Towards a better understanding of soil multi-trophic biodiversity through the use of environmental DNA metabarcoding

Irene Calderon Sanou

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant
Spécialité : Biodiversité-Ecologie-Environnement
Unité de recherche : Laboratoire d'ECologie Alpine

Vers une meilleure compréhension de la biodiversité multi-trophique du sol grâce à l'utilisation du métabarcodage de l'ADN environnemental

Towards a better understanding of soil multi-trophic biodiversity through the use of environmental DNA metabarcoding

Présentée par :

Irene CALDERON SANOU

Direction de thèse :

Wilfried THUILLER Directeur de recherche 1 classe, directeur adjoint du LECA, Université Grenoble Alpes	Directeur de thèse
Tamara MUNKEMULLER Université Grenoble Alpes	Co-encadrante de thèse

Rapporteurs :

Marc BUEE
DIRECTEUR DE RECHERCHE, INRAE Centre Grand Est - Nancy
Jérôme CHAVE
DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Ouest

Thèse soutenue publiquement le **20 juin 2022**, devant le jury composé de :

Wilfried THUILLER DIRECTEUR DE RECHERCHE, CNRS délégation Alpes	Directeur de thèse
Marc BUEE DIRECTEUR DE RECHERCHE, INRAE Centre Grand Est - Nancy	Rapporteur
Jérôme CHAVE DIRECTEUR DE RECHERCHE, CNRS délégation Occitanie Ouest	Rapporteur
Nathalie FROMIN CHARGE DE RECHERCHE, CNRS délégation Occitanie Est	Examinatrice
Eric COISSAC MAITRE DE CONFERENCES HDR, Université Grenoble Alpes	Examinateur
François MUNOZ PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président

Invités :

Tamara MüNKEMüLLER
CHARGE DE RECHERCHE, CNRS

Table of Contents

ABSTRACT

Although soil organisms represent one-quarter of the whole biodiversity on earth, our current understanding of the main drivers of soil biodiversity along environmental gradients is mostly restricted to a limited set of aboveground macro-organisms. In light of increasing global threats to ecosystems, the inclusion of soil organisms into macroecological studies is crucial to improve predictions of ecological responses of terrestrial ecosystems to global changes and support their conservation. Moreover, multitrophic approaches that account for multiple groups of interacting organisms in the ecosystem allow a more holistic understanding of soil biodiversity and its drivers.

In my PhD, I aimed at getting a better understanding of the response of soil multitrophic diversity to rapid environmental changes at regional and local scales, by combining soil environmental DNA (eDNA) metabarcoding data, and mathematical and statistical tools derived from network theory, and food web ecology.

The thesis is developed in five chapters. First, since most of soil data rely on eDNA metabarcoding approaches, I needed to clarify the uncertainties underlying the use of eDNA in empirical analyses. In my first chapter, I thus benchmarked the different curation steps commonly used when using eDNA and tested their influence on specific ecological analyses. In particular, I showed that the use of Shannon diversity led to more reliable results from different ecological analyses. I also proposed a roadmap and decision tree to optimise the curation steps in function of the ecological question. Second, to simplify the complexity of the soil diversity, I developed a workflow to categorize taxonomically annotated sequences into trophic groups and to further build soil food webs (chapter 2). Next, I studied how soil multitrophic diversity vary along environmental gradients using a large-scale biodiversity observatory in the French Alps (chapters 3 and 4). In the third chapter, I conducted a comparative analysis across major soil trophic groups to assess the drivers of soil diversity in the light of well-known macro-ecological hypotheses applied specifically here to the soil context. I found that the energy and physiological tolerance hypotheses were particularly relevant in explaining the spatial variation in soil biodiversity. In the fourth chapter, I described how soil food web structure and composition varied along environmental gradients and assessed the main drivers of this variation. Finally, using eDNA soil data from subarctic birch forests of Northern Norway, I showed that the effect of severe moth outbreaks has cascaded locally from plant communities to the entire soil food web, creating a shift in the ecosystem state (chapter 5).

I believe my PhD has opened new research avenues in the understanding of multi-trophic soil biodiversity. Zooming out from the species level to a meaningful definition of trophic and functional groups allows a larger inclusion of multiple groups and to reach the ultimate goal of understanding all-in-end soil biodiversity distribution and composition.

RÉSUMÉ

Bien que les organismes du sol représentent un quart de l'ensemble de la biodiversité sur terre, notre compréhension actuelle des principaux moteurs de la biodiversité du sol le long des gradients environnementaux est principalement limitée à un ensemble restreint de macro-organismes de surface. À la lumière des menaces mondiales croissantes qui pèsent sur les écosystèmes, l'inclusion des organismes du sol dans les études macroécologiques est cruciale pour améliorer les prévisions des réponses écologiques des écosystèmes terrestres aux changements globaux et pour soutenir leur conservation. De plus, les approches multi-trophiques qui tiennent compte de plusieurs groupes d'organismes en interaction dans l'écosystème permettent une compréhension plus holistique de la biodiversité du sol et de ses moteurs.

Dans ma thèse, j'ai cherché à mieux comprendre la réponse de la diversité multi-trophique du sol aux changements environnementaux rapides à l'échelle régionale et locale, en combinant les données de métabarcodage de l'ADN environnemental du sol (ADNe) et les outils mathématiques et statistiques dérivés de la théorie des réseaux et de l'écologie des réseaux trophiques.

La thèse est développée en quatre cinq. Tout d'abord, puisque la plupart des données sur les sols reposent sur des approches de métabarcodage de l'ADNe, j'ai dû clarifier les incertitudes qui sous-tendent l'utilisation de l'ADNe dans les analyses empiriques. Dans mon premier chapitre, j'ai donc évalué les différentes étapes de curation couramment utilisées lors de l'utilisation d'ADNe et testé leur influence sur des analyses écologiques spécifiques. En particulier, j'ai montré que l'utilisation de la diversité de Shannon conduisait à des résultats plus fiables pour différentes analyses écologiques. J'ai également proposé une feuille de route et un arbre de décision pour optimiser les étapes de nettoyage des données en fonction de la question écologique. Ensuite, pour simplifier la complexité de la diversité du sol, j'ai développé un workflow pour catégoriser les séquences annotées taxonomiquement en groupes trophiques et pour construire les réseaux trophiques du sol (chapitre 2). Ensuite, j'ai étudié comment la diversité multi-trophique du sol varie le long de gradients environnementaux en utilisant un observatoire de la biodiversité à grande échelle dans les Alpes françaises (chapitres 3 et 4). Dans le troisième chapitre, j'ai mené une analyse comparative entre les principaux groupes trophiques du sol afin d'évaluer les moteurs de la diversité du sol à la lumière d'hypothèses macro-écologiques bien connues, appliquées ici spécifiquement au contexte du sol. J'ai constaté que les hypothèses de énergie et de tolérance physiologique étaient particulièrement pertinentes pour expliquer la variation spatiale de la biodiversité des sols. Dans le quatrième chapitre, j'ai décrit comment la structure et la composition du réseau trophique du sol varient le long des gradients environnementaux et j'ai évalué les principaux facteurs de cette variation. Enfin, à l'aide de données pédologiques d'ADNe provenant de forêts de bouleaux subarctiques du nord de la Norvège, j'ai montré que l'effet de graves épidémies de chenilles s'est propagé localement des communautés végétales à l'ensemble du réseau trophique du sol, créant un changement dans l'état de l'écosystème (chapitre 5).

Je pense que mon doctorat a ouvert de nouvelles voies de recherche dans la compréhension de la biodiversité multi-trophique des sols. Passer du niveau de l'espèce à une définition significative des groupes trophiques et fonctionnels permet d'inclure davantage de groupes multiples et d'atteindre l'objectif ultime de comprendre la distribution et la composition de la biodiversité du sol dans son ensemble.

GENERAL INTRODUCTION

Biodiversity encompasses the variety of life in all its forms, but the concept of biodiversity often evokes the macro-organisms that are most visible to the human eye and the most charismatic, such as mammals, birds, plants or butterflies. However, these macro-organisms constitute only a small fraction of the biodiversity that can be found on Earth. A large part of the organisms that make up the Earth's biodiversity escape our eyes either because they are microscopic in size or because they live hidden in elusive environments (e.g., the depths of the ocean, the forest canopy, the soil). Soils contain much of this hidden biodiversity, harbouring as much as a quarter of the species described on Earth. Although once considered a black box, general awareness of the importance of soil biodiversity has increased in recent decades, especially because of its fundamental role in the functioning of terrestrial ecosystems and nature's contribution to people.

The increased awareness of soil biodiversity has also highlighted gaps in scientific knowledge. One important gap concerns our understanding of how soil biodiversity, including its richness, composition and functional linkages, is structured across large spatial scales and what are its main drivers. Indeed, ecological disciplines that seek to understand the main drivers of biodiversity (i.e., macroecology, biogeography, community ecology) have mostly focused on aboveground macro-organisms, but it remains unclear whether the hypotheses tested on aboveground organisms are valid for belowground soil diversity. This fundamental knowledge is an essential building block in the process of preventing or redressing the biodiversity crisis and the threats to ecosystem integrity and functioning caused by global changes. Yet, for a number of soil organisms, this fundamental knowledge remains unresolved. Describing and understanding how the whole soil communities respond to environmental stress is necessary to predict future changes and identify critical transitions and effects on ecosystem functioning.

In response to this gap, research on soil biodiversity has increased in recent years (including the time when I undertook this PhD), and so has our understanding of the processes that shape soil biodiversity across spatial scales. This PhD is part of this common effort to improve our understanding of soil biodiversity patterns. We build on DNA metabarcoding analysis of environmental samples, a method that has recently emerged unravelling novel cross-taxon macroecological patterns for soil, to answer the following questions (1) How can we make better use of environmental DNA metabarcoding data to study soil biodiversity in its totality and integrity? (2) How do soil multi-trophic assemblages vary in space and are structured in response to the environment? We build on the existing theoretical framework, primarily designed for aboveground organisms, and apply it to soil biodiversity.

1. Integration of soil biodiversity into macroecological studies

1.1. Learning from diversity patterns

Biodiversity has many dimensions, including the diversity 'within species, between species, and of ecosystems' (UN of Convention on Biological Diversity). Biodiversity comprises thus not only species richness, but multiple dimensions describing different ways of relating living organisms, such as the genetic, phylogenetic, functional, interaction and trophic diversity of ecological communities (Naeem *et al.* 2012). Understanding what controls the structure of biodiversity, in all its dimensions, across spatial and temporal scales is of central interest for ecologists. At the end of the 18th century, Alexander von Humboldt introduced the first scoops that would give rise to disciplines such as biogeography, macroecology and community ecology, through his expeditions in the quest to understand what determines the distribution of life on Earth. Since then, and as a result of decades of research, the study of the spatial variation of biodiversity at different scales gave rise to a large number of observable macroecological patterns such as latitudinal patterns (Hillebrand 2004), altitudinal patterns (McCain & Grytnes 2010), the species-area relationship (Drakare *et al.* 2006), and the distance decay of similarity (Nekola & White 1999). Diversity patterns are at the origin of the main rules in ecology that determine the structure of biodiversity and its formation (Gaston 2000; Pontarp *et al.* 2019; Rahbek *et al.* 2019).

Identifying patterns of diversity and the mechanisms responsible for those patterns remains topical and challenging in ecology. Contemporary researchers interested in this question are driven not only by curiosity and intrigue to understand the spatial organization of nature but recognise also its essential importance in the quest to predict the consequences of global changes (Thuiller *et al.* 2013). Predicting how current global changes such as climate warming and changes in disturbance regimes affect the biodiversity of our planet, needs to first understand what are the environmental drivers of biodiversity. In addition, understanding the spatial structure of biodiversity is necessary to build conservation and mitigation strategies that are more than urgent in the face of the biodiversity crisis (IPBES 2019; Pollock *et al.* 2020).

However, the ecological theories aiming to explain biodiversity patterns have mostly been tested on aboveground macro-organisms, and rarely on soil organisms, with the exception of the strong developments in microbial macroecology in the last decades (O'Malley 2007; Soininen 2012), and the special attention given to some macroinvertebrate groups like earthworms (Decaëns 2010; Rutgers *et al.* 2016; Phillips *et al.* 2019). Soil biodiversity becomes thus a missing piece in our understanding of how biodiversity is structured on Earth. In light of increasing global threats to ecosystems, several studies and papers advocate the inclusion of

soil organisms into macroecological studies to improve predictions of ecological responses of the whole ecosystems to global changes and support their conservation (Cameron *et al.* 2018; Shade *et al.* 2018; Guerra *et al.* 2020, 2021).

1.2. The "missing" patterns of soil biodiversity

Soils harbour a large complexity of living organisms belonging to all kingdoms of life, ranging in size from micro-organisms such as bacteria and fungi to macro-organisms such as earthworms and insects, and with diverse life strategies (Orgiazzi *et al.* 2016)(Fig. 1). Soil biodiversity encompasses a significant proportion of the described species on terrestrial ecosystems. Moreover, soil biodiversity not only comprises the number of species inhabiting soils and their abundances but also their genetic, functional and trophic diversity. Within this functional and trophic diversity, there are a number of groups with major implications for society, such as decomposers, essential for nutrient recycling in terrestrial ecosystems (Crowther *et al.* 2019), or pathogens, of general interest to agriculture and public health (Wall *et al.* 2015).

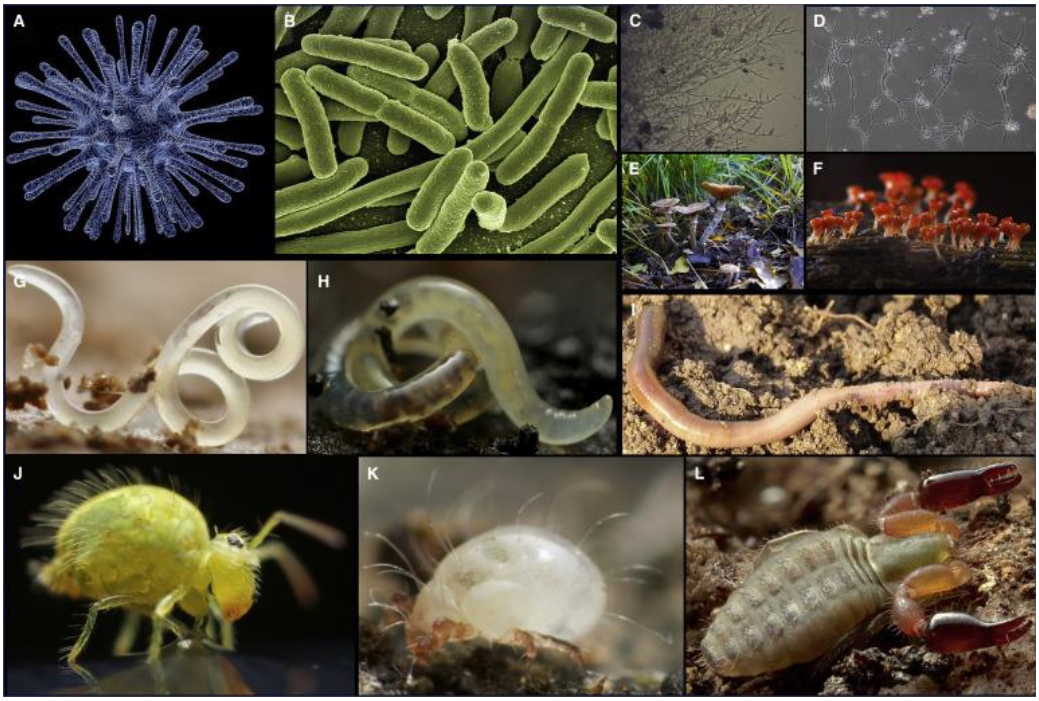


Figure 1. Soil biodiversity pictures illustrating the diversity of organisms living in soils**.** The pictures correspond to soil biota of different sizes including microorganisms represented by virus (A), bacteria (B), fungal hyphae (C), protist amoeba (D), fruiting bodies of fungi (E), protists slime mould (F), microfauna represented by nematodes (G), mesofauna represented by enchytraeids (H), springtails (J) and mites (K), and macrofauna represented by earthworms (I) and pseudoscorpions (L). Figure from Geisen *et al.* (2019b).

If soil biodiversity has traditionally been less studied in ecological disciplines such as macroecology, this is partly due to its cryptic nature, which makes it difficult to study (Geisen *et al.* 2019b; White *et al.* 2020). The study of soil-dwelling organisms mostly relies on specialized techniques of extraction that vary for organisms in different size categories (i.e., microbes, microfauna, mesofauna and macrofauna, Geisen *et al.* 2019a), and thus was historically limited by technological development for some groups, e.g. the microscope and molecular analyses for the study of micro-organisms diversity (Ferris *et al.* 2012). Additionally, morphological assessments are time-consuming and require a high level of taxonomic expertise, but the number of taxonomists dedicated to soil biota is limited, adding that normally one taxonomist is specialized in a single taxon and that taxonomists are unequally distributed across countries (~ 80% of taxonomists are based in northern countries, Gaston & May 1992). This limits the availability of community datasets at the species level and can create knowledge gaps in some geographic regions (Cameron *et al.* 2018; Guerra *et al.* 2020). In reality, the taxonomic diversity of soils is largely undescribed (Decaëns 2010). To this, can be added the difficulty of identifying organisms at immature states, as is the case for Acari and Diptera. Moreover, the scale of approach in soils can be different than the one used for aboveground macro-organisms, due to the high degree of heterogeneity at incredibly small grains that exist in soils (Ettema & Wardle 2002; Young & Crawford 2004). For these and other reasons, most soil biodiversity studies have been conducted at local scales and have focused on individual taxa, making difficult the generalization of spatial patterns for soil biodiversity (Decaëns 2010; White *et al.* 2020). While cryptic and elusive organisms are not exclusive to soil (e.g., aboveground leaf microbes), most soil organisms fall in this category, reducing the number of studies addressing soil biodiversity.

In order to gain a better understanding of how large-scale climatic variation or regional-scale environmental change affect soil biodiversity, we need to describe both the diversity of local communities (α-diversity) and the composition turnover between communities (β-diversity). Standard diversity indices allow to take into account quantitative data based on organism's abundance, and their phylogenetic or functional relationship (Chao *et al.* 2014). For soil organisms, the traditional measures of abundance can vary across taxonomic subfields and can represent a real challenge for some organisms, e.g., delimiting fungal individuals. Biomass or relative abundances retrieved form DNA sequencing methods could be more adequate measures of abundance to be compare across soil organisms from different kingdoms (Fierer *et al.* 2009; Shade *et al.* 2018).

Our ability to study soil biodiversity at large spatial scales has largely improved in the last decades with joint taxonomic efforts, the development of new sampling technologies (e.g., eDNA metabarcoding) and the increase of collaborative databases and initiatives focusing on soil taxa (e.g., Drilobase, Earth microbiome project, Global Soil Biodiversity Initiative) or functions (e.g., The Biological and Ecological Traits of Soil Invertebrates database, BactoTraits, FungalTraits). Large scale diversity patterns have thus recently been revealed for some soil organism groups (e.g., Tedersoo *et al.* 2014; Delgado-Baquerizo *et al.* 2018; Phillips *et al.* 2020), starting to unveil their environmental drivers. But, in order to have an integral vision of soil biodiversity, we need to understand not only the spatial distribution of certain representatives of soil biodiversity but to integrate the whole prism of soil organisms, and include not only their richness but also their taxonomic and functional structure. However, as important as it is to consider all possible taxa, it is also important to see these taxa not independently of each other but in interaction with each other (Albert *et al.* 2021).

2. A multitrophic approach to unifying biodiversity

2.1. The importance of multitrophic approaches

Given the complexity of the living world, most attempts to explore the causes of ecological diversity focus on single trophic levels and/or taxonomic groups, ignoring the added complexity of biotic interactions across different trophic levels (Seibold *et al.* 2018). Yet, the biodiversity of an ecosystem is structured across trophic levels that constantly interact through the flow of matter and energy, and thus a complete understanding of the general patterns and mechanisms that structure biodiversity needs to take these interactions into account (Seibold *et al.* 2018; Münkemüller *et al.* 2020; Thakur 2020). Hence, much seminal ecological work aiming at understanding the drivers of biodiversity is based on the development of the trophic structure of ecosystems (Lindeman 1942; Hutchinson 1959), and the same is true for the predictions of known hypotheses explaining the diversity of organisms. For example, the 'Energy-diversity hypotheses' predict that an increase in the amount of energy or resources available in the system promotes diversity across trophic levels (Wright 1983; Evans *et al.* 2005), and implies that the diversity of a trophic level is determined by the energy available at the lower trophic level. It is now recognized that trophic interactions play a major role in shaping the diversity of ecological communities over large spatial scales and that should be accounted for in macroecological studies (Guisan & Thuiller 2005; Gravel *et al.* 2011). It may even sound urgent given that

cascading effects of one trophic level to another could occur through trophic interactions, and this could result in rapid and irreversible state shifts of ecosystems (Estes *et al.* 2011; Scheffer *et al.* 2012).

In a multi-trophic approach, diversity can be addressed from two dimensions. A first dimension constitutes the diversity within trophic groups (e.g., species diversity). A second dimension constitutes the diversity across trophic groups (e.g., the number of trophic groups, the diversity of trophic interactions). This two-dimensional view of trophic networks resembles the concept of horizontal and vertical diversity (Duffy *et al.* 2007). Horizontal and vertical diversity can affect the functioning and stability of multi-trophic communities through different mechanisms and can respond differently to environmental changes or disturbances (Duffy *et al.* 2007; Kardol *et al.* 2016; Martinez-Almoyna *et al.* 2019; Zhao *et al.* 2019). In this thesis, we will be interested in studying soil multi-trophic communities through these two dimensions. The second dimension is also referred to as food web diversity or food web structure throughout this manuscript.

The multi-trophic complexity of ecological communities can be approached by the study of food webs. Food webs are complex networks of trophic interactions among species, 'trophospecies', guilds, functional or trophic groups, distributed across different trophic levels (Dunne 2006). The first representation of a food web dates back to Elton (1927, as cited in Tylianakis & Morris 2017), who classified the species into trophic groups having both similar functional roles within the food web and similar impacts on the environment (The *Eltonian* niche concept). Following an increase in the available documentation on the feeding behaviour of species, the representation of food webs has been largely developed in terms of diversity and resolution (Kéfi *et al.* 2012; O'Connor *et al.* 2020). In parallel, the incorporation of methods developed in network theory into the study of ecological networks has improved our ability to compare networks along environmental gradients (Pellissier *et al.* 2017; Tylianakis & Morris 2017; Botella *et al.* 2022). These advances are also valid for soil biodiversity and constitute a great opportunity to provide an integrative perspective in the study of soil biodiversity patterns.

2.2. The soil food web

The soil food web allows the unification of the very different taxa, functions and trophic levels that make up soil biodiversity, while accounting for the complex interactions between these groups. In the soil food web, organisms are categorized across trophic levels based on the resources they consume. Functional and/or phylogenetic information is often used to group soil

organisms into the trophic groups that will represent the nodes of the food web (Moore & de Ruiter 1991; Scheu 2002; Berg & Bengtsson 2007). The main basal resources of soil food webs can be classified into plants, soil organic matter and direct sources of energy for autotrophs such as solar and chemical energy. Energy is transferred from these basal resources through primary producers and consumers and to high trophic levels represented by predators and/or animal parasites (Fig 2). Different representations of the soil food web exist, from the highly influential version proposed by Hunt and colleagues (1987), to more recent and more resolutive elaborations that have emerged as a consequence of the increased empirical knowledge on the trophic relationships of soil organisms (e.g., Potapov 2022). Traditionally, energy flowing through the soil food web has been categorized into energy 'channels' based on the basal resource at the origin of the channel, e.g., green channel (from plants) vs brown channel (from organic matter), or fast channel (bacteria-based) vs low channel (fungi-based) (Moore & de Ruiter 1991; de Vries *et al.* 2013). Despite the fact that the terminology of 'channel' is widely used in the literature to relate food webs to functions or processes, recent evidence points out the existence of reticulate channels because groups in low trophic levels can feed on multiple energy channels (e.g. omnivore protists that feed on both bacteria and fungi), complexifying the separation of the food web into the traditional binary categories (Geisen *et al.* 2016; Potapov *et al.* 2021). For this thesis, I only referred to the concept of channels for discussion purposes, as the quantification of energy fluxes is out of the scope of this PhD.

The representation of the soil food web that I used for this thesis (Fig. 2) includes organisms ranging from microorganisms to macrofauna (thus excluding vertebrates and viruses) and includes various types of interactions, e.g., mutualisms, parasitism, predation, which all constitute trophic interactions as they represent a carbon transfer that is needed for the growth and development of the groups in the subsequent trophic levels. For example, plant symbionts such as mycorrhizal fungi are a major component of the soil food web and have a trophic interaction with plants as most of the carbon they obtained comes from this mutualistic association (Antunes & Koyama 2017).

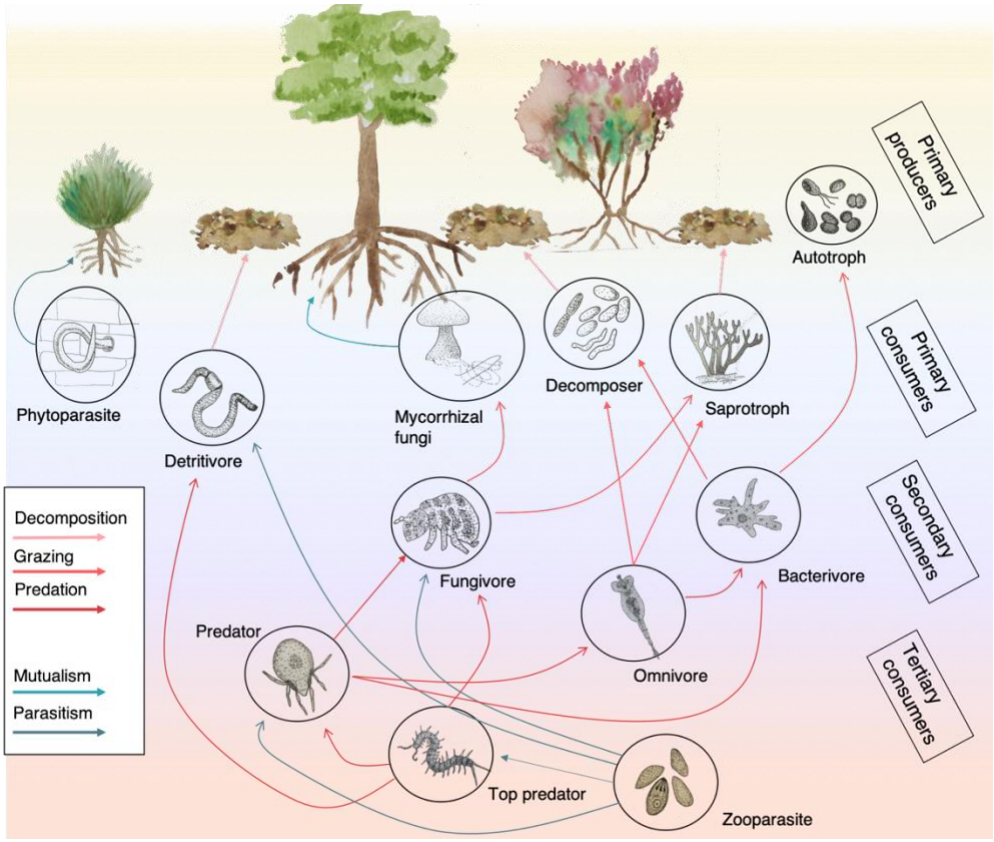


Figure 2. Representation of a soil food web showing the position of major soil trophic groups across trophic levels and the different types of trophic interactions. Basal resources of the soil food web are plants, organic matter and solar/chemical energy (for autotrophs).

3. Disentangling the drivers of diversity

3.1. Community assembling rules and ecological filters

Biodiversity patterns are driven by multiple ecological and evolutionary processes acting across spatial and temporal scales. The concept of ecological filters provides a conceptual framework to understand how different eco-evolutionary processes lead to the realization of local community assemblages (Keddy 1992; Cornwell *et al.* 2006)(Fig. 3A). From the global or regional pool of species, the local composition of a realized community is the consequence of both the dispersal ability of species and their biogeographic history, i.e., 'dispersal filter' (Sexton *et al.* 2009), and, the capacity of the species to establish and reproduce under the local environmental conditions, i.e., 'niche filter' (Cornwell *et al.* 2006). The local environment includes both the abiotic and the biotic environments. Both filters operate at different dimensions, i.e., geographic and ecological space respectively, are not necessarily hierarchical and are influenced by multiple interacting eco-evolutionary processes such as species physiology and biotic interactions (Thuiller *et al.* 2013). The main ecological processes

determining the diversity and composition of the different organisms that locally coexist are expected to differ between trophic and functional groups or taxa because of differences related to evolutionary history, dispersal traits, and habitat requirements (Hillebrand *et al.* 2001; De Bie *et al.* 2012).

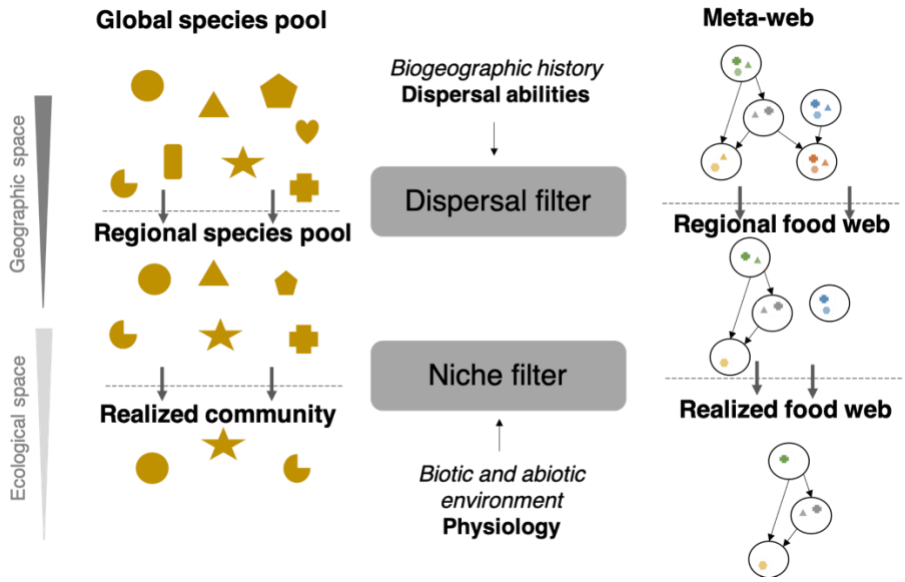


Figure 3. Ecological filters of community assembling processes from global or regional pool to local realized communities, at two levels of organisation, species within trophic groups (right) and food webs (left). The figure was adapted from Thuiller *et al.* (2013).

The concept of ecological filters can also be applied at the food web structure level (Pellissier *et al.* 2017; Tylianakis & Morris 2017)(Fig. 3B). In this sense, the composition of the realized local food web would be, in part, the result of the ecological filters acting on the taxa composing each of the trophic groups in the food web, thus the sum of the eco-evolutionary process shaping species diversity across the trophic groups locally coexisting. In addition, selection not only occurs on taxa but also on the realized interactions. In theory, this could be even partly independent of taxa, e.g. some interaction between partners occur just under certain abiotic conditions. Moreover, from a food web perspective, primary filtering out of species could lead to secondary "extinctions" or filtering out of other species dependent of the interaction (Gravel *et al.* 2011). In practice, the trophic interactions of a realized community also indicate the co-occurrence of two interacting groups, bringing thus information on how groups of species are co-selected by the environment or how they co-influence each other distributions. At this level of organization, the global pool of species is replaced by the metaweb, which represents the global or regional pool of trophic groups and their potential interactions (Dunne 2006;

Tylianakis & Morris 2017). The relative influence of the ecological filters on the structure of local food webs could allow evaluating to which extent the ecological processes acting on species and interactions translate into changes at a macroscopic scale of organization.

In this thesis, I did not evaluate directly the effect of dispersal limitations on the different soil trophic groups. The geographic space and its influence on soil biodiversity were indirectly accounted for in some chapters to control for spatial autocorrelation (chapters 3 and 5), and directly tested in chapter 4. Instead, I was principally interested in studying the effect of the abiotic and biotic environmental filters on soil trophic group diversity and food web structure. Testing the effect of ecological filters on natural communities can be complex. Ideally, it would be necessary to observe how community composition assembles under changing conditions. However, this could take an incredibly long time and effort to obtain unbiased results. An alternative is to use existing environmental gradients providing natural space-for-time settings to assess in situ responses to environmental change.

3.2. Analysing ecological filters along gradients

Ecologists conducting empirical research aim at describing the co-variation between diversity and environment to further link it with ecological theory explaining the mechanisms behind the resulting patterns of diversity (Münkemüller *et al.* 2020; Grainger *et al.* 2021). For this, empirical research builds on existing environmental gradients at different spatial scales that act as natural observatories to study the spatial distribution of biodiversity (see *Box 1. Importance of spatial scale*). Two of the commonly used environmental gradients, which are those used in this thesis, are the elevational and the disturbance gradients (Fig. 4). Elevational gradients are well suited to test empirically large-scale drivers of biodiversity as they encompass wide ranges of environmental gradients in abiotic and biotic conditions over a reduced spatial scale (Fig. 4A) (McCain & Grytnes 2010). Instead, disturbance gradients are widely used to understand the role of natural disturbances in maintaining biodiversity at local scales, because disturbances promote local heterogeneity, control spatio-temporal dynamics and drive successional trajectories (Fig.4B) (Thom & Seidl 2016).

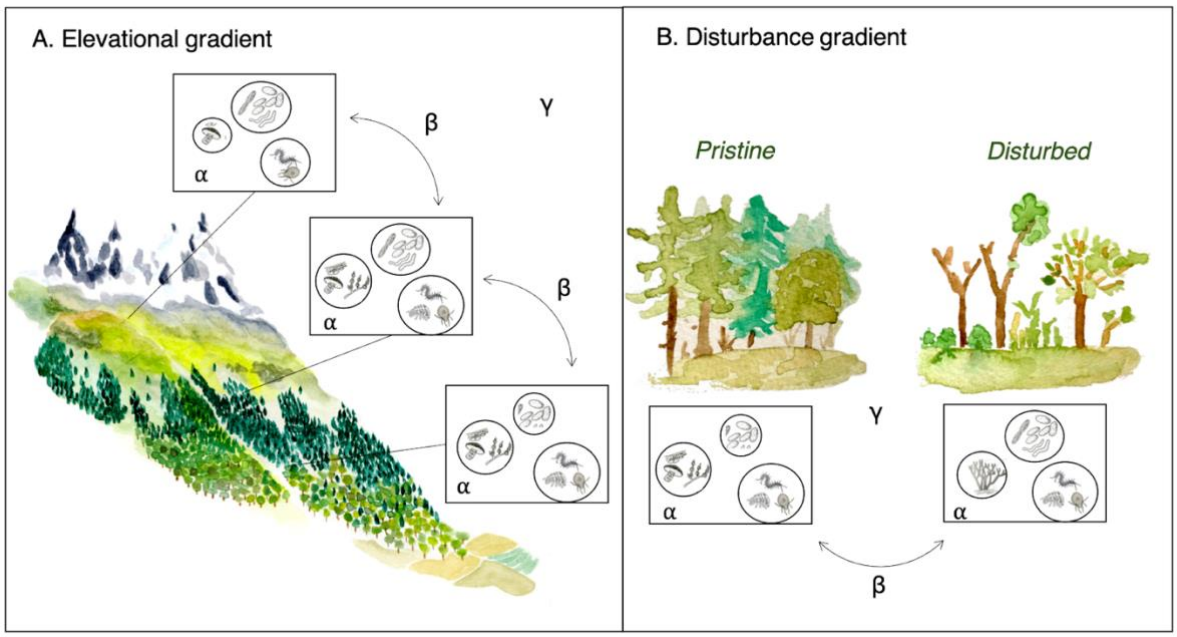


Figure 4. The two types of gradient that were used in this thesis: the elevational gradient (A) and the disturbance gradient (B). From ecological gradients, we can estimate the regional diversity (γ-diversity) representing the total diversity across the studied gradient(s), the local diversity (α-diversity) characterizing the diversity at a given location, and the compositional turnover of the community from one site to another (β-diversity). Mountain drawing in (A) was made by Camille Martinez-Almoyna.

Elevational patterns in diversity are commonly used to differentiate between competing hypotheses of diversity (e.g., Peters *et al.* 2016; Nottinghan *et al.* 2018), or to gain a better understanding of the potential effects of global changes on diversity (Sundqvist *et al.* 2013), among other multiple applications. Indeed, mountains have inspired ecologists through different generations and are at the origin of several biogeographical theories of biodiversity (Lomolino 2001). Examples of hypotheses that are commonly tested through elevational gradients are the energy-diversity hypothesis (Wright 1983) and the stress-diversity hypothesis (Grime 1973; Louthan *et al.* 2015). These hypotheses predict how diversity, but also biotic interactions, change in relation to available resources or abiotic conditions. The relative strength of biotic and abiotic filters can be also tested in the lenses of ecological theory across environmental gradients. For example, along stress gradients, competition filters are expected to be especially strong in benign conditions, while environmental filters are expected to be dominant under stressful conditions (Louthan *et al.* 2015). In stressful conditions, biotic interactions can play an important role trough facilitation, making the conditions more easy for the establishment of other species, e.g., plants can create microclimate favourable to microbes (Roy *et al.* 2013).

Box 1. Importance of spatial scale

It is widely recognized that the drivers of diversity and their relative importance vary with the spatial scale of the study. A spatial pattern of diversity depends on the characteristics of the study area, the size and spacing of the samples, and the organism under study. Two particular attributes of spatial scale in ecological studies are the spatial grain and the spatial extent (Guisan *et al.* 2017). The **spatial grain** corresponds to the unit of sampling or the area/volume covered by each data point. The **spatial extent** corresponds to the geographical space covered by the study, and is thus related to the range of the environmental gradients considered.

Soil are highly complex habitats with nested levels of heterogeneity. Spatial distribution of soil organisms can occur both vertically across soil layers, and horizontally, the latter being more documented. Thakur *et al.* (2020) proposed a framework describing spatial grain at which ecological theories can be studied representing different soil compartments: a coarse grain (S) where all organisms can be sampled, an intermediate grain (S') that can be represented by a hotspot such as the rhizosphere or the litter layer, where intermediate sized soil organisms can be sampled, and a fine grain or microsite (S'') ranging from the tip of a plant to a single aggregate where microorganisms are the main focus.

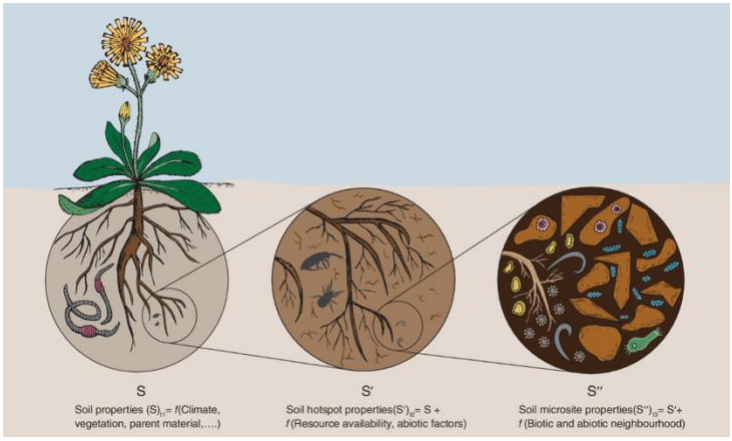


Figure from Thakur *et al.* (2020)

Soil biota are spatially structured over distances of tens to hundred meters and can present patchy distributions at the scale of centimeters to meters, depending on the organism. Using a nested spatial sampling design is thus recommended to explore the spatial aggregation of soil biota among a range of scales representing the heterogeneity of the landscape studied (Ettema & Wardle 2002).

In this thesis, where the aim was to sample the whole soil multi-trophic communities, a coarse spatial grain was preferred to detect spatial patterns over geographic areas ranging from 12 to 10,000 km^2.

Disturbance gradients had guided the understanding of the diversity-disturbance relationship, which also exerts a major influence on ecological theory. For example, the Intermediate Disturbance Hypothesis postulates that biodiversity peaks at intermediate levels of disturbance, where both colonist and climax species are able to coexist and decline at low and high levels of disturbance where they are mutually excluded by exploitative competition (Grime 1973; Horn 1974; Connell 1978). Also, for more than a century forest ecologists have investigated post-disturbance successional dynamics across disturbance gradients (Clements 1916). The importance of biotic interactions can also change along the succession process following a disturbance. For example, symbiotic associations with nitrogen-fixing bacteria and mycorrhiza can be crucial for plant establishment during early succession (Nara 2006). The diversity-disturbance relationship is also of interest in sustainable management (e.g., in practices that mimic natural disturbances, Harvey *et al.* 2002)) and in the prediction of the future scenarios for biodiversity and ecosystems under global change when accounting for the increase in the frequency and intensity of disturbances (Seidl *et al.* 2011).

As the construction and analyses of food webs become more accessible, its inclusion in macroecological studies increases in the search for understanding of what drives multitrophic community structure. There has been an increasing interest in evaluating empirically how the structure of food webs varies along environmental gradients (Pellissier *et al.* 2017). For this, one approach consists of summarising the structure of the food web through network metrics (e.g. connectance, modularity, vulnerability, etc) and related them to ecological processes (Braga *et al.* 2019), but the real meaning of these metrics or the ecological information they provide still in debate (Thompson *et al.* 2012). Other approaches consist at comparing the composition of local food webs across environmental gradients and assessing how much of the variance is explained by environmental or geographic predictors (Poisot *et al.* 2012; Pellissier *et al.* 2017). Changes in the structure of food webs along environmental gradients is of great interest as they are key to assess the functioning of the ecosystems and the stability of communities (Thompson *et al.* 2012; Eisenhauer *et al.* 2019).

4. Soil diversity patterns: state of knowledge

The concept of the 'black box' designating the soil compartment has begun to be left behind thanks to the increasing research illuminating our knowledge on soil biodiversity patterns and its drivers (Orgiazzi *et al.* 2016; FAO *et al.* 2020). It is now known that soil biota is spatially and temporally structured at different scales and respond to rules of community assemblage, such as niche-based process, although the relative contribution of different ecological processes is unclear and variable across soil organisms. The span of body sizes, phylogenetic history, life-history traits strategies and mobility capacities characterising soil organisms is reflected in a wide range of dispersal abilities and physiological adaptations to different environments. Yet, most studies looking at diversity patterns of soil biota focus on certain representatives of soil such as bacteria, fungi, earthworms and ants, while the knowledge we have for other organisms such as mites, enchytraeids, and rotifers remains scarce, making it difficult to draw general conclusions (Orgiazzi *et al.* 2016). In the following section, I aimed at providing a short but broad picture of the current knowledge on the drivers of soil diversity patterns at large scales.

The dispersal of soil organisms is particularly poorly understood, as measuring the dispersal of soil organisms might be very challenging (Ettema & Wardle 2002). Historically, microbes were thought to occur "everywhere" due to their high dispersibility and large population size, minimizing the importance given to geographic dispersal barriers in microbial macroecology

(O'Malley 2007). This view has been repeatedly challenged by the observations of a strong spatial structure of microbial communities across scales, raising questions about the importance of the dispersal constraints in shaping microorganisms diversity patterns (Zhou *et al.* 2013; Evans *et al.* 2017). Contrary, the dispersion of larger organisms such as soil meso- and macro-fauna is thought to be more limited due to the complexity of the soil environment limiting their movement, and their longer reproduction times (Ettema & Wardle 2002). In line with this, previous studies have shown that larger body sized organisms have more stochastic distribution patterns compared to smaller organisms that are less limited by dispersal or drift and are more strongly structured by the environment (De Bie *et al.* 2012; Zinger *et al.* 2018).

Several studies report the predominant importance of niche-based processes on the community assembly of soil biota (Decaëns 2010; Wu *et al.* 2011; Aslani *et al.* 2022). Multiple abiotic and biotic factors jointly determine the structure of soil communities. In the one hand, climatic factors such as temperature and precipitation combined with soil properties such as pH, organic matter content, C/N ratio and soil texture have been shown to co-vary with the diversity of soil taxa such as fungi (Tedersoo *et al.* 2014; Glassman *et al.* 2017), earthworms (Rutgers *et al.* 2016; Phillips *et al.* 2019), bacteria (Ramirez *et al.* 2014; Delgado-Baquerizo *et al.* 2018; Karimi *et al.* 2018), and protists (Bates *et al.* 2013; Fiore-Donno *et al.* 2020) at different spatial scales. On the other hand, the main biotic factors structuring soil communities are related to plant communities. The diversity, composition and biomass of plant communities have a major influence in shaping the soil environment. Plants provide direct resources to the soil through the roots, but also indirectly through the litter, and can shape soil biota habitats at different scales, from microhabitats to landscapes (Scherber *et al.* 2010; Eisenhauer *et al.* 2013; Roy *et al.* 2013; Prober *et al.* 2015; Leff *et al.* 2018). At larger spatial scales, the vegetation type characterizing an habitat, e.g., grassland vs forest, can be a determinant of the soil community structure (Ramirez *et al.* 2014; Fiore-Donno *et al.* 2020).

While macroecological studies mostly focus on soil taxonomic groups, the effect of abiotic and biotic factors on soil diversity can vary for different soil functional groups. For example, plant symbionts such as mycorrhizal fungi or parasitic nematodes are mainly structured by plant communities, following a co-distribution with their plant hosts and peak on diversity or abundance where their hosts are more diverse or abundant (Tedersoo *et al.* 2012; van den Hoogen *et al.* 2019; Wang *et al.* 2019). Other biotic factors that can influence soil diversity patterns are the biotic interactions occurring within the soil food web. Trophic interactions can affect soil communities and their multitrophic interactions through top-down or bottom-up controls (Scherber *et al.* 2010; Schuldt *et al.* 2017). Moreover, antagonistic interactions

between trophic groups within a same trophic level can limit their distribution, for example, between bacteria and fungi (Bahram *et al.* 2018) or between different fungal guilds (e.g., the 'Gadgil effect' between ectomycorrhizal fungi and saprotrophs, Bending 2003). Otherwise, past studies suggest that the community structure of soil fauna is not driven by competitive exclusion, although competition can take place for larger organisms such as earthworms in highly productive systems (reviewed in Decaëns 2010). The physical constraints and heterogeneity of the soil environment might also limit the direct interactions between soil organisms, including trophic interactions (Erktan *et al.* 2020). Indeed, the highly heterogeneous nature of the soil matrix provides a great diversity of niches that may allow high levels of local diversity (Nielsen *et al.* 2010).

At the global scale, diversity patterns have been recently described for a number of soil taxa, e.g., Fierer *et al.* 2009 for microbes, Phillips *et al.* 2019 for earthworms, van den Hoogen *et al.* 2019 nematodes, Oliverio *et al.* 2020 for protists, revealing that soil biodiversity might have different distribution patterns than aboveground macro-organisms biodiversity at this scale. However, few studies have analysed how different guilds or trophic groups within a taxa change across the latitudinal gradients (Bahram *et al.* 2018; Wang *et al.* 2019; van den Hoogen *et al.* 2020). At regional and local scales, divergent responses have been found across and within soil groups or taxa (Hendershot *et al.* 2017; George *et al.* 2019; Looby & Martin 2020). Overall, the existing studies suggest that local species diversity of soil organisms is highly limited by stress (e.g., resource availability, pH) and disturbance (e.g., fire, wind) (Decaëns 2010; Orgiazzi *et al.* 2016; Coyle *et al.* 2017; Glassman *et al.* 2017), similarly than for aboveground macro-organisms. Yet, it is still difficult to conclude at which extent the macroecological patterns of soil biota mirror those of above ground macro-organisms, because the existing publications are too scarce and biased to some representatives in both the aboveground and belowground compartments. Moreover, few studies have looked at soil biodiversity patterns in the light of ecological theory. Thakur and (2020) investigated how some main ecological theories could explain soil biodiversity patterns and found that less than 6% of studies addressing the reviewed theories included soil organisms. While some support was found for the tested ecological hypotheses, the studies diverged in the focal soil group and in the spatial scale considered, making it difficult to make comparisons and to draw robust conclusions.

The effect of environmental conditions and land use change on the soil food web structure has received a lot of attention with the goal of understanding how changes in soil food web structure affect ecosystem functions such as nutrient cycling and plant productivity (Hunt *et al.* 1987;

Berg & Bengtsson 2007; de Vries *et al.* 2013). Most of these studies have been conducted in arable systems (Berg & Bengtsson 2007; Morriën 2016), while less studies have investigated what are the community assemblage processes that drive soil food web structure and diversity in natural systems. The same abiotic and biotic factors driving soil biodiversity patterns may influence the structure of soil food webs across spatial and temporal scales, and thereby influence ecosystem functions. While the reconstruction of more resolutive food webs becomes more accessible, studying their spatial patterns across large spatial scales can bring new insights into the community assembling processes acting on the structure of soil multitrophic communities. For example, Morriën and colleagues (2017) studied the change in soil food webs structure during the restoration of an abandoned arable land, and found that the structure of the soil food web changed through time, becoming more connected, and that those changes were related to an enhanced efficiency of carbon uptake by the soil food web. Studying macroecological patterns with a food web approach gives the promise of getting a better understanding of soil biodiversity complexity and its multidimensionality (Eisenhauer *et al.* 2019). For example, the vulnerability of soil trophic interactions face to environmental changes could be identified allowing to better predict the cascading effects of global changes (Hedlund *et al.* 2004; Eisenhauer *et al.* 2013).

Despite the increasing body of literature dedicated to exploring soil diversity patterns, it is still difficult to draw general conclusions because most studies focused on single taxa or considered different spatial scales. Global diversity maps to assess latitudinal patterns are at the cutting-edge of the macroecological patterns of soil biodiversity that have emerged in recent years, but sampling gaps across the world, for example across tropical regions and northern latitudes, and also across taxa still constitute a challenge for these studies and their generalization at the global scale (Cameron *et al.* 2018; Guerra *et al.* 2020). Otherwise, regional studies with an intensive soil sampling and covering a wide range of environmental conditions can provide the resolution required to disentangle confounding effects of different predictors, leading to robust conclusions on the drivers of soil biodiversity (e.g., Rutgers *et al.* 2016; Karimi *et al.* 2018). In order to enlarge the sampling to several taxa at the same time and across large-spatial scales, the use of environmental DNA metabarcoding seems a promising opportunity, e.g., Wu *et al.* 2011; Bastida *et al.* 2020. Furthermore, we could go beyond describing the patterns of multiple taxa with eDNA data and bridge this data with other ecological meaningful frameworks, to include other dimensions of soil biodiversity such as the ones accounted for with a food web approach.

