UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Environmental Sciences
XXXIV cycle

Dip. of Environmental Sciences and Policy

Reconstructing community dynamics and functions in deglaciated areas through an *ad hoc* environmental DNA metabarcoding technique

PhD thesis

Alessia Guerrieri

Tutor: Prof. Gentile Francesco Ficetola

Academic Year
2020-2021

CONTENTS

ABSTRACT

Deglaciated areas are increasing worldwide due to the global glacier shrinkage and are undergoing rapid colonization by multiple lifeforms, thus representing experimental systems to study community dynamics and ecosystem functioning. Our understanding of biotic colonization after glacier retreat has greatly advanced in recent years thanks to the integration of methodological innovations and ecological theories. The development of standardized, robust and previously tested molecular tools can shed light on fundamental gaps in ecological research. Here, I show how classical approaches can be combined with environmental DNA (eDNA) metabarcoding and functional trait analysis to document the formation of multitrophic communities, improving our understanding of the biotic processes that occur following the retreat of glaciers. This work contributed to both a methodological advancement of the eDNA metabarcoding technique and a conceptual advancement in the understanding of the dynamics of communities in deglaciated areas. In the first half of my thesis, I show how some important methodological approaches can be optimized to improve the strength and reliability of datasets built using eDNA metabarcoding. Specifically, I compare several approaches to soil preservation, and show that both low temperature and desiccation can enable long-term soil conservation before eDNA extraction. Subsequently, I compare thresholds for the bioinformatics clustering of metabarcoding data, and evaluate how optimal thresholds can be defined on the basis of study aims and markers. In the second half of the thesis, I analyse how environmental factors influence the colonization of glacier forelands by several taxonomic groups (bacteria, fungi, microeukaryotes and animals). By using a multi-taxa approach, I investigate how soil depth and time since glacier retreat influence the colonization of ice-free areas, and show that different groups colonize the forelands in a similar way, with communities homogenizing through different depths with time. Finally, I evaluate how time since glacier retreat, soil features and climate interact in determining the velocity of colonization of nematodes, and show how they shape community structure and functionality through succession. Understanding how metabarcoding data can be properly used to assess the colonization of glacier forelands and how different factors influence this

process is particularly important in the context of ongoing climate change and for the extrapolation of current observations to the future. Besides helping to understand the biological consequence of the retreat of glaciers, the advancements provided here may allow the development of scenarios on how mountain environments can change over the next century.

CHAPTER 1. INTRODUCTION[1]

Glaciers are retreating on all continents at a rate that has accelerated in recent decades (Hock et al. 2019; Zemp et al. 2019). For instance, in the European Alps, glaciers have lost 25-30% of their surface area over the last 60 years, and the rate of glacier reduction is now 200-300% faster than twenty years ago (Smiraglia & Diolaiuti 2015; Sommer et al. 2020). The retreat of glaciers since the mid-19th Century is attributable to a combined effect of natural and anthropogenic causes, but the importance of anthropogenic factors has increased in recent years, and human activities are the main drivers of the retreats observed since 1990 (Marzeion et al. 2014).

The study of environmental changes occurring after the retreat of glaciers provides a great opportunity to understand the formation of communities, and has been a classic approach to understanding the temporal dynamics of ecosystems (Clements 1916; Connell & Slatyer 1977; Matthews 1992). Ice-free areas are rapidly colonized by organisms and, in less than a century, complex ecosystems, including forests, can sometimes develop in areas previously occupied by glaciers. The possibility of having multiple freshly exposed surfaces, the fact that these areas are clearly delimited by the glacier path, and that these dated study sites are in chronological sequence (a chronosequence – **Fig. 1**), make them ideal for ecological succession studies. When glacier retreat is well described and chronologically constrained, it is possible to provide a maximum date for the ecosystems (the time elapsed since the glacier retreat), and thus to analyze the chronosequences, assuming that time is a major factor in the evolution of ecosystems. The studies of deglaciated areas carried out since the beginning of the 20th Century were a milestone in the development of ecological succession theory, with a strong emphasis on plants and invertebrates.

Although classic studies of ecological succession in deglaciated areas provided critical insight into ecosystem formation, they tended to focus on just one trophic level (often plants). More recent studies

[1] This section includes my contribution to the following paper: Ficetola, G. F., Marta, S., Guerrieri, A., Gobbi, M., Ambrosini, R., Fontaneto, D., … Thuiller, W. (2021). Dynamics of Ecological Communities Following Current Retreat of Glaciers. Annual Review of Ecology, Evolution, and Systematics, 52(1), 405-426. doi: 10.1146/annurev-ecolsys-010521-040017

have provided crucial advances in our understanding of biotic colonization after glacier retreats, with a growing interest on the overall community and on how it is structured. On the one hand, ecologists are now using community, functional and network theories and tools to go beyond the mere description of biodiversity patterns and to tease apart the role of different processes (i.e., habitat filtering, dispersal and biotic interactions) in the formation of species and functional communities, and in the emergence of functional diversity. On the other hand, recently developed molecular tools have opened a new window on the invisible, providing information on the formation of multitrophic communities, including the microbes that play crucial roles in soil food webs.

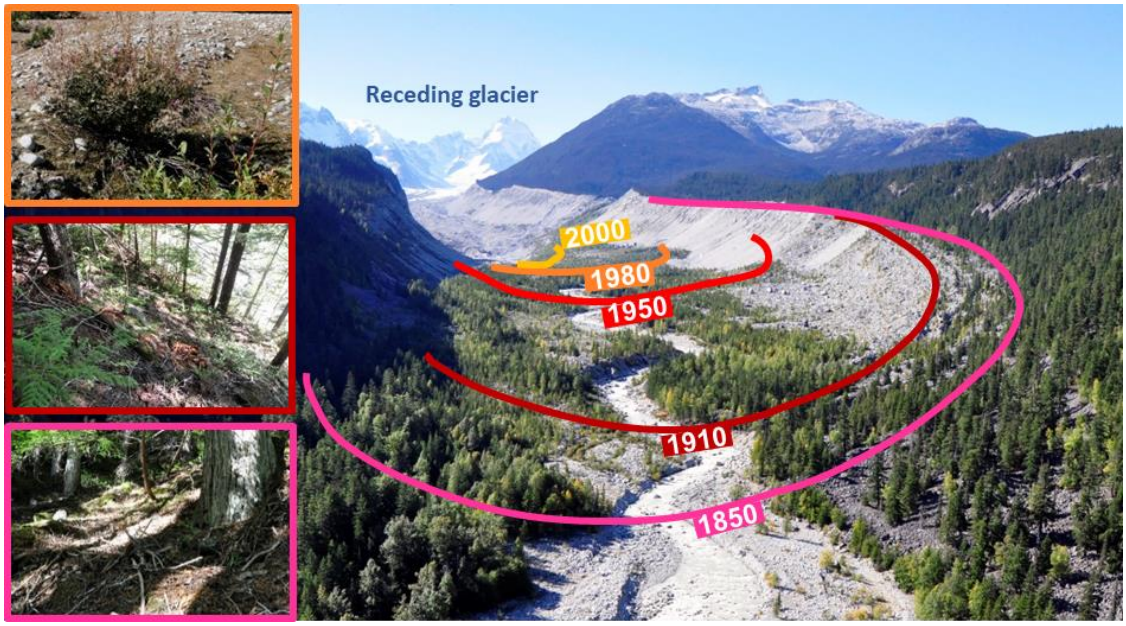


Figure 1. In a glacier foreland, a "chronosequence" is a sequence of dated sites for which time after retreat has been measured using a wide range of approaches (e.g. photographs, radionuclides…). In this figure, the Tiedemann Glacier chronosequence (Canada), with strongly different communities between early and late development stages.

Newly deglaciated areas represent increasingly important components of mountain and high-latitude ecosystems due to climate change, and rates of glacier retreat are predicted to accelerate in the coming decades (Hock et al. 2019). Understanding the consequence of glacier retreat is therefore becoming an increasingly important topic for both ecology and global change studies, with researchers trying to understand the consequence of retreat and to develop scenarios on how mountain environments may change over the next century.

1.1 Patterns of colonization rate

How long does it take for a community to be established? This question remains essential for fundamental and applied ecology, and is particularly important should we want to predict the long-term consequences of global change. The colonization of glacier forelands is assumed to start immediately after the glacier retreat (but see **Box 1**) with local-scale environmental factors and feedbacks driving community development over time. Thousands of years can be required for the full establishment of mature ecosystems (Delgado-Baquerizo et al. 2019). However, in some cases, colonization can be surprisingly fast, as the rate at which different taxa colonize glacier forelands is not constant, and there are clear differences both between taxonomic groups and regions of the world.

BOX 1. Are glacier forelands an example of primary succession?

Community assembly after glacier retreat is often described as the "ideal example" of primary succession. However, glaciers are not lifeless environments, as they are ecosystems teeming with microbial and eukaryotic life and supporting active ecological processes (Hodson et al. 2008; Hotaling et al. 2017; Gobbi et al. 2021). Glaciers host rich communities, which are connected by the flow of meltwater from the glacier through the ice fractures to the subglacial sediments (Hotaling et al. 2017). Glacier meltdown exposes subglacial sediments; moreover, supraglacial sediments, together with the communities they host, are added to subglacial till. Bacterial communities in recently deglaciated terrains can be more similar to those of supra- and subglacial sediments than to those found in atmospheric deposition (snow, rain and wind-dispersed dust). This indicates that bacteria inherited from the glacier are important starting points of communities in newly deglaciated surfaces (Rime et al. 2016). Similar processes have been proposed for other glacier-dwelling organisms, such as invertebrates and algae (Stibal et al. 2020; Rosero et al. 2021). Thus, even the most recently deglaciated areas are not devoid of life, as they host many life forms from glacier environments. In addition, melting glaciers release a substantial amount of ancient organic matter, which pioneer species can feed on. As this ancient carbon can be the basis of food webs in recently deglaciated grounds, the term "primary succession" is questionable (Hågvar & Ohlson 2013). Complex feedbacks between glaciers and their forelands are also possible, as communities in environments close to glaciers are sources of colonization of supraglacial environments (Franzetti et al. 2017; Gobbi et al. 2017). Hence, glacier and near-glacier ecosystems cannot be disentangled; rather, they must be considered together for a comprehensive understanding of ecological responses to glacier retreat (Stibal et al. 2020).

Factors triggering differences in the rate of biotic colonization include i) successional stage (e.g. early successional vs mid- and late-successional stage), ii) traits of the surrounding species pool (e.g. dispersal

ability, diet) and iii) abiotic factors, such as climate, geomorphic stability of surfaces, and altitudinal gradients along glacier forelands (**Section 1.2**).

1.1.1. Changes of colonization rate through time. The rate of colonization is not constant over time, with a rapid development of pioneer communities in the first 20-50 years, followed by a decrease in colonization rate at later successional stages (Kaufmann 2001; Gobbi et al. 2006a; Gobbi et al. 2010; Schlegel & Riesen 2012). The deceleration of colonization rate is consistent with the predictions of the facilitation and inhibition hypotheses (Connell & Slatyer 1977). Indeed, biotic interactions affect colonization rate in multiple ways, with facilitation in the early colonization stages and an increasing importance of competition, parasitism, or predation, decelerating colonization at later stages (Koffel et al. 2018; Benavent-González et al. 2019). This change in the type of biotic interactions can create complex patterns when glacier forelands are wide and include multiple altitudinal zones (e.g. both areas above and below the tree line; Gobbi et al. 2007; Schlegel & Riesen 2012; Vater & Matthews 2015).

1.1.2. Differences in colonization rate across taxa. The rate of colonization strongly varies from one taxonomic group to another, and even between related species. Some taxa (e.g. bacteria, rotifers, bacterivorous nematodes, springtails, ground beetles, spiders) can colonize the sediments immediately after the glacier retreat, whereas the rate of colonization can be much slower for other taxa (Matthews 1992; Gobbi et al. 2006b; Hagvar 2011; Brown & Jumpponen 2014; Lei et al. 2015; Hågvar et al. 2020; Rosero et al. 2021). Differences are determined by the interplay between dispersal ability and tolerance to environmental conditions. Aerial transport by windy tides allows very rapid colonization by pioneer species. Most of microscopic animals (rotifers, nematodes, tardigrades), and several species of mosses, springtails, mites, spiders and insects, can reach recently ice-free lands as "aeroplankton" (Coulson et al. 2003; Flø & Hågvar 2013; Gobbi et al. 2017; Fontaneto 2019; Hågvar et al. 2020). Aerial transport is an effective dispersal mechanism if the transported species are able to survive and reproduce on the colonized substrate, and is also frequent for many non-flying organisms. These species are capable of rapidly colonizing newly

deglaciated terrains and, since they also have specific traits (e.g. ability to perform ballooning, fast growth, short larval stages), they determine the functional differences between early- and late-stage communities (**Section 1.3** and **Section 1.4**).

Dispersal by land tends to be a slower process than by air, although some ground dwelling arthropods can colonize new areas rapidly (Lei et al. 2015; Hågvar et al. 2020). Capture-recapture data on ground beetles (Nebriinae) suggest they only move few meters per day (Kaufmann & Gobbi Unpublished; Hågvar et al. 2020), possibly because they have to walk between stones and across the three-dimensional space (Tenan et al. 2016). Consequently, for some taxa there is a lag between the availability of suitable habitats and colonization (Brambilla & Gobbi 2014). Sites where suitable habitats are already available, but which have been ice-free for limited periods (e.g. <100 years), are mainly occupied by species with high dispersal abilities (e.g. winged carabid beetles and ballooning spiders; Gobbi et al. 2017), while slow dispersers may take more than 100 years to colonize suitable habitats. Large animals such as birds and mammals can quickly disperse to the ice-free areas for foraging, but the establishment of self-sustaining populations is generally delayed by the need of stable and abundant food resources that can require several decades (Rosvold 2016; Williams et al. 2019). Overall, the rate of dispersal of different taxa is determined by the combined effects of environmental constraints, species characteristics and interspecific interactions. All these factors need to be integrated for an effective understanding of the processes involved. Finally, this also depends on the available species pool nearby the glacier, as species-poor areas of the world will likely be colonized more slowly and will form less complex communities than more biodiverse regions. More data are needed to ascertain this hypothesis.

1.2. The processes determining colonization

Early studies on the colonization of deglaciated terrains assumed that time after glacier retreat is the main factor in colonization (Johnson & Miyanishi 2008). However, chronosequences may not always be able to clearly capture effects of time on colonization, given many factors (e.g. differences in site characteristics, altitude, differences between local species pool) can have profound consequences on communities (Johnson & Miyanishi 2008). We now know that time is only one driver, with colonization

rates influenced by multiple stochastic and environmental factors, such as the availability of potential propagules, the surrounding species pool, microclimate, topography, solar radiation, bedrock type and (proto-)soil properties (Rydgren et al. 2014; Tampucci et al. 2015; Vater & Matthews 2015; Makoto & Wilson 2016). What is the relative importance of time, abiotic features/processes, biotic interactions and dispersal limitations? Understanding how environmental factors shape community changes and functions is pivotal to predict impacts of ongoing and future environmental changes (Johnson & Miyanishi 2008; Bjorkman et al. 2018). Even though biotic processes are certainly important, here we will focus on the role of abiotic processes, such as climatic differences and soil features.

1.2.1. Abiotic factors and processes: time, soil and disturbance. Traditionally, time since glacier retreat (age) was considered as a major predictor of succession along glacier forelands, based on the assumption that sites along the sequence share the same biotic and abiotic history and differ only in age (space-for-time substitution; Johnson & Miyanishi 2008). Dating of deglaciation steps is readily available for hundreds of glaciers around the world (Leclercq & Oerlemans 2012; Marta et al., 2021), and can be a good indicator of many key environmental properties. Nevertheless, it is essential to distinguish between the direct effects of time on successional processes (e.g. some species take some time to colonize; Brambilla & Gobbi 2014), its indirect effects on environmental properties that can affect community assembly, and factors that are simply collinear to age (e.g. within a foreland older terrain often are at the lowest elevations; **Fig. 1**).

Most abiotic features change rapidly over time, influencing biogeochemical cycles, the rate of colonization, and the identity of colonizers (Chapin et al. 1994; Raffl et al. 2006; Mori et al. 2017). Increasing age is related to higher surface stability and related soil development (Erhart 1951; Raffl et al. 2006; Erschbamer et al. 2008), including accumulation of organic matter, due to positive feedback loops between sediments/soils, vegetation and fauna (Chapin et al. 1994; Eichel et al. 2016; Whitesides & Butler 2016; Mori et al. 2017; Vega et al. 2020; Khedim et al. 2021). At the same time, soil pH and nutrient content can change with age (Rydgren et al. 2014; Tampucci et al. 2015; Khedim et al. 2021), soil texture becomes finer, and organic debris form organic topsoil

horizons (Schumann et al. 2016). In addition, geological, geomorphological and climatic differences between regions influence the kind and rate of rock weathering and thus the resulting physical and chemical properties of soils (Jenny 1994; Walker et al. 2010; Khedim et al. 2021), determining divergent trajectories of soil development.

In turn, soil characteristics that vary with time can impact all components of communities, with multiple feedbacks. For instance, soil stability favors plant germination and invertebrate establishment (Erschbamer et al. 2008; Cauvy-Fraunié & Dangles 2019), while increased nitrogen availability in older soils facilitates plant growth (Chapin et al. 1994) and soil development influences the structure of both bacterial and fungal communities (Dong et al. 2016; Gorniak et al. 2017; Kim et al. 2017). Direct relationships between animal communities and soil characteristics are more difficult to identify, given their frequent association with plant communities. Still, fine-scale analyses revealed that grain size, micro-topographic position, litter and edaphic characteristics create multiple microhabitats that drive the distribution of arthropods (Schlegel & Riesen 2012; Hågvar et al. 2020).

Finally, glacier forelands are often subject to important disturbances such as flooding, geomorphological instability, cryoturbation and grazing (Kaufmann 2002; Garibotti et al. 2011; Chapin et al. 2016; Schumann et al. 2016), all of which may alter, slow down or even stop succession, resulting in non-linear trajectories of community development (Walker et al. 2010). Geomorphic instability (e.g. debris flow, erosion) is particularly frequent and strongly limits the development of vegetation but, also in this case, feedbacks are frequent, as some pioneer plants greatly increase the stability of slopes, promoting the development of soil and of plant communities (Eichel et al. 2016).

It is difficult to measure the relative importance of time since glacier retreat vs. abiotic factors, as the effects can be highly scale- and context-dependent. In some cases, abiotic factors may have a comparable, and sometimes stronger effect than that of time (Raffl et al. 2006; Garibotti et al. 2011), particularly at fine (microhabitat) scale (Burga et al. 2010). For instance, recently

deglaciated but wet microhabitats can reach high plant diversity 20 years earlier than sunny areas of the same foreland (Raffl et al. 2006), and the effects of soil characteristics may outweigh those of time at later (century) stages of succession (Rydgren et al. 2014). Nevertheless, generalizations remain difficult, with other researchers observing a limited impact of local abiotic conditions, and suggesting that time since deglaciation remains the best driver of community richness and structure, especially over large spatial scales (Burga et al. 2010; Nascimbene et al. 2017).

1.2.2. The potential role of climate. Micro- and macro-climatic characteristics are major drivers of biodiversity patterns (Ficetola et al. 2018; Thuiller et al. 2020; Zellweger et al. 2020) and play also a key role in the colonization of deglaciated terrains. Mountain environments have complex topographies. Along a single chronosequence, interactions between topography (mainly slope and aspect) and local climate, altitude and solar irradiance generate different micrometeorological conditions determining differences in species distributions and colonization rates (Garibotti et al. 2011; Schumann et al. 2016; Feldmeier et al. 2020). Some studies suggested that warm climatic conditions accelerate colonization (Vater & Matthews 2015; Fridley & Wright 2018). For instance, in the Alps, a 0.6°C increase in summer temperatures can double the colonization rate of early successional stages by invertebrates (Kaufmann 2002). It has therefore been proposed that the current colonization by plants is significantly faster than a century ago, due to the acceleration of climate change (Fickert et al. 2017). Faster development under warmer climates has been proposed for other features of succession in glacier forelands (e.g. accumulation of soil organic matter; Khedim et al. 2021), probably because of increased primary productivity of autotrophs and/or faster metabolism of heterotrophs.

Temperature is not the only important climatic factor. Precipitation regimes and climate continentality affect the water balance and may consequently alter the successional trajectories (Vater & Matthews 2015). Additionally, season length and solar irradiance vary with latitude and altitude, conditioning the time and energy available for life processes (Körner 2000). Finally, climate interacts with other abiotic features. Insolation and temperature influence snow cover

persistence, melting-water discharge and the length of growing season (Tampucci et al. 2015). This, in turn, may have additional effects on decomposition rate, C and N mineralization, CO_2 and NO_2 fluxes and net nitrification (Williams et al. 1998; Khedim et al. 2021). Microclimate, soil texture and aspect also influence soil moisture, affecting the development of plant communities and nitrogen mineralization (Raffl et al. 2006; Rydgren et al. 2014; Schumann et al. 2016).

Understanding how colonization is influenced by climate is particularly important in the context of ongoing climate change, as this would complicate the extrapolation of current observations to the future. Despite substantial efforts to understand the drivers of colonization, most studies focused on a single, or a limited number of chronosequences without assessing the effect of climatic and environmental differences between areas (reviewed in Cauvy-Fraunié & Dangles 2019). However, even limited variation across areas can lead to drastic divergence in the trajectories of ecological communities (Vater & Matthews 2015; Chapin et al. 2016), thus more data are needed to understand the dependence of colonization on climate, with comparisons of several glacier forelands through meta-analytic approaches (Cauvy-Fraunié & Dangles 2019), or with new broad-scale studies.

1.3. Multitrophic linkages and functional diversity

The recent broadening of taxonomic focus has revealed the importance of multitrophic interactions to community assembly. Although plant colonization is considered a major precondition for the arrival of other organisms (Gobbi et al. 2006b; Vater 2012; Schmidt et al. 2014), not all animals need an established plant community to colonize (Hagvar 2011; Vater 2012; Hågvar et al. 2020). In many cases spiders, predatory insects, bacterivorous nematodes and detritivores colonize deglaciated terrains well before plants (Hågvar & Ohlson 2013; Lei et al. 2015; Sint et al. 2019; Hågvar et al. 2020). The paradox of many different predators inhabiting areas without primary producers and herbivores was initially explained by the input of allochthonous material blown by wind, but analyses of species traits and food webs revealed more complex mechanisms (Raso et al. 2014).

Strong abiotic filters promote the early establishment of a relatively simple functional pool dominated by heterotrophic organisms (Hodkinson et al. 2002; Raso et al. 2014; Mori et al. 2017; Hågvar et al. 2020). So far, little is known about how these simple food webs develop into complex networks during succession (König et al. 2011; Raso et al. 2014), but the recent increase in multitrophic studies and the application of molecular tools allow unprecedented reconstructions of how the functioning of these communities change. The communities establishing immediately after the retreat of glaciers seem to mostly feed on indigenous materials produced locally by microorganisms, or on old organic matter released by glaciers. Intraguild predation can be frequent; wind inputs apparently have a limited importance even though uncertainties exist on the role of allochthonous detritus (Hågvar & Ohlson 2013; Raso et al. 2014; Azzoni et al. 2016; Sint et al. 2019). Subsequently, the colonization by arthropods can be an important catalyst for plant colonization, by providing additional nutrients on recently exposed terrains, thus accelerating the transition from heterotrophic to productive communities capable of supporting more complex food webs, and to stock large amount of organic matter (König et al. 2011; Mori et al. 2017; Benavent-González et al. 2019). At late successional stages, intraguild predation remains present, but the availability of prey at low trophic levels increases (e.g. increased frequency and diversity of herbivores), thus predators have a broader prey spectrum and show increased specialization and niche differentiation, enabling the emergence of more complex food webs (Raso et al. 2014; Sint et al. 2019; Hågvar et al. 2020). Complex feedbacks also occur for herbivorous vertebrates that exploit environments. Several mammals and birds can attain high biomass nearby glaciers; their feces increase the availability of nutrients and dissolved organic matter, thus influencing both bacterial and plant communities (Rosvold 2016; Vega et al. 2020).

Interactions between plants and animals are not limited to consumption. Analyses assessing taxonomic and functional diversity of plants and interacting insects showed that, in recently deglaciated areas, communities host a limited number of species and only a few functional groups, the flowers being mainly visited by pollinators. In late successional stages, the same flower species are visited by a larger number of insects belonging to more functional groups, including predators, parasitoids and phytophagous

(Losapio et al. 2015). The increase of species richness along chronosequences also results in increased levels of functional and trophic diversity and, sometimes, redundancy, with plants and their pollinating insects forming increasingly complex networks at late successional stages. In turn, the growing complexity of these networks probably confers stability and robustness to the entire food web (Albrecht et al. 2010; Losapio et al. 2015; Losapio et al. 2016; Ricotta et al. 2016).

Some colonizing animals can even act as ecosystem engineers, with cascading effects on multiple components of communities. For example, mammals that feed on fungi frequently have spores in their fecal pellets, thus providing inoculum to accelerate the spread and diversify populations of mycorrhizal fungi for early successional plants, with potential feedbacks on the whole community (Cázares & Trappe 1994). Second, burrowing animals such as marmots and gophers can disturb large areas of forelands. Soils disturbed by these animals have lower compactness and higher nutrient contents, and in these soils conifers have better germination and establishments (Whitesides & Butler 2016). Despite these advances, most studies have so far focused on a few components of ecological networks (e.g. arthropod predators vs prey; mammals vs plants), calling for more integrative studies analyzing a broader set of guilds. For instance, the increased taxonomic and functional diversity along chronosequences is probably paralleled by a growth of parasitic interactions (see Lei et al. 2015 for an example with nematodes), but very few studies addressed variation of parasites in these environments.

1.4. From species identity to functional diversity

Multitrophic studies enable a more complete understanding of community patterns, but also increase the complexity of analyses. Documenting all the potential pairwise interactions that occur in a community is prohibitive. Instead, traits are common currencies, allowing comparisons across communities, even with different regional pools (Gravel et al. 2016). Trait-based approaches can allow for generalization on the structure and dynamics of ecological communities (Gravel et al. 2016; Moretti et al. 2016), and have been used to approximate some aspects of ecosystem functioning (Naeem et al. 2012; Gravel et al. 2016). Early analyses of functional traits on deglaciated grounds mostly focused on plants (e.g. Chapin et al. 1994), but

we can now analyze traits for a broad range of taxa, from bacteria to animals (Brbić et al. 2016; Moretti et al. 2016). A growing number of studies has revealed the functional variation during successions after glacier retreat, showing that the increase of species richness along chronosequences also results in higher functional diversity and, sometimes, redundancy, with consistent patterns across multiple taxa, including plants, arthropods and nematodes (Losapio et al. 2016; Ricotta et al. 2016; Brown et al. 2017; Cauvy-Fraunié & Dangles 2019). Understanding which traits prevail in the early stages of community formation and how they change during successions helps assessing how species characteristics determine assembly rules and community dynamics (Chang & Turner 2019). Trait variation can also be used to discriminate the potential processes that explain colonization, and to measure the trajectories of functional diversity and ecosystem functioning (**Fig. 2**; Naeem et al. 2012; Gravel et al. 2016).

Many mechanisms proposed to explain community dynamics in deglaciated terrains provide clear predictions on trait variation over time (**Fig. 2**). For instance, the hypothesis that limited dispersal determines the rate of colonization is confirmed by the prevalence of taxa with traits favoring dispersal immediately after glacier retreat, while a lag time is often observed for poor dispersers (**Section 1.2**). Similarly, the tolerance model predicts that early colonizers are replaced by species with better competitive ability, and has been confirmed by the observation of ruderal, fast growing species in recently deglaciated areas, while later stages host stress tolerant species with slower growth and denser and smaller leaves (Erschbamer & Caccianiga 2016; Ricotta et al. 2020). The analysis of traits can thus provide efficient tests of ongoing processes. Importantly, the proposed processes are often non-exclusive, and some can prevail at early stages while becoming less important subsequently. The rate of turnover of functional traits could be used to evaluate how the importance of different processes changes through time (Brown et al. 2017). For instance, we could imagine strong turnover in dispersal-related traits during the first decades after glacier retreat. Then, after a given age, poor dispersers can become more frequent, and trends in competitive ability could emerge. Functional traits representing growth, dispersal or competition can be obtained for most of the components of communities, enabling to compare the importance of different processes across functional levels, and might even be used to assess at which stage of the

succession communities are, predicting their stability and future development (Erschbamer & Mayer 2011; Ricotta et al. 2020).

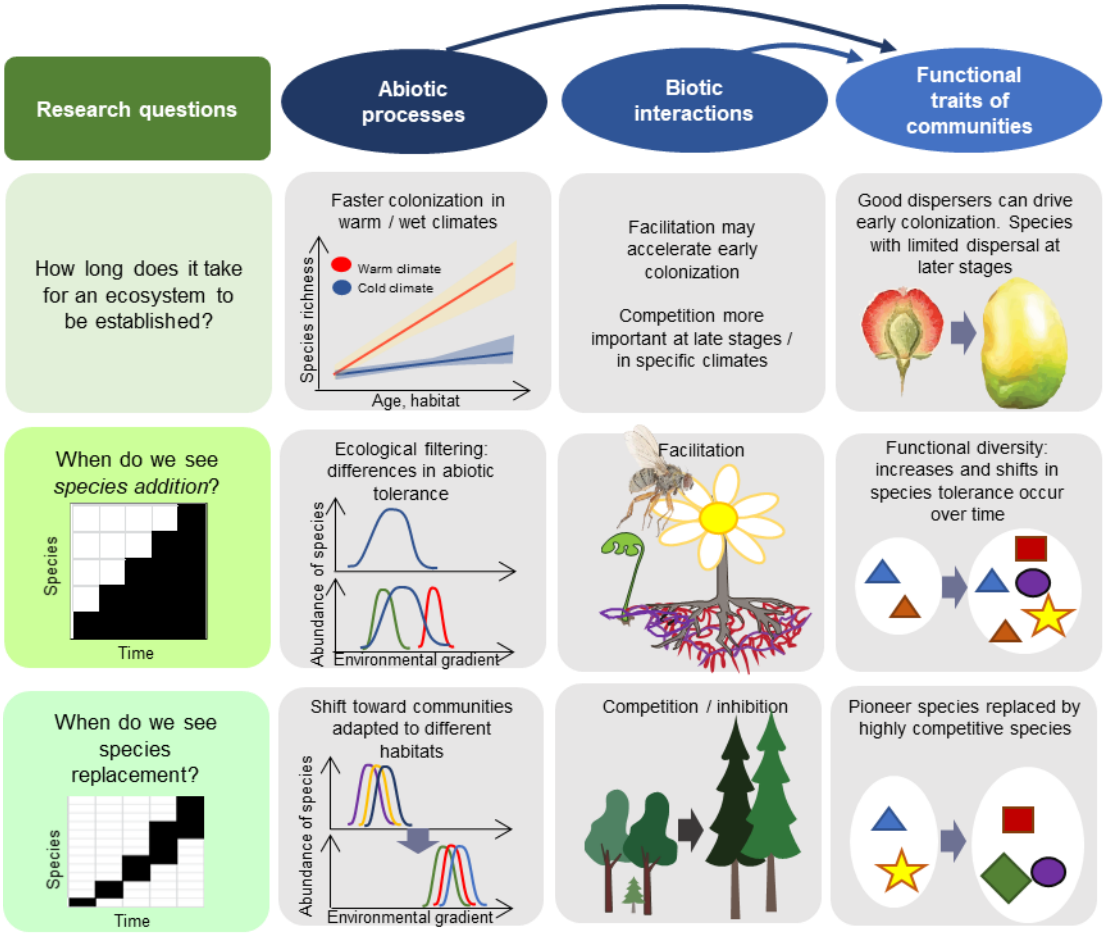


Figure 2. Examples of how analyses on abiotic factors, biotic interactions and functional traits can help to understand the dynamics of colonization after glacier retreats.

1.5. Environmental DNA and metabarcoding for the reconstruction of communities

Increasing evidence that all trophic levels contribute to the development of biodiversity in deglaciated terrains has led to an increase in studies targeting a broad range of taxa and trophic levels. However, the morphological identification of species is often laborious and time-consuming. Metabarcoding of environmental DNA (eDNA) is increasingly used to reconstruct communities of soil and water organisms, and can provide a more complete understanding of community dynamics (Taberlet et al., 2018). Environmental DNA is the DNA that can be extracted from environmental samples (e.g., water, sediments, soil) without isolating the macroscopic remains of the organisms (Taberlet et al. 2018). A given environmental sample generally contains the eDNA of many organisms. Multiple species of a given

taxonomic group (e.g. insects, annelids, bacteria) can be amplified using "universal primers", compared with reference databases and eventually identified in a procedure called eDNA metabarcoding (Taberlet et al. 2018). Recent metabarcoding analyses have revealed the complex patterns of microbial diversity in deglaciated terrains (e.g. Schmidt et al. 2008; Blaalid et al. 2012; Rime et al. 2015; Dong et al. 2016; Matsuoka et al. 2019). Although fewer studies have thus far used metabarcoding to study macro-organisms in deglaciated terrains, the application of these approaches to a wide range of taxa could improve our understanding of biodiversity dynamics and of multitrophic relationships (Rosero et al. 2021).

A typical eDNA study relies on some critical steps (**Fig. 3**). An opportune definition of the experimental design is needed and should include a pilot study in order to test whether the experiment is suitable for addressing specific scientific questions (Dickie et al., 2018; Taberlet et al., 2018). Environmental DNA can be collected from a wide range of environmental samples (including soil, water, sediments, but also faeces and organism bulk samples; Taberlet et al., 2018) and habitats. The sampling strategy must be standardized and adjusted according to many factors. The type of environment (e.g., aquatic vs. terrestrial) and its physical, chemical and biological characteristics affects the eDNA persistence in the environment (Levy-Booth et al., 2007; Pedersen et al., 2015; Pietramellara et al., 2009). Within a single sample, different organisms can be retrieved (e.g. microorganisms vs. macro-organisms) and so the optimal size of the studied area, the size of the sampling units and the suitable number of biological replicates could differ among taxa. However, the resulting sampling strategy is usually a compromise between these aspects and also consider logistical and financial aspects. After sample collection, the raw material is often transported to lab facilities and preserved until lab processing (but immediate *in situ* DNA extraction is sometimes possible; see Zinger, Taberlet, et al., 2019b). Samples are usually preserved below 0°C, at 0-4°C, at room temperature (with or without the addition of preservatives), or desiccated (e.g., using silica gel; Dickie et al., 2018). Ideally, communities recovered from preserved samples should match those retrieved if samples had been processed immediately after sampling. However, inappropriate preservation conditions can impact the ecological results of metabarcoding studies by causing DNA degradation and / or proliferation of certain taxonomic groups with respect to others (Cardona et al., 2012; Orchard, Standish, Nicol, Dickie, & Ryan, 2017). The eDNA

can then be extracted from the environmental samples following opportune standardized protocols to allow a reliable comparison of samples (Taberlet et al., 2012; Eichmiller et al., 2016; Zinger et al., 2016; Lear et al., 2018; Capo et al., 2021). The DNA extraction strategy is likely to impact ecological results by affecting DNA yield and community composition in a different way, depending on the target groups (e.g. prokaryotes vs eukaryotes; Dopheide et al., 2019). The extraction should include technical replicates (i.e., several extractions of the same sample) in order to account for the heterogeneity of the environmental samples, and extraction negative controls (i.e., an extraction where all reagents are used but no eDNA is added), in order to monitor for potential contamination during the extraction protocol (Taberlet et al., 2018). Subsequently, eDNA can undergo Polymerase Chain Reaction (PCR) amplification of one or multiple genetic markers, or alternatively shotgun sequencing (i.e., sequencing the DNA without PCR amplification before library preparation). The aim of the PCR step is to synthetize a large number of copies (amplicons) of the target DNA fragments. Several factors such as the presence of polymerase inhibitors, a wrong hybridization temperature or a wrong number of cycles (Taberlet et al., 2018) can affect the outcome of this step. Thus, technical choices need to be optimized for the experiment considering the initial concentration of target DNA as well as the primers set selected for the study.

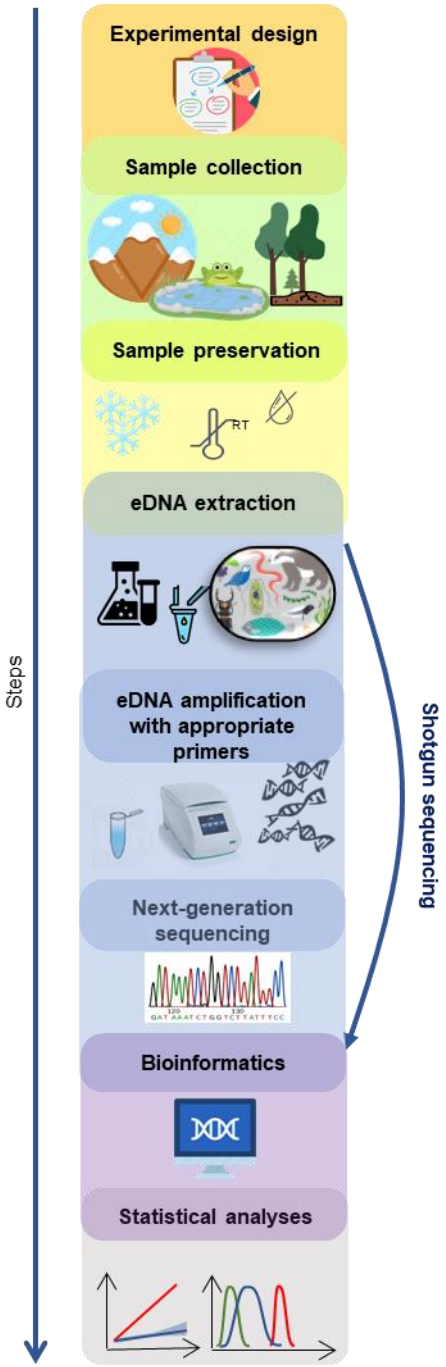


Figure 3. Simplified workflow for a typical eDNA study.

PCR amplification should include technical replicates (i.e., several amplification of the same sample) in order to account for this variability. Appropriate number of PCR negative controls (i.e., a reaction where template DNA is replaced by DNA-free water) and PCR positive controls (i.e., a reaction where template DNA is replaced by a mock community) should also be included in order to monitor for potential contamination during the amplification (Taberlet et al., 2018; De Barba et al., 2014; Lopes et al., 2015). Negative and positive controls are important to adjust the filtering strategy at the data analysis step so that PCR and sequencing errors are detected and removed, and the alpha diversity is not overestimated (Coissac et al., 2012).

To date, eDNA studies mostly rely on the Illumina technology and perform sequencing with a single-read approach (i.e., sequencing of a single strand) or a paired-end approach (i.e., sequencing of both one strand and its complementary strand; Nichols et al., 2018, Taberlet et al., 2018, Bohmann et al., 2022). While different protocols exist for eDNA shotgun sequencing, amplicon sequencing can be performed without including additional PCR steps during library preparation. This strategy is called “MetaFast” protocol (www.fasteris.com/metafast) and reduce the risk of producing sequencing artifacts (e.g., chimeras and tag-jumps; Schnell et al. 2015a; Taberlet et al 2018). High-throughput sequencing of amplicons produce huge quantities of DNA sequence data that can be processed using bioinformatics tools (e.g., the OBITools; Boyer et al., 2016). Sequence data are usually curated with the final scope of filtering out molecular artefacts and reducing the size of the data so that the final taxonomic inventories are as representative as possible of the actual biodiversity (Taberlet et al., 2018). However, the curation strategies strongly depend on the study design and on the ecological questions to be answered, and inappropriate choices can bias biodiversity estimates and lead to misinterpretation of results. The final output can be translated into ecological information using appropriate statistical approaches (Paliy and Shankar 2016, Chen and Ficetola 2020).

Although this section did not provide an exhaustive review of this complex technique (Taberlet et al., 2018), it aimed to provide some clues to help understand the main steps of an eDNA metabarcoding study and their associated issues. Each step is critical to obtain sounds ecological conclusions because

they require to make methodological choices that can heavily influence the reliability and interpretation of results (Alberdi et al., 2018, Zinger et al., 2019). An increasing number of studies has already assessed how methodological choices across the different steps could influence the conclusions of a study (Calderón-Sanou, Münkemüller, Boyer, Zinger, & Thuiller, 2020; Cantera et al., 2019; Chen & Ficetola, 2020; Nichols et al., 2018; Taberlet et al., 2018). Still, there is an urgent need for the development, standardization and optimization of methods dedicated to metabarcoding.

1.6. Aims and structure of the thesis

Our understanding of biotic colonization after glacier retreat has greatly advanced in recent years thanks to the integration of methodological innovations and ecological theories. Understanding how the development of standardized, robust and previously tested molecular tools can be properly used to assess the colonization of glacier forelands is particularly important in the context of ongoing climate change. Therefore, this work contributes to both a methodological advancement of the eDNA metabarcoding technique and a conceptual advancement in the understanding of the biotic processes that occur following the retreat of glaciers. More specifically, this thesis has the aims of: 1) improving our current understanding of the impact of methodological choices on the reliability and reproducibility of metabarcoding data, in the context of increasing the (appropriate) use of this tool for current and future ecology research; 2) improving our knowledge about how different taxonomic groups colonize recently deglaciated areas, by exploiting the power of metabarcoding for the broad-scale production of biodiversity data and 3) understanding how abiotic parameters (e.g. soil depth, climatic differences between areas of the world) influence the colonization of recently deglaciated areas by evaluating their impacts on both taxonomic and functional diversity of communities. Besides helping to understand the consequence of glacier retreat, these advancements may allow the development of scenarios on how mountain environments will change over the next century.

The thesis is composed of two functionally complementary sections. The first section, comprising Chapters 2 and 3, analyses the impact of some methodological choices on the production of biodiversity data through metabarcoding. Specifically, **Chapter 2 (paper A)** evaluates the impact of 5 different

strategies of soil sample preservation for metabarcoding studies on the resulting estimates of taxon diversity and community composition of three different groups (bacteria, fungi and eukaryotes), and in three different habitats (forest, river bank and grassland). Based on the results of this study, guidelines were proposed to optimize soil samples preservation conditions and reduce the negative impact on the results of metabarcoding studies.

Chapter 3 (paper B) provides new insights on methodological choices related to the bioinformatics analysis of sequence data. Using sequences from a public database and for eight metabarcoding markers, from generalist markers (i.e., targeting entire superkingdoms or kingdoms) to more specific markers (i.e., targeting a class or a subclass), this study evaluates different clustering thresholds and provide practical guidelines to minimize the error rates associated with this step.

The second section, comprising Chapters 4 and 5, shows how metabarcoding data can be used to understand biotic responses to the retreat of glaciers. Specifically, **Chapter 4 (paper C**) evaluates the response of six taxonomic groups (Eukaryota, Bacteria, Mycota, Collembola, Insecta, Oligochaeta) to glacier retreat across the forelands of five Alpine glaciers. This study assesses whether communities associated to surface (0-5 cm) and deep (7.5-20 cm) soils show similar trajectories through time, testing the hypothesis that soil development determines vertical homogenization of communities.

Chapter 5 (paper D) evaluates how climatic differences between regions could affect the emergence of taxonomic and functional diversity after glacier retreat, with a specific focus on nematodes, which are among the most abundant animals in soil and play a keystone role in soil functioning. Using a DNA metabarcoding dataset which includes 48 glaciers located on five continents, this study evaluates how the taxonomic and functional diversity of nematodes change in ecosystems developing after the retreat of glaciers, and identifies the key mechanisms driving the changes of communities.

CHAPTER 2. ARTICLE A[2]

Effects of soil preservation for biodiversity monitoring using environmental DNA

Running title: Preserving soil for eDNA analyses

Alessia Guerrieri[1,] *, Aurélie Bonin[1], Tamara Münkemüller[2], Ludovic Gielly[2], Wilfried Thuiller[2], Gentile Francesco Ficetola[1], [2]

Affiliations:

[1] Department of Environmental Science and Policy, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy

[2] Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000, Grenoble, France

* Correspondence author. Email: alessia.guerrieri@unimi.it

[2] Published in Molecular Ecology (2021)

ABSTRACT

Environmental DNA metabarcoding is becoming a key tool for biodiversity monitoring over large geographical or taxonomic scales and for elusive taxa like soil organisms. Increasing sample sizes and interest in remote or extreme areas often require the preservation of soil samples and thus deviations from optimal standardized protocols. However, we still ignore the impact of different methods of soil sample preservation on the results of metabarcoding studies and there is no guideline for best practices so far. Here, we assessed the impact of four methods of soil sample preservation that can be conveniently used also in metabarcoding studies targeting remote or difficult to access areas. Tested methods include: preservation at room temperature for 6h, preservation at 4°C for three days, desiccation immediately after sampling and preservation for 21 days, and desiccation after 6h at room temperature and preservation for 21 days. For each preservation method, we benchmarked resulting estimates of taxon diversity and community composition of three different taxonomic groups (bacteria, fungi and eukaryotes) in three different habitats (forest, river bank and grassland) against results obtained under ideal conditions (i.e. extraction of eDNA right after sampling). Overall, the different preservation methods only marginally impaired results and only under certain conditions. When rare taxa were considered, we detected small but significant changes in MOTU richness of bacteria, fungi and eukaryotes across treatments, but MOTUs richness was similar across preservation methods if rare taxa were not considered. All the approaches were able to identify differences in community structure among habitats, and the communities retrieved using the different preservation conditions were extremely similar. We propose guidelines on the selection of the optimal soil sample preservation conditions for metabarcoding studies, depending on the practical constraints, costs and ultimate research goals.

