Metabarcoding data reveal vertical multitaxa variation in topsoil communities during the colonization of deglaciated forelands

Alessia Guerrieri1 | Alexis Carteron1 | Aurélie Bonin1,2 | Silvio Marta1 | Roberto Ambrosini1 | Marco Caccianiga1 | Isabel Cantera1 | Chiara Compostella3 | Guglielmina Diolaiuti1 | Diego Fontaneto4 | Ludovic Gielly5 | Fabrizio Gili1 | Mauro Gobbi6 | Jerome Poulenard7 | Pierre Taberlet5,8 | Andrea Zerboni3 | Wilfried Thuiller5 | Gentile Francesco Ficetola1,5

Abstract

Ice-free areas are expanding worldwide due to dramatic glacier shrinkage and are undergoing rapid colonization by multiple lifeforms, thus representing key environments to study ecosystem development. It has been proposed that the colonization dynamics of deglaciated terrains is different between surface and deep soils but that the heterogeneity between communities inhabiting surface and deep soils decreases through time. Nevertheless, tests of this hypothesis remain scarce, and it is unclear whether patterns are consistent among different taxonomic groups. Here, we used environmental DNA metabarcoding to test whether community diversity and composition of six groups (Eukaryota, Bacteria, Mycota, Collembola, Insecta, and Oligochaeta) differ between the surface (0–5 cm) and deeper (7.5–20 cm) soil at different stages of development and across five Alpine glaciers. Taxonomic diversity increased with time since glacier retreat and with soil evolution. The pattern was consistent across groups and soil depths. For Eukaryota and Mycota, alpha-diversity was highest at the surface. Time since glacier retreat explained more variation of community composition than depth. Beta-diversity between surface and deep layers decreased with time since glacier retreat, supporting the hypothesis that the first 20 cm of soil tends to homogenize through time. Several molecular operational taxonomic units of bacteria and fungi were significant indicators of specific depths and/or soil development stages, confirming the strong functional variation of microbial communities through time and depth. The complexity of community patterns highlights the importance of integrating information from multiple taxonomic groups to unravel community variation in response to ongoing global changes.

[Correction added on 29-November-2022, after first online publication: funding statement has been added.]
Assessing both the vertical and the horizontal composition and distribution of topsoil colonizers is pivotal to infer the key ecological processes of the succession occurring after the early years of glacier retreat. Rime et al. (2015) performed a rare attempt to integrate soil depth into the study of Alpine primary succession (see also Bajerski & Wagner, 2013; Schütte et al., 2009). They assessed the structure of microbial communities along one glacier foreland and found that soil depth and development stage interact in shaping the biodiversity of bacteria and fungi. Differences between communities from surface and deep layers were particularly strong immediately after glacier retreat but decreased at older soil development stages, with a homogenization of communities through time. However, Rime et al. (2015) focused only on microorganisms and considered just one glacier foreland. As different topsoil organisms can have very different responses (Cauvy-Fraunié & Dangles, 2019; Donald et al., 2021; Ficetola et al., 2021; Rosero et al., 2021), the study of multiple taxa is needed for a better understanding of the ecological processes governing community development after glacier retreat. With appropriate technical precautions, environmental DNA (eDNA) metabarcoding allows us to analyse both micro- and macro-organisms of soils across a wide range of natural systems (Taberlet et al., 2018). The possibility of combining multiple metabarcodes makes eDNA a particularly powerful tool for estimating soil diversity across multiple taxa (Donald et al., 2021). This approach helps to overcome several limitations of conventional sampling and enables relatively fast and cost-efficient data production. Furthermore, the analyses can be conducted over broad geographical scales and from remote areas (e.g., Zinger, Taberlet, et al., 2019), and taxonomic inventories can be related to environmental characteristics in order to infer ecological processes.

Here, we used metabarcoding data from soil eDNA to study the vertical distribution of animals and microorganisms within the top 20 cm of soil, where most of the microbial diversity has been retrieved and where most soil invertebrates live (Fierer et al., 2003; Menta, 2012). We studied communities at different stages of soil development to understand how their vertical distribution varies through time in deglaciated areas. First, we tested whether and how the alpha-diversity of multiple taxa changes with soil depth and development stage, and whether the pattern is similar among different depths through time. We expected the alpha-diversity of communities to decrease in deep soils, as observed for microbes (Zhao et al., 2021), and increase through time, as soil develops (Khedim et al., 2021). Second, we evaluated the differences in community composition between surface and deep layers, and tested for potential taxonomic characteristics of the different depths or stages of soil development. Rime et al. (2015) proposed that differences between surface and deep soil decrease along successions, with homogenization of communities through time, but the generality of these
conclusions needs to be tested. We analysed the beta-diversity between surface and deep layers for six taxonomic groups representing a large proportion of biodiversity. If Rime's homogenization hypothesis applies to the whole biota, we expect that beta-diversity between surface and deep layers decreases from recent to more developed terrains, with a consistent pattern across taxa. Such information will provide key insights into the long-term responses of multiple topsoil organisms to glacier retreat.