5. General methodology - Studying the complexity of soil biodiversity by combining soil eDNA metabarcoding and trophic or functional information.

5.1. Environmental DNA metabarcoding: a monitoring tool for soil biodiversity

Environmental DNA (eDNA) metabarcoding consists in amplifying and sequencing a genomic marker – or DNA barcode – of the DNA contained in environmental samples such as soil, water or faeces (Taberlet *et al.* 2018). In this thesis, we will focus on the eDNA coming from soil samples. The detection of DNA in a soil sample may occur because the living organism is present in the sample in an active or dormant stage (e.g. bacteria), or because traces remain to attest to the presence of the organism in the sample or in its vicinity revealing its presence in the community (e.g. carcasses, skin, faeces, body fluid, etc) (Barnes & Turner 2016). Thus, from an eDNA sample and combined with high-throughput sequencing, the diversity of the whole multitrophic community can be assessed. The rapid advancements of eDNA metabarcoding make it now possible to tackle unresolved questions that could not be addressed with traditional biodiversity surveys so far and to study far elusive taxa diversity, like soil microbial organisms, thereby improving our understanding of their community assembly processes and their main drivers at large scales (Wu *et al.* 2011; Drummond *et al.* 2015; Deiner *et al.* 2017). While eDNA metabarcoding was initially developed for micro-organisms (Tiedje *et al.* 1999), the ability of this method to efficiently monitor larger organisms is now recognized (Deiner *et al.* 2017). The use of eDNA metabarcoding alone or in complement with conventional methods has revealed that soil diversity is greater than previously thought, e.g. for protists (Geisen *et al.* 2016), fungi (Buée *et al.* 2009), earthworms (Bienert *et al.* 2012) and rotifers (Robeson *et al.* 2011).

5.2. Bias and pitfalls in eDNA metabarcoding data

In eDNA metabarcoding surveys, the obtained data consist of hundreds to millions of DNA sequencing reads from the multiple species co-occurring within soil samples. The process to obtain this list of sequences includes several methodological steps of fieldwork, laboratory treatment and bioinformatics processing, which can be subject to potential biases (Zinger *et al.* 2019). The basic steps are: 1) soil sampling in the field, 2) DNA extraction from soil 3) amplification of a specific DNA region with the use of a DNA marker 4) sequencing the DNA amplicons 5) processing the retrieved sequences through a bioinformatic pipeline. These different steps and the potential biases introduced at each step are described in *Box 2. DNA metabarcoding processing and sources of errors.* The bioinformatics pipeline intends to detect and correct these potential 'errors' that accumulate along with the eDNA processing and that

correspond to artefactual DNA that may lead to inflated diversity estimates (Bálint *et al.* 2016; Zinger *et al.* 2019). However, decisions regarding the bioinformatics process can be subjective, e.g., subject to laboratory or author personal preferences, and therefore there is much interest in understanding how variations in the bioinformatic pipeline can influence the ecological results across studies. This problem was studied in Chapter 1 of this thesis.

At the end of the bioinformatics process, a list of sequences and their abundances, i.e., the identity and number of sequencing read counts is obtained. The sequences are then usually grouped by DNA sequence similarity into Molecular Operational Taxonomic Units (MOTUs), which are next assigned to known taxa after comparison to reference databases when such databases are available. Curated data has thus the form of a community matrix that lists the taxonomically annotated MOTUs found in each environmental sample, and their sequencing read counts. Compared to conventional methods species are replaced by MOTUs, and species abundances are replaced by the number of sequencing reads. MOTUs are not necessarily transposable to the classic taxonomy due to the different evolutionary rates of DNA barcodes amongst clades (e.g., Schoch *et al.* 2012). Still, they are often considered pragmatic proxies of species in biodiversity assessments. In the same way, the number of sequencing reads can't be interpreted as a measure of species abundance. Some studies have found a positive correlation between the relative abundance of sequencing reads and the biomass across samples in experimental studies or through simulations (Deiner *et al.* 2017; Kelly *et al.* 2019). However, several factors related to the 'nature' of the eDNA (e.g., origin, stability) can affect this relationship (Barnes & Turner 2016).

Box 2. DNA metabarcoding processing and sources of errors

1. Soil sampling is carried out in a delimited sampling area of interest. The sampling strategy is designed to obtain diversity estimates that are representative of the sampling area, and may vary depending on the research question (e.g., diversity estimation per se vs. studying patterns of diversity). Within this sampling area, a number of biological replicates are collected. In terrestrial ecosystems, each replicate usually corresponds to a soil core.

2. DNA extraction is conducted on each biological replicate, preferably right after sampling to avoid DNA degradation or changes on microbial communities. Alternatively, they should be frozen, dried, or conserved in particular buffers when possible to inhibit any biological activity prior DNA extraction.

3. A particular barcoding region D is then amplified from the DNA extracted by using primers targeting priming sites common to the clade of interest (e.g., universal primers for bacteria, primers specific to plants) yet flanking regions variable enough to discriminate taxa. In our study, these primer pairs are equipped with a short sequence label in the 5' end of each primer (tag), of which combination is unique to each sample. This enables retrieving the sample of origin of each sequencing reads in downstream analyses. As most DNA is highly fragmented in the soil (**degraded DNA**), the DNA barcode must be short enough to be successfully amplified by PCR (Polymerase Chain Reaction). Working with relatively short barcodes is furthermore necessary due to the sequencing length limits of most HTS sequencers (e.g., ca. 150 bp for a HiSeq Illumina platform and 500 bp for a MiSeq Illumina platform). These constraints inherently come with a loss of the phylogenetic/taxonomic resolution in downstream analyses.

During PCR amplification, common **PCR errors** are produced by the DNA polymerase, which can substitute a nucleotide by another during DNA replication. Such variants from the genuine DNA fragments can also be amplified during subsequent PCR cycles, and subjected to new PCR errors. During PCR amplification, the formation of chimeric DNA fragments can also occur through recombination of two or more parent DNA fragments that are aborted extension products from an earlier cycle of PCR. **Chimeras** can represent a significant proportion of all produced amplicons. These two types of errors inherently inflate biodiversity estimates.

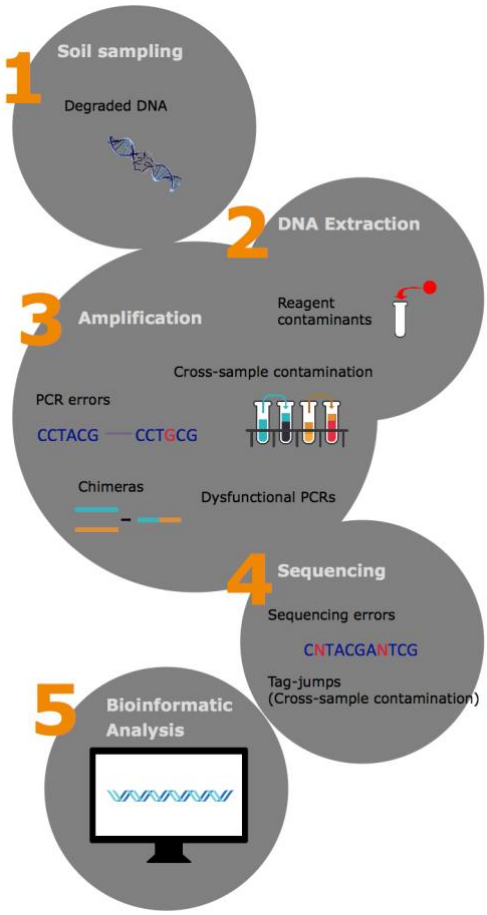
Because DNA extracts can contain PCR inhibitors (e.g., humic acids), some PCR can fail (**dysfunctional PCRs**) and produce a majority of artefactual amplicons (e.g. primer dimers, partial DNA fragments, etc.). This may inflate diversity estimates too and further lead to spurious ecological conclusions. To control for such artefacts, it is often recommended to conduct several technical PCR replicates for each biological sample.

At both 2. and 3. steps, **reagent contaminants** coming from consumable/equipment (e.g., DNA extraction or PCR commercial kits) or any external source can be introduced. Even if these contaminants come in low proportions and lab protocols are well respected, the use of universal primers and the high sensitivity of HTS may lead to a non-negligible amount of such contaminants in DNA metabarcoding data, and hence, to diversity inflation. The systematic sequencing of negative controls (i.e., blanks of DNA extraction and PCR amplification) enable identifying such contaminants. They are indeed better amplified and more detectable in negative controls, as they are not in competition with the DNA templates of interest.

Internal contaminants can also occur when DNA from one sample accidentally passed from one sample to another (e.g., through aerosol produced when pipetting). This phenomenon can be referred to as **cross-sample contamination** and may lead to false positives. Given the biases mentioned above, technical replicates and positive/negative controls are often conducted to evaluate them and improve the accuracy of downstream analysis.

4. After PCR amplifications, a sequencing library is prepared: all amplicons from the different samples are pooled together and ligated to sequencing primers. This library is then subjected to HTS sequencing (on Illumina sequencers in most cases). Either at the library preparation step, or during the sequencing, the formation of chimeras can occur across sequences belonging to different samples. This can lead to what is now often referred as "**tag-jumps**", "tag-switch", or "cross-talks", i.e., a chimera containing a genuine sequence, but for which the tag combination is artificial. In certain cases, these artificial tag combinations correspond to the tag combinations already associated to different samples in the experimental design, hence leading to a spurious assignment of this sequence to these samples. In downstream analyses, this bias looks like a cross-sample contamination.

Also, during sequencing process, the identification of some nucleotide can be ambiguous. These nucleotides will appear as Ns in the sequencing output and be interpreted as **sequencing errors**.



5. Once the sequencing data obtained, bioinformatics analyses are conducted to transform sequences data into a MOTU community matrix ready for ecological analyses. During the bioinformatics processing, the sequencing reads are, amongst other, reassign to their samples and assigned to a taxa by comparison with reference databases. It is also during this process, that the different errors accumulated during the previous steps can be removed using algorithms available in many different software. At present, it is often let to the discretion of the user to choose what are the appropriate data curation steps to include in the bioinformatic analysis.

Reference: Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: for biodiversity research and monitoring*. New York: Oxford University Press

Note: This box was modified from the Appendix 1 provided in the original publication corresponding to Chapter 1:
Calderón-Sanou I, *et al*. (2020) From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *J Biogeogr 47:193–20*

5.3. Making ecological sense from a bunch of sequences

So far, most studies using eDNA metabarcoding to monitor cross-kingdom biodiversity have focused on describing the diversity patterns of broad taxonomic groups, such as bacteria, eukaryotes, and fungi (e.g., Wu *et al.* 2011; Drummond *et al.* 2015; George *et al.* 2019). However, if we aim at obtaining an integrated knowledge of the functions of soil biodiversity, i.e., the ecological roles of soil organisms, we need to move away from pure taxon-based biodiversity assessments. There is a growing interest in applying trait-based approaches to the study of biodiversity with eDNA data, in which taxonomic annotations of sequences are complemented with information on traits (Crowther *et al.* 2014). Body size is a commonly used trait in soil ecology when dealing with the whole multitrophic community that can be used to disentangle ecological processes acting on soil communities such as dispersal limitations (Zinger *et al.* 2018). Functional or trophic information might also be needed if the aim is to build ecological networks such as food webs from eDNA metabarcoding data (Roslin & Majaneva 2016). The construction of heuristic food webs from eDNA data combined with the ecological knowledge of soil organisms seems a promising avenue. The application of this method from eDNA data has been limited (Compson *et al.* 2018), and to my knowledge not yet applied to soil organisms. Different databases with functional or trophic information on soil organisms exist, e.g. FUNGuild database for fungi (Nguyen *et al.* 2016), and might be useful for building heuristic soil food webs from eDNA data. The methodology used in the construction of soil food webs from eDNA data and its related challenges are addressed in Chapter 2.

OBJECTIVES

The aim of this PhD was to improve our understanding of how soil biodiversity responds to environmental changes through the use of eDNA metabarcoding.

This general objective is reached through two specific objectives, one addressing the methodological constraints and one addressing ecological questions:

1. To improve the use of environmental DNA metabarcoding data to get robust ecological conclusions and an integrative representation of soil biodiversity.
2. To test ecological hypotheses to understand how different dimensions of soil biodiversity (from MOTUs diversity to the soil food web structure) respond to the environment by using empirical data at different spatial extents and in different contexts.

STRUCTURE OF THE MANUSCRIPT

The methodological objectives were addressed in Chapters 1 and 2. First, because the diversity analyses in my PhD were based on eDNA metabarcoding data, we needed to gain a better understanding of the uncertainties associated with the use of eDNA metabarcoding in empirical analyses. Can we obtain reliable biodiversity patterns when using eDNA data? How sensitive are different ecological analyses (i.e., spatial diversity partitioning, distance-decay) to the methodological choices in the eDNA data curation process? Which are the curation steps that introduce more variability in the results? These questions were addressed in **Chapter 1**. After confirming that we could obtain reliable results using a stringent pipeline and adequate measures of diversity, we developed a workflow to categorize taxonomically annotated sequences into trophic groups and to further build a metaweb. This methodology is described in **Chapter 2** and was used in the further chapters.

The ecological questions were addressed in Chapters 3, 4 and 5. First, in **Chapter 3,** at the scale of the French Alps, we studied how the diversity within the different soil trophic groups responded to environmental changes based on the predictions of existing ecological hypotheses. Second, in **Chapter 4**, we quantified how the structure of soil food webs varied across several elevational transects in the French Alps and deciphered the importance of

geographic distance and environmental factors to explain spatial soil food web turnover. Third, at a smaller scale, we study the cascading effects of moth outbreaks on soil food webs along a disturbance gradient comprising undisturbed and defoliated forests in the Varanger region at Northeastern Norway (**Chapter 5**). Figure 5 describes the positioning of these different chapters according to the spatial scale studied and the biodiversity dimension considered.

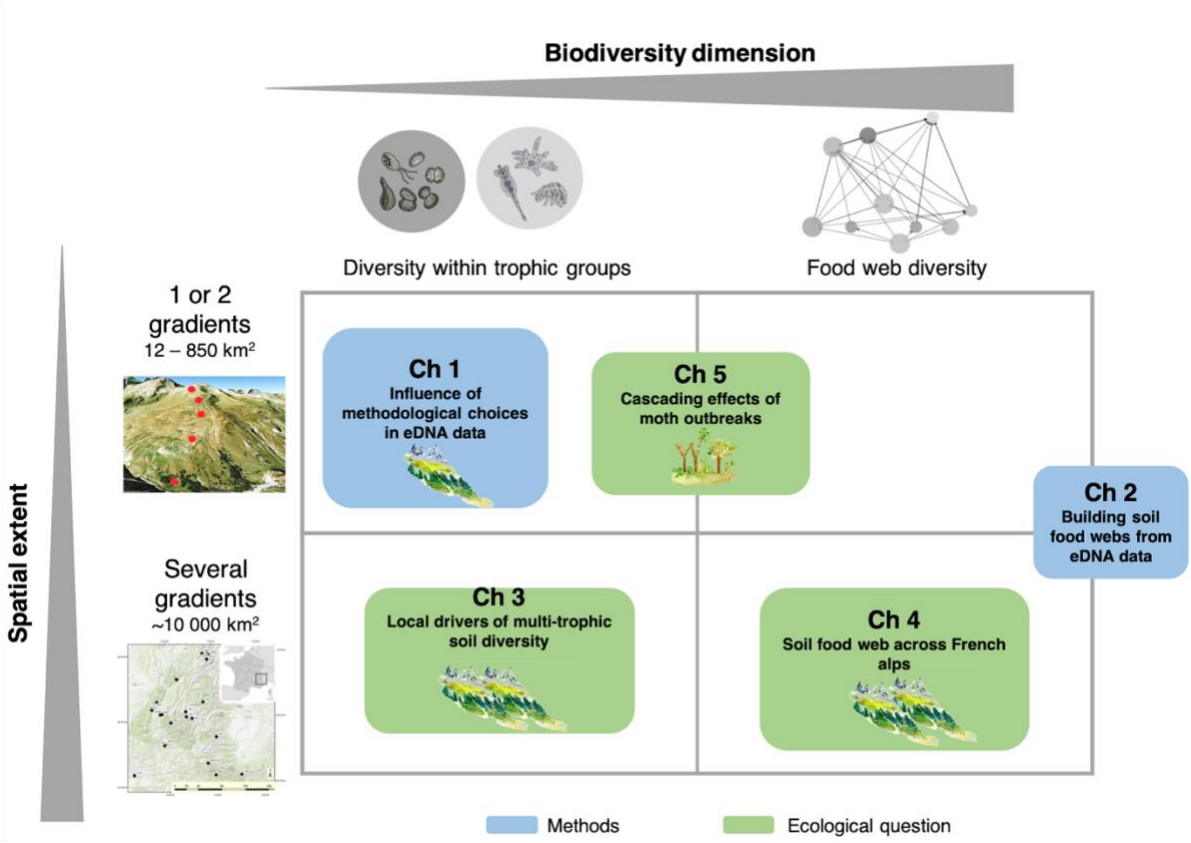


Figure 5. Summary of the research chapters developed in this PhD thesis positioned according to the spatial extent and the biodiversity dimension considered used in the study. The type of gradient (elevational or disturbance gradient) used in the chapter is indicated with a pictogram. Colors indicate if the chapter was related to the methodological (blue) or the ecological (green) objective.

SCIENTIFIC PUBLICATIONS

This thesis led to the production of four scientific papers, from which three are already published, and one is in preparation and should be submitted during the summer:

- Calderón-Sanou I, *et al.* From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *J Biogeogr 47:193–206 (2020).* https://doi.org/10.1111/jbi.13681
- Calderón-Sanou, I., *et al.* Cascading effects of moth outbreaks on subarctic soil food webs. *Sci Rep 11, 15054* (2021). https://doi.org/10.1038/s41598-021-94227-z
- Calderón-Sanou, I., *et al.* Energy and physiological tolerance explain multi-trophic soil diversity in temperate mountains. *Divers Distrib*. https://doi.org/10.1111/ddi.13529
- Calderón-Sanou, I., *et al.* Spatial turnover of soil food webs along environmental gradients. *In preparation.*

During my PhD, I also collaborated with colleagues on other research projects, leading to the production of two scientific papers. For the first paper, my contribution was mostly related to my consistent participation in the botanical surveys conducted:

- Bektaş, B., Thuiller, W., Renaud, J., Gueguen, M., **Calderón-Sanou, I.**, Valay, J-G. Colace, M-P, Münkemüller, T. A spatially explicit trait-based approach uncovers changes in assembly processes under warming. *In revision with Eco Lett.*

For the second paper, I contributed conceptually by providing ideas and suggestions based on my own reflections on the definition of soil trophic groups:

- Hedde, M., Blight, O., Briones, M.J.I., Bonfanti, J., Brauman, A., Brondani, M., **Calderón Sanou, I.**, Clause, J., *et al.* Avoiding cacophony in soil fauna classifications. Under review in *Geoderma*.

Other academic productions in which I was involved during my thesis were the internship reports of four master students that I co-supervised. None of them has led to a scientific publication yet.

REFERENCES OF THE INTRODUCTION

Albert, G., Gauzens, B., Loreau, M., Wang, S. & Brose, U. (2021). The hidden role of multi-trophic interactions in driving diversity–productivity relationships. *Ecology Letters*, 00, 1–11.

Antunes, P.M. & Koyama, A. (2017). Chapter 9 - Mycorrhizas as Nutrient and Energy Pumps of Soil Food Webs: Multitrophic Interactions and Feedbacks. In: *Mycorrhizal Mediation of Soil Fertility, Structure, and Carbon Storage* (eds. Johnson, N.C., Gehring, C. & Jansa, J.). Elsevier, pp. 149–173.

Aslani, F., Geisen, S., Ning, D., Tedersoo, L. & Bahram, M. (2022). Towards revealing the global diversity and community assembly of soil eukaryotes. *Ecology Letters*, 25, 65–76.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., *et al.* (2018). Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.

Bálint, M., Bahram, M., Eren, A.M., Faust, K., Fuhrman, J.A., Lindahl, B., *et al.* (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, 40, 686–700.

Barnes, M.A. & Turner, C.R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17, 1–17.

Bastida, F., Eldridge, D.J., Abades, S., Alfaro, F.D., Gallardo, A., García-Velázquez, L., *et al.* (2020). Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Mol Ecol*, 29, 752–761.

Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., *et al.* (2013). Global biogeography of highly diverse protistan communities in soil. *ISME J*, 7, 652–659.

Bending, G.D. (2003). Litter decomposition, ectomycorrhizal roots and the 'Gadgil' effect. *New Phytologist*, 158, 228–229.

Berg, M.P. & Bengtsson, J. (2007). Temporal and spatial variability in soil food web structure. *Oikos*, 116, 1789–1804.

Bienert, F., Danieli, S. de, Miquel, C., Coissac, E., Poillot, C., Brun, J.J., *et al.* (2012). Tracking earthworm communities from soil DNA. *Molecular Ecology*, 21, 2017.

Botella, C., Dray, S., Matias, C., Miele, V. & Thuiller, W. (2022). An appraisal of graph embeddings for comparing trophic network architectures. *Methods Ecol Evol*, 13, 203–216.

Braga, J., Pollock, L.J., Barros, C., Galiana, N., Montoya, J.M., Gravel, D., *et al.* (2019). Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe. *Glob Ecol Biogeogr*, 28, 1636–1648.

Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., *et al.* (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, 184, 449–456.

Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., *et al.* (2018). Global gaps in soil biodiversity data. *Nat Ecol Evol*, 2, 1042–1043.

Chao, A., Chiu, C.-H. & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45, 297–324.

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

Compson, Z.G., Monk, W.A., Curry, C.J., Gravel, D., Bush, A., Baker, C.J.O., *et al.* (2018). Chapter Two - Linking DNA Metabarcoding and Text Mining to Create Network-Based Biomonitoring Tools: A Case Study on Boreal Wetland Macroinvertebrate Communities. In: *Advances in Ecological Research*, Next Generation Biomonitoring: Part 2 (eds. Bohan, D.A., Dumbrell, A.J., Woodward, G. & Jackson, M.). Academic Press, pp. 33–74.

Connell, J.H. (1978). Diversity in Tropical Rain Forests and Coral Reefs. *Science*, 199, 1302–1310.

Cornwell, W.K., Schwilk, D.W. & Ackerly, D.D. (2006). A trait-based test for habitat filtering: convex hull volume. *Ecology*, 87, 1465–1471.

Coyle, D.R., Nagendra, U.J., Taylor, M.K., Campbell, J.H., Cunard, C.E., Joslin, A.H., *et al.* (2017). Soil fauna responses to natural disturbances, invasive species, and global climate change: Current state of the science and a call to action. *Soil Biology and Biochemistry*, 110, 116–133.

Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., *et al.* (2019). The global soil community and its influence on biogeochemistry. *Science*, 365, eaav0550.

Crowther, T.W., Maynard, D.S., Crowther, T.R., Peccia, J., Smith, J.R. & Bradford, M.A. (2014). Untangling the fungal niche: the trait-based approach. *Frontiers in Microbiology*, 5.

De Bie, T., Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., *et al.* (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters*, 15, 740–747.

Decaëns, T. (2010). Macroecological patterns in soil communities. *Glob Ecol Biogeogr*, 19, 287–302.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., *et al.* (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., *et al.* (2018). A global atlas of the dominant bacteria found in soil. *Science*, 359, 320–325.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, 9, 215–227.

Drummond, A.J., Newcomb, R.D., Buckley, T.R., Xie, D., Dopheide, A., Potter, B.C., *et al.* (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4.

Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thébault, E. & Loreau, M. (2007). The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol Letters*, 10, 522–538.

Dunne, J.A. (2006). The network structure of food webs. In: *Ecological networks: linking structure to dynamics in food webs* (eds. Pascual, M. & Dunne, J.A.). Oxford University Press, pp. 27–86.

Eisenhauer, N., Dobies, T., Cesarz, S., Hobbie, S.E., Meyer, R.J., Worm, K., *et al.* (2013). Plant diversity effects on soil food webs are stronger than those of elevated CO2 and N deposition in a long-term grassland experiment. *Proc Natl Acad Sci U S A*, 110, 6889–6894.

Eisenhauer, N., Schielzeth, H., Barnes, A.D., Barry, K.E., Bonn, A., Brose, U., *et al.* (2019). A multitrophic perspective on biodiversity–ecosystem functioning research. *Adv Ecol Res.*, 61, 1–54.

Elton, C.S. (1927). *Animal ecology*. Macmillan Co., New York,.

Erktan, A., Or, D. & Scheu, S. (2020). The physical structure of soil: Determinant and consequence of trophic interactions. *Soil Biol Biochem*, 148, 107876.

Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J., *et al.* (2011). Trophic Downgrading of Planet Earth. *Science*, 333, 301–306.

Ettema, C.H. & Wardle, D.A. (2002). Spatial soil ecology. *Trends Ecol Evol*, 17, 177–183.

Evans, K.L., Greenwood, J.J.D. & Gaston, K.J. (2005). Dissecting the species–energy relationship. *Proc Biol Sci*, 272, 2155–2163.

Evans, S., Martiny, J.B.H. & Allison, S.D. (2017). Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J*, 11, 176–185.

FAO, ITPS, GSBI, SCBD & EC. (2020). *State of knowledge of soil biodiversity: status, challenges and potentialities, Report 2020.* Rome.

Ferris, H., Griffiths, B.S., Porazinska, D.L., Powers, T.O., Wang, K.-H. & Tenuta, M. (2012). Reflections on Plant and Soil Nematode Ecology: Past, Present and Future. *J Nematol*, 44, 115–126.

Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A. & Cleveland, C.C. (2009). Global patterns in belowground communities. *Ecol Lett*, 12, 1238–1249.

Fiore-Donno, A.M., Richter-Heitmann, T. & Bonkowski, M. (2020). Contrasting Responses of Protistan Plant Parasites and Phagotrophs to Ecosystems, Land Management and Soil Properties. *Front Microbiol*, 11.

Gaston, K.J. (2000). Global patterns in biodiversity. *Nature*, 405, 35012228.

Gaston, K.J. & May, R.M. (1992). Taxonomy of taxonomists. *Nature*, 356, 281–282.

Geisen, S., Briones, M.J.I., Gan, H., Behan-Pelletier, V.M., Friman, V.-P., de Groot, G.A., *et al.* (2019a). A methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry*, 136, 107536.

Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T. & Bonkowski, M. (2016). The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biology and Biochemistry*, 94, 10–18.

Geisen, S., Wall, D.H. & van der Putten, W.H. (2019b). Challenges and Opportunities for Soil Biodiversity in the Anthropocene. *Current Biology*, 29, R1036–R1044.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., *et al.* (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, 10, 1107.

Glassman, S.I., Wang, I.J. & Bruns, T.D. (2017). Environmental filtering by pH and soil nutrients drives community assembly in fungi at fine spatial scales. *Molecular Ecology*, 26, 6960–6973.

Grainger, T.N., Senthilnathan, A., Ke, P.-J., Barbour, M.A., Jones, N.T., DeLong, J.P., *et al.* (2021). An Empiricist's Guide to Using Ecological Theory. *The American Naturalist*, 000–000.

Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters*, 14, 1010–1016.

Grime, J.P. (1973). Competitive Exclusion in Herbaceous Vegetation. *Nature*, 242, 344–347.

Guerra, C.A., Bardgett, R.D., Caon, L., Crowther, T.W., Delgado-Baquerizo, M., Montanarella, L., *et al.* (2021). Tracking, targeting, and conserving soil biodiversity. *Science*, 371, 239–241.

Guerra, C.A., Heintz-Buschart, A., Sikorski, J., Chatzinotas, A., Guerrero-Ramírez, N., Cesarz, S., *et al.* (2020). Blind spots in global soil biodiversity and ecosystem function research. *Nat Commun*, 11, 3870.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.

Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017). *Habitat Suitability and Distribution Models: With Applications in R*. Ecology, Biodiversity and Conservation. Cambridge University Press, Cambridge.

Harvey, B.D., Leduc, A., Gauthier, S. & Bergeron, Y. (2002). Stand-landscape integration in natural disturbance-based management of the southern boreal forest. *Forest Ecology and Management*, 155, 369–385.

Hedlund, K., Griffiths, B., Christensen, S., Scheu, S., Setälä, H., Tscharntke, T., *et al.* (2004). Trophic interactions in changing landscapes: responses of soil food webs. *Basic Appl Ecol*, 5, 495–503.

Hendershot, J.N., Read, Q.D., Henning, J.A., Sanders, N.J. & Classen, A.T. (2017). Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecology*, 98, 1757–1763.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *Am Nat*, 163, 192–211.

Hillebrand, H., Watermann, F., Karez, R. & Berninger, U.-G. (2001). Differences in species richness patterns between unicellular and multicellular organisms. *Oecologia*, 126, 114–124.

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

van den Hoogen, J., Geisen, S., Wall, D.H., Wardle, D.A., Traunspurger, W., de Goede, R.G.M., *et al.* (2020). A global database of soil nematode abundance and functional group composition. *Sci Data*, 7, 103.

Horn, H.S. (1974). The Ecology of Secondary Succession. *Annual Review of Ecology and Systematics*, 5, 25–37.

Hunt, H.W., Coleman, D.C., Ingham, E.R., Ingham, R.E., Elliott, E.T., Moore, J.C., *et al.* (1987). The detrital food web in a shortgrass prairie. *Biology and Fertility of Soils*, 3, 57–68.

Hutchinson, G.E. (1959). Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *Am Nat*, 93, 145–59.

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES secretariat, Bonn, Germany, IPBES secretariat, Bonn, Germany.

Kardol, P., Throop, H.L., Adkins, J. & Graaff, M.-A. de. (2016). A hierarchical framework for studying the role of biodiversity in soil food web processes and ecosystem services. *Soil Biol Biochem*, Special issue: Food web interactions in the root zone: influences on community and ecosystem dynamics, 102, 33–36.

Karimi, B., Terrat, S., Dequiedt, S., Saby, N.P.A., Horrigue, W., Lelièvre, M., *et al.* (2018). Biogeography of soil bacteria and archaea across France. *Sci Adv*, 4, eaat1808.

Keddy, P.A. (1992). Assembly and Response Rules: Two Goals for Predictive Community Ecology. *Journal of Vegetation Science*, 3, 157–164.

Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., *et al.* (2012). More than a meal… integrating non-feeding interactions into food webs. *Ecology Letters*, 15, 291–300.

Kelly, R.P., Shelton, A.O. & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Sci Rep*, 9, 12133.

Leff, J.W., Bardgett, R.D., Wilkinson, A., Jackson, B.G., Pritchard, W.J., De Long, J.R., *et al.* (2018). Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits. *ISME J*, 12, 1794–1805.

Lindeman, R.L. (1942). The Trophic-Dynamic Aspect of Ecology. *Ecology*, 23, 399–417.

Lomolino, Mark.V. (2001). Elevation gradients of species-density: historical and prospective views. *Global Ecology and Biogeography*, 10, 3–13.

Looby, C.I. & Martin, P.H. (2020). Diversity and function of soil microbes on montane gradients: the state of knowledge in a changing world. *FEMS Microbiology Ecology*, 96, fiaa122.

Louthan, A.M., Doak, D.F. & Angert, A.L. (2015). Where and When do Species Interactions Set Range Limits? *Trends Ecol Evol*, 30, 780–792.

Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., Ohlmann, M., Foulquier, A., Clément, J., *et al.* (2019). Multi-trophic β-diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. *Funct Ecol*, 33, 2053–2064.

McCain, C.M. & Grytnes, J.-A. (2010). Elevational Gradients in Species Richness. In: *Encyclopedia of Life Sciences* (ed. John Wiley & Sons, Ltd). John Wiley & Sons, Ltd, Chichester, UK, p. a0022548.

Moore, J.C. & de Ruiter, P.C. (1991). Temporal and spatial heterogeneity of trophic interactions within below-ground food webs. *Agriculture, Ecosystems & Environment*, 34, 371–397.

Morriën, E. (2016). Understanding soil food web dynamics, how close do we get? *Soil Biology and Biochemistry*, 102, 10–13.

Morriën, E., Hannula, S.E., Snoek, L.B., Helmsing, N.R., Zweers, H., de Hollander, M., *et al.* (2017). Soil networks become more connected and take up more carbon as nature restoration progresses. *Nature Communications*, 8, 14349.

Münkemüller, T., Gallien, L., Pollock, L.J., Barros, C., Carboni, M., Chalmandrier, L., *et al.* (2020). Dos and don'ts when inferring assembly rules from diversity patterns. *Global Ecol Biogeogr*, 29, 1212–1229.

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

Nara, K. (2006). Ectomycorrhizal networks and seedling establishment during early primary succession. *New Phytol*, 169, 169–178.

Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *J Biogeogr*, 26, 867–878.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., *et al.* (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol*, 20, 241–248.

Nielsen, U.N., Osler, G.H.R., Campbell, C.D., Neilson, R., Burslem, D.F.R.P. & Wal, R. van der. (2010). The Enigma of Soil Animal Species Diversity Revisited: The Role of Small-Scale Heterogeneity. *PLOS ONE*, 5, e11567.

Nottinghan, A., Fierer, N., Turner, B., Whitaker, J., Ostle, N., McNamara, N., *et al.* (2018). Microbes follow Humboldt: temperature drives plant and soil microbial diversity patterns from the Amazon to the Andes, 0 Bytes.

O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J Biogeogr*, 47, 181–192.

Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L. & Fierer, N. (2020). The global-scale distributions of soil protists and their contributions to belowground systems. *Sci Adv*, 6, eaax8787.

O'Malley, M.A. (2007). The nineteenth century roots of "everything is everywhere." *Nat Rev Microbiol*, 5, 647–651.

Orgiazzi, A., Bardgett, R.D. & Barrios, E. (2016). Global soil biodiversity atlas. *Global soil biodiversity atlas.*

Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2017). Comparing species interaction networks along environmental gradients. *Biol Rev*, 93, 785–800.

Peters, M.K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., *et al.* (2016). Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nat Commun*, 7, 13736.

Phillips, H.R.P., Guerra, C.A., Bartz, M.L.C., Briones, M.J.I., Brown, G., Crowther, T.W., *et al.* (2019). Global distribution of earthworm diversity. *Science*, 366, 480–485.

Phillips, H.R.P., Heintz-Buschart, A. & Eisenhauer, N. (2020). Putting soil invertebrate diversity on the map. *Molecular Ecology*, 29, 655–657.

Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecol Lett*, 15, 1353–1361.

Pollock, L.J., O'Connor, L.M.J., Mokany, K., Rosauer, D.F., Talluto, M.V. & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends in Ecology & Evolution*, 35, 1119–1128.

Pontarp, M., Bunnefeld, L., Cabral, J.S., Etienne, R.S., Fritz, S.A., Gillespie, R., *et al.* (2019). The Latitudinal Diversity Gradient: Novel Understanding through Mechanistic Eco-evolutionary Models. *Trends Ecol Evol*, 34, 211–223.

Potapov, A.M. (2022). Multifunctionality of belowground food webs: resource, size and spatial energy channels. *Biological Reviews*, n/a.

Potapov, A.M., Rozanova, O.L., Semenina, E.E., Leonov, V.D., Belyakova, O.I., Bogatyreva, V.Y., *et al.* (2021). Size compartmentalization of energy channeling in terrestrial belowground food webs. *Ecology*, 102, e03421.

Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S., *et al.* (2015). Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters*, 18, 85–95.

Rahbek, C., Borregaard, M.K., Colwell, R.K., Dalsgaard, B., Holt, B.G., Morueta-Holme, N., *et al.* (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, 365, 1108–1113.

Ramirez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W., *et al.* (2014). Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B.*, 281, 20141988.

Robeson, M.S., King, A.J., Freeman, K.R., Birky, C.W., Martin, A.P. & Schmidt, S.K. (2011). Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 4406–4410.

Roslin, T. & Majaneva, S. (2016). The use of DNA barcodes in food web construction—terrestrial and aquatic ecologists unite! *Genome*, 59, 603–628.

Roy, J., Albert, C., Choler, P., Clément, J.-C., Ibanez, S., Lavergne, S., *et al.* (2013). Microbes on the Cliff: Alpine Cushion Plants Structure Bacterial and Fungal Communities. *Front Microbiol*, 4, 64.

Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., *et al.* (2016). Mapping earthworm communities in Europe. *Applied Soil Ecology*, Soil biodiversity and ecosystem functions across Europe: A transect covering variations in bio-geographical zones, land use and soil properties, 97, 98–111.

Scheffer, M., Carpenter, S.R., Lenton, T.M., Bascompte, J., Brock, W., Dakos, V., *et al.* (2012). Anticipating Critical Transitions. *Science*, 338, 344–348.

Scherber, C., Eisenhauer, N., Weisser, W.W., Schmid, B., Voigt, W., Fischer, M., *et al.* (2010). Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature*, 468, 553–556.

Scheu, S. (2002). The soil food web: structure and perspectives. *European Journal of Soil Biology*, 38, 11–20.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., *et al.* (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109, 6241–6246.

Schuldt, A., Bruelheide, H., Buscot, F., Assmann, T., Erfmeier, A., Klein, A.-M., *et al.* (2017). Belowground top-down and aboveground bottom-up effects structure multitrophic community relationships in a biodiverse forest. *Scientific Reports*, 7, 4222.

Seibold, S., Cadotte, M.W., MacIvor, J.S., Thorn, S. & Müller, J. (2018). The Necessity of Multitrophic Approaches in Community Ecology. *Trends in Ecology & Evolution*, 33, 754–764.

Seidl, R., Fernandes, P.M., Fonseca, T.F., Gillet, F., Jönsson, A.M., Merganičová, K., *et al.* (2011). Modelling natural disturbances in forest ecosystems: a review. *Ecological Modelling*, 222, 903–924.

Sexton, J.P., McIntyre, P.J., Angert, A.L. & Rice, K.J. (2009). Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, 40, 415–436.

Shade, A., Dunn, R.R., Blowes, S.A., Keil, P., Bohannan, B.J.M., Herrmann, M., *et al.* (2018). Macroecology to Unite All Life, Large and Small. *Trends Ecol Evol*, 33, 731–744.

Soininen, J. (2012). Macroecology of unicellular organisms – patterns and processes. *Environmental Microbiology Reports*, 4, 10–22.

Sundqvist, M.K., Sanders, N.J. & Wardle, D.A. (2013). Community and Ecosystem Responses to Elevational Gradients: Processes, Mechanisms, and Insights for Global Change. *Annual Review of Ecology, Evolution, and Systematics*, 44, 261–280.

Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. (2018). *Environmental DNA: for biodiversity research and monitoring*. Oxford University Press, New York.

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N.S., Wijesundera, R., *et al.* (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688–1256688.

Tedersoo, L., Bahram, M., Toots, M., Diédhiou, A.G., Henkel, T.W., Kjøller, R., *et al.* (2012). Towards global patterns in the diversity and community structure of ectomycorrhizal fungi. *Molecular Ecology*, 21, 4160–4170.

Thakur, M.P. (2020). Climate warming and trophic mismatches in terrestrial ecosystems: the green–brown imbalance hypothesis. *Biol Lett*, 16, 20190770.

Thakur, M.P., Phillips, H.R.P., Brose, U., Vries, F.T.D., Lavelle, P., Loreau, M., *et al.* (2020). Towards an integrative understanding of soil biodiversity. *Biol Rev*, 95, 350–364.

Thom, D. & Seidl, R. (2016). Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests: Disturbance impacts on biodiversity and services. *Biol Rev*, 91, 760–781.

Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., *et al.* (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol Evol*, 27, 689–697.

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K., *et al.* (2013). A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters*, 16, 94–105.

Tiedje, J.M., Asuming-Brempong, S., Nüsslein, K., Marsh, T.L. & Flynn, S.J. (1999). Opening the black box of soil microbial diversity. *Applied Soil Ecology*, 13, 109–122.

Tylianakis, J.M. & Morris, R.J. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics*, 48, 25–48.

de Vries, F.T., Thébault, E., Liiri, M., Birkhofer, K., Tsiafouli, M.A., Bjørnlund, L., *et al.* (2013). Soil food web properties explain ecosystem services across European land use systems. *Proceedings of the National Academy of Sciences*, 110, 14296–14301.

Wall, D.H., Nielsen, U.N. & Six, J. (2015). Soil biodiversity and human health. *Nature*, 528, 69–76.

Wang, P., Chen, Y., Sun, Y., Tan, S., Zhang, S., Wang, Z., *et al.* (2019). Distinct Biogeography of Different Fungal Guilds and Their Associations With Plant Species Richness in Forest Ecosystems. *Front Ecol Evol*, 7, 216.

White, H.J., León-Sánchez, L., Burton, V.J., Cameron, E.K., Caruso, T., Cunha, L., *et al.* (2020). Methods and approaches to advance soil macroecology. *Glob Ecol Biogeogr*, 29, 1674–1690.

Wright, D.H. (1983). Species-Energy Theory: An Extension of Species-Area Theory. *Oikos*, 41, 496–506.

Wu, T., Ayres, E., Bardgett, R.D., Wall, D.H. & Garey, J.R. (2011). Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences*, 108, 17720–17725.

Young, I.M. & Crawford, J.W. (2004). Interactions and Self-Organization in the Soil-Microbe Complex. *Science*, 304, 1634–1637.

Zhao, Q., Brink, P.J.V. den, Carpentier, C., Wang, Y.X.G., Rodríguez-Sánchez, P., Xu, C., *et al.* (2019). Horizontal and vertical diversity jointly shape food web stability against small and large perturbations. *Ecology Letters*, 22, 1152–1162.

Zhou, J., Liu, W., Deng, Y., Jiang, Y.-H., Xue, K., He, Z., *et al.* (2013). Stochastic Assembly Leads to Alternative Communities with Distinct Functions in a Bioreactor Microbial Community. *mBio*, 4, e00584-12.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., *et al.* (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Mol Ecol*, 28, 1857–1862.

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., *et al.* (2018). Body size determines soil community assembly in a tropical forest. *Molecular Ecology*.

CHAPTER 1:

From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices?

Received: 23 January 2019 | Revised: 12 July 2019 | Accepted: 15 July 2019

DOI: 10.1111/jbi.13681

RESEARCH PAPER

Journal of Biogeography | WILEY

From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices?

Irene Calderón-Sanou[1] | Tamara Münkemüller[1] | Frédéric Boyer[1] | Lucie Zinger[2] | Wilfried Thuiller[1]

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

[2]Institut de Biologie de l'Ecole Normale Superieure (IBENS), CNRS, Inserm, PSL Research University, Paris, France

Correspondence
Irene Calderón-Sanou, Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France.
Email: irecalsa@gmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-10-LAB-56, ANR-15-IDEX-02 and ANR-16-CE02-0009

Handling Editor: Holger Kreft

Abstract

Aim: Environmental DNA (eDNA) is increasingly used for analysing and modelling all-inclusive biodiversity patterns. However, the reliability of eDNA-based diversity estimates is commonly compromised by arbitrary decisions for curating the data from molecular artefacts. Here, we test the sensitivity of common ecological analyses to these curation steps, and identify the crucial ones to draw sound ecological conclusions.

Location: Valloire, French Alps.

Taxon: Vascular plants and fungi.

Methods: Using soil eDNA metabarcoding data for plants and fungi from 20 plots sampled along a 1000-m elevational gradient, we tested how the conclusions from three types of ecological analyses: (a) the spatial partitioning of diversity, (b) the diversity–environment relationship, and (c) the distance–decay relationship, are robust to data curation steps. Since eDNA metabarcoding data also comprise erroneous sequences with low frequencies, diversity estimates were further calculated using abundance-based Hill numbers, which penalize rare sequences through a scaling parameter, namely the order of diversity q (Richness with q = 0, Shannon diversity with q ~ 1, Simpson diversity with q = 2).

Results: We showed that results from different ecological analyses had varying degrees of sensitivity to data curation strategies and that the use of Shannon and Simpson diversities led to more reliable results. We demonstrated that molecular operational taxonomic unit clustering, removal of polymerase chain reaction errors and of cross-sample contaminations had major impacts on ecological analyses.

Main conclusions: In the Era of Big Data, eDNA metabarcoding is going to be one of the major tools to describe, model and predict biodiversity in space and time. However, ignoring crucial data curation steps will impede the robustness of several ecological conclusions. Here, we propose a roadmap of crucial curation steps for different types of ecological analyses.

KEYWORDS

data curation strategies, distance–decay, environmental DNA, Hill numbers, metabarcoding, sensitivity analysis, spatial partitioning of diversity

Note: Owing to a production error, this article was accidentally omitted from issue 47:1 of *Journal of Biogeography* (a special issue containing papers presented at the meeting 'Macroecology in the age of Big Data') when the issue was published online on 27 January 2020. This article was subsequently added to the issue on 29 January 2020. The publishers apologize for this error and the inconvenience caused.

1 | INTRODUCTION

Understanding the structure and distribution of biodiversity across space and time is a critical goal in ecology. The development of environmental DNA (eDNA) metabarcoding approaches now facilitate the monitoring of species at biogeographical scales and across the whole tree of life (Drummond et al., 2015; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). It is now possible to tackle unresolved questions that could not be addressed with traditional biodiversity surveys so far. For example, eDNA-based biodiversity studies have enabled the spatial partitioning of diversity (i.e. gamma, alpha and beta diversity) of so far elusive taxa in both terrestrial and marine environments (e.g. marine viruses and protists, soil fungi and bacteria), thereby improving our understanding of their community assembly processes and of their role in structuring communities and networks at global scales (e.g. Lima-Mendez et al., 2015; Tedersoo et al., 2014). However, while the eDNA metabarcoding approach promises substantial advances in macroecology and multi-taxa studies, it requires an appropriate and careful processing of the tremendous amount of sequences generated to draw robust and ecologically meaningful conclusions.

Indeed, the analyses of diversity patterns (e.g. alpha- and beta-diversity; Whittaker, 1960) across space and of the processes generating these patterns are traditionally based on community matrices representing the presence/abundance of species across samples. In eDNA metabarcoding surveys, the data consist of hundreds to millions of DNA sequencing reads from the hundreds to thousands of species co-occurring within samples. Using bioinformatics, these data are then transformed in community matrices, but with species replaced by DNA sequences, and species abundance replaced by a number of sequencing reads. While, in an ideal world, one sequence should correspond to a single species, in practice, it can correspond to several species if the DNA region has a low taxonomic resolution, and more critically, one species can be represented by tens to thousands of variant sequences. Amongst those variants, a few are biologically meaningful (e.g. intraspecific variability), but the large majority of them are technical errors produced at the different stages of the laboratory treatments, from DNA extraction to sequencing (see Table 1 and Appendix S1; Bálint et al., 2016; Taberlet, Bonin, Zinger, & Coissac, 2018). These errors can represent more than 70% of the sequences in raw metabarcoding datasets, and have usually low frequencies (e.g. singletons; Brown, Veach, et al., 2015). If interpreted as genuine, these sequences can, therefore, inflate diversity by several orders of magnitude and lead to flawed ecological interpretations (Kunin, Engelbrektson, Ochman, & Hugenholtz, 2010). Molecular protocols are thus applied to reduce and/or control specific technical errors accumulated during the data production. For example, replicated polymerase chain reaction (PCR) amplification and use of negative controls allow identifying artefactual sequences resulting from random errors introduced by DNA polymerases or sequencers, as well as reagent contaminants (de Barba et al., 2014). However, error rates remain high even with the most stringent molecular protocols (Bálint et al., 2016; Taberlet et al., 2018), which has led to the development of bioinformatics algorithms aiming at detecting errors known to occur during data generation (e.g. PCR errors or chimeric sequences). Also, most of these tools require specifying thresholds and parameter values, which are usually based on arbitrary decisions and visual assessments. An example is the classification of sequence variants into MOTUs (Molecular Operational Taxonomic Units) based on the similarity of sequences. While this step is critical because MOTUs are used as a proxy for species in the majority of DNA metabarcoding studies (Appendix S1), MOTUs are commonly defined using a 97% sequence similarity threshold, a value historically defined as the similarity level of full-length 16S rRNA barcodes below which bacterial strains necessarily belong to different species (Stackebrandt & Goebel, 1994). However, the optimal threshold value to define MOTUs depends on the focal taxa and polymorphism/length of the DNA marker used (e.g. Brown, Chain, Crease, MacIsaac, & Cristescu, 2015; Kunin et al., 2010). It also depends on the PCR/sequencing error rate, which varies across molecular protocols, and depends on the amount of target DNA: when it is low, each genuine DNA fragment has a higher probability of being amplified at each PCR cycle (Taberlet et al., 2018).

Hence, using DNA metabarcoding requires making several methodological choices. Beyond those related to molecular protocols and bioinformatics software, one of the most critical choice is to decide which data curation steps to include in the curation procedure. Indeed, each step directly affects the community matrix obtained, by influencing the final list of MOTUs and/or their frequencies within samples. Previous methodological studies have thus underlined the importance of data curation steps on the reliability of ecological analyses and provided guidelines for bioinformatics decision-making (e.g. Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Schloss, 2010). However, most of these studies tested the influence of data curation procedures on a single metric or ecological question. However, questions related to local community richness can be very sensitive to errors (Flynn, Brown, Chain, MacIsaac, & Cristescu, 2015), while comparisons of communities' composition might be less affected (Leray & Knowlton, 2015; Taberlet et al., 2018). In addition, most studies have focused on microbial communities (bacteria or fungi), and few have addressed such questions to macro-organisms. Finally, most published tests have so far relied on mock communities (i.e. positive controls) usually made of DNA extracts for few known species. While mock communities are useful to identify errors and estimate error rates, the conclusions cannot easily be translated to realistic environments with rich and complex communities (Alberdi et al., 2018).