KEYWORDS: eDNA metabarcoding, sample storage, MOTU richness, α and β diversity, microbial communities, eukaryotes

INTRODUCTION

Environmental DNA (hereafter referred to as "eDNA") can be defined as the mixture of complex, sometimes degraded, DNA that microorganisms (e.g. bacteria and fungi) or macro-organisms (e.g. animals, plants) leave behind in their environment (i.e. soil, water, sediments, etc.). By studying short, taxonomically-informative DNA fragments obtained from eDNA samples, it is possible to identify the associated taxa and therefore to survey biodiversity. Coined as “eDNA metabarcoding”, this approach has revolutionized several branches of ecology and environmental sciences during the last decade, by providing relatively quick and non-invasive assessments of present or past biodiversity of animals, plants and microorganisms (Taberlet, Bonin, Zinger, & Coissac, 2018). Metabarcoding is particularly valuable for monitoring biodiversity over large geographical or taxonomic scales (De Vargas et al., 2015; Delgado-Baquerizo et al., 2018; Zinger et al., 2019b). Furthermore, it gives access to biodiversity components that are elusive to conventional survey methods. For instance, it allows the rapid assessment of microbial soil biodiversity, which is extremely complex, time-consuming and imperfect when using direct observations, culturing techniques or microscopy (Giovannoni, Britschgi, Moyer, & Field, 1990; Ward, Weller, & Bateson, 1990).

Metabarcoding relies on a succession of several steps: 1) sampling; 2) preservation of the collected material until lab processing; 3) DNA extraction; 4) PCR amplification of a selected genomic region; 5) high-throughput sequencing of amplicons; and 6) analysis of sequences using bioinformatics and statistical tools (Zinger, Bonin, et al., 2019). Each step is critical to obtain robust taxonomic inventories and diversity estimates, and an increasing number of studies has assessed how methodological choices across the different steps could influence the conclusions of a study (Calderón-Sanou, Münkemüller, Boyer, Zinger, & Thuiller, 2020; Cantera et al., 2019; Chen & Ficetola, 2020; Nichols et al., 2018; Taberlet et al., 2018). Despite this growing body of literature, so far little attention has been devoted to the effect of different preservation conditions of the collected environmental material before lab processing (i.e. step 2). We thus know little about the optimal storage conditions of the collected material, and how long samples can be stored to limit biases in taxonomic inventories.

Some recent studies have analysed the preservation of sampling material obtained from water (see e.g. Kumar, Eble, & Gaither, 2020; Majaneva et al., 2018). Conversely, methodological analyses on the effects of sample preservation are largely scarce for soil, perhaps because the majority of metabarcoding studies have so far been performed in temperate areas where access to lab facilities is often easy (Hoffmann, Schubert, & Calvignac-Spencer, 2016; Huerlimann et al., 2020). In such cases, sample preservation is sometimes not necessary at all, or at least not over long periods of time. However, one great promise of metabarcoding is its potential for providing biodiversity data for remote areas, where biodiversity monitoring is essential but difficult. When sampling in remote or inaccessible areas (e.g. tropical and arctic areas; mountain chains), samples are rarely collected nearby lab facilities and an immediate in situ DNA extraction is generally not possible due to logistic constraints (but see Zinger, Taberlet, et al., 2019b for a notable exception). More generally, with the ever-increasing number of samples analyzed during a typical metabarcoding study, sample preservation is more and more indispensable, and the time lag between sample collection and subsequent molecular processing makes it particularly relevant to understand the impact of sample preservation, and to identify preservation strategies that do not bias the conclusions of studies.

In an ideal metabarcoding study, communities recovered from preserved samples should match those retrieved if samples had been processed immediately after sampling. However, inappropriate preservation conditions can cause both DNA degradation and the proliferation of certain taxonomic groups with respect to others, before DNA extraction (Cardona et al., 2012; Orchard, Standish, Nicol, Dickie, & Ryan, 2017). This can in turn affect taxa detection and also the relative contributions of different taxonomic groups to the estimated biodiversity. A recent review suggested that the majority of eDNA metabarcoding studies does not provide accurate information about sample treatment before processing (Dickie et al., 2018). Almost half of the studies do not report how samples were stored and conserved, and 30% of them store samples at 0-4°C, and thus at a temperature where many bacteria and fungi continue to be active and potentially affecting the whole sample. About 15% of the studies stored samples in a range of 5-35°C, which can be considered a poor practice when no preservatives are added (Dickie et al., 2018), and only 10% stored them below 0°C (Dickie et al., 2018).

So far, the consequences of preservation practices and the resulting deviations from immediate processing and analyses have rarely been studied quantitatively. Some studies suggested that liquid-based preservatives have limited usefulness for soil eDNA (Tatangelo, Franzetti, Gandolfi, Bestetti, & Ambrosini, 2014). Yet, Lauber, Zhou, Gordon, Knight, & Fierer (2010) tested the effect of storing samples from soil, human gut and skin at different temperatures and did not detect any significant effect on bacterial communities, while Orchard et al. (2017) found that storage time and temperature can affect colonization by arbuscular mycorrhizal fungi, with subsequent impacts on the reconstruction of communities. Differences between these studies may be due to their different protocols. However, they also focused on different taxonomic groups, which may react differently to storage period and temperature. Consequently, these studies are difficult to compare, highlighting the importance of formal assessments of preservation methods. Desiccation is a further approach that can efficiently conserve high-quality DNA for genomic studies (e.g. Alsos et al., 2020; Chase & Hills, 1991). Although not widely used for metabarcoding samples, desiccation through silica gel has good potential for soil, as it allows removing >25% of its weight in water in a few hours (P. Taberlet, pers. communication), is cost-effective, easy to transport, and is not an issue for aircraft transportation (no flammable or dangerous preservatives). A clear understanding of the effect of different preservation methods, especially across various groups of taxa, is thus pivotal for a robust application of eDNA metabarcoding to biodiversity monitoring in general, and that of remote areas in particular.

Here, using eDNA metabarcoding of different taxonomic groups in soil systems, we tested: (i) how preservation methods influence overall richness estimates and what the role of rarely observed taxa is; (ii) how preservation methods influence identified community structure and its turnover between different habitats; and (iii) what the best practices are under limited laboratory access. More specifically, we first selected three soil preservation methods (room temperature, 4°C, desiccation in silica gel) because they are commonly used in the literature (room temperature and 4°C) or because they are easy to implement in the field (desiccation and room temperature). Then, we assessed the impact of these preservation methods applied to different durations in order to mimic logistic constraints (see Supplemental Information, Appendix A for details on experimental design), and compared the communities obtained with those

observed in ideal conditions, i.e. when eDNA is extracted immediately after sampling (within less than one hour). We examined bacterial, fungal and eukaryotic communities to cover a broad taxonomic range, since different taxa can be differentially affected by sample preservation conditions (Cardona et al., 2012; Orchard et al., 2017).

MATERIALS AND METHODS

Soil preservation and experimental treatments

In April 2019, we collected soil samples from three habitats: a grassland (N 45.194° E 5.776°), a broadleaved forest (N 45.196° E 5.774°), and a vegetated river bank (N 45.195° E 5.780°). The study design was optimized to allow DNA extraction immediately after sampling, which hampered using distant study sites. All sites were within 400 m from the Laboratoire d'Écologie Alpine (LECA) in Grenoble, France. We choose habitats with some differentiation to allow different communities but not too extreme and relatively close together so that we expect some overlap between communities. This mimics what is commonly done in the field when gradients are sampled. We established five plots within each habitat; the minimum distance between nearby plots was about 20 m. Within each plot, we collected four soil samples (with a minimum distance of one meter) at a depth of 0–20 cm and then pooled them together, for a total of five pooled samples per habitat (approx. 200 g each pooled sample). Soil litter was not included in the samples. Pooled samples (15 in total) were homogenized; subsequently, from each of them we took five subsamples of 15 g of soil (total: 75 subsamples; Fig. 1).

The five soil subsamples of each pooled sample were subjected to five different treatments: 1) eDNA was extracted immediately after sampling (within 1 h; treatment hereafter referred to as "control"); 2) samples were preserved at room temperature (21-23°C) and eDNA was extracted 6 h after sampling; 3) samples were inserted in sterile 50-mL falcon tubes and preserved at 4°C. eDNA was extracted three days after sampling; 4) samples were inserted in hermetic, sterile boxes with 20 g of silica gel immediately after sampling, then stored at room temperature, and eDNA was extracted 21 days after sampling; 5) samples

were inserted in hermetic, sterile boxes with 20 g of silica gel 6h after sampling, then stored at room temperature, and eDNA was extracted 21 days after sampling.

We did not test full-factorial combinations of different preservation periods and conditions, which was not feasible in terms of time and costs. We did not consider freezing or storage in liquid nitrogen, which is unrealistic when dealing with large numbers and / or volumes of samples, as is the case for large-scale metabarcoding studies. Furthermore, freezing is generally impossible in remote areas (Dickie et al., 2018), where maintaining a cold chain cannot be ensured given the logistical challenges and this is often replaced by preservation in a cool box (4°C). Previous studies showed that preservative solutions have limited usefulness for soil or sediment samples (Rissanen, Kurhela, Aho, Oittinen, & Tiirola, 2010; Tatangelo et al., 2014).

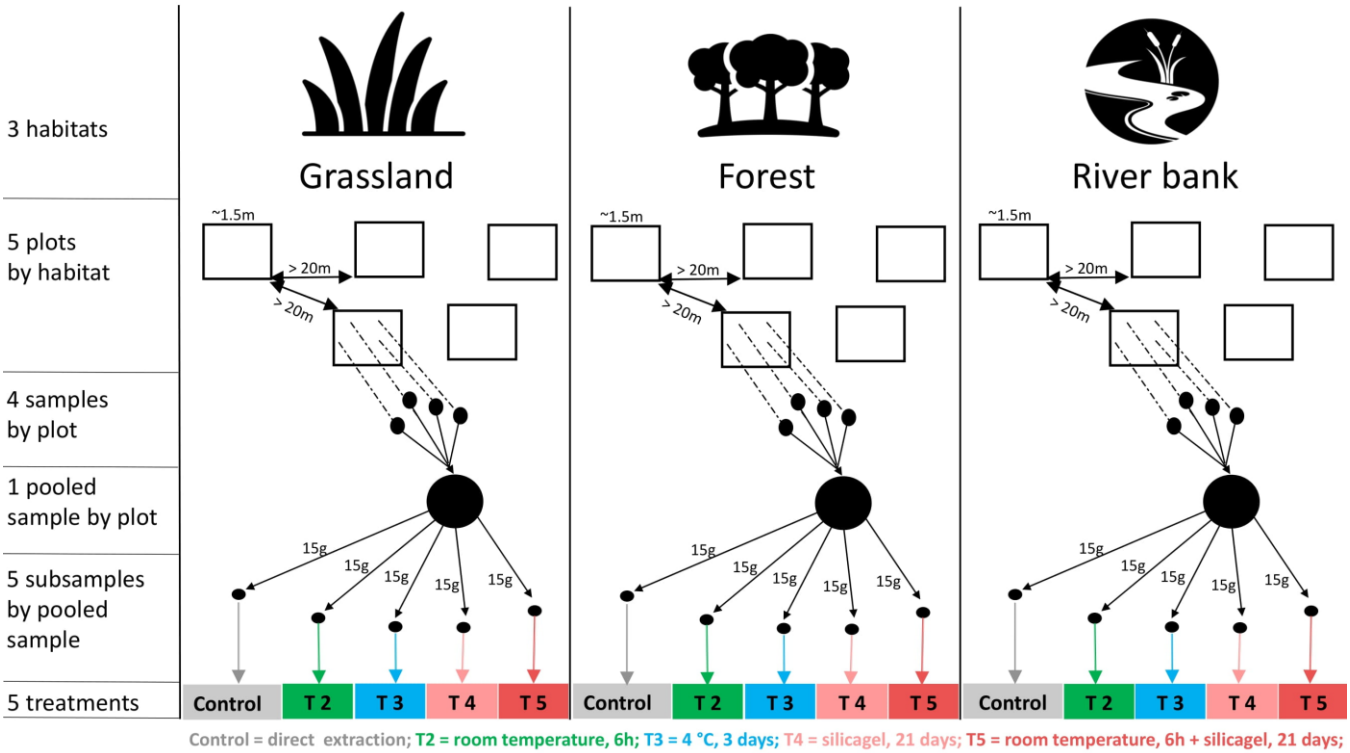


Figure 1. Experimental sampling design.

Molecular analyses

For all sample treatments, eDNA extraction was performed in a dedicated room using the NucleoSpin® Soil Mini Kit (Macherey-Nagel, Germany), after a preliminary step where 15 g of soil were mixed with

20 ml of phosphate buffer for 15 minutes as described in (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012); and with a final elution in 150 µl. We also included one extraction negative control per treatment.

Environmental DNA of bacteria, fungi and eukaryotes was amplified using primers designed for markers Bact02 (Taberlet et al., 2018), Fung02 (Epp et al., 2012; Taberlet et al., 2018) and Euka02 (Guardiola et al., 2015), respectively. Bact02 and Fung02 amplify fragments of about 220-250 bp, while Euka02 generally amplifies fragments <150 bp. The three markers are well suited for metabarcoding analyses, as all have a very low number of mismatches in the priming region across target organisms, and the relatively short length of amplified fragments allows their use with potentially degraded DNA (Taberlet et al., 2018). To allow bioinformatic discrimination of PCR replicates after sequencing, eight-nucleotide long tags were added on the 5' end of both forward and reverse primers, so that each PCR replicate was represented by a unique combination of forward and reverse tags. Tags had at least five nucleotide differences among them (Coissac, 2012). Samples were randomized on a 96-well plate, along with the five extraction controls, eight bioinformatic blanks, six PCR negative controls and two PCR positive controls. PCR positive controls were included to check for potential cross-contaminations and to monitor amplification and sequencing performances. The positive control was a 1:10 dilution of the ZymoBIOMICS™ Microbial Community DNA Standard II (Zymo Research, USA) constituted 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.

In order to avoid over-amplification of template DNA and to limit chimera formation, we determined the optimal number of amplification cycles and DNA extract dilution using qPCR. The qPCR assay was conducted on 48 randomly selected samples, using 1 µl of 1:1000 diluted SYBR® Green I nucleic acid gel stain (Invitrogen™, USA), with a real-time PCR thermal cycler set to standard mode. qPCR was performed for both 1:10 diluted and undiluted template eDNA.

For Bact02 and Fung02, PCR reactions were performed on 1:10 diluted template DNA, using 32 and 44 cycles respectively. For Euka02, we performed 34 cycles on undiluted DNA. All PCR reactions consisted of 10 µl of AmpliTaq Gold 360 Master Mix 2X (Applied Biosystems™, Foster City, CA, USA), 2 µl of primers mix at initial concentration of 5 µM of each primer, 0.16 µl of Bovine Serum Albumin (corresponding to 3.2 µg; Roche Diagnostic, Basel, Switzerland) and 2 µl of DNA extract, for a final volume of 20 µl. The PCR profiles had an initial step of 10 min at 95°C, followed by several cycles of a 30 s denaturation at 94°C, a 30 s annealing at 53°C (Bact02), 56°C (Fung02) or 45°C (Euka02), and a 90 s elongation for Bact02 and Fung02, or a 60 s elongation for Euka02 at 72°C, followed by a final elongation at 72°C for 7 minutes. The amplification was performed in 384-well plates, with four replicates for each sample. After amplification, PCR products of the same marker were pooled together in equal volumes and a 5-µl aliquot of the pooled amplicons was visualized by high-resolution capillary electrophoresis (QIAxcel Advanced System, QIAGEN, GERMANY) to verify the expected fragments length and to monitor primer dimers. Pooled amplicons were purified using the MinElute PCR Purification Kit (QIAGEN, GERMANY) following the manufacturer's protocol. Six subsamples of the pool of amplicons were purified separately for each marker, and then combined again before being sent for library preparation and sequencing to Fasteris (SA, Geneva, Switzerland). One library was prepared per marker using the MetaFast protocol (Taberlet et al., 2018) and then sequenced using the MiSeq (Fung02 and Bact02) or HiSeq 2500 (Euka02) platforms (Illumina, San Diego, CA, USA) with a paired-end approach (2 × 250 bp for Fung02 and Bact02, and 2 × 150 bp for Euka02).

Bioinformatic treatment

The bioinformatic treatment of sequence data was performed using the OBITools software suite (Boyer et al., 2016). First, forward and reverse reads were assembled using the illuminapairedend program, keeping only sequences with an alignment score higher than 40. Aligned sequences were assigned to the corresponding PCR replicate using the program ngsfilter, by allowing two and zero mismatches on primers and tags, respectively. After sequence dereplication using obiuniq, bad-quality sequences (i.e. containing

"N"), sequences whose length fell outside the expected size interval (below 45 bp for Bact02, below 68 bp Fung02 and below 36 bp for Euka02) and singletons were filtered out. The obiclean program was run to detect potential PCR or sequencing errors with the -r option set at 0.5: in a PCR reaction, sequences are tagged as "heads" when they are at least twice as abundant as other related sequences differing by one base. Only the sequences tagged as "heads" in at least one PCR were kept.

Taxonomic assignment was conducted using the ecotag program based on a reference database constructed from EMBL (version 136) by running the ecoPCR program (Ficetola et al., 2010). More specifically, ecoPCR carried out an in-silico PCR with the primer pair used for the experiment and allowing three mismatches per primer. The obtained reference databases were further curated by keeping only the sequences assigned at least at the family level.

Further data filtering was performed in R version 3.6.1 (R Core Team, 2018) to remove spurious sequences that can bias ecological conclusions drawn from DNA metabarcoding data (Calderón-Sanou et al., 2020). More specifically, we discarded all MOTUs with best identity <85% (Fung02, Bact02) or <80% (Euka02). These MOTUs were indeed rare (31%, 1.7% and 7.3% of reads for Fung02, Bact02 and Euka02, respectively) and their effect on the NMDS was marginal, as observed in other studies (e.g. Botnen et al., 2018). Furthermore, we removed MOTUS with less than five occurrences in the overall dataset, detected in more than one extraction or PCR negative control (Zinger, Bonin, et al., 2019a), or that were detected in less than two PCR replicates of the same sample, as they often represent false positives (Ficetola et al., 2015).

Statistical analyses

For all taxonomic groups, we used Generalized Linear Mixed Models (GLMMs) to test if the different treatments lead to differences in the observed MOTU richness. In GLMMs, the number of MOTUs per sample was calculated and used as a dependent effect, the five treatments were used as predictors, and sample identity was used as a random factor. The model was performed with the generalized poisson

distribution error using the R package glmmTMB (Brooks et al., 2017), in order to consider overdispersion (Consul & Famoye, 1992). If GLMM detected significant differences among treatments, we used treatment contrasts to test if each treatment led to communities significantly different from those unraveled by the "control" condition. Treatment contrasts are standard non-orthogonal contrasts, in which each category (treatment) is compared to a user-defined reference category, and are appropriate to compare multiple treatments against one single control category (in this case, immediate extraction; (Field, Miles, & Field, 2015). The uncorrected number of MOTUs tends to overestimate the actual taxonomic richness (Calderón-Sanou et al., 2020). Therefore, we repeated this analysis twice: considering all the observed MOTUs, and considering only MOTUs with frequency ≥ 1% in each sample (hereafter referred to as "common MOTUs").

Subsequently, we used multivariate analyses to assess the variation of bacteria, fungi and eukaryotic communities across habitats and treatments. Before running multivariate analyses, we calculated the proportion of reads of each MOTU in each sample. Relative abundance values were then transformed using the Box-Cox transformation, which simultaneously solves the double-zero problem and improves the multivariate normality of data (Legendre & Borcard, 2018).

First, we used Nonmetric Multi-Dimensional Scaling (NMDS) to describe differences in communities among the three habitats, and check whether different treatments yield different interpretations of ecological relationships among samples. NMDS uses an optimization process to find a configuration of points (samples) in a space with a small number of dimensions, and is suitable for metabarcoding analyses that aim to reconstruct variation in community composition as well as possible, without preserving any particular distance measure among objects (Borcard, Gillet, & Legendre, 2011; Chen & Ficetola, 2020; Paliy & Shankar, 2016). Given its robustness and flexibility, NMDS is often used as the first step to characterize the similarity of communities in metabarcoding studies (Chen & Ficetola, 2020; Paliy & Shankar, 2016). NMDS was run on the Euclidean distance computed on Box–Cox-chord-transformed data (Legendre & Borcard, 2018), by building 1,000 ordinations.

Second, we used ProcMod, a Procrustes-based analysis (Coissac & Gonindard-Melodelima, 2019), to measure the multivariate correlations between the communities obtained using the different treatments. ProcMod can be used to measure the shared variation between matrices, and is particularly appropriate to test relationships between datasets obtained through DNA metabarcoding and metagenomics (Coissac & Gonindard-Melodelima, 2019). Procrustes analyses tend to overfit the data; therefore, we used a modified version of Procrustes correlation that is robust to highly-dimensional data and allows a correct estimation of the shared variation between data sets (Coissac & Gonindard-Melodelima, 2019). The Procrustes-based correlation tests were performed using the corls function in the R package ProcMod, using 1,000 randomizations to test the mean covariance between random matrices (Coissac & Gonindard-Melodelima, 2019).

Third, we used redundancy analysis (RDA) to measure the amount of variation among communities that is explained by differences in habitat and treatments (Legendre & Legendre, 2012; Ter Braak, 1986). With habitat typology and treatment as constraining matrices, we used treatment contrasts to test if each treatment led to communities significantly different from those unraveled by the control treatment. Thus, significant treatment contrasts indicate that results between control and experimental treatments differ in an important way, while non-significant results mean that deviation from ideal conditions is not specifically pronounced. Significance of RDA and treatment contrasts was tested through 10,000 permutations using the vegan package in R (Borcard et al., 2011; Oksanen et al., 2019).

For bacteria only, RDA detected significant differences between the control and some of the treatments. We thus ran a similarity percentage analysis with the simper R function (Clarke, 1993) from vegan to identify the taxa contributing to the overall pairwise treatment difference (Geyer et al., 2014). Significance was tested using 50,000 permutations. Given the large number of tests performed, the significance of tests was corrected using the False Discovery Rate (FDR) method with the fdrtool package (Strimmer, 2008). FDR has greater power than traditional approaches (e.g. Bonferroni correction) when performing multiple comparisons (Benjamini & Hochberg, 1995). All statistical analyses were performed in the R environment.

RESULTS

A total of 6.3, 7.9 and 25.7 million reads were obtained from the Bact02, Fung02 and Euka02 libraries, respectively. After read assembly, quality filtering, spurious sequence and contaminant removal, 481,411; 2,511,721 and 13,232,441 good-quality sequences remained, consisted of 660 (Bact02), 1,075 (Fung02) and 3,611 (Euka02) unique sequences (i.e. MOTUs).

Differences in MOTU richness among treatments

Generalized Linear Mixed Models allowed identifying shifts in the richness of observed MOTUs. When we considered all the detected MOTUs, GLMM detected significant differences in MOTUs richness among treatments for all the markers considered (Bact02: $\chi_4 = 38.9$, $P < 0.001$; Fung02: $\chi_4 = 18.2$, $P = 0.001$; Euka02: $\chi_4 = 21.7$, $P < 0.001$; Fig. 2). Compared to the control, contrasts showed small but significant changes in MOTUs richness under the 4°C treatment (Bact02: $z = 2.54$, $P = 0.010$; Fung02: $z = -2.17$, $P = 0.029$; Euka02: $z = 2.65$, $P = 0.008$), the silicagel treatment (Bact02: $z = -2.93$, $P = 0.003$; Fung02: $z = -3.99$, $P < 0.001$; Euka02: $z = 3.92$, $P < 0.001$), and the silicagel+6h treatment (Bact02: $z = -3.74$; Fung02: $z = -4.02$; Euka02: $z = 4.18$; all $P < 0.001$). The 6h treatment caused a small but significant decrease in MOTUs richness for fungi ($z = -2.42$; $P = 0.015$), but not for bacteria and eukaryotes ($P = 0.456$, $P = 0.283$, respectively; for all contrasts: Table S1).

Nevertheless, when we repeated analyses by excluding MOTUs with a frequency <1%, differences in richness were much smaller, and were only significant for bacteria and fungi (Bact02: $\chi_4 = 9.69$, $P = 0.045$; Fung02: $\chi_4 = 14.1$, $P = 0.006$; Euka02: $\chi_4 = 2.22$, $P = 0.693$; Fig. 2). Compared to the control, MOTUs richness decreases for Bact02 under the 4°C treatment ($z = -2.91$; $P = 0.003$) and increases for Fung02 under the two silicagel treatments ($z = 2.77$; $P = 0.005$; $z = 1.75$; $P = 0.080$; respectively), while no significant effect was detected for Euka02 under any of the treatments (all $P > 0.170$; for all contrasts: Table S1).

Habitat caused a significant effect in MOTUs richness only for Fung02 both before and after removing rare MOTUs (before: $\chi_1 = 11.8$, $P < 0.001$; after: $\chi_1 = 20.5$, $P < 0.001$).

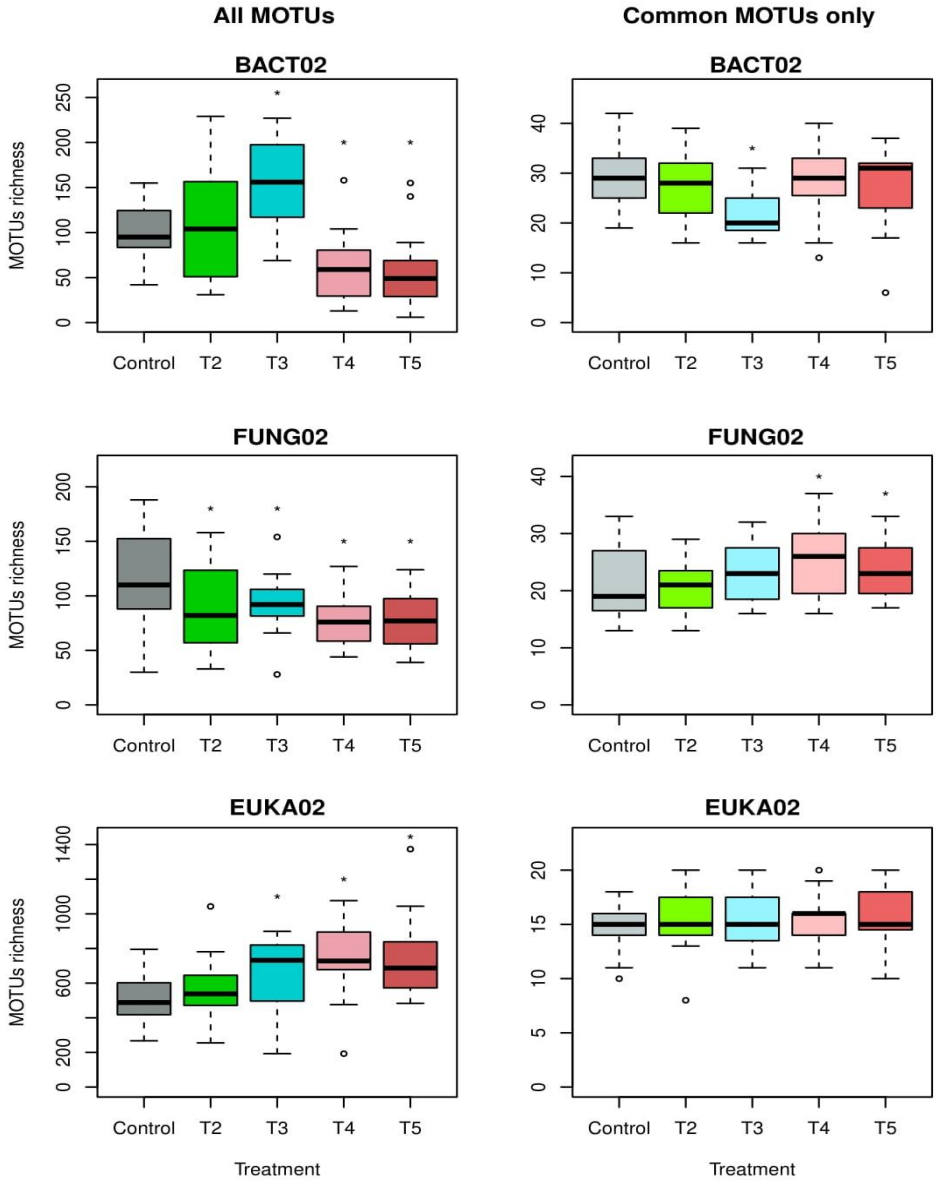


Figure 2. MOTU richness across the different treatments (control: immediate extraction; T2: extraction after 6 hr at room temperature; T3: extraction after 3 days at 4°C; T4: immediate preservation in tubes with silica gel, extraction after 21 days; T5: preservation in tubes with silica gel after 6 hr at room temperature, extraction after 21 days) before (left) and after (right) removing MOTUs with frequency < 1% in each sample.

Ecological similarity of communities among treatments

Nonmetric Multi-Dimensional Scaling showed a stress value of 0.13 for Bact02, 0.14 for Fung02 and 0.12 for Euka02. For each of the three markers, the NMDS plots obtained for the five sample preservation

treatments were extremely similar, and the ecological differences among the three habitats were clearly identified by all the preservation treatments (Fig. 3).

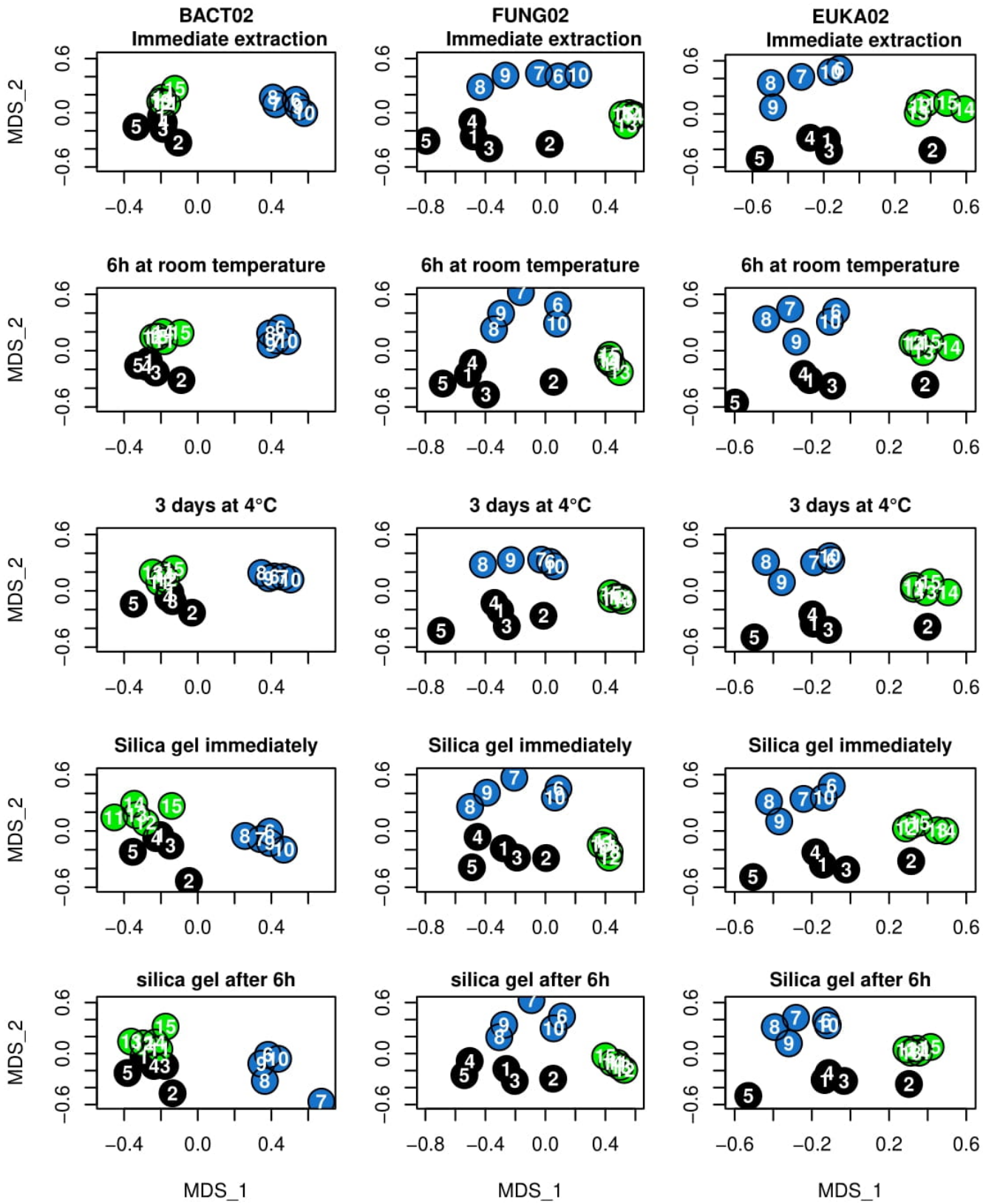


Figure 3. Plots of nonmetric dimensional scaling showing dissimilarities of communities among the three habitats: broadleaved forest (black); grassland (green); vegetated riverbank (blue). Each plot shows the results of metabarcoding analysis based on soil samples subjected to five different treatments.

The multivariate correlation between the communities obtained with the five treatments was always very strong (Procrustes-modified correlation: for all comparisons between "control" and treatments $r \geq 0.84$, $P < 0.0001$; Fig. 4) indicating, for all markers, that most of the variation of retrieved communities was shared across all the treatments. Procrustes correlations were particularly high for Fung02 and Euka02 (all $r \geq 0.9$), and between the control and the treatments 6h and 4°C (all $r \geq 0.93$; Fig. 4).

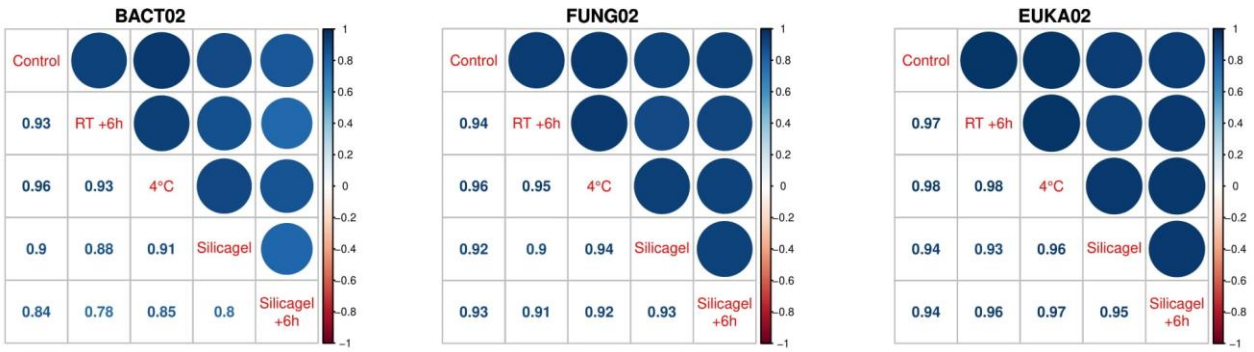


Figure 4. Procrustes correlation between communities obtained from metabarcoding analyses based on soil samples across environmental conditions subjected to five sample treatments (control: immediate extraction; RT + 6 hr: extraction after 6 hr at room temperature; 4°C: extraction after 3 days at 4°C; silicagel: immediate preservation in tubes with silica gel, extraction after 21 days; silicagel + 6 hr: preservation in tubes with silica gel after 6 hr at room temperature, extraction after 21 days. All correlation coefficients are highly significant (all $p < .0001$).

Differences between the obtained communities

Redundancy analysis allowed us to measure the amount of variation explained by differences among habitats and by treatments. Overall, 33%, 24%, and 33% of variability was explained by differences in habitat for bacteria, fungi, and eukaryotes, respectively. The community differences among habitats were strongly significant for the three taxonomic groups (permutation test: all $P \leq 0.001$). Differences among treatments were much weaker, and explained 9%, 2% and 2% of variation only for bacteria, fungi and eukaryotes, respectively. Differences between treatments were significant for bacteria (permutation test: $P < 0.0001$), but not for fungi and eukaryotes (both $P = 1$).

For bacteria, contrasts did not detect significant differences between control and the 6h or 4°C treatments. Differences between control and the two silicagel treatments were significant but explained a limited amount of variation (for both treatments, ≈3% of variation explained; $P < 0.0001$; Table 1). We thus used similarity percentage analysis to identify the MOTUs significantly contributing to these differences. Only one single MOTU showed a significant contribution ($P = 0.03$ after FDR correction) to the differences between control and silicagel treatment; this MOTU (belonging to the Bacteroidetes phylum) showed a very limited frequency under the silicagel treatment (Fig. S1). After FDR correction, no MOTU showed a significant contribution to the differences between control and the silicagel+6h treatment.

Table 1. Treatment contrasts assessing differences between the control (immediate extraction) and four approaches to soil conservation before eDNA extraction. Each conservation treatment was compared against the control in order to determine the percentage of explained variability.

	Bact02		Fung02		Euka02	
	Explained variability (%)	*p*	Explained variability (%)	*p*	Explained variability (%)	*p*
Treatment 2: room temperature, extraction after 6 hr	0.58	.956	0.53	1	0.44	.993
Treatment 3:4°C, extraction after 3 days	0.81	563	0.58	.976	0.54	.949
Treatment 4: silica gel immediately inserted, extraction after 21 days	3.14	<.001	1.03	.518	0.63	.876
Treatment 5: silica gel inserted after 6 hr, extraction after 21 days	3.16	<.001	0.73	.880	0.75	.725

DISCUSSION

Monitoring soil biodiversity with eDNA metabarcoding over large geographical and taxonomic scales and sometimes in remote places is increasingly important in ecological research. Understanding how preservation conditions affect estimates of taxonomic richness and community composition is essential to ensure sound conclusions. Our study shows that soil metabarcoding results are surprisingly robust to preservation conditions, as we observed limited differences in community structure and diversity estimates when samples were preserved using different strategies. However, some taxonomic groups and diversity

components are more sensitive than others to certain preservation conditions. This allowed us to develop guidelines for preservation depending on the aims of monitoring programs and on focal taxa.

The aim of this study was comparing realistic approaches to soil preservation against an ideal situation. Immediate extraction was our reference approach, as it avoids both DNA degradation (i.e. potential under-representation of certain taxa) and continued growth of certain taxonomic groups (i.e. potential over-representation of other taxa). Unfortunately, immediate extraction is only possible if sampling occurs nearby facilities, or when a mobile eDNA laboratory is available (e.g. Zinger, Taberlet, et al. 2019b), and logistical constraints often hamper its application in remote areas. We selected preservation conditions among the most achievable, cost-effective and frequent practices to sampling soil for eDNA studies (Dickie et al., 2018), although additional storage methods (e.g. liquid nitrogen, dry ice, RNA later) are available.

Influence of preservation methods on richness estimates

Preservation methods generated some small but significant differences in MOTUs richness compared to what is observed in the "control", with some contrasting effects across taxa. When considering all the MOTUs, none of the preservation conditions yielded estimates of alpha-diversity identical to the "control". For instance, just six hours at room temperature caused a significant decrease of MOTUs richness in fungi. It has been shown that estimates of alpha-diversity using metabarcoding are extremely sensitive to methodological choices (Calderón-Sanou et al., 2020). Our study underlines that even preservation for a very short time can affect the detection of rare MOTUs and highlights the sensitivity of fungi to preservation at room temperature (Delavaux, Bever, Karppinen, & Bainard, 2020). MOTUs richness of all the taxa was also affected by preservation at 4°C, which caused a slight increase of MOTUs richness for bacteria and eukaryotes, and a slight decrease for fungi. The effect of temperature and time storage in fungal and bacterial growth has already been proven (see e.g. Orchard et al., 2017; Pettersson

& Bååth, 2003). Despite this, in addition to temperature, we can expect that other parameters such as initial soil moisture and pH influence bacterial growth (Bååth & Arnebrant, 1994; Drenovsky, Vo, Graham, & Scow, 2004; Fernández-Calviño & Bååth, 2010; Kaiser et al., 2016) with a combined effect. Finally, drought affects the richness of microbial communities in soil ecosystems with differential effects across taxa depending on their ecology (Evans, Wallenstein, & Burke, 2014; Meisner, Jacquiod, Snoek, Ten Hooven, & van der Putten, 2018; Ochoa-Hueso et al., 2018), and three weeks of preservation with silica gel generally reduced the observed MOTUs richness in bacteria and fungi, while it increased the richness of eukaryotes.

However, our study also shows that specific caution is mostly necessary when rare MOTUs are of interest. The exclusion of rare and uncommon MOTUs strongly reduced differences between optimal conditions and different preservation. The remaining effects were much weaker for bacteria and fungi, while disappeared for eukaryotes (Fig. 2), and can be due to the heterogeneous growth across taxonomic groups, or to differential DNA degradation under different preservation conditions. This suggests that the effect of preservation approach on taxonomic richness mostly occurs on rare species, as already suggested for microbial communities (Meisner et al., 2018). Several authors have shown that eDNA metabarcoding does not represent the best tool for the detection of rare MOTUs, as some rare MOTUs remain undetected, while many sequences detected at rare frequency are artifacts (Brown et al., 2015). Estimates of α-diversity should therefore always be taken with caution, and indices that underweight rare MOTUs (e.g. Shannon or Simpson diversity) can provide more robust estimates (Brown et al., 2015; Calderón-Sanou et al., 2020; Bálint et al., 2016).

Differences in community structure

If the study interest is in community structure and not in richness estimates, then preservation choices become even less important. In fact, the similarity of communities obtained through the different

preservation conditions is surprisingly high (see Procrustes correlation coefficients; Fig. 4); the amount of variation explained by preservation conditions was much lower than the observed differences among habitats (see redundancy analysis), and multivariate analyses consistently allowed to detect community differences among habitats (Fig. 3). In other words, metabarcoding is able to identify the ecological differences among sites, independently of the preservation approach. Even though metabarcoding analyses are sensitive to methodological choices, estimates of relationships between diversity and the environment are often robust (Calderón-Sanou et al., 2020; Ji et al., 2013), and this is a very good news if we want to apply these approaches to broad-scale monitoring programs, aiming at assessing the effects of environmental changes.

Bacteria were the only taxon for which we detected significant differences between the "control" and the preservation conditions, with ≈3% of variability explained by differences between the "control" and the desiccation treatments. The observed differences most likely refer to some taxa that are affected by the dry conditions and could lead to an overrepresentation of some taxa that are more resistant under these conditions. We expect some taxa to better survive and grow in dry treatments with respect to others, and this would make their DNA more available in the samples. Differences between the desiccation treatments and the control were small, and only one out of 660 MOTUs showed a significant variation in abundance with the control. This MOTU (belonging to the Bacteroidetes phylum of bacteria, see Supporting Information) was generally abundant in the control and preservation conditions 2 and 3 (average frequency of reads around 10%) while it drastically decreased under preservation conditions 4 and 5 (Fig. S1). This agrees with studies showing that different genera belonging to this phylum respond differently to drought (Meisner et al. 2018). In fact, the Bacteroidetes Flavobacterium shows an increase in abundance over dried-treatments, even though differences after three weeks were not significant (see Fig. S2). The significant differences observed for some taxa and preservation conditions stress the importance of selecting the preservation method before starting a monitoring program and using it consistently through the whole monitoring, to avoid confusion between the effects of methods and of environmental changes.

Conclusions: guidelines for optimizing preservation conditions

Standardized protocols are essential for repeatable and reliable biodiversity monitoring, and our results allow to propose guidelines to improve and standardize the preservation of soil samples for eDNA metabarcoding analyses (Fig. 5):

1) If sampling occurs close to lab facilities, or a mobile lab is available, extracting DNA as soon as possible is the best approach. Storing samples a few hours at room temperature does not have major impact on the outcome of analyses, especially if the focus is not on rare MOTUs;

2) If lab facilities are available after a short-time transportation, storing samples in the fridge (0-4°C) for a few days is a safe approach as it does not have a significant impact on community composition, and only moderately affects MOTUs richness. However, this approach can be problematic if the aim is to estimate MOTUs richness, and particularly the occurrence of rare MOTUs. The feasibility of this strategy also depends on the number and volume of samples, and to the possibility of maintaining the cold chain;

3) If monitoring in remote areas, sample desiccation (e.g. using silica gel) and long-term preservation at room temperature is a reasonable approach, and it is particularly convenient when working with a large volume of samples. This approach preserves ecological signal, but can affect the detection of some taxa, particularly among the rarest ones. Therefore, this approach is suboptimal for monitoring programs aiming at detecting rare MOTUs.

An effective application of eDNA metabarcoding to biodiversity monitoring is complex, and protocols of sample preservation are key methodological choices that have to be considered when designing a metabarcoding-based monitoring. When working in difficult and remote environments researchers are faced with the trade-offs between a faithful representation of biodiversity, and multiple logistic constraints in the field. Accurate a-priori planning is often the basis of successful monitoring programs and our

guidelines can help researchers and practitioners to identify the best approach to sample preservation, depending on the studied taxa and research goals.