2 | MATERIALS AND METHODS

2.1 | Sample collection and preservation

In Summer 2018, we collected 280 soil samples from five Alpine forelands (Figure 1): Amola (coordinates for the centre of the foreland: 46.215°N, 10.697°E), Morteratsch (46.438°N, 9.936°E), Rutor (46.669°N, 6.992°E), Sforzellina (46.351°N, 10.510°E) and Grande di Verra (45.895°N, 7.749°E). For each foreland, we selected three to eight sites corresponding to the line occupied by the glacier front at different dates. The position of the glacier at a given time point is known on the basis of the literature, dated images and field surveys; we focused on the period between the end of the Little Ice Age (~1850) and recent years of glacier retreat (Marta et al., 2021). Sites were thus representative of different stages of soil development depending on the time elapsed between sampling activities and the retreat of glaciers (hereafter referred to as “time since glacier retreat,” ranging from 12 to 168 years).

Along the line that represents each site, we established five regularly spaced plots at distances of about 20m. All plots within the same site were approximately at the same distance from the glacier front and had the same age since glacier retreat (see also figure 1 of Ficetola et al., 2021). At each plot, we collected five soil cores within 1-m distance and we kept the 0–5 and 7.5–20-cm portions to be representative of two different soil depths, hereafter called “surface” and “deep” soils, respectively (Figure 1). For each of the five cores, we pooled portions of the same depth together (~40g) to form one composite sample of ~200g per plot and homogenized it. This strategy maximizes the detection of differences between soil layers (see Edwards & Cook, 2015) while limiting the number of analysed

FIGURE 1  Study area and sampling design. Maps with pseudocolour representation of altitude (source: 30 arc-sec digital elevation model) and land cover (Source: Copernicus Sentinel data 2019) created by AG, SM and GFF.
samples. Then, we took 15 g of soil from each composite sample and desiccated it immediately in sterile boxes with 40 g of silica gel. This approach enables cost-effective and long-term preservation of soil eDNA (Guerrieri et al., 2021). Soil collection was performed wearing gloves and the sampling tools were decontaminated with a portable blow torch (>1000°C) before the collection of each sample. We did not include soil litter and avoided roots, leaves and other large plant organs.

Soils were sampled “by depth” rather than “by horizons,” as is common practice in eDNA-based studies (Dickie et al., 2018) and in soil monitoring programmes involving multiple glacier forelands (e.g., Khedim et al., 2021; Orgiazzi et al., 2018; Rime et al., 2015; Schweizer et al., 2018), because soil horizons are not yet differentiated at early stages of soil development, and because this approach provides a standardized pattern that can be applied across soils from multiple areas at very different development stages (Dickie et al., 2018; Khedim et al., 2021; Rime et al., 2015). Thus, the two categories “surface” (0–5 cm) and “deep” (7.5–20 cm) are used to define soil samples collected at two different depths, regardless of the horizons.

At each site and for each plot we also collected one composite sample (depth: 0–20 cm) to measure organic carbon. For each sample, the organic carbon was measured by a Thermo Fisher Flash sample (depth: 0–20 cm) to measure organic carbon. For each sample, organic soil samples collected at two different depths, regardless of the horizons (Figure S1).

2.2 Molecular analyses

In a dedicated room, we mixed the 15 g of soil with 20 ml of phosphate buffer for 15 min as described in Taberlet et al. (2012); then we extracted eDNA using the NucleoSpin Soil Mini Kit (Macherey-Nagel) with a final elution in 150 μl and with one negative extraction control every 23 samples (total: 12).

First, we amplified eDNA using generalist primers, targeting a broad range of soil organisms (bacteria and eukaryotes) to obtain an overview of the overall variation in biodiversity across the whole tree of life. The generalist primers were: Bact02 (16S rDNA, amplifying Bacteria; Taberlet et al., 2018) and Euka02 (18S rDNA, Eukaryota; Guardiola et al., 2015). Subsequently, we added four specific primers targeting key taxa within eukaryotes (fungi, springtails, insects and earthworms; Table 1) to obtain a better resolution of organisms that have a particularly important role in soil food webs. The specific primers were: Fung02 (ITS1, amplifying Mycota; Epp et al., 2012), Coll01 (16S, Collombola, i.e., springtails; Janssen et al., 2018), Inse01 (16S, Insecta; Taberlet et al., 2018) and Olig01 (16S, Oligochaeta, i.e., earthworms; Bienert et al., 2012). All these markers are well suited for metabarcoding analyses thanks to the very conserved priming regions across target organisms, and they perform well with degraded DNA due to the relatively short length of amplified fragments (Taberlet et al., 2018; Table 1). Inse01 is particularly suited for metabarcoding of insects as it shows consistently high amplification success across most orders, and is well represented in reference databases (Ficetola et al., 2021). We used forward and reverse primers tagged on the 5′-end with eight-nucleotide-long tags having at least five nucleotide differences among them (Coissac, 2012). All PCR (polymerase chain reaction) replicates were represented by a unique combination of forward and reverse primer tags. This allowed us to uniquely identify each PCR replicate after sequencing. We randomized all samples on 96-well plates and included 24 bioinformatic blanks, 12 PCR-negative controls and one PCR-positive control. The positive control was composed of genomic DNA of eight bacterial and two fungal strains (i.e., Pseudomonas aeruginosa, Escherichia coli, Salmonella enterica, Lactobacillus fermentum, Enterococcus faecalis, Staphylococcus aureus, Listeria monocytogenes, Bacillus subtilis, Saccharomyces cerevisiae, Cryptococcus neoformans) at known concentrations (ZymoBIOMICs Microbial Community DNA Standard II, Zymo Research; diluted 1:10) and we used it to check for potential cross-contaminations and to monitor amplification and sequencing performance.