Here, we address how methodological choices related to the DNA metabarcoding data curation strategy influence the results for different types of ecological analyses and their related diversity metrics. We used soil eDNA data from an elevational gradient in the French Alps, and focused on plants and soil fungi to represent both macro- and microorganisms, as well as DNA markers with different length (Table 2). Patterns of plant diversity have been extensively studied in this area (e.g. Chalmandrier, Münkemüller, Lavergne, &

TABLE 1 Brief description of classical technical errors occurring in DNA metabarcoding data, the associated data curation steps tested in the present study and the curation methodology

Target error	Definition	Curation step (abbreviation) and methodology
Mixed	Common obvious molecular/sequencing errors such as mispaired reads, sequences with ambiguous bases, that are too short or singletons.	Common basic filtering: Removal of sequences meeting these criteria. This step is not tested here and has been applied systematically.
PCR error	Base misincorporation by the DNA polymerase during the PCR amplification.	PCR errors removal (PCR error): Identification of PCR errors using a model-based classification of sequences based on their similarities and abundances. The model reflects the accumulation of base misincorporation across PCR cycles, where genuine sequences remain more abundant than their respective errors.
Highly spurious sequences	Chimeras from multiple parents, primers dimers, etc. or sequences from highly degraded DNA fragments that largely differ from any known sequence.	Highly spurious sequences removal (spurious): Removal of sequences of whose similarity with their closest match in public reference databases is below 70% (plants) or 50% (fungi).
Chimeras	Sequences obtained from the recombination of two or more parent sequences	Chimera detection and removal (chimeras): Removal of sequences that have a high probability to be a subsequence from other, more abundant sequences in the dataset.
Remaining PCR errors/ Biological variation	Sequences from the same species either resulting from a PCR error that could not be filtered above, or from intraspecific variability	MOTU clustering (clustering): Clustering of sequences into MOTUs on the basis of their pairwise similarity. Here done at different sequence similarity thresholds.
External contaminants	DNA coming from an external source other than the biological sample	Reagent contaminants cleaning (reagent): Removal of sequences that are more abundant in negative controls relative to biological samples because of the absence of other competing DNA fragments during the amplification process.
Cross-contaminations or tag-jumps	Genuine sequences present in a sample where actually absent, either due to cross-contaminations at the bench, or due to tag-jumps occurring during the library preparation or the sequencing, that is, switches of nucleotidic labels used to assign the sequencing reads to their samples. These contaminants are usually of much lower abundance than their sample of origin.	Cross-sample contamination curation (cross): If the abundance of a given MOTU in a given sample is below 0.03% of the total MOTU abundance in the entire dataset, it is considered as absent in this sample.
Dysfunctional PCRs	PCRs that are too different in comparison with their technical replicates.	Dysfunctional PCR removal (DysPCR): Removal of PCR replicates from a single biological sample that are more dissimilar to each other in MOTUs composition and structure than are the PCR obtained from other biological sample.

Abbreviations: MOTU, molecular operational taxonomic unit; PCR, polymerase chain reaction.
Note: Target errors make reference to the errors described further in Appendix S1. See also Table S2.4 for more details on the curation steps used in this study.

Thuiller, 2015) and serve as a good reference to evaluate the results estimated from eDNA metabarcoding data. We subjected these data to 256 different data curation strategies, which correspond to all possible combinations of seven critical data curation steps. We then tested how the curation strategies influence the inferences drawn from three different ecological analyses: (a) a spatial partitioning of diversity (i.e. gamma, alpha and beta diversities) to estimate the regional and local diversity of the gradient, (b) a diversity–environment relationship, to analyse the influence of environment on the local community diversity (alpha), and (c) a distance–decay analysis, to evaluate if similarities between communities (beta) decrease with increasing geographic distances. To this end, we first checked the accuracy of eDNA metabarcoding data in detecting ecological patterns by comparing the eDNA-based diversity patterns with the expected values based on mock communities and traditional botanical surveys (only available for plants). Second, we did an overall sensitivity analysis to test the sensitivity of ecological results to the data curation strategy. Finally, with a variance partitioning analysis we identified the crucial curation steps (i.e. those that introduced more variance to the results) to include or consider in the curation procedure.

To achieve these objectives, we built on Hill numbers (Hill, 1973) to estimate diversity, which unifies mathematically the best known diversity measures in ecology through a unique parameter q (i.e. Richness at q = 0, the exponential of Shannon entropy at q ~ 1 and the inverse of Simpson at q = 2). In this framework, the weight of the rare species decreases when increasing the value of the parameter

TABLE 2 Characteristics of the DNA markers used to estimate eDNA-based diversity in this study

DNA Marker	Target taxa	Forward primer (5′–3′)	Reverse primer (5′–3′)	Length [range] (bp)	References
P6 loop of the chloroplast *trn*L intron	Vascular plants	g:GGGCAATCCTGAGCCAA	h: CCATTGAGTCTCTG CACCTATC	48 [10–220]	Taberlet et al., 2007
Nuclear ribosomal DNA Internal Transcribed Spacer 1 (ITS1)	Fungi	ITS5: GGAAGTAAAAGTCG TAACAAGG	Fung02:CCAAGAGATC CGTTGYTGAAAGTK	226 [68–919]	White, Bruns, Lee, & Taylor, 1990; Taberlet et al., 2018

q. This feature is particularly relevant for DNA metabarcoding data, since artefactual sequences are usually rare compared to the genuine ones (Bálint et al., 2016; Taberlet et al., 2018). Hill numbers can thus penalize these rare sequences at different degrees: q = 1 is the order of diversity that levels the MOTUs exactly according to their relative abundances, while q < 1 overweigh rare MOTUs and q > 1 overweight abundant MOTUs. As a result, we could expect that diversity measures that give less importance to rare sequences (i.e. q > 0) are less sensitive to the data curation strategy, because they penalize the artefactual sequences targeted by the curation steps.

2 | MATERIALS AND METHODS

2.1 | Sample data

Soil cores were sampled at 10 different elevations equally distributed across an elevational gradient in the northern French Alps (from 1,748 m to 2,725 m a.s.l.) in 2012. At each elevation, two 10 m × 10 m plots were selected (20 plots in total). In each plot, 21 soil cores distributed along the two diagonals were sampled. Soil corers were cleaned and sterilized between each sample collection. Extracellular DNA was then extracted twice, from 15 g as described in Taberlet, Prud'homme, et al. (2012). Aboveground plant community information (hereafter observed plant diversity) was obtained in each plot with a botanical survey conducted during the annual productivity peak (mid-July) using the Braun-Blanquet cover-abundance scale (Braun-Blanquet, 1946).

2.2 | Molecular analyses

eDNA-based plant diversity was estimated by targeting a vascular plant-specific marker (P6 loop of chloroplast trnL, Table 2). It targets highly conserved priming sites across vascular plants and amplifies a short region, which is desired when working with degraded DNA. eDNA-based fungal diversity was assessed using the nuclear ribosomal Internal Transcribed Spacer 1 (ITS1; Table 2). For each DNA extract, PCRs were run in duplicate leading to four technical replicates per core sample and DNA marker. PCR thermocycling conditions and mixture composition and purification can be found in Table S2.1 in Appendix S2. To control for potential contaminants, extraction and PCR blank controls were included in the experiment. To control for false positives caused by tag-switching events, we also defined "sequencing blank controls", that is, tag combinations not used in our experimental design, but that could be formed at the library preparation or sequencing stage (See Appendix S1). We also included positive controls in this experiment, which consisted of a mix of DNA extracted from 16 plant species. For this, genomic DNA was extracted from leaf tissue using the DNeasy Plant Kit (Qiagen GmbH), quantified, diluted at different concentrations for each species and mixed to form a mock community (species composition provided in Table S2.2, Appendix S2). Positive controls allow for quantification of technical biases introduced by PCR and sequencing. Illumina sequencing was performed on a HiSeq platform (2 × 100 bp paired-end reads) for plant amplicons and on a MiSeq (2 × 250 bp paired-end reads) for fungi amplicons, both using the paired-end technology.

2.3 | Bioinformatics analyses

The Illumina sequencing paired-end reads (Table S2.3) were preprocessed for each marker with three procedures: (a) assembling forward and reverse paired-end reads based on their overlapping 3'-end sequences, (b) assigning each read to its respective sample (demultiplexing) and (c) combining strictly identical sequences into unique DNA sequences while keeping information on their abundance (number of sequencing reads) in each sample (dereplication). Then we systematically processed the dereplicated sequences following common data curation procedures that included removal of sequences with low paired-end alignment scores, removal of singletons, removal of short sequences and removal of sequences containing ambiguous bases (not to be confounded with a phred-quality filtering; Figure 1a; Table 1; Table S2.4). Singletons are sequences that occur only once in the whole dataset and many studies agree that their removal is necessary to reduce data complexity/computational time and because they mostly correspond to molecular artefacts that may inflate disproportionately diversity indices (Brown, Veach, et al., 2015; Kunin et al., 2010). In our data, they represented 70%–80% of the total number of sequences but only 1%–15% of the total number of sequencing reads for plants and fungi respectively (Table S2.3 in Appendix S2). We finally assigned each remaining sequence to a taxonomic clade with the *ecotag* command from the OBITools software package (Boyer et al., 2016) that uses a lowest common ancestor algorithm for the assignment, and the EMBL database version 133 as a reference.

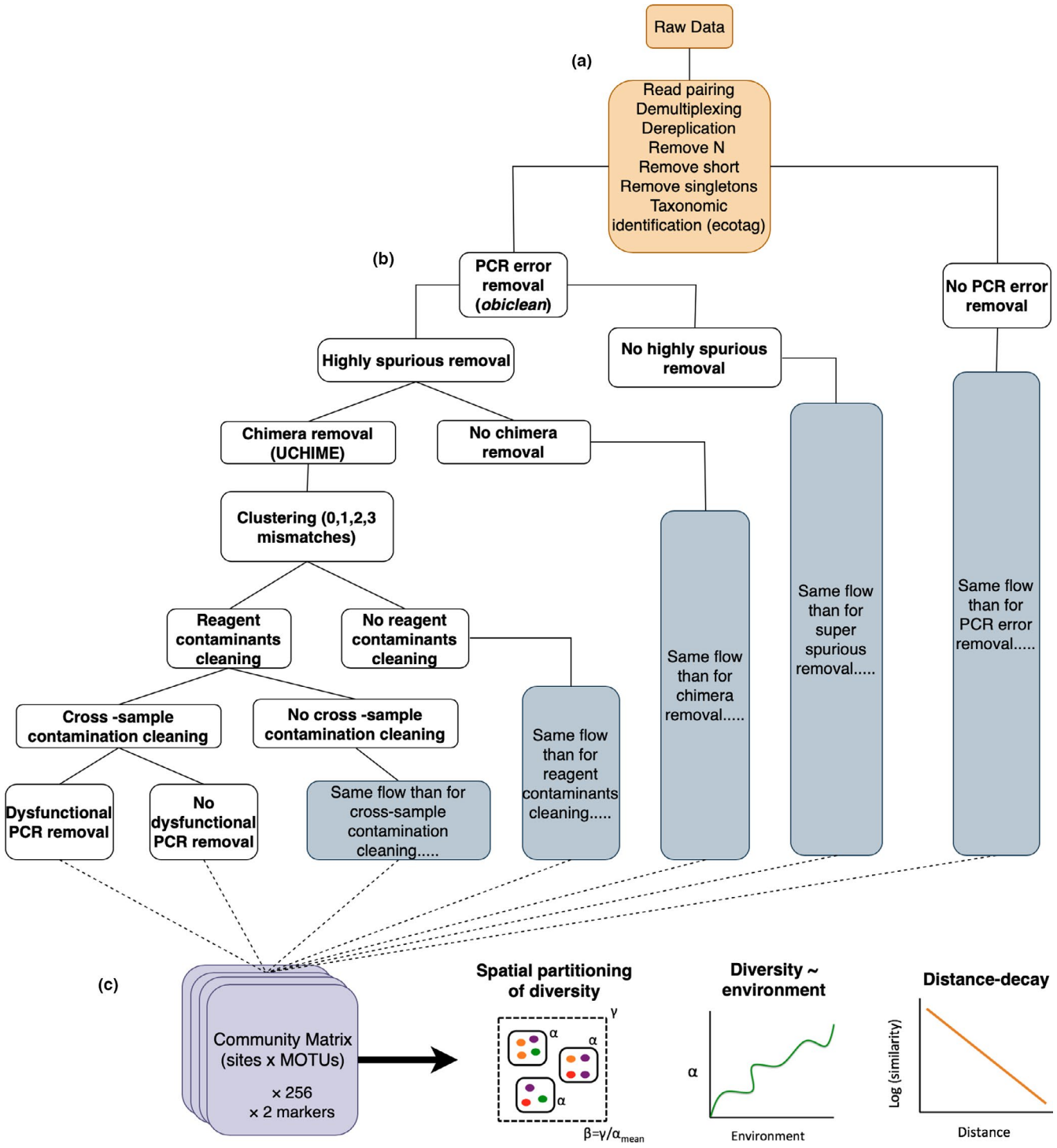


FIGURE 1 Workflow of the sensitivity analysis. (a) Raw data are curated with basic filtering steps for each DNA marker (plants: trnL-P6 loop, fungi: internal transcribed spacer 1). (b) Filtered data are processed using seven curation steps that were varied or removed in each data curation strategy making a total of 256 possible combinations. As a result, 256 community matrices are obtained per DNA marker and used to (c) conduct three types of ecological analyses. The range of values obtained for each ecological analysis and diversity metric represents the variance due to the data curation strategy [Colour figure can be viewed at wileyonlinelibrary.com]

Next, data from each marker were processed following a range of different data curation strategies to test the sensitivity of ecological analyses to different methodological choices (Figure 1b). To do so, we selected seven important steps: (a) removal of PCR errors, (b) filtering of highly spurious sequences, (c) removal of chimeras, (d) sequence classification into MOTUs (MOTU clustering), (e) removal of reagent contaminants, (f) cross-sample contamination cleaning and (g) dysfunctional PCRs filtering (see Table 1; Appendix S1; Table S2.4 in Appendix S2 for target errors and step descriptions). Curation steps were either kept or excluded, and

were always performed in the same order in each data curation strategy. For the MOTU clustering step, when kept, three clustering thresholds were tested (1, 2 or 3 mismatches allowed between pairwise aligned sequences). We used here raw mismatches rather than percentages of dissimilarities because the DNA markers used are short (< 100 bp) and/or highly polymorphic in length. Using the percentages of dissimilarity in this case would penalize more little differences when alignments are short than when they are long.

All different possible combinations of these curation strategies were implemented (Figure 1b). Most of the curation steps were done using the software OBITOOLS (Boyer et al., 2016). Chimera detection was performed with UCHIME (Edgar, Haas, Clemente, Quince, & Knight, 2011) and we used SUMACLUST (Mercier, Boyer, Bonin, & Coissac, 2013) for MOTU clustering due to its ability in handling large datasets and its flexibility for defining the clustering threshold (see Table S2.4 for more details on the algorithm). After data curation, PCR replicates were summed and standardized by the total number of reads in each core sample. We then pooled the samples for each of the 20 plots to obtain a single community per plot. For this, MOTUs abundance (already standardized by the number of reads) were summed and standardized by the number of samples in each plot. For each of the data curation strategies, we obtained a community matrix with rows representing plots and columns representing all the MOTUs obtained after curation, which we used here as a proxy for species. Therefore, our sensitivity analysis was conducted on a total of 256 matrices for each DNA marker (Figure 1c).

2.4 | Ecological questions

We tested the sensitivity of the results for three common ecological analyses to the above-mentioned data curation strategies using MOTUs as equivalent of species:

2.4.1 | Spatial partitioning of diversity

We used the multiplicative diversity partitioning approach (Whittaker, 1960) to analyse gamma (here the diversity across the entire gradient), alpha (diversity of local communities) and beta diversity (diversity between communities). In the Hill numbers framework, gamma diversity is the effective number of species in the pooled meta-community (i.e. across all plots), alpha diversity is the effective number of species per community (i.e. plot) and beta diversity is the effective number of communities, calculated as the ratio of gamma diversity to alpha diversity. We followed Chao, Chiu, and Jost, (2014)'s definition where beta diversity is independent of alpha and ranges from 1 (all communities are identical) to the total number of communities *N* (when *N* = 20 all communities are different). We limited our study to taxonomic diversity, because the DNA markers we used here are rather short (Table 2) and are highly variable in length, which make them not suitable for inferring accurate phylogenetic relationships at the scale of the community.

2.4.2 | Diversity–environment relationship (alpha ~ soil organic matter content)

Diversity is often linked to abiotic drivers, and a common ecological research question is how alpha diversity changes along an environmental gradient. Here, we fitted a linear model to determine changes in alpha diversity along a gradient of soil organic matter content (SOM content), known to be a strong predictor of diversity changes in the study site (Ohlmann et al., 2018).

2.4.3 | Distance–decay relationship (similarity ~ geographic distance)

Species' distributions and resulting diversity patterns are controlled by both species dispersal abilities and spatial turnover of environmental conditions (Tuomisto, 2003). One hypothesis is thus that spatially distant communities are more different than close communities ("distance-decay", Green et al., 2004; Tuomisto, 2003). We used the Jaccard-type overlap (U_{qN}) as a measure of similarity (Chao et al., 2014) and we fitted a linear model using the log transformation of similarity against the geographic distance to evaluate the distance–decay. The geographic distance between plots was calculated with Euclidean distances using the elevation values of the plots.

For each DNA marker (plant and fungi), we calculated the gamma, alpha and beta diversities (spatial partitioning of diversity) for each of the 256 community matrices obtained from the different metabarcoding data curation strategies using Hill numbers with values of q = {0,0.5,1,2}. For the diversity–environment and the distance–decay relationships, we fitted our models to each community matrix and extracted the slopes and the R-squares of the models. Alpha diversity and community similarity were calculated using Hill numbers with values of q = {0,1,2}.

2.5 | Sensitivity analyses

2.5.1 | Detectability of ecological patterns

To test the ability of eDNA metabarcoding data and of the different data curation strategies to detect ecological patterns we (a) evaluated the completeness of the sampling unit (plot), and (b) used the observed plant diversity and positive controls as references to evaluate the accuracy of the ecological results. We acknowledge that eDNA-based diversity is expected to slightly diverge from observed diversity (see discussion) but they should follow similar trends (Hiiesalu et al., 2012; Träger, Öpik, Vasar, & Wilson, 2019; Yoccoz et al., 2012). The sampling completeness of each plot was evaluated with rarefaction curves for the different orders of diversity q = {0,1,2} and for three data curation strategies with varying filtering stringency: a "no data curation" strategy with no curation step at all; a "basic curation" strategy including only the chimera removal and a traditional clustering threshold allowing three mismatches between clustered sequences and, a "rigorous curation" strategy, including all the curation steps considered here and a clustering threshold allowing two mismatches.

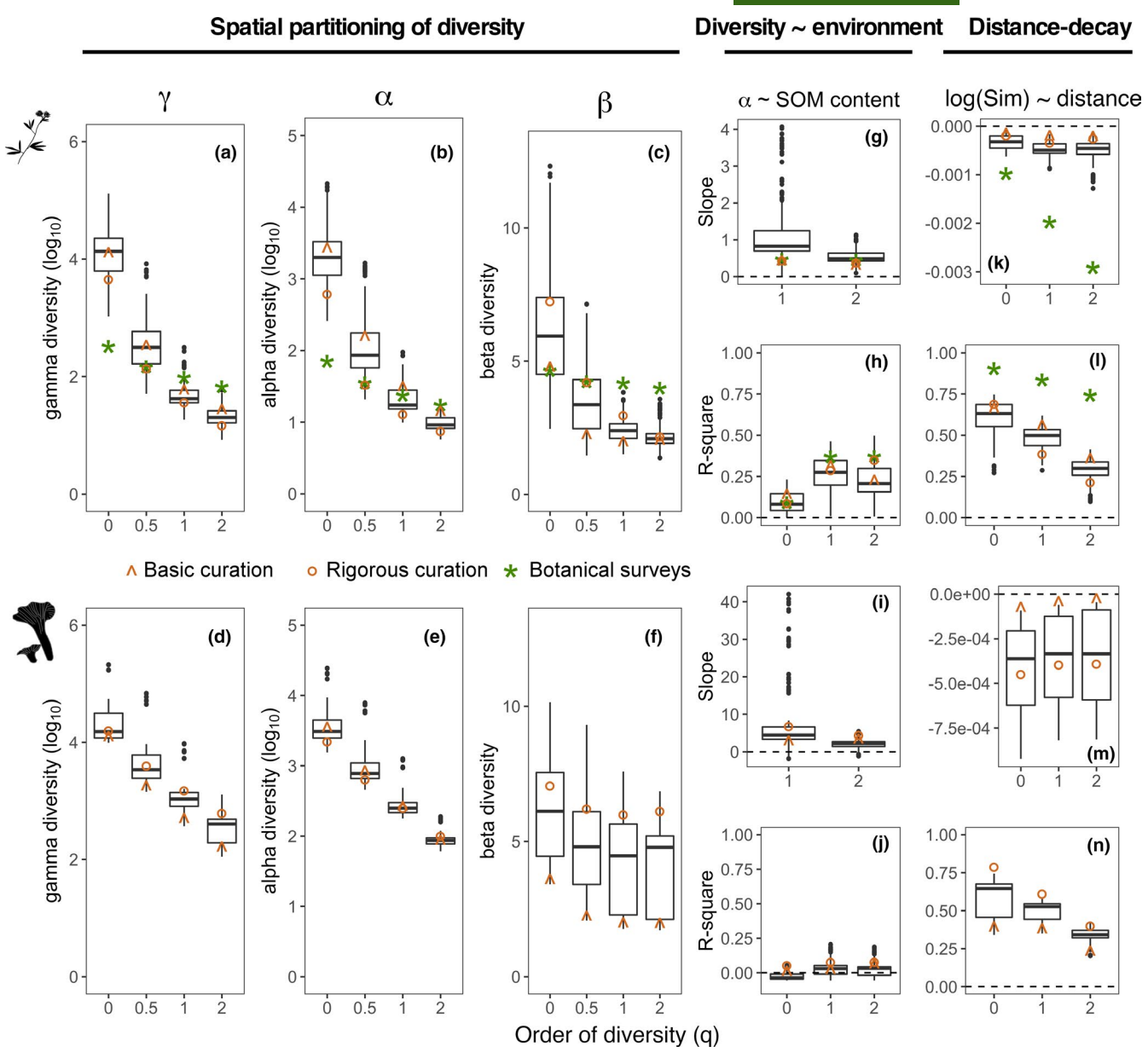


FIGURE 2 Estimated values of the spatial partitioning of diversity components (a-f), of the regression parameters from the diversity–environment (g-j), and of distance–decay (k-n) relationships across the 256 curation strategies for different diversity metrics (Hill numbers, q = {0,0.5,1,2}). The top row (a-c, g, h, k, and l) corresponds to the plant DNA marker (trnL-P6 loop) and bottom row (d-f, i, j, m, and n) to the fungi DNA marker (internal transcribed spacer 1). Size of each box (including whiskers) represents the sensitivity of the diversity metrics or the model parameters to the data curation strategy. The circle and the triangle symbols indicate the values obtained from a rigorous and a basic curation strategy respectively. The star symbol indicates the values calculated from botanical survey (only represented for plants, top row) [Colour figure can be viewed at wileyonlinelibrary.com]

2.5.2 | Overall sensitivity analyses

To test the sensitivity of the results for the different ecological analyses and their related diversity metrics to the data curation strategy, we used the variance of each diversity estimate, obtained across the 256 community matrices and for each marker (Figure 1c). For the diversity–environment and the distance–decay relationships, we looked at the variance in the slope and the R-square of the linear regression across the 256 models for each marker. In addition, we used "the rigorous" and "the basic" curation strategies explained above, that correspond to commonly used pipelines, to exemplify how results can differ between studies.

2.5.3 | Identifying the crucial steps of the curation procedure

To identify the crucial steps we did a variance partitioning analysis for each diversity metric. For the spatial partitioning of diversity, the diversity metrics (gamma, alpha and beta diversities) were used as the response variable in function of the curation

steps. For the diversity–environment and the distance–decay relationships we used the slope and the R-square of the models as the response variable in function of the curation steps. Variance partitioning analyses were done with the R package RELAIMPO (Grömping, 2006).

3 | RESULTS

3.1 | Detectability of ecological patterns with eDNA metabarcoding data

3.1.1 | Sampling completeness of the plots

For both markers/taxa, the total diversity was well represented by the number of reads sequenced, when considering the diversity at q = {1,2} (Figure S2.1 and S2.2 in Appendix S2). At q = {0}, the rarefaction curve rarely saturated, but we obtained more asymptotic curves when increasing the stringency of the data curation strategy.

3.1.2 | Spatial partitioning of diversity

Overall, we found that alpha diversity estimates at q = {1,2} were closer to the observed plant diversity (Figure 2b) and to the positive controls composition (Figure 3) than at q = {0,0.5}. However, diversity at q = {1} slightly underestimated gamma (Figure 2a) and beta (Figure 2c) while all diversity components were underestimated for most curation strategies at q = {2} (Figure 2a-c). Richness (q = 0) was always overestimated. While we obtained very accurate results for diversity at q = {0.5} when using a rigorous pipeline, a basic pipeline led to a substantial overestimation.

3.1.3 | Diversity–environment relationship

While the expected positive slope was in most cases detected (Figure 2g) and its value was on average very similar to the one obtained for observed plant diversity, especially when using a rigorous pipeline, it was highly overestimated for some data curation strategies at q = {0,1}.

3.1.4 | Distance–decay relationship

The expected negative slope of the distance–decay curve was always detected (Figure 2k). However, independently of the data curation strategy, the slope was always underestimated compared to the curve calculated with observed plant diversity. Also, the R-square of the distance–decay relationship was reduced at q = {2} (Figure 2l).

3.2 | Overall sensitivity of ecological questions and diversity metrics

The results of different ecological questions had varying degrees of sensitivity to the data curation strategies. While the estimates in all ecological questions were highly sensitive (width of the boxplots in Figure 2), the main signal of the diversity–environment and the distance–decay relationships was consistent across most curation strategies.

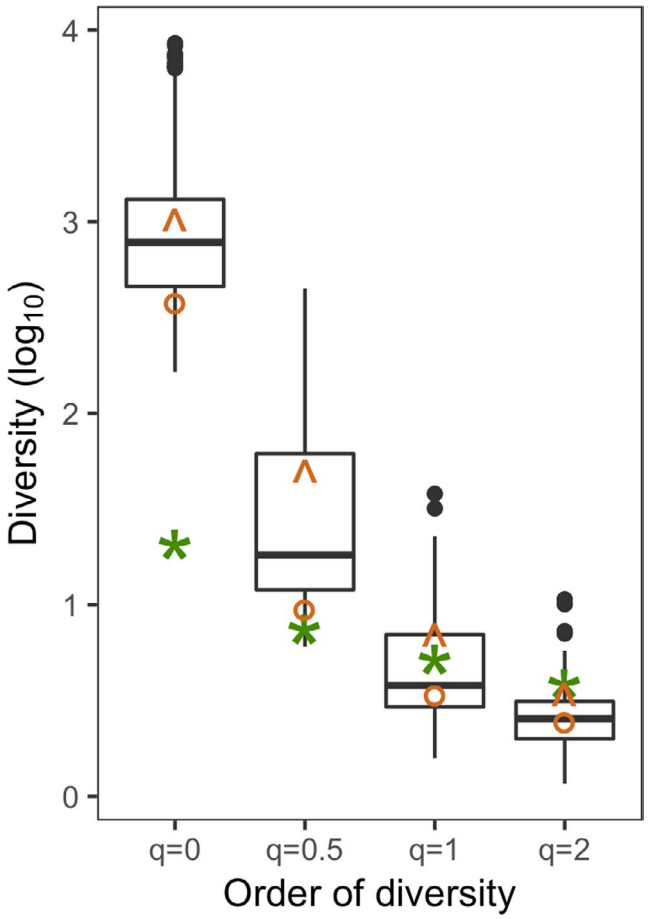


FIGURE 3 Mean diversity estimated in positive controls across the 256 data curation strategies for different diversity metrics (Hill numbers, q = {0,0.5,1,2}). Size of each box (including whiskers) represents the sensitivity of the diversity metrics to the data curation strategy. The star symbol indicates the values calculated from the known species composition in positive controls, the other symbols are as in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

3.2.1 | Spatial partitioning of diversity

Sensitivity of gamma, alpha and beta diversity decreased for higher values of q, that is, weighing down rare MOTUs (Figure 2a-f). Diversity estimates at q = {0} were the most sensitive, with more than two orders of magnitude for both gamma and alpha (Figure 2a,b) diversities of plants. Likewise, the rigorous and basic curation strategies (circles and triangles in Figure 2) exhibited a steep difference at q = {0}, which decreased when using higher values of q in the majority of cases.

3.2.2 | Diversity–environment relationship

The interpretation of the alpha-SOM content relationship could change depending on the data curation strategy used. However,

the alpha-SOM content relationship was more robust when using q = {1,2}, that is, a positive relation between alpha diversity and SOM content was detected independently of the data curation strategy used (Figure 2g,h). Patterns in fungi diversity were more robust, that is, no relation between fungi diversity and SOM content was detected across the different pipelines. A very weak positive relation between fungi diversity and SOM content was observed for q = {1,2}. The rigorous and the basic strategies led to very similar results for both DNA markers/taxa.

3.2.3 | Distance–decay relationship

In contrast, a significant distance–decay relationship was always detected from eDNA metabarcoding data independently of the data curation strategy, but the rate at which similarity decays with increasing distance between plots (i.e. slope) slightly changed across strategies. While very similar results were found between the rigorous and the basic strategies for the distance–decay curve of plants, the slope of the distance–decay curve for fungi was very low when using a basic instead of a rigorous strategy.

3.3 | Crucial steps of the curation procedure

Overall, we found that two curation steps, the removal of PCR error and the clustering to define MOTUs, explained most of the variation in diversity estimates across data curation strategies (more than 15% each and usually more than 40% in total) for most of the diversity metrics in the ecological analyses and for both markers/taxa (Figure 4 and Figure S2.3 in Appendix S2). Also, cross-sample contamination removal explained large parts of the variance of beta diversity in the spatial partitioning of diversity analyses (Figure 4a,b) and of R-squares and slopes in the diversity–environment (Figure 4c,d) and distance–decay (Figure 4e,f) relationships analyses.

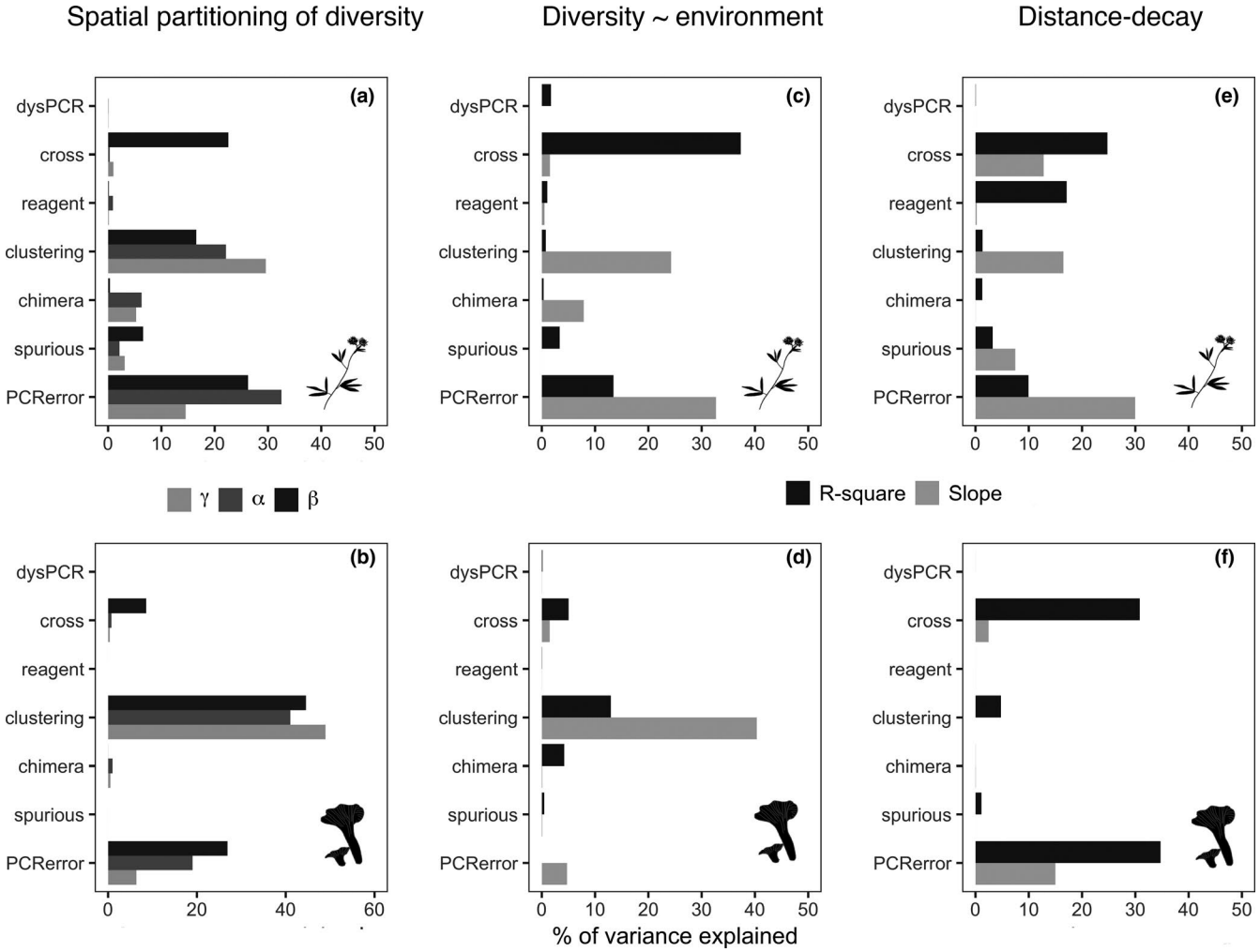


FIGURE 4 Relative importance (% of variance explained) of the data curation steps on the variability of estimated values of the spatial partitioning of diversity components (a, b) and of the parameters from the diversity–environment (c, d) and distance–decay (e, f) relationships, using Hill numbers at q = {1} (see Figure S2.3 for the other q values). The top row (a, c, and e) corresponds to the plant DNA marker (trnL-P6 loop) and bottom row (b, d, and f) to the fungi DNA marker (internal transcribed spacer 1). A model was fitted independently for each diversity component (a, b) or model parameter (c-f) as response variable, with curation steps as main effects

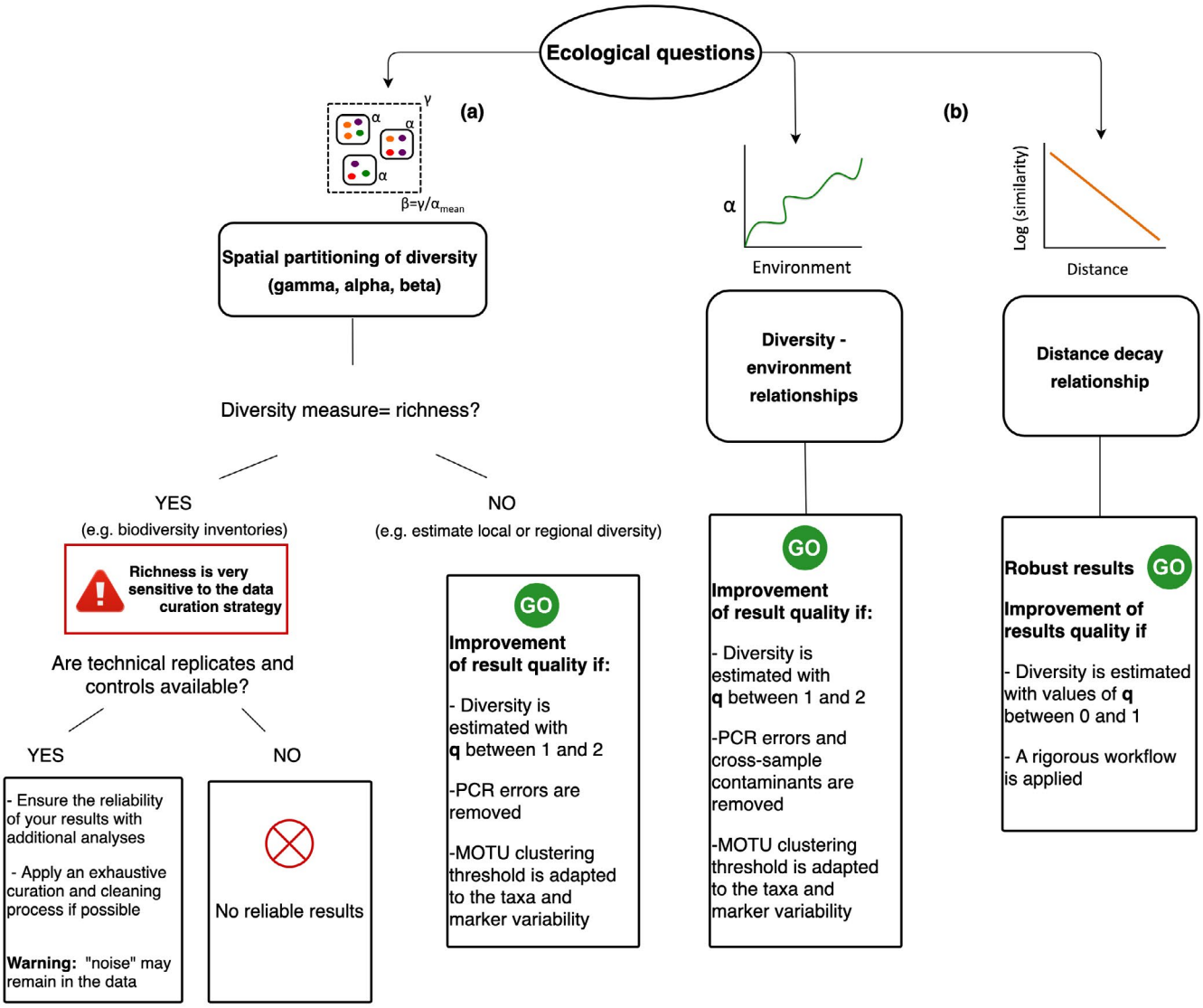


FIGURE 5 Guidelines to improve the reliability of ecological results when analysing environmental DNA metabarcoding data [Colour figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

Ecologists do now increasingly rely on DNA metabarcoding to measure biodiversity as this approach holds the promise of allowing testing long-standing hypotheses at spatial, temporal and taxonomic scales that were hitherto inaccessible with traditional approaches. However, the technique is still hampered by a substantial amount of technical errors (Table 1; Appendix S1; Bálint et al., 2016; Taberlet et al., 2018). Here, we sought at testing the sensitivity of the conclusions drawn from different ecological analyses and diversity metrics to the steps commonly used to curate DNA metabarcoding data from such errors. We show that ecological conclusions had varying degrees of sensitivity to the data curation strategies and that the use of metrics that are less sensitive to rare species/MOTUs (i.e. Shannon and Simpson diversity) leads to more robust diversity estimates. Also, we demonstrated that MOTU clustering, removal of PCR errors and removal of cross-sample contaminations have a major influence on ecological results, and must always be carefully included when curating DNA metabarcoding data.

The breadth of our study makes our findings generalizable to other systems. Indeed, we found similar trends in the sensitivity of gamma and alpha diversity estimates for both our observed plant diversity and the mock community (Figure 2 vs Figure 3). Second, our study focuses on both plants and fungi, that widely differ in their ecological properties and the length of their markers (on average 50 bp for plants vs 225 bp for fungi). Still, while they do not share the same diversity patterns, their sensitivity to data curation strategies were comparable. Furthermore, we expect that our study and the experimental testing design we developed will stimulate further methodological studies (e.g. for tropical or aquatic systems and other markers/taxa) and that they will serve as a guide to prioritize some curation steps when deciding for a curation strategy.

4.1 | Linking methodological choices with ecological questions

The ecological question(s) underlying a study should lead the prioritization of the curation steps to be included in the data curation procedure, as well as the selection of appropriate diversity metrics (Figure 5). If the aim of the study is to estimate the spatial partitioning of diversity (Figure 5a), it is important to keep in mind that all diversity components are biased by the data curation steps. Richness is highly sensitive to error accumulation, and was hence the metric responding the strongest to the data curation strategy. Consequently, if measuring richness is crucial for the study, and, thus, rare species are important, the reliability of the results must be confirmed with additional analyses. For example, a more conservative strategy (i.e. keeping only MOTUs present in more than a certain number of PCR replicates) can improve the reliability of final results, but with the risk of missing species represented by few sequences in only a few samples due to the sampling process occurring when preparing aliquots of one DNA extract (Alberdi et al., 2018). Verifying the pertinence of species detected by looking in detail into the taxonomic assignments can also improve the reliability of results, even though this could be problematic for poorly known taxa with incomplete reference databases (Cristescu, 2014). Also, positive controls (with mock communities) and numerous negative controls (extraction, PCR) must be included in all the phases of sequence generations to ensure the accuracy of richness estimates (Bálint et al., 2016). In any cases, a certain degree of uncertainty will always remain because of the complexity of deciding objectively which sequences are genuine and which are artefactual.

We corroborated that richness is a very sensitive metric and is always overestimated (Figure 2a-c). The intrinsic properties of eDNA can inflate the diversity compared to traditional surveys because eDNA can persist in the environment or be transported through space depending on the abiotic conditions (e.g. water transport, temperature, UV, or microbial activity; Barnes & Turner, 2016). This means that the diversity eDNA estimates not only encompass local and current species, but also species that are dormant (Hiiesalu et al., 2012), that were present in the recent past (Yoccoz et al., 2012) or that are present in the vicinity of the studied area (Taberlet et al., 2018). In other words, the spatio-temporal window captured by local eDNA diversity estimates may be larger than that captured by traditional approaches, a property that can be desirable or not depending on the question addressed. Distinguishing this feature from methodological bias remains at this stage difficult, as it may look like cross-contamination, and also because the cycle of eDNA in the environment remains poorly understood (Barnes & Turner, 2016). However, it is crucial to account for eDNA properties when interpreting richness-based studies to avoid meaningless conclusions.

When the detection of rare species is not of importance, Hill numbers are a promising solution to increase the robustness of results and to avoid the inflation of diversity estimates. The Hill numbers approach has been already proposed to better estimate microbial diversity (e.g. Bálint et al., 2016; Chiu & Chao, 2016), and we corroborate its efficiency for estimating plant diversity and potentially other macro-organisms from metabarcoding data. Both, Shannon and Simpson diversity measures led to a satisfying representativeness of the sampling unit diversity and were robust to the different data curation strategies tested here, but Shannon diversity was less biased. In the same way that richness overestimated diversity, Simpson diversity tended to underestimate diversity. Diversity measures, other than richness (i.e. $q > 0$), account for species/MOTUs abundance structure. The factors determining species' abundances in a community are not the only factors determining the MOTUs' abundances. These correspond to a pool of DNA fragments from current, dormant, or past populations (e.g. microbes) down to one (or part of one) single multicellular individual that are besides amplified by PCR. Consequently, a highly abundant MOTU does not necessarily imply that more individuals of the corresponding taxon were present, it could also be due to for example, higher body mass, larger root systems, or slower DNA decomposition. Besides, given the exponential nature of the PCR amplification, abundant taxa become even more abundant in this step and this could lead to an underestimation of Simpson diversity. Hence, interpreting MOTUs frequency directly as species abundance can be highly misleading, and estimating species abundance in terms of number of individuals or biomass from eDNA is still a major challenge in the field (Deiner et al., 2017). However, MOTUs frequency correlates to a certain extent to species relative abundance, and more importantly, errors are usually rarer than genuine sequences (reviewed in Taberlet et al., 2018). Accordingly, Shannon diversity from eDNA samples appears here as a balanced diversity measure, robust to the data curation strategy, and hence, to rare errors. This can be generalized to all ecological analyses tested in this study. Given these results, we argue that using a complete diversity profile (for example, with q values between 0 and 2) may allow improving confidence in diversity estimates from eDNA data while getting information about MOTUs structure of abundances.

Another important outcome of our assessment is that despite the above-mentioned limits, robust conclusions can be obtained from eDNA metabarcoding data if the aim is to link local diversity (alpha) or community similarity (beta) to environmental or geographic gradients (Figure 5b). Changes in local diversity across an environmental gradient were more sensitive to the data curation strategies than the distance–decay relationship. Our results thus corroborate other studies that demonstrated the robustness of beta diversity to bioinformatics analyses (Botnen, Davey, Halvorsen, & Kauserud, 2018; Deiner et al., 2017). However, the slope of the distance–decay was always underestimated compared to that obtained from observed plant diversity. On one hand, this could result from a lack of phylogenetic resolution of the genetic marker used here, which is relatively short. In alpine ecosystems, it is common to see abundant species

replaced by closely related species across an elevational gradient (Chalmandrier et al., 2015). A genetic marker with a low phylogenetic resolution would not detect these changes and as a consequence, gamma and beta diversities would be underestimated. However, the underestimation of gamma diversity relative to alpha diversity is not strong enough, suggesting that other reasons may also explain the lower slope of the distance–decay curve for eDNA-based plant diversity. Botanical surveys used in this study represent just a local snapshot of the visible plant diversity at the sampling time, and, unlike the eDNA approach, may miss species with an offset phenology or present only in the vicinity of the sampling area (Hiiesalu et al., 2012). We can expect that the larger spatio-temporal window captured by the eDNA metabarcoding approach would thus result in higher similarity among the sites, which could be tested by increasing the botanical sampling effort across seasons and years to reduce botanical surveys biases related to the differentiated phenology of the species.

4.2 | Crucial steps for designing a careful curation protocol

While we included here curation steps that are common to most bioinformatic tools (e.g. QIIME, USEARCH), we acknowledge that algorithms within OBITOOLS have their own particularities, as each of the other packages, and that the results obtained here may not be directly transferable. However, we expect that the differences from a specific software are minor compared to the differences caused by the choice of specific curation steps (Bonder, Abeln, Zaura, & Brandt, 2012). In general, we corroborate past studies concluding that the clustering threshold used for defining MOTUs leads to significant changes in diversity estimates and that this is especially important for alpha and gamma diversities, but less so for beta diversity (Botnen et al., 2018; Brown, Veach, et al., 2015; Kunin et al., 2010). Additionally, we found that PCR errors and cross-sample contaminations are critical steps and that including them leads to more realistic spatial diversity patterns and estimates of diversity components. These two steps correct the diversity at local levels (i.e. sample level) and are especially important when comparing communities. To our knowledge, this is the first study testing in a systematic way the effect of these curation steps on results across different types of ecological analyses. We recommend carefully choosing the MOTU clustering threshold, for example, empirical means can be estimated for each marker or targeted taxa using in silico methods with reference databases (Taberlet et al., 2018) or experimentally, using mock communities (Brown, Veach, et al., 2015), and considering removing PCR errors and cross-sample contaminations when designing a curation protocol to study biodiversity patterns. Furthermore, a rigorous data curation strategy including all the curation steps of the present study allowed obtaining accurate diversity estimates and diversity–environment and distance–decay relationships. This demonstrates that the other curation steps should not be neglected.

ACKNOWLEDGEMENTS

We thank the large team of researchers and students that helped collect the data. The research received funding from the French Agence Nationale de la Recherche (ANR) through the GlobNets (ANR-16-CE02-0009) project, and from 'Investissement d'Avenir' grants managed by the ANR (Trajectories: ANR-15-IDEX-02; Montane: OSUG@2020: ANR-10-LAB-56). All computations were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr).

DATA AVAILABILITY STATEMENT

Prefiltered sequencing data as well as associated metadata are available on the Dryad Digital Repository (https://doi.org/10.5061/dryad.0t39970).

ORCID

Irene Calderón-Sanou https://orcid.org/0000-0003-4608-1187
Lucie Zinger https://orcid.org/0000-0002-3400-5825
Wilfried Thuiller https://orcid.org/0000-0002-5388-5274

REFERENCES

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*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Bálint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., … Tedersoo, L. (2016). Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, *40*(5), 686–700. https://doi.org/10.1093/femsre/fuw017

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bonder, M. J., Abeln, S., Zaura, E., & Brandt, B. W. (2012). Comparing clustering and pre-processing in taxonomy analysis. *Bioinformatics*, *28*(22), 2891–2897. https://doi.org/10.1093/bioinformatics/bts552

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. https://doi.org/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. https://doi.org/10.1111/1755-0998.12428

Braun-Blanquet, J. (1946). Über den Deckungswert der Arten in den Pflanzengesellschaften der Ordnung Vaccinio-Piceetalia. *Jahresbericht Der Naturforschenden Gesellschaft Graubünden*, *130*, 115–119.

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*(11), 2234–2251. https://doi.org/10.1002/ece3.1485

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. https://doi.org/10.1016/j.funeco.2014.08.006

Chalmandrier, L., Münkemüller, T., Lavergne, S., & Thuiller, W. (2015). Effects of species' similarity and dominance on the functional and phylogenetic structure of a plant meta-community. *Ecology*, *96*(1), 143–153. https://doi.org/10.1890/13-2153.1

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 297–324. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Chiu, C.-H., & Chao, A. (2016). Estimating and cmparing microbial diversity in the presence of sequencing errors. *PeerJ*, *4*, e1634. https://doi.org/10.7717/peerj.1634

Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, *29*(10), 566–571. https://doi.org/10.1016/j.tree.2014.08.001

de Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, *14*(2), 306–323. https://doi.org/10.1111/1755-0998.12188

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., … Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, *4*(1), https://doi.org/10.1186/s13742-015-0086-1

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Flynn, J. M., Brown, E. A., Chain, F. J. J., MacIsaac, H. J., & Cristescu, M. E. (2015). Toward accurate molecular identification of species in complex environmental samples: Testing the performance of sequence filtering and clustering methods. *Ecology and Evolution*, *5*(11), 2252–2266. https://doi.org/10.1002/ece3.1497

Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., … Beattie, A. J. (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, *432*(7018), 747–750. https://doi.org/10.1038/nature03034

Grömping, U. (2006). Relative Importance for Linear Regression in *R*: The Package relaimpo. *Journal of Statistical Software*, *17*(1). https://doi.org/10.18637/jss.v017.i01

Hiiesalu, I., Öpik, M., Metsis, M., Lilje, L., Davison, J., Vasar, M., … Pärtel, M. (2012). Plant species richness belowground: Higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology*, *21*(8), 2004–2016. https://doi.org/10.1111/j.1365-294X.2011.05390.x

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352

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*(1), 118–123. https://doi.org/10.1111/j.1462-2920.2009.02051.x

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., … Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073–1262073. https://doi.org/10.1126/science.1262073

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. Abstract, 27–29. Citeseer.

Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L., … Thuiller, W. (2018). Mapping the imprint of biotic interactions on β-diversity. *Ecology Letters*, *21*(11), 1660–1669. https://doi.org/10.1111/ele.13143

Schloss, P. D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, *6*(7), e1000844. https://doi.org/10.1371/journal.pcbi.1000844

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *44*(4), 846–849. https://doi.org/10.1099/00207713-44-4-846

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. New York: Oxford University Press.

