from T0 underwent  
215 different molecular manipulations during library preparation than samples from T1 and T2. The NEBNext  
216 adaptor from Illumina was diluted 10-fold for samples from T1 and T2, and 50-fold for samples from T0,  
217 to prevent the occurrence of dimers. Size selection for samples from T1 and T2 was performed with  
218 Agencourt AMPure XP beads (Beckman Coulter, USA), using the volumes selecting for libraries with 500-  
219 700 bp inserts. No size selection was applied for samples from T0 due to low DNA concentrations of the  
220 libraries. PCR amplification was performed with 15 cycles for samples from T1 and T2, and 18 cycles for  
221 samples from T0. Primers used for samples from T1 and T2 were diluted 2-fold. Primers used for samples  
222 from T0 were not diluted. Libraries were pooled equimolarly, and 15 pM of the mixture was spiked with  
223 1 % PhiX. Sequencing was carried out on a MiSeq sequencer using a MiSeq Reagent Kit v3 for 600 cycles  
224 (Illumina, USA). Raw sequencing data obtained from the MiSeq is available at the sequencing read archive  
225 (SRA) under the accession number PRJNA509545.

226

227 *Bioinformatical analysis of sequencing data*

228 The raw sequencing data was processed as described by Vestergaard et al. [54]. Removal of remnant  
229 adaptor sequences, trimming of terminal nucleotides with Phred quality scores less than 15, and removal  
230 of reads shorter than 50 bp was carried out using AdapterRemoval [55]. Reads containing more than 1 %  
231 ambiguous bases (N) were removed by means of PRINSEQ-lite (version 0.20.4) [56]. DeconSeq (version  
232 0.4.3) [57] was used to remove PhiX contamination. Sufficient coverage of the metagenomic datasets was  
233 confirmed by means of Nonpareil (version 2.4) [58] with default settings (Supplementary material 1: Fig.  
234 S1).

235 Metagenomes obtained from bulk soils (T0) comprised reads on average 106 bp shorter than  
236 metagenomes created from BSCs (T1 and T2). To test whether the difference in read length affects the  
237 accuracy of annotations, T1 and T2 reads were trimmed *in-silico* in a randomized manner to resemble the  
238 length distributions of T0 reads. A comparison of the length distributions of exemplary “short” and “long  
239 reads” metagenomes, before and after trimming, is presented in Supplementary material 1: Fig. S2. The  
240 metagenomes with trimmed sequences were analyzed taxonomically together with the original  
241 metagenomes. Principal coordinates analysis (PCoA) ordination plots (Supplementary material 1: Fig. S3)  
242 showed that the taxonomic annotations were not notably biased by the difference in read length.  
243 Consequently, further analyses were performed on the metagenomes with original read lengths.

244 For taxonomic classification, metagenomic reads were aligned against the National Center for  
245 Biotechnology Information Non-Redundant (NCBI-NR) protein sequences database (January 2017) using  
246 Kaiju (version 1.4.4) [59] in Greedy mode with 5 allowed mismatches. Additionally, bacterial 16S rRNA  
247 gene sequences were extracted from the metagenomic datasets and annotated using SortMeRNA  
248 (version 2.0) [60] with the SILVA SSU database (release 132).

249 Subsequent functional annotations were performed for bacterial reads identified by Kaiju only. COG  
250 (Clusters of Orthologous Groups) functional categories were assigned based on the eggNOG (evolutionary  
251 genealogy of genes: Non-supervised Orthologous Groups) database (version 4.5) [61]. Assignment of  
252 genes specific for EPS and LPS biosynthesis and excretion, which were the focus of the current study, was  
253 carried out according to Cania et al. [37] by hidden Markov model (HMM) searches combined with blasts  
254 against protein sequences derived from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database  
255 (October 2016). Briefly, HMMs were obtained from the TIGRFAMs (version 15) [62] and Pfam (version 30)  
256 [63] databases. FragGeneScan (version 1.19) [64] was used to predict open-reading frames, which were  
257 subsequently scanned with HMMER (version 3) (hmmer.org). Matching reads (E-value threshold =  $10^{-5}$ )  
258 were mapped to KEGG Orthology (KO) numbers. A KO number was assigned to those reads for which the  
259 top 25 blast results were consistent. Blasting was carried out using Diamond (version 0.8.38) [65] with  
260 more-sensitive parameters. HMMs and KO numbers used for the analysis are listed in Table 1. Genes *algE*,  
261 *epsA* and *epsG* were not included in the analysis due to very low relative abundances ( $< 5 \times 10^7$ ). As most  
262 reads (> 50 %) assigned to the genes of interest using the HMM-KEGG pipeline were classified into the  
263 COG category “Function unknown”, this study was based mainly on the targeted approach proposed by  
264 Cania et al. [37]. The eggNOG pipeline was employed only for a general overview of the data.

265

266 *Statistical analysis and data visualization*

267 Analyses of the sequencing data were based on relative abundances of reads. These were obtained by  
268 dividing the number of reads assigned to a gene, COG functional category or bacterial family, by the total  
269 number of bacterial reads per sample, and multiplying by 100.

270 Statistical analyses and data visualization were conducted using R (version 3.4.4) [66]. Effects of soil  
271 substrate, incubation time, and their possible interaction, were determined according to Field et al. [67].

272 Briefly, significant differences were detected by a robust 2-way independent analysis of variance (ANOVA)  
273 based on the median as M-estimator, with 2000 bootstrap samples. For this purpose, the pbad2way  
274 function from the WRS package [68] was used. The influence was counted as significant if the p-value was  
275 below 5 % ( $p < 0.05$ ). Benjamini-Hochberg procedure was used to control the false discovery rate in data  
276 derived from the metagenomic datasets. Omega squared ( $\omega^2$ ) was calculated as an effect size to estimate  
277 the magnitude of observed influences of the analyzed factors. It can be interpreted as the percentage of  
278 variation in the dependent variable explained by the independent variable [69].

279 To detect global differences between samples, principal coordinate analysis (PCoA) ordinations of Bray-  
280 Curtis dissimilarity matrices were created using the pcoa function from the ape package [70]. Corrections  
281 for negative eigenvalues were performed by means of the Cailliez procedure. Bray-Curtis distances were  
282 calculated as an appropriate measure for community abundance data [71] using the vegdist function from  
283 the vegan package [72].

284 Spearman's rank correlation coefficient was used to identify whether the relative abundances of bacterial  
285 families and their functional genes were correlated. For this purpose, the function cor.test was used. The  
286 correlation was considered to be significant if  $p < 0.05$ . The average Rho was calculated based on absolute  
287 values.

288

## 289 **Results**

### 290 *Initial soil substrate parameters*

291 Bulk soils collected from both sites had similar low content of DOC and DON. DOC values were in the range  
292 of  $4.57 \pm 1.67 \mu\text{g/g}$  in samples collected from Chicken Creek, and  $6.63 \pm 0.46 \mu\text{g/g}$  in those from Lieberose,  
293 while DON was below detection limit in samples from both sites. Conversely, pH values differed between

294 soils from both locations. Soil from Chicken Creek was slightly alkaline ( $7.31 \pm 0.30$ ), whereas soil from  
295 Lieberose was rather acidic ( $\text{pH} = 5.42 \pm 0.39$ ). Initial soil substrate parameters are presented in Table 2.

296

### 297 *BSC development*

298 BSCs developed in the microcosm experiment were in the initial stage of development. They consisted  
299 mostly of bacterial and algal biofilms, which enmeshed soil particles and formed patches on the soil  
300 surface. Mosses were also observed, but they did not form a dense surface cover yet. For BSCs developed  
301 on the Chicken Creek soil, mosses and algae were already visible after the first four months of incubation  
302 (T1). For the Lieberose soil, mostly biofilms around single soil particles were visible at T1, whereas mosses  
303 and distinct BSC structures appeared after ten months of incubation (T2). Representative pictures are  
304 presented in Supplementary material 1: Fig. S4.

305 ANOVA revealed a significant influence of incubation time on DOC ( $p < 0.001$ ,  $\omega^2 = 0.70$ ) and DON ( $p <$   
306  $0.001$ ,  $\omega^2 = 0.81$ ). They accumulated over time and increased by one order of magnitude in BSCs at the  
307 end of the experiment compared to the bulk soils at the beginning of the experiment. The water  
308 repellency index at T2 was comparable between BSCs grown on soils taken from both locations. It  
309 amounted to  $1.12 \pm 0.15$  for BSCs originating from Chicken Creek, and  $1.16 \pm 0.25$  for those from  
310 Lieberose. BSC parameters are summarized in Table 2.

311 The exemplary CT images (Supplementary material 1: Fig. S5A-D) of BSCs from T2 grown on soil from  
312 Chicken Creek showed a layer of smaller particles in the crust horizon compared to the underlying soil  
313 substrate. Positive Euler numbers (Supplementary material 1: Fig. S5E) for both BSCs and the underlying  
314 soil indicate more isolated pores than connections in the pore network. The connectivity of the pore space  
315 was lower for the BSCs, especially when small pores ( $46 \mu\text{m}$ ) were considered (Euler number of  $8.6 \text{ mm}^{-3}$ ).  
316 In the underlying soil, the connectivity was the lowest for pore size class of  $107 \mu\text{m}$  (Euler number of

317 4.6 mm<sup>-3</sup>). The connectivity then increased towards larger pore sizes as indicated by decreasing Euler  
318 numbers.

319

### 320 *Major characteristics of the shotgun sequencing libraries*

321 Shotgun sequencing of 18 libraries made from bulk soils from the beginning of the experiment (T0) and  
322 BSCs from the four-months (T1) and ten-months (T2) samplings generated 18.3 Gbases of data in total.  
323 This corresponded to 59,710,640 filtered reads. The number of filtered reads per sample varied between  
324 2.1 and 5.3 million. Mean lengths of sequences after trimming ranged from 120 to 250 bp. Details of the  
325 raw and filtered sequencing data are summarized in Supplementary material 2: Table S1.

326 The coverage of the microbial diversity by the metagenomic datasets, which was calculated using  
327 Nonpareil, varied from 16.5 % to 67.3 % (Supplementary material 1: Fig. S1). As expected, metagenomes  
328 from T0 (nonpareil diversity index of  $19.24 \pm 0.07$ ) had higher coverage ( $41.9 \pm 12.8$  %) compared to  
329 metagenomes obtained from T1 and T2 (nonpareil diversity index of  $20.44 \pm 0.31$ , coverage of  $25.5 \pm 6.0$   
330 %).

331

### 332 *Taxonomic analysis*

333 42.83 % of all metagenomic reads were assigned to *Bacteria*, which could be further differentiated into  
334 366 families. Only these reads were further analyzed, as the main focus of this study was on EPS and LPS  
335 producers of bacterial origin, and molecular data on other microbial polysaccharide producers in the  
336 employed databases is poor. The principal coordinate analysis (PCoA) ordination plot (Fig. 1a) showed  
337 that bacterial communities were highly similar at the family level in bulk soils, and underwent  
338 differentiation during the development of BSCs. Dominant families were identified by selecting the five

339 most abundant families from each location at each time point, and sorting them according to their relative  
340 abundance of all metagenomes. Relative abundances of the dominant families are shown in  
341 Supplementary material 1: Fig. S6. As confirmed by ANOVA, the most characteristic families for T0 were  
342 *Burkholderiaceae*, *Comamonadaceae* and *Moraxellaceae*. *Flavobacteriaceae* were also highly abundant  
343 at T0, but showed additional differences between the two substrates, and had generally higher relative  
344 abundance in samples from Chicken Creek. Similarly, *Sphingomonadaceae* were typical for samples from  
345 Chicken Creek, but their relative abundance did not change significantly between the sampled time points.  
346 *Streptomyetaceae* had generally higher relative abundance in samples from Lieberose, and occurred  
347 mostly at T1 and T2. *Ktedonobacteraceae* and *Acidobacteriaceae* were typical at T1 and T2 for BSCs grown  
348 on soil substrate from the sand dune near Lieberose. *Bradyrhizobiaceae* were also characteristic for BSCs  
349 originating from Lieberose, but their abundance increased there only at T2. Cyanobacteria, including  
350 *Leptolyngbyaceae*, *Tolypothrichaceae* and *Nostocaceae*, were most abundant in BSCs grown on soil  
351 substrate from Chicken Creek at T1 and T2, while *Oscillatoriaceae* and *Microcoleaceae* dominated there  
352 at T1. Significance levels and  $\omega^2$  values are presented in Supplementary material 2: Table S2. Overall, the  
353 relative abundances of 13 families were influenced only by location, 125 – only by time, 63 – by both  
354 factors, and 130 – by interaction of both factors. The full list of impacted families can be taken from  
355 Supplementary material 2: Table S3.

356 The results of the taxonomic analysis of the whole metagenomic datasets based on the NCBI-NR database  
357 were supported by the 16S rRNA gene annotations with SILVA. Although only 0.0062 % of all metagenomic  
358 reads were assigned to the bacterial 16S rRNA gene, bacterial community composition did not differ when  
359 data from the analysis of the complete metagenomics datasets was compared to the phylogenetic analysis  
360 of subsampled 16S rRNA fragments (data not shown).

361

362 *Functional annotation of metagenomic datasets*

363 General function prediction in the metagenomic datasets was performed by means of the eggNOG  
364 database. In total, 73.08 % of bacterial reads were assigned to COG functional categories. The “function  
365 unknown” category was most abundant (~ 20 %), followed by “replication, recombination and repair” as  
366 well as “amino acid transport and metabolism” (each ~ 6 %). Relatively low abundant (< 0.5 %), but with  
367 a special importance to the initiation of BSC formation, were the “cell motility” and “extracellular  
368 structures” categories. ANOVA showed that these two categories were more abundant in bulk soils  
369 compared to BSCs. COG functional classification is presented in Supplementary material 1: Fig. S7, and  
370 significance levels and  $\omega^2$  values are listed in Supplementary material 2: Table S4.

371 Genes specific for the biosynthesis and excretion of alginate, colonic acid, levan and other EPSs as well as  
372 LPSs, which were identified using an approach combining HMM searches with blasts against sequences  
373 derived from the KEGG database, comprised 0.018 % of bacterial reads (Fig. 2). Key genes, with the overall  
374 relative abundance in all metagenomes in the range between 0.002 % and 0.005 %, were *wza*, *wcaB* and  
375 *wcaF* of the Wzy-dependent EPS synthesis pathway, and *lptF* and *lptG* of the LPS synthesis pathway.  
376 Moderately abundant ( $\geq 0.001$  %) were *kpsE* of the ABC transporter-dependent EPS synthesis pathway,  
377 and *wzt* of the LPS synthesis pathway. Genes *wcaK/amsJ*, *algJ*, *sacB* and *lptC* were the least abundant ( $\leq$   
378 0.0003 %).

379 ANOVA revealed that the relative abundances of most investigated genes changed mainly between T0  
380 and T1. However, the differences in the relative numbers of gene copies were also driven by the  
381 underlying soil substrate (Supplementary material 2: Table S5). In particular, the genes *wza* and *wcaF*  
382 increased at T1, and the increase was more pronounced in samples originating from Chicken Creek  
383 compared to those from Lieberose. Moreover, *wzt* increased in BSCs grown on soil substrate from Chicken  
384 Creek already at T1, while the increase was observed in BSCs grown on bulk soil taken from Lieberose only

385 at T2. Conversely, *kpsE* and *lptC* decreased at T1. Additionally, *kpsE* was relatively more abundant in  
386 samples from Lieberose, whereas *lptG* was dominating in samples from Chicken Creek. Finally, *wcaK/amsJ*,  
387 *algJ*, *sacB* and *lptF* were not significantly affected by either incubation time or soil substrate.

388

### 389 *Investigation of potential EPS/LPS producers*

390 The investigated genes were found in 210 different bacterial families, of which 11 families were found  
391 harboring the genes in samples originating from both locations, taken at all three sampling time points  
392 (Supplementary material 1: Fig. S8). The number of families harboring genes related to EPS and LPS  
393 formation was higher at T1 and T2 compared to T0 (Fig. 3). At T0, the investigated genes were associated  
394 with 33 families in bulk soil from Chicken Creek, and in 34 – in bulk soil from Lieberose. These numbers  
395 increased at T1 to 150 families in samples originating from Chicken Creek, and 100 – in samples from  
396 Lieberose. At T2, 146 families harbored the investigated genes in samples from Chicken Creek, and 87 –  
397 in samples from Lieberose.

398 Taxonomy of bacteria potentially capable of synthesis and excretion of EPSs and LPSs is presented in Fig.  
399 3 at the level of phylum or class (in case of *Proteobacteria*). At T0, the investigated genes were harbored  
400 mainly by *Betaproteobacteria*, whereas at T1 and T2, the major potential producers of EPSs and LPSs were  
401 members of *Cyanobacteria*, *Alphaproteobacteria* and *Chloroflexi*. Interestingly, differences were also  
402 found in the diversity pattern of potential EPS and LPS producers in response to the different soil  
403 substrates. In particular, *Cyanobacteria* were typical for BSCs grown on soil taken from Chicken Creek,  
404 while *Chloroflexi* and *Acidobacteria* were characteristic for BSCs originating from Lieberose.

405 ANOVA identified a significant impact on the overall relative abundance of the investigated genes caused  
406 by location alone in three families, by time alone – in 14 families, and by interaction of both factors – in  
407 23 families. The full list of affected families can be taken from Supplementary material 2: Table S3. The

408 PCoA plot (Fig. 1b) indicated that the distribution pattern of the analyzed genes among bacterial families  
409 resembled that of the total bacterial community (Fig. 1a). In fact, Spearman's rank correlation analysis  
410 revealed a positive correlation between the total abundance of a given family and the amount of  
411 sequences related to EPS and LPS formation harbored by that family for 57 families of potential EPS and  
412 LPS producers (average Rho = 0.69, minimum Rho = 0.47, maximum Rho = 0.97). Three families showed a  
413 negative correlation (average Rho = 0.56, minimum Rho = 0.49, maximum Rho = 0.61), and 150 exhibited  
414 no correlation (average Rho = 0.23, minimum Rho = 0.00, maximum Rho = 0.46).

415 Of the 57 families that showed a positive correlation, 25 exceeded an abundance of 1 %, and encompassed  
416 altogether 43.26 % of all bacterial reads. Both the relative abundance as well as the potential for EPS and  
417 LPS synthesis and export of these families were strongly influenced by both incubation time and  
418 underlying soil substrate (Fig. 4). In fact, these factors selected the key producers of EPSs and LPSs already  
419 at the phylum level. *Betaproteobacteria* (especially *Burkholderiaceae*), as well as *Gammaproteobacteria*  
420 (*Moraxellaceae*) and *Bacteroidetes* (*Flavobacteriaceae*) were prevalent at T0, although most of their  
421 members were found also at T1 and T2. *Deltaproteobacteria* (*Myxococcaceae* and *Archangiaceae*) and  
422 *Planctomycetes* (*Gemmataceae* and *Planctomycetaceae*) occurred mainly at T1 and T2 in BSCs grown on  
423 soil taken from Chicken Creek. However, *Gemmataceae* were relatively abundant also at T2 in BSCs  
424 originating from Lieberose. *Cyanobacteria* were characteristic for Chicken Creek samples from T1 and T2,  
425 but some of their members (*Oscillatoriaceae* and *Leptolyngbyaceae*) could also be important for EPS and  
426 LPS production in Lieberose samples from T1 and T2. Typical for Lieberose samples from T1 and T2 were  
427 *Chloroflexi* (*Ktedonobacteraceae* and *Thermogemmatissporaceae*) and *Acidobacteria* (*Acidobacteriaceae*).  
428 *Alphaproteobacteria* were prevalent at T1 and T2 in general, but some of their members were more  
429 characteristic for one of the underlying substrates (e.g. *Sphingomonadaceae* for soil from Chicken Creek,  
430 and *Acetobacteraceae* for that from Lieberose).