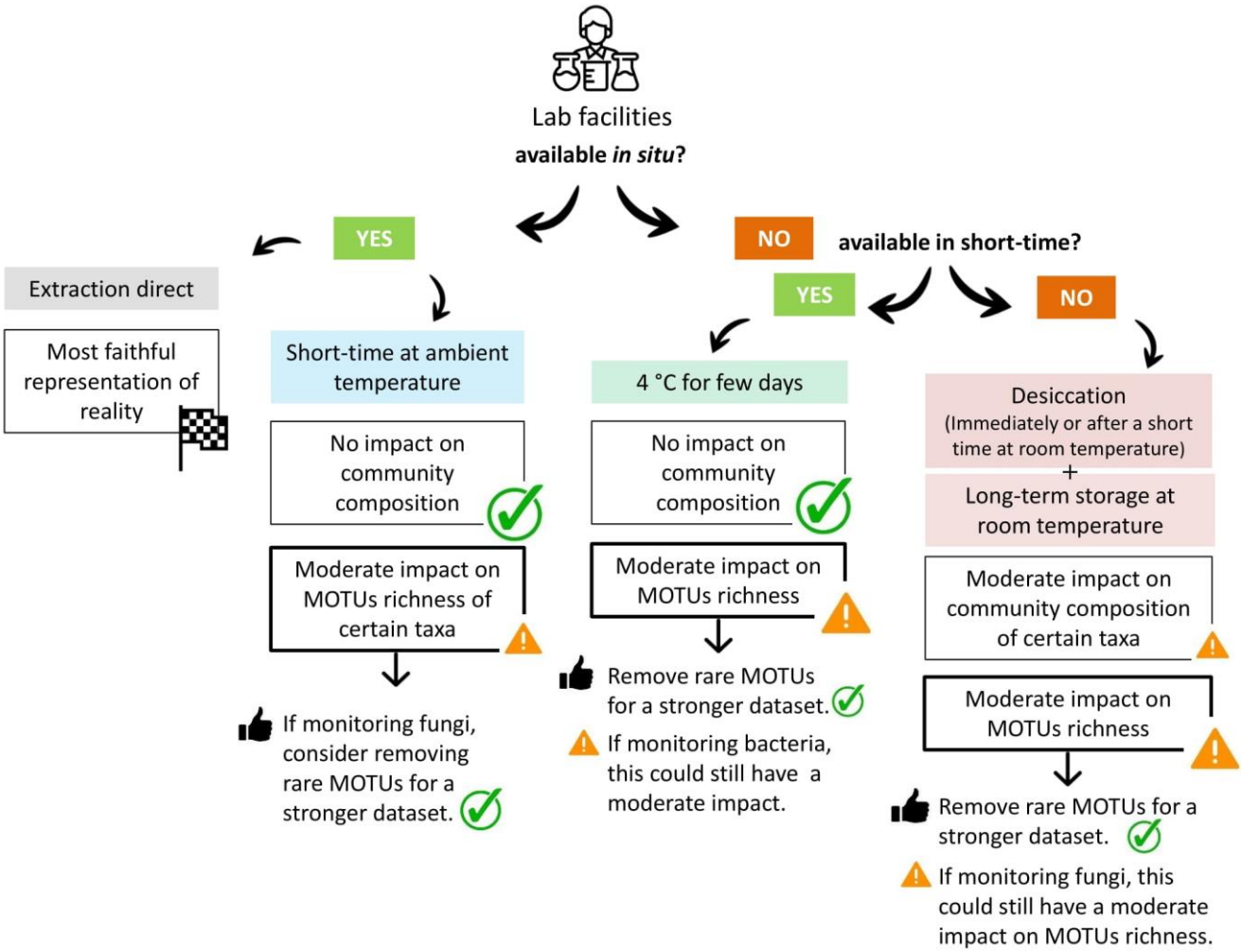


Figure 5. Guidelines for improving monitoring strategies with eDNA from soil.

ACKNOWLEDGMENTS

We thank P. Taberlet for constructive discussions. This study is funded by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). WT, TM and LG also thank the rom the French Agence Nationale de la Recherche (ANR) through the GlobNets (ANR-16-CE02-0009) project and from 'Investissement d'Avenir' grant managed by the ANR (Montane: OSUG@2020: ANR-10-LAB-56).

Supplemental Information for:

Effects of soil preservation for biodiversity monitoring using environmental DNA

Alessia Guerrieri, Aurélie Bonin, Tamara Münkemüller, Ludovic Gielly, Wilfried Thuiller, Gentile Francesco Ficetola

APPENDIX A: Design of soil preservation conditions

We did not test full-factorial combinations of different preservation periods and conditions, and we did not consider freezing, which is unrealistic when dealing with large numbers and / or volumes of samples, as is the case for more and more metabarcoding studies. Furthermore, freezing is generally impossible when sampling remote areas, where maintaining a cold chain cannot be ensured given the logistical challenges and, in the best cases, it is replaced by preservation in a cool box (i.e. 4°C or more).

In the design of treatments, we considered approaches allowing preservation at different temperatures and for different periods. For preservation condition 2, we accounted for a certain delay (six hours) between sampling and extraction, that could correspond to local transportation from the sampling area to the nearest base station. Sometimes, even in the case of an *in situ* extraction, samples remain at ambient temperature for hours prior to extraction, especially when monitoring a large area or a tricky ground, but this can have an impact on the final results (Delavaux et al., 2020). Preservation at 4°C is among the most frequent approaches to soil preservation (Dickie et al., 2018). This can be attained through portable refrigerators and requires the cold chain not to be interrupted at any point during transportation, which is only possible when lab facilities are accessible in a relatively short time (Hoffmann et al., 2016; Huerlimann et al., 2020). However, one of the most attractive characteristics of eDNA metabarcoding is its capacity to provide biodiversity data for understudied areas (e.g. tropical and arctic areas; mountain chains). Preservation conditions 4 and 5 of this study refer to the situations where samples are located far from lab facilities, in areas for which traditional biomonitoring is particularly challenging. When sampling these areas, the time lag between sample collection and subsequent molecular analysis can be particularly relevant. The use of silica gel to preserve soil samples is then a good solution (Chase & Hills, 1991; Guo, Yang, Chen, Li, & Guo, 2018): silica gel allows removing up to 26% of its weight in water in a relatively short time (overnight; Taberlet, pers. communication), is cost-effective, easy to transport, and is not an issue for aircraft transportation (no flammable or dangerous preservatives). Silica gel can be added *in situ* during sampling (protocol 4) or within a short time lag, allowing for local transportation from the sampling area to the nearest facilities (protocol 5). Afterwards, samples can be stored at room temperature, protected from sunlight to avoid photo-degradation of DNA.

REFERENCES

Chase, M. W., & Hills, H. H. (1991). Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *TAXON*, *40*(2), 215–220. doi: 10.2307/1222975

Delavaux, C. S., Bever, J. D., Karppinen, E. M., & Bainard, L. D. (2020). Keeping it cool: Soil sample cold pack storage and DNA shipment up to 1 month does not impact metabarcoding results. *Ecology and Evolution*, (November 2019), 1–13. doi: 10.1002/ece3.6219

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., … Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources*, *18*(5), 940–952. doi: 10.1111/1755-0998.12907

Guo, Y., Yang, G. Q., Chen, Y., Li, D., & Guo, Z. (2018). A comparison of different methods for preserving plant molecular materials and the effect of degraded DNA on ddRAD sequencing. *Plant Diversity*, *40*(3), 106–116. doi: 10.1016/j.pld.2018.04.001

Hoffmann, C., Schubert, G., & Calvignac-Spencer, S. (2016, February 1). Aquatic biodiversity assessment for the

lazy. *Molecular Ecology*, Vol. 25, pp. 846–848. doi: 10.1111/mec.13535
Huerlimann, R., Cooper, M. K., Edmunds, R. C., Villacorta-Rath, C., Le Port, A., Robson, H. L. A., … Jerry, D. R. (2020). Enhancing tropical conservation and ecology research with aquatic environmental DNA methods: an introduction for non-environmental DNA specialists. *Animal Conservation*. doi: 10.1111/acv.12583

SUPPLEMENTARY FIGURES

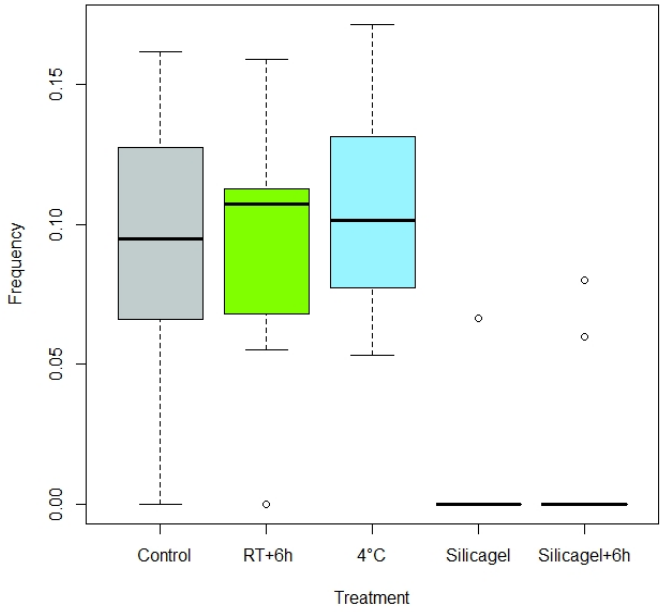


Figure S1. Frequency of MOTU n° 21 across the different treatments. This MOTU showed a significant contribution to the differences between the control treatment and the treatment "immediate silicagel", and a nearly-significant contribution to the differences between the control treatment and the treatment "silicagel after 6h".

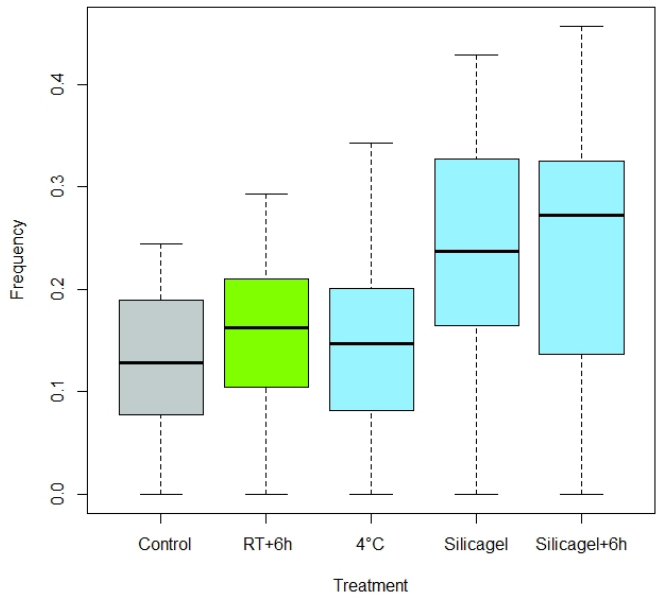


Figure S2. Frequency of MOTU n° 3 across the different treatments. This MOTU, belonging to the Bacteroidetes *Flavobacterium*, showed a weak and non-significant increase in abundance between the control treatment and the treatment "immediate silicagel" and "silicagel after 6h".

SUPPLEMENTARY TABLES

Table S1. Results of generalized linear mixed models for all MOTUs and for MOTUs with frequency (f) ≥ 1%.

	Bact02		Fung02		Euka02		Bact02 (f ≥ 1%)		Fung02 (f ≥ 1%)		Euka02 (f ≥ 1%)	
	z	P	Z	P	z	P	Z	P	z	P	z	P
Treatment 2	-0.74	0.456	-2.42	0.015	1.07	0.283	-0.78	0.438	-0.82	0.413	1.00	0.313
Treatment 3	2.54	0.010	-2.17	0.029	2.65	0.008	-2.91	0.003	1.33	0.184	0.98	0.323
Treatment 4	-2.93	0.003	-3.99	<0.001	3.92	<0.001	-0.28	0.77	2.77	0.005	0.50	0.613
Treatment 5	-3.74	<0.001	-4.02	<0.001	4.18	<0.001	-0.84	0.403	1.75	0.080	1.37	0.170

CHAPTER 3. ARTICLE B[3]

Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies

Aurélie Bonin[1,2]*, Alessia Guerrieri[1], G. Francesco Ficetola[1,3]

[1] Department of Environmental Science and Policy, University of Milan. Via Celoria 10, 20126 Milano Italy

[2] Argaly, Bâtiment CleanSpace, 354 Voie Magellan, 73800 Sainte-Hélène-du-Lac, France

[3] Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France

* Corresponding author: aurelie.bonin@argaly.com

[3] Submitted to Molecular Ecology Resources (Oct-2021)

ABSTRACT

Clustering approaches are pivotal to handle the many sequence variants obtained in DNA metabarcoding datasets, therefore they have become a key step of metabarcoding analysis pipelines. Clustering often relies on a sequence similarity threshold to gather sequences in Molecular Operational Taxonomic Units (MOTUs) that ideally each represent a homogeneous taxonomic entity, e.g. a species or a genus. However, the choice of the clustering threshold is rarely justified, and its impact on MOTU over-splitting or over-merging even less tested. Here, we evaluated clustering threshold values for several metabarcoding markers under different criteria: limitation of MOTU over-merging, limitation of MOTU over-splitting, and trade-off between over-merging and over-splitting. We extracted sequences from a public database for eight markers, ranging from generalist markers targeting Bacteria or Eukaryota, to more specific markers targeting a class or a subclass (e.g. Insecta, Oligochaeta). Based on the distributions of pairwise sequence similarities within species and within genera and on the rates of over-splitting and over-merging across different clustering thresholds, we were able to propose threshold values minimizing the risk of over-splitting, that of over-merging, or offering a trade-off between the two risks. For generalist markers, high similarity thresholds (0.96-0.99) are generally appropriate, while more specific markers require lower values (0.85-0.96). These results do not support the use of a fixed clustering threshold (e.g. 0.97). Instead, we advocate a careful examination of the most appropriate threshold based on the research objectives, the potential costs of over-splitting and over-merging, and the features of the studied markers.

KEYWORDS: metabarcoding marker; sequence variant; analysis parameter; MOTU over-splitting, MOTU over-merging; alpha diversity

INTRODUCTION

Metabarcoding studies are typically based on a succession of experimental steps governed by important methodological choices (Zinger et al. 2019). These include a) the definition of sampling design and selection of sampling sites (Dickie et al. 2018), b) the approach used for the preservation of starting material (Tatangelo et al. 2014, Guerrieri et al. 2021), c) the protocol used for DNA extraction (Taberlet et al. 2012, Eichmiller et al. 2016, Zinger et al. 2016, Lear et al. 2018, Capo et al. 2021), d) the selection of appropriate primers to amplify a taxonomically-informative genomic region (Elbrecht et al. 2016, Fahner et al. 2016, Ficetola et al. 2021), e) the strategy adopted for DNA amplification and high-throughput sequencing of amplicons (Nichols et al. 2018, Taberlet et al. 2018, Bohmann et al. 2022), f) the pipeline selected for bioinformatics analyses (Boyer et al. 2016, Calderón-Sanou et al. 2020, Capo et al. 2021, Couton et al. 2021, Macher et al. 2021, Mächler et al. 2021), and g) the statistical approach used to translate metabarcoding data into ecological or evolutionary information (Paliy and Shankar 2016, Chen and Ficetola 2020). Each of these methodological choices can heavily influence the reliability and interpretation of results (Alberdi et al. 2018, Zinger et al. 2019), and there is thus a critical need for the development, proper assessment and optimization of methods specially dedicated to metabarcoding.

When analysing metabarcoding data, bioinformatic pipelines generally produce a list of detected sequences, that can be assigned to a given taxon with a more or less precise taxonomic resolution. However, the number of unique sequences obtained after bioinformatic treatment is generally much higher than the number of taxa actually present in the sample (Calderón-Sanou et al. 2020, Mächler et al. 2021). This stems from multiple reasons including genuine intraspecific diversity of the selected markers and errors occurring during the amplification or sequencing steps. Consequently, sequence clustering approaches are often used to collapse very similar sequences into one single Molecular Operational Taxonomic Unit (MOTU), which does not necessarily correspond to a traditional species (Kopylova et al. 2016, Froslev et al. 2017, Bhat et al. 2019, Antich et al. 2021).

Sequence clustering can be performed using similarity thresholds, Bayesian approaches, or through single-linkage (Antich et al. 2021). Approaches based on similarity thresholds can have excellent performance and they display several advantages such as flexibility and easy implementation (Kopylova et al. 2016, Wei et al. 2021). However, two key parameters have to be determined *a priori* when performing clustering based on sequence similarity. The first one is the sequence to be selected as representative of the cluster. In the case of metabarcoding studies, keeping the most abundant sequence of the cluster as the cluster representative is a convenient way of merging sequence variants generated during the PCR or sequencing steps with the original sequence they derive from (Mercier et al. 2013). The second parameter is the similarity threshold (clustering threshold) used to build MOTUS (Clare et al. 2016, Calderón-Sanou et al. 2020, Wei et al. 2021). Choosing this threshold is delicate without prior knowledge on the maker and its intrinsic level of diversity. A too low threshold can collapse different taxa into the same MOTU (over-merging), while a too high threshold can create too many MOTUs (over-splitting), compared to the actual diversity levels (Clare et al. 2016, Roy et al. 2019, Schloss 2021).

Some works suggest that the ecological interpretation of metabarcoding data can be relatively robust to the threshold selected for sequence clustering. For instance, Botnen et al. (2018) used thresholds ranging from 0.87 to 0.99 of sequence similarity to analyse multiple microbial communities, and obtained community structures highly coherent across thresholds. Nevertheless, levels of alpha diversity can be heavily impacted by the threshold selection. Ideally, the threshold used for clustering would depend on a trade-off between MOTU over-splitting and MOTU over-merging. A growing number of markers are currently being used in metabarcoding studies (Taberlet et al. 2018), with some allowing broad-scale biodiversity assessment but having limited taxonomic resolution (e.g. 18S rDNA primers amplifying all eukaryotes; Guardiola et al. 2015) and others being highly specific to one single class or even family (e.g. Baamrane et al. 2012, Ficetola et al. 2021). Biodiversity surveys generally aim to generate a set of MOTUs that are each associated with a unique taxon, and with all taxa situated at the same level in the taxonomic tree, to facilitate comparisons. In

these conditions, optimal clustering thresholds probably strongly differ across markers. One can for example expect high similarity thresholds for highly conserved markers, and lower clustering thresholds for markers showing high intraspecific variability (Kunin et al. 2010, Brown et al. 2015). However, there is limited quantitative assessment of how optimal clustering thresholds vary across markers (but see Alberdi et al. 2018).

In this study, we analysed sequences from a public database (EMBL) to identify clustering thresholds for different markers and under different criteria. We considered eight metabarcoding markers (Table 1), ranging from generalist ones (e.g. a 16S rDNA-based marker targeting all Bacteria and a 18S rDNA-based marker targeting all Eukaryota) to more specific markers (e.g. markers specific of earthworms, insects or springtails). We evaluated how clustering thresholds can change for each taxonomic group, depending on the criterion adopted to set the threshold. We used two alternative strategies to identify thresholds, each time with different objectives in mind. First, following a procedure similar to the one adopted in barcoding studies (Meyer and Paulay 2005), we compared the distribution probabilities of sequence similarities among different species of the same genus, and among different individuals of the same species to identify thresholds: *i*) minimizing the risk that different sequences of the same species are split in different MOTUs (i.e. risk of over-splitting); *ii*) minimizing the risk that distinct but related species are clustered in the same MOTU (i.e. risk of over-merging); *iii*) balancing the risk of over-splitting and over-merging (Figure 1A). Second, we calculated the over-splitting and over-merging rates of the studied markers for a range of clustering thresholds, to identify values that minimize the two error rates (Figure 1B). We expect that, if researchers want to minimize over-splitting, they should select lower clustering thresholds than if they want to minimize over-merging. Furthermore, we expect higher clustering threshold values for generalist markers compared to markers targeting one class or more restricted taxonomic groups, because of the lower taxonomic resolution and slower evolutionary rate of the former.

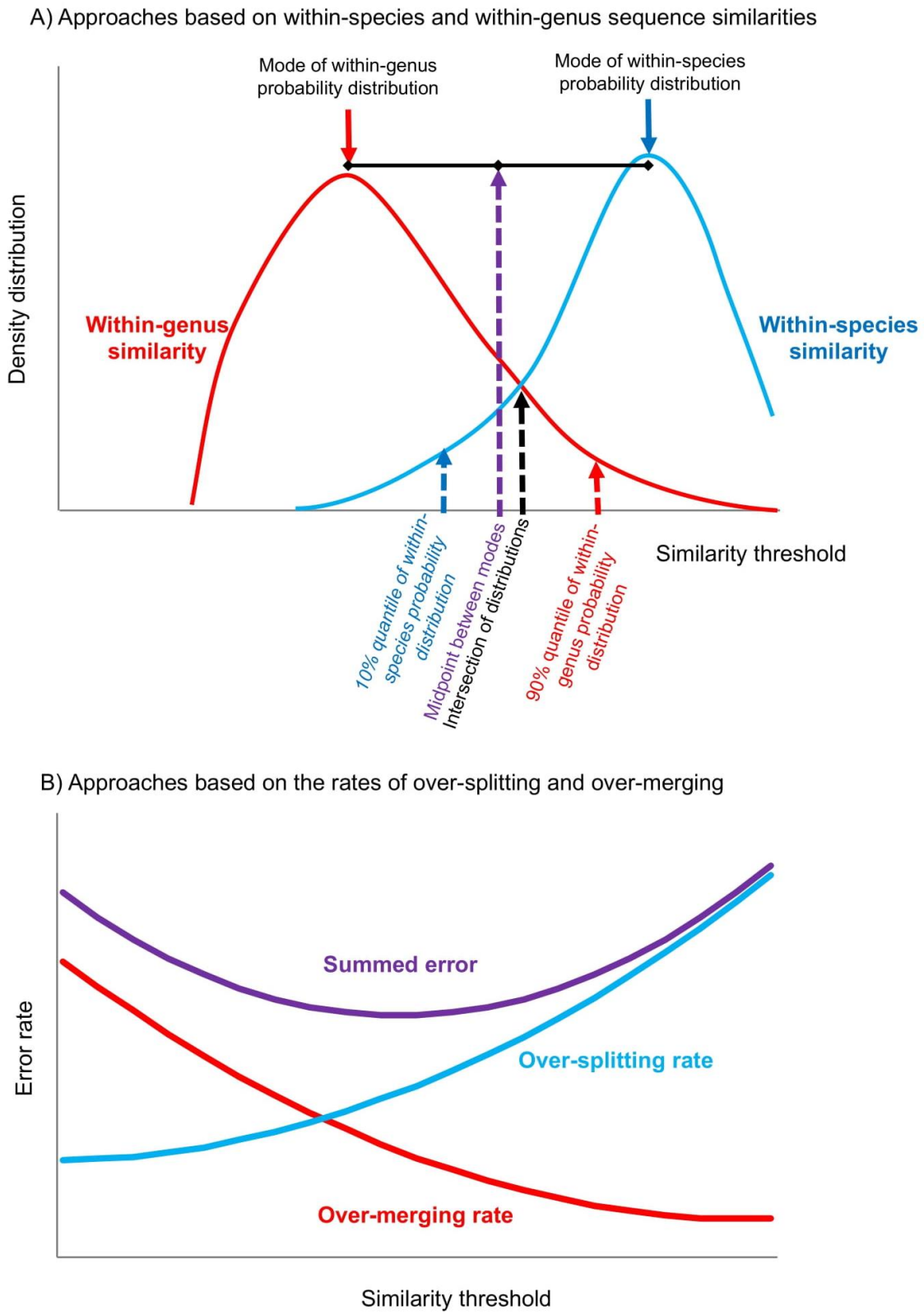


Figure 1. Different approaches to identify the most appropriate clustering thresholds. A): approaches based on similarities between sequences belonging to different individuals from the same species (blue curve), and similarities between sequences belonging to different species from the same genus (red curve). One can choose to minimize the risk that different sequences from the same species are split in different MOTUs (over-splitting risk; e.g. 10% quantile of the distribution of within-species similarities), the risk that sequences from different species belonging to the same genus are

clustered in the same MOTU (over-merging risk; e.g. 90% quantile of within-genus similarities), or one can try to find a balance between the risks of over-splitting and over-merging (e.g. with the intersection between probability distributions, or the midpoint between the modes of both distributions). B) Approaches based on rates of over-splitting and over-merging. One can compare the over-splitting (blue) and the over-merging (red) rates, and/or one can identify the thresholds minimizing the sum of these rates (violet).

METHODS

Markers examined and construction of sequence datasets

We focused on a set of eight DNA metabarcoding markers (Bact02, Euka02, Fung02, Sper01, Arth02, Coll01, Inse01, Olig01) targeting a wide range of organisms (Table 1). Four of these markers can be considered as generalist, i.e. targeting entire superkingdoms or kingdoms: Bact02 targeting Bacteria; Euka02 targeting Eukaryota; Fung02 targeting Fungi; Sper01 targeting Spermatophyta (vascular plants). One marker was intermediate (Arth02; targeting arthropods, i.e. the most species-rich phylum on Earth). Finally, three were more specific, i.e. targeting groups from subclasses to classes: Coll01 targeting Collembola (springtails); Inse01 targeting Insecta; Olig01 targeting Oligochaeta (earthworms).

For each of these markers, a sequence database was built from EMBL release 140 as follows. An *in silico* PCR was first carried out by running the program *ecoPCR* (Ficetola et al. 2010) using the corresponding primers (Table S1). Three mismatches per primer were allowed (-e option), and the amplified amplicon length without primers was restricted (-l and -L options) to the expected length interval (Table S1). The amplified sequences were further filtered by keeping only those belonging to the target taxonomic group, showing a taxonomic assignment (i.e. taxid) at the species and genus levels and having no ambiguous nucleotides. This allowed assembling a working dataset, from which we extracted two sub-datasets. The "within-species" dataset was built by keeping only species for which at least two sequences (identical or not) were available; if >2 sequences were available for a given species, we randomly selected two sequences for that species. The "within-genus" dataset was

built by keeping only genera for which at least two sequences were available; if >2 sequences were available for a given genus, we randomly selected two sequences for that genus. For some markers (Bact02, Euka02, Fung02, Inse01, Sper01), the within-species dataset and sometimes the within-genus dataset still contained a very large number of sequences (>10,000). To limit computation time for these markers, we randomly selected a subset of 5000 different taxa, to reach a final number of sequences equal to 10,000. Table S2 summarizes the number of sequences in the different datasets.

Table 1. Target groups and taxonomic resolution of the eight studied markers.

Marker	Target group	Taxonomic level	Taxonomic resolution *				Reference(s)
			Species level	Genus level	Family level	Order level	
Bact02	Bacteria	Superkingdom	19.6%	55.7%	55.1%	60.2%	Taberlet et al. (2018)
Euka02	Eukaryota	Superkingdom	47.0%	59.5%	68.3%	67.1%	Guardiola et al. (2015)
Fung02	Fungi	Kingdom	72.5%	90.2%	87.7%	85.5%	Epp et al. (2012), Taberlet et al. (2018)
Sper01	Spermatophyta	Clade < kingdom	21.5%	36.9%	77.4%	89.6%	Taberlet et al. (2007)
Arth02	Arthropoda	Phylum	68.6%	89.6%	97.5%	100.0%	Taberlet et al. (2018)
Coll01	Collembola	Class	80.5%	87.2%	75.0%	NA	Janssen et al. (2018)
Inse01	Insecta	Class	87.8%	96.8%	95.4%	79.3%	Taberlet et al. (2018)
Olig01	Oligochaeta	Subclass	89.3%	95.7%	100.0%	100.0%	Bienert et al. (2012), Taberlet et al. (2018)

* Estimated as the percentage of discriminated taxa among amplified taxa; reported from Taberlet et al. (2018).

Calculation of sequence similarities and probability distributions

As a measure of sequence similarity, we computed the pairwise LCS (Longest Common Subsequence) scores between pairs of sequences in the within-species and the within-genus datasets using the *sumatra* program (Mercier et al. 2013). Methodological comparisons showed that this algorithm provides an excellent balance between performance and computation efficiency (Jackson et al. 2016, Kopylova et al. 2016, Bhat et al. 2019). The similarity scores resulting from the within-species dataset were then filtered in R (R Core Team 2020) to keep only those representing similarities

between sequences of the same species, while the scores resulting from the within-genus dataset were filtered to keep only those representing similarities between different species of the same genus.

Approaches to identify clustering thresholds on the basis of within-species and within-genus sequence similarities

We first assessed within-species and within-genus sequence similarities to evaluate four different strategies to determine the corresponding appropriate clustering threshold (Figure 1A) that: *i*) avoid over-splitting; *ii*) avoid over-merging; *iii*) find a balance between over-splitting and over-merging, with two distinct procedures based on the intersection (*iii*-a) or on modes (*iii*-b) of the density probability distributions. These strategies are analogous to those adopted in traditional barcoding studies to set the limit between intra-specific and inter-specific diversity (Meyer and Paulay 2005).

i) **Avoid over-splitting**

In this case, the aim is to avoid distributing different sequences belonging to the same species in different clusters (i.e. limiting the probability of generating additional spurious MOTUs). For this approach, we selected as clustering threshold the 10% quantile of the distribution of similarities between sequences from the same species (within-species dataset). With this approach, sequences attributed to the same species in EMBL are gathered in the same cluster in 90% of the cases.

ii) **Avoid over-merging**

In this case, the aim is to avoid gathering sequences attributed to different species of the same genus in the same cluster (i.e. limiting the probability of merging related species in the same MOTU). For this approach, we selected as clustering threshold the 90% quantile of the distribution of similarities between different species belonging to the same genus. With this approach, sequences attributed to different species belonging to the same genus are assigned to different clusters in 90% of the cases.

iii) **Find a balance between over-splitting and over-merging**

In this case, the aim was to minimize both over-splitting and over-merging. We considered two distinct approaches. First, we obtained the probability distribution of within-species and within-genus sequence pairwise similarities using the *density* function from R, with biased cross-validation (bw="bcv") as smoothing bandwidth selector and a Gaussian smoothing kernel (kernel="gaussian"; Venables and Ripley 2002). Other possible smoothing bandwidth selectors were tested, but biased cross-validation was the approach best fitting the score histograms for all markers and all datasets (data not shown). The balance threshold *iii-a* was then identified as the intersection between the probability distributions of the within-species and within-genus similarities. As an alternative approach to balance over-merging and over-splitting (*iii-b*), we calculated the midpoint between the modes of the within-species and within-genus probability distributions.

Rates of over-merging and over-splitting

For each marker, over-merging and over-splitting rates were evaluated at different clustering thresholds using the within-species dataset described in the paragraph "Markers examined and construction of sequences datasets". This dataset contains two sequences at random, identical or not, for a number of species belonging to the taxonomic group of interest.

For each within-species dataset, clustering was performed using the *sumaclust* program (Mercier et al. 2013) with the *-n* option (normalization by alignment length) based on the sequence similarities first calculated using the *sumatra* program (see above; Mercier et al. 2013). Threshold values (*-t* option) ranging from 0.90 to 1 at 0.01 steps were tested for all markers except Coll01 and Olig01 for which wider ranges ([0.70 – 1] and [0.80 – 1], respectively) were selected based on the within-genus and within-species sequence similarity probability distributions determined previously (see Figure 2). Clustered datasets were then explored to calculate five different variables at each clustering threshold: 1) the number of clusters; 2) the percentage of MOTUs containing one single species; 3) the percentage of MOTUs containing one single genus; 4) the percentage of species

gathered in one single MOTU; 5) the percentage of genera gathered in one single MOTU. Variables 2 and 3 are indicative of appropriate MOTU merging of sequences at the species and genus levels, respectively, while variables 4 and 5 are indicative of appropriate MOTU splitting at the species and genus levels, respectively.

These values were also used to calculate three measures of error. We defined the over-merging rate as 1 - the percentage of MOTUs containing one single species; and the over-splitting rate as 1 - the percentage of species gathered in one single MOTU. The summed error rate was then calculated as the sum of the over-merging and over-splitting rates. It should be noted that for this estimate, we assigned the same weight to over-splitting and over-merging.

RESULTS

Our *in-silico* PCRs amplified between 17,000 (Coll01) and 3,2000,000 (Bact02) sequences per marker (Table S2). After data filtering, we retained between 510 (Coll01) and 708,000 (Bact02) sequences per marker. The within-species dataset comprised between 118 (Coll01) and 10,000 (Bact02, Euka02, Fung02 and Sper01) sequences, while the within-genus dataset comprised between 74 (Coll01) and 10,000 (Euka02 and Sper01) sequences per marker.

Clustering thresholds determined from probability distributions of within-species and within-genus sequence similarities

The probability distributions of within-species and within-genus sequence similarities showed very contrasting patterns between the generalist and the specific markers (Figure 2). For the five markers targeting a phylum or broader taxonomic groups (Bact02, Euka02, Fung02, Sper01, and Arth02), the distributions of within-species and within-genus similarities were rather similar, both showing a mode at very high similarity values (Figure 2). Fung02 showed a slightly different pattern, as the within-

genus similarities had a very broad distribution. Conversely, for the more specific markers, the distributions of sequence similarities were very different, with two clearly distinct peaks.

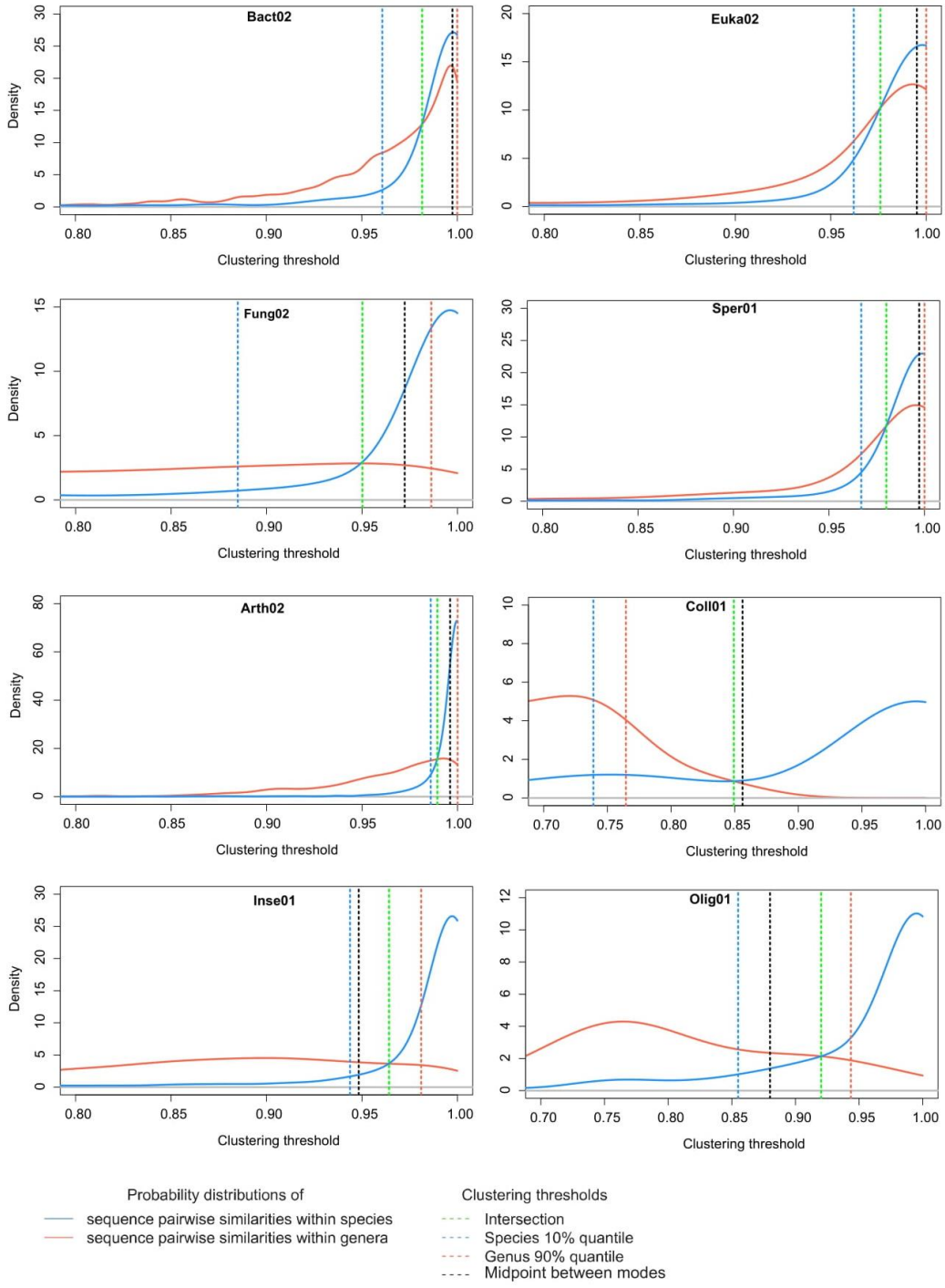


Figure 2. Density probability distributions of sequence pairwise similarities within species (blue lines) and within genera (red lines) for the eight studied markers. For each marker, dotted lines represent the 10% quantile of the within-species probability distribution (blue; threshold limiting over-splitting), the 90% quantile of the within-genus probability distribution (red; threshold limiting over-merging), the intersection of the within-species and within-genus probability distributions (green, balance-a) and the midpoint between modes (black, balance-b).

Within-species similarities remained very high (mostly above 0.95), while within-genus similarities generally showed lower values (mode around 0.90 for Inse01, and below 0.80 for Olig01 and Coll01).

For all markers, criterion i (avoid over-splitting) yielded the lowest thresholds (Figure 3, Table S3), with very low levels for Coll01 and Olig01. Conversely, criterion ii (avoid over-merging) yielded extremely high values. For all generalist markers, avoiding over-merging would require setting clustering thresholds at 0.99 or higher. Unexpectedly, for Coll01, criterion ii resulted in a rather low threshold (0.765), because many within-genus comparisons showed very low similarity values.

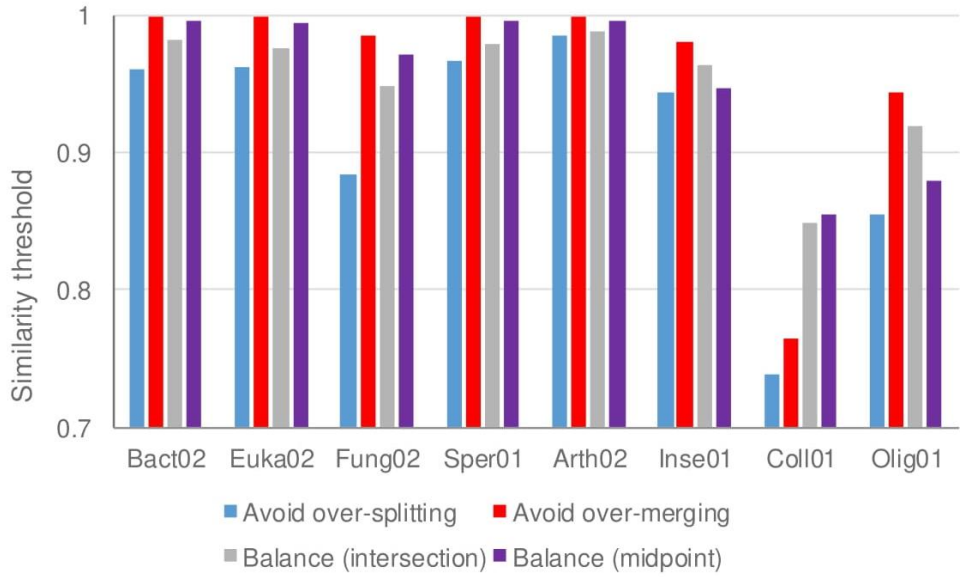


Figure 3. Different possible clustering thresholds for the eight studied markers, depending on the selected criterion.

Criteria *iii-a* and *iii-b* searching a balance between over-merging and over-splitting yielded somehow contrasting results across markers. For the three specific markers (Coll01, Inse01, and Olig01), the within-genus and within-species similarities showed clearly distinct peaks (Figure 2). As a consequence, the intersection between the two curves could effectively represent the point minimizing both over-merging and over-splitting (see discussion), and the midpoint between the modes also identified rather similar threshold values. On the contrary, for the generalist markers, the within-species and within-genus similarities showed very high overlap and similar modes, and the density distributions actually intersected at values lower than both modes. The midpoint between the

modes continued to identify threshold values intermediate between the peaks of within-species and within-genus similarities.

Rates of over-splitting and over-merging

For all the markers, whatever the clustering threshold examined (values ≥ 0.70 for Coll01, ≥ 0.80 for Olig01 and ≥ 0.90 for the other markers), the percentage of MOTUs containing one single species was higher than 50%, and that of MOTUs containing one single genus was higher or close to 70% (Figure 4). Overall, for the generalist and intermediate markers, these two percentages showed a regular increase with the clustering threshold, and for the specific markers, they tended to values close to 100% for high thresholds. Unsurprisingly, the two percentages tended to be lower for the generalist markers than for the specific markers at a given threshold, indicating that the former are more sensitive to over-merging. Fung02 was a notable exception, since about 87% and 97% of MOTUs contained one single species and one single genus, respectively, at the 0.97 threshold, which is a frequently adopted clustering threshold for fungal ITS sequences. These values were comparable to those observed for the specific markers, for which > 85% and > 98% of MOTUs contained one single species or one single genus, respectively, for thresholds ≥ 0.95.

For all markers, the percentages of species and genera gathered in one single MOTU decrease both at a similar rate with the clustering threshold, with generally a sharp drop at high thresholds (≥ 0.98; Figure 4). However, the pattern of MOTU splitting was less characteristic of generalist vs. specific markers. For some markers (Euka02, Sper01, Arth02, Inse01), the percentage of species or genera gathered in a single MOTU remained higher or close to 50% up to high thresholds (0.98). On the contrary, for Bact02, Fung02, Coll01, Olig01, these percentages dropped quickly when the clustering threshold increased, indicating that these markers are susceptible to over-splitting.

For all markers, the number of clusters generally increased regularly with the clustering threshold up to 0.97-0.98 (Figure 4), followed by a sharp rise up to 1 (which was however less obvious for Euka02

and Olig01). For example, for Bact02, the number of clusters more than doubled between 0.97 (2862) and 1 (6461).

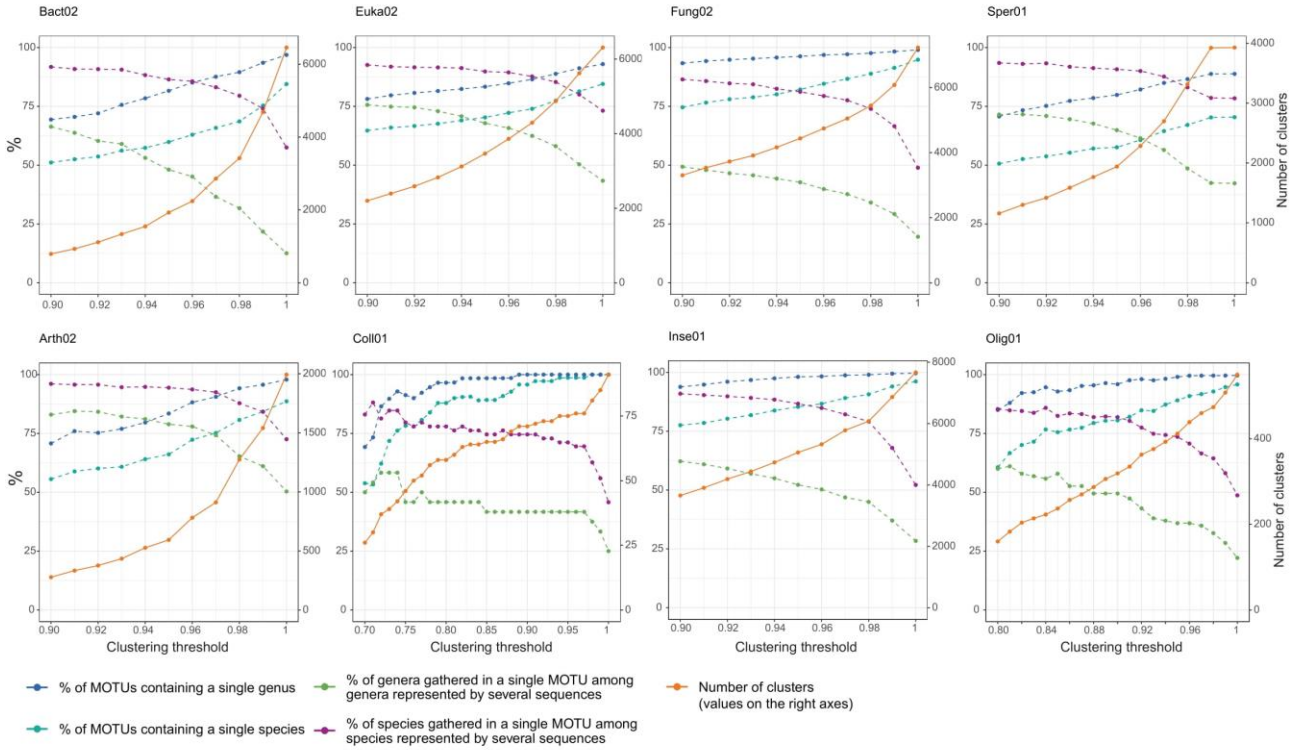


Figure 4. Evolution of over-splitting and over-merging rates for a range of clustering thresholds, for the eight studied markers. The left y-axes report percentage values; the right y-axes indicate the number of obtained clusters.