We determined the optimal number of amplification cycles and DNA dilution by conducting a qPCR essay on 48 randomly selected samples, using 1 μl of 1:1000 diluted SYBR Green I nucleic acid gel stain (Invitrogen), and both undiluted and 1:10 diluted DNA, with a real-time PCR thermal cycler set to standard mode. This step is useful to avoid over-amplifying eDNA and to limit chimera formation. Based on qPCRs results, we finally performed 42 (Bact02), 45 (Euka02, Fung02) or 55 (Coll01, Inse01, Olig01) amplification cycles on diluted (Euka02, Coll01) or undiluted (Bact02, Fung02, Inse01, Olig01) DNA. Amplification consisted of 20-μl reactions with 10 μl of AmpliTaq Gold 360 Master Mix 2x (Applied Biosystems), 2 μl of forward and reverse primer mix (initial concentration of each primer: 5 μM), 0.16 μl of bovine serum albumin (i.e., 3.2 μg; Roche Diagnostic) and 2 μl of eDNA. We performed reactions in 384-well plates, with four PCR replicates per sample (Ficetola et al., 2015), setting the following PCOR profiles: an initial step of 10 min at 95°C; several cycles of 30 s denaturation at 95°C; 30 s annealing at 53°C (Bact02), 45°C (Euka02), 56°C (Fung02), 51°C (Coll01), 55°C (Olig01) or 52°C (Inse01); 90s elongation for Bact02 and Olig02, or 60s elongation for all the others markers at 72°C; and final elongation at 72°C for 7 min. After amplification, we pooled together all amplicons of the same marker and visualized a 5 μM aliquot by high-resolution capillary electrophoresis (QIAxcel Advanced System, Qiagen) to check fragment lengths and monitor primer dimers. Finally, for each marker, we purified six subsamples of the pooled amplicons separately, using the MinElute PCR Purification Kit (Qiagen) as per the manufacturer’s instructions and combined them. Libraries were prepared following the MetaFast protocol (Taberlet et al., 2018) and sequenced using the MiSeq (Bact02 and Fung02) or HiSeq 2500 (all others) Illumina platforms (Illumina) with a paired-end approach (2 × 250 bp for Bact02 and Fung02, and 2 × 150 bp for the others markers) at Fasteris. For each marker, the average sequence depth was ~10,000 reads per PCR replicate.
TABLE 1 Characteristics of metabarcodes used for this study

<table>
<thead>
<tr>
<th>Marker</th>
<th>Target group</th>
<th>Target gene</th>
<th>Forward primer/Reverse primer</th>
<th>Min. metabarcode length (bp)</th>
<th>Max. metabarcode length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bact02</td>
<td>Bacteria</td>
<td>V4 region, 16S rDNA gene</td>
<td>GCCAGCMGCGCCGCTAA/</td>
<td>45</td>
<td>748</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GGAAGATCAGGCTTTAGGGAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euka02</td>
<td>Eukaryota</td>
<td>V7 region, 18S rDNA gene</td>
<td>TTTGTCGTTAAAATAG/G</td>
<td>36</td>
<td>892</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAAGAGATCAGGCTTTAGGGAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fung02</td>
<td>Fungi</td>
<td>ITS1 nuclear rDNA gene</td>
<td>GGAAGATGCTACCTTACGAAAGG/</td>
<td>68</td>
<td>919</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAAGAGATCAGGCTTTAGGGAAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coll01</td>
<td>Collembola</td>
<td>16S mitochondrial rDNA gene</td>
<td>ACCGTGTTATCCTTACAG/G</td>
<td>76</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CACGATAACCGCTTATAGTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inse01</td>
<td>Insecta</td>
<td>16S mitochondrial rDNA gene</td>
<td>RGAGAGAAGAGACCCCTATATA/</td>
<td>75</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACGCTGGTATACCTTACAGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olig01</td>
<td>Oligochaeta</td>
<td>16S mitochondrial rDNA gene</td>
<td>CAAGAGGACCCCTATAGAGCTTC/</td>
<td>93</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCTGTTATCCCTAAGGACTR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Excluding primers.

