Received: 29 April 2023 | Accepted: 8 August 2023

DOI: 10.1111/1462-2920.16497

RESEARCH ARTICLE

ENVIRONMENTAL MICROBIOLOGY Applied Microbiology International

Microbial dynamics in soils of the Damma glacier forefield show succession in the functional genetic potential

Maomao Feng[1,2] | **Gilda Varliero**[1] | **Weihong Qi**[3,4] | **Beat Stierli**[1] | **Arwyn Edwards**[5] | **Serina Robinson**[6] | **Marcel G. A. van der Heijden**[2,7] | **Beat Frey**[1]

[1]Rhizosphere Processes Group, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

[2]Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

[3]Functional Genomics Center Zurich, ETH Zurich and University of Zurich, Zurich, Switzerland

[4]Swiss Institute of Bioinformatics SIB, Geneva, Switzerland

[5]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK

[6]Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Dübendorf, Switzerland

[7]Plant-Soil Interactions, Agroscope, Zurich, Switzerland

Correspondence
Beat Frey, Rhizosphere Processes Group, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland.
Email: beat.frey@wsl.ch

Funding information
Chinese Government Scholarship, Grant/Award Number: 201904910436; Horizon 2020 Framework Programme (Marie Sklodowska Curie Actions program), Grant/Award Number: 675546; Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Grant/Award Number: 5233.00388.001.01

Abstract

Glacier retreat is a visible consequence of climate change worldwide. Although taxonomic change of the soil microbiomes in glacier forefields have been widely documented, how microbial genetic potential changes along succession is little known. Here, we used shotgun metagenomics to analyse whether the soil microbial genetic potential differed between four stages of soil development (SSD) sampled along three transects in the Damma glacier forefield (Switzerland). The SSDs were characterized by an increasing vegetation cover, from barren soil, to biological soil crust, to sparsely vegetated soil and finally to vegetated soil. Results suggested that SSD significantly influenced microbial genetic potential, with the lowest functional diversity surprisingly occurring in the vegetated soils. Overall, carbohydrate metabolism and secondary metabolite biosynthesis genes overrepresented in vegetated soils, which could be partly attributed to plant–soil feedbacks. For C degradation, glycoside hydrolase genes enriched in vegetated soils, while auxiliary activity and carbohydrate esterases genes overrepresented in barren soils, suggested high labile C degradation potential in vegetated, and high recalcitrant C degradation potential in barren soils. For N-cycling, organic N degradation and synthesis genes dominated along succession, and gene families involved in nitrification were overrepresented in barren soils. Our study provides new insights into how the microbial genetic potential changes during soil formation along the Damma glacier forefield.

INTRODUCTION

Global temperatures are expected to increase by 2–6°C by 2085 compared with values in 1981–2010 (Collins et al., 2013; Gobiet et al., 2014). Glaciers are retreating worldwide with accelerated climate warming (Haeberli et al., 2002). Newly exposed rocks are colonized by microbial pioneers, a process considered fundamental in shaping the physical and biological development of these ecosystems (Donhauser &

Frey, 2018). Microorganisms in glacier forefield soils play important roles in nitrogen (N) fixation and in the weathering of rock minerals, and they contribute to the accumulation of soil organic matter (SOM) by utilizing ancient carbon (C) sources and photosynthetically derived autochthonous organic C (Bardgett et al., 2007; Frey et al., 2010; Rime, Hartmann, & Frey, 2016; Smittenberg et al., 2012). As soil develops, soil properties, such as texture, pH and available nutrient contents, change steadily, as does the microbial abundance associated with the concomitant restructuring of soil microbial communities (Duc et al., 2009; Rime et al., 2015). Therefore, retreating glacier forefields have been used extensively to investigate changes in microbial communities during soil formation, based on space-for-time substitution (e.g., Bajerski & Wagner, 2013; Bernasconi et al., 2011; Rime et al., 2015; Walker & del Moral, 2010; Zumsteg et al., 2012).

The Damma glacier, located in the Central Alps in Switzerland, is a model system to study ecosystem formation and is probably the most studied glacier forefield in the world. Research on the soil microbial communities have mostly been focused on the microbial dynamics along the chronosequence, which are mainly determined by the change in soil pH, clay content and plant establishment during soil development (Bernasconi et al., 2011; Frey et al., 2013; Rime et al., 2015; Zumsteg et al., 2012). For example, recently deglaciated soils in the Damma glacier forefield were reported to be dominated by photosynthetic and N_2-fixing bacteria (e.g., *Geobacter*, Cyanobacteriota) and by microorganisms active in weathering, which can increase C and nutrients in soils, thereby facilitating microbial succession (Bajerski & Wagner, 2013; Frey et al., 2013; Rime et al., 2015; Schmidt et al., 2008). Vegetated soils, in contrast, were found to be inhabited by microbes capable of degrading complex organic compounds including lignocellulolytic and ectomycorrhizal fungi (Rime et al., 2015; Zumsteg et al., 2012). While the taxonomic responses of the soil microbiomes along the Damma glacier forefield have been studied extensively, changes in functional diversity and the microbial genetic potential with soil development have been poorly addressed. Information about such changes is necessary for a thorough understanding of ecosystem formation in glacier forefields.

C availability is a factor limiting the growth of microbial biomass in C-depleted glacier forefields (Bradley et al., 2014; Göransson et al., 2011). Generally, recently deglaciated soils are dominated by an ancient, thermally stable soil organic carbon (SOC) fraction, while there is more newly accumulated SOC (mostly of plant origin) in older soils (Khedim et al., 2021; Smittenberg et al., 2012). A previous study by Schimel and Schaeffer (2012) suggested that microbial community composition, substrate quality and physical access to substrates can all influence microbial control over C degradation in soil. C dynamics, such as C accumulation (Smittenberg et al., 2012), C fluxes (Guelland et al., 2012) and C mineralization (Guelland et al., 2013), have been investigated in the Damma glacier forefield. For example, a study using ^{13}C-labelled C sources to identify the C-degrading microorganisms in the Damma glacier forefield suggested that Betaproteobacterial taxa affiliated with the families Oxalobacteraceae and Comamonadaceae were important in C utilization in recently deglaciated soils, while C-degrading fungal taxa varied in the different soil environments (Rime, Hartmann, Stierli, et al., 2016). In addition, in an investigation using ^{13}C-DNA stable isotope probing, *Flavobacterium* sp. was found to predominantly incorporate fungal-derived C, while Acidobacteria and Pseudomonadota were found to mainly incorporate algal-derived C (Zumsteg et al., 2013). However, studies on the microbial genetic potential regarding C-cycling, especially genes involved in C degradation, have not been performed in glacier forefields. Considering the importance of soil microbiomes in C-cycling, a comprehensive study on the degradative capacity of the microorganisms present in the Damma glacier forefield would be a valuable contribution.

Along with C, N availability is a primary limiting factor for plant productivity (Duc et al., 2009) and microbial growth and metabolism (Göransson et al., 2011). Therefore, a thorough understanding of the shifts in the microbial-meditated N-cycling processes along soil chronosequences is of great importance. Previous studies have suggested that N_2 fixation in the initial soils of the Damma glacier forefield is relatively low (Brankatschk et al., 2011; Duc et al., 2009). Instead, decomposition and mineralization of ancient or allochthonous organic matter have been found to be the dominant N transformation processes in recently deglaciated barren soils in the Damma glacier forefield (Brankatschk et al., 2011). As plants become established with soil development, increasing SOM can promote the energy-demanding N_2 fixation process (Duc et al., 2009). In addition, in vegetated soils, N transformation processes such as nitrification and denitrification can be more important because nitrate is more abundant in the soils (Brankatschk et al., 2011). Previous studies have also indicated a shift from P limitation to N limitation for plant growth along the soil chronosequence, suggesting that the initial N content (derived from glacier melting and atmospheric deposition) was sufficient for the establishment of the first pioneer plants in the early stages of soil development and that N_2-fixation was not essential during that stage because P availability limited plant growth (Göransson et al., 2016). Further, uptake and retention of N by the plant–soil system have been reported to remain constant along the Damma glacier forefield (De Vries et al., 2021). The studies mentioned above provided

important information regarding N-cycling processes along the Damma glacier forefield, and how plants and microbiomes are interlinked to mitigate N poverty in initial stages of soil formation. However, a more complete understanding of how microbially meditated N-cycling pathways change along the soil chronosequence is still needed.

There are a few attempts using shotgun metagenomics from glacier forefields in the Arctic to better understand N-cycling processes (Nash et al., 2018; Varliero et al., 2021). For example, Nash et al. (2018) investigated *nifH* genes across four glacier forefields, and identified metabolically diverse diazotrophs, such as *Nostoc* and *Geobacter*. In addition, Varliero et al. (2021) identified functionally important taxa (e.g., *Fimbriiglobus*, *Streptomyces*) involved in the N_2 fixation process in three Arctic proglacial systems. As far as we know, however, there has been no investigation of this kind from the Central Alps.

Apart from C and N-cycling, microbial-meditated rock weathering, play a pivotal role in nutrient acquisition and soil formation in early stages of glacier forefield (Uroz et al., 2009, 2015; Welch et al., 2002). Rock weathering bacteria (Frey et al., 2010; Lepleux et al., 2012; Liu et al., 2012; Olsson-Francis et al., 2015) and fungi (Brunner et al., 2011) are commonly found in glacier forefields, where they produce organic acids (e.g., oxalate) and hydrogen cyanide (HCN) to mobilize the nutrients (El-Tarabily et al., 2008) and siderophore to import iron into cells (Frey et al., 2010; Olsson-Francis et al., 2015). However, there are still limited studies from glacier forefield although weathering has been found an important process in early stages of soil formation.

Here, we used shotgun metagenomics to study the microbial genetic potential related to weathering and C- and N-cycling pathways in the Damma glacier forefield (Figure S1A). Specifically, we studied samples from four different stages of soil development (SSDs), characterized by an increasing vegetation cover from barren (0% vegetation cover), to greyish biological soil crusts (10%), to sparsely vegetated (35%) and to vegetated (95%) (Figure S1B). We hypothesized that:

i. Functional gene diversity (e.g., C- and N-cycling genes) increases with advancing SSD along the glacier forefield, due to the increasing occurrence of vegetation and environmentally heterogeneous microniches.
ii. Functional categories related to C degradation are more abundant in vegetated soils because C is more abundant and more heterogenous C sources are available.
iii. Genes related to the degradation of labile C (e.g., starch, oligosaccharides) are more abundant in vegetated soils, which feature more plant litter and greater microbial necromass input, while genes related to the degradation of recalcitrant C (e.g., ancient C) are more abundant in barren soils.
iv. Genes involved in N_2 fixation (*nif*-genes) are more abundant in vegetated soils, where the presence of plant litter may favour energy-demanding N_2 fixation, and genes related to nitrification (e.g., *amoA*-genes) are also enriched in vegetated soils, due to greater ammonium input via N_2 fixation.
v. Rock-weathering bacteria differ between stages of soil development with higher relative abundances in the barren soils.

EXPERIMENTAL PROCEDURES

Study site

The Damma glacier forefield, located in the central Swiss Alps (N46°380, E8°280), has been intensively studied (Bernasconi et al., 2011; Frey et al., 2010, 2013; Rime et al., 2015; Rime, Hartmann, & Frey, 2016; Zumsteg et al., 2012). The glacier has been retreating since ~1850 at an approximate rate of 10 m per year (Zumsteg et al., 2012). The climate at the Damma glacier is typical for a high-alpine environment, with an annual mean temperature of 2.2°C, annual mean precipitation of 2300 mm and annual water runoff ~2700 mm (Hindshaw et al., 2011). The bedrock material along the forefield consists mainly of coarse-grained metamorphic granite (Bernasconi et al., 2011). In the youngest soils the soil texture is typical of coarse sand, while there is more loamy sand in the vegetated soils (Bernasconi et al., 2011).

Sampling

In 2016, three parallel transects spanning four different stages of soil development were set up, one along the center of the sites and one on each side of the center transect at about 30 m distance. The four SSDs have been deglaciated for 4, 12, 45 and 120 years (Bernasconi et al., 2011). The sites were characterized by an increasing vegetation cover with 0%, 10%, 35% and 95% of the site surface, respectively (Figure S1B) (Bernasconi et al., 2011). The 4-year-old soils are located near the glacier terminus and consist mainly of barren sandy rock. The 12-year-old soils are mainly colonized by mosses and lichens. The 45-year-old soils are sparsely vegetated, and their vegetations mainly consist of grass (e.g., *Agrostis gigantea* and *Festuca rubra*) and shrubs (*Salix* spp.). The 120-year-old soils are densely vegetated and are characterized by shrubs (e.g., *Rhododendron ferrugineum* and *Salix* spp.) and grass (e.g., *Festuca rubra* and *Agrostis gigantea*; De Vries et al., 2021; Göransson et al., 2016; Rime et al., 2015). Surface soils at a depth of 0–10 cm,

including biocrust and vegetation, were collected within a 1-m^2 area in each site along each transect, then pooled to form one replicate. In total, 12 independent samples (three transects and four SSDs) were collected. Soil samples were transported to WSL in a polystyrene box cooled with cold packs. Soil samples were sieved through a 2-mm sieve to remove fine roots and stones and then stored until further processing, at −20°C for biological analyses and at 4°C for physico-chemical analyses.

Soil physico-chemical and biotic characteristics

Soil pH was measured using a glass electrode linked to a pH meter (FEP20-FiveEasy Plus, Mettler-Toledo GmbH, Greifensee, Switzerland). Soil texture (sand, silt and clay content) was determined using the hydrometer technique (Gee & Bauder, 1986). About 2 g of homogenized soil was ground with a Teflon ball mill, and about 40 mg of dried soil was subsequently weighed into tin caps for measuring total carbon (TC) and total nitrogen (TN) concentrations with a CHN analyser (Shimadzu, Tokyo, Japan). Dried soil was extracted with milliQ water (1:10 m/v) in 250-mL PE bottles, placed on an overhead shaker overnight at room temperature, and passed through a pleated paper filter (0790 1/2 Whatman paper filter, Cytiva, Marlborough, MA, USA). Dissolved organic carbon (DOC) and dissolved organic nitrogen (DON) concentrations were measured with a TOC-V analyser (Shimadzu, Tokyo, Japan). Soil nitrate (NO_3^-) concentration was measured by ion chromatography with an IC:DX-120 chromatograph (Dionex, Thermo Fisher Scientific, Waltham, MA, USA). Soil ammonium (NH_4^+) concentration was determined photometrically with a FIAS 300 (PerkinElmer, Waltham, MA, USA). Water holding capacity (WHC) was measured according to Schimel et al. (1999). Roots in soil samples were collected during sieving and dried overnight at 60°C before measuring root biomass. Microbial biomass C (MBC) was determined using the chloroform fumigation–extraction method (Vance et al., 1987).

