Luísa Fraga Dornellas Cysneiros

EVALUATING SOIL MACROFAUNA USING MORPHOLOGICAL AND DNA APPROACHES ALONG A GRADIENT OF FARMING INTENSIFICATION

Dissertação no âmbito do Mestrado Internacional em Ecologia Aplicada orientada pelo Doutor Luís das Neves Cunha e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e tecnologias da Universidade de Coimbra

Julho de 2023

Evaluating soil macrofauna using morphological and DNA approaches along a gradient of farming intensification

Luísa Fraga Dornellas Cysneiros

This work was carried out at the Centre for Functional Ecology - Science for People & the Planet (CFE) of the Department of Life Sciences of the University of Coimbra and was funded by the Fundação Calouste Gulbenkian through a "Novos Talentos Científicos - Sustentabilidade Ambiental" Grant, research project "CULTIVAR - Rede de competências para o desenvolvimento sustentável e inovação no setor Agroalimentar" (CENTRO-01-0145-FEDER-000020) co-financed by the Regional Operational Programme Centro 2020, Portugal 2020, and the European Union, through the European Social Fund (ESF) and the strategic plan of CFE (UIDB/04 004/2020), financed by FCT/MCTES through national funds (PIDDAC).

FUNDAÇÃO
CALOUSTE
GULBENKIAN

CENTRE FOR
FUNCTIONAL
ECOLOGY

INTERNATIONAL
MASTER in
APPLIED
ECOLOGY

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

ACKNOWLEDGEMENTS

To Dr. Luís Cunha a special thanks. Thank you for guiding me in this adventure that was the thesis, for all the knowledge transmitted and opportunities created. You took into account my personal tastes, and this thesis gained shape under your careful guidance. A permanent and interested support that makes me want to pursue my studies in this area of research.

Prof. Dr. José Paulo Sousa, thank you for the passage of knowledge, mainly statistical, and for the opportunities provided.

To Sara Mendes for the company in the sorting room, for the passage of taxonomic knowledge, and for allowing the inclusion of primary parts to this thesis.

To CIBIO-InBIO, Dr. Vanessa Mata and Maria João Magalhães who allowed me a fundamental collaboration in the development of this thesis, with an unparalleled interest, monitoring and availability.

This academic marathon would not have been possible without the support of my family. Mom, your support was essential from the day I was born and the education, encouragement and strength you gave me the tools I needed to pursue my dreams. Without your support and the support of my father, I am sure I would never have reached this. My biggest 'thank you' goes to you two. To my sister Rita. Both of us have accomplished something great this year, so I would like to mark my achievement with yours. Congratulations, sister! I could not be prouder. You were my role model while growing up and I have learnt so much from you. I am sure I would not be the person I am today, thank you for believing in me before I did. To my brother Vasco and sister Taia (and Lara), for always being there and for the invaluable support and humour over the years.

To the Real República Rápo-Táxo, a special thanks for having provided so many moments of leisure as well as lunches and dinners, some more, others less rushed. Without your support, availability, and friendship, this would have been a very different year.

Real República dos LySOS for having given me a stay in Porto, allowing me not only to carry out part of this thesis in Invicta, but also to get to know a house and its people in the process.

Beatriz Gonçalves, Gabriela Teixeira, Inês Morgado, Madalena Sottomayor, Mafalda Albuquerque, and Manuel Sampaio, thank you for all your support during the past years.

Ursula Heinze and Panteha Ardani, colleagues and companions of the masters, and especially great friends and colleagues, that I will always take with me. The late work nights, the coffees in different countries, and all the incredible and intense experiences we shared in these two

years. The experience of this master's programme would never have been the same without you.

I appreciate the environment created by the research group, with special attention to the help given by Diogo Fonseca in the laboratory and his company at lunchtime. And to Sofia Ramos, a big 'thank you' in particular, who accompanied me in the final stretch of this thesis in an intense way and without whom this phase would have been much more complicated.

The CULTIVAR project that allowed me to develop this thesis and provided the necessary resources for its idealization and consequent execution.

Calouste Gulbenkian Foundation for awarding me a research initiation grant 'Novos Talentos Cientficos - Sustentabilidade Ambiental', which allowed me to explore new themes and deepen my knowledge of the academic environment.

A special thanks to the tutor who was assigned under the aforementioned Scholarship, 'Novos Talentos Científicos - Sustentabilidade Ambiental '', Dr. Ruben Heleno, and Dr. Filipa Bessa, both of whom accompanied me and contributed attentively and closely in the elaboration of the project developed here.

ABSTRACT

The concept of One Health emphasizes the interconnections between human, animal, and environmental health, highlighting the need for multidisciplinary collaboration for a sustainable future. Soil health is a critical foundation of terrestrial ecosystems, supporting essential functions for life. Soil biodiversity, which consists of various species, plays a vital role in these, such as nutrient cycles, decomposition, and maintenance of soil structure. Research on soil biodiversity remains limited compared to aboveground organisms, which poses challenges to conservation. Regarding agroecosystems, intensive practises are the major threats soil biodiversity is under. While extensive and traditional, such as agroforestry, could promote biodiversity, contribute to food security, climate change mitigation, and ecosystem services, being an ally towards a more secure and sustainable world. Advanced molecular methods, such as barcoding and metabarcoding, offer promising results in assessing soil biodiversity more efficiently. This study aims to evaluate and compare morphotaxonomy and metabarcoding approaches to assess soil macrofauna diversity in agroecosystems with varying levels of intensity of management and to explore the validation of the methods used. Using an integrative approach, morphotaxonomy and molecular techniques were combined to assess soil invertebrate fauna in different agroecosystems. The analysis of 192 samples, through morphotaxonomy, revealed a total of 9418 individuals from 13 different taxonomic groups. Of the sequenced samples, 716 operational taxonomic units (OTU) belonging to the soil macrofauna were identified. Primer pairs, such as Folmer, BF3/BR2, were employed for DNA barcoding and metabarcoding of the COI barcode. Carabids were identified at the species level to assess recovery efficiency of the molecular methods, with both methods recovering OTUs from carabid species with high percentages of similarity. However, discrepancies between the two methods were observed, indicating potential primer bias and/or poorly curated data bases, thus the need for further exploration. Regarding the assessment of biodiversity of agroecosystems, metabarcoding indicated the need to intensify sampling efforts to reveal the true biodiversity of the systems, while morphotaxonomy showed greater sampling completeness. The integrated data set was congruent in revealing that higher levels of biodiversity were found in traditional agroforests and montados, compared to improved pastures.

The study emphasizes the need for highly integrative approaches to overcome the pitfalls inherent in each method and improve the precision of species identification. Further analysis should assess the impact of specific taxonomic groups on the dynamics of agroecosystems and evaluate the influence of environmental and management practises on biodiversity richness and community composition. Additionally, concentration effects in mock communities and

comparisons with similar studies are essential to refine the metabarcoding pipeline and increase genetic references. However, the collected data have the potential for further development and more in-depth conclusions in the future. This study sheds light on the challenges and opportunities of using molecular-based approaches, highlighting the importance of integrating morphotaxonomy knowledge with molecular data to curate existing databases and enhance reliability. Conservation of biodiversity requires reliable methods to understand biodiversity loss and identify areas at risk. This research demonstrates how an integrative strategy that uses both morphotaxonomy and molecular techniques can provide comprehensive insight into species richness and composition. Molecular methods, such as DNA barcoding and metabarcoding, can offer precise species identification, especially when morphological identification is difficult or ambiguous. However, there are pitfalls associated with molecular approaches, including primer bias and mismatches between morphological and molecular methods. To address these challenges, collaboration between researchers and taxonomy experts is vital to ensure accurate species assignments and reveal hidden diversity.

RESUMO

O conceito de 'One Health' evidencia a interdependência entre a saúde humana, animal e ambiental, enfatizando uma colaboração multidisciplinar para um futuro sustentável. A saúde do solo é essencial para a saúde dos ecossistemas terrestres, suportando funções vitais para a produção e vida. A biodiversidade do solo, composta por diversas espécies, desempenha um funções essenciais nos ciclos de nutrientes, na decomposição e na manutenção da estrutura do solo. No entanto, as práticas agrícolas intensivas nos agroecossistemas ameaçam a biodiversidade do solo, enquanto os métodos extensivos e tradicionais promovem a biodiversidade, segurança alimentar, mitigação das alterações climáticas e serviços dos ecossistemas. A investigação da biodiversidade do solo é reduzida quando comparada com o conhecimento existente sobre os organismos terrestres, o que representa enormes desafios à sua conservação. No contexto agrícola, práticas intensivas representam uma das principais ameaças à biodiversidade do solo, enquanto práticas extensivas e tradicionais, como a agrofloresta, podem promover a biodiversidade, contribuindo para a segurança alimentar, mitigação das alterações climáticas e serviços dos ecossistemas, tornando-se assim aliados à criação de um mundo mais seguro e sustentável.

Métodos moleculares, como barcoding e metabarcoding, podem oferecer resultados promissores na avaliação eficiente da biodiversidade do solo. Este estudo tem como objetivo comparar diferentes abordagens, nomeadamente, morfotaxonomia e metabarcoding de forma a avaliar a diversidade da macrofauna do solo em agroecossistemas com diferentes níveis de intensidade de gestão, assim como explorar a eficácia e fidelidade dos métodos utilizados.

Através de uma abordagem integrativa, morfotaxonomia e técnicas moleculares foram combinadas de forma a avaliar a macrofauna do solo em diferentes agroecossistemas. A análise de 192 amostras revelou um total de 9418 indivíduos pertencentes a 13 grupos taxonómicos diferentes. Foram ainda identificadas 716 unidades taxonómicas operacionais (OTUs). Para a análise genética do gene mitocondrial citocromo C (subunidade I) foram usados pares de primers: Folmer e BF3/BR2, para o barcoding e o metabarcoding. Ambos os métodos recuperaram OTUs de espécies de carabídeos com altas percentagens de similaridade. No entanto, foram observadas discrepâncias entre os três métodos, morfotaxonomia, barcoding e metabarcoding, demonstrando a necessidade de melhorar a fiabilidade dos métodos. Relativamente avaliação da biodiversidade nos diferentes agroecossistemas, o metabarcoding indicou a necessidade de aumentar os esforços de amostragem de forma a revelar a verdadeira biodiversidade existente nos agroecossistemas, já a abordagem morfológica não demonstrou a mesma necessidade, mostrando-se como uma abordagem mais superficial. Neste estudo revelaram-se níveis mais elevados de riqueza de biodiversidade em agroecossistemas tradicionais e montados quando comparados com pastagens geridas.

Aqui realçamos a necessidade de abordagens integrativas e multidisciplinares de forma a superar as dificuldades inerentes a cada método, ultimamente melhorando a precisão e rapidez na identificação de espécies. Investigações futuras deverão avaliar o impacto dos diferentes grupos taxonómicos presentes assim como a sua função e impacto nos agroecossistemas. O estudo dos efeitos da concentração das comunidades modelo, assim como a comparação com estudos semelhantes são essenciais para melhorar os métodos e o pipeline de "metabarcoding" aqui explorado, assim como aumentar e retificar as referências genéticas nas bases de dados locais. Os dados aqui obtidos permitem ainda uma maior exploração no futuro.

Este estudo enaltece tanto os desafios como as oportunidades do uso de abordagens moleculares para a avaliação da biodiversidade, destacando ainda importância da integração do conhecimento de morfotaxonomia de forma a aumentar a confiança nos resultados. A conservação da biodiversidade requer métodos fiáveis para avaliar a perda de biodiversidade e identificar áreas em especial risco. Aqui demonstramos como uma estratégia integrada que utiliza tanto morfotaxonomia e técnicas moleculares pode dar conhecimentos sobre a riqueza e a composição das espécies. Métodos moleculares, como barcoding e metabarcoding de DNA, podem oferecer identificação precisa de espécies, especialmente quando a identificação morfológica é difícil ou ambígua. Porém, existem dificuldades associadas a estas abordagens moleculares, como erros associados aos primers escolhidos como discordâncias entre métodos morfológicos e moleculares. Para enfrentar estes desafios, a colaboração interdisciplinar é fundamental de forma a garantir a correta atribuição da sequência genética à espécie e assim permitir a obtenção de resultados corretos e revelar diversidade anteriormente oculta.

Table of Contents

Acknowledgements 1

Abstract 3

Resumo 4

Introduction 7

Soil Health is Human Health 7

Soil Biodiversity 8

Soil Biodiversity in Agroecosystems 9

Soil Biodiversity Threats 11

Biodiversity Assessment Tools 12

Aims of the study and expected results 14

Materials and Methods 15

Study site and sample collection 15

Biodiversity Assessment Tools 18

Statistical Analysis 22

Results 25

Methodological Comparison 26

Hill Numbers for Diversity Assessment and Sample Completeness 28

Biodiversity Assessment Based on Morphotaxonomy 28

Biodiversity Assessment Based on OTUs 32

Discussion 36

The Value of Morphotaxonomy and its Challenges 36

Traditional and Organic Agroecosystems Hold High Levels of Biodiversity 38

Challenges and Opportunities for Improvement of DNA-Based Approaches 39

OTU Communities as Determinants of Agroecosystem Separation and Clustering 39

Future Research: Implications for Conservation and Agroecosystems Management 40

Conclusion 41

References 42

Supplementary Materials 49

INTRODUCTION

SOIL HEALTH IS HUMAN HEALTH

The One Health (Figure 1) concept underlines and recognizes that the health of humans, animals, and the environment is closely linked and interdependent, thus the importance and necessity of close collaboration across multiple disciplines to ensure a sustainable and healthy future (OHHLEP et al., 2022).

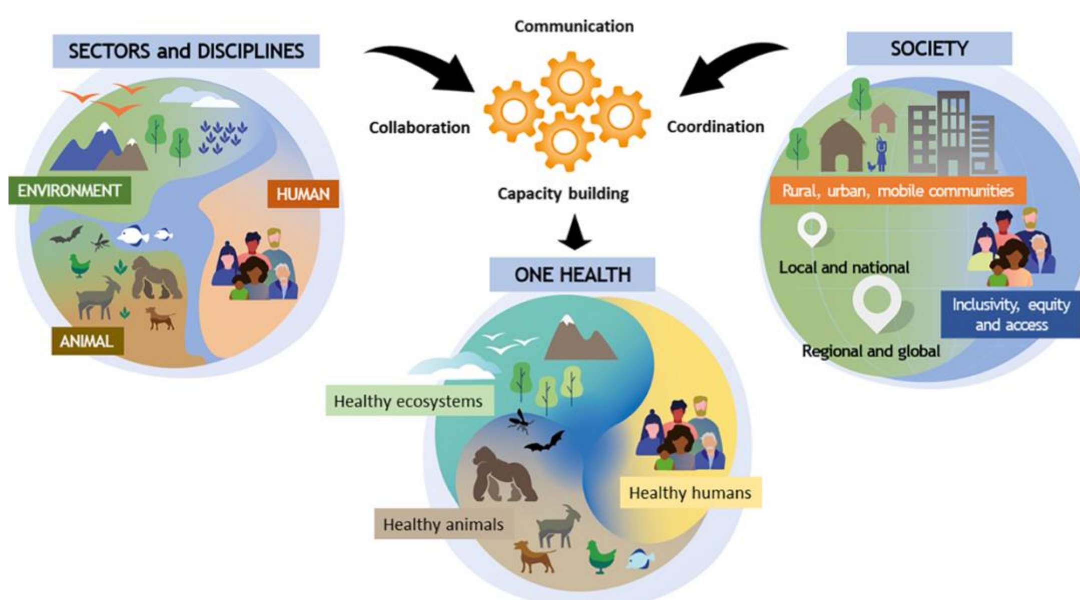


Figure 1: One Health concept as the source for a sustainable and healthy future for ecosystems, animals, and humans. Image from: (One Health High-Level Expert Panel (OHHLEP) et al., 2022).

This conceptual framework has clear advantages in improving the overall health of humans, animals, and ecosystems by combining social and environmental protection and supporting sustainable and resilient economic development (OHHLEP et al., 2022). However, to implement such a vision, multiple challenges and constraints must be faced that range from legal, ethical, financial, and social capacity. Therefore, measures addressing human, animal, and ecosystem health, including loss of biodiversity, clean air and energy, the impact of climate change, food and water security, and social inequalities, are essential (OHHLEP et al., 2022).

Soil is the foundation of every terrestrial biome, supporting multiple functions for the balanced functioning of ecosystems. It provides essential goods and services to sustain and maintain life. As defined by Doran & Parkin (1994), healthy soil is described as 'the capacity

of a soil to function within ecosystem boundaries to sustain biological productivity, maintain environmental quality, and promote plant and animal health'. This definition greatly emphasizes the interdependent relationships between human, animal, and environmental health.

The European Union Commission has recently released the 'EU Soil Strategy for 2030', where the importance of soils as the premise of food chains, both human and aboveground biodiversity, is acknowledged (European Commission, 2021). This sets up initiatives and political support for research and monitoring of soils, enforcing the establishment of correct management, better land use practises, and restoration initiatives of degraded soils, hopefully inverting the recent tendency of disregarding the importance of soil organisms in agricultural management (Bender et al., 2016).

Focussing on soil health, several threats have been highlighted in previous research (Stolte et al., 2016). Unfortunately, soils' high complexity and functioning is not fully understood, so establishing actions to tackle or mitigate these can be difficult. Some threats identified were soil erosion by water and/or wind, organic matter decline, soil compaction, soil sealing, soil contamination, soil salinization, desertification, flooding and landslides, and soil biodiversity decline (European Commission, 2021; Stolte et al., 2016).

Overall, soil is highly heterogeneous, capable of having severely different chemical and physical properties, thus holding different ecological niches providing an astonishing range of biodiversity (FAO, 2020; Stolte et al., 2016).

Soil Biodiversity

Soils hold almost ¼ of all species on earth, being one of the most biodiverse habitats (Arribas et al., 2021; Guerra et al., 2021) and a global reservoir of biodiversity (Arribas et al., 2021; Guerra et al., 2021). According to FAO (2020) soil biodiversity can be defined as 'the variety of life belowground, from genes and species to the communities they form, as well as the ecological complexes to which they contribute and to which they belong, from soil microhabitats to landscapes'. Soil biota compromises organisms that depend on the soil throughout their life cycle, but also organisms that rely on it during part of their life cycle, for laying eggs, overwintering or during certain life stages (Orgiazzi et al., 2016).