2.3 Bioinformatic treatment

We used the OBITools software suite (Boyer et al., 2016) to perform the bioinformatic treatment of raw sequence data, as follows. First, we assembled the forward and reverse reads using the illuminapiarender program and kept only sequences with an alignment score > 40 (corresponding to a 10-nucleotide overlap of the forward and reverse reads). Second, we assigned aligned sequences to the corresponding PCR replicate using the program ncssfilter and allowed two and zero mismatches on primers and tags, respectively. Third, we dereplicated sequences using obiuniq and discarded low-quality sequences (i.e., containing "N"). Sequences whose length was lower or higher than expected (based on the minimum and maximum metabarcode length; Table 1) and singletons (i.e., spurious sequences only occurring once; Bálint et al., 2014). Fourth, we ran the obiclean program with the option -r set at 0.5 to detect potential PCR or sequencing errors and keep only the sequences tagged as "heads" in at least one PCR. Sequences are tagged as "heads" when they are at least twice (-r option set at 0.5) as abundant as other related sequences differing by one base in the same PCR. Fifth, we clustered sequences at a threshold of 96% (Bact02, Euka02, Inse01), 95% (Fung02), 92% (Olig01) or 85% (Coll01) sequence similarity using the sumaclust program (https://git.metabarcoding.org/obitools/sumaclust/wiki/home). These thresholds were selected for each taxon on the basis of the distributions of pairwise sequence similarities within and between species of the same genus for each marker (Bonin et al., 2022). The selected thresholds allow us to minimize the risk that multiple individuals belonging to the same species are assigned to different molecular operational taxonomic units (MOTUs), while limiting the probability that distinct species are collapsed in one single MOTU (Bonin et al., 2022). Differences in thresholds are related to differences in taxonomic resolution across markers.

For the taxonomic assignment, we built for each marker a sequence reference database from EMBL (version 140), as follows. First, we ran the ecopcr program (Ficetola et al., 2010) to carry out an in silico PCR with the primer pairs used for the experiment, allowing three mismatches per primer. Then, we curated the obtained reference databases by keeping only sequences assigned at the species, genus and family levels. Finally, the taxonomic assignment was performed by the ecotag program on each sequence using the reference database, and only considering MOTUs with a best identity ≥ 80% (Bact02, Euka02, Fung02) or ≥ 60% (Coll01, Inse01, Olig01).

When working with eDNA metabarcoding, many sequences detected at low frequency are artefacts produced by PCR, contaminants and sequencing errors that can inflate the rare biosphere (Bálint et al., 2016; Brown et al., 2015; Zinger, Bonin, et al., 2019). To remove spurious sequences and avoid bias in ecological conclusions (Calderón-Sanou et al., 2020) we performed additional filtering in version 4.0 (R Core Team, 2018) and discarded MOTUs observed less than five (Bact02, Fung02, Inse01), 10 (Olig01), 11 (Coll01) or 12 (Euka02) times overall. These values correspond to the minimum number of reads that removed ≥ 99.99% of sequences detected in the blanks (i.e., tag-jump errors) of each experiment (i.e., each marker) and thus differ among markers. Furthermore, we discarded MOTUs detected in only one sample, as they might represent spurious sequences (Bálint et al., 2016), MOTUs detected in fewer than two PCR replicates of the same sample, as they often represent false positives (Ficetola et al., 2015), and MOTUs detected in more than one extraction or PCR-negative control, as they might represent contaminants (Zinger, Bonin, et al., 2019).

2.4 Statistical analyses

At each plot and for each depth, we measured the alpha-diversity of Bacteria, Eukaryota, Mycota, Collembola, Insecta and Oligochaeta, through Hill numbers. In metabarcoding studies, the joint use of multiple Hill numbers allows biodiversity measures to be obtained that are robust to bioinformatic treatments and other methodological...
choices (Alberdi & Gilbert, 2019; Calderón-Sanou et al., 2020; Mächler et al., 2021). We used the parameters \( q = 0 \) and \( q = 1 \) in the \textit{hill_taxa} function of the \textit{hillr} package (Chao et al., 2014). A value of \( q = 0 \) returns the taxonomic richness and is insensitive to MOTU frequency, while \( q = 1 \) returns the exponential of the Shannon diversity, and limits the weight of rare MOTUs (Alberdi & Gilbert, 2019). We could not use values of \( q > 1 \) because they cannot be applied to communities with richness = 0.

We used univariate Bayesian Generalized Linear Mixed Models (GLMMs) to assess the relationship between time since glacier retreat and depth, and the alpha-diversity of Bacteria, Eukaryota, Mycota, Collembola, Insecta and Oligochaeta. In GLMMs, the alpha-diversity of each sample (log-transformed and scaled: mean = 0, SD = 1) was the dependent variable, and we used a Gaussian error, considering both the parameters \( q = 0 \) and \( q = 1 \). As independent variables, we considered time since glacier retreat (log-transformed and scaled: mean = 0, SD = 1) and depth. We included glacier identity and plot identity as random factors. GLMMs also included the interaction between depth and time since glacier retreat, to test the hypothesis that depth affects the colonization rate of the studied groups. We used the widely applicable information criterion (WAIC) to compare models with and without interactions (Gelman et al., 2013). Models using a log-normal error distribution and untransformed alpha-diversity values yielded identical results.

Soil nutrient content changes at different stages of soil development in deglaciated areas, with the amount of organic carbon increasing through time (Khedim et al., 2021) and potentially influencing the alpha-diversity of communities (Guo et al., 2018). We therefore re-analysed the pattern of alpha-diversity using organic carbon as an independent variable, instead of time since glacier retreat. This analysis was run for a subset of samples (\( N = 276 \)) for which data of total organic carbon content were available. Organic carbon was strongly related to time since glacier retreat (GLMM with organic carbon as dependent variable and age as independent variable: \( R^2 = .66 \)), so it was impossible to include organic carbon and age as independent variables in the same model. Organic carbon data were representative of the overall soil core (0–20 cm); thus, the analysis did not allow us to test the role of variation in carbon content between surface and deep layers. Two plots (i.e., four samples in total) from one site of the Amola glacier foreland were excluded from this analysis because no soil data were available.