Our results showed clear patterns for over-merging and over-splitting rates, with over-splitting quickly increasing and over-merging quickly decreasing at high clustering thresholds (Figure 5). For several markers, the summed error showed a relatively clear minimum at specific clustering thresholds (Figure 5): 0.96-0.99 for Bact02 and Euka02, 0.97-0.99 for Arth02, 0.94-0.96 for Inse01, and 0.96-0.98 for Sper01. The minimum was much less evident for Fung02, Coll01 and Oligo01, these markers showing relatively similar summed error rates over a broad range of clustering thresholds (Fung02: 0.91-0.98; Coll01: 0.82-0.96, with multiple minima; Oligo01: 0.84-0.96, with multiple minima).

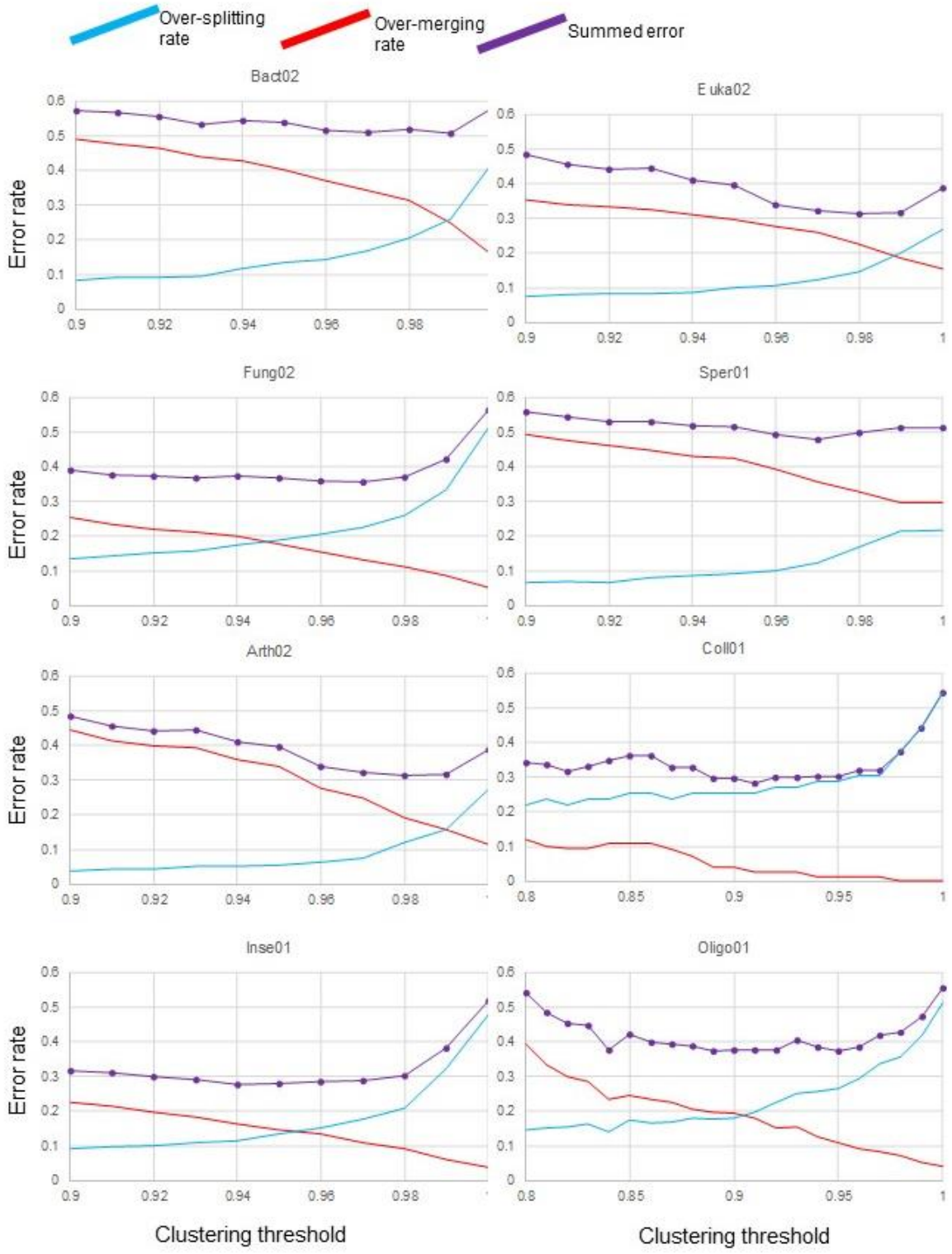


Figure 5. Over-splitting (blue) and over-merging (red) rates, as well as the summed error rate (i.e. over-splitting rate + over-merging rate; violet), for the eight studied markers across a range of clustering thresholds.

DISCUSSION

Sequence clustering approaches are routinely used for the identification of MOTUs in metabarcoding studies, and they often resort to methods based on similarity values. Still, selecting a clustering threshold for a given marker more than often relies on common practices and rules of thumb rather than on proper scientific argument. By analysing extensive sequence data deposited in public databases for a range of generalist and specialist markers, we showed that different threshold values can be selected depending on the marker and on the criterion favoured by researchers. All the markers we examined are situated in non-protein coding genes (Table S1), which has an influence on levels of sequence intraspecific diversity. The 10% quantile of the within-species similarity probability distribution was almost always lower than the 0.97 clustering threshold traditionally used in barcoding for markers targeting protein-coding genes like COI (Hebert et al. 2003), or for microbial MOTU delimitation (Bálint et al. 2016), indicating that some level of over-splitting can occur at this threshold.

Although for all the markers the within-genus similarity values were generally lower than the within-species similarities, the overlap between the two distributions was dependent on the generalist vs. specific nature of the marker. For some specific markers (e.g. Coll01 and Olig01), distinct peaks were visible for the two similarity metrics (Figure 2). Within-species similarities generally were >0.90, while within-genus values were lower, frequently below 0.80. Such a pattern is not unexpected for markers with an excellent taxonomic resolution and designed to identify taxa at the species level. Conversely for the generalist markers, within-species and within-genus similarity probability distributions largely overlapped and the differences between the peaks were minimal. Nevertheless, even for these markers, the density of within-species similarity was consistently higher than that of within-genus similarity at high clustering thresholds, indicating that the probability of observing the corresponding similarity value is higher within species than within genus. In other words, at high clustering thresholds, a MOTU is more likely to represent a species than a genus. This result is

confirmed by the fact that the percentage of MOTUs containing a single species is always higher than 50%, whatever the clustering threshold or the marker considered (Figure 4).

The sequences used as a primary source of information in this study were downloaded from EMBL, and our results are thus highly dependent on the quality of the data deposited in this public database. Even though broad-scale analyses suggest that these data are generally reliable (Leray et al. 2019), errors in the sequence itself (e.g. wrong nucleotide, or more complex errors like insertions, deletions, inversions, duplications or pseudogene sequences) and taxonomic mislabelling can occur in public sequence databases, especially for organisms which are difficult to identify based on the morphology (Bridge et al. 2003, Bidartondo 2008, Valkiūnas et al. 2008, Mioduchowska et al. 2018). While the first type of error will affect within-species sequence similarity negatively, sometimes substantially, the effect of the second type is more diffuse. For example, in a group like springtails where species delimitation is tricky (Porco et al. 2012), the existence of cryptic species will decrease within-species sequence similarity while increasing over-splitting rates. In a group like Bacteria, type strains are sometimes entered at the species level in the NCBI (EMBL) taxonomy (Federhen 2015), leading to an inflation of within-genus similarity and over-merging rates. In every case though, database errors will make within-species and within-genus similarities distributions more difficult to distinguish and clustering thresholds trickier to identify, thus the over-splitting or over-merging rates reported here could be artificially higher than in reality.

In this work, we came up with a global measure of the error associated with a given clustering threshold, that we called the "summed error". We calculate it by summing over-splitting and over-merging rates, assuming both have the same cost for biodiversity studies. However, it is possible to assign a differential weight to over-splitting and over-merging. For instance, if the aim is to reach conservative estimated of alpha diversity (i.e. avoid over-splitting), more weight can be assigned to over-splitting rate. Conversely, if the aim is to tease apart closely related species, that differ in their sensitivity to environmental stressors or in threat levels, one may prefer to avoid over-merging, particularly when extensive reference databases are available (Roy et al. 2019, Lopes et al. 2021).

For most of the markers we examined, the summed error approach provided relatively clear results, and identified a range of threshold values that minimized the summed error. For instance, for Euka02, the summed error was relatively low at thresholds between 0.96 and 0.99 (Figure 5), indicating a good trade-off between over-merging and over-splitting. Interestingly, this range of values was also highlighted by the analysis of probability distributions (Figure 3, Table S3). Indeed, 0.96 is the threshold minimizing over-splitting for Euka02 while 0.99 is the balance (midpoint) threshold. The consistency of values obtained with very different approaches supports the robustness of our conclusions.

However, for a few markers, the threshold values minimizing summed error yielded somewhat less clear patterns. For Fung02, the summed error rate was rather constant (36-37%) at all the thresholds between 0.92 and 0.98, while it quickly increased for higher clustering thresholds. For Coll01 and Oligo01, the summed error rate showed multiple minima, some of which at very low clustering thresholds (Figure 5). In principle, increasing the threshold value should determine a monotone decrease of over-merging, and a monotone increase of over-splitting (Figure 1B). However, at low similarity values this was not always the case (Figure 5). This probably occurs because a very large number of sequences have pairwise similarities of 0.80-0.85 (Figure 2), and this might affect the identification of clusters, with some sequences clustering together e.g. at 0.85 but not at 0.86 similarity values. We also note that these similarity values match the ones corresponding to the intersection between the within-genus and within-species similarities for these markers (Figure 3). It is also possible that, at this level of sequence similarity, there is strong uncertainty between MOTUs representing different hierarchical levels of taxonomy.

Our results provide quantitative data that can help researchers set their optimal clustering thresholds, and understand the consequences of choosing low or high clustering thresholds. If a clear minimum exists for the summed error rate, it probably represents an excellent trade-off between over-merging and over-splitting. In this sense, a threshold value ranging from 0.96 to 0.99 is probably appropriate for both Bact02 and Euka02, while Arth02 should accommodate a slightly higher range

(0.98-0.99) and a fixed threshold of 0.97 seems to be more suitable for Sper01. For Inse01, lower threshold values (0.94-0.96) are more judicious. All these values match with those obtained on the basis of within-species and within-genus similarities (Figure 3). However, for Coll01, Oligo01 and Fung02, the summed error rate does not provide clear indications, and within-species and within-genus similarity distributions (e.g. midpoint between modes) might be more informative to set the threshold value (Figures 2 and 3).

The selection of clustering thresholds can have strong effect in the estimates of MOTUs richness (Figure 4), still it is important to remember that it often does not have a tremendous effect on the ecological message conveyed by metabarcoding data. For instance, Clare et al. (2016) examined different clustering thresholds to analyse dietary overlap between skinks and shrews in Mauritius. Although high clustering thresholds yielded a larger number of MOTUs, ecological conclusions remained rather consistent overall. Therefore, provided that appropriate parameters are considered (e.g. alpha diversity measured using Hill's numbers with $q > 0$ instead of richness, beta diversity estimates), the interpretation of data can be relatively robust (Clare et al. 2016, Roy et al. 2019, Calderón-Sanou et al. 2020, Mächler et al. 2021). Nevertheless, we discourage the blind application of one single clustering threshold like the classical 0.97, as it can have very different meaning across markers, and can inflate MOTU richness for fast-evolving markers. Instead, we advocate the ad-hoc definition of the most appropriate thresholds, on the basis of research aims, on the cost of over-splitting and over-merging, and on the features of the studied markers.

Acknowledgments

This study was supported by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities).

Supplementary online material for:

Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies

Aurélie Bonin, Alessia Guerrieri, G. Francesco Ficetola

Table S1. Characteristics of the DNA metabarcoding markers examined in this study.

Marker	Target group	Target gene	Forward primer	Reverse primer	Minimum amplicon length (bp*)	Maximum amplicon length (bp*)
Bact02	Bacteria	V4 region of the 16S rDNA gene	GCCAGCMGCCGCGGTAA	GGACTACCMGGGTATCTAA	45	748
Euka02	Eukaryota	V7 region of the 18S rDNA gene	TTTGTCTGSTTAATTSCG	CACAGACCTGTTATTGC	36	892
Fung02	Fungi	ITS1 nuclear rDNA gene	GGAAGTAAAAGTCGTAACAAGG	CAAGAGATCCGTTGYTGAAAGTK	68	919
Sper01	Spermatophyta	P6 loop of the intron of the chloroplastic trnL gene	GGGCAATCCTGAGCCAA	CCATTGAGTCTCTGCACCTATC	10	220
Arth02	Arthropoda	16S mitochondrial rDNA gene	GATAGAAACCRACCTGGYT	AARTTACYTTAGGGATAACAG	76	168
Coll01	Collembola	16S mitochondrial rDNA gene	ACGCTGTTATCCCTWAGG	GACGATAAGACCCTWTAGA	76	192
Inse01	Insecta	16S mitochondrial rDNA gene	RGACGAGAAGACCCTATARA	ACGCTGTTATCCCTAARGTA	75	265
Olig01	Oligochaeta	16S mitochondrial rDNA gene	CAAGAAGACCCTATAGAGCTT	CCTGTTATCCCTAAGGTARCT	93	196

*Excluding primers.

Table S2. Characteristics of the sequence datasets used in this study.

Marker	*N* sequences amplified *in silico*	*N* sequences after filtration in the working dataset	*N* sequences in the within-species dataset	*N* sequences in the within-genus dataset
Bact02	3,202,507	707,874	10,000	5,464
Euka02	247,843	135,162	10,000	10,000
Fung02	209,223	64,615	10,000	3,164
Sper01	148,819	129,885	10,000	10,000
Arth02	101,955	19,727	4,036	3,014
Coll01	17,4635	510	118	74
Inse01	356,740	58,997	10,000	9,398
Olig01	213,741	4,884	764	258

Table S3. Values of the different thresholds estimated for the eight studied markers on the basis of within-species and within-genus sequence similarities.

Target	**Bact02**	**Euka02**	**Fung02**	**Sper01**	**Arth02**	**Coll01**	**Inse01**	**Olig01**
Criterion *i*: Avoid over-splitting (10% quantile of within-species probability distribution)	0.961	0.962	0.885	0.967	0.986	0.739	0.944	0.855
Criterion *ii*: Avoid over-merging (90% quantile of within-genus probability distribution)	1.000	1.000	0.986	1.000	1.000	0.765	0.981	0.944
Criterion *iii*-a: Balance-a (intersection of within-species and within-genus probability distributions)	0.982	0.976	0.949	0.980	0.989	0.849	0.964	0.920
Criterion *iii*-b: Balance-b (midpoint between modes)	0.997	0.995	0.972	0.997	0.996	0.856	0.948	0.880

CHAPTER 4. ARTICLE C[4]

Metabarcoding data reveal vertical multi-taxa variation in topsoil communities during the colonization of deglaciated forelands

Running title: multi-taxa variation of topsoil communities

Alessia Guerrieri[1], Alexis Carteron[1,*], Aurélie Bonin[1,2], Silvio Marta[1], Roberto Ambrosini[1], Marco Caccianiga[1], Isabel Cantera[1], Chiara Compostella[3], Guglielmina Diolaiuti[1], Diego Fontaneto[4], Ludovic Gielly[5], Fabrizio Gili[1], Mauro Gobbi[6], Jerome Poulenard[7], Pierre Taberlet[5,8], Andrea Zerboni[3], Wilfried Thuiller[5], Gentile Francesco Ficetola[1,5]

[1] Department of Environmental Science and Policy, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy

[2] Argaly, Bâtiment CleanSpace, 354 Voie Magellan, 73800 Sainte-Hélène-du-Lac, France

[3] Department of Earth Sciences "A. Desio", Università degli Studi di Milano, Via L. Mangiagalli 34, 20133 Milano, Italy

[4] CNR - Water Research Institute, Largo V. Tonolli 50, 28922, Verbania, Italy

[5] Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000, Grenoble, France

[6] Research & Museum Collections Office, Climate and Ecology Unit, MUSE-Science Museum, Corso del lavoro e della Scienza, 3, 38122 Trento, Italy

[7] Univ. Savoie Mont Blanc, CNRS, EDYTEM, F-73000 Chambéry

[8] UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

* Correspondence author. Email: alexis.carteron@unimi.it

[4] Submitted to Molecular Ecology (Dec-2021)

ABSTRACT

Ice-free areas are increasing worldwide due to the dramatic glacier shrinkage and are undergoing rapid colonization by multiple lifeforms, thus representing key environments to study ecosystem development. Soils have a complex vertical structure. However, we know little about how microbial and animal communities differ across soil depths and development stages during the colonization of deglaciated terrains, how these differences evolve through time, and whether patterns are consistent among different taxonomic groups. Here, we used environmental DNA metabarcoding to describe how community diversity and composition of six groups (Eukaryota, Bacteria, Mycota, Collembola, Insecta, Oligochaeta) differ between surface (0-5 cm) and relatively deep (7.5-20 cm) soils at different stages of development across five Alpine glaciers. Taxonomic diversity increased with time since glacier retreat and with soil evolution; the pattern was consistent across different groups and soil depths. For Eukaryota, and particularly Mycota, alpha-diversity was generally the highest in soils close to the surface. Time since glacier retreat was a more important driver of community composition compared to soil depth; for nearly all the taxa, differences in community composition between surface and deep soils decreased with time since glacier retreat, suggesting that the development of soil and/or of vegetation tends to homogenize the first 20 cm of soil through time. Within both Bacteria and Mycota, several molecular operational taxonomic units 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.

KEYWORDS: environmental DNA, glacier retreat, Hill's number, beta-diversity, soil depth, springtails, earthworms, insects, fungi

INTRODUCTION

The worldwide shrinkage of glaciers is causing a fast increase in ice-free areas on all continents, thus providing more and more potential habitats for multiple organisms (Ficetola et al., 2021). After ice loss, organisms with high dispersal abilities can colonize the newly exposed terrains relatively quickly (Gobbi et al., 2017; Hågvar et al., 2020; Kaufmann, 2001; Rosero et al., 2021). Both micro- and macro-organisms (e.g. bacteria, fungi and soil fauna) influence soil development being involved in many biogeochemical processes such as soil nutrient cycling (Bardgett 2005; Bardgett & Van Der Putten, 2014), and can interact with each other in determining ecosystem functioning (Ingham, Trofymow, Ingham, & Coleman, 1985). Assessing community variation in this biodiversity compartment and across multiple glacier forelands is important to understand how these ecosystems develop after the retreat of glaciers, and is a key topic of global change biology (Cauvy-Fraunié & Dangles, 2019).

To date, most of our knowledge about soil ecology focuses on the top 10 cm of soil and on microbial communities (Bahram et al., 2015). However, soil characteristics vary vertically (Khokon, Schneider, Daniel, & Polle, 2021). In particular, physical features (e.g., pH, soil moisture, micro-climatic characteristics) typically change through space, and the availability of nutrients (e.g., organic carbon, total nitrogen), together with the enzyme activities of the associated microorganisms, decrease from the topsoil to deeper soil layers (Herold et al., 2014; Moradi et al., 2020). Such variation of habitat conditions can strongly influence the community structure of inhabiting taxa (Carteron, Beigas, Joly, Turner, & Laliberté, 2021; Franzetti et al., 2020; Orwin, Kirschbaum, John, & Dickie, 2011; Rime et al., 2015) because different organisms can be associated with different soil conditions (Khokon, Schneider, Daniel, & Polle, 2021; Mundra et al., 2021). Several studies showed differences in microbial communities across soil depths in several terrestrial habitats including grasslands, forests, high elevation, post-mining and reforested-soils, and agreed that depth significantly affects the abundance, composition and diversity of bacteria and fungi, with the richest communities often

associated to surface layers (e.g. Zhao, Zheng, Zhang, Gao, & Fan, 2021; Carteron et al., 2021; Chu et al., 2016; Moradi et al., 2020; Chen, Jiao, Li, & Du, 2020). These differences can be attributable to the decrease in nutrients content at increasing depths (Chu et al., 2016), or to differences in microclimatic conditions and water availability. However, the spatial structuring and micro-habitats conditions of soil communities are yet poorly known, and most studies only focused on very few limited taxonomic groups (Doblas-Miranda, Sánchez-Piñero, & González-Megías, 2009; Moradi et al., 2020; Sadaka & Ponge, 2003), making it difficult to compare the responses of functionally different taxa.

During soil formation after the retreat of glaciers, many features of the substrate change through time, with modifications of physical properties and nutrients content, and a progressive vertical stratification of developed soils (Schaetzl & Anderson, 2005; Mavris, Egli, Plotze, Blum, Mirabella and Giaccai, 2010; Khedim et al., 2021; Wietrzyk-Pełka, Rola, Szymański, & Węgrzyn, 2020). Despite many studies investigating the biotic colonization after glacier retreat, the majority of them focused on organisms living above or just below the surface (reviewed in Ficetola et al. 2021), while limited information is available about the vertical distribution of different topsoil organisms across stages of soil development. Assessing the vertical as well as the horizontal composition and distribution of topsoil colonizers along the glacier forelands is pivotal to infer the key ecological processes under the primary succession that occurs since the early years after glacier retreat. Rime et al. (2015) performed a rare attempt of integrating soil depth into the study of Alpine primary successions (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, while decreased at older soil development stages, with a homogenization through time. However, Rime et al. (2015) only focused 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 processing governing community development after glacier retreat.

Approaches based on the metabarcoding of environmental DNA (eDNA; Taberlet, Bonin, Zinger, & Coissac, 2018) help overcoming several limitations of conventional sampling and are increasingly used because of their relatively fast and cost-efficient data production. Environmental DNA metabarcoding allows the monitoring of communities of micro- and macro-organisms in a wide range of natural systems (Bohmann et al., 2014). With appropriate technical precautions (Guerrieri et al., 2020), soil communities can be sampled and studied via metabarcoding over broad geographic scales and from remote areas (e.g. Zinger, Taberlet, et al., 2019), and data can be related to environmental characteristics in order to infer ecological processes. The combination of multiple metabarcodes makes eDNA particularly powerful tool for estimating the multi-taxa soil diversity (Donald et al., 2021). Here, we used metabarcoding data from soil eDNA in order to study the vertical distribution of microbes and animals within the top 20 cm of soil, where most microbial diversity has been retrieved (Fierer, Schimel, & Holden, 2003) and where most soil invertebrates spend their life cycle (Menta, 2012). First, we tested whether and how the overall taxonomic diversity of multiple taxa changes with soil depth and time since glacier retreat. We expected that the alpha-diversity of communities increases through time and decreases with depth, especially in the youngest soils. Moreover, we tested whether the changes in alpha-diversity through time are consistent between surface and deep layers. Second, we evaluated the differences in community composition between different depths and tested for potential taxa characteristic of the different depths or stages of soil development. Rime et al (2015) observed that differences between surface and deep soil decrease at older soil development stages, with a homogenization of communities through time, but these conclusions were only based on microorganisms from one single glacier foreland. We analysed the beta-diversity between surface and deep layers for six taxonomic groups representing a large proportion of biodiversity. If the 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.

MATERIALS AND METHODS

Samples collection and preservation

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

At each site, we established five regularly spaced plots at distances of about 20 m. At each plot, we collected five soil cores within one-meter distance and we kept the 0-5 cm and 7.5-20 cm portions to be representative of two different soil depths, hereafter called "surface" and "deep" soils, respectively (Fig. 1). For each of the five cores, we pooled portions of the same depth together to form one composite sample of ~200 g and we homogenized it. We took 15 g of soil from each composite sample and desiccated it immediately in sterile boxes with 40 g of silica gel. Previous analyses showed that this approach enables a cost-effective and long-term preservation of soil eDNA (Guerrieri et al., 2021). Soil eDNA 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.

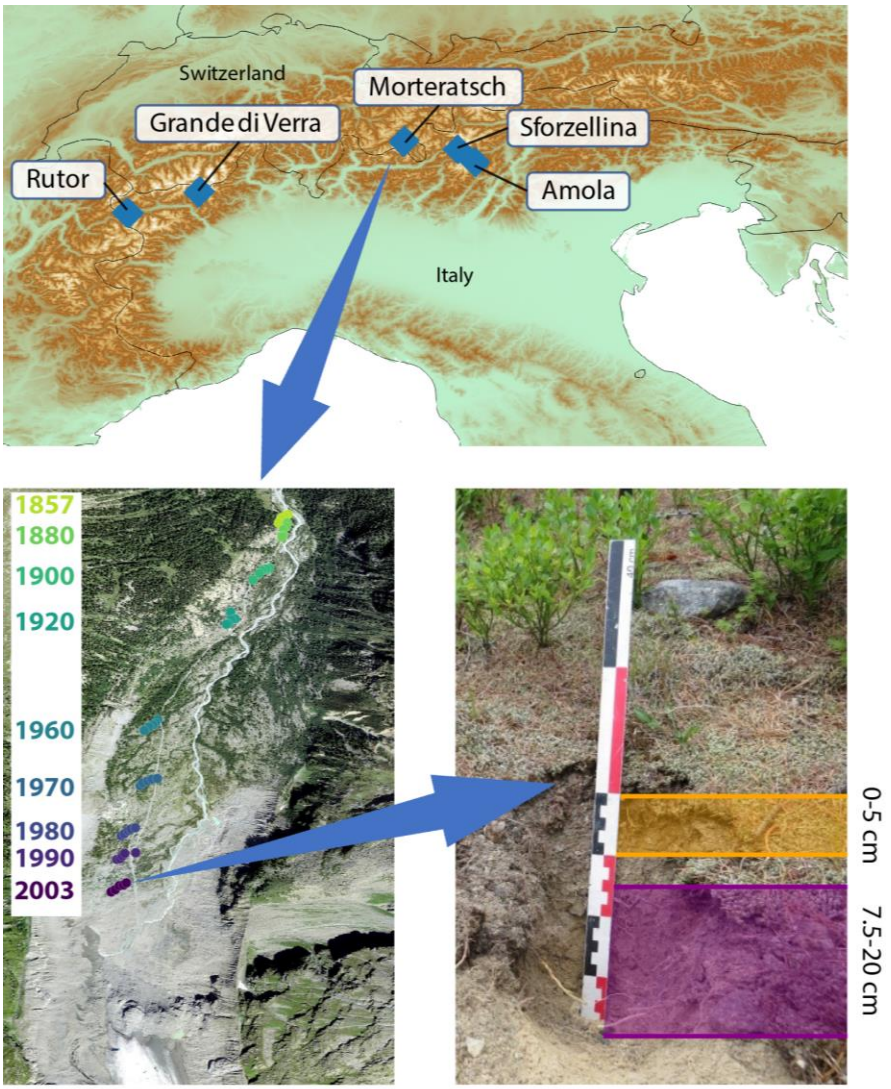


Figure 1. Study area and sampling design. Maps are pseudo-colour representation of altitude (source: 30 arc-sec digital elevation model) and land cover (Source: Copernicus Sentinel data 2019) created by AG, SM and GFF.

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 networks involving multiple glacier forelands (e.g., Khedim et al., 2021; Orgiazzi, Ballabio, Panagos, Jones, & Fernández-Ugalde, 2018; Rime et al., 2015; Schweizer, Hoeschen, Schlüter, Kögel-Knabner, & Mueller, 2018) because soil horizons are not yet differentiated in early stages of soil development, and because this approach allows obtaining 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 soil depths, regardless of the horizons.

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, Coissac, Pompanon, Brochmann, & Willerslev (2012); then we extracted eDNA using the NucleoSpin® Soil Mini Kit (Macherey-Nagel, Germany) with a final elution in 150 µl for both soil samples and with one negative extraction control every 23 samples (total: 12).

We amplified eDNA of bacteria, eukaryotes, fungi, springtails, insects and earthworms using primers designed for markers Bact02 (Bacteria and Archaea: Taberlet et al., 2018), Euka02 (Eukaryota: Guardiola et al., 2015), Fung02 (Mycota: Epp et al., 2012; Taberlet et al., 2018), Coll01 (Collembola, i.e. springtails: Janssen et al., 2018), Inse01 (Insecta: Taberlet et al., 2018), and Olig01 (Oligochaeta, i.e. earthworms: Bienert et al., 2012; Taberlet et al., 2018). We selected this set of markers to cover a wide range of organisms at different taxonomic resolution as we included three generalist markers (targeting entire superkingdoms or kingdoms: Bact02, Euka02 and Fung02) and three more specific markers (targeting from classes to subclasses: Coll01, Inse01, Olig01). All these markers are well suited for metabarcoding analyses thanks to the low number of mismatches in the priming regions across target organisms, and they perform well with potentially degraded DNA due to the relatively short length of amplified fragments (Taberlet et al., 2018; Table S1). We used forward and reverse primers tagged on the 5'-end with eight-nucleotide long tags with at least five nucleotide differences among them (Coissac, 2012) and combined them in a way that all PCR replicates were represented by a unique combination of forward and reverse 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 constituted 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, USA; 1:10 diluted) and we used it to check for potential cross-contaminations and to monitor amplification and sequencing performances.

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™, USA), 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™, Foster City, CA, USA), 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, Basel, Switzerland) 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 PCR profiles: an initial step of 10 min at 95°C; several cycles of a 30 s denaturation at 95°C; a 30 s annealing at 53°C (Bact02), 45°C (Euka02), 56°C (Fung02), 51°C (Coll01), 55°C (Olig01) or 52°C (Inse01); a 90 s elongation for Bact02 and Fung02, or a 60 s elongation for all the others markers at 72°C; a final elongation at 72°C for 7 min. After amplification, we pooled together all amplicons of the same marker and visualized a 5-µl aliquot by high-resolution capillary electrophoresis (QIAxcel Advanced System, QIAGEN, GERMANY) in order to check fragments length and monitor primer dimers. Finally, for each marker, we purified six subsamples of the pooled amplicons separately, using the MinElute PCR Purification Kit (QIAGEN, GERMANY) as per the manufacturer's instructions and combined them again. 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, San Diego, CA, USA) with a paired-end approach (2 × 250 bp for Bact02 and Fung02, and 2 × 150 bp for the others markers) at Fasteris (SA, Geneva, Switzerland).

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 *illuminapairedend* program and kept only sequences with an alignment score greater than 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 *ngsfilter* and allowed two and zero mismatches on primers and tags, respectively. Third, we dereplicated sequences using *obiuniq* and discarded bad-quality sequences (i.e., containing "N"), sequences whose length was lower or higher than expected (based on the minimum and maximum metabarcode length; Table S1) and singletons. Fourth, we ran the *obiclean* program with the option -r set at 0.5 to detect potential PCR or sequencing errors and kept 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/wikis/home). These thresholds minimize the risk that sequences attributed to the same species are clustered in different MOTUs and were selected on the basis of preliminary bioinformatics analyses (Bonin, Guerrieri, & Ficetola, 2021).

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.

In order to remove spurious sequences and avoid bias in ecological conclusions (Calderón-Sanou, Münkemüller, Boyer, Zinger, & Thuiller, 2020) we performed additional filtering in R (version 4.0). We discarded MOTUs with a best identity < 80% (Bact02, Euka02, Fung02) or < 60% (Coll02, Inse01, Olig01) and MOTUs observed less than five (Bact02, Fung02, Inse01), ten (Olig01), eleven (Coll01) or twelve (Euka02) times overall. The latter corresponds to the minimum number of reads that removed ≥ 99.99% of sequences detected in our blanks (i.e., tag-jump errors). Furthermore, we discarded MOTUs detected in only one sample, as they represent singletons, MOTUs detected in less 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).

Statistical analyses

At each plot and for each depth, we measured alpha-diversity of Bacteria, Eukaryota, Mycota, Collembola, Insecta and Oligochaeta, through Hill numbers. The joint use of different Hill numbers allows to obtain biodiversity measures in metabarcoding studies that are robust to bioinformatic treatments and other methodological choices (Alberdi & Gilbert, 2019; Calderón-Sanou et al., 2020; Mächler, Walser, & Altermatt, 2021). We used the parameters q = 0 and q = 1 in the *hill_taxa* function of the *hillR* package (Chao, Chiu, & Jost, 2014). Values of q = 0 returns the taxonomic richness and are insensitive to MOTUs 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 variation in alpha-diversity of Bacteria, Eukaryota, Mycota, Collembola, Insecta and Oligochaeta with time since glacier retreat and depth. We ran mixed models with the alpha-diversity of each sample (log-transformed) as dependent variable, and used a Gaussian error to attain normality of model residuals, 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. These models also included interactions 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, Hwang, & Vehtari, 2013).

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), potentially influencing alpha diversity of soil communities (Guo et al., 2018). We therefore re-analysed the pattern of alpha diversity using organic carbon as an independent variable, instead of age since glacier retreat. This analysis was run for a subset of samples (N = 276) for which data of total organic carbon content were obtained by elemental analysis (Organic Elemental Analyzer, model Flash 2000, Thermo Fisher; Khedim et al., 2021; Lacchini 2020). Organic carbon was strongly related to age since glacier retreat (GLMM with organic carbon as dependent variable and age as independent variable: $R^2 = 0.66$) thus it was impossible to include organic carbon and age together in the same model. Organic carbon data were representative of the overall soil core (0-20 cm); thus, the analysis did not allow testing the role of variation in carbon content between surface and deep layers. Two plots (i.e. four samples in total) 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 *beta.multi* function of the *betapart* package with the Sorensen 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 of nestedness and turnover when one of the compared communities has no taxon (Baselga, 2010). For each taxonomic group, we built GLMMs 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 the analyses for the turnover and nestedness components of beta diversity. We ran all generalized mixed models with three MCMC chains, 5,000 iterations and a burn-in of 5,000 in the *brms* R package (Bürkner, 2017). After processing, c-hat values were always <1.01, indicating convergence.

To visualize the variation of the structure of belowground communities across different stages of soil development, we used distance-based Principal Component Analysis (db-PCA). We calculated distance among samples using the Hellinger distance that allows us to control for the double zero issue (Legendre & Legendre, 2012). Prior to ordination, count data were normalized with a shift-log transformation in order to stabilize extreme values and variance inflation. 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 *adonis* function of the *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), 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 years since glacier retreat (i.e., < 40 years; 40-95 years 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. Thus, for this analysis we only retained MOTUs with a relative abundance > 0.1% for each taxonomic group. We computed the IndVal index using the *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 its fidelity (i.e., the relative frequency) within a class. Each MOTU could be an indicator of at maximum two environmental classes (De Cáceres, Legendre, & Moretti, 2010), so that a MOTU could result as indicator of e.g. one or both depths at a given soil stage, or of one or two consecutive stages at a given soil depth. This choice allowed keeping the number and the 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 FDR method (Benjamini & Hochberg, 1995). We used the packages *ggplot2* (Wickham, 2016), *ggpubr* (Kassambara, 2020), *phyloseq* (McMurdie & Holmes, 2013) and *vegan* (Oksanen et al., 2019) for multivariate statistical analyses and visualization.

RESULTS

A total of 7,335,969 (Bact02), 9,655,151 (Euka02), 15,649,401 (Fung02), 8,941,690 (Coll01), 7,537,839 (Inse01) and 6,107,034 (Olig01) reads were obtained after bioinformatic filtering. After clustering and spurious sequence removal, DNA metabarcoding yielded 1,825 (Bact02), 753 (Euka02), 1,483 (Fung02), 118 (Coll01), 396 (Inse01) and 97 (Olig01) MOTUs.

How is alpha-diversity related to soil depth, time since deglaciation and soil features?

Overall, the alpha-diversity was highest for the generalist markers (Euka02, Bact02, Fung02) compared to the specialist ones (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 of alpha-diversity with time since glacier retreat for all the taxonomic groups. For Eukaryota and Mycota, alpha-diversity was significantly higher in communities retrieved at surface layers with depth of 0-5 cm, compared to the communities detected

in the deeper layer of soil (Fig. 2; Table S3). Furthermore, for Collembola we detected an interaction between depth and time since glacier retreat. For this group, the alpha-diversity index was close to

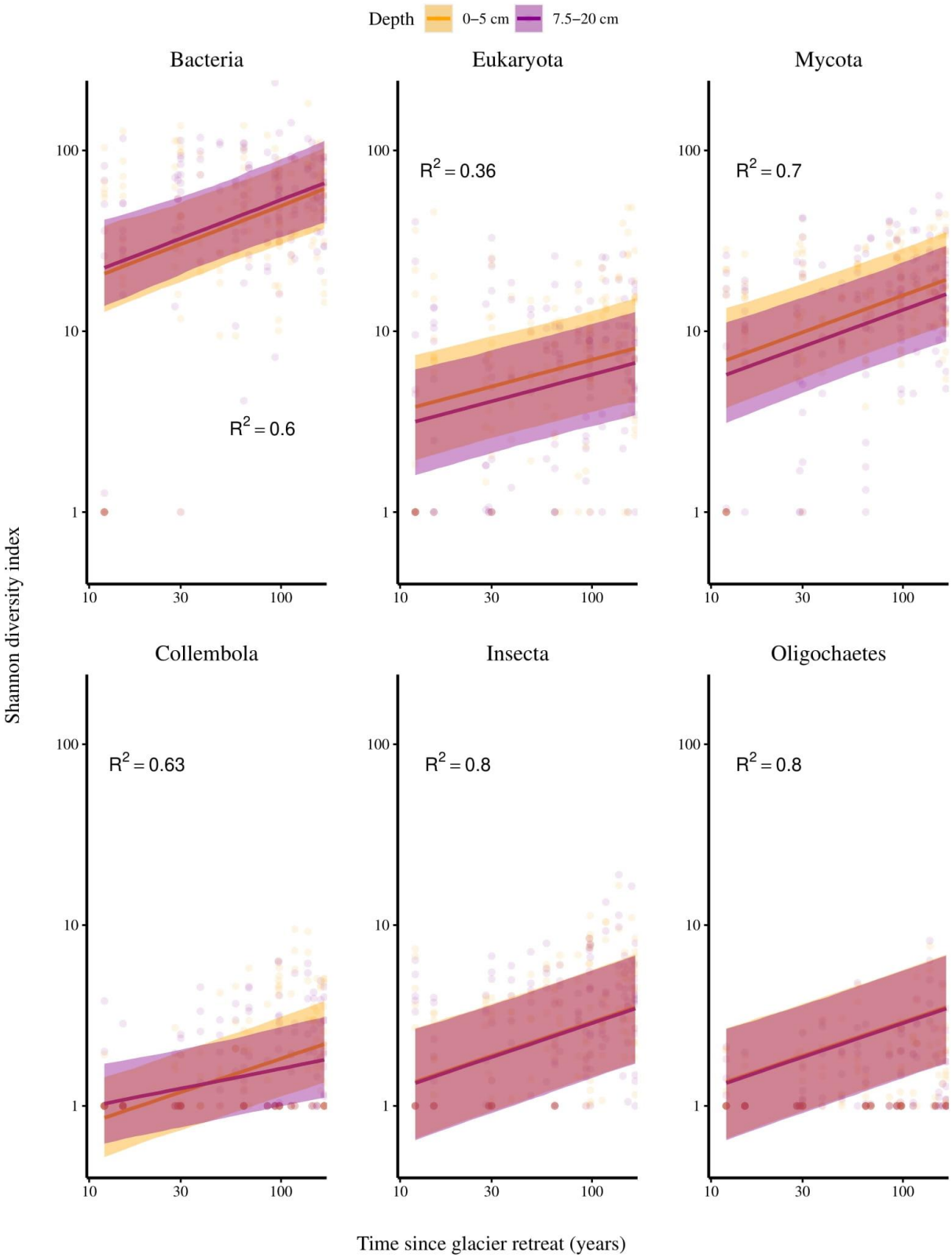


Figure 2. Variation of alpha-diversity (measured using q = 1) through time and depth in six taxonomic groups. Results obtained using q = 0 can be seen in Table S3.

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 we repeated analyses using q = 0 (Table S3). Results were very similar when we used soil carbon content as a predictor variable instead of time since glacier retreat, as GLMMs showed a significant increase in alpha-diversity with average organic carbon content of the plot, even though the R^2 values of these models were generally lower than the R^2 of models with age as independent variable (Fig. S1; Table S4).

Changes in beta diversity through time

In order to assess the beta diversity between surface and deep soils, plots having zero MOTUs in at least one depth were removed, corresponding to 4.28% (Bacteria), 14.28% (Eukaryota), 6.43% (Mycota), 39.28% (Collembola), 13.57% (Insecta) and 37.85% (Oligochaeta) of total plots.

GLMMs allowed us to detect changes in the beta-diversity of communities between surface and deep soil. Differences in community composition between the two depths decreased with time since glacier retreat for Bacteria, Eukaryota, Mycota and Insecta, indicating homogenization of communities, while we did not detect significant changes through time for the beta-diversity of Collembola and Oligochaeta (Fig. 3; Supplementary Table S5). Collembola and Oligochaeta were also the taxa for which the largest number of sites were discarded because of a lack of MOTUs. Overall, our models did not show significant changes in the turnover or nestedness components of the beta diversity measures through time, with the only exception of Oligochaeta, for which nestedness between surface and deep soils tended to increase through time (Supplementary Table S5 and Fig. S2).

Within each deglaciated foreland, the structure of communities was primarily related to time since glacier retreat (Fig. 4). Time significantly affected community structure for Bacteria, Mycota and Eukaryota (PERMANOVA: $p < 0.05$; Table 1); the amount of variance explained by time ranged from 2.4% to 5.7%. For Bacteria, Mycota, Eukaryota, as well as for Insecta, community structure

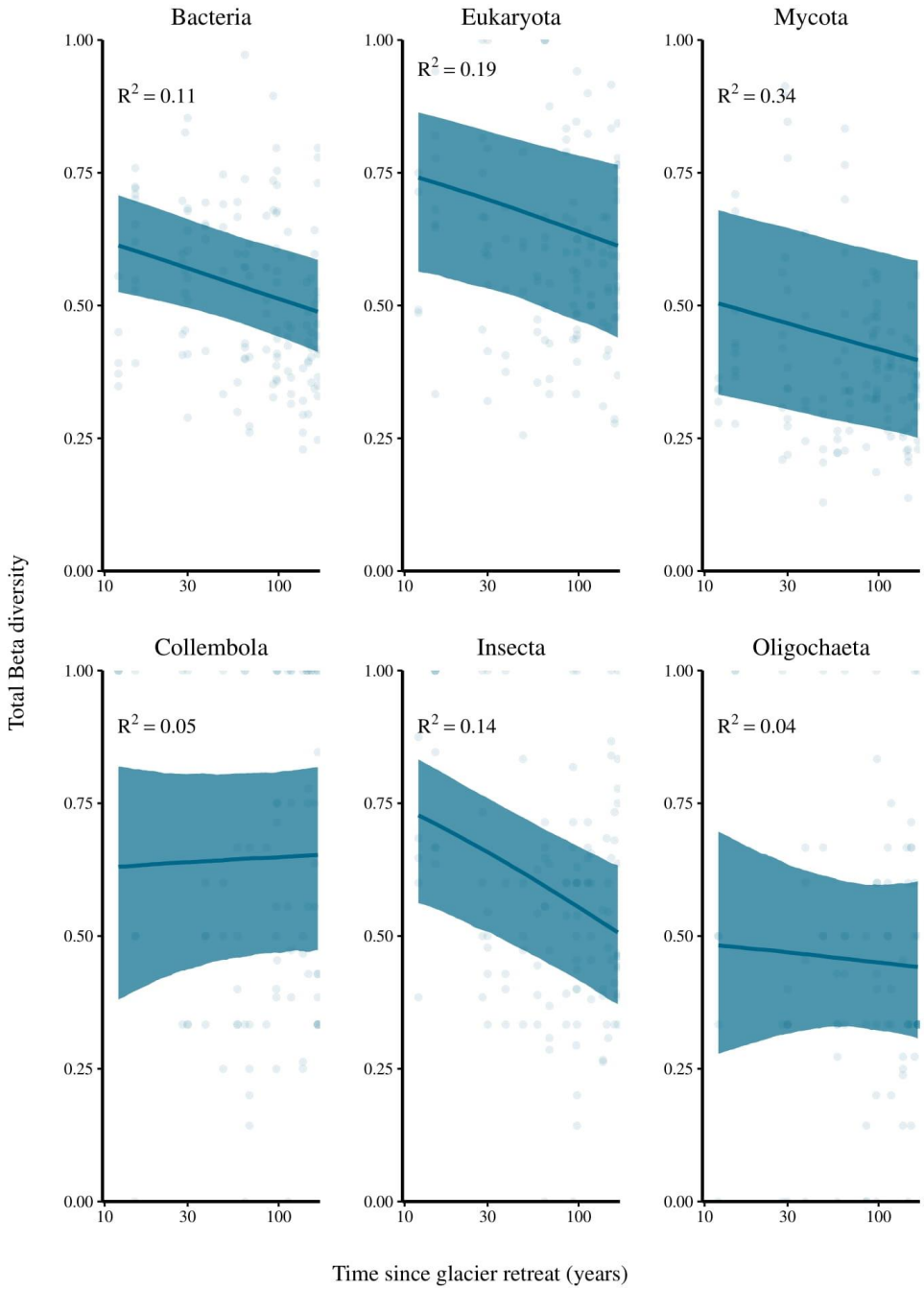


Figure 3. Differences in total beta diversity between soil depths through time.

also differed significantly between soil depths, but the explained variance was smaller (< 1%; Table 1). For none of the groups, we detected a significant interaction between time and soil depth (Table 1), suggesting that the effect of time was consistent between surface and deep soils. Differences in multivariate dispersions were never significant between soil depths, but were significant across time except for Collembola (Table 1). Bacterial community structure was the most strongly related to time

and depth (R^2 = 5%; Table 1). Differences among deglaciated forelands were marked but tended to follow similar trends across the taxonomic groups (Fig. 4).

