UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

DNA metabarcoding in terrestrial biodiversity assessment and monitoring: the case of nocturnal insects in NE Portugal

Rebeca Mateus Ramos de Campos

Mestrado em Ecologia e Gestão Ambiental

Dissertação orientada por:

Doutor Pedro Rui Correia de Oliveira Beja
CIBIO, Faculdade de Ciências da Universidade do Porto (FCUP)

Professora Maria Filomena de Magalhães
Faculdade de Ciências da Universidade de Lisboa (FCUL)

2019

A human being is a part of the whole called by us Universe, a part limited in time and space. He experiences himself, his thoughts and feelings as something separated from the rest, a kind of optical delusion of his consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons nearest to us. Our task must be to free ourselves from this prison by widening our circle of compassion to embrace all living creatures and the whole of nature in its beauty.

Albert Einstein

Acknowledgements

I am sorry, but these will go in mixed idioms, Portuguese and English.

Em primeiro lugar agradeço ao meu orientador, Pedro Beja, por teres concedido a oportunidade de estudar este tema e trabalhar com uma equipa tão fantástica. Obrigada por pelas tuas imensas criativas boas ideias e pelo pragmatismo e praticidade na resolução de problemas que fazem de trabalhar contigo uma experiência enriquecedora.

À Filomena Magalhães minha professora e orientadora, que foi quem inicialmente me encaminhou para esta decisão de subir ao norte do país para estudar estes bichinhos, obrigada!

À Vanessa Mata, que foi de longe quem mais trabalho teve comigo. Obrigada pelo teu apoio, por seres imprestável, por estares lá para qualquer pergunta, por mais idiota que fosse, sempre com calma, serenidade e paciência. O que me ofereceste ao longo deste trabalho é impagável!

À Sónia Ferreira por todos os bons momentos no campo e a respirar pó de insecto. Obrigada por me ensinares (ou pelo menos tentares) a ter paciência e a importância de ouvir, respirar e deixar pensar!

To Martin Corley, who brought up an incomparable knowledge about moths and whose passion on moths is contagious, you taught me so much! Thank you for coming to Portugal all these years to study moths! Above all, thank you for being so kind and helpful, whether in person to sort and identify hundreds of moths of all imaginable sizes in Tua with 40ºC during our only after-lunch-rest-time, whether by promptly answering all my emails with so many answers to my questions.

Ao João Nunes pelos momentos no campo e dicas fora dele. Obrigada por seres um verdadeiro adolescente, por saberes muito mais que eu sobre borboletas e por sermos parvos no campo!

À Joana Verissímo pela ajuda e boa disposição! Obrigada por tornares até o trabalho de laboratório divertido e por me explicares tantas coisas que não entendia, foi um contributo enriquecedor!

A toda a família, começando pelo meu pai que me incentivou sempre a estudar e a fazer o melhor possível em cada situação, por se preocupar mais com o meu bem estar que com o seu próprio! Obrigada pai, porque sabes que sem ti dificilmente seria possível ter embarcado em mais esta jornada!

Ao Rodri, meu mano, que mesmo distante sempre presente!

À Wanda Caio, minha madrinha, pelos conselhos, bons momentos, por querer sempre saber e por se preocupar só um bocadinho mais comigo agora que comigo no futuro! Obrigada por me ajudares a dar um passo de cada vez, e lembrares que é preciso disfrutar do caminho, mais chegar à meta.

À tia, tio e primos, especialmente à minha prima, Filipa Palmeirim que sempre tem sido uma fonte de inspiração biológica. E à Andreia Palmeirim e à baby Ema, cujo primeiro ano de vida foi também o ano desta tese!

À equipa do EnvMetaGen lá em cima, Bastian Egeter, Aina Garcia Raventós, Mafalda Galhardo.

A toda a malta do Porto! Do CiBIO, Ana Buchadas, Aspassia Anagnostopoulou, Fulvio Licata, Yuri Simone, Cristian Pizzigalli, Paula Tralma! À minha família barcelense, especialmente à Vânia Silva, com quem foi tão bom viver, pelos melhores pequenos almoços e sobrevivência na autodestruição do nosso mega apartamento – remebering Forrester, sempre! Mas também à Natália, ao Mauro e ao Afonso, que me ensinou que paciência e persistência dão sempre frutos! À Mariana e ao Bruno, porque cumbiar ajudou à pressão aguentar! Ao Rafa que acompanhou a jornada do portuense duma ponta à outra! À malta do Rés-da-rua e do Espaço Compasso, que foram dos sítios mais lindos, com pessoas maravilhosas que encontrei nesta cidade. E a toda a malta da Azurara surf school, nomeadamente ao Lagoa que me

ensiou a surfar e foi quem deu o empurrão para apanhar a minha primeira onda – vital neste ano de ondas várias!

À Raquel Andrade, minha irmã, pela omnipresência ao longo dos últimos 16 anos, são 2/3 das nossas vidas!

Aos Lilos, minha família escolhida, Madalena Gaspar, Íris Mota, Inês Santos, Carlos Silva, Sofia Moinhos obrigada! Foram muitos anos a biologar, guardar hortas e muito mais que isso juntos, nem há palavras para escrever ou metade da tese seria um rolo de agradecimentos.

À Diana Niepce que pensa que eu sou uma adolescente e não percebe porque é que matamos borboletas e ainda assim contribuiu com observações muito pertinentes e momentos de descontração.

Ao Aboo Gani pelas conversas e sugestões, em especial por sugerires que nunca pare de perguntar e que falasse com o João Santos e o António Barata a quem agradeço os *statistical insights* e a muita paciência para as minhas perguntas.

Àqueles que me acompanharam nos bons e maus momentos de fazer uma tese em ecologia e gestão ambiental e com quem me tinha preparado no primeiro ano para o que se ia seguir, até ao fim, Inês Bom, Miguel Azevedo e Pedro Neves!

À HortaFCUL, e todos os seus actuais e antigos guardiões, os que foram, os que ficaram e os que vieram, que tanto me ensinaram e que tão útil foi esse conhecimento para o sucesso do meu percurso académico. Nomeadamente, ao Pedro Moreira que me incentivou a candidatar-me ao mestrado e ao Daniel Lopes pela amizade omnipresente e pelos contributos gráficos.

Ao André Garcia, pelos contributos gráficos e não só, ao apoio incondicional à positividade – a encarar os problemas como oportunidades! A toda a restante malta da Jazzy Dance Studios que nesta fase final proporcionou os momentos de descontração e concentração no movimento e na dança que muito trouxeram ao bem-estar geral e ajudaram a acabar esta tarefa que por momentos pode ter parecido uma missão impossível!

Agora vou voar com as borboletas!

Abstract

Biodiversity is declining worldwide, and one of its main causes is the expansion and intensification of agriculture. This process has been studied mainly for vertebrates and a few invertebrate groups, but information is missing for most insects despite their deep relationship with agriculture. This is partly due to difficulties in species identification, which requires taxonomic expertise and is costly and time consuming, especially for highly diverse insect communities. In traditional species identification, external or internal morphological characters are used to distinguish between different biological entities, one specimen at a time. Moreover, for many insect species only adult males (or females) can be reliably identified, and identification of larval stages or specimens in poor condition is often not possible. Recently, the development of next generation sequencing (NGS) of DNA is revolutionizing the ability to study highly diverse biological communities, by providing a relatively simple and inexpensive method for species identification. However, this approach remains little used in biodiversity assessment and monitoring, particularly in the case of nocturnal insects. Here we develop and test a workflow based on DNA metabarcoding for understanding how agricultural land uses affect the diversity and composition of nocturnal insect communities. Samples were collected in a complex mosaic landscape in NE Portugal in July (70 sites) and September (78) 2017 using UV light traps, covering four habitats with decreasing level of human management and increasing vegetation cover: vineyards, olive groves, cork oak woodland and riparian zones. Each site yielded a bulk sample of insects, which was processed using standard molecular and bioinformatics pipelines to produce a list of taxa identified at the lowest possible taxonomic level. Molecular procedures involved the testing of four metabarcoding primers, one of which produced the best results and was used in subsequent metabarcoding amplification. Validation of the method involving 12 samples revealed a close matching, albeit not perfect, between the species of moth (Lepidoptera) visually identified by a taxonomist and those recovered through metabarcoding. Metabarcoding of the bulk samples retrieved 1130 taxa, most of which were Lepidoptera (429 OTUs), Diptera (244) and Coleoptera (166). Despite this large number of taxa, accumulation curves revealed that sampling effort was still insufficient to capture the entire diversity within each habitat. There was significant variation in richness among habitats and between seasons, with vineyards showing consistently the lowest number of taxa. In contrast, the richness of olive groves was comparable, or even higher than that of the more natural cork oak and riparian habitats. The assemblage composition varied markedly across seasons, while variation among habitats was less marked, though there were differences in both characteristic and exclusive species. Variation in moth assemblage composition among habitats and seasons was related to functional traits, with for instance larger species recorded in cork oak and riparian habitats and in September, species with specialized diets in cork oak and September, and detritivore species in September. Overall, this workflow provides an efficient tool to assess and monitor nocturnal insects, with the potential for greatly advancing our understanding of agricultural impacts on biodiversity.

Keywords

Community ecology, invertebrate diversity, NGS sequencing, agroforest landscapes, moths (Lepidoptera)

Resumo

A disrupção de sistemas naturais está associada à perda de biodiversidade e actualmente põe em risco o equílibro de toda a biosfera. Apesar do desenvolvimento e implementação de medidas de protecção de biodiversidade, muitas espécies continuam em perigo. Devido maioritariamente à acção humana as ameaças à diversidade natural são inúmeras, potenciando a perda de espécies ainda desconhecidas, uma vez que grande parte da diversidade biológica continua por descrever ou descobrir. Desta forma, evidencia-se a necessidade de conhecer e monitorizar esta diversidade às escalas local e global. Só é possível desenvolver ações de conservação adequadas e medir o sucesso das mesmas ao encontrar formas de conhecer e quantificar a diversidade e de a monitorizar ao longo do tempo e do espaço. Neste contexto, são os insectos e outros invertebrados que mais carecem de estudo, monitorização e acções de conservação. Com milhões de espécies desconhecidas que interagem em sistemas complexos, o estudo das comunidades de insectos torna-se desafiante. Ainda assim, os poucos estudos existentes indicam que, à semelhança do que ocorre com os vertebrados, muitas espécies de insectos estão em declínio. No entanto, sabe-se que os insectos desempenham um papel essencial nos ecossistemas, sendo responsáveis por diversos serviços e desserviços, nomeadamente em ecossistemas agroflorestais. Por um lado, participam em processos essenciais como a polinização e reciclagem de nutrientes, e por outro lado, constituem potenciais pragas capazes de causar graves danos à produtividade e economia agrícola.

Neste contexto, estudar insectos é particularmente relevante e um dos desafios tem sido a identificação das espécies. Esta constitui uma das componentes inerentes a estudos de biologia e ecologia, bem como ao desenvolvimento de planos de monitorização eficientes. Tradicionalmente, a identificação de espécies de insetos tem sido feita com base em características morfológicas externas ou internas, um espécime de cada vez. Contudo, este processo depende do trabalho de taxonomistas especializados, tornando-se dispendioso em termos de custo e tempo. Mais ainda, a maioria das espécies sofre metamorfose completa e muitos organismos são virtualmente impossíveis de identificar nas fases larvares, o que dificulta o processo. Hoje em dia, existem novos métodos moleculares que cada vez mais permitem ultrapassar alguns destes problemas. Entre estes, salientam-se os métodos que surgiram com o desenvolvimento de sequenciadores de nova geração (NGS), nomeadamente um conjunto de técnicas designado *DNA metabarcoding*.

O *DNA metabarcoding* faz uso de *primers* universais na amplificação de *barcodes* em massa, provenientes de espécimes contidos em várias amostras complexas. Estas amostras são posteriormente marcadas com curtas sequências distintas, o que permite juntar (*pooling*) e sequenciar múltiplas amostras e posteriormente recuperar a sua composição individual. O uso desta metodologia tem cada vez mais vindo a ser reconhecido como particularmente útil em amostras cujo estado não permite identificação, como água ou guano, ou em amostras complexas compostas por misturas de espécies. Contudo tem ainda muitas limitações associadas, entre as quais a falta de protocolos standardizados, a necessidade de desenhar primers universais, a morosidade dos processos dos quais dependem a construção de bibliotecas de *barcodes* e algumas limitações bioinformáticas. Ainda assim, o método tem potencial de produzir resultados com elevada resolução taxonómica, permitindo identificar múltiplas espécies de uma só vez, sem a dependência de taxonomistas experientes.

Neste estudo utilizamos o *DNA metabarcoding* para estimar a diversidade de artrópodes terrestres num parque natural recentemente designado, o Parque Natural e Regional do Vale do Tua (PNRVT). O parque natural, está inserido numa paisagem fortemente fragmentada que resulta num complexo mosaico agroflorestal, típico do Mediterrâneo. No total, foram recolhidas 144 amostras de insectos no PNRVT em dois períodos distintos, Julho (n=68) e Setembro (n=76). A amostragem ocorreu numa área geográfica vasta, que se estende por todo o parque natural, em 2 habitats agrícolas e 2 habitats semi-

naturais/naturais representativos da paisagem agroflorestal fragmentada que lhe é característica. Os tipos de habitat foram definidos por ordem crescente de cobertura vegetal e decrescente de intervenção humana: vinhas, olivais, sobreirais e galerias ripícolas. Nas amostragens utilizaram-se armadilhas de luz ultravioleta (UV), e o conteúdo de cada armadilha foi processado separadamente em laboratório até ao *pooling* das amostras para a sequenciação. Especial atenção foi dada aos lepidópteros que são um grupo de insectos altamente diverso e pouco estudado. Para colmatar a falta de informação centralizada sobre muitos aspectos da sua biologia foram compilados um conjunto de *traits* de espécies desta ordem recorrendo a várias fontes online. Nomeadamente, compilaram-se *traits* morfológicos, fisiológicos, comportamentais e de distribuição geográfica. Estes dados foram utilizados para completar as análises do ponto de vista ecológico. Desenvolveu-se e utilizou-se um protocolo dividido em 4 fases para analisar os dados obtidos através da sequenciação Miseq i) teste piloto, e da sequenciação Hiseq ii) validação do método, iii) design para a monitorização ecológica e iv) análise de *traits*. Inicialmente utilizaram-se apenas 3 amostras para um teste piloto que incluiu testar a eficácia de vários primers. Todas as amostras foram sequenciadas posteriormente com o primer que produziu melhores resultados, o BF2/BR2. Este par de primers permitiu produzir mais identificações ao nível da espécie e menos sequências únicas, com menos erros e reduzindo o tempo de análise. A validação do método *i.e.* comparação dos resultados obtidos com metabarcoding e identificação visual permitiu verificar que a maioria das espécies de lepidópteros foi detectada. A sequenciação das restantes amostras recolhidas foi levada a cabo na mesma corrida (Hiseq). Estas foram recolhidas segundo o desenho de um plano de monitorização que incluiu a montagem e recolha de um número considerável de amostras nos 4 habitats representativos da paisagem, fornecendo dados para várias análises de riqueza e composição de espécies. As restantes análises foram efectuada para todas as unidades taxonómicas operacionais (OTUs) com excepção da última, a análise de *traits*, que se focou apenas nos lepidópteros, dada a sua importância ecológica e também pela disponibilidade de informação.

Apesar de terem sido recolhidas 144 amostras, o esforço de amostragem provou-se insuficiente para atingir um número de espécies representativo da riqueza actual. Ainda assim, e apesar de não ser possível obter informação relativa à abundância das espécies, o método permitiu obter resultados com bastante resolução taxonómica. Deste modo foi possível captar as diferenças na riqueza e composição de espécies entre épocas e habitats através de análises exploratórias e testes de significância estatística. Verificou-se que as vinhas tendem a apresentar os valores de riqueza média mais baixos, em conjunto com os sobreirais em Julho. Por outro lado, nos olivais ocorreram alguns dos valores mais elevados riqueza de OTUs, bem como nos sobreirais em Setembro. As maiores diferenças entre a riqueza global de insectos e de lepidópteros registaram-se nas galerias ripícolas, indicando que este habitat é muito rico em outros *taxa* de insectos. Pelo contrário, em zonas de sobreiral destacou-se uma elevada riqueza em lepidópteros, especialmente em Setembro. Ao nível da composição das comunidades, as diferenças captadas parecem estar maioritariamente relacionadas com a época de recolha das amostras como seria expectável, dado que a maioria das espécies de insectos só está em fase adulta e activa numa determinhada época do ano. Contudo as análises indicam que o habitat também tem influência na composição das comunidades. As galerias ripícolas foram o habitat mais distinto em Julho enquanto que em Setembro os sobreirais tinham uma composição mais única de OTUs. A maioria das análises aponta para maiores semelhanças na composição das comunidades presentes em olivais e vinhas, quer ao nível geral de composição das amostras quer por partilharem algumas espécies dominantes. Foi possível relacionar certos *traits* das espécies de lepidópteros com a sua ocorrência na paisagem e nas duas épocas de amostragem. Nomeadamente, registou-se a presença de espécies de maiores dimensões em sobreirais e galerias ripícolas, a ocorrência de espécies de menores dimensões, com menor número de gerações e tendencialmente polífagas no mês de Julho e de espécies detrítivoras em Setembro. Ainda é difícil explicar muitos destes padrões e perceber o seu significado ecológico uma vez que a biologia destas

espécies é pouco conhecida. Simultaneamente, são muitas espécies para avaliar e conhecer, pelo que o seu estudo além de necessário e desafiante, abre as portas a uma série de questões científicas que permanecem por responder.