431

432 **Discussion**

433 *Bacterial communities of initial soils*

434 In the present study, initial BSCs developed from indigenous communities of free-living microbes, which  
435 were highly similar in bulk soils from both sites. As carbon and nitrogen availability are one of the most  
436 important factors shaping bacterial community structure [73, 74], their low concentrations could be the  
437 primary influence selecting only the best-adapted bacteria in nutrient-poor habitats such as the Chicken  
438 Creek catchment and the Lieberose sand dune. In fact, the most abundant bacterial families in the bulk  
439 soils from our study were *Burkholderiaceae*, *Comamonadaceae*, *Moraxellaceae* and *Flavobacteriaceae*.  
440 These families exhibit oligotrophic traits, as their metabolic versatility and ability to degrade a wide range  
441 of compounds, such as various polymers, polycyclic aromatic compounds, phenols and halogenated  
442 aromatics, enables them to thrive even in environments with limited nutritional opportunities [75-78].  
443 Consequently, these groups were isolated from habitats such as crude oil, desert soil, glacier ice or distilled  
444 water. Furthermore, many members of these families possess fimbriae and exhibit motility. This is in line  
445 with the higher amount of corresponding reads found in the bulk soils compared to the initial BSCs. These  
446 traits are especially important for free-living bacteria, as they assist in the first steps of cell attachment to  
447 a surface and establishment of biofilms [79]. In contrast, genes involved in the formation of EPSs and LPSs,  
448 which are particularly relevant in the later stages of biofilm development, were generally more abundant  
449 in the initial BSCs compared to the bulk soils.

450

451 *Influence of initial BSC on soil stability and hydrological properties*

452 EPSs and LPSs have protective functions, bind and mediate penetration of micronutrients into the cell,  
453 and function in cell-to-surface and cell-to-cell interactions, which are critical for biofilm development [16,  
454 80, 81]. The prevalence of genes related to EPS and LPS synthesis and export in initial BSCs was therefore

455 expected. EPSs and LPSs also play an important role in improving soil stability, especially in initial BSCs  
456 that harbor large amounts of bacteria, like in our study. Bacterial polysaccharides adhere around soil  
457 particles, connecting them and cementing into larger aggregates [82]. Several studies demonstrated that  
458 bacterial polysaccharides increased the amount of stable soil aggregates [83-85] and reduced rainfall-  
459 induced erosion up to 98% [35]. Using the exemplary XCMT images of the ten-months-old samples from  
460 Chicken Creek, we also confirmed the ability of initial BSCs to trap surface soil particles. Similar activity of  
461 cyanobacterial crusts was captured on XCMT images for example by Raanan et al. [86]. Moreover, the  
462 increase of the potential for EPS and LPS formation in the initial BSCs compared to the bulk soils was  
463 correlated in our study with an accumulation of dissolved organic carbon (data not shown). Altogether,  
464 these point to an increased production of adhesive bacterial polysaccharides in our BSCs.

465 Additionally, we measured the influence of BSCs on soil hydrological properties, as the key role in altering  
466 soil moisture dynamics seems to be played by polysaccharides [1]. On one hand, they tend to clog pores  
467 through swelling, which may reduce soil infiltrability [32, 34, 49, 87]. On the other hand, they can increase  
468 soil porosity, which is known to positively affect water penetration [88, 89]. Some researchers also  
469 postulate that polysaccharides alter the hydrophobicity of BSC surfaces [90]. In our study, the water  
470 repellency of BSCs incubated for ten months on both substrates was close to ideal wettability. Similar  
471 water repellency was reported for very young BSCs also in other studies [49, 91]. In initial BSCs, the effect  
472 on hydrological processes highly depends on the transient amount and chemical nature of polysaccharides  
473 building the bacterial biofilms [92]. For example, water molecules as well as nutrients are bound mainly  
474 by the hydrophilic polysaccharide fractions, while the hydrophobic fractions increase the stability of BSCs  
475 and their ability to adhere to solid surfaces [93]. Furthermore, polysaccharides in bacterial biofilms are  
476 subjected to constant modification and degradation processes, both enzymatic and abiotic [4]. Colica et  
477 al. [94] underlined that polysaccharide content cannot be directly correlated with BSC age, as the transient  
478 amount of polysaccharides in BSCs depends on the activity of both polysaccharide producers as well as

479 chemoheterotrophic organisms that use polymeric carbohydrates as a carbon source. Thus, the  
480 hydrological properties of BSCs are highly dynamic and may fluctuate during BSC development, as shown  
481 previously [49]. Comparing the structure of bacterial communities in BSCs with the composition and  
482 chemical properties of bacterial polysaccharides throughout the whole development of BSCs would surely  
483 shed more light on this issue. However, more research on the methods of extracting bacterial  
484 polysaccharides from BSC needs to be done before such measurements will be reliable and give additional  
485 information compared to the repellency index [4, 95].

486

#### 487 *Genes related to EPS and LPS formation*

488 Although the total relative abundance of genes involved in the formation of EPSs and LPSs increased in  
489 the initial BSCs compared to the bulk soils, the individual genes showed different responses. Especially  
490 abundant and showing the strongest increase were genes from the Wzy-dependent EPS synthesis  
491 pathway and the LPS synthesis pathway. Most bacterial reads in our study belonged to phyla well-known  
492 for LPS production, such as *Proteobacteria* (40 %), *Cyanobacteria* (20 %) and *Bacteroidetes* (5 %) [96].  
493 Moreover, recent evidence shows that LPS producers can be found even in phyla that are commonly  
494 considered as lacking LPSs [97, 98]. Therefore, the relatively high abundance of genes from the LPS  
495 synthesis pathways in our study was expected. Similarly, the relative abundance of genes from the Wzy-  
496 dependent pathway was expected, as it is the most widely distributed mechanism of EPS assembly and  
497 export [99, 100]. In particular, the *wza* gene encodes for an outer membrane protein Wza, which  
498 participates in the translocation across the outer membrane of a variety of EPSs in many different taxa  
499 [27]. In comparison, genes belonging to the other pathways of EPS assembly and export (ABC-dependent  
500 and synthase-dependent), as well as to the extracellular EPS synthesis, were less abundant in our  
501 metagenomes. However, these genes are found only in a limited number of bacteria [101-103].

502 In contrast to the other investigated genes, the relative abundances of the *kpsE* gene, which is part of the  
503 ABC-dependent EPS synthesis pathway, and the *lptC* gene of the LPS synthesis pathway, decreased in the  
504 BSCs compared to the bulk soils. The gene *kpsE* is associated with the synthesis of capsular  
505 polysaccharides, which enhance survival of bacterial cells in harsh environments [104]. This could explain  
506 the high relative abundance of *kpsE* in the low-nutrient bulk soils of the Chicken Creek catchment and the  
507 Lieberose sand dune. The LptC protein is part of the LptBFGC LPS export complex together with LptF and  
508 LptG. However, unlike LptF and LptG, LptC is not well-conserved among Gram-negative bacteria [105,  
509 106], and may not even be essential for LPS formation [107].

510 The differences in the relative abundances of genes associated with EPS and LPS formation were observed  
511 mainly between the bulk soils and the BSCs. Conversely, very few differences in the relative abundances  
512 of the investigated genes were found between samples originating from Chicken Creek and Lieberose.

513

514 *Differentiation of potential key producers of EPS and LPS during initial development of BSC on different soil*  
515 *substrates*

516 Even though the soil substrate had little impact on the relative abundance of the investigated genes, it  
517 shaped the composition of bacterial communities in the developing BSCs. In fact, bacterial communities  
518 that were highly similar in the bulk soils underwent differentiation once BSCs started to develop.  
519 Furthermore, the taxonomic affiliation of the investigated genes reflected the overall composition of the  
520 bacterial communities in our study, and thus the differentiation of the overall bacterial communities was  
521 accompanied by the differentiation of the communities of potential producers of adhesive  
522 polysaccharides. This is in line with the theory about functional redundancy, which states that important  
523 functions are preserved by a community even if the community changes its composition [108]. Our results

524 indicate that the potential to form EPSs and LPSs is an important trait for initial BSCs, as it is maintained  
525 despite the different development of bacterial communities on the two investigated substrates.

526 The importance of the potential to produce “soil glue” in the initial stage of BSC development is further  
527 underlined by the fact that the highest numbers of sequences related to EPS and LPS biosynthesis were  
528 harbored by the families dominating the initial BSCs. The potential key producers of adhesive  
529 polysaccharides found in BSCs grown on soil from the Chicken Creek catchment and the Lieberose sand  
530 dune were distinct already at the phylum level. In the Chicken Creek BSCs, the most abundant potential  
531 producers of EPSs and LPSs were *Cyanobacteria*. They are well known for their capability to form external  
532 polysaccharidic layers that enable them to survive in extreme environments [13]. In fact, the genetic  
533 machinery of the LPS synthesis as well as the Wzy-dependent pathway of the EPS synthesis were both  
534 found in *Cyanobacteria* before [27]. This explains the dominance of these particular polysaccharide  
535 biosynthesis pathways in the metagenomes from the Chicken Creek BSCs. However, *Cyanobacteria* played  
536 only a minor role in the community of potential EPS and LPS producers in the Lieberose BSCs, possibly  
537 because they prefer alkaline environments [1]. In the BSCs grown on the soil from Lieberose,  
538 *Cyanobacteria* were replaced by *Chloroflexi* and *Acidobacteria*, which favor acidic habitats [109-113].  
539 While *Chloroflexi* lack the ability to synthesize LPSs, *Acidobacteria* are known LPS producers [96].  
540 Furthermore, even though the information on the proficiency of both phyla in EPS formation is still limited,  
541 sequences related to EPS synthesis were previously found in *Acidobacteria*, and a recent report suggests  
542 that some members of this phylum produce large amount of EPSs [114]. *Acidobacteria* and *Chloroflexi* are  
543 also members of communities that embed themselves in an EPS matrix, such as biofilms, microbial mats  
544 and BSCs [115-117]. The low relative abundance of *Cyanobacteria* in BSCs grown on the soil from  
545 Lieberose suggests that, besides *Chloroflexi*, the other major phototrophic organisms there could have  
546 been algae, which are also well-known producers of EPSs. Algae dominate acidic soils, and are major  
547 components of the natural BSCs found at Lieberose, except for the terminal successional stage that is

548 dominated by mosses and fungi [1, 44, 45, 49]. However, the identification of eukaryotes involved in  
549 polysaccharide production is difficult using short-read shotgun sequencing, and would require a different  
550 approach [118, 119]. In any case, our results show that potential producers of EPSs and LPSs dominate  
551 bacterial communities of BSCs during the initial stage of BSC development. Consequently, the  
552 differentiation of overall bacterial communities leads to the emergence of distinct potential key producers  
553 of “soil glue”.

554 The differentiation of bacterial communities in our study could have been on one hand triggered by soil  
555 properties. For example, the two soil substrates used in our study differed in pH, which is one of the most  
556 important edaphic parameters determining the composition of bacterial communities in soil [120], but  
557 usually signifies that other edaphic parameters (e.g. micronutrient availability) also differ [121]. Therefore,  
558 the experimental design of the current study prevents us from making any definite conclusions on the  
559 influence of edaphic parameters on the community structure of potential “soil glue” producers. On the  
560 other hand, the observed differentiation of bacterial communities could have resulted from various rare  
561 species that were too low abundant to detect in the bulk soils, and started dominating during the initial  
562 development of BSCs. To identify the main drivers shaping the community composition of potential  
563 producers of EPSs and LPSs in initial BSCs, future experiments should involve multiple sterile soil  
564 substrates with diverse edaphic parameters, inoculated with the same initial bacterial community.

565

## 566 **Conclusions**

567 Our study indicates that the potential to produce EPSs and LPSs is an important trait for bacterial  
568 communities forming BSCs in the initial stage of BSC development, as (i) the relative abundance of genes  
569 related to the biosynthesis of adhesive polysaccharides increases in the bacterial communities of initial  
570 BSCs compared to the indigenous bacterial communities of bulk soils, (ii) the relative abundances of EPS

571 and LPS genes remain similar in initial BSCs with different composition of bacterial communities, and (iii)  
572 the highest numbers of sequences related to the “soil glue” production is found in families dominating  
573 initial BSCs. Furthermore, we demonstrate that the community composition of potential producers of  
574 EPSs and LPSs reflects the overall structure of bacterial communities in initial BSCs, and thus initial BSCs  
575 with different bacterial community composition harbor distinct potential key producers of adhesive  
576 polysaccharides. Whether the ability of BSCs to improve soil development in the long term is  
577 compromised by differences in the efficiency of polysaccharide formation, or the adhesive properties of  
578 EPSs and LPSs produced by different taxa, needs further investigation. Similarly, whether the  
579 differentiation of bacterial communities during the initial development of BSCs is primarily triggered by  
580 soil properties, or results from various rare species present in the initial bacterial community of bulk soil,  
581 remains to be determined.

582

### 583 **Compliance with Ethical Standards**

584 Conflict of Interest: The authors declare that they have no conflict of interest.

585 This article does not contain any studies with human participants or animals performed by any of the  
586 authors.

587

### 588 **References**

- 589 1. Belnap J, Lange OL (2003) Biological soil crusts: structure, function, and management. Springer,  
590 New York
- 591 2. Sancho LG, Maestre FT, Büdel B (2014) Biological soil crusts in a changing world: introduction to  
592 the special issue. *Biodivers Conserv* 23: 1611–1617.
- 593 3. Wu Y, Rao B, Wu P, Liu Y, Li G, Li D (2013) Development of artificially induced biological soil crusts  
594 in fields and their effects on top soil. *Plant Soil* 370: 115-124.
- 595 4. Rossi F, Mugnai G, De Philippis R (2018) Complex role of the polymeric matrix in biological soil  
596 crusts. *Plant Soil* 429: 19-34.

- 597 5. Totsche KU, Amelung W, Gerzabek MH, Guggenberger G, Klumpp E, Knief C, Lehndorff E, Mikutta  
598 R, Peth S, Prechtel A (2018) Microaggregates in soils. *J Plant Nutr Soil Sc* 181: 104-136.
- 599 6. Six J, Bossuyt H, Degryze S, Denef K (2004) A history of research on the link between (micro)  
600 aggregates, soil biota, and soil organic matter dynamics. *Soil Till Res* 79: 7-31.
- 601 7. Abu-Lail NI, Camesano TA (2003) Role of lipopolysaccharides in the adhesion, retention, and  
602 transport of *Escherichia coli* JM109. *Environ Sci Technol* 37: 2173-2183.
- 603 8. Veste M, Littmann T, Breckle S-W, Yair A (2001) The role of biological soil crusts on desert sand  
604 dunes in the northwestern Negev, Israel. In: Breckle, S-W, Veste, M, Wucherer, W (eds.)  
605 Sustainable land use in deserts. Springer, Heidelberg, pp. 357-367
- 606 9. Selbmann L, Stingle F, Petruccioli M (2003) Exopolysaccharide production by filamentous fungi:  
607 the example of *Botryosphaeria rhodina*. *Anton Leeuw Int J G* 84: 135-145.
- 608 10. Martínez-Cánovas MJ, Quesada E, Martínez-Checa F, del Moral A, Bejar V (2004) *Salipiger*  
609 *mucescens* gen. nov., sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium  
610 isolated from hypersaline soil, belonging to the  $\alpha$ -Proteobacteria. *Int J Syst Evol Micr* 54: 1735-  
611 1740.
- 612 11. Suela Silva M, Naves Sales A, Teixeira Magalhães-Guedes K, Ribeiro Dias D, Schwan RF (2013)  
613 Brazilian Cerrado soil Actinobacteria ecology. *Biomed Res Int* 2013.
- 614 12. Wu N, Zhang Y, Pan H, Zhang J (2010) The role of nonphotosynthetic microbes in the recovery of  
615 biological soil crusts in the Gurbantunggut Desert, Northwestern China. *Arid Land Res Manag* 24:  
616 42-56.
- 617 13. Pereira S, Zille A, Micheletti E, Moradas-Ferreira P, De Philippis R, Tamagnini P (2009) Complexity  
618 of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative  
619 genes involved in their biosynthesis and assembly. *FEMS Microbiol Rev* 33: 917-941.
- 620 14. Seviour R, Stasinopoulos S, Auer D, Gibbs P (1992) Production of pullulan and other  
621 exopolysaccharides by filamentous fungi. *Crc Cr Rev Biotechn* 12: 279-298.
- 622 15. Mahapatra S, Banerjee D (2013) Fungal exopolysaccharide: production, composition and  
623 applications. *Microbiol Insights* 6: 1-16.
- 624 16. Suresh Kumar A, Mody K, Jha B (2007) Bacterial exopolysaccharides—a perception. *J Basic Microb*  
625 47: 103-117.
- 626 17. Zaady E, Kuhn U, Wilske B, Sandoval-Soto L, Kesselmeier J (2000) Patterns of CO<sub>2</sub> exchange in  
627 biological soil crusts of successional age. *Soil Biol Biochem* 32: 959-966.
- 628 18. Fischer T, Gypser S, Subbotina M, Veste M (2014) Synergic hydraulic and nutritional feedback  
629 mechanisms control surface patchiness of biological soil crusts on tertiary sands at a post-mining  
630 site. *J Hydrol Hydromech* 62: 293-302.
- 631 19. Fischer T, Veste M (2018) Carbon cycling of biological soil crusts mirrors ecological maturity along  
632 a Central European inland dune catena. *Catena* 160: 68-75.
- 633 20. West NE (1990) Structure and function of microphytic soil crusts in wildland ecosystems of arid to  
634 semi-arid regions. *Adv Ecol Res* 20: 179-223.
- 635 21. Eldridge D, Greene R (1994) Microbiotic soil crusts—a review of their roles in soil and ecological  
636 processes in the rangelands of Australia. *Soil Res* 32: 389-415.
- 637 22. Kidron G, Barzilay E, Sachs E (2000) Microclimate control upon sand microbiotic crusts, western  
638 Negev Desert, Israel. *Geomorphology* 36: 1-18.
- 639 23. Yair A, Almog R, Veste M (2011) Differential hydrological response of biological topsoil crusts  
640 along a rainfall gradient in a sandy arid area: Northern Negev desert, Israel. *Catena* 87: 326-333.
- 641 24. Kidron GJ, Vonshak A (2012) The use of microbiotic crusts as biomarkers for ponding, subsurface  
642 flow and soil moisture content and duration. *Geoderma* 181: 56-64.
- 643 25. Belnap J (2006) The potential roles of biological soil crusts in dryland hydrologic cycles. *Hydrol*  
644 *Process* 20: 3159-3178.