Table 1. Differences in community structures (Bray-Curtis dissimilarities) 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. * *p*-value ≤ 0.05; ** *p*-value ≤ 0.01, * *p*-value ≤ 0.001.**

Taxonomic group	Variables	Variance explained (R^2)	*P*-value	Dispersion homogeneity *P*-value
Bacteria	Time	0.057	**0.003****	**<0.001*****
	Depth	0.006	**0.001****	0.585
	Time × Depth	0.002	0.354	-
Eukaryota	Time	0.024	**0.004****	**<0.001*****
	Depth	0.004	**0.001*****	0.310
	Time × Depth	0.002	0.716	-
Mycota	Time	0.044	**0.005****	**<0.001*****
	Depth	0.002	**0.005****	0.314
	Time × Depth	0.001	0.102	-
Collembola	Time	0.023	0.501	**0.035***
	Depth	0.002	0.692	0.408
	Time × Depth	0.004	0.321	-
Insecta	Time	0.029	0.102	**<0.001*****
	Depth	0.002	**0.047***	0.827
	Time × Depth	0.002	0.354	-
Oligochaeta	Time	0.035	0.09	**0.002****
	Depth	0.003	0.086	0.890
	Time × Depth	0.003	0.176	-

Based on the specificity and fidelity of each MOTU, 86 were identified as indicators (47 Bacteria, 34 Mycota and five Eukaryota; Table S6). For Bacteria, 22 taxa were strongly associated with young foreland soils, including members of the genera *Roseiflexus*, *Herbaspirillum*, *Novosphingobium* that exhibited particularly high IndVal, while no one was strictly associated to the intermediate ages. Seventeen taxa of Bacteria were indicators of both surface and deep soil layers in older forelands, including members from the genera *Actinoallomurus* and *Ferrimicrobium* that showed the highest

IndVal. Six taxa were indicators of the deep soil layers at both intermediate and old age. For Eukaryota, five taxa were considered as indicators; three were fungi related to old soils, while one mite (genus *Gamasina*) was associated with the intermediate age class. For Mycota, 18 taxa were indicators of both surface and deep layers in older forelands, including members of the genus *Cladophialophora* and the family *Glomeraceae.* Ten Mycota taxa were indicators of both surface and deep layers in young forelands while intermediate forelands contained less indicators, with only five taxa. Only one Mycota taxon was representative of a specific soil layer (the MOTU identified as *Golovinomyces sordidus*, associated to the surface layer of young forelands).

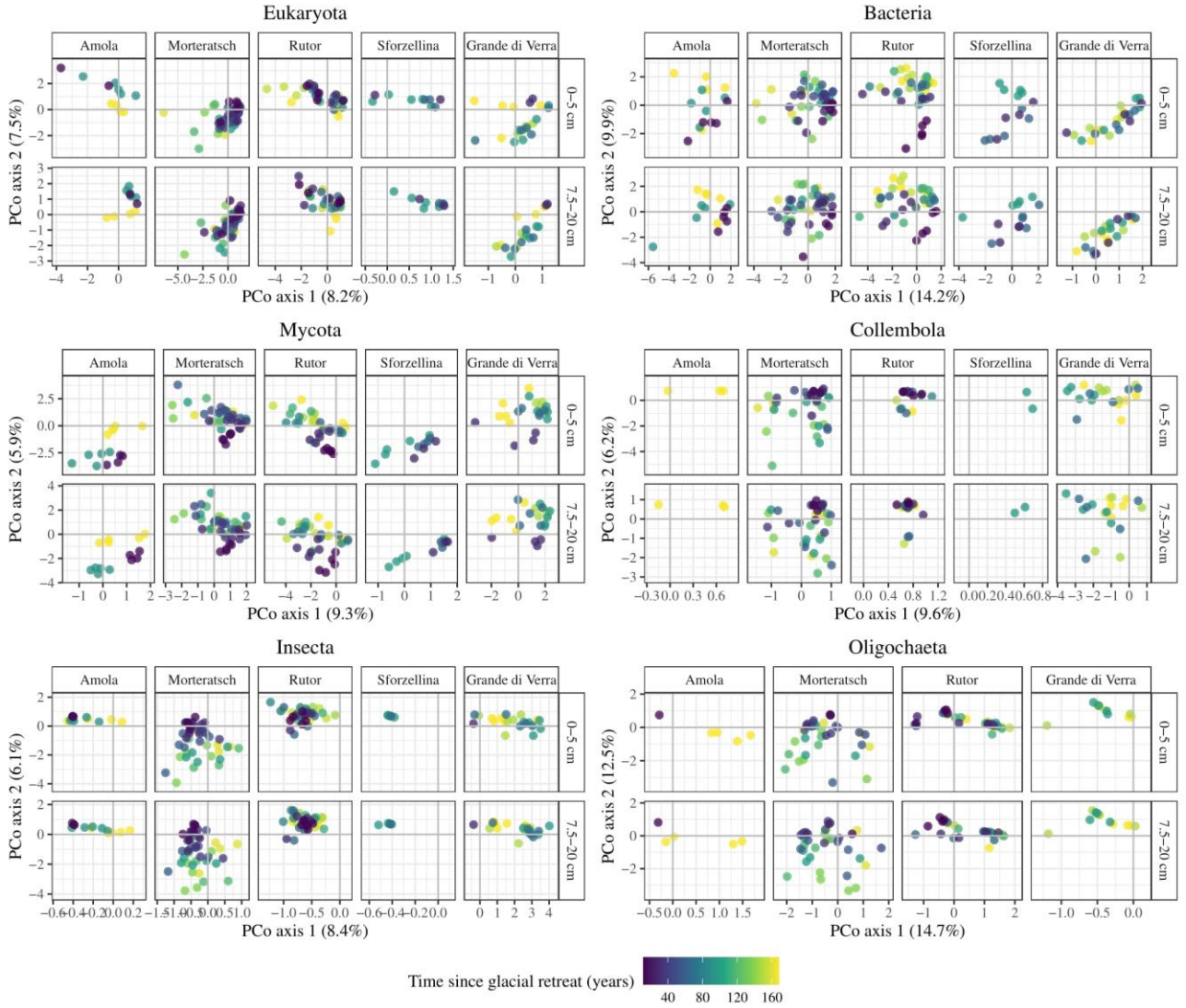


Figure 4. Ordination of the community structures (Hellinger distances dissimilarities) of the six taxonomic groups in the five proglacial plains at two sampling depths (0-5 cm and 7.5-20 cm). The first two axes of the distance-based Principal Component Analyses are displayed with corresponding percentage of explained variance. Sample points are displayed with color representing time since glacier retreat.

DISCUSSION

Our work provides new insights on the colonization and primary succession patterns in deglaciated terrains, by integrating soil depth in the primary succession studies and by implementing a multi-taxa approach across multiple forelands. Here, the multi-taxa approach allowed characterizing patterns for a wide range of topsoil organisms involved in colonization and successional processes. In order to cover the largest proportion of biodiversity, we considered both generalist (for Bacteria, Mycota and Eukaryota) and more specific (for Collembola, Insecta and Oligochaeta) markers. Alpha- and beta-diversity variation through time showed a strikingly consistent pattern across these taxa. The considered depths did not strongly affect the alpha diversity of some taxa at any stage of soil development, even though communities inhabiting surface and deep soil layers were not exactly the same. Importantly, beta-diversity between surface and deep soil decreased through time across most of taxa, supporting the hypothesis of homogenization between surface and deep soil along the succession (Rime et al., 2015).

Changes in alpha-diversity with soil age 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 a similar pattern across all the study groups. For the whole Eukaryota and, within them, for Mycota, the highest alpha-diversity was found in the surface soils, supporting our hypothesis that the richness of communities decreases toward deep soils. 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, soils tend to have higher water holding capacity, more exchangeable cations, carbon and nutrient contents toward the surface (Rime et al., 2015). These properties are vital for most belowground organisms,

especially in those resource-limited ecosystems, resulting in higher bacterial activity, DNA concentration, 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 of depth). In principle, the deep layer might hold larger environmental heterogeneity, given that it is the thickest one. Thus, 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 for this group taxonomic richness increased at different rates between the two soil depths. Springtails were nearly absent in soils aged less than 30 years (Fig. 2). Then, the alpha-diversity increased, but the increase was faster in the surface layer compared to the deep layer, probably because the fast accumulation of organic matter in surface soils (Herold et al., 2014; Moradi et al., 2020) allows the establishment of these organisms, which have multiple trophic roles, from detritivore to herbivorous. For the other taxa, we did not detect significant interaction between soil depth and soil development stage, suggesting that alpha-diversity increases through time with a similar pattern between surface and deep layers, except for Collembola and perhaps in very young soils (see Fig. 2).

In glacier forelands, the amount of organic matter consistently increases through time (Khedim et al., 2021). By repeating the analyses of alpha-diversity, considering the organic carbon content as independent variable instead of time since glacier retreat, we confirmed that our conclusions are not biased by the issues of using different sites as substitutes of time (issues of space-for-time substitution in successional studies; Johnson & Miyanishi, 2008). Soil carbon content is a major driver of soil biodiversity changes (Chu et al., 2016); consistently with this idea, alpha-diversity tended to increase with organic carbon. Nevertheless, models with time showed slightly higher R^2 values than the ones with soil organic carbon (Fig. 2, supplementary Figure S1), 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 in order to disentangle the role of both time and soil features as drivers of primary succession.

Communities 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 time. As seen for the alpha diversity, this pattern was consistent across nearly all taxa. Collembola and Oligochaeta are the only taxa for which this was not evident, but these animals were nearly absent from recently deglaciated soils (and particularly from the deep layers; Fig. 2), probably because many of them require well developed soils, with abundant organic matter to find resources (Phillips et al., 2019). Therefore, for Collembola and Oligochaeta, many plots at early development stages were excluded from this analysis, reducing statistical power. In principle, the variation of beta diversity between surface and deep layer can be attributable to both species gain/loss (nestedness) and replacement (Baselga, 2010). Turnover was more important than nestedness for invertebrates (Collembola, Insecta, Oligochaeta), while for microorganisms (Bacteria and Mycota) turnover and nestedness showed a similar importance (Supplementary fig. S2), and the relevance of these two components of beta-diversity remained similar through time (Supplementary Fig. S2; 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 animals responded the same way (Fig. 3). Community homogenization is probably related to the structural modifications observed during the development of soil horizons (e.g. Schaetzl & Anderson, 2005). The study of sites at different stages of soil formation has shown a differentiation of organic horizon immediately after glacial retreat (O), followed by the development of an organo-mineral horizon (A) during the first 150 years (Crocker & Major, 1955; Mavris, Egli, Plotze, Blum, Mirabella and Giaccai, 2010). The strong vertical variation of physical, chemical and structural features (e.g. light, temperature, pH; Moradi et al., 2020; Mundra

et al., 2021) clearly affects communities, which show a particularly strong response to fine-scale environmental heterogeneity (Rime et al., 2015; Moradi et al., 2020; Mundra et al., 2021). For example, immediately after glacier retreat, the amount of fine sediments is the highest at the surface (Rime et al., 2015). This can determine differences in humidity between the surface and the deeper layers, that in turn affect communities (Rime et al., 2015). The decrease of beta-diversity can be explained by the progressive deepening of the organo-mineral horizon (Mavris et al., 2010), where abundant resources favour the establishment of complex communities. Plant richness and cover quickly increase during the first decades after glacier retreat, and 40 years after glacier retreat plants cover generally rises above 50% (Rime et al., 2015). Plant roots generally influence the first 20 cm of soils and more, and could have determined the homogenization of superficial and deep samples of our study (Rime et al., 2015). Differences between surface and deep layers would probably be stronger if a larger vertical gradient is analysed (e.g. from surface to 50 cm deep; Moradi et al., 2020), and this is certainly an important aspect that deserves future studies. However, in glacier forelands the study of deep layers by eDNA analysis is sometimes problematic because rock outcrops are frequently a few centimetres below the surface. In any case, for all the taxa considered here, time since glacier retreat remained the main determinant of community variation, as it explained much more variation in community composition compared to depth (Table 1). 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 after glacier retreat (Ficetola et al., 2021; Rime et al., 2015).

For microorganisms, the significant community differences between soil layers (Table 1) likely are determined by taxa that are specialists of given environmental features. This idea is supported by the observation that all MOTUs identified as indicators of surface or deep soils are bacteria or fungi (supplementary Table S6). Conversely, for invertebrates, soil depth explained a very limited amount of variation in community composition (Table 1). 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). 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. Taxa identified as clear indicators both here and by Rime et al. (2015) include the *Clostridium* bacteria, known as anaerobic, which were consistently found as indicators of the earliest stage of soil development. Similarly, several fungal saprotrophs were indicators across the different stages of soil development, while *Lachnum* was a microfungus consistently associated with the most developed soils (Nguyen et al., 2016). *Gemmatimonas* tend to be copiotrophic bacteria (Ho, Di Lonardo, & Bodelier, 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 also indicators of later stages of development. Contrary to Rime et al. (2015), arbuscular mycorrhizal fungi (i.e., Glomeromycetes) tended to be associated with oldest forelands, confirming the growing importance of plant-associated fungi along community development (Davey et al., 2015). The indicative species represented only a small fraction of the community and very limited information is available for most of the taxa either because they are poorly studied or because the taxonomic level is not precise, still the analysis suggested a strong functional variation of microbial communities through time and depth. Based on these results, further studies should use the entire taxonomy and perform more accurate functional traits analyses in order to understand the ecological role of communities at different stage of soil development.

Conclusion

Understanding the development of communities in primary 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 on ecosystem functioning, for example for the sequestration of organic carbon in these soils (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. Future studies are required to identify possible factors driving biotic colonization within the same system (e.g. microclimate, soil features, etc.) and patterns of biotic interactions.

ACKNOWLEDGMENTS

This study is funded by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities).

Supplementary Information for:

Metabarcoding data reveal vertical multi-taxa variation in topsoil communities during the colonization of deglaciated forelands

Alessia Guerrieri, Alexis Carteron, Aurélie Bonin, Silvio Marta, Roberto Ambrosini, Marco Caccianiga, Isabel Cantera, Chiara Compostella, Guglielmina Diolaiuti, Diego Fontaneto, Ludovic Gielly, Fabrizio Gili, Mauro Gobbi, Jerome Poulenard, Pierre Taberlet, Andrea Zerboni, Wilfried Thuiller, Gentile Francesco Ficetola

SUPPLEMENTARY TABLES

Table S1. Characteristics of Metabarcode used for this study.

Marker	Target group	Target gene	Forward primer/ Reverse primer	Min. metabarcode length (bp*)	Max. metabarcode length (bp*)
Bact02	Bacteria	V4 region of the 16S rDNA gene	GCCAGCMGCCGCGGTAA/ GGACTACCMGGGTATCTAA	45	748
Euka02	Eukaryota	V7 region of the 18S rDNA gene	TTTGTCTGSTTAATTSCG/ CAAGAGATCCGTTGYTGAAAGTK	36	892
Fung02	Fungi	ITS1 nuclear rDNA gene	GGAAGTAAAAGTCGTAACAAGG/ CAAGAGATCCGTTGYTGAAAGTK	68	919
Coll01	Collembola	16S mitochondrial rDNA gene	ACGCTGTTATCCCTWAGG/ GACGATAAGACCCTWTAGA	76	192
Inse01	Insecta	16S mitochondrial rDNA gene	RGACGAGAAGACCCTATARA/ ACGCTGTTATCCCTAARGTA	75	265
Olig01	Oligochaeta	16S mitochondrial rDNA gene	CAAGAAGACCCTATAGAGCTT/ CCTGTTATCCCTAAGGTARCT	93	196

*Excluding primers

Table S2. Pearson's product moment correlations (r) between alpha-diversity metrics obtained with different Hill numbers (q = 0 and q = 1). Significant correlations are in bold; for all the significant correlations, P < 0.01.

	Bacteria: h1	Eukaryota: h1	Mycota: h1	Collembola: h1	Insecta: h1	Oligochaeta: h1
Bacteria: h0	**0.95**					
Eukaryota: h0		**0.87**				
Mycota: h0			**0.78**			
Collembola: h0				**0.93**		
Insecta: h0					**0.90**	
Oligochaeta: h0						**0.87**

Table S3. Results of Bayesian univariate models assessing the relationship between the two alpha diversity metrics of the six groups, time since glacier retreat and soil depth. Parameters with 95% credible intervals non-overlapping zero are highlighted in bold. Interactions were included in the model for Collembola and Mycota as the model without interactions showed higher WAIC than the model with interactions [For Collembola: 341.8 vs 333.9 (Shannon) and 930.9 vs 925.9 (Richness); for Mycota: 2578.8 vs 2574.5 (Richness)].

Taxonomic group	Dependent variable	Independent variables	*B*	Q2.5	Q97.5
Bacteria	Shannon index	Time	**0.3384**	**0.2042**	**0.4697**
		Depth	0.0735	-0.0735	0.2201
	Richness	Time	**0.1328**	**0.0276**	**0.2435**
		Depth	0.0839	-0.1115	0.2737
Eukaryota	Shannon index	Time	**0.2374**	**0.1193**	**0.3556**
		Depth	-0.1877	-0.3753	0.0009
	Richness	Time	**0.2891**	**0.1558**	**0.4305**
		Depth	**-0.2427**	**-0.4640**	**-0.0276**
Mycota	Shannon index	Time	**0.3264**	**0.2128**	**0.4383**
		Depth	**-0.1832**	**-0.2979**	**-0.0677**
	Richness	Time	**0.3742**	**0.2469**	**0.5061**
		Depth	**-0.1510**	**-0.2523**	**-0.0497**
		Time x depth	**0.1142**	**0.0052**	**0.2252**
Collembola	Shannon index	Time	**0.3000**	**0.2100**	**0.3809**
		Depth	-0.0583	-0.1389	0.0233
		Time x depth	**-0.1220**	**-0.2015**	**-0.0381**
	Richness	Time	**0.7008**	**0.5225**	**0.8930**
		Depth	-0.1187	-0.3095	0.0691
		Time x depth	**-0.2477**	**-0.4612**	**-0.0300**
Insecta	Shannon index	Time	**0.3016**	**0.2181**	**0.3834**
		Depth	-0.0143	-0.0911	0.0637
	Richness	Time	**0.4581**	**0.3413**	**0.5772**
		Depth	-0.0656	-0.1560	0.0222
Oligochaeta	Shannon index	Time	**0.1194**	**0.0645**	**0.1760**
		Depth	-0.0253	-0.0881	0.0369
	Richness	Time	**0.4077**	**0.2726**	**0.5425**
		Depth	**-0.0316**	**-0.1991**	**0.1382**

Table S4. Results of Bayesian univariate models assessing the relationship between Shannon diversity of the six groups and soil organic carbon content of the entire soil core. Parameters with 95% credible intervals non-overlapping zero are highlighted in bold.

Taxonomic group	Dependent variable	Independent variable	*B*	Q2.5	Q97.5
Bacteria	Shannon index	Organic carbon	**0.3958**	**0.2383**	**0.5596**
Eukaryota	Shannon index	Organic carbon	**0.3289**	**0.1903**	**0.4642**
Mycota	Shannon index	Organic carbon	**0.4263**	**0.2904**	**0.5591**
Collembola	Shannon index	Organic carbon	**0.3526**	**0.2751**	**0.4270**
Insecta	Shannon index	Organic carbon	**0.4235**	**0.3298**	**0.5181**
Oligochaeta	Shannon index	Organic carbon	**0.1602**	**0.0925**	**0.2285**

Table S5. Results of Bayesian univariate models assessing the relationship between the beta diversity measures (total diversity, turnover, nestedness) of the six groups and time since glacier retreat. Parameters with 95% credible intervals non-overlapping zero are highlighted in bold.

Taxonomic group	Dependent variable	Independent variable	*B*	Q2.5	Q97.5
Bacteria	Total	Time	**-0.1512**	**-0.2604**	**-0.0440**
	Turnover	Time	-0.0831	-0.2431	0.0786
	Nestedness	Time	-0.1020	-0.2655	0.0638
Eukaryota	Total	Time	**-0.1702**	**-0.3209**	**-0.0184**
	Turnover	Time	-0.0042	-0.2156	0.2139
	Nestedness	Time	-0.0236	-0.2153	0.1773
Mycota	Total	Time	**-0.0474**	**-0.0735**	**-0.0217**
	Turnover	Time	0.0653	-0.0672	0.1951
	Nestedness	Time	-0.1251	-0.2858	0.0387
Collembola	Total	Time	0.0360	-0.2232	0.3006
	Turnover	Time	0.0934	-0.1955	0.3998
	Nestedness	Time	0.0262	-0.1960	0.2567
Insecta	Total	Time	**-0.2824**	**-0.4717**	**-0.1018**
	Turnover	Time	-0.2543	-0.4663	-0.0421
	Nestedness	Time	0.0525	-0.1178	0.2275
Oligochaeta	Total	Time	-0.0460	-0.3122	0.2111

	Turnover	Time	-0.1728	-0.4462	0.1153
	Nestedness	Time	**0.2623**	**0.0290**	**0.4982**

Table S6**.** Indicator value (IndVal) analysis aiming at identifying taxon as indicator of group of deglaciated forelands that are young (Y), intermediate (I) or older (O) in two soil depths of 0-5 cm (1) or 7.5-20 cm (2). Abbreviated such as Y-1 is the group of young deglaciated forelands of 0-5 cm soil depth. Only indicator taxa with *p*-value ≤ 0.01 are reported and their associated group or combination of groups is shown with *. *P*-values were adjusted for multiple comparison tests using the FDR method. Color shading represents the year class (i.e., Y: below 40; I: between 40 and 95; O: above 95).

Group	ID	Most close taxon	IndVal	*P*-value	Y-1	Y-2	I-1	I-2	O-1	O-2
Bacteria	Bact02-00053	uncultured *Roseiflexus* sp.	0.62	0.002	*	*				
Bacteria	Bact02-00070	uncultured *Acidisphaera* sp.	0.48	0.008	*	*				
Bacteria	Bact02-00077	Bacteroidetes	0.53	0.002	*	*				
Bacteria	Bact02-00080	Acidimicrobiales	0.55	0.002	*	*				
Bacteria	Bact02-00088	Novosphingobium	0.57	0.002	*	*				
Bacteria	Bact02-00099	*Segetibacter aerophilus*	0.5	0.003	*	*				
Bacteria	Bact02-00145	uncultured *Ohtaekwangia* sp.	0.51	0.004	*	*				
Bacteria	Bact02-00154	*Herbaspirillum* sp.	0.58	0.002	*	*				
Bacteria	Bact02-00174	uncultured *Conexibacter* sp. 1	0.57	0.002	*	*				
Bacteria	Bact02-00189	uncultured *Elusimicrobium* sp.	0.4	0.005	*	*				
Bacteria	Bact02-00198	Angustibacter	0.42	0.003	*	*				
Bacteria	Bact02-00214	uncultured *Gemmatimonas* sp. 1	0.46	0.009	*	*				
Bacteria	Bact02-00225	uncultured *Legionella* sp.	0.42	0.004	*	*				
Bacteria	Bact02-00238	Sphingomonadales 1	0.51	0.002	*	*				
Bacteria	Bact02-00245	uncultured *Sphingomonas* sp.	0.48	0.002	*	*				
Bacteria	Bact02-00250	uncultured *Xanthomonas* sp.	0.44	0.002	*	*				
Bacteria	Bact02-00271	uncultured *Acidothermus* sp.	0.44	0.008	*	*				
Bacteria	Bact02-00273	Acetobacteraceae	0.42	0.002	*	*				
Bacteria	Bact02-00297	uncultured *Nitrosospira* sp.	0.47	0.002	*	*				
Bacteria	Bact02-00338	Clostridium	0.47	0.003	*	*				
Bacteria	Bact02-00377	Sphingomonadales 2	0.41	0.009	*	*				
Bacteria	Bact02-00421	Oxalobacteraceae	0.4	0.009	*	*				
Bacteria	Bact02-00614	Unidentified 1	0.39	0.007		*		*		
Bacteria	Bact02-00136	uncultured *Verrucomicrobium* sp.	0.59	0.005				*		*
Bacteria	Bact02-00149	uncultured *Pedomicrobium* sp.	0.54	0.008				*		*
Bacteria	Bact02-00206	Solirubrobacterales	0.63	0.002				*		*
Bacteria	Bact02-00362	uncultured *Gemmatimonas* sp. 2	0.52	0.003				*		*
Bacteria	Bact02-00634	uncultured *Geobacter* sp.	0.43	0.006				*		*
Bacteria	Bact02-00754	Unidentified 2	0.42	0.01				*		*
Bacteria	Bact02-00072	uncultured *Acidobacterium* sp.	0.59	0.01					*	*
Bacteria	Bact02-00085	uncultured *Actinoallomurus* sp.	0.63	0.002					*	*
Bacteria	Bact02-00100	uncultured *Gemmatimonas* sp. 3	0.64	0.002					*	*
Bacteria	Bact02-00139	uncultured *Conexibacter* sp. 2	0.55	0.003					*	*
Bacteria	Bact02-00159	Verrucomicrobiaceae	0.57	0.002					*	*

Bacteria	Bact02-00196	uncultured *Holophaga* sp. 1	0.6	0.002				*	*
Bacteria	Bact02-00208	uncultured *Ferrimicrobium* sp. 1	0.62	0.002				*	*
Bacteria	Bact02-00212	Unidentified 3	0.54	0.005				*	*
Bacteria	Bact02-00226	*Puia dinghuensis*	0.56	0.002				*	*
Bacteria	Bact02-00278	Terrabacteria group	0.41	0.005				*	*
Bacteria	Bact02-00296	uncultured *Acidisphaera* sp.	0.54	0.002				*	*
Bacteria	Bact02-00306	uncultured *Actinomyces* sp.	0.55	0.002				*	*
Bacteria	Bact02-00322	*Actinocrinis puniceicyclus*	0.43	0.006				*	*
Bacteria	Bact02-00401	uncultured *Ferrimicrobium* sp. 2	0.41	0.003				*	*
Bacteria	Bact02-00509	uncultured *Holophaga* sp. 2	0.41	0.009				*	*
Bacteria	Bact02-00535	Unidentified 4	0.46	0.002				*	*
Bacteria	Bact02-00593	uncultured *Conexibacter* sp. 3	0.43	0.005				*	*
Bacteria	Bact02-00669	Unidentified 5	0.45	0.008				*	*
Eukaryota	Euka02-00073	Unidentified	0.46	0.005	*	*			
Eukaryota	Euka02-00268	*Gamasina*	0.49	0.005			*	*	
Eukaryota	Euka02-00052	Tremellomycetes	0.57	0.005				*	*
Eukaryota	Euka02-00119	*Umbelopsis* sp.	0.49	0.005				*	*
Eukaryota	Euka02-00291	Agaricomycotina	0.4	0.008					*
Mycota	Fung02-00133	*Golovinomyces sordidus*	0.36	0.007	*				
Mycota	Fung02-00006	Leotiomyceta	0.61	0.001	*	*			
Mycota	Fung02-00032	Melanommataceae	0.53	0.009	*	*			
Mycota	Fung02-00092	Microbotryomycetes	0.41	0.008	*	*			
Mycota	Fung02-00139	Sebacinales	0.44	0.003	*	*			
Mycota	Fung02-00146	Leotiomycetes	0.39	0.003	*	*			
Mycota	Fung02-00154	*Dioszegia* sp.	0.51	0.001	*	*			
Mycota	Fung02-00252	*Mortierella*	0.41	0.004	*	*			
Mycota	Fung02-00266	Helicobasidium	0.45	0.001	*	*			
Mycota	Fung02-00341	*Tomentella* sp.	0.41	0.001	*	*			
Mycota	Fung02-00493	*Pleotrichocladium opacum*	0.37	0.003	*	*			
Mycota	Fung02-00024	*Laccaria* sp.	0.47	0.009			*	*	
Mycota	Fung02-00119	*Sebacina*	0.48	0.003			*	*	
Mycota	Fung02-00223	*Mortierella* sp.	0.49	0.001			*	*	
Mycota	Fung02-00944	*Entorrhiza aschersoniana*	0.43	0.001			*	*	
Mycota	Fung02-00979	*Exophiala*	0.44	0.009			*	*	
Mycota	Fung02-00009	*Pezoloma ericae*	0.55	0.009				*	*
Mycota	Fung02-00027	*Archaeorhizomyces borealis*	0.4	0.004				*	*
Mycota	Fung02-00085	Glomeraceae 1	0.48	0.001				*	*
Mycota	Fung02-00090	*Oidiodendron*	0.44	0.005				*	*
Mycota	Fung02-00093	Glomeraceae 2	0.49	0.001				*	*
Mycota	Fung02-00100	Mortierella 1	0.45	0.001				*	*
Mycota	Fung02-00160	*Cladophialophora chaetospira* 1	0.61	0.001				*	*
Mycota	Fung02-00197	*Lachnum sp.*	0.52	0.001				*	*
Mycota	Fung02-00261	*Solicoccozyma* sp.	0.49	0.009				*	*
Mycota	Fung02-00313	Glomeraceae 3	0.43	0.001				*	*
Mycota	Fung02-00382	*Saitozyma podzolica*	0.49	0.001				*	*

Mycota	Fung02-00411	Helotiales	0.43	0.01	*	*
Mycota	Fung02-00440	*Mortierella* 2	0.46	0.003	*	*
Mycota	Fung02-00468	*Hygrophorus* sp.	0.44	0.001	*	*
Mycota	Fung02-00682	*Cladophialophora chaetospira* 2	0.39	0.005	*	*
Mycota	Fung02-00712	Leotiomycetes	0.52	0.001	*	*
Mycota	Fung02-00976	*Hamatocanthoscypha* sp.	0.51	0.001	*	*
Mycota	Fung02-01782	Hypocreales	0.38	0.003	*	*

SUPPLEMENTARY FIGURES

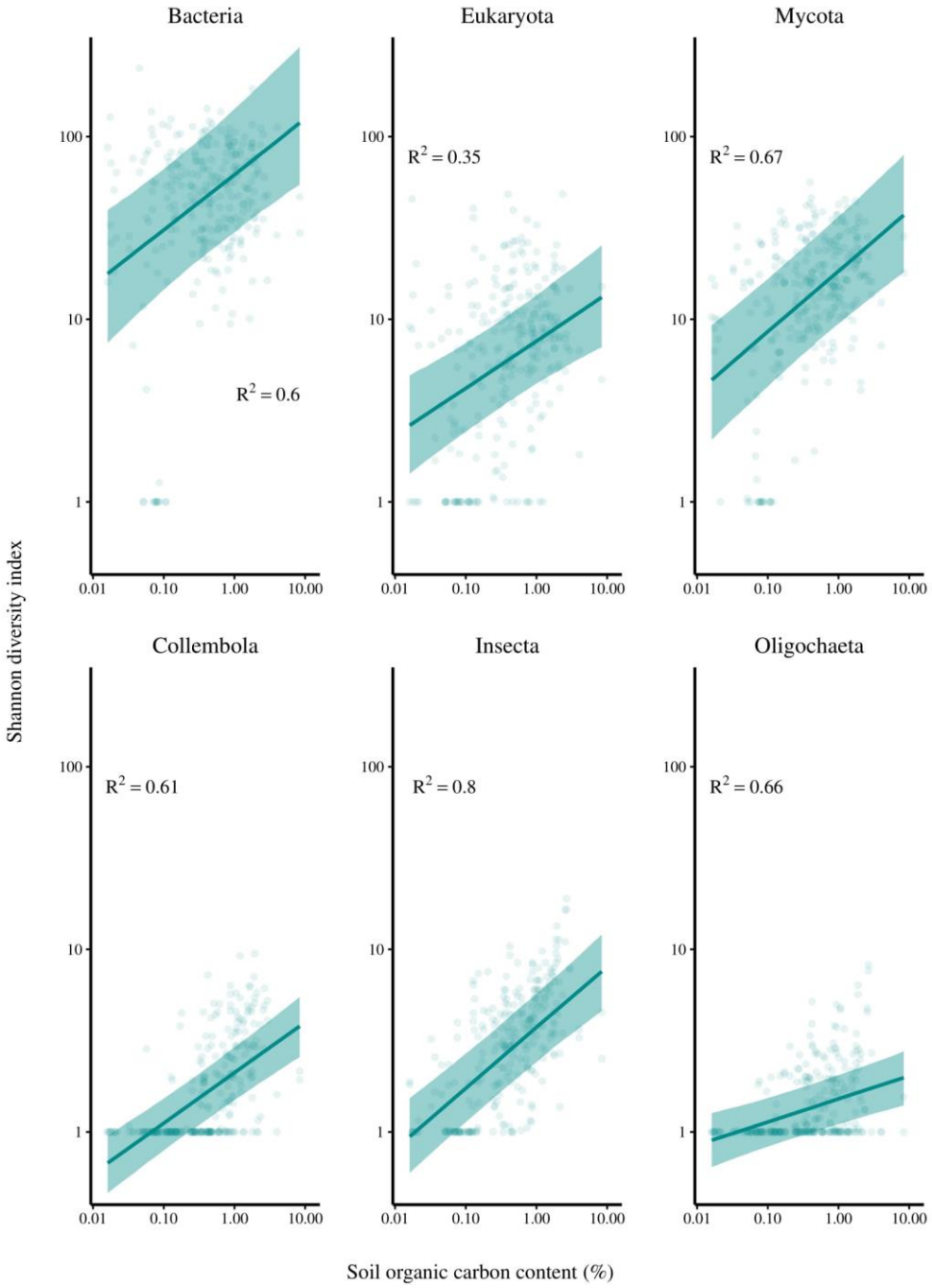


Figure S1. Variation of alpha-diversity (measured using q = 1) with soil organic carbon in six taxonomic groups.

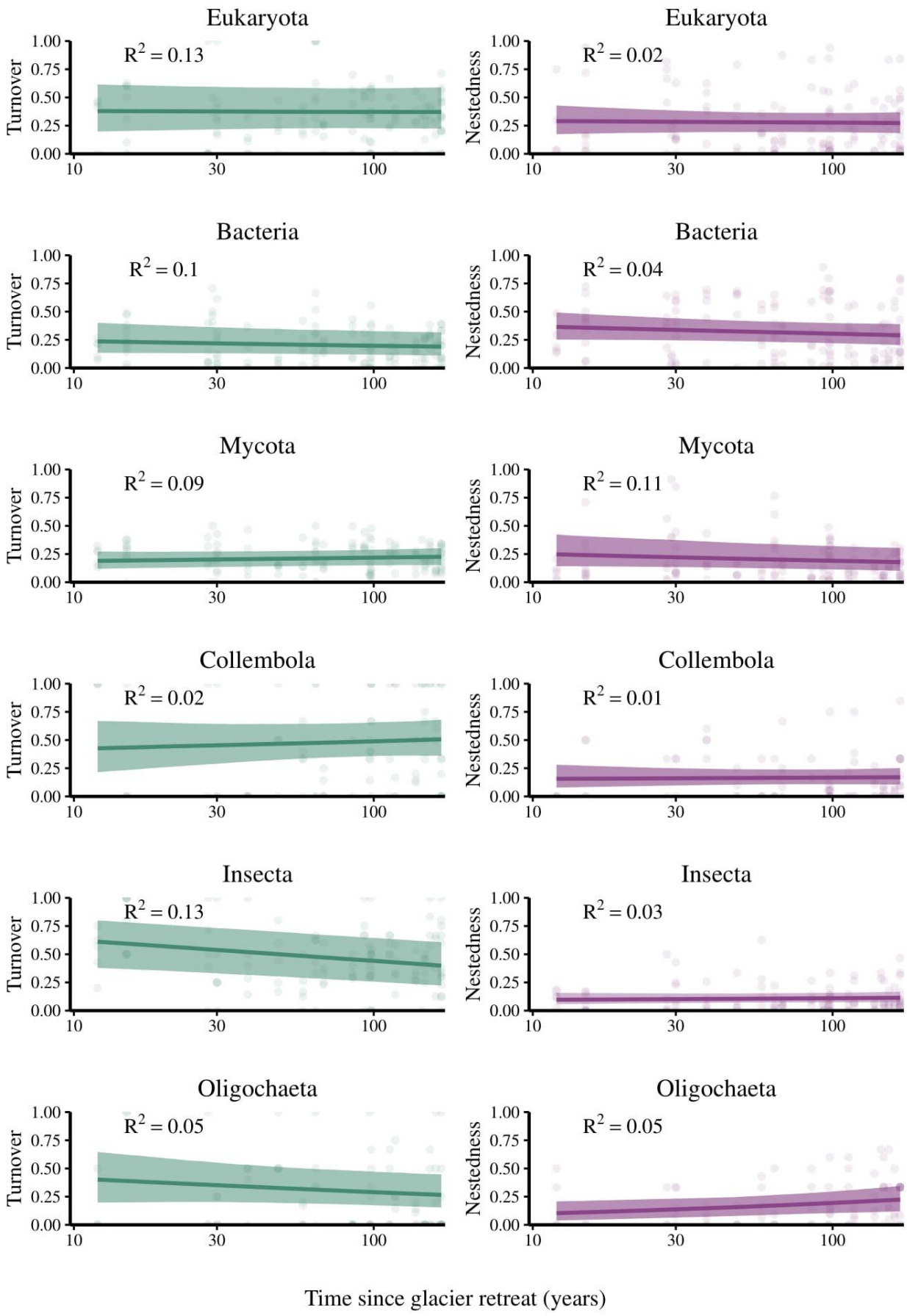


Figure S2. Differences in turnover and nestedness components of beta diversity between soil depths through time.

CHAPTER 5. ARTICLE D[5]

Local climates modulate soil biodiversity dynamics after glacier retreat

Alessia Guerrieri*[1], Isabel Cantera*[1], Silvio Marta*[1], Aurélie Bonin*[1], Roberto Ambrosini[1], Marco Caccianiga[5], Fabien Anthelme[4], Roberto Sergio Azzoni[1], Peter Almond[2], Pablo Alviz[3] Gazitúa, Sophie Cauvy-Fraunié[6], Jorge Luis Ceballos Lievano[7], Pritam Chand[8], Milap Chand Sharma[9], John Clague[10], Justiniano Alejo Cochachín Rapre[11], Chiara Compostella[12], Rolando Cruz Encarnación[11], Olivier Dangles[13], Philip Deline[14], Andre Eger[15], Sergey Erokhin[16], Andrea Franzetti[17], Ludovic Gielly[18], Fabrizi Gili[1,19], Mauro Gobbi[20], Sigmund Hågvar[21], Norine Khedim[14], Gwendolyn Peyre[23], Francesca Pittino[17], Angela Proietto[12], Antoine Rabatel[24], Nurai Urseitova[16], Yan Yang[25], Vitalii Zaginaev[16], Andrea Zerboni[12], Anaïs Zimmer[26], Pierre Taberlet[18,22], Guglielmina Adele Diolaiuti[1], Jerome Poulenard[14], Diego Fontaneto*[27], Wilfried Thuiller*[18], Gentile Francesco Ficetola*[1,18]

[1] Dipartimento di Scienze e Politiche Ambientali, Università degli Studi di Milano, via Celoria 10, 20133 Milano, Italy

[2] Department of Soil and Physical Sciences, Lincoln University, Lincoln 7647, New Zealand

[3] Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Osorno, Chile

[4] Laboratory AMAP, IRD, University of Montpellier, CIRAD, CNRS, INRA, Montpellier, France

[5] Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 10, 20133 Milano, Italy

[6] INRAE, UR RIVERLY, Centre de Lyon-Villeurbanne, Villeurbanne, France

[7] Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM, Bogotá, Colombia

[8] Department of Geography, School of Environment and Earth Sciences, Central University of Punjab, VPO- Ghudda, Bathinda -151401, Punjab, India

[9] Centre for the Study of Regional Development - School of Social Sciences, Jawaharlal Nehru University, New Mehrauli Road, 110067 New Delhi, India

[10] Department of Earth Sciences, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6, Canada

[11] Área de Evaluación de Glaciares y Lagunas, Autoridad Nacional del Agua, Huaraz, Peru

[12] Dipartimento di Scienze della Terra "Ardito Desio", Via L. Mangiagalli 34, 20133 Milano, Italy

[5] Ready for submission to Nature

[13] CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France

[14] Université Savoie Mont Blanc, Université Grenoble Alpes, EDYTEM, 73000 Chambéry, France

[15] Mannaki Whenua - Landcare Research, Soils and Landscapes, 54 Gerald St, Lincoln 7608, New Zealand

[16] Institute of Water Problems and Hydro-Energy, Kyrgyz National Academy of Sciences, 720033, Frunze, 533, Bishkek, Kyrgyzstan

[17] Department of Earth and Environmental Sciences (DISAT) - University of Milano-Bicocca, Milano, Italy

[18] Laboratoire d'Ecologie Alpine, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Grenoble, France

[19] Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy

[20] Section of Invertebrate Zoology and Hydrobiology, MUSE-Science Museum, Corso del Lavoro e della Scienza, 3, 38122 Trento, Italy

[21] Department of Ecology and Natural Resource Management (INA), Norwegian University of Life Sciences, Universitetstunet 3, 1433 Ås, Norway

[22] UiT – The Arctic University of Norway, Tromsø Museum, Tromsø, Norway

[23] Department of Civil and Environmental Engineering, University of the Andes, Bogotá, Colombia

[24] Univ. Grenoble Alpes, CNRS, IRD, Grenoble-INP, Institut des Géosciences de l'Environnement (IGE, UMR 5001), Grenoble, France

[25] Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, 610041, China

[26] Department of Geography and the Environment, University of Texas at Austin, 78712 Austin, TX, USA

[27] CNR - Water Research Institute, Largo V. Tonolli 50, 28922, Verbania, Italy

* These authors contributed equally to this work.

The worldwide retreat of glaciers causes growing availability of ice-free areas that are colonized by life forms and develop into new ecosystems, and we urgently need to understand the dynamics of these environments. Climatic differences between regions could affect the emergence of biodiversity and functionality after glacier retreat, but global tests of this hypothesis are lacking. Nematode worms are the most abundant soil animals, with a keystone role in soil functioning. We used a global dataset based on DNA metabarcoding to evaluate how nematode's taxonomic and functional diversity evolves in ecosystems developing after the retreat of glaciers. We show that nematodes colonize terrains immediately after glacier retreat; their taxonomic and functional diversity quickly increase through time, but the colonization pattern strongly interacts with local climates, with very rapid initial colonization in forelands under mild summer temperatures. Colder forelands initially host poorer communities, but then colonization rate accelerates, levelling biodiversity differences between climatic regimes. Communities showed functional shifts during colonization, from small-sized colonizers to large-sized persisters, indicating that environmental filtering affects succession patterns of soil animals. The retreat of glaciers is followed by heterogeneous colonization dynamics depending on local climates. Recognizing the complexity of biotic responses will allow to understand how these environments evolve and to predict cascading effects at multiple scales.

The world's mountain glaciers are retreating at an unprecedented rate due to ongoing climate change, and the pace of glacier retreat is accelerating across the globe (Hugonnet *et al.* 2021). This dramatic glacier shrinkage will have multiple and significant impacts on both the biotic and abiotic components of ecosystems and the services they provide (Intergovernmental Panel on Climate Change 2021). Deglaciated areas undergo rapid geomorphological changes and a quick colonization by multiple lifeforms, and the interactions among colonizers have complex impacts on the dynamics and the functioning of these emerging ecosystems (Ficetola *et al.* 2021b). The loss of mountain glaciers is likely an irreversible process (Intergovernmental Panel on Climate Change 2021), thus we urgently

need standardized, global-scale studies on how local-to-regional differences (e.g. differences in climate) determine the development of different biodiversity components in these ecosystems under a changing climate (Crowther *et al.* 2019; Ficetola *et al.* 2021b). This can allow to understand how biodiversity develops in different climatic regimes, to predict the future dynamics of these ecosystems and to identify common patterns in biotic colonization (Cauvy-Fraunié & Dangles 2019).

Soil formation is a key step of the development of ecosystems after glacier retreat (Cauvy-Fraunié & Dangles 2019; Ficetola *et al.* 2021b; Khedim *et al.* 2021). Nematode worms are a dominant component of soil communities where they are the most abundant animals, with a crucial role in the development and functioning of ecosystems (Crowther *et al.* 2019; Fontaneto 2019; van den Hoogen *et al.* 2019; Delgado-Baquerizo *et al.* 2020; Zawierucha *et al.* 2021). Nematodes have very broad range of functional roles, and operate at all trophic levels of food webs. Their biodiversity affects a broad range of ecosystem services, from productivity to nutrient cycling and decomposition of organic matter, and they are excellent indicators of soil biological activity (Neher 2001; van den Hoogen *et al.* 2019; Delgado-Baquerizo *et al.* 2020). Their small size allows rapid dispersal and colonization through wind and / or water, still so far studies assessing colonization of nematodes after glacier retreat have generally focused on small areas (Lei *et al.* 2015; Devetter *et al.* 2021; Rosero *et al.* 2021). Understanding the factors affecting nematode colonization of deglaciated terrains at the global scale will provide key information on the formation of soil, on biotic communities that will follow, and on their long-term responses to climate change. Emerging molecular approaches, such as environmental DNA (eDNA) metabarcoding (Taberlet *et al.* 2018) enable rapid and cost-effective assessment of nematodes (Geisen *et al.* 2018; Delgado-Baquerizo *et al.* 2020) and provide consistent estimates of soil biodiversity, even over broad spatial scales (White *et al.* 2020). Furthermore, taxonomic data can be combined with available databases on functional traits to obtain estimates of functional diversity (Aglieri *et al.* 2020; White *et al.* 2020).