DNA extraction, quantitative PCR and shotgun sequencing

Total DNA of each of the 12 samples was extracted from 10 g soil using the DNeasy PowerMax Soil Kit (Qiagen GmbH, Hilden, Germany), following the manufacturer's instructions (Frey et al., 2016). DNA was quantified using a fluorescence emission procedure with PicoGreen (Molecular Probes, Eugene, OR, USA). Bacterial and fungal biomass were estimated by quantitative PCR (qPCR) on a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA). qPCR reactions were prepared using 6.6 μL of the DNA extracts and the primer pairs 27F/519R and ITS3/ITS4, which amplify the V1–V3 region of the 16S rRNA gene in bacteria and the ITS2 genomic region in fungi, respectively. qPCR programs were performed as previously described (Frey et al., 2020, 2021). An aliquot of DNA template (600–1000 ng) was sheared to a peak target size of 350 bp using a Bioruptor (Diagenode, S.A., Liège, Belgium). Shotgun library preparation was carried out with a Nextera XT Library Preparation kit (Illumina, San Diego, CA, USA). Sequencing reactions were carried out on the Illumina HiSeq 2500 platform (2 × 100 cycles) at the IBERS Aberystwyth Translational Genomics Facility. The raw sequences were deposited in the NCBI Sequence Read Archive under accession number PRJNA796041.

Metagenome assembly and functional annotation

A customized pipeline was applied to pre-process raw reads, assemble pre-processed reads into contigs and annotate contigs for functionality and taxonomy (Frey, Rast, et al., 2022; Frey, Varliero, et al., 2022). In short, the quality of raw reads was checked using FastQC v0.11.8 (Andrews, 2018), followed by quality filtering, read trimming and removal of Illumina adapters using Trimmomatic v0.36 (Bolger et al., 2014), which resulted in pre-processed reads. The pre-processed reads were then assembled into contigs (>200 bp) by iteratively building de Bruijn graphs using k-mers of increasing size with MEGAHIT v1.2.9 (-k-min 27-k-step 10; [Li et al., 2015]). MetaGeneMark v3.38 was used to predict protein-coding genes in the contigs (Zhu et al., 2010). Predicted genes were annotated to the eggNOG (Huerta-Cepas et al., 2016), Carbohydrate-Active enZYmes (CAZy; [Cantarel et al., 2009]) and NCyc (Tu et al., 2019) databases. EggNOG, which classifies genes into orthologous groups (OGs) of proteins and assigns OGs to general functional categories, was used to evaluate the microbial genetic potential for general metabolic and cellular functions. The annotation to the eggNOG v4.5.1 database was carried out with eggnog-mapper v1.0.3 operated with the DIAMOND search mode against all protein sequences (Huerta-Cepas et al., 2017). C- and N-cycling genes were annotated with the CAZy and NCyc databases to reveal the microbial genetic potential related to the C- and N-cycle. The annotation of the predicted genes to the CAZy (downloaded on 20.07.2017) and NCyc (curated sequences clustered at 100% sequence identity) databases was done with SWORD v1.0.3 ($-v\ 10^{-6}$) (Anwar et al., 2019). In addition to the categorization by enzyme class implemented in CAZy, a manual categorization of CAZy genes into different C substrates was performed as previously outlined (Frey, Rast, et al., 2022; Rüthi et al., 2023).

We also identified weathering genes (e.g., *obcA* genes that are involved in the first step of oxalate biosynthesis, genes involved in cyanide synthesis and genes involved in siderophore synthesis and transport) and their taxonomic affiliation following procedure in Varliero et al. (2021).

Abundance of protein-coding genes

BWA aligner v0.7.15 (bwa-mem; [Li, 2013]) was used to map pre-processed read pairs to assembled contigs. Gene abundances were obtained by counting the reads that mapped to the predicted protein-coding genes using the 'featureCounts' function from the package Subread v1.6.2 (-minOverlap 10, Q = 10, -primary; [Liao et al., 2014]). Read abundances of predicted genes were calculated as the mean depth of the mapped reads within the gene-coding regions (e.g., [sum per base read depth]/[gene length]), and counts were normalized to transcripts per million (TPM) (Abrams et al., 2019).

Taxonomic classification

Predicted protein-coding genes annotated to the functional databases (e.g., eggNOG) were assigned taxonomically using Kaiju v1.7.4 (Menzel et al., 2016) with the precompiled NCBI BLAST nr + euk database (version from 24.02.2021) and default settings. The helper program kaiju-addTaxonNames was utilized to convert NCBI taxon IDs to taxonomy. Additionally, the CheckM v1.1.2 (Parks et al., 2015) function -ssu_finder was used to identify 16S and 18S rDNA sequences from the contigs. These sequences were assigned to the SILVA taxonomy database (release 138; [Quast et al., 2013]) using SINA v1.2.12 (Pruesse et al., 2012). To estimate the abundance of the 16S and 18S rRNA genes, the corresponding read counts per contig were normalized to the contig length in kbp.

Statistical analyses

Alpha-diversity indices of functional and taxonomic genes were analysed using the *phyloseq* R package (McMurdie & Holmes, 2013) (R version 4.2.2, [R Core team, 2022]) based on sequences rarefied to even depth (1,038,288 for eggNOG dataset, 37,723 for CAZy dataset, 6892 for NCyc dataset and 8297 for SSU dataset). The effect of SSD on soil properties, relative abundances of different phyla, relative abundances of weathering genes and alpha-diversity indices of functional and taxonomic genes were assessed by one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) test using the *agricolae* package in R. All variables were tested for normality and homogeneity of variances using Shapiro–Wilk and Bartlett's tests, respectively. In case of non-normality and/or heteroscedasticity, variables were transformed by taking either the natural logarithm or the square root of the original value. In addition, the relationship between soil characteristics and microbial functional diversity and soil age were tested using linear regression, implemented in R. To investigate the differences in functional and ribosomal gene structures between SSDs, Permutational multivariate analysis of variance (PERMANOVA), implemented in PRIMER v7, was used to assess the statistical significance of the gene structure of the different SSDs. Differences in multivariate dispersion among functional and taxonomic communities were analysed using the PERMADISP function in PRIMER v7. To investigate the differences in functional and ribosomal gene structures between SSDs, principal component analysis (PCA) based on Euclidean matrices was performed using TPM counts (genes annotated using the eggNOG, CAZy and NCyc databases) and the relative abundances of ribosomal counts (SSU dataset) using the ggfortify and dplyr packages in R. The environmental parameters were regressed against the PCA ordination scores using the 'envfit' function in the *vegan* package (Oksanen et al., 2011) in R. To assess changes in functional genes (annotated using eggNOG, CAZy and NCyc database) with increasing SSD, $\log_2$-fold changes (LFCs) were calculated between each pair of SSDs using the *DESeq2* R package (Love et al., 2014; R version 4.3.3, [R Core team, 2022]). LFCs with adjusted *P*-value <0.05 (using Benjamini-Hochberg method) were considered significant in the DESeq2 analysis. All plots (except PCA plots for community structures) were generated with the R package *ggplot2* (Wickham, 2011).

RESULTS

Changes in soil and microbial characteristics with SSD

Soil pH ranged from 4.7 to 6.2 and decreased significantly ($P < 0.05$) with increasing SSD (soil age), whereas most of the other soil physicochemical variables such as total C, total N, DOC, DON, NO_3^- and NH_4^+ increased with increasing SSD (Table 1; Figure S2). Regarding biotic (plant and microbial) properties, root biomass, MBC, DNA concentration, prokaryotic 16S rRNA gene copies, fungal ITS copies and the 16S/ITS2 ratio increased with increasing SSD and were greatest in the vegetated soil (Table 1; Figure S2).

TABLE 1 Effect of stage of soil development (SSD) along the Damma glacier forefield on soil physico-chemical and biotic (plant and microbial) characteristics

	SSD (soil age in years)			
	4	**12**	**45**	**120**
Soil chemical properties				
pH (H_2O)	6.1 ± 0.1**a**	5.5 ± 0.1**b**	5.7 ± 0.1**b**	4.9 ± 0.2**c**
C (%)	0.07 ± 0.02**b**	0.09 ± 0.02**b**	0.32 ± 0.09**b**	2.53 ± 0.64**a**
N (%)	0.01 ± 0**c**	0.01 ± 0**c**	0.04 ± 0.01**b**	0.14 ± 0.02**a**
DOC (μg g^{-1} soil)	6.5 ± 1.5**c**	13.6 ± 2.1**bc**	17.8 ± 3.2**b**	38.5 ± 6.3**a**
DON (μg g^{-1} soil)	0.31 ± 0.03**b**	0.90 ± 0.20**b**	1.90 ± 0.47**b**	4.74 ± 1.03**a**
NO_3^- (μg g^{-1} soil)	0.08 ± 0.02**c**	0.17 ± 0.06**c**	0.47 ± 0.06**b**	0.95 ± 0.15**a**
NH_4^+ (μg g^{-1} soil)	0.03 ± 0.01**b**	0.04 ± 0.01**b**	0.26 ± 0.06**b**	7.96 ± 0.66**a**
Soil physical properties				
Sand (%)	90 ± 3**a**	71 ± 2**c**	46 ± 2**d**	77 ± 2**b**
Silt (%)	9 ± 3**d**	28 ± 2**b**	52 ± 2**a**	19 ± 2**c**
Clay (%)	1 ± 0**c**	2 ± 0.**b**	2 ± 0**b**	4 ± 0**a**
WHC (%)	17.5 ± 2.6**c**	32.0 ± 3.2**b**	39.5 ± 4.6**b**	55.6 ± 4.6**a**
Biological properties				
RB (mg g^{-1} soil)	0 ± 0**b**	1.3 ± 0.2**b**	5.2 ± 2.3**b**	46.8 ± 7.3**a**
MBC (μg g^{-1} soil)	8.9 ± 1.2**c**	19.5 ± 3.9**c**	60.5 ± 11.0**b**	151.9 ± 5.9**a**
DNA (ng g^{-1} dry soil)	153.9 ± 18.4**c**	738.9 ± 348.2**bc**	1821.2 ± 234.4**b**	8761.1 ± 743.5**a**
16S (g^{-1} dry soil)	(3.9 ± 2.1) × 10^6**b**	(6.4 ± 1.9) × 10^7**b**	(7.3 ± 4.1) × 10^8**b**	(1.5 ± 0.5) × 10^9**a**
ITS (g^{-1} dry soil)	(5.0 ± 2.4) × 10^6**b**	(3.1 ± 0.8) × 10^7**a**	(3.1 ± 1.3) × 10^7**a**	(4.6 ± 1.0) × 10^7**a**
16S/ITS	1.0 ± 1.1**b**	2.2 ± 0.9**b**	25.6 ± 8.2**ab**	36.2 ± 21.0**a**

Note: SSDs were characterized by vegetation cover and soil age: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The effect of SSD was analysed by one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) test. *P*-values were corrected for multiple testing with the Benjamini–Hochberg method. Values are means ± SD ($n = 3$). Significant differences between SSD are annotated by different lowercase letters in bold. C, total carbon; N, total nitrogen; DOC, dissolved organic carbon; DON, dissolved organic nitrogen; WHC, water holding capacity; RB, root biomass; MBC, microbial biomass carbon; DNA, DNA content; 16S, copy numbers of bacterial 16S gene; ITS, copy numbers of fungal ITS region.

TABLE 2 Effect of stage of soil development (SSD) along the Damma glacier forefield on the total number of sequences and the percentage of protein-coding genes and contigs.

	SSD (soil age in years)			
	4	**12**	**45**	**120**
Raw reads (×10^6)	20.2 ± 3.9**a**	15.0 ± 8.8**a**	22.8 ± 2.8**a**	16.2 ± 4.7**a**
HQ reads (×10^6)	20.0 ± 3.9**a**	14.9 ± 8.7**a**	22.6 ± 2.8**a**	16.1 ± 4.7**a**
No. reads mapped to CDS genes (×10^6)	4.0 ± 0.9**a**	3.5 ± 2.3 **a**	3.6 ± 0.6**a**	2.0 ± 0.4**a**
% reads mapped to CDS genes	20 ± 0**a**	23 ± 2**a**	16 ± 1**b**	13 ± 1**b**
No. reads mapped to contigs (×10^6)	6.1 ± 1.5**a**	4.6 ± 3.0**a**	5.9 ± 1.2**a**	3.4 ± 1.0**a**
% reads mapped to contigs	30 ± 2**a**	31 ± 2**a**	25 ± 3**b**	22 ± 1**b**

Note: SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The effect of SSD was analysed by one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) test. *P*-values were corrected for multiple testing with the Benjamini–Hochberg method. Values are means ± SD ($n = 3$). Significant differences based on the LSD test are annotated by different lowercase letters in bold. Raw reads, total number of sequences after sequencing; HQ reads, high-quality sequences after quality control; CDS genes, coding DNA sequence, genes that encode proteins; Contig, a set of overlapping DNA segments that together represent a consensus region of DNA.

Overall metagenome sequencing results

After quality control, 221,040,668 high-quality reads (8,327,037 to 25,436,792 reads per sample) were obtained (Table 2; Table S1). MEGAHIT assembly of reads into contigs produced a total of 1,477,538 contigs of 571 bp on average, ranging from 200 to 75,697 bp, with an N50 value of 577 and a GC content of 55% (Table S2). In total, 1,875,711 predicted genes were found among the contigs, 934,602 of which could be

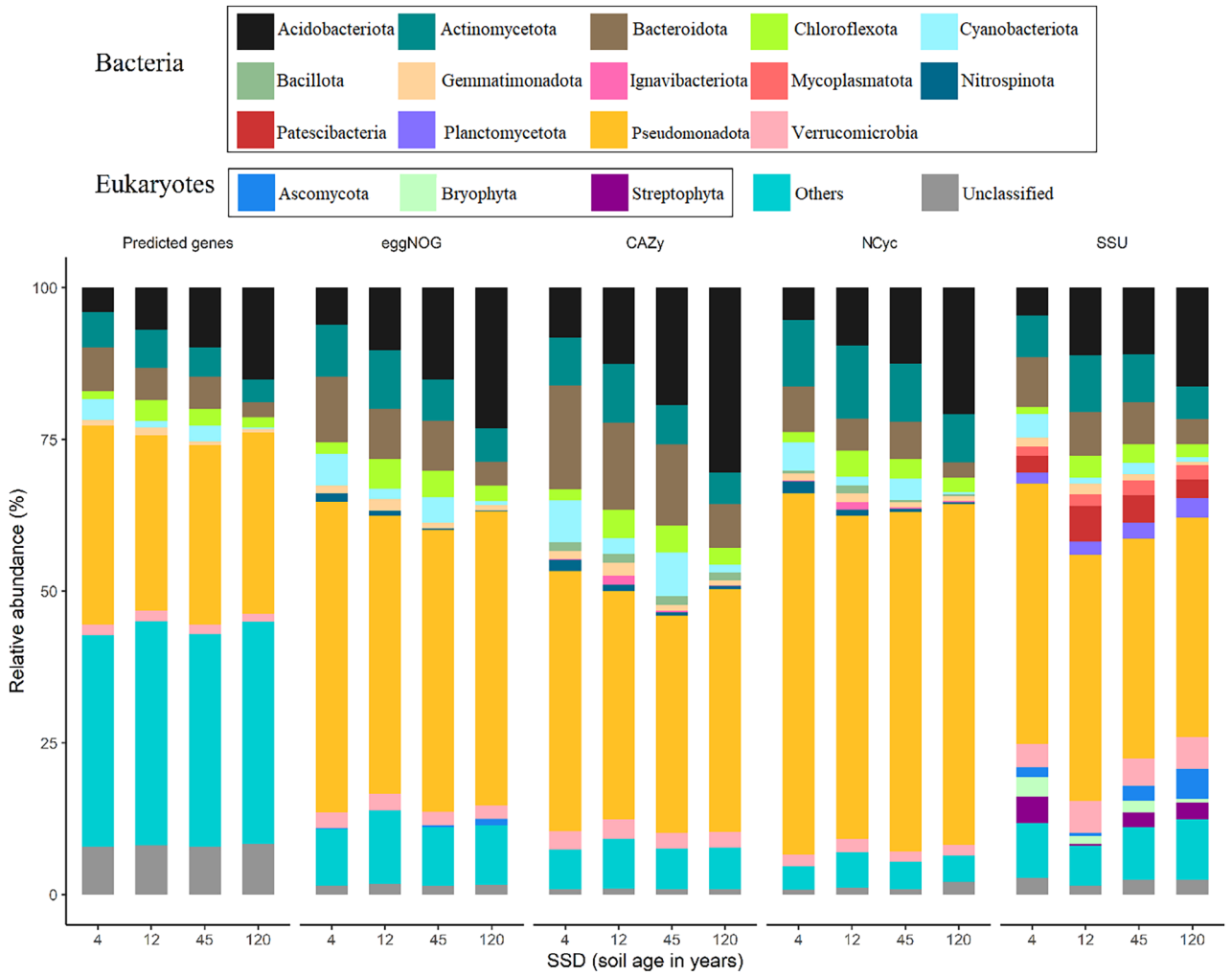

FIGURE 1 Changes in taxonomic composition with stage of soil development (SSD) along the Damma glacier forefield, based on predicted, functional (annotated using eggNOG, CAZy and NCycDB) and ribosomal (SSU) genes. SSDs were characterized by vegetation cover and soil age: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). Relative abundances are the mean of three replicates. Only phyla with a relative abundance >0.01 are shown.

annotated with the eggNOG database (including only 1940 COGs), 9202 with the CAZy database, and 2105 with the NCyc database, corresponding to 49.8%, 1.0% and 0.2% of the reads mapped to predicted genes, respectively (Table S2).