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. https://doi.org/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*(3), e14. https://doi.org/10.1093/nar/gkl938

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*(8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N. S., Wijesundera, R., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*(6213), 1256688–1256688. https://doi.org/10.1126/science.1256688

Träger, S., Öpik, M., Vasar, M., & Wilson, S. D. (2019). Belowground plant parts are crucial for comprehensively estimating total plant richness in herbaceous and woody habitats. *Ecology*, *100*(2), e02575. https://doi.org/10.1002/ecy.2575

Tuomisto, H. (2003). Dispersal, Environment, and Floristic Variation of Western Amazonian Forests. *Science*, *299*(5604), 241–244. https://doi.org/10.1126/science.1078037

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols a guide to methods and applications* (pp. 315–322). New York: Academic Press.

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 279–338. https://doi.org/10.2307/1943563

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., … Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, *21*(15), 3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x

BIOSKETCH

Irene Calderón-Sanou is a PhD student aiming at a better understanding of multi-trophic assemblages through the use of environmental DNA.

Author contributions: WT initiated the overall idea, and together with ICS, LZ and TM conceived the overall analyses. ICS, LZ and FB conceptualized the data curation strategies, ICS ran the curation procedures and analysed all the results, and led the writing with significant contributions from all co-authors.

SUPPORTING INFORMATION

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

How to cite this article: Calderón-Sanou I, Münkemüller T, Boyer F, Zinger L, Thuiller W. From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *J Biogeogr.* 2020;47: 193–206. https://doi.org/10.1111/jbi.13681

CHAPTER 2:

Building trophic groups and soil food webs from environmental DNA metabarcoding data

Introduction

Studying the enormous span of soil organisms living in the soil and their interactions is challenging (Geisen *et al.* 2019b). Soil taxa are highly diverse ranging from microorganisms such as bacteria to animals including different phyla as nematodes and arthropods. Moreover, soil organisms interact in a number of ways including mutualistic, predatory and parasitic interactions across different trophic levels (Orgiazzi *et al.* 2016). Monitoring the span of soil organisms present in soil multi-trophic communities through conventional monitoring methods is a hard task that needs specialized extraction techniques for each organism's category size, and the identification of numerous taxa by soil specialists (Geisen *et al.* 2019a). Moreover, the role of most soil biota in the ecosystems remain undescribed (Geisen *et al.* 2016; Gongalsky 2021). Therefore, the complexity of soil biodiversity can be best approached by focusing on groups of soil organisms that have similar ecological roles in the ecosystem using the fact that related organisms often share functions (Bardgett & van der Putten 2014; Eisenhauer *et al.* 2019; Potapov *et al.* 2019a). Food webs propose an integrative vision of soil biodiversity as they consider simultaneously several functionally important groups and their functional or trophic linkages (Dunne 2006; Barnes *et al.* 2018). Studying the diversity of food webs through different dimensions i.e. horizontal (within trophic groups) and vertical (across trophic groups) diversity, can bring complementary insights into the understanding of how soil biodiversity responds to environmental changes (Duffy *et al.* 2007; Martinez-Almoyna *et al.* 2019).

The challenges to the study of soil foodwebs are especially limiting at large spatial scales, but the rapid development of high throughput sequencing and the gain on trophic and functional knowledge for soil organisms may allow overcoming these challenges (Roslin & Majaneva 2016; Bohan *et al.* 2017). Nowadays, the development of environmental DNA (eDNA) metabarcoding facilitates the complete monitoring of soil biodiversity at biogeographical scales and across the whole tree of life (Taberlet *et al.* 2012; Deiner *et al.* 2017). Moreover, knowledge about the trophic preferences of soil organisms has been growing in recent years as a consequence of decades of research and the development of new methods allowing to assess the feeding preferences of soil organisms (e.g., stable isotopes, diet-based eDNA metabarcoding, Roslin & Majaneva 2016; Potapov *et al.* 2019). There has also been considerable development in the creation of databases assembling the functional and/or trophic information of some soil organisms and making it easily available to the scientific community (e.g., Wardeh *et al.* 2015; Põlme *et al.* 2020). Combining the eDNA metabarcoding monitoring method with the existing knowledge on trophic and functional relationships of soil organisms,

enables the use of a food web approach to get a better representation of the whole soil multi-trophic community. While some attention has been given to the construction of heuristic food webs from the data generated by eDNA metabarcoding (e.g., for macroinvertebrates in freshwater ecosystems, Compson *et al.* 2018), this has seldom been applied to soil organisms. Assigning a large number of taxa to trophic guilds and identifying all the possible trophic interactions is challenging, and needs 1) the development of a common trophic framework for the whole soil biota, 2) the availability of information on how soil organisms use available resources, and 3) the definition of the resolution to be used to define these guilds, from a few broad 'trophic classes' (e.g. decomposers, predators) to several refined 'trophic groups' (e.g. bacterivorous nematodes, arbuscular mycorrhizal or saprotrophic fungi), depending on the research question.

Here, I tackle this challenge by providing a systematic framework combining soil eDNA metabarcoding data with databases and information on soil organisms resource acquisition strategies to build heuristic soil food webs (Fig. 1). The aim of this chapter was to provide the procedure used 1) to classify the taxonomically annotated Molecular Operational Taxonomic Units (MOTUs) into different trophic groups (or classes), and, 2) to build the metaweb (i.e., regional food web) by providing the trophic links between these groups. Finally, I present the two versions of the soil metaweb that were obtained and used in this thesis, at two levels of resolution. I decided to use two levels of resolution to be able to study soil biodiversity across hierarchical levels of biodiversity organization, which reveal different aspects of the multi-trophic community and are related to different ecosystem processes.

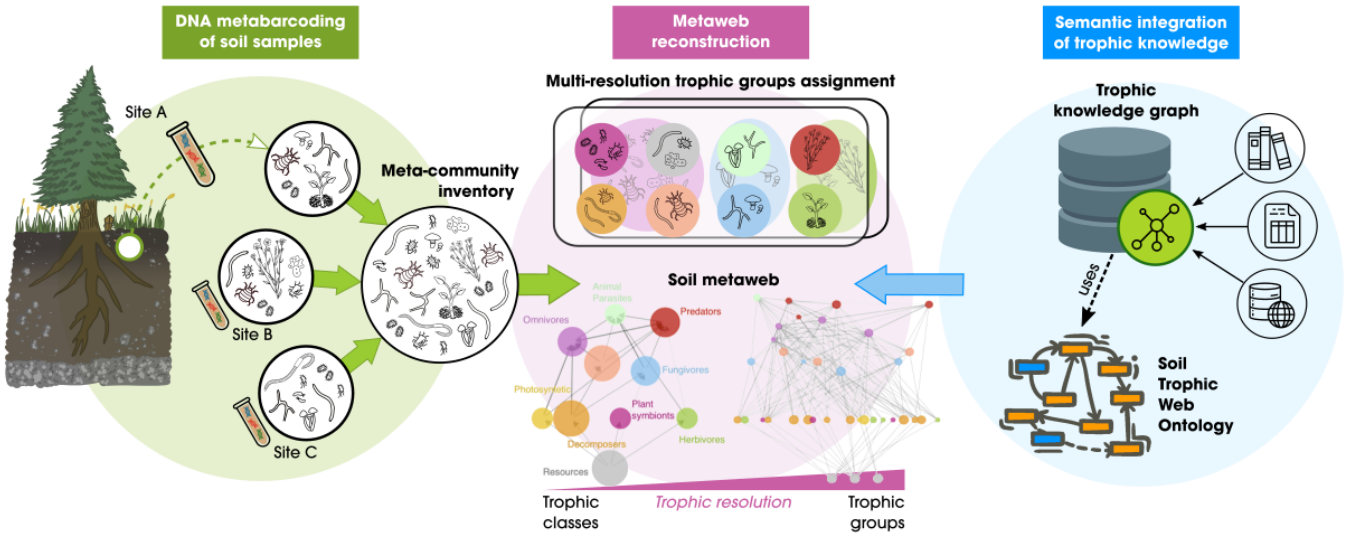


Figure 1. Workflow used to classify MOTUs retrieved from eDNA metabarcoding data into trophic groups and build soil food webs. In a first step, soil samples are collected and processed to obtain a clean list of taxonomically annotated MOTUs (left panel). In a second step, the functional or feeding information of the main sampled taxa is assembled from expert knowledge, existing databases and literature. An ontology to build the soil food webs is defined. This ontology describes the distribution of trophic groups across trophic levels and includes the definition of some rules for interactions (right panel). Finally, the trophic information is matched to the eDNA data, which allows to categorize the MOTUs into trophic groups (or classes) and to build a metaweb containing all the trophic groups and their potential interactions (central panel). The metaweb can be conceived at different levels of resolution (e.g. trophic groups and trophic classes). Figure provided by Nicolas Leguillarme.

Material and Methods

eDNA metabarcoding data description – The data used here comes from the GlobNets project (2016-2022). Multi-trophic assemblage datasets of soil biodiversity were sampled across multiple forest and grassland plots along environmental or disturbances gradients in different biomes using eDNA metabarcoding. My PhD focuses on two specific datasets: the TROMSO dataset from soils sampled along a disturbance gradient in the subarctic birch forests of the Varanger region, in Northeastern Norway (Chapter 5), and, 2) the ORCHAMP dataset from

the Orchamp observatory consisting of soil samples collected along elevational transects in the French Alps (Chapters 1, 3 and 4). The general bioinformatic pipeline used to clean these datasets is detailed in Chapter 1, and the differences in the cleaning processes associated with each dataset are described in the corresponding chapters. Overall, a complete OBITools pipeline was performed (Boyer *et al.* 2016), followed by a pre-processing using the pipeline described in the 'metabaR' R package (Zinger *et al.* 2020). For each of these datasets, sequences were clustered into MOTUs using 'sumaclust' (Mercier *et al.* 2013), which were taxonomically assigned using the *ecotag* command from the OBITools, and marker-specific databases built from the EMBL database version for clade specific markers (fung02, inse01, olig01, coll02, Taberlet *et al.* 2018), and with the SILVAngs pipeline (Quast *et al.* 2013), using the SILVA version 132 for ribosomal universal markers (euka02, bact01). The taxonomic annotation of the MOTUs was used to classify the MOTUs into trophic groups and trophic classes and to build the metaweb.

Classification of MOTUs into trophic groups and trophic classes – In the first instance, I identified the major trophic classes commonly associated with soil groups from the literature (e.g., Moore & de Ruiter 1991; Orgiazzi *et al.* 2016; Barnes *et al.* 2018). The trophic classes included autotrophs, decomposers, detritivores, phytophagous or phytoparasites, plant mutualists, bacterivores, fungivores, omnivores, predators and zooparasites. Next, I defined finer trophic groups by separating phylogenetic distant groups that could have a different set of prey/predators (e.g., bacterivore mites *vs.* bacterivore nematodes) or groups differing in their resources acquisition strategy (e.g. different types of mycorrhiza and saprotrophs). In this thesis, the taxonomic rank that I used to delimitate phylogenetic distant groups was at different levels and comprised Bacteria, Fungi, Protista, the different phyla within Metazoa, and the different classes or orders within Arthropoda and Annelida (Fig. 2). For each of these high-rank taxa, the taxonomically annotated MOTUs were assigned to the different trophic classes using specific assigning tools (e.g., Faprotax, NINJA, Funguild) or sources from the literature and different criteria, which are detailed in Table 1. The fine trophic groups thus consisted of a mixture between trophic classes and taxonomic high-rank taxa (Fig. 2). Trophic groups were defined mainly on the basis of the taxonomic resolution of the marker and trophic or functional information available in the literature (see discussion).

Table 1. Methods and criteria used to assign and define the trophic groups. Assigning tools or databases, a detailed description on the assigned criteria and references used to assign the taxa to trophic groups is provided for each database and for each high rank taxa.

Kingdom (eDNA marker)	Assigning tools	Description of the methods and assigning criteria		References
		TROMSO dataset	ORCHAMP dataset	
Fungi (Fung02)	FUNGuild[1] (FG) Fungal Traits[2] (FT)	Trophic annotations were done using FG. Guilds were kept based on the following criteria: (1) reflecting the diversity of broad trophic groups found in fungi (saprotrophs, symbionts and plant pathogens), (2) that could respond differently to disturbances for the fine groups definition, e.g. we distinguish between the different types of resources used by the saprotroph (wood, soil, undefined) and the different types of mycorrhiza because these could change differently following the moth outbreaks, and (3), that together they represent more than 70% of the reads in the dataset.	Trophic annotations were made using FT and FG. All MOTUs annotated at the genus level were assigned using the primary lifestyle from FT. MOTUs annotated at higher taxonomic levels were assigned using FG and classified to correspond to the categories obtained from FT. Only groups having more than 0.1 % of the total reads from the marker were kept as trophic groups.	(1) Nguyen *et al.* 2016 (2) Põlme *et al.* 2020
Bacteria (Bact02/ Bact01)	Faprotax[4], Wardeh database[5] (WDB), Literature	Bacteria were divided in Heterotrophic and Photosynthetic. - Photosynthetic included Cyanobacteria, Chloroflexi and taxa identified as phototrophs from Faprotax. - Heterotrophic bacteria identified in Faprotax as part of the N-cycle (i.e. nitrifying, N-fixing), pathogens (absent in our dataset) or predatory bacteria (i.e. Myxobacteria) were considered as different groups. The rest of the taxa were classified as copiotrophs and oligotrophs using the classification in [6].	Bacteria were annotated using Faprotax. Autotrophic bacteria were separated into chemolitoautotroph and photolitoautotroph based on Faprotax annotations. Cyanobacteria and Chloroflexi were also included in the photolitoautotroph group. Pathogens (zoo- and plantparasites) were identified using both WDB and Faprotax. The rest of bacteria was classified as heterotrophic bacteria and considered decomposers. The pipeline for Bacteria assignment was developed and automatized by Lucie Zinger.	(4) Louca *et al.* 2016 (5) Wardeh *et al.* 2015 (6) Ho *et al.* 2017
Protist (Euka01/ Euka02)	Literature	MOTUs of protists were classified in all the broad trophic classes definitions. Protists were considered Eukarvore or Protistivore (i.e. Predator) when feeding mainly on protists but not bacteria, Bacterivores when feeding mainly on bacteria, and Omnivores when feeding on both bacteria and eukaryotes like in [7]. Only completely phototroph protists were classified as Photosynthetic. We based mainly on [8] for trophic groups assignments and we complemented with compiled databases ([7] for Cercozoa, and a general database compiled by colleages in the GlobNets project). For groups presenting very variable feeding modes (e.g. Dinoflagellata, Cilliophora) we avoided doing generalizations at higher taxonomic levels. We remove protists that were exclusively parasites on vertebrates, because vertebrates were not included in our soil network.		(7) Fiore-Dono *et al.* 2019 (8) Adl *et al.* 2019

Metazoa (Euka01/ Euka02)	NEMAGuild[9], NEMAPLEX[10], Literature	We kept the following phyla: Annelida (O. Haplotaxida), Arthropoda, Mollusca (C. Gastropoda), Nematoda, Rotifera and Tardigrada. The phylum Arthropoda was divided in the classes: Arachnida (mites and spiders), Collembola, Chilopoda and Insecta. Insects were very poorly represented in the dataset (1 MOTU, few reads) or non resolutive for the marker, thus we excluded them from the analyses. For taxa with conserved trophic behavior and/or with no enough taxonomic resolution (due to the DNA marker), we did generalizations concerning their trophic group. For example, although Collembola have varying trophic behaviors (e.g. fungivores, predators, detritivores) we generalized them all as Fungivores as they have been historically classified in soil food webs because no enough resolution of the marker. Nematodes and mites were classified into the different trophic classes using NEMAGuild, NEMAPLEX, and more literature for specific taxa not represented in the databases. For Nematoda, no generalizations were made at higher taxonomic levels than family because of the variability of feeding habits within an order. For mites, generalizations were made depending on the group.	We kept the following phyla: Arthropoda, Mollusca (C. Gastropoda), Nematoda, Rotifera and Tardigrada. The phylum Arthropoda was divided in the classes: Arachnida (separated into mites, spiders and pseudoscorpions), Diplopoda, Chilopoda and Protura. Insects, springtails and oligochaetes were removed and classified based on the data from the specific marker for insects. Nematodes and mites were classified into the different trophic classes using NEMAGuild, NEMAPLEX, and more literature for specific taxa not represented in the databases. For Nematoda, no generalizations were made at higher taxonomic levels than family because of the variability of feeding habits within an order. For mites, generalizations were made depending on the group.	(9) Nguyen *et al.* 2016 (10) http://nemaplex.ucdavis.edu/
		ORCHAMP dataset		
Collembola (Coll01)	Literature	Collembola were classified based on the trophic groups described in [11]. Only MOTUs annotated at least to the family level were assigned. Data from [12] was used to determine the habitat layer of the genus in families present in different trophic groups, to assess their trophic group based on [11].		(11) Potapov *et al.* 2016 (12) Saifutdinov *et al.* 2020
Oligochaeta (Olig01)	Literature	Oligochaetes were divided into Enchytraeids and Earthworms. Earthworms were classified into the ecological categories described in [13], and further verified by an expert (Mickael Hedde).		(13) Bottinelli *et al.* 2020
Insecta (Inse01)	Literature	Insects were classified into trophic groups based on [14] and a general local database compiled by collaborators of the GlobNets project, and the local database was prioritized. Orders of insects having an aquatic larvae and flying adult (e.g. Odonata, Ephemeroptera) were removed from the dataset. Also insects spending most time of their cycle aboveground, such as Hemiptera, Lepidoptera, Orthoptera and some families of Hymenoptera were removed. Trophic assignment of larvae was used to assign insect taxa with flying adults (e.g. Diptera). For some few families with different trophic behavior between the larvae and the adult, larvae trophic group was preferred.		(14) Rainford & Mayhew 2015

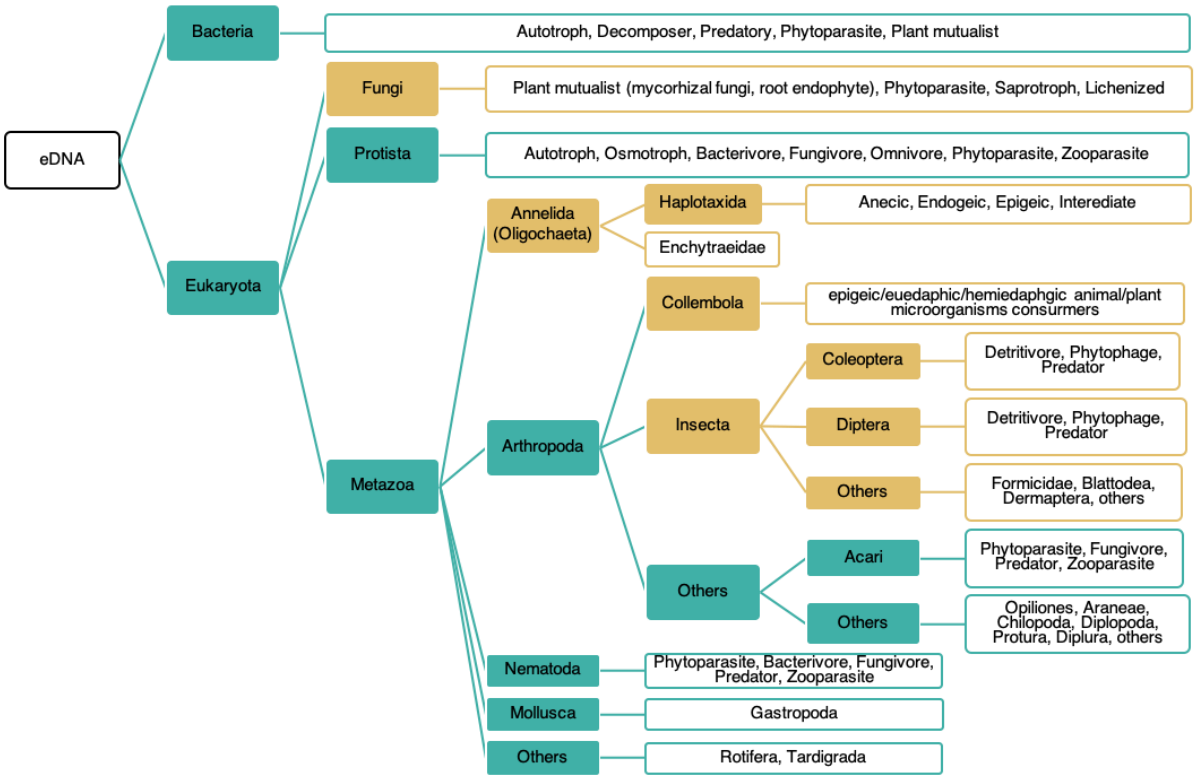


Figure 2. Schematic representation of the classification of MOTUs into trophic groups from different high-rank taxa. Colours show to the taxa sampled with universal markers (in green: euka02/euka01 for eukaryote, bact02/bact01 for bacteria) and with clade specific markers (in yellow: fung02 for fungi, inse01 for insect, olig01 for oligochaete, and coll02 for collembola). Note: not all clade specific markers were available for the TROMSO datasets, thus information from universal markers were used instead. Open boxes contain the defined trophic groups from the last high rank taxa in filled boxes.

From trophic groups and trophic classes to metaweb – A metaweb is a theoretical network containing all trophic groups (or classes) and their potential interactions of the large scale ecosystem under study (for example in this thesis, subarctic and alpine systems, respectively represented by TROMSO and ORCHAMP datasets). Observable local food webs are then subsets of this theoretical metaweb. In this thesis, I first built a fine resolution metaweb at the level of trophic groups, and then deduced the metaweb at the trophic class level from the first one. This was possible given the hierarchical nature of the groups, i.e., each trophic group was assigned to a unique trophic class. To do this, I used the R packages metanetwork

(https://gitlab.com/marcohlmann/metanetwork) and econetwork (Miele *et al.* 2021), which allow to aggregate networks at different resolutions.

Three main basal resources were established for the construction of the metaweb at the trophic group level: energy (solar or chemical), plants, and organic matter. In the metaweb at the trophic class level, these resources were grouped into an unique node accounting for the three basal resources. Resource nodes were added to the metaweb with a structural purpose.

I added trophic links between trophic groups based on the main feeding preferences of the group. Therefore, plant symbiont groups (i.e., plant mutualists and phytoparasites) were associated with the plant resource, detritivores and decomposers were associated with the organic matter resource and autotrophs were associated with the energy resource. Next, bacterivores were associated with all trophic groups containing bacteria, and fungivores with all trophic groups containing fungi (but see additional constraints below). Omnivore protists, by definition (Table 1), were associated with all trophic groups containing bacteria or fungi. Finally, I added the trophic interactions of the remaining trophic groups of omnivores, predators and zooparasites within the metaweb thanks to a literature review based on the dietary preferences of the majority of taxa within each of these groups. That is, I chose the taxa that constituted more than 90% of the group's abundance (i.e., read counts) and did a literature review for each of these taxa. All the possible interactions between these taxa and the other trophic groups were attributed to the whole group.

Some additional constraints were added when assigning the trophic interactions between trophic groups based on (1) the organism's size, i.e. predators fed only on smaller prey, with some exceptions like animal parasites and omnivore nematodes that can eat larger prey, and macro-organisms did not feed on microorganisms (except for bacterivores and fungivores), (2) habitat differentiation, i.e. strict plant endoparasites (i.e. protists) were not considered as prey of other free-living predators, and (3) feeding preferences, e.g. fungivores fed only on saprotrophic fungi and Ectomycorrhizal, which are preferred to arbuscular mycorrhizal fungi.

From metaweb to local food webs – From the metaweb, I deduced the composition and structure of the local soil food webs, based on the trophic classes or groups detected locally, and assuming that co-occurring classes or groups interact as in the metaweb (i.e. that the local web is a strict subset of the metaweb). For the soil food webs at the trophic group resolution, trophic groups were weighted by their relative abundance and trophic interactions were binary (i.e., present or absent). For the soil food webs at the trophic class resolution, trophic class weight was estimated as the sum of the relative abundances of the trophic groups inside the trophic class, and the

interactions were weighted by the probability of interaction between two classes given the interactions between their respective trophic groups and the relative abundances of these groups (estimated as in Ohlmann *et al.* 2019).

Results

TROMSO metaweb - The metaweb was composed of 10 trophic classes and 32 interactions at the coarser resolution (Fig. 3a), corresponding to 40 trophic groups (of which 3 are resources) and 194 potential interactions (Fig. 3b) at the finer resolution.

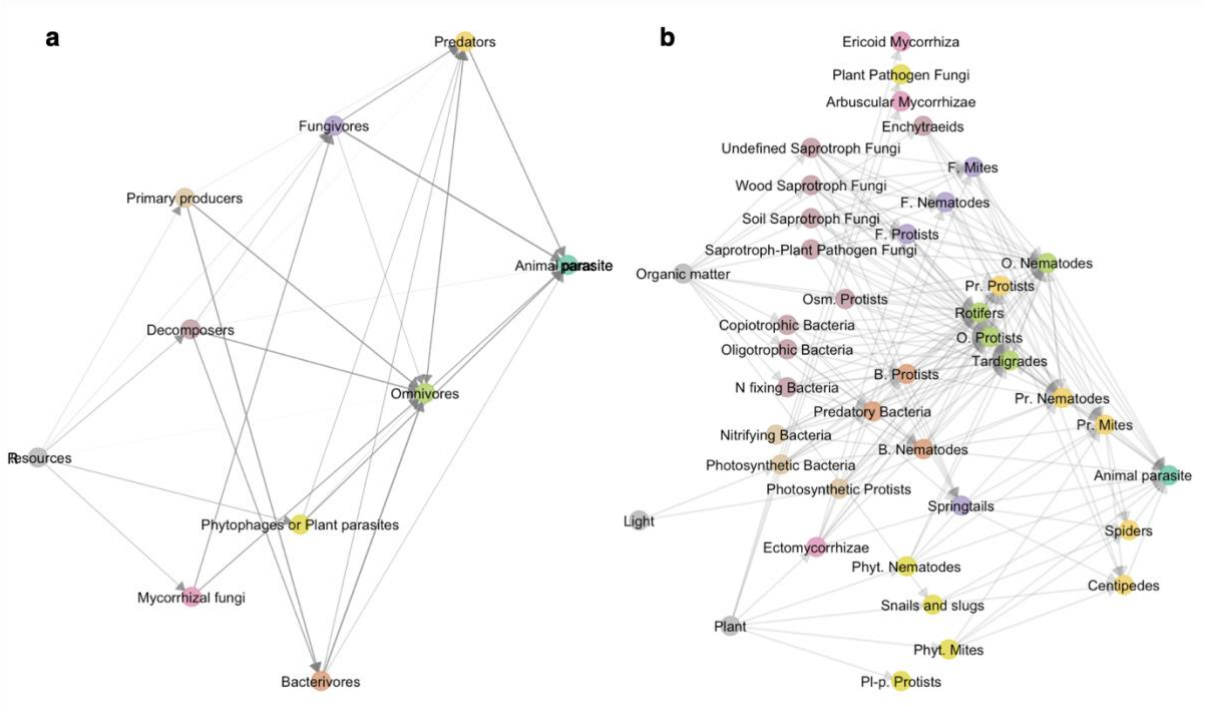


Figure 3. Metaweb or regional soil food web of subarctic birch forests (TROMSO dataset) at two levels of resolution: trophic class resolution (a) and trophic group resolution (b). Colour of the trophic groups in (b) correspond to the trophic classes in (a). The nodes are distributed horizontally based on their trophic level from the left (basal levels) to the right (higher levels). The figure and the trophic level calculations were done using the R package 'metanetwork'.

ORCHAMP metaweb - The metaweb was composed of 11 trophic classes and 45 interactions at the coarser resolution (Fig. 2a), corresponding to 55 trophic groups (of which 3 are resources) and 383 potential interactions (Fig.2a) at the finer resolution.

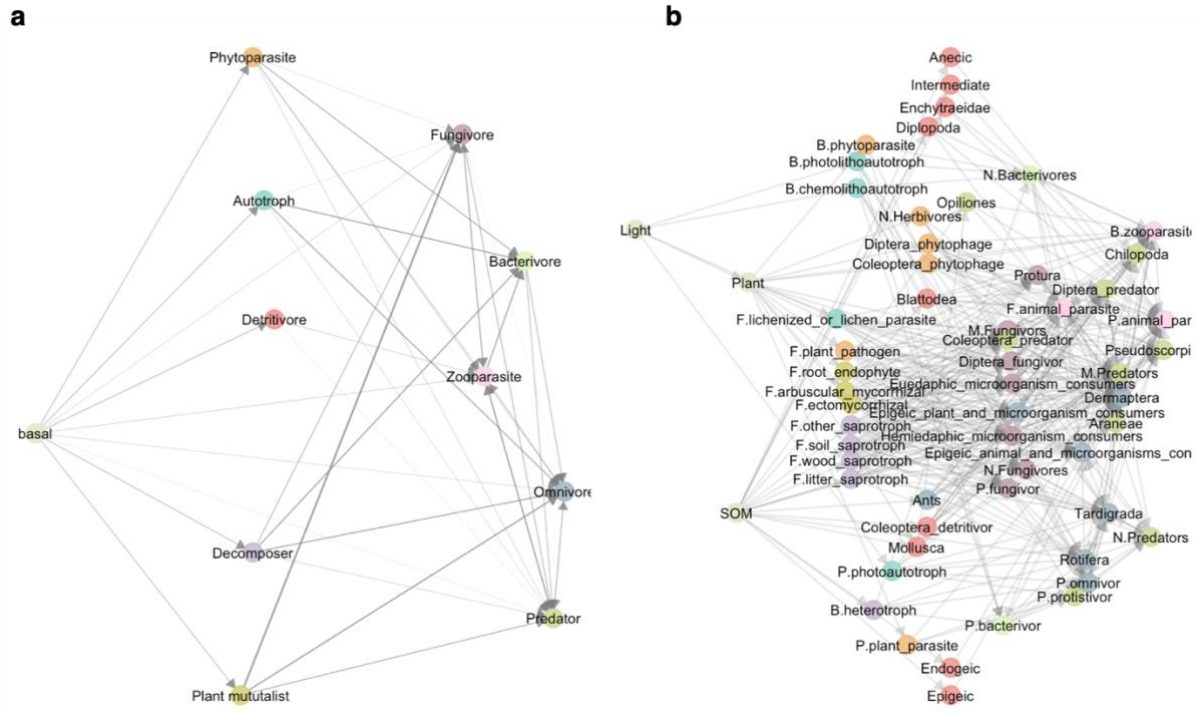


Figure 4. Metaweb or regional soil food web of the French Alps (ORCHAMP dataset) at two levels of resolution: trophic class resolution (a) and trophic group resolution (b). Colour of the trophic groups in (b) correspond to the trophic classes in (a). The nodes are distributed horizontally based on their trophic level from the left (basal levels) to the right (higher levels). The figure and the trophic level calculations were done using the R package 'metanetwork'.

Discussion

In this chapter I detailed the procedure used to analyse eDNA data by grouping all soil taxa through their ecological similarity and by structuring these groups based on their feeding relations in food webs. The construction of heuristic food webs allows for an integrative and ecological representation of soil multi-trophic assemblages in terrestrial ecosystems. Several challenges related to the construction of heuristic food webs from eDNA data were identified and addressed in the methodology presented here. In this discussion I justify some of the choices made and present their potential biases or limitations.

The construction of food webs from eDNA data is limited both by the taxonomic resolution of the marker and by the trophic or functional information available in the literature. Both of these factors can influence the procedure and decisions made throughout the process of constructing trophic networks. For example, for the TROMSO dataset, I kept all springtails within a single trophic group because the Euka02 marker used to sample this taxon was not sufficiently resolute to make finer trophic groups. In contrast, for the ORCHAMP dataset (Chapters 3 and 4), the use of a specific marker for springtails (Coll02) provided higher taxonomic resolution (i.e. down to the family or genus level), which allowed the use of a finer trophic classification (e.g., epigeic animal and microorganisms consumer). For other taxa, such as rotifers and tardigrades, specific trophic groups are poorly defined, so each of these phyla was kept as a trophic group (Potapov *et al.* 2022) .Furthermore, in the process of assigning MOTUs to trophic groups, some data may be lost, for example, corresponding to MOTUs without sufficient taxonomic resolution or taxa that could not be assigned to a trophic group. Therefore, a balance has to be found between the use of ecologically significant groups while losing some information, and the use of broad taxa as groups that include most of the sequences but are ecologically less informative. Checking the amount of information to be lost, for example, the percentage of unassigned MOTUs reads, is an important step during the process and can be decisive in the definition or resolution to be used to construct the groups. For further analyses, I re-defined the trophic groups to fit the requirements of the statistical analyses to be performed. For example, in Chapters 3 and 5, an individual model was fitted for each trophic group to estimate how group diversity varied as a function of environmental predictors. In this case, some fine trophic groups were merged to create a group with sufficient variation in diversity or with sufficient occurrences (e.g. all earthworms were merged into a single group).

Another challenge when building heuristic soil food webs is to find the desired resolution to define the trophic groups, from a few broad 'trophic classes' (e.g. decomposers, predators) to several refined 'trophic groups' (e.g. bacterivorous nematodes, saprotrophic fungi). The definition of trophic groups and their resolution is linked to the ecological questions being addressed. When the ecological hypotheses being tested are explicitly related to bottom-up processes (e.g. the effect of basal resources on diversity across trophic levels), a high resolution of basal groups may be preferred (e.g. different types of mycorrhizal or saprotrophic fungi, or separating bacterial phyla like in Morriën *et al.* 2017), however, some studies specifically interested in assessing top-down effects or relating trophic structure to ecosystem functions often choose to keep groups at a higher resolution (e.g., Schuldt *et al.* 2017). In this thesis, I have chosen to consider different resolutions to construct the trophic groups. The integration of different resolutions in ecological network analyses allows the study of changes in the structure of food webs by taking into account trophic redundancy and/or trophic complementarity. Here, I rely on the conceptual framework found in the literature to define groups at different resolutions. However, other models or algorithms that detect structural equivalence in networks, such as stochastic block models, could be considered in future studies to identify 'modules' of taxa or fine trophic groups and aggregate them into larger groups (Gauzens *et al.* 2015; O'Connor *et al.* 2020; Bloor *et al.* 2021).

Finally, soil food web construction needs the collaboration of multiple soil specialists working with different subgroups of taxa that need to agree on different concepts, and this represented a major challenge in this thesis. Existing classifications and concepts to describe feeding behaviours can be very different across soil taxonomic subgroups (Hedde *et al. unpublished*), complexifying the task of having a common trophic framework for the whole soil biota. During my thesis I participated in a collaborative work within a soil ecologists' community with the aim to define an ontology with an homogenized vocabulary to build soil food webs (https://github.com/nleguillarme/soil_food_web_ontology). The process to review inconsistencies in trophic and functional vocabulary and to find a common agreement between the different parts can take time and need of common effort to surpass conceptual limitations. Moreover, databases and literature are constantly actualizing. Developing tools for the integration of functional and trophic information of the soil biota can be very useful and may save time to ecological research (e.g., GRATIN gratin.nova.u-ga.fr:7200). Providing an standardized or automatized method to build heuristic soil food webs from eDNA data would thus need the stabilisation of the vocabulary and concepts of trophic ecology across soil

organisms, the common effort to integrate and homogenize this information and the development of 'friendly-user' bioinformatic tools to make it accessible and exploitable to the scientific community. But in principle, the improvement of the method would depend on the continued and joint effort of soil ecologists (including taxonomists and naturalists) that sample, identify and describe species, but also that contribute to assign traits, functional and trophic information to species or broader taxa.

Acknowledgments

This work would not have been possible without the collaboration of fellow soil specialists who guided me through fruitful discussions. For this I would especially like to thank Lucie Zinger, Mickael Hedde, Heidy Schimann, Mélanie Roy, Jean Trap and Stéphane Bec. I also thank Camille Martinez-Almoyna and Marc Ohlmann who started the conceptual framework to classify MOTUs into trophic groups and contributed to its improvement. Finally, I would like to acknowledge the work of Nicolas Leguillarme and Sara Si-moussi in providing informatic but also conceptual tools that facilitated the gathering of information and their use to build the metawebs.

References

Bardgett, R.D. & van der Putten, W.H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, 515, 505–511.

Barnes, A.D., Jochum, M., Lefcheck, J.S., Eisenhauer, N., Scherber, C., O'Connor, M.I., *et al.* (2018). Energy Flux: The Link between Multitrophic Biodiversity and Ecosystem Functioning. *Trends Ecol Evol*, 33, 186–197.

Bloor, J.M.G., Si-Moussi, S., Taberlet, P., Carrère, P. & Hedde, M. (2021). Analysis of complex trophic networks reveals the signature of land-use intensification on soil communities in agroecosystems. *Sci Rep*, 11, 18260.

Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017). Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends Ecol Evol*, 32, 477–487.

Bottinelli, N., Hedde, M., Jouquet, P. & Capowiez, Y. (2020). An explicit definition of earthworm ecological categories – Marcel Bouché's triangle revisited. *Geoderma*, 372, 114361.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. (2016). OBITOOLS : a UNIX - inspired software package for DNA metabarcoding. *Mol Ecol Resour*, 16, 176–182.

Compson, Z.G., Monk, W.A., Curry, C.J., Gravel, D., Bush, A., Baker, C.J.O., *et al.* (2018). Chapter Two - Linking DNA Metabarcoding and Text Mining to Create Network-Based Biomonitoring Tools: A Case Study on Boreal Wetland Macroinvertebrate Communities. In: *Advances in Ecological Research*, Next Generation Biomonitoring: Part 2 (eds. Bohan, D.A., Dumbrell, A.J., Woodward, G. & Jackson, M.). Academic Press, pp. 33–74.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., *et al.* (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26, 5872–5895.

Duffy, J.E., Cardinale, B.J., France, K.E., McIntyre, P.B., Thébault, E. & Loreau, M. (2007). The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecol Letters*, 10, 522–538.

Dunne, J.A. (2006). The network structure of food webs. In: *Ecological networks: linking structure to dynamics in food webs* (eds. Pascual, M. & Dunne, J.A.). Oxford University Press, pp. 27–86.

Eisenhauer, N., Schielzeth, H., Barnes, A.D., Barry, K.E., Bonn, A., Brose, U., *et al.* (2019). A multitrophic perspective on biodiversity–ecosystem functioning research. *Adv Ecol Res.*, 61, 1–54.

Gauzens, B., Thébault, E., Lacroix, G. & Legendre, S. (2015). Trophic groups and modules: two levels of group detection in food webs. *J. R. Soc. Interface*, 12, 20141176.

Geisen, S., Briones, M.J.I., Gan, H., Behan-Pelletier, V.M., Friman, V.-P., de Groot, G.A., *et al.* (2019a). A methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry*, 136, 107536.

Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T. & Bonkowski, M. (2016). The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biology and Biochemistry*, 94, 10–18.

Geisen, S., Wall, D.H. & van der Putten, W.H. (2019b). Challenges and Opportunities for Soil Biodiversity in the Anthropocene. *Current Biology*, 29, R1036–R1044.

Gongalsky, K.B. (2021). Soil macrofauna: Study problems and perspectives. *Soil Biology and Biochemistry*, 159, 108281.

Hedde, M., Blight, O., Briones, M.J.I., Bonfanti, J., Brauman, A., Brondani, M., Calderón Sanou, I., Clause, J., *et al.* Avoiding cacophony in soil fauna classifications. Under review in *Geoderma*.

Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., Ohlmann, M., Foulquier, A., Clément, J., *et al.* (2019). Multi-trophic β-diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. *Funct Ecol*, 33, 2053–2064.

Mercier, C., Boyer, F., Bonin, A. & Coissac, E. (2013). SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. In: *Programs and Abstracts of the SeqBio 2013 workshop. Abstract*. Citeseer, pp. 27–29.

Miele, V., Matias, C., Dray, S., Botella, C., Ohlmann, M., Poggiato, G., *et al.* (2021). *econetwork: Analyzing Ecological Networks*.

Moore, J.C. & de Ruiter, P.C. (1991). Temporal and spatial heterogeneity of trophic interactions within below-ground food webs. *Agriculture, Ecosystems & Environment*, 34, 371–397.
Morriën, E., Hannula, S.E., Snoek, L.B., Helmsing, N.R., Zweers, H., de Hollander, M., *et al.* (2017). Soil networks become more connected and take up more carbon as nature restoration progresses. *Nature Communications*, 8, 14349.
Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., *et al.* (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol*, 20, 241–248.
O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J Biogeogr*, 47, 181–192.
Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecol Lett*, 22, 737–747.
Orgiazzi, A., Bardgett, R.D. & Barrios, E. (2016). Global soil biodiversity atlas. *Global soil biodiversity atlas.*
Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B.D., Clemmensen, K.E., Kauserud, H., *et al.* (2020). FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers*, 105, 1–16.
Potapov, A.A., Semenina, E.E., Korotkevich, A.Yu., Kuznetsova, N.A. & Tiunov, A.V. (2016). Connecting taxonomy and ecology: Trophic niches of collembolans as related to taxonomic identity and life forms. *Soil Biology and Biochemistry*, 101, 20–31.
Potapov, A.M., Beaulieu, F., Birkhofer, K., Bluhm, S.L., Degtyarev, M.I., Devetter, M., *et al.* (2022). Feeding habits and multifunctional classification of soil-associated consumers from protists to vertebrates. *Biological Reviews*, brv.12832.
Potapov, A.M., Scheu, S. & Tiunov, A.V. (2019a). Trophic consistency of supraspecific taxa in below-ground invertebrate communities: Comparison across lineages and taxonomic ranks. *Functional Ecology*, 33, 1172–1183.
Potapov, A.M., Tiunov, A.V. & Scheu, S. (2019b). Uncovering trophic positions and food resources of soil animals using bulk natural stable isotope composition: Stable isotopes in soil food web studies. *Biol Rev*, 94, 37–59.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, 41, D590–D596.
Rainford, J.L. & Mayhew, P.J. (2015). Diet Evolution and Clade Richness in Hexapoda: A Phylogenetic Study of Higher Taxa. *Am Nat*, 186, 777–791.
Roslin, T. & Majaneva, S. (2016). The use of DNA barcodes in food web construction—terrestrial and aquatic ecologists unite! *Genome*, 59, 603–628.
Saifutdinov, R.A., Sabirov, R.M. & Zaitsev, A.S. (2020). Springtail (Hexapoda: Collembola) functional group composition varies between different biotopes in Russian rice growing systems. *European Journal of Soil Biology*, 99, 103208.
Schuldt, A., Bruelheide, H., Buscot, F., Assmann, T., Erfmeier, A., Klein, A.-M., *et al.* (2017). Belowground top-down and aboveground bottom-up effects structure multitrophic community relationships in a biodiverse forest. *Scientific Reports*, 7, 4222.
Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. (2018). *Environmental DNA: for biodiversity research and monitoring*. Oxford University Press, New York.
Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L.H. (2012). Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
Wardeh, M., Risley, C., McIntyre, M.K., Setzkorn, C. & Baylis, M. (2015). Database of host-pathogen and related species interactions, and their global distribution. *Sci Data*, 2, 150049.
Zinger, L., Lionnet, C., Benoiston, A.-S., Donald, J., Mercier, C. & Boyer, F. (2020). metabaR : an R package for the evaluation and improvement of DNA metabarcoding data quality, Preprint at http://biorxiv.org/lookup/doi/10.1101/2020.08.28.271817.

CHAPTER 3:

Energy and physiological tolerance explain multi-trophic soil diversity in temperate mountains

Received: 25 November 2021 | Revised: 17 March 2022 | Accepted: 22 March 2022

DOI: 10.1111/ddi.13529

RESEARCH ARTICLE

Diversity and Distributions WILEY

Energy and physiological tolerance explain multi-trophic soil diversity in temperate mountains

Irene Calderón-Sanou[1] | Lucie Zinger[2] | Mickael Hedde[3] | Camille Martinez-Almoyna[1] | Amelie Saillard[1] | Julien Renaud[1] | Ludovic Gielly[1] | Norine Khedim[4,1] | Clement Lionnet[1] | Marc Ohlmann[1] | Orchamp Consortium | Tamara Münkemüller[1] | Wilfried Thuiller[1]

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

[2]Département de biologie, Institut de Biologie de l'ENS (IBENS), École normale supérieure, CNRS, INSERM, Université PSL, Paris, France

[3]Eco&Sols, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France

[4]Univ. Savoie Mont-Blanc, Univ. Grenoble Alpes, CNRS, EDYTEM, Chambéry, France

Correspondence
Irene Calderón-Sanou, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France.
Email: irecalsa@gmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-10-LAB-56, ANR-15-IDEX-02 and ANR-16-CE02-0009

Editor: Alfried Vogler

Abstract

Aim: Although soil biodiversity is extremely rich and spatially variable, both in terms of species and trophic groups, we still know little about its main drivers. Here, we contrast four long-standing hypotheses to explain the spatial variation of soil multi-trophic diversity: energy, physiological tolerance, habitat heterogeneity and resource heterogeneity.

Location: French Alps.

Methods: We built on a large-scale observatory across the French Alps (Orchamp) made of seventeen elevational gradients (~90 plots) ranging from low to very high altitude (280–3,160 m), and encompassing large variations in climate, vegetation and pedological conditions. Biodiversity measurements of 36 soil trophic groups were obtained through environmental DNA metabarcoding. Using a machine learning approach, we assessed (1) the relative importance of predictors linked to different ecological hypotheses in explaining overall multi-trophic soil biodiversity and (2) the consistency of the response curves across trophic groups.

Results: We showed that predictors associated with the four hypotheses had a statistically significant influence on soil multi-trophic diversity, with the strongest support for the energy and physiological tolerance hypotheses. Physiological tolerance explained spatial variation in soil diversity consistently across trophic groups, and was an especially strong predictor for bacteria, protists and microfauna. The effect of energy was more group-specific, with energy input through soil organic matter strongly affecting groups related to the detritus channel. Habitat and resource heterogeneity had overall weaker and more specific impacts on biodiversity with habitat heterogeneity affecting mostly autotrophs, and resource heterogeneity affecting bacterivores, phytophagous insects, enchytraeids and saprotrophic fungi.

Main Conclusions: Despite the variability of responses to the environmental drivers found across soil trophic groups, major commonalities on the ecological processes

structuring soil biodiversity emerged. We conclude that among the major ecological hypotheses traditionally applied to aboveground organisms, some are particularly relevant to predict the spatial variation in soil biodiversity across the major soil trophic groups.

KEYWORDS
environmental DNA metabarcoding, French Alps, macroecology, random forest, soil biodiversity, trophic groups

1 | INTRODUCTION

With the ever-increasing availability of biodiversity information, a global synthesis on the major ecological determinants of broad-scale biodiversity patterns is starting to emerge (Belmaker & Jetz, 2015; Braga et al., 2019; Pontarp et al., 2019; Thuiller et al., 2020). This general understanding is pivotal to forecast how biodiversity responds to natural and anthropogenic changes (McGill et al., 2015; Urban et al., 2016). Yet, most of the empirical support is grounded on specific aboveground macroorganisms, in particular vertebrates and plants. Comparatively, soil biodiversity has been largely less studied (Guerra et al., 2020), although it represents one quarter of global diversity and is essential for decomposition, nutrient cycling or carbon sequestrations (Delgado-Baquerizo et al., 2020; Wagg et al., 2014). Therefore, it remains unclear whether the ecological hypotheses that hold true for aboveground systems, such as the energy or the habitat heterogeneity hypotheses, also apply to the massive bulk of belowground biodiversity (Bardgett et al., 2005; Decaëns, 2010).

Historically, the complexity of studying the soil compartment, for example, complex physical structure (Young & Crawford, 2004), taxonomic impediment (Decaëns, 2010), scale of approach (Bardgett et al., 2005; Ettema & Wardle, 2002; Thakur et al., 2020), has hindered the integration of soil biodiversity into a broader ecological hypothesis testing framework. Yet, our ability to study soil biodiversity at large spatial scales is constantly improving with joint taxonomic efforts, the development of new sampling technologies (e.g. eDNA metabarcoding) and the increase of collaborative databases and initiatives (e.g. Drilobase, Earth microbiome project, Global Soil Biodiversity Initiative). Global-scale analyses have thus recently emerged for several soil organism groups (e.g. Tedersoo et al., 2014 for fungi; Delgado-Baquerizo et al., 2018 for bacteria; Phillips et al., 2019 for earthworms; van den Hoogen et al., 2019 for nematodes; Oliverio et al., 2020 for protists), unveiling their environmental drivers. Yet, whether soil biodiversity at all its taxonomic and trophic levels responds to the same ecological drivers as aboveground diversity and follows similar trends remains to be tested. For such tests, the integration of spatial scales and the scale at which organisms are analysed together is pivotal (Thakur et al., 2020; White et al., 2020). Indeed, the way environmental parameters drive local diversity can depend on the spatial extent (e.g. Steiner & Leibold, 2004), or the taxonomic or trophic groups being studied (e.g. Boyero et al., 2011; Peters et al., 2016; Tedersoo et al., 2014).

Among the hypotheses formulated to explain the spatial variation of biodiversity, theory and support from empirical studies on plants and other aboveground organisms have led to four major ecological hypotheses: the "energy hypothesis", the "physiological tolerance hypothesis", the "habitat heterogeneity hypothesis" and the "resource heterogeneity hypothesis" (Figure 1). Yet, these hypotheses have been seldom tested in a single framework for soil organisms (Decaëns, 2010; Thakur et al., 2020), and even less at the scale of the whole soil biota. Observing diversity patterns of soil organisms in nature, that is, the relationship between various relevant predictors and soil diversity, is a first step to test whether these ecological hypotheses apply to the wide range of soil organisms (Shade et al., 2018).

The "*energy hypothesis*" predicts a positive relationship between diversity and energy. An increasing amount of energy (i.e. thermic, solar or chemical) promotes diversity across trophic levels by increasing speciation rates and/or the number of species populations, and thereby reducing local extinction (Evans et al., 2005; Wright, 1983). An extension of the hypothesis predicts a hump-shaped relationship with a decrease in diversity at high energy levels due to exclusive competition (Mittelbach et al., 2001). Plant productivity is traditionally used as a primary energy measure, because it accounts for water limitations in the transformation of solar energy into available resources, and because plants are the main basal resource (primary producers) for aboveground organisms (Currie et al., 2004; Evans et al., 2005). Yet, in the soil compartment, soil organic matter (SOM) is also a major source of energy fuelling the soil food web (Moore et al., 2004). The local amount and content of SOM is driven by multiple drivers such as plant community composition, climate or parent material (Wiesmeier et al., 2019), and not only by plant productivity. Considering both solar energy and SOM, hereafter referred as primary and secondary energy, respectively, is thus essential to test the energy-diversity relationship for the soil biota. Therefore, since most soil organisms are thought to be weakly limited by competition due to their limited mobility and the complexity of the soil matrix (Ettema & Wardle, 2002; Wardle, 2006), it could be expected that soil diversity increase monotonously with available energy.