Conclui-se que o *DNA metabarcoding* tem potencial para integrar estudos ecológicos e como ferramenta para monitorização rápida e eficaz, reproduzível no tempo e no espaco, para identificação de espécies em amostras complexas de insectos. Aqui demonstramos que esta ferramenta pode ser particularmente útil em contextos de comparação de habitats em vastas zonas fragmentadas num curto período temporal, mas que pode também ser extensível com resultados reproduzíveis e comparáveis ao longo do tempo. A elevada resolução taxonómica do método permitiu detectar diferenças entre riqueza e composição de espécies em cada habitat. Nomeadamente, o método permitiu identificar espécies dominantes, e até relaccionar a ocorrência de certas espécies com os seus *traits* morfológicos, fisiológicos e comportamentais. Além disso, a estrutura de análise em 4 fases pode ser utilizada noutros contextos com as devidas adaptações. Na verdade, as várias fases do processo incluem testes e validações e servem para auto-calibrar e adaptar o método, tornando a sua aplicação possível em diversos contextos e paisagens. Ainda assim, várias limitações inerentes método mantêm-se, quer anível laboratorial, quer bioinformático, e por isso é necessário continuar a investigar o seu uso e promover a sua implementação como uma prática corrente em estudos ecológicos.

Palavras-chave

Ecologia de comunidades, diversidade de invertebrados, sequenciadores de nova geração, paisagens agroflorestais, borboletas nocturnas (Lepidoptera)

Table of contents

List of figures

List of tables

List of Abbreviations

BOLD Barcode of Life Data Systems

bp base-pairs

CA Correspondence Analysis

CBD Convention on Biological Diversity

CBOL Consortium for the Barcode of Life

COI Cytochrome C Oxidase

DNA Deoxyribonucleic Acid

eDNA Environmental DNA

GLM Generalized Linear Model

HNVF High Nature Value Farmland

ID Identification

IUCN International Union for Conservation of Nature

MEA Millennium Ecosystem Assessment

NCBI National Center of Biotechnology Information

NGS Next Generation Sequencing

OTU Operational Taxonomic Unit

PCR Polymerase Chain Reaction

PNRVT Parque Natural e Regiona do Vale do Tua

PCoA Principal Coordinate Analysis

PermANOVA Permutational Analysis of Variance

SIMPER Similarity Percentage Analysis

UNESCO United Nations Educational, Scientific and Cultural Organization

UV Ultra violet

1. Introduction

1.1. The decline of biodiversity and the unknown reality of insects

Natural systems are under severe disruption, changing at an unprecedent rate and threatening biological communities by shifting their interactions and consequently the balance of the whole biosphere[1]. Accentuated changes in the composition of biological communities, as species losses or introductions, may affect ecosystem processes which together determine the regulation of many Earth systems[1,2]. Simultaneously, ecosystem degradation and alteration of ecosystem processes affects biodiversity[3–5].

Biodiversity loss is a reality and the end of the threats that make it true is far from sight[6]. As biodiversity is deeply connected with ecosystem functioning and services, it ultimately relates to human well-being[7]. Still, human action clearly is at the root of the problem, having its effect through habitat loss and degradation, overexploitation, pollution, climate change and invasive alien species[8,9]. These drivers act together, and most times a combination of many simultaneously endangers biodiversity and ecosystems over a given area. Despite the increase of conservation initiatives to address these threats, the efforts have not been enough to halt biodiversity loss nor the pressures that affect it[10]. In fact, the scenarios for future extinction rates exceed those in recent fossil record, mostly due to human action and how it interferes with ecosystems and their species[6,10]. Could it be that we are on the way to the 6th mass extinction? Some call it Anthropocene[11–13] and there is a chance that species are going extinct before we get to know them[14–16], as still millions of them remain unknown or undescribed out there[10,17,18].

Most of the biological diversity of metazoans is made of insects and other invertebrate species, thus its protection and conservation is implicit in most biodiversity agreements as the Convention on Biological Diversity[19]. At its sixth meeting in 2002, it was set the goal to achieve a significant reduction of the biodiversity loss rate by 2010[20]. Globally it was clear that by the deadline that goal was far from achieved[21,22]. Other conventions followed CBD in establishing goals to protect biodiversity in one way or another. However, invertebrate diversity has received still less attention from researchers, hence fewer monitoring and conservation programs have been designed to target its taxonomic groups. While most conservation measures still privilege a small number of flagship species, it is well-known that invertebrates play a fundamental role in ecosystems as they participate in many ecosystem processes and provide important services[19,23,24]. However, most described species lack assessment and their populational trends are unknown globally[25]. Under this context, the importance of assessing the status of their ecosystems, the composition of their communities and how they interact becomes increasingly relevant.

Although only a few local or national studies are available[19], these examples indicate that terrestrial invertebrates are declining and under threat. Only 1% of the all described insect species were evaluated by IUCN, 40% were stated as threatened[19,25] and 67% of monitored populations show 45% mean abundance decline[12]. Habitat conversion and fragmentation due to changes in land use are some of the main causes of change in species assemblages (Figure 1-1)[8,26]. Among them, it turns out that agriculture is one of the main drivers of terrestrial biodiversity loss and ecosystem degradation[8,27–29]. Farmers are the main land managers at a global scale[28] as croplands and pastures extend for about 40% of the Earth's land cover, making up one of the largest terrestrial biomes[29,30].

However, not all kinds of agriculture have the same effect on habitats and biodiversity, namely on insect diversity[31–35]. Different kinds of systems are associated with different levels of habitat degradation. Intensive monoculture systems that extend over large areas are generally associated with biodiversity declines[36]. For example, the conversion of traditional coffee plantations consisting of agroforest systems with tree shades to unshaded coffee monocultures was associated with a reduction of arthropod

diversity[37]. The intensification of agriculture has been grounded in a reduction of habitat the complexity, which can thus be seen as simplified versions of natural ecosystems, where the number of species is often lower than in the surrounding natural habitats[33,38,39]. This simplification results in artificial ecosystems that are only maintained by a continued human intervention: mechanized seed-bed preparation and sowing replace natural seed dispersal, synthetic pesticides replace the action of natural pest control by predators or parasites, and genetic manipulation replaces natural selection, evolution and speciation[33]. Pesticides are particularly detrimental to insect communities as most of the times their action is nonspecific, affecting not only the target species but also its predators[40,41]. Other causes of biodiversity decline that are directly or indirectly linked to agriculture include eutrophication[42], species introduction[8] and climate change[43–46].

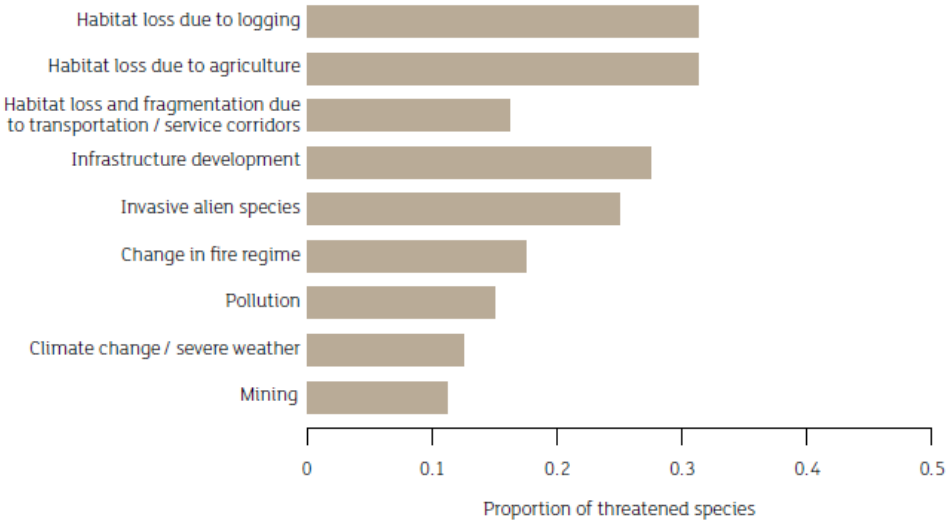


Figure 1-1: Representation of the main threats that globally affect terrestrial invertebrates on the IUCN Red List. The bars represent the proportion of affected threatened species (within the categories of Critically Endangered (CR), Endangered (EN) and Vulnerable (VU)). Source: Collen et al. 2010.

Whereas intensive agriculture exploitation relates to biodiversity loss, extensive agricultural systems as those used in low-intensity farming, may cause diversity to increase by introducing heterogeneity in the landscape[36,47–49]. Particularly in Europe, some agricultural landscapes have high nature conservation and cultural/aesthetical value[50]. These systems have existed for centuries and biodiversity has managed to cope and evolve within this framework. Where there used to be continuous homogenous woodlands, the fragmentation caused by the introduction of agriculture introduced habitat diversity creating these designated High Nature Value farmlands (HNVF)[47]. Contrary to what one might think, here the depopulation of the interior and agricultural land abandonment is expected to cause biodiversity to decline.

Insects strongly interact with agricultural landscapes, being responsible for several ecosystem services and disservices[23,24]. On one hand, insects have a fundamental ecological role in nutrient cycling, plant pollination, preserving soil structure and quality, controlling proportions of populations of prey and predators as they are major food source for several organisms[19,23]. Ultimately these processes relate to the stability of economic systems. For example, it is estimated that insect pollination is worth 9,5% of the economic value of the world's entire food supply[51]. On the other hand, some represent potential pests which can have devastating effects in crop productivity[23] with estimates of 18 to 20% of the worldwide annual production being lost due to the action of arthropods[52].

1.2. Monitoring complex insect communities with DNA metabarcoding

Being the most diverse animal group on the planet, insect communities have large numbers of species interacting in complex ways, making accurate species-level identification challenging. Historically, it has been done by observation of morphological and behavioral characteristics, requiring a high level of taxonomic expertise and being time-consuming and expensive[53]. Moreover, besides a general lack of knowledge about insect species and communities, there are also few specialized taxonomists and even they often struggle with species identifications[54]. Several factors as phenotypic plasticity within the same taxa, similarities among different taxa and the occurrence of cryptic species are likely to cause errors in species identification[54]. Most insects undergo complete metamorphosis and most species are hardly impossible to identify while in their larval stages. Frequently it requires examining internal characteres which implies microscopic preparation and observation of their genitalia. For some of those, it can only be done by collecting either adult male or female specimens as only one of the sexes has relevant characters for species identification. Finally, the species concept itself if ambiguous**,** and the boundary/threshold between different species is many times hard to observe[13,55]. Recently more and more molecular methods and technology have been developed and tested and an outbreak in molecular identification of animal samples occurred when DNA barcoding was introduced.

DNA barcoding makes use of diversity in short DNA fragments occurring between conserved regions of the genome to assign taxonomy to specimens[54,56]. Barcoding can be very useful as it allows to identify samples whose conditions prevent reliable morphological identification such as those containing remains or damaged specimens[53]. Two important properties should characterize the genes used for barcoding. Primarily they should mutate at a similar rate among different species in a way that the barcode sequence differs by a minimum percentage of base-pairs (at least 2% in closely related species) while individuals of the same species should differ very little (less than 2%)[57]. Secondly the flank regions of the barcode sequences should have a low mutation rate and low variation among most living organisms so that it becomes easy to design universal primers[57]. It has been widely done[58–61] and the Consortium for the Barcode of Life (CBOL, http://www.barcodeoflife.org/) made efforts to standardize the process through the definition of typical markers to be used for each taxon. For animals, a 658bp region of the mitochondrial cytochrome c oxidase subunit I gene (COI) is the standardized barcode marker[54], while others are proposed for plants, fungi and other life forms[57]. Because amplifying and sequencing short barcodes is not only faster but also cheaper, it became another way to document biodiversity information and over the few last years, millions of barcodes were sequenced[56]. These sequences are tied in a curated online database, the barcode of life data system (BOLD), thus kept accessible and verifiable[56]. Traditionally, barcoding has been done with Sanger sequencing which allows for low error rate when dealing with only one specimen at the time. However, with this technique it becomes virtually impossible to efficiently identify hundreds or thousands of specimens simultaneously.

The recent development of next generation sequencing allowed to characterize whole communities by pooling all samples in one sequencing run. Each of these runs allows for tens of millions of sequence-reads per experiment, enabling rapid assessments and direct measurements of biological diversity by merging DNA taxonomy and next generation DNA sequencing. Here the prefix ‘meta-’ is added to barcoding and it becomes metabarcoding, although nomenclature is variable at this stage. DNA metabarcoding has revolutionized biodiversity surveys of both bacteria, fungi, and invertebrate communities. The high level of taxonomic resolution makes it a powerful tool with versatile applications for ecological studies and it can also be faster and cheaper[57,62,63]. Massive sample collection and homogenization followed by DNA amplification and sequencing makes it possible to identify multiple species from more complex mixtures as bulk samples consisting of entire specimens[63,64]. Furthermore, it allows to explore diversity in environmental samples with degraded DNA[65] as in soil[66,67], water[68,69],

faeces[70,71], gut and stomach content[72], altogether defined as environmental DNA (eDNA). Genetic identification may also be advantageous to explore poorly known taxa and ecosystems, as well as distinguishing among cryptic species[58,64] and finding rare ones[68].

However, the method is highly dependent on polymerase chain reaction (PCR) which can induce amplification errors and relies on finding suitable metabarcoding markers[73]. Identifying many organisms from a single PCR product is subjected to primer bias and the selected region will influence the results, highlighting the importance of the choice of markers as it is easy to over or under amplify certain sequences[53]. The bioinformatic analysis of millions of produced sequences is still challenging and requires computation skills and sequencing errors are frequent[53]. Moreover, there is a lack of standardized and automatized protocols for assigning taxonomy to the obtained sequences. Additionally, this step requires comparison with a reference collection with sequences in high-quality taxonomic barcode libraries, which sometimes are unavailable. Building a good quality reference library is resource and time consuming, and it will allways still require a combination of efforts between a taxonomist and the geneticist.

DNA metabarcoding seems promising for ecological research, although some of its limitations prevent it from being considered as fully reliable source of ecological understanding to support decision making. Therefore, it becomes increasingly relevant to test and calibrate this method since it holds the potential to integrate a more standardized way of developing monitoring programs and other management actions. For example, it is necessary to estimate how does the species richness detected with metabarcoding relates to the actual species pool of the study sites[57,66], this is how well the method performs. Still, this technique has the power to provide deeper knowledge on local communities which is essential to take decisions through a preventive approach, balancing environmentally sustainable management with species conservation and local economic development.

DNA metabarcoding may be particularly useful for the assessment and monitoring of hiper-diverse insect communities, which are difficult to study otherwise. Because species are organized in communities, focusing on a single species can be a reductive approach to reality and single-species studies have progressively given place to these integrative community ecology studies[74–78]. When communities are disrupted, sets of species are replaced by others with different traits and new assemblages are formed[79]. Some species get lost by chance, others cannot cope with the new set of environmental conditions, or depended on the presence of some lost species, and finally some find in the new scenario an opportunity to thrive[79]. It is time to set up the mechanisms to determine which species are being lost and the factors that determine their presence in these habitats. Furthermore, it is important to understand which functional traits make species more likely to be present or absence from a given habitat, or whether it is more or less likely to get extinct in diverse communities [80–82].

1.3. The study case: nocturnal insects in complex mosaic landscapes of NE Portugal

To overcome the difficulties inherent to species identification in an era marked by technology, this thesis focuses on the development and test of approaches to incorporate NGS technology in designing efficient methods to rapidly assess the composition of wild assemblages of insects, and to establish monitoring programs to assess their fluctuations over time. Although there is an increasing number of publications on the use of DNA metabarcoding to identify species in bulk samples and eDNA, few studies test the potential of the methodology outside the laboratory. As this method may be an alternative to the time-consuming and expensive process of conventional morphological identification of specimens as well as more replicable across different spatial and temporal scales and verifiable by third parties. It becomes

more and more necessary to calibrate and test it to further apply it in monitoring programs and support decision making.

The study was carried out in the Natural and Regional Park of the Tua Valey (Parque Natural e Regional do Vale do Tua – PNRVT), which is characterised by a complex landscape mosaic with multiple land uses (Figure 1-2). Previous studies shown very diverse insect communities to exist in the area and resulted in new occurrences and finding novel species[83–88]. Here, we used DNA metabarcoding to characterize insect communities in two agricultural habitats – vineyards and olive groves – and two natural or semi-natural habitats – cork oak woodlands and riparian galleries. These habitats resemble different stages of the ecological sucession with by presenting a crescent order of vegetable cover and are representative of the landscape. The ecological characteristics and anthropogenic pressures felt in each habitat widely vary, thus we expect marked differences in biological communities. As insect communities may show considerable seasonal dynamics, sampling was carried out at two time periods to evaluate the ability of the method to detect temporal changes in community composition.

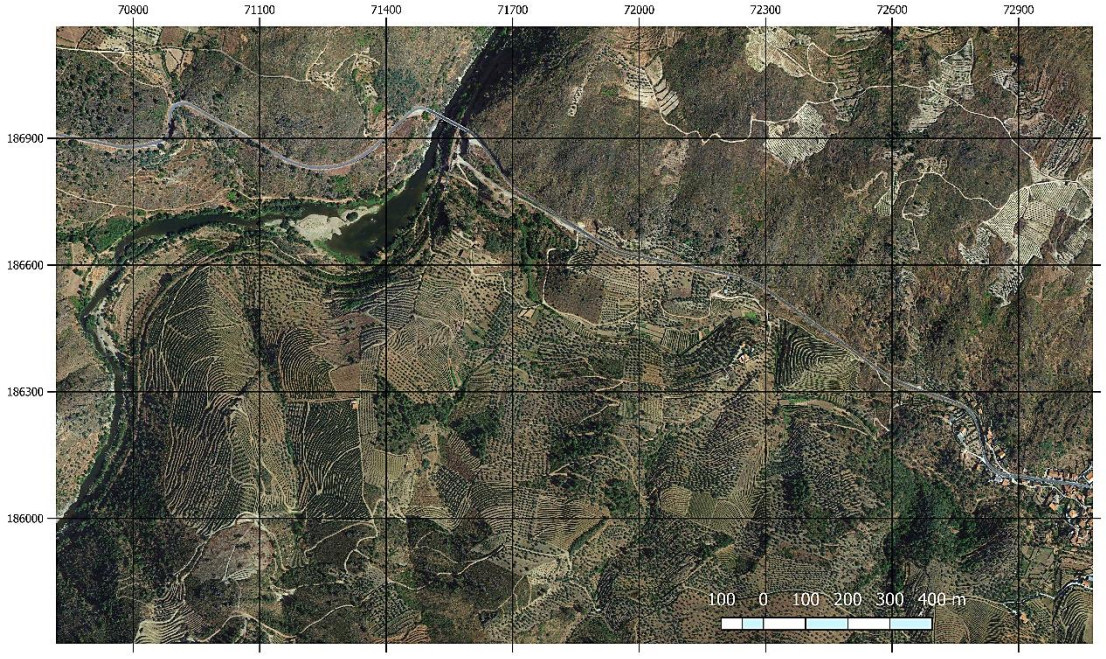


Figure 1-2: Satellite image of the fragmented landscape of the Tua Valley (image edit with QGIS with data from google maps).