Taxonomic composition of the metagenomes

Contigs were classified into different taxa, whereas reads mapped to classified contigs were used to quantify the abundance of the taxa. Bacteria were the most abundant organisms in all SSDs, with relative abundances >90%. Diverse bacterial taxa were identified by analysing the 16S rRNA genes within contigs. Among the Bacteria, the phylum Pseudomonadota dominated in all SSDs, followed by Acidobacteriota, Actinobacteriota and Bacteroidota (Figure 1; Table S3). The relative abundance of Acidobacteriota increased steadily from barren to vegetated soil (Figure 1; Table S3). For Eukaryotes, high relative abundances of Ascomycota, Bryophyta and Streptophyta were only found in ribosomal (SSU) genes within contigs, with the highest relative abundances of Bryophyta and Streptophyta occurring in the barren soil (Figure 1; Table S3).

Changes in the diversity of protein-coding genes with SSD

There was a significant effect of SSD on alpha-diversity for both functional and ribosomal genes (Figure 2A-H; Table S4). After rarefaction, the richness and Shannon index of functional genes annotated with eggNOG, CAZy and NCyc database decreased significantly with

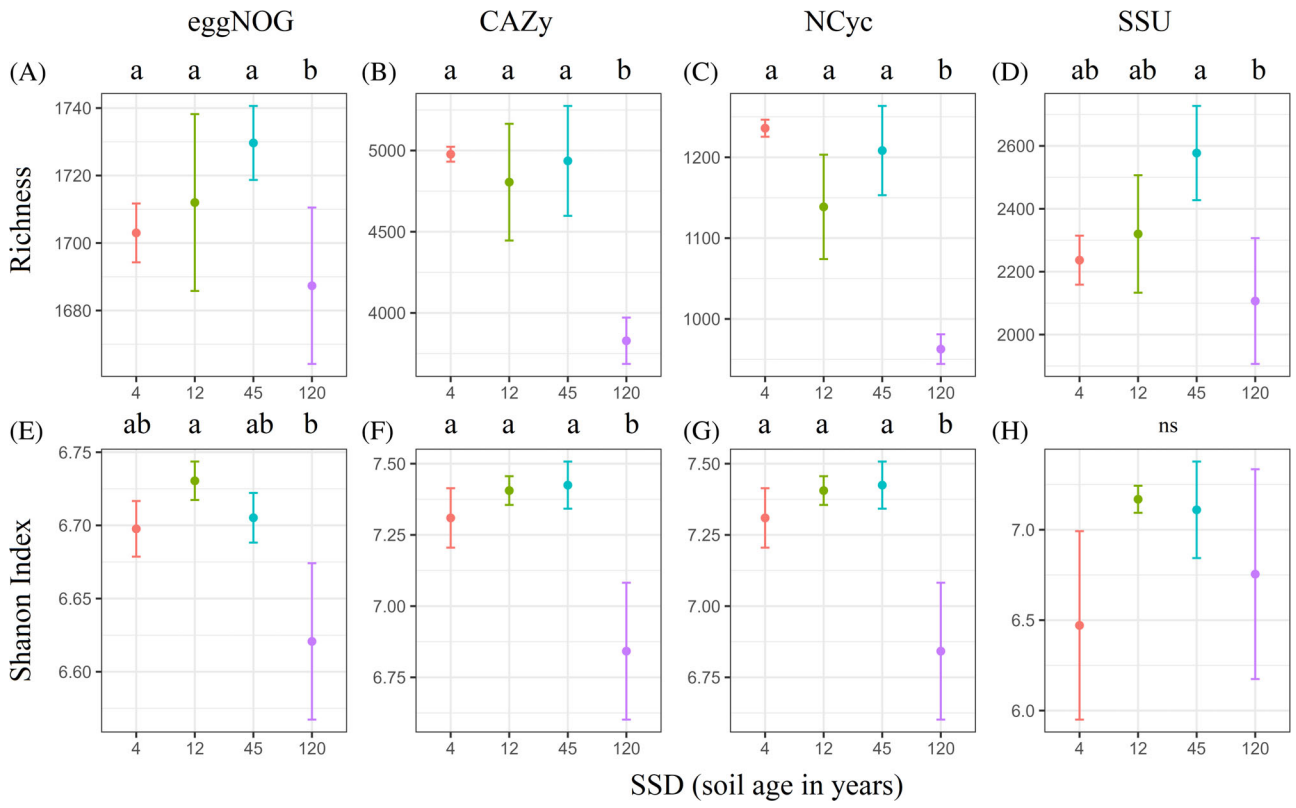

FIGURE 2 Changes in the alpha-diversity of functional (annotated using eggNOG, CAZy and NCyc database) and ribosomal (SSU) genes with stage of soil development (SSD) along the Damma glacier forefield. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). Richness (A–D) and Shannon index (E–H) are shown for each SSD. Data are shown as mean and standard deviation ($n = 3$). Significance was calculated using one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) test. *P*-values were corrected for multiple testing with the Benjamini–Hochberg method. Significant differences between SSDs are annotated by different lowercase letters. ns, not significant.

increasing soil age (Figure S3). For CAZy enzyme classes, alpha-diversity tended to be lower in the vegetated soils. In particular, significant ($P < 0.05$) differences between SSDs were found for the richness of carbohydrate-binding modules (CBMs), glycoside hydrolases (GHs) and glycoside transferases (GTs) and for the Shannon index of the CBMs and GHs (Figures S5 and S8). Similarly, the lowest alpha-diversity of N-cycling families, such as those involved in denitrification and dissimilatory nitrate reduction (DNR) and in nitrification, was found in the vegetated soils (Figures S6 and S9). Details of the alpha-diversity of functional genes annotated with eggNOG, CAZy and NCyc database at the category/class/family level can be found in the Figures S4–S9.

Shifts in functional gene structure with SSD and links with environmental variables

SSD had a significant impact on the structure of functional (eggNOG: $P = 0.0002$; CAZy: $P = 0.0007$; NCyc: $P = 0.0004$) and ribosomal genes ($P = 0.004$) (Table 3; Figure S10A–D). There was a significant difference in functional gene structure between the different SSDs, with the exception of pairwise comparisons involving the 45-year-old soils analysed with CAZy and NCyc database. In contrast, the taxonomic structure was only significantly different between the 4- and the 120-year-old soils (Table 3; Figure S10A–D). In addition, results from beta-dispersion suggested that the differences in microbial gene structure in the different SSDs were not caused by within-group differences (Table S5). Results of the 'envfit' analysis demonstrated soil pH, clay, N, DOC, DON, nitrate and WHC contributed most to the differences in functional and taxonomic community structure (Table 4).

Shifts of eggNOG genes with SSD

Differential abundance analysis indicated shifts in the overall functional genes between two SSDs (only COG genes were considered here). The total number of

TABLE 3 Effect of stage of soil development (SSD) along the Damma glacier forefield on the structure of functional (annotated using eggNOG, CAZy and NCyc database) and ribosomal (SSU) genes.

	eggNOG		CAZy		NCyc		SSU	
Overall effect								
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
SSD	5.36	**0.0002**	3.41	**0.0007**	3.46	**0.0004**	2.71	**0.004**
Soil age (years)								
	F	*P*	*F*	*P*	*F*	*P*	*F*	*P*
12 over 4	2.44	**0.0225**	1.98	**0.0384**	2.17	**0.0275**	1.76	0.059
45 over 4	2.05	**0.0328**	1.80	0.0517	1.72	0.0579	1.53	0.093
120 over 4	2.70	**0.0130**	2.23	**0.0211**	2.08	**0.0260**	1.97	**0.034**
45 over 12	2.12	**0.0322**	1.54	0.0906	1.68	0.0641	1.40	0.143
120 over 12	2.87	**0.0099**	2.08	**0.0270**	2.12	**0.0227**	1.76	0.053
120 over 45	1.50	0.1144	1.38	0.1375	1.31	0.1764	1.33	0.157

Note: SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The effect of SSD was analysed by one-way analysis of variance (ANOVA) followed by a least significant difference (LSD) test. *P*-values were corrected with the Benjamini–Hochberg method. Significant differences ($P < 0.05$) are marked in bold.

TABLE 4 Abiotic and biotic variables that influenced microbial functional (annotated using eggNOG, CAZy and NCyc database) and ribosomal (SSU) genes structure.

	eggNOG		CAZy		NCyc		SSU	
	R^2	*P*	R^2	*P*	R^2	*P*	R^2	*P*
pH (H_2O)	0.69	**0.0072**	0.67	**0.0099**	0.69	**0.0063**	0.81	**0.0015**
C (%)	0.66	**0.0126**	0.72	**0.0080**	0.65	**0.0127**	0.75	**0.0075**
N (%)	0.78	**0.0041**	0.79	**0.0031**	0.78	**0.0061**	0.86	**0.0007**
DOC (μg g^{-1} soil)	0.76	**0.0043**	0.73	**0.0056**	0.75	**0.0052**	0.83	**0.0012**
DON (μg g^{-1} soil)	0.83	**0.0017**	0.8	**0.0022**	0.82	**0.002**	0.9	**0.0002**
NO_3^- (μg g^{-1} soil)	0.84	**0.0004**	0.87	**0.0003**	0.84	**0.0013**	0.86	**0.0006**
NH_4^+(μg g^{-1} soil)	0.65	**0.0156**	0.67	**0.0126**	0.64	**0.0147**	0.76	**0.0100**
WHC (%)	0.93	**0.0001**	0.93	**0.0001**	0.93	**0.0001**	0.88	**0.0003**
Clay (%)	0.75	**0.0051**	0.73	**0.0055**	0.74	**0.0067**	0.84	**0.0008**
RB (mg g^{-1} soil)	0.69	**0.0118**	0.69	**0.0116**	0.68	**0.0115**	0.79	**0.0041**
MBC (μg g^{-1}soil)	0.88	**0.0005**	0.88	**0.0004**	0.88	**0.0011**	0.93	**0.0002**
DNA (ng g^{-1} dry soil)	0.75	**0.0077**	0.75	**0.0068**	0.74	**0.0089**	0.85	**0.0008**
16S	0.84	**0.0004**	0.79	**0.0011**	0.84	**0.0004**	0.89	**0.0001**
ITS	0.84	**0.0005**	0.85	**0.0005**	0.85	**0.0003**	0.87	**0.0002**
16S/ITS	0.68	**0.0028**	0.57	**0.0219**	0.67	**0.0065**	0.71	**0.0048**

Note: Significant variables ($P < 0.05$) are marked in bold.C, total carbon; N, total nitrogen; DOC, dissolved organic carbon; DON, dissolved organic nitrogen; WHC, water holding capacity; RB, root biomass; MBC, microbial biomass carbon; 16S, copy number of bacterial 16S gene; ITS, copy number of fungal ITS region.

differentially abundant genes annotated to eggNOG, and the abundance of most of these genes, differed among SSDs (Table 5; Figure S11; Table S12). The comparison between barren soils and biological soil crusts indicated 443 differentially abundant genes (22.8% of the total number of COG genes) with 191 under- and 252 overrepresented genes (Table 5). The comparison between barren and vegetated soils showed 558 differentially abundant genes (28.8%), with 266 under- and 292 overrepresented genes (Table 5). Between sparsely vegetated and vegetated soils only a minor fraction of differentially abundant COG genes (42 genes; 2.2% of the total number of COGs) were detected. Here, we aimed to characterize only the most salient shifts between SSDs. The 45-year-old soils were not included because of the few significant differences in the comparisons between 45-year-olds and other SSDs (Table 5; Figure S10A); however, we provide the complete test statistics of all functional genes at various levels in Tables S5, along with a more detailed description of the results (Supporting Results).

The comparison between barren soils and biological soil crusts showed that genes related to energy production and conversion and carbohydrate transport

TABLE 5 Numbers of functional genes annotated using eggNOG (only COGs were included), CAZy and NCyc database that represent Differentially Abundant Genes (DAGs) between each two stages of soil development (SSDs) along the Damma glacier forefield.

		SSDs (soil age in years)					
		12 over 4[a]	**45 over 4**	**120 over 4**	**45 over 12**	**120 over 12**	**120 over 45**
eggNOG[b]	Total	1940	1940	1940	1940	1940	1940
	DAGs	443	271	558	260	629	42
	Up	252	153	292	135	320	24
	Down	191	118	266	125	309	18
CAZy	Total	9202	9202	9202	9202	9202	9202
	DAGs	874	401	772	203	642	6
	Up	621	199	301	36	160	4
	Down	253	202	471	167	482	2
NCyc	Total	2105	2105	2105	2105	2105	2105
	DAGs	259	69	145	43	196	3
	Up	208	28	49	1	24	3
	Down	51	41	96	42	172	0

Note: SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years).

[a]The $\log_2$-fold change (LFC) value of '12 over 4' is the $\log_2$ of (gene abundance of 12/gene abundance of 4-year-old soils). 'Total' means the total number of genes used in the DESeq2 analysis between 12 and 4-year-old soils. 'DAGs' means the total number of significantly different genes between 12 and 4-year-old soils. 'Up' means the number of genes that were significantly more abundant (LFC > 0, $P < 0.05$) in the 12-year-old soils relative to the 4-year-old soils, while 'Down' means the number of genes that were significantly more abundant in the 4-year-old soils than in the 12-year-old soils (LFC < 0, $P < 0.05$). The same stands for '120 over 4' and '120 over 12'.