For each plot, we estimated the beta-diversity between the two soil depths based on incidence data. We used the \textit{beta.multi} function of the \textit{betapart} package with the Sörensken family (Baselga & Orme, 2012). This function partitions the total beta-diversity (beta.SOR) into its nestedness (beta.SNE) and turnover (beta.SIM) components, reflecting the species gain/loss and replacement, respectively (Baselga, 2010). We excluded plots having zero MOTUs in at least one depth, given that the formula of Baselga’s partitioning retrieves undefined values when one of the compared communities has no taxa (Baselga, 2010). For each taxonomic group, we built a GLMM to test the hypothesis that beta-diversity between the two soil depths decreases with time since glacier retreat. We ran mixed models with rescaled indices (Smithson & Verkuilen, 2006) to avoid fixed zeros and ones, using a beta distribution, and included glacier identity as a random factor. Models for beta-diversity were limited to plots with at least one detected MOTU in both depths. We then repeated this analysis for the turnover and nestedness components of beta-diversity. We ran all GLMMs with three Markov chain Monte Carlo (MCMC) chains, 5000 iterations and a burn-in of 2500 in the \textit{brms} package (Bürkner, 2017). After processing, \( \phi \) values were always < 1.01, indicating convergence.

To visualize the variation in the composition of below-ground communities across different stages of soil development, we used distance-based principal component analysis (db-PCA). We calculated distances between samples using the Sørensen index (Legendre & Legendre, 2012). As for the beta-diversity analysis, we removed plots having zero MOTUs in at least one depth. To test for differences in communities across time, depth and their potential interaction, we performed a permutational multivariate analysis of variance (PERMANOVA) using the \textit{adonis} function of the \textit{vegan} package (Oksanen et al., 2019) with glaciers as strata and permutations set to 9999. Time was log-transformed. Results of PERMANOVA can be sensitive to differences in multivariate dispersion (Anderson, 2001), and therefore we computed the homogeneity of variance among groups (Anderson, 2006) and tested for its significance by permutations (\( n = 9999 \)). We used data visualization in ordination plots to support the interpretation of the statistical tests.

Finally, we used the indicator value (IndVal; Dufrêne & Legendre, 1997) approach to identify MOTUs that were characteristic for particular stages of soil development and/or soil depth. Prior to the analysis, we grouped samples into three classes based on the time since glacier retreat (i.e., \(< 40, 40–95 \) and \( > 95 \) years) and depth (surface/deep), for a total of six environmental classes. Metabarcoding approaches can lead to a very large number of MOTUs, and very rare MOTUs could not be appropriate as indicators (Caro & O’doherty, 1999). Thus, for this analysis only, we focused on MOTUs with a relative abundance > 0.1% for each taxonomic group, to reduce the number of candidate indicator species and increase statistical power. We computed the IndVal index using the \textit{indicspecies} package (De Cáceres & Legendre, 2009). For a given taxon, the IndVal index is based on its specificity (i.e., the concentration of abundance) and fidelity (i.e., the relative frequency) within a class. Each MOTU could be an indicator of a maximum of two environmental classes (De Cáceres et al., 2010), so that a MOTU could be an indicator of, for example, one or both depths at a given soil stage, or of one or two consecutive stages at a given depth. This choice allowed us to keep the number and ecological meaningfulness of the combinations reasonable. The significance of indicator values was tested through random permutations (\( n = 9999 \)) and \( p \)-values were adjusted for multiple comparison tests using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995). We used the R packages \textit{ggplot2} (Wickham, 2016), \textit{ggpubr} (Kassambara, 2020), \textit{phyloseq} (McMurtrie & Holmes, 2013) and \textit{vegan} (Oksanen et al., 2019) for multivariate analyses and visualization.
RESULTS

The samples used for this study were part of a larger data set of 1488 samples, for which Illumina sequencing yielded a total of 141,463,933 (Bact02), 169,700,440 (Euka02), 183,563,263 (Fung02), 155,801,809 (Coll01), 136,279,617 (Inse01) and 134,857,799 (Olig01) raw sequence reads. After clustering and removal of spurious sequences, DNA metabarcoding yielded 1825 (Bact02), 753 (Euka02), 1483 (Fung02), 118 (Coll01), 396 (Inse01) and 97 (Olig01) high quality MOTUs available for the purposes of the present study (see Table S1).

3.1 How is alpha-diversity related to soil depth, time since glacier retreat and soil features?

Overall, alpha-diversity was highest for the generalist markers (Euka02, Bact02, Fung02) compared to the specialist markers (Coll01, Inse01, Olig01). Estimates of alpha-diversity obtained with different Hill numbers (q = 0 and q = 1) were strongly correlated (for all taxonomic groups, r > 0.78; Table S2).