There is growing evidence that the impacts of climate change on biodiversity are uneven at the global scale (Freeman *et al.* 2021), and it has been proposed that the ecological responses to

glacier retreat are affected by local climatic conditions (Ficetola *et al.* 2021b). Macro-climatic differences in temperature and, to a lesser extent precipitation and solar radiation, influence the evolution of ecosystems after glacier retreat, with earlier ecosystem formation and colonization in relatively warm areas (Kaufmann 2002; Khedim *et al.* 2021). Different rates of colonization between glacier forelands under different climates can be predicted, but lack of global data has hampered testing this hypothesis.

Here, we tackle this issue by studying nematode colonization on multiple chronosequences of glacier retreat (Supplementary Fig. 1), from the Little Ice Age to recent, from five continents and from the equator to polar regions (Fig. 1), and investigate the variation of both taxonomic and functional diversity of communities. These are key drivers of ecosystem multifunctionality and allow a more complete overview of biotic responses (Le Bagousse-Pinguet *et al.* 2019). With this unique dataset, we assess if taxonomic and functional colonization are different in glacier forelands with different macro-climatic conditions, testing the hypothesis that milder temperatures favour the early development of biodiverse communities.

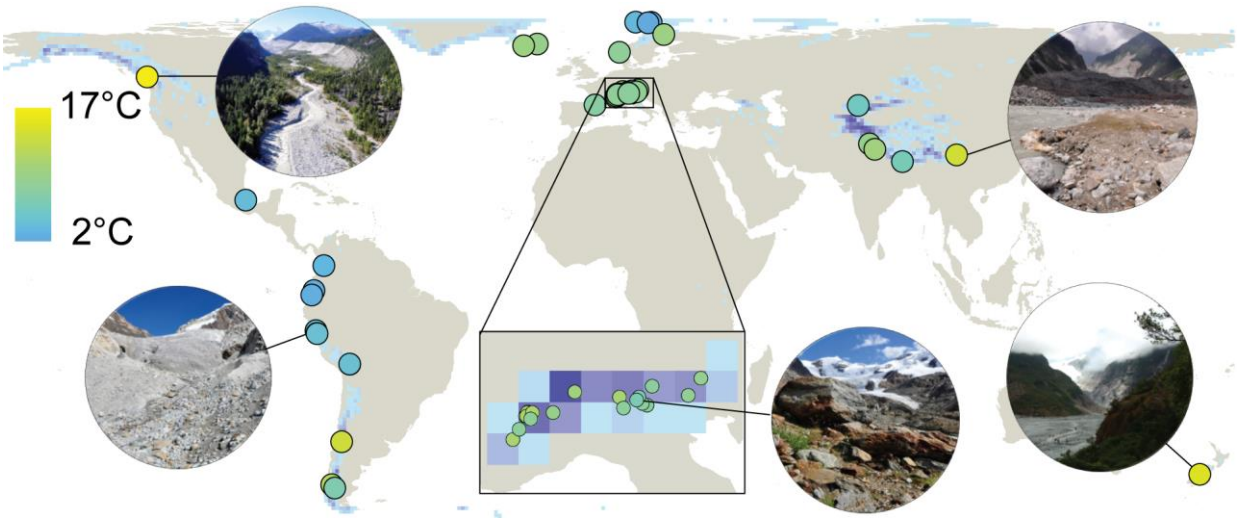


Figure 1. Global distribution of the studied glacier forelands. The colour of the dots represents the average temperature of forelands during the warmest quarter. The background blue grid represents the number of glaciers per each 1 x 1 degree cell according to the GLIMS database (www.glims.org) and ranges from 1 (pale blue) to 5,000 glaciers (darkest blue). Insets present pictures of a subset of forelands.

We obtained data from 240 dated sites (2-9 sites per chronosequence; average: 5) from 48 glacier forelands (Fig 1). Forelands with mild temperature (average temperature in the warmest quarter of the year > 7.8°C; hereafter: summer temperature) did not necessarily occur in tropical regions, but were located in areas ranging from the subtropics (e.g. South China) to relatively high latitudes (e.g. Canada). After data cleaning and clustering, DNA metabarcoding yielded a total of 3969 Molecular Operational Taxonomic Units (MOTUs) of eukaryotes, 172 of which were identified as nematodes, with a maximum of 22 nematode MOTUs per site. We assessed taxonomic and Faith's functional richness using the first-order jackknife estimator from field and PCR replicates, to limit underestimations of actual diversity. Estimated values of taxonomic richness suggest that the detected MOTUs represented about 71% of the actual richness (range of coverage: 56-100%), with an estimated diversity of 0 to 34 MOTUs per site. The coverage was high also for Faith's functional richness (83% of actual diversity values detected). For both richness variables, detected and estimated diversity values were strongly correlated (in all cases, Pearson's $r \geq 0.96$; Supplementary Fig. 2). Different measures of functional diversity (Faith's functional richness and Rao functional diversity accounting for MOTU relative abundance) were strongly intercorrelated, and also correlated with taxonomic diversity (all $r \geq 0.72$; Supplementary Table 1), whereas functional uniqueness was weakly correlated with taxonomic diversity and with the other measures of functional diversity ($r \leq 0.34$; Supplementary Table 1).

The taxonomic richness of nematodes strongly increased with time since glacier retreat and was significantly higher in forelands with mild summer temperature (Fig. 2a-b, Supplementary Table 2). Furthermore, we detected a strong interaction between time and temperature. In forelands with mild summers, nematodes were able to colonize the terrain in a short time, being already present four years after glacier retreat. On the other hand, in forelands with cold summer periods, taxonomic richness was essentially zero at all sites aged less than 15 years (Fig. 2a). Afterwards, taxonomic

richness generally increased with time, but the increase was relatively faster in cold forelands, so that the diversity was similar across all forelands 150 years after glacier retreat. Segmented regressions did not detect thresholds in the increase in taxonomic richness either in warm or cold summer environments (Supplementary Table 3). All results were robust to the inclusion of soil features (pH and total organic carbon) in the models (Supplementary Table 4). Highly consistent results were found irrespective of different choices of Hill's numbers (Supplementary Fig. 4).

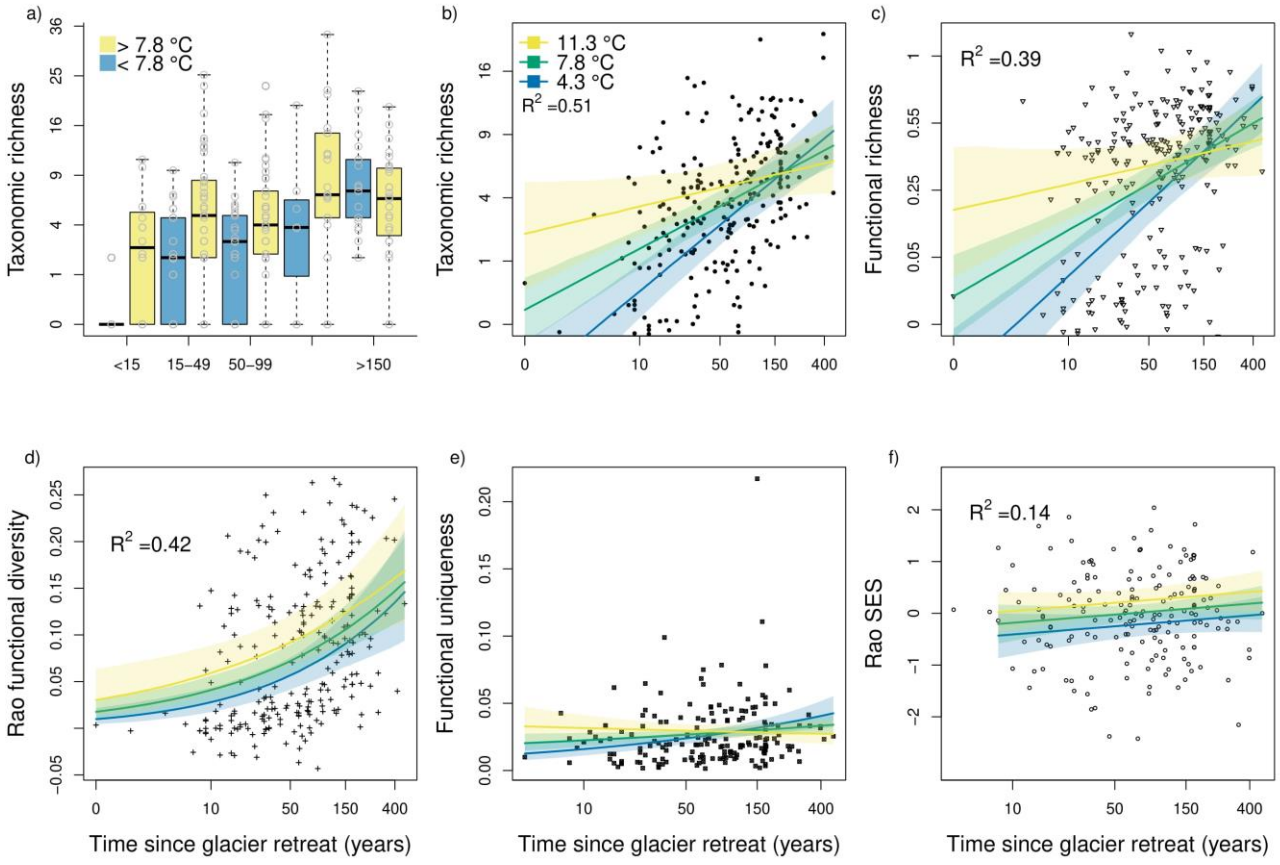


Figure 2. Changes in taxonomic and functional diversity of nematodes in deglaciated terrains with age. a): Taxonomic richness (first-order jackknife estimator). Blue boxes: forelands with mean temperature during the warmest quarter (summer temperature) below the average (7.8°C); Yellow boxes: forelands with summer temperature > 7.8°C. b-e): Relationship between different measures of taxonomic and functional diversity and time since glacier retreat, in forelands with different temperature conditions: 7.8°C is the average summer temperature across all the forelands; 4.3° and 11.3°C are the average ± one standard deviation. b): taxonomic richness; c): Faith functional richness; d): Rao functional diversity; e): functional uniqueness; f): standard effect sizes of Rao functional diversity. b-f represent conditional plots of Bayesian mixed models; partial residuals were calculated using fixed effects only; shaded areas are 95% credible intervals for forelands with different summer temperatures. In panels a-c, $N = 240$ communities. In d-e), $N = 177$ communities with observed richness > 0. Results for phylogenetic diversity are shown in Supplementary Fig. 3.

Similarly, values of functional diversity increased with time (Figure 2c-d; Supplementary Table 2) for both Faith and Rao indices. Again, functional diversity of soils exposed less than 150 years ago was consistently higher in forelands with mild summers. Time since glacier retreat also interacted with temperature in determining the occurrence of functionally unique species. In forelands with cold summers, functional uniqueness was very low immediately after glacier retreat, indicating that young communities were characterized by species sharing the same functional traits among them. In these cold forelands, functional uniqueness increases through time, reaching values similar to other forelands approx. 100 years after glacier retreat (Fig. 2e). All measures of taxonomic and functional diversity were unrelated to differences in precipitation across forelands.

We also evaluated whether taxonomic richness consistently increases within the different functional groups of nematodes, defined on the basis of their consumption traits. All consumption traits were already present soon after glacier retreat, and the richness of most of them (bacterivores, fungivores, omnivores, predators, parasites of plants and detritivores) increased through time ($R^2 >$ 0.2; Supplementary Figure 5, Supplementary Table 5). Bacterivores, fungivores and detritivores also showed higher richness in forelands with mild summers (Supplementary Table 5); we did not find support for an interaction between climatic features and time since glacier retreat (model with interactions: WAIC =2253.7, model without interactions: WAIC =2250.8).

We then attempted to identify the coexistence mechanisms affecting the functional assembly of communities. In so doing, we compared the observed values of Rao functional diversity to the ones obtained with random communities using null models. Under environmental filtering, communities are expected to be functionally homogeneous, with species harbouring specific traits that allow them to tolerate specific conditions. This coexistence mechanism would thus lead to a functional diversity lower than expected on the basis of observed taxonomic diversity. Conversely, under strong competitive interactions, which are reflected by a limiting similarity assembly process, colonization should start with a few groups that are functionally different (e.g. use different resources) (Weiher &

Keddy 1999; Götzenberger *et al.* 2012). We predict a stronger effect of environmental filtering in harsh conditions (e.g. very cold climates and/or soon after glacier retreat), whereas limiting similarity could be more important under milder conditions and/or in more developed communities. The observed variation of functional diversity was similar to what we would expect on the basis of taxonomic diversity [Rao Standardized Effect Sizes (SES) significant ($|SES| > 1.96$) only in 1.4% of sites], suggesting that the accumulation of different traits during colonization is mostly caused by the accumulation of new MOTUs. Nevertheless, the values of Rao SES were positively related to temperature and, to a lesser extent, to time since glacier retreat (Fig. 2f; Supplementary Table 2), supporting the hypothesis that environmental filtering limits the functional diversity of the coldest forelands for a few years after glacier retreat.

Finally, we quantified the functional changes of nematode communities along the chronosequences using multivariate models. We expected functional changes of communities through time, with a shift from small-sized colonizer species to large-bodied and more persistent species (Tom 1990). During colonization, communities changed their average life history and topological traits. Soon after glacier retreat, communities were dominated by small species with low colonizer-persister values (i.e. colonizer species with *r* ecological strategy) (Ferris *et al.* 2001). In later stages, we found more persister species with longer generation times and larger body size (Supplementary Table 6; Fig. 3a-d). For these traits, we found no relationships with climatic parameters nor an interaction between climate and time (model with interaction: WAIC=1156.9, model without interactions: WAIC=1146.3). Fitting communities in the functional space using PCoA (Fig. 3e-f) confirmed that, immediately after glacier retreat, communities are dominated by small-bodied and colonizer species, mostly feeding on bacteria, fungi, unicellular eukaryotes or plants, while at later stages communities are trophycally richer with more omnivores and predators, and are dominated by more persistent species with larger body size (Fig. 3e-f, Supplementary Table 7).

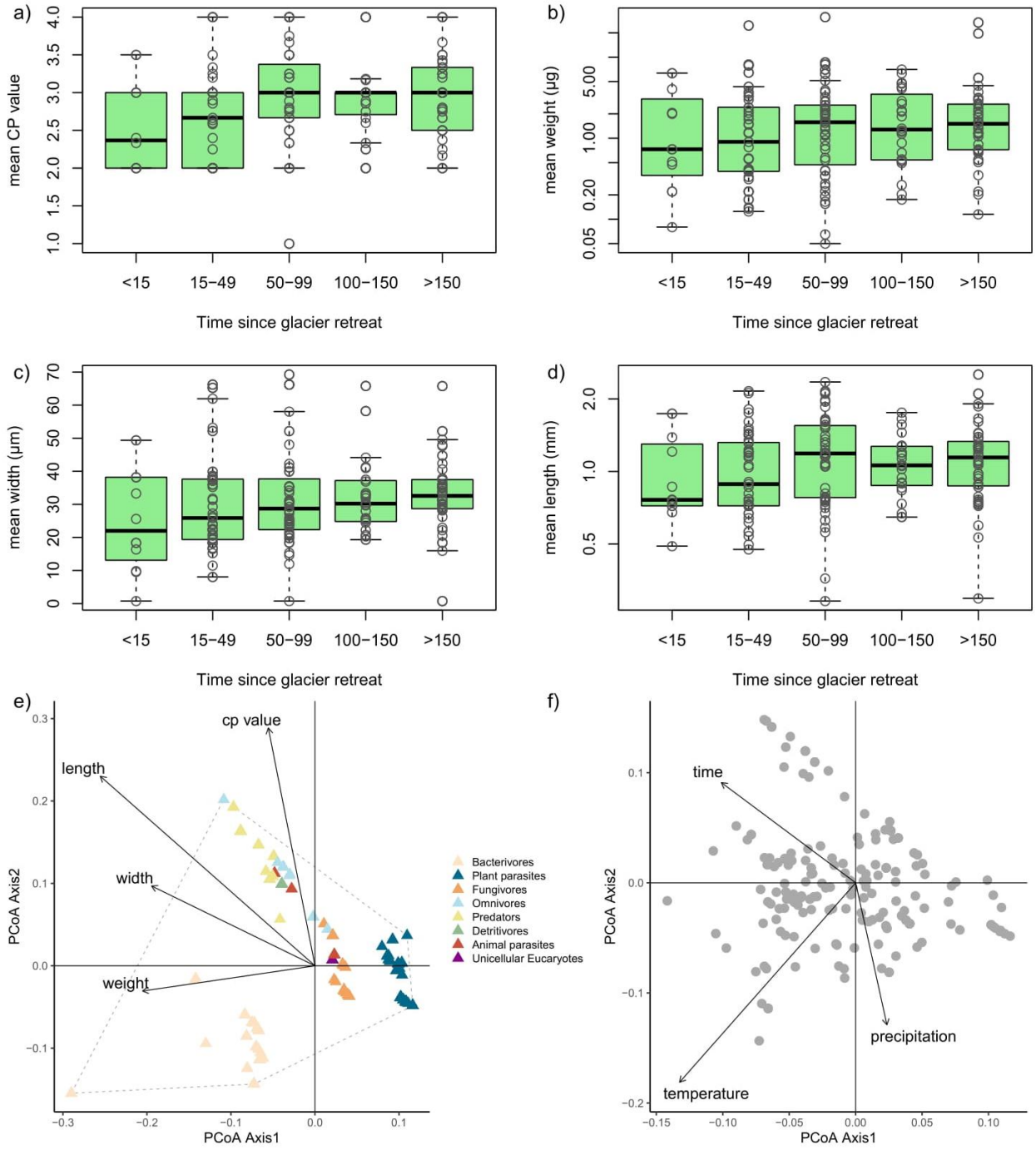


Figure 3. Changes in functional traits of nematode communities with nematode colonization. a-d: boxplots of traits in sites at different times after glacier retreat. e-f: Results of Principal Coordinate Analysis showing variation of functional traits across glacier forelands. *N* = 177 communities with observed richness > 0.

Understanding how different components of biodiversity develop under different climatic regimes and assessing the spatial variation of different functional groups in response to glacier retreat are pivotal to predicting the future of these emerging ecosystems. After glacier retreat, soil

colonization by keystone animals, such as nematodes, was affected by macro-climatic differences among forelands (Fig. 2), supporting the hypothesis that the retreat of glaciers impacts soil biodiversity in different ways depending on regional climates, which influence the response of ecosystems to climate change (Kaufmann 2002; Khedim *et al.* 2021).

Nematodes are excellent dispersers because of their small size (Fig. 3) and resistant propagules, and are able to quickly colonize recently deglaciated areas (< 15 years) (Lei *et al.* 2015; Fontaneto 2019). The colonization seems to start earlier in forelands with mild temperature during the growing season; in these terrains, we found high taxonomic and functional richness a few years after glacier retreat, whereas in cold areas 10-15 years may be required for soil ecosystems to develop enough for nematode communities to become established (Fig. 2a). Forelands in different climates have distinct trajectories, probably determined by the balance between colonization and competition, and after approximately 150 years even the coldest forelands achieve diversity values comparable to the warmest ones, where competition is probably stronger (Fig. 2) (Lei *et al.* 2015; Losapio *et al.* 2021). Our global sampling design allowed to tease apart the actual role of climatic conditions from that of biogeographical context, because foreland temperature also depends on their elevations, thus mild forelands occurred across a broad ranges of latitudes (Fig. 1). Therefore, the observed patterns are not biased by differences between tropical, temperate and boreal areas (Freeman *et al.* 2021) (Supplementary Fig. S6), suggesting that local climate is a better proxy of factors driving community building than latitude.

The biomass of below-ground animals, including nematodes, is tightly linked to the accumulation of organic matter into soil (Fierer *et al.* 2009; Lei *et al.* 2015; van den Hoogen *et al.* 2019), and recent studies have shown particularly fast increase of soil organic matter through time in forelands with mild summer temperature (Khedim *et al.* 2021). The complex interplay between climate and soil evolution probably affects all components of the biota, determining different patterns of responses to glacier retreat. The dispersal ability of nematodes allows their propagules to reach

deglaciated terrains from nearby sites where communities are already structured. This explains why nematodes belonging to most of the trophic groups are present from the early stages of colonization(Lei *et al.* 2015) (Supplementary Fig. 4). Nevertheless, we are clearly not in a neutral (Hubbell 2001) mechanism of assembly here as, after accounting for the effect of species richness, functional diversity remains higher in mild forefields. In other words, we observe a strong environmental filtering in cold climates and in recently deglaciated areas(Cauvy-Fraunié & Dangles 2019), where functional diversity was lower than expected by null models (Fig. 2f). Those results, however, might be affected by the fact that we only considered average conditions in each foreland, as we aimed at testing whether climatic differences across forelands determine divergent colonization patterns. Forelands are heterogeneous landscapes, and fine-scale studies evaluating the impact of micro-meteorological conditions (e.g. soil humidity and below-surface temperature) and soil features will be required to understand processes acting locally(Ficetola *et al.* 2021b). Still our results are robust, because the inclusion of key fine-scale features of soil (e.g. soil pH and organic carbon content) did not affect our overall conclusions (Supplementary Table 4).

Even though all the nematode guilds may be present soon after glacier retreat, the average traits of the nematode communities shifted through time (Fig. 3). The small-sized nematodes with a *r* reproductive strategy found soon after glacier retreat (Fig. 3 a-d) probably have high dispersal abilities and / or better fitness in cold environments (Ferris *et al.* 2001; Cauvy-Fraunié & Dangles 2019). With time, functional richness increases through the addition of more persistent species with larger body size, and the richness of many trophic groups (bacterivores, fungivores, omnivores, predators and plants parasites) increases, probably because of the growing number of available niches and the weaker environmental filtering. Overall, the increase in functionality mostly occurs through the addition of new functions, instead of the replacement of pioneers by more competitive species (Supplementary Fig. S7). This supports the hypothesis that, for soil animals, differences in colonization and tolerance are more important than competition in determining community turnover,

resulting in different succession patterns from the ones observed for plants (Gobbi *et al.* 2017; Ficetola *et al.* 2021b).

A key next step of research will be the development of multi-trophic analyses that evaluate the complex relationships between habitat evolution and different components of biodiversity (e.g. bacteria, fungi, plants, other animals). This will allow a better understanding of how climate change modifies habitats and trophic links in high-mountain environments and lead to a rapid development of new ecosystems (Martinez-Almoyna *et al.* 2019). Due to ongoing climate change, newly deglaciated terrains are expected to be increasingly important ecosystems in this century (Intergovernmental Panel on Climate Change 2021). Recognizing the complexity of biodiversity responses across climates will allow understanding how these environments evolve and to predict cascading effects at both global and local scale.

METHODS

Field activities. From 2014 to 2020, we collected 1197 soil samples in 48 glacier forelands, covering five continents from the equator to polar regions (Fig 1). In these forelands, information on the dates of glacier retreat since the end of the Little Ice Age is available from the literature, dated images and field surveys (Marta *et al.* 2021). For each foreland, we selected 2-9 sites along the chronosequences of glacier retreat (that is the chronological sequence of specific geomorphological features, such as moraines or inter-moraine flat areas, along deglaciated terrains for which the date of glacier retreat is known; Supplementary Fig. 1). A site was generally chosen at a line delimiting the position of the glacier's front on a given date. At each site, we established 3-10 plots (mean = 5, SD = 0.6), regularly spaced for a more repeatable biodiversity estimate (Dickie *et al.* 2018) and, if possible, at distances of 20 m. At each plot, we collected five soil subsamples within one meter and at a depth of 0–20 cm and pooled them together to form a composite sample of ~200 g per plot. We did not include soil litter and avoided roots, leaves and other large plant organs. Composite samples were homogenized;

from each composite sample we took 15 g of soil and desiccated it immediately in sterile boxes with 40 g of silica gel (Guerrieri *et al.* 2021). Desiccation is a cost-effective approach allowing long-term preservation of eDNA for biodiversity assessments (Guerrieri *et al.* 2021). Before the collection of each sample, all the sampling equipment underwent strict decontamination protocols (burned at >1000°C with a portable blow torch).

Molecular analyses. Environmental DNA from the 1197 soil samples was extracted in a dedicated room using the NucleoSpin Soil Mini Kit (Macherey-Nagel). We added a preliminary step where the soil was mixed with 20 ml of phosphate buffer for 15 min (Taberlet *et al.* 2012) and eluted in 150 µl. To control for contamination in the extraction room, we also included one extraction control every ~10 samples (total: 101 extraction controls) (Zinger *et al.* 2019). We used the Euka02 primers (Guardiola *et al.* 2015; Taberlet *et al.* 2018) to amplify a ~120 bp fragment of the V7 region of the 18S rDNA gene. This marker amplifies all the metazoans with very good performance, and its resolution often allows genus- or family level identification (Guardiola *et al.* 2015; Taberlet *et al.* 2018; Ficetola *et al.* 2021a). Primers included 8-nucleotide-long tags on the 5′ end. Each tag had at least five nucleotide differences with the others, thus allowing bioinformatic discrimination of PCR replicates after sequencing (Taberlet *et al.* 2018). DNA extracts were randomized in 96-well plates together with extraction controls, bioinformatic blanks (i.e. tagging-system controls), PCR negative and positive controls (total across all plates: 291 blanks, 90 negative and 53 positive controls). Positive controls consisted of ten-times diluted genomic DNA of eight bacterial and two fungal strains (ZymoBIOMICS Microbial Community DNA Standard II; Zymo Research). In eDNA metabarcoding-based analyses, extraction and PCR negative controls are pivotal to monitor contamination; blanks allow identification of tag-jump issues, and positive controls allow monitoring of potential cross-contamination of samples, amplification and sequencing performance (Zinger *et al.* 2019). We used quantitative PCR (qPCR) to determine the optimal number of PCR cycles. We randomly selected 48 DNA samples and used 2 µl of undiluted or 1:10 diluted DNA, and 1 µl of

1:1,000 diluted SYBR Green I nucleic acid gel stain (Invitrogen), with a real-time PCR thermal cycler set to standard mode. On the basis of qPCR results and for all samples, we performed 45 amplification cycles of 2 µl undiluted DNA in a 20-µl reaction volume with 10 µl of AmpliTaq Gold 360 Master Mix 2X (Applied Biosystems), 2 µl of primers mix (5 µM of each primer) and 0.16 µl of bovine serum albumin (Roche Diagnostic).

PCR amplification of samples was performed in 384-well plates and consisted of an initial step of 10 min at 95°C, followed by 30 s denaturation at 95°C, 30 s annealing at 45°C, 60 s elongation at 72°C, and 7 min final elongation at 72°C. All samples and controls underwent four PCR replicates (Ficetola *et al.* 2015). PCRs were performed in four distinct batches. All amplicons with a unique combination of forward and reverse tags within each batch were pooled. We used 5-µl aliquots of pooled amplicons to monitor the amplified fragment length and check for primer dimers using high-resolution capillary electrophoresis (QIAxcel Advanced System, Qiagen). Then, we purified six subsamples of the pooled amplicons using the MinElute PCR Purification Kit (Qiagen) following the manufacturer's protocol. Finally, we combined subsamples and sent them to Fasteris (Switzerland), where library preparation and sequencing were performed using the MetaFast protocol (Taberlet *et al.* 2018) and Illumina platforms (paired-end approach, 2x150 bp), respectively.

Bioinformatic treatment. We performed the bioinformatic treatment of sequence data using the OBITools software suite (Boyer *et al.* 2016). First, we used the illuminapairedend program to assemble forward and reverse reads and the ngsfilter program to assign sequences with an alignment score > 40 to the corresponding sample. For this step, we allowed two mismatches on primers and zero mismatches on tags. Then, we dereplicated sequences using the obiuniq program and filtered out those containing “N” or with an unexpected sequence length (e.g. <36 bp) and singletons. Subsequently, we used the obiclean program to keep sequences present in at least one PCR and that were at least twice as abundant as other related sequences differing by one base (hereafter "head sequences"). This step allowed us to remove PCR and sequencing errors. At this point, sequences

from different experiments were concatenated into one file and clustered at a threshold of 96% sequence similarity using the sumaclust program (https://git.metabarcoding.org/obitools/sumaclust/wikis/home). The 96% threshold was selected on the basis of preliminary bioinformatics analyses; this similarity threshold minimizes the risk that sequences attributed to the same species are clustered in different MOTUs (Bonin *et al.* 2021). Finally, we performed taxonomic assignment of sequences based on the EMBL reference database (version 140). The reference database was curated by carrying out an *in-silico* PCR with the ecopcr program (Ficetola *et al.* 2010). We used the same primers as in the experiment, allowing three mismatches per primer, and kept sequences assigned at least at the family level. Next, we assigned detected sequences to molecular operational taxonomic units (MOTUs) using the ecotag program, following the procedure described in (Boyer *et al.* 2016). ecotag matches each sequence in the dataset against the reference database and then uses a lowest common ancestor algorithm to identify the taxonomic level of the assignation (e.g. genus, family, order) (Boyer *et al.* 2016). We then performed additional filtering in R (version 4.0) to remove contaminants and tag-jump errors on the basis of sequences detected in controls and in blanks (Zinger *et al.* 2019). Specifically, we discarded MOTUs with best identity < 80%, detected less than ten times in the overall data set, or detected in at least one extraction or PCR-negative control (Zinger *et al.* 2019). We also removed MOTUs detected in only one sample, as they represent singletons, and MOTUs that were never found in at least two replicated PCRs per sample, as they are possible contaminants (Ficetola *et al.* 2015). For subsequent analyses, we only kept sequences assigned at the nematode phylum. See supplementary Table S8 for the number of sequences kept at each step in the analyses. The number of nematode MOTUs detected per site was unrelated to the number of DNA reads (Spearman's correlation: $r_s = 0.08$, $N = 240$, $P > 0.05$).

Assessing taxonomic and functional diversity. Due to the difficulty of relating the number of reads to the relative abundance of species in metabarcoding studies, we first measured taxonomic diversity

of PCR replicates as the total number of taxonomic units detected. We obtained functional traits of each MOTU on the basis of the nemaplex database (http://nemaplex.ucdavis.edu/), considering three topological traits (body length, body weight and body width), one life-history index [colonizer-persister value, ranging from 1 (colonizer species with short generation time and r reproductive strategy) to 5 (persistent species with long generation time and k strategy)] and eight consumption traits (parasites of animals, parasites of plants, bacterivores, fungivores, detritivores, omnivores, predators, feeding on unicellular eukaryotes) (Tom 1990; Ferris *et al.* 2001; Ferris 2010). The nemaplex database reports functional traits at genus-level resolution. Thus, we directly assigned traits only for MOTUs identified at the genus level; for MOTUs identified at the family level, traits were calculated as the average of the genera within the family.

The number of detected taxa generally underestimates the actual diversity of a given site, because some species often remain undetected, thus several approaches have been developed to estimate the actual number of taxa on the basis of replicated sampling of a given site (Colwell & Coddington 1994; Gotelli & Colwell 2011). We estimated biodiversity values with the first-order jackknife estimators for incidence data; this is among the best-performing approaches for evaluating the completeness of biodiversity inventories. Analyses of datasets of completely surveyed areas and simulations suggest that the approach can provide robust estimates of the actual species richness if a sufficiently large number of replicated samples is analysed (Colwell & Coddington 1994; Chazdon *et al.* 1998; Gotelli & Colwell 2011). To extrapolate biodiversity values at each site, we considered the different plots and the four PCRs for each plot as replicated samples. On average, we had 20 replicates per site (~5 plots with 4 PCR replicates each); this replication level is appropriate for biodiversity estimation in environments with moderate diversity levels (Gotelli & Colwell 2011). We estimated site-level taxonomic diversity using the alpha.accum function of the BAT R package (Cardoso *et al.* 2015). Hill numbers are an alternative approach to biodiversity measurement and are particularly useful in metabarcoding studies as they avoid issues related to excessive weighting of rare MOTUs and are extremely robust to different filtering strategies. Furthermore, Hill numbers > 0

take into account differences in read abundance that might be related to differences in biomass (Calderón-Sanou *et al.* 2020; Mächler *et al.* 2021). We therefore repeated the analyses of taxonomic diversity using Hill numbers (q = 0 and q = 1, equivalent to richness and Shannon diversity, respectively) to estimate taxonomic diversity; increasing q values indicate higher weight for the MOTUs with proportionally more reads. We could not perform this analysis using Hill numbers > 1 because at several sites we detected zero nematode MOTUs.

We used Faith diversity and Rao quadratic entropy to estimate functional diversity (Petchey & Gaston 2006), because approaches based on measures of the multidimensional trait space (e.g. N-dimensional hypervolumes) cannot be applied to communities with less than three functionally diverse taxa and we detected < 3 functionally diverse MOTUs at 61% of sites. For the Faith index, we first estimated pairwise trait dissimilarities using Gower distances, with equal weights for each trait (i.e., downweighting consumption traits based on the number of categories used: 1/8) (Laliberté & Legendre 2010). We then built a UPGMA dendrogram using functional distances for the full set of MOTUs and estimated diversity using the first-order jackknife estimator with the alpha.accum function (Cardoso *et al.* 2015).

The Rao index of quadratic entropy was computed based on the relative abundance of MOTUs in samples (proportion of reads) and trait dissimilarities between MOTUs following the script in ref. (De Bello *et al.* 2009). Using the proportion of reads can yield diversity measures that are less sensitive to parameters that are used for data filtering (Calderón-Sanou *et al.* 2020; Mächler *et al.* 2021). This index measures functional diversity by integrating information about functional richness and functional divergence (Mouchet *et al.* 2010). Finally, we evaluated functional uniqueness of communities using the R package funrar (Grenié *et al.* 2017). For each MOTU, we calculated its functional uniqueness index and averaged them across MOTUs present in each community. We calculated functional uniqueness as the functional distance of a focal MOTU from its nearest neighbour (Grenié *et al.* 2017) in the global functional space including all the studied MOTUs.

We used null-models to exclude the effects of taxonomic richness on functional diversity (Aros-Mualin *et al.* 2021), for all sites with at least one MOTU. We compared observed values of Rao index to ones obtained by randomizing MOTU identities 999 times, while keeping the number of detected MOTUs per community fixed. We then calculated standardized effect size (SES) values as the difference between the estimated values and the mean of randomly generated values of Rao functional diversity divided by the standard deviation of the 999 null values. Negative values of SES indicate that the functional diversity is lower than expected by chance given the observed taxonomic richness, whereas positive values indicate that functional diversity is higher than expected under random assembly (Gotelli & McCabe 2002).

In a preliminary analysis, we also assessed phylogenetic diversity (Supplementary Material, Supplementary Text). However, estimates of phylogenetic diversity showed an almost perfect correlation with taxonomic diversity ($r = 0.98$). The relationship between phylogenetic diversity, time since glacier retreat and climatic variables was identical to that observed for taxonomic diversity (see Supplementary Material Figure 3 and Table 1).

Drivers of changes in diversity and traits. We first used univariate Bayesian generalized mixed models (GMMs) to assess how time since glacier retreat and macroclimatic conditions determine nematode diversity. We ran mixed models with the different biodiversity metrics as dependent variables [taxonomic richness, Faith and Rao functional diversity, functional uniqueness, richness-independent functional diversity (SES)]. As independent variables we considered time since glacier retreat (log-transformed), and mean temperature and total precipitation (log-transformed) during the warmest season. These analyses were limited to forelands where we obtained nematode data for >2 sites. A main aim of this study was to assess whether colonization patterns are different between forelands in different climatic regimes, thus we considered the average conditions of all the sites within the foreland, calculated on the basis of ChelsaClim dataset (Karger *et al.* 2017). Mixed models also included interactions between climate and time since glacier retreat, to test the hypothesis that

climate affects colonization rate. We included the identity of glacier forelands as a random factor. Time and climatic variables were scaled (mean = 0, SD = 1) to allow comparison of their estimated effects. Models were run with three MCMC chains using 10,000 iterations and a burn-in of 5,000 in the brms R package (Bürkner 2018). Rao functional diversity and functional uniqueness were rescaled following ref. (Smithson & Verkuilen 2006) to avoid zeros and ones, and models were run using Beta distribution in brms. For all models, c-hat was <1.01, indicating convergence. Several soil characteristics, such as nutrient contents, can change along chronosequences and influence soil organisms (Ficetola *et al.* 2021b). To confirm that our conclusions were not biased by differences in soil features, we repeated analyses for a subset of sites (N = 163) for which two key soil features were available: total organic carbon and pH (ref. (Khedim *et al.* 2021) and Ficetola et al. unpublished data; see Supplementary Table 4).

We also used segmented regression to assess whether there are thresholds in the relationships between taxonomic richness and time since glacier retreat (Muggeo 2003; Ficetola & Denoël 2009). We used maximum likelihood to build linear mixed models with one breakpoint (segmented package in R)(Muggeo 2003). We then compared models with the breakpoint with linear mixed models without the breakpoint on the basis of the Bayesian Information Criterion (BIC). As analyses suggested that the relationship between time since glacier retreat and diversity are affected by local climatic conditions (Fig. 2), we analysed separately glaciers with summer temperature above the mean and those with summer temperature below the mean, with 7.8°C being the mean summer temperature of forelands in our dataset.

We next built two multivariate generalized mixed models to assess the impact of time and climatic variables on 1) the richness of MOTUs belonging to eight trophic groups, defined on the basis of MOTU consumption traits and 2) the average traits of communities (colonizer-persistent value and topological traits). In the case of consumption traits, the number of MOTUs with each consumption trait at each site was used in the model as a dependent variable; we used a negative

binomial distribution and considered all sites. For the colonizer-persistent value and for each topological trait (body length, body width and weight), we averaged the values of all of the MOTUs present in a given community, and considered them as dependent variables; we used a Gaussian error and transformed morphological features using logarithms (weight and body length) or square-root (body width) to attain normality. The model for average community traits was limited to communities with at least one detected MOTU. Multivariate models included glacier identity as a random factor and were run using the same parameters as univariate GMMs. We used the widely applicable information criterion (WAIC) to compare models with and without interactions (Gelman *et al.* 2014).

To further assess the variation of functional traits, we visualized trait and community distributions in the functional space. We built a multivariate functional space by ordinating MOTUs on the basis of Gower trait distances using Principal Coordinates Analysis (PCoA). Then, we built a global functional space with the first two axes of the PCoA (Borcard *et al.* 2011; Oksanen *et al.* 2019). We then used the *envifit* function from the vegan package to fit variables onto the PCoA. A determination coefficient (R^2) was calculated to assess the strength of the correlations of the axes with traits, as well as time and climatic variables. Variables having high R^2 correspond to strong predictors of the ordination axis. The significance of R^2 values was assessed by comparing them with R^2 values obtained from 999 random permutations of the data. Continuous variables were transformed into vectors with directions representing the correlation with the axes and length proportional to the strength of the correlation between the axis and the variables. For categorical variables, we computed average ordination scores for each category of the variables to locate the different categories within the functional space. All statistical analyses were performed in the R environment (version 4.0; www.r-project-org).

Finally, to determine if functional changes are caused by the addition of new traits or by a replacement of traits, we partitioned the functional beta-diversity among sites within the same glacier foreland using the *beta* function in the package BAT (Cardoso *et al.* 2015). When using abundance

data, this function partitions total beta diversity (FBD-total) into the beta diversity explained by replacement of species alone (FBD-Repl) and beta diversity explained by abundance loss/gain (FBD-rich) alone. We calculated functional diversity based on an UPGMA dendrogram built by clustering Gower functional distances between MOTUs.

Supplementary Text for

Local climates modulate soil biodiversity dynamics after glacier retreat

Alessia Guerrieri, Isabel Cantera, Silvio Marta, Aurélie Bonin, Roberto Ambrosini, Marco Caccianiga, Fabien Anthelme, Roberto Sergio Azzoni, Peter Almond, Pablo Alviz Gazitúa, Sophie Cauvy-Fraunié, Jorge Luis Ceballos Lievano, Pritam Chand, Milap Chand Sharma, John Clague, Justiniano Alejo Cochachín Rapre, Chiara Compostella, Rolando Cruz Encarnación, Olivier Dangles, Philip Deline, Andre Eger, Sergey Erokhin, Andrea Franzetti, Ludovic Gielly, Fabrizi Gili1, Mauro Gobbi, Sigmund Hågvar, Norine Khedim, Gwendolyn Peyre, Francesca Pittino, Angela Proietto, Antoine Rabatel, Nurai Urseitova, Yan Yang, Vitalii Zaginaev, Andrea Zerboni, Anaïs Zimmer, Pierre Taberlet, Guglielmina Adele Diolaiuti, Jerome Poulenard, Diego Fontaneto, Wilfried Thuiller, Gentile Francesco Ficetola

ANALYSES OF PHYLOGENETIC DIVERSITY

DNA metabarcoding produces very short sequences (< 150 bp in our study case), that are often deemed insufficient for accurate estimates of phylogenetic diversity. Nevertheless, the obtained sequences often matched the published ones with limited similarity, and several sequences matched more than one genus, thus it was impossible estimating phylogenetic diversity on the basis of published phylogenies. We thus relied on the obtained sequences, by first evaluating if they provide phylogenetic reconstructions that are consistent with the known relationships between taxa. Nematode sequences were aligned using MAFFT v7 with Q-INS-i settings to account for the secondary structure of ribosomal markers (Katoh et al. 2019); we obtained a 151 bp alignment. The phylogenetic reconstruction was obtained from the alignment with Bayesian Inference using BEAST v2.5 (Bouckaert et al. 2019) through the CIPRES online portal (Miller et al. 2010). We obtained an ultrametric rooted tree, with terminals equidistant from the root and with edge lengths proportional to evolutionary rate and time, on a relative scale. Parameters in BEAST were set as default, except for GTR+I+G as evolutionary model, estimated random local clock, and 50,000,000 generations with a burnin of 20%. The obtained three was highly consistent with the known taxonomic relationships between identified taxa, the clades showed good support and the effective sample size of parameters was always >1000, thus we assumed that the obtained tree provides an acceptable approximation of the relationships between these taxa (Alberdi and Gilbert 2019). We then calculated Faith's phylogenetic diversity of each site as the total branch length of the phylogenetic tree linking all MOTUs represented in a given community, using the alpha.accum function in BAT (Petchey and Gaston 2006, Cardoso et al. 2015). The obtained correlation between phylogenetic and taxonomic diversity was extremely high (r = 0.98; Supplementary Table 1). The relationship between

phylogenetic diversity, time since glacier retreat and climatic variables was identical to the one observed for taxonomic diversity (Supplementary Figure 3).

REFERENCES

Alberdi. A. and Gilbert. M. T. P. 2019. A guide to the application of Hill numbers to DNA-based diversity analyses. — Mol. Ecol. Resour. 19: 804-817.

Bouckaert. R. et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. — PLoS Comput. Biol. 15:

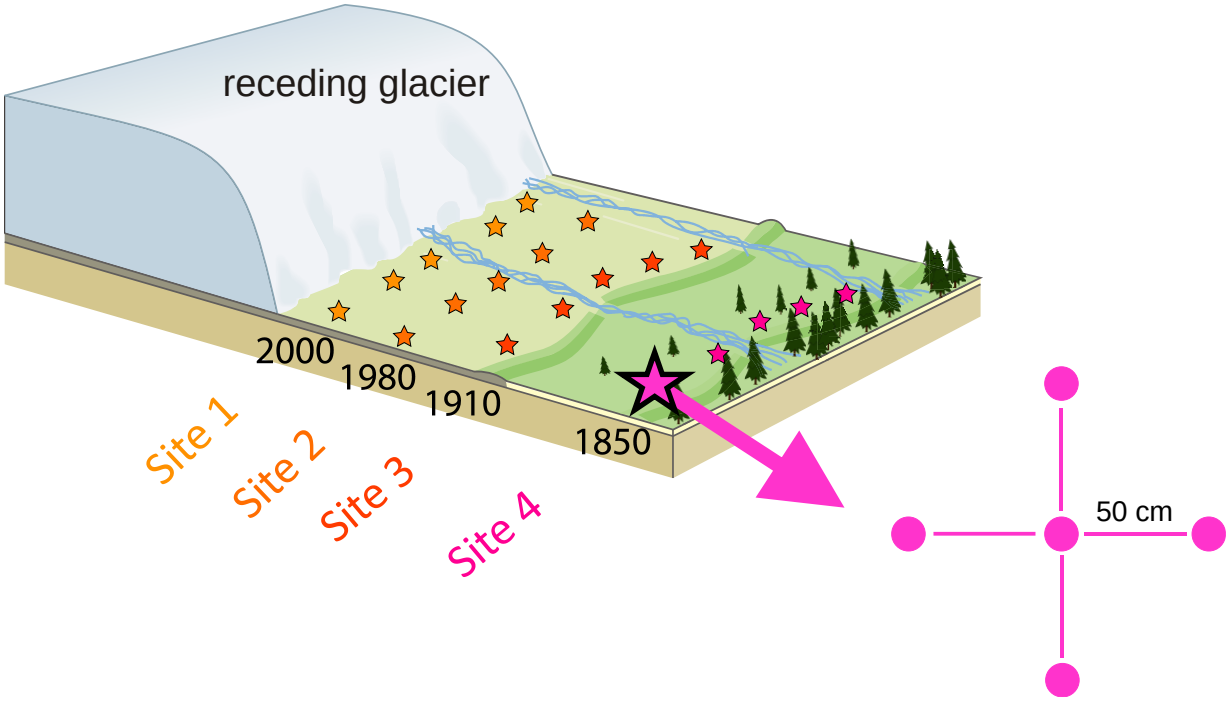
Cardoso. P. et al. 2015. BAT – Biodiversity Assessment Tools. an R package for the measurement and estimation of alpha and beta taxon. phylogenetic and functional diversity. — Methods Ecol. Evol. 6: 232-236.