[b]Only with COG ID/annotations.

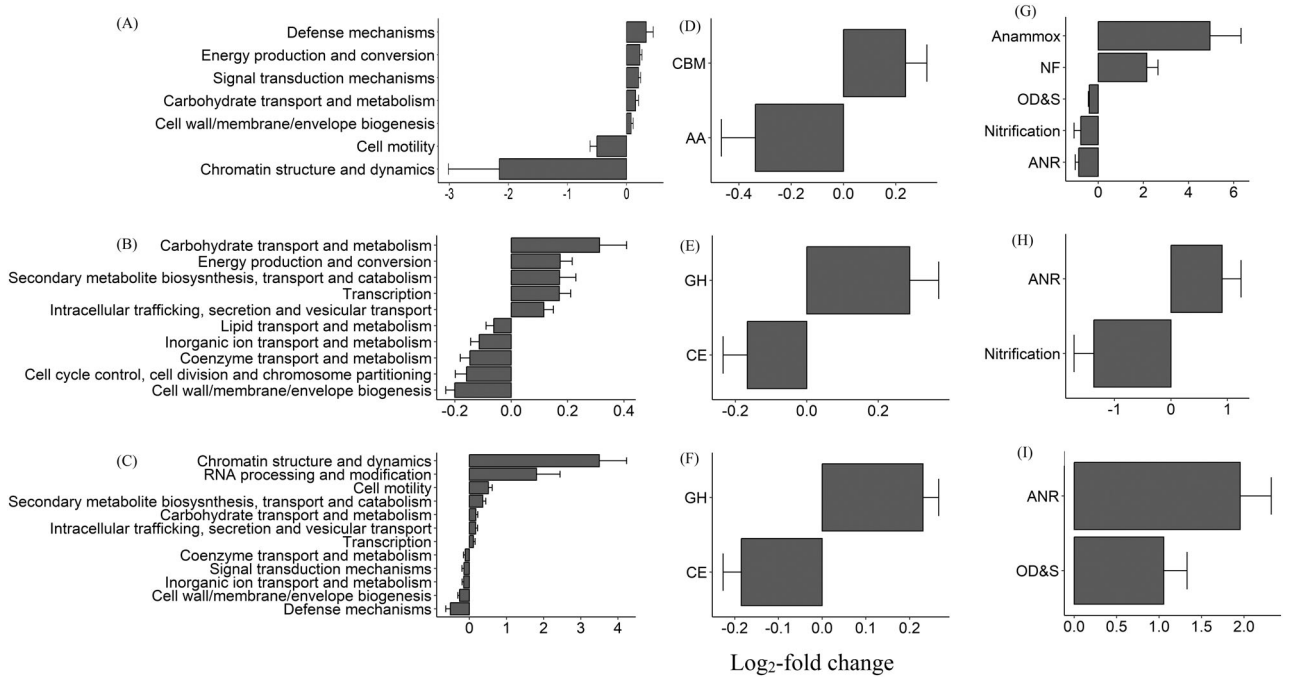

FIGURE 3 Effect of stage of soil development (SSD) on the functional genetic potential along the Damma glacier forefield. Bars represent $\log_2$-fold changes (LFCs) in functional genes annotated with eggNOG (A–C), CAZy (D–F) and NCyc (G–I) database, aggregated over functional categories that differed in abundance ($P < 0.05$) between SSDs with soil ages of 4 and 12 years (A, D, G), 4 and 120 years (B, E, H) and 12 and 120 years (C, F, I). Positive LFCs mean that genes are enriched in the latter soils, e.g., in (A) defence mechanisms were more abundant in 12- than 4-year-old soils. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). AA, auxiliary activities; ANR, assimilatory nitrate reduction; CBM, carbohydrate-binding modules; CE, carbohydrate esterases; GH, glycoside hydrolases; NF, nitrogen fixation; OD&S, organic degradation and synthesis.

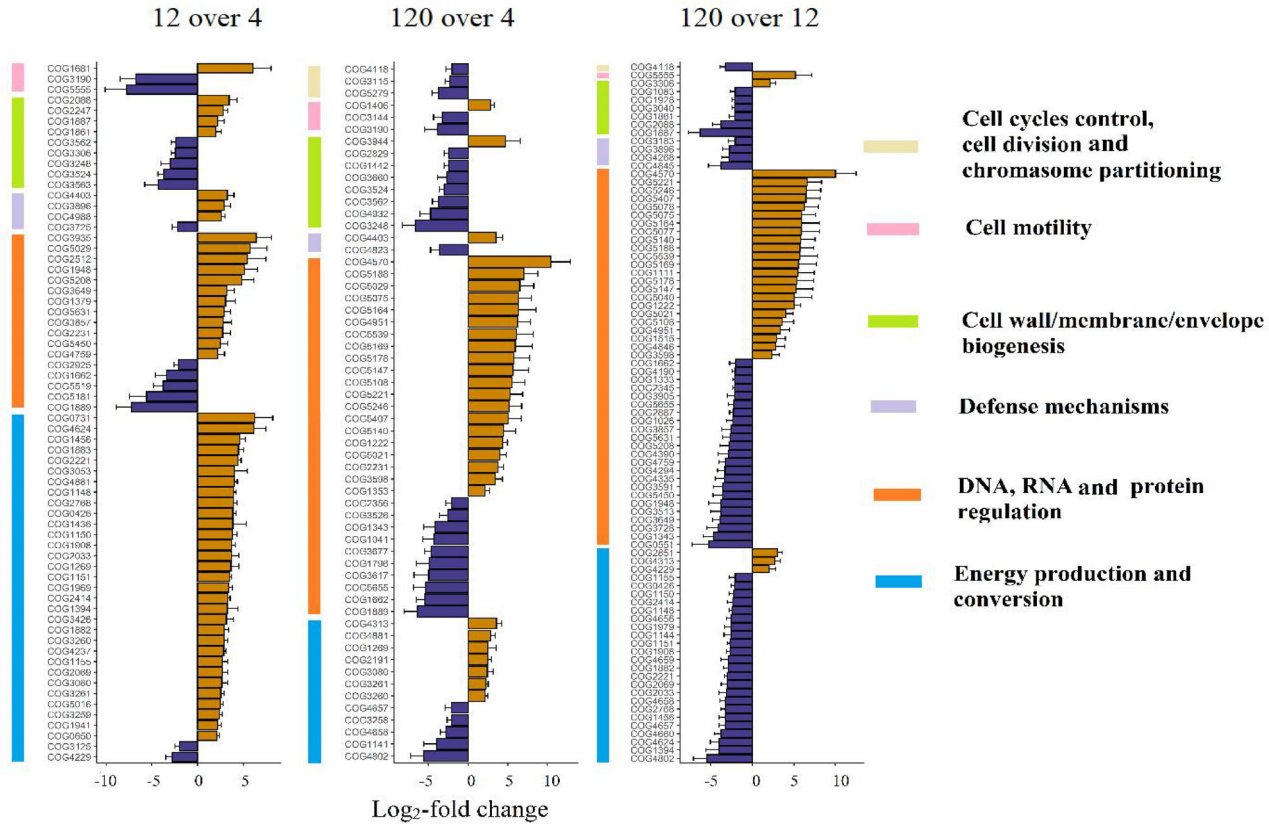

FIGURE 4 Effect of stage of soil development (SSD) along the Damma glacier forefield on functional genes. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The $\log_2$-fold change (LFC) value '12 over 4' is the $\log_2$ of gene abundance of 12/gene abundance of 4-year-old soils. The same stands for '120 over 4' and '120 over 12'. Bars represent LFCs in functional genes annotated with the eggNOG database between different SSDs. Only significantly differing COGs ($P < 0.05$) with $|LFC| > 2$ in the categories DNA, RNA, protein regulation, energy production and conversion, cell wall/membrane/envelope biogenesis, cell motility, defence mechanisms, cell cycle control, cell division and chromosome partitioning are depicted. Orange indicates overrepresented COGs (LFC > 0) and violet indicates underrepresented COGs (LFC < 0). For example, COG1681 is overrepresented in the 12-year-old soils relative to the 4-year-old soils, while COG3190 is overrepresented in the 4-year-old soils. All the COGs and their annotations can be found in Table S6. The 45-year-old soils are not included here because, based on the principal coordinate analysis of the functional gene structure, the 45-year-old soils are very similar to the 120-year-old soils (see Figure S10A) and have the fewest differentially abundant genes in the overall comparisons (see Table 5).

and metabolism (functional type Metabolism) and to defence mechanisms, signal transduction mechanisms and cell wall/membrane/envelope biogenesis (Cellular processes and signalling) were enriched in the biological soil crusts, while genes related to cell motility (Cellular processes and signalling) were overrepresented in the barren soils (Figure 3A; Table S6). At the individual gene level, differentially abundant genes involved in defence mechanisms (i.e., COG3896: chloramphenicol, COG4988: ABC transporter), and energy production and conversion (i.e., COG4624: 4Fe-4S binding domain protein, COG 1456: CO dehydrogenase acetyl-CoA synthase gamma subunit, COG1882: formate acetyltransferase) were enriched in the biological soil crusts, while genes involved in cell motility (i.e., COG5555: metallophosphoesterase, COG3190: flagellar biosynthesis protein fliO) were overrepresented in barren soils (Figure 4).

By comparing barren and vegetated soils, differential abundance analysis showed that functional categories related to carbohydrate transport and metabolism, energy production and conversion and secondary metabolite biosynthesis (Metabolism) were overrepresented in vegetated soils, while functional categories related to inorganic ion transport and metabolism and to lipid transport and metabolism (Metabolism) and related to cell wall/membrane/envelope biogenesis (Cellular processes and signalling) were enriched in barren soils (Table 6; Figure 3B). At the individual gene level, genes involved in energy production and conversion (i.e., COG3080: fumarate reductase subunit D, COG3260: NADH ubiquinone oxidoreductase 20 kDa subunit) were enriched in vegetated soils, while genes in the categories cell motility (i.e., COG3190: flagellar biogenesis proteins, COG3144: flagellar hook-length control protein) and

TABLE 6 Summary of the overrepresented/underrepresented functional genes annotated using eggNOG, CAZy and NCyc database that differed significantly between SSDs.

Database	Category/class/family[a]	Gene ID	Function
eggNOG	G↑[b]	COG1455↑	PTS system
		COG3280↑	alpha amylase
		COG3469↑	chitinase
		COG3934↑	cellulase
		COG4678↑	lysozyme
	M↓	COG2951↑	lytic murein transglycosylase
		COG3134↓	surface antigen
		COG3248↓	channel-forming protein
		COG3307↓	antigen polymerase
		COG3944↑	biosynthesis protein
		COG4623↓	murein-degrading enzyme
	Q↑	COG3319↑	synthetase (specify what kind of synthetase in which pathways)
		COG3648↑	uricase
		COG4689↑	decarboxylase
CAZy	AA	AA1↓	laccase (lignin)
		AA5↓	oxidase (lignin)
		AA5_2↓	oxidase (lignin)
	CE↓	CE3↓	acetyl xylan esterase (hemicellulose)
		CE8↓	pectin methylesterase (pectin)
		CE9↓	N-acetylglucosamine 6-phosphate deacetylase (chitin)
	GH↑	GH19↑	chitinase (chitin)
		GH38↑	α-mannosidase (starch/oligosacchride)
		GH43_24↑	exo-β-1,3-galactanase (hemicellulose)
		GH115↑	xylan α-1,2-glucuronidase (hemicellulose)
NCyc	NF	nifD↑	nitrogenase molybdenum-iron protein alpha chain
		nifK↑	nitrogenase molybdenum-iron protein beta chain
	Nitrification↓	amoB_B↓	ammonia monooxygenase subunit B (bacteria)
		amoC_B↓	ammonia monooxygenase subunit C (bacteria)
		hao↓	hydroxylamine dehydrogenase
	OD&S	ansB↑	glutamin-(asparagin-)ase
		gdh_K00260↑	glutamate dehydrogenase
		gdh_K00262↑	glutamate dehydrogenase
		gdh_K15371↓	glutamate dehydrogenase
		glsA↑	glutaminase
		nmo↓	nitronate monooxygenase

Note: SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). For the original data used to create this table, refer to Tables S6, S7 and S8. M, cell wall/membrane/envelope biogenesis; G, carbohydrate transport and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism.
Abbreviations: AA, auxiliary activities; CE, carbohydrate esterases; GH, glycoside hydrolases; NF, nitrogen fixation; OD&S, organic degradation and synthesis.
[a]Significantly differently abundant genes between barren and vegetated soils (most of them were also significantly differed between other comparisons) and with important functions in glacial forefield soils were selected and listed. Only comparison between barren (4-year-old) and vegetated (120-year-old) soils was included here, as it is a good representation of how microbial functional potential changed as soil develops. The table is formatted as gene categories (eggNOG)/classes (CAZy)/families (NCyc) followed by corresponding individual genes (eggNOG and NCyc)/gene families (CAZy) within the above gene category/class/family). Genes at category/class/family level are underlined, corresponding genes at individual/family level are not underlined. Genes at individual/family level are formatted as gene ID followed by their annotated functions.
[b]The $\log_2$-fold change (LFC) value of '120 over 4' is the $\log_2$ of (gene abundance of 120/gene abundance of 4-year-old soils). '↑' means the genes were significantly more abundant (LFC >0, $P < 0.05$) in the 120-than 4-year-old soils, '↓' means the genes were significantly more abundant in the 4- than 120-year-old soils (LFC <0, $P < 0.05$). For example, functional category 'G' in eggNOG dataset is overrepresented in 120-year-old soils, and COG3280 within category 'G', which has the gene potential to coding alpha amylase, is also overrepresented in 120- than 4-year-old soils.

cell wall/membrane/envelope biogenesis (i.e., COG1442: glycosyl transferase family) were overrepresented in barren soils (Figure 4).

Comparisons at the functional category level between biological soil crusts and vegetated soils showed that several functional categories within the main functional type such as 'Metabolism': secondary metabolite biosynthesis and carbohydrate transport and metabolism; 'Information storage and processing': RNA processing and modification; and 'Cellular processing and signaling': cell motility, were overrepresented in vegetated soils, while functional categories within 'Metabolism: that is, inorganic ion transport and metabolism; and 'Cellular processing and signaling': that is, defence mechanisms and cell wall/membrane/envelope biogenesis were enriched in the biological soil crusts (Figure 3C). At the individual gene level, genes involved in defence mechanisms (i.e., COG3896: chloramphenicol phosphotransferase-like protein, COG4845: chloramphenicol acetyltransferase), cell wall/membrane/envelope biogenesis (i.e., COG1928: glycosyl transferase family; COG1083: CMP-N-acetylneuraminic acid synthetase) and energy production and conversion (i.e., COG1150: heterodisulfide reductase; COG2414: ferredoxin oxidoreductase) were overrepresented in the biological soil crusts (Figure 4). Complete test statistics of all eggNOG genes at various levels, along with a more detailed description of the results, are provided in the Supplementary Information (Table S6; Supporting Results).