- 645 26. Schmid J, Sieber V, Rehm B (2015) Bacterial exopolysaccharides: biosynthesis pathways and  
646 engineering strategies. *Front Microbiol* 6: 496.
- 647 27. Pereira SB, Mota R, Santos CL, De Philippis R, Tamagnini P (2013) Assembly and export of  
648 extracellular polymeric substances (EPS) in cyanobacteria: a phylogenomic approach. *Adv Bot Res*  
649 65: 235-279.
- 650 28. Hunt F (1985) Patterns of LPS synthesis in gram negative bacteria. *J Theor Biol* 115: 213-219.
- 651 29. Wang X, Quinn PJ (2010) Lipopolysaccharide: Biosynthetic pathway and structure modification.  
652 *Prog Lipid Res* 49: 97-107.
- 653 30. Whitfield C, Trent MS (2014) Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev*  
654 *Biochem* 83: 99-128.
- 655 31. Costa OY, Raaijmakers JM, Kuramae EE (2018) Microbial extracellular polymeric substances:  
656 ecological function and impact on soil aggregation. *Front Microbiol* 9: 1-14.
- 657 32. Mazor G, Kidron GJ, Vonshak A, Abeliovich A (1996) The role of cyanobacterial exopolysaccharides  
658 in structuring desert microbial crusts. *FEMS Microbiol Ecol* 21: 121-130.
- 659 33. HuiXia P, ZhengMing C, XueMei Z, ShuYong M, XiaoLing Q, Fang W (2007) A study on an  
660 oligotrophic bacteria and its ecological characteristics in an arid desert area. *Sci China Ser D* 50:  
661 128-134.
- 662 34. Colica G, Li H, Rossi F, Li D, Liu Y, De Philippis R (2014) Microbial secreted exopolysaccharides  
663 affect the hydrological behavior of induced biological soil crusts in desert sandy soils. *Soil Biol*  
664 *Biochem* 68: 62-70.
- 665 35. Kheirfam H, Sadeghi SH, Darki BZ, Homae M (2017) Controlling rainfall-induced soil loss from  
666 small experimental plots through inoculation of bacteria and cyanobacteria. *Catena* 152: 40-46.
- 667 36. Mugnai G, Rossi F, Felde VJMNL, Colesie C, Büdel B, Peth S, Kaplan A, De Philippis R (2018) The  
668 potential of the cyanobacterium *Leptolyngbya ohadii* as inoculum for stabilizing bare sandy  
669 substrates. *Soil Biol Biochem* 127: 318-328.
- 670 37. Cania B, Vestergaard G, Krauss M, Fliessbach A, Schloter M, Schulz S (2019) A long-term field  
671 experiment demonstrates the influence of tillage on the bacterial potential to produce soil  
672 structure-stabilizing agents such as exopolysaccharides and lipopolysaccharides. *Environ*  
673 *Microbiome* 1: 1-14.
- 674 38. Gerwin W, Schaaf W, Biemelt D, Fischer A, Winter S, Hüttl RF (2009) The artificial catchment  
675 "Chicken Creek"(Lusatia, Germany)—A landscape laboratory for interdisciplinary studies of initial  
676 ecosystem development. *Ecol Eng* 35: 1786-1796.
- 677 39. Russell DJ, Hohberg K, Elmer M (2010) Primary colonisation of newly formed soils by actinidid  
678 mites. *Soil Org* 82: 237-251.
- 679 40. Zaplata MK, Winter S, Fischer A, Kollmann J, Ulrich W (2012) Species-driven phases and increasing  
680 structure in early-successional plant communities. *Am Nat* 181: E17-E27.
- 681 41. Lukešová A (2001) Soil algae in brown coal and lignite post-mining areas in central Europe (Czech  
682 Republic and Germany). *Restor Ecol* 9: 341-350.
- 683 42. Gypser S, Herppich WB, Fischer T, Lange P, Veste M (2016) Photosynthetic characteristics and  
684 their spatial variance on biological soil crusts covering initial soils of post-mining sites in Lower  
685 Lusatia, NE Germany. *Flora* 220: 103-116.
- 686 43. Dümig A, Veste M, Hagedorn F, Fischer T, Lange P, Spröte R, Kögel-Knabner I (2014) Organic  
687 matter from biological soil crusts induces the initial formation of sandy temperate soils. *Catena*  
688 122: 196-208.
- 689 44. Fischer T, Veste M, Bens O, Hüttl RF (2012) Dew formation on the surface of biological soil crusts  
690 in central European sand ecosystems. *Biogeosciences* 9: 4621-4628.

- 691 45. Fischer T, Veste M, Eisele A, Bens O, Spyra W, Hüttl RF (2012) Small scale spatial heterogeneity of  
692 Normalized Difference Vegetation Indices (NDVIs) and hot spots of photosynthesis in biological  
693 soil crusts. *Flora* 207: 159-167.
- 694 46. Döhring T, Koefflerlein M, Thiel S, Seidlitz HK (1996) Spectral shaping of artificial UV-B irradiation  
695 for vegetation stress research. *J Plant Physiol* 148: 115-119.
- 696 47. Thiel S, Döhring T, Köfferlein M, Kosak A, Martin P, Seidlitz HK (1996) A phytotron for plant stress  
697 research: how far can artificial lighting compare to natural sunlight? *J Plant Physiol* 148: 456-463.
- 698 48. Brankatschk R, Töwe S, Kleineidam K, Schloter M, Zeyer J (2011) Abundances and potential  
699 activities of nitrogen cycling microbial communities along a chronosequence of a glacier forefield.  
700 *ISME J* 5: 1025-1037.
- 701 49. Fischer T, Veste M, Wiehe W, Lange P (2010) Water repellency and pore clogging at early  
702 successional stages of microbiotic crusts on inland dunes, Brandenburg, NE Germany. *Catena* 80:  
703 47-52.
- 704 50. Hallett P, Young I (1999) Changes to water repellence of soil aggregates caused by substrate-  
705 induced microbial activity. *Eur J Soil Sci* 50: 35-40.
- 706 51. Urbanek E, Hallett P, Feeney D, Horn R (2007) Water repellency and distribution of hydrophilic  
707 and hydrophobic compounds in soil aggregates from different tillage systems. *Geoderma* 140:  
708 147-155.
- 709 52. Köhne JM, Schlüter S, Vogel H-J (2011) Predicting solute transport in structured soil using pore  
710 network models. *Vadose Zone J* 10: 1082-1096.
- 711 53. Vogel H-J, Weller U, Schlüter S (2010) Quantification of soil structure based on Minkowski  
712 functions. *Comput Geosci* 36: 1236-1245.
- 713 54. Vestergaard G, Schulz S, Schöler A, Schloter M (2017) Making big data smart—how to use  
714 metagenomics to understand soil quality. *Biol Fert Soils* 53: 479-484. doi: 10.1007/s00374-017-  
715 1191-3
- 716 55. Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming,  
717 identification, and read merging. *BMC Res Notes* 9: 88.
- 718 56. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets.  
719 *Bioinformatics* 27: 863-864.
- 720 57. Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from  
721 genomic and metagenomic datasets. *PloS One* 6: e17288.
- 722 58. Rodriguez-R LM, Konstantinidis KT (2014) Estimating coverage in metagenomic data sets and why  
723 it matters. *ISME J* 8: 2349-2351.
- 724 59. Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics  
725 with Kaiju. *Nat Commun* 7: 11257.
- 726 60. Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in  
727 metatranscriptomic data. *Bioinformatics* 28: 3211-3217.
- 728 61. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR,  
729 Sunagawa S, Kuhn M (2015) eggNOG 4.5: a hierarchical orthology framework with improved  
730 functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:  
731 D286-D293.
- 732 62. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E (2013) TIGRFAMs and genome  
733 properties in 2013. *Nucleic Acids Res* 41: D387-D395.
- 734 63. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M,  
735 Sangrador-Vegas A (2016) The Pfam protein families database: towards a more sustainable future.  
736 *Nucleic Acids Res* 44: D279-D285.
- 737 64. Rho M, Tang H, Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads.  
738 *Nucleic Acids Res* 38: e191-e191.

- 739 65. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat*  
740 *Methods* 12: 59-60.
- 741 66. R Core Team (2016) R: a language and environment for statistical computing. R Foundation for  
742 Statistical Computing, Vienna.
- 743 67. Field A, Miles J, Field Z (2012) *Discovering statistics using R*. Sage publications, Thousand Oaks,  
744 California
- 745 68. Wilcox RR, Schönbrodt FD (2014) The WRS package for robust statistics in R. R package version  
746 0.24.
- 747 69. Tunks T (1978) The use of omega squared in interpreting statistical significance. *B Coun Res Music*  
748 *Ed*: 28-34.
- 749 70. Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language.  
750 *Bioinformatics* 20: 289-290.
- 751 71. Legendre P, Legendre LF (2012) *Numerical ecology*. Elsevier, Amsterdam
- 752 72. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,  
753 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) *vegan: community ecology*  
754 *package*. R package version 2.5-1.
- 755 73. Sul WJ, Asuming-Brempong S, Wang Q, Turlousse DM, Penton CR, Deng Y, Rodrigues JL, Adiku  
756 SG, Jones JW, Zhou J (2013) Tropical agricultural land management influences on soil microbial  
757 communities through its effect on soil organic carbon. *Soil Biol Biochem* 65: 33-38.
- 758 74. Cederlund H, Wessén E, Enwall K, Jones CM, Juhanson J, Pell M, Philippot L, Hallin S (2014) Soil  
759 carbon quality and nitrogen fertilization structure bacterial communities with predictable  
760 responses of major bacterial phyla. *Appl Soil Ecol* 84: 62-68.
- 761 75. Coenye T (2014) The family Burkholderiaceae. In: Rosenberg, E (ed.) *The prokaryotes:*  
762 *Alphaproteobacteria and Betaproteobacteria*. Springer, Heidelberg, pp. 759-776
- 763 76. Willems A (2014) The family Comamonadaceae. In: Rosenberg, E (ed.) *The prokaryotes:*  
764 *Alphaproteobacteria and Betaproteobacteria*. Springer, Heidelberg, pp. 777-851
- 765 77. Teixeira LM, Merquior VLC (2014) The family Moraxellaceae. In: Rosenberg, E (ed.) *The*  
766 *prokaryotes: Gammaproteobacteria*. Springer, Heidelberg, pp. 443-476
- 767 78. McBride MJ (2014) The family Flavobacteriaceae. In: Rosenberg, E (ed.) *The prokaryotes: Other*  
768 *Major Lineages of Bacteria and the Archaea*. Springer, Heidelberg, pp. 643-676
- 769 79. Vogeleeer P, Tremblay YD, Mafu AA, Jacques M, Harel J (2014) Life on the outside: role of biofilms  
770 in environmental persistence of Shiga-toxin producing *Escherichia coli*. *Front Microbiol* 5: 317.
- 771 80. Lindhout T, Lau PCY, Brewer D, Lam JS (2009) Truncation in the core oligosaccharide of  
772 lipopolysaccharide affects flagella-mediated motility in *Pseudomonas aeruginosa* PAO1 via  
773 modulation of cell surface attachment. *Microbiology+* 155: 3449-3460.
- 774 81. Kierek K, Watnick PI (2003) The *Vibrio cholerae* O139 O-antigen polysaccharide is essential for  
775 Ca<sup>2+</sup>-dependent biofilm development in sea water. *P Natl Acad Sci USA* 100: 14357-14362.
- 776 82. Huang Q, Wu H, Cai P, Fein JB, Chen W (2015) Atomic force microscopy measurements of bacterial  
777 adhesion and biofilm formation onto clay-sized particles. *Sci Rep-UK* 5: 16857.
- 778 83. de Caire GZ, De Cano MS, De Mule MZ, Palma R, Colombo K (1997) Exopolysaccharide of *Nostoc*  
779 *muscorum* (Cyanobacteria) in the aggregation of soil particles. *J Appl Phycol* 9: 249-253.
- 780 84. Carrasco L, Caravaca F, Azcón R, Roldán A (2009) Soil acidity determines the effectiveness of an  
781 organic amendment and a native bacterium for increasing soil stabilisation in semiarid mine  
782 tailings. *Chemosphere* 74: 239-244.
- 783 85. Rossi F, Li H, Liu Y, De Philippis R (2017) Cyanobacterial inoculation (cyanobacterisation):  
784 perspectives for the development of a standardized multifunctional technology for soil  
785 fertilization and desertification reversal. *Earth-Sci Rev* 171: 28-43.

- 786 86. Raanan H, Felde VJ, Peth S, Drahorad S, Ionescu D, Eshkol G, Treves H, Felix-Henningsen P,  
787 Berkowicz SM, Keren N (2016) Three-dimensional structure and cyanobacterial activity within a  
788 desert biological soil crust. *Environ Microbiol* 18: 372-383.
- 789 87. Issa OM, Défarge C, Trichet J, Valentin C, Rajot J-L (2009) Microbiotic soil crusts in the Sahel of  
790 Western Niger and their influence on soil porosity and water dynamics. *Catena* 77: 48-55.
- 791 88. Greene R (1992) Soil physical properties of three geomorphic zones in a semi-arid mulga  
792 woodland [*Acacia aneura*]. *Aust J Soil Res* 30: 55-69.
- 793 89. Eldridge DJ (2003) Biological soil crusts and water relations in Australian deserts. In: Belnap, J,  
794 Lange, OL (eds.) *Biological soil crusts: structure, function, and management*. Springer, Berlin pp.  
795 327-337
- 796 90. Felde VJMNL, Rossi F, Colesie C, Uteau-Puschmann D, Horne R, Felix-Henningsen P, De Philippis R,  
797 Peth S (2016) Pore characteristics in biological soil crusts are independent of extracellular  
798 polymeric substances. *Soil Biol Biochem* 103: 294-299.
- 799 91. Pluis J (1994) Algal crust formation in the inland dune area, Laarder Wasmeer, the Netherlands.  
800 *Vegetatio* 113: 41-51.
- 801 92. Flemming H-C, Wingender J (2010) The biofilm matrix. *Nat Rev Microbiol* 8: 623–633.
- 802 93. Rossi F, Micheletti E, Bruno L, Adhikary SP, Albertano P, De Philippis R (2012) Characteristics and  
803 role of the exocellular polysaccharides produced by five cyanobacteria isolated from phototrophic  
804 biofilms growing on stone monuments. *Biofouling* 28: 215-224.
- 805 94. Colica G, Li H, Rossi F, Philippis RD, Liu Y (2015) Differentiation of the characteristics of excreted  
806 extracellular polysaccharides reveals the heterogeneous primary succession of induced biological  
807 soil crusts. *J Appl Phycol* 27: 24-32.
- 808 95. Redmile-Gordon M, Brookes P, Evershed R, Goulding K, Hirsch P (2014) Measuring the soil-  
809 microbial interface: Extraction of extracellular polymeric substances (EPS) from soil biofilms. *Soil*  
810 *Biol Biochem* 72: 163-171.
- 811 96. Lagier J-C, Million M, Hugon P, Armougom F, Raoult D (2012) Human gut microbiota: repertoire  
812 and variations. *Front Cell Infect Mi* 2: 136.
- 813 97. Antunes LCS, Poppleton D, Klingl A, Criscuolo A, Dupuy B, Brochier-Armanet C, Beloin C, Gribaldo  
814 S (2016) Phylogenomic analysis supports the ancestral presence of LPS-outer membranes in the  
815 Firmicutes. *Elife* 5: e14589.
- 816 98. Poppleton DI, Duchateau M, Hourdel V, Matondo M, Flechsler J, Klingl A, Beloin C, Gribaldo S  
817 (2017) Outer membrane proteome of *veillonella parvula*: a diderm firmicute of the human  
818 microbiome. *Front Microbiol* 8: 1215.
- 819 99. Whitfield C, Larue K (2008) Stop and go: regulation of chain length in the biosynthesis of bacterial  
820 polysaccharides. *Nat Struct Mol Biol* 15: 121-123.
- 821 100. Whitfield C (2010) Polymerases: glycan chain-length control. *Nat Chem Biol* 6: 403-404.
- 822 101. Rosenow C, Esumeh F, Roberts IS, Jann K (1995) Characterization and localization of the KpsE  
823 protein of *Escherichia coli* K5, which is involved in polysaccharide export. *J Bacteriol* 177: 1137-  
824 1143.
- 825 102. Bachtiar BM, Coloe PJ, Fry BN (2007) Knockout mutagenesis of the *kpsE* gene of *Campylobacter*  
826 *jejuni* 81116 and its involvement in bacterium–host interactions. *FEMS Immunol Med Mic* 49: 149-  
827 154.
- 828 103. Muhammadi, Ahmed N (2007) Genetics of bacterial alginate: alginate genes distribution,  
829 organization and biosynthesis in bacteria. *Curr Genomics* 8: 191-202.
- 830 104. Rendueles O, Garcia-Garcerà M, Néron B, Touchon M, Rocha EP (2017) Abundance and co-  
831 occurrence of extracellular capsules increase environmental breadth: Implications for the  
832 emergence of pathogens. *PLoS Pathog* 13: e1006525.

- 833 105. Putker F, Bos MP, Tommassen J (2015) Transport of lipopolysaccharide to the Gram-negative  
834 bacterial cell surface. *FEMS Microbiol Rev* 39: 985-1002.
- 835 106. Ruiz N, Gronenberg LS, Kahne D, Silhavy TJ (2008) Identification of two inner-membrane proteins  
836 required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *P Natl*  
837 *Acad Sci USA* 105: 5537-5542.
- 838 107. Benedet M, Falchi FA, Puccio S, Di Benedetto C, Peano C, Polissi A, Dehò G (2016) The lack of the  
839 essential LptC protein in the trans-envelope lipopolysaccharide transport machine is  
840 circumvented by suppressor mutations in LptF, an inner membrane component of the *Escherichia*  
841 *coli* transporter. *PloS One* 11: e0161354.
- 842 108. Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in microbial communities.  
843 *P Natl Acad Sci USA* 105: 11512-11519.
- 844 109. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N (2009) A comprehensive survey  
845 of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* 3: 442-453.
- 846 110. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a  
847 predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol*  
848 75: 5111-5120.
- 849 111. Wilhelm RC, Niederberger TD, Greer C, Whyte LG (2011) Microbial diversity of active layer and  
850 permafrost in an acidic wetland from the Canadian High Arctic. *Can J Microbiol* 57: 303-315.
- 851 112. Santofimia E, González-Toril E, López-Pamo E, Gomariz M, Amils R, Aguilera Á (2013) Microbial  
852 diversity and its relationship to physicochemical characteristics of the water in two extreme acidic  
853 pit lakes from the Iberian Pyrite Belt (SW Spain). *PLoS One* 8: e66746.
- 854 113. Jones DS, Lapakko KA, Wenz ZJ, Olson MC, Roepke EW, Sadowsky MJ, Novak PJ, Bailey JV (2017)  
855 Novel microbial assemblages dominate weathered sulfide-bearing rock from copper-nickel  
856 deposits in the Duluth complex, Minnesota, USA. *Appl Environ Microbiol* 83: e00909-00917.
- 857 114. Kielak AM, Castellane TC, Campanharo JC, Colnago LA, Costa OY, Da Silva MLC, Van Veen JA, Lemos  
858 EG, Kuramae EE (2017) Characterization of novel Acidobacteria exopolysaccharides with potential  
859 industrial and ecological applications. *Sci Rep-UK* 7: 41193.
- 860 115. Rampadarath S, Bandhoa K, Puchooa D, Jeewon R, Bal S (2017) Early bacterial biofilm colonizers  
861 in the coastal waters of Mauritius. *Electron J Biotechn* 29: 13-21.
- 862 116. Prieto-Barajas CM, Valencia-Cantero E, Santoyo G (2017) Microbial mat ecosystems: Structure  
863 types, functional diversity, and biotechnological application. *Electron J Biotechn* 31: 48-56.
- 864 117. Mogul R, Vaishampayan P, Bashir M, McKay CP, Schubert K, Bornaccorsi R, Gomez E, Tharayil S,  
865 Payton G, Capra J (2017) Microbial community and biochemical dynamics of biological soil crusts  
866 across a gradient of surface coverage in the central Mojave desert. *Front Microbiol* 8: 1974.
- 867 118. De Vries M, Schöler A, Ertl J, Xu Z, Schloter M (2015) Metagenomic analyses reveal no differences  
868 in genes involved in cellulose degradation under different tillage treatments. *FEMS Microbiol Ecol*  
869 91: fiv069.
- 870 119. Wooley JC, Godzik A, Friedberg I (2010) A primer on metagenomics. *PLoS Comput Biol* 6: 1-13.
- 871 120. Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome.  
872 *Nat Rev Microbiol* 15: 579-590.
- 873 121. Lammel DR, Barth G, Ovaskainen O, Cruz LM, Zanatta JA, Ryo M, de Souza EM, Pedrosa FO (2018)  
874 Direct and indirect effects of a pH gradient bring insights into the mechanisms driving prokaryotic  
875 community structures. *Microbiome* 6: 106.