The study focused on nocturnal insects as among the hidden diversity of insects, these have been hiding in the dark, and so much remains to find out. Particular attention was given to Lepidoptera, because they are a highly diverse and poorly studied group of insects, apart from diurnal butterflies which correspond to only about 5% of the species of this order. The remaining 95% are nocturnal butterflies, also known as moths[89] and very little is known about their status. The dynamics of their communities may have a deep influence in agricultural and forest systems as most species have an herbivorous diet during their initial life stages. On one hand they may assume pest behavior and cause damage in crops and stored food items either by preventing plants from carrying on photosynthesis at an adequate rate for sustaining their growth and productivity, by damaging the reproductive parts of the plant as flowers and fruits or by becoming pests in warehouses. On the other hand, some species may have a beneficial role in agriculture as they regulate the proliferation of weeds that otherwise would be eliminated with herbicides. Moreover, certain species have detritivore larvae that have similar roles to those of fungi, influencing decomposition processes[90]. Others may act as pollinators when adult and benefit the proliferation of birds, bats and predatory insects as they constitute a major source of prey for these groups, namely certain endangered species as many bats[25,70]. Moth communities are usually

understudied as most insects, not only due to their high diversity, but also because identification requires a high level of taxonomic expertise. However, because the cytochrome c oxidase I (COI) gene is frequently used for DNA barcoding, vast libraries are already available for this marker on BOLD and in the CiBIO/InBIO the reference collection.

1.4. Objectives

The overall goal of this study is to develop and test an approach to the use of DNA metabarcoding for assessment and monitoring of insect communities, focusing specifically on nocturnal insects. To achieve this goal, we sampled nocturnal insect communities in different habitats (olive groves, vineyeards, cork oak woodlands, and riparian galleries) within a complex mosaic landscape, aiming at:

1. Identify the assemblages of species present in invertebrate bulk samples (directly collected from the field), using DNA metabarcoding
2. Perform a comparative analysis of the morphologic ID versus the genetic ID to assess how well does it describe the present communities
3. Analyze the differences between insect communities in agricultural and natural habitats in a heterogeneous fragmented landscape
4. Relate the occurrence of Lepidoptera within the landscape and time period to their morphophysiological and ecological traits

2. Methods

2.1. Study area

This study was conducted within the PNRVT located in the region of Trás-os-Montes and Alto Douro in Northeast Portugal (Figure 2-1). The natural park has an area of about 24.750 ha, extending over a total of 5 municipalities: Alijó and Murça on the right margin of the Tua river, and Vila Flor, Carrazeda de Ansiães and Mirandela on the left margin. It was recently designated as a compensatory measure of the environmental impact caused by the construction of a dam at the mouth of the Tua river. Due to the flooding of the valley (from 2016 to 2017), the landscape has recently undergone through profound changes not only in appearance but also in some of its physical-chemical parameters such as humidity.

Figure 2-1: Detail of the municipalities in which the PNRVT is inserted and its location within Portugal, in the NE of the country.

The landscape is characterized by a rugged topography, with steep areas, plateaus, and the enclosed valleys of the rivers Douro, Tua and Tinhela. Inserted in the depopulated interior of north Portugal, the Alto Douro region was recognized as a World Heritage Site by UNESCO in 2001 and has been labeled as an iconic for its wine and olive production[91]. A long history of low density human occupation has transformed the landscape which thereby appears highly fragmented, as a mosaic of alternated natural plant zones (55%) and agricultural areas of various types (42%)[92]. In the latter, subsistence farming associated with vineyards and olive groves in small properties predominates. In this fragmented landscape and many animal and plant species subsist among vine terraces, olive groves, almonds, figs and other fruits, oak woodland patches, streams and rivers. Not surprising as the Mediterranean climate is associated with a high species diversity[93] and the variety of microclimates in the valleys boosts biological diversity, namely of insects. Problems such as soil erosion, lack of incentives to maintain traditional practices, and fire lead to a progressive depopulation of the interior. Human populations age and their children run away from these harsh conditions to the cities. Still, the fewer inhabitants left, kept with traditional management practices and low-intensity farming, preserving in many areas the heterogeneity of the landscape and its associated diversity. In this context the sampling design covers

both forest and agricultural areas in the trapping scheme, as an attempt to better represent this high landscape diversity.

Vineyards and olive groves are agricultural areas of medium to high intensity management whereas cork oak woodlands also had some degree of human intervention but of lower intensity level. Riparian zones were considered as representatives of a more naturalized habitat, although the sampling points were always located in areas of access to the rivers, thus frequented by bathers and fishermen, among other users. Although there was an effort to sample areas that were less affected by the flooding of the dam reservoir, some of the riparian zones characteristic vegetation was already changing. The agricultural zones were well represented in the sampling design, since they cover most of the accessible areas in the park. However, cork oak patches and riparian zones were unevenly distributed in the landscape and many times occurred in steep inaccessible places.

2.2. Sample collection

Samples were collected using custom UV light traps designed to target night flying insects, specially moths by taking advantage of their attraction to light (Figure 2-2-A). Light traps were equipped with three 30cm long UV LED strips of 395-405nm powered by a 12Ah 12V lead battery. The lights were also connected to a solar light sensor that activated the circuit at sunset and shut it down on sunrise, enabling the placement of several traps at different hours while assuring an equal functioning time for each of them throughout the night (Figure 2-2-B). In July traps were on for about 9 to 10 hours each night whereas in September, the reduction of the photoperiod disconnected the traps only after 11 to 12h. Each trap was equipped with a breathable fabric bag containing card eggboxes for the insects to hide. Traps were collected on sunrise up to a maximum of 2.5h after daylight, by simply losing each bag with a rubber band. Bags containing the insects were frozen at -20ºC degrees at least for 3h before sorting.

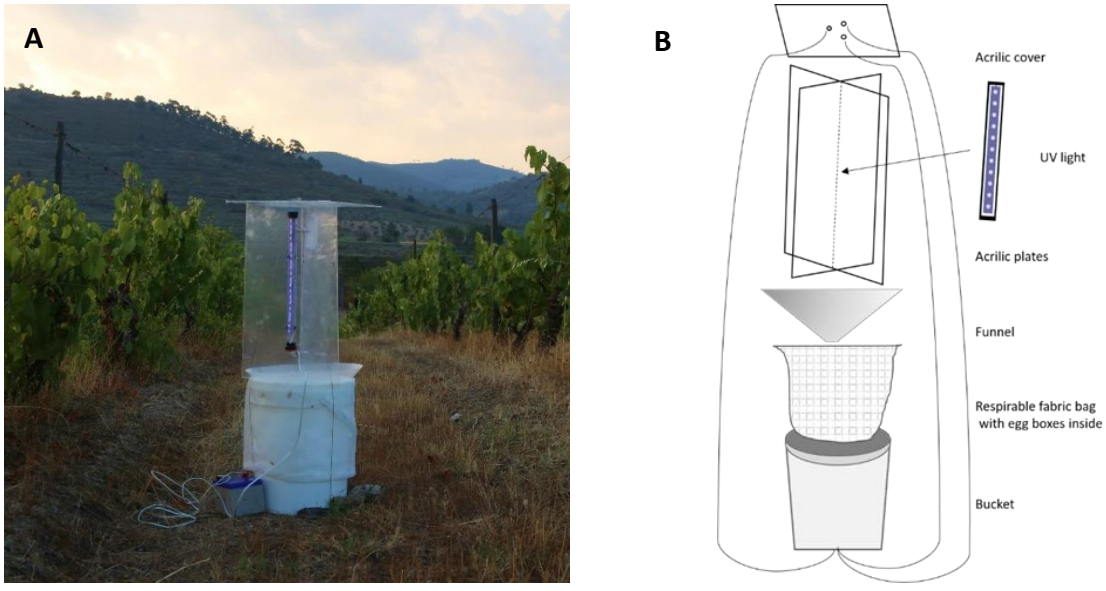


Figure 2-2: A) Picture of a complete light trap set and ready to use and B) schematic representation of the assembly of all the pieces that make up the complete trapping system.

2.3. Sampling design

To compare if the data obtained with DNA metabarcoding produced similar results to a traditional visual identification based on external morphological characters, twelve traps were collected and identified by a moth specialist. Martin Corley, who identified these moths has been studying Lepidoptera in Portugal over the past 20 years and has several published articles about them[83–86,88]. These samples were collected between the two main sampling seasons, in the end of July, as this process was dependent on the taxonomists' availability and it was the only time to do it. The intention was to test the amplification success of some primer sets (protocol testing) and to validate the metabarcoding approach.

For the monitoring plan design a total of 148 samples were set along 35 localities within the study area: 70 samples were collected in July and 78 in September (Figure 2-3). The choice of both sampling seasons – July and September – is due to the occurrence of a summer peak in insect activity for many species. On both seasons, four different kinds of habitats were sampled: vineyards, olive groves, cork oak woodland, riparian zones (Figure 2-4), as they should well represent its overall diversity, gathering this way information that could integrate a monitoring plan. On the first sampling season 2 traps (1 in olive grove and 1 in vineyard) were found disconnected by human action. Whereas on the second season, 2 traps placed in olive groves had technical issues that prevented them from working properly. A total of 144 samples end up being collected in both seasons (68 in July + 76 in September) (Table 2-1). The locaties of each trap and its geographic coordinates can be seen on Appendix I-A and the corresponding sample codes are ordered in Appendix I-B.

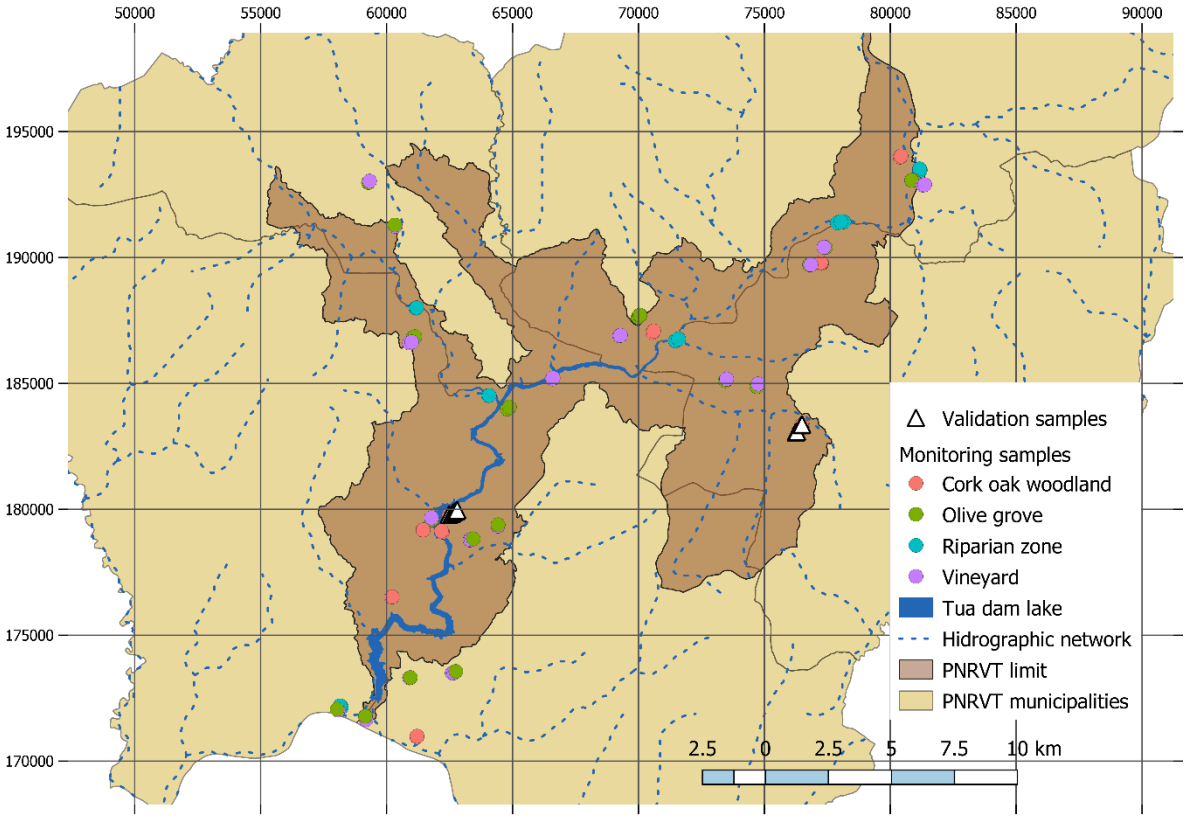


Figure 2-3: Geographic distribution of the sampling points within the PNRVT boundaries (dark brown) that are inserted in a larger area that comprises all the municipalities (light brown).

Table 2-1: Number of traps set and recovered on each sampled habitat – vineyard, olive groves, cork oak woodland and riparian zones – on both seasons, July and September 2017.

SAMPLED HABITATS		JULY 2017		SEPTEMBER 2017	
		Traps set	Traps recovered	Traps set	Traps recovered
Medium/high intensity management	VINEYARD	22	21	22	22
	OLIVE GROVE	22	21	22	20
Low intensity management	CORK OAK WOODLAND	14	14	22	22
	RIPARIAN ZONES	12	12	13	13
	TOTAL	70	68	78	76

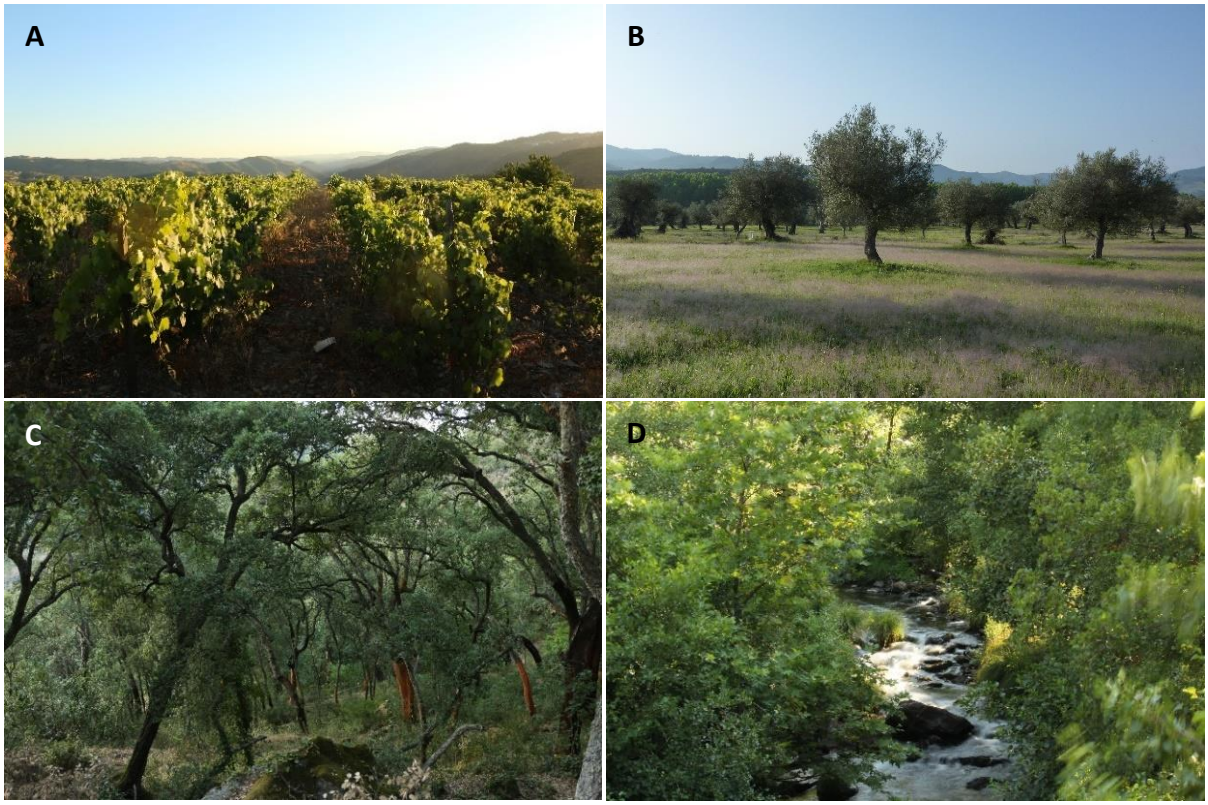


Figure 2-4: Examples of the four habitats sampled under this study: A) vineyards, B) olive groves, C) cork oak woodlands and D) riparian zones.

2.3. Sample processing and laboratorial analysis

2.3.1. Sample preparation

Frozen bags were examined to record which orders of insects were present in each trap. To avoid contamination all the material used (tweezers and spatulas) was always cleaned with ethanol 96% between samples. The contents of each bag were preserved in individual labeled falcon tubes (50mL) – one tube per sample – with ethanol 96% until DNA extraction in the lab. A few samples contained insects larger than ~5cm which were kept in separate tubes to avoid future bias in DNA amplification[94], for these only a leg was added to the bulk in the lab. Prior to DNA extraction all samples were dried by filtering out most of the ethanol and incubated at 56ºC for about 2 days. To determine if the samples

were completely dry, tubes with a large volume of insects were weighted along the drying time and samples were removed from the incubator only when those weights were constant for about half a day. Afterwards, samples were homogenized into a fine powder using the Bullet Blender 50-DX homogenizer (Next Advance, USA), with 4 glass beads of 8mm diameter per sample during 15min.

2.3.2. DNA extraction

The DNA extraction was done using the E.Z.N.A.® Tissue DNA Kit, following an adapted protocol (Appendix II). We performed up to 3 DNA extractions per sample (replicates), each one using 70 to 100 mg of the homogenized insect powder to increase species detection. For some samples with less biomass only one or two replicates were possible (Appendix III).