To safeguard balanced ecosystems, it is imperative to conserve soil biodiversity, as these hold fundamental roles in supporting soil functions of both natural and managed ecosystems (Orgiazzi et al., 2016) and, therefore, the goods and services we deeply rely on. Essential services and functions supported by soil biodiversity include, for instance, nutrient and carbon

cycles, decomposition, soil structure vegetation dynamics, and eco-evolutionary responses of ecosystems to global change (Bardgett & Van Der Putten, 2014; Bender et al., 2016; Ferreira et al., 2022; Orgiazzi et al., 2016).

Soil macrofauna comprises organisms with more than 2mm, and these are functionally diverse, contributing to ecosystem processes in different ways. They can be decomposers (beetles, isopods, gastropods, etc), predators (spiders, beetles, chilopods, opiliones, etc), herbivores, and ecosystem engineers (earthworms, ants, termites) (Jeffery et al., 2010; Swift et al., 1979).

Soil Biodiversity in Agroecosystems

Regarding biodiversity levels, the insurance hypothesis suggests that higher levels will confer more resilience to the ecosystem against disturbance or stress. Despite this, some research defends that the soil biodiversity is functionally redundant. Hence, less diversity will not affect the soil functionality so much, and the goods and services will still be provided in a less biodiverse ecosystem. However, crossing a certain biodiversity threshold, beyond which functions begin to diminish, could have abrupt and devastating effects on the ecosystem and its productivity (Yachi & Loreau, 1999).

Agroecosystems consist of ecosystems that humans modify to produce goods, such as food and fibre. They can be managed in different ways, going from extensive to intensive practises. The latter relies on several management practices, such as tillage, the use of chemicals, mechanization, and plant breeding. These practices impact the ecosystem, which can lead to the degradation of soils (Jeffery et al., 2010) and consequently to the loss of productivity. Sustainable agriculture practices consist of the timeframe within which plant and animal production is maintained without production decline, the time needed for the resources to be maintained or renewed, and farmers' economic viability and maintenance (Brussaard et al., 2007).

Inherent to agroecosystems are their biological resources; these include soil biodiversity, and different management practises will impact this resource, for example, in the species richness, abundance, and communities present (Brussaard et al., 2007). Main management options comprise tillage, crop rotation, sequence, and organic input. Intensive and hyperintensive agriculture decreases species richness and the dominance of some species (Brussaard et al., 2007). On the contrary, management characterized by crop rotation, no-tillage, organic fertilizers, and preserving natural elements in the field increases species

richness and overall population density (Figure 2) (Brussaard et al., 2007). Some practices are known to have positive impacts, such as drainage and irrigation; others, such as soil cultivation or application of organic matter, can positively or negatively affect soil species richness (Brussaard et al., 2007). Thus, it is vital to understand the potential consequences on soil biodiversity to make informed management decisions regarding agricultural practices.

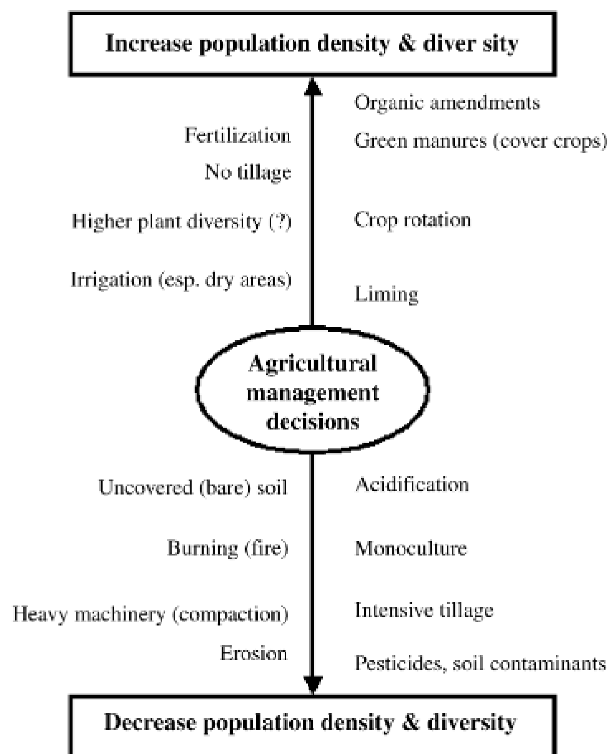


Figure 2: Effect of different agricultural management practises on the density and diversity of soil animal populations. From: (Brown et al., in press; after Hendrix et al., 1990) copied from (Brussaard et al., 2007).

On the one hand, some groups of soil macrofauna, such as slugs and snails (Orgiazzi et al., 2016), can be considered pests in agricultural crops, whereby they may harm agricultural production. On the other hand, soil organisms can assist in sustainable agricultural production (Bender et al., 2016; Creamer et al., 2016). For instance, some groups can be pest control agents (biocontrol), replacing the use of pesticides. Soil biodiversity is also essential for maintaining soil fertility and nutrient redistribution (Orgiazzi et al., 2016). Hence, maximizing their contribution can help to ensure food security while minimizing adverse environmental effects (Bender et al., 2016).

An agroforest is an extensively managed ecosystem. It takes advantage of the ecological interactions between trees and agricultural elements by combining them with animals, crops, and/or trees, thus bringing several benefits; for example, it increases biological activity, nutrient supply, and biodiversity (Mupepele et al., 2021; Torralba et al., 2016). Consequently, this practice improves soil health and is considered a more sustainable agricultural practice (Mupepele et al., 2021; Torralba et al., 2016). Thus, it contributes to income and food security, biodiversity conservation, and ecosystem services, being considered a mitigation and adaptation technique for climate change for agriculture.

Therefore, understanding the state of the existing macroinvertebrate communities in the different agroecosystems will allow the drawing of conclusions and possible management adaptations at the local scale to increase beneficial biodiversity and support ecosystem functions.

SOIL BIODIVERSITY THREATS

Land use change, overexploitation, and pollution are some of the multiple threats soil fauna faces. Intensive agriculture is considered one of the significant drivers of soil biodiversity loss, as it has been found to decrease taxonomic diversity, functional groups, and complexity of the food web, where the latter contained fewer trophic levels and fewer species of larger body mass (Tsiafouli et al., 2015). On top of this, climate change is also predicted to influence soil biodiversity as it impacts both abiotic and biotic factors that regulate soil diversity and communities (Stolte et al., 2016).

Despite their clear ecological importance, supporting both natural and managed ecosystems, soil biodiversity remains understudied when compared to other taxa, such as aboveground organisms (Orgiazzi et al., 2016; Wall, 2012), further complicating soil conservation (Orgiazzi et al., 2016). Specifically, species belonging to the classes of Arachnida and Insecta belong to the most underrepresented in the GBIF (Global Biodiversity Information Facility) database when compared with the number of known species (Troudet et al., 2017), demonstrating a taxonomic bias when it comes to research. Notwithstanding the acknowledgement of the existing bias, most classes over and underrepresented in 1950 remained the same by 2017 (Troudet et al., 2017), while the bulk of biodiversity remains to be reported.

Coupled with this gap in knowledge of the existing soil biodiversity, understanding the main influences of their distribution also requires the scientific community's attention, as these are far less understood than those driving the distributions of aboveground biodiversity (Orgiazzi et al., 2016). The main drivers of abiotic factors are climate, such as temperature and moisture, soil texture, salinity, and pH. Regarding abiotic ones, vegetation composition, diversity and

trophic interactions are the most influential (Orgiazzi et al., 2016). Thus, changes in the ecosystems will consequently have consequences on soil-dependent organisms.

Therefore, monitoring biodiversity communities and populations in agroecosystems is fundamental to making informed decisions to maintain or improve biodiversity, ensuring the long-term sustainability of agroecosystems.

Biodiversity Assessment Tools

Assessing soil biodiversity, namely, macrofauna, using traditional approaches such as taxonomic identification is regarded as both costly and time-demanding, requiring high taxonomic expertise (Watts et al., 2019). Taxonomists are becoming scarce (Cao et al., 2016; Valdecasas & Camacho, 2003), making it even more challenging to identify entomological specialists. Moreover, several cryptic species and larvae can be wrongly identified, leading to erroneous results. Also, damaged individuals might not be able to be identified. Nowadays, the use of molecular techniques, such as metabarcoding, a developing method is becoming more common to assess biodiversity communities in different ecosystems, including, for example, freshwater (Baselga et al., 2013; Bista et al., 2018) marine (Fonseca et al., 2010; Leray & Knowlton, 2015), terrestrial (Arjona et al., 2022; Ji et al., 2013; Martoni et al., 2023; Mata et al., 2021; Watts et al., 2019) and also paleoenvironments (Cao et al., 2020). Besides biodiversity assessments, for instance, metabarcoding can be applied in dietary studies (Mata et al., 2019). The method combines DNA taxonomy and high-throughput DNA sequencing (HTS) as it can sequence mixed samples, thus identifying multiple species at once, efficiently reducing the time necessary to assess the whole community of an ecosystem.

Sanger sequencing, or DNA barcoding, allows the identification of species through short sequences of genes, the barcodes. These are genetic fragments neighboured by highly conserved regions used routinely to identify species (Liu et al., 2020). The cytochrome c oxidase subunit I (COI) barcode is a 658 bp segment of the mitochondrial gene that is commonly used to identify animal species (Hebert et al., 2003; Liu et al., 2020) and is thus widely used to establish reference databases.

As shown in Figure 3, DNA barcoding can be used to bridge taxonomy and molecular ecology approaches, such as metabarcoding, by generating molecular reference databases of specimens curated by specialists. Similarly to DNA barcoding, the identification relies on short DNA sequences, the barcode.

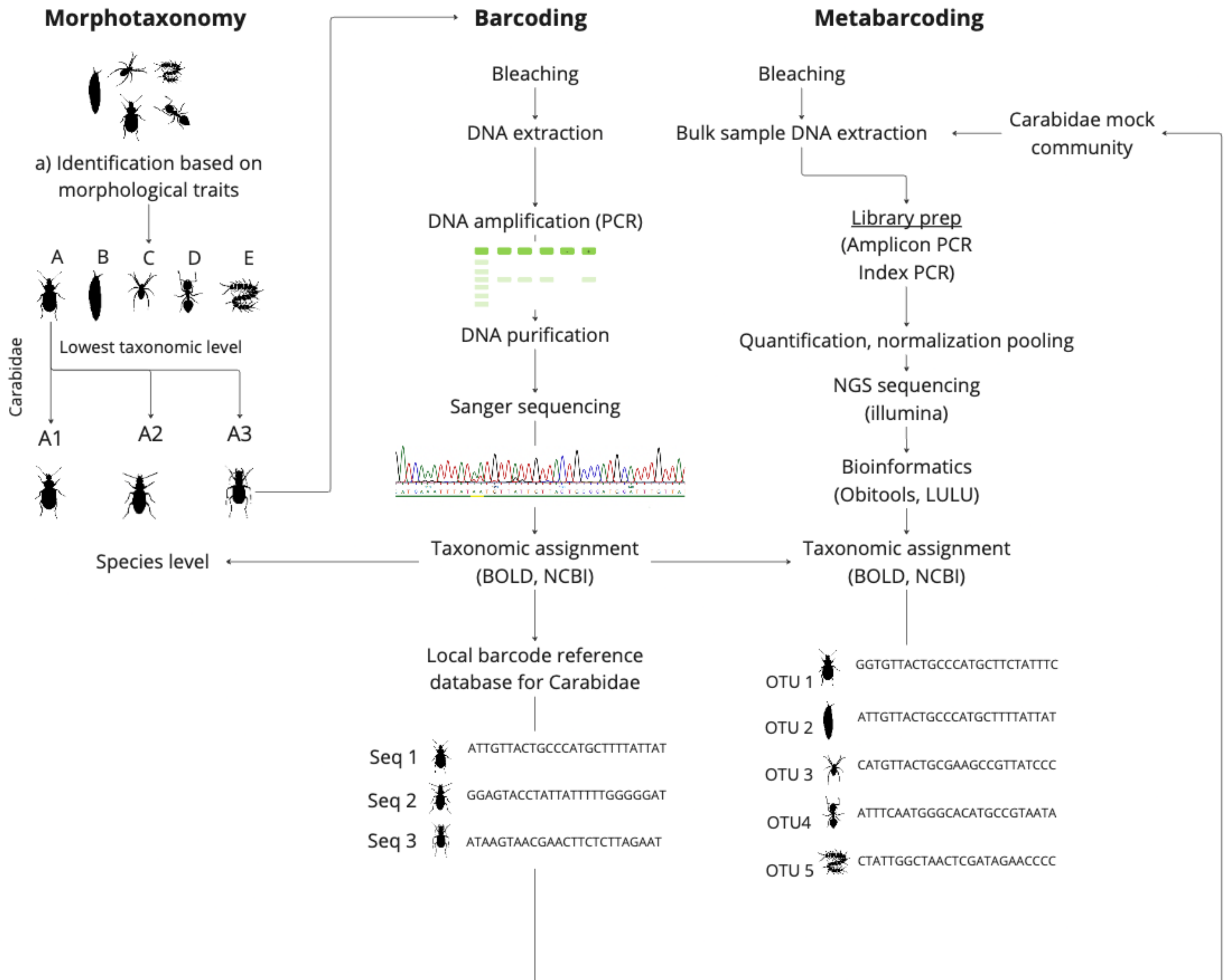


Figure 3: Schematic summary of the three methods used (morphotaxonomy, barcoding, and metabarcoding), highlighting their interdependence. Arthropod silhouettes extracted from PhyloPic (www.phylopic.org).

Researchers can overcome the limits of each method individually by combining different methodologies, improving the accuracy and reliability of biodiversity evaluations (Cristescu, 2014; Janzen et al., 2005). This integrative method (Figure 3) enables a more thorough and robust understanding of ecosystems, allowing for discovering hidden or cryptic species, the identification of genetic variants within species, and the assessment of overall biodiversity patterns (Janzen et al., 2005).

AIMS OF THE STUDY AND EXPECTED RESULTS

In Portugal, the lack of information on soil biodiversity is exacerbated by the lack of reference collections for several taxa groups. Regarding DNA barcodes, the InBIO Barcoding Initiative contains separated records of reference databases; these are usually specific to order (Ferreira et al., 2020; Schweiger et al., 2007), family, genus, or even species-specific (Corley et al., 2020), thus lacking a coherent and more exhaustive taxa information.

The existing knowledge gap also fosters the opportunity for the validation of a coherent methodological approach to identify soil macrofauna consistently. This is particularly important for the case of central Portugal, where the only existing studies were done on arthropod morphospecies (da Silva et al., 2019).

Therefore, in this study, two overarching goals of different nature were aimed with several secondary objectives and hypotheses: one focused on the ecological impact of agroecosystems with varying levels of disturbance intensity and the other on developing methods involving the integration of classical taxonomy with molecular approaches.

Integrating Classical Taxonomy with Molecular Approaches

In this approach, we aimed to evaluate and compare the value and complementarity of morphotaxonomy, barcoding, and metabarcoding approaches, including the establishment of protocols to collect, preserve, and assess soil macrofauna using NGS (Next Generation Sequencing) methods, while contributing to the enlargement of the local reference database of COI segments.

A mock community was built to test the usefulness of the metabarcoding pathway for assessing the richness of carabid soil macrofauna. The carabid family was chosen for this analysis for two main reasons: it is ubiquitous, and species-level identification was achievable as a specialist was available. This widely-known group is abundant in several ecosystems, from forests to wetlands, and in anthropogenic environments such as agroecosystems (Jeffery et al., 2010). These can be used as bioindicators for several purposes such as soil quality assessments

and cropland management (Jeffery et al., 2010). Using a ubiquitous, functionally and genetically diverse group, a carabid mock community was established by combining morphotaxonomy with Sanger sequencing (barcoding) of specimens. It was made up of 31 carabid sequences derived from curated specimens. The recovery percentage of this simulated community was used to calculate the efficiency of the metabarcoding pipeline. This mock sample was then added to pre-selected samples to assess species-level recovery efficiency. This approach allowed us to test the usefulness, efficiency, and reliability of the metabarcoding pathway for assessing soil macrofauna richness of carabids.

Determining Soil Biodiversity in Agroecosystems with Different Levels of Distress

We intended to evaluate soil macrofauna composition and abundance in different agroecosystems with different levels of anthropogenic disturbance. Considering the various management practices, biodiversity levels and the presence of functional groups are expected to differ among varying levels of disturbance intensity. We hypothesized that higher biodiversity richness and diversity would be retained in more extensively managed agroecosystems than in intensive and Hyperintensive ones.

The species richness and composition of each agroecosystem were assessed using morphotaxonomy and metabarcoding; the workflow is represented in Figure 3.

MATERIALS AND METHODS

STUDY SITE AND SAMPLE COLLECTION

Under the scope of this study, we focused on soil invertebrate communities from 8 agroecosystems with different management practices. This study was carried out in Idanha-a-Nova, Portugal (Figure 4 (A)) at eight distinct sites with different types of agroecosystems, namely four agroforests, one improved pasture and three monocultures (one intensive and two hyperintensive). For each system, three plots were selected with a minimum of 500m between each other; furthermore, considering that soil macrofauna is the targeted group, independence is ensured. At each plot, three transects (each with 50 m) were defined, and three pitfall traps in an equidistant (25m * 25m) manner were deployed per transect, with a total of nine traps per plot (Figure 4 (B)) and making up a total of 216 pitfall traps. Sampling was done during the fall season, from late November until the beginning of December 2022. Pitfall traps remained in the field for at least 13 days and up to 17 days, table 1. Pitfall traps were filled with ethylene glycol, and lids were placed to reduce rainwater accumulation and consequent

reduction of ethylene glycol concentration, hindering degradation of organisms. Once collected, the samples were stored in 96% ethanol at room temperature until further analysis.