877 Table 1. Proteins related to exo- and lipopolysaccharide production with corresponding KO numbers,  
 878 HMM IDs and genes

Protein	KO number	HMM ID	Gene
polysaccharide export outer membrane protein Wza	K01991	PF02563	<i>wza</i>
colanic acid biosynthesis acetyltransferase WcaB	K03819	TIGR04016	<i>wcaB</i>
colanic acid biosynthesis acetyltransferase WcaF	K03818	TIGR04008	<i>wcaF</i>
colanic acid/amylovoran biosynthesis pyruvyl transferase WcaK/AmsJ	K16710	TIGR04006	<i>wcaK/amsJ</i>
capsular polysaccharide export system permease KpsE	K10107	TIGR01010	<i>kpsE</i>
alginate biosynthesis acetyltransferase AlgJ	K19295	PF16822	<i>algJ</i>
levansucrase SacB	K00692	PF02435	<i>sacB</i>
lipopolysaccharide transport system ATP-binding protein Wzt	K09691	PF14524	<i>wzt</i>
LptBFGC lipopolysaccharide export complex inner membrane protein LptC	K11719	TIGR04409, PF06835	<i>lptC</i>
LptBFGC lipopolysaccharide export complex permease LptF	K07091	TIGR04407	<i>lptF</i>
LptBFGC lipopolysaccharide export complex permease LptG	K11720	TIGR04408, PF03739	<i>lptG</i>

879

880 Table 2. DNA concentration, dissolved organic carbon (DOC), water repellency, dissolved organic nitrogen  
 881 (DON) and pH values. The mark “-” signifies that the parameter was not measured for the respective  
 882 samples, while “bdl” stands for “below detection limit”

Location	Time	DNA [ng/g]	DOC [ $\mu$ g/g]	DON [ $\mu$ g/g]	pH	Water repellency
Chicken Creek	T0	1.14 $\pm$ 0.29	4.57 $\pm$ 1.67	bdl	7.31 $\pm$ 0.30	-
	T1	28.89 $\pm$ 8.58	36.50 $\pm$ 4.31	1.32 $\pm$ 0.36	-	nd
	T2	30.95 $\pm$ 7.73	48.02 $\pm$ 18.06	1.14 $\pm$ 0.36	-	1.12 $\pm$ 0.15
Lieberose	T0	2.25 $\pm$ 1.04	6.63 $\pm$ 0.46	bdl	5.42 $\pm$ 0.39	-
	T1	5.79 $\pm$ 2.14	42.02 $\pm$ 6.97	0.83 $\pm$ 0.09	-	-
	T2	24.65 $\pm$ 4.63	81.03 $\pm$ 26.52	1.11 $\pm$ 0.15	-	1.16 $\pm$ 0.25

883

884

885

## Figure titles and legends

886 **Fig. 1** PCoA plots depicting differences on the family level in a – bacterial community composition, and b  
887 – taxonomic affiliation of genes related to EPS and LPS formation. Ellipses drawn around triplicates  
888 represent a 95 % confidence level

889 **Fig. 2** Relative abundances of genes specific for the formation of EPSs and LPSs. Error bars show standard  
890 deviations

891 **Fig. 3** Comparison of relative abundances of bacteria with and without the potential for EPS and LPS  
892 formation (labeled as „Present” and „Absent”). The distinction between the potential producers and non-  
893 producers was performed on the level of family. The families were then pooled according to their  
894 respective phyla or, in case of *Proteobacteria*, classes. Values above bars represent total numbers of  
895 displayed families

896 **Fig. 4** Potential key families of EPS and LPS formation based on their relative abundance and the relative  
897 abundance of their genes related to EPS and LPS biosynthesis. Note the different color intensities between  
898 bacterial and gene abundances

899

900

## Supplementary materials

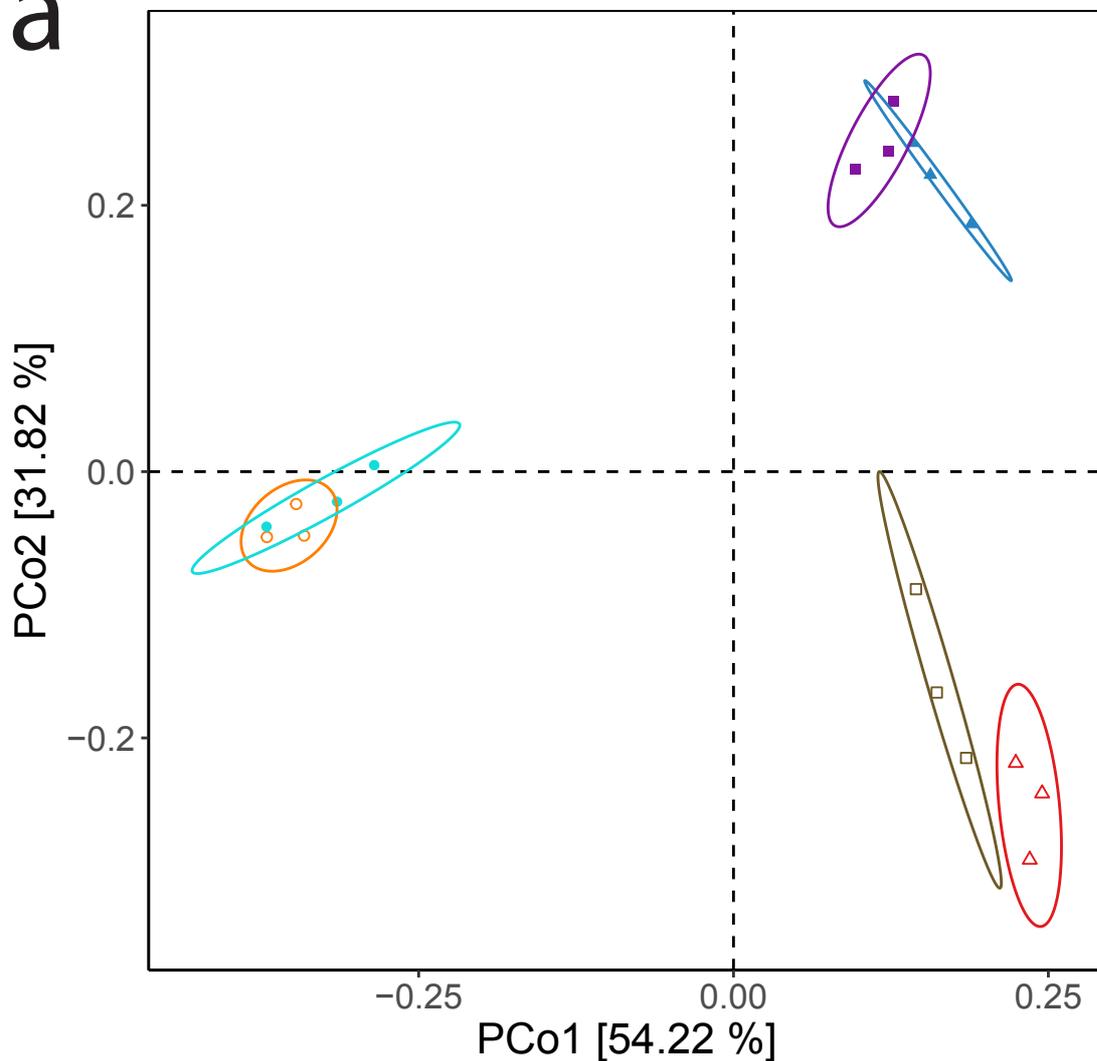
901 **Supplementary material 1:** Supplementary figures S1-8

902 **Supplementary material 2:** Supplementary tables S1-5

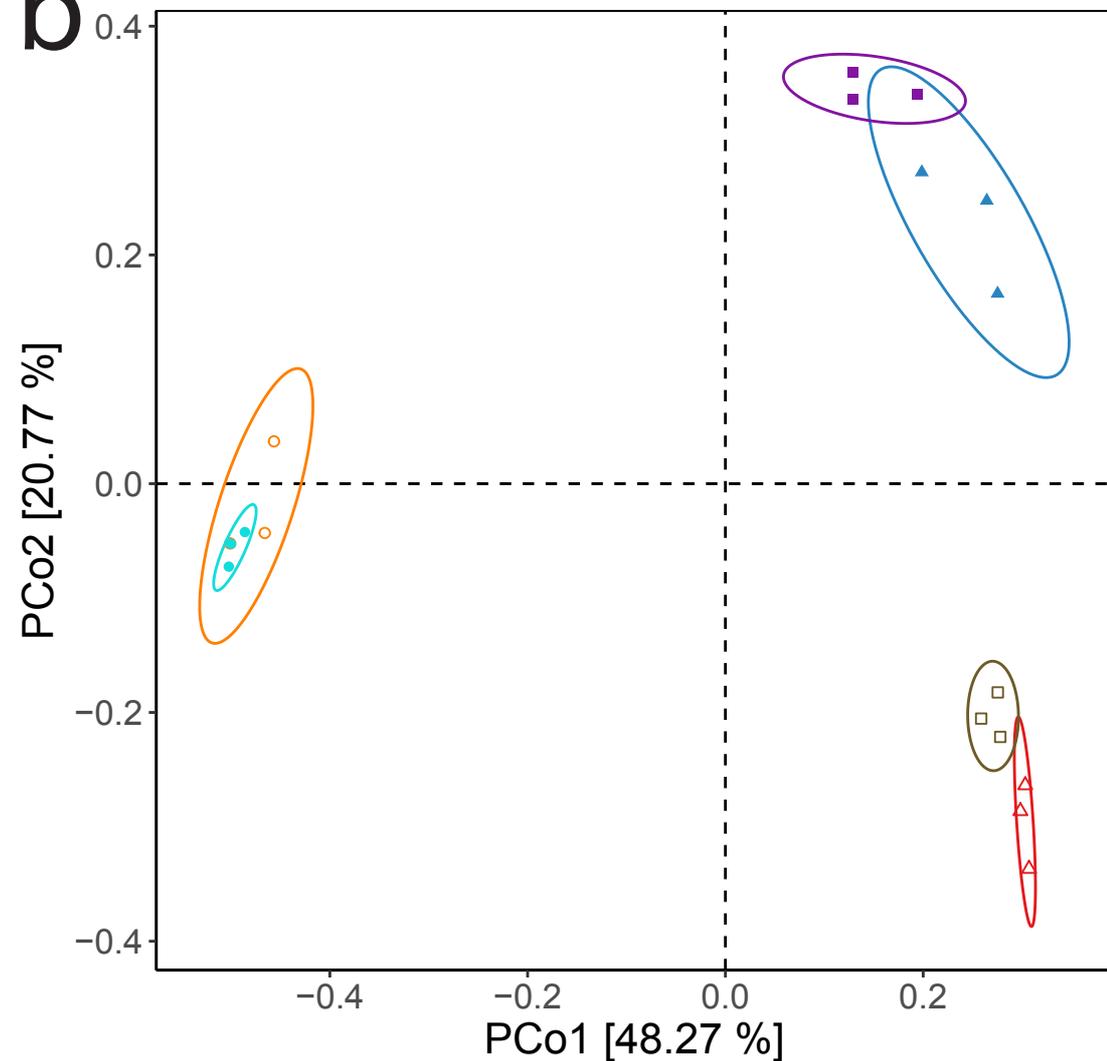
Figure 1

**a**

PCoA - PCo1 vs PCo2

**b**

PCoA - PCo1 vs PCo2

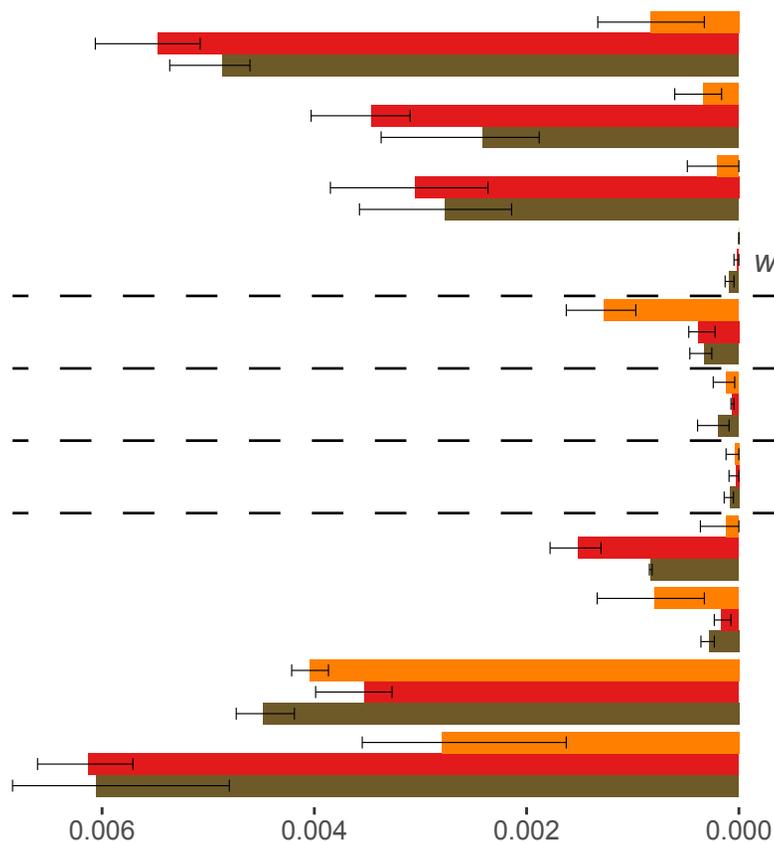
**Location, time**

- Chicken Creek, T0
- △ Chicken Creek, T1
- Chicken Creek, T2
- Lieberose, T0
- ▲ Lieberose, T1
- Lieberose, T2