Katoh. K. et al. 2019. MAFFT online service: multiple sequence alignment. interactive sequence choice and visualization. — Briefings in Bioinformatics 20: 1160-1166.

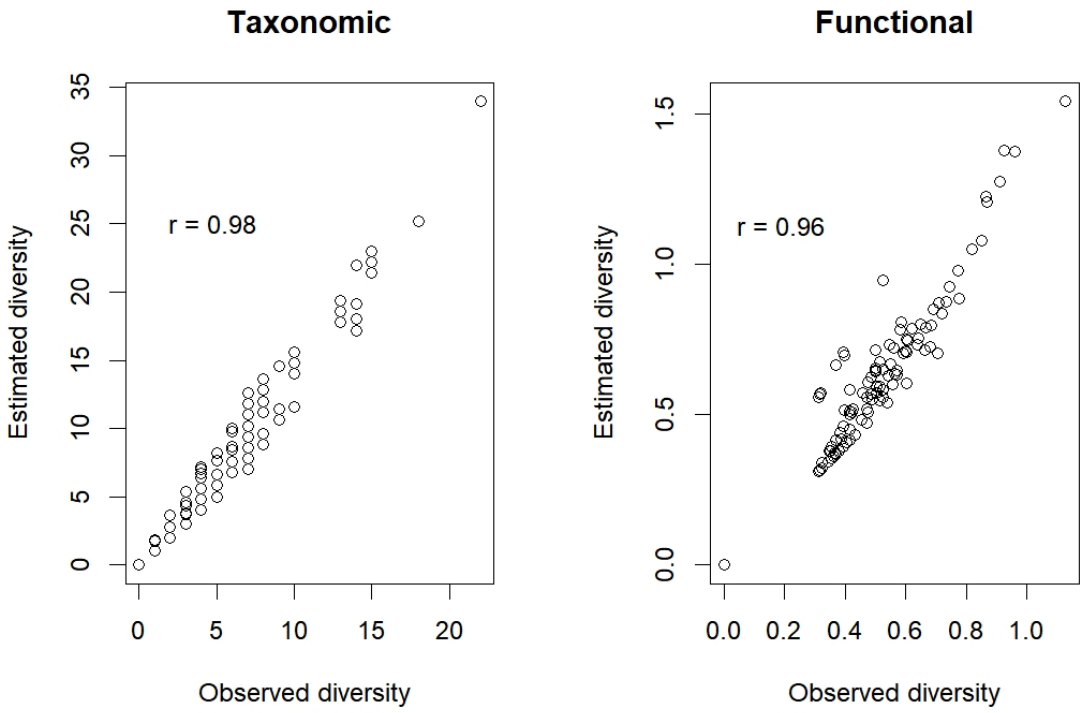
Miller. M. A. et al. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE). pp. 1-8.

Petchey. O. L. and Gaston. K. J. 2006. Functional diversity: back to basics and looking forward. — Ecol. Lett. 9: 741-758.

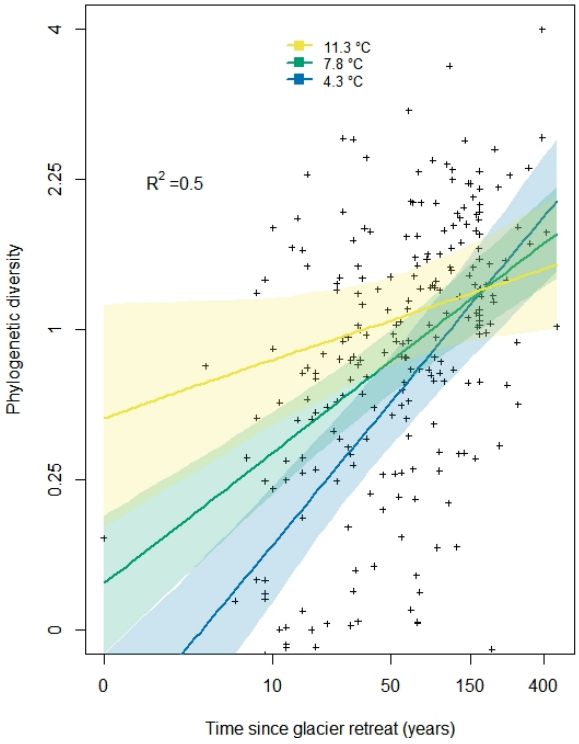
SUPPLEMENTARY FIGURES



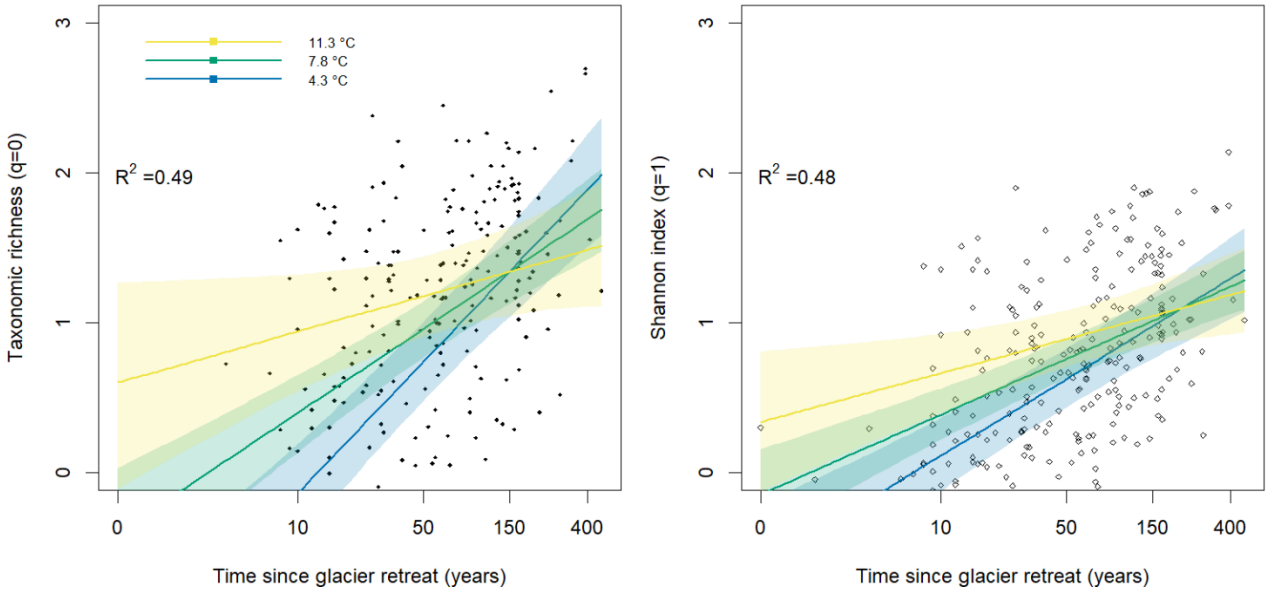
Supplementary Figure 1. Sampling scheme. For each receding glacier, we identified 2-9 sites, representing the past position of the glacier at known dates. For each site, we established ~5 plots, at regular intervals of 20 m. Within each plot, we collected 5 soil subsamples within 1 m^2, the distribution of subsamples is shown by the pink inset. The five sub-samples were then pooled to form one composite sample per plot.



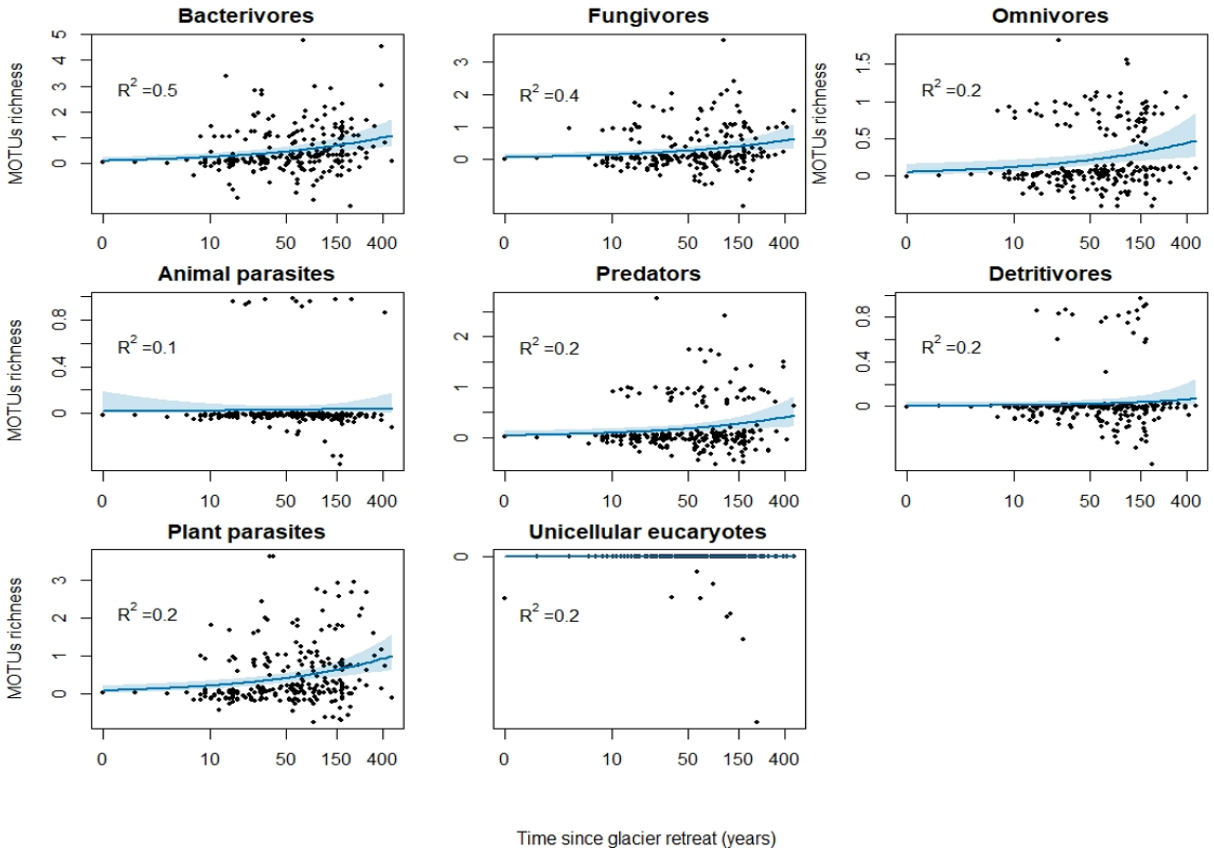
Supplementary Figure 2. Pearson's product moment correlations (r) between observed and estimated diversity. Taxonomic and Faith functional diversities were estimated with the first-order jackknife estimators for incidence data (see methods).



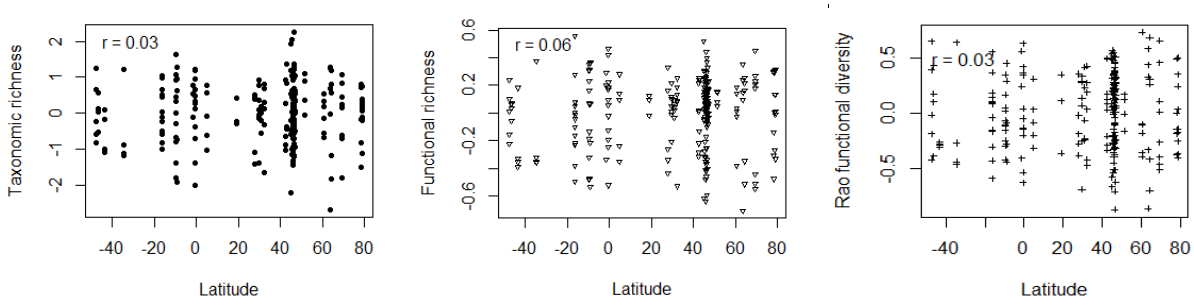
Supplementary Figure 3. Changes in phylogenetic diversity of nematodes in deglaciated terrains with different age. Conditional plot of Bayesian mixed model; shaded areas are 95% credible intervals for forelands with different summer temperature; N = 177 communities with observed richness > 0.



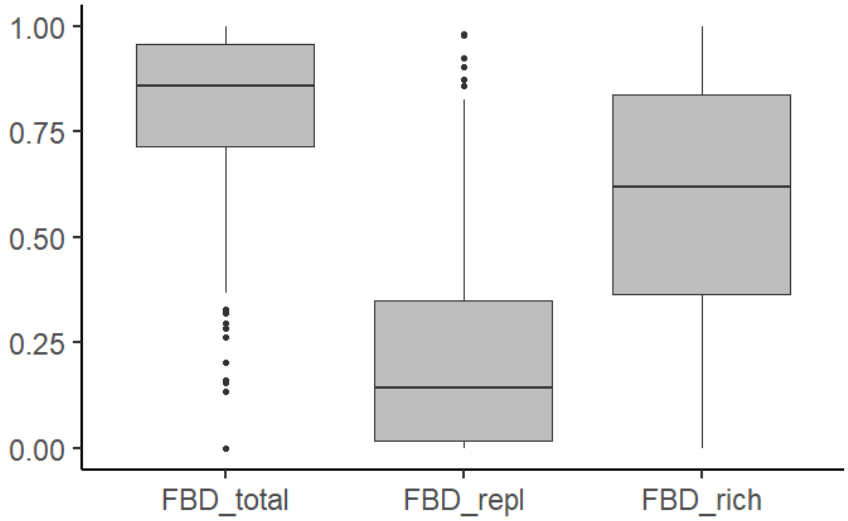
Supplementary Figure 4. Changes in taxonomic diversity of nematodes in deglaciated terrains with different age. Here taxonomic diversity is estimated using different Hill's numbers (q = 0 and q = 1; diversity estimates are log-transformed to improve normality). We show the conditional plots of Bayesian mixed model; shaded areas are 95% credible intervals for forelands with different summer temperature; N = 240 communities. Note that Hill's numbers > 1 cannot be computed because in several sites we detected zero nematode MOTUs.



Supplementary Figure 5. Variation of the richness of eight functional groups of nematodes through time along chronosequences in deglaciated terrains. We show the conditional plots of Bayesian mixed model; shaded areas are 95% credible intervals; N = 240 communities.



Supplementary Figure 6. Relationship between latitude of sampled sites and the residuals of: taxonomic diversity, Faith functional richness; Rao functional diversity.



Supplementary Figure 7. Partitioning of functional beta-diversity among sites within glacier forelands. Total beta diversity (FBD-total) is decomposed into the beta diversity explained by replacement of species alone (FBD-Repl) and beta diversity explained by abundance loss/gain (FBD-rich) alone.

SUPPLEMENTARY TABLES

Supplementary Table 1. Pearson's product moment correlations (r) between all diversity metrics. Significant correlations are in bold; for all the significant correlations, $P < 0.01$.

	Phylogenetic Diversity	Faith Functional Diversity	Rao Functional Diversity	Functional Uniqueness	Rao SES
Taxonomic Diversity	**0.98**	**0.84**	**0.79**	**0.33**	0.09
Phylogenetic Diversity		**0.85**	**0.77**	**0.33**	0.08
Faith Functional Diversity			**0.72**	**0.34**	0.08
Rao Functional Diversity				**0.29**	**0.69**
Functional Uniqueness					0.00

Table 2. Results of Bayesian univariate models assessing the relationship between nematode biodiversity [Taxonomic, Faith and Rao Functional diversity, Functional Uniqueness and Rao's standard effect sizes (SES)] and time since glacier retreat, precipitation and temperature. Parameters with 95% confidence intervals non-overlapping zero are highlighted in bold. Interactions were not included in the model for Rao SES as the model without interactions showed lower WAIC than the model with interactions (361 vs 366).

Dependent variable	Independent variables	*B*	Q2.5	Q97.5
Taxonomic diversity	Time since glacier retreat	**0.444**	**0.320**	**0.568**
	Summer temperature	**0.271**	**0.035**	**0.531**
	Summer precipitation	0.066	-0.157	0.303
	Time x Temperature	**-0.246**	**-0.375**	**-0.107**
	Time x Precipitation	-0.070	-0.202	0.068
Faith functional diversity	Time since glacier retreat	**0.113**	**0.077**	**0.148**
	Summer temperature	**0.062**	**0.003**	**0.123**
	Summer precipitation	0.024	-0.031	0.081
	Time x Temperature	**-0.068**	**-0.108**	**-0.030**
	Time x Precipitation	-0.012	-0.051	0.027
Rao Functional diversity	Time since glacier retreat	**0.400**	**0.270**	**0.530**
	Summer temperature	**0.250**	**0.030**	**0.470**
	Summer precipitation	0.130	-0.070	0.350
	Time x Temperature	-0.080	-0.210	0.050
	Time x Precipitation	-0.050	-0.180	0.090
Functional uniqueness	Time since glacier retreat	**0.110**	**0.010**	**0.200**
	Summer temperature	0.060	-0.050	0.170
	Summer precipitation	0.020	-0.090	0.140
	Time x Temperature	**-0.140**	**-0.240**	**-0.050**
	Time x Precipitation	-0.050	-0.160	0.060
Rao SES	Time since glacier retreat	0.100	-0.050	0.240
	Summer temperature	**0.230**	**0.060**	**0.400**
	Summer precipitation	0.040	-0.120	0.210

Supplementary Table 3. Comparison between segmented and linear mixed models, assessing relationships between time since glacier retreat and taxonomic richness. The models with breakpoint showed Bayesian Information Criterion (BIC) consistently higher than the respective models without the breakpoint, indicating weak support for the breakpoint.

	Forelands with T > 7.8°C	Forelands with T ≤ 7.8°C
Model with breakpoint (BIC)	445.5	296.2
Model without breakpoint (BIC)	438.3	293.3

Supplementary Table 4. Results of Bayesian univariate models with the inclusion of soil features (total organic carbon and pH) assessing the variation of Taxonomic and Functional diversity of nematode communities (R^2 = 0.5 and R^2 = 0.4, respectively). Parameters with 95% confidence intervals non-overlapping zero are highlighted in bold.

Dependent variable	Independent variables	*B*	Q2.5	Q97.5
Taxonomic diversity	Time since glacier retreat	**0.3431**	**0.1615**	**0.5254**
	Summer temperature	0.1224	-0.1993	0.4343
	Summer precipitation	0.0646	-0.2902	0.4264
	Soil organic C	0.0214	-0.2272	0.2927
	Soil pH	-0.1757	-0.4136	0.0766
	Time x Temperature	**-0.3153**	**-0.4921**	**-0.1396**
	Time x Precipitation	-0.0711	-0.2451	0.1014
	Time x Soil organic C	-0.0269	-0.1555	0.1016
	Time x pH	0.0521	-0.1021	0.2062
Faith's functional diversity	Time since glacier retreat	**0.1095**	**0.0548**	**0.1639**
	Summer temperature	0.0090	-0.0730	0.0894
	Summer precipitation	0.0379	-0.0526	0.1228
	Soil organic C	-0.0160	-0.0825	0.0524
	Soil pH	-0.0511	-0.1176	0.0164
	Time x Temperature	**-0.1019**	**-0.1565**	**-0.0483**
	Time x Precipitation	-0.0080	-0.0619	0.0447
	Time x Soil organic C	0.0062	-0.0322	0.0444
	Time x pH	0.0016	-0.0453	0.0499

Supplementary Table 5. Results of Bayesian multivariate models assessing the variation of consumption traits of nematode communities. Parameters with 95% confidence intervals non-overlapping zero are highlighted in bold. Interactions were not included in the models as the model without interactions showed lower WAIC (2253.7 vs 2250.8).

Dependent variable	Independent variables	*B*	Q2.5	Q5	Q95	Q97.5
Bacterivores	Time since glacier retreat	**0.399**	**0.223**	**0.250**	**0.550**	**0.577**
	Summer temperature	**0.462**	**0.123**	**0.180**	**0.750**	**0.807**
	Summer precipitation	0.189	-0.167	-0.110	0.490	0.547
Fungivores	Time since glacier retreat	**0.396**	**0.174**	**0.210**	**0.590**	**0.626**
	Summer temperature	**0.368**	**0.020**	**0.080**	**0.660**	**0.730**
	Summer precipitation	0.058	-0.298	-0.240	0.380	0.447
Omnivores	Time since glacier retreat	**0.375**	**0.108**	**0.150**	**0.610**	**0.657**
	Summer temperature	0.296	-0.032	0.020	0.570	0.633
	Summer precipitation	0.204	-0.134	-0.080	0.500	0.558
Parasites of animals	Time since glacier retreat	0.151	-0.483	-0.380	0.720	0.842
	Summer temperature	0.190	-0.507	-0.390	0.770	0.907
	Summer precipitation	-0.101	-0.842	-0.700	0.510	0.649
Predators	Time since glacier retreat	**0.422**	**0.132**	**0.180**	**0.680**	**0.725**
	Summer temperature	-0.254	-0.638	-0.570	0.040	0.097
	Summer precipitation	-0.258	-0.630	-0.560	0.030	0.095
Detritivores	Time since glacier retreat	0.558	-0.016	0.070	1.090	1.208
	Summer temperature	**0.960**	**0.138**	**0.260**	**1.750**	**1.949**
	Summer precipitation	0.381	-0.482	-0.340	1.170	1.365
Parasites of plants	Time since glacier retreat	**0.409**	**0.203**	**0.240**	**0.590**	**0.622**
	Summer temperature	0.165	-0.097	-0.050	0.380	0.436
	Summer precipitation	0.173	-0.102	-0.050	0.410	0.456
Feeding on Unicellular eucaryotes	Time since glacier retreat	-0.249	-1.408	-1.170	0.670	0.880
	Summer temperature	0.739	-1.170	-0.810	2.520	3.124
	Summer precipitation	0.921	-1.027	-0.710	2.930	3.470

Supplementary Table 6. Results of Bayesian multivariate models assessing the variation of life history and topological average traits of nematode communities. Parameters with 95% confidence intervals non-overlapping zero are highlighted in bold. Interactions were not included in the models as the model without interactions showed lower WAIC (1146.3 vs 1156.9).

Dependent variable	Independent variables	*B*	Q2.5	Q5	Q95	Q97.5
Colonization /Persistence	Time since glacier retreat	**0.142**	**0.055**	**0.069**	**0.215**	**0.230**
Value	Summer temperature	0.018	-0.080	-0.065	0.097	0.114
	Summer precipitation	-0.049	-0.140	-0.124	0.029	0.044
Weight	Time since glacier retreat	**0.194**	**0.019**	**0.045**	**0.342**	**0.372**
	Summer temperature	-0.039	-0.230	-0.200	0.121	0.156
	Summer precipitation	-0.014	-0.200	-0.169	0.139	0.166
Body width	Time since glacier retreat	**0.213**	**0.024**	**0.056**	**0.364**	**0.389**
	Summer temperature	-0.005	-0.215	-0.177	0.168	0.201
	Summer precipitation	-0.101	-0.290	-0.262	0.055	0.088
Body length	Time since glacier retreat	0.058	-0.001	0.010	0.106	0.117
	Summer temperature	-0.027	-0.091	-0.080	0.026	0.036
	Summer precipitation	-0.020	-0.080	-0.069	0.032	0.042

Supplementary Table 7. Results of Principal Coordinates Analysis assessing the variation of functional traits in the functional space. Contribution of each continuous trait to the PCoA 1 and PCoA 2 axes of the functional space of nematodes communities. The table shows the determination coefficient R^2 of the correlation between each trait and the ordination, *P* values based on random permutations of the data indicate if observed R^2 are higher than R^2 with randomly permuted data and the direction cosines of the continuous variables on the PCoA axes.

Variable	Axis 1	Axis 2	*P*	R^2
Consumption trait	NA	NA	0.001	0.67
Life history and topological traits:				
Colonization /Persistence value	-0.72	0.7	0.001	0.73
Width	-0.86	-0.5	0.001	0.45
Weight	-0.57	-0.82	0.001	0.45
Length	-0.99	-0.11	0.001	0.85

Supplementary Table 8. Number of sequences retained at different steps of bioinformatic treatment and data filtering.

	Number of reads	Number of different sequences
Raw sequences	169 700 440	
Kept after alignment	163 481 209	
Kept after ngsfilter	124 634 214	
Kept after filtering	97 845 376	72437
Kept after clustering	97 845 376	16033
Kept after removal of contaminants and of sequences detected in just one sample	30 473 982	3969
Referred to Nematoda	1 550 956	172

CHAPTER 6. CONCLUSIONS

This thesis represents both a step toward a better integration of recently-developed molecular tools in ecology research, and a step toward a better understanding of the dynamic of communities in deglaciated areas, as it operated in two directions. The first direction is a methodological advancement of the eDNA metabarcoding technique. Despite eDNA is transforming our ability to measure biodiversity, standardized protocols are needed to allow comparison of datasets and to encourage this technique as an environmental surveying tool for government regulation. The second direction is an advancement in our ecological understanding of community dynamics in response to climate change. Today, an important topic for ecology and global change studies is to understand the consequence of glacier retreat and to develop scenarios on how mountain environments may change over the next century. Recently-developed molecular tools have opened a new window on the invisible, providing information on the formation of multitrophic communities. Based on these aspects, this thesis answered the following scientific questions:

1) how do methodological choices such as soil preservation protocols and data curation procedures impact the reliability and reproducibility of metabarcoding data and what should the scientific community do to avoid such problems?

2) how do different taxonomic groups colonize recently deglaciated areas through time and do they respond in a similar way depending on the soil depth they are associated with?

3) how do climatic differences between areas of the world influence the colonization of recently deglaciated areas and what is the impact on both taxonomic and functional diversity?

As for any emerging tool, many studies have been published so far in order to consider the many problems that are likely to occur at any step of the analysis and to ensure a correct use of the information stored in eDNA, thus proving a great improvement of this technique already. However, several methodological questions remain unaddressed, and a better understanding of the

consequences of each methodological choice applied to eDNA metabarcoding on the reliability and reproducibility of data is needed.

Chapter 2 answered the first question by showing how several preservation approaches for soil samples affect estimates of taxonomic richness and community composition of different taxonomic groups. The study showed that soil metabarcoding data are surprisingly robust to preservation conditions, as limited differences in community structure and diversity estimates were observed when samples were preserved using different strategies. Nevertheless, some taxonomic groups and diversity components are more sensitive than others to certain preservation conditions. Thus, the choice should be adapted to the research goal and to the focal taxa, but also depends on the field constrains, especially if remote and / or difficult-to-access areas are investigated. Furthermore, the study underlines the strong interplay between the different steps of the eDNA metabarcoding workflow (**Chapter 1, Fig. 3**). Soil preservation is only one of the many methodological issues that can affect the outcome of metabarcoding analyses. For instance, in the following works of this thesis, eDNA was recovered from large volumes of soil (i.e. 15 g) in order to reduce sample heterogeneity (Taberlet et al., 2012), preserved using a desiccation strategy and extracted following a phosphate-buffer approach (Taberlet et al., 2012), thus mostly containing extracellular DNA. This strategy might result in lower DNA yield for certain groups in spite of others but congruent community composition, while others have used different extraction methods from smaller sample volumes and found higher biodiversity estimates for prokaryotes (Dopheide et al., 2019). In this context, it is worth to reminding that studies based on eDNA metabarcoding require a combined use of different biodiversity metrics to draw sounds ecological conclusions and a trade-off between each step of the workflow is always needed, especially when a multi-taxa approach is used. Furthermore, a qPCR assay was conducted using a subset of samples in order to determine the optimal number of PCR amplification cycles and DNA extract dilution to be used for each of the primers set. This allowed to ensure DNA of each taxonomic group to be properly amplified, to avoid over-amplification of template DNA and to limit chimera formation. Importantly, the same methodology should be consistently applied across

taxonomic groups, allowing the internal comparison of biodiversity among samples and studies. Overall, even though methodological choices at the sample preservation or extraction steps could partially limit the accuracy of some of data, thus reducing our ability to answer certain research questions (depending on the studied taxa), sometimes an opportune solution can be found by adapting the methodological choices at the subsequent steps. For example, at the data curation step, the choice of removing rare MOTUs from the dataset increases its strength by allowing to answer specific questions in spite of the biases produced by the previous choice made on the preservation strategy (**Chapter 2, Fig. 5**).

For this reason, **Chapter 3** analysed another important step of eDNA metabarcoding studies, the sequence clustering approach, which is routinely used for the identification of MOTUs during the bioinformatic pipeline. Traditionally, sequence clustering approaches rely on similarity values between the obtained sequences, and often on common practices rather than scientific arguments to define the clustering thresholds. This study showed that different thresholds can be selected depending on the studied markers and on the criterion favored by researchers (e.g. a trade-off between over-merging and over-splitting of sequences) to attain research goals. For instance, low similarity values are generally appropriate for specialist markers (e.g. metabarcodes amplifying only one class), or if we want to avoid overestimating the values of alpha-diversity. In the following works of this thesis, clustering thresholds represented the intersection between the within-species and within-genus probability distributions (**Chapter 3, Fig. 2**), thus resulting in a balance between over-merging and over-splitting of sequences. Despite sequence clustering thresholds can have strong effects in the estimates of MOTUs richness, also in this case it is important to remember that the interpretation of data remains relatively robust if subsequent statistical analyses rely on appropriate parameters to measure alpha and beta diversity (e.g., Hill's numbers with $q > 0$; Clare et al. 2016, Roy et al. 2019, Calderón-Sanou et al. 2020, Mächler et al. 2021).

In recent years, the use of eDNA metabarcoding technique has spread among the scientific community especially dedicated (but not only) to ecology, thanks to its capability of providing a huge

amount of information about all types of organisms, and thus to its potential role for enhancing biodiversity and monitoring research. Despite recent advances in our understanding of the drivers and rates of community assembly, many important questions remain unanswered because most studies performed so far only considered one or a few taxonomic groups and had limited geographic focus. Different organisms can have very different responses (Cauvy-Fraunié & Dangles, 2019; Donald et al., 2021; Ficetola et al., 2021; Rosero et al., 2021), thus the study of multiple taxa is needed for a better understanding of the ecological processes governing community development.

Chapter 4 answered the second question by analyzing how microbial and animal communities differ across soil depths and development stages. The study showed that taxonomic diversity always increased with time since glacier retreat, with a consistent pattern among different groups and across soil depths, suggesting that time since glacier retreat is a more important driver of community composition compared to soil depth. Some differences in community composition among different soil depths exist, with a lower taxonomic richness in deep soils, and several taxa (especially Bacteria and Fungi) associated to specific depths. Nevertheless, these differences decreased at older stages of soil development, suggesting that communities go through a process of homogenization, as already observed by Rime et al. (2015) for Fungi and Bacteria. is probably related to the complex process of soil development that occurs with time. The interesting point of this study is that by using a multi-taxa approach, the investigation allowed to generalize these conclusions to (almost) the whole belowground biota. Furthermore, this study provided evidence of an early establishment of organisms belonging to different taxonomic groups (from micro-organisms to animals) that occurred since the early years after the retreat of glaciers, albeit with some differences in terms of their alpha diversity. For instance, Bacteria had higher alpha diversity than Insecta or Oligochaeta, still both taxonomic groups were present soon after glacier retreat.

As reviewed in **Chapter 1**, ecologists are now using community, functional and network theories and tools to go beyond the mere description of biodiversity patterns and to tease apart the role of different factors, such as biotic and abiotic parameters, in the formation of communities, and

in the emergence of functional diversity. **Chapter 5** answered the third question by analyzing how time and regional climate influence the emergence of taxonomic and functional diversity of nematode communities in ecosystems developing after the retreat of glaciers in different regions of the world. The study showed that nematodes colonize the forelands immediately after glacier retreat, in accordance with results observed in **Chapter 4** for several Alpine glaciers. Both taxonomic and functional diversity of nematodes increase rapidly through time but the colonization starts sooner in forelands under mild summer temperatures, whom host richer communities. Despite the evidence that climate influences the colonization rate of nematodes, still after few years the colonization accelerates in colder forelands levelling biodiversity differences between climatic regimes. Furthermore, communities of nematodes showed functional shifts during colonization (e.g., from small-sized colonizers to large-sized persisters) indicating a strong role of habitat filtering in the emergence of functional diversity.

As showed by the multiple analyses presented here, the combination of multi-taxa approaches and broad-scale biodiversity data produced via metabarcoding is an effective tool for studying the biotic responses to the retreat of glaciers. Further studies are now needed in order to understand what is the role of other abiotic factors such as microclimate and dispersal limitations and the role of biotic interactions in determining functionally complex ecosystems. For example, the rate of dispersal of different taxa is determined by the combined effects of environmental constraints, species characteristics and interspecific interactions (Ficetola et al., 2021). Biotic interactions can affect colonization rate in multiple ways, for instance with facilitation at the early stages and competition, parasitism, or predation, at later stages, thus decelerating colonization rate (Koffel et al. 2018; Benavent-González et al. 2019). The dataset produced during my thesis work comprises taxonomic information for the whole biota, from microorganisms to animals and plants, in 48 glacier forelands, covering five continents from the equator to polar regions (**Chapter 5, Fig. 1**), for a more complete overview of biotic responses to climate change. Recognizing the complexity of biotic responses will thus allow to better understand how these environments evolve through time and to provide key

insights into the mechanisms that drive their modifications in order to enable the advancement of a more complete theory of ecosystem dynamics (Chang & Turner 2019). Such knowledge will allow to predict cascading effects of the ongoing climate change in these highly dynamic environments at multiple scales.

REFERENCES

Aglieri, G., Baillie, C., Mariani, S., Cattano, C., Calò, A., Turco, G., …, Milazzo, M. (2020). Environmental DNA effectively captures functional diversity of coastal fish communities. Molecular Ecology, 00:1–13. doi: org/10.1111/mec.15661

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

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9, 134-147.

Albrecht, M., Riesen, M., and Schmid, B. (2010). Plant–pollinator network assembly along the chronosequence of a glacier foreland. Oikos 119:1610-1624.

Alsos, I. G., Lavergne, S., Merkel, M. K. F., Boleda, M., Lammers, Y., Alberti, A., … Coissac, E. (2020). The treasure vault can be opened: Large-scale genome skimming works well using herbarium and silica gel dried material. Plants, 9(4), 432. doi: 10.3390/plants9040432

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecology, 26(1), 32–46. doi: 10.1111/J.1442-9993.2001.01070.PP.X

Anderson, M. J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions. Biometrics, 62(1), 245–253. doi: 10.1111/J.1541-0420.2005.00440.X

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC Bioinformatics, 22, 177.

Aros-Mualin, D., Noben, S., Karger, D.N., Carvajal-Hernández, C.I., Salazar, L., Hernández-Rojas, A., …, Kessler, M. (2021). Functional diversity in ferns is driven by species richness rather than by environmental constraints. Frontiers in Plant Science, 11:615723. doi: 10.3389/fpls.2020.615723

Azzoni, R. S., Senese, A., Zerboni, A., Maugeri, M., Smiraglia, C. and Diolaiuti, G. A. (2016). Estimating ice albedo from fine debris cover quantified by a semi-automatic method: the case study of Forni Glacier, Italian Alps. Cryosphere 10:665-679.

Baamrane, M. A. A., Shehzad, W., Ouhammou, A., Abbad, A., Naimi, M., Coissac, E., … Znari, M. (2012). Assessment of the food habits of the Moroccan dorcas gazelle in M'Sabih Talaa, West Central Morocco, using the trnL approach. PLoS ONE, 7, e35643.

Bååth, E., & Arnebrant, K. (1994). Growth rate and response of bacterial communities to pH in limed and ash treated forest soils. Soil Biology and Biochemistry, 26(8), 995–1001. doi: 10.1016/0038-0717(94)90114-7

Bajerski, F., & Wagner, D. (2013). Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. FEMS Microbiology Ecology, 85(1), 128–142. doi: 10.1111/1574-6941.12105

Bálint, M., Bahram, M., Murat Eren, A., Faust, K., Fuhrman, J. A., Bj¨, B., … Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. FEMS Microbiology Reviews, 017, 686–700. doi: 10.1093/femsre/fuw017

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

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography, 19(1), 134–143. doi: 10.1111/j.1466-8238.2009.00490.x

Baselga, A., & Orme, C. D. L. (2012). Betapart: An R package for the study of beta diversity. Methods in Ecology and Evolution, 3(5), 808–812. doi: 10.1111/j.2041-210X.2012.00224.x

Benavent-González, A., Raggio, J., Villagra, J., Blanquer, J. M., Pintado, A., Rozzi, R., Green, T. G. A., and Sancho, L. G. (2019). High nitrogen contribution by Gunnera magellanica and nitrogen transfer by mycorrhizas drive an extraordinarily fast primary succession in sub-Antarctic Chile. New Phytologist 223:661-674.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bhat, A. H., Prabhu, P., & Balakrishnan, K. (2019). A critical analysis of state-of-the-art metagenomics OTU clustering algorithms. Journal of Biosciences, 44, 9.

Bidartondo, M. I. (2008). Preserving accuracy in GenBank. Science, 319, 1616.

Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J. J., & Taberlet, P. (2012). Tracking earthworm communities from soil DNA. Molecular Ecology, 21, 2017-2030. doi: 10.1111/j.1365-294X.2011.05407.x

Bjorkman, A.D., Myers-Smith, I.H., Elmendorf, S.C., Normand, S., Rüger, N., Beck, P.S.A., …, Weiher, E. (2018). Plant functional trait change across a warming tundra biome. Nature 562, 57–62. doi: org/10.1038/s41586-018-0563-7

Blaalid, R., Carlsen, T., Kumar, S., Halvorsen, R., Ugland, K. I., Fontana, G. and Kauserud, H. (2012). Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. Molecular Ecology 21:1897-1908.

Bohmann, K., Elbrecht ,V., Carøe, C., Bista, L., Leese, F., Bunce, M., Yu, D. W., … Creer, S. (2021). Strategies for sample labelling and library preparation in DNA metabarcoding studies. Molecular Ecology Resources.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., … de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology and Evolution. doi: 10.1016/j.tree.2014.04.003

Bonin, A., Guerrieri, A., & Ficetola, G. F. (2021). Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies. Authorea Preprints. doi: 10.22541/AU.163398436.69595421/V1

Borcard, D., Gillet, F., & Legendre, P. (2011). Numerical Ecology with R. In Numerical Ecology with R. doi: 10.1007/978-1-4419-7976-6

Botnen, S. S., Davey, M. L., Halvorsen, R., & Kauserud, H. (2018). Sequence clustering threshold has little effect on the recovery of microbial community structure. Molecular Ecology Resources, 18(5), 1064–1076. doi: 10.1111/1755-0998.12894

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. doi: 10.1111/1755-0998.12428

Brambilla, M., and Gobbi, M. (2014). A century of chasing the ice: delayed colonisation of ice-free sites by ground beetles along glacier forelands in the Alps. Ecography 37:33-42.

Brbić, M., Piškorec, M., Vidulin, V., Kriško, A., Šmuc, T. and Supek, F. (2016). The landscape of microbial phenotypic traits and associated genes. Nucleic Acids Research 44:10074-10090.

Bridge, P. D., Roberts, P. J., Spooner, B. M., & Panchal, G. (2003). On the unreliability of published DNA sequences. New Phytologist, 160, 43-48.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R Journal, 9(2), 378–400. doi: 10.32614/rj-2017-066

Brown, E. A., Chain, F. J. J., Crease, T. J., MacIsaac, H. J., & Cristescu, M. E. (2015). Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? Ecology and Evolution, 5, 2234-2251.

Brown, L. E., Khamis, K., Wilkes, M., Blaen, P., Brittain, J. E., Carrivick, J. L., …, Milner, A. M. (2017). Functional diversity and community assembly of river invertebrates show globally consistent responses to decreasing glacier cover. Nature Ecology & Evolution 2:325–333.

Brown, S. P., and Jumpponen, A. (2014). Contrasting primary successional trajectories of fungi and bacteria in retreating glacier soils. Molecular Ecology 23:481-497.

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., … Jumpponen, A. (2015). Scraping the bottom of the barrel: Are rare high throughput sequences artifacts? Fungal Ecology, 13, 221–225. doi: 10.1016/j.funeco.2014.08.006

Burga, C. A., Krüsi, B., Egli, M., Wernli, M., Elsener, S., Ziefle, M., Fischer, T., and Mavris, C. (2010). Plant succession and soil development on the foreland of the Morteratsch glacier (Pontresina, Switzerland): Straight forward or chaotic? Flora - Morphology, Distribution, Functional Ecology of Plants 205:561-576.

Bürkner, P.C. (2018) Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10, 395-411.

Bürkner, P. C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. R Journal, 10(1), 395–411. doi: 10.32614/rj-2018-017

Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L., & Thuiller, W. (2020). From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? Journal of Biogeography, 47(1), 193–206. doi: 10.1111/jbi.13681

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., … Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. Scientific Reports, 9(1), 1–11. doi: 10.1038/s41598-019-39399-5

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P., Vuillemin, A. …Parducci, L. (2021). Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: Overview and recommendations. Quaternary 4, 6.

Cardona, S., Eck, A., Cassellas, M., Gallart, M., Alastrue, C., Dore, J., … Manichanh, C. (2012). Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC Microbiology, 12. doi: 10.1186/1471-2180-12-158

Cardoso, P., Rigal, F., Carvalho, J.C. (2015). BAT – Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. Methods in Ecology and Evolution, 6, 232-236.

Carteron, A., Beigas, M., Joly, S., Turner, B. L., & Laliberté, E. (2021). Temperate Forests Dominated by Arbuscular or Ectomycorrhizal Fungi Are Characterized by Strong Shifts from Saprotrophic to Mycorrhizal Fungi with Increasing Soil Depth. Microbial Ecology, 82(2), 377–390. doi: 10.1007/s00248-020-01540-7

Cauvy-Fraunié, S., & Dangles, O. (2019). A global synthesis of biodiversity responses to glacier retreat. Nature Ecology and Evolution, 3(12), 1675–1685. doi: 10.1038/s41559-019-1042-8

Cázares, E., and Trappe, J. M. (1994). Spore Dispersal of Ectomycorrhizal Fungi on a Glacier Forefront by Mammal Mycophagy. Mycologia 86:507-510.

Chang, C. C., & Turner, B. L. (2019). Ecological succession in a changing world. Journal of Ecology, 107(2), 503–509. doi: 10.1111/1365-2745.13132

Chang, C. C., and Turner, B. L. (2019). Ecological succession in a changing world. Journal of Ecology 107:503-509.

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. Http://Dx.Doi.Org/10.1146/Annurev-Ecolsys-120213-091540, 45, 297–324. doi: 10.1146/ANNUREV-ECOLSYS-120213-091540

Chapin, F. S., Conway, A. J., Johnstone, J. F., Hollingsworth, T. N., and Hollingsworth, J. (2016). Absence of net long-term successional facilitation by alder in a boreal Alaska floodplain. Ecology 97:2986-2997.

Chapin, F. S., Walker, L. R., Fastie, C. L., and Sharman, L. C. (1994). Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. Ecological Monographs 64:149-175.

Chase, M. W., & Hills, H. H. (1991). Silica gel: An ideal material for field preservation of leaf samples for DNA studies. TAXON, 40(2), 215–220. doi: 10.2307/1222975

Chazdon, R.L., Colwell, R.K., Denslow, J.S., Guariguata, M.R. (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of

NE Costa Rica. In: Forest biodiversity research, monitoring and modeling: Conceptual background and Old World case studies (eds. Dallmeier F, Comiskey JA), pp. 285-309. Parthenon Publishing, Paris.

Chen, W., & Ficetola, G. F. (2020). Numerical methods for sedimentary-ancient-DNA-based study on past biodiversity and ecosystem functioning. Environmental DNA, edn3.79. doi: 10.1002/edn3.79

Chen, W., Jiao, S., Li, Q., & Du, N. (2020). Dispersal limitation relative to environmental filtering governs the vertical small-scale assembly of soil microbiomes during restoration. Journal of Applied Ecology, 57(2), 402–412. doi: 10.1111/1365-2664.13533

Chu, H., Sun, H., Tripathi, B. M., Adams, J. M., Huang, R., Zhang, Y., & Shi, Y. (2016). Bacterial community dissimilarity between the surface and subsurface soils equals horizontal differences over several kilometers in the western Tibetan Plateau. Environmental Microbiology, 18(5), 1523–1533. doi: 10.1111/1462-2920.13236

Clare, E. L., Chain, F. J. J., Littlefair, J. E., & Cristescu, M. E. (2016). The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding data. Genome, 59, 981-990.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. Austral Ecology, 18(1), 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Clements, F. E. (1916). Plant succession: an analysis of the development of vegetation. Carnegie Institution of Washington.