Shifts in CAZy genes with SSD

Overall, we found diverse CAZy gene in all SSDs (Figure S12A). Numerous genes related to the degradation of lignin, cellulose, hemicellulose and chitin indicated a high degradative capacity of the microorganisms present in the Damma glacier forefield (Table S7). The total number of differentially abundant genes annotated to CAZy, and the abundance of most of these genes, differed between two SSDs (Table 5; Table S12; Figure S12B). The comparison between barren soils and biological soil crusts indicated 874 differentially abundant genes (9.5% of the total number of CAZy genes), with 253 under- and 621 overrepresented genes (Table 5). The comparison between barren and vegetated soils showed 772 differentially abundant genes (8.4% of the total number of CAZy genes), with 471 under- and 301 overrepresented genes (Table 5). Between sparsely vegetated and vegetated soils only a minor fraction of differentially abundant CAZy genes (6 genes; 0.07% of the total number of CAZy genes) were detected.

At the CAZy family level, the comparison between barren soils and biological soil crusts indicated that genes involved in auxiliary activities (AA7) related to the degradation of lignin were overrepresented in the biological soil crusts (Figure 5). Similarly, gene families involved in the degradation of chitin (i.e., GH19, GH20, CBM54, CE14), cellulose (i.e., GH6, GH9, CBM3), hemicellulose (i.e., GH16, GH27, CBM35), pectin (i.e., PL4_1, PL9_1, PL11_1) and starch/oligosaccharides (i.e., GH3, GH13_7, GH95) were overrepresented in the biological soil crusts (Figure 5). In contrast, we found several gene families related to the degradation of various labile and recalcitrant C sources (i.e., GH5_5, GH5_39, GH10, GH17; GH43_24, GH113, CE7 and PL4_2) that were overrepresented in the barren soils.

The comparison between barren and vegetated soils revealed that several genes involved in auxiliary activities (i.e., AA1, AA5, AA5_2) were overrepresented in the barren soils (Figure 5; Table 6). Similarly, several genes involved in the degradation of cellulose (i.e., GH5_5, GH8, CBM3), hemicellulose (i.e., GH67, GH113, CE6, CE16), chitin (i.e., CE9, CE11, CBM5, CBM18, CBM54) and pectin (i.e., PL1, GH105, PL9, CE8, GH28, PL1_2) were overrepresented in barren soils. In contrast, several gene families involved in the degradation of starch/oligosaccharides (i.e., CBM42, CBM57, GH13_11, GH13_23, GH13_26, GH32, GH38) were overrepresented in the vegetated soils (Figure 5). Complete test statistics of all CAZy genes at various levels, along with a more detailed description of the results, are provided in the Supporting Information (Table S7; Supporting Results).

Shifts in N-cycling genes with SSD

The most abundant N-cycling gene family was that involved in organic degradation and synthesis (OD&S; relative abundance of 67.3% across all habitats), followed by denitrification and dissimilatory nitrate reduction (22.0%), while genes involved in N_2 fixation (1.5%) and nitrification (0.7%) were relatively less abundant (Figure S13A). The total number of differentially abundant genes annotated to the NCyc database, and the abundance of most of these genes, differed between two SSDs (Table 5; Figure S13B; Table S12). The comparison between barren soils (4 years old) and biological soil crusts (12 years old) indicated 259 differentially abundant genes (12.3% of the total number of N-cycling genes), with 51 under- and 208 overrepresented genes (Table 5). The comparison between barren (4 years old) and vegetated soils (120 years old) showed 145 differentially abundant genes (6.9%), with 96 under- and 49 overrepresented genes (Table 5). Between sparsely vegetated (45 years old) and vegetated soils (120 years old) only three differentially abundant N-cycling genes (0.14%) were detected (Table 5).

At the individual gene level, the comparison between barren soils and biological soil crusts showed that genes related to anammox (*hzsC*), assimilatory N

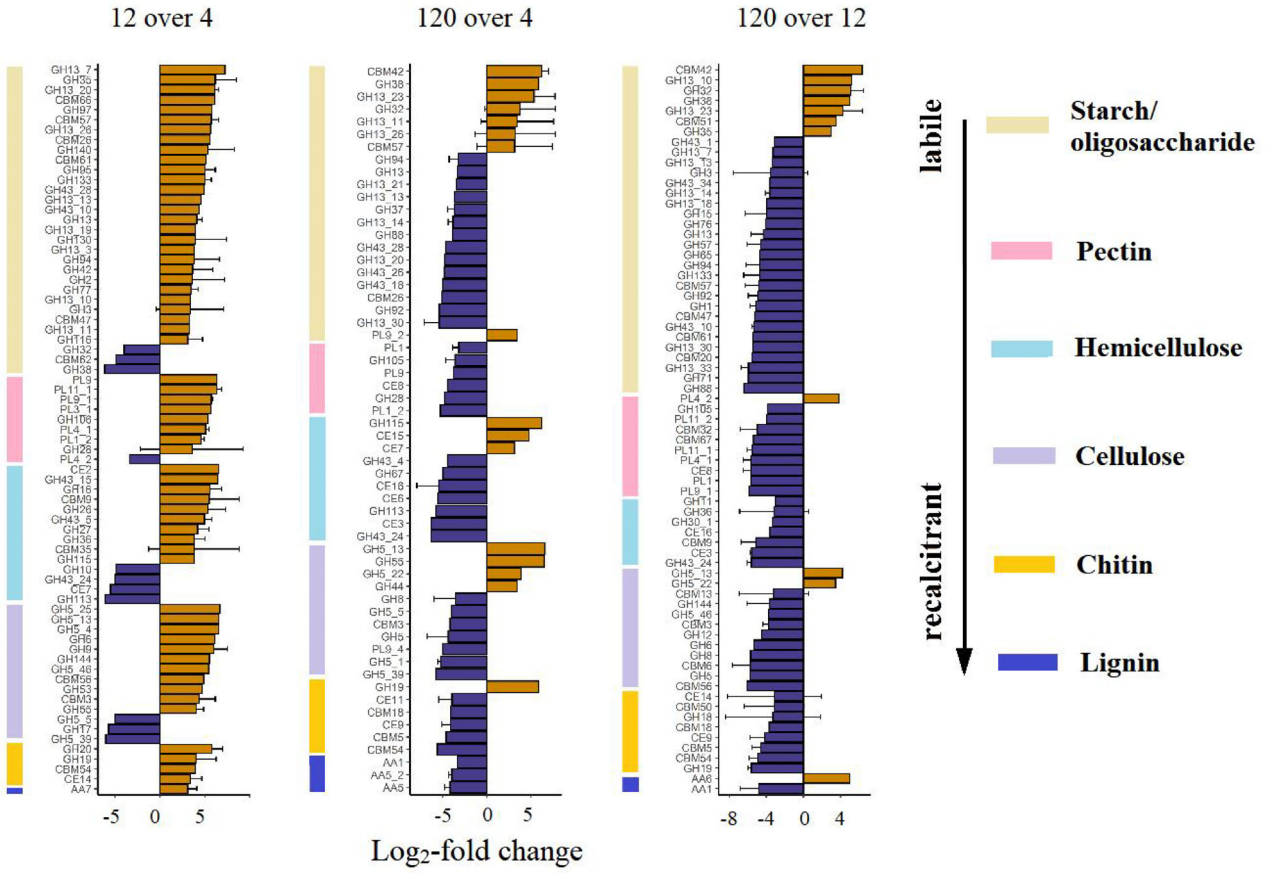

FIGURE 5 Effect of stage of soil development (SSD) along the Damma glacier forefield on CAZy genes. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The $\log_2$-fold change (LFC) value '12 over 4' is the $\log_2$ of gene abundance of 12/gene abundance of 4-year-old soils. The same stands for '120 over 4' and '120 over 12'. Bars represent LFCs in C-degrading genes annotated using the CAZy database between different SSDs. Only significantly differing genes ($P < 0.05$) with $|LFC| > 3$ are depicted. Orange indicates overrepresented (LFC > 0) genes and violet indicates underrepresented genes (LFC < 0). For example, GH13_7 is overrepresented in the 12-year-old soils relative to the 4-year-old soils, while GH38 is overrepresented in the 4-year-old soils. AA, auxiliary activities; CBM, carbohydrate-binding modules; CE, carbohydrate esterases; GH, glycoside hydrolases; PL, polysaccharide lyases. Detailed information on all the significantly differing C-degrading genes can be found in Table S7. The 45-year-old soils are not included here because, based on the principal coordinate analysis of the functional gene structure, the 45-year-old soils are very similar to the 120-year-old soils (see Figure S10B) and have the fewest differentially abundant genes in the overall comparisons (see Table 5).

reduction (ANR) (*NR*, *narC*, *nasB* and *nirA*), denitrification and DNR (i.e., *narG*, *nirK*, *nirS*, *nosZ*), nitrification (*hao*), N_2 fixation (*nifK*, *nifD*, *nifH*) and OD&S (i.e., *glsA*, *asnB*, *gdh_00260*) were overrepresented in the biological soil crusts (Figure 6; Table 6). In contrast, *nirB* and *nirD* genes related to denitrification and DNR were overrepresented in the barren soils (Figure 6).

By comparing barren and vegetated soils, differential abundance analysis revealed that genes involved in ANR (i.e., *NR*, *nirA*), denitrification and DNR (i.e., *nosZ*, *narY*, *nirB*) and nitrification (*hao*, *amoB_B, amoC_B*) were overrepresented in barren soils, whereas *nifD* and *nifK*, responsible for N_2 fixation, and *gdh_K00260*, *ansB*, *gdh_Koo262* and *glsA*, involved in OD&S, were enriched in the vegetated soils (Figure 6; Table 6).

In the comparison between biological soil crusts and vegetated soils, genes involved in denitrification and DNR (*norB*, *nirK*, *napC*), nitrification (*hao*) and nitrogen fixation (*nifH* and *nifK*) were overrepresented in the biological soil crusts. Most of the genes involved in OD&S (except for *ureA* and *ansB*) and ANR (except for *narB* and *nasA*) were overrepresented in the vegetated soils (Figure 6). Complete test statistics of all N-cycling genes at various levels, along with a more detailed description of the results, are provided in the Supporting Information (Table S8; Supporting Results).

Barren soils have a high capacity for rock weathering

The percentage of weathering genes was lowest in the biological soil crusts (17.6 ± 0.7%) and highest (21.3 ± 1.2%) in barren soils (Table S9). We further studied

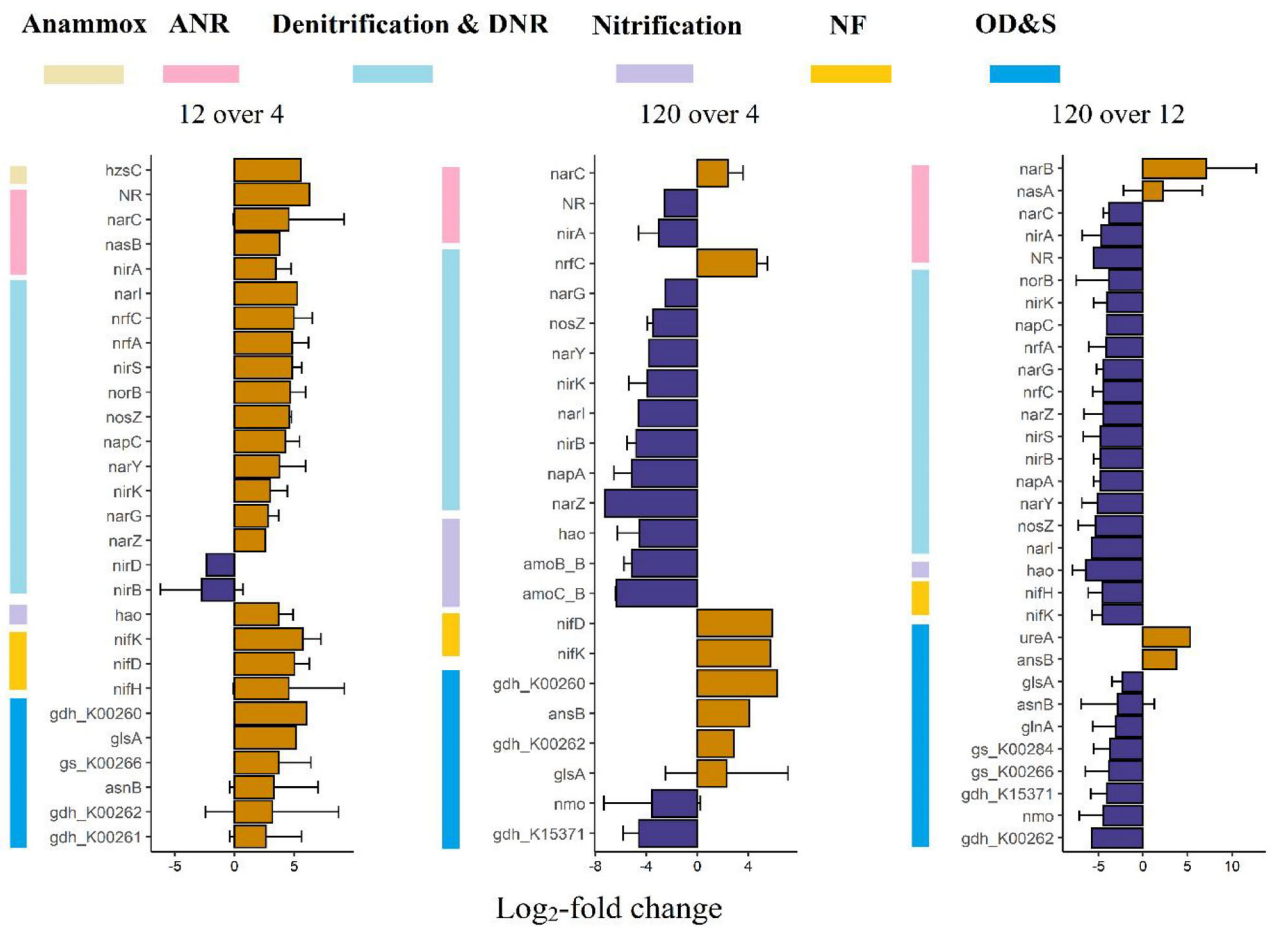

FIGURE 6 Effect of stage of soil development (SSD) along the Damma glacier forefield on N-cycling genes. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The $\log_2$-fold change (LFC) value '12 over 4' is the $\log_2$ of gene abundance of 12/gene abundance of 4-year-old soils. The same stands for '120 over 4' and '120 over 12'. Bars represent LFCs in N-cycling genes annotated using the NCyc database between different SSDs. Only significantly differing abundant genes ($P < 0.05$) with $|LFC| > 2$ are depicted. Orange indicates overrepresented genes (LFC > 0) and violet indicates underrepresented genes (LFC < 0). For example, *hzsC* is overrepresented in the 12-year-old soils relative to the 4-year-old soils, while *nirB* is overrepresented in the 4-year-old soils. Detailed information on all significantly differing abundant N-cycling genes can be found in Table S8. The 45-year-old soils are not included here because, based on the principal coordinate analysis of the functional genetic structure, the 45-year-old soils are very similar to the 120-year-old soils (see Figure S10C) and have the fewest differentially abundant genes in the overall comparisons (see Table 5). ANR, assimilatory nitrate reduction; DNR, dissimilatory nitrate reduction; NF, nitrogen fixation; OD&S, organic degradation and synthesis.

the taxonomic affiliation of the microbiomes associated with rock weathering in the different SSDs. At the class level, Betaproteobacteria was the most abundant (relative abundance 35% on average across all SSDs), with the highest abundance in barren soils (40.8 ± 2.4%) and biological soil crusts (40.8 ± 1.4%). The second most abundant class was Alphaproteobacteria (24.8% on average), with the highest abundance in sparsely vegetated (28.6 ± 4.3%) and vegetated (33.6 ± 4.7%) soils (Figure 7A; Table S9). At the genus level, *Bradyrhizobium* was the most abundant (7.9%), with the highest relative abundance in sparsely vegetated (10.4 ± 2.5%) and vegetated (12.6 ± 2.0%) soils (Figure 7B), followed by *Variovorax* (2.6%) and *Rhizobacter* (1.9%), which were most abundant in barren soils (Figure 7B; Table S10). The relative abundance of the top 20 weathering genes at the family level is listed in Table S11.