Figure 2

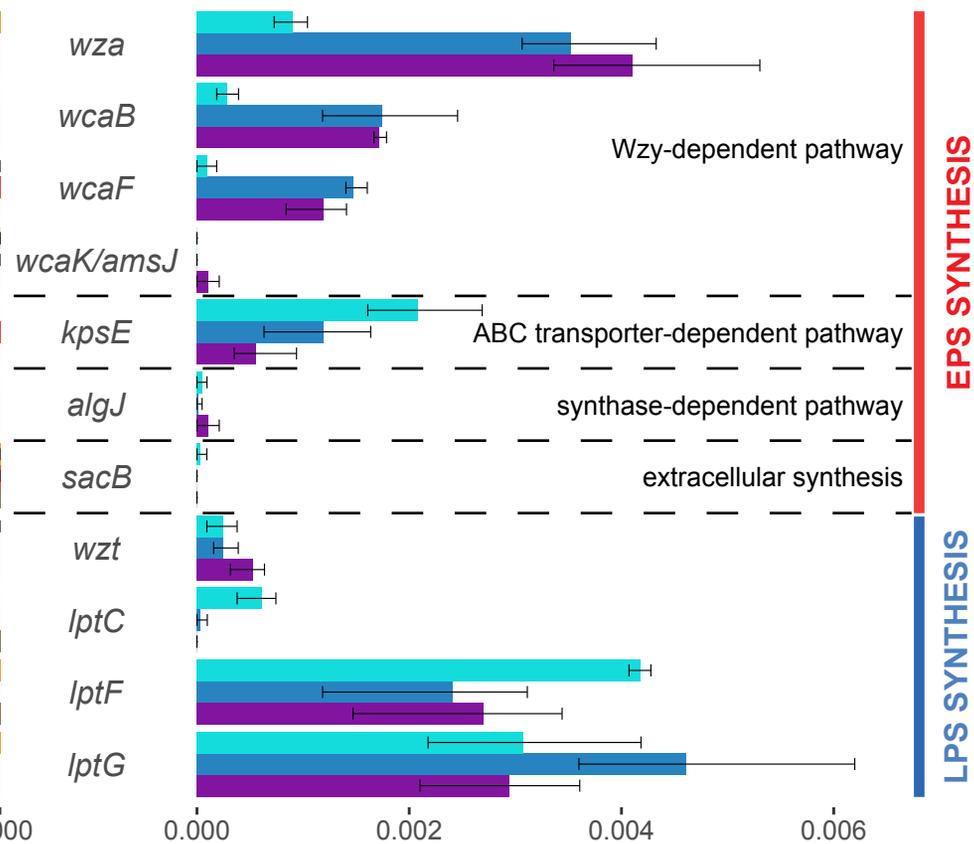
## Chicken Creek

T0 T1 T2



## Lieberose

T0 T1 T2



Wzy-dependent pathway

ABC transporter-dependent pathway

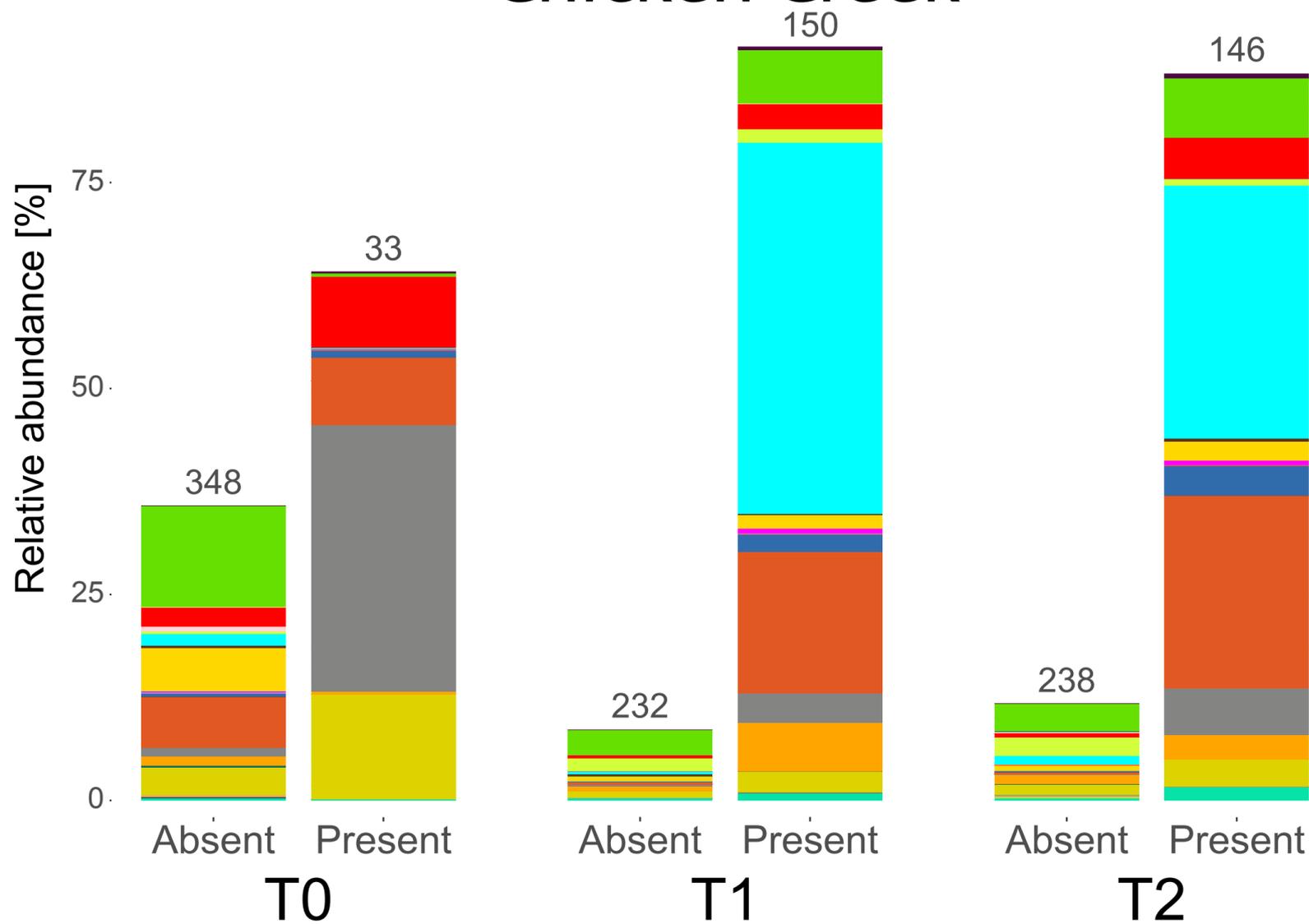
synthase-dependent pathway

extracellular synthesis

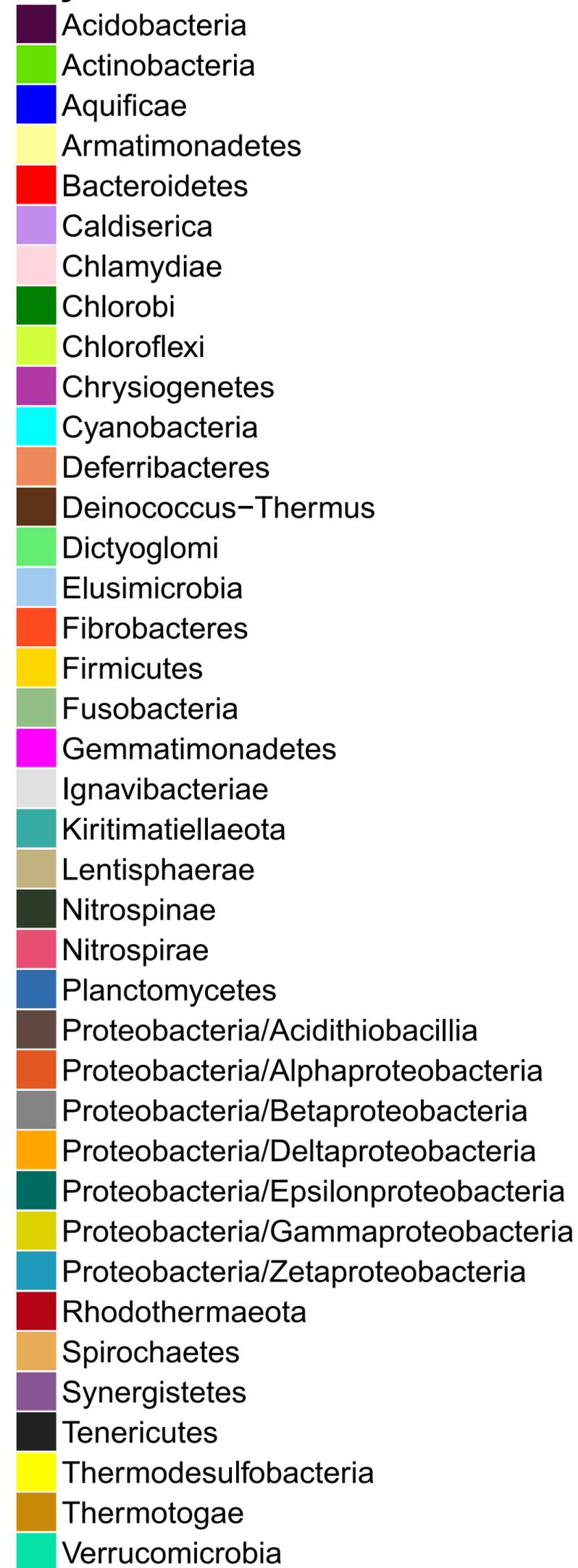
EPS SYNTHESIS

LPS SYNTHESIS

## Chicken Creek



## Phyla/classes



## Lieberose

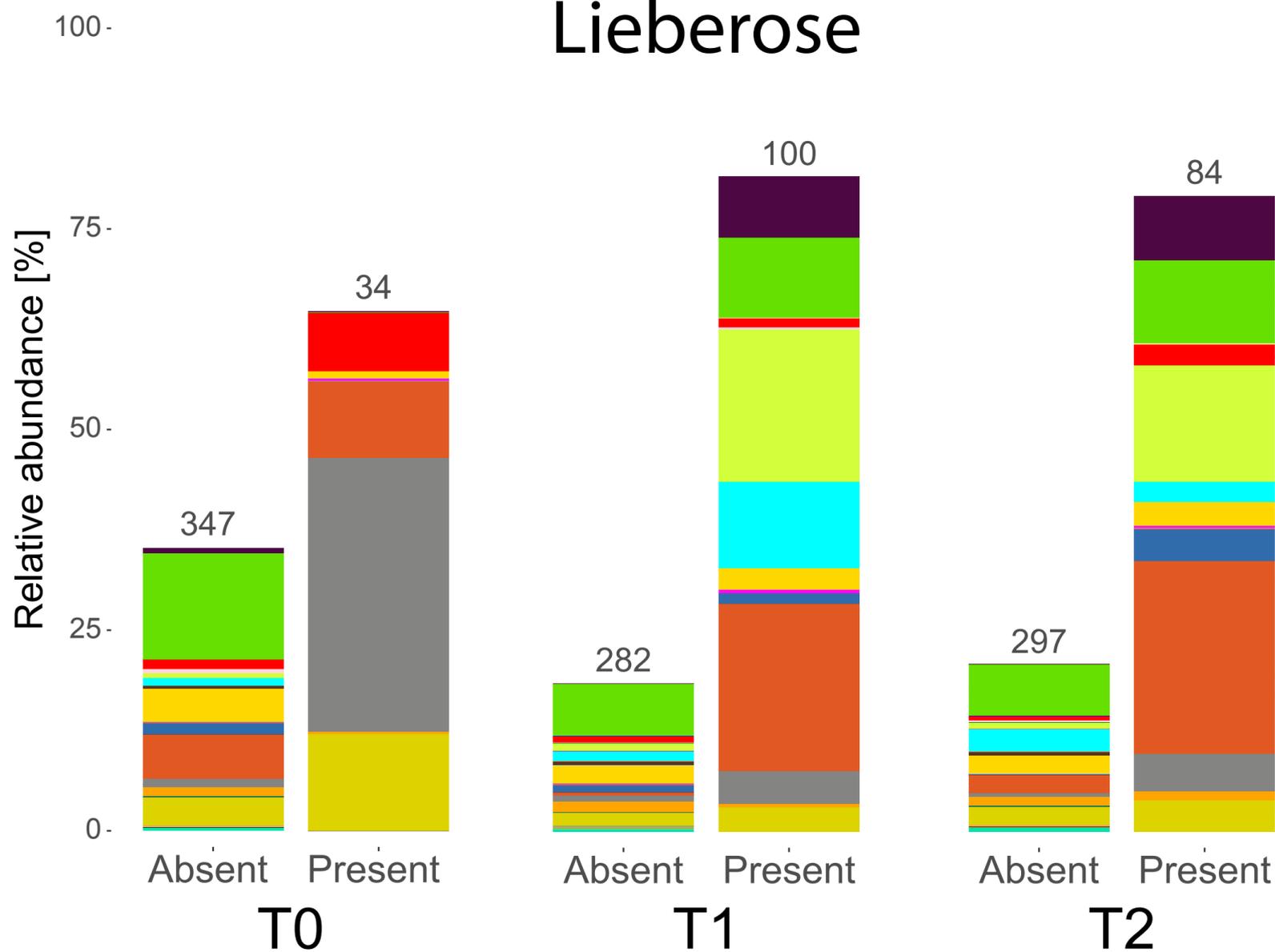
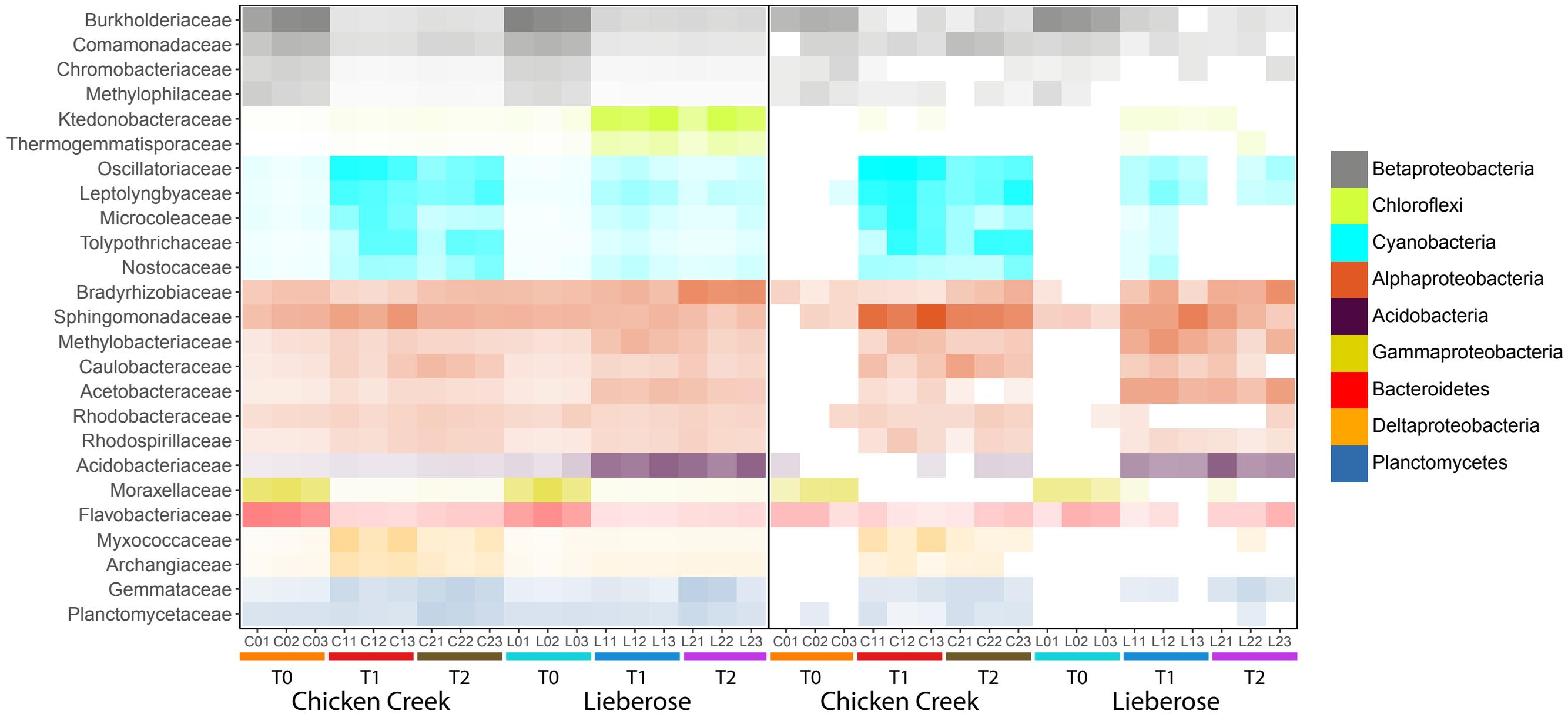


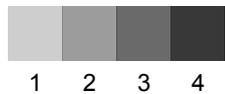
Figure 4

### Bacterial abundance

### Gene abundace



sqrt(relative abundance [%])



sqrt(relative abundance [%])





Click here to access/download  
**Supplementary Material**  
Supplementary material 1.pdf





Click here to access/download  
**Supplementary Material**  
Supplementary material 2.pdf



## Response to Reviewers

### Manuscript ID MECO-D-19-00199

We would like to thank the reviewers for their detailed comments and helpful suggestions. We believe that their feedback has enabled us to improve and clarify the manuscript. We will now address the specific comments of each reviewer.

#### Reviewer #1

Comment 1: Introduction. Lines 37-38. depends on structure of BSC organisms

Response 1: The referred sentence has been changed accordingly (lines 70-71).

Comment 2: Lines 53. Polysaccharide-producing

Response 2: The dash has been added, as corrected by the reviewer (lines 99-100).

Comment 3: Lines 46-52. Which family/genus of cyanobacteria or bacteria are the main producers of EPS and LPS? Could you add few sentences with this information.

Response 3: While LPSs are present in most Gram-negative bacteria, EPSs are exuded by a wide range of taxa. Among the most-recognized producers of EPSs are cyanobacterial members of *Oscillatoria*, *Nostoc*, *Lyngbya* and *Schizothrix*, as well as bacterial members of *Microbacterium*, *Pseudomonas*, *Bacillus*, *Paenibacillus* and *Streptomyces*. This information has been added to the revised manuscript (lines 88-91).

Comment 4: Lines 53-54. Are there other studies about EPS and LPS producers in different developmental stages of BSC? If yes, add few sentences. If not, indicate that there are no such studies.

Response 4: To our knowledge, no previous study focused on the communities of EPS and LPS producers in any developmental stage of BSCs. This has been stated now in the revised manuscript (lines 98-99).

Comment 5: Lines 57-58. "as opposed to most studies" - add at least few references

Response 5: This part of the Introduction section has been revised, and the focus has been changed according to the suggestions of the reviewers. In the revision process, the referred sentence has been deleted.

Comment 6: Methods. Lines 80-82. Contradicting sentence. Write "and" instead of "while"

Response 6: This part of the Materials and Methods section has been revised. Instead of comparing the two sites, now they are described separately. In the revision process, the referred sentence has been deleted.

Comment 7: Line 83. "were" instead of "are"

Response 7: The referred sentence has been corrected accordingly (line 291).

Comment 8: Lines 82-86. You are introducing chemical measurements in Line 119. This information should go to the results section.

Response 8: Chemical measurements of the initial soil substrates have been moved to the Results section (subsection "Initial soil substrate parameters"), as suggested by the reviewer (lines 290-295).

Comment 9: Line 86. pH 5.42 is not slightly acidic, but rather acidic (it's in the range between medium and strong)

Response 9: The word "slightly" has been exchanged for "rather" in accordance with the reviewer's suggestion (line 295).

Comment 10: Lines 87-90. This paragraph should go to the next section with sampling description.

Response 10: The referred paragraph has been moved from the "Site description" subsection to the next "Sampling and incubation experiment" subsection in accordance with the reviewer's suggestion (lines 152-158).

Comment 11: Line 92. Sampling and incubation experiment

Response 11: The title of the subsection has been change in accordance with the reviewer's suggestion (line 151).

Comment 12: Lines 95-96. How many pots were in total? From each site?

Response 12: In total, the microcosm experiment consisted of 18 pots (9 per site). This information has been added to the revised manuscript (lines 160-161).

Comment 13: Line 139. DNA extraction, PCR and sequencing

Response 13: The title of the subsection has been changed from "DNA extraction and sequencing" to "DNA extraction, library preparation and sequencing" (line 203). The term "library preparation" was used instead of the term "PCR" that was suggested by the reviewer because PCR was only one of several steps during library preparation, while library preparation describes the whole process that was performed to prepare the extracted DNA for sequencing.

Comment 14: Line 154-155. Indicate why you used different number of cycles.

Response 14: As stated in the manuscript (lines 212-214), library preparation was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina following the protocol of the producer. Illumina's NEBNext Ultra DNA Library Prep protocol recommends modifications to the standard procedure for samples with low DNA concentrations. For this reason, samples from T0, which had very low DNA concentrations,

underwent different molecular manipulations during library preparation compared to samples from T1 and T2, which had higher DNA concentrations. This explanation has been added to the revised manuscript (lines 214-215). The DNA concentrations have also been added as part of Table 2.

Comment 15: Line 155. Which primers did you use? This is very important information.

Response 15: As stated in the manuscript (line 213), we used the NEBNext Multiplex Oligos for Illumina. The reason is that shotgun sequencing was used in this study instead of amplicon sequencing. In amplicon sequencing, specific primers are used to target specific genes. In this case, the information about the primer sequences is important to know which regions of the genes were sequenced. However, shotgun sequencing is a method used for sequencing of random DNA regions. Specifically, DNA is broken up randomly into short fragments, and special adaptors are attached to the ends of these fragments. The adaptors contain a unique sequence to which primers can bind. The NEBNext Multiplex Oligos for Illumina used in this study contain short tag sequences used to identify the origin of every DNA fragment during parallel sequencing of different samples.

Comment 16: Lines 150-160. I would like to see clear explanation why molecular manipulations for T0 were performed differently than for T1 and T2.

Response 16: See our response to comment 14.

Comment 17: Line 160. I cannot find your sequences in database.

Response 17: The sequences are already uploaded and will become publically available after this manuscript is accepted. The following picture shows the confirmation email received from the database after uploading the sequences.



Dear Barbara Cania,

This is an automatic acknowledgment that your recent submission to SRA database has been successfully processed and will be released on the date specified.

Please reference PRJNA509545 in your publication.

SRA accession: PRJNA509545  
Temporary Submission ID: SUB4944632  
Release date: 2020-01-20

Your SRA records will be accessible with the following link after the indicated release date:  
<https://www.ncbi.nlm.nih.gov/sra/PRJNA509545>

Send questions and update requests to [sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov); include the SRA accession PRJNA509545 in any correspondence.

Regards,

NCBI SRA Submissions Staff  
Bethesda, Maryland USA

Comment 18: Line 165. What do you mean by trimming of the reads here? Were they trimmed to the same number of reads in each sample? Or is it primers and adapters removal? It is not clear.

Response 18: By trimming of the reads, we referred to the removal of terminal nucleotides with low Phred quality scores (less than 15). This has been specified in the revised manuscript (line 229).