Coissac, E. (2012). OligoTag: A program for designing sets of tags for next-generation sequencing of multiplexed samples. Methods in Molecular Biology, 888, 13–31. doi: 10.1007/978-1-61779-870-2_2

Coissac, E., & Gonindard-Melodelima, C. (2019). Assessing the shared variation among high-dimensional data matrices: a modified version of the Procrustean correlation coefficient. BioRxiv Bioinformatics, 1–16. doi: 10.1101/842070

Colwell, R.K., Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society B: Biological Sciences, 345, 101-118.

Connell, S. H., and Slatyer, R. O. (1977). Mechanisms of Succession in Natural Communities and Their Role in Community Stability and Organization. The American Naturalist 111:1119-1144.

Consul, P. C., & Famoye, F. (1992). Generalized poisson regression model. Communications in Statistics - Theory and Methods, 21(1), 89–109. doi: 10.1080/03610929208830766

Coulson, S. J., Hodkinson, I. D., and Webb, N. R. (2003). Aerial dispersal of invertebrates over a high-Arctic glacier foreland: Midtre Lovénbreen, Svalbard. Polar Biology 26:530-537.

Couton, M., Baud, A., Daguin-Thiébaut, C., Corre, E., Comtet, T., & Viard, F. (2021). High-throughput sequencing on preservative ethanol is effective at jointly examining infraspecific and taxonomic diversity, although bioinformatics pipelines do not perform equally. Ecology and Evolution, 11, 5533-5546.

Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., …, Maynard, D.S. (2019). The global soil community and its influence on biogeochemistry. Science, 365(6455):eaav0550. doi: 10.1126/science.aav0550. PMID: 31439761.

Davey, M., Blaalid, R., Vik, U., Carlsen, T., Kauserud, H., & Eidesen, P. B. (2015). Primary succession of Bistorta vivipara (L.) Delabre (Polygonaceae) root-associated fungi mirrors plant succession in two glacial chronosequences. Environmental Microbiology, 17(8), 2777–2790. doi: 10.1111/1462-2920.12770/SUPPINFO

De Barba, M., Miquel, C., Boyer, F., Rioux, D., Coissac, E., Taberlet, P. (2014). DNA metabarcoding multiplexing for omnivorous diet analysis and validation of data accuracy. Molecular Ecology Resources, 14, 306–323.

De Bello, F., Thuiller, W., Lepš, J., Choler, P., Clément, J.-C., Macek, P., …, Lavorel, S. (2009) Partitioning of functional diversity reveals the scale and extent of trait convergence and divergence. Journal of Vegetation Science, 20, 475-486.

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. Ecology, 90(12), 3566–3574. doi: 10.1890/08-1823.1

De Cáceres, M., Legendre, P., & Moretti, M. (2010). Improving indicator species analysis by combining groups of sites. Oikos, 119(10), 1674–1684. doi: 10.1111/j.1600-0706.2010.18334.x

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237). doi: 10.1126/science.1261605

Delavaux, C. S., Bever, J. D., Karppinen, E. M., & Bainard, L. D. (2020). Keeping it cool: Soil sample cold pack storage and DNA shipment up to 1 month does not impact metabarcoding results. Ecology and Evolution, (November 2019), 1–13. doi: 10.1002/ece3.6219

Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., et al. (2020) Multiple elements of soil biodiversity drive ecosystem functions across biomes. Nature Ecology & Evolution, 4, 210-220.

Delgado-Baquerizo, M., Bardgett, R. D., Vitousek, P. M., Maestre, F. T., Williams, M. A., Eldridge, D. J., Fierer, N. (2019). Changes in belowground biodiversity during ecosystem development. Proceedings of the National Academy of Sciences of the United States of America 116:6891.

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

Devetter, M., Hanel, L., Raschmanova, N., Bryndova, M., Schlaghamersky, J. (2021) Terrestrial invertebrates along a gradient of deglaciation in Svalbard: Long-term development of soil fauna communities. Geoderma, 383, 11.

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., … Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. Molecular Ecology Resources, 18(5), 940–952. doi: 10.1111/1755-0998.12907

Doblas-Miranda, E., Sánchez-Piñero, F., & González-Megías, A. (2009). Vertical distribution of soil macrofauna in an arid ecosystem: Are litter and belowground compartmentalized habitats? Pedobiologia, 52(6), 361–373. doi: 10.1016/J.PEDOBI.2008.11.006

Dopheide, A., Xie, D., Buckley, T.R., Drummond, A.J., Newcomb, R.D. (2019). Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution;10:120–133. doi: org/10.1111/2041-210X.13086

Donald, J., Murienne, J., Chave, J., Iribar, A., Louisanna, E., Manzi, S., … Zinger, L. (2021). Multi-taxa environmental DNA inventories reveal distinct taxonomic and functional diversity in urban tropical forest fragments. Global Ecology and Conservation, 29, e01724. doi: 10.1016/J.GECCO.2021.E01724

Dong, K., Tripathi, B., Moroenyane, I., Kim, W., Li, N., Chu, H. Y., and Adams, J. (2016). Soil fungal community development in a high Arctic glacier foreland follows a directional replacement model, with a mid-successional diversity maximum. Scientific Reports 6:9.

Drenovsky, R. E., Vo, D., Graham, K. J., & Scow, K. M. (2004). Soil water content and organic carbon availability are major determinants of soil microbial community composition. Microbial Ecology, 48(3), 424–430. doi: 10.1007/s00248-003-1063-2

Eichel, J., Corenblit, D., and Dikau, R. (2016). Conditions for feedbacks between geomorphic and vegetation dynamics on lateral moraine slopes: a biogeomorphic feedback window. Earth Surface Processes and Landforms 41:406-419.

Eichmiller, J. J., Miller L. M., & Sorensen, P.W. (2016). Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. Molecular Ecology Resources, 16, 56-68.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J. N., … Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ, 4, 12.

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., … Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. Molecular Ecology, 21(8), 1821–1833. doi: 10.1111/j.1365-294X.2012.05537.x

Erhart, H. (1951). La genèse des sols en tant que phénomène géologique. Esquisse d'une théorie géologique et géochimique. Biostasie et rhéxistasie. Masson, Paris.

Erschbamer, B., and M. Caccianiga (2016). Glacier Forelands: Lessons of Plant Population and Community Development. Progress in Botany 78:259-284.

Erschbamer, B., and Mayer, R. (2011). Can successional species groups be discriminated based on their life history traits? A study from a glacier foreland in the Central Alps. Plant Ecology & Diversity 4:341-351.

Erschbamer, B., Schlag, R. N., and Winkler, E. (2008). Colonization processes on a central Alpine glacier foreland. Journal of Vegetation Science 19:855-862.

Evans, S. E., Wallenstein, M. D., & Burke, I. C. (2014). Is bacterial moisture niche a good predictor of shifts in community composition under long-term drought. Ecology, 95(1), 110–122. doi: 10.1890/13-0500.1

Fahner, N. A., Shokralla, S., Baird, D. J., & Hajibabaei, M. (2016). Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. PLoS ONE, 11, e0157505.

Federhen, S. (2015). Type material in the NCBI Taxonomy Database. Nucleic Acids Research, 43, D1086-D1098.

Feldmeier, S., Schmidt, B. R., Zimmermann, N. E., Veith, M., Ficetola, G. F., and Lotters, S. (2020). Shifting aspect or elevation? The climate change response of ectotherms in a complex mountain topography. Diversity and Distributions:1483-1495.

Fernández-Calviño, D., & Bååth, E. (2010). Growth response of the bacterial community to pH in soils differing in pH. FEMS Microbiology Ecology, 73(1), 149–156. doi: 10.1111/j.1574-6941.2010.00873.x

Ferris, H. (2010) Form and function: Metabolic footprints of nematodes in the soil food web. European Journal of Soil Biology, 46, 97-104.

Ferris, H., Bongers, T., de Goede, R.G.M. (2001) A framework for soil food web diagnostics: extension of the nematode faunal analysis concept. Applied Soil Ecology, 18, 13-29.

Ficetola, G.F., Boyer, F., Valentini, A., Bonin, A., Meyer, A., Dejean, A., …, Taberlet, P. (2021a) Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. Molecular Ecology, 30, 3189–3202.

Ficetola, G.F., Coissac, E., Zundel, S., et al. (2010) An In silico approach for the evaluation of DNA barcodes. BMC Genomics, 11, 434.

Ficetola, G.F., Denoël, M. (2009) Ecological thresholds: an assessment of methods to identify abrupt changes in species-habitat relationships. Ecography, 32, 1075-1084.

Ficetola, G.F., Marta, S., Guerrieri, A., et al. (2021b) Dynamics of ecological communities following current retreat of glaciers. Annual review of Ecology Evolution and Systematics, 52, 405-426.

Ficetola, G.F., Pansu, J., Bonin, A., et al. (2015) Replication levels, false presences, and the estimation of presence / absence from eDNA metabarcoding data. Molecular Ecology Resources, 15, 543–556.

Ficetola, G. F., Boyer, F, Valentini, A. Bonin, Meyer, A., Dejean, T., … Taberlet, P. (2021b). Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. Molecular Ecology, 30, 3189–3202.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 434. doi: 10.1186/1471-2164-11-434

Ficetola, G. F., Lunghi, E., Canedoli, C., Padoa-Schioppa, E., Pennati, R., and Manenti, R. (2018). Differences between microhabitat and broad-scale patterns of niche evolution in terrestrial salamanders. Scientific Reports 8:10575.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., … Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources. doi: 10.1111/1755-0998.12338

Ficetola, G. F., Marta, S., Guerrieri, A., Gobbi, M., Ambrosini, R., Fontaneto, D., … Thuiller, W. (2021). Dynamics of Ecological Communities Following Current Retreat of Glaciers. Annual Review of Ecology, Evolution, and Systematics, 52(1), 1–51. doi: 10.1146/annurev-ecolsys-010521-040017

Fickert, T., Grüninger, F., and Damm, B. (2017). Klebelsberg revisited: did primary succession of plants in glacier forelands a century ago differ from today? Alpine Botany 127:17-29.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics Using r. London: SAGE Publications.

Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C. (2009) Global patterns in belowground communities. Ecology Letters, 12, 1238-1249.

Fierer, N., Schimel, J. P., & Holden, P. A. (2003). Variations in microbial community composition through two soil depth profiles. Soil Biology and Biochemistry, 35(1), 167–176. doi: 10.1016/S0038-0717(02)00251-1

Flø, D., and Hågvar, S. (2013). Aerial Dispersal of Invertebrates and Mosses Close to a Receding Alpine Glacier in Southern Norway. Arctic Antarctic and Alpine Research 45:481-490.

Fontaneto D (2019) Long-distance passive dispersal in microscopic aquatic animals. Movement Ecology, 7, 10.

Franzetti, A., Pittino, F., Gandolfi, I., Azzoni, R.S., Diolaiuti, G., Smiraglia, C., Pelfini, M., Compostella, C., Turchetti, B., Buzzini, P., Ambrosini, R. (2020). Early ecological succession patterns of bacterial, fungal and plant communities along a chronosequence in a

recently deglaciated area of the Italian Alps. FEMS Microbiol Ecol, 96(10):fiaa165. doi: 10.1093/femsec/fiaa165. PMID: 32815995.

Franzetti, A., Navarra, F., Tagliaferri, I., Bestetti, G., Minora, U., Azzoni, R. S., Diolaiuti, G., Smiraglia, C., and Ambrosini, R. (2017). Potential sources of bacteria colonizing the cryoconite of an Alpine glacier. PLoS ONE 12: e0174786.

Freeman, B.G., Song, Y., Feeley, K.J., Zhu, K. (2021). Montane species track rising temperatures better in the tropics than in the temperate zone. Ecology Letters, n/a.

Fridley, J. D., and Wright, J. P. (2018). Temperature accelerates the rate fields become forests. Proceedings of the National Academy of Sciences of the United States of America 115:4702-4706.

Froslev, T. G., Kjoller, R., Bruun, H. H., Ejrnaes, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications, 8, 11.

Garibotti, I. A., Pissolito, C. I., and Villalba, R. (2011). Spatiotemporal Pattern of Primary Succession in Relation to Meso-topographic Gradients on Recently Deglaciated Terrains in the Patagonian Andes. Arctic Antarctic and Alpine Research 43:555-567.

Geisen, S., Snoek, L.B., ten Hooven, F.C., et al. (2018). Integrating quantitative morphological and qualitative molecular methods to analyse soil nematode community responses to plant range expansion. Methods in Ecology and Evolution, 9, 1366-1378.

Gelman, A., Hwang, J., Vehtari, A. (2014) Understanding predictive information criteria for Bayesian models. Statistics and Computing, 24, 997-1016.

Gelman, A., Hwang, J., & Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. Statistics and Computing 2013 24:6, 24(6), 997–1016. doi: 10.1007/S11222-013-9416-2

Geyer, K. M., Altrichter, A. E., Takacs-Vesbach, C. D., Van Horn, D. J., Gooseff, M. N., & Barrett, J. E. (2014). Bacterial community composition of divergent soil habitats in a polar desert. FEMS Microbiology Ecology, 89(2), 490–494. doi: 10.1111/1574-6941.12306

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. Nature, 345(6270), 60–63. doi: 10.1038/345060a0

Gobbi, M., Ballarin, F., Brambilla, M., et al. (2017) Life in harsh environments: carabid and spider trait types and functional diversity on a debris-covered glacier and along its foreland. Ecological Entomology, 42, 838-848.

Gobbi, M., Rossaro, B., Vater, A., De Bernardi, F., Pelfini, M., and Brandmayr, P. (2007). Environmental features influencing Carabid beetle (Coleoptera) assemblages along a recently deglaciated area in the Alpine region. Ecological Entomology 32:682-689.

Gobbi, M., Ballarin, F., Brambilla, M., Compostella, C., Isaia, M., Losapio, G., … Caccianiga, M. (2017). Life in harsh environments: carabid and spider trait types and functional diversity on a debris-covered glacier and along its foreland. Ecological Entomology, 42(6), 838–848. doi: 10.1111/EEN.12456

Gobbi, M., Fontaneto, D., and De Bernardi, F. (2006b). Influence of climate changes on animal communities in space and time: the case of spider assemblages along an alpine glacier foreland. Global Change Biology 12:1985-1992.

Gobbi, M., Ambrosini, R., Casarotto, C., Diolaiuti, G., Ficetola, G. F., Lencioni, V., …, Caccianiga, M. (2021). Vanishing permanent glaciers: Climate change is threatening a European Union habitat (Code 8340) and its poorly known biodiversity. Biodiversity and Conservation: DOI: 10.1007/s10531-10021-02185-10539.

Gobbi, M., De Bernardi, F., Pelfini, M., Rossaro, B., and Brandmayr, P. (2006a). Epigean arthropod succession along a 154-year glacier foreland cronosequence in the Forni Valley (Central Italian Alps). Arctic Antarctic and Alpine Research 38:357-362.

Gobbi, M., Caccianiga, M., Cerabolini, B., De Bernardi, F., Luzzaro, A., and Pierce, S. (2010). Plant adaptive responses during primary succession are associated with functional adaptations in ground beetles on deglaciated terrain. Community Ecology 11:223-231.

Gorniak, D., Marszalek, H., Kwasniak-Kominek, M., Rzepa, G., and Manecki, M. (2017). Soil formation and initial microbiological activity on a foreland of an Arctic glacier (SW Svalbard). Applied Soil Ecology 114:34-44.

Gotelli, N.J., Colwell, R.K. (2011) Estimating species richness. In: Biological diversity. Frontiers in measurement and assessment (eds. Magurran AE, McGill BJ), pp. 379-391. Oxford University Press, Oxford.

Gotelli, N.J., McCabe, D.J. (2002) Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. Ecology, 83, 2091-2096.

Götzenberger, L., de Bello, F., Bråthen, K.A., et al. (2012) Ecological assembly rules in plant communities—approaches, patterns and prospects. Biological Reviews, 87, 111-127.

Gravel, D., Albouy, C., and Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. Philosophical Transactions of the Royal Society B-Biological Sciences 371:14.

Grenié, M., Denelle, P., Tucker, C.M., Munoz, F., Violle, C. (2017) funrar: An R package to characterize functional rarity. Diversity and Distributions, 23, 1365-1371.

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., & Turon, X. (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. PLOS ONE, 10(10), e0139633. doi: 10.1371/journal.pone.0139633

Guerrieri, A., Bonin, A., Münkemüller, T., Gielly, L., Thuiller, W., & Francesco Ficetola, G. (2020). Effects of soil preservation for biodiversity monitoring using environmental DNA. Molecular Ecology. doi: 10.1111/mec.15674

Guo, Y., Chen, X., Wu, Y., Zhang, L., Cheng, J., Wei, G., & Lin, Y. (2018). Natural revegetation of a semiarid habitat alters taxonomic and functional diversity of soil microbial communities. Science of The Total Environment, 635, 598–606. doi: 10.1016/J.SCITOTENV.2018.04.171

Guo, Y., Yang, G. Q., Chen, Y., Li, D., & Guo, Z. (2018). A comparison of different methods for preserving plant molecular materials and the effect of degraded DNA on ddRAD sequencing. Plant Diversity, 40(3), 106–116. doi: 10.1016/j.pld.2018.04.001

Hagvar, S. (2011). Primary Succession in Glacier Forelands: How Small Animals Conquer New Land Around Melting Glaciers. Pages 151-172. International Perspectives on Global Environmental Change.

Hågvar, S., and Ohlson, W. (2013). Ancient carbon from a melting glacier gives high 14C age in living pioneer invertebrates. Scientific Reports 3:2820.

Hågvar, S., Gobbi, M., Kaufmann, R., Ingimarsdóttir, M., Caccianiga, M., Valle, B., … Vater, A. (2020). Ecosystem Birth Near Melting Glaciers: A Review on the Pioneer Role of Ground-Dwelling Arthropods. 11, 644. doi: 10.3390/insects11090644

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B-Biological Sciences, 270, S96-S99.

Herold, N., Schöning, I., Berner, D., Haslwimmer, H., Kandeler, E., Michalzik, B., & Schrumpf, M. (2014). Vertical gradients of potential enzyme activities in soil profiles of European beech, Norway spruce and Scots pine dominated forest sites. Pedobiologia, 57(3), 181–189. doi: 10.1016/J.PEDOBI.2014.03.003

Ho, A., Di Lonardo, D. P., & Bodelier, P. L. E. (2017). Revisiting life strategy concepts in environmental microbial ecology. FEMS Microbiology Ecology, 93(3). doi: 10.1093/FEMSEC/FIX006

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y. et al. (2019). High mountain areas. In Pörtner HO, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M,

Poloczanska E, Mintenbeck K, Nicolai M, Okem A, Petzold J, Rama B, Weyer N, editors, The Ocean and Cryosphere in a Changing Climate. Intergovernmental Panel on Climate Change. p. 2-1 - 2-90

Hodkinson, I. D., Webb, N. R., and Coulson, S. J. (2002). Primary community assembly on land - the missing stages: why are the heterotrophic organisms always there first? Journal of Ecology 90:569-577.

Hodson, A., Anesio, A. M., Tranter, M., Fountain, A., Osborn, M., Priscu, J., Laybourn-Parry, J. and Sattler, B. (2008). Glacial ecosystems. Ecological Monographs 78:41-67.

Hoffmann, C., Schubert, G., & Calvignac-Spencer, S. (2016). Aquatic biodiversity assessment for the lazy. Molecular Ecology, Vol. 25, pp. 846–848. doi: 10.1111/mec.13535

Hotaling, S., Hood, E., and Hamilton, T. L. (2017). Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. Environmental Microbiology 19:2935-2948.

Hubbell, S.P. (2001) The Unified Neutral Theory of Biodiversity and Biogeography Princeton University Press.

Huerlimann, R., Cooper, M. K., Edmunds, R. C., Villacorta-Rath, C., Le Port, A., Robson, H. L. A., … Jerry, D. R. (2020). Enhancing tropical conservation and ecology research with aquatic environmental DNA methods: an introduction for non-environmental DNA specialists. Animal Conservation. doi: 10.1111/acv.12583

Hugonnet, R., McNabb, R., Berthier, E. et al. (2021). Accelerated global glacier mass loss in the early twenty-first century. Nature 592, 726–731. doi: org/10.1038/s41586-021-03436-z

Ingham, R. E., Trofymow, J. A., Ingham, E. R., & Coleman, D. C. (1985). Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. Ecological Monographs, 55(1), 119–140. doi: 10.2307/1942528

Intergovernmental Panel on Climate Change (2021) Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge.

Jackson, M. A., Bell, J. T., Spector, T. D., & Steves, C. J. (2016). A heritability-based comparison of methods used to cluster 16S rRNA gene sequences into operational taxonomic units. PeerJ, 4, 19.

Janssen, P., Bec, S., Fuhr, M., Taberlet, P., Brun, J.-J., & Bouget, C. (2018). Present conditions may mediate the legacy effect of past land-use changes on species richness and composition of above- and below-ground assemblages. Journal of Ecology, 106(1), 306–318. doi: 10.1111/1365-2745.12808

Jenny, H. (1994). Factors of soil formation: a system of quantitative pedology. Dover Publications, New York.

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters, 16(10), 1245–1257. doi: 10.1111/ele.12162

Jiao, S., Chen, W., Wang, J., Du, N., Li, Q., & Wei, G. (2018). Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. Microbiome, 6(1), 1–13. doi: 10.1186/s40168-018-0526-0

Johnson, E. A., & Miyanishi, K. (2008). Testing the assumptions of chronosequences in succession. Ecology Letters, 11(5), 419–431. doi: 10.1111/J.1461-0248.2008.01173.X

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., … Daniel, R. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. Scientific Reports, 6(August), 1–12. doi: 10.1038/srep33696

Karger, D., Conrad, O., Böhner, J. et al. (2017). Climatologies at high resolution for the earth's land surface areas. Sci Data 4, 170122. doi: org/10.1038/sdata.2017.122

Kassambara, A. (2020). “ggplot2” Based Publication Ready Plots [R package ggpubr version 0.4.0]. Retrieved from https://cran.r-project.org/package=ggpubr

Kaufmann, R. (2002). Glacier foreland colonisation: distinguishing between short-term and long-term effects of climate change. Oecologia, 130, 470-475.

Kaufmann, R. (2001). Invertebrate succession on an alpine glacier foreland. Ecology, 82(8), 2261–2278. doi: 10.1890/0012-9658(2001)082[2261:ISOAAG]2.0.CO;2

Khedim, N., Cécillon, L., Poulenard, J., Barré, P., Baudin, F., Marta, S., … Ficetola, G. F. (2021). Topsoil organic matter build-up in glacier forelands around the world. Global Change Biology, 27(8), 1662–1677. doi: 10.1111/gcb.15496

Khokon, A. M., Schneider, D., Daniel, R., & Polle, A. (2021). Soil Layers Matter: Vertical Stratification of Root-Associated Fungal Assemblages in Temperate Forests Reveals Differences in Habitat Colonization. Microorganisms 2021, Vol. 9, Page 2131, 9(10), 2131. doi: 10.3390/MICROORGANISMS9102131

Kim, M., Jung, J. Y., Laffly, D., Kwon, H. Y., and Lee, Y. K. (2017). Shifts in bacterial community structure during succession in a glacier foreland of the High Arctic. Fems Microbiology Ecology 93:9.

Koffel, T., Boudsocq, S., Loeuille, N., and Daufresne, T. (2018). Facilitation- vs. competition-driven succession: the key role of resource-ratio. Ecology Letters 21:1010-1021.

König, T., Kaufmann, R., and Scheu, S. (2011). The formation of terrestrial food webs in glacier foreland: Evidence for the pivotal role of decomposer prey and intraguild predation. Pedobiologia 54:147-152.

Kopylova, E., Navas-Molina, J. A., Mercier, C., Xu, Z. Z., Mahe, F., He, Y., … Knight, R. (2016). Open-source sequence clustering methods improve the state of the art. mSystems, 1, 16.

Körner, C. (2000). Why are there global gradients in species richness? Mountains might hold the answer. Trends in Ecology & Evolution 15:513-514.

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. Molecular Ecology Resources, 20(1), 29–39. doi: 10.1111/1755-0998.13107

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology, 12, 118-123.

Laliberté, E., Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. Ecology, 91, 299-305.

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiology Letters, 307(1), 80–86. doi: 10.1111/j.1574-6968.2010.01965.x

Le Bagousse-Pinguet, Y., Soliveres, S., Gross, N., et al. (2019) Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. Proceedings of the National Academy of Sciences, 116, 8419.

Lear, G., Dickie, I., Banks, J., Boyer, S., Buckley, H. L., Buckley, T. R., … Holdaway, R. (2018). Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. New Zealand Journal of Ecology, 42, 10.

Leclercq, P. W., and Oerlemans, J. (2012). Global and hemispheric temperature reconstruction from glacier length fluctuations. Climate Dynamics 38:1065-1079.

Legendre, P., & Borcard, D. (2018). Box-Cox-chord transformations for community composition data prior to beta diversity analysis. Ecography, 41(11), 1820–1824. doi: 10.1111/ecog.03498

Legendre, P., & Legendre, L. (2012). Numerical Ecology Ch 6 - Multidimensional qualitative data. Developments in Environmental Modelling, 24, 337–424. doi: 10.1016/B978-0-444-53868-0.50008-3

Lei, Y.B., Zhou, J., Xiao, H.F., et al. (2015) Soil nematode assemblages as bioindicators of primary succession along a 120-year-old chronosequence on the Hailuogou Glacier forefield, SW China. Soil Biology & Biochemistry, 88, 362-371.

Leray, M., Knowlton, N., Ho, S. L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. Proceedings of the National Academy of Sciences of the United States of America, 116, 22651-22656.

Levy-Booth, D.J., Campbell, R.G., Gulden, R.H., et al. (2007). Cycling of extracellular DNA in the soil environment. Soil Biology & Biochemistry, 39, 2977–2991.

Lopes, C. M., Baêta, D., Valentini, A., Lyra, M. L., Sabbag, A. F., Gasparini, J. L., … Zamudio, R. K. (2021). Lost and found: Frogs in a biodiversity hotspot rediscovered with environmental DNA. Molecular Ecology, 30, 3289-3298.

Losapio, G., Cerabolini, B.E.L., Maffioletti, C., et al. (2021) The Consequences of Glacier Retreat Are Uneven Between Plant Species. Frontiers in Ecology and Evolution, 8, 520.

Losapio, G., Gobbi, M., Marano, G. et al. (2016). Feedback effects between plant and flower-visiting insect communities along a primary succession gradient. Arthropod-Plant Interactions 10, 485–495. doi: org/10.1007/s11829-016-9444-x

Losapio, G., Jordan, F., Caccianiga, M., and Gobbi, M. (2015). Structure-dynamic relationship of plant-insect networks along a primary succession gradient on a glacier foreland. Ecological Modelling 314:73-79.

Macher, T.-H., Beermann, A. J., & Leese, F. (2021). TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. Molecular Ecology Resources, 21, 1705-1714.

Mächler, E., Walser, J.-C., & Altermatt, F. (2021). Decision-making and best practices for taxonomy-free environmental DNA metabarcoding in biomonitoring using Hill numbers. Molecular Ecology, 30(13), 3326–3339. doi: 10.1111/MEC.15725

Majaneva, M., Diserud, O. H., Eagle, S. H. C., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. Scientific Reports, 8(1), 4682. doi: 10.1038/s41598-018-23052-8

Makoto, K., and Wilson, S. D. (2016). New Multicentury Evidence for Dispersal Limitation during Primary Succession. American Naturalist 187:804-811.

Marta, S., Azzoni, R. S., Fugazza, D., Tielidze, L., Chand, P., Sieron, K., … Zimmer, A. (2021). The Retreat of Mountain Glaciers since the Little Ice Age : A Spatially Explicit Database. 4–11.

Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., et al. (2019) Multi-trophic β-diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. Functional Ecology, 33, 2053-2064.

Marzeion, B., Cogley, J. G., Richter, K., and Parkes, D. (2014). Attribution of global glacier mass loss to anthropogenic and natural causes. Science 345:919-921.

Matsuoka, S., Ogisu, Y., Sakoh, S., Hobara, S., and Osono, T. (2019). Taxonomic, functional, and phylogenetic diversity of fungi along primary successional and elevational gradients near Mount Robson, British Columbia. Polar Science 21:165-171.

Matthews, J. A. (1992). The Ecology of Recently-deglaciated Terrain: A Geoecological Approach to Glacier Forelands and Primary Succession. Cambridge University Press, Cambridge.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE, 8(4), e61217. doi: 10.1371/JOURNAL.PONE.0061217

Meisner, A., Jacquiod, S., Snoek, B. L., Ten Hooven, F. C., & van der Putten, W. H. (2018). Drought legacy effects on the composition of soil fungal and prokaryote communities. Frontiers in Microbiology, 9(MAR), 294. doi: 10.3389/fmicb.2018.00294

Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. Programs and Abstracts of the SeqBio 2013 Workshop, 27-29.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. PLoS Biology, 3, 2229-2238.

Mioduchowska, M., Czyz, M. J., Goldyn, B., Kur, J., & Sell, J. (2018). Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too "universal" ? PLoS ONE, 13, e0199609.

Moradi, J., John, K., Vicentini, F., Veselá, H., Vicena, J., Ardestani, M. M., & Frouz, J. (2020). Vertical distribution of soil fauna and microbial community under two contrasting post mining chronosequences: Sites reclaimed by alder plantation and unreclaimed regrowth. Global Ecology and Conservation, 23. doi: 10.1016/j.gecco.2020.e01165

Moretti, M., Dias, A. T. C., de Bello, F., Altermatt, F., Chown, S. L., Azcárate, F. M., …, Berg, M. P. (2016). Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. Functional Ecology:DOI: 10.1111/1365-2435.12776.

Mori, A. S., Osono, T., Cornelissen, J. H. C., Craine, J., and Uchida, M. (2017). Biodiversity-ecosystem function relationships change through primary succession. Oikos 126:1637-1649.

Mouchet, M.A., Villéger, S., Mason, N.W.H., Mouillot, D. (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. Functional Ecology, 24, 867-876.

Muggeo, V.M.R. (2003). Estimating regression models with unknown break-points. Statistics in Medicine, 22, 3055-3071.

Mundra, S., Kjønaas, O. J., Morgado, L. N., Krabberød, A. K., Ransedokken, Y., & Kauserud, H. (2021). Soil depth matters: shift in composition and inter-kingdom co-occurrence patterns of microorganisms in forest soils. FEMS Microbiology Ecology, 97(3). doi: 10.1093/FEMSEC/FIAB022

Naeem, S., Duffy, J. E., and Zavaleta, E. (2012). The Functions of Biological Diversity in an Age of Extinction. Science 336:1401-1406.

Nascimbene, J., Mayrhofer, H., Dainese, M., and Bilovitz Peter, O. (2017). Assembly patterns of soil-dwelling lichens after glacier retreat in the European Alps. Journal of Biogeography 44:1393-1404.

Neher, D.A. (2001) Role of nematodes in soil health and their use as indicators. Journal of nematology, 33, 161-168.

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., … Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology, 20, 241–248. doi: 10.1016/J.FUNECO.2015.06.006

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., … Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources, 18(5), 927–939. doi: 10.1111/1755-0998.12895

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., … Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources, 18, 927-939.

Ochoa-Hueso, R., Collins, S. L., Delgado-Baquerizo, M., Hamonts, K., Pockman, W. T., Sinsabaugh, R. L., … Power, S. A. (2018). Drought consistently alters the composition of soil fungal and bacterial communities in grasslands from two continents. Global Change Biology, 24(7), 2818–2827. doi: 10.1111/gcb.14113

Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L., and Thuiller, W. (2019). Diversity indices for ecological networks: a unifying framework using Hill numbers. Ecology Letters 22:737-747.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., … Maintainer, H. W. (2019). vegan: Community Ecology Package Version 2.5-6. www.r-project.org

Orchard, S., Standish, R. J., Nicol, D., Dickie, I. A., & Ryan, M. H. (2017). Sample storage conditions alter colonisation structures of arbuscular mycorrhizal fungi and, particularly, fine root endophyte. Plant and Soil, 412(1–2), 35–42. doi: 10.1007/s11104-016-2867-4

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: a review. European Journal of Soil Science, 69(1), 140–153. doi: 10.1111/EJSS.12499

Orwin, K. H., Kirschbaum, M. U. F., John, M. G. S., & Dickie, I. A. (2011). Organic nutrient uptake by mycorrhizal fungi enhances ecosystem carbon storage: a model-based assessment. Ecology Letters, 14(5), 493–502. doi: 10.1111/J.1461-0248.2011.01611.X

Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. Molecular Ecology, Vol. 25, pp. 1032–1057. doi: 10.1111/mec.13536

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., et al. (2015). Ancient and modern environmental DNA. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 370, 20130383.

Petchey, O.L., Gaston, K.J. (2006). Functional diversity: back to basics and looking forward. Ecology Letters, 9, 741-758.

Pettersson, M., & Bååth, E. (2003). Temperature-dependent changes in the soil bacterial community in limed and unlimed soil. FEMS Microbiology Ecology, 45(1), 13–21. doi: 10.1016/S0168-6496(03)00106-5

Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., … Eisenhauer, N. (2019). Global distribution of earthworm diversity. Science, 366(6464), 480–485. doi: 10.1126/SCIENCE.AAX4851/SUPPL_FILE/AAX4851-PHILLIPS-SM.PDF

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M.T., Guerri, G., Nannipieri, P. (2009). Extracellular DNA in soil and sediment: fate and ecological relevance. Biology and Fertility of Soils, 45, 219–235.

Porco, D., Bedos, A., Penelope, G., Janion, C., Skarżyński, D., Stevens, M. I., … Deharveng, L. (2012). Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebrate Systematics, 26, 470-477.

R Core Team (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Raffl, C., Mallaun, M., Mayer, R., and Erschbamer, B. (2006). Vegetation Succession Pattern and Diversity Changes in a Glacier Valley, Central Alps, Austria. Arctic, Antarctic, and Alpine Research 38:421-428.

Raso, L., Sint, D., Mayer, R., Plangg, S., Recheis, T., Brunner, S., Kaufmann, R., and Traugott, M. (2014). Intraguild predation in pioneer predator communities of alpine glacier forelands. Molecular Ecology 23:3744-3754.

Ricotta, C., Acosta, A. T. R., Caccianiga, M., Cerabolini, B. E. L., Godefroid, S. and Carboni, M. (2020). From abundance-based to functional-based indicator species. Ecological Indicators 118:106761.

Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B. E. L., and Pavoine, S. (2016). Measuring the functional redundancy of biological communities: a quantitative guide. Methods in Ecology and Evolution 7:1386-1395.

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(5), 1091–1108. doi: 10.1111/mec.13051

Rime, T., Hartmann, M., and 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.

Rissanen, A. J., Kurhela, E., Aho, T., Oittinen, T., & Tiirola, M. (2010). Storage of environmental samples for guaranteeing nucleic acid yields for molecular microbiological studies. Applied Microbiology and Biotechnology, 88(4), 977–984. doi: 10.1007/s00253-010-2838-2

Rosero, P., Crespo-Pérez, V., Espinosa, R., Andino, P., Barragán, Á., Moret, P., … Cauvy-Fraunié, S. (2021). Multi-taxa colonisation along the foreland of a vanishing equatorial glacier. Ecography, 44(7), 1010–1021. doi: 10.1111/ecog.05478

Rosvold, J. (2016). Perennial ice and snow-covered land as important ecosystems for birds and mammals. Journal of Biogeography 43:3-12.

Roy, J., Mazel, F., Sosa-Hernández, M. A., Dueñas, J. F., Hempel, S., Zinger, L., & Rillig, M. C. (2019). The relative importance of ecological drivers of arbuscular mycorrhizal fungal distribution varies with taxon phylogenetic resolution. New Phytologist, 224, 936-948.

Rydgren, K., Halvorsen, R., Töpper, J. P., and Njøs, J. M. (2014). Glacier foreland succession and the fading effect of terrain age. Journal of Vegetation Science 25:1367-1380.

Sadaka, N., & Ponge, J. F. (2003). Soil animal communities in holm oak forests: influence of horizon, altitude and year. European Journal of Soil Biology, 39(4), 197–207. doi: 10.1016/J.EJSOBI.2003.06.001

Schlegel, J., and Riesen, M. (2012). Environmental gradients and succession patterns of carabid beetles (Coleoptera: Carabidae) in an Alpine glacier retreat zone. Journal of Insect Conservation 16:657-675.

Schloss, P. D. (2021). Amplicon sequence variants artificially split bacterial genomes into separate clusters. mSphere, 6, e00191-00121.

Schmidt, S. K., Nemergut, D. R., Darcy, J. L., and Lynch, R. (2014). Do bacterial and fungal communities assemble differently during primary succession? Molecular Ecology 23:254-258.

Schmidt, S. K., Reed, S. C., Nemergut, D. R., Grandy, A. S., Cleveland, C. C., Weintraub, M. N., …, Martin, A.M. (2008). The earliest stages of ecosystem succession in high-elevation (5000 metres above sea level), recently deglaciated soils. Proceedings of the Royal Society B-Biological Sciences 275:2793-2802.

Schnell, I.B., Bohmann, K., Gilbert, M.T.P. (2015a). Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources, 15, 1289–1303.

Schumann, K., Gewolf, S., and Tackenberg, O. (2016). Factors affecting primary succession of glacier foreland vegetation in the European Alps. Alpine Botany 126:105-117.

Schütte, U. M. E., Abdo, Z., Bent, S. J., Williams, C. J., Schneider, G. M., Solheim, B., & Forney, L. J. (2009). Bacterial succession in a glacier foreland of the High Arctic. The ISME Journal 2009 3:11, 3(11), 1258–1268. doi: 10.1038/ismej.2009.71

Schweizer, S. A., Hoeschen, C., Schlüter, S., Kögel-Knabner, I., & Mueller, C. W. (2018). Rapid soil formation after glacial retreat shaped by spatial patterns of organic matter accrual in microaggregates. Global Change Biology, 24(4), 1637–1650. doi: 10.1111/GCB.14014

Sint, D., Kaufmann, R., Mayer, R., and Traugott, M. (2019). Resolving the predator first paradox: Arthropod predator food webs in pioneer sites of glacier forelands. Molecular Ecology 28:336-347.

Smiraglia, C., and Diolaiuti, G. A. (2015). The New Italian Glacier Inventory. Ev-K2-CNR Publ., Bergamo, Italy.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods, 11(1), 54–71. doi: 10.1037/1082-989X.11.1.54

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., … Hellström, M. (2017). Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution, 8(5), 635–645. doi: 10.1111/2041-210X.12683

Stibal, M., Bradley, J. A., Edwards, A., Hotaling, S., Zawierucha, K., Rosvold, J., …, Anesio, A. M. (2020). Glacial ecosystems are essential to understanding biodiversity responses to glacier retreat. Nature Ecology & Evolution 4:686-687.

Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics (Oxford, England), 24(12), 1461–1462. doi: 10.1093/bioinformatics/btn209

Taberlet, P., Prud'homme, S.M., Campione, E., et al. (2012) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology, 21, 1816-1820.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, É. (2018). Environmental DNA: For Biodiversity Research and Monitoring. In Environmental DNA: For Biodiversity Research and Monitoring. doi: 10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21(8), 2045–2050. doi: 10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35, e14.

Taberlet, P., Prud'homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., … Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology, 21, 1816-1820.

Tampucci, D., Gobbi, M., Boracchi, P., Cabrini, E., Compostella, C., Mangili, F., Marano, G., Pantini, P., and Caccianiga, M. (2015). Plant and arthropod colonisation of a glacier foreland in a peripheral mountain range. Biodiversity 16:213-223.

Tatangelo, V., Franzetti, A., Gandolfi, I., Bestetti, G., & Ambrosini, R. (2014). Effect of preservation method on the assessment of bacterial community structure in soil and water samples. FEMS Microbiology Letters, 356(1), 32–38. doi: 10.1111/1574-6968.12475

Tenan, S., Maffioletti, C., Caccianiga, M., Compostella, C., Seppi, R., and Gobbi, M. (2016). Hierarchical models for describing space-for-time variations in insect population size and sex-ratio along a primary succession. Ecological Modelling 329:18-28.

Ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology, 67(5), 1167–1179. doi: 10.2307/1938672

Thuiller, W., Gravel, D., Ficetola, G. F., Lavergne, S., Munkemuller, T., Pollock, L. J., Zimmermann, N. E., and Mazel, F. (2020). Productivity begets less phylogenetic diversity but higher uniqueness than expected. Journal of Biogeography 47:44-58.

Tom, B. (1990) The Maturity Index: An Ecological Measure of Environmental Disturbance Based on Nematode Species Composition. Oecologia, 83, 14-19.

Valkiūnas, G., Atkinson, C. T., Bensch, S., Sehgal, R. N., & Ricklefs, R. E. (2008). Parasite misidentifications in GenBank: how to minimize their number? Trends in Parasitology, 24, 247-248.

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., …, Crowther, T. W. (2019) Soil nematode abundance and functional group composition at a global scale. Nature, 572, 194-198.

Vater, A. E. (2012). Insect and arachnid colonization on the Storbreen glacier foreland, Jotunheimen, Norway: Persistence of taxa suggests an alternative model of succession. Holocene 22:1123-1133.

Vater, A. E., and Matthews, J. A. (2015). Succession of pitfall-trapped insects and arachnids on eight Norwegian glacier forelands along an altitudinal gradient: Patterns and models. Holocene 25:108-129.

Vega, E. N., Navarro, M. B., and Modenutti, B. (2020). Goose and hare faeces as a source of nutrients and dissolved organic matter for bacterial communities in the newly formed proglacial lake Ventisquero Negro (Patagonia, Argentina). Hydrobiologia 847:1479-1489.

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S. Fourth Edition. Springer, New York.

Walker, L. R., Wardle, D. A., Bardgett, R. D., and Clarkson, B. D. (2010). The use of chronosequences in studies of ecological succession and soil development. Journal of Ecology 98:725-736.

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature, 345(6270), 63–65. doi: 10.1038/345063a0

Wei, Z.-G., Zhang, X.-D., Cao, M., Liu, F., Qian, Y., & Zhang, S.-W. (2021). Comparison of methods for picking the operational taxonomic units from amplicon sequences. Frontiers in Microbiology, 12, 644012.

Weiher, E., Keddy, P. (1999) Ecological Assembly Rules: Perspectives, Advances, Retreats. Cambridge University Press, Cambridge.

White, H.J., León-Sánchez, L., Burton, V.J.,' Cameron, E. K., Caruso, T., Cunha, L., …, Caplat, P. (2020) Methods and approaches to advance soil macroecology. Global Ecology and Biogeography, 29, 1674-1690.

Whitesides, C. J., and Butler, D. R. (2016). Bioturbation by gophers and marmots and its effects on conifer germination. Earth Surface Processes and Landforms 41:2269-2281.

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

Wietrzyk-Pełka, P., Rola, K., Szymański, W., & Węgrzyn, M. H. (2020). Organic carbon accumulation in the glacier forelands with regard to variability of environmental conditions in different ecogenesis stages of High Arctic ecosystems. Science of The Total Environment, 717, 135151. doi: 10.1016/J.SCITOTENV.2019.135151

Williams, M. W., Brooks, P. D., and Seastedt, T. (1998). Nitrogen and Carbon Soil Dynamics in Response to Climate Change in a High-elevation Ecosystem in the Rocky Mountains, U.S.A. Arctic and Alpine Research 30:26-30.

Williams, P. J., Hooten, M. B., Esslinger, G. G., Womble, J. N., Bodkin, J. L., and Bower, M. R. (2019). The rise of an apex predator following deglaciation. Diversity and Distributions 25:895-908.

Zawierucha, K., Porazinska, D.L., Ficetola, G.F., Ambrosini, R., Baccolo, G., Buda, J., …,Takeuchi, N. (2021) A hole in the nematosphere: tardigrades and rotifers dominate the cryoconite hole environment, whereas nematodes are missing. Journal of Zoology, 313, 18-36.

Zellweger, F. , De Frenne, P., Lenoir, J., Vangansbeke, P., Verheyen, K., Bernhardt-Römermann, M., Coomes, D. (2020). Forest microclimate dynamics drive plant responses to warming. Science 368:772.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., …, Cogley, J. G. (2019). Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. Nature, 568, 382–386. doi: org/10.1038/s41586-019-1071-0

Zhao, H., Zheng, W., Zhang, S., Gao, W., & Fan, Y. (2021). Soil microbial community variation with time and soil depth in Eurasian Steppe (Inner Mongolia, China). Annals of Microbiology 2021 71:1, 71(1), 1–12. doi: 10.1186/S13213-021-01633-9

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., … Taberlet, P. (2019a). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology, Vol. 28, pp. 1857–1862. doi: 10.1111/mec.15060

Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., … Taberlet, P. (2016). Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biology & Biochemistry, 96, 16-19.

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., … Chave, J. (2019b). Body size determines soil community assembly in a tropical forest. Molecular Ecology, 28(3), 528–543. doi: 10.1111/mec.14919

Zumsteg, A., Luster, J., Göransson, H., Smittenberg, R. H., Brunner, I., Bernasconi, S. M., … Frey, B. (2011). Bacterial, Archaeal and Fungal Succession in the Forefield of a Receding Glacier. Microbial Ecology 2011 63:3, 63(3), 552–564. doi: 10.1007/S00248-011-9991-8