(A)

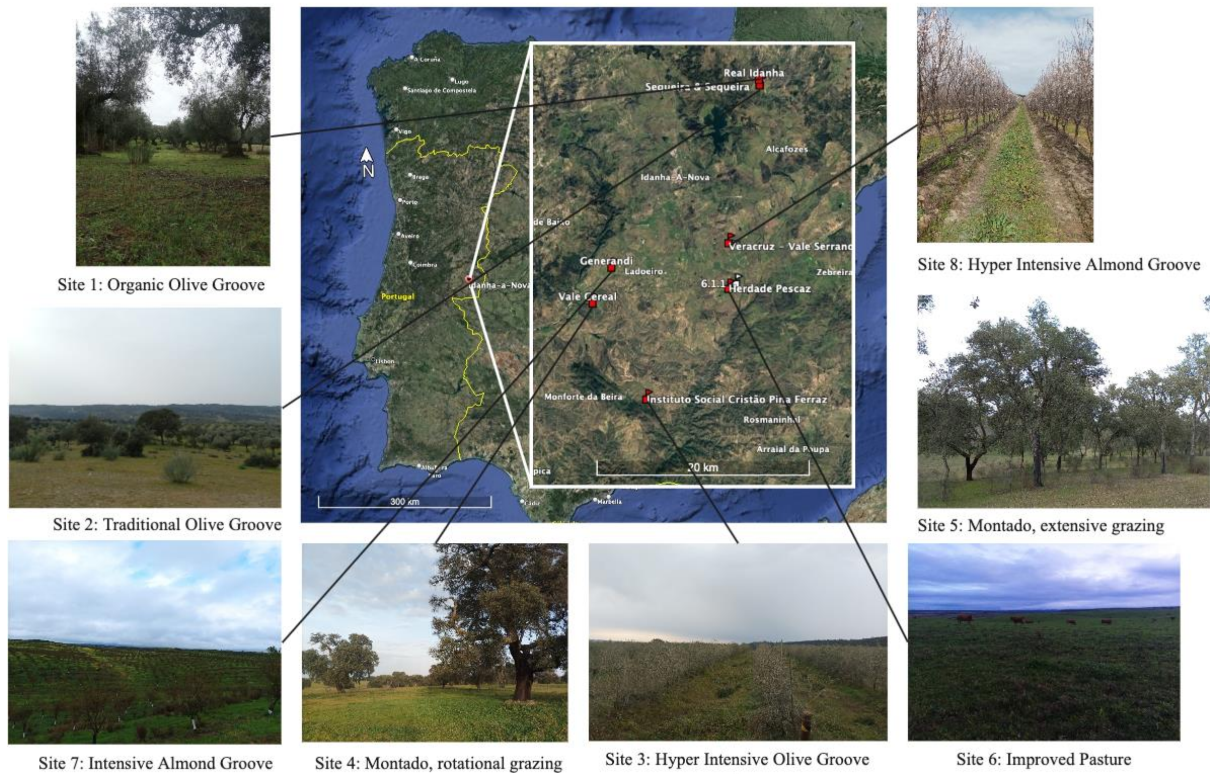
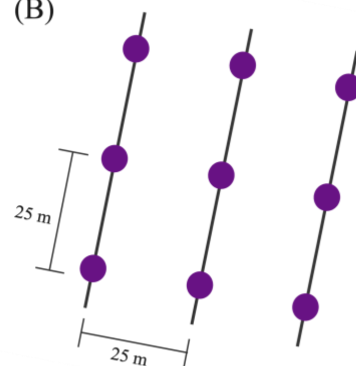


Figure 4: (A) Location of the eight sampling sites of different agroecosystems and management practices were sampled. In each site, 27 pitfalls traps were during 13 to 17 days. (B) Plot setup: transects are represented by the black line, and the purple circle represents pitfall traps.



Of the 216 pitfalls placed in the field, only 192 were recovered (table 1), and 23 samples were lost due to wildlife, potentially due to wild boar activity. As for DNA metabarcoding, 235

samples were sequenced, 191 simple samples, 30 samples with added mock sample , one mock sample, nine negatives with one per each 25 sequenced samples, and four BPcrr. Three samples were lost in the processing, two from site three and one from site 5. PCR negative controls did not show signs of contamination. Hence, 220, 189 simple samples (table 1), 30 samples with added mock samples and one mock sample contributed with OTUs (Operational Taxonomic Units) after bioinformatic filtering. Table 1 summarises the number of samples (n) per sampling site for each methodology: morphotaxonomy and metabarcoding; and the time they remained in the field.

Table 1: Number of samples recovered by site and used in morphotaxonomy and metabarcoding approaches, with the respective number of days they remained in the field.

Site	Agroecosystem	Days in the field	n (morphotaxonomy)	n (metabarcoding)
1	Organic Olive Grove	14	25	25
2	Traditional Olive Grove	14	27	27
3	Hyperintensive Olive Grove	14	27	25
4	Montado Rotational grazing	14	27	27
5	Montado Extensive grazing	15	13	12
6	Improved Pasture	13	27	27
7	Intensive Almond Grove	17	21	21
8	Hyperintensive Almond Grove	17	25	25
Total			**192**	**189**

Particularly in the Iberian Peninsula we can find the Montado/Dehesa (Portuguese/Spanish), a traditional agro-silvio-pastoral agroecosystem (Figure 4 (A), sites 4 and 5). These are considered a cultural legacy being culturally and economically important. They are related to low disturbance levels and to an extensive production system of several production activities (Pinto-Correia et al., 2011), thus compiling a heterogeneous landscape.

An improved pasture (Figure 4 (A), site 6) is characterized by the presence of selected grass and clover species with high grazing value. These are maintained by livestock grazing and the use of lime and fertilizers (Jeffery et al., 2010). The improved pasture sampled is irrigated and has livestock throughout the year. The grass species is selected and planted, and fertilizers are used.

A conventional olive plantation and one that was recognized as an organic olive crop Figure 4 (A), sites 2 and 1) were both evaluated. In addition to these, intensive and hypertensive monocultures, an almond (intensive and hyperintensive) Figure 4 (A), sites 7 and 8) and an olive plantation (hyperintensive) Figure 4 (A), site 3) were also assessed. Regarding intensive and hyperintensive agricultural practices, several interventions are explored for the most

profitable production possible, such as the use of pesticides and fertilizers, space between trees, use of machinery, tillage, etc.

BIODIVERSITY ASSESSMENT TOOLS

Morphotaxonomy

Traditional taxonomic identification of macroinvertebrates was carried out using dichotomous keys relying on the morphological characteristics of the organisms. Samples were analyzed under a low-power microscope, and organisms were sorted to the lowest possible taxonomic level using morphological characters based on available taxonomic expertise and literature.

DNA Barcoding of the Carabidae Mock Community

To build the local reference database, specimens belonging to the Carabidae family were identified at the species-level based on morphological traits by a specialist assisted by a dichotomous key (Aguiar & Serrano, 2012), and an external expert was contacted when needed. When handling specimens, to avoid cross-contamination among samples, all materials used, such as tweezers, were cleaned and went through flame sterilization for disinfection.

We generated a local barcode reference database for the universal arthropod COI barcode region for the DNA of the 31 Carabidae species collected. For each extraction, we used one specimen and, when possible, two, making up to 51 extractions.

Previously to DNA extraction, specimens were bleached to remove external DNA. Individuals were removed from 96% ethanol to dry for 10 minutes on cellulose paper and then bleached with a 0,5 % sodium hypochlorite solution ('Klorix'). By gently shaking in a 2ml tube for three minutes. Subsequently, the bleach was discarded and then the individuals were washed with distilled water three times for one minute each.

DNA extraction followed an adapted protocol using the Qiagen DNeasy® Blood & Tissue Kit with overnight lysis. A nondestructive approach was used for small individuals, <1cm, was used, where the whole organism went through the DNA extraction process. While for bigger specimens, >1cm, three legs of one side were removed and macerated. DNA was quantified and checked for contamination using Nanodrop apparatus (Thermofisher).

PCRs of samples containing > 10 ng/ml comprised 10 µl of Master Mix, 1 µl of each 10 nM primer, 4 µl of H_2O, and 4 µl of DNA template, while extractions containing 10 - 100 ng/ml comprised 10 µl of Master Mix, 1 µl of each 10 nM primer, 6 µl of H_2O, and 2 µl of

DNA template, samples with DNA concentration >100 ng/ml comprised 10 µl of Master Mix, 1 µl of each 10 nM primer, 7 µl of H2O and 1 µl of DNA template.

Cycling conditions were the same despite the DNA extraction concentration and consisted of initial denaturation at 95 ° C for 3 min and 5 cycles of denaturation at 95 ° C for 30 s, annealing at 46 ° C for 30 s, and extension at 72 ° C for 45 s, followed by 32 cycles of denaturation at 95 ° C for 30 s, annealing at 51 ° C for 30 s, and extension at 72 ° C for with a final elongation at 72 ° C for 10 min. The PCR products were tested on 2% agarose gel to verify the amplification success. We sequenced the 710 bp COI barcode region using the Folmer primers, LCO1490: 5'-ggtcaacaaatcataaagatattgg-3' and HC02198: 5'-taaacttcagggtgaccaaaaaatca-3' (Folmer et al., 1994). Subsequently, the PCR products were thereafter purified following the ExoSAP-IT™ Express PCR Product Cleanup protocol. Sanger sequencing of the purified PCR products was performed by Eurofins (Germany).

All COI barcode sequences were used as a query in a search against the BOLD (Barcode of Life Data System: https://boldsystems.org/) database and were taxonomically assigned to matching species identification where possible. References with no matching sequences were also analyzed through blast (Altschul et al., 1990) using the NCBI (National Centre for Biotechnology Information: https://www.ncbi.nlm.nih.gov/) database.

The mock community sample was built by pooling the individual DNA extractions of each Carabidae species in equimolar proportions, measured through a Qubit (Thermofisher). The pooled sample was used to evaluate whether the recovery percentage of taxonomic units would vary according to sample diversity complexity (Higher vs Low) and proportion of the added mock sample (0.1, 0.25 and 0.50). The mock sample was added to five samples per treatment combination, making a total of 30 samples that could then be compared to the same matching samples without the added mock.

Metabarcoding of Bulk Soil Fauna Samples

Formicideae (Hymenoptera) and Carabidae (Coleoptera) families did not take part in the metabarcoding sequencing process as the first was required for a different project while the second was used for the barcoding analysis and mock sample creation to evaluate the efficiency of the metabarcoding methodology.

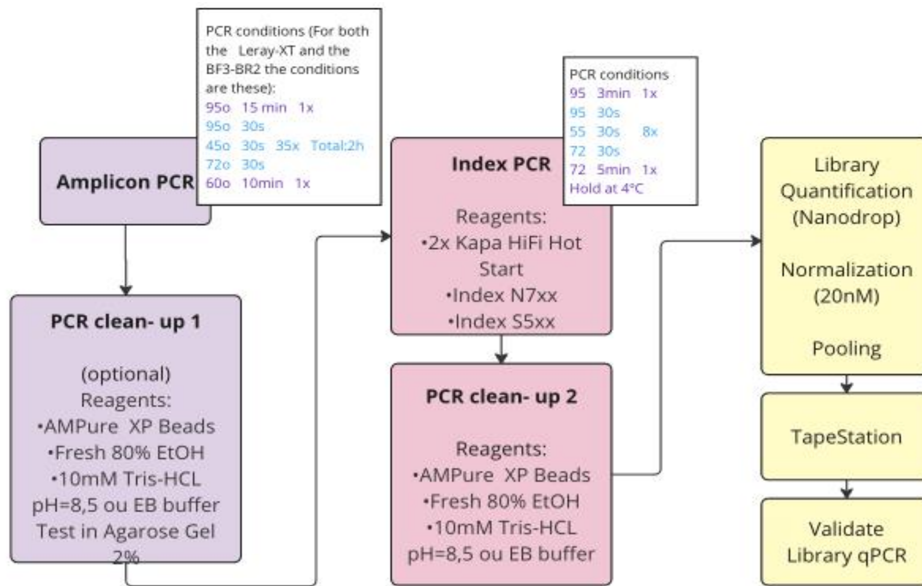
Sample Processing and DNA Extraction

All specimens from each sample were transferred to a falcon to constitute a bulk sample, making up 191 bulk samples. When handling specimens, all materials, such as tweezers and spatulas, were cleaned with sodium hypochlorite (3%) and 96% ethanol. Before DNA extraction, half of the bulk samples underwent decontamination with a 3% sodium hypochlorite

solution for one minute to reduce exogenous DNA and were washed three times with distilled water for 1 minute (Hausmann et al., 2021). Bulk samples were then dried in an incubator at 56ºC overnight. Homogenization followed to turn each bulk sample into a fine powder using the Bullet Blender 50-DX homogenizer (Next Advance), with one to four glass beads (according to the amount of biological content) of 8 mm diameter during 15 min or more, when required. DNA extraction was performed using the E.Z.N.A.® Tissue DNA Kit, following an adapted protocol (Mata et al., 2021). DNA extractions were performed using 70 mg of homogenized insect powder whenever possible.

Metabarcoding Library Preparation

The primer pair was the BF3-BR2 for amplification, which amplifies the 458bp amplicon fragment of the cytochrome C oxidase I mitochondrial gene (Elbrecht et al., 2019). Figure 5 schematizes the dual-PCR protocol followed for Illumina MiSeq library preparation. Amplicon PCR comprised 5 µl of Qiagen Multiplex Master Mix, 0.3 µl of each 10 nM primer, 3.4 µl of H2O, and 1 µl of diluted DNA. Cycling conditions consisted of initial denaturation at 95 ° C for 15 min, followed by 35 cycles of denaturation at 95 ° C for 30 s, annealing at 45 ° C for 30 s, and extension at 72 ° C for 30 s, with final elongation at 60 ° C for 10 min. The PCR products were tested on 2% agarose gel to verify the amplification success. All field samples produced visible amplification bands, while extraction and PCR negative controls showed no signs of amplification. To incorporate 7-bp-long identification tags and Illumina P5 and P7 adapters, a second PCR (index PCR) was performed. Index PCR conditions were similar to the first PCR, except that 7 µl of Kapa HiFi Hot Start was used, as well as 0.7 µl of each 10 nM indexing primer. Cycling conditions consisted of initial denaturing at 95 ° C for 3 min, followed by eight cycles of denaturation at 95 ° C for 30 s, annealing at 55 ° C for 30 s, and extension at 72 ° C for 30 s, with a final elongation at 72 ° C for 5 min. At the end of each PCR, the products were purified using Agencourt AMPure XP beads (Beckman Coulter), quantified using Nanodrop, and diluted to 20 nM. Purified and normalized PCR products were further pooled into a single library and quantified using qPCR (KAPA Library Quant Kit qPCR Mix, Bio-Rad iCycler). The final library was diluted to 4 nM and sequenced on an Illumina Novaseq Platform (Novogene, Cambridge, UK) using a 2 × 250 bp for an average coverage of 160,000 paired reads per PCR product.



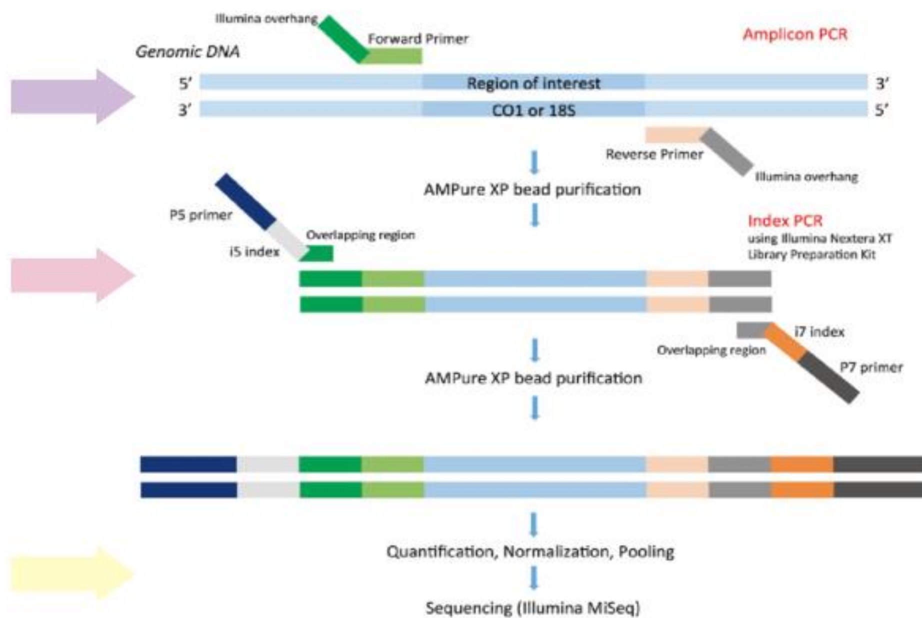


Figure 5: Dual-PCR procedure for Illumina library preparation (Amplicon PCR + Index PCR). The first PCR step, in purple, uses amplicon-specific primers (BF3 and BR2) including Illumina adapter overhangs. Second PCR allows the incorporation of Illumina index adapters i5 and i7. Following each process, bead purification is performed. Prior to sequencing on the Illumina Novaseq, quantification, normalization, and pooling are performed. Adapted from Bourlat et al.,(2016) and Illumina, 2013.

Bioinfomatic Pipeline

OBITools (Boyer et al., 2016) was used for the sequence processing, coupled with VSEARCH (Rognes et al., 2016) and LULU (Frøslev et al., 2017) for denoising. With the "illuminapairedend" command, paired-end readings were aligned and deleted if the overlapping quality exceeded 40. Second, reads were assigned to samples and primer sequences were deleted using "ngsfilter," resulting in four mismatches to the expected primer sequence. Finally, using the "obiuniq" command, reads were collapsed into haplotypes and singletons (haplotypes with only one read per sample) were deleted. Each sample's remaining haplotypes were combined into a single file and once more dereplicated. The data set was denoised using the VSEARCH command "--cluster_unoise" to remove extraneous sequences caused by PCR and sequencing errors and then "--uchime3_denovo" to exclude possible chimeric sequences. The remaining sequences were then grouped using a 99% similarity criterion, and the original readings were remapped to the remaining haplotypes. Finally, we removed co-occurring haplotypes that shared genetic similarities using LULU, greatly reducing the number of mitochondrial, nuclear copies that would otherwise have been present in the final dataset and artificially increasing the number of molecular units and taxa. All haplotypes retained after bioinformatic processing were identified to the lowest possible taxonomic level. This allowed to visually define clusters (or bins) of haplotypes that corresponded to the same taxon and identify chimeric sequences and PCR errors that remained in the dataset. Additionally, sequences of non-target taxa, such as bacteria, plants, and fungi, using the annotation outcome, were removed from the data set. Taxa were identified by comparing the representative haplotypes of each cluster against online databases (BOLD and NCBI).

STATISTICAL ANALYSIS

Despite the possible identification to lower taxonomic, based on morphological analysis, some groups were joined together for analysis, these were Coleoptera (other Coleoptera and Carabidae), Hymenoptera (other Hymenoptera and Formicidae), and Myriapoda (Chilipoda and Diplopoda). The larvae and unidentifiable roups were not considered for further analysis due to the inherent challenges associated with their accurate identification.