Comment 19: Results. Lines 212-216. Would be interesting to know detailed description of developmental process of BSCs. For example, coverage area of visible BSC components (e.g. mosses, algae, cyanobacteria). Which mosses were present in the BSCs? Were these mosses also in the fields? Were there lichens? I think this information is important since the presence/absence of BSC organisms might affect diversity and abundance of potential EPS/LPS producers.

Response 19: We agree that the presence/absence of BSC organisms might affect diversity and abundance of potential EPS/LPS producers. We believe that this issue, as well as the influence of edaphic parameters on the EPS/LPS producers, should be addressed in further studies. However, the aim of the current study was to investigate if different types of bacterial BSCs harbor distinct potential bacterial producers of EPSs/LPSs, and not to elucidate the influence of different abiotic and biotic factors on their potential differentiation. Large parts of the manuscript have been revised to make this point clear. For this reason, organisms other than bacteria have not been analyzed as part of this study. Similarly, the coverage areas of visible BSC components have not been calculated, although we have now added more pictures of the BSC surface to Supplementary material 1: Fig. S4. This should give the readers a better idea of the general structure and coverage of the crusts. Additionally, we have added a detailed description of the BSCs present in the fields (lines 123-133 and 136-149). The added description suggests that the developmental process of the BSCs cultivated in the microcosm experiment resembled the one of the analogous BSCs present in the fields. Therefore, one could assume that the same species of mosses, fungi, algae and lichens that were identified in the fields were either already present in the cultivated BSCs, or would show in the later stages of development of those BSCs.

Comment 20: Line 214. Mostly or only?

Response 20: To avoid confusion, we have decided to use the word “mostly”, as it better conveys the message of the referred sentence (line 302).

Comment 21: Line 215. Pictures instead of details.

Response 21: The referred sentence has been changed in accordance with the reviewer’s suggestion (line 303).

Comment 22: Lines 218-225. To demonstrate soil chemistry better, I would suggest to prepare a table or graph with studied chemical parameters and how they follow the BSC development.

Response 22: Following the reviewer’s suggestion, we have added a table showing the investigated parameters (Table 2).

Comment 23: Lines 234-253. I find that this paragraph is very confusing and partially could be moved to the Methods section.

Response 23: We agree that the referred paragraph contained a lot of technical details that were more suitable for the Materials and Methods section. Following the reviewer's suggestion, the description of testing the accuracy of annotations has been moved to the Materials and Methods section (lines 235-243), and only the basic characteristics of the sequencing data have been left in the Results section (lines 321-330), and were additionally separated into two paragraphs for better understanding.

Comment 24: Line 237. (2 soil substrates x 3 sampling time points x 3 independent replicates) - it has been already present in methods. Delete it from the results section.

Response 24: The referred sentence has been changed in accordance with the reviewer's suggestion (lines 321-322).

Comment 25: Line 239. "number" instead of "amount"

Response 25: The referred sentence has been changed in accordance with the reviewer's suggestion (line 323).

Comment 26: Line 240. 120 - 250 bp is very short. Did you expect such short reads?

Response 26: The maximum length of raw reads that can be obtained using MiSeq is 300 bp. Therefore, after removal of remnant adaptor sequences and trimming of terminal nucleotides with low Phred quality scores, the mean read length of 250 bp, which characterized the samples from T1 and T2, is normal. However, we agree that the mean read length of 120 bp, which characterized the samples from T0, is rather short. The shorter read length of the samples from T0 resulted from different molecular manipulations that these samples underwent during library preparation compared to the samples from T1 and T2. The library preparation procedure is described in detail in the manuscript (lines 212-222). Specifically, no size selection was performed for samples from T0 as recommended in the protocol from Illumina for samples with low DNA concentrations (Table 2). Consequently, the metagenomic libraries of the samples from T0 were mixtures of different fragment sizes. As shorter fragments have priority over longer fragments during sequencing with MiSeq, the rather short mean read length of samples from T0 was expected. To assure that the read length did not affect the accuracy of bacterial annotations, we compared the taxonomic annotations of the metagenomes with long reads before and after *in-silico* trimming them to resemble the metagenomes with short reads. This approach is described in detail in the manuscript (lines 235-243).

Comment 27: Lines 242-249. Were all samples trimmed to 120 reads?

Response 27: We did not trim all reads to 120 bp. However, as we shared the concern about different reads lengths, we confirmed that the read length did not affect the accuracy of bacterial annotations by comparing the taxonomic annotations of the metagenomes with long reads before and after *in-silico* trimming them to resemble the metagenomes with short reads. This approach is described in detail in the manuscript (lines 235-243). As the taxonomic annotations were not notably biased by the read length,

further analyses were performed on the metagenomes with original read lengths, which is communicated in the manuscript (lines 241-243).

Comment 28: Lines 257-258. Did these reads belong to 16S rRNA gene sequences? Could you add this information.

Response 28: The National Center for Biotechnology Information Non-Redundant (NCBI-NR) protein sequences database does not include 16S rRNA gene sequences. All taxonomic results presented in the manuscript are based on annotations of the whole metagenomic datasets using the NCBI-NR database, while 16S rRNA gene sequences were extracted from the metagenomic datasets (they comprised 0.0062 % of all metagenomic reads) and annotated with SILVA SSU database only to confirm these results. We tried to make this clearer in the referred sentence as well as in other parts of the revised manuscript to avoid confusion (lines 246-248, 333-334, 356-360).

Comment 29: Line 263. to their total relative abundance of all metagenomes

Response 29: The referred sentence has been changed in accordance with the reviewer's suggestion (line 340).

Comment 30: Line 268. "And" instead of "but"

Response 30: In the referred sentence, "but" signifies opposition. We agree that it could be replaced with "and", but we fear that such replacement would distort the meaning of the referred sentence. Therefore, we have not made the replacement.

Comment 31: Line 270. Delete "additionally"

Response 31: The referred sentence has been changed in accordance with the reviewer's suggestion (line 347).

Comment 32: Lines 273-274. Cyanobacteria, including Leptolyngbyaceae, Tolypothrichaceae and Nostocaceae, were abundant in BSC from Chicken Creek at T1 and T2, while Oscillatoriaceae and Microcoleaceae dominated at T1.

Response 32: The referred sentence has been changed in accordance with the reviewer's suggestion (lines 350-352).

Comment 33: Line 280. Such a little percentage? Was it on a species level? Or is it a number based on all metagenomes? It is not clear.

Response 33: The referred percentage (0.0062 %) is a ratio of reads assigned to the bacterial 16S rRNA gene to all metagenomic reads. We have changed the referred sentence accordingly (line 357).

Comment 34: Lines 288-291. Indicate the name of genes in brackets.

Response 34: In the referred fragment, we describe the abundances of COG functional categories and not of genes. As each category may contain thousands of genes, it is not possible to list the names of the genes included in them. Therefore, we have revised the referred fragment, avoiding mentioning genes, and underlining that we refer to COG functional categories, to prevent confusion (lines 363-368).

Comment 35: Line 292. All above genes or only genes involved in cell motility, and formation of extracellular structures?

Response 35: We have improved the referred sentence and clarified that we mean only the two categories “cell motility” and “extracellular structures” (lines 368-369).

Comment 36: Lines 296-299. This should be in Methods section.

Response 36: The referred sentence has been moved to the Materials and Methods section in accordance with the reviewer’s suggestion (lines 261-264).

Comment 37: Lines 300-308. No need to explain the function of the genes in the results. It is clear from the Table 1.

Response 37: The functions of the genes have been removed from the Results section in accordance with the reviewer’s suggestion (lines 373-378).

Comment 38: Lines 301-305. According to Figure 2, it is not true. For example, abundance of *wza* in T0 is less than 0.002, or *kpsE* is less than 0.001 in T1 and T2.

Response 38: We agree that the previous version of the referred sentence, which stated that the relative abundance of the key genes ranged between 0.002 % and 0.005 %, might have been confusing and stimulated the idea that we refer to the relative abundance of the genes in subsets of the data, as shown in Figure 2. What we actually refer to, is the overall relative abundance of the key genes in all metagenomes. Therefore, the presented percentages are correct. This has been now clarified in the manuscript (lines 373-374).

Comment 39: Lines 309-311. The sentence is hard to understand. Rather divide it into two separate sentences.

Response 39: The referred sentence has been split into two sentences in accordance with the reviewer’s suggestion (lines 379-381).

Comment 40: Lines 315-316. Was it significant? Does not seem significant from Fig. 2

Response 40: We double-checked the results, and the influence of location on the relative abundances of *kpsE* and *lptG* turned out to be indeed significant. We agree that this significance is not striking in Fig.2, although it is still visible. The explanation for that can be found in Supplementary material 1: Table S5 that contains significance levels and effect sizes for all investigated genes. While p-values for both genes were below 0.05, omega squared equaled 0.18 and 0.17 for *kpsE* and *lptG*, respectively. Omega squared was calculated as an effect size to estimate the magnitude of observed influences of the analyzed factors, and

can be interpreted as the percentage of variation in the dependent variable explained by the independent variable, as described in the manuscript (lines 276-278). Therefore, although the influence of location on both genes is significant, the magnitude of this influence is small and not as obvious as the influence of time, which yielded higher omega squared values (0.53 and 0.57 for *kpsE* and *lptG*, respectively).

Comment 41: Lines 317-318. "although *IptF* showed a decreasing trend in Lieberose at T1" - delete it. There were other trends, therefore I do not understand why you point out on this one particularly.

Response 41: The referred sentence has been deleted in accordance with the reviewer's suggestion.

Comment 42: Lines 322-328. Maybe you could create a Venn diagram to better illustrate numbers of shared families.

Response 42: A Venn diagram has been created in accordance with the reviewer's suggestion (Supplementary material 1: Fig. S8).

Comment 43: Discussion. Line 373. Write few examples

Response 43: *Burkholderiaceae*, *Comamonadaceae*, *Moraxellaceae* and *Flavobacteriaceae* are metabolically versatile and can degrade a wide range of compounds, such as various polymers, polycyclic aromatic compounds, phenols and halogenated aromatics. This information has been added to the manuscript (lines 440-442).

Comment 44: Line 380. Do you mean they prevailed in BSCs? If you mean they prevailed in comparison to other genes abundance, it is contradicting with Fig2, since their abundance is less than 0.01%.

Response 44: We agree that the previous version of the referred sentence, which stated that genes involved in the formation of EPSs and LPSs were generally more abundant in BSCs, might have been confusing and triggered the idea that we compare the abundance of these genes to other genes. What we actually compare is the abundance of genes involved in the formation of EPSs and LPSs in BSCs and bulk soils. This has been clarified in the revised manuscript (line 449).

Comment 45: Line 394-396. Do you have any idea why?

Response 45: We think that the accumulation of dissolved organic carbon, which corresponded to the increase of the relative abundance of genes related to polysaccharide production, as well as the visual entrapment of soil particles in the upper soil layer, suggests an increased production of adhesive bacterial polysaccharides in the studied BSCs. This has already been communicated in the original version of the manuscript. However, the reviewer's question indicates that we were not clear enough about this issue before. For this reason, the whole paragraph has been restructured and revised, partially also to include the references suggested by the reviewer 3. In the revision process, the referred paragraph has been separated into two paragraphs, and we hope that the current version presents our ideas more clearly (lines 452-485).

Comment 46: Line 410. What was the proportion of gram-negative bacterial reads in your study? And their proportion with LPS-producing genes?

Response 46: To answer the reviewer's first question, we have calculated the ratio of reads assigned to phyla considered as Gram-negative to all bacterial reads, and to answer the second question, we have calculated the ratio of reads assigned to the LPS genes investigated in this study to reads assigned to phyla considered as Gram-negative. It is important to mention that phyla with candidatus status (1.4 %) and bacterial reads not assigned at the phylum level (3.7 %) were not regarded as either Gram-negative or Gram-positive. In total, Gram-negative bacterial reads comprised 81.1 % of all bacterial reads, and LPS reads comprised 0.012 % of all Gram-negative reads. However, not all members of phyla considered as Gram-negative have the ability to produce LPSs, while some members of phyla considered as Gram-positive have been recently found to have the ability to produce LPSs as well. Moreover, the LPS genes investigated in this study are not the only genes involved in LPS production, although they are the most conserved and specific ones. Therefore, we have decided that the values calculated on the reviewer's request are too biased to be included in the manuscript. However, the information about LPS-producing bacteria that we found while searching for a way to answer the reviewer's questions have been used to improve the discussion on the abundance of LPS genes in the revised manuscript (lines 491-495.)

Comment 47: Lines 415-418. Make shorter sentences. In the current state this sentence is too complex and confusing.

Response 47: We agree that the referred sentence was too complex and confusing, and thus it has been split into three separate sentences in accordance with the reviewer's suggestion (lines 497-501).

Comment 48: Lines 427-428. Why was it of secondary importance? Not clear.

Response 48: The referred sentence was part of the Discussion section that has been revised (510-528) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. In the revision process, the referred sentence has been deleted.

Comment 49: Moreover, you cannot talk about edaphic parameters in general, since you measured only pH, DOC and DON.

Response 49: We agree with the reviewer and have revised the manuscript accordingly. In the previous version of the manuscript, the possible influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs on two different soil substrates was discussed broadly. However, demonstrating that edaphic parameters are the main driver of the differentiation was not the initial aim of this study. In the first place, the development of BSCs from indigenous communities of free-living microbes, even though these communities seemed highly similar, made it impossible to distinguish whether the differentiation of bacterial communities was primarily triggered by edaphic variables, or resulted from various rare species present in the indigenous communities of the soil substrates. For this reason, we measured only a few basic soil parameters, and did not attempt to perform any multivariate analyses. In the current version of the manuscript, we focus on the main aim of the study, which was to investigate if different types of bacterial BSCs harbor distinct potential producers of EPSs and LPSs. Large parts of the manuscript have been either deleted or revised in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial

development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. The possible influence of edaphic parameters is now mentioned only briefly as one of the possible influences on the potential EPS and LPS producers of bacterial BSCs that needs to be addressed in future studies (lines 554-559, 578-581).

Comment 50: Line 438. Which edaphic parameters? You studied only pH, DOC and DON.

Response 50: The referred sentence was part of the discussion that has been revised (lines 510-528) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. In the revision process, the referred sentence has been deleted.

Comment 51: Line 453-455. You cannot claim with confidence that low abundance of cyanobacteria indicates presence of algae. You did not study that, therefore you might just speculate.

Response 51: We agree with the reviewer and have changed the referred sentence to underline that we are just speculating (lines 544-546).

Comment 52: Line 457. "In addition" instead of "In conclusion". Conclusion is the next paragraph.

Response 52: The referred sentence has been revised to summarize the paragraph (lines 550-553), while the final conclusion has been moved to the Conclusions section, as suggested by the reviewer (lines 576-578).

Comment 53: Conclusion. Line 466. It does not show it, since you studied only few edaphic parameters.

Response 53: The referred sentence was part of the Conclusions section that has been revised (lines 567-581) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. In the revision process, the referred sentence has been deleted.

Comment 54: Line 468. You can only say that pH seems to influence, but not that it is primary edaphic factor for EPS and LPS producers.

Response 54: The referred sentence was part of the Conclusions section that has been revised (lines 567-581) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and

LPSs. In the revision process, the referred sentence has been deleted from the Conclusions section. Now the possible influence of pH is mentioned only briefly in the Discussion section (lines 555-557).

Comment 55: Line 474. I do not find it surprising. As you explain in the next sentence, some genes might not be expressed.

Response 55: The referred sentence was part of the Conclusions section that has been revised (lines 567-581) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. In the revision process, the referred sentence has been deleted

Comment 56: Figures. Leave only the legend for the figures. The explanations should be placed to the Method section.

Response 56: Figure legends have been changed in accordance with the reviewer's suggestion (lines 886-898).

Comment 57: Figure 1. "...development of BSC based PCoA: a - overall community, and b - genes related to EPS and LPS synthesis."

Response 57: The referred figure legend has been shortened as suggested by the reviewer. However, a different wording has been used, as we believe that the one suggested by the reviewer implies that plot b shows the abundances of genes related to EPS and LPS synthesis instead of the taxonomic affiliation of genes related to EPS and LPS formation (line 886-888).

Comment 58: Figure 2. Why are *sacB* and *kpsE* in grey colour? Would be better to illustrate genes in reverse sequence: T0 → T1 → T2. Blue colours in Lieberose T0 and T1 are very hard to distinguish.

Response 58: We used gray coloring to highlight that the genes *sacB* and *kpsE* belong to different EPS assembly and export pathways. However, we admit that this might have caused confusion. Therefore, the different EPS assembly and export pathways are now separated by dashed lines. Moreover, the referred graph already illustrates genes in the sequence requested by the reviewer, from the top of the graph to the bottom. Therefore, we assume that the reviewer would prefer to see the aforementioned sequence from the bottom of the graph to the top. However, we believe that it is more intuitive to read graphs from the top to the bottom, and thus we have not changed the sequence. Thirdly, we had chosen the color palette that is colorblind-friendly, and thus we have decided not to change the basic colors. However, we have adjusted the intensity of the colors to make them more distinguishable.

Comment 59: Figure 3. In my opinion, the "Absent" columns are useless and make the figure too big and complex. If you leave only potential producers, it will be enough to understand that other bacteria are not potential producers.

Response 59: We believe that the “Absent” columns make the comparison of the amount and relative abundance of potential polysaccharide-producers to non-producers as well as following their dynamics easier. Furthermore, these columns serve the purpose of showing that the taxonomic composition of potential polysaccharide-producers changes drastically during the initial development of BSCs as opposed to the non-producers. Finally, the “Absent” columns illustrate that not all members of BSCs have the potential to produce EPSs and LPSs, as could be falsely assumed by some readers (compare with reviewer 2, comment 7). Therefore, we have decided to leave the referred graph unchanged.

Comment 60: Figure 4. Something like this: "... formation in BSC and bulk soils based on their relative abundances. Note the different colour intensities between bacterial and gene abundances."

Response 60: The referred figure legend has been changed accordingly (lines 896-898).

## Reviewer #2

Comment 1: The title is a little confusing. Did the biological soil crusts grown on both soil substrates favour the establishment of bacteria with potential to produce exopolysaccharides and lipopolysaccharides? Or was one substrate more benefit to the establishment?

Response 1: We agree that the previous title was confusing and raised questions that we did not intended to answer through our study. Therefore, the current title summarizes the main finding of the study, which is that biological soil crusts grown on different soil substrates harbor distinct communities of bacteria with potential to produce exopolysaccharides and lipopolysaccharides.

Comment 2: At the end of Abstract, the authors concluded that biological soil crusts harbour a high number of potential producers of adhesive polysaccharides. Could the results supported this conclusion? Did the authors compared the amount of bacteria between bulk soil and biological soil crusts? I think they just got the relative abundance results.

Response 2: The abstract has been rewritten as suggested by reviewer 3 (comment 6). In the process of rewriting, the referred sentence has been deleted.

Comment 3: In the manuscript, the authors should clearly clarify what are biological soil crusts. Are they occur as a transient phase during the plant-free season? Did reference 2 describe this phenomenon? Did real crusts develop in the cultivation experiment? Or just cyanobacteria and mosses growing on soil surface?