The "*physiological tolerance hypothesis*" states that favourable environmental conditions support higher biodiversity because a wider range of strategies can persist under such conditions (i.e. tighter niche packing), while only a few well-adapted species can tolerate stressful conditions (Currie et al., 2004; Spasojevic & Suding, 2012).

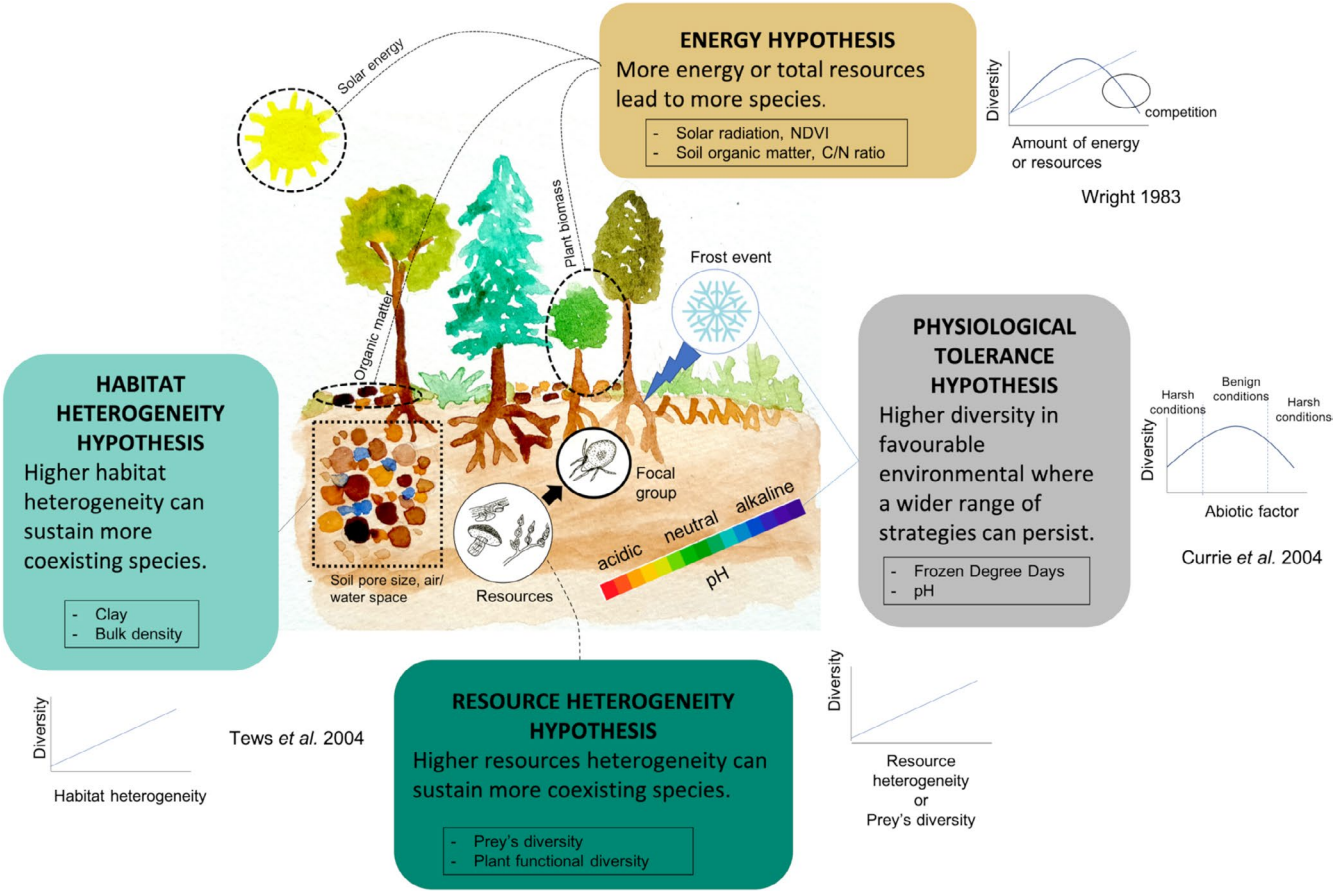


FIGURE 1 Overview of the four big ecological hypotheses and theoretical predictions tested in this study within the soil biodiversity context. Each hypothesis is introduced in a coloured box, the predictors used to represent each hypothesis are given at the end of the boxes in a frame

Temperature is one of the most acknowledged factors constraining the "thermal niche" of organisms. Yet, compared to aboveground temperature, soil temperatures are buffered making it more difficult to isolate its effect on soil biodiversity. For example, in mountain environments, soil temperature is strongly regulated by snow cover and duration (Carlson et al., 2015). In the absence of snow, soil frost might impact the structure and activity of soil communities (Schostag et al., 2019; Sulkava & Huhta, 2003). In addition, soil organisms often rely on other abiotic conditions such as water availability, heavy metal content and pH that can generate stressful conditions at extreme values, for example, drought, toxicity, acidity (Gans, 2005; Xu et al., 2012). Indeed, soil pH is recognized as a major driver of soil microorganisms diversity (Fierer & Jackson, 2006). While the stressful environmental factors may differ, the general response form to stress should be the same for above and belowground diversity.

The "*habitat heterogeneity hypothesis*" postulates that increasing habitat heterogeneity provides larger niche space or dimensionality that can be finely partitioned and sustain more coexisting species (Stein et al., 2014; Tews et al., 2004). Traditionally, the "*habitat heterogeneity hypothesis*" is tested at the landscape scale where biodiversity increases with habitat or vegetation diversity (Stein et al., 2014). However, soils can harbour a high degree of heterogeneity at much smaller grains than those considered aboveground (Young & Crawford, 2004), and this partly explains their remarkably high biodiversity (Ettema & Wardle, 2002; Nielsen et al., 2010). On a microscale, habitat heterogeneity can be structural, that is, associated with the size distribution of the pores, which is controlled by soil texture and compaction (i.e. bulk density). Pore size distribution varies within and between soil types, and can influence habitat conditions by modulating nutrient availability, gas diffusion and soil water holding capacity (Ranjard & Richaume, 2001; Six et al., 2004), parameters that may affect the diversity of soil organisms (Nielsen et al., 2010; Xia et al., 2020) and their interactions (Erktan et al., 2020). The effects of soil texture and compaction on the diversity might vary between soil organisms with different sizes or life-history strategies (Seaton et al., 2020) or whether there are ecosystem engineers able to modify the soil structural properties (Decaëns, 2010; Six et al., 2004).

The "*resource heterogeneity hypothesis*" follows the same rationale as the habitat hypothesis. An increase in resource heterogeneity can lead to an increase in diversity (Steiner, 2001; Heidrich et al., 2020; Dal Bello et al., 2021). We acknowledge that resource heterogeneity can be intrinsically linked to the habitat

heterogeneity, which makes it difficult to separate them. As for aboveground, soil basal resources can take different forms, but their heterogeneity can be well approximated by plant functional diversity since it explains variation in SOM composition, type of potential mycorrhiza, root exudates and direct resources for phytophages (Anderson, 1978; Eviner & Chapin, 2003; Hooper et al., 2000). For higher trophic level groups (secondary and tertiary consumers), the diversity in potential prey might be taken as a proxy for resource heterogeneity.

Here, we tested the above outlined macroecological biodiversity hypotheses and estimated their relative importance in explaining soil biodiversity patterns across most soil trophic groups. We built on a large-scale observatory network across the French Alps (Orchamp) that provides soil biodiversity measurements from environmental DNA metabarcoding across seventeen elevational gradients ranging from low to very high altitude (280–3160 m), and harbouring very contrasting climatic, vegetation and pedological conditions (Figure S1). Mountainous systems are well suited to test empirically large-scale drivers of biodiversity as they include wide ranges of environmental conditions and high biotic turnover over a reduced spatial scale (McCain & Grytnes, 2010). Instead of focusing on specific taxonomic orders, we followed a multi-trophic approach to test the above hypotheses on most trophic groups representative of soil ecosystems. After selecting the predictors related to the ecological hypotheses, we used a machine learning approach to account for complex interactions between predictors and soil biodiversity and corrected for remaining spatial dependencies that may originate from processes that have not been considered, such as missing abiotic factors or dispersal limitations. More specifically, we used biodiversity patterns to assess (1) the relative importance of predictors linked to different ecological hypotheses in explaining overall multi-trophic soil biodiversity and (2) the consistency of the response curves across trophic groups.

2 | MATERIAL AND METHODS

2.1 | Study site and sampling design

The data come from the French Alps long-term observatory, Orchamp (www.orchamp.osug.fr, Appendix S1), made of multiple elevational gradients distributed across the whole French Alps (ca. 40,500 km^2) and representative of the environmental conditions of the region. Each elevational gradient has a homogenous exposure and slope, and consists of four to nine 30×30 m plots separated by 200 m of altitude, on average. In this study, we used data gathered from 2016 to 2018, corresponding to 17 gradients (Figure S1), 90 plots and 540 soil samples. Plant species abundances were quantified at the vegetation peak (mostly in July or August) along a linear transect crossing each plot using the pin-point method (Jonasson, 1988). A second 4-m-wide transect was dedicated to soil sampling at the end of the summer season. Soil was sampled from 3 subplots (2×2 m) selected across the transect where we collected around ten soil cores of 5 cm in diameter that were separated into two soil layers, that is, surface (ca. 1–8 cm depth) and subsurface (ca. 8–16 cm depth), which could be differentiated in most cases by a change in the colour. The ten soil cores were pooled together and homogenized by separating the two layers to make a biological sample per soil layer per subplot, to obtain a total of six samples per plot.

2.2 | Soil sample processing

Each soil sample was separated into two components. The main part was sieved at 2mm and used to measure soil physicochemical properties (soil pH, SOM content and soil C/N) as described in (Martinez-Almoyna et al., 2020). The other part was used for environmental DNA, where DNA was extracted from a 15 g aliquot and processed in the field using the procedure described in Taberlet et al. (2012), Taberlet et al. (2018). We used six DNA markers to have a complete overview of the soil biota, including two universal markers (euka02 for eukaryote, bact01 for bacteria) and fourth clade-specific markers (fung02 for fungi, inse01 for insect, olig01 for oligochaete and coll02 for collembola). Information on the markers and molecular analyses including PCR, library preparation and sequencing steps are detailed in Appendix S2. A standardized bioinformatic pipeline was then applied (Calderón-Sanou et al., 2020), using the OBITools software (Boyer et al., 2016) and the R package "metabaR" (Zinger et al., 2021), to remove contaminants and errors and to get the taxonomic composition in terms of Molecular Operational Taxonomic Unit (MOTU) of each sample (Appendix S2).

2.3 | Diversity of trophic groups

The obtained MOTUs were classified into 36 trophic groups. We chose to distinguish not only trophic levels but also phylogenetic distant groups of the same trophic level, as they may have different preys/predators or exhibit different resource acquisition strategies (e.g. bacterivorous nematodes vs. protists, or predatory mites versus insects, Potapov et al., 2019), following Calderón-Sanou et al. (2021). The databases used for the trophic and functional assignments were FungalTraits (Põlme et al., 2020), for fungal MOTUs assigned at the genus level and FUNGuild (Nguyen et al., 2016) for the rest of fungi, FAPROTAX (Louca et al., 2016) and Wardeh et al. (2015) database of host–pathogen interactions for bacteria, NEMAguild and Nemaplex (http://nemaplex.ucdavis.edu/) for nematodes. The main references used included Adl et al. (2019) for protists, Rainford and Mayhew (2015) for insects and Potapov et al. (2016) for Collembola. The most abundant taxonomic clades composing each trophic group are shown in Table S1. The MOTU diversity of each trophic group was estimated per sample using the exponential of the Shannon entropy (i.e. Shannon diversity), which represents "the effective number of MOTUs" as it penalizes rare sequences that could be artefacts in eDNA data. Shannon diversity leads to more robust ecological results and to diversity estimates that are more similar to those

assessed from conventional sampling approaches (Calderón-Sanou et al., 2020).

2.4 | Environmental predictors

We used two environmental predictors to represent each ecological hypothesis (Figure 1), with the condition of having a final set of weakly correlated predictors (see Figure S2 for a visualization of the correlation between all initially considered parameters).

2.4.1 | Energy hypothesis

It was separated into primary (solar energy) and secondary energy (SOM), and two predictors were selected for each category. Solar radiation and the Normalized Difference Vegetation Index (NDVI) were used to represent the primary energy predictors. Solar radiation directly measures the amount of solar energy arriving into the Earth's surface, while NDVI estimates the amount of solar energy that is transformed by photoautotrophic organisms into available resources accounting for water limitations (Evans et al., 2005). We did not add mean annual temperature as sometimes done to represent energy (Clarke & Gaston, 2006) since it was strongly correlated to NDVI (Figure S2). Solar radiation was calculated per plot as the sum of the daily surface incident direct and diffuse shortwave radiation accumulated over 10 years, from 2008 to 2018. NDVI was estimated from the surface spectral reflectance at a resolution of 250 m from MODIS (Moderate Resolution Imaging Spectroradiometer), available online: https://lpdaac.usgs.gov/products/mod09q1v006/. Raw NDVI times series were pre-processed following Choler (2015), and we kept the mean yearly sum of NDVI greater than 0.2 over 2009–2019, as the final predictor for the analyses measured at the plot level. To represent secondary energy, we used the SOM content and the C/N ratio, measured from the soil samples. The former indicates the total amount of organic matter available in the soil, while the latter is a proxy for nutrient availability or SOM decomposability (Cleveland & Liptzin, 2007), meaning that soils with low C/N rates have potentially more readily available energy than soils with high C/N ratio, if we account for nutrient stoichiometric constraints.

2.4.2 | Physiological tolerance hypothesis

We used soil pH and the freezing degree days (FDD) to represent potential sources of abiotic physiological stress for soil organisms. The pH has been described as an important limiting physiological factor of soil communities (Fierer & Jackson, 2006; Räty & Huhta, 2003). The FDD summarizes the duration and intensity of ground freezing events and it has been addressed as a good candidate to model the thermal niches (Choler, 2018). FDD was calculated per plot as the annual sum of average daily degrees below zero, modelled within the first soil horizon (1 cm depth) and averaged over 2008–2018.

2.4.3 | Habitat heterogeneity hypothesis

Clay percentage in soil and bulk density were selected to represent the microscale habitat heterogeneity. Clay percentage characterizes the soil texture and thus reflects the granulometry distribution, the aeration, ability of soil to retain water and more globally the physical properties of the soil (Hao et al., 2007; Seaton et al., 2020). Soil texture might affect diversity differently across trophic groups with different sizes or life-history strategies (Seaton et al., 2020; Vreeken-Buijs et al., 1998). For example, the diversity of mesofauna could be expected to increase in coarse-textured soils (i.e. with low clay percentage), where the higher availability of larger pores provides more different habitats to be potentially colonized by these organisms (Vreeken-Buijs et al., 1998). Bulk density reflects soil compaction and porosity as it accounts for the amount of soil per volume unit when removing water and air spaces (Hao et al., 2007). Compact soils, with higher values of bulk density, have relatively lower total pore space and organic matter content, thus providing a lower heterogeneity of habitats. Both variables were measured from a soil pit carried out next to the plot). Three soil replicates were collected with a volumetric cylinder (100 cm3) from the superficial horizon. They were dried at 105°C for 24 h and sieved to 2 mm. The mass of dry soil (mS) contained in the cylinder as well as the mass of coarse elements greater than 2 mm (mEG) were measured. The formula applied for the calculation of bulk density is as follows (Equation 1), with V_{cyl} for the volume of the cylinder. The bulk density of the three replicates were averaged.

$$Da = \frac{mS\text{-}mEG}{Vcyl} \quad (1)$$

2.4.4 | Resource heterogeneity hypothesis

For decomposers, detritivores and plant symbionts, we used two metrics of plant functional diversity as predictors, that is, the functional richness and the functional divergence (Villéger et al., 2008), calculated for each plot using the R package "FD" (Laliberté & Legendre, 2010). Functional richness represents the total trait space filled by all the plant species present in the community (here the plot). Functional divergence describes how specie's abundances are distributed within the functional trait volume. To estimate these two metrics, we used our own trait measurement values for species (median values across individuals) present in our botanical surveys. We included the following traits: specific leaf area (SLA), leaf carbon and nitrogen ratio, root depth (extracted from Landolt et al., 2010), vegetative plant height and woodyness index. For the rest of the soil groups (except autotrophs), we selected two predictors measuring prey diversity (exponential of the Shannon entropy) of the focal trophic group. For omnivores (i.e. tardigrades, rotifers and protists), we used the MOTU's diversity of bacteria and the MOTU's diversity of fungi. For bacterivores, we used the MOTU's diversity and phylum's diversity of bacteria. For fungivores and zooparasites, we used the MOTU's diversity and class diversity of fungi and metazoans,

respectively. We used the diversity at these two taxonomic levels, because MOTUs diversity might be redundant depending on the level of generalism of the focal trophic group, that is, a predator might be indifferent to two closely related species. For predators, we used the MOTU's diversity and class diversity of a subgroup of metazoans (or protists for protistivores), in which we excluded the focal group and we only considered the category of size that could potentially be a prey for the focal group (e.g. only micro-metazoans for predatory nematodes).

Solar radiation and FDD were calculated from the SAFRAN-SURFEX/ISBA-Crocus-MEPRA reanalysis (Durand et al., 2009; Vannier & Braud, 2012), a model which addresses meteorological and snow conditions in mountainous regions based on large-scale topographical features.

2.5 | Spatial structure

Given the hierarchical sampling design of the data (two soil layers within plots within gradients), we accounted for the overall spatial structure of the samples to avoid having spatial autocorrelation issues (Dray et al., 2012). We defined a set of spatial predictors representing the residual spatial structure (i.e. the left-out spatial structure not explained by the environmental predictors) to include in the models. This approach aims to reduce the spatial autocorrelation that could remain in the residuals and to identify potential spatial structures with a strong influence on soil diversity. We did so using Moran's eigenvector maps (MEM), a method based on computing the principal coordinates of a matrix of geographic neighbours (Dray et al., 2006). The obtained eigenvectors are orthogonal and have a straightforward interpretation as each of them represents a spatial pattern at a given scale that can be ranked from broad spatial structures to fine spatial structures. We identified 18 MEM-variables describing significant spatial autocorrelation (only positive eigenvalues, Dray et al., 2006) based on the Euclidean geographic distances between each subplot using the function *dbmem* from the R package "adespatial" (Dray et al., 2021). MEM 1 to 8 described broad scale spatial structures, while MEM 9 to 18 represented intermediate to fine spatial structures (Figure S3). To remove the imprint of the environment on these MEMs, we modelled with a random forest each of the 18 MEMs as a function of our environmental predictors and extracted the residuals of these relationships. These residuals thus represented the spatial structure not explained by our environmental predictors (e.g. missing predictors, dispersal limitations). This approach differs from partialling out the spatial component of diversity and compare the pure effect of environment, the pure effect of space and the shared explained variance (Borcard et al., 1992). Here, we argue that space is likely affecting environment and that environment is then affecting biodiversity. The shared explained variance of space and environment is thus relevant for our hypotheses. We treat the pure effect of space as a statistical nuisance as we cannot link it to ecological processes, given that we jointly analyse taxa with very different dispersal abilities. We made sure that it was properly accounted for to avoid residual spatial autocorrelation (Dray et al., 2012).

2.6 | Random forest

To model the diversity of each trophic group as a function of the predictors representing our four hypotheses and the residual spatial structure, we used random forest models (Breiman, 2001), which are particularly suited when nonlinear relationships and complex interactions among predictors are expected. Random forest analyses were run with the R package "party" (Hothorn et al., 2006) with the *cforest_unbiased* function, which avoids bias introduced by heterogeneity in scale and number of categories among predictors (Strobl et al., 2007). The number of trees was set to 1,000 and the number of variables randomly sampled as candidates at each split (*mtry*) was tuned using the function *train* of the R package "caret" (Kuhn, 2020; Table S2). Variable importance was estimated as the mean decrease in accuracy using the function *varimp*. The method allows assessing relative variable importance, by identifying the covariates which, when removed, ensure a significant drop of prediction power (Strobl et al., 2007). It thus avoids any over-fitting and allows sound inference. Overall explained variance (r-square) was calculated by extracting the coefficient of determination between predictions and observations. The shape of the relationship between the diversity and the predictors was assessed with partial dependent plots obtained from the R package "iml" (Molnar et al., 2018), which estimate the marginal effect of a given predictor while accounting for the average effect of the other predictors in the model. We considered that a relationship was relevant, when the predictor had a predictive importance higher than 25%. The predictive importance was assessed by permuting each predictor one by one and then evaluating how the prediction was affected.

A single random forest model was run for each trophic group with the same set of predictors, that is, solar radiation, NDVI, SOM, C/N ratio, percentage of clay, bulk density, two variables corresponding to resource heterogeneity (variable across trophic groups, and excluded for autotrophs) and the 18 residual spatial structure predictors. All analyses were run in the R statistical environment (R Core Team, 2020).

3 | RESULTS

We identified 222,739 bacterial and 50,241 eukaryotic (including 5,467 metazoans and 11,115 protists, Figure A2-1 in Appendix S2) MOTUs from the universal markers, corresponding to 13,173,466 and 28,645,720 reads respectively. From the clade-specific markers, we recovered 48,127, 2,799, 3,113, 5,128 MOTUs and 29,022,014, 1,507,963, 5,558,110, 16,738,061 reads of fungi, insects, collembola and oligochaetes respectively (see Table A2-3 in Appendix S2 for the statistics per year). From the identified sequences 1,333,857 MOTUs corresponding to 50,770,784 reads were assigned to the trophic

groups. Table 1 presents the number of reads, families and MOTUs retrieved for each trophic group and the estimated Shannon diversity.

The predictors underlying the tested ecological hypotheses explained a significant part of the spatial variation of diversity of most trophic groups. The overall explained variance varied from 29% for detritivorous insects to 79% for arbuscular mycorrhizal fungi (Figure 2a, Table S2). The residual spatial structure explained much less variance than the environmental predictors, confirming the relevance of the latter to predict soil biodiversity. Only the diversity of predatory and phytophagous insects, and photoautotrophic protists was better explained by pure broad residual spatial structures than by the environment (MEM7, Figure S3).

We found that predictors associated with the energy and the physiological tolerance hypotheses were generally the most important, even so the relative importance of the predictors did vary between soil trophic groups in different trophic positions or from different body size categories (Figure 2b, Figure 3). The energy hypothesis was particularly important for consumers, that is, tertiary and secondary consumers and plant symbionts, and less important for autotrophs (Figure 3a). In particular, the secondary energy predictors related to SOM explained a large part of the diversity of most fungivores and detritivorous insects well-linked to the detritus channel. When looking at the tendencies per category of body size, the energy hypothesis was more important for metazoans of all sizes and fungi diversity, while the physiological tolerance hypothesis explained most variation for bacteria, protists and microfauna (Figure 3b). The habitat heterogeneity hypothesis had a higher importance for autotrophs compared to the other groups. The resource heterogeneity hypothesis was especially important for bacterivores (both protists and nematodes), phytophagous insects, enchytraeids and soil saprotrophic fungi.

In general, we found that the partial response curves of diversity to predictors were consistent across most soil trophic groups (Figure 4) and in agreement with predictions (Figure 1), with some few exceptions. The diversity of most trophic groups including zooparasites protists and fungi, metazoans consumers and ectomycorrhizal fungi, strongly increased with NDVI, but decreased for photolithoautotroph bacteria, phytophagous protists and earthworms (Figure 4a). The steepest changes in soil diversity across the NDVI gradient occurred in the transition from forest (high NDVI) to alpine grasslands (low NDVI). Groups for which diversity strongly increased with solar radiation included zooparasite bacteria, phytophagous protists and earthworms. All trophic groups primarily feeding on detritus positively increased in diversity with SOM (Figure 4b). The diversity of several groups was also influenced by the C/N ratio: diversity decreased for herbivorous and bacterivorous nematodes, and root endophyte and arbuscular mycorrhizal fungi, but increased for ectomycorrhizal fungi and fungivorous nematodes (Table S3). With the exception of rotifers and tardigrades, all trophic groups responding to pH increased in diversity in more alkaline soils (Figure 4c). This positive relationship had a sigmoid form for all groups, but both the inflection points and associated slopes strongly varied across trophic groups. Saprotrophic, root endophytes and phytoparasitic fungi, and also photolithoautotrophic bacteria were positively affected by the soil clay content, and chemolithoautotrophic bacteria were positively affected by soil bulk density (Table S3). All phytophagous insects, saprotrophic fungi and bacterivore groups responded positively to resource heterogeneity, that is, plant functional richness and bacteria diversity respectively (Figure 4d). Enchytraeids responded positively to plant functional divergence (aka. resource heterogeneity).

4 | DISCUSSION

Testing ecological hypotheses has largely contributed to our understanding on how biodiversity is structured on Earth. However, generality can only be claimed if a significant part of biodiversity is covered. Here, we add an important missing piece to the general picture by testing several major ecological hypotheses simultaneously on the majority of trophic groups inhabiting the soil and along sharp environmental gradients which allow some generalization to be made. Our results confirm that the main environmental drivers of soil biodiversity are variable across soil trophic groups and depend on their resource or physiological requirements. Yet, we also find major commonalities in the ecological processes structuring soil biodiversity. Overall, the energy and physiological tolerance hypotheses had the strongest support from soil multi-trophic biodiversity.

Our results are in agreement with previous studies finding that an increase in primary energy increases the diversity of soil organisms such as protists (Oliverio et al., 2020), metazoans (Peters et al., 2016), soil predators (Binkenstein et al., 2018) and fungi (Tedersoo et al., 2014, Figure 2b). Our results also reveal that secondary energy, related to soil organic matter, has a positive effect on soil biodiversity, especially for fungivorous and detritivorous animals, in agreement with earlier work (Binkenstein et al., 2018; Canedoli et al., 2020; Caruso et al., 2019). We found that the relative importance between primary and secondary energy varies across trophic groups, with no clear trends across trophic levels, suggesting that both energy channels are at play across the soil food web. However, some groups responded to specific energy predictors in a way that differs from the predictions of the "*energy hypothesis*" (Figure 1). For example, the diversity of earthworms, phytophagous fungi and photolithoautotroph bacteria decreased with increasing NDVI. Part of these divergent trends between diversity and NDVI might be explained by the transition from forest to grassland in the NDVI gradient in our study system, for example, alpine grassland soils are more suitable for autotrophic bacteria adapted to high elevation stressful conditions (Guo et al., 2015). Otherwise, a negative interaction between ectomycorrhizal fungi and phytophagous fungi could explain the decrease in diversity of the latter (Figure 4a). Indeed, ectomycorrhizal fungi can provide protection against pathogens to their plant hosts, thus reducing the incidence of phytophagous fungi and their diversity (Antunes & Koyama, 2017; Wang et al., 2019). Other divergent, but not unexpected, trends were found along the

TABLE 1 Information on the environmental DNA data characterizing each trophic group, including the DNA marker used to sample each group and the final number of reads, families (orders for protists), MOTUs and Shannon diversity obtained in total across the French alps and per sample (mean ± SD)

Trophic position	Trophic group	DNA marker	Total reads (per sample)	Total families/orders (per sample)	Total MOTUs (per sample)	Total Shannon diversity (per sample)
Autotrophs	B.chemolithoautotroph	bact01	269,418 (517.1 ± 606.8)	20 (9.5 ± 1.9)	3,148 (80.1 ± 47.3)	108.4 (23.1 ± 9.8)
	B.photolithoautotroph	bact01	51,400 (98.8 ± 123.8)	20 (3.9 ± 1.8)	715 (22.3 ± 14.8)	43.7 (10.3 ± 5.6)
	P.photoautotroph	euka02	34,083 (65 ± 91.3)	12 (2.4 ± 1)	374 (13.2 ± 7.3)	55.5 (8.7 ± 4.1)
Decomposers	B.heterotroph	bact01	10,082,382 (19,352 ± 12,905.5)	432 (161 ± 23.9)	94,308 (2,980.1 ± 1,148.9)	2,352.1 (611.9 ± 205.1)
	F.saprotroph (litter)	fung02	1,588,239 (3,013.7 ± 4,770.2)	62 (12.2 ± 4.6)	1,305 (37.4 ± 17.7)	200.2 (9.8 ± 5.6)
	F.saprotroph (other)	fung02	437,395 (830 ± 1,602.1)	58 (9.6 ± 4.3)	969 (25.7 ± 16.8)	185.7 (10.4 ± 6.2)
	F.saprotroph (soil)	fung02	7,438,380 (14,114.6 ± 15,124.5)	44 (14 ± 4.2)	3,587 (75.5 ± 37.1)	435 (14 ± 8)
	F.saprotroph (wood)	fung02	691,399 (1,319.5 ± 2,263.9)	102 (11.4 ± 5.2)	1,163 (20.1 ± 11.9)	207 (6.5 ± 4.3)
Detritivores	Earthworms	olig01	5,274,333 (13,386.6 ± 24,467.6)	6 (1.6 ± 0.6)	1,151 (10.8 ± 8.6)	70.4 (2.5 ± 2.6)
	Enchytraeidae	olig01	8,802,809 (21,418 ± 23,679.7)	1 (1 ± 0)	2,386 (24.3 ± 19.5)	60.8 (3 ± 2.3)
	I.detritivore	inse01	290,140 (1,458 ± 4,538.5)	29 (1.9 ± 1)	240 (3.1 ± 2.3)	57.6 (1.6 ± 0.8)
Plant mutualists	F.arbuscular_mycorrhizal	fung02	3,463,460 (7,141.2 ± 16,896.2)	10 (2.9 ± 1.7)	2,426 (36.6 ± 36.8)	254.1 (13.4 ± 12)
	F.ectomycorrhizal	fung02	5,304,179 (11,119.9 ± 13,724.7)	38 (7 ± 4.8)	2,795 (27.5 ± 28.5)	336.6 (6.1 ± 5.4)
	F.root_endophyte	fung02	127,187 (247.9 ± 650.7)	6 (2.9 ± 1.2)	174 (5.6 ± 3.4)	39.8 (3.2 ± 1.8)
Phytophagous	B.phytoparasite	bact01	526,809 (1,011.1 ± 1,029)	38 (20 ± 3.5)	5,539 (159.3 ± 89.3)	140.3 (35.5 ± 16.6)
	F.phytoparasite	fung02	213,181 (405.3 ± 522)	73 (11.3 ± 5.4)	806 (21 ± 12.6)	168.4 (9.5 ± 5.4)
	I.phytophagous	inse01	210,868 (1,270.3 ± 3,706.3)	10 (1.5 ± 0.8)	172 (2.2 ± 1.7)	55 (1.4 ± 0.7)
	N.herbivore	euka02	11,518 (22.6 ± 26.2)	10 (2.6 ± 1.3)	135 (5 ± 3.1)	41.3 (3.6 ± 2)
	P.phytoparasite	euka02	16,035 (31.9 ± 68)	18 (3 ± 1.6)	167 (5.1 ± 3.6)	37.1 (3.6 ± 2.2)
Bacterivores	N.bacterivore	euka02	67,829 (130.4 ± 174.1)	15 (6.5 ± 2.1)	176 (10.2 ± 4.3)	17.8 (5.1 ± 2.1)
	P.bacterivore	euka02	232,544 (446.3 ± 485.7)	22 (10.6 ± 2.1)	2,764 (108 ± 53.3)	370.5 (62.9 ± 23.6)
Fungivores	C.epigeic	coll01	1,956,200 (4,347.1 ± 6,678.4)	5 (2.7 ± 1.1)	704 (7.9 ± 7.4)	66.7 (2.7 ± 1.4)
	C.euedaphic_hemiedaphic	coll01	1,723,259 (4,351.7 ± 7,281.1)	3 (1.8 ± 0.5)	636 (8.7 ± 11)	53.1 (2.8 ± 1.9)
	M.fungivore	euka02	233,881 (449.8 ± 484.9)	22 (4.1 ± 2)	170 (10.7 ± 5)	22.8 (5.4 ± 2.4)
	N.fungivore	euka02	4,594 (10.3 ± 26.6)	7 (1.4 ± 0.7)	73 (2.4 ± 1.4)	21.7 (2 ± 1.1)
	P.fungivore	euka02	16,436 (31.7 ± 38.3)	2 (1.6 ± 0.5)	237 (6.6 ± 4.6)	27.1 (4.3 ± 2.6)
Omnivores	Rotifera	euka02	4,327 (9.5 ± 11.6)	N/A	39 (2.4 ± 1.5)	5.8 (2 ± 1)
	Tardigrada	euka02	19,822 (42.3 ± 67.1)	N/A	52 (3.1 ± 1.9)	5.8 (2.1 ± 1)
	P.omnivore	euka02	308,176 (581.5 ± 537.3)	7 (4 ± 0.5)	1,677 (66.2 ± 25)	97.7 (26.9 ± 11)

TABLE 1 (Continued)

Trophic position	Trophic group	DNA marker	Total reads (per sample)	Total families/orders (per sample)	Total MOTUs (per sample)	Total Shannon diversity (per sample)
Predators	I.predator	inse01	207,077 (1,327.4 ± 4,557.6)	20 (1.7 ± 0.8)	142 (2 ± 1.3)	56.5 (1.3 ± 0.5)
	M.predator	euka02	492,098 (928.5 ± 1,055.5)	32 (6.1 ± 2.6)	336 (13.8 ± 6.8)	9.4 (3.6 ± 2.2)
	N.predator	euka02	3,344 (8.8 ± 14.8)	7 (1.6 ± 0.8)	19 (1.8 ± 1)	7.9 (1.6 ± 0.8)
	P.protistivore	euka02	36,815 (69.5 ± 75.8)	4 (1.6 ± 0.7)	334 (15.8 ± 8.4)	61.1 (11.2 ± 5.5)
Zooparasites	B.zooparasite	bact01	536,202 (1,029.2 ± 1,234.2)	78 (18.7 ± 4.9)	4,329 (142.7 ± 80.9)	112.1 (34.2 ± 19.7)
	F.zooparasite	fung02	74,803 (147.5 ± 256.1)	12 (3.6 ± 1.5)	238 (8.7 ± 5.2)	45.5 (4.9 ± 2.6)
	P.zooparasite	euka02	20,162 (38.8 ± 60.2)	10 (3.1 ± 1.2)	371 (8.4 ± 5.4)	93.5 (5.7 ± 3.5)

Abbreviation: N/A: not applicable.

C/N ratio gradient, that is, ectomycorrhizal fungi and fungivorous nematodes were more diverse in soils with more recalcitrant organic matter (i.e. higher C/N ratio). This result reflects the differences in the energetic requirements or life-history traits of the different groups that may complexify generalizations of energy-related mechanisms. Contrary to other decomposers, ectomycorrhizal fungi can degrade recalcitrant organic complexes by using energy from their hosts (Lindahl & Tunlid, 2015). An increase in ectomycorrhizal fungi diversity could presumably cascade on fungivore nematodes diversity. Furthermore, while we show that energy has mainly a positive influence on soil biodiversity, the underlying mechanisms remain to be tested. For example, the more individual hypothesis states that greater energy availability allows a community to contain a larger number of individuals, and hence of a larger number of species with viable population size (Wright, 1983). Quantifying species abundance or biomass would be needed to test this hypothesis, but this information is unfortunately not yet available with eDNA metabarcoding data (Taberlet et al., 2018), and would be extremely challenging to obtain for the wide range of organisms studied here.

Physiological tolerances, mainly to soil pH, were also a strong predictor of the diversity of soil organisms, especially for organisms living in the aqueous phase of the soils. Indeed, in the study system, the diversity of groups of bacteria, protists and microfauna was more constrained by pH-induced stress rather than limited by energy or habitat and resource heterogeneity (Figure 3b), in accordance with previous studies highlighting the importance of pH for soil microbes (Fierer & Jackson, 2006; Karimi et al., 2018) and invertebrates (Bastida et al., 2020; Räty & Huhta, 2003). The sigmoid trend observed between diversity and pH might correspond to the first half of the humpback curve expected from the theory (Figure 1). Indeed, our sampling had relatively few sites with alkaline soils, and did not include soils with pH >8, levels from which other studies have observed a decrease of diversity (e.g. Fierer & Jackson, 2006). Our results revealed consistent decreases of diversity in more acidic soils, but also different tolerance thresholds across soil trophic groups. The strong effect of soil pH might also be the sum of multiple linked factors not considered in this study including bedrock type and plant communities (Roy et al., 2013). Contrarily, FDD had a minor effect on soil biodiversity. Limited effect of freezing events on soil biodiversity has previously been reported, and may result from the frost resistance (Männistö et al., 2018; Stres et al., 2010) or the rapid recovery of soil communities (Sulkava & Huhta, 2003). Theoretically, this low importance could be due to a scale mismatch between the measured soil communities (subplots are $4m^2$ large) and the climatic data resolution (~300m). However, between the available in situ temperature HOBOs and the climatic data used here showed very consistent patterns, rendering the scale mismatch hypothesis unprobeable. Otherwise, a change in composition or activity, without changes in local diversity, might also have occurred and remains to be tested (Schostag et al., 2019; Stres et al., 2010).

The "*habitat heterogeneity*" and the "*resource heterogeneity*" hypotheses weakly explained the spatial variation in diversity of soil trophic groups compared to "energy" and "physiological tolerance",

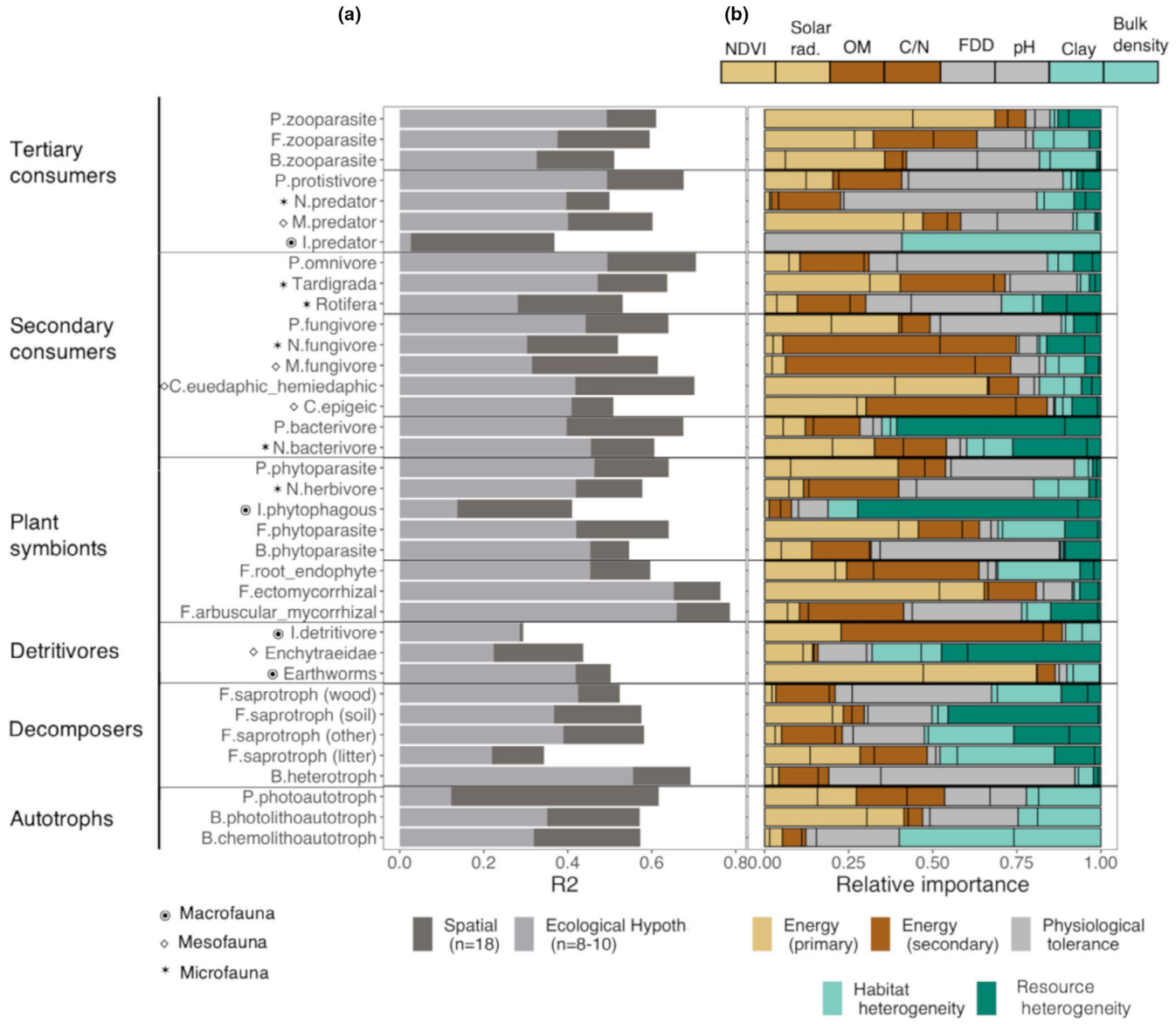


FIGURE 2 Relative importance of competing hypotheses in explaining the alpha diversity of soil trophic groups. (a) Total r-squared of the random forest model for each trophic group. Colours represent the relative importance of the environmental versus the spatial predictors. Environmental predictors correspond to all the biotic and abiotic variables used to test the ecological hypotheses, and spatial predictors correspond to the residuals of the spatial structure when removing the effect of the environment. (b) Relative importance of the environmental predictors used to test the ecological hypotheses (colour key). The relative variable importance is the mean decrease in squared error, rescaled to sum the total r2 (a) or 1 (b). Letters correspond to broad taxonomic groups: Bacteria (B.), Protozoa (P.), Metazoa (C.: Collembola, I.: Insects, M.: Mites, N: Nematodes) and Fungi (F.). Symbols indicate the size category for fauna groups

with notable exceptions. Saprotrophic, root endophytes and phytoparasitic fungi, as well as autotrophic bacteria were highly affected by habitat heterogeneity. We found that these groups tended to be more diverse in fine-textured soils (higher clay percentage), which usually exhibit greater water retention capacity but also more recalcitrant and stable organic matter (Ranjard & Richaume, 2001; Six et al., 2004). Previous studies have shown that soil texture can influence bacterial and fungal diversity, with subgroups of taxa responding differently to the proportion of soil particles (i.e. clay, sand, silt) (Karimi et al., 2018; Seaton et al., 2020; Xia et al., 2020). Our results showed that such differences are also visible when considering different trophic groups of fungi and bacteria. The importance of "*habitat heterogeneity*" could be expected to vary across soil trophic groups, as the spatial scale at which heterogeneity is perceived by organisms of different sizes or different lifestyles can be highly variable (Heidrich et al., 2020). Here again, perhaps the scale at which we measured heterogeneity was not relevant for some specific groups. When looking at the effect of resource heterogeneity, prey's diversity was remarkably important for bacterivores. Strong associations between bacterivore protists

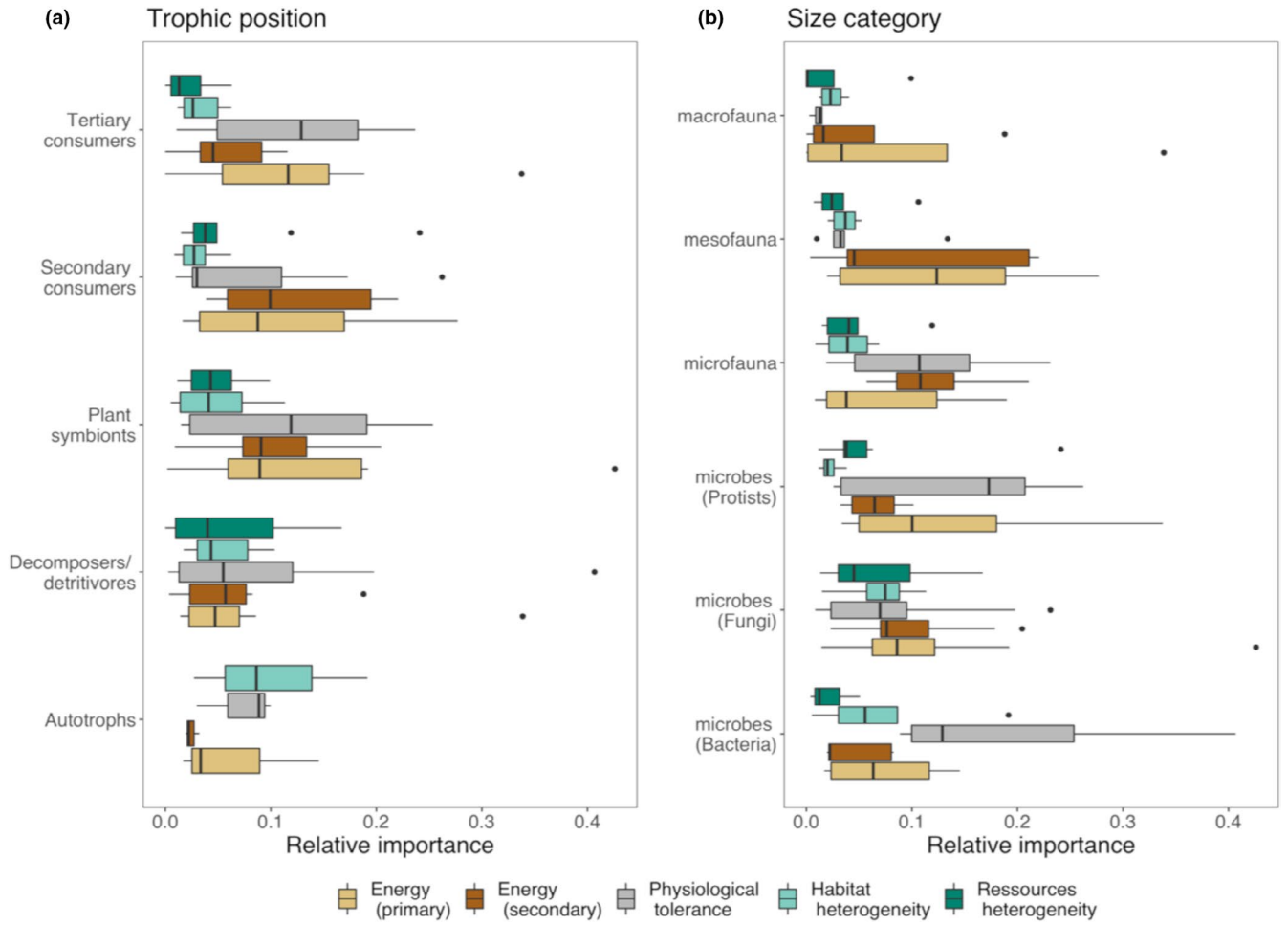


FIGURE 3 Boxplots of the relative importance of ecological hypotheses by trophic position and body size category. Relative importance of the four ecological hypotheses tested in this study across groups categorized by trophic position (a) or body size category (b). The values of relative importance correspond to the mean decrease in squared error from the random forest per trophic group, rescaled to sum the total r-square

and bacteria diversity have been recently reported (Oliverio et al., 2020; Xiong et al., 2021), and could indicate a degree of trophic specialization in bacterivorous protists. Co-variation in diversity might also indicate shared habitat preferences between protists (or nematodes) and bacteria, but our results and previous studies point to noticeable differences in the factors shaping the diversity of these groups (Oliverio et al., 2020; Xiong et al., 2021). Moreover, the strong response of saprotrophic fungi to plant functional diversity could be explained by a trophic specialization, in accordance with a recent study showing a high degree of specialization to specific soil and litter compounds for some saprotrophic fungi (Algora Gallardo et al., 2021). The significant association does not necessarily imply the realization of a trophic interaction, but it is a first step in assessing whether such interactions exist, leave signals in diversity distribution and can give us insights into the degree of food speciation in the focus trophic group.

To conclude, our near-complete coverage of soil biodiversity across trophic groups and across large and steep environmental gradients provides consistent and novel insights on the macroecological rules shaping the distribution of belowground biodiversity. Building on the efficiency of environmental DNA analyses combined with the wealth of existing knowledge on soil organisms, we showed that energy and physiological tolerance are the most plausible hypotheses to explain the spatial distribution of soil diversity at a regional scale. Interestingly, we found strong commonalities between trophic groups in their response to environmental drivers that should be later compared to aboveground organisms living in the same locations (e.g. ground-dwelling arthropods, pollinators or herbivores). Should belowground and aboveground compartments respond differently to environmental drivers, it will complexify their management under human-induced pressures. Finally, identifying how these patterns in local diversity translate into compositional changes and interaction network structuration in space will be of crucial importance to understand soil biodiversity assembly and how it might be affected by ongoing environmental changes.