Regarding the analysis of metabarcoding data, some OTUs were not considered, as they compromised groups that did not belong to the soil macrofauna group or did not gather enough information. These were groups belonging to phyla: Actinobacteria; Amoebozoa; Ascomycota; Basidiomycota; Bryophyta; Chlorophyta; Cnidaria; Echinodermata; Heterokontophyta; Magnoliophyta; Ochrophyta; Proteobacteria; Rhodophyta; Rotifera; Zygomycota and no matches. OTUs from the phylums Arthropoda, Mollusca, Nematoda, and Annelida required a

more careful assessment. The Class Collembola and the orders Mesostigmata; Sarcoptiformes; Trombidiformes (class Arachnida) were also removed, as these are considered mesofauna. Similarly, flying insects such as Diptera, Lepidoptera; Neuroptera; Strepsiptera; Thysanoptera were not considered. OTUs only with taxonomic assignments higher than Class were not considered for analysis. Only OTUs above 85% similarity were used for the analysis.

Methodology Assessment

Carabid species-level taxonomic assignments were used for methodological comparisons between traditional, barcoding, and metabarcoding methods. Regarding molecular-based approaches, species-level identifications were only considered for similarity values above 98.5% (da Silva et al., 2019).

To assess whether sample richness between decontaminated / nondecontaminated groups was comparable, a two-sample t-test between groups was performed for each type of agroecosystem.

Diversity Indices

Both morphotaxonomic and metabarcoding data were transformed into sample-based (incidence) data (Chao et al., 2014), thus each pitfall trap samples were the sampling unit, and not the individual or each OTU. Therefore, two species-by-sampling-unit incidence matrix were formed by the presence/absence (non-detection) of each order/class within each sampling unit. For morphological analysis , diversity estimates were based on the effective number of taxonomic groups present, for the analysis based on metabarcoding data, estimates were made based on OTUs presence.

Hill numbers were used here to characterize the taxonomic diversity of the different agroecosystems (Chao et al., 2014). The first three Hill numbers: species richness ($q = 0$), the exponential of Shannon's entropy index ($q = 1$), and the inverse Simpson's concentration index ($q = 2$) (Chao et al., 2014) were calculated for each methodological approach. These differ on how rare species are scaled (Roswell et al., 2021), sensitivity towards rare species decreases with higher q orders. Thus, when $q = 0$, the analysis is highly sensitive to rare species, while with $q = 2$ this sensitivity is highly decreased, when $q = 1$ it lies midway in between the two (Roswell et al., 2021).

Sample size-based rarefaction and extrapolation curve, for orders $q = 0$ were visualized to assess sampling completeness. It is considered that the observed diversity is equal to the "true" diversity if the sample size-based rarefaction and extrapolation curve stabilizes, approaching an asymptote, and if sample coverage is close to one (Roswell et al., 2021).

Agroecosystems Biodiversity Composition

For both traditional and molecular methods, a generalized linear model (GLM), following a poisson family, was used to assess diversity differences within taxonomic groups at each agroecosystem. Identification based on morphological traits compared the groups with lowest taxonomic level possible, while OTU based analysis allowed the comparison of different taxonomic levels, Class and order, present per site compared.

To compare OTUs composition between the different sampling sites, a Jaccard distance matrix based on OTU richness records per sample was calculated. Subsequently, to identify differences between agroecosystems, permutational multivariate analysis of variance (PerMANOVA) with 999 permutations was performed coupled to a pairwise Pillai test with Benferroni p-value corrections to test differences between sites.

All statistical analyses and data visualization were performed in R Version 2023.06.0 + 421 using packages' iNEXT' (v3.0.0.) (Hsieh et al., 2016), for diversity estimates, ggplot2 v3.4.2 and ggpubr, for data visualization, vegan v2.6.2 (Oksanen et al., 2015) and RVAideMemoire v0.9-83 (Herve, 2023) for the permutational multivariate analysis of variance.

RESULTS

Based on morphological characteristics, from the 192 samples, 9418 individuals from the targeted groups, 7343 ($\bar{x}$ =38.484 $\pm$ 41.506) belonged to 13 different taxonomic groups, of different levels: 974 Aranae $\bar{x}$ =5.073 $\pm$ 4.520); 22 Archaeognatha ($\bar{x}$ =0.115 $\pm$ 0.527); 22 Blattodea ($\bar{x}$ = 0.115 $\pm$ 0.500); 1810 Coleoptera ($\bar{x}$ =9.427 $\pm$ 6.117); 38 Dermaptera ($\bar{x}$ = 0.198 $\pm$ 0.525); 150 Gastropoda ($\bar{x}$ = 0.781 $\pm$ 3.018); 521 Hemiptera ($\bar{x}$ = 2.714 $\pm$ 5.072); 3375 Hymenoptera ($\underline{x}$ = 17.578 $\pm$ 3.810); 137 Myriapoda ($\bar{x}$ =0.714 $\pm$ 1.334); 121 Isopoda ($\bar{x}$ = 0.630 $\pm$ 2.542); 134 Opilione ($\bar{x}$ = 0.698 $\pm$ 1.729); 28 Orthoptera ($\bar{x}$ = 0.146 $\pm$ 0.474) and 11 Pseudoscopiones ($\bar{x}$ = 0.057 $\pm$ 0.259).

The 220 sequenced samples that contributed with molecular operational units totalized 1625 OTUs (11522252 reads), with an average of 52373.873 ± 33134.384 (± SD, standard deviation) reads per PCR product. Of these only 716 OTUs (9643624 reads), with an average of 43834.655 ± 32145.006 (± SD) reads per PCR corresponded our target group, soil macrofauna, and were used for the under the scope of this study which reflected on the presence of seven classes and 22 orders identified by metabarcoding. Figure 6 summarises the number of OTU richness per order identified.

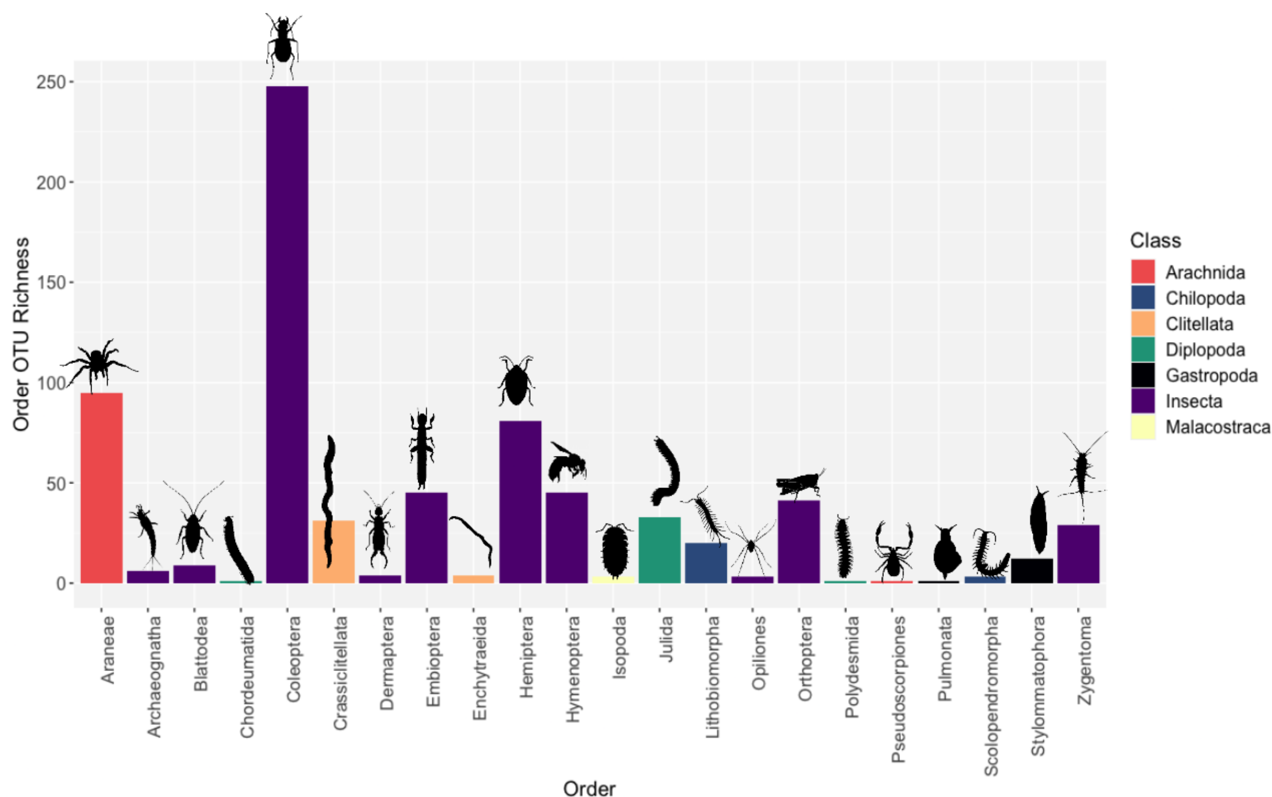


Figure 6: Number of OTUs by order, colored by Class. Arachnida: Araneae (95), Pseudoscorpiones (1), Arachnida; Insecta: Archaeognatha (6), Blattodea (9), Coleoptera (248) s, Dermaptera (4), Embioptera (45), Hemiptera (81), Hymenoptera (45), Opiliones (3), Orthoptera (41), Zygentoma (29); Chilopoda: Lithobiomorpha (20), Scolopendromorpha(3); Diplopoda: Chordeumatida (1), Julida (33), Polydesmida (1); Malacostraca: Isopoda (3); Annelida, Clitellata: Crassiclitellata (31), Enchytraeida (4); Mollusca, Gastropoda: Pulmonata (1), Stylommatophora (12). Silhouettes extracted from PhyloPic (www.phylopic.org).

METHODOLOGICAL COMPARISON

Decontamination

The KOH decontamination method did not show significant differences ($p > 0.05$) in the number of OTU richness at most agroecosystems, besides at the Traditional Olive Grove site where $p \leq 0.05$ ($p = 0.0137$), Figure 7 summarises the t-test results at each site. P-values and number of samples decontaminated/non-decontaminated are exposed in the Supplementary Materials (Table S1).

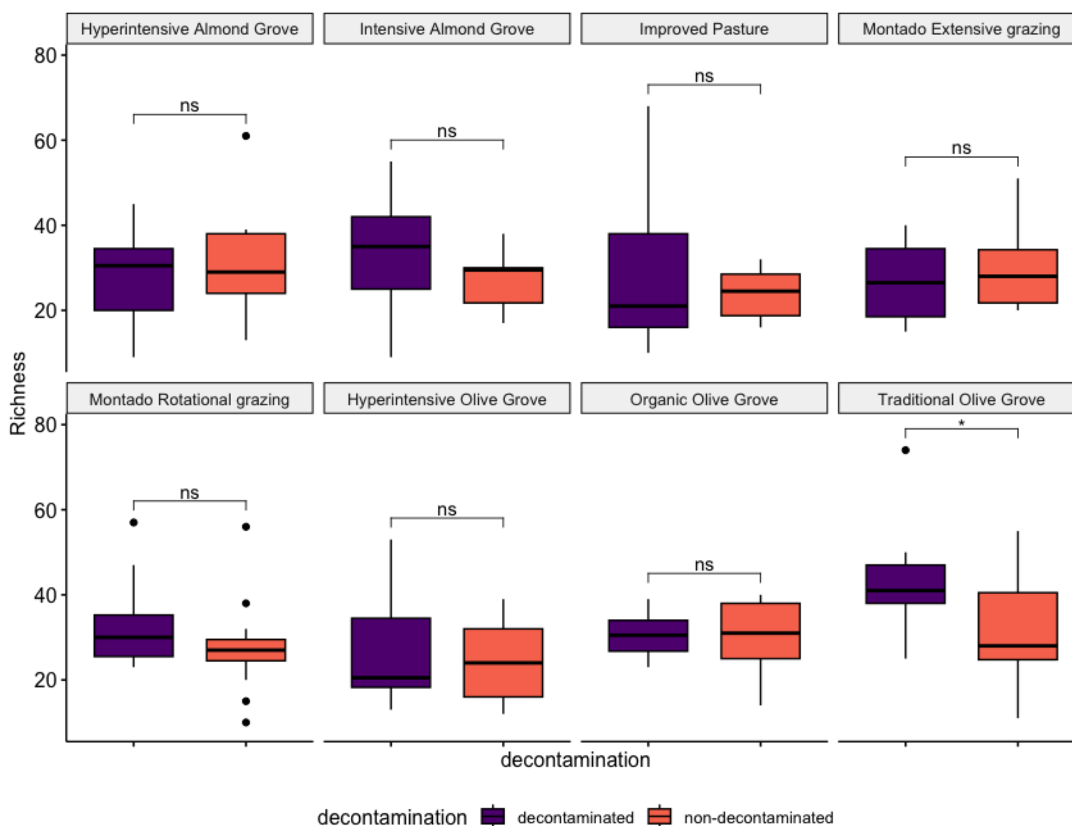


Figure 7: Boxplots of OTUs richness for decontaminated/non-decontaminated samples for the different agroecosystems. The results of a - test are represented in each graph, 'ns' stands for non-significant differences (p-value > 0.05), and '' stands for statistically significant differences (p-values ≤ 0.05).*

Using a Carabidae Mock Community to Evaluate OTU Recovery

For the methodological comparison, carabids were the target group. Regarding morphotaxonomy, 32 different species were identified, while barcoding and metabarcoding identified 17 and 16, respectively (table 2). Only eight species were equally identified by the three methods, and 17 species were only identified morphologically.

Table 2: Carabidae species identified by watch method, morphotaxonomy, barcoding, and metabarcoding. Regarding OTU based approaches, only species with a > 95.8 similarity percentage were considered.

Family	Genus	Species	Morphotaxonomy	Barcoding	Metabarcoding	Comments
Carabidae	Acupalpus	*A. cantabricus*	0	1	1	Possible taxonomic conflict
	Agonum	*A. muelleri*	1	1	1	
		A. nigrum	1	1	1	
	Amara	*A. aenea*	1	1	1	
	Bembemdion	*B. lampros*	1	0	0	
		B. tethys	1	0	1	
	Brachinus	*B. bodemeyeri*	1	0	0	
		B. sclopeta	0	0	1	
	Calathus	*C. circumseptus*	1	1	1	
		C. granatensis	1	1	1	
		C. melanocephalus	1	0	0	
		C. mollis	1	1	0	Not integrated in the mock community sample
	Calosoma	*C. maderae*	1	0	0	
	Carabus	*C. rugosus*	1	0	0	
		C. lusitanicus	1	0	0	
		C. melancholicus submeridionalis	1	1	0	
		C. macrocephalus	0	0	1	
	Harpalus	*H. distinguendus*	1	0	1	
		H. oblitus patruelis	1	0	0	
		H. rubripes	1	1	0	
		H. rufipes	0	0	1	
		H. smaradinus	0	1	0	
	Leistus	*L. fulvibarbis*	1	1	0	
	Licinus	*L. punctulatus*	0	1	1	Possible taxonomic conflict
		L. punctatulus granulatus	1	0	0	
	Nebria	*N. salina*	1	1	1	
	Notiophilus	*N. geminatus*	1	0	0	
		N. marginatus	1	0	0	
		N. quadripunctatus	1	1	0	
	Olisthopus	*O. hispanicus*	1	0	0	
	Ophonus	*O. subquadratus*	1	0	0	
	Poecilus	*P. purpurascens*	1	0	0	
	Pseudophonus	*P. rufipes*	1	0	0	
	Pterostichus	*P. globosus*	0	1	1	Possible taxonomic conflict
	Steropus	*S. globosus ebenus*	1	0	0	
	Syntomus	*S. foveatus*	1	1	1	
	Tachys	*T. algiricus*	1	0	0	
	Trechus	*T. obtusus*	1	1	1	
		T. quadristriatus	1	0	0	
	Total		32	17	16	

Of the 31 DNA extractions added to the mock sample, after sequencing and filtering, it returned 66 different OTUs (Figure 8), meaning that more OTUs were identified than the ones added. Overall, the proportion of OTUs recovered by the complex samples was 0.675 ± 0.066 (±SD) (Supplementary Material, table S2).

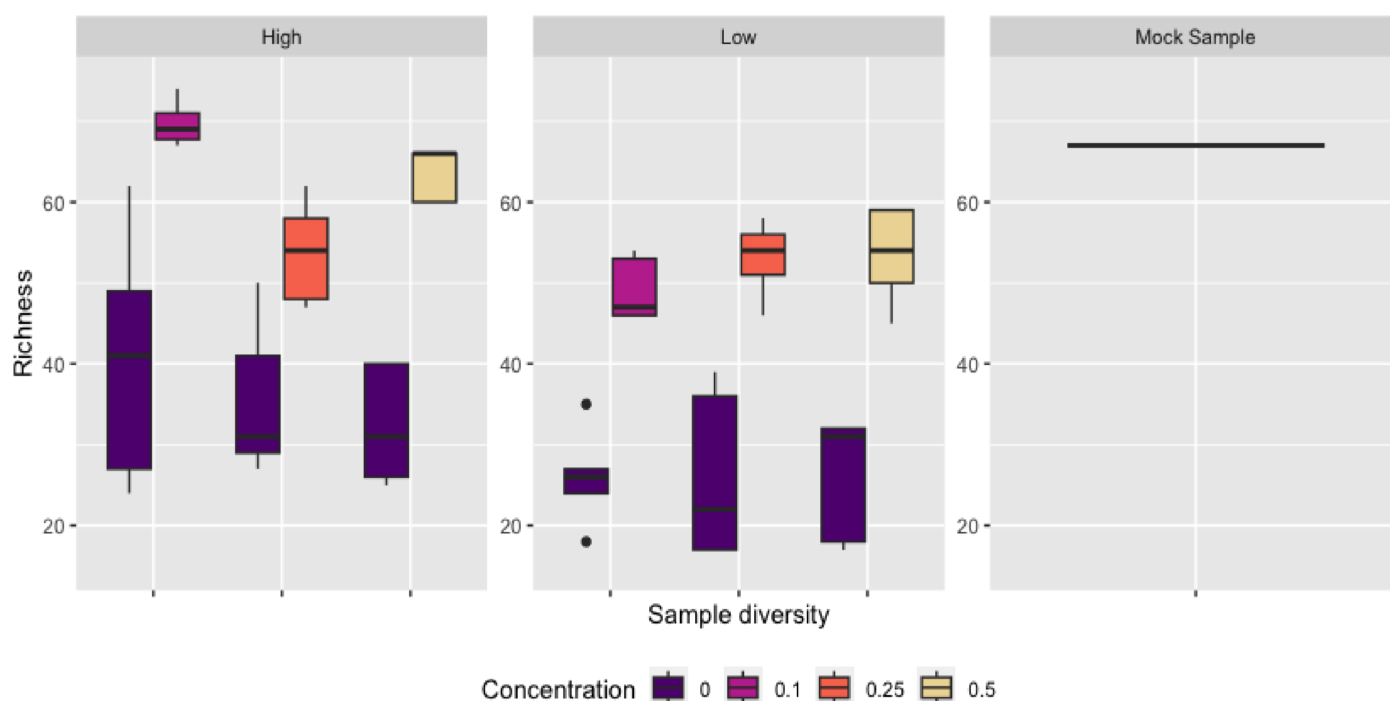


Figure 8: Boxplots of OTU richness of each diversity (high / low) group (n = 15). Samples are divided by their diversity, high or low, based on the morphological identification of specimens and for comparison, the mock sample (n = 1). Within their diversity, samples differ in the proportion of added mock sample: 0 (n = 5), 0.1 (n = 5); 0.25 (n = 5) and 0.50 (n = 5).