2.3.3. Protocol testing

To define the best protocol, a series of initial tests were done. We tested four primer sets, B2[95], Ar5[64], LCO[96] and ZBJ [97] to amplify different fragments of the Folmer region of the cytochrome oxidase I (COI) mitochondrial gene (Table 2-2). DNA amplifications were carried through polymerase chain reaction (PCR) performed over three different dilutions (1:10, 1:100, 1:200) of the DNA extractions in a total volume of 10 µL on three replicates of two random samples plus one morphologically identified sample (n=9). PCR conditions consisted of initial denaturing at 95 ºC for 15 min, followed by 35 cycles of denaturing at 95 ºC for 30 s, annealing at 45 ºC for 30 s and extension at 72 ºC for 30 s, with final elongation at 60 ºC for 10 min. The PCR products were tested in 2% agarose gel to check for the amplification success. Particularly, the strength of the bands, as well as their expected size were analyzed. After ensuring that the negative controls of the PCR were not contaminated, each successfully amplified sample went through an indexing PCR where unique indices of 7bp with a minimum of 3bp difference among them were added. This step was done using two combinations of indices per PCR product to further test the effect of sequencing depth on species diversity recovery. This way, one index combination would be sequenced at a high sequencing depth (~160,000 reads/sample), while the other would be sequenced with low sequencing depth (~80,000 reads/sample). Some samples were tested in agarose to check if the indices had been incorporated, as these indices also included the regions that blind to the flow cell of the sequencer, a 100bp difference was expected. Amplicons were cleaned and purified by using AMPure XP beads (Beckman Coulter) in a ratio of 1:0.8. Purified products were then quantified in Nanodrop, normalized to 15 nM and pooled per primerset and sequencing depth. Pooled libraries were then quantified using qPCR (KAPA Library Quant Kit qPCR Mix, Bio-Rad iCycler), diluted to 4nM each, and finally pooled together according to the desired proportions. The final pool was run on a MiSeq platform from Illumina, using a 2x250bp v2 kit.

Table 2-2: The four tested primer sets.

	Primerset	Sequence (5'-3')	Fragment size (bp)	Reference
B2	BF2	GCHCCHGAYATRGCHTTYC	461	Elbrecht & Leese 2017
	BR2	TCDGGRTGNCCRAARAAYC		
LC	LCO1490	GGTCAACAAATCATAAAGATATTGG	370	Folmer et al. 1994
	C R	GGIGGRTAIACIGTTCAICC		Shokralla et al. 2015
Ar5	ArF5	GCICCIGAYATRKCITTYCCICG	356	Gibson et al. 2014
	ArR5	GTRATDGCDCCDGCDARDACDGG		
ZBJ	ZBJ-ArtF1c	AGATATTGGAACWTTATATTTTATTTTTGG	211	Zeale et al. 2011
	ZBJ-ArtR2c	ACTAATCAATTWCCAAATCCTCC		

2.3.4. Final sequencing

The initial protocol testing revealed that using the combination of the BF2/BF2 primer-set on the 1:100 DNA dilution worked better. When sequenced at a high sequencing depth, it retrieved the highest number of different OTUs, therefore the process was replicated for all the samples. All the remaining lab procedures followed the protocol used for testing and described in the previous section, but instead of sequencing on a Miseq platform, the final library was sequenced on two lanes of the Illumina HiSeq 2500 Platform using a 2x250bp RapidRun kit.

2.4. Bioinformatic analysis

Bioinformatic processing of sequencing reads was done using OBITools[99], a set of python programs that was designed to simplify the treatment of NGS output data. For the protocol testing, a separate analysis was carried for each primer set. All the remaining bioinformatic steps were similar in the analysis of the samples generated by the HiSeq run. First, paired-end reads were aligned using the command 'illuminapairedend' and discarded if overlapping quality was less than 40. Second, reads were assigned to samples and primer sequences were removed using 'ngsfilter', allowing a total of 4 mismatches to the expected primer sequence. Finally, reads were collapsed into haplotypes and singletons (haplotypes with only one read per sample) were removed. Haplotype diversity and read count per fragment length, as well as bibliographic information of each marker was used to discard haplotypes shorter and/or longer than expected. The remaining haplotypes left per sample were all joined in a unique file and the 'obiuniq' command was run again to obtain the set of unique sequences across the entire dataset. The command 'obiclean' was then used to remove potentially spurious sequences with an 'r' level of one, meaning that any 'A' haplotype differing one base-pair from a 'B' haplotype, with an absolute read count lower than 'B', and that was not found without the presence of 'B' in any sample, was removed as it was most likely a PCR or sequencing error.

The taxonomic assignment of each haplotype to a taxon was done with the support of a neighbour-joining phylogenic tree based on an alignment of the haplotypes for each primerset sorted by their read count. This allowed to visually define clusters of haplotypes that corresponded to the same taxon, as well as identify chimeric sequences and PCR errors that were unfiltered in the previous bioinformatic steps. Taxa were identified by comparing the representative haplotypes of each cluster against online databases (BOLD), as well as unpublished sequences of arthropods collected in northern Portugal for CIBIO-InBIO Barcoding Initiative (here referred to as reference collection). Species level identifications were made for similarity values of above 98.5%. Whenever a haplotype matched several species, genus, or families at similar identity levels we tried to select the most inclusive taxonomic rank. For example, if a haplotype matched with 95% similarity two species of different genus belonging to the same family, we identified it only to family level. Those that best matched the same taxa were collapsed into a single taxon as we assumed that they belonged to the same OTU. Haplotypes whose identification was only possible up to family, order or class level were clustered according to their similarity into distinct taxa (e.g., Carabidae 1, Carabidae 2, and so on). After identifying all the haplotypes, we excluded the ones not belonging to the phylum Arthropoda.

2.5. Trait database

The species functional trait database was built by compiling information on traits of moth species known to occur in the Tua Valley from several online sources[100–116]. A total of twelve different traits distributed among five main categories were collected, namely moths' size (wingspan), phenology (flight time, migratory behavior, number of generations per year, winter status), diet (polyphagia index, detritivores

index, leafminer behavior, list of host plants and part of the plant consumed), habitat (preferred habitats) and distribution (within Europe, due to lack of global data for many species). Polyphagia index ranged from 1 to 3, with 1 being attributed to monophagous species, i.e. those feeding on only one species or genus, 2 to oligophagous species, i.e. those that feed on host plants from an entire family, and 3 to those that feed on more than one family of host plants. Both migratory behavior, detritivore index and leafminer behavior were binary variables meaning 1 presents the behavior, 0 does not present the behavior. Most information on moth's phenology (flight time and number of generations) were from other countries of Europe. When different results were found, the values correspondent to the country closest to the Iberian Peninsula were selected. For example, when a certain moth species had 1 generation in Belgium but 3 in France, the value for France was taken.

2.6. Data analysis

The analysis was designed in 4 stages to cover up all the objectives (Figure 2-5). Stage 1 consisted of a simple protocol testing: some samples from both the validation and the monitoring sample collection were used to test the efficiency of different primers, as well as sequencing depth, in species detection in a test run**.** This enabled the choice of an adequate primer set before sequencing all samples, as well as the number of sequencing reads needed to properly characterize each sample.

Stage 2 involved testing if metabarcoding produces similar results to those generated by visual identification and examine the differences between both methods and if the produced results were similar. The list of species obtained with the metabarcoding approach was compared with the list of species identified by morphology. To compare the identifications produced by each method on the twelve samples collected for the validation step, we performed a permutational multivariate analysis of variance (PermANOVA). This analysis was carried with the function 'adonis' implemented in the R package 'vegan' with 999 permutations.

Stage 3 consisted in using the monitoring samples to assess the diversity and species composition on the four different habitat categories, with increasing vegetable cover and decreasing level of management: vineyards, olive groves, cork oak woodlands and riparian zones. First, rarefaction curves of species richness/sample coverage using 'iNEXT' R package (Hill number with q=0) to calculate the expected species richness. Sample size-based rarefaction and samples completeness curves were used to assess if both sample effort and sample coverage (number of reads) were enough to estimate species richness. Then, the average species richness was calculated for each habitat and season. Moreover, we calculated the taxonomic distinctness index (delta +) proposed by Clarke and Warwick[117]. This index uses the path length in a Linnaean taxonomy to estimate the average 'distance' between pairs of taxonomical units in a sample, capturing taxonomic diversity instead of plain richness. To test for the differences in average species richness and taxonomic diversity among habitats and seasons we used the R package 'MASS' to calculate GLMs. A poisson distribution was used for species richness, while normal distribution was used for the taxonomic diversity index. Moreover, we used multiple comparison tests with Bonferroni correction to assess which habitats were statistically different from each other. This was implemented in the package 'lsmeans'.

The most common species across each habitat and the set of habitats together (across the landscape) were identified. For that purpose, we used barplots to represent the relative number of occurrences of the top 20 most found OTUs in each habitat and in all habitats. Moreover, we built up Venn diagrams to check the number of OTUs shared among habitats. The Jaccard similarity matrix (β-diversity) between all pairs of samples was calculated and used as an input to perform a principal coordinate analysis (PCoA) also known as classical multidimensional scaling (CMS). This analysis allows to a calculate the axis of major variation in the composition of the communities and to re-represent the data

spatially in a way that each unit is a point and their distance is representative of their differences. Here, similar samples will tend to cluster and samples that differ the most will appear more distant in the chart. Points were colored by their habitat and shaped by their season, allowing to visualize which of these variables relates to the variation in the assemblages on each sample. The differences in species composition between each group (habitat and season) indicated by the PCoA were tested through a PermANOVA, using 999 permutations. A percentage of similarities analysis (SIMPER) was performed to check which species were contributing the most to the recorded differences of community composition between habitats and seasons. Finally, a matrix of geographical distance between samples was calculated using the R package 'rgdal'. We used a mantel test (999 permutations) to compare it with the Jaccard similarity matrix (β diversity) used to calculate the PCoA in order to assess if species composition was spatially correlated.

Finally, in stage 4 we use statistical models to relate the data on species occurrence combined with the collected traits to infer whether there were functional traits associated with certain habitats. This was carried out using a "fourth corner analysis"[118] calculated using a binomial regression with the 'traitglm' function within the R package 'mvabund'. This analysis uses 3 input matrices describing species occurrence per sampling unit, environmental covariates per sampling unit and traits of the occurring species to estimate a 4th matrix that relates the environmental covariates with species traits[118]. In this case we intend to relate the 4 habitats and 2 sampled seasons with the following species traits: average wingspan, maximum generation number, migratory behavior (0/1) and 3 diet traits – polyphagia (1 to 3), detritivity (0/1) and leafminer behavior (0/1). In this case we chose to use a function added in the 'mvabund' package for fitting GLMs using a LASSO penalty which uses model selection to evaluate which variables are significantly contributing to the model.

Figure 2-5: Schematic representation of the 4-stage study design.

3. Results

3.1. Protocol testing

When the PCR products were run in agarose gel, clean and strong bands were obtained for both B2 and LCO primer sets, while Ar5 did not produce any bands and ZBJ produced faint ones. Therefore, the procedure for sequencing was carried only with the primers that seemed to have successfully amplified the samples – B2 and LCO. The 1:100 dilution was selected as it seemed to produce the stronger bands for both primer pairs. The data produced by the Miseq run revealed that the B2 primer set allowed to detect 26 more Lepidoptera species than LCO (Table 0-1, Appendix III). Moreover, B2 produced less errors as chimeric sequences, thus it generated only 1204 haplotypes after the bioinformatic filtering while LCO produced 3096 haplotypes, enabling a more efficient analysis. The PCR products sequenced at higher sequencing (~160,000 reads/sample) depth seemed to better capture the diversity of each sample, while the ones sequenced at lower depths (~80,000 reads/sample) did not always reach the asymptote of expected number species, especially in more diverse samples like IT141 (Figure 3-1). The coverage test results allowed to estimate the number of reads that would be necessary to aim for in the Hi Seq run to detect most species, even in high biomass samples.

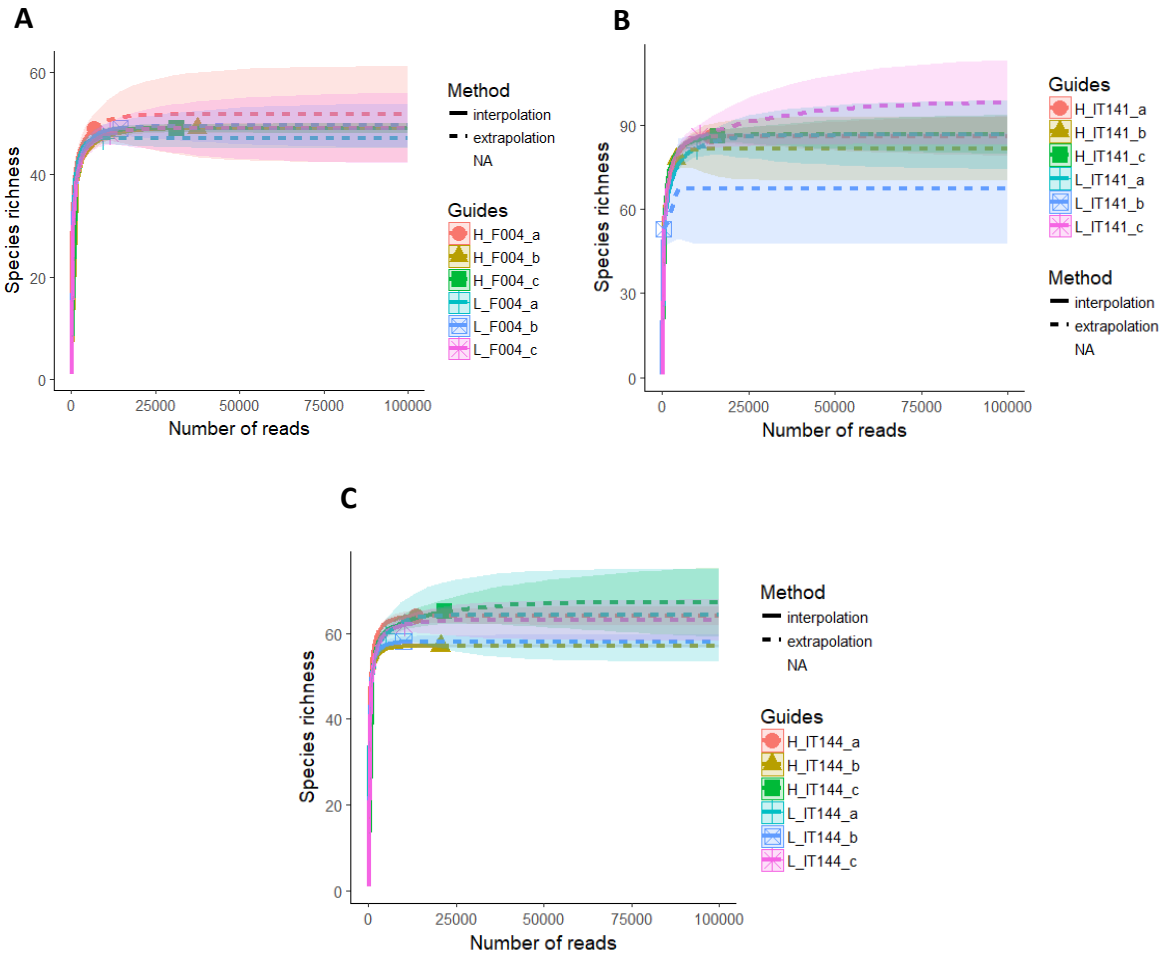


Figure 3-1: Coverage test results from the Mi Seq run for the three extractions (a, b and c) of the test samples: A) F004 which was visually identified, B) IT141 a "regular" sample and C) IT144 a sample with high insect biomass.

3.2. Validation of metabarcoding

The metabarcoding and visual identification of the 12 test traps allowed the recovery of 108 and 101 species, respectively. Although most species records were shared between both methods (268 out of 366 – 73%, Table 3-1, Appendix IV), there were some species either missing in the metabarcoding analysis or identified differently (e.g. same genus, but different species) by visual ID. In other cases, the lack of a complete reference database did not allow a species level identification of retrieved DNA sequences, with barcodes identified only at genus, family or order level. Nevertheless, the PerMANOVA analysis revealed that the species composition obtained using both methods did not differ significantly (df = 1, F = 0.9856, R^2 = 0.0429, P = 0.492).

Table 3-1: Lepidoptera families detected using either only DNA metabarcoding but not through visual identification and the opposite: families detected through visual identification but not when using DNA metabarcoding. Values refer to the percentage of occurences of species in the twelve samples identified exclusively by metabarcoding or visual ID as well as by both methods.

Family	Metabarcoding ID (%)	Visual ID (%)	Both methods (%)	Total
Autostichidae	0.55	0.27	0.82	1.64
Brachodidae	1.64	1.64	0.27	3.55
Crambidae	0.27	0.27	4.64	5.19
Depressaridae	0.55	0.55	0.82	1.91
Drepanidae	0.00	0.27	1.37	1.64
Erebidae	0.27	0.55	9.84	10.66
Gelechiidae	0.82	0.55	3.55	4.92
Geometridae	3.28	4.10	17.49	24.86
Lasiocampidae	0.00	0.00	3.01	3.01
Lecithoceridae	0.27	0.55	0.27	1.09
Momphidae	0.00	0.27	0.00	0.27
Noctuidae	1.64	1.91	12.02	15.57
Nolidae	0.00	0.27	0.82	1.09
Notodontidae	0.27	0.00	1.37	1.64
Nymphalidae	0.00	0.00	0.27	0.27
Oecophoridae	0.00	0.27	1.37	1.64
Pterophoridae	0.00	0.00	0.55	0.55
Pyralidae	0.82	1.91	10.38	13.11
Scythridae	0.82	1.09	0.00	1.91
Tineidae	0.00	0.00	1.09	1.09
Tortricidae	0.00	0.00	3.28	3.28
Unknown	0.82	0.00	0.00	0.82
Ypsolophidae	0.00	0.27	0.00	0.27
Total	12.02	14.75	73.22	100.00

3.3. Species richness

A total of 1130 final identifications were attributed to OTUs belonging to 163 different families splited among 21 arthropod orders. The habitats with higher number of different OTUs recorded in total were olive groves (735 OTUs) followed by vineyards (658), cork oak woodlands (634) and riparian zones (578).