When we used q = 1, we observed an increase in alpha-diversity with time since glacier retreat for all the taxonomic groups. For Eukaryota and Mycota, alpha-diversity was significantly higher in communities retrieved in the surface, compared to communities retrieved in the deep layer (Figure 2; Table S3). For Collembola we detected an interaction between depth and time since glacier retreat. For this group, alpha-diversity was close to one (mean: 1.19 ± 0.51; corresponding to richness ~0) at relatively young sites (<30 years after glacier retreat) and increased with time, but the increase was faster in communities at 0–5 cm of depth. All results were highly consistent when repeating the analyses using q = 0 (Table S3). Similarly, when we used the average soil carbon content of the plot as predictor instead of time since glacier retreat, GLMMs showed a significant increase in alpha-diversity with carbon, even though the R² values of these models were generally lower than the R² of models where time was the predictor (Figure S3; Table S4).

3.2 How does beta-diversity change through time?

To assess the beta-diversity between surface and deep soils, plots having zero MOTUs in at least one depth were removed, corresponding to 4.3% (Bacteria), 14.3% (Eukaryota), 6.4% (Mycota), 39.3% (Collembola), 13.6% (Insecta) and 37.9% (Oligochaeta) of total plots.

GLMMs showed that community dissimilarity between the two depths decreased with time since glacier retreat for Bacteria, Eukaryota, Mycota and Insecta, indicating a homogenization of communities. We did not detect significant changes through time for the beta-diversity of Collembola and Oligochaeta (Figure 3; Table S5), for which the largest number of sites were discarded because of a lack of MOTUs, especially in young soils (Figure S2).

Overall, our models did not show significant changes in the turnover of nestedness components of beta-diversity through time, with the only exception of Oligochaeta, for which nestedness between surface and deep soils increased through time (Table S5 and Figure S4).

Within each foreland, the composition of communities was primarily related to time since glacier retreat (Figure 4). Time significantly affected community composition of Bacteria, Mycota and Eukaryota (PERMANOVA: p < .05; Table 2); the amount of variance explained by time ranged from 2.3% to 5.9%. For Bacteria, Mycota and Eukaryota, community composition also differed significantly between soil depths, but the explained variance was smaller (<1%; Table 2). For none of the groups the interaction between time and soil depth was significant (Table 2), suggesting that the pattern was consistent between surface and deep soils. Differences in multivariate dispersion were never significant across depths, but were significant across time, except for Collembola (Table 2). Bacterial community composition was the most closely related to time and depth (R² = 7%; Table 2). Differences among deglaciated forelands were marked but tended to follow similar trends across the taxonomic groups (Figure 4).

Based on the specificity and fidelity of each MOTU for the selected classes of soil development stage and depth, 86 taxa were identified as indicators (47 Bacteria, 34 Mycota and five Eukaryota; Table S6). For Bacteria, 22 taxa were strongly associated with the young stage of soil development, including members of the genera *Roseiflexus*, *Herbaspirillum* and *Novosphingobium* that exhibited particularly high IndVal, while no taxa were strictly associated with the intermediate class. Seventeen taxa of Bacteria were indicators of both surface and deep soil layer in the old class, including members from the genera *Actinoallomurus* and *Ferrimicrobium* that showed the highest IndVal. Six taxa were indicators of the deep layer at both intermediate and old stages. For Eukaryota, five taxa were considered as indicators; three were fungi associated with the old class, while one mite (genus *Gamasina*) was associated with the intermediate class. For Mycota, 18 taxa were indicators of both surface and deep layer in the older stage, including members of the genus *Cladophialophora* and the family Glomeraceae. Ten Mycota taxa were indicators of both surface and deep layer in the young stage while the intermediate stage contained fewer indicators, with only five taxa. Only one MOTU of Mycota, identified as *Golovinnomyces sordidus*, was representative of the surface layer in young stages of soil development.

DISCUSSION

Our work provides new insights into the spatial structuring and evolution of soil communities in recently deglaciated terrains as it evaluates the responses of functionally different taxa to glacier retreat and assesses how community composition changes in relation to soil depth and development stage. The considered depths did not strongly affect the alpha-diversity of some taxa at any stage of soil...
development, but the communities inhabiting surface and deep soil layers were not exactly the same. Importantly, beta-diversity between surface and deep soils decreased through time across most taxa, supporting the hypothesis of homogenization of communities along the succession (Rime et al., 2015). Furthermore, the variation of alpha- and beta-diversity showed a consistent pattern across taxa, suggesting a high generality of our conclusions.

4.1 | Changes in alpha-diversity with soil development stage and the impact of depth

Alpha-diversity increased through time, as previously observed in successional studies of microorganisms, plants and soil invertebrates (Erschbamer & Caccianiga, 2016; Ficetola et al., 2021; Matthews, 1992), with similar responses across all the study groups. For the whole Eukaryota and, within them, for Mycota, the highest alpha-diversity was consistently found in the surface layer, supporting our hypothesis that the richness of communities decreases with depth. This observation agrees with the idea that the highest soil biodiversity is hosted close to the surface, as already observed for fungi, bacteria and some faunal groups (Carteron et al., 2021; Chen et al., 2020; Chu et al., 2016; Jiao et al., 2018; Moradi et al., 2020; Mundra et al., 2021; Rime et al., 2015). In glacier forelands, surface soils tend to have higher water-holding capacity, carbon and nutrient contents, and more exchangeable cations (Rime et al., 2015). These properties are vital for most below-ground organisms, especially in resource-limited ecosystems, and determine higher bacterial activity, DNA concentration, and fungal and root biomass in the first centimetres of soil (Rime et al., 2015). We highlight that, in our sampling design, the surface sampling covered a thinner layer compared to the deep one (from 0 to 5 cm vs. from 7.5 to 20 cm depth). In principle, the deep layer might hold larger environmental heterogeneity, given that it is the thicker one. Thus, differences in alpha-diversity between layers might be even larger had we sampled layers with similar thicknesses.