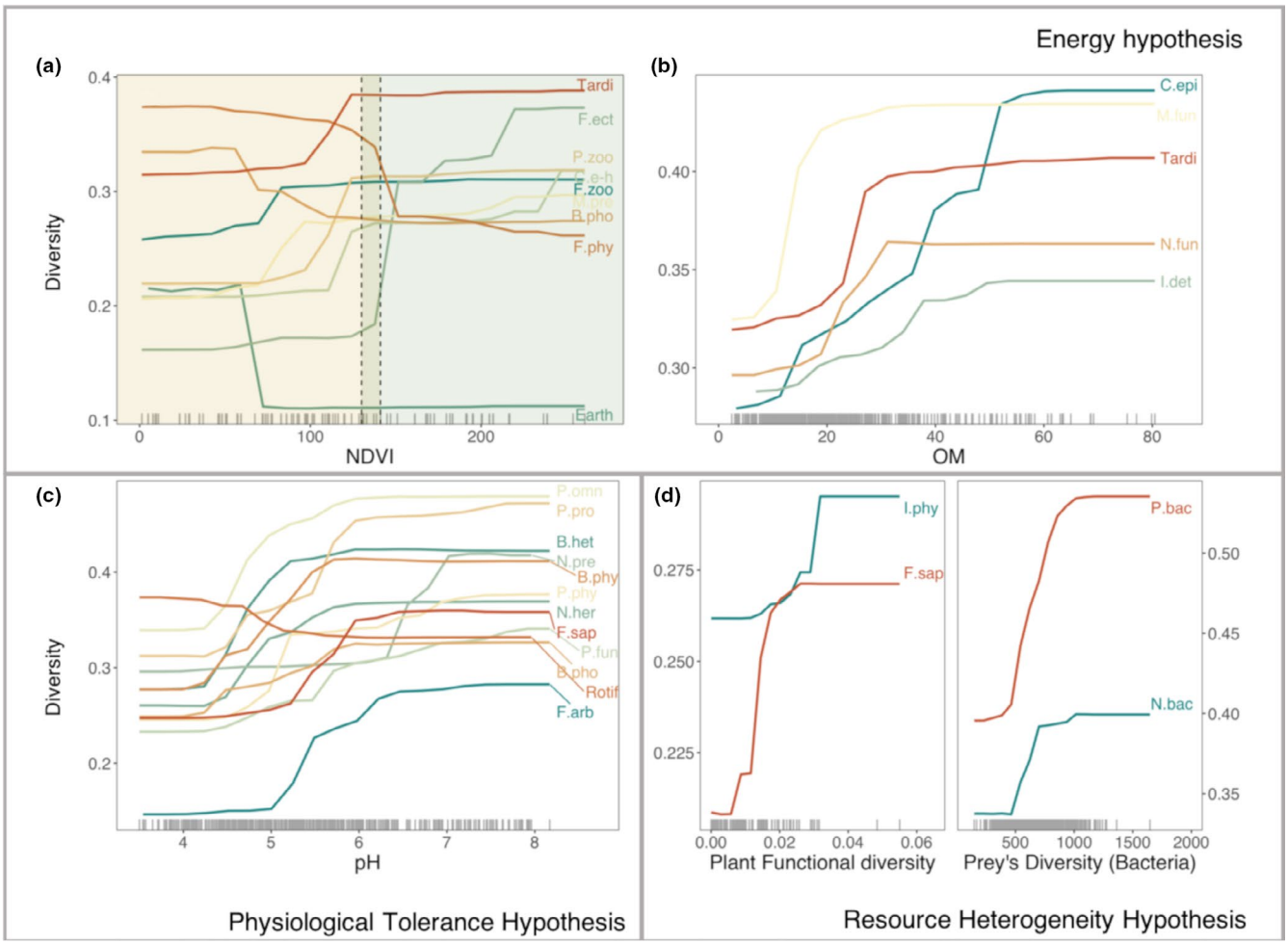


FIGURE 4 Predicted diversity of soil trophic groups as a function of the environmental predictors representing the ecological hypotheses. Partial dependence plots showing the marginal effect of the predictors representing the ecological hypotheses on the diversity of soil trophic groups based on the random forest model results. The predictors represented are (a) NDVI, with the colours corresponding to the transition from alpine grassland (yellow) to forest (green), (b) Soil organic matter, (c) pH and (d) plant functional richness and Bacteria MOTUs diversity. Diversity was standardized by the maximum diversity observed per group to have a comparable scale across groups. Only groups for which the predictors had a predictive importance higher than 25% were represented. The predictive importance of the predictor was assessed by permuting each predictor one by one and then evaluating how the prediction was affected. Taxonomic groups are abbreviated as Bacteria (B.), Collembola (C.), Earthworms (Earth), Fungi (F.), Insects (I.), Mites (M.), Nematodes (N.) Protozoa (P.), Rotifera (Rotif) and Tardigrada (Tardi). The rest of the trophic component is abbreviated as arbuscular mycorrhizal (arb), bacterivore (bac), detritivore (det), ectomycorrhizal (ect), epigeic (epi), euedaphic-hemiedaphic (e-h), fungivore (fun), herbivore (her), heterotroph (het), omnivore (omn), photolitoautotroph (pho), phytoparasite or phytophageous (phy), predator (pre), protistivore (pro), saprotroph (sap) and zooparasite (zoo)

ACKNOWLEDGEMENTS

This project was possible thanks to the many people that participated in the field work and in the laboratory procedures, including students, interns, postdoctoral fellows, field assistants, park managers and rangers. We thank Cindy Arnoldi for conducting the soil physicochemical analyses and, Nicolas Leguillarme and Giovanni Poggiato for support with the development of bioinformatic pipelines. We thank Sara Si-Moussi and Gemma Rutten for fruitful and constructive discussions. This research was funded by the French Agence Nationale de la Recherche (ANR) through the GlobNet project (ANR-16-CE02-0009), from two "Investissement d'Avenir" grants (Trajectories: ANR-15-IDEX-02; Montane: OSUG@2020: ANR-10-LAB-56) and the METRO Grenoble Alpes and the Isere department. Support was also received from the Lautaret Garden-UMS 3370 (Univ. Grenoble Alpes, CNRS, SAJF, 38000 Grenoble, France), member of AnaEE-France (ANR-11-INBS-0001AnaEE-Services, Investissements d'Avenir frame) and of the LTSER Zone Atelier Alpes, a member of the eLTER-Europe network.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting the results of this study are publicly available in the DRYAD repository https://doi.org/10.5061/dryad.q573n5tm6.

ORCID

Irene Calderón-Sanou https://orcid.org/0000-0003-4608-1187
Wilfried Thuiller https://orcid.org/0000-0002-5388-5274

REFERENCES

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., … Zhang, Q. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology*, *66*, 4–119. https://doi.org/10.1111/jeu.12691

Algora Gallardo, C., Baldrian, P., & López-Mondéjar, R. (2021). Litter-inhabiting fungi show high level of specialization towards biopolymers composing plant and fungal biomass. *Biology and Fertility of Soils*, *57*, 77–88. https://doi.org/10.1007/s00374-020-01507-3

Anderson, J. M. (1978). Inter- and intra-habitat relationships between woodland cryptostigmata species diversity and the diversity of soil and litter microhabitats. *Oecologia*, *32*, 341–348. https://doi.org/10.1007/BF00345112

Antunes, P. M., & Koyama, A. (2017). Chapter 9 - Mycorrhizas as nutrient and energy pumps of soil food webs: Multitrophic interactions and feedbacks. In N. C. Johnson, C. Gehring, & J. Jansa (Eds.) *Mycorrhizal mediation of soil fertility, structure, and carbon storage* (pp. 149–173). Elsevier.

Bardgett, R. D., Yeates, G. W., & Anderson, J. M. (2005). Patterns and determinants of soil biological diversity. In D. Hopkins, M. Usher, & R. Bardgett (Eds.), *Biological diversity and function in soils*, *ecological reviews* (pp. 100–118). Cambridge University Press.

Bastida, F., Eldridge, D. J., Abades, S., Alfaro, F. D., Gallardo, A., García-Velázquez, L., García, C., Hart, S. C., Pérez, C. A., Santos, F., Trivedi, P., Williams, M. A., & Delgado-Baquerizo, M. (2020). Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Molecular Ecology*, *29*, 752–761. https://doi.org/10.1111/mec.15299

Belmaker, J., & Jetz, W. (2015). Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecology Letters*, *18*, 563–571. https://doi.org/10.1111/ele.12438

Binkenstein, J., Klein, A.-M., Assmann, T., Buscot, F., Erfmeier, A., Ma, K., Pietsch, K. A., Schmidt, K., Scholten, T., Wubet, T., Bruelheide, H., Schuldt, A., & Staab, M. (2018). Multi-trophic guilds respond differently to changing elevation in a subtropical forest. *Ecography*, *41*, 1013–1023. https://doi.org/10.1111/ecog.03086

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, *73*, 1045–1055. https://doi.org/10.2307/1940179

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*, 176–182.

Boyero, L., Pearson, R. G., Dudgeon, D., Graça, M. A. S., Gessner, M. O., Albariño, R. J., Ferreira, V., Yule, C. M., Boulton, A. J., Arunachalam, M., Callisto, M., Chauvet, E., Ramírez, A., Chará, J., Moretti, M. S., Gonçalves, J. F., Helson, J. E., Chará-Serna, A. M., Encalada, A. C., … Pringle, C. M. (2011). Global distribution of a key trophic guild contrasts with common latitudinal diversity patterns. *Ecology*, *92*, 1839–1848. https://doi.org/10.1890/10-2244.1

Braga, J., Pollock, L. J., Barros, C., Galiana, N., Montoya, J. M., Gravel, D., Maiorano, L., Montemaggiori, A., Ficetola, G. F., Dray, S., & Thuiller, W. (2019). Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe. *Global Ecology and Biogeography*, *28*, 1636–1648. https://doi.org/10.1111/geb.12981

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

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*, 193–206. https://doi.org/10.1111/jbi.13681

Calderón-Sanou, I., Münkemüller, T., Zinger, L., Schimann, H., Yoccoz, N. G., Gielly, L., Foulquier, A., Hedde, M., Ohlmann, M., Roy, M., Si-Moussi, S., & Thuiller, W. (2021). Cascading effects of moth outbreaks on subarctic soil food webs. *Scientific Reports*, *11*, 15054. https://doi.org/10.1038/s41598-021-94227-z

Canedoli, C., Ferrè, C., Abu El Khair, D., Comolli, R., Liga, C., Mazzucchelli, F., Proietto, A., Rota, N., Colombo, G., Bassano, B., Viterbi, R., & Padoa-Schioppa, E. (2020). Evaluation of ecosystem services in a protected mountain area: Soil organic carbon stock and biodiversity in alpine forests and grasslands. *Ecosyst Serv*, *44*, 101135. https://doi.org/10.1016/j.ecoser.2020.101135

Carlson, B. Z., Choler, P., Renaud, J., Dedieu, J.-P., & Thuiller, W. (2015). Modelling snow cover duration improves predictions of functional and taxonomic diversity for alpine plant communities. *Annals of Botany*, *116*, 1023–1034. https://doi.org/10.1093/aob/mcv041

Caruso, T., Schaefer, I., Monson, F., & Keith, A. M. (2019). Oribatid mites show how climate and latitudinal gradients in organic matter can drive large-scale biodiversity patterns of soil communities. *Journal of Biogeography*, *46*, 611–620. https://doi.org/10.1111/jbi.13501

Choler, P. (2015). Growth response of temperate mountain grasslands to inter-annual variations in snow cover duration. *Biogeosciences*, *12*, 3885–3897. https://doi.org/10.5194/bg-12-3885-2015

Choler, P. (2018). Winter soil temperature dependence of alpine plant distribution: Implications for anticipating vegetation changes under a warming climate. *Perspectives in Plant Ecology, Evolution and Systematics*, *30*, 6–15. https://doi.org/10.1016/j.ppees.2017.11.002

Clarke, A., & Gaston, K. J. (2006). Climate, energy and diversity. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 2257–2266. https://doi.org/10.1098/rspb.2006.3545

Cleveland, C. C., & Liptzin, D. (2007). C:N: P stoichiometry in soil: Is there a "Redfield ratio" for the microbial biomass? *Biogeochemistry*, *85*, 235–252. https://doi.org/10.1007/s10533-007-9132-0

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guegan, J.-F., Hawkins, B. A., Kaufman, D. M., Kerr, J. T., Oberdorff, T., O'Brien, E., & Turner, J. R. G. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, *7*, 1121–1134. https://doi.org/10.1111/j.1461-0248.2004.00671.x

Dal Bello, M., Lee, H., Goyal, A., & Gore, J. (2021). Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nature Ecology & Evolution*, 1–11.

Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and Biogeography*, *19*, 287–302. https://doi.org/10.1111/j.1466-8238.2009.00517.x

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, *359*, 320–325. https://doi.org/10.1126/science.aap9516

Delgado-Baquerizo, M., Reich, P. B., Trivedi, C., Eldridge, D. J., Abades, S., Alfaro, F. D., Bastida, F., Berhe, A. A., Cutler, N. A., Gallardo, A., García-Velázquez, L., Hart, S. C., Hayes, P. E., He, J.-Z., Hseu, Z.-Y., Hu, H.-W., Kirchmair, M., Neuhauser, S., Pérez, C. A., … Singh, B. K. (2020). Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution*, *4*, 210–220. https://doi.org/10.1038/s41559-019-1084-y

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2021). *adespatial: Multivariate Multiscale Spatial Analysis*.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*, 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P. R., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., & Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, *82*, 257–275. https://doi.org/10.1890/11-1183.1

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., & Mérindol, L. (2009). Reanalysis of 44 Yr of climate in the french alps (1958–2002): Methodology, model validation, climatology, and trends for air temperature and precipitation. *Journal of Applied Meteorology and Climatology*, *48*, 429–449. https://doi.org/10.1175/2008JAMC1808.1

Erktan, A., Or, D., & Scheu, S. (2020). The physical structure of soil: Determinant and consequence of trophic interactions. *Soil Biology & Biochemistry*, *148*, 107876. https://doi.org/10.1016/j.soilbio.2020.107876

Ettema, C. H., & Wardle, D. A. (2002). Spatial soil ecology. *Trends in Ecology & Evolution*, *17*, 177–183. https://doi.org/10.1016/S0169-5347(02)02496-5

Evans, K. L., Warren, P. H., & Gaston, K. J. (2005). Species–energy relationships at the macroecological scale: A review of the mechanisms. *Biological Reviews*, *80*, 1–25. https://doi.org/10.1017/S1464793104006517

Eviner, V. T., & Chapin, F. S. III (2003). Functional matrix: A conceptual framework for predicting multiple plant effects on ecosystem processes. *Annual Review of Ecology Evolution and Systematics*, *34*, 455–485. https://doi.org/10.1146/annurev.ecolsys.34.011802.132342

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *PNAS*, *103*, 626–631. https://doi.org/10.1073/pnas.0507535103

Gans, J., Wolinsky, M., & Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, *309*, 1387–1390. https://doi.org/10.1126/science.1112665

Guerra, C. A., Heintz-Buschart, A., Sikorski, J., Chatzinotas, A., Guerrero-Ramírez, N., Cesarz, S., Beaumelle, L., Rillig, M. C., Maestre, F. T., Delgado-Baquerizo, M., Buscot, F., Overmann, J., Patoine, G., Phillips, H. R. P., Winter, M., Wubet, T., Küsel, K., Bardgett, R. D., Cameron, E. K., … Eisenhauer, N. (2020). Blind spots in global soil biodiversity and ecosystem function research. *Nature Communications*, *11*, 3870. https://doi.org/10.1038/s41467-020-17688-2

Guo, G., Kong, W., Liu, J., Zhao, J., Du, H., Zhang, X., & Xia, P. (2015). Diversity and distribution of autotrophic microbial community along environmental gradients in grassland soils on the Tibetan Plateau. *Applied Microbiology and Biotechnology*, *99*, 8765–8776. https://doi.org/10.1007/s00253-015-6723-x

Hao, X., Ball, B. C., Culley, J. L. B., Carter, M. R., & Parkin, G. W. (2007). Soil density and porosity. In M. R. Carter, & E. G. Gregorich (Eds.), *Soil sampling and methods of analysis* (pp. 743–759). CRC Press.

Heidrich, L., Bae, S., Levick, S., Seibold, S., Weisser, W., Krzystek, P., Magdon, P., Nauss, T., Schall, P., Serebryanyk, A., Wöllauer, S., Ammer, C., Bässler, C., Doerfler, I., Fischer, M., Gossner, M. M., Heurich, M., Hothorn, T., Jung, K., … Müller, J. (2020). Heterogeneity–diversity relationships differ between and within trophic levels in temperate forests. *Nature Ecology & Evolution*, *4*, 1204–1212. https://doi.org/10.1038/s41559-020-1245-z

Hooper, D. U., Bignell, D. E., Brown, V. K., Brussard, L., Mark dangerfield, J., Wall, D. H., Wardle, D. A., Coleman, D. C., Giller, K. E., Lavelle, P., Van der putten, W. H., De ruiter, P. C., Rusek, J., Silver, W. L., Tiedje, J. M., & Wolters, V. (2000). Interactions between aboveground and belowground biodiversity in terrestrial ecosystems: Patterns, mechanisms, and feedbacks. *BioScience*, *50*, 1049–1061.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*, 651–674. https://doi.org/10.1198/106186006X133933

Jonasson, S. (1988). Evaluation of the point intercept method for the estimation of plant biomass. *Oikos*, *52*, 101–106. https://doi.org/10.2307/3565988

Karimi, B., Terrat, S., Dequiedt, S., Saby, N. P. A., Horrigue, W., Lelièvre, M., Nowak, V., Jolivet, C., Arrouays, D., Wincker, P., Cruaud, C., Bispo, A., Maron, P.-A., Bouré, N. C. P., & Ranjard, L. (2018). Biogeography of soil bacteria and archaea across France. *Science Advances*, *4*, eaat1808. https://doi.org/10.1126/sciadv.aat1808

Kuhn, M. (2020). *caret: Classification and Regression Training*.

Laliberté, E., & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, *91*, 299–305. https://doi.org/10.1890/08-2244.1

Landolt, E., Bäumler, B., Ehrhardt, A., Hegg, O., Klötzli, F., Lämmler, W. et al (2010). *Flora indicativa: Okologische Zeigerwerte und biologische Kennzeichen zur Flora der Schweiz und der Alpen*. Haupt.

Lindahl, B. D., & Tunlid, A. (2015). Ectomycorrhizal fungi – potential organic matter decomposers, yet not saprotrophs. *New Phytologist*, *205*, 1443–1447. https://doi.org/10.1111/nph.13201

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*, 1272–1277. https://doi.org/10.1126/science.aaf4507

Männistö, M., Vuosku, J., Stark, S., Saravesi, K., Suokas, M., Markkola, A., Martz, F., & Rautio, P. (2018). Bacterial and fungal communities in boreal forest soil are insensitive to changes in snow cover conditions. *FEMS Microbiology Ecology*, *94*. https://doi.org/10.1093/femsec/fiy123

Martinez-Almoyna, C., Piton, G., Abdulhak, S., Boulangeat, L., Choler, P., Delahaye, T., Dentant, C., Foulquier, A., Poulenard, J., Noble, V., Renaud, J., Rome, M., Saillard, A., Thuiller, W., & Münkemüller, T. (2020). Climate, soil resources and microbial activity shape the distributions of mountain plants based on their functional traits. *Ecography*, *43*, 1550–1559. https://doi.org/10.1111/ecog.05269

McCain, C. M., & Grytnes, J.-A. (2010). Elevational gradients in species richness. In John Wiley & Sons, Ltd (Ed.). *Encyclopedia of life sciences*. (p. a0022548). John Wiley & Sons, Ltd.

McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology & Evolution*, *30*, 104–113. https://doi.org/10.1016/j.tree.2014.11.006

Mittelbach, G. G., Steiner, C. F., Scheiner, S. M., Gross, K. L., Reynolds, H. L., Waide, R. B., Willig, M. R., Dodson, S. I., & Gough, L. (2001). What is the observed relationship between species richness and productivity? *Ecology*, *82*, 2381–2396.

Molnar, C., Bischl, B., & Casalicchio, G. (2018). iml: An R package for interpretable machine learning. *JOSS*, *3*, 786. https://doi.org/10.21105/joss.00786

Moore, J. C., Berlow, E. L., Coleman, D. C., Ruiter, P. C., Dong, Q., Hastings, A., Johnson, N. C., McCann, K. S., Melville, K., Morin, P. J., Nadelhoffer, K., Rosemond, A. D., Post, D. M., Sabo, J. L., Scow, K. M., Vanni, M. J., & Wall, D. H. (2004). Detritus, trophic dynamics and biodiversity. *Ecology Letters*, *7*, 584–600. https://doi.org/10.1111/j.1461-0248.2004.00606.x

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

Nielsen, U. N., Osler, G. H. R., Campbell, C. D., Neilson, R., Burslem, D. F. R. P., & van der Wal, R. (2010). The enigma of soil animal species

diversity revisited: The role of small-scale heterogeneity. *PLoS One*, *5*, e11567. https://doi.org/10.1371/journal.pone.0011567

Oliverio, A. M., Geisen, S., Delgado-Baquerizo, M., Maestre, F. T., Turner, B. L., & Fierer, N. (2020). The global-scale distributions of soil protists and their contributions to belowground systems. *Science Advances*, *6*, eaax8787. https://doi.org/10.1126/sciadv.aax8787

Peters, M. K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., Ensslin, A., Ferger, S. W., Frederiksen, S. B., Gebert, F., Haas, M., Helbig-Bonitz, M., Hemp, C., Kindeketa, W. J., Mwangomo, E., Ngereza, C., Otte, I., Röder, J., Rutten, G., … Steffan-Dewenter, I. (2016). Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nature Communications*, *7*, 13736. https://doi.org/10.1038/ncomms13736

Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., Ferlian, O., Gongalsky, K. B., van den Hoogen, J., Krebs, J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E. M., Bennett, J. M., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., … Eisenhauer, N. (2019). Global distribution of earthworm diversity. *Science*, *366*, 480–485. https://doi.org/10.1126/science.aax4851

Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B. D., Clemmensen, K. E., Kauserud, H., Nguyen, N., Kjøller, R., Bates, S. T., Baldrian, P., Frøslev, T. G., Adojaan, K., Vizzini, A., Suija, A., Pfister, D., Baral, H.-O., Järv, H., Madrid, H., Nordén, J., … Tedersoo, L. (2020). FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Diversity*, *105*, 1–16. https://doi.org/10.1007/s13225-020-00466-2

Pontarp, M., Bunnefeld, L., Cabral, J. S., Etienne, R. S., Fritz, S. A., Gillespie, R., Graham, C. H., Hagen, O., Hartig, F., Huang, S., Jansson, R., Maliet, O., Münkemüller, T., Pellissier, L., Rangel, T. F., Storch, D., Wiegand, T., & Hurlbert, A. H. (2019). The latitudinal diversity gradient: Novel understanding through mechanistic eco-evolutionary models. *Trends in Ecology & Evolution*, *34*, 211–223. https://doi.org/10.1016/j.tree.2018.11.009

Potapov, A. M., Scheu, S., & Tiunov, A. V. (2019). Trophic consistency of supraspecific taxa in below-ground invertebrate communities: Comparison across lineages and taxonomic ranks. *Functional Ecology*, *33*, 1172–1183.

Potapov, A. A., Semenina, E. E., Korotkevich, A. Y., Kuznetsova, N. A., & Tiunov, A. V. (2016). Connecting taxonomy and ecology: Trophic niches of collembolans as related to taxonomic identity and life forms. *Soil Biology and Biochemistry*, *101*, 20–31. https://doi.org/10.1016/j.soilbio.2016.07.002

R Core Team (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rainford, J. L., & Mayhew, P. J. (2015). Diet evolution and clade richness in Hexapoda: A phylogenetic study of higher taxa. *American Naturalist*, *186*, 777–791. https://doi.org/10.1086/683461

Ranjard, L., & Richaume, A. (2001). Quantitative and qualitative microscale distribution of bacteria in soil. *Research in Microbiology*, *152*, 707–716. https://doi.org/10.1016/S0923-2508(01)01251-7

Räty, M., & Huhta, V. (2003). Earthworms and pH affect communities of nematodes and enchytraeids in forest soil. *Biology and Fertility of Soils*, *38*, 52–58. https://doi.org/10.1007/s00374-003-0614-5

Roy, J., Albert, C. H., Ibanez, S., Saccone, P., Zinger, L., Choler, P., Clément, J.-C., Lavergne, S., & Geremia, R. A. (2013). Microbes on the cliff: Alpine cushion plants structure bacterial and fungal communities. *Frontiers in Microbiology*, *4*, 64. https://doi.org/10.3389/fmicb.2013.00064

Schostag, M., Priemé, A., Jacquiod, S., Russel, J., Ekelund, F., & Jacobsen, C. S. (2019). Bacterial and protozoan dynamics upon thawing and freezing of an active layer permafrost soil. *ISME Journal*, *13*, 1345–1359. https://doi.org/10.1038/s41396-019-0351-x

Seaton, F. M., George, P. B. L., Lebron, I., Jones, D. L., Creer, S., & Robinson, D. A. (2020). Soil textural heterogeneity impacts bacterial but not fungal diversity. *Soil Biology & Biochemistry*, *144*, 107766. https://doi.org/10.1016/j.soilbio.2020.107766

Shade, A., Dunn, R. R., Blowes, S. A., Keil, P., Bohannan, B. J. M., Herrmann, M., Küsel, K., Lennon, J. T., Sanders, N. J., Storch, D., & Chase, J. (2018). Macroecology to unite all life, large and small. *Trends in Ecology & Evolution*, *33*, 731–744. https://doi.org/10.1016/j.tree.2018.08.005

Six, J., Bossuyt, H., Degryze, S., & Denef, K. (2004). A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research*, *79*, 7–31. https://doi.org/10.1016/j.still.2004.03.008

Spasojevic, M. J., & Suding, K. N. (2012). Inferring community assembly mechanisms from functional diversity patterns: The importance of multiple assembly processes. *Journal of Ecology*, *100*, 652–661. https://doi.org/10.1111/j.1365-2745.2011.01945.x

Stein, A., Gerstner, K., & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, *17*, 866–880. https://doi.org/10.1111/ele.12277

Steiner, C. F. (2001). The effects of prey heterogeneity and consumer identity on the limitation of trophic-level biomass. *Ecology*, *82*, 2495–2506.

Steiner, C. F., & Leibold, M. A. (2004). Cyclic assembly trajectories and scale-dependent productivity-diversity relationships. *Ecology*, *85*, 107–113. https://doi.org/10.1890/03-3010

Stres, B., Philippot, L., Faganeli, J., & Tiedje, J. M. (2010). Frequent freeze–thaw cycles yield diminished yet resistant and responsive microbial communities in two temperate soils: A laboratory experiment. *FEMS Microbiology Ecology*, *74*, 323–335. https://doi.org/10.1111/j.1574-6941.2010.00951.x

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Sulkava, P., & Huhta, V. (2003). Effects of hard frost and freeze-thaw cycles on decomposer communities and N mineralisation in boreal forest soil. *Applied Soil Ecology*, *22*, 225–239. https://doi.org/10.1016/S0929-1393(02)00155-5

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

Taberlet, P., Prud'homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & 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. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*, 1256688. https://doi.org/10.1126/science.1256688

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M. C., Schwager, M., & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. *Journal of Biogeography*, *31*, 79–92. https://doi.org/10.1046/j.0305-0270.2003.00994.x

Thakur, M. P., Phillips, H. R. P., Brose, U., De Vries, F. T., Lavelle, P., Loreau, M., Mathieu, J., Mulder, C., Van der Putten, W. H., Rillig, M. C., Wardle, D. A., Bach, E. M., Bartz, M. L. C., Bennett, J. M., Briones, M. J. I., Brown, G., Decaëns, T., Eisenhauer, N., Ferlian, O., … Cameron, E. K. (2020). Towards an integrative understanding of soil biodiversity. *Biological Reviews*, *95*, 350–364. https://doi.org/10.1111/brv.12567

Thuiller, W., Gravel, D., Ficetola, G. F., Lavergne, S., Münkemüller, T., Pollock, L. J., Zimmermann, N. E., & Mazel, F. (2020). Productivity

begets less phylogenetic diversity but higher uniqueness than expected. *Journal of Biogeography*, *47*, 44–58. https://doi.org/10.1111/jbi.13630

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe'er, G., Singer, A., Bridle, J. R., Crozier, L. G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J. J., Holt, R. D., Huth, A., Johst, K., Krug, C. B., Leadley, P. W., Palmer, S. C. F., Pantel, J. H., … Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, *353*, aad8466. https://doi.org/10.1126/science.aad8466

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., de Goede, R. G. M., Adams, B. J., Ahmad, W., Andriuzzi, W. S., Bardgett, R. D., Bonkowski, M., Campos-Herrera, R., Cares, J. E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S. R., Creamer, R., … Crowther, T. W. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, *572*, 194–198. https://doi.org/10.1038/s41586-019-1418-6

Vannier, O., & Braud, I. (2012). *Calcul d'une évapotranspiration de référence spatialisée pour la modélisation hydrologique à partir des données de la réanalyse SAFRAN de Météo-France*. irstea.

Villéger, S., Mason, N. W. H., & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, *89*, 2290–2301. https://doi.org/10.1890/07-1206.1

Vreeken-Buijs, M. J., Hassink, J., & Brussaard, L. (1998). Relationships of soil microarthropod biomass with organic matter and pore size distribution in soils under different land use. *Soil Biology & Biochemistry*, *30*, 97–106. https://doi.org/10.1016/S0038-0717(97)00064-3

Wagg, C., Bender, S. F., Widmer, F., & van der Heijden, M. G. A. (2014). Soil biodiversity and soil community composition determine ecosystem multifunctionality. *PNAS*, *111*, 5266–5270. https://doi.org/10.1073/pnas.1320054111

Wang, P., Chen, Y., Sun, Y., Tan, S., Zhang, S., Wang, Z., Zhou, J., Zhang, G., Shu, W., Luo, C., & Kuang, J. (2019). Distinct biogeography of different fungal guilds and their associations with plant species richness in forest ecosystems. *Frontiers in Ecology and Evolution*, *7*, 216. https://doi.org/10.3389/fevo.2019.00216

Wardeh, M., Risley, C., McIntyre, M. K., Setzkorn, C., & Baylis, M. (2015). Database of host-pathogen and related species interactions, and their global distribution. *Scientific Data*, *2*, 150049. https://doi.org/10.1038/sdata.2015.49

Wardle, D. A. (2006). The influence of biotic interactions on soil biodiversity. *Ecology Letters*, *9*, 870–886. https://doi.org/10.1111/j.1461-0248.2006.00931.x

White, H. J., León-Sánchez, L., Burton, V. J., Cameron, E. K., Caruso, T., Cunha, L., Dirilgen, T., Jurburg, S. D., Kelly, R., Kumaresan, D., Ochoa-Hueso, R., Ordonez, A., Phillips, H. R. P., Prieto, I., Schmidt, O., & Caplat, P. (2020). Methods and approaches to advance soil macroecology. *Global Ecology and Biogeography*, *29*, 1674–1690. https://doi.org/10.1111/geb.13156

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., & Kögel-Knabner, I. (2019). Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. *Geoderma*, *333*, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026

Wright, D. H. (1983). Species-energy theory: An extension of species-area theory. *Oikos*, *41*, 496–506. https://doi.org/10.2307/3544109

Xia, Q., Rufty, T., & Shi, W. (2020). Soil microbial diversity and composition: Links to soil texture and associated properties. *Soil Biology & Biochemistry*, *149*, 107953. https://doi.org/10.1016/j.soilbio.2020.107953

Xiong, W. U., Jousset, A., Li, R., Delgado-Baquerizo, M., Bahram, M., Logares, R., Wilden, B., de Groot, G. A., Amacker, N., Kowalchuk, G. A., Shen, Q., & Geisen, S. (2021). A global overview of the trophic structure within microbiomes across ecosystems. *Environment International*, *151*, 106438. https://doi.org/10.1016/j.envint.2021.106438

Xu, G.-L., Kuster, T. M., Günthardt-Goerg, M. S., Dobbertin, M., & Li, M.-H. (2012). Seasonal exposure to drought and air warming affects soil collembola and mites. *PLoS One*, *7*, e43102. https://doi.org/10.1371/journal.pone.0043102

Young, I. M., & Crawford, J. W. (2004). Interactions and self-organization in the soil-microbe complex. *Science*, *304*, 1634–1637. https://doi.org/10.1126/science.1097394

Zinger, L., Lionnet, C., Benoiston, A.-S., Donald, J., Mercier, C., & Boyer, F. (2021). metabaR: An r package for the evaluation and improvement of DNA metabarcoding data quality. *Methods in Ecology and Evolution*, *12*, 586–592.

BIOSKETCH

Irene Calderón-Sanou is a PhD student aiming at a better understanding on how environmental gradients and major disturbances influence soil multi-trophic diversity through the use of environmental DNA metabarcoding.

Author contributions: W.T., T.M. and I.C. conceptualized the study with the help of L.Z. and M.H. A.S. and W.T. led the field work, with the participation of M.O., C.M-A., J.R., C.L., L.G., I.C.S. and several members of the Orchamp consortium. L.G. executed on field eDNA extractions and all laboratory samples processing. C.L. and I.C.S. processed and cleaned the sequenced eDNA data. J.R. and C.M-A. provided the code for extracting the environmental data. NK sampled and measured the pit soil data. I.C.S., L.Z. and M.H. participated in the assignment of the taxa to trophic groups. I.C.S. ran the analyses. All co-authors participated in the writing led by I.C.S., W.T. and T.M.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Calderón-Sanou, I., Zinger, L., Hedde, M., Martinez-Almoyna, C., Saillard, A., Renaud, J., Gielly, L., Khedim, N., Lionnet, C., Ohlmann, M., Consortium, O., Münkemüller, T., & Thuiller, W. (2022). Energy and physiological tolerance explain multi-trophic soil diversity in temperate mountains. *Diversity and Distributions*, *00*, 1–16. https://doi.org/10.1111/ddi.13529

CHAPTER 4:

Spatial turnover of soil food webs along environmental gradients

Status: In preparation.

Collaborators: Marc Ohlmann, Lucie Zinger, Mickael Hedde, Clément Lionnet, Camille Martinez-Almoyna, Amélie Saillard, Julien Renaud, Nicolas Le Guillarme, Ludovic Gielly, Orchamp Consortium, Tamara Münkemüller & Wilfried Thuiller

Abstract: While soil food webs are key drivers of ecosystem functioning and associated services, we largely ignore how they change along large environmental gradients. Given their inherent complexity, responses of organisms may occur at different resolutions: broad trophic classes would respond if they were selected by different environments, and refined trophic groups would respond in addition if they were functionally complementary (and not redundant) within classes. Thus, looking at different resolutions when studying soil food webs may reveal different patterns of assembly. Here, we aimed at quantifying and investigating the spatial variation of soil food webs along the sharp environmental gradients of the whole French Alps. Using network dissimilarity metrics applied over 451 local soil food webs along 24 elevational transects, we (1) quantified soil food web turnover at two resolutions, and (2) deciphered the importance of geographic distance and environmental factors to explain spatial soil food web turnover. We found spatial variability of trophic groups and trophic interactions at both resolutions, but turnover between trophic classes was much weaker than between refined trophic groups. This confirmed the existence of a backbone of soil food webs (i.e. trophic classes that are always present). Environment variation explained much more of the soil food web turnover between sites than spatial distance. Soil C/N ratio and NDVI were the most important variables at both resolutions, while plant taxonomic turnover only influenced food web structure at the finer trophic group resolution. Our results illuminate the spatial structure of soil food webs at a large spatial scale, and their nested structure with a strong turnover of trophic groups determined by environmental filtering (i.e. trophic group sorting) ultimately constrained by a universal backbone of soil trophic interactions.

Keywords: food web structure, network turnover, trophic group resolution, trophic interactions, environmental DNA, mountain systems.

Introduction

Soils harbour an enormous diversity of functionally distinct organisms that coexist and interact at all trophic levels providing crucial ecosystem services such as carbon storage, organic matter decomposition, mineralization, nitrogen fixation, plant performance and resistance to pests and stress (Bardgett & van der Putten 2014; Delgado-Baquerizo *et al.* 2020). Understanding the ecological processes that shape soil biodiversity and its spatial turnover is thus fundamental to predicting the consequences of global changes on terrestrial ecosystems and guiding more integrative conservation strategies (Soliveres *et al.* 2016; Guerra *et al.* 2021). Although our knowledge of the spatial distribution of soil biodiversity, especially for specific groups like earthworms, has improved in the last decades (Decaëns 2010; Orgiazzi *et al.* 2016; Rutgers *et al.* 2016), a complete picture of how the whole soil biodiversity responds to environmental drivers is still missing. Yet quantifying changes in the spatial structure of soil biodiversity and quantifying the drivers of those changes, like space or environmental variation, should bring crucial knowledge on the ecological processes structuring soil communities, and will reveal spatial changes in the functioning of terrestrial ecosystems (Eisenhauer *et al.* 2021).

Ecological networks, such as food webs, provide a suited representation of multitrophic communities as it considers simultaneously several functionally important groups and their linkages across trophic levels (Thompson *et al.* 2012). In soil food webs, nodes are groups of organisms sharing the same set of prey and predators or with similar functions in the ecosystem (Eltonian niche, Elton 1927), and edges represent their trophic interactions. However, due to the inherent complexity of soil biodiversity, the definition of these groups can vary from refined trophic groups (e.g. nematode bacterivore and predatory coleopteres, Potapov 2022) to broad trophic classes (e.g. herbivores and decomposers, Buzhdygan *et al.* 2020), changing the lens at which we express and quantify soil food web structure. Detecting changes in the structure of soil food webs along environmental gradients can critically depend on the resolution at which groups are built (conceptual figure to be done) broad trophic classes would respond if they were selected by different environments, and refined trophic groups would respond in addition if they were functionally complementary (and not redundant) within classes. While this context-specific definition might be seen as a weakness, we rather see it as a strength. First, it allows the grouping of organisms that could be resolved at different taxonomic precisions. Last but not least, grouping organisms at increasing resolutions allows the analysis of soil food web structures while zooming in and zooming out on the web. As such, detecting strong soil food web turnover at a very fine trophic group resolution that vanishes at coarser resolution (i.e.

trophic class) implies the identification of a common backbone of interactions underlying soil food web, with a turnover only within its low-level classes (Bramon Mora *et al.* 2018). Another interesting feature of analysing soil food webs is that while trophic groups (or classes) and interaction diversity are inherently correlated, they might vary differently along with the environment (Poisot *et al.* 2012; Thompson *et al.* 2012), and their joint analysis should illuminate the processes structuring food webs. Indeed, differences in food web structures in terms of trophic interactions depend on the degree of interactions of the trophic groups, e.g., the removal of a trophic group interacting with many other groups will lead to higher structural turnover than a trophic group poorly connected (conceptual figure to be done).

ß-diversity metrics that quantify the compositional dissimilarity between pairs of communities are known to give considerable insights into the ecological processes controlling the spatial variation in community structure along ecological gradients (Baselga 2010; Chase & Myers 2011; Ohlmann *et al.* 2018; Martinez-Almoyna *et al.* 2019). For example, a decay in community similarity with spatial or environmental distance (distance decay of similarity, Nekola & White 1999) can reveal community assembling processes such as dispersal limitation and/or environmental filtering (i.e. species or group sorting), respectively. While this pattern has been widely documented across organisms in different ecosystems (Astorga *et al.* 2012; Graco-Roza *et al.* 2021), it has mostly been investigated within a single trophic level or multiple ones but taken independently of each other. In other words, how food web similarity changes with both spatial and environmental similarity has been poorly addressed so far. Even more interesting, we might expect different distance decays when focusing on either trophic or interaction diversity, and when varying the trophic resolution at which soil food webs are expressed.

Here, we aimed at quantifying the spatial variation of soil food web structure (i.e. soil food web turnover) along the sharp environmental gradients of the French Alps. We compared the structure of 451 soil food webs located at nested spatial distances (from 8m to 250km), disposed along 24 elevational gradients and along with various environmental conditions across the French alps (Fig. 1A). Food web turnover was measured using network dissimilarity metrics (Ohlmann *et al.* 2019), which consider the dissimilarity of groups and interactions separately, at two levels of resolution: trophic groups (fine resolution) and trophic classes (coarse resolution). First, we assessed whether the structure of soil food webs varied across the French Alps and how it depended on the resolution at which the food web was considered. We expected

less variation at the trophic class level, which we hypothesised to be a backbone soil food web that should be almost invariant in space. Second, we compared the strength of the geographic vs. the environmental distances in shaping soil food web structure at both levels of resolution. We then identified the main environmental variables explaining this variation, among climatic, soil and plant-related variables. We expected soil food webs to be strongly structured by the environment as a consequence of environmental filtering acting on trophic groups. Functional approaches, such as the one used to create the trophic groups, assume that environmental filtering selects species with suites of traits that allow them to coexist under similar environmental conditions (Ackerly & Cornwell 2007). At short spatial distances, we could expect two contrasting results. On one hand, soil organisms might be seen as highly dispersive and thus neighbour soil food webs should have a similar structure. On the other hand, the soil itself is highly heterogeneous and might drastically change over small spatial distances implying strong environmental filtering on organisms.

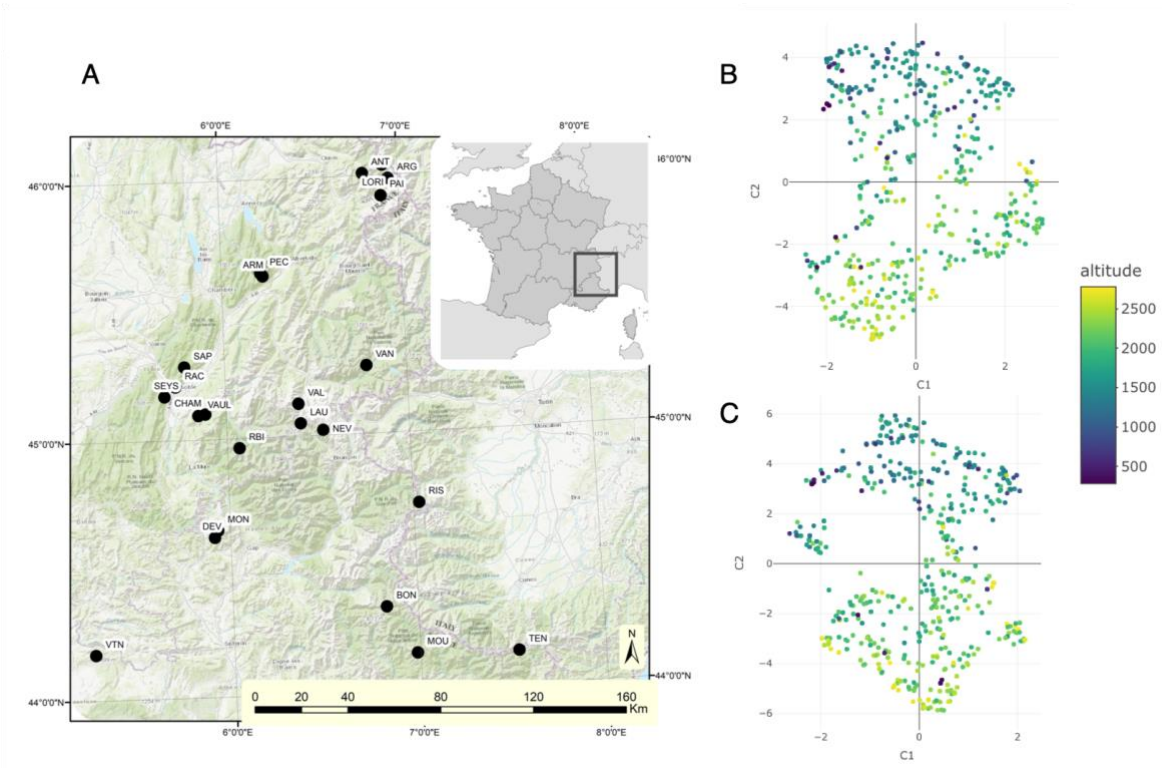


Figure 1. Map of the study site showing the distribution of the 24 elevational transects sampled in the French Alps (A). UMAP 2-D plane representing the similarity between soil food webs (each dot is a food web) from the point of view of trophic group composition (B) and trophic group interactions (C), with a colour scale representing the altitude. Dots that are close to each other in the 2D plane have similar structures in trophic group composition or interactions.

Material and Methods

Study site

The data was obtained from the long-term observatory, Orchamp (www.orchamp.osug.fr, Fig.1A), made of multiple elevational transects distributed across the whole French Alps, with contrasting climatic, vegetation and pedological conditions. Each elevational transect consisted of four to nine 30 x 30 m permanent plots separated by 200 m of altitude, on average. In this study, we used data gathered from 2016 to 2020, corresponding to 24 elevational transects and 113 plots. Plant species abundances were quantified at the vegetation peak (mostly in July or August) along a linear transect crossing each plot using the pin-point method (Jonasson 1988). A second 4 m wide transect was dedicated to soil sampling at the end of the summer season. The soil was sampled from 3 subplots (2 x 2 m) selected across the transect. Around ten soil cores of 5 cm in diameter were collected per subplot and pooled together to make a biological sample. Some elevational transects were sampled two times (i.e., resurveyed), in two different years. Soil samples from the same subplot but sampled in different years were considered as separate samples in the analyses, and their spatial dependency was considered indirectly through the spatial coordinates of the plot. A total of 451 soil samples were thus treated in this study, equivalent to 415 soil food webs.

Metabarcoding

The retrieved 451 samples were processed following the same procedure described in Chapter 3, but different clustering thresholds were applied to obtain the Molecular Operational Taxonomic Unit (MOTU) of the specific markers (i.e., 85% for Coll01, 88% for Olig01 and 95% for Inse01), while the clustering threshold of 97% was conserved for the universal markers (i.e., Euka02, Fung01, Bacte01), following the recommendations in Bonin *et al.* (2021). Metazoan taxa not registered in the European region were removed using the GBIFfilter tool (https://github.com/nleguillarme/gbif-filter-python).

Food web construction

The retrieved taxonomically annotated sequences were assigned to trophic groups and trophic classes, to further build the soil metaweb at two levels of resolution as described in Chapter 2.

Environmental variables selection

We selected a set of weakly correlated variables representing the climatic, soil and vegetation environmental categories. For climate, we retained growing degree days at 0°C (GDD), total

annual precipitation and frost degree days (FDD). For soil, we used the soil pH, the amount of organic matter and the C/N ratio in the organic matter. For vegetation we used plant taxonomic dissimilarity, plant functional dissimilarity and NDVI. The environmental variables were calculated as in chapter 3. Plant taxonomic dissimilarity was estimated with the Jaccard pair-wise dissimilarity index and calculated using the *beta.pair* command from the R package 'betapart' (R Core Team 2020; Baselga *et al.* 2022), and plant functional dissimilarity with the Jaccard-like functional index from the *beta.fd.multidim* command from the R package 'mFD' (Magneville *et al.* 2021).

Statistical analyses

We quantified the dissimilarity in soil food webs using a set of network diversity metrics that generalise the Hill numbers to networks (Ohlmann *et al.* 2019). These network metrics allow calculating both the diversity of trophic groups or classes and the diversity of interactions. Additionally, relying on Hill numbers, these network metrics allow to weight the diversity measures by the relative abundance of organisms, a highly desirable property, especially when focusing on trophic classes that can have large differences in the relative abundance of trophic groups within them. In traditional community diversity analyses, this weighting parameter, called *q*, distinguishes species richness (q=0), Shannon entropy (q=1) and Simpson diversity (q=2). Here, we used both q=0 and q=1 to account for the relative abundances of sequences in trophic groups and the relative frequencies of trophic groups within the trophic classes. The relative abundance of sequences were estimated using a double-transformation, where first, total read counts were transformed into proportions within the sample, and second, the resulting proportions were standardized by the largest observed proportion observed across samples for each trophic group. Relative abundances of trophic groups varied thus between 0 (absent) to 1 (largest observed proportion), allowing to have a comparable measure across trophic groups. Relative frequencies of trophic classes were the sum of the relatives abundances of the trophic groups within the class, standardized across the whole food web to sum 1. Network dissimilarity was thus calculated for all pairs of samples, using q=0 and q=1, and at the two resolutions (trophic group and trophic class), using the R package 'econetwork' (Miele *et al.* 2021).

To quantify the relative importance of environmental and spatial distances to explain food web turnover, we used Generalized Dissimilarity Models (GDM, Ferrier *et al.* 2007). We built a single GDM for each of the dissimilarity metrics (i.e. trophic group diversity, trophic group interaction, trophic class and trophic class interaction for both q=0 and q= 1, so 8 GDMs) using

the spatial coordinates of the samples and all selected environmental variables as predictors. The models were run using the R package "gdm" (Fitzpatrick *et al.* 2022). Variance partitioning between environment and space was assessed using the function *gdm.partition.deviance*. A variance partitioning analysis was also applied to assess the relative importance of the three environmental categories, i.e., climate, soil, plant.

Results

Variability of soil food webs across the French alps

The entire soil metaweb across the 451 soil samples was composed of 58 trophic groups and 383 potential interactions, which were aggregated into 11 trophic classes with 45 interactions. Local food webs were made of 41 ± 4 SD trophic groups, totalling around 204 ± 37 SD interactions, which corresponded to 11 trophic classes with 41 ± 2 SD interactions. The absence of variability in the number of trophic classes composing the local food webs supports our hypothesis on the existence of a backbone in soil food web structure at the trophic class level. The only exception was a single local food web where the class fungivore was absent. However, differences in the food web structure at the trophic class level were detected when accounting for relative abundances, although the mean dissimilarity at this level was rather low (trophic class dissimilarity: 0.09 ± 0.05 SD , interactions dissimilarity: 0.28 ± 0.12 SD).

Overall, we found that soil food web turnover was particularly well structured and varied along the elevational range covered in this study. This was the case for both trophic groups and their interactions (Fig.1. B,C). In general, mean dissimilarity per pair of samples was higher for interactions than for groups and classes but relationships varied at the two resolutions (Fig. 2). Interestingly, for a given dissimilarity in trophic groups or classes of soil food webs, the dissimilarity of interactions was highly variable, and this variability was higher at the trophic class level and when accounting for trophic group relative abundances (i.e., with q=1; Fig. 2).

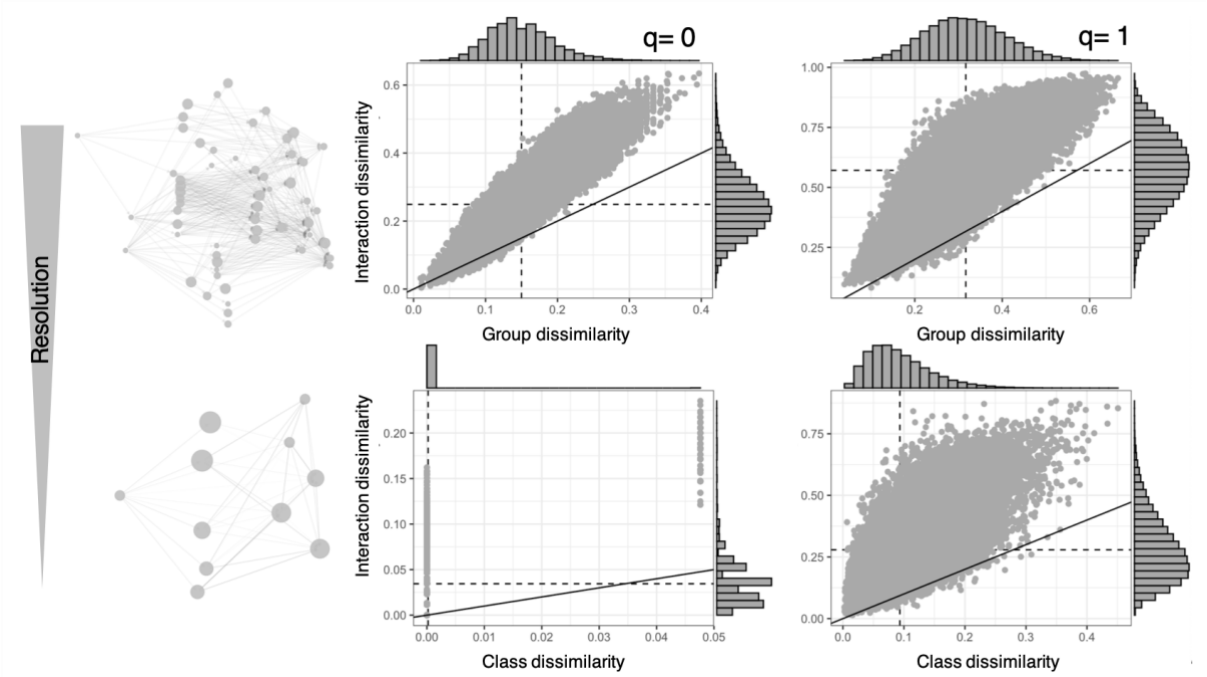


Figure 2. Relationship between trophic interaction dissimilarity and trophic group/class dissimilarity for all pairwise comparisons among the 451 soil food webs sampled across the French Alps. The relationship is showed for the dissimilarity metrics calculated using q=0 (left panels) and q=1 (right panels). The solid line represents the 1:1: relationship. Dashed lines correspond to the mean dissimilarity of groups/classes (vertical) and interactions (horizontal) and the histograms represent the variation in dissimilarity across pairwise comparisons between soil food webs.