DISCUSSION

Greater microbial genetic potential in younger soils

In this study, we investigated the metabolic capabilities of the soil microbiomes along a chronosequence of the Damma glacier forefield. In contradiction to our first hypothesis, the alpha-diversity of functional and ribosomal genes were lower in the vegetated soils.

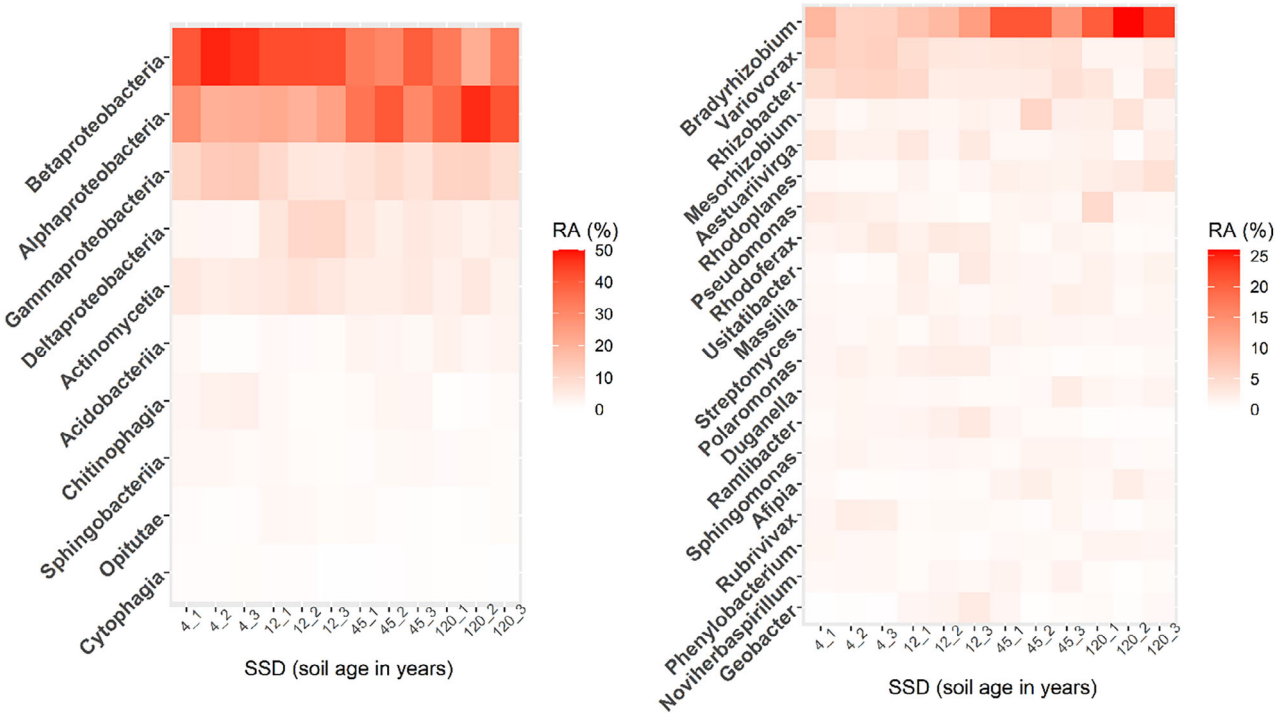

FIGURE 7 Effect of stage of soil development (SSD) along the Damma glacier forefield on the rock-weathering microbiome at the class (A) and the genus (B) level. Classes with relative abundances (RAs) ranked in the top 10 and genera with relative abundances ranked in the top 20 are shown. SSDs were characterized by vegetation cover: barren (0% vegetation cover, age 4 years), greyish biological soil crusts (10%, 12 years), sparsely vegetated (35%, 45 years) and vegetated (95%, 120 years). The sample IDs are formatted as SSD (soil age) followed by replicate number.

Previous findings have indicated that taxonomic diversity increases along soil chronosequences because of the increased C and N contents in soils, resulting from plant litter input and root exudates, which can sustain the growth of fast-growing microorganisms (Bajerski & Wagner, 2013; Rime, Hartmann, Stierli, et al., 2016; Schutte et al., 2009; Zumsteg et al., 2012). In addition, an increase in the number of potential microniches, habitat heterogeneity and resource changes with soil age, soil crust development and plant growth may support a more complex soil microbiome (Duran et al., 2021; Edwards & Cook, 2015; Rime et al., 2015; Schutte et al., 2010). The higher functional diversity in barren soils and biological soil crusts in our study could be explained by the selective pressure imposed by low nutrient and high ultraviolet (UV) conditions (Ni et al., 2022). In the vegetated soils, the successful establishment of highly competitive organisms and the increase in interspecific competition may have led to a decline in diversity (Ortiz-Alvarez et al., 2018; Pothula & Adams, 2022). In addition, microorganisms stored in glacier habitats (e.g., ice, sub- or supraglacial sediments and glacier streams) and released into barren soil after glacier retreat may have contributed to the higher microbial diversity in the initial soils (Rime, Hartmann, & Frey, 2016; Zhong et al., 2021). As for community structure, the microbial functional and ribosomal gene structures were significantly influenced by similar abiotic variables, such as soil pH, clay, N, DOC, DON, nitrate and WHC. Significant influences of soil pH, silt content and gravimetric soil moisture on the bacterial and fungal community structures in the Damma glacier forefield were also reported by Rime et al. (2015). Interestingly, we found strong changes in the functional gene structure with SSDs but no or only small shifts (between 4- and 120-year-old soils) in the ribosomal gene structure (Table 3). The variation of functional gene structure was more evident, which suggested that, though the taxonomy was not significantly changed, the microbial function differed. This means that functional redundancy (similar microorganisms capable of a particular metabolic function [Louca et al., 2018]) exists in the Damma glacier forefield which has so far never reported from a forefield of a retreating glacier.

Shifts of functional pathways along the Damma glacier forefield

Functional genes annotated to EggNOG revealed that the majority of the predicted genes in the Damma

glacier forefield were categorized as replication, recombination and repair, energy production and conversion, translation, ribosomal structure and biogenesis and amino acid transport and metabolism, indicating that soil microbiomes were metabolically active (Donhauser et al., 2021). We observed a statistically significant increase in the relative abundances of various functional pathways with increasing SSD. Differential abundance analysis demonstrated that genes related to carbohydrate transport and metabolism were generally enriched in vegetated soils, which suggests functional adaptation of the microbial community to utilize varying C resources, such as cellulose, hemicellulose, starch, pectin and chitin, accumulating in the soil due to the growth of plants and other microbial communities. These findings support our second hypothesis, that the microbial functional capability related to C dynamics increases with soil development, and may be attributed to the gradual accumulation of soil C derived from root exudates and plant litter input (Bardgett et al., 2005; Knelman et al., 2012). Previous studies from the Damma glacier forefield (Smittenberg et al., 2012) and a meta-analysis from 10 glacier forefields (Khedim et al., 2021) also showed that soil organic C increased with increasing SSD, where labile C was mainly of plant origin (Kelly et al., 2021). In addition, genes related to secondary metabolite biosynthesis, transport and catabolism (i.e., COG3648: uricase, COG4689 [decarboxylase] and COG2162 [acetyltransferase]) were overrepresented in vegetated soils, which may be attributed to the enhanced root–microbiome interactions in later SSDs. Further, genes related to cell wall/membrane/envelope biogenesis were enriched in barren soils and biological soil crusts, which is consistent with previous findings that microbiomes in nutrient-poor environments invest energy predominantly in cell metabolism (Bhattacharya et al., 2022).

As glacier forefields are generally reported to be harsh environments for the survival of soil microorganisms (low nutrients, high UV radiation; [Bradley et al., 2014]), we particularly focused on genes within the category of post-translational modification, protein turnover and chaperones, which contains many functional genes related to stress responses (Dey et al., 2022; Donhauser et al., 2021). In fact, we found many genes related to stress responses that were overrepresented in barren soils (Table S6), although at the category level we did not find a significant enrichment in barren soils as expected. As an example, COG1066, which plays a role in the repair of endogenous alkylation damage, was enriched in barren soils compared with vegetated soils. Overall, our study demonstrates that the soil microbiomes in recently deglaciated soils are functionally very diverse, with an enrichment of functional pathways related to cell metabolism in barren soils and an enrichment of pathways related to C dynamics and secondary metabolism in vegetated soils.

Shifts in the microbial carbon degradation potential with SSD

Carbohydrate active enzymes (CAZymes) are key in the degradation of soil organic matter (SOM), and the pool size of SOC depends on the balance between the formation of SOM from the decomposition of plant litter and its mineralization to inorganic C by releasing CO_2 to the atmosphere (Lopez-Mondejar et al., 2018). We selected specific CAZyme families that are involved in the decomposition of SOM, including plant-derived biomass, such as starch, hemicellulose, carbohydrate esters, pectin and lignin; fungal-derived biomass, such as chitin and glucan; and bacteria-derived biomass, such as peptidoglycan (Lombard et al., 2014). Among the functional groups of CAZymes at the class level, the gene abundances of glycoside hydrolases (GHs), carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs) and auxiliary activities (AAs) were the most abundant CAZymes involved in the microbial degradation of various C sources in all the SSDs, which is in accordance with findings from high-alpine soils from the eastern Swiss Alps (Donhauser et al., 2021; Perez-Mon et al., 2022; Rüthi et al., 2023). Enzymes contributing to plant-derived biomass decomposition are classified into GH, CE, AA and Polysaccharide Lyase (PL) families. Enzymes contributing to both the fungal- and bacteria-derived biomass are classified as GH families. Differential abundance analysis showed that genes related to GHs were overrepresented in the vegetated soils, suggesting that genes related to the degradation of labile C sources derived from plant biomass increased with increasing SSD. The more abundant fresh C sources available from plant root exudates (Khedim et al., 2021) and microbial necromass (Miltner et al., 2011) may explain the higher abundance of genes related to GHs observed in older soils.

CAZyme families of AAs were enriched in barren soils and biological soil crusts, suggesting a high microbial degradation capacity of recalcitrant C in recently deglaciated soils, supporting our third hypothesis. We found three main enzyme families that participated in the decomposition of lignin: oxidases (AA5, AA7), manganese peroxidases and laccases (AA1), with manganese peroxidases (AA2) being the least abundant among them. In younger soils, in contrast, a high abundance of genes related to CEs and AAs may be attributed to ancient recalcitrate C and lignin accumulation. These results agree with previous finding that microbial communities can use ancient and recalcitrant C as an energy source in the initial stage of soil formation in glacier forefields (Bardgett et al., 2007; Khedim et al., 2021).

Similarly, C-degrading genes related to CEs were enriched in barren soils and biological soil crusts, suggesting a high microbial metabolic capability for complex carbohydrate deconstruction in recently

deglaciated soils. In particular, gene families involved in hemicellulose (CE2, CE3, CE7), pectin (CE8) and chitin (CE9, CE11, CE14) deconstruction were enriched in barren soils. The acetylation of plant polysaccharides can protect plant cell walls from microbial degradation. CEs are involved in the degradation of acetyl groups, following the genes involved in GHs, which participate in the extracellular hydrolysis of plant polysaccharides, that is, cellulose and chitin (Biely, 2012). Overall, in our study the microbiomes of recently deglaciated soils show a high genetic potential for complex carbohydrate deconstruction, with an enrichment of genes related to recalcitrant C degradation in barren soils and an enrichment of genes related to labile C degradation in vegetated soils. These genes mostly belong to Bacteria, as reads ascribed to Bacteria made up >90% of the dataset. However, fungi are also important C-degrading organisms, even if only between 2.5% and 3.6% of the reads in our dataset were ascribed to fungi. It has previously been reported that eukaryotes may be underrepresented in metagenome datasets due to database and gene prediction biases (Jacquiod et al., 2016). Further, our approach relied on forming functional and taxonomic profiles by mapping reads back to the co-assembly. Because fungal genomes are larger and more complex than bacterial genomes, and therefore assemble more poorly, this choice of approach could have contributed the observed bias. The low abundances of unicellular eukaryotes compared with bacteria have been found in other studies (Delmont et al., 2011; Tveit et al., 2013).

SSD affects the microbial genetic potential for N-cycling

Organic N transformation (OD&S) was the most abundant N pathway in our forefield soils, indicating that microorganisms can degrade organic N compounds to provide reactive N for growth and build-up microbial biomass. Our results suggest that this function is more widespread among glacier forefield microbial communities than the more specialized respiratory N transformations. Organic N transformation was also the most important N-cycling process in studies from high-alpine soils (Donhauser et al., 2021; Perez-Mon et al., 2022; Rüthi et al., 2023).

In our study, OD&S genes were overrepresented in the barren soils compared with the biological soil crusts. This finding can be partly attributed to the amount of allochthonous organic material stored in the subglacial sediments, but released after glacier retreat (Bardgett et al., 2007; Khedim et al., 2021). This possibility has also been supported by previous observations that mineralization, mainly the decomposition of deposited organic material, is the main driver of N-cycling in the recently deglaciated soils in the Damma glacier forefield (Brankatschk et al., 2011). However, at the individual gene level, particular genes involved in amino acid biosynthesis, such as *glsA* (glutaminase), *asnB* (asparagine synthase) and *gs* (glutamate synthase), were enriched in the biological soil crusts in our study, suggesting that genes within OD&S may differ in their response to SSD. When comparing recently deglaciated soils (barren soils and biological soil crusts) with vegetated soils (45 and 120 years old), genes related to this N pathway (OD&S) were overrepresented in older soils, which may be explained by the greater input of microbial necromass and plant litter (Table 1). Such a tendency of N mineralization along Damma glacier forefield agreed the review in Schulz et al. (2013).