Response 3: We agree that the previous version of the manuscript was not completely clear on what BSCs are. In the revised version of the manuscript, we clarify in the Introduction section that BSCs consist of organisms that live in a close association with soil particles, forming a coherent layer within the uppermost few millimeters of the topsoil, or directly on the soil surface (lines 60-61). We also explain that bacteria highly colonize BSCs in the initial stage of BSC development. They are the first colonizers of bare soils, and EPSs and LPSs produced by them stimulate the consolidation of soil particles in preparation for the establishment of cryptogamic surface cover that appears during later stages of BSC development (lines 91-94). Therefore, as long as they form a coherent layer within the topsoil, bacterial crusts are still “real” BSCs even though their structure and properties differ from mature cryptogamic BSCs. For this reason,

we now use the term “initial BSCs” throughout the manuscript to distinguish the BSCs that developed in the macrocosm experiment from mature BSCs. To further address the reviewer’s question if real crusts developed in the cultivation experiment, we now state clearly in the Results section that BSCs developed in the microcosm experiment were in the initial stage of development. Moreover, we describe more explicitly that the observed bacterial biofilms enmeshed soil particles and formed coherent patches on the soil surface (lines 298-300). To support this, we have added more pictures of soil surface to Supplementary material 1: Fig. S4. Further evidence is provided by the CT pictures of 10-months-old samples from Chicken Creek that show the entrapment of small soil particles at the surface (Supplementary material 1: Fig. S5).

Comment 4: At the end of first paragraph of Introduction, the authors describe the best studied producers of polysaccharides are cyanobacteria and algae, but also some others. Have all of them been well studied about the polysaccharide producing?

Response 4: The non-photosynthetic microbial members of BSCs, including fungi, proteobacteria and actinobacteria, are also prominent producers of polysaccharides, but they are not as thoroughly studied as cyanobacteria and algae. This information has been added to the revised manuscript (line 67).

Comment 5: At the end of second paragraph of introduction, the references should be given at the end of each point, including the developmental stage of BSC, environmental and edaphic factors.

Response 5: We have corrected the references in accordance with the reviewer’s suggestion (lines 77-79).

Comment 6: Are there biological soil crusts in the sampling sites? Again, did real crusts develop in the cultivation experiment?

Response 6: We agree with the reviewer that the information if BSCs occur at the sites from which the soil used in the microcosm experiment originated, is essential to understand why we expected that different types of BSCs would develop in the microcosm experiment. Therefore, the information that BSCs occur at both sampling sites, together with the detailed description of the BSCs there, has been added to the Materials and Methods section (lines 123-133 and 136-149). The added description shows that the initial BSCs developed in the microcosm experiment resembled their naturally occurring analogs at the respective sites.

Comment 7: Actually, all the microorganisms in biological soil crusts could produce polysaccharide, so what is the meaning of potential to produce exopolysaccharides and lipopolysaccharides? If all the microorganisms can produce exopolysaccharides and lipopolysaccharides, could we understand the results as bacterial composition shift during the cultivation process? Just different bacterial community produced the exopolysaccharides and lipopolysaccharides at different stages of cultivation.

Response 7: The reviewer seems to be under the impression that all microorganisms in BSCs could produce EPSs and LPSs. It is true that many BSC organisms produce these compounds, but not all of them, as explained for example by Rossi et al. (2018). Even in this manuscript, Figure 3 shows that the bacterial communities building initial BSCs can be divided into potential polysaccharide-producers and non-producers. To avoid misunderstandings, we now clearly state in the Introduction section that not all members of BSCs have the ability to produce polysaccharides (line 80). However, we agree with the

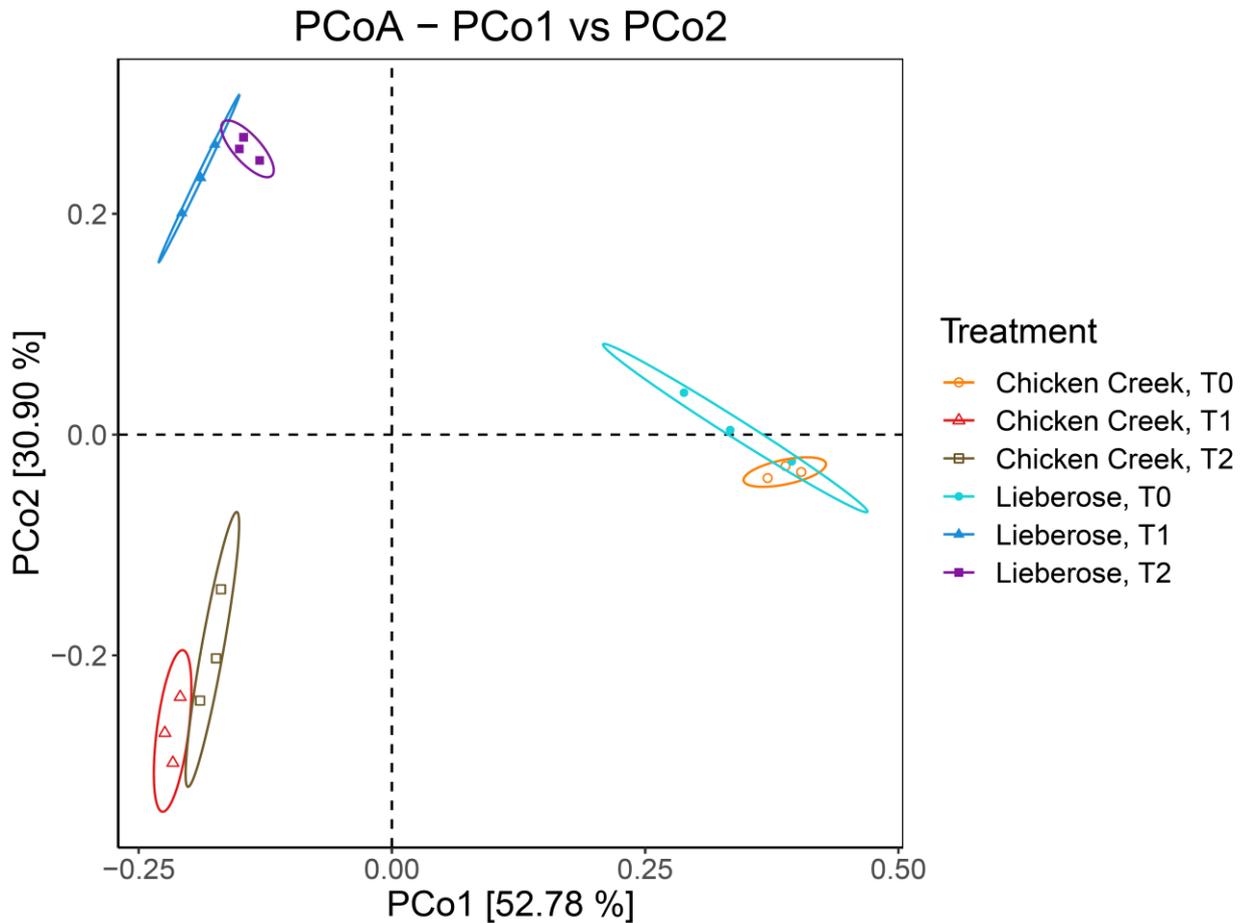
reviewer that EPS and LPS producers undergo a compositional shift during the development of BSCs, as has been shown in the manuscript. Furthermore, different types of initial BSCs harbor distinct producers of EPSs and LPSs. In the Introduction section, we now explain that polysaccharides produced by different taxa can have different adhesive properties (lines 71-77). In the Conclusions section, we state that whether the ability of BSCs to improve soil development in the long term is compromised by differences in the efficiency of polysaccharide formation, or the adhesive properties of EPSs and LPSs produced by different taxa, needs further investigation (lines 576-578).

### **Reviewer #3**

Comment 1: The authors took samples from two distinct habitats, an artificial catchment and a moving dune. Despite a similar composition pattern of soil bacteria at the family level, these soils dominated by different species is expectable. Although abiotic factors were well studied and were supposed to shape bacterial assemblage, the species pool of local sites is also a key parameter on the dynamic of community composition. Each of two soils certainly consisted of adaptable bacteria regarding to the environment. In my view, using a bit of mixed fresh soils from both sites, then inoculating the mixed soil on the sterile substrates, respectively, might be a better choice. I understand this cannot be fixed easily, but if the composition pattern of bacteria at the higher taxonomic level, for instance genus or OTU, could be demonstrated to be highly similar, it may reduce the dispute. After all, it's possible that the differentiation of the taxonomic composition arises from that various rare species in the beginning dominate during the later development.

Response 1: We would like to thank the reviewer for this and the following two comments, which gave us a clearer idea what were the weakest points of the storyline in the previous version of the manuscript, and how to present our main messages better in the new version of the manuscript. We believe that in the previous version of the manuscript, we discussed too broadly the possible influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs on two different soil substrates. This gave the wrong impression that demonstrating that edaphic parameters are the main driver of the observed differentiation was the initial aim of the study. If that was the case, we agree that the study design proposed by the reviewer would be more appropriate than the current one. However, the study was designed to investigate if different types of bacterial BSCs harbor distinct potential producers of EPSs and LPSs. In order to be able to follow the dynamics of bacterial communities in developing BSCs, and compare BSCs of the same age, we cultivated BSCs in a microcosm experiment. We cultivated them on unprocessed soil originating from sites where different types of BSCs occur naturally to increase the chance that different types of BSCs would develop also in the microcosm experiment. The high similarity of bacterial communities in the initial soils was not planned, and has been discussed in detail already in the previous version of the manuscript. Furthermore, we agree with the reviewer that the current study design makes it impossible to distinguish whether the differentiation of bacterial communities observed once BSCs started to develop was primarily triggered by edaphic variables, or resulted from various rare species present in the indigenous communities of the soil substrates. We believe that this holds true even though the communities showed highly similar patterns also at the species level (see the following figure), as the coverage of the microbial diversity by the metagenomic datasets from T0 equaled  $41.9 \pm 12.8 \%$  (Supplementary material 1: Fig. S1). Therefore, large parts of the manuscript have been either deleted or revised in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSC. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. Edaphic parameters are mentioned

now only briefly together with rare species as possible influences on the potential EPS and LPS producers of bacterial BSCs that need to be addressed in future studies (lines 554-564, 578-581).



Comment 2: The definition of BSCs in this study is one of another major concerns. Is that all topsoil with bacterial communities that producing extracellular polymeric matrix can be called as BSCs? Namely, was there a coherent layer with an intimate association between soil particles and microorganisms already formed? I prefer to call this as the pre-phase before the crust forming, or the preliminary stage of BSC. I think it's important because it's the EPS/LPS producers that construct the crustal structure, and it will be confused as claimed in the title that BSC favors its builder...It's like the logical relation between eggs and hen. I suggest to check on this issue throughout the manuscript and to make the story line around the bacterial assembling during the preliminary phase of BSC forming.

Response 2: We agree that the previous version of the manuscript was not completely clear on what BSCs are. In the revised manuscript, we clarify in the Introduction section that BSCs consist of organisms that live in a close association with soil particles, forming a coherent layer within the uppermost few millimeters of the topsoil, or directly on the soil surface (lines 60-61). We also explain that bacteria highly populate BSCs in the initial stage of BSC development. They are the first colonizers of bare soils, and EPSs and LPSs produced by them stimulate the consolidation of soil particles in preparation for the establishment of cryptogamic surface cover that appears during later stages of BSC development (lines 91-94). Therefore, as long as they form a coherent layer within the topsoil, bacterial crusts are still “real” BSCs even though their structure and properties differ from mature cryptogamic BSCs. For this reason, we now use the term “initial BSCs” throughout the revised manuscript to distinguish the BSCs that developed in the macrocosm experiment from mature BSCs. We now also state clearly in the Results

section that BSCs developed in the microcosm experiment were in the initial stage of development. Moreover, we describe more explicitly that the observed bacterial biofilms enmeshed soil particles and formed coherent patches on the soil surface (298-300). To support this, we have added more pictures of soil surface to Supplementary material 1: Fig. S4.

Comment 3: The physicochemical analyses on the samples from different habitats were insufficient and fragmental. If the authors attempt to demonstrate that soil variables are the main driver on the differentiation of bacterial communities between two experimental sets, these abiotic factors need to be investigated in detail, rather than putting them all into a black box. For example, the multivariate regression is one of the useful tools to handle the variables. Also, the set of only DOC, DON, and soil pH is limited to describe the nature of soil. Particle size, water-holding capacity, salinity, as well as TOC, are crucial parameters. Furthermore, the analysis of pore structure of the upper layer by CT is a solid evidence to depict the development of BSCs, but I don't understand why only sample of T2 from Chicken Creek was measured. It made the CT analysis meaningless because no dynamic change of the pore structure was elucidated. Is there any solution when the thickness of crust is low?

Response 3: We fully agree with the reviewer and encourage in the revised manuscript further experiments designed to investigate the influence of soil variables on the potential EPS and LPS producers in bacterial BSCs (lines 562-564). However, we believe that such experiments should involve multiple sterile soil substrates with diverse edaphic parameters, inoculated with the same initial bacterial community. Therefore, the current study does not meet the conditions to make any definite conclusions on the influence of edaphic parameters on the community structure of potential polysaccharide producers. In any case, this was not the aim of the study, as explained in detail in the answer to comment 1 of the reviewer. For this reason, we measured only a few basic soil parameters, and did not attempt to perform any multivariate analyses. Similarly, only the results of the analysis of pore structure of the upper layer by CT for samples of T2 from Chicken Creek were included in the manuscript because these were the only samples thick enough that we were able to measure with the techniques that we had at hand. BSCs in other samples were thinner, and thus not detectable in the CT images. In fact, the image processing technique used in this study has been already improved compared to the original technique by Schlüter et al. (2010), but would need to be improved further to analyze even thinner crusts. In the meanwhile, we have decided to include the results of the analysis as a supporting information.

Comment 4: The phylogenetic names at the family level should be the regular font. Only the genus' names are Italic, please check it thoroughly. Most importantly, please code the number of lines and pages before submitting to the journal! It made the reviewing very hard...

Response 4: We would like to apologize for forgetting to code the line numbers before submitting to the journal. In the revised manuscript, the line numbers have been set to continuous. However, we did not change the formatting of bacterial names, as the guidelines of the American Society for Microbiology and the Journal of Bacteriology state that the names of all microbial taxa (from kingdom to subspecies) should be italicized.

Comment 5: The title needs to be changed, the current one is confusing.

Response 5: We agree that the previous title was confusing. Therefore, the current title summarizes in a simple way the main finding of the study, which is that biological soil crusts grown on different soil

substrates harbor distinct communities of bacteria with potential to produce exopolysaccharides and lipopolysaccharides.

Comment 6: The abstract needs to be improved. "However, ...", what's meant by "the potential"? The changes of EPS/LPS producers? The scientific question needs to be straightforward and clear.

Response 6: We agree that the previous abstract needed improvement. Furthermore, it would have not been suitable for the new version of the manuscript. In the reviewing process, the sentence that confused the reviewer has been deleted. The introductory part has been made more focused on the bacterial members of initial BSC that improve soil stability through the production of EPSs and LPSs that "glue" soil particles together. The scientific question about the dynamics of these bacteria during the initial development of BSCs has been stated, followed by a clear and straightforward hypothesis that different types of initial BSCs harbor distinct potential producers of EPSs and LPSs. In the revised version, we only present the most essential results that support the final conclusion that the potential to form EPSs and LPSs is an important trait for initial BSCs, as the relative abundance of genes related to EPS and LPS production showed similar increasing trend in different types of initial BSCs despite the different community composition of their EPS and LPS producers.

Comment 7: Abstract, "Interestingly, ...", cryptic.

Response 7: The abstract has been rewritten, and the referred term has been deleted.

Comment 8: Introduction, "plant-free season", please use other words.

Response 8: The referred part of the Introduction section has been revised (lines 58-61), and the referred term has been deleted.

Comment 9: Line 5 in the first paragraph, please add a reference to "EPM".

Response 9: The reference has been added in accordance with the reviewer's request (line 63).

Comment 10: Page 4 line 5, "furthermore" instead of "overall", and add one or two references that demonstrated the effect of abiotic factors on EPM, as well as the relations between EPM composition and microbial composition.

Response 10: The referred paragraph has been partially revised to accommodate the reviewer's request (lines 74-79). Now it states that different taxa produce polysaccharides with different sugar composition. As the slightest differences in the sugar composition can result in completely different physical traits of the polysaccharide, the properties of EPM could be influenced by any factor that changes the structure of polysaccharide-producing communities. This is supported by appropriate references. The word "overall" has been deleted during the revision of this paragraph.

Comment 11: Page 4 line 7, "Thus, ..." confusing, what's the link with the previous sentence.

Response 11: We agree that the referred sentence in its previous form was not conclusive, and therefore the word "thus" was inappropriate and now has been deleted (lines 80-81).

Comment 12: Page 4 line 13, these pathways may not be the EPS/LPS-specific ones, how to elucidate the gene abundances in the pathway linked with the content of EPS/LPS. These three pathways all related to the excretion?

Response 12: As the reviewer suggested, the referred pathways are all related to excretion. In the original manuscript, we explained that these pathways are used to export polysaccharides after their intracellular synthesis. To avoid misunderstandings, we have replaced the word “export” with “excretion”, as suggested by the reviewer (line 84). Moreover, we agree with the reviewer that most genes related to EPS/LPS synthesis is involved also in the basic intracellular metabolism of sugars, and thus is not specific for EPS/LPS production. This problem has been described before by Cania et al. (2019) who established the bioinformatics pipeline employed in this study. They tested 81 genes related to EPS/LPS production, but only 14 of them were EPS/LPS-specific. In fact, most of the specific genes, which were used in the pipeline, were part of the aforementioned pathways. The exception was *sacB* responsible for the extracellular synthesis of levan. Therefore, by using the pipeline by Cania et al. (2019), we target only genes specific for EPS and LPS production, as we underline in the Introduction section (lines 108-110). However, it was not our intention to link gene abundances to EPS/LPS content, as suggested by the reviewer. Bacteria capable of producing EPSs and LPSs could perform this function with different efficiency, and regulation could take place on the transcription level. To make such a link, measuring the bacterial production of polysaccharides would be necessary, although difficult due to technical reasons that have been discussed by Rossi et al. (2018) and Redmile-Gordon et al. (2014) (lines 483-485). Therefore, throughout the revised manuscript, we use the word “potential” when referring to EPS/LPS production.

Comment 13: I think the introduction section did not show the advantage of metagenomic data and why the author employed it, please add a brief paragraph to claim it.

Response 13: Information on the advantages of metagenomic analysis for the study has been added in accordance with the reviewer’s request (lines 97-99).

Comment 14: Page 4 line 18, synthesis? excretion?

Response 14: We agree that the referred sentence in the previous form could have been confusing, and thus we have specified that we mean the relative abundance of genes related to EPS and LPS formation (line 101).

Comment 15: Page 4 line 20, it's an interesting hypothesis, but please mention the effect of soil variables on microbial assembling earlier.

Response 15: It has been now specified in an earlier paragraph that, among other, soil variables like soil pH, texture and nutrient content influence the composition of BSC organisms (lines 77-79).

Comment 16: Page 4 line 23, the underlying soil might not be the only source of variability.

Response 16: The referred sentence was part of the Introduction section that has been revised (lines 99-106) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial

communities during the initial development of BSCs. Instead, the focus is now on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs. During the process of revision, the referred sentence has been deleted

Comment 17: Page 5 line 1, need to be more careful when using the term BSC.

Response 17: We agree with the reviewer and now use the terms “initial BSCs” throughout the revised manuscript to distinguish the BSCs that developed in the macrocosm experiment from mature BSCs.