Spatial vs environment drivers of soil food web turnover

The GDM explained 10.5 to 20.5% of the variance in soil food web dissimilarity across samples (Fig.3A). The GDMs for the trophic class resolution at q=0 could not be run due to the low variability. The total variance explained was higher for food web structure at the finer resolution, which is to be expected due to the greater variation at this resolution. The variance explained when using presence/absence data (i.e., q=0) was less than 10% for all metrics), so further we considered only the results for the abundance data. The environment alone explained most part of the variance for all dissimilarity metrics (Fig.3.A), and this was mostly due to variables associated with soil and plant communities (Fig. 3B). Climatic variables explained less than 3% of the variance in soil food web structure dissimilarity, including combined effects with plant and soil variables. The C/N ratio in soil organic matter and the NDVI were the most important factors explaining soil food web turnover across the French Alps for all the metrics

and at both resolutions (Fig. 3B). The rate of turnover of groups and classes was higher along the lower part of the C/N ratio gradient (~0-10), but the interaction's turnover was constant along the gradient. The rate of food web turnover was constant along the NDVI gradient. The dissimilarity of trophic groups and their interactions increased constantly with the plant taxonomic dissimilarity, but not the dissimilarity of trophic classes and their interaction. The geographic distance had an effect on food web dissimilarity only at large spatial scales.

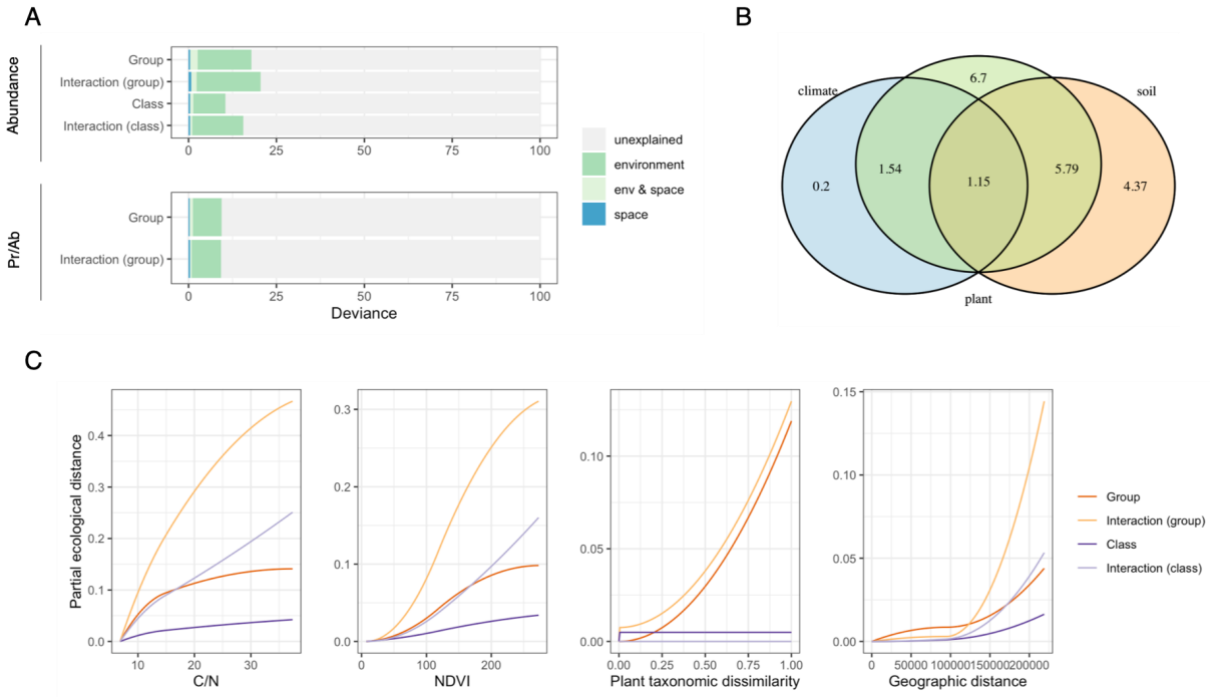


Figure 3. Summary of GDM results. (A) Variance partitioning of the deviance is explained by environmental vs spatial distances for the different dissimilarity metrics of food web structure at the level of groups and classes. (B) Variance partitioning between the environmental variables categorized into climate, soil and plant for the dissimilarity of interactions at the trophic group level for q=1. (C) The selected most important predictors of the food web structure dissimilarity based on the GDM for q=1. Each panel shows the partial ecological distance, or food web turnover, as a function of an environmental predictor when holding all other variables constant. The slope at any point on the curve indicates the rate of food web turnover at that position along the environmental gradient (x-axis), while the total height reached by the function indicates the total amount of food web turnover due to that environmental predictor.

Discussion

Studying how different metrics of food web structure co-vary with the environment is needed to gain a better understanding of the processes structuring soil multitrophic communities. In this study, we showed that the local structure of soil food webs varies along environmental gradients in the French Alps, especially when using a fine resolution food web. Instead, the variability was low when using a coarser resolution food web, showing a conserved trophic structure at this level of resolution. We found that the local food web dissimilarity across samples was better explained by the environment than by the geographic space. Interestingly we found that similar environmental variables explained most part of the variation in the food web structure at both levels of resolution (i.e., groups, classes, interactions between groups, the interaction between classes), except for plant taxonomic dissimilarity, which only influenced the food web structure at the finer resolution.

Aspects to discuss:

- The idea of a backbone at the trophic class level.
- The effect of C/N and NDVI on food web structure based on other studies. Changes in C/N ratio are also related to the type of habitat (forests vs grassland). We need to explore more the effect of habitat on soil food web in the analyses. Include also discussion about the most important turnover at the low range of the C/N ratio gradient and the differences in the curves of groups and interactions.
- Relate the results with the results from Chapter 3 → the drivers of trophic group's diversity. Energy, mostly NDVI, was also a main determinant of group's diversity.
- Mentioned that when analysing the alpha diversity of food webs, pH was the most important factor → pH limits the alpha diversity of the food webs, while energetic constrains defines the composition of food webs.
- Discuss the influence of plant composition on food web structure only at the fine resolution level → the broad trophic structure is conserved but the identity of the groups within class depends on plant identity (?).
- Discuss the low importance found for geographic distances. Maybe coordinates used are not very representative. Also, the large differences in the distances across samples (within plot vs across gradients)--> the effect of large distances may obscure the effect of small distances. Rescaling the distances could be a potential solution to be tested. Spatial distances might be less important at larger scales when accounting for trophic

groups because species dispersal limitations are obscured by grouping the species into trophic groups. But, at smaller distance we could expect that food webs more close are more similar. Dispersal limitation can be very important for larger organisms such as oribatids for which few centimetres of unsuitable habitat can limit the dispersion of most species.

References

Ackerly, D.D. & Cornwell, W.K. (2007). A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecol Letters*, 10, 135–145.

Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R. & Muotka, T. (2012). Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules?: Decay of similarity in freshwater communities. *Global Ecology and Biogeography*, 21, 365–375.

Bardgett, R.D. & van der Putten, W.H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, 515, 505–511.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity: Partitioning beta diversity. *Global Ecology and Biogeography*, 19, 134–143.

Baselga, A., Orme, D., Villeger, S., Bortoli, J.D., Leprieur, F., Logez, M., *et al.* (2022). *betapart: Partitioning Beta Diversity into Turnover and Nestedness Components*.

Bonin, A., Guerrieri, A. & Ficetola, G.F. (2021). *Optimal sequence similarity thresholds for clustering of molecular operational taxonomic units in DNA metabarcoding studies* (preprint). Preprints.

Bramon Mora, B., Gravel, D., Gilarranz, L.J., Poisot, T. & Stouffer, D.B. (2018). Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nat Commun*, 9, 2603.

Buzhdygan, O.Y., Meyer, S.T., Weisser, W.W., Eisenhauer, N., Ebeling, A., Borrett, S.R., *et al.* (2020). Biodiversity increases multitrophic energy use efficiency, flow and storage in grasslands. *Nat Ecol Evol*, 4, 393–405.

Chase, J.M. & Myers, J.A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Phil. Trans. R. Soc. B*, 366, 2351–2363.

Decaëns, T. (2010). Macroecological patterns in soil communities. *Glob Ecol Biogeogr*, 19, 287–302.

Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., Eldridge, D.J., Abades, S., Alfaro, F.D., *et al.* (2020). Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nat Ecol Evol*, 4, 210–220.

Eisenhauer, N., Buscot, F., Heintz-Buschart, A., Jurburg, S.D., Küsel, K., Sikorski, J., *et al.* (2021). The multidimensionality of soil macroecology. *Glob Ecol Biogeogr*, 30, 4–10.

Elton, C.S. (1927). *Animal ecology*. Macmillan Co., New York,.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13, 252–264.

Fitzpatrick, M., Mokany, K., Manion, G., Nieto-Lugilde, D., Ferrier, S., Lisk, M., *et al.* (2022). *gdm: Generalized Dissimilarity Modeling*.

Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A.T.R., Alahuhta, J., Altman, J., *et al.* (2021). Distance decay 2.0 – a global synthesis of taxonomic and functional turnover in ecological communities. *bioRxiv*.

Guerra, C.A., Bardgett, R.D., Caon, L., Crowther, T.W., Delgado-Baquerizo, M., Montanarella, L., *et al.* (2021). Tracking, targeting, and conserving soil biodiversity. *Science*, 371, 239–241.

Jonasson, S. (1988). Evaluation of the Point Intercept Method for the Estimation of Plant Biomass. *Oikos*, 52, 101–106.

Magneville, C., Loiseau, N., Albouy, C., Casajus, N., Claverie, T., Escalas, A., *et al.* (2021). *mFD: Compute and Illustrate the Multiple Facets of Functional Diversity*.
Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., Ohlmann, M., Foulquier, A., Clément, J., *et al.* (2019). Multi-trophic β-diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. *Funct Ecol*, 33, 2053–2064.
Miele, V., Matias, C., Dray, S., Botella, C., Ohlmann, M., Poggiato, G., *et al.* (2021). *econetwork: Analyzing Ecological Networks*.
Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *J Biogeogr*, 26, 867–878.
Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L., *et al.* (2018). Mapping the imprint of biotic interactions on β-diversity. *Ecology Letters*, 21, 1660–1669.
Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L. & Thuiller, W. (2019). Diversity indices for ecological networks: a unifying framework using Hill numbers. *Ecol Lett*, 22, 737–747.
Orgiazzi, A., Bardgett, R.D. & Barrios, E. (2016). Global soil biodiversity atlas. *Global soil biodiversity atlas.*
Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecol Lett*, 15, 1353–1361.
Potapov, A.M. (2022). Multifunctionality of belowground food webs: resource, size and spatial energy channels. *Biological Reviews*, n/a.
R Core Team. (2020). *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., *et al.* (2016). Mapping earthworm communities in Europe. *Applied Soil Ecology*, Soil biodiversity and ecosystem functions across Europe: A transect covering variations in bio-geographical zones, land use and soil properties, 97, 98–111.
Soliveres, S., van der Plas, F., Manning, P., Prati, D., Gossner, M.M., Renner, S.C., *et al.* (2016). Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature*, 536, 456–459.
Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., *et al.* (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol Evol*, 27, 689–697.

CHAPTER 5:

Cascading effects of moth outbreaks on subarctic soil food webs

scientific reports

Check for updates

OPEN

Cascading effects of moth outbreaks on subarctic soil food webs

Irene Calderón-Sanou[1✉], Tamara Münkemüller[1], Lucie Zinger[2], Heidy Schimann[3], Nigel Gilles Yoccoz[4], Ludovic Gielly[1], Arnaud Foulquier[1], Mickael Hedde[5], Marc Ohlmann[6], Mélanie Roy[7,8], Sara Si-Moussi[5] & Wilfried Thuiller[1]

The increasing severity and frequency of natural disturbances requires a better understanding of their effects on all compartments of biodiversity. In Northern Fennoscandia, recent large-scale moth outbreaks have led to an abrupt change in plant communities from birch forests dominated by dwarf shrubs to grass-dominated systems. However, the indirect effects on the belowground compartment remained unclear. Here, we combined eDNA surveys of multiple trophic groups with network analyses to demonstrate that moth defoliation has far-reaching consequences on soil food webs. Following this disturbance, diversity and relative abundance of certain trophic groups declined (e.g., ectomycorrhizal fungi), while many others expanded (e.g., bacterivores and omnivores) making soil food webs more diverse and structurally different. Overall, the direct and indirect consequences of moth outbreaks increased belowground diversity at different trophic levels. Our results highlight that a holistic view of ecosystems improves our understanding of cascading effects of major disturbances on soil food webs.

Natural disturbances, such as fires, droughts, or insect outbreaks, are key drivers of ecosystem dynamics and community structure[1]. Global change could exacerbate their severity and frequency worldwide with potential extensive impacts on biodiversity, ecosystems and human societies[2,3]. Understanding the effect of disturbances on the dynamics and structure of biodiversity is therefore more than ever a crucial issue in ecology. Yet, the high variability of local biodiversity trends in response to global changes asks for more integrative analyses, going beyond mere measures of species richness and accounting for the multiple components of the ecosystems[4,5]. Particularly, soil organisms are rarely included when synthesizing biodiversity trends in the face of disturbances, despite their recognized and well documented influence on multiple ecosystem functions (e.g. nutrient cycling) and nature contributions to people (e.g. carbon storage or depollution)[6–9].

Most studies quantifying the effect of disturbances on biodiversity have focused on a single trophic or taxonomic group, often directly affected by the disturbance, like plants[9]. However, much less is known on how the effects propagate across trophic levels ultimately affecting the entire ecosystem. Plants and soil organisms are tightly linked through direct and indirect interactions, including mutualism, parasitism or predation, which promote the exchange and supply of nutrients and ensure multiple ecosystem processes[6,7]. Ignoring these trophic interactions and how resource deprivation in one trophic level can cascade to other levels may obscure the true consequences of disturbances for ecosystems[10]. Furthermore, misleading conclusions could be drawn if resulting disturbance effects differ between trophic levels[11]. Most natural disturbances cause immediate fluctuations in the quantity and quality of available soil resources[1]. Extreme winds can remove or deposit organic matter on the forest floor, while insect outbreaks increase soil nutrient inputs through defoliation and insect faeces and corpses. These local changes in basal resource availability can have important consequences on the abundance

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Ecologie Alpine, 38000 Grenoble, France. [2]Institut de Biologie de L'ENS (IBENS), Département de biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. [3]INRA EcoFoG (AgroParisTech, CNRS, CIRAD, INRA, Université Des Antilles, Université de Guyane), Kourou, France. [4]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway. [5]Eco&Sols, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, 34398 Montpellier, France. [6]Université Savoie Mont-Blanc, LAMA, 73000 Chambéry, France. [7]Laboratoire Évolution Et Diversité Biologique, CNRS, UMR 5174 UPS CNRS IRD, Université Toulouse 3 Paul Sabatier, Toulouse, France. [8]Instituto Franco-Argentino Para El Estudio del Clima Y Sus Impactos (UMI IFAECI/CNRS-CONICET-UBA-IRD), Dpto. de Ciencias de La Atmosfera Y Los Oceanos, FCEN, Universidad de Buenos Aires, Intendente Guiraldes 2160 - Ciudad Universitaria (C1428EGA), Ciudad Autónoma de Buenos Aires, Argentina. ✉email: irecalsa@gmail.com

and diversity of primary producers (e.g. plants or nitrifying bacteria) and primary consumers (e.g. decomposers or herbivores), but also subsequently on the whole soil food web through bottom-up cascading effects[12–14]. Predicting whether the effects would vanish or amplify remains challenging due to the complexity of soil food webs in real ecosystems. Stoichiometry-based studies have provided numerous evidences that such indirect effects propagate across soil food webs from the microfauna to the macrofauna in terms of composition and biomass[15–17]. However, these approaches don't include the microbial part of the soil food web, and often lack resolution or breadth for the micro and macro fauna when describing the diversity and composition of these complex communities. In addition, changes in the abundance and diversity of organisms across the food web are likely to induce structural changes in the entire interaction network, potentially leading to alternative ecosystem states[8,18,19]. Thus, quantifying cascading effects of disturbances on ecosystems requires a holistic view of biodiversity with not only exhaustive sampling methods to capture all-in-end biodiversity, but also a suitable analytic approach to analyze changes in trophic levels and interactions.

To meet this challenge, we combined the power of environmental DNA metabarcoding (eDNA)[20] to obtain a nearly complete view of the belowground biodiversity, with a food web approach and network theory. Grouping species with the same trophic position (i.e. shared predators and preys/resources) in ecological networks facilitates the study of complex multitrophic communities[21–23]. In such an approach, the focus is not on species, but rather on trophic groups and trophic interactions. The definition of the trophic groups depends both on the resolution of the observation units (e.g. the taxonomic resolution) and the information available on their diet or trophic position[24–26], and is also related to the ecological question. When studying the large-scale consequences of disturbances on biodiversity, there is a trade-off between sufficiently fine resolution to reliably and meaningfully measure cascading effects[22,27], and sufficiently broad resolution to avoid knowledge gaps and cope with heterogeneity of taxonomic resolution in the data[25,28]. Once a food web is built, diversity can be measured within trophic groups (e.g. species diversity) and between trophic groups (e.g. trophic diversity or diversity of interactions), allowing the integration of ecological processes occurring at different dimensions of the food web (e.g. competition and predation)[29,30]. For this, network theory provides appropriate metrics to describe and compare the diversity and structure of ecological networks, accounting for both group abundances and interactions[31,32].

Here, we study the effect of moth outbreaks on soil food webs of subarctic birch forests in Northern Fennoscandia. These forests have experienced moth outbreaks of unprecedented scale and severity in recent decades, which have led to a sudden and persistent vegetation change -from birch forests with understory dominated by dwarf shrubs to grass-dominated systems associated with high tree mortality- that was still visible 8 years after the disturbance[33–36] (Fig. 1). Moth outbreaks is a good model for assessing the cascading effects of disturbance on soil food webs, as the larvae only attack the foliage of the dominant primary producers, i.e. the birch tree (*Betula pubescens*), and some abundant species of erect and dwarf shrubs in the understory layer (e.g. *Betula nana, Empetrum nigrum, Vaccinium* spp.). In parallel, soil organic matter is enriched through dead plants and N addition from larval faeces and corpses[37,38]. We can therefore expect that impacts on the whole soil food web arise from bottom-up effects from changes in the vegetation and basal resources to the other trophic compartments[12]. Drastic shifts in the composition of biological communities following defoliation have been already reported in these nutrient-limited soils where the dominance of the allelopathic dwarf shrub *Empetrum nigrum* in the understory leads to regressive succession that may inhibit soil microbial activity, organic matter decomposition, and thus nutrient availability[39–41]. These shifts correspond to a replacement of *Empetrum nigrum* by the grass *Avenella flexuosa*[34] with subsequent effects on the diversity and abundance of organisms directly relying on plants, including vertebrate herbivores[33], birds[42], saproxylic beetles[38], and fungal communities[43,44]. However, we still ignore whether moth outbreaks induced indirect effects across the soil food web, whether these effects are of comparable magnitude to those observed for vegetation, and finally, whether these effects have significant consequences on trophic interactions and ultimately on the whole soil food web structure.

We used eDNA data obtained from 86 soil samples from two well-studied areas in northeastern Norway (i.e., Tana and Kirkenes). This study design allowed for appropriate pairwise comparisons between coupled undamaged and defoliated forest based on well-documented defoliation patterns from both remote sensing and field methods (Fig. 1). The sampling design aimed at capturing the environmental heterogeneity at different spatial scales of the landscape within these areas. We then classified both microorganisms and macroinvertebrates into 9 broad trophic classes and 37 finer trophic groups to build metawebs[45] at two levels of resolution for the study area (Fig. 2). The metawebs were then used to infer local soil food webs based on taxa detected locally in each soil sample. The trophic class resolution corresponds to what is commonly used in soil food web ecology (e.g.[22,27]), but we additionally included the trophic group resolution because a finer resolution is needed to capture specific effects of disturbance on groups that are hidden at a coarser resolution. For instance, different types of mycorrhizal fungi like arbuscular mycorrhizal fungi and ectomycorrhizal fungi may have opposite responses to tree defoliation, the former increasing and the later decreasing in their proportion following disturbances[46].

Using this approach, we tested three hypotheses about the cascading effects of moth defoliation on the local soil food webs at different levels of organization. First, (H_1) moth defoliation changes the diversity in MOTUs (Molecular Operational Taxonomic Unit) and the relative abundances of most trophic groups. We expected positive effects on most decomposers and their consumers through the impulse in soil resources availability[47,48] from both moth outbreaks and the decreased abundance of the allelopathic species *Empetrum nigrum*. In parallel, we expected negative effects on e.g., ectomycorrhizal and ericoid mycorrhizal fungi, as the result of the decline of birch and ericaceous shrub roots. Second, (H_2) the magnitude of the effect differs among trophic groups across the soil food web. We expected the effect of defoliation to be stronger for primary consumers and decomposers that are directly affected by changes in basal resources availability and plant composition, and then to decrease toward higher trophic levels (attenuation of the effects). Third, (H_3), moth defoliation changes the overall structure of the local soil food webs[10,49]. We expected to observe differences in the trophic groups and links diversity and composition of the local food webs between defoliated and undamaged forests.

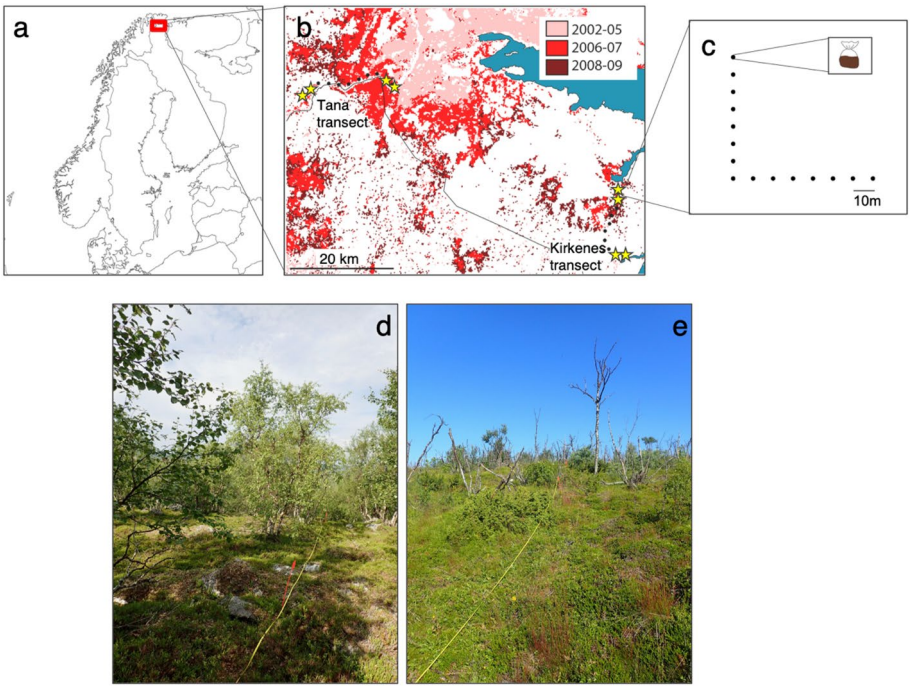


Figure 1. Sampling design in undamaged and defoliated forests. Map of the study location in the Northeastern Norway (**a**), in the Varanger region (**b**). Red areas (**b**) represent birch forest that experienced severe defoliation during the most recent moth outbreak. Yellow stars indicate the stations that were sampled in this study, at each extreme of two pre-established transects (black dots) going from highly defoliated forests stands to undamaged stands. Soil sampling was conducted in each station along an L-shaped transect (**c**). Photos illustrate the stations from undamaged (**d**) and defoliated (**e**) forests. Red flags in the photos indicate the sampling points represented in (**c**). Undamaged forests were characterized by living birch trees (*Betula pubescens*) and a shaded understory dominated by ericaceous shrubs (e.g. *Empetrum nigrum*). Defoliated forests were characterized by dead birch trees, patches of remaining ericaceous shrubs and a soil covered by light-tolerant grass and herbs including the dominant *Avenella flexuosa*. Photo credits: Heidy Schimann. Map (**a**) was created using ArcGIS® software 10.4.1 by Esri (www.esri.com). Map (**b**) was modified from[38] (https://doi.org/10.1371/journal.pone.0099624.g001).

Results and discussion

Fitting a multilevel linear model for each trophic group, we found that moth defoliation increased MOTU diversity and the relative abundances of most trophic groups (Fig. 3). This is consistent with H_1 and food web theory predictions, i.e. the effect of disturbances should propagate up the food web levels when resources are enriched through bottom-up processes[14,50]. Overall, diversity and relative abundance followed similar trends within trophic groups (Fig. 3a,b).

The basal groups directly linked to plants or basal resources (e.g., soil organic matter and light), i.e., mycorrhizal fungi, phytophagous or plant parasites, decomposers and primary producers, were expected to respond to changes in the composition of plant communities and nutrient enrichment following the outbreak. Here, comparing undamaged and defoliated forests, we observed a radical shift from ectomycorrhizal to arbuscular mycorrhizal fungal communities. This is consistent with the reduction of birch fine woody roots in defoliated forests, which are obligate hosts for most ectomycorrhizal fungi, and with the increases of herb and grass roots that are mostly associated with arbuscular mycorrhizal fungi[43,44,51] (Supplementary Fig. 2). The increased diversity and relative abundance of slugs, snails and plant pathogen protists could be in part explained by the increased palatability of the plant assemblages. Indeed, grasses like *Avenella flexuosa,* which is dominant in the defoliated forests, are more palatable as compared to allelopathic species like *Empetrum nigrum*[33,52]. Photosynthetic protists diversity and relative abundance also increased in defoliated sites which are more open, hence allowing more light to reach the soil (Fig. 1).

Among the decomposers, defoliation led to an increase in the diversity of heterotrophic bacteria, protists, saprotroph-plant pathogen fungi and wood saprotroph fungi. Similarly, the relative abundance of protists, saprotroph-plant pathogen fungi and enchytraeids increased. Differences in plant litter chemistry between undamaged and defoliated forests (Supplementary Fig. 2) might drive the communities of decomposers[53] and could explain these changes. For instance, the litter produced by *Empetrum nigrum,* which dominates undamaged forests, releases of phenolic compounds[52] that can strongly reduce plant species diversity[40,41]. Such detrimental effects might also hold true for the diversity and abundance of most decomposers. Soils from defoliated forests had lower C/N ratios, suggesting that defoliation promote more labile, easily decomposable organic matter inputs (Supplementary Fig. 3) but more precise soil nutrient measurements would be needed to confirm this.

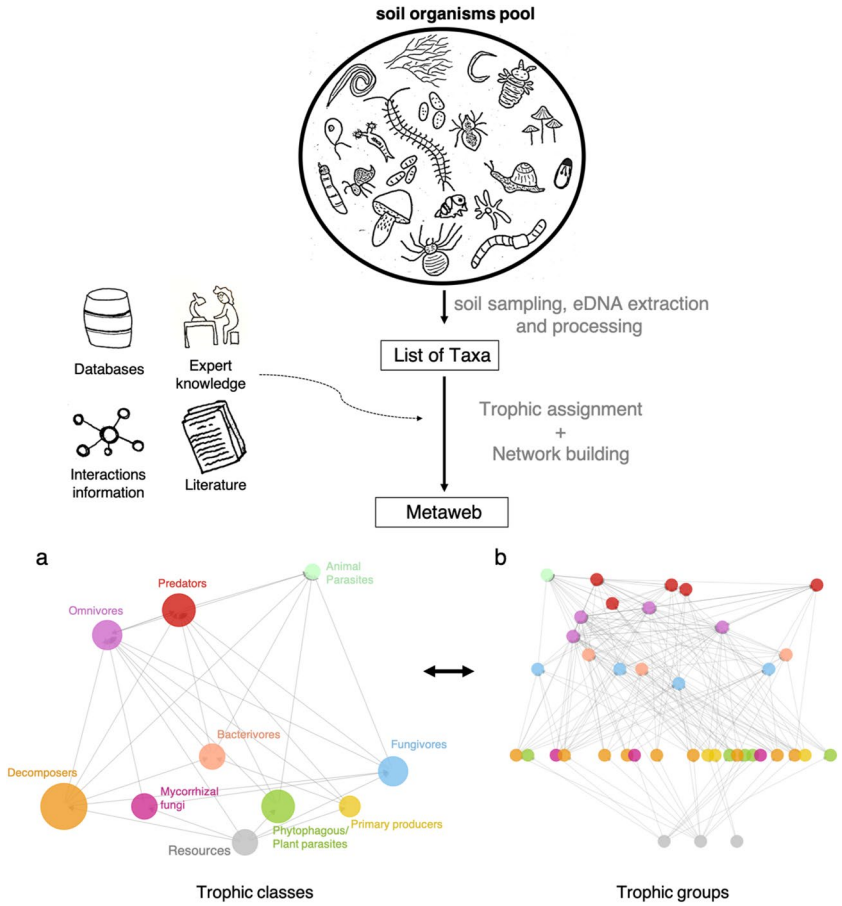


Figure 2. Methodology used to build the *metaweb* from soil eDNA. First, eDNA was extracted and processed from the 86 soil samples to obtain a list of taxa for the study area. Second, using an extensive collection of trophic knowledge from databases, literature and experts, taxa were assigned to broad trophic classes and then to finer trophic groups, which separate distant phylogenetic groups or groups that differ in their resources acquisition strategy. Main trophic links were collected from literature and current knowledge to build the *metaweb* at two levels of resolution (**a**,**b**). The colours correspond to the trophic classes (**a**) that are refined and split in the highly resolved *metaweb* (**b**).

Contrary to our expectation, the magnitude of the effect of defoliation did not decrease further up the food web (Fig. 3a,b), but was instead equally important at all trophic levels. This result did not depend on the number of sequences obtained for each group (Fig. 3c). This rejects the hypothesis of a mitigation of the effects of the disturbance when moving up to higher trophic levels in the soil food web (H_2). For example, the indirect effect of defoliation on the diversity of copiotrophic bacteria was as strong as the effect on their protist predators, and as strong as the effect on nematodes feeding on protists. In addition, the effect of defoliation on animal parasites, which are at the top of the soil food web, was similar to the effect on mycorrhizal fungi. Our findings are consistent with other studies pointing out that species-poor ecosystems, like subarctic birch forests, could be more prone to the propagation of bottom-up disturbances along food webs[54]. Furthermore, while some groups were affected by defoliation, other groups within the same trophic class were not (e.g. herbivore mite *vs.* plant pathogens protists, or ectomycorrhizal *vs.* ericoid mycorrhizal fungi). Other studies have highlighted the challenge of predicting the effect of an environmental stressor on overall biodiversity due to the variety of responses that organisms can have, associated with attributes such as dispersal abilities or resistance structures (e.g. cysts in protists)[55,56]. This is particularly important in soil food webs consisting of organisms with large differences in body size, life-span and life history strategies, and therefore in their response time to disturbance, which can vary from seconds to decades[7,48,56]. This complexity hampers our ability to detect consistent patterns when studying soil food webs at fixed sampling times.

We then examined how changes in trophic groups relative abundances influenced the network structure of local soil food webs, using network diversity indices[31]. Following H_3, moth defoliation significantly altered the whole soil food web structure in terms of node and link abundances, both for the trophic class and group resolutions (Supplementary Fig. 4). An increase in local diversity (α-diversity) of trophic groups and links in defoliated forests partially explained the changes in food web structure (Fig. 4). When zooming out to trophic classes, differences in the α-diversity of soil food webs were less obvious but food webs were nevertheless slightly more diverse for defoliated forests (Supplementary Fig. 5). This reflects that within a trophic class, trophic groups can

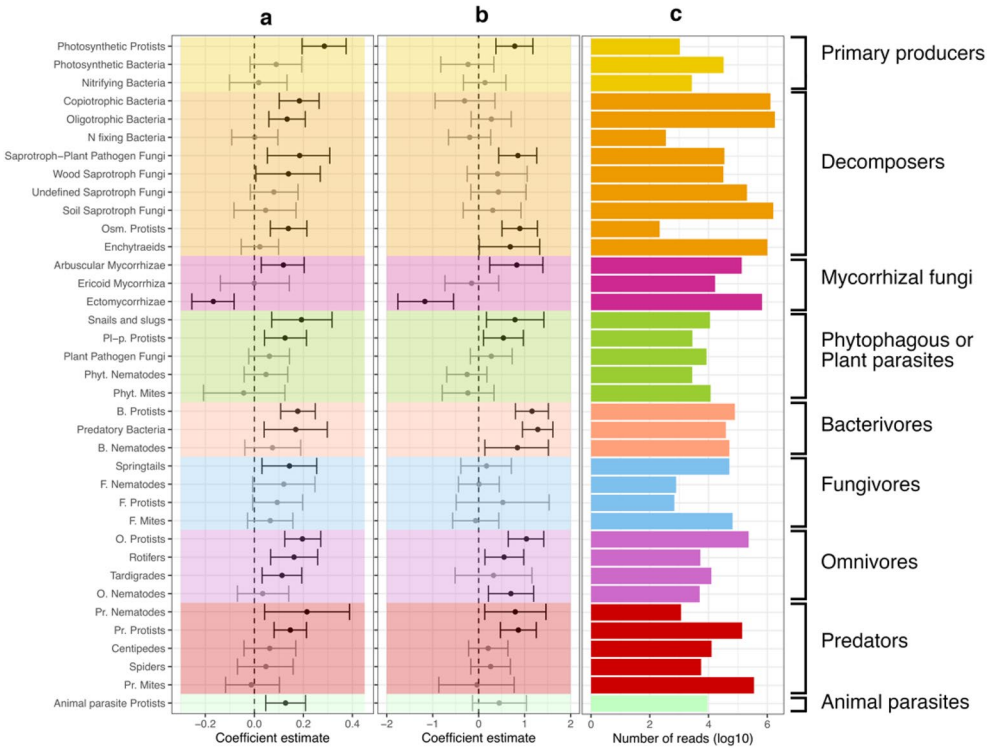


Figure 3. Effect of defoliation on diversity and relative abundance of trophic groups. Effect size of defoliation on MOTU diversity (**a**) and relative abundance of reads (**b**) for each trophic group with 90% credible intervals. The barplot (**c**) shows the total number of reads (logarithmic scale) of each trophic group in the overall dataset. A multilevel linear model was fitted individually for each trophic group with a dummy variable for defoliation as predictor and a random factor accounting for the nested sampling design. MOTU diversity was standardized by the maximum value observed within each trophic group to obtain comparable effect sizes between groups. The colours correspond to the trophic class definitions (see Fig. 2).

have opposite responses (Fig. 3) that are averaged out when only considering trophic class, and highlights the importance of using a finer trophic resolution than what is often used in the literature to understand the variability of cascading effects in the different components of the soil food web. On average, we observed a decrease in the proportion of most classes of primary consumers (i.e. plant mycorrhiza, herbivores/plant pathogens, decomposers) within the soil food webs in defoliated forests, and an increase in the proportion of higher trophic level classes (i.e. bacterivores, omnivores, predators), which were rare in the undamaged forests (Fig. 5). These changes in relative abundance proportions within the soil food web are not to be confounded with the individual changes in the relative abundances of the trophic groups (Fig. 3b). For instance, a decrease in the proportion of some classes might be related to weaker increase in average of the relative abundance of the groups within the class from undamaged to defoliated forests, compared to a stronger average increase for classes in higher trophic levels.

The observed shifts in the structure of soil food webs could translate into impacts on multiple ecosystem functions, including carbon and nutrient fluxes, and plant productivity[21,22]. Recent studies have observed a slowdown in soil C and N cycles following severe outbreaks in these forests and have related this result to the decrease in the below-ground C-allocation to the rhizosphere and the decrease of ectomycorrhizal fungi[37,51]. An interesting avenue would be to relate how other components of the food web diversity (e.g. decomposer channel) contribute to the C:N stoichiometry to derive predictions on the long-term effects of these important disturbances on biogeochemical cycles.

The spatial extent of the study was limited to two landscape areas of ca 20 km extent, and we acknowledge that further monitoring would be required to assess the full extent of soil food webs responses to moth outbreaks in subarctic birch forests. Previous studies have found that the effect of moth outbreaks on biological communities can vary depending on local productivity and climatic conditions (as represented by the two areas Tana and Kirkenes)[33,34,42]. We found, however, a consistent response for most soil organisms across the two areas that translated into significant local changes in the whole soil food web diversity and composition. The consistency and strength of the effects of defoliation on the different facets of local soil food webs point to general conclusions on the bottom-up cascading effects of moth outbreaks on soil communities in these subarctic birch forests, despite the heterogeneity in environmental condition of the studied system.

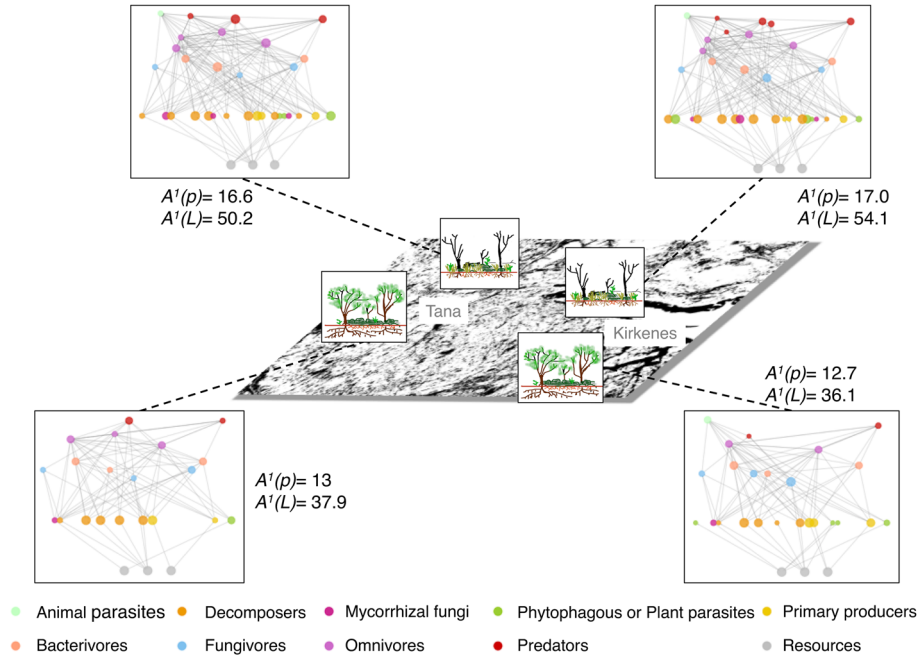


Figure 4. Topology and α-diversity of local food webs in undamaged *vs.* defoliated forests. The values represent the α-diversity of the soil food webs for each area and category of defoliation at the trophic group resolution: $A^1(\mathbf{p})$ is the diversity in trophic group abundances (nodes) and $A^1(\mathbf{L})$ the diversity in trophic links abundances (edges) using Shannon diversity. Nodes of the local food webs corresponded to the local relative abundances of the groups varying from 0 (when the group was absent) to 1 (when the group was at its maximum observed abundance). Links were binary links (i.e. present or absent) assuming an interaction when the two groups concerned were present. For the visualization, four local soil food webs (with an average value of $A^1(\mathbf{p})$) were selected to highlight the differences in diversity between undamaged and defoliated forests of each area. The colours correspond to the trophic classes and the nodes are distributed vertically based on their trophic level from the bottom (basal levels) to the top (higher levels).

Conclusion

The multitrophic approach used in this study, which combines an exhaustive diversity sampling (here eDNA data) with current trophic knowledge, an extended soil food web approach and ecological network theory, allows understanding the cascading effects of disturbances on soil biodiversity. We demonstrated that recent moth outbreaks in birch forests of Northern Fennoscandia caused major local shifts in the diversity and relative abundance of most trophic groups, ultimately changing the structure of the soil food web. We found more diverse soil food webs in defoliated forests compared to undamaged forests, accompanied by an increase in the proportion of groups in higher trophic levels.

We emphasize the need to consider different levels of resolution to ensure the robustness of conclusions and improve our understanding of how soil diversity responds to disturbances. Highly resolved food webs allow to map the cascading effects by revealing the variability of organisms' responses. In contrast, low resolution food webs provide a general picture on how these changes affect the food web structure. Our study opens new prospects in understanding the response of complex and diverse food webs to disturbance.

Material and methods

Sampling. The study took place in the Varanger region at approximately 70° N, 29° E, Northeastern Norway. This region is located in the transition between subarctic deciduous forests and the arctic tundra. Periodic outbreaks of the autumn moth (*Epirrita autumnata*) and more recently the winter moth (*Operophtera brumata*) have occurred in the region with a 9–10-years frequency approximately. Recently, the consecutive episodes by the two species caused a severe mortality of birch trees[35]. Sampling was replicated in two areas located approximately 70 km apart, both at the border of the outbreak range, but with slight differences in the defoliation year: Tana (70°03′ N, 27°45′ E.), defoliated during 2006–2007, and Kirkenes (69°46′ N, 29°20′ E) defoliated during 2007–2009. Differences in the forest characteristics between these two areas allow to control for the influence of the initial forest characteristic on the effect of defoliation, that has been proved to be non-negligible in past studies[34,38,42]. In each area, stations along a linear transect were previously established from highly impacted forest stands to undamaged stands[38,42]. In order to maximise the differences between defoliated and undamaged forests we selected the two stations at one extreme of the transect corresponding to defoliated forest, i.e. almost all tree stems dead or heavily damaged, and the two stations at the other extreme of the transect corresponding to undamaged forest, i.e. all trees alive, based on the damage-scores measured in[38,42] (Fig. 1). The two adjacent

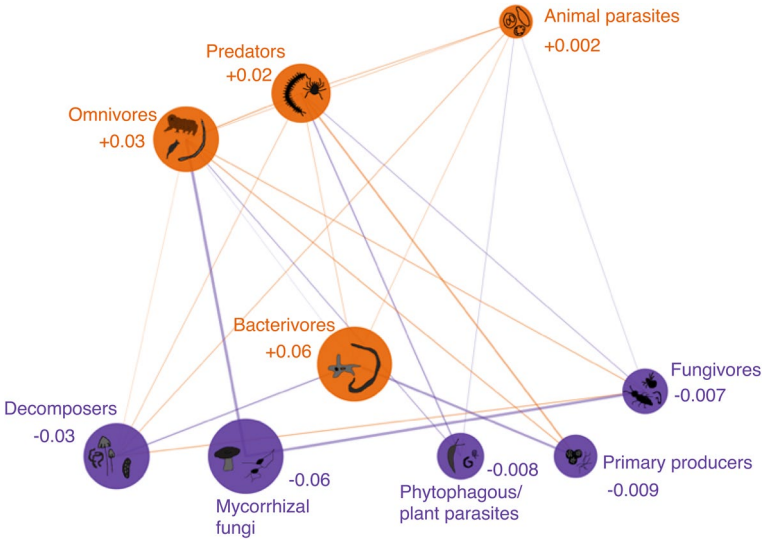


Figure 5. Structural differences among the local soil food webs from undamaged to defoliated forests at the trophic class resolution. Orange colour represents an increase, and purple colour a decrease in the relative abundance proportion within the local food webs of trophic classes (nodes) and link probability between classes (edges) from undamaged to defoliated forest. Relative abundance proportion corresponds to the sum of the relative abundances of the trophic groups inside the trophic class normalized within the local food web to sum one. Link probability corresponds to the probability of interaction between two classes given the links between their respective trophic groups and the relative abundances of these groups. The widths of the edges are scaled by the square root of the changes in link probability. Size of the nodes are proportional to the value of change in relative abundances proportion within the soil food web, indicated with numbers. Nodes are distributed vertically based on their trophic level from the bottom (basal levels) to the top (higher levels).

stations, separated from at least 2 km within defoliated or undamaged forests, were considered as local replicates and were surrounded by a large area of forest in their same condition, i.e. defoliated or undamaged. Defoliated and undamaged stations within an area were ca. 20 km apart. In July 2017, we sampled in each station 15 soil cores along an L-shaped transect with 10 m distances between neighbouring cores, corresponding to the biological replicates at the plot scale and aiming to account for microhabitat heterogeneity. This sampling design allowed to account for the local heterogeneity at different spatial scales (from meters to kilometres) and it was a good compromise for covering sufficiently local diversity across groups of varying spatial distributions[57], while already minimizing spatial autocorrelation as it has been shown for earthworms and bacteria (> 5 m between soil samples[58,59]). Soil corers were cleaned and flame sterilized between each sample collection. Extracellular DNA was then extracted from 15 g as described in[60,61]. Botanical surveys were conducted and consisted of annotating the species present in the vicinity (1 m^2) of each soil core.

Laboratory analyses. DNA extractions were conducted at the field on a mobile field unit. PCR, sequencing and soil physico-chemical analyses were performed at the Laboratoire d'Écologie Alpine (LECA) in Grenoble, France. Physicochemical soil properties were quantified from soil cores, including soil organic matter content (%), pH, soil moisture and C (%), N (%) and P content.

DNA extraction, PCR and sequencing negative controls were included in the experiment and used to identify potential contaminants and to control for false positives caused by tag-switching events. In order to set extracellular DNA (eDNA) free from clay and silica particles, each sample was rotatively shaken for 15 min in a 15 ml saturated phosphate buffer solution (Na_2HPO_4; 0.12 M; pH ≈ 8). Two ml of sediment/buffer mixture were then sampled and centrifuged for 10 min at 10,000 g. A 400 µl aliquot of supernatant was recovered and used as starting material for eDNA extraction using NucleoSpin® Soil extraction kit (Macherey–Nagel GmbH, Düren, Germany), following manufacturer's instructions except skipping the lysis cell step[60]. After elution, DNA extracts were diluted 10 times before being used as template for amplification. Eight negative extraction controls were also performed.

DNA amplification and sequencing. To assign the sequence reads to their relevant samples after high-throughput sequencing, we added unique eight base-long tags (with at least five differences between each other) to the 5' end of each primer (modified from[62,63]). DNA amplifications were carried out in a final volume of 20 µl containing 2 µl of DNA sample, 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 and 0.16 µl of Bovine Serum Albumin. A total of 10 PCR negative and six positive PCR controls were included. Each sample (including all controls) was amplified in quadruplicate. Eukaryotes, Fungi and Protists were targeted using the respective DNA markers: Euka02 (18S rRNA gene), Fung02 (ITS1) and Bact01 (16S rRNA gene) described in[20]. PCR thermo-

cycling conditions were as follow: after an initial step of 10 min at 95 °C, the mixtures underwent 45 cycles of 30 s at 95 °C, 30 s at 57–55–45 °C (Bact01, Fung02, Euka02, respectively) and 60 s at 72 °C, followed by a final elongation at 72 °C for 7 min. The amplification success was checked using capillary electrophoresis (QIAxcel System; Qiagen). PCR products were mixed in an equi-volume way (15 µl each) and 8 aliquots of 100 µl of the resulting mix were then purified using MinElute Purification kit (Qiagen GmbH, Hilden, Germany). Purified products were then pooled together before sequencing. This later was performed by pair-end sequencing on Illumina HiSeq 2000 platform (2*125 for Euka02, and 2*250 for both Bact01 and Fung02) at Fasteris, Geneva, Switzerland.

Bioinformatics. Sequences from the three libraries were pre-processed using the OBITools software[64]. Forward and reverse paired-end reads were assembled based on their overlapping 3'-end sequences, demultiplexed and dereplicated. We then removed sequences with low paired-end alignment scores, singletons, short sequences and sequences containing ambiguous bases, as well as PCR errors using the *obiclean* command. Molecular Operational Taxonomic Units were built by clustering sequences at 97% of similarity using SUMACLUST[65]. Taxonomic annotations were performed with the SILVAngs pipeline (Quast et al. 2013), using the SILVA version 132 for Bact02 and Euka01. For Fung02 and Euka01 (only metazoa), we used the *ecotag* command from the OBITools, and the EMBL database version 136. Taxonomic annotations with > 75% identities were retained. Cross-sample contaminations and reagent contaminants were removed on the basis of negative and empty controls, and dysfunctional PCRs were detected and removed following the procedures described in[66] with the metabaR R package[67,68]. For each marker, non-targeted taxa were eliminated. For Euka01 marker, we also excluded MOTUs identified as fungi, plants, and non-soil animals. After curation, PCR replicates were pooled together into samples. Only remaining common samples between the three MOTU tables were retained (n = 86). Number of reads, MOTUs, PCR replicates and samples before and after the curation process are available in Supplementary Table 2.

Soil food webs. Using current knowledge on soil organisms, we classified the MOTUs, based on their taxonomic annotations, into 9 broad trophic classes, using a classic soil food web backbone (e.g.[22,27]). These trophic classes included primary consumers, decomposers, phytophagous or plant parasites, mycorrhizal fungi, bacterivores, fungivores, omnivores, predators and animal parasites (Fig. 2a). Next, we defined 37 finer trophic groups by separating phylogenetic distant groups that could have a different set of prey/predators (e.g., bacterivore mites and bacterivore nematodes) or groups differing in their resources acquisition strategy (e.g. different types of mycorrhiza and saprotrophs). The definition on the trophic groups was made in accordance with the information available and the taxonomic resolution of the marker (Fig. 2b, Supplementary Fig. 1, Supplementary Table 1). For example, we kept collembola as a unique trophic group because the marker Euka02 was not resolutive enough to assign the MOTUs of this group to the family level, which was needed to a finer trophic classification. We kept both levels of resolution for the analyses, i.e., trophic class and trophic group. The databases used for the taxonomic assignment were FUNGuild[69] for fungi, FAPROTAX[70] for bacteria, NEMAguild[69] and Nemaplex (http://nemaplex.ucdavis.edu/) for nematodes, and the main references used included[71] for protists (and[72] for cercozoa), and[73] for heterotrophic bacteria (i.e. copiotrophic and oligotrophic classification). The main taxonomic clades composing the trophic classes and groups are in Supplementary Table 3. Specific criteria used to define the trophic classes and groups for each kingdom are in Supplementary Table 4. A table for each kingdom including the list of taxa, the trophic groups assignment, the taxonomic level of assignment and the references or databases used is available on Supplementary files.