Hill Numbers for Diversity Assessment and Sample Completeness

Biodiversity Assessment Based on Morphotaxonomy

The Hill number of order q = 0 (species richness), Figure 9 (A), demonstrates that Hyperintensive and Traditional Olive Grove have the highest group diversity per sampling unit. Besides the Hyperintensive and the Organic Olive Grove sites, all sites reach a plateau with interpolation, demonstration sampling completeness. The Hill number of order q = 0, Figure 9 (B), sample completeness of morphological-based analysis reached a plateau in every

agroecosystem (Figure 9 (C)), furthermore, the interpolation and extrapolation curve of the sample coverage (SC) of most sites reach SC = 1).

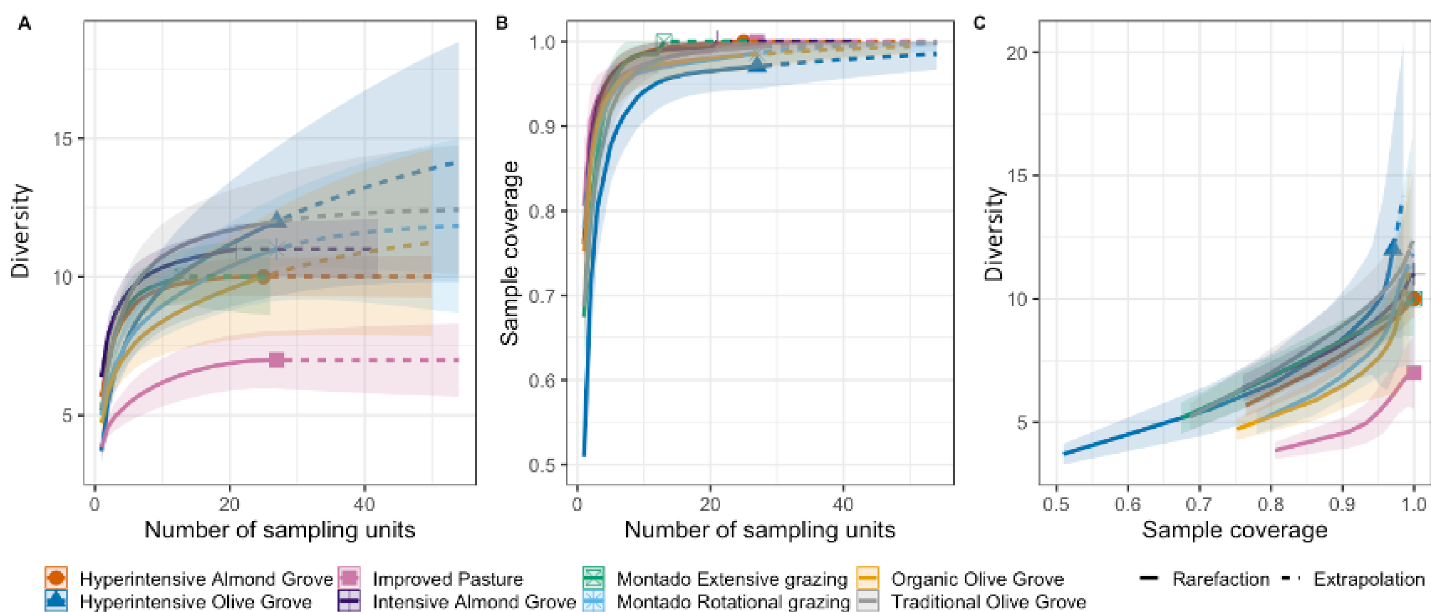


Figure 9: Hill number plots q = 0 (species richness) of the groups morphologically identified by sampling unit based on an incidence matrix per sampling site. Solid lines represent interpolation, while dotted lines represent extrapolation. (A) Sample-size-based curves; (B) Sample completeness curve: accumulation of species with increasing sampling effort; (C) Coverage-based curves: expected diversity as a function of expected coverage.

The Hill number of order q = 1 and q = 2 are less influenced by rare species, thus sample completeness is reached in both and the confidence interval is much more compact (Figure 10).

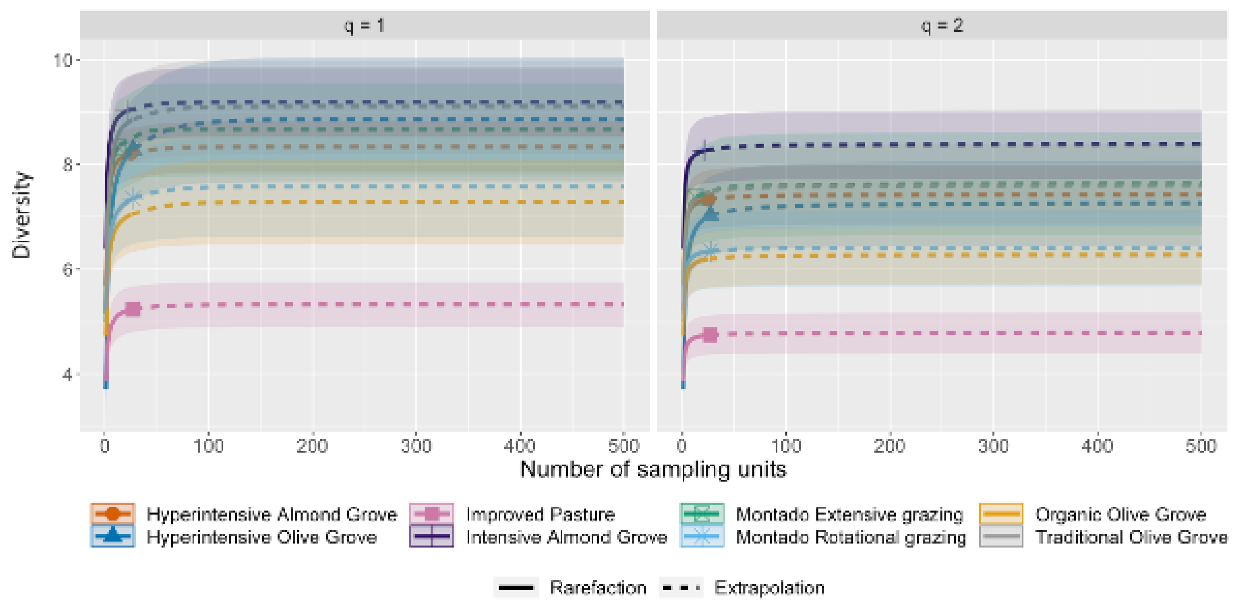


Figure 10: Morphologically identified groups diversity sample size-based rarefaction (solid line) and extrapolation (dotted line) for each sampling site for Hill number order q = 1 (Shannon's entropy index), left panel and q = 2 (inverse Simpson's concentration index), right panel. Shaded areas represent a confidence interval of 95%.

Regarding the assessment of morphologically identified groups (Figure 11; Supplementary Materials, table S3) the total mean abundance (Figure 11 (A)) was higher at the Hyperintensive Almond Grove (67.360 ± 72.995 SD), being significantly different than any other sampling site (Supplementary Materials). It is followed by the Intensive Almond Grove (44.190 ± 21.669 SD) and the rotational montado (41.074 ± 18.953 b). The Montado extensive (22.692 ± 14.471 SD) and Hyperintensive Olive Grove (21.296 ± 34.822 SD) were the least mean abundant sites. Taking into account the identified groups, Aranea, Coleoptera and Hymenoptera were the most abundant taxa in all sites.

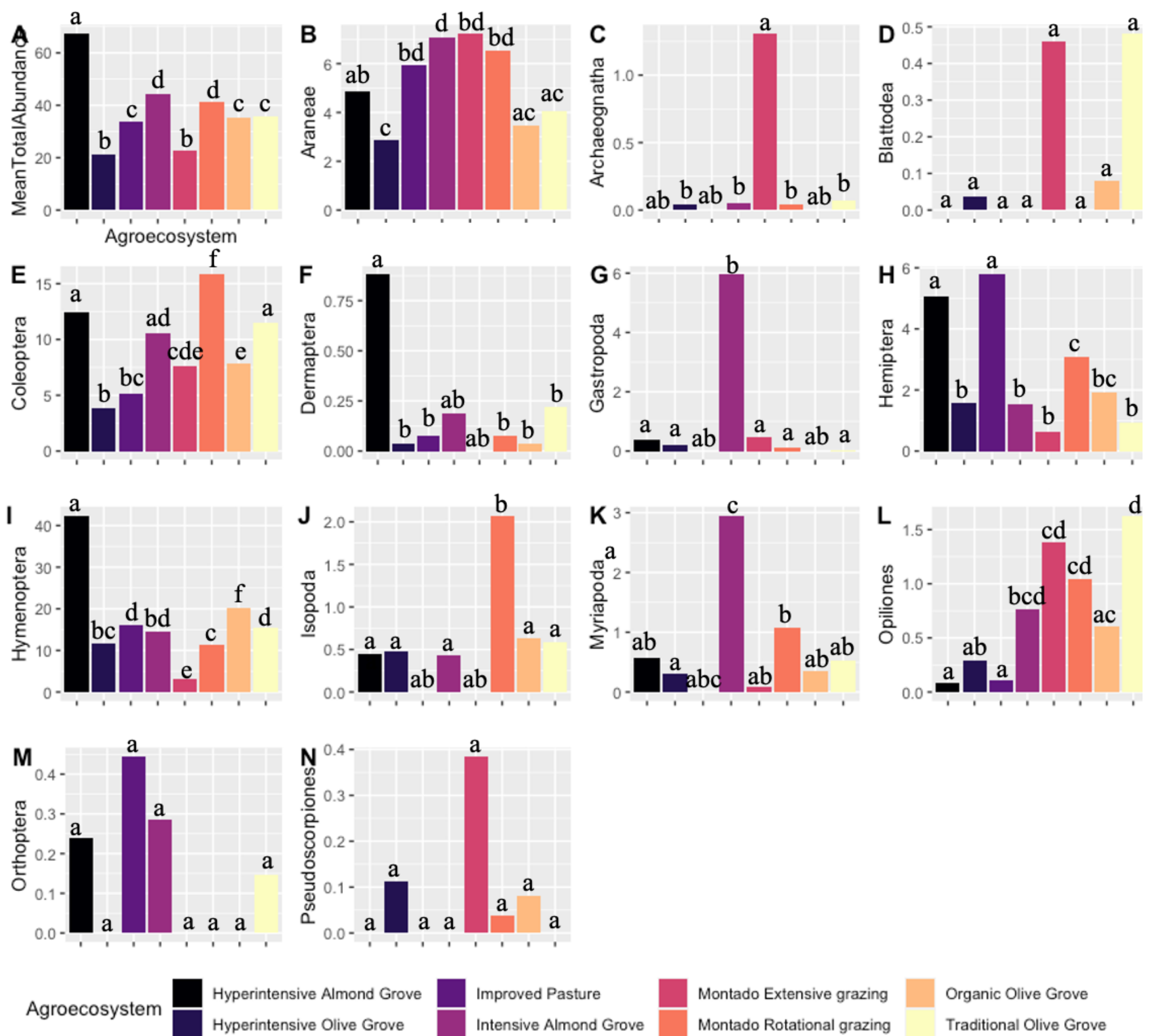


Figure 11: Histograms of mean abundance of groups identified using traditional taxonomy, per sample in each agroecosystem. (A) Total mean abundance containing all groups; (B) mean of Araneae; (C) mean of Archaeognatha; (D) mean of Blattodea; (E) mean of Coleoptera; (F) mean of Dermaptera; (G) mean of Gastropoda; (H) mean of Hemiptera; (I) mean of Hymenoptera; (J) mean of Isopoda;(K) Mean of Myriapoda; (L) mean of Opiliones; (M) mean of Orthopetra; (N) mean of Pseudoscorpiones. Due to highly different mean abundances consider the y axis discrepancies. Different letters indicate significantly different mean abundances (p ≤ 0.05) per agroecosystem based on a GLM model with Poisson family with a post hoc Tuckey test with 0.95 confidence interval in the pairwise tests, with a 0.05 significance level.

Biodiversity Assessment Based on OTUs

Based on incidence frequency, the richness of OTUs was compared across the different sampling sites, considering Hill numbers of order q=0. The Tradicional Olive Grove shows the highest diversity, followed by the Montado with rotational grazing and the organic olive Grove (Figure 12 (A)). On the other hand, the improved pasture is the least diverse site regarding OTUs order richness. Intensive and Hyperintensive almond Grove follow similar interpolation and extrapolation curves. Sample completeness of OTU based analysis did not reach a plateau, (Figure 12 (B)), furthermore the closer the coverage based values are to one the better are the sites represented by the sampling effort, in Figure 12 (C).

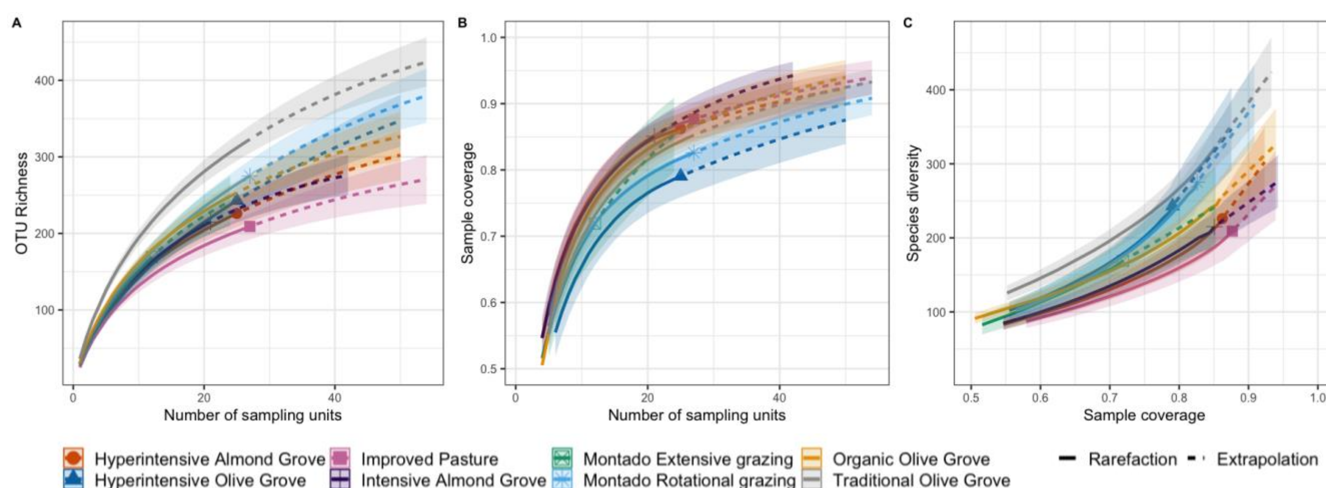


Figure 12: Plots of Hill number q = 0 (species richness) of the richness of OTUs identified groups by sampling unit based on an incidence matrix per sampling sites. Solid lines represent interpolation, while dotted lines represent extrapolation. (A) Sample-size-based curves; (B) sample completeness curve: accumulation of species with increasing sampling effort; (C) coverage-based curves: expected diversity as a function of expected coverage.

Regarding the Hill number of order q = 1 and q = 2 (Figure 13), sample completeness is reached in both.

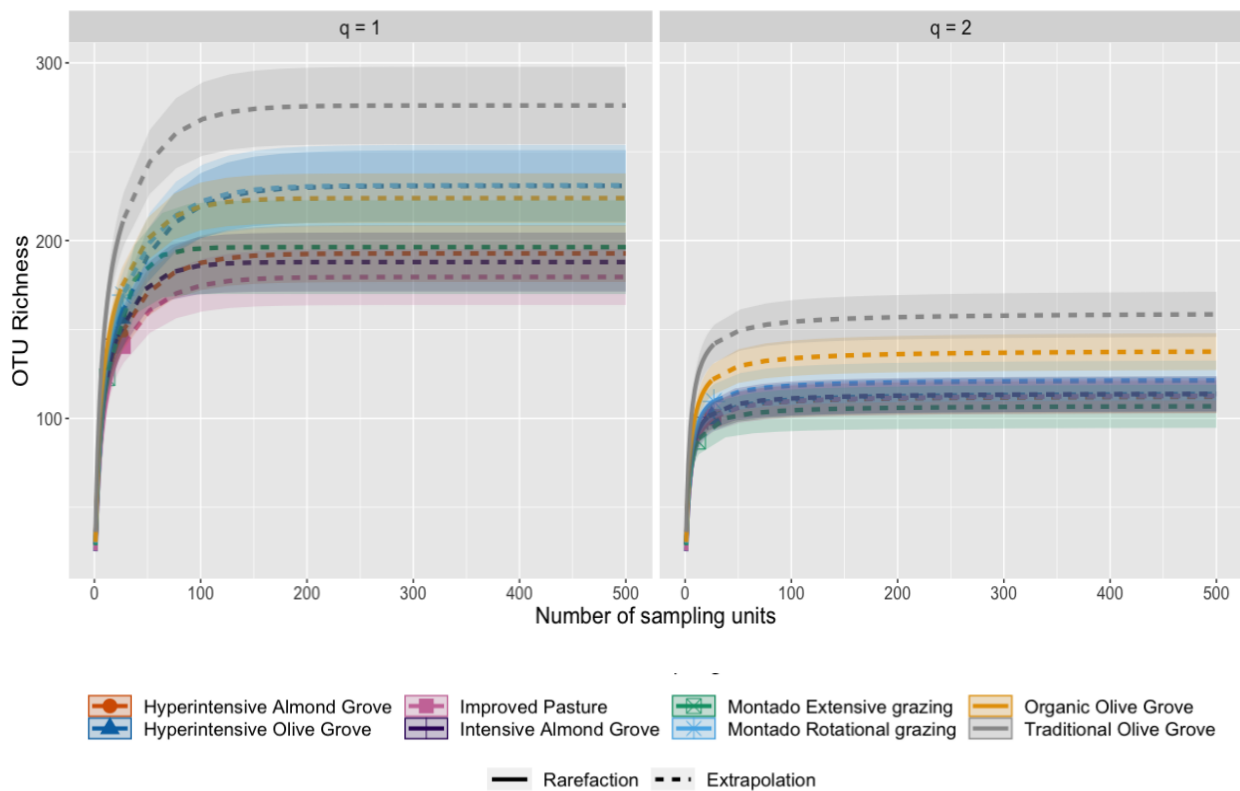


Figure 13: OTU diversity sample size-based rarefaction (solid line) and extrapolation (dotted line) for each sampling site Hill number order q = 1 (Shannon's entropy index), left panel, and q = 2 (inverse Simpson's concentration index), right panel. Shaded areas represent a confidence interval of 95%.