3.3.1. Species richness according to sample size and sample completeness

Species accumulation curves indicate that the sampling effort was not enough to capture the total species richness for most habitats, either considering all insect OTUs (Figure 3-2) or just Lepidoptera (Figure 3-3). Sample coverage was similar for all habitats, meaning that although the total number of species occurring in each habitat was not captured, the diversity found can be directly compared among habitats. Still, a total of 1130 insect OTUs and from which 400 were Lepidoptera were found within the samples.

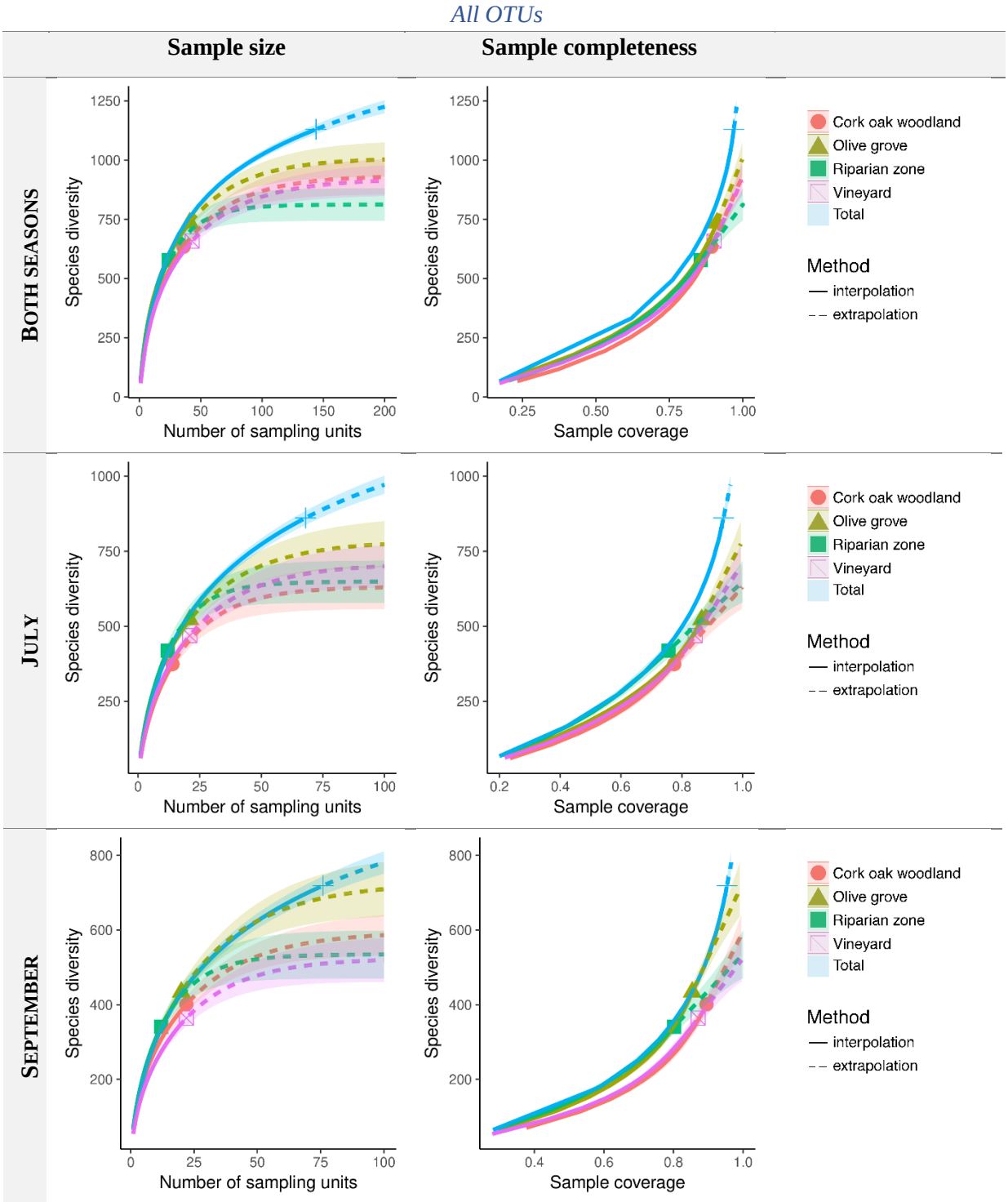


Figure 3-2: Hi Seq coverage test to check how sampling effort (number of sampling units) and sample coverage performed in order to get a good proportion of total species richness.

Lepidoptera

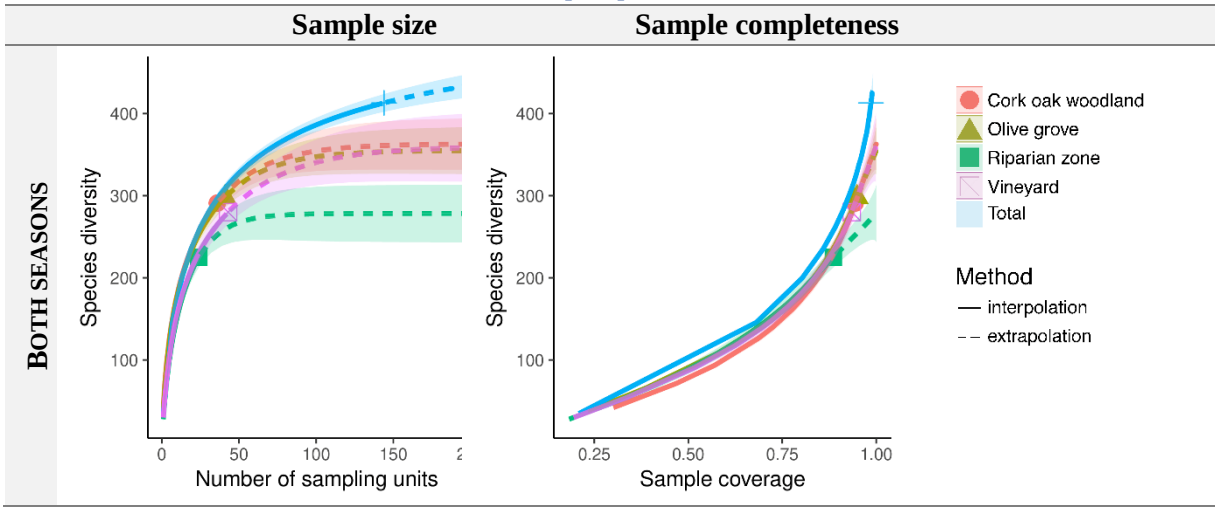
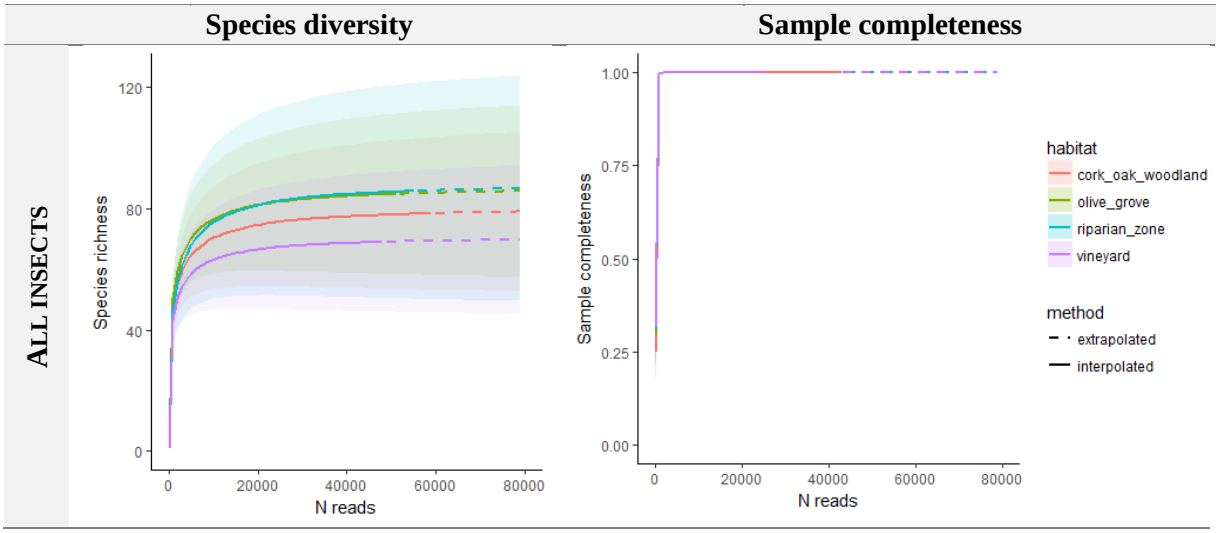


Figure 3-3: Hi Seq coverage test to check if sampling effort (number of sampling units) and sample coverage were enough to capture Lepidoptera total richness.

3.3.2. Species richness and sample completeness according to sample coverage

Contrarly to sample number, sample coverage was enough to capture species richness, this means the number of reads performed by the sequencer enabled to capture most of the species present in the samples (Figure 3-4). The extrapolation starts when the diversity curves are already stable, which means that more reads would not enable to detect more OTUs, whether for all insects or only Lepidoptera. The sample completeness analysis indicates that after filtering, about 1000 reads per sample would be enough to detect all species within each trap.



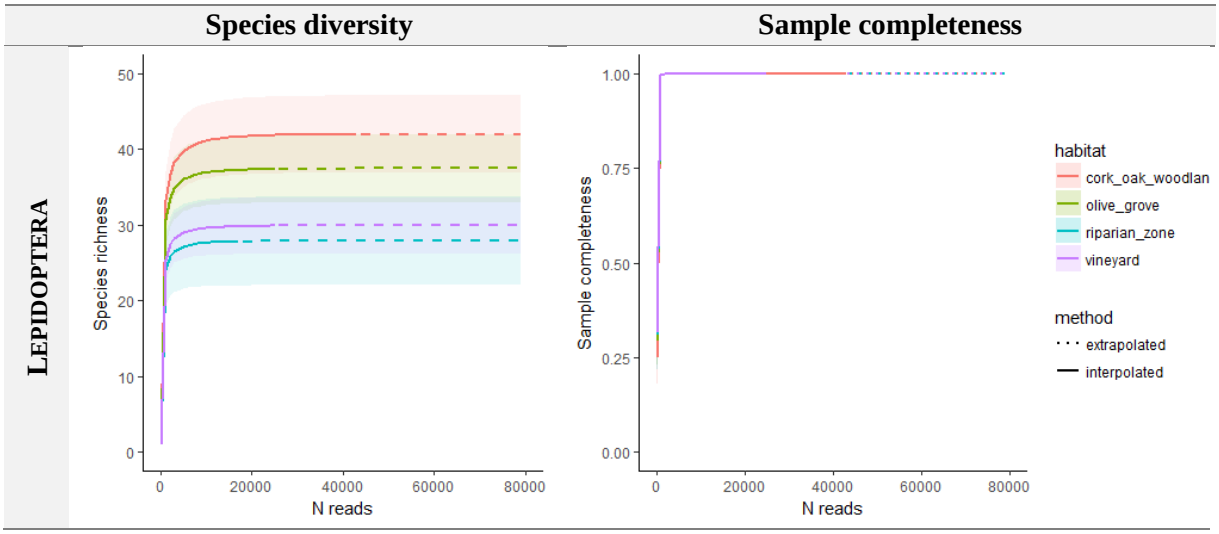


Figure 3-4: Hi Seq coverage test to check if read coverage (number of reads per habitat) and sample completeness were enough to capture Lepidoptera total richness.

3.3.3. Mean species richness

The average number of species found per sample varied according to the habitat and season sampled in different ways, either considering all insects or just moths (Table 3-2). For all insect OTUs in July, olive groves and riparian zones were the habitats with highest richness (Figure 3-5-A) and did not differ from each other (Appendix V-A). They both exhibited higher richness than cork oak woodlands and vineyards, although for riparian zones the difference with vineyards was not significant. In September this pattern changed slightly as the average richness of cork oak woodlands increased to similar levels of those of olive groves and riparian zones, and the richness of vineyards decreased to even lower values. For Lepidoptera OTUs in July, cork oak woodlands and olive groves exhibited the highest mean species richness (Figure 3-5-B), although only olive groves had significantly higher richness than riparian zones and vineyards (Appendix V-B). As for all insects, in September this pattern changed, but this time cork oak woodlands became the richest habitat, followed by olive groves, riparian zones and vineyards. Overall, the pattern observed in September can be characterized by an increase in the average number of Lepidoptera species found on each habitat, especially in cork oak woodlands and olive groves.

Table 3-2: GLM results highlighting the variables whose variation had significant effects in species richness ($p < 0.05$).

		LR Chisq	*Df*	*p-value*	
All insects richness	Habitat	71.377	3	**<0.00001**	***
	Season	3.183	1	0.0744	
	Habitat*Season	29.425	3	**<0.00001**	***
Lepidoptera richness	Habitat	110.631	3	**<0.00001**	***
	Season	108.477	1	**<0.00001**	***
	Habitat*Season	14.966	3	**0.001846**	**

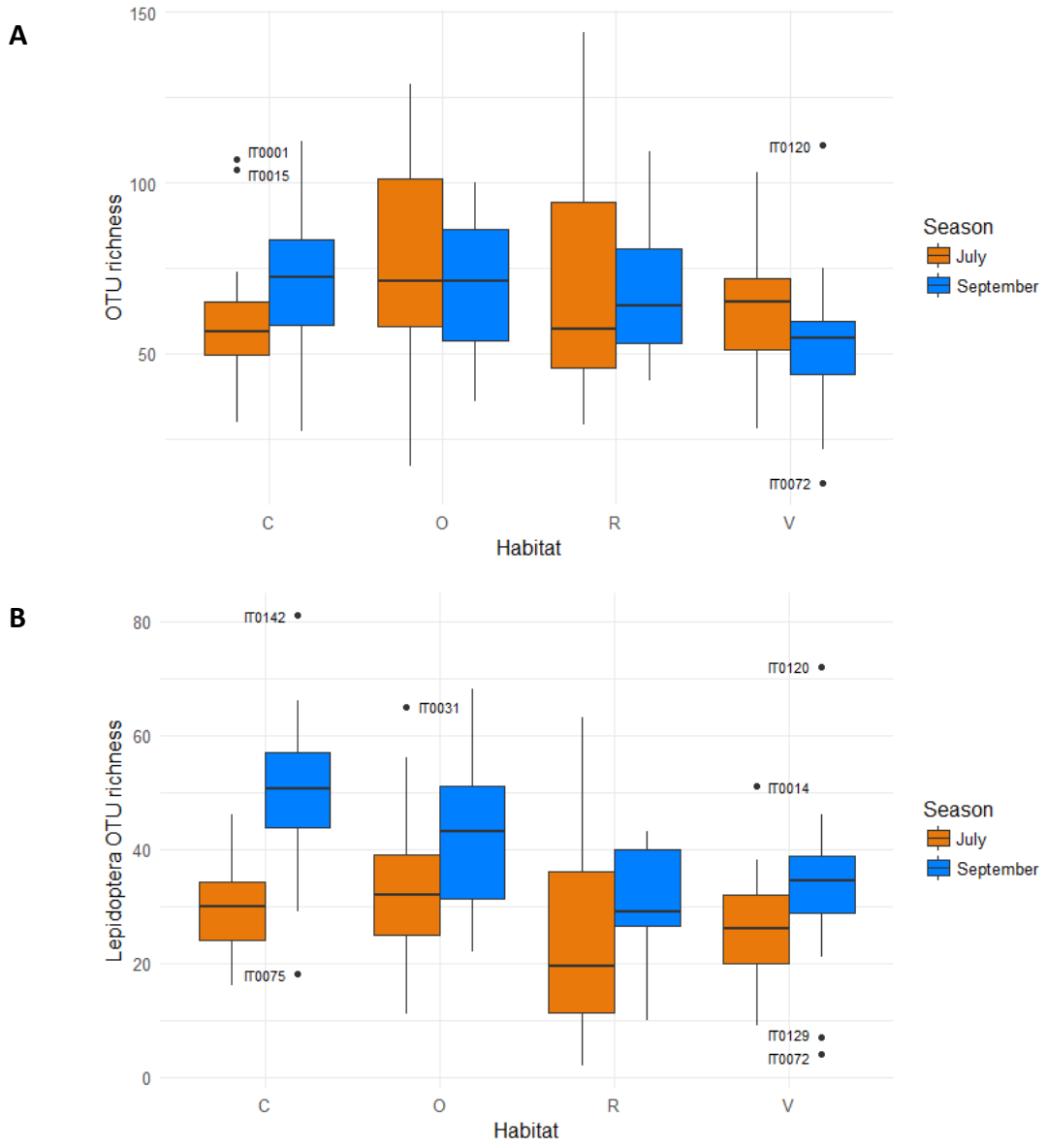


Figure 3-5: Box-plot diagram showing the variation of the OTU richnes for A) all insect orders and B) Lepidoptera within the 4 sampled habitats C – Cork oak woodland, O – Olive grove, R – Riparian zone and V – Vineyard on both seasons, July (orange) and September (blue). Outliers are represented as black dots and the correspondent samples labeled.

3.3.4. Taxonomic distinctiveness

The index of taxonomic distinctiveness (delta +) was not correlated with OTU richness (pearson coef = -0.1070) nor with Lepidoptera richness (pearson coef = -0.0703) (Appendix VI). The GLM results indicated that there was no significant interaction (habitat*season) for any of the delta+ values, indicating that unlike species richness, it varied similarly among habitats in both seasons (Table 3-3). Delta+ was significantly higher in July than September when considering all insect orders (Figure 3-6-A) but remained constant when considering only moths (Figure 3-6-B). Delta+ also varied among habitats when considering all insects, and to a much lower magnitude when considering just moths. For all insects, delta+ was highest in riparian zones, followed by vineyards, olive groves and cork oak woodlands. Multiple comparison tests showed that the intermediate levels of olive groves did not differ significantly though from both vineyards and cork oak woodlands (Appendix VII-A). Looking at the variation of taxonomic distinctiveness in communities of Lepidoptera, the values of the index were

overall constant across habitats and seasons, except that olive groves had significantly lower delta+ than vineyards, although the signal was weak (Appendix VII-B).

Table 3-3: GLM results highlighting the variables whose variation had significant effects in the variation of the taxonomic index Delta+ ($p < 0.05$).