Another way for the ecosystem to obtain reactive N is N_2 fixation. Nitrogenase genes were found in all SSDs but at low abundance, which is in line with reports from Arctic glacier forefields (Nash et al., 2018; Varliero et al., 2021). Interestingly, N-cycling genes related to N_2 fixation were less abundant in the barren soils and highest in biological soils crust, which supports our fourth hypothesis. A low N_2-fixation potential in newly exposed barren soils of the Damma glacier forefield has been reported previously (Brankatschk et al., 2011; Duc et al., 2009). However, we did find that the autotrophic N_2-fixing bacteria (e.g., Cyanobacteria) was more abundant in 4-year-old soil, which was also suggested by the finding of Frey et al. (2013). A plausible explanation may be that the more nutritious conditions in older soils favour energy-demanding non-autotrophic N_2-fixing microbiomes (Duc et al., 2009). Another explanation could be that biological N_2 fixation is not needed in the initial SSD, as plants and possibly also microorganisms are mainly limited by phosphorus and not nitrogen (Göransson et al., 2016). Plant colonization has been identified as a driving factor in influencing the abundance of *nifH* genes and the diazotrophic community composition (Brankatschk et al., 2011; Duc et al., 2009; Frey et al., 2013; Wang et al., 2021). In biological soil crusts, where few plants (e.g., mosses) grow and soil nutrient contents increase, N_2-fixing microorganisms have a competitive advantage over plants because they can overcome the elevated demand for nutrients while their activity is not hindered by energy limitation (Brankatschk et al., 2011). In vegetated soils, in contrast, plants are strong competitors for nutrient acquisition because they have a longer lifespan than microorganisms and better ability to retain the assimilated N (Hodge et al., 2000).

We also found that the relative abundances of genes related to the nitrification pathway were generally very low, which is in accordance with other reports from high-alpine soils of the Central Alps (Donhauser et al., 2021; Perez-Mon et al., 2022) and from Arctic glacier forefields (Tian et al., 2023). Interestingly, gene families involved in nitrification were overrepresented in

barren soils compared with vegetated soils in our study. This result is in contrast to our expectation that nitrification genes would not be favoured by the low N content in recently deglaciated soils. However, our findings agree with those reported by Brankatschk et al. (2011), who used quantitative PCR to analyse *amoA* genes along the Damma glacier chronosequence and found that the gene copy numbers of *amoA* genes declined towards vegetated soils. One possible explanation might be the soil pH, which is highest in the barren soils and decreases with the presence of vegetation. Acidity has been reported to be a crucial factor inhibiting autotrophic nitrification (Nicol et al., 2008). Alternatively, the higher nitrification genetic potential in barren soils compared with older soils could be caused by the higher oxygen availability in the shallow oxic or partially oxic sandy glacial sediments, as nitrification is an aerobic process (Opitz et al., 2014).

Overall, we did find changes in the abundance of N-cycling genes along the Damma glacier forefield: OD&S dominated N-cycling processes during the process of soil development, genes involved in nitrogen fixation were more abundant in the vegetated soils, and genes related to nitrification were enriched in the barren soils. Despite having distinct requirements (e.g., for oxygen), many reactions in the N-cycle tend to co-occur in the environment, leading to efficient N recycling, competition for the same resource (e.g., nitrate) and coupled processes, such as nitrification and denitrification (Mosley et al., 2022). Brankatschk et al. (2011) suggested, however, that the number of genes involved in major N-cycling pathways may not be consistent with the corresponding enzyme activity, so a deeper understanding of N-cycling genes at the transcriptional and translational level may still be necessary in the Damma glacier forefield.

Distribution of rock-weathering microbiome along Damma glacier forefield

Rock weathering creates gradients in proglacial environments. Whereas the ground in the ice margins is dominated by rocks, SOM increases with an associated increase in vegetation. Here, we focused on rock-weathering processes that shape forefield dynamics and nutrient bioavailability, specifically considering the *obcA* genes that are involved in the first step of oxalate biosynthesis, genes involved in cyanide biosynthesis, and genes involved in siderophore biosynthesis and transport (Ferreira et al., 2019; Frey et al., 2010; Varliero et al., 2021; Welch et al., 1999). Weathering genes were abundant in alpha- and betaproteobacterial taxa along the Damma glacier forefield, with Betaproteobacteria more abundant in barren soil and Alphaproteobacteria more abundant in vegetated soils. Such results agreed with out fifth hypothesis and suggested that these microorganisms play important roles in rock weathering in the initial SSD (Frey et al., 2010; Lapanje et al., 2012; Lepleux et al., 2012). The genus with most of the genes involved in oxalate and siderophore biosynthesis was *Bradyrhizobium*. Bacteria belonging to this genus are mainly recognized as plant symbionts but can also be present as free-living organisms (VanInsberghe et al., 2015). The high relative abundance of genes involved in siderophore metabolism relates to the fact that a considerable production of siderophores is needed to uptake iron released from the rock dissolution performed by oxalate release. Diazotrophs such as *Bradyrhizobium* require iron as a nitrogenase cofactor (Rubio & Ludden, 2008). Weathering-active alphaproteobacterial taxa have also been isolated from the granitic surfaces of rock sediments of recently deglaciated barren soils in the Damma glacier forefield (Frey et al., 2010; Lapanje et al., 2012). The involvement of *Bradyrhizobium* and *Variovorax* in oxalate and siderophore biosynthesis has also been reported in Arctic proglacial systems (Varliero et al., 2021). Microorganisms can grow on rock surfaces or within rocks, where plants are unable to develop, in particular through microbial chemosynthesis supported by mineral weathering (de los Rios et al., 2003; Frey et al., 2010; Garrido-Benavent et al., 2020; Varliero et al., 2021; Wierzchos et al., 2018). In these environments, microorganisms together with lichens pave the way for subsequent plant colonization, as inorganic nutrients are released that can eventually be used by plants (Wild et al., 2022). *Bradyrhizobium* have been reported to fix N_2 (Nash et al., 2018) and supply it to lichens. Further, lichens support mineral weathering by producing fungal organic acids (Brunner et al., 2011; Seneviratne & Indrasena, 2006). The association of *Bradyrhizobium* with pioneering plants may explain its more important role in rock weathering in the vegetated soils. *Variovorax* sp. have been demonstrated to be able to yield microbial isolates that could utilize rock-bound organic matter sources (Samuels, 2018) and have been isolated from the fine granitic sand at the terminus of the Damma glacier (Frey et al., 2010; Lapanje et al., 2012). With an increasing pace of global warming, microbiomes containing rock-weathering capabilities may play more important roles in nutrient accumulation and soil formation in the Damma glacier forefield.

CONCLUSIONS

Our shotgun metagenomics study demonstrates that the microbial communities present in the initial SSD of the Damma glacier forefield harbour a high microbial functional potential for several C and N metabolic pathways. Genes involved in adaptation to multiple stressful conditions were found to be widely distributed in the

communities along the chronosequence. Metabolic flexibility and diverse adaptation strategies are conducive to the survival of microbial communities in the extreme environment of glacier forefields. Our chronosequence-based analysis of microbial succession shows that early SSDs are enriched with autotrophs and recalcitrant compound degraders, indicative of a community influenced by the common ice microbiome, whereas later SSDs have a larger heterotrophic microbial component. Our results confirm the presence of nitrogenase and rock-weathering genes in known N_2-fixing and rock-weathering organisms in the glacier forefield soils. Overall, our results highlight how a deeper information on these and other microbially driven environmental processes is necessary for comprehending microbial succession with the retreat of glaciers. Microbial functional potential, metabolic strategies and impacts of microbial first colonizers in the process of primary succession of glacier forefields remain poorly explored. However, continuing technological progress (e.g., metatranscriptomics) and the incorporation of chemical labeling or microcosm studies will help improve our understanding of how these pioneer microorganisms survive and how their activities shape their environment to enable colonization by successive organisms.

AUTHOR CONTRIBUTIONS

Maomao Feng: Data curation (lead); formal analysis (lead); investigation (supporting); methodology (supporting); writing – original draft (lead). **Gilda Varliero:** Data curation (supporting); formal analysis (equal); methodology (supporting); visualization (supporting); writing – review and editing (supporting). **Weihong Qi:** Data curation (supporting); formal analysis (supporting); methodology (supporting); writing – review and editing (supporting). **Beat Stierli:** Investigation (supporting); methodology (supporting); writing – review and editing (supporting). **Arwyn Edwards:** Methodology (supporting); writing – review and editing (supporting). **Serina Robinson:** Writing – review and editing (supporting). **Marcel G. A. van der Heijden:** Supervision (supporting); writing – review and editing (supporting). **Beat Frey:** Conceptualization (lead); data curation (supporting); formal analysis (supporting); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (lead); supervision (lead); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

We acknowledge Noureddine Hajjar, Alessandro Schlumpf and Jonathan Donhauser from the WSL for sample processing and analyses. We also acknowledge Aberystwyth University for completing shotgun sequencing. We thank Melissa Dawes for her valuable contribution to the editing of this article. This study was funded by a WSL competitive grant (Metagenomics 5233.00388.001.01), the China Scholarship Council (grant number 201904910436) and a MicroArctic Innovative Training Network grant supported by the European Commission's Marie Sklodowska Curie Actions program (project number 675546). Open access funding provided by ETH-Bereich Forschungsanstalten.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. All sequence data have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA796041 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA796041).

ORCID

Maomao Feng https://orcid.org/0009-0005-6266-6283
Arwyn Edwards https://orcid.org/0000-0003-1762-8593
Beat Frey https://orcid.org/0000-0002-6391-3574

REFERENCES

Abrams, Z.B., Johnson, T.S., Huang, K., Payne, P.R.O. & Coombes, K. (2019) A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinformatics*, 20, 679.

Andrews, S. (2018) Fast QC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Anwar, M.Z., Lanzen, A., Bang-Andreasen, T. & Jacobsen, C.S. (2019) To assemble or not to resemble—a validated Comparative Metatranscriptomics Workflow (CoMW). *GigaScience*, 8, 1–10.

Bajerski, F. & Wagner, D. (2013) Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. *FEMS Microbiology Ecology*, 85, 128–142.

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

Bardgett, R.D., Richter, A., Bol, R., Garnett, M.H., Baumler, R., Xu, X. et al. (2007) Heterotrophic microbial communities use ancient carbon following glacial retreat. *Biology Letters*, 3, 487–490.

Bernasconi, S.M., Bauder, A., Bourdon, B., Brunner, I., Bünemann, E., Christl, I. et al. (2011) Chemical and biological gradients along the Damma Glacier Soil Chronosequence, Switzerland. *Vadose Zone Journal*, 10, 863–887.

Bhattacharya, P., Tiwari, P., Talukdar, G. & Rawat, G.S. (2022) Shifts in bacterial community composition and functional traits at different time periods post-deglaciation of Gangotri Glacier, Himalaya. *Current Microbiology*, 79, 91.

Biely, P. (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances*, 30, 1575–1588.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Bradley, J.A., Singarayer, J.S. & Anesio, A.M. (2014) Microbial community dynamics in the forefield of glaciers. *Proceedings of the Biological Sciences*, 281, 20140882.

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

Brunner, I., Plotze, M., Rieder, S., Zumsteg, A., Furrer, G. & Frey, B. (2011) Pioneering fungi from the Damma glacier forefield in the Swiss Alps can promote granite weathering. *Geobiology*, 9, 266–279.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research*, 37, D233–D238.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P. et al. (2013) Long-term climate change: projections, commitments and irreversibility. In: *Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 1029–1136.

de los Rios, A., Wierzchos, J., Sancho, L.G. & Ascaso, C. (2003) Acid microenvironments in microbial biofilms of antarctic endolithic microecosystems. *Environmental Microbiology*, 5, 231–237.

De Vries, F.T., Thion, C., Bahn, M., Bergk Pinto, B., Cécillon, S., Frey, B. et al. (2021) Glacier forelands reveal fundamental plant and microbial controls on short-term ecosystem nitrogen retention. *Journal of Ecology*, 109, 3710–3723.

Delmont, T.O., Patrick, R., Ian, C., Pascal, S. & Timothy, M.V. (2011) Metagenomic comparison of direct and indirect soil DNA extraction approaches. *Journal of Microbiological Methods*, 86, 397–400.

Dey, P., Malik, A., Singh, D.K., Haange, S.B., von Bergen, M. & Jehmlich, N. (2022) Insight into the molecular mechanisms underpinning the mycoremediation of multiple metals by proteomic technique. *Frontiers in Microbiology*, 13, 872576.

Donhauser, J. & Frey, B. (2018) Alpine soil microbial ecology in a changing world. *FEMS Microbiology Ecology*, 94, fiy099.

Donhauser, J., Qi, W., Bergk-Pinto, B. & Frey, B. (2021) High temperatures enhance the microbial genetic potential to recycle C and N from necromass in high-mountain soils. *Global Change Biology*, 27, 1365–1386.

Duc, L., Noll, M., Meier, B.E., Burgmann, H. & Zeyer, J. (2009) High diversity of diazotrophs in the forefield of a receding alpine glacier. *Microbial Ecology*, 57, 179–190.

Duran, J., Rodriguez, A., Heiethmarsson, S., Lehmann, J.R.K., Del Moral, A., Garrido-Benavent, I. et al. (2021) Cryptogamic cover determines soil attributes and functioning in polar terrestrial ecosystems. *Science of the Total Environment*, 762, 143169.

Edwards, A. & Cook, S. (2015) Microbial dynamics in glacier forefield soils show succession is not just skin deep. *Molecular Ecology*, 24, 963–966.

El-Tarabily, K.A., Nassar, A.H. & Sivasithamparam, K. (2008) Promotion of growth of bean (Phaseolus vulgaris L.) in a calcareous soil by a phosphate-solubilizing, rhizosphere-competent isolate of Micromonospora endolithica. *Applied Soil Ecology*, 39, 161–171.

Ferreira, C.M.H., Vilas-Boas, A., Sousa, C.A., Soares, H. & Soares, E.V. (2019) Comparison of five bacterial strains producing siderophores with ability to chelate iron under alkaline conditions. *AMB Express*, 9, 78.

Frey, B., Bühler, L., Schmutz, S., Zumsteg, A. & Furrer, G. (2013) Molecular characterization of phototrophic microorganisms in the forefield of a receding glacier in the Swiss Alps. *Environmental Research Letters*, 8, 015033.

Frey, B., Carnol, M., Dharmarajah, A., Brunner, I. & Schleppi, P. (2020) Only minor changes in the soil microbiome of a subalpine forest after 20 years of moderately increased nitrogen loads. *Frontiers in Forests and Global Change*, 3, 77.

Frey, B., Rast, B.M., Qi, W., Stierli, B. & Brunner, I. (2022) Long-term mercury contamination does not affect the microbial gene potential for C and N cycling in soils but enhances detoxification gene abundance. *Frontiers in Microbiology*, 13, 1034138.

Frey, B., Rieder, S.R., Brunner, I., Plotze, M., Koetzsch, S., Lapanje, A. et al. (2010) Weathering-associated bacteria from the Damma glacier forefield: physiological capabilities and impact on granite dissolution. *Applied and Environmental Microbiology*, 76, 4788–4796.

Frey, B., Rime, T., Phillips, M., Stierli, B., Hajdas, I., Widmer, F. et al. (2016) Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiology Ecology*, 92, fiw018.

Frey, B., Varliero, G., Qi, W., Stierli, B., Walthert, L. & Brunner, I. (2022) Shotgun metagenomics of deep forest soil layers show evidence of altered microbial genetic potential for biogeochemical cycling. *Frontiers in Microbiology*, 13, 828977.