Considering OTU richness by Class (Figure 14 (A); Supplementary Material table S4) the traditional olive Grove was the site with highest OTU richness (36.0741 ± 13.255 SD), followed by the organic olive Grove, (30.400 ± 7.101 SD), followed rotational Montado (29.926 ± 10.433 SD), intensive almond Grove (29.571 ± 11.016 SD) and the Hyperintensive almond Grove (28.920 ± 11.832 SD).

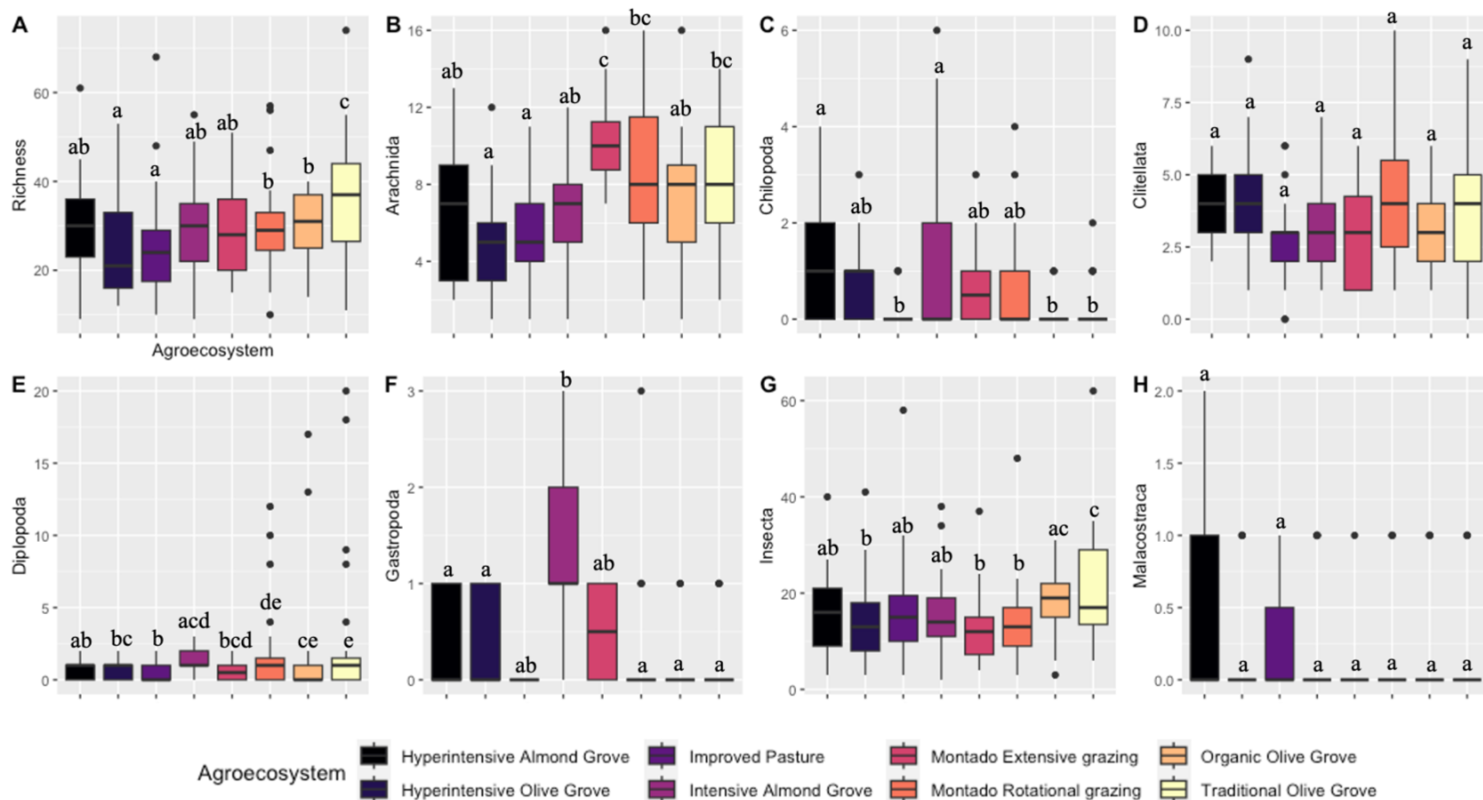


Figure 14: Boxplots of OTU richness by Class identified per sample at each sampling site. GLM model with Poisson family, post hoc Tuckey test, 0.95 confidence interval in the pairwise tests, compact letter display with a 0.05 significance level. (A) Total OTU richness; (B) Arachnina OTU richness; (C) Chilipoda OTU richness; (D) Clitellata OTU richness; (E) Diplopoda OTU richness; (F) Gaspropoda OTU richness; (G) Insecta OTU richness; (H) Malacostraca OTU richness. Due to highly different OTUs richness levels, consider the y axis discrepancies. Different letters indicate significantly different mean abundances (p ≤ 0.05) per agroecosystem based on a GLM model with Poisson family with a post hoc Tuckey test with 0.95 confidence interval in pairwise tests, with a 0.05 significance level.

Permutational multivariate analysis of variance (PerMANOVA) indicated that the diversity of OTUs per sample differed significantly according to the agroecosystem type (Supplementary Materials, table S5), this factor explains 9.8% of the total variance of OTUs ($p < 0.001$). The Pillai pairwise post hoc test showed significant differences in all paired comparisons of agroecosystems ($p < 0.05$). The dispersion of the sites regarding OTU richness per sample can be visualized in Figure 15.

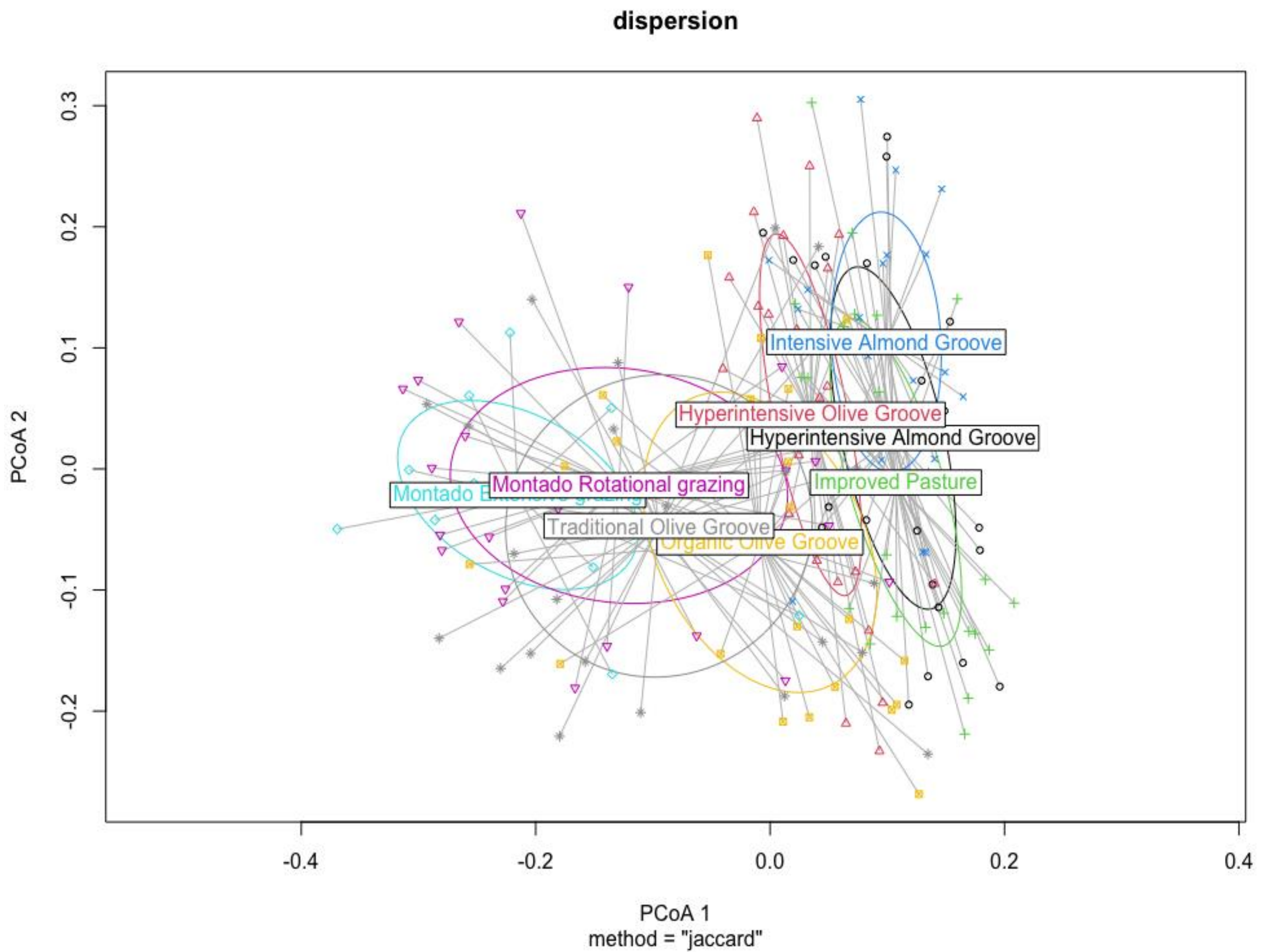


Figure 15: Permutational multivariate analysis of variance (PerMANOVA) based on Jacccard distances of OTU richness per sample, with 95% confidence ellipses for each site.

DISCUSSION

THE VALUE OF MORPHOTAXONOMY AND ITS CHALLENGES

Using an integrative strategy, researchers can gain a thorough understanding of species richness and composition by combining the advantages of morphotaxonomy (traditional taxonomic identification based on physical characteristics) and building on the knowledge that has been developed through past centuries with molecular techniques (such as DNA barcoding or metabarcoding) (Cristescu, 2014).

Our results show that only eight species were common in all three methodological approaches, demonstrating the difficulty in obtaining a reliable identification (table 2). In addition, 17 species were only identified morphologically, which may be due to the lack of available sequences on the databases, with no match being found. On the contrary, some species were only identified based on one or both molecular methods, which may indicate some taxonomic conflicts. Taxonomic conflicts can be due to the lack of well-curated databases leading to erroneous identification-based genetic sequences or the existence of highly difficult and ambiguous species to identify. On the one hand, DNA-based methods allow for the detection and analysis of genetic data, enabling precise species identification, exceptionally when morphological identification may be complex or ambiguous. However, this relies on available genetic data for a given species and proper curation of those sequences. On the other hand, highly specialized and trained researchers can offer physical validation and insight to reach the species level.

Primer bias can be an issue concerning molecular approaches (Elbrecht et al., 2019; Taberlet et al., 2018). The choice of primer pairs could influence results, as some primers may have more affinity and higher PCR efficiency toward specific species or taxa, possibly competing with others (Elbrecht et al., 2019; Taberlet et al., 2018). This might become even more difficult when the target group is a wide variety of taxa, such as when dealing with bulk samples. Thus, choosing primers for DNA metabarcoding is deeply based on previous knowledge of the target group and the characteristics of the available primers. Here, the Folmer primer pairs (Folmer et al., 1994) were the choice for the barcoding approach, while the BF3 and BR2 primer pairs (Elbrecht et al., 2019) for metabarcoding. Both recovered carabid species level OTUs, with high similarity percentages. Still, some discrepancies were found, as some species were not recovered equally, being present in one but not in the other method. This might exhibit some bias associated with the primer pair complementarity or even due to primer slippage and should be further explored (Elbrecht et al., 2018).

Despite the usefulness and time expenditure when compared to traditional methods, the use of molecular-based approaches through the use of public databases such as BOLD or NCBI may lead to mismatches between morphological and molecular methods, as some information may not have been curated by taxonomy specialists(Collins & Cruickshank, 2013). These problems have been recognized for a long time; however, they still lead to incongruences (Waugh, 2007). The use of DNA-based species identification methods is growing, and these difficulties should not be ignored, as they can lead to identification errors and, therefore, reduce results quality and reliability. To overcome the challenges found in each methodology, the alliance between morphotaxonomy knowledge with molecular accurate data must be fostered to curate the already published databases, enhancing their reliability (Janzen et al., 2005). Furthermore, gathering joint efforts to publish new molecular data of prime quality would significantly improve the quality of molecular-based research.

The cases where both the barcoding and metabarcoding match a different species from what was identified using classical taxonomy may reveal taxonomic conflicts. This offers an opportunity for full cooperation with taxonomy leaders and potentially revealing hidden diversity as the presence of a complex of cryptic species, not uncommon in soil invertebrates (Cunha et al., 2014; Hlebec et al., 2023; Janzen et al., 2005; King et al., 2008; Novo et al., 2010). The latest can be due to the difficulties underpinning morphologically identifying the highest taxonomic level, such as subjective parameters, e.g. colours. Furthermore, this could also happen when the OTUs present in the used database (BOLD and NCBI) are wrongly assigned to the species or there is a conflict with the identification based on morphology (Collins & Cruickshank, 2013).

These results highlight the importance of setting highly integrative approaches to overcome the pitfalls inherent to each method, thus allowing the correct assignment of species to a molecular sequence (Conrado et al., 2023). The molecular approach will become progressively more reliable with the development of high-quality reference databases, namely local reference databases. It has also been suggested to provide, coupled with the DNA barcode of the identified specimens, key characters and a bibliography for identifying each putative species (Cao et al., 2016; Collins & Cruickshank, 2013).

Science is not and should never be static. New and better methods will eventually surge, and researchers should be open to these and help build up new knowledge and methods based on those previously used to provide new technologies with the reliability needed.

Traditional and Organic Agroecosystems Hold High Levels of Biodiversity

Any effective conservation strategy relies on the availability and reliability of methods that allow us to truly understand the underlying effects and reasons for possible biodiversity loss and pinpoint the areas in danger. This study revealed that molecular-based approaches disclose different conclusions regarding species diversity richness according to the agroecosystem.

Different agroecosystem sites are identified as the most biodiverse by morphological-based techniques and metabarcoding (Figures 9 and 12); the first method identified Traditional and Hyperintensive Olive Groves, while the second has Traditional Olive Grove and the Montado with rotational grazing. Despite the shared result, in which both biodiversity assessment methods demonstrate higher biodiversity at the Traditional Olive Grove, the permutational analysis of variance (Figure 15) further distinguishes them concerning the diversity of OTUs per sample. These findings are in accordance with the literature, which states that extensive management practices used in agroforestry and traditional methods, are considered to hold higher biodiversity levels (Brussaard et al., 2007; Mupepele et al., 2021; Torralba et al., 2016) and our results based on OTUs diversity, are in accordance. Agroecosystems that rely on more extensive practices display higher diversity levels when compared to improved pasture and intensively managed agroecosystems (Bender et al., 2016; Mupepele et al., 2021; Tsiafouli et al., 2015). Improved pasture systems showed, in both approaches, the lowest diversity. These improved pastures are characterized by their landscape homogeneity with no to very few vegetation layers that could contribute to lower levels of biodiversity. Traditional agro-silvio-pastoral practises, such as Montado, are related to low disturbance levels and to an extensive production system of several production activities (Pinto-Correia et al., 2011), compiling a heterogeneous landscape, thus providing different ecological niches necessary for the establishment of other species (FAO, 2020; Stolte et al., 2016).

Moreover, while the first method showed sampling completeness as sample completeness reached a plateau and the sample coverage (SC) reached SC = 1 for most samples (Figure 9 (B) and (C)), metabarcoding revealed the necessity of intensifying sampling efforts to reveal the true biodiversity of the systems, as sample completeness did not reach a plateau nor did the sample coverage (SC) interpolation or extrapolation curve reach SC = 1 (Figure 12 (B)) and (C)) demonstrating insufficient samples to reveal the OTUs "true" richness of the different agroecosystems.

This demonstrates that traditional species identification and surveying approaches lack depth, thus being unable to fully represent the diversity in agroecosystems.

Challenges and Opportunities for Improvement of DNA-Based Approaches

Studies have shown that sodium bleaching procedures do not affect the sequencing success of COI barcoding (Hausmann et al., 2021), thus being used in the decontamination process of environmental DNA samples (Liu et al., 2020). These steps ensure the recovery of the targeted OTUs and control possible contamination, most notably between samples. This procedure can be time-consuming and, as our results suggest, not significantly affect the number of OTUs reads in most of the agroecosystem types studied (Figure 7). Accordingly, including or not this step depends on the research question, the sampling design, field and laboratory conditions, and of time availability. Regardless, it is beneficial to increase the reliability of the sequenced data and results. In that sense, whenever possible, this step should not be overlooked. A negative control sample should have been added since the field sampling to assess better the existence of field, laboratory, and cross-sample contamination (Liu et al., 2020).

The Carabidae mock community was employed to validate the metabarcoding pipeline. The results demonstrated that while both barcoding and metabarcoding techniques identified carabid species, some discrepancies were found, highlighting the importance of integrating different methods for accurate species identification (Janzen et al., 2005). The use of DNA-based methods, such as metabarcoding, revealed higher levels of biodiversity in traditional and organic agroecosystems compared to improved pastures, showcasing the potential of molecular approaches to unveil unprecedented levels of soil biodiversity and even revealing levels of' hidden diversity such as the presence of cryptic speciation.

OTU Communities as Determinants of Agroecosystem Separation and Clustering

According to the permutational multivariate analysis of variance (perMANOVA), all agroecosystems have significantly different OTU communities (Supplementary Materials). Moreover, it is possible to observe in Figure 15 that sites extensively managed, such as Montados, the Traditional and Organic Olive Groves, are more similar to each other than the more intensively managed ones clustered together in a different area of the plot. This demonstrates that according to management practises, the communities held at each agroecosystem will differ, hosting different biological communities.