		LR Chisq	*Df*	*p-value*	
All insects Delta+	Habitat	43.380	3	**2.044e-09**	***
	Season	43.806	1	**3.627e-11**	***
	Habitat*Season	1.639	3	0.6506	
Lepidoptera Delta+	Habitat	8.0932	3	**0.04412**	*
	Season	0.4922	1	0.48293	
	Habitat*Season	5.2480	3	0.15451	

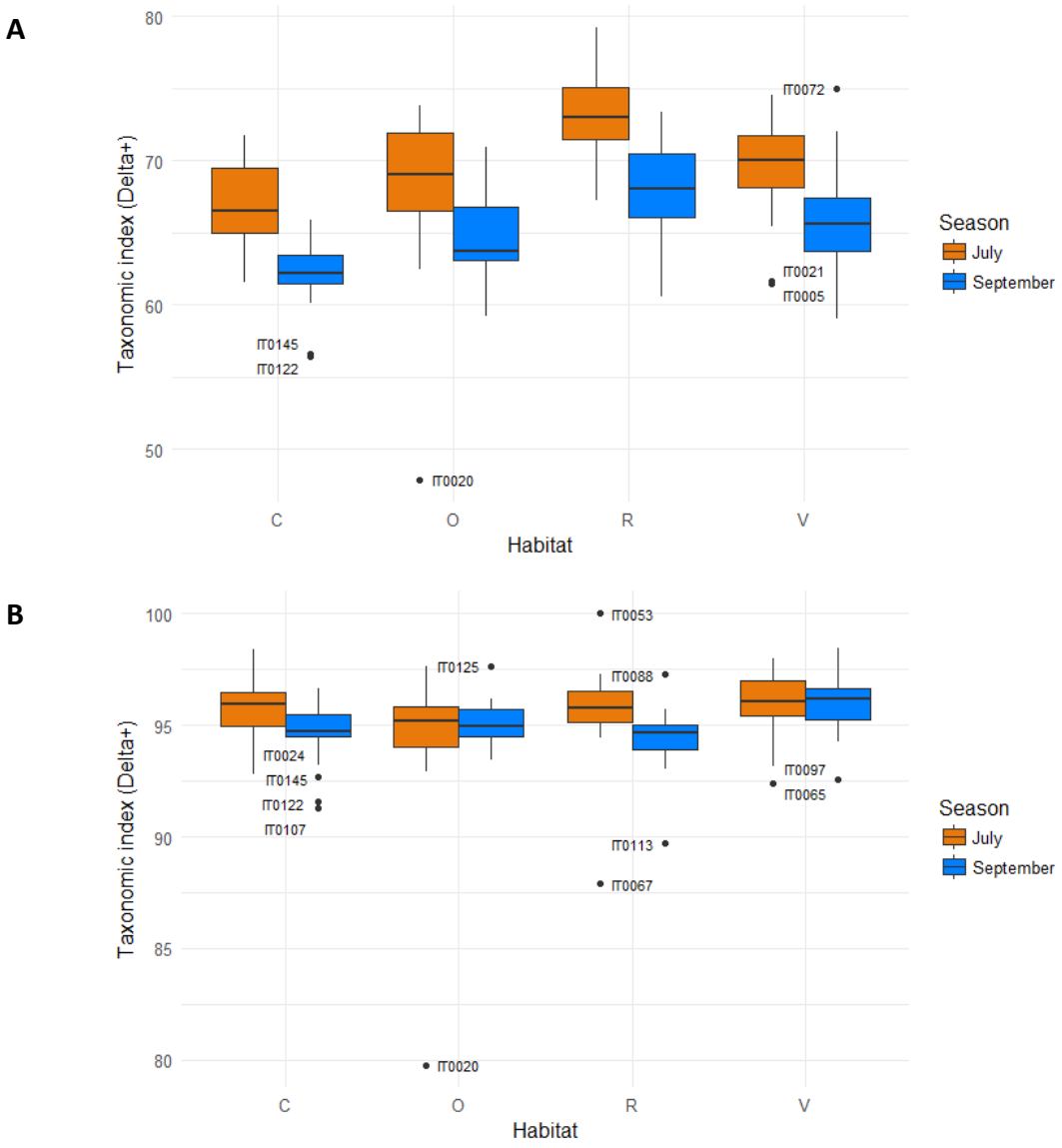


Figure 3-6: Box-plot diagram showing the variation of the index of taxonomic distinctiveness for A) all insect orders and B) Lepidoptera within the 4 sampled habitats C – Cork oak woodland, O – Olive grove, R – Riparian zone and V – Vineyard on both seasons, July (orange) and September (blue). Outliers are represented as black dots and the darkred dots mark the average values.

3.4. Community composition

3.4.1. Common species

The structure of the OTU assemblages seems to be similar across habitats when it comes to relative frequency of occurrence (i.e. proportion of the traps of each habitat in which a certain OTU appeared) (Figure 3-7). Still, the most common OTUs varied between habitats. For example, if we consider dominant OTUs those that appear in more than 50% of the traps, olive groves had 3 dominant OTUs: *Dysspastus fallax* (78.1%), *Chironomidae sp. 4* (65.9%) and *Ancylosis oblitella* (63.4%). Vineyards also had three dominant species and the first 2 are shared with olive groves: *Chironomidae sp. 4* (69.8%) and *Dysspastus fallax* (55.8%), the 3rd was a complex of species *Eurodachtha siculella/canigella* (51.2%). The dominant species in cork oak forest were all Lepidoptera: *Watsonalla uncinula* which appeared in 86,1% of the traps set in this habitat followed by *Cydia fagiglandana* (69,4%) and *Acrobasis glaucela* (66.7%) with much lower proportion of occurrence. In riparian zones the 3 most frequent species were all Diptera from the family Chironomidae with similar percentages.

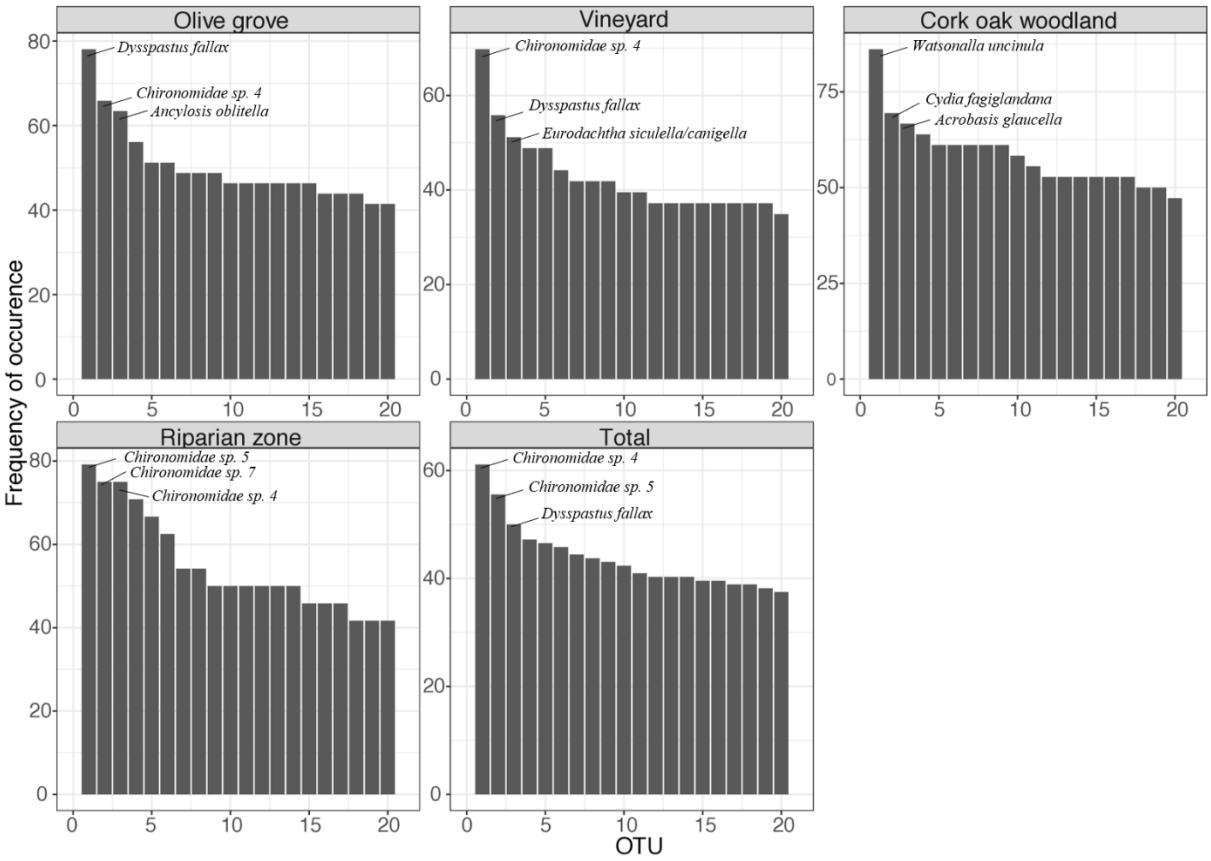


Figure 3-7: Bar charts represent the relative frequency of occurrence of the top 20 most frequent OTUs in each habitat and in all habitats together (landscape level). The values were obtained by dividing the number of traps were the OTU was detected (absolute frequency of occurrence) by the total number of traps collected in that habitat. The three most frequent OTUs of each habitat and total are mentioned in the charts.

Lepidoptera OTUs have slightly different patterns (Figure 3-8). Both olive groves and vineyards had 2 dominant species with the first being *Dysspastus fallax* on both, although the percentage of occurrence in olive groves (78.0%) was much higher than in vineyards (55.8%). The second species for olive groves and vineyards were respectively *Ancylosis oblitella* (63.4%) and *Eurodachtha siculella/canigella* (51.2%). In cork oak woodland there was a clear dominance of *Watsonalla uncinula* (86.1%) although many other species occurred in over 50% of the traps placed in this habitat. In riparian zones *Mythimna sicula* was the only species occurring in more than 50% of the traps. Only one species, *Dysspastus fallax*,

occurred in exactly 50% of all habitats combined. *Caradrina flavirena* was not dominant in any of the habitats when considered singularly, but it appeared in 47.2% of the traps of all habitats.

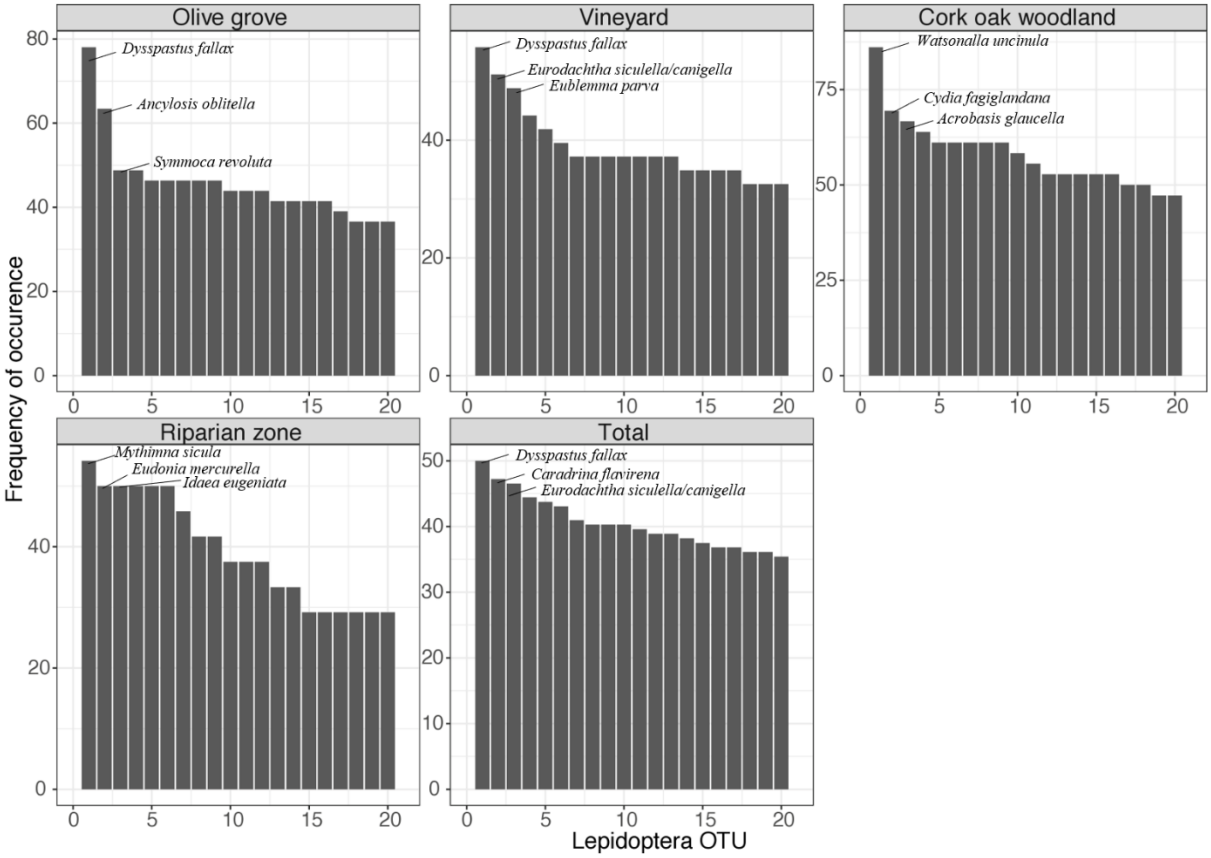


Figure 3-8: Bar charts represent the relative frequency of occurrence of the top 20 most frequent Lepidoptera OTUs in each habitat and in all habitats together (landscape level). The values were obtained by dividing the number of traps were the Lepidoptera OTU was detected (absolute frequency of occurrence) by the total number of traps collected in that habitat. The three most frequent Lepidoptera of each habitat and total are mentioned in the charts.

3.4.2. Shared species

The following Venn diagrams show the number of insect OTUs (Figure 3-9-A) and Lepidoptera OTUs (Figure 3-9-B) occurring in all combinations of the four sampled habitats. Cork oak woodland is the habitat that presents most unique OTUs (113 unique OTU and 36 unique Lepidoptera OTU). Riparian zones had the largest difference between unique OTUs and unique Lepidoptera OTUs, indicating that most unique species do not belong to this order. The percentage of shared species among the 4 habitats was 21,77% (246 from 1130) of all OTUs and 33,41% (138 from 413) of the Lepidoptera.

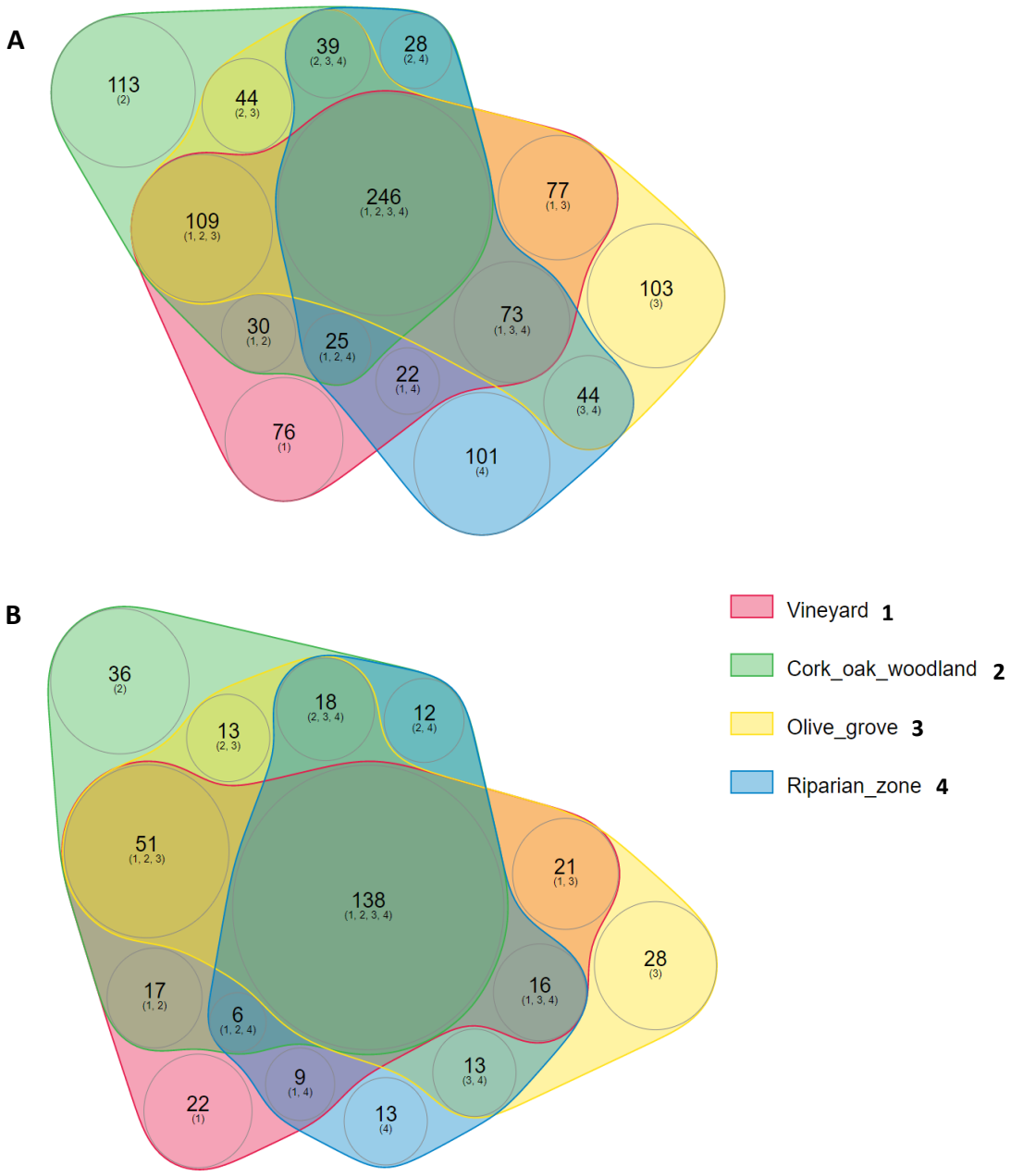


Figure 3-9: Venn diagrams showing the number of species present in each habitat, as well as their overlap in all possible combinations of the four habitats. The large values at the center of each circle correspond to the number of A) all insect OTUs and B) Lepidoptera OTUs. Each circle represents one habitat or a combination of 2, 3 or 4 habitats which appear as small numbers in brackets.