Comment 18: Page 5 line 2, "targeted"?

Response 18: The referred sentence has been rewritten and now explains that the employed bioinformatics pipeline was targeting genes specific for EPS and LPS production (lines 108-110).

Comment 19: Page 5 line 8, moving dune.

Response 19: The referred sentence has been changed in accordance with the reviewer’s suggestion (line 116).

Comment 20: Page 5 line 17, if the soil texture of two sites is similar, how to conclude that the soil variables lead to the differentiation of microbial community.

Response 20: According to Fierer et al. (2017), soil texture has less importance in structuring soil bacterial communities than variables such as nitrogen and phosphorus availability, soil moisture availability, soil [O<sub>2</sub>] and redox status, organic carbon quality and quantity, and most of all pH. However, we agree that the experimental design of the current study prevents us from making any definite conclusions on the influence of edaphic parameters on the bacterial community structure. Therefore, large parts of the manuscript have been either deleted or revised.

Comment 21: Page 6 line 5, what was the temperature and how long time of the storage until the experiment.

Response 21: The soil was transported and afterwards stored at room temperature in the dark for approximately six months before the incubation experiment. During that time, pre-experiments to adjust the incubation conditions for BSC growth were performed. This information has been added to the revised manuscript (lines 155-158).

Comment 22: Page 7 line 17, please briefly introduce how water repellency was measured and what's the unit of it.

Response 22: The basic idea behind the method is that infiltration of a non-polar liquid into the surface of a porous body is determined solely by the physical properties of that liquid and the pore dimensions, whereas infiltration of a polar liquid in addition is determined by polar interactions with the polar surfaces of mineral particles. Hence, infiltration data of two liquids with differing polarity can be used to

characterize the surface polarity, where non-polar surfaces won't wet with (polar) water. This approach very much resembles contact angle measurements to determine surface energy, but, due to infiltration, only supercritical contact angles can be determined for soils. The repellency index as used in this study allows to characterize surface wettability in situations when water drops won't persist at the surface long enough to determine contact angles or to measure the water drop penetration time. The term „repellency index” is somewhat misleading, because it implies pronounced repellency where water drops persevere on the surface. In fact, in case of pronounced repellency (or zero water infiltration) it cannot be determined, because the water infiltration term is in the denominator of the equation. However, we continue to use the term „repellency index” to characterize wettability because it is common. The “repellency index” has no dimension. In case of ideal wettability, it amounts to 1, where values exceeding 1 characterize less wettable (or more repellent) soils. A short description how water repellency was measured and what is the unit of it has been added to the manuscript (lines 190-192).

Comment 23: Page 8 line 3, "XCMT", when the abbreviation arises at the first time, please use the full spelling.

Response 23: The full spelling has been added in accordance with the reviewer's request (line 199).

Comment 24: Page 8 line 8, delete "Buffer SL1...".

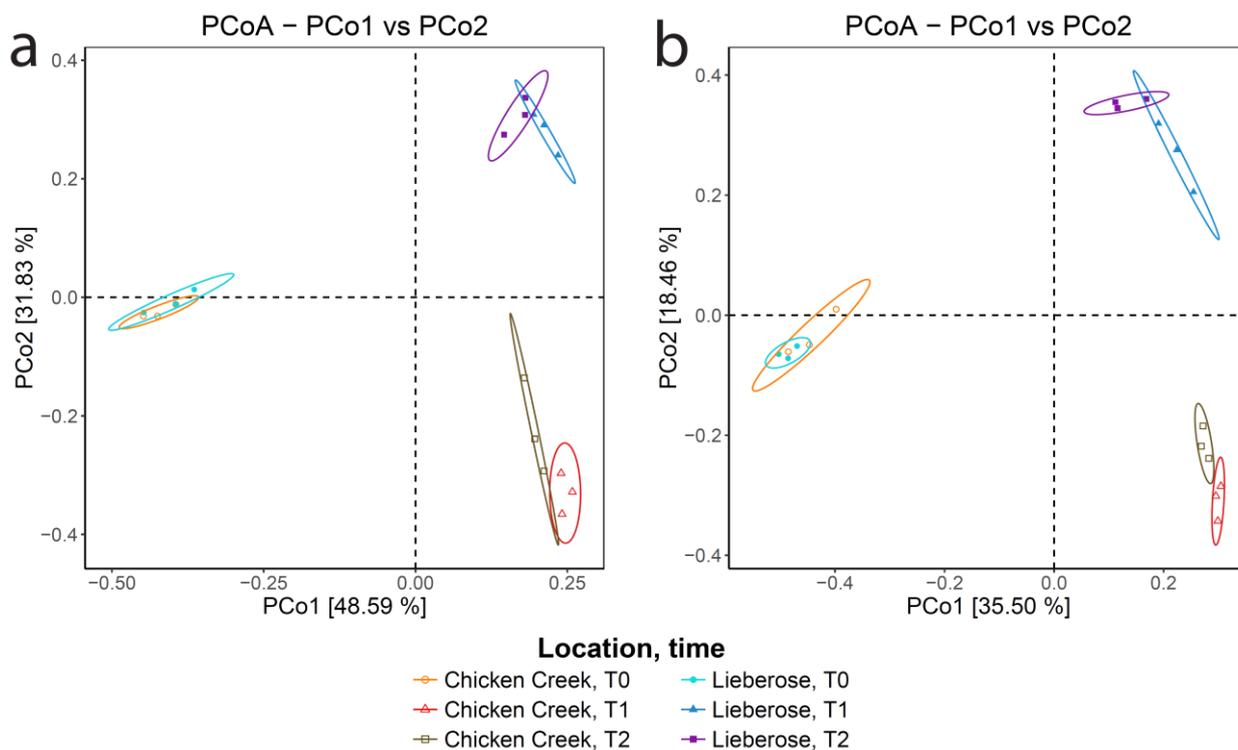
Response 24: The 'Genomic DNA from soil' NucleoSpin Soil Kit offers two lysis buffers, Buffer SL1 and Buffer SL2, but only one of them is used during the DNA extraction procedure. The manufacturer's protocol suggests that both lysis buffers should be tested in parallel for each new sample material to obtain best results in DNA yield and purity. In our study, the buffer that gave better results was Buffer SL1. Therefore, we believe that the information which specific buffer was used is potentially of interest to the readers, and thus we have decided not to delete the referred sentence. To avoid confusion, we have additionally specified in the revised manuscript that the Buffer SL1 was chosen based on a pretest performance (lines 205-206).

Comment 25: Page 9 line 10, I suggest to organize the supplementary files into two or three integrated appendixes, and list as figures and tables. This may benefit for the readers to understand.

Response 25: We agree with the reviewer, and thus we have joined all supplementary files into two supplementary files (Supplementary material 1 and 2 – one with figures and the other one with tables), in a similar manner to some recent publications in Microbial Ecology.

Comment 26: Page 10, line 19, use Jaccard's dissimilarity or NMDS method to proof the results by the Bray-Curtis matrix because the former two based on the presence of species or the ranking matrix of the quantitative data.

Response 26: As suggested by the reviewer, we used Jaccard dissimilarity to prove the results by the Bray-Curtis matrix (compare original Figure 1 based on Bray-Curtis distances with the following figure based on Jaccard distances). The two figures show the same pattern, and thus we have decided to keep in the manuscript only the original Figure 1, as Bray-Curtis distances are a more appropriate measure for community abundance data that was analyzed in this study. This has been specified in the revised manuscript (lines 281-282).



Comment 27: The result section, in my view, a constrained ordination that analysis the detail influences of each physicochemical parameters on the pattern of bacteria could be helpful. And please add a figure or table that containing the data of soil variables.

Response 27: Following the reviewer's request, we have added a table with studied parameters (Table 2). However, as explained in detail in the answers to comments 1 and 3 of the reviewer, analyzing the influences of physicochemical parameters on the pattern of bacteria was not the aim of this study, and thus only a few basic parameters were measured. We believe that an experiment aiming at determining the influence of soil variables on the pattern of bacteria should include multiple sterile soil substrates with diverse edaphic parameters, inoculated with the same initial bacterial community. Furthermore, as pointed out by the reviewer in comment 3, parameters such as particle size, water-holding capacity, salinity, as well as TOC could also be crucial for bacterial dynamics. Therefore, a constrained ordination using the parameters measured in this study would be strongly biased, and could lead to false conclusions. Therefore, we have decided not to include such an analysis in the revised manuscript.

Comment 28: Page 11 line 6, why the contents of EPS/LPS were not measured, they are direct evidences on the development.

Response 28: We agree with the reviewer that measuring the EPS and LPS content in the developing BSCs would be necessary to link the gene abundance with the actual EPS and LPS production. However, although the research on methods for extracting polysaccharides from BSCs progressed considerably in recent years, as summarized by Rossi et al. (2018), no common technique has been established yet. Furthermore, the existing methodologies still have biases, and, most importantly, they are not suitable for measuring polysaccharides produced specifically by the bacterial members of BSCs. Therefore, the data on the contents of EPSs/LPSs is missing. This information has been added in the revised manuscript (lines 483-485). Instead, the repellency index, which gives hints about the actual EPS/LPS content, as increasing EPS/LPS contents reduce water infiltration, has been measured and discussed (lines 465-483).

Comment 29: Page 11 line 11, add a new figure 1 to exhibit the map of sites and the photographs of the process of your experiment.

Response 29: As suggested by the reviewer, we have added more pictures of the sites and the BSCs cultivated in the microcosm experiment to Supplementary material 1: Fig. S4. However, we believe that the exact location of the sites is of secondary importance compared to the description of BSCs occurring there naturally, which has been also added in the revised version of the manuscript (lines 123-133 and 136-149). Furthermore, the location of the sites, as well as the experiential procedures are already described in detail in the Materials and Methods section, and we are convinced that these descriptions are sufficient to repeat our experiment.

Comment 30: Page 11 line 15, the sample from Lieberose developed more slowly, but posed higher DOC content?

Response 30: The trend described in the referred sentence was not statistically significant, and thus the referred sentence has been deleted to avoid misunderstandings. Furthermore, by the “slower development” we meant that mosses were visible later in the samples from Lieberose compared to the samples from Chicken Creek. However, the speed of development of BSC on both soils was comparable. The description of BSC development has been revised to avoid confusion (lines 298-304).

Comment 31: Page 11 line 19, what's the number indicate? The unit?

Response 31: A detailed answer to this comment has been already given in the answer to the comment 22 of the reviewer.

Comment 32: Page 12 line 2, if the substrate can be treated, why other samples were not measured?

Response 32: Pre-experiments indicated that only BSCs from T2 grown on substrate from Chicken Creek developed a thickness sufficient for visualization by computed tomography. Thus, only these samples were used to determine connectivity of the three-dimensional pore system of the BSCs and the underlying soil as described previously. This answer has been implemented into the manuscript (lines 193-196). For further details, please check the answer to the comment 3 of the reviewer.

Comment 33: Page 12 line 11, "GB" instead.

Response 33: We agree that the word gigabases can be abbreviated in the referred sentence. However, we also believe that the abbreviation “GB” suggested by the reviewer could mean in the referred context (GB of data) either gigabases or gigabytes, which could cause misunderstandings. Therefore, we have decided to use the abbreviation “Gbases” instead (line 322).

Comment 34: Page 12 line 12, "x106" instead of "mln".

Response 34: We agree that the word “million” is not commonly abbreviated in scientific writing. Therefore, we have decided not to abbreviate it in the referred sentence (line 324).

Comment 35: Page 13 line 16, "was".

Response 35: Scientific names of families, such as the referred "*Flavobacteriaceae*", are Latin, and not English. The Latin name *Flavobacteriaceae* refers to a group of organisms, and thus is considered to be plural. Therefore, stating that "*Flavobacteriaceae* were..." is correct.

Comment 36: Page 14 line 2, use " $\omega^2$ " directly.

Response 36: The referred sentence has been changed in accordance with the reviewer's suggestion (line 352).

Comment 37: Page 14 line 4, a cycle tree with a few layers with annotation can be more readable using the represent sequence of each family.

Response 37: The referred sentence states the number of families influenced by the origin of soil substrate, the incubation time and the interaction of those two factors. We would like to thank the reviewer for the suggestion to try to represent the sequences of each influenced family on a graph, but we believe that it would only make the message of that sentence less straightforward. Specifically, the aim of the referred sentence was to statistically support Figure 1, which shows the differentiation of bacterial communities. Moreover, the full list of impacted families is already provided in Supplementary material 2: Table S3. Therefore, we believe that creating another supplementary graph that would not provide any new information that could be discussed in the manuscript, is unnecessary.

Comment 38: Page 14 line 11, I suggest to add a global pattern of the gene dynamic and point out the genes involved in EPS/LPS pathway (a volcano plot?).

Response 38: We admit that the idea provided by the reviewer is interesting. Unfortunately, it is impossible to realize, as the analyzed genes involved in EPS/LPS production have been identified using a very hypothesis-driven bioinformatics pipeline. This approach have enabled us to identify genes specific for EPS/LPS production, but is not suitable to analyze global patterns of the gene dynamics. For this reason, in order to get an idea about the dynamics of important functions in our metagenomic data, we employed the eggNOG database-based pipeline instead of the EPS/LPS pipeline. However, as most reads (> 50 %) assigned to the EPSs/LPSs-related genes using the targeted pipeline were classified into the COG category "Function unknown", this study was based mainly on the hypothesis-driven approach designed for EPS/LPS genes, and the eggNOG pipeline was employed only to get a general overview of the data. This has been underlined in the revised manuscript (lines 261-264). Therefore, showing EPS/LPS genes in a global pattern of the gene dynamics is impossible.

Comment 39: Online source 12, a negative value of  $\omega^2$ ?

Response 39: We would like to thank the reviewer for this comment. As already stated in the manuscript, values of omega squared can be interpreted as the percentage of variation in the dependent variable explained by the independent variable. Therefore, the values of omega squared range from 0 to 1. Tunks (1978) explains that calculation can yield a negative value of omega squared, but in such case it is interpreted as 0. Therefore, the negative values of omega squared are instead presented as 0 in the

manuscript. The negative value of omega squared pointed out by the reviewer was thus a mistake and has been corrected in the revised manuscript.

Comment 40: Page 15 line 11, the fewest?

Response 40: Instead of stating that the least abundant genes had just a few annotated reads, we now disclose the gene abundance ( $\leq 0.0003\%$ ) (lines 377-378). Therefore, stating that the fewest reads were annotated to those genes is no longer necessary.

Comment 41: Page 16 line 1, clarify the sentence.

Response 41: The referred sentence now clarifies that 11 families were found harboring the investigated genes in samples originating from both locations, taken at all three sampling time points, instead of stating that 11 families harbored the investigated genes independently of soil substrate and incubation time (lines 390-391).

Comment 42: Page 16 line 14, is it Alpha-proteobacteria or Acidobacteria?

Response 42: We confirm *Acidobacteria* were meant, as is stated in the referred sentence that lists families characteristic for BSCs originating from Lieberose. *Alphaproteobacteria* were also highly abundant in BSCs originating from Lieberose, but they were highly abundant in BSCs originating from Chicken Creek as well, which makes them not characteristic for Lieberose only.

Comment 43: Please divide the discussion section into several subsections.

Response 43: The Discussion section has been divided into several subsections, in accordance with the reviewer's request.

Comment 44: Page 18 line 16, add more reference, e.g. Colica et al., 2014 SBB.

Response 44: The suggested references have been added (line 467).

Comment 45: Page 18 line 19, add Rossi et al. 2017 ESR as reference.

Response 45: The suggested reference has been added (line 458).

Comment 46: Page 18 line 23, inferred from X-ray data?

Response 46: We have clarified that the exemplary pictures of the ten-months-old samples from Chicken Creek were taken using X-ray computed microtomography, as suggested by the reviewer (line 459).

Comment 47: Page 19 line 10, some references need to be discussed: 1. Raanan et al., 2016 EM; 2. Felde et al., 2016 SBB; 3. Colica et al., 2015 JAP.

Response 47: The suggested references are now discussed in several places of the revised manuscript (lines 460-461, 468-469, 476-479).

Comment 48: Page 19 line 17, it's awkward using "this is not surprising", please indicate which taxa of bacteria that concerned in this study are gram-negative bacteria, e.g., cyanobacteria.

Response 48: We agree that the previous version of the referred fragment could have been considered superficial and awkward. Therefore, the referred fragment has been revised. In the current version, we properly discuss which Gram-negative bacteria were highly abundant in the study, and how their presence could be connected with the abundance of LPS genes (lines 491-495).

Comment 49: Page 20 line 14, confusing sentence.

Response 49: The referred sentence was part of the Discussion section that has been revised (lines 510-528) in order to take the focus off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSC. In the revision process, the referred sentence has been deleted.

Comment 50: Page 21 lines 6-7, add references.

Response 50: References have been added, as requested (lines 537, 538)

Comment 51: Page 21 lines 14-16, how? Chloroflexi is also phototrophs.

Response 51: We agree that *Chloroflexi* should have been mentioned in the referred sentence. This has been corrected in the revised version of the manuscript (line 545).

Comment 52: Page 21 line 16, delete "in conclusion".

Response 52: The referred sentence has been rewritten to summarize the paragraph, while the final conclusion has been moved to the Conclusions section (lines 576-578). In the process of rewriting, "in conclusion" has been deleted in accordance with the reviewer's suggestion (lines 550-553).

Comment 53: Page 21 line 20, move the last sentence to the conclusion section.

Response 53: The referred sentence has been moved to the Conclusions section, as suggested by the reviewer (lines 576-578).

Comment 54: The conclusion section should be rewritten.

Response 54: We agree that the previous version of the Conclusions section required rewriting. In the current version of the Conclusions section, the focus has been put off the influence of edaphic parameters on the differentiation of bacterial communities during the initial development of BSCs. Instead, the focus has been put on the importance of the potential to form EPSs and LPSs for bacterial BSCs, and the link

between the overall composition of bacterial communities of initial BSCs and the community structure of potential producers of EPSs and LPSs (lines 567-581).

Comment 55: Page 22 line 2, why multivariate analysis was not concerned.

Response 55: As explained already in the answers to the reviewer's comments 3 and 27, the design of the study makes it impossible to make any definite conclusions on the influence of edaphic parameters on the community structure of potential polysaccharide producers. In any case, this was not the aim of the study, as explained in detail in the answer to comment 1 of the reviewer. For this reason, we measured only a few basic soil parameters. As pointed out by the reviewer in comment 3, parameters such as particle size, water-holding capacity, salinity, as well as TOC could also be crucial for bacteria. Therefore, a multivariate analysis using the parameters measured in this study would be incomplete and strongly biased, and could lead to false conclusions. Therefore, we have decided not to include it in the revised manuscript.

Comment 56: Page 22 line 9, the crustal layer may be not exactly formed...

Response 56: We agree that the structure of BSCs changes during their development, and while initial BSCs dominated by bacteria still form a layer within the top millimeters of topsoil, the crustal layer is more obvious in the later stages of BSC development, when they are dominated by mosses and lichens. For this reason, we now use the terms "initial BSCs" throughout the revised manuscript to distinguish the BSCs that developed in the macrocosm experiment from mature BSCs.

Comment 57: The last part of the conclusion seems like cryptic and meaningless discussion, rather than conclusion.

Response 57: As the Conclusions section has been rewritten, the referred part has been deleted.