Frey, B., Walthert, L., Perez-Mon, C., Stierli, B., Köchli, R., Dharmarajah, A. et al. (2021) Deep soil layers of drought-exposed forests Harbor poorly known bacterial and fungal communities. *Frontiers in Microbiology*, 12, 674160.

Garrido-Benavent, I., Perez-Ortega, S., Duran, J., Ascaso, C., Pointing, S.B., Rodriguez-Cielos, R. et al. (2020) Differential colonization and succession of microbial communities in rock and soil substrates on a maritime Antarctic glacier Forefield. *Frontiers in Microbiology*, 11, 126.

Gee, G.W. & Bauder, J. (1986) Methods of soil analysis, part 1, physical and mineralogical methods. In: Klute, A. (Ed.) *Soil Science Society of America*. Madison: American Society of Agronomy.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J. & Stoffel, M. (2014) 21st century climate change in the European Alps—a review. *Science of the Total Environment*, 493, 1138–1151.

Göransson, H., Olde Venterink, H. & Bååth, E. (2011) Soil bacterial growth and nutrient limitation along a chronosequence from a glacier forefield. *Soil Biology and Biochemistry*, 43, 1333–1340.

Göransson, H., Welc, M., Bünemann, E.K., Christl, I. & Venterink, H.O. (2016) Nitrogen and phosphorus availability at early stages of soil development in the Damma glacier forefield, Switzerland; implications for establishment of N_2-fixing plants. *Plant and Soil*, 404, 251–261.

Guelland, K., Esperschütz, J., Bornhauser, D., Bernasconi, S.M., Kretzschmar, R. & Hagedorn, F. (2013) Mineralisation and leaching of C from 13C labelled plant litter along an initial soil chronosequence of a glacier forefield. *Soil Biology and Biochemistry*, 57, 237–247.

Guelland, K., Hagedorn, F., Smittenberg, R.H., Göransson, H., Bernasconi, S.M., Hajdas, I. et al. (2012) Evolution of carbon fluxes during initial soil formation along the forefield of Damma glacier, Switzerland. *Biogeochemistry*, 113, 545–561.

Haeberli, W., Maisch, M. & Paul, F. (2002) Mountain glaciers in global climate-related observation networks. *WMO Bulletin*, 51, 18–25.

Hindshaw, R.S., Tipper, E.T., Reynolds, B.C., Lemarchand, E., Wiederhold, J.G., Magnusson, J. et al. (2011) Hydrological control of stream water chemistry in a glacial catchment (Damma glacier, Switzerland). *Chemical Geology*, 285, 215–230.

Hodge, A., Robinson, D. & Fitter, A. (2000) Are microorganisms more effective than plants at competing for nitrogen? *Trends in Plant Science*, 5, 304–308.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C. et al. (2017) Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, 34, 2115–2122.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C. et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44, D286–D293.

Jacquiod, S., Stenbaek, J., Santos, S.S., Winding, A., Sorensen, S.J. & Prieme, A. (2016) Metagenomes provide valuable comparative information on soil microeukaryotes. *Research in Microbiology*, 167, 436–450.

Kelly, C.N., Schwaner, G.W., Cumming, J.R. & Driscoll, T.P. (2021) Metagenomic reconstruction of nitrogen and carbon cycling pathways in forest soil: influence of different hardwood tree species. *Soil Biology and Biochemistry*, 156, 108226.

Khedim, N., Cecillon, L., Poulenard, J., Barre, P., Baudin, F., Marta, S. et al. (2021) Topsoil organic matter build-up in glacier forelands around the world. *Global Change Biology*, 27, 1662–1677.

Knelman, J.E., Legg, T.M., O'Neill, S.P., Washenberger, C.L., González, A., Cleveland, C.C. et al. (2012) Bacterial community structure and function change in association with colonizer plants during early primary succession in a glacier forefield. *Soil Biology and Biochemistry*, 46, 172–180.

Lapanje, A., Wimmersberger, C., Furrer, G., Brunner, I. & Frey, B. (2012) Pattern of elemental release during the granite dissolution can be changed by aerobic heterotrophic bacterial strains isolated from Damma Glacier (central Alps) deglaciated granite sand. *Microbial Ecology*, 63, 865–882.

Lepleux, C., Turpault, M.P., Oger, P., Frey-Klett, P. & Uroz, S. (2012) Correlation of the abundance of betaproteobacteria on mineral surfaces with mineral weathering in forest soils. *Applied and Environmental Microbiology*, 78, 7114–7119.

Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M. et al. (2015) Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME Journal*, 9, 2490–2502.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*.

Liao, Y., Smyth, G.K. & Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923–930.

Liu, G.X., Hu, P., Zhang, W., Wu, X., Yang, X., Chen, T. et al. (2012) Variations in soil culturable bacteria communities and biochemical characteristics in the Dongkemadi glacier forefield along a chronosequence. *Folia Microbiologia (Praha)*, 57, 485–494.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42, D490–D495.

Lopez-Mondejar, R., Brabcova, V., Stursova, M., Davidova, A., Jansa, J., Cajthaml, T. et al. (2018) Decomposer food web in a deciduous forest shows high share of generalist microorganisms and importance of microbial biomass recycling. *The ISME Journal*, 12, 1768–1778.

Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I. et al. (2018) Function and functional redundancy in microbial systems. *Nature Ecology and Evolution*, 2, 936–943.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

McMurdie, P.J. & Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8, e61217.

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

Miltner, A., Bombach, P., Schmidt-Brücken, B. & Kästner, M. (2011) SOM genesis: microbial biomass as a significant source. *Biogeochemistry*, 111, 41–55.

Mosley, O.E., Gios, E., Close, M., Weaver, L., Daughney, C. & Handley, K.M. (2022) Nitrogen cycling and microbial cooperation in the terrestrial subsurface. *The ISME Journal*, 16, 2561–2573.

Nash, M.V., Anesio, A.M., Barker, G., Tranter, M., Varliero, G., Eloe-Fadrosh, E.A. et al. (2018) Metagenomic insights into diazotrophic communities across Arctic glacier forefields. *FEMS Microbiology Ecology*, 94, fiy114.

Ni, G., Lappan, R., Hernandez, M., Santini, T., Tomkins, A.G. & Greening, C. (2022) Functional basis of primary succession: traits of the pioneer microbes. *Environmental Microbiology*, 25, 171–176.

Nicol, G.W., Leininger, S., Schleper, C. & Prosser, J.I. (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environmental Microbiology*, 10, 2966–2978.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R. et al. (2011) vegan: Community Ecology Package (R package version 2.0. 2). http://CRAN.R-projectorg/package=vegan

Olsson-Francis, K., Boardman, C.P., Pearson, V.K., Schofield, P.F., Oliver, A. & Summers, S. (2015) A culture-independent and culture-dependent study of the bacterial community from the bedrock soil interface. *Advances in Applied Microbiology*, 05, 842–857.

Opitz, S., Kusel, K., Spott, O., Totsche, K.U. & Herrmann, M. (2014) Oxygen availability and distance to surface environments determine community composition and abundance of ammonia-oxidizing prokaryotes in two superimposed pristine limestone aquifers in the Hainich region, Germany. *FEMS Microbiology Ecology*, 90, 39–53.

Ortiz-Alvarez, R., Fierer, N., de Los Rios, A., Casamayor, E.O. & Barberan, A. (2018) Consistent changes in the taxonomic structure and functional attributes of bacterial communities during primary succession. *The ISME Journal*, 12, 1658–1667.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25, 1043–1055.

Perez-Mon, C., Stierli, B., Plotze, M. & Frey, B. (2022) Fast and persistent responses of alpine permafrost microbial communities to in situ warming. *Science of the Total Environment*, 807, 150720.

Pothula, S.K. & Adams, B.J. (2022) Community assembly in the wake of glacial retreat: a meta-analysis. *Global Change Biology*, 28, 6973–6991.

Pruesse, E., Peplies, J. & Glockner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28, 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590–D596.

R Core team. (2022) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rime, T., Hartmann, M., Brunner, I., Widmer, F., Zeyer, J. & Frey, B. (2015) Vertical distribution of the soil microbiota along a successional gradient in a glacier forefield. *Molecular Ecology*, 24, 1091–1108.

Rime, T., Hartmann, M. & Frey, B. (2016) Potential sources of microbial colonizers in an initial soil ecosystem after retreat of an alpine glacier. *The ISME Journal*, 10, 1625–1641.

Rime, T., Hartmann, M., Stierli, B., Anesio, A.M. & Frey, B. (2016) Assimilation of microbial and plant carbon by active prokaryotic and fungal populations in glacial forefields. *Soil Biology and Biochemistry*, 98, 30–41.

Rubio, L.M. & Ludden, P.W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology*, 62, 93–111.

Rüthi, J., Rast, B.M., Qi, W., Perez-Mon, C., Pardi-Comensoli, L., Brunner, I. et al. (2023) The plastisphere microbiome in alpine soils alters the microbial genetic potential for plastic degradation and biogeochemical cycling. *Journal of Hazardous Materials*, 441, 129941.

Samuels, T. (2018) *Microbial weathering of shale rock in natural and historic industrial environments*. Edinburgh: University of Edinburgh.
Schimel, J.P., Gulledge, J.M., Clein-Curley, J.S., Lindstrom, J.E. & Braddock, J.F. (1999) Moisture effects on microbial activity and community structure in decomposing birch litter in the Alaskan taiga. *Soil Biology and Biochemistry*, 31, 831–838.
Schimel, J.P. & Schaeffer, S.M. (2012) Microbial control over carbon cycling in soil. *Frontiers in Microbiology*, 3, 348.
Schmidt, S.K., Reed, S.C., Nemergut, D.R., Grandy, A.S., Cleveland, C.C., Weintraub, M.N. et al. (2008) The earliest stages of ecosystem succession in high-elevation (5000 metres above sea level), recently deglaciated soils. *Proceedings of the Biological Sciences*, 275, 2793–2802.
Schulz, S., Brankatschk, R., Dümig, A., Kögel-Knabner, I., Schloter, M. & Zeyer, J. (2013) The role of microorganisms at different stages of ecosystem development for soil formation. *Biogeosciences*, 10, 3983–3996.
Schutte, U.M., Abdo, Z., Bent, S.J., Williams, C.J., Schneider, G.M., Solheim, B. et al. (2009) Bacterial succession in a glacier foreland of the High Arctic. *The ISME Journal*, 3, 1258–1268.
Schutte, U.M., Abdo, Z., Foster, J., Ravel, J., Bunge, J., Solheim, B. et al. (2010) Bacterial diversity in a glacier foreland of the high Arctic. *Molecular Ecology*, 19, 54–66.
Seneviratne, G.I. & Indrasena, I.K. (2006) Nitrogen fixation in lichens is important for improved rock weathering. *Journal of Biosciences*, 31, 639–643.
Smittenberg, R.H., Gierga, M., Göransson, H., Christl, I., Farinotti, D. & Bernasconi, S.M. (2012) Climate-sensitive ecosystem carbon dynamics along the soil chronosequence of the Damma glacier forefield, Switzerland. *Global Change Biology*, 18, 1941–1955.
Tian, C., Lv, Y., Yang, Z., Zhang, R., Zhu, Z., Ma, H. et al. (2023) Microbial community structure and metabolic potential at the initial stage of soil development of the glacial Forefields in Svalbard. *Microbial Ecology*, 86, 933–946.
Tu, Q., Lin, L., Cheng, L., Deng, Y. & He, Z. (2019) NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics*, 35, 1040–1048.
Tveit, A., Schwacke, R., Svenning, M.M. & Urich, T. (2013) Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *The ISME Journal*, 7, 299–311.
Uroz, S., Calvaruso, C., Turpault, M.P. & Frey-Klett, P. (2009) Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends in Microbiology*, 17, 378–387.
Uroz, S., Kelly, L.C., Turpault, M.P., Lepleux, C. & Frey-Klett, P. (2015) The mineralosphere concept: mineralogical control of the distribution and function of mineral-associated bacterial communities. *Trends in Microbiology*, 23, 751–762.
Vance, E.D., Brookes, P.C. & Jenkinson, D.S. (1987) An extraction method for measuring soil microbial biomass C. *Soil Biology and Biochemistry*, 19, 703–707.
VanInsberghe, D., Maas, K.R., Cardenas, E., Strachan, C.R., Hallam, S.J. & Mohn, W.W. (2015) Non-symbiotic Bradyrhizobium ecotypes dominate North American forest soils. *The ISME Journal*, 9, 2435–2441.
Varliero, G., Anesio, A.M. & Barker, G.L.A. (2021) A taxon-wise insight into rock weathering and nitrogen fixation functional profiles of proglacial systems. *Frontiers in Microbiology*, 12, 627437.
Walker, L.R. & del Moral, R. (2010) *Primary succession and ecosystem rehabilitation*. Cambridge: Cambridge University Press.
Wang, Y., Chen, Y., Xue, Q., Xiang, Q., Zhao, K., Yu, X. et al. (2021) The abundance of the nifH gene became higher and the nifH-containing diazotrophic bacterial communities changed during primary succession in the Hailuogou Glacier Chronosequence, China. *Frontiers in Microbiology*, 12, 672656.
Welch, S.A., Barker, W.W. & Banfield, J.F. (1999) Microbial extracellular polysaccharides and plagioclase dissolution. *Geochimica et Cosmochimica Acta*, 63, 1405–1419.
Welch, S.A., Taunton, A.E. & Banfield, J.F. (2002) Effect of microorganisms and microbial metabolites on apatite dissolution. *Geomicrobiology Journal*, 19, 343–367.
Wickham, H. (2011) ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3, 180–185.
Wierzchos, J., Casero, M.C., Artieda, O. & Ascaso, C. (2018) Endolithic microbial habitats as refuges for life in polyextreme environment of the Atacama Desert. *Current Opinion in Microbiology*, 43, 124–131.
Wild, B., Gerrits, R. & Bonneville, S. (2022) The contribution of living organisms to rock weathering in the critical zone. *npj Materials Degradation*, 6, 98.
Zhong, Z.P., Tian, F., Roux, S., Gazitua, M.C., Solonenko, N.E., Li, Y.F. et al. (2021) Glacier ice archives nearly 15,000-year-old microbes and phages. *Microbiome*, 9, 160.
Zhu, W., Lomsadze, A. & Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, 38, e132.
Zumsteg, A., Luster, J., Goransson, H., Smittenberg, R.H., Brunner, I., Bernasconi, S.M. et al. (2012) Bacterial, archaeal and fungal succession in the forefield of a receding glacier. *Microbial Ecology*, 63, 552–564.
Zumsteg, A., Schmutz, S. & Frey, B. (2013) Identification of biomass utilizing bacteria in a carbon-depleted glacier forefield soil by the use of 13C DNA stable isotope probing. *Environmental Microbiology Reports*, 5, 424–437.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Feng, M., Varliero, G., Qi, W., Stierli, B., Edwards, A., Robinson, S. et al. (2023) Microbial dynamics in soils of the Damma glacier forefield show succession in the functional genetic potential. *Environmental Microbiology*, 25(12), 3116–3138. Available from: https://doi.org/10.1111/1462-2920.16497