Future Research: Implications for Conservation and Agroecosystems Management

The data collected within the scope of this research still has excellent potential for further analysis with promising results. Several pathways can be explored in future research, such as:

1. Exploration of Taxonomic Group Impacts Based on Its Functions: Further analysis should be conducted to identify which taxonomic groups and associated functions most influence the characterization of different agroecosystems and ultimately affect the productivity of the systems and delivery of ecosystem services. Understanding the impact of specific taxa on the agroecosystem dynamics can inform targeted conservation efforts and management strategies. Furthermore, a more in-depth analysis of the identified OTUs should be performed to identify species with high ecological and agricultural value. Species with roles as ecosystem engineers or pests could have significant implications for agricultural practices and ecosystem functioning. In addition, performing this analysis considering different taxonomic levels, such as genus and family, could be interesting to explore if the general pattern observed in Figure 15 is similar to a broader view of the communities present.
2. Integration of Environmental and Management Practises: Collecting additional information on environmental and management practises in the studied agroecosystems will allow for the assessment of their influence on biodiversity richness and community composition. Understanding the relationship between agricultural practices and soil biodiversity can lead to more sustainable and biodiversity-friendly farming methods.
3. Concentration Effects in the Mock Community: The significance of the concentration of the added mock sample should be further evaluated to understand potential PCR biases better. Adjusting the concentration levels in the mock community could provide valuable insights into the reliability of the metabarcoding pipeline.
4. Further Comparison with Similar Studies and Increasing Genetic References: To establish a robust metabarcoding approach for soil fauna, more comparisons with other large functional groups, such as ants and spiders or understudied taxa, should be conducted, including increasing the number of species with available barcodes in genetic databases. Assessing the proportion of unidentified organisms below the order/class level and comparing it with similar studies can help identify potential areas for improvement.

CONCLUSION

This study used an integrated approach to investigate soil macrofauna in various agroecosystems using both morphological taxonomy and molecular methods. A broad analysis of 192 samples revealed 9418 individuals representing 13 different taxonomic groups. Additionally, 716 operational taxonomic units (OTUs) related to soil macrofauna were identified from the 220 sequence samples. Although the findings of this research are significant, they suggest potential for further study and a deeper understanding. One promising avenue of future research will be to study the communities at each site defined at OTU level and determine their impact on separating and clustering the agricultural ecosystems. Coupled to this, collecting more information to integrate environmental conditions and management practises to assess possible influences on biodiversity richness, and communities could give valuable information on the impacts of agricultural management practices on soil biodiversity, thus providing insights into sustainable land management and conservation practices.

This study also sheds light on the challenges and pitfalls of using molecular-based approaches. Primer bias and discrepancies between morphological and molecular methods were observed, emphasizing, once more, the need for highly integrative approaches and the curation of existing databases to ensure reliable species assignments. Cooperation between researchers and taxonomy experts is crucial to address taxonomic conflicts and reveal hidden diversity, such as cryptic species complexes. Furthermore, the comparison between decontaminated and non-decontaminated samples showed no significant differences in most agroecosystems, emphasizing that careful decontamination steps should be taken if possible to ensure data accuracy, but might not be crucial.

This research offers a valuable foundation for future exploration and in-depth analysis. By taking an integrative approach, it is possible to gain a comprehensive understanding of soil biodiversity and its dynamics in different agroecosystems. Future research efforts should focus on refining and enhancing the metabarcoding pipeline, expanding the analysis to include other taxa, and investigating the influence of environmental and management practices on soil fauna diversity. Such research will contribute to our understanding of the complex interactions between agriculture and biodiversity and inform strategies for sustainable land management and conservation practices. By embracing new technologies and collaborative efforts, research can continue to advance the knowledge and application of molecular-based approaches for assessing biodiversity with particular relevance to soil fauna.

REFERENCES

Aguiar, C. A. S., & Serrano, A. R. M. (2012). *Coleópteros carabídeos (Coleoptera, Carabidae) de Portugal Continental: Chaves para a sua identificação*. Sociedade Portuguesa de Entomologia.

Arjona, Y., Arribas, P., Salces-Castellano, A., López, H., Emerson, B. C., & Andújar, C. (2022). Metabarcoding for biodiversity inventory blind spots: A test case using the beetle fauna of an insular cloud forest. *Molecular Ecology*, mec.16716. https://doi.org/10.1111/mec.16716

Arribas, P., Andújar, C., Salces-Castellano, A., Emerson, B. C., & Vogler, A. P. (2021). The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. *Molecular Ecology*, *30*(1), 48–61. https://doi.org/10.1111/mec.15591

Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, *515*(7528), 505–511. https://doi.org/10.1038/nature13855

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nat. Commun.* https://doi.org/10.1038/ncomms2881

Bender, S. F., Wagg, C., & van der Heijden, M. G. A. (2016). An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends in Ecology and Evolution*, *31*(6), 440–452. https://doi.org/10.1016/j.tree.2016.02.016

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, *18*(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Bourlat, S. J., Haenel, Q., Finnman, J., & Leray, M. (2016). Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Dual-PCR Method. *Methods in Molecular Biology (Clifton, N.J.)*, *1452*, 197–207. https://doi.org/10.1007/978-1-4939-3774-5_13

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

Brussaard, L., de Ruiter, P. C., & Brown, G. G. (2007). Soil biodiversity for agricultural sustainability. *Agriculture, Ecosystems and Environment*, *121*(3), 233–244. https://doi.org/10.1016/j.agee.2006.12.013

Cao, X., Liu, J., Chen, J., Zheng, G., Kuntner, M., & Agnarsson, I. (2016). Rapid dissemination of taxonomic discoveries based on DNA barcoding and morphology. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep37066

Cao, X., Xu, X., Bian, R., Wang, Y., Yu, H., Xu, Y., Duan, G., Bi, L., Chen, P., Gao, S., Wang, J., Peng, J., & Qu, J. (2020). Sedimentary ancient DNA metabarcoding delineates the contrastingly temporal change of lake cyanobacterial communities. *Water Research*, *183*, 116077. https://doi.org/10.1016/j.watres.2020.116077

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Collins, R. A., & Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, *13*(6), 969–975. https://doi.org/10.1111/1755-0998.12046

Conrado, A. C., Demetrio, W. C., Stanton, D. W. G., Bartz, M. L. C., James, S. W., Santos, A., da Silva, E., Ferreira, T., Acioli, A. N. S., Ferreira, A. C., Maia, L. S., Silva, T. A. C., Lavelle, P., Velasquez, E., Tapia-Coral, S. C., Muniz, A. W., Segalla, R. F., Decaëns, T., Nadolny, H. S., … Cunha, L. (2023). Amazonian earthworm biodiversity is heavily impacted by ancient and recent human disturbance. *Science of The Total Environment*, *895*, 165087. https://doi.org/10.1016/j.scitotenv.2023.165087

Corley, M., Ferreira, S., Lvovsky, A. L., Rosete, J., Sousa, P., & Beja, P. (2020). *The InBIO Barcoding Initiative Database: Borkhausenia crimnodes Meyrick, 1912 (Lepidoptera, Oecophoridae), a southern hemisphere species resident in Portugal*. https://doi.org/10.15468/8yge73

Creamer, R. E., Hannula, S. E., Leeuwen, J. P. V., Stone, D., Rutgers, M., Schmelz, R. M., Ruiter, P. C. de, Hendriksen, N. B., Bolger, T., Bouffaud, M. L., Buee, M., Carvalho, F., Costa, D., Dirilgen, T., Francisco, R., Griffiths, B. S., Griffiths, R., Martin, F., Silva, P. M. da, … Lemanceau, P. (2016). Ecological network analysis reveals the inter-connection between soil biodiversity and ecosystem function as affected by land use across Europe. *Applied Soil Ecology*, *97*, 112–124. https://doi.org/10.1016/j.apsoil.2015.08.006

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

Cunha, L., Montiel, R., Novo, M., Orozco-terWengel, P., Rodrigues, A., Morgan, A. J., & Kille, P. (2014). Living on a volcano's edge: Genetic isolation of an extremophile terrestrial metazoan. *Heredity*, *112*(2), 132–142. https://doi.org/10.1038/hdy.2013.84

da Silva, L. P., Heleno, R. H., Costa, J. M., Valente, M., Mata, V. A., Gonçalves, S. C., da Silva, A. A., Alves, J., & Ramos, J. A. (2019). Natural woodlands hold more diverse, abundant, and unique biota than novel anthropogenic forests: A multi-group assessment. *European Journal of Forest Research*, *138*(3), 461–472. https://doi.org/10.1007/s10342-019-01183-5

Doran, J. W., & Parkin, T. B. (1994). Defining and Assessing Soil Quality. In *Defining Soil Quality for a Sustainable Environment* (pp. 1–21). John Wiley & Sons, Ltd. https://doi.org/10.2136/sssaspecpub35.c1

Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, *2019*(10). https://doi.org/10.7717/peerj.7745

Elbrecht, V., Hebert, P. D. N., & Steinke, D. (2018). Slippage of degenerate primers can cause variation in amplicon length. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-29364-z

European Commission. (2021). *EU Soil Strategy for 2030: Reaping the benefits of healthy soils for people, food, nature and climate*. European Comission. https://www.eea.europa.eu/data-and-maps/dashboards/land-take-statistics#tab-based-on-data

FAO. (2020). *State of knowledge of soil biodiversity—Status, challenges and potentialities*. FAO. https://doi.org/10.4060/cb1928en

Ferreira, C. S. S., Seifollahi-Aghmiuni, S., Destouni, G., Ghajarnia, N., & Kalantari, Z. (2022). Soil degradation in the European Mediterranean region: Processes, status and consequences. *Science of The Total Environment*, *805*, 150106. https://doi.org/10.1016/j.scitotenv.2021.150106

Ferreira, S., Bohn, H., Sousa, P., & Beja, P. (2020). *The InBIO Barcoding Initiative Database: Blattodea 01*. https://doi.org/10.15468/m7f8sg

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. In *Molecular Marine Biology and Biotechnology* (Vol. 3, Issue 5, pp. 294–299).

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., Packer, M., Blaxter, M. L., Lambshead, P. J. D., Thomas, W. K., & Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, *1*(1), Article 1. https://doi.org/10.1038/ncomms1095

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-01312-x

Guerra, C. A., Bardgett, R. D., Caon, L., Crowther, T. W., Delgado-Baquerizo, M., Montanarella, L., Navarro, L. M., Orgiazzi, A., Singh, B. K., Tedersoo, L., Vargas-Rojas, R., Briones, M. J. I., Buscot, F., Cameron, E. K., Cesarz, S., Chatzinotas, A., Cowan, D. A., Djukic, I., Van Den Hoogen, J., … Eisenhauer, N. (2021). Tracking, targeting, and conserving soil biodiversity: A monitoring and indicator system can inform policy. *Science*, *371*(6526), 239–241. https://doi.org/10.1126/science.abd7926

Hausmann, A., Höcherl, A., Niessner, A., Zakharov, E., Dolynskyj, S., Hausmann, J. D., Höcherl, A., Zakharov, A., Dolynskyi, E., Diller, J., & Dolynskyi, S. (2021). Accurate decontamination of insects from bulk samples does not affect DNA sequencing success. *Spixana*, *44*(1), 71–76. https://doi.org/10.5883/DS-PANLARVA

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Herve, M. (2023). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics*. https://cran.r-project.org/web/packages/RVAideMemoire/index.html

Hlebec, D., Podnar, M., Kučinić, M., & Harms, D. (2023). Molecular analyses of pseudoscorpions in a subterranean biodiversity hotspot reveal cryptic diversity and microendemism. *Scientific Reports*, *13*(1), Article 1. https://doi.org/10.1038/s41598-022-26298-5

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

Illumina. (2013). *16S metagenomic sequencing library preparation guide*. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Janzen, D. H., Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., & Hebert, P. D. N. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1835–1845. https://doi.org/10.1098/rstb.2005.1715

Jeffery, S., Gardi, C., Jones, A., Montanarella, M., Marmo, Miko, L., Ritz, K., Pérès, G., Römbke, J., & H. van der Putten, W. (2010). *European atlas of soil biodiversity.* European Commission. Joint Research Centre. Publications Office of the European Union. https://doi.org/10.2788/94222

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. https://doi.org/10.1111/ele.12162

King, R. A., Tibble, A. L., & Symondson, W. O. C. (2008). Opening a can of worms: Unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, *17*(21), 4684–4698. https://doi.org/10.1111/j.1365-294X.2008.03931.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

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, *45*(3), 373–385. https://doi.org/10.1111/EEN.12831

Martoni, F., Smith, R. L., Piper, A. M., Nancarrow, N., Aftab, M., Trebicki, P., Kimber, R. B. E., Rodoni, B. C., & Blacket, M. J. (2023). Non-destructive insect metabarcoding as a surveillance tool for the Australian grains industry: A first trial for the iMapPESTS smart trap. *Metabarcoding and Metagenomics*, *7*. https://doi.org/10.3897/mbmg.7.95650

Mata, V. A., Ferreira, S., Campos, R. M., Silva, L. P., Veríssimo, J., Corley, M. F. V., & Beja, P. (2021). Efficient assessment of nocturnal flying insect communities by combining automatic light traps and DNA metabarcoding. *Environmental DNA*, *3*(2), 398–408. https://doi.org/10.1002/edn3.125

Mata, V. A., Rebelo, H., Amorim, F., McCracken, G. F., Jarman, S., & Beja, P. (2019). How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. *Molecular Ecology*, *28*(2), 165–175. https://doi.org/10.1111/mec.14779

Mupepele, A.-C., Keller, M., & Dormann, C. F. (2021). European agroforestry has no unequivocal effect on biodiversity: A time-cumulative meta-analysis. *BMC Ecology and Evolution*, *21*(1), 193. https://doi.org/10.1186/s12862-021-01911-9

Novo, M., Almodóvar, A., Fernández, R., Trigo, D., & Díaz Cosín, D. J. (2010). Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Molecular Phylogenetics and Evolution*, *56*(1), 507–512. https://doi.org/10.1016/j.ympev.2010.04.010

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., O'Hara, B., Simpson, G., Solymos, P., Stevens, H., & Wagner, H. (2015). Vegan: Community Ecology Package. *R Package Version 2.2-1*, *2*, 1–2.

One Health High-Level Expert Panel (OHHLEP), Adisasmito, W. B., Almuhairi, S., Behravesh, C. B., Bilivogui, P., Bukachi, S. A., Casas, N., Cediel Becerra, N., Charron, D. F., Chaudhary, A., Ciacci Zanella, J. R., Cunningham, A. A., Dar, O., Debnath, N., Dungu, B., Farag, E., Gao, G. F., Hayman, D. T. S., Khaitsa, M., … Zhou, L. (2022). One Health: A new definition for a sustainable and healthy future. *PLOS Pathogens*, *18*(6), e1010537. https://doi.org/10.1371/journal.ppat.1010537

Orgiazzi, A., Bardgett, R. D., Barrios, E., Behan-Pelletier, V., Briones, M. J. I., Chotte, J.-L., Deyn, G. B. D., Eggleton, P., Fierer, N., Fraser, T., Hedlund, K., Jeffery, S., Johnson, N. C., Jones, A., Kandeler, E., Kaneko, N., Lavelle, P., Lemanceau, P., Miko, L., … Wall, D. H. (2016). *Global Soil Biodiversity Atlas*. Publications Office of the European Union. https://doi.org/10.2788/799182

Pinto-Correia, T., Ribeiro, N., & Sá-Sousa, P. (2011). Introducing the montado, the cork and holm oak agroforestry system of Southern Portugal. *Agroforestry Systems*, *82*(2), 99–104. https://doi.org/10.1007/s10457-011-9388-1

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Roswell, M., Dushoff, J., & Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, *130*(3), 321–338. https://doi.org/10.1111/oik.07202

Schweiger, O., Musche, M., Bailey, D., Billeter, R., Diekötter, T., Hendrickx, F., Herzog, F., Liira, J., Maelfait, J. P., Speelmans, M., & Dziock, F. (2007). Functional richness of local hoverfly communities (Diptera, Syrphidae) in response to land use across temperate Europe. *Oikos*, *116*(3), 461–472. https://doi.org/10.1111/J.2007.0030-1299.15372.X

Stolte, J., Tesfai, M., Øygarden, L., Kværnø, S., Keizer, J., Verheijen, F., Panagos, P., Ballabio, C., Hessel, R., & European Commission. Joint Research Centre. Institute for Environment and Sustainability. (2016). *Soil threats in Europe.* Publications Office.

Swift, M. J., Heal, O. W., & Anderson, J. M. (1979). *Decomposition in Terrestrial Ecosystems*. Blackwell.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Torralba, M., Fagerholm, N., Burgess, P. J., Moreno, G., & Plieninger, T. (2016). Do European agroforestry systems enhance biodiversity and ecosystem services? A meta-analysis. *Agriculture, Ecosystems & Environment*, *230*, 150–161. https://doi.org/10.1016/j.agee.2016.06.002

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-09084-6

Tsiafouli, M. A., Thébault, E., Sgardelis, S. P., de Ruiter, P. C., van der Putten, W. H., Birkhofer, K., Hemerik, L., de Vries, F. T., Bardgett, R. D., Brady, M. V., Bjornlund, L., Jørgensen, H. B., Christensen, S., Hertefeldt, T. D., Hotes, S., Gera Hol, W. H., Frouz, J., Liiri, M., Mortimer, S. R., … Hedlund, K. (2015). Intensive agriculture reduces soil biodiversity across Europe. *Global Change Biology*, *21*(2), 973–985. https://doi.org/10.1111/gcb.12752

Valdecasas, A. G., & Camacho, A. I. (2003). Conservation to the rescue of taxonomy. *Biodiversity & Conservation*, *12*(6), 1113–1117. https://doi.org/10.1023/A:1023082606162

Wall, D. H. (Ed.). (2012). *Soil ecology and ecosystem services*. Oxford University Press.