The MOTU diversity of each trophic group was estimated per sample using the Shannon diversity (i.e. the exponential of the Shannon entropy) since this is a relevant measure for eDNA data[74]. In eDNA metabarcoding studies, changes in the abundance/biomass of an individual taxon may be inferred, in some extents, from changes in their relative abundances across samples, although this correspondence can be noised by different biological or technical factors (reviewed in[20]). However, some taxon can exhibit higher gene copies than others, making these changes in relative abundance more difficult to compare across groups contrary to other abundance standardized measures such as biomass. Relative abundances were thus estimated using a double-transformation. First, the total read counts of each trophic group were converted to proportions within a sample, and second, the resulting proportions were standardized by the largest observed proportion across all samples for each trophic group. Relative abundance of each group varied from 0 (absent) to 1 (largest observed diversity/proportion), allowing to obtain comparable measures across groups. Relative abundances of trophic classes were calculated by summing the relative abundances of the trophic group included in the trophic class[31].

The *metaweb*, which contains the potential trophic interactions of the soil food webs of the system under study, was built for trophic classes and trophic groups[45]. Trophic links between trophic groups and trophic classes were added based on the main feeding preferences. Some constraints were added when assigning the trophic links between trophic groups based on (1) the organisms size, i.e. predators fed only on smaller preys, with some exceptions like animal parasites and omnivore nematodes that can eat larger preys, and macroorganisms did not interact with microorganisms, (2) habitat differentiation, i.e. strict plant endoparasites (i.e. protists) were not considered as prey of other free living predators, and (3) feeding preferences, e.g. fungivores fed only on saprotrophic fungi and Ectomycorrhizal, which are preferred to arbuscular mycorrhizal fungi[75]. The complete *metaweb* of trophic groups can be found in Supplementary Fig. 1 and the adjacency matrix is available in Supplementary files. Resource nodes were added to the food web representations with a structural purpose and corresponded to the main resources of the soil food web, i.e., sunlight, organic matter and plants, but were excluded from the diversity analyses, because the aim was to quantify the diversity of organisms within the soil food web. Differences in resources and plant composition between undamaged and defoliated forests were evaluated

aside with multivariate analyses (see below). The *metaweb* was then used to characterize the composition and structure of the local soil food webs based on the trophic classes or groups detected locally in each soil sample (n = 86), assuming that classes or groups present locally interact as in the *metaweb*. For the local soil food webs at the trophic group resolution, nodes corresponded to the local relative abundance of the groups and links were binary (i.e., present or absent) assuming an interaction when the two groups concerned were present. For the trophic class resolution, nodes corresponded to the sum of the relative abundances of the trophic groups inside the trophic class and the links were weighted by the probability of interaction between two classes given the links between their respective trophic groups and the relative abundances of these groups as a proxy for the probability of an encounter[31].

Statistical analyses. Differences in resources and plant composition between undamaged and defoliated forests were evaluated with multivariate analyses. A correspondence analysis was run to evaluate the differences in plant community composition. Plant communities from undamaged forests were mostly associated with ericaceous dwarf shrubs such as *Empetrum nigrum* and *Vaccinium* spp., but also of other shrubs and herbs in Kirkenes, e.g., *Salix* sp., *Betula nana*, *Equisetum* sp. (Supplementary Fig. 2). In defoliated forests plant composition was more variable among samples, including several species of grass and herbs, such as the dominant *Avenella flexuosa*. For the soil physico-chemical characteristics that we measured, the first two axes of a Principal Component Analysis explained 74.7% of the variance. The first axis was related to soil organic matter (SOM) and the second axis was related to the litter quality (measured with the C/N ratio) and inversely to soil acidity (i.e. pH) (Supplementary Fig. 3). Samples from defoliated forests were related to higher values of SOM, C, N, P and pH and lower C/N values.

To assess the effect of moth defoliation on MOTU diversity and the relative abundance of the trophic groups, a multilevel linear model was applied separately to each trophic group using the function 'stan_lmer' from the R package Rstanarm[76] with the default priors. In each model, a fixed effect for defoliation was included as a dummy variable (0 corresponding to the undamaged forest and 1 to the defoliated forests). To account for the structure of the sampling design, i.e. soil cores clustered within stations and stations clustered within areas, we added a nested random term for stations within area to the intercept, where station was a factor with 8 levels and area a factor with 2 levels. Note that random factors allowed for borrowing information from each station and area, and that using a Bayesian approach led to non-zero estimates of area and station random effects, contrary to approaches using REML. Even if we suspected that the effect of defoliation could vary between the areas due to the contrasting habitat characteristics of Kirkenes and Tana, preliminary analyses showed that the effect was similar for both areas (i.e. the coefficient of the interaction between area and defoliation was small and 95% CI widely overlapped with 0 for most groups). MOTU diversity was standardized by the largest diversity observed across samples for each trophic group, to obtain comparable effect sizes across groups. A Yeo-Johnson transformation was applied to the relative abundances to improve the distribution of the residuals. Each model was run with 4 parallel MCMC chains with 15,000 iterations each. Model convergence was assessed visually and by checking Rhat < 1.10 for all the parameters. The normality of residuals was evaluated visually by using quantile–quantile (Q–Q) plots, and residuals were plotted against fitted values to assess outliers or influential values.

To study changes in the structure of local food webs, we estimated network diversity indices using the R package econetwork[31]. It allows computing several diversity indices on groups and link abundances using a viewpoint parameter that control the importance given to low *vs.* high relative abundances. We used a measure of dissimilarity of node and link compositions at different resolutions (trophic group and trophic class) to analyse whether there was a change in the structure of local soil food webs due to defoliation. A mixed multivariate distance matrix regression was then run using the dissimilarity matrix as the response, including a dummy variable for defoliation as a predictor and accounting for the nested sampling design as a random effect using the R package MDMR[77]. Local diversity (α-diversity) was estimated as the generalised mean of local diversity within each category of defoliation (i.e. defoliated forest and undamaged forest) within each area (Tana and Kirkenes). Both network local diversity and dissimilarity were computed using 1 as viewpoint parameter (eta in the package). Using this value of parameter, local diversity is the exponential of Shannon entropy. Figures 2, 3, 4, 5 were made using the R software (R 3.6.3)[68].

Received: 10 February 2021; Accepted: 6 July 2021
Published online: 23 July 2021

References

1. Pickett, S. T. A. & White, P. S. *The Ecology of Natural Disturbance and Patch Dynamics* (Academic Press, 1985).
2. IPBES. *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (IPBES Secretariat, 2019).
3. Brun, P. *et al.* Large-scale early-wilting response of Central European forests to the 2018 extreme drought. *Glob. Change Biol.* **00**, 1–15 (2020).
4. Cardinale, B. J., Gonzalez, A., Allington, G. R. H. & Loreau, M. Is local biodiversity declining or not? A summary of the debate over analysis of species richness time trends. *Biol. Conserv.* **219**, 175–183 (2018).
5. Vellend, M. *et al.* Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19456–19459 (2013).
6. Bardgett, R. D. & Wardle, D. A. *Aboveground-Belowground Linkages: Biotic Interactions, Ecosystem Processes, and Global Change* (Oxford University Press, 2010).
7. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
8. Bardgett, R. D. & Caruso, T. Soil microbial community responses to climate extremes: Resistance, resilience and transitions to alternative states. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190112 (2020).

9. Thom, D. & Seidl, R. Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests: Disturbance impacts on biodiversity and services. *Biol. Rev.* **91**, 760–781 (2016).
10. van der Putten, W. H. *et al.* Trophic interactions in a changing world. *Basic Appl. Ecol.* **5**, 487–494 (2004).
11. Lafferty, K. D. & Suchanek, T. H. Revisiting Paine's 1966 sea star removal experiment, the most-cited empirical article in the American Naturalist. *Am. Nat.* **188**, 365–378 (2016).
12. Scherber, C. *et al.* Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature* **468**, 553–556 (2010).
13. Barnes, A. D. *et al.* Direct and cascading impacts of tropical land-use change on multi-trophic biodiversity. *Nat. Ecol. Evol.* **1**, 1511–1519 (2017).
14. Barbier, M. & Loreau, M. Pyramids and cascades: A synthesis of food chain functioning and stability. *Ecol. Lett.* **22**, 405–419 (2019).
15. Mancinelli, G. & Mulder, C. Chapter three—detrital dynamics and cascading effects on supporting ecosystem services. In *Advances in ecological research* Vol. 53 (eds Woodward, G. & Bohan, D. A.) 97–160 (Academic Press, 2015).
16. Mulder, C., Vonk, J. A., Hollander, H. A. D., Hendriks, A. J. & Breure, A. M. How allometric scaling relates to soil abiotics. *Oikos* **120**, 529–536 (2011).
17. Allen, A. P. & Gillooly, J. F. Towards an integration of ecological stoichiometry and the metabolic theory of ecology to better understand nutrient cycling. *Ecol. Lett.* **12**, 369–384 (2009).
18. de Ruiter, P. C., Neutel, A.-M. & Moore, J. C. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260 (1995).
19. Estes, J. A. *et al.* Trophic downgrading of planet earth. *Science* **333**, 301–306 (2011).
20. Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. *Environmental DNA: For Biodiversity Research and Monitoring* (Oxford University Press, 2018).
21. Gravel, D., Albouy, C. & Thuiller, W. The meaning of functional trait composition of food webs for ecosystem functioning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150268 (2016).
22. Barnes, A. D. *et al.* Energy flux: The link between multitrophic biodiversity and ecosystem functioning. *Trends Ecol. Evol.* **33**, 186–197 (2018).
23. Elton, C. S. *Animal Ecology* 1–256 (Macmillan Co., 1927). https://doi.org/10.5962/bhl.title.7435.
24. Bohan, D. A. *et al.* Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends Ecol. Evol.* **32**, 477–487 (2017).
25. Roslin, T. & Majaneva, S. The use of DNA barcodes in food web construction—terrestrial and aquatic ecologists unite!. *Genome* **59**, 603–628 (2016).
26. Cohen, J. E. *et al.* Improving food webs. *Ecology* **74**, 252–258 (1993).
27. Buzhdygan, O. Y. *et al.* Biodiversity increases multitrophic energy use efficiency, flow and storage in grasslands. *Nat. Ecol. Evol.* **4**, 393–405 (2020).
28. Martinez, N. D. Effects of resolution on food web structure. *Oikos* **66**, 403 (1993).
29. Thompson, R. M. *et al.* Food webs: Reconciling the structure and function of biodiversity. *Trends Ecol. Evol.* **27**, 689–697 (2012).
30. Kardol, P., Throop, H. L., Adkins, J. & de Graaff, M.-A. A hierarchical framework for studying the role of biodiversity in soil food web processes and ecosystem services. *Soil Biol. Biochem.* **102**, 33–36 (2016).
31. Ohlmann, M. *et al.* Diversity indices for ecological networks: A unifying framework using Hill numbers. *Ecol. Lett.* **22**, 737–747 (2019).
32. Pellissier, L. *et al.* Comparing species interaction networks along environmental gradients. *Biol. Rev.* **93**, 785–800 (2017).
33. Jepsen, J. U. *et al.* Ecosystem impacts of a range expanding forest defoliator at the forest-tundra ecotone. *Ecosystems* **16**, 561–575 (2013).
34. Karlsen, S. R., Jepsen, J. U., Odland, A., Ims, R. A. & Elvebakk, A. Outbreaks by canopy-feeding geometrid moth cause state-dependent shifts in understorey plant communities. *Oecologia* **173**, 859–870 (2013).
35. Jepsen, J. U., Hagen, S. B., Ims, R. A. & Yoccoz, N. G. Climate change and outbreaks of the geometrids *Operophtera brumata* and *Epirrita autumnata* in subarctic birch forest: Evidence of a recent outbreak range expansion. *J. Anim. Ecol.* **77**, 257–264 (2008).
36. Vindstad, O. P. L., Jepsen, J. U., Ek, M., Pepi, A. & Ims, R. A. Can novel pest outbreaks drive ecosystem transitions in northern-boreal birch forest?. *J. Ecol.* **107**, 1141–1153 (2019).
37. Sandén, H. *et al.* Moth outbreaks reduce decomposition in subarctic forest soils. *Ecosystems* **23**, 151–163 (2019).
38. Vindstad, O. P. L. *et al.* Numerical responses of saproxylic beetles to rapid increases in dead wood availability following geometrid moth outbreaks in sub-arctic mountain birch forest. *PLoS ONE* **9**, e99624 (2014).
39. Nilsson, M.-C. & Wardle, D. A. Understory vegetation as a forest ecosystem driver: Evidence from the northern Swedish boreal forest. *Front. Ecol. Environ.* **3**, 421–428 (2005).
40. Bråthen, K. A. & Ravolainen, V. T. Niche construction by growth forms is as strong a predictor of species diversity as environmental gradients. *J. Ecol.* **103**, 701–713 (2015).
41. Bråthen, K. A., Gonzalez, V. T. & Yoccoz, N. G. Gatekeepers to the effects of climate warming? Niche construction restricts plant community changes along a temperature gradient. *Perspect. Plant Ecol. Evol. Syst.* **30**, 71–81 (2018).
42. Vindstad, O. P. L., Jepsen, J. U. & Ims, R. A. Resistance of a sub-arctic bird community to severe forest damage caused by geometrid moth outbreaks. *Eur. J. For. Res.* **134**, 725–736 (2015).
43. Parker, T. C. *et al.* Slowed biogeochemical cycling in sub-arctic birch forest linked to reduced mycorrhizal growth and community change after a defoliation event. *Ecosystems* **20**, 316–330 (2017).
44. Saravesi, K. *et al.* Moth outbreaks alter root-associated fungal communities in subarctic mountain birch forests. *Microb. Ecol.* **69**, 788–797 (2015).
45. Dunne, J. A. The network structure of food webs. In *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds Pascual, M. & Dunne, J. A.) 27–86 (Oxford University Press, 2006).
46. Rodriguez-Ramos, J. C. *et al.* Changes in soil fungal community composition depend on functional group and forest disturbance type. *New Phytol.* **00**, 1–13 (2020).
47. Decaëns, T. Macroecological patterns in soil communities. *Glob. Ecol. Biogeogr.* **19**, 287–302 (2010).
48. Bardgett, R. D., Yeates, G. W. & Anderson, J. M. Patterns and determinants of soil biological diversity. In *Biological Diversity and Function in Soils* (eds Hopkins, D. *et al.*) 100–118 (Cambridge University Press, 2005).
49. Worm, B. & Duffy, J. E. Biodiversity, productivity and stability in real food webs. *Trends Ecol. Evol.* **18**, 628–632 (2003).
50. Ponsard, S., Arditi, R. & Jost, C. Assessing top-down and bottom-up control in a litter-based soil macroinvertebrate food chain. *Oikos* **89**, 524–540 (2000).
51. Kristensen, J. Å., Rousk, J. & Metcalfe, D. B. Below-ground responses to insect herbivory in ecosystems with woody plant canopies: A meta-analysis. *J. Ecol.* **108**, 917–930 (2020).
52. González, V. T. *et al.* Batatasin-III and the allelopathic capacity of *Empetrum nigrum*. *Nord. J. Bot.* **33**, 225–231 (2015).
53. Veen, G. F. *et al.* The role of plant litter in driving plant-soil feedbacks. *Front. Environ. Sci.* **7**, 168 (2019).
54. Calizza, E., Rossi, L., Careddu, G., Sporta Caputi, S. & Costantini, M. L. Species richness and vulnerability to disturbance propagation in real food webs. *Sci. Rep.* **9**, 19331 (2019).
55. Antiqueira, P. A. P., Petchey, O. L., dos Santos, V. P., de Oliveira, V. M. & Romero, G. Q. Environmental change and predator diversity drive alpha and beta diversity in freshwater macro and microorganisms. *Glob. Change Biol.* **24**, 3715–3728 (2018).

56. Hedlund, K. *et al.* Trophic interactions in changing landscapes: Responses of soil food webs. *Basic Appl. Ecol.* **5**, 495–503 (2004).
57. Ettema, C. H. & Wardle, D. A. Spatial soil ecology. *Trends Ecol. Evol.* **17**, 177–183 (2002).
58. O'Brien, S. L. *et al.* Spatial scale drives patterns in soil bacterial diversity. *Environ. Microbiol.* **18**, 2039–2051 (2016).
59. Jiménez, J. J., Decaëns, T., Lavelle, P. & Rossi, J.-P. Dissecting the multi-scale spatial relationship of earthworm assemblages with soil environmental variability. *BMC Ecol.* **14**, 26 (2014).
60. Taberlet, P. *et al.* Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* **21**, 1816–1820 (2012).
61. Zinger, L. *et al.* Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. *Soil Biol. Biochem.* **96**, 16–19 (2016).
62. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**, e197 (2007).
63. Valentini, A. *et al.* New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. *Mol. Ecol. Resour.* **9**, 51–60 (2009).
64. Boyer, F. *et al.* OBITOOLS: A UNIX-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 176–182 (2016).
65. Mercier, C., Boyer, F., Bonin, A. & Coissac, E. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. in *Programs and Abstracts of the SeqBio 2013 workshop. Abstract* 27–29 (Citeseer, 2013).
66. Zinger, L. *et al.* DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* **28**, 1857–1862 (2019).
67. Zinger, L. *et al.* metabaR : an R package for the evaluation and improvement of DNA metabarcoding data quality. https://doi.org/10.1101/2020.08.28.271817 (2020).
68. R Core Team. *A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).
69. Nguyen, N. H. *et al.* FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **20**, 241–248 (2016).
70. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).
71. Adl, S. M. *et al.* Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).
72. Fiore-Donno, A. M. *et al.* Functional traits and spatio-temporal structure of a major group of soil protists (Rhizaria: Cercozoa) in a temperate grassland. *Front. Microbiol.* **10**, 1332 (2019).
73. Ho, A., Lonardo, D. P. D. & Bodelier, P. L. E. Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol. Ecol.* **93**, 6 (2017).
74. Calderón-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L. & Thuiller, W. From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices?. *J. Biogeogr.* **47**, 193–206 (2020).
75. Antunes, P. M. & Koyama, A. Chapter 9 - Mycorrhizas as Nutrient and Energy Pumps of Soil Food Webs: Multitrophic Interactions and Feedbacks. in *Mycorrhizal Mediation of Soil Fertility, Structure, and Carbon Storage* (eds. Johnson, N. C., Gehring, C. & Jansa, J.) 149–173 (Elsevier, 2017).
76. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. *rstanarm: Bayesian applied regression modeling via Stan*. (R package version 2.21.1, 2020).
77. McArtor, D. B., Lubke, G. H. & Bergeman, C. S. Extending multivariate distance matrix regression with an effect size measure and the asymptotic null distribution of the test statistic. *Psychometrika* **82**, 1052–1077 (2017).

Acknowledgements

This research was funded by the French Agence Nationale de la Recherche (ANR) through the GlobNet (ANR-16-CE02-0009) and EcoNet (ANR-18-CE02-0010) projects. We thank Clement Lionnet and Nicolas Leguillarme for support with the development of bioinformatic pipelines, Cindy Arnoldi for the soil physicochemical analyses and Julien Renaud for providing the map in Fig.1b. Jane Jepsen et Ole Petter Vindstad contributed to the study design and information about the defoliation and ecosystem changes.

Author contributions

W.T. and N.G.Y. conceptualize the overall idea, and together with I.C.S. and T.M. conceived the research aims and analyses. L.Z., H.S., L.G. and N.G.Y. conducted the field sampling and L.G. ran the laboratory procedures. I.C.S., L.Z., H.S., M.H. and S.S.M. conducted the trophic data collection. I.C.S. conducted the bioinformatics and statistical analyses and led the writing with significant contributions from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94227-z.

Correspondence and requests for materials should be addressed to I.C.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

GENERAL DISCUSSION

One of the greatest challenges in ecology is to integrate the enormous diversity of living organisms that inhabit the planet into the existing ecological theories. Only with this integration we can gain a more general understanding of the factors that shape and structure ecological communities. This is necessary if we are to make accurate predictions about the effects of global changes and adapt our conservation strategies to address the current biodiversity crisis. However, due to its inherent complexity, a large part of the biodiversity that is contained in soils has been overlooked in macroecological studies. Nowadays, with the development of new technologies such as eDNA metabarcoding and the accumulation of knowledge on the functions and interactions of soil organisms from decades of research, it is possible to study diversity patterns of soil organisms across large scales and/or across a variety of taxa distributed across the multiple trophic levels characterizing terrestrial ecosystems.

This PhD thesis aimed at improving our understanding of how soil biodiversity responds to environmental changes through the use of eDNA metabarcoding. The objective was carried out (1) by critically assessing the methodology necessary to obtain reliable results knowing the biases of the eDNA metabarcoding approach, and by developing an integrative approach to simplify the complexity of all soil organisms and include their trophic or functional linkages; (2) by carrying out ecological research to answer topical questions and test ecological hypotheses originally developed on aboveground taxa. In the following discussion, I summarize the contributions of this PhD to our state of knowledge on soil biodiversity in two sections: a first one discussing the methodological contributions for the study of soil biodiversity, and a second one suggesting how these results can be incorporated into the current knowledge on soil ecology.

1. Methodological contributions of my thesis.

When I started this PhD, I realized how much confusion there could be on the conception of the bioinformatic pipeline to deal with the biases associated with the eDNA metabarcoding, especially for ecologists like me to whom this method was primarily a tool and not a research field in itself. Despite the enormous amount of literature available on the subject, there was still no clear guideline on which curation steps should be included in the pipeline and how these choices could influence the results of common ecological analyses used in macroecology or community ecology. In response to this issue, I conducted the research developed in Chapter 1 in collaboration with colleagues with specific expertise on laboratory and/or bioinformatic work

to treat eDNA data (Calderón-Sanou *et al.* 2020). The idea was to provide a concise roadmap for ecologists interested in studying biodiversity patterns using eDNA metabarcoding. At the end of my PhD, I believe that this research was a substantial contribution to the target audience, based on the high citing rate the paper received (51 citations in April 2022 in Google Scholar). The main conclusion of this study was that reliable and robust ecological results can be obtained when using both a stringent bioinformatic pipeline (described in Chapter 1) and the use of diversity measures that allow to weight the importance given to rare sequences (Fig. 5 in chapter 1) (Hill numbers, Hill 1973). Specifically, we found that Shannon diversity gave robust results to the bioinformatic pipeline and led to similar results than conventional sampling methods (but we only tested this for plants). Recent studies also advocate the use of Hill numbers in eDNA metabarcoding studies. For example, Mächler and colleagues (2021) conducted a similar study on freshwater ecosystems and also concluded that the combined use of a stringent treatment and Hill numbers with q=1 or q=2, i.e., Shannon and Simpson diversity, lead to more robust results and recommended its use to increase comparability across studies using eDNA data. Alberdi and Gilbert (2019) also proposed a detailed guideline for the use of Hill numbers in different types of studies using DNA sequencing-based techniques to estimate diversity. The methods to account for eDNA metabarcoding biases are constantly being improved, and the efficiency and utility of eDNA metabarcoding to monitor biodiversity are increasingly recognized (Coissac *et al.* 2012; Deiner *et al.* 2017; Ruppert *et al.* 2019). While MOTUs diversity estimated from eDNA cannot be expected to give exactly the same results as species diversity estimated from morphological assessments, our results suggest that similar trends can be found if data are cleaned with care and adequate diversity metrics are used. Yet, richness estimates can be very sensitive to the curation strategy and give high estimates relative to conventional methods, thus we recommended the use of additional controls and analysis to ensure the reliability of this metric. The effort from both conventional soil diversity monitoring and eDNA metabarcoding should be unified to improve our knowledge of soil biodiversity patterns and their drivers (Orgiazzi *et al.* 2015; Bastida *et al.* 2020; Phillips *et al.* 2020).

Another methodological contribution of this thesis was the framework used for the construction of soil food webs from eDNA data. Most studies analyzing large-scale biodiversity patterns with eDNA focus on overall biodiversity of taxa including broad taxonomic groups (e.g., Wu *et al.* 2011; Drummond *et al.* 2015; George *et al.* 2019) or on functional groups within a single taxon (e.g., Tedersoo *et al.* 2014; van den Hoogen *et al.* 2019). Here, we proposed to analyze eDNA data by grouping all soil taxa through their ecological similarity and by structuring these

groups based on their feeding relations in food webs. Heuristic food webs allow for an integrative and ecological representation of soil multi-trophic assemblages in terrestrial ecosystems (Thompson *et al.* 2012), but their use has been limited by the difficult task of sampling entire multi-trophic communities and collecting trophic and/or functional information on their components. This is true for aboveground organisms (but see Maiorano *et al.* 2020) as well but even more for belowground organisms, of which most are cryptic and difficult to observe. Heuristic food web construction from DNA-based data has already been implemented by Compson and colleagues (2018), who combined DNA metabarcoding data with a text mining routine to extract trait information from the literature to construct food webs of freshwater benthic macroinvertebrates in conjunction with a river system. To my knowledge, this DNA-based approach has not yet been applied to soil communities, with the exception of the papers from this thesis and Bloor and colleagues. (2021), who used a similar approach that the one I used, but classified the MOTUs into a large number of trophic groups with links and then used a stochastic block model to simplify the food web. In the heuristic food webs of Compson and colleagues (2018), nodes were represented by genera and only included macroinvertebrates. In the case of soil food webs, if we want to integrate very distant phylogenetic groups (e.g. bacteria and eukaryotes, unicellular and multicellular organisms), trophic groups might be a better option to avoid extremely complex food webs and redundancy in interactions (Scheu 2002; Potapov 2022). Furthermore, since the trophic information available for some groups is at a lower taxonomic resolution (e.g., bacterivores), the use of fine taxa as nodes would lead to repeated information (or trophic redundancy) that could bias the results. Recently, Blackman and colleagues (2022) also implemented heuristic food web construction from eDNA data by assigning MOTUs to trophic groups based on the literature, and applied it to study changes in the food web structure of freshwater communities across temporal and spatial gradients.

The methodology used in this thesis allowed the construction of a metaweb for the different study systems and at different resolutions. The use of different resolutions allowed us to assess the compositional and structural spatial variation of soil food webs while accounting for potential trophic redundancy. In Chapter 4, an interesting finding was that the coarser resolution of the metaweb constituted a common skeleton for soil trophic networks, although there were small variations in group abundance and interactions between different local communities. In contrast, trophic networks at the finest resolution showed strong variability across different local communities. This variability reflects that from one locality to another entire fine groups can be reduced drastically or ‘go extinct’, which means that there is no replacement of taxa.

However, at the coarser resolution, we observe that trophic redundancy across groups makes that at least one group per trophic class is present, buffering the variability and potentially the effects of this variability on the functioning of the soil food web.

2. Discussion on the contribution of this thesis to the general knowledge of soil biodiversity patterns

In addition to providing a conceptual and comprehensive framework for studying soil multi-trophic communities, this thesis applied this framework to answer specific ecological questions: 1) do the ecological hypotheses on drivers of biodiversity gradients largely tested on aboveground macro-organisms hold for soil organisms? (Chapter 3) 2) at which extent the structure of soil food webs varies along environmental gradients? (Chapter 4) 3) Do the effects of moth outbreaks cascade into the soil food web? (Chapter 5). In this section I discuss how the results addressing these questions can be integrated to our general knowledge on soil biodiversity patterns, while highlighting specific limitations and future perspectives.

The results of this thesis support the general knowledge stating that local diversity is driven by energy input (Evans *et al.* 2005) and stress, i.e., resource availability and environmental harshness (Grime 1973; Huston 1979; Decaëns 2010). I found that energy, in particular the organic matter content of soils and plant biomass, and environmental harshness explained to a large extent the spatial structuring of the diversity of soil trophic groups, at both local and regional scales. The effect of energy and environmental harshness was explicitly tested in chapter 3, where we found a strong effect of NDVI, soil organic matter and pH on most soil trophic groups diversity. Moreover, in Chapter 5, we showed that an increase in the productivity of the system through plant composition turnover and nutrient enrichment following moth outbreaks lead to an increase in the diversity of soil organisms across the whole food web. Ongoing global changes can be responsible for changes in soil resources availability and the abiotic soil environment, and the fundamental question how these changes would cascade across trophic levels in both the belowground and aboveground compartments remains (Bardgett & van der Putten 2014; Thakur 2020). The results of this thesis suggest that changes in basal resources might influence the diversity not only of basal groups, but also of higher trophic levels, with similar strength (Fig. 3 in chapter 3, Fig. 3 in chapter 5), although a more mechanistic understanding is needed to provide precise predictions (Barbier & Loreau 2019). Moreover, we found that not only taxon diversity but the whole structure of the food web was

influenced by environmental change (Fig. 1 in chapter 4, Fig. 5 in chapter 5). This could translate into changes in the ecosystem functions and/or the stability of the soil communities that are needed to be further explored (Thompson *et al.* 2012; Eisenhauer *et al.* 2019; Zhao *et al.* 2019).

While our results suggest that soil biodiversity follows general ecological rules, the specific environmental drivers of soil biodiversity were not necessarily the same as those of aboveground macro-organisms, and were variable among soil trophic groups. In general, climatic variables, which are the main drivers used to explain spatial variation in aboveground diversity (Currie *et al.* 2004; Clarke & Gaston 2006; Braga *et al.* 2019), were less important in soil, at least their direct effect, which supports previous studies on soil diversity (Ramirez *et al.* 2014; Karimi *et al.* 2018; Caruso *et al.* 2019). Instead, plant community characteristics such as biomass, composition, diversity and soil properties such as pH, organic matter and C/N ratio were better predictors of soil diversity change. As expected, responses to environmental factors varied between soil trophic groups (Fig. 2,3 in chapter 3, Fig. 3 in chapter 5), and this was mainly related to differences in resource requirements of soil biota, e.g., plant-based versus detritus-based resources, or to different sensitivity to the abiotic environment, e.g., pH had a strong influence on organisms inhabiting water films. Changes in plant communities and soil properties not only led to changes in the diversity of soil organisms within trophic levels, but also in the structure of the soil food web. The relative importance of environmental factors influencing the local diversity of the soil food web (alpha diversity) was different from that influencing soil food web turnover (beta diversity). Interestingly, pH was an important factor limiting food web diversity, in terms of trophic groups and trophic interactions (side analysis in chapter 4), which is expected to be more limited by energy constraints or resource availability (Baiser *et al.* 2019). In contrast, plant communities and soil C/N ratio explained soil food webs turnover (Fig. 3 in chapter 4). Further analyses are needed to better understand these variations in structure and composition in terms of the group and interaction identity responsible for food web turnover across these environmental gradients. It is also important to recognize that all the biotic and abiotic factors considered interact and influence each other in complex ways. For example, the spatial variability of NDVI, one of the main predictors of soil biodiversity according to the results of this thesis, is influenced by climatic variables related to water availability and temperature (Choler 2015), so the indirect role of climate on soil biodiversity cannot be neglected (Bardgett & van der Putten 2014; Martinez-Almoyna *et al.* 2019). Similarly, pH is a complex variable that depends on several factors, such as the type of soil

bedrock and the composition of the plant community (Roy *et al.* 2013). In this thesis, I was particularly interested in understanding the relative influence of these different variables on the diversity of soil food webs. Indeed, identifying the main environmental predictors of soil biodiversity is a key step in predicting the consequences of global changes in biodiversity (see 'General perspectives' section).

Another important contribution of this thesis was the study of the variability of soil food webs along broad environmental gradients (chapter 4) or in response to disturbances (chapter 5). The variation of soil food webs along large spatial scales and in natural systems is scarce in the literature, and our results bring new insights in unraveling the influence of the environment on soil food web structure. An interesting avenue would be the study of how trophic interactions drive the spatial structure of the trophic groups. This could be done by analyzing how the diversity of a given trophic group or class depends on the diversity of the other groups or classes (both alpha diversity and beta diversity). This was partially done in chapter 3, when I tested how the diversity of a trophic group responded to the diversity of its resource ('Resource Heterogeneity Hypothesis', Fig. 1 in chapter 3). Yet, we could also look at how the turnover or local diversity of one group is related to the turnover or local diversity of the other trophic groups, to reveal spatial co-dependencies between the different components of the soil food web. Ohlmann and colleagues (2018) approached this question using a probabilistic graphical model (graphical lasso), which allows identifying partial correlations between soil trophic groups across samples while accounting simultaneously for the effect of the abiotic environment. Their method could be extended to the trophic groups built from eDNA data to infer a network of partial correlations. Further, the inferred network could be compared to the heuristic food web (the metaweb) to assess which trophic interactions have an impact in driving the spatial structure of soil food webs.

Finally, I found necessary to mention that a major limitation of the eDNA metabarcoding method to study the drivers of soil biodiversity is its limited ability to estimate abundances. Yet, many ecological theories state that the effect of the environment on diversity is driven by changes in abundances (e.g. the more individual hypothesis, Wright 1983). Thus, eDNA metabarcoding opens the door to test the predictions of ecological theories at large scales and for a broad range of organisms, but testing some mechanisms of these hypotheses needs complementary sampling methods or experimental setups. Abundance information would be crucial not only for testing ecological hypotheses but also to make a direct link between food web structure and ecosystem functioning. For example, the energy flux framework proposed

by Barnes and colleagues (2018) allows to do this by combining metabolic scaling theory and food-web energy dynamics, but needs a measure of abundance across the trophic groups. The principle is to calculate metabolic rates that are dependent on body mass, environmental temperature and phylogenetic grouping, combined with resource-specific assimilation efficiencies and energy loss to predation, to obtain energy flux across the trophic levels of a food web, and use it as a unified measure of multitrophic ecosystem functioning (Barnes *et al.* 2014, 2018). However, is it realistic to obtain the abundance of all the soil trophic groups, from microorganisms to macrofauna, at large spatial scales? For now, it is not, or at least it depends on abundance data gathered from multiple sources, which could present biases from heterogeneous sampling methods or not be resolutive enough at the taxonomic level to couple with our eDNA data. Otherwise, an approach used to quantify the abundance of species through DNA detection is the quantitative PCR, in which the number of copies of a target DNA marker are directly quantified. However, this method relies on species-specific DNA markers, and thus is limited to one or a small set of species. We are thus far for being able to estimate cross-kingdom organism abundances with DNA-based methods, but smaller scale studies in complement with large-scale assessments could allow to link soil food webs retrieved from eDNA metabarcoding with abundances of some key soil biota to related to ecosystem functions (this is further discussed in the 'General perspectives' section). Future methodological advances may improve the way we interpret abundances from eDNA (Shelton *et al.* 2022), although it would take time until we can relate the abundance of DNA found in the environment with the real abundance of organisms, especially for multicellular organisms.

3. General perspectives

Some perspectives were already discussed in the previous sections of the discussion, but here I would like to expand first, on what I considered the next logical step of the analyses to be done following this thesis, and second on a potential avenue to improve our mechanistic understanding on how global changes will affect terrestrial ecosystems.

From patterns to predictions and conservation strategies

The study of diversity patterns has essential applications today, such as the implementation of predictive models in the face of ongoing global changes, and the improvement of conservation strategies. These applications were beyond the scope of this thesis, but my results may provide

some clues for future studies. Interestingly, we were able to explain a large part of the variation in diversity of most trophic groups across the French Alps with only a few factors that are easy to measure and extrapolate at large scales, such as pH and NDIV (Fig. 2a in chapter 3). This gives good prospects for modeling major groups of soil biota from eDNA data, which is necessary if one wants to extrapolate to create maps or predict the consequences of environmental changes. Yet, the occurrences of some organisms, such as insects, were very low in the samples, signaling the need to improve the sampling for these organisms by expanding the sampling area or sampling effort, or complementing the sampling with other methods such as pitfalls. I acknowledge that making spatial or temporal predictions would require testing different existing predictive models and better assessing their predictive capacities (e.g., cross validation), and although it is out of my personal expertise it should be easily done from the data produced in this thesis (on-going work in the team). Contrary, the variance on the food web structure explained by the environment was low (less than 20% in chapter 4), questioning the ability to predict soil biodiversity at this level of organization. However, joint models could be used to jointly predict the local diversity or abundance of trophic groups (Pollock *et al.* 2014; but see Poggiato *et al.* 2021), and local food webs could be deduced next. Also, joint modelling of groups could be useful to increase the ability to predict the diversity or abundance or groups with low occurrences such as insects. If we were able to predict the spatial distribution of the trophic groups we could predict biodiversity change face to potential scenarios, but we could also use the predicted maps to assess the conservation status of soil organisms and their coverage within the existing protected areas (Thuiller *et al.* 2014; O'Connor *et al.* 2021). Indeed, most protected areas are based on aboveground habitat types, and little knowledge exist on how these areas are effective in protecting soil biodiversity (Cameron *et al.* 2019; Ciobanu *et al.* 2019). It is thus crucial to map soil food webs on specific regions, e.g., the French Alps, to assess the protected status of soil biodiversity and inform stakeholders. Food web are being more and more considered in conservation strategies with the goal of not only conserving species but also their interactions and ecosystem functions (Harvey *et al.* 2017).

Towards a more mechanistic understanding on the effect of global change on soil food webs

In this thesis, I studied the drivers of soil biodiversity using a space-for-time substitution approach. However, a more mechanistic understanding is needed in order to improve our predictions on the consequences of global change on soil biodiversity and ecosystem functions. As a consequence of ongoing climate change but also annual variability and extreme events,

assemblages and resulting food webs are expected to experience temporal dynamics that could be transient or not (Ryo *et al.* 2019). A conceptual framework on the many direct effects of warming on specific soil trophic or functional groups and specific ecosystem processes have emerged in the last decades (e.g., Bardgett & Caruso 2020; Zhou *et al.* 2020). However, we still know little about the cascading effects across the soil food webs and ecosystem functions, and how these will develop over time when an entire ecosystem is warmed in the field (Schwarz *et al.* 2017). Understanding these transient dynamics is of major importance for predicting net effects of warming on the different ecosystem compartments, but also on their recovery and resistance to land use changes happening in parallel in a world where a continuously changing climate prevents ecosystems from settling in stable states (Bardgett & Caruso 2020). Applying the methods developed in this thesis to experimental setups simulating warming in the field, such as transplant experiments where both the below and aboveground compartments are transplanted (Bektaş *et al.* 2021). In addition, at this scale it would be feasible to obtain abundance data for soil organisms, which would allow us to go further into the dynamics of soil communities in the face of warming and to link networks to functions. This is something I am going to work on during a 2-y postdoc starting in September, where I aim to contrast my previous results to results obtained from a transplant experiments carried out along an Orchamp gradients.

CONCLUSION

This thesis provides a new key to deepen our understanding on soil biodiversity and its drivers. Through methodological development with the aim of improving the use of eDNA metabarcoding data to have a more comprehensive view of soil multi-trophic communities, I believe it provides an important piece of knowledge in our understanding on soil ecology. To move forward, we would need to advance in parallel in technical upgrades of eDNA metabarcoding data processing and interpretation, the conceptual integration of soil biodiversity into ecological networks, and the theoretical development of ecological models allowing us to integrate this information to get a more mechanistic understanding of soil biodiversity.

REFERENCES OF THE DISCUSSION

Alberdi, A. & Gilbert, M.T.P. (2019). A guide to the application of Hill numbers to DNA-based diversity analyses. *Molecular Ecology Resources*, 19, 804–817.

Baiser, B., Gravel, D., Cirtwill, A.R., Dunne, J.A., Fahimipour, A.K., Gilarranz, L.J., *et al.* (2019). Ecogeographical rules and the macroecology of food webs. *Global Ecology and Biogeography*, 28, 1204–1218.

Barbier, M. & Loreau, M. (2019). Pyramids and cascades: a synthesis of food chain functioning and stability. *Ecol Lett*, 22, 405–419.

Bardgett, R.D. & Caruso, T. (2020). Soil microbial community responses to climate extremes: resistance, resilience and transitions to alternative states. *Philos Trans R Soc Lond B Biol Sci*, 375, 20190112.

Bardgett, R.D. & van der Putten, W.H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, 515, 505–511.

Barnes, A.D., Jochum, M., Lefcheck, J.S., Eisenhauer, N., Scherber, C., O'Connor, M.I., *et al.* (2018). Energy Flux: The Link between Multitrophic Biodiversity and Ecosystem Functioning. *Trends Ecol Evol*, 33, 186–197.

Barnes, A.D., Jochum, M., Mumme, S., Haneda, N.F., Farajallah, A., Widarto, T.H., *et al.* (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat Commun*, 5, 5351.

Bastida, F., Eldridge, D.J., Abades, S., Alfaro, F.D., Gallardo, A., García-Velázquez, L., *et al.* (2020). Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Mol Ecol*, 29, 752–761.

Bektaş, B., Thuiller, W., Saillard, A., Choler, P., Renaud, J., Colace, M.-P., *et al.* (2021). Lags in phenological acclimation of mountain grasslands after recent warming. *Journal of Ecology*, 109, 3396–3410.

Blackman, R.C., Ho, H.-C., Walser, J.-C. & Altermatt, F. (2022). Spatio-temporal patterns of multi-trophic biodiversity and food-web characteristics uncovered across a river catchment using environmental DNA. *Commun Biol*, 5, 1–11.

Bloor, J.M.G., Si-Moussi, S., Taberlet, P., Carrère, P. & Hedde, M. (2021). Analysis of complex trophic networks reveals the signature of land-use intensification on soil communities in agroecosystems. *Sci Rep*, 11, 18260.

Braga, J., Pollock, L.J., Barros, C., Galiana, N., Montoya, J.M., Gravel, D., *et al.* (2019). Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe. *Glob Ecol Biogeogr*, 28, 1636–1648.

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? *J Biogeogr*, 47, 193–206.

Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., *et al.* (2019). Global mismatches in aboveground and belowground biodiversity. *Conservation Biology*, 33, 1187–1192.

Choler, P. (2015). Growth response of temperate mountain grasslands to inter-annual variations in snow cover duration. *Biogeosciences*, 12, 3885–3897.

Ciobanu, M., Eisenhauer, N., Stoica, I.-A. & Cesarz, S. (2019). Natura 2000 priority and non-priority habitats do not differ in soil nematode diversity. *Applied Soil Ecology*, 135, 166–173.

Clarke, A. & Gaston, K.J. (2006). Climate, energy and diversity. *Proc Biol Sci*, 273, 2257–2266.

Compson, Z.G., Monk, W.A., Curry, C.J., Gravel, D., Bush, A., Baker, C.J.O., *et al.* (2018). Chapter Two - Linking DNA Metabarcoding and Text Mining to Create Network-Based Biomonitoring Tools: A Case Study on Boreal Wetland Macroinvertebrate Communities. In: *Advances in Ecological Research*, Next Generation Biomonitoring: Part 2 (eds. Bohan, D.A., Dumbrell, A.J., Woodward, G. & Jackson, M.). Academic Press, pp. 33–74.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., *et al.* (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol Lett*, 7, 1121–1134.

Decaëns, T. (2010). Macroecological patterns in soil communities. *Glob Ecol Biogeogr*, 19, 287–302.

Drummond, A.J., Newcomb, R.D., Buckley, T.R., Xie, D., Dopheide, A., Potter, B.C., *et al.* (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4.

Eisenhauer, N., Schielzeth, H., Barnes, A.D., Barry, K.E., Bonn, A., Brose, U., *et al.* (2019). A multitrophic perspective on biodiversity–ecosystem functioning research. *Adv Ecol Res.*, 61, 1–54.

Evans, K.L., Warren, P.H. & Gaston, K.J. (2005). Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biol Rev*, 80, 1–25.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., *et al.* (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, 10, 1107.

Grime, J.P. (1973). Competitive Exclusion in Herbaceous Vegetation. *Nature*, 242, 344–347.

Harvey, E., Gounand, I., Ward, C.L. & Altermatt, F. (2017). Bridging ecology and conservation: from ecological networks to ecosystem function. *J Appl Ecol*, 54, 371–379.

Hill, M.O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54, 427–432.

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

Huston, M. (1979). A General Hypothesis of Species Diversity. *The American Naturalist*, 113, 81–101.

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, 3326–3339.

Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods. *Global Ecol Biogeogr*, 29, 1452–1457.

Martinez-Almoyna, C., Thuiller, W., Chalmandrier, L., Ohlmann, M., Foulquier, A., Clément, J., *et al.* (2019). Multi-trophic β-diversity mediates the effect of environmental gradients on the turnover of multiple ecosystem functions. *Funct Ecol*, 33, 2053–2064.

O'Connor, L.M.J., Pollock, L.J., Renaud, J., Verhagen, W., Verburg, P.H., Lavorel, S., *et al.* (2021). Balancing conservation priorities for nature and for people in Europe. *Science*, 372, 856–860.

Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L., *et al.* (2018). Mapping the imprint of biotic interactions on β-diversity. *Ecology Letters*, 21, 1660–1669.

Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A. & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biology and Biochemistry*, 80, 244–250.

Phillips, H.R.P., Heintz-Buschart, A. & Eisenhauer, N. (2020). Putting soil invertebrate diversity on the map. *Molecular Ecology*, 29, 655–657.

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J.S. & Thuiller, W. (2021). On the Interpretations of Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, 36, 391–401.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.

Potapov, A.M. (2022). Multifunctionality of belowground food webs: resource, size and spatial energy channels. *Biological Reviews*, n/a.

Roy, J., Albert, C., Choler, P., Clément, J.-C., Ibanez, S., Lavergne, S., *et al.* (2013). Microbes on the Cliff: Alpine Cushion Plants Structure Bacterial and Fungal Communities. *Front Microbiol*, 4, 64.

Ryo, M., Aguilar-Trigueros, C.A., Pinek, L., Muller, L.A.H. & Rillig, M.C. (2019). Basic Principles of Temporal Dynamics. *Trends in Ecology & Evolution*, 34, 723–733.

Scheu, S. (2002). The soil food web: structure and perspectives. *European Journal of Soil Biology*, 38, 11–20.

Schwarz, B., Barnes, A.D., Thakur, M.P., Brose, U., Ciobanu, M., Reich, P.B., *et al.* (2017). Warming alters energetic structure and function but not resilience of soil food webs. *Nature Clim Change*, 7, 895–900.

Shelton, A.O., Gold, Z.J., Jensen, A.J., D'Agnese, E., Allan, E.A., Cise, A.V., *et al.* (2022). Toward quantitative metabarcoding.

Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N.S., Wijesundera, R., *et al.* (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688–1256688.

Thakur, M.P. (2020). Climate warming and trophic mismatches in terrestrial ecosystems: the green–brown imbalance hypothesis. *Biol Lett*, 16, 20190770.

Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., *et al.* (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol Evol*, 27, 689–697.

Thuiller, W., Guéguen, M., Georges, D., Bonet, R., Chalmandrier, L., Garraud, L., *et al.* (2014). Are different facets of plant diversity well protected against climate and land cover changes? A test study in the French Alps. *Ecography*, 37, 1254–1266.

Wright, D.H. (1983). Species-Energy Theory: An Extension of Species-Area Theory. *Oikos*, 41, 496–506.

Wu, T., Ayres, E., Bardgett, R.D., Wall, D.H. & Garey, J.R. (2011). Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences*, 108, 17720–17725.

Zhao, Q., Brink, P.J.V. den, Carpentier, C., Wang, Y.X.G., Rodríguez-Sánchez, P., Xu, C., *et al.* (2019). Horizontal and vertical diversity jointly shape food web stability against small and large perturbations. *Ecology Letters*, 22, 1152–1162.

Zhou, Z., Wang, C. & Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nat Commun*, 11, 3072.

AKNOWLEDGMENTS

En primer lugar quisiera agradecer a mi familia, Fausto, Ofelia y Laura por su apoyo durante todos estos años de estudio. Esta tesis les está dedicada y fue posible gracias a su apoyo y transmisión. Gracias por motivarme a estudiar lo que me apasionara, por inspirarme a siempre buscar nuevos horizontes con una visión interdisciplinaria. Gracias por impulsarme a tener una mente abierta y curiosa, un corazón dedicado y dirigido a los demás, por enseñarme el respeto y la tolerancia, por darme todo para crecer en una atmósfera de abundancia y amor.

Je remercie mes superviseurs Wilfried et Tamara pour leur confiance, transmission et soutien, et pour m'avoir permis de apprendre et développer mes compétences de chercheur dans une atmosphère de bienveillance, coopération et haute qualité scientifique.

Mercie aussi à Lucie Zinger et Mickael Hedde pour être là tout le long de ma thèse comme des superviseurs aussi et m'apprendre ce que je sais maintenant sur l'écologie des sols. Également merci aux autres chercheurs de Globnet , Heidy, Mélanie, Julian, Anne-Sophie, et tous les autres avec qui j'ai interagi et qui m'ont beaucoup appris.

Je remercie tous mes collègues de travail pour leur soutien. Les amis doctorants ou postdocs, Camille, Marc, Louise, Billur, Giovanni, Xavier, Manu, Marie, Martin, Pierre, Christophe, Gemma, Laura, Marta, Ceres, Joao et tous les autres pour créer une atmosphère de soutien collectif, d'entraide, de respect et de partage, favorable au développement en académique mais aussi personnel. Les amis et collègues ingénieurs, Juju, Maya, Amélie, Clément, Nicolas, Fred et Sara pour leur patience et pour aider à résoudre tout un tas de problèmes et questions tout au long de la thèse (et pour les apéros spontanés et nécessaires dans le bureau 205 ou au bar). Aux chercheurs, Arnaud, Stéphane, Laure, Flo pour leurs conseils amicales et transmission de connaissances. Merci à Christian, Ludo, Cindy et tous les autres qui ont participé aux manips qui ont donné lieu aux données utilisées dans cette thèse.

Merci à mes sœurs Camille et Louise pour ses beaux moments de partage dans la coloc Calousir de l'amitié et l'amour qui restera dans mes bons souvenirs comme une étape fleuri, en couleurs, musicale, éphémère, émotionnelle, changeante, de sororité, avec larmes et rires et plein des cœurs…

Merci à tous les amis pour exister !

Merci à Théophile pour son amour, soutien à tout moment, patience, bienveillance, compréhension, et pour me donner tous les ingrédients nécessaires pour finaliser ce projet de thèse dans les meilleurs de conditions. Merci au petit Jafar pour absorber tout le stress de fin de thèse et nettoyer l'atmosphère avec ces ronronnements.

Gracias a la tierra por todo lo que nos da y gracias a todos esos organismos que trabajan escondidos en el suelo para limpiar, reciclar, y transformar la materia en flores de todos los colores, en alimento, en más vida y diversidad. Soñando por que como humanidad lleguemos a apreciar el valor de todos los organismos vivos y la importancia vital de mantener suelos sanos, diversos y funcionales.