3.4.3. Community similarity

The mantel test results revealed that the geographical distance between sampling points did not correlate to the Jaccard similarity matrix or β diversity ($p = 0.089$). From the considered variables – season and habitat – the sampling period seems to be the one that most affects community composition, both when considering all insects or just Lepidoptera (Figure 3-10-A, Figure 3-11-A). When looking at each season separately (discarding season effect) it seems that communities were more homogeneous in July than in September both for all OTUs (Figure 3-10-B, Figure 3-10-C) and just Lepidoptera (Figure 3-11-B, Figure 3-11-C). In July it seems like the riparian zone had more distinct communities comparing to other habitats (Figure 3-10-B) whereas in September cork oak woodlands were more unique (Figure 3-10-C).

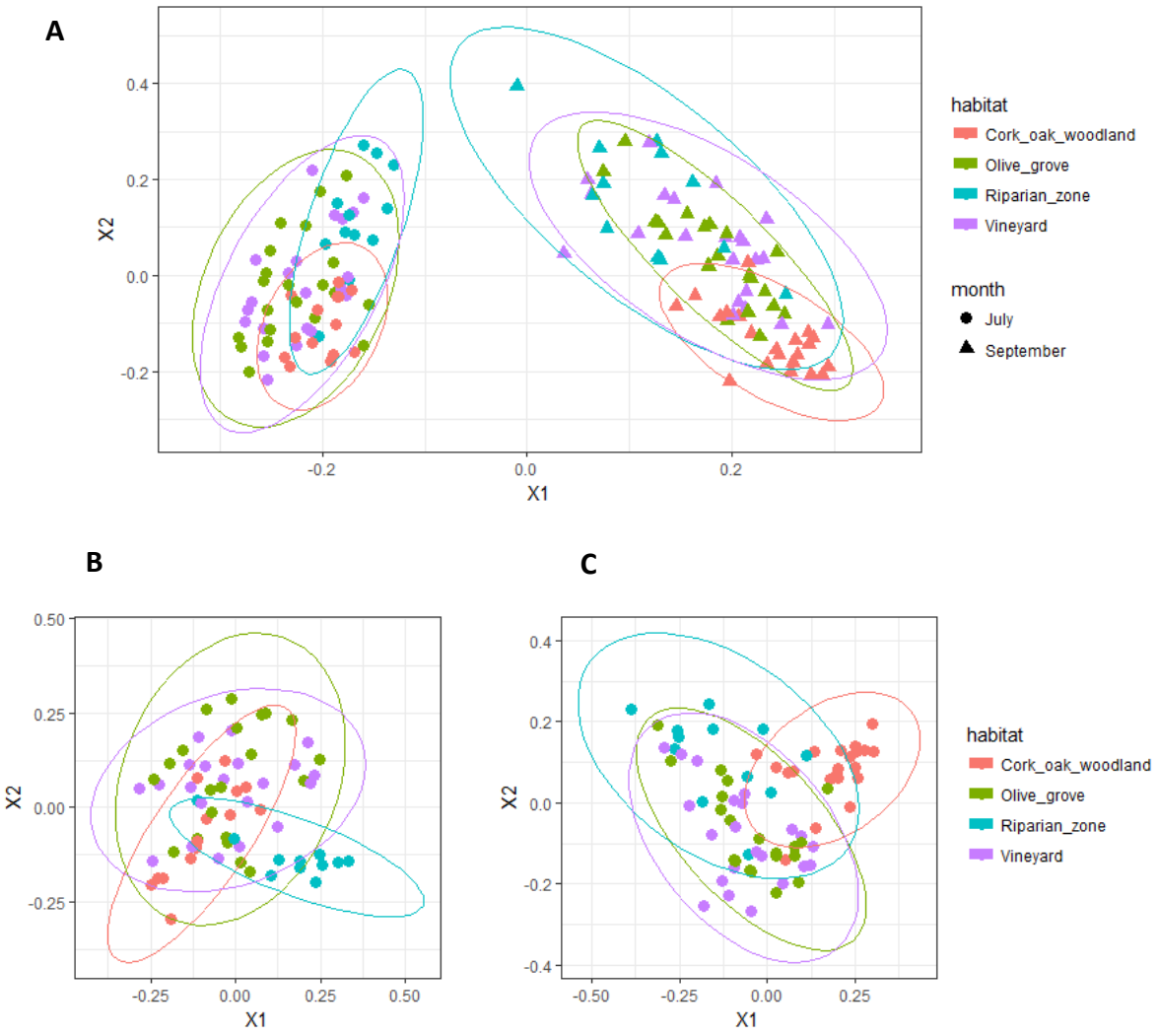


Figure 3-10: Principal coordinates analysis of Jaccard distances. Each point represents a sample and the corresponding binary composition of all insects OTUs. The samples in A) are defined in color and shape according to the habitat and season in which it was collected. The two lower graphs represent samples of different seasons, B) July and C) September separately and points are colored according to the habitat. Ellipses surround 95% of the points of each habitat/season.

Lepidoptera communities had similar OTU composition patterns to those of all insects both in July (with riparian zones standing out, Figure 3-11-B) and in September when the most unique areas were cork oak woodlands (Figure 3-11-C).

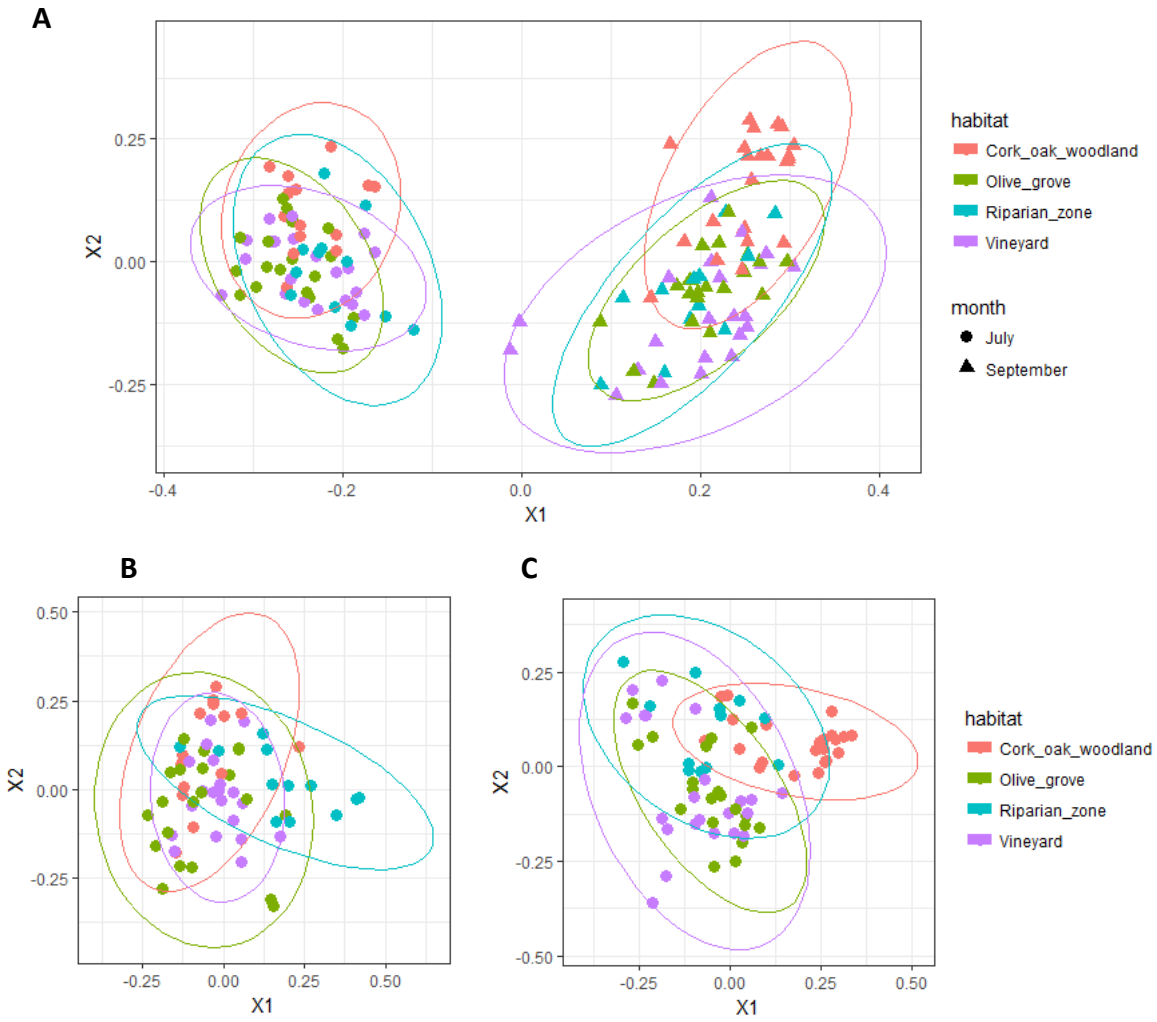


Figure 3-11: Princinpal coordinates analysis of Jaccard distances. Each point represents a sample and the corresponding binary composition of Lepidoptera OTUs. The samples in A) are defined in color and shape according to the habitat and season in which it was collected. The two lower graphs represent samples of different seasons, B) July and C) September separately and points are colored according to the habitat. Ellipses surround 95% of the points of each habitat/season.

The strong seasonal effect extendes to family level, although there is a higher overlap of communities than at the OTU level (Figure 3-12-A). When it comes to habitat, riparian zones appear to present a more distinct group of families in July (Figure 3-12-B), although they present the largest seasonal overlap. This is, the insect communities in riparian zones in July and September were more homogeneous in terms on families found, although quite distinct from the other habitats on each season. By looking at each season separately we can better observe this seasonal effect, also for cork woodland areas in September (Figure 3-12-C). Finally, at order level the communities completely overlap and there is no degree of separation either at habitat or seasonal level (Figure 3-13).

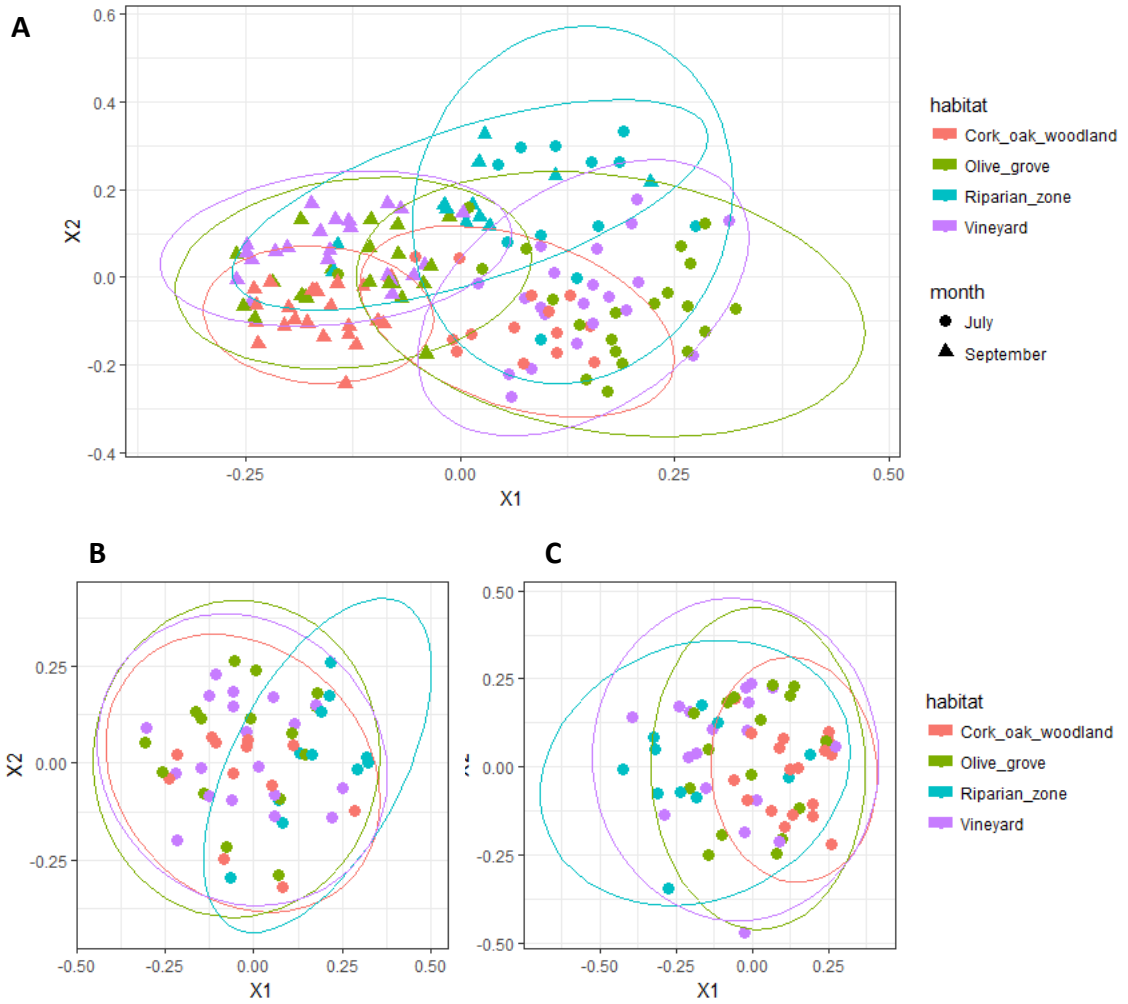


Figure 3-12: Principal coordinates analysis of Jaccard distances. Each point represents a sample and the corresponding community composition at family level. The samples in A) are defined in color and shape according to the habitat and season in which it was collected. The two lower graphs represent samples of different seasons, B) July and C) September separately and points are colored according to the habitat. Ellipses surround 95% of the points of each habitat/season.

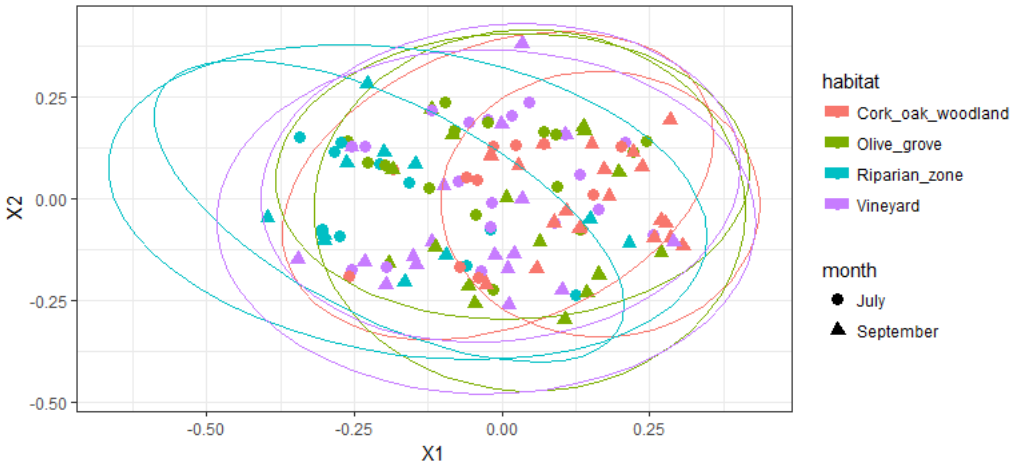


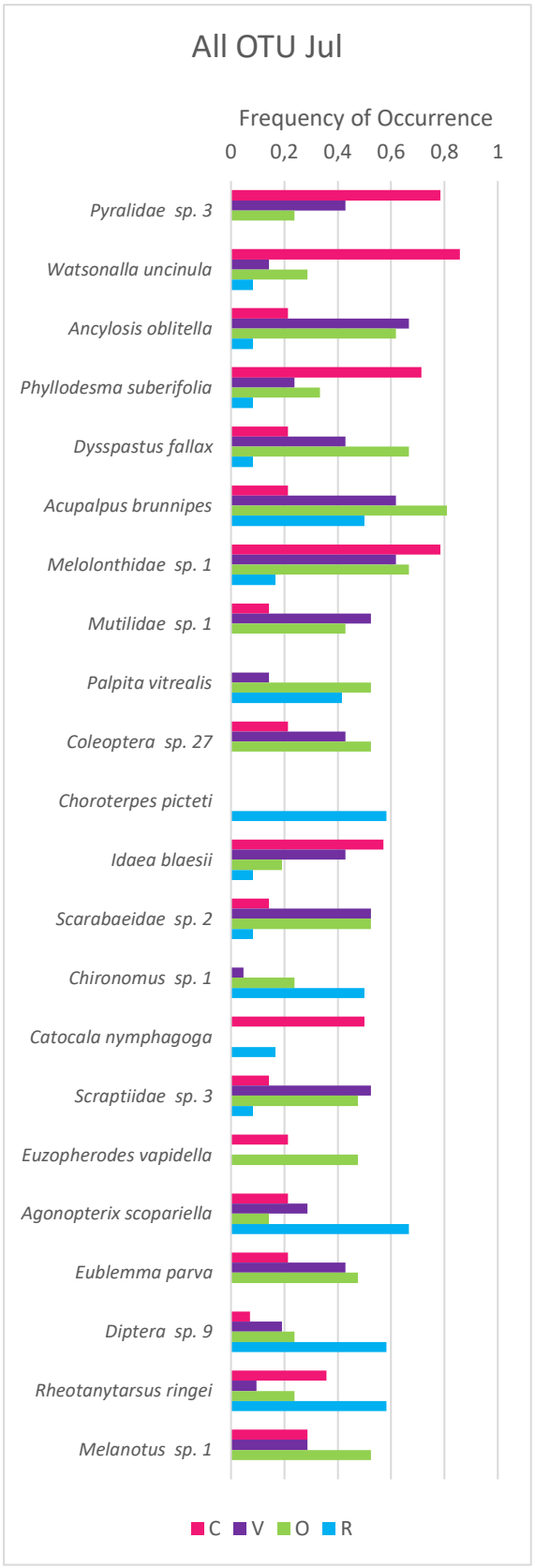
Figure 3-13: Principal coordinates analysis of Jaccard distances. Each point represents a sample and the corresponding composition community at order level. The samples in the top graph are defined in color and shape according to the habitat and season in which it was collected. Ellipses surround 95% of the points of each habitat/season.