For springtails only, the interaction between soil depth and development stage had a significant effect on alpha-diversity, indicating that the richness of this group increased at different rates between the two soil depths. Springtails were nearly absent in surface and deep layer of soils within the first 30 years since glacier retreat (Figure 2; see also Figure S2). The alpha-diversity of springtails increased with time, but the increase was faster in the surface layer compared to the deep layer, probably because the rapid accumulation of organic matter near the surface (Herold et al., 2014; Moradi et al., 2020) facilitates the establishment of these organisms (Phillips et al., 2019), which have multiple trophic roles, from detritivores to herbivores. In recently deglaciated terrains, springtails can be extremely abundant above the soil surface (see Hägvar & Gobbi, 2022); the very low richness observed immediately after glacier retreat might occur because we focused on soil samples, and the time required for colonization can be longer for taxa specializing in deep environments (Figure 2). For the other taxa, we did not detect a significant interaction between soil depth and development stage, suggesting that alpha-diversity increases through time at a similar rate between surface and deep layers (see Figure 2).

In glacier forelands, the amount of organic matter consistently increases through time (Khedim et al., 2014; Zumsteg et al., 2018). Soil carbon content is a major driver of soil biodiversity changes (Chu et al., 2016); consistent with this idea, alpha-diversity tended to increase with the amount of organic carbon. Nevertheless, models including time as the independent variable showed slightly higher $R^2$ values than those with soil organic carbon (Figure 2; Figure S3), suggesting that time since glacier retreat is a better predictor of alpha-diversity than organic carbon, even though these parameters are strongly correlated (Rime et al., 2015; Zumsteg et al., 2011). Further studies are needed to disentangle the role of both time and soil features as drivers of community variation in glacier forelands.

4.2 | Community differences between surface and deep soils change through time

The beta-diversity between surface and deep layers was particularly high soon after the retreat of glaciers and then decreased with a consistent pattern across most of taxa. This pattern was not observed for Collembola and Oligochaeta, but the taxonomic richness of these animals was often zero in recently deglaciated soils (and particularly in the deep layers; Figures S2 and S3). Therefore, for Collembola and Oligochaeta, many plots at early development stages were excluded from this analysis, reducing the statistical power.

In principle, the variation of beta-diversity between the surface and the deep layer can be attributed to both species gain/loss (nestedness) and replacement (turnover; Baselga, 2010). Turnover was much more important than nestedness for invertebrates (Collembola, Insecta, Oligochaeta), while these two components of beta-diversity showed a similar importance for microorganisms (Bacteria and Mycota), and their relevance remained similar through time (Figure S4; Table S5).
The decrease in beta-diversity between surface and deep layers through time confirms the hypothesis of homogenization of communities (Rime et al., 2015) and extends it to the whole soil biota, as bacteria, microeukaryotes and some animals responded in the same way (Figure 3). Community homogenization is probably related to the structural modifications observed during the development of organic horizon through time (Mavris et al., 2010), where growing resources favour the establishment of complex communities. Furthermore, plant richness and cover quickly increase with time after glacier retreat (Rime et al., 2015) and could have contributed to the homogenization of superficial and deep samples, as plant roots generally influence the first 20 cm of soils (Rime et al., 2015). On the studied glacier forelands, a marked increase of plant cover can be observed throughout the succession. Immediately after glacier retreat only scattered individuals of pioneer species are present, generally with a cover of 5% or less. With time, plant communities become increasingly complex and at late stages (Little Ice Age: ~1850) vegetation cover can reach values from 50% (Sforzellina) to 70% (Amola), and even 90% on the most favourable sites such as the Rutor foreland (Burga, 1999; Caccianiga et al., 2006; Caccianiga & Andreis, 2004; Gobbi et al., 2017). Regardless, for all the taxa considered, time since glacier retreat remained the main determinant of community
variation, as it explained much more variation compared to depth (Table 2). This confirms the idea that, even though fine-scale heterogeneity certainly has a role, time since glacier retreat remains the main determinant of community evolution (Ficetola et al., 2021; Rime et al., 2015).

For microorganisms, the significant community differences between soil layers (Table 2) are probably determined by taxa that are specialists of fine-scale environmental features. This idea is supported by the observation that all the MOTUs identified as indicators of surface or deep soil layers are bacteria or fungi (Table S6). Conversely, for invertebrates, soil depth explained a very limited amount of variation in community composition (Table 2). This could be due to the lower richness of these taxa (which limits statistical power), or to the fact that a broader vertical profile would be required to identify specialists (Moradi et al., 2020). However, in glacier forelands the study of deep layers by eDNA analysis is sometimes problematic because rock outcrops are frequent >20 cm below the surface. Several taxa identified as indicators in Rime et al. (2015) showed similar patterns across the different locations of our study, confirming the strong functional variation of communities through time. For instance, bacteria of the genus Clostridium, known to be anaerobic, were consistently found as indicators of the earliest stage of soil development here and by Rime et al. (2015). Similarly, several fungal saprotrophs were indicators across the different stages, while Lachnum was a microfungus consistently associated with the most developed soils (Nguyen et al., 2016). Gemmatimonas tend to be copiotrophic bacteria (Ho et al., 2017) and include multiple MOTUs that were found as indicators of different stages. Interestingly, fungi such as Laccaria or Hygrophorus (i.e., potential ectomycorrhizal taxa; Nguyen et al., 2016) were indicators of old stages of soil development. Contrary to Rime et al. (2015), arbuscular mycorrhizal fungi (i.e., Glomeromycetes) tended to be associated with the old stage, confirming the growing importance of plant-associated fungi along community development (Davey et al., 2015).