Watts, C., Dopheide, A., Holdaway, R., Davis, C., Wood, J., Thornburrow, D., & Dickie, I. A. (2019). DNA metabarcoding as a tool for invertebrate community monitoring: A case study comparison with conventional techniques. *Austral Entomology*, *58*(3), 675–686. https://doi.org/10.1111/aen.12384

Waugh, J. (2007). DNA barcoding in animal species: Progress, potential and pitfalls. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *29*(2), 188–197. https://doi.org/10.1002/bies.20529

Yachi, S., & Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proceedings of the National Academy of Sciences*, *96*(4), 1463–1468. https://doi.org/10.1073/pnas.96.4.1463

Supplementary Materials

Table S1: Two-sample t-tests of "decontaminated"/"non-decontaminated", the table presents the sample sizes (n1 and n2), t-statistic, degrees of freedom (df), p-values (p), and the corresponding significance level notation (p.signif). Non-significant different "ns" (p-value ≥ 0.05), and statistically significant differences with p-values ≥ 0.05, are represented by "", p-values ≥ 0.01 "**", and p-values < 0.001 '***'.".*

Culture	n1 (decontaminated)	n2 (non-decontaminated)	statistic	df	p	p.signif
Hyperintensive Almond Grove	14	11	-0.792	19.804	0.438	ns
Intensive Almond Grove	9	12	1.244	10.007	0.242	ns
Improved Pasture	13	14	0.608	14.392	0.553	ns
Montado Extensive grazing	6	6	-0.579	9.812	0.576	ns
Montado Rotational grazing	12	15	1.315	23.818	0.201	ns
Hyperintensive Olive Grove	12	13	0.527	20.114	0.604	ns
Organic Olive Grove	12	13	0.124	18.946	0.902	ns
Traditional Olive Grove	11	16	2.688	21.084	0.014	*

Table S2: Samples with mock community with high diversity (n = 15) and low diversity (n = 15) based on the number of taxonomic groups identified based on morphological characteristic present per sample. Proportion of mock: 0.10 (n = 5), 0.25 (n = 5) and 0.5 (n = 5) per diversity group. The number of total OTUs present and in common with the mock community sample are available at the table followed by the proportion of recovered OTUs from the mock sample per sample. Mean ± SD per proportion of mock sample, diversity and total OTUs recovered.

Sample id	Diversity	Proportion of mock	Total OUTS	OTUs in common with MOCK	Proportion of recovered OTUs	Average of common OTUs	SD	Average of common OTUs for each diversity group	SD	Total average of common OTUs recovered	SD
22M	High	0.1	75	51	0.773	0.7728	± 0.019	0.689	± 0.079	**0.675**	**± 0.066**
72M			67	49	0.742						
76M			69	51	0.773						
78M			73	52	0.788						
9M			66	52	0.788						
108M		0.25	47	37	0.561	0.6034	±0.045				
109M			61	43	0.652						
123M			53	43	0.652						
124M			57	38	0.576						
87M			46	38	0.576						
142M		0.5	59	42	0.636	0.6908	± 0.038				
145M			65	45	0.682						
165M			65	45	0.682						
177M			65	48	0.727						
61M			59	48	0.727						
19M	Low	0.1	53	45	0.682	0.6334	± 0.062	0.661	± 0.049		
26M			52	47	0.712						
28M			45	41	0.621						
37M			46	38	0.576						
63M			45	38	0.576						
130M		0.25	53	45	0.682	0.6788	± 0.035				
70M			45	41	0.621						
77M			50	46	0.697						
84M			57	47	0.712						
112M			55	45	0.682						
149M		0.5	58	49	0.742	0.6698	± 0.045				
152M			58	44	0.667						
171M			49	43	0.652						
189M			53	44	0.667						
132M			44	41	0.621						
Mock sample	NA	1	66	66	1						

Table S3: Abundance of morphologically identified soil macrofauna specimens, (mean ± SD, n=192; individuals per pitfall). Samples from 8 agroecosystems in Idanha-a-Nova, Portugal. Individuals identified to the lowest taxonomic groups. Different letters indicate significant differences among environments (p-value < 0.05, GLM following a Poisson family with Tukey HSD Test).

	Hyperintensive Almond Grove	**Hyperintensive Olive Grove**	**Improved Pasture**	**Intensive Almond Grove**	**Montado Extensive grazing**	**Montado Rotational grazing**	**Organic Olive Grove**	**Traditional Olive Grove**
Aranae	4.880 ± 2.934 ab	2.852± 2.612 c	5.963 ± 7.867 bd	7.095 ± 4.949 d	7.231 ± 3.811 bd	6.519 ± 4.255 bd	3.440 ± 2.417 ac	4.037 ± 2.157 ac
Archaeognata	0.0 ± 0.0 ab	0.037 ± 0.192 b	0.0 ± 0.0 ab	0.048 ± 0.218 b	1.308 ± 1.702 a	0.037 ± 0.192 b	0.0 ± 0.0 ab	0.074 ± 0.267 b
Blattodea	0.0 ± 0.0 a	0.037 ± 0.192 a	0.0 ± 0.0 a	0.0 ±0.0 a	0.462 ± 0.877 a	0.0 ± 0.0 a	0.080 ± 0.400 a	0.481 ± 1.122 a
Coleoptera	12.440 ± 10.125 a	3.815 ± 4.161 b	5.185 ± 3.690 bc	10.571 ± 11.071 ad	7.615 ± 10.571 cde	15.815 ± 11.266 f	7.840 ± 6.094 e	11.556 ± 7.143 a
Dermaptera	0.880 ± 1.054 a	0.037 ± 0.192 b	0.074 ± 0.267 b	0.190 ± 0.512 ab	0.0 ± 0.0 ab	0.074 ± 0.267 b	0.040 ± 0.200 b	0.222 ± 0.506 b
Gastropoda	0.360 ± 0.569 a	0.222 ± 0.506 a	0.0 ± 0.0 ab	5.952 ± 8.022 b	0.462 ± 0.519 a	0.111 ± 0.424 a	0.0 ± 0.0 ab	0.037 ± 0.192 a
Hemiptera	5.080 ± 9.055 a	1.556 ± 2.006 b	5.778 ± 9.070 a	1.524 ± 1.569 b	0.615 ± 1.121 b	3.074 ± 3.551 c	1.920 ± 2.060 bc	0.926 ± 1.328 b
Hymenoptera	42.400 ± 73.756 a	11.556 ± 31.294 bc	16.000 ± 16.758 d	14.381 ± 11.240 bd	3.154 ± 3.313 e	11.222 ± 13.042 c	20.160 ± 17.951 f	15.593 ± 10.173 d
Isopoda	0.440 ± 0.821 a	0.481 ± 1.553 a	0.0 ± 0.0 ab	0.429 ± 0.507 a	0.0 ± 0.0 ab	2.074 ± 6.742 b	0.640 ± 0.810 a	0.593 ± 0.971 a
Myriapoda	0.560 ± 0.712 ab	0.296 ± 0.542 a	0.0 ± 0.0 abc	2.952 ± 3.138 c	0.077 ± 0.277 ab	1.074 ± 1.940 b	0.360 ± 0.569 ab	0.519 ± 0.753 ab
Opiliones	0.080 ± 0.277 a	0.296 ± 0.609 ab	0.111 ± 0.424 a	0.762 ± 1.786 bcd	1.385 ± 2.663 cd	1.037 ± 2.084 cd	0.600 ± 1.848 ac	1.630 ± 2.884 d
Orthoptera	0.240 ± 0.523 a	0.0 ± 0.0 a	0.444 ± 1.013 a	0.286 ± 0.561 a	0.0 ± 0.0 a	0.0 ± 0.0 a	0.0 ± 0.0 a	0.148 ± 0.362 a
Pseudoscorpiones	0.0 ± 0.0 a	0.111 ± 0.424 a	0.0 ± 0.0 a	0.0 ± 0.0 a	0.385 ± 0.650 a	0.037 ± 0.192 a	0.080 ± 0.277 a	0.0 ± 0.0 a
Total abundance	**67.360 ± 72.995 a**	**21.296 ± 34.822 b**	**33.556 ± 22.128 c**	**44.190 ± 21.669 d**	**22.692 ± 14.471 b**	**41.074 ± 18.953 d**	**35.160 ± 23.946 c**	**35.815 ± 16.373 c**

Table S4: OTU based analysis, Class and Order richness (mean ± SD, n=191; OTU richness per pitfall), of soil macrofauna. Samples from 8 agroecosystems in Idanha-a-Nova, Portugal. Different letters indicate significant differences among environments (p-value < 0.05, GLM following a Poisson family with Tukey HSD Test).

Class	Order	Hyperintensive Almond Grove	Hyperintensive Olive Grove	Improved Pasture	Intensive Almond Grove	Montado Extensive grazing	Montado Rotational grazing	Organic Olive Grove	Traditional Olive Grove
Arachnida		**6.720 ± 3.062 ab**	**5.080 ± 2.548 a**	**5.444 ± 2.207 a**	**6.476 ± 2.768 ab**	**10.333 ± 2.605 c**	**8.889 ± 3.806 bc**	**7.040 ± 3.195 ab**	**8.407 ± 3.377 bc**
	Araneae	6.720 ± 3.062 ab	5.040 ± 2.541 a	5.444 ± 2.207 a	6.476 ± 2.768 ab	10.333 ± 2.605 c	8.852 ± 3.780 bc	7.00 ± 3.189 ab	8.370 3.410 bc
	Pseudoscorpiones	0.000 ± 0.000 a	0.040 ± 0.200 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.037 ± 0.192	0.040 ± 0.200 a	0.037 ± 0.192 a
Insecta		**15.720 ± 8.749 ab**	**14.000 ± 9.074 b**	**16.407 ± 10.966 ab**	**15.905 ± 8.905 ab**	**13.417 ± 9.501 b**	**14.000 ± 8.806 b**	**18.120 ± 6.852 ac**	**21.074 ± 12.016 c**
	Archaeognatha	0.040 ± 0.200 a	0.000 ± 0.000 ab	0.037 ± 0.192 a	0.095 ± 0.436 a	1.167 ± 1.467 b	0.037 ± 0.192 a	0.000 ± 0.000 ab	0.037 ± 0.192 a
	Blattodea	0.040 ± 0.200 a	0.040 ± 0.200 a	0.000 ± 0.000 ab	0.000 ± 0.000 ab	1.000 ± 2.216 bc	0.037 ± 0.192 a	0.280 ± 1.400 ac	1.148 ± 2.445 b
	Coleoptera	9.840 ± 6.517 ab	8.880 ± 5.183 ac	9.259 ± 6.780 acd	11.810 ± 7.891 bd	7.000 ± 2.374 a	10.259 ± 7.789 bc	11.200 ± 4.787 bc	12.333 ± 5.657 b
	Dermaptera	0.920 ± 0.909 a	0.160 ± 0.374 b	0.296 ± 0.465 ab	0.333 ± 0.577 ab	0.083 ± 0.289 ab	0.259 ± 0.594 ab	0.040 ± 0.200 b	0.296 ± 0.609 ab
	Embioptera	0.000 ± 0.000 ab	1.200 ± 6.000 b	0.000 ± 0.000 ab	0.095 ± 0.436 a	1.750 ± 6.062 b	0.000 ± 0.000 ab	0.040 ± 0.200 a	1.667 ± 7.721 b
	Hemiptera	1.800 ± 1.936 ab	1.440 ± 1.446 b	2.815 ± 2.497 ac	1.381 ± 1.284 b	0.583 ± 0.793 b	1.852 ± 1.610 ab	4.360 ± 3.872 c	1.630 ± 2.963 ab
	Hymenoptera	0.520 ± 0.510 a	0.800 ± 0.957 a	0.741 ± 1.196 a	0.190 ± 0.512 a	0.167 ± 0.389 a	0.519 ± 0.893 a	0.800 ± 0.817 a	0.370 ± 0.688 a
	Opiliones	0.000 ± 0.000 a	0.040 ± 0.200 a	0.074 ± 0.267 a	0.095 ± 0.436 a	0.167 ± 0.577 a	0.148 ± 0.456 a	0.080 ± 0.277 a	0.111 ± 0.320 a
	Orthoptera	2.560 ± 4.891 a	0.760 ± 0.879 b	3.185 ± 5.421 a	1.905 ± 3.562 ac	0.667 ± 0.985 bc	0.667 ± 0.679 b	1.320 ± 1.145 bc	3.259 ± 4.981 a
	Zygentoma	0.000 ± 0.000 a	0.680 ± 2.839 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.833 ± 2.588 a	0.222 ± 0.641 a	0.000 ± 0.000 a	0.222 ± 1.155 a
Chilopoda		**1.280 ± 1.137 a**	**0.680 ± 0.802 ab**	**0.148 ± 0.362 b**	**1.048 ± 1.687 a**	**0.750 ± 0.965 ab**	**0.704 ± 1.103 ab**	**0.120 ± 0.332 b**	**0.259 ± 0.526 b**
	Lithobiomorpha	1.280 ± 1.137 a	0.680 ± 0.802 ab	0.148 ± 0.362 b	1.048 ± 1.687 a	0.500 ± 0.674 ab	0.593 ± 0.888 ab	0.120 ± 0.332 b	0.259 ± 0.526 b
	Scolopendromorpha	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.250 ± 0.866 a	0.111 ± 0.577 a	0.000 ± 0.000 a	0.000 ± 0.000 a
Clitellata		**3.920 ± 1.222 a**	**4.320 ± 1.819 a**	**2.778 ± 1.450 a**	**3.238 ± 1.578 a**	**3.000 ± 1.809 a**	**4.185 ± 2.418 a**	**3.280 ± 1.275 a**	**3.370 ± 2.022 a**
	Crassiclitellata	3.840 ± 1.248 a	4.200 ± 1.732 a	2.778 ± 1.450 a	3.190 ± 1.601 a	3.000 ± 1.809 a	4.185 ± 2.418 a	3.280 ± 1.275 a	3.370 ± 2.022 a
	Enchytraeida	0.0800 ± 0.277 a	0.120 ± 0.332 a	0.000 ± 0.000 a	0.0477 ± 0.218 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a
Diplopoda		**0.600 ± 0.646 ab**	**0.720 ± 0.792 bc**	**0.407 ± 0.694 b**	**1.333 ± 0.658 acd**	**0.583 ± 0.669 bcd**	**1.815 ± 3.163 de**	**1.600 ± 4.123 ce**	**2.667 ± 5.226 e**
	Chordeumatida	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.037 ± 0.192 a
	Julida	0.600 ± 0.646 a	0.640 ± 0.757 a	0.407 ± 0.694 a	0.714 ± 0.717 ab	0.583 ± 0.669 ac	1.815 ± 3.163 cd	1.600 ± 4.123 bcd	2.593 ± 5.242 d
	Polydesmida	0.000 ± 0.000 a	0.080 ± 0.277 a	0.000 ± 0.000 a	0.619 ± 0.498 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.037 ± 0.192 a
Gastropoda		**0.320 ± 0.476 a**	**0.280 ± 0.459 a**	**0.000 ± 0.000 ab**	**1.333 ± 0.966 b**	**0.500 ± 0.522 ab**	**0.222 ± 0.641 a**	**0.080 ± 0.277 a**	**0.111 ± 0.320 a**
	Pulmonata	0.000 ± 0.000 a	0.040 ± 0.200 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a	0.000 ± 0.000 a
	Stylommatophora	0.320 ± 0.476 a	0.240 ± 0.436 a	0.000 ± 0.000 ab	1.333 ± 0.966 b	0.500 ± 0.522 ab	0.222 ± 0.641 a	0.080 ± 0.277 a	0.111 ± 0.320 a
Malacostraca		**0.360 ± 0.569 a**	**0.160 ± 0.374 a**	**0.259 ± 0.447 a**	**0.238 ± 0.436 a**	**0.083 ± 0.289 a**	**0.111 ± 0.320 a**	**0.160 ± 0.374 a**	**0.185 ± 0.396 a**
	Isopoda	0.360 ± 0.569 a	0.160 ± 0.374 a	0.259 ± 0.447 a	0.238 ± 0.436 a	0.083 ± 0.289 a	0.111 ± 0.320 a	0.160 ± 0.374 a	0.185 ± 0.396 a
Total Richness		**28.920 ± 11.832 ab**	**25.240 ± 10.760 a**	**25.444 ± 12.280 a**	**29.571 ± 11.016 ab**	**28.667 ± 10.637 ab**	**29.926 ± 10.433 b**	**30.400 ± 7.101 b**	**36.074 ± 13.255 c**

Table S5: Permutational multivariate analysis of variance (PerMANOVA) with a pairwise Pillai post-hoc. The table presents the sites being compared, "agroecosystem 1" and "agroecosystem 2", the p.values of the test "p-values (p)" and the corrected p-value "pvalue.adj" by the Benferroni p-value corrections Non-significant different "ns" (p-value ≥ 0.05), and statistically significant differences with p-values ≤ 0.05, are represented by "", p-values ≤ 0.01 "**", and p-values < 0.001 '***'.*

Agroecosystem1	Agroecosystem2	pvalue	pvalue.adj
Hyperintensive Olive Grove	Hyperintensive Almond Grove	0.001***	0.028*
Improved Pasture	Hyperintensive Almond Grove	0.001***	0.028*
Intensive Almond Grove	Hyperintensive Almond Grove	0.001***	0.028*
Montado Extensive grazing	Hyperintensive Almond Grove	0.001***	0.028*
Montado Rotational grazing	Hyperintensive Almond Grove	0.001***	0.028*
Organic Olive Grove	Hyperintensive Almond Grove	0.001***	0.028*
Traditional Olive Grove	Hyperintensive Almond Grove	0.001***	0.028*
Improved Pasture	Hyperintensive Olive Grove	0.001***	0.028*
Intensive Almond Grove	Hyperintensive Olive Grove	0.001***	0.028*
Montado Extensive grazing	Hyperintensive Olive Grove	0.001***	0.028*
Montado Rotational grazing	Hyperintensive Olive Grove	0.001***	0.028*
Organic Olive Grove	Hyperintensive Olive Grove	0.001***	0.028*
Traditional Olive Grove	Hyperintensive Olive Grove	0.001***	0.028*
Intensive Almond Grove	Improved Pasture	0.001***	0.028*
Montado Extensive grazing	Improved Pasture	0.001***	0.028*
Montado Rotational grazing	Improved Pasture	0.001***	0.028*
Organic Olive Grove	Improved Pasture	0.001***	0.028*
Traditional Olive Grove	Improved Pasture	0.001***	0.028*
Montado Extensive grazing	Intensive Almond Grove	0.001***	0.028*
Montado Rotational grazing	Intensive Almond Grove	0.001***	0.028*
Organic Olive Grove	Intensive Almond Grove	0.001***	0.028*
Traditional Olive Grove	Intensive Almond Grove	0.001***	0.028*
Montado Rotational grazing	Montado Extensive grazing	0.001***	0.028*
Organic Olive Grove	Montado Extensive grazing	0.001***	0.028*
Traditional Olive Grove	Montado Extensive grazing	0.002**	0.056 ns
Organic Olive Grove	Montado Rotational grazing	0.001***	0.028*
Traditional Olive Grove	Montado Rotational grazing	0.001***	0.028*
Traditional Olive Grove	Organic Olive Grove	0.002**	0.056 ns