### 5 | CONCLUSION

Understanding the development of communities in successions remains a major task of ecological studies. Our study suggests that, even though time since glacier retreat is a more important driver than depth in shaping the diversity of communities, patterns are not identical for superficial and deeper samples. This can have important consequences for ecosystem functioning, for example for the sequestration of organic carbon within glacier forelands (Khedim et al., 2021). Nevertheless, differences between depths tend to decrease through time with a consistent pattern in both microorganisms and animals, possibly because of the increasing role of plants along successional stages. Further studies are needed to identify possible factors driving biotic colonization within the same system (microclimate, soil features, etc.) and patterns of biotic interactions.

### AUTHOR CONTRIBUTIONS

GFF, JP, WT, PT, GAD, DF, RA and MC designed the experiment. SM, GFF, RA, MC, FG and MG conducted the field work. AG and AB and LG conducted molecular analyses. AG and AB performed the

### TABLE 2 Differences in community composition (Sørensen index) of the six taxonomic groups across time, depth and their interaction using permutational multivariate analysis of variance (PERMANOVA) and tested for the homogeneity of multivariate dispersions; p-values were determined using 9999 permutations

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Variable</th>
<th>Variance explained (partial $R^2$)</th>
<th>$p$</th>
<th>Dispersion homogeneity ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td>Time</td>
<td>.059</td>
<td>.002</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.006</td>
<td>.001</td>
<td>.545</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.002</td>
<td>.278</td>
<td>—</td>
</tr>
<tr>
<td><strong>Eukaryota</strong></td>
<td>Time</td>
<td>.025</td>
<td>.007</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.005</td>
<td>.001</td>
<td>.267</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.002</td>
<td>.831</td>
<td>—</td>
</tr>
<tr>
<td><strong>Mycota</strong></td>
<td>Time</td>
<td>.047</td>
<td>.013</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.002</td>
<td>.013</td>
<td>.325</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.001</td>
<td>.143</td>
<td>—</td>
</tr>
<tr>
<td><strong>Collembola</strong></td>
<td>Time</td>
<td>.023</td>
<td>.577</td>
<td>.076</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.002</td>
<td>.654</td>
<td>.374</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.004</td>
<td>.453</td>
<td>—</td>
</tr>
<tr>
<td><strong>Insecta</strong></td>
<td>Time</td>
<td>.03</td>
<td>.114</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.002</td>
<td>.064</td>
<td>.654</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.002</td>
<td>.339</td>
<td>—</td>
</tr>
<tr>
<td><strong>Oligochaeta</strong></td>
<td>Time</td>
<td>.035</td>
<td>.114</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>.003</td>
<td>.135</td>
<td>.766</td>
</tr>
<tr>
<td></td>
<td>Time × Depth</td>
<td>.003</td>
<td>.184</td>
<td>—</td>
</tr>
</tbody>
</table>

Bold indicates significant $p$-values.
bioinformatic treatment of sequence data. CC, AZ and JP performed edaphic analyses. AG, AC and IC ran statistical analyses. AG, AC and GFF drafted the manuscript. All the authors contributed substantially to the revision process and accepted the final version.

ACKNOWLEDGEMENTS
This study was funded by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities), Open Access Funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

CONFLICT OF INTEREST
The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT
Raw sequence data and filtered data are deposited in https://doi.org/10.5281/zenodo.6620359 and https://datadryad.org/stash/share/ACsdND52MV8uRoYtvjzo0jTTFNYYMG5as6yQGnW6x4 digital repositories, respectively.

ORCID
Alessia Guerrieri https://orcid.org/0000-0002-1519-3517
Alexis Carteron https://orcid.org/0000-0003-4162-9254
Audrey Bonin https://orcid.org/0000-0001-7800-8609
Silvio Marta https://orcid.org/0000-0001-8850-610X
Roberto Ambrosini https://orcid.org/0000-0002-7148-1468
Isabel Canera https://orcid.org/0000-0003-3161-1878
Ludovic Gielly https://orcid.org/0000-0001-5164-6512
Fabrizio Gili https://orcid.org/0000-0002-1817-0193
Mauro Gobbi https://orcid.org/0000-0002-1704-4857
Jerome Poulenard https://orcid.org/0000-0003-0810-0308
Pierre Taberlet https://orcid.org/0000-0002-3554-5954
Gentile Francesco Ficetola https://orcid.org/0000-0003-3414-5155

REFERENCES


Wickham, H. (2016). *ggplot2*. doi: https://doi.org/10.1007/978-3-319-24277-4


**SUPPORTING INFORMATION**

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