The PermANOVA analysis confirmed the seasonal and habitat effect suggested by the PCoA representations (Table 3-4). Whether looking at all OTU's beta diversity, or just Lepidoptera, it is possible to observe that community composition varies with habitat and season. Moreover, as the interaction of the two variables was also significant, suggesting that these differences in composition among different habitats are dependent on the variation of the season and vice-versa.

Table 3-4: Permutational multivatiate analysis of variable (PermANOVA) with 999 permutations, based on the Jaccard similarity coefficient measure for presence-absence data.

		Df	*SS*	*MS*	*F*	*R2*	*p-value*	
All insect OTU's	Season	1	5.695	5.6953	15.9916	0.09662	0.001	***
	Habitat	3	3.148	1.0494	2.9466	0.05341	0.001	***
	Season*Habitat	3	1.664	0.5546	1.5571	0.02823	0.001	***
	Residuals	136	48.435	0.3561		0.82174		
	Total	143	58.943			1.00000		
Lepidoptera OTU's	Season	1	7.367	7.3671	22.1949	0.12933	0.001	***
	Habitat	3	2.894	0.9645	2.9058	0.05079	0.001	***
	Season*Habitat	3	1.563	0.5209	1.5692	0.02743	0.002	**
	Residuals	136	45.142	0.3319		0.79245		
	Total	143	56.965			1.00000		

The SIMPER analysis revealed which OTUs were contributing the most to the variation of the assemblages among habitats on each season. The frequency of occurrence of the OTUs that contributed the most to the variation of assemblages, either considering all insects or only Lepidoptera is represented on Figure 3-14. Among these, it can be highlighted that there was only one species unique of a single habitat, *Choroterpes picteti* in riparian zones. Overall most OTUs contributing to the difference among habitats simply seem to have different frequencies of occurrence. Some OTUs were dominant in certain habitats and occur in low proportions in others, such as *Watsonalla uncinula* which was present in most cork oak woodland areas and *Psychomyia* sp. 1 mostly in riparian zones. Moreover, it is of notice that these OTUs are different in the two seasons, either for insects in general or just Lepidoptera, which was expectable since the two seasons presented very different species compositions. Most OTUs with high frequency of occurrence in riparian zones were diptera, while cork oak woodlands showed a high frequency of Lepidoptera species.


All OTU Jul
Frequency of Occurrence
0
0,2
0,4
0,6
0,8
1
Pyralidae sp. 3
Watsonalla uncinula
Ancylosis oblitella
Phyllodesma suberifolia
Dysspastus fallax
Acupalpus brunnipes
Melolonthidae sp. 1
Mutilidae sp. 1
Palpita vitrealis
Coleoptera sp. 27
Choroterpes picteti
Idaea blaesii
Scarabaeidae sp. 2
Chironomus sp. 1
Catocala nymphagoga
Scraptiidae sp. 3
Euzopherodes vapidella
Agonopterix scopariella
Eublemma parva
Diptera sp. 9
Rheotanytarsus ringei
Melanotus sp. 1
C V O R

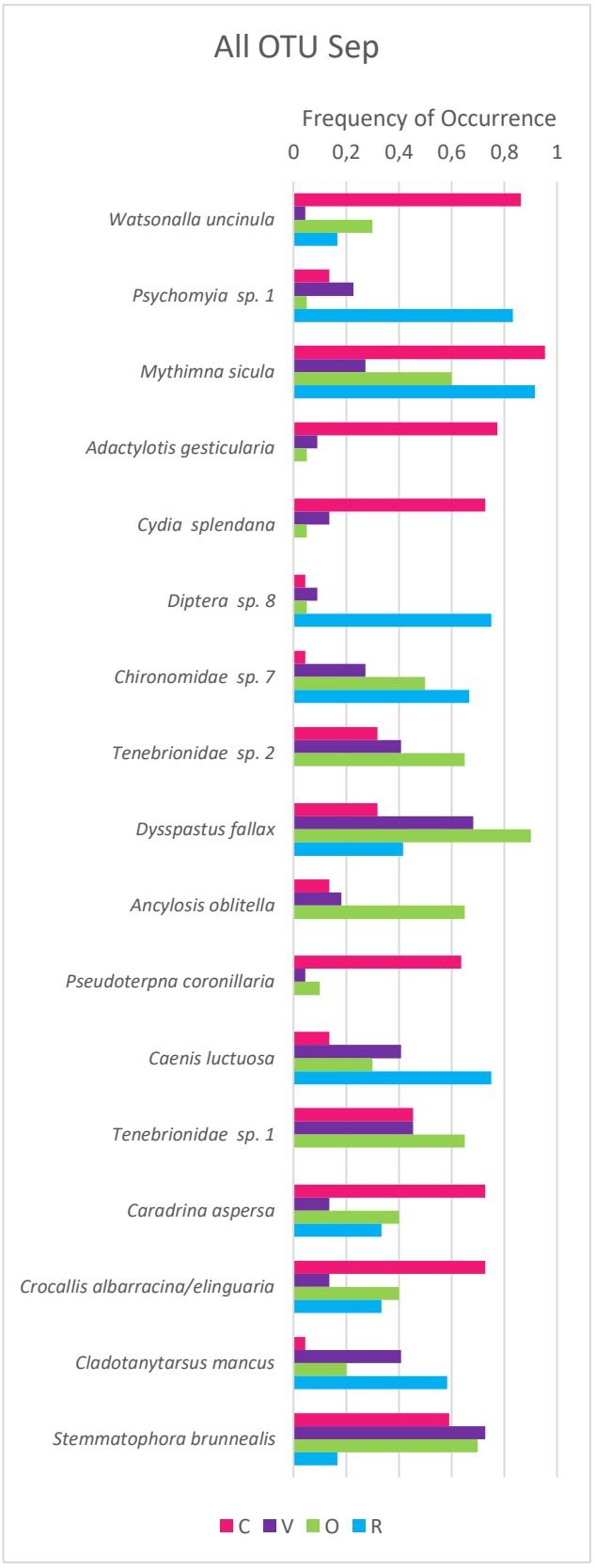

All OTU Sep
Frequency of Occurrence
0
0,2
0,4
0,6
0,8
1
Watsonalla uncinula
Psychomyia sp. 1
Mythimna sicula
Adactylotis gesticularia
Cydia splendana
Diptera sp. 8
Chironomidae sp. 7
Tenebrionidae sp. 2
Dysspastus fallax
Ancylosis oblitella
Pseudoterpna coronillaria
Caenis luctuosa
Tenebrionidae sp. 1
Caradrina aspersa
Crocallis albarracina/elinguaria
Cladotanytarsus mancus
Stemmatophora brunnealis
C V O R

Figure 3-14: SIMPER analysis. The two top plots represent the frequency of occurrence of the insect OTUs that contribute to about 10% of the variation between habitats of each season according to the results of the SIMPER analysis. The lower two plots represent Lepidoptera OTUs that contribute with about 20% for most of the differences recorded for each habitat.

3.5. Associations between traits and habitats

Mean wingspan was positively related with cork oak and riparian habitats, and negatively related with olive grove habitats. Also, mean wingspan was negatively associated with July, indicating that species recorded in September were generally larger (Figure 3-15). Polyphagia (diet) was negatively related with cork oak woodland habitats, indicating that they contain more specialist species (Figure 3-15). There was also a positive association between polyphagy and July; whereas the inverse happened with detritivore species, which were more likely to occur in September (Figure 3-15).

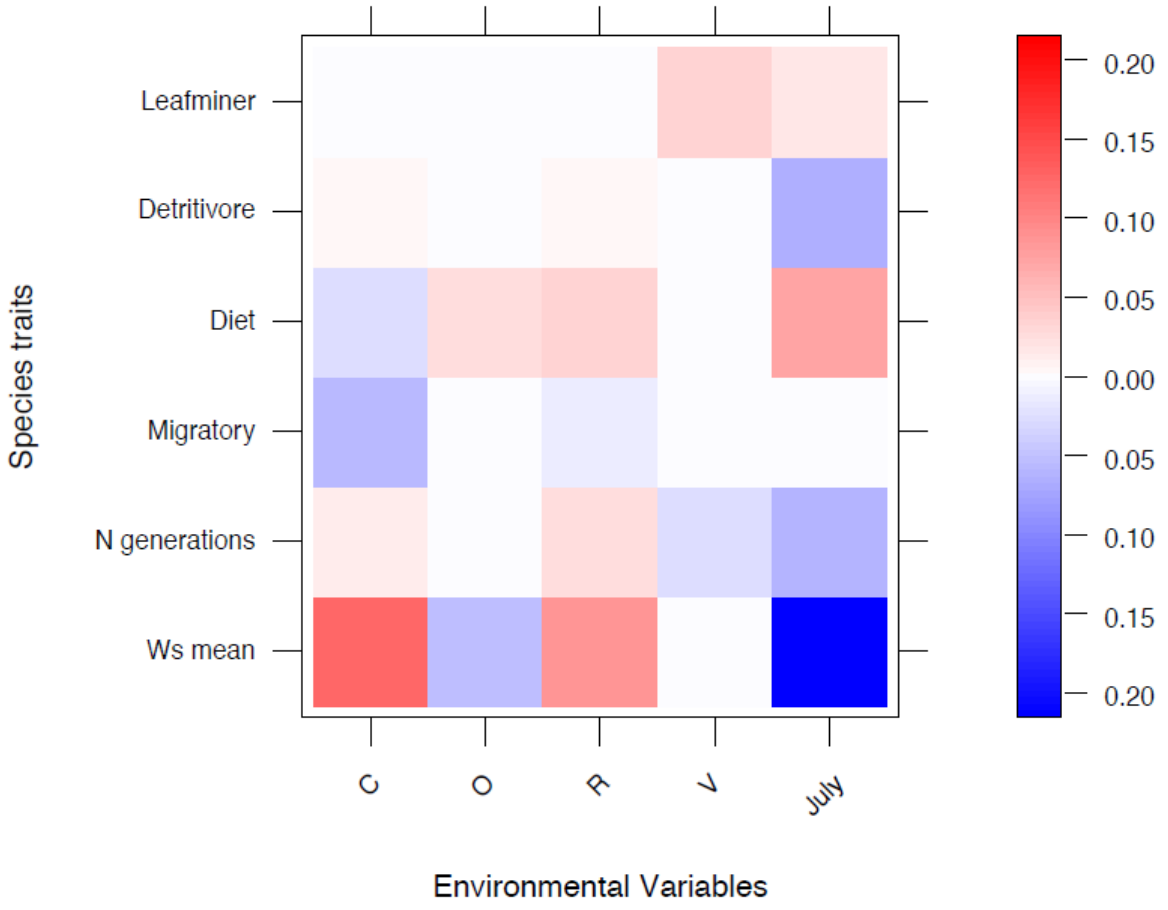


Figure 3-15: Visual representation of the relationship between species occurrence and their traits estimated with a binomial regression. Values that are larger in magnitude suggest stronger interactions.

4. Discussion

4.1. The importance of protocol testing and how the primer pair influences results

Previous studies like the ones conducted by Elbrecht and Leese (2017) and Yu *et al* (2012) have highlighted the importance of finding adequate primer pairs *i.e.* those that can ideally amplify all taxa in a sample, as essential to improve the use of DNA metabarcoding in ecological studies[53,63,95,119]. More and more studies have focused on evaluating primer bias and the proportion of undetected taxa, showing that primer choice influences the accuracy of identifications up to species level[95,119] and that variation in protocols reflect on the success of the whole process. However, few have experimentally compared the performance of different markers for the study of natural moth communities directly from the field. Here we support the importance of primer testing when passing from 'lab-made' communities to natural communities.

From four initial primer pairs, two were discarded after the first PCR and another one after sequencing the two leftover primer pairs, LCO and B2. Therefore, we end up choosing to proceed using only the best performing primer pair B2 which allowed to identify more species. Simultaneously, it produced cleaner data, easier and less time consuming to analyze, and thus allowed to optimize the proportion of species found in relation to the time it takes to filter and blast all sequences, improving overall cost-effectiveness. The COI barcoding primer pair BF2/BR2 has a high level of base degeneracy which probably helped to amplify as many insect taxa as possible[95]. Although it was designed for aquatic macroinvertebrate assessment[95] it performed remarkably well in the amplification of Lepidoptera species, providing an overall better cost-effectiveness. One of the advantages of NGS sequencing is that it enables to extract massive amounts of data from each sequenced sample. However, alignment time increases a lot with number of sequences[53], increasing the analysis' cost and difficulty by requiring high performance computers. We confirmed the importance of evaluating the primers performance prior to mass sequencing of samples using a relatively simple protocol, avoiding unnecessary risks and expenses that could result from 'blindly' sequencing all samples in a HiSeq run.

4.2. Validation of metabarcoding

The validation of metabarcoding's ability to recover the species contained in visually identified bulk samples was an important step to understand the efficiency and shortcomings of the technique. Overall, metabarcoding was able to detect most of the species that were also detected visually, although in some cases small-bodied species seemed to have gone undetected. In other cases, we were not able to reach species-level identification due to the lack of proper DNA references, confirming how crucial well-curated and complete reference databases are for metabarcoding studies. Nevertheless, most of the mismatches seemed to be related to differences in species identity, rather than to undetected or non-identified species. Although moth samples were identified by a specialist, this was rapidly done in the field, without proper set up of specimens and genitalia validation for each identification. The naked-eye visual identification of moths relies mostly on observation of the wing patterns and many of the specimens were not in ideal conditions for taxonomic identification, for instance due to wing scales worn off. This means that some of the visual species records might not be completely correct which justifies why in some cases a certain species was found using metabarcoding and a different one of the same genus was found in visual identification. Still, most species occurrence was detected with the two methods, and the PerMANOVA analysis detected no statistical differences in community composition between both methods, which means that metabarcoding was able to recover the species composition of bulk field samples. Nonetheless, ecological studies usually require many samples and rigorous

taxonomic validation of every specimen is simply not feasible, meaning that the use of metabarcoding can exponentially accelerate the speed and accuracy of insect ecological studies.

4.3. Extracting ecological information from molecular methods

The developed analysis pipeline performed well in converting NGS data into ecological interpretable information. First, we were able to verify that sampling effort was not enough to capture total species richness (alpha diversity). Still, differences were detected in species richness and species composition – alpha and beta diversity – among different habitats across the two seasons. Alpha and beta diversity metrics are useful to measure diversity in a broad and efficient way[63], allowing comparisons across time and space. Although sampling effort was below desirable, the detailed taxonomic resolution achieved with NGS sequence analysis still seems to compensate in producing results prone to undergo ecological analysis. In this case, it allowed to relate species occurrence to their traits by using an adequate statistical framework, but the produced data can be used in many other kinds of analysis.

4.3.1. OTU richness

Overall, metabarcoding was able to recover an outstanding number of arthropod OTUs, with 1130 OTUs found in the Tua Valley, of which over 400 were moths. The analysis of the cumulative curves and expected Hill numbers shows that the obtained OTU richness values do not reach the asymptote of the curves as it would be desirable. That is, the number of samples is not enough to fully describe the diversity of each habitat, hence subtle patterns caused by less common species might have been missed. Still, a high number of OTUs was identified, and the analysis of the produced data provides an overview of how diversity varied over seasons and habitats.

We were able to observe that OTU richness varied with habitat with the highest average richness per sample recorded in olive groves in July. Olive groves are a kind of orchard, and it is not new that maintaining trees in the landscape generally relates to higher diversity[33,120,121]. Finding such high levels of richness in olive groves suggests that further studies could be carried on, exploring the possibility of designating certain zones within the PNRVT as High Nature Value Farmlands (HNVF), as lately there has been an attempt to define, map and conserve these important areas for biodiversity[47]. Thus, DNA metabarcoding could be used to monitor diversity in these areas and evaluate the effect of their protection status.

The lowest richness values were recorded in vineyards (both seasons) and cork oak forests (July). A low richness in vineyards is not surprising given the human intervention that these areas are subjected to. Besides having none or almost no trees, most vineyards are tilled. Soil tillage eliminates not only many species that rely on soil for part of their life cycle, but also a considerable part of understory plant diversity as herbaceous and shrubs, as well as their associated insect diversity[122–124]. The even lower average richness recorded in September could relate to an even higher human impact during the sampling period as the dry and hot summer anticipated the start of the harvesting season. While sampling we noticed some of the vines had already been harvested and there was higher movement of people in general.

The opposite happened in cork oak woodlands where the record of low richness in July was followed by a sudden increase in September. This is rather odd since there were no significant changes on riparian zones and olive groves with season in overall insect richness. In fact, the levels of average OTU richness in September in cork oak woodlands grew to become similar to the levels of richness in riparian zones

and olive groves. This was probably caused by the great increase in Lepidoptera richness recorded for this habitat towards the end of summer.

Riparian zones presented above average diversity in both months, suggesting that these areas hold unique features that enable a high diversity of species to prosper, such as high plant species diversity, including larger and older trees. Riparian zones are ecotones and by being the interface of aquatic and terrestrial environments[125] they albergate species that are typical from one of the patches, as well as some that are characteristic of riparian zones themselves. Moreover, their unique dynamics are related to water currents, such as flood regimes and water availability even in summer[126], that in Mediterranean regions are usually characterized by a temporary water flow, thus leading to seasonal changes in aquatic and terrestrial plant communities and probably insects.

When considering only Lepidoptera OTUs, although the total number of species found in each habitat is similar, the average number of species per sample was higher in olive groves and cork oak woodlands than in the remaining habitats. This diversity is probably boosted by the presence of trees[127], namely cork oaks and olives trees, which are typically Mediterranean and have existed there for millenia[128]. As lepidoptera are mostly herbivore, many species that rely on these trees, or associated plant species, might have established themselves along time. Old cork and olive trees have a variety of associated lichens on which some moth species feed[129]. Besides cork oaks and olive trees, these systems present high plant diversity at understory levels[128,130] such as bushes and herbaceous plants that determine the presence of many other moth species. This type of agroforesty habitats are many times suitable for both species that prefer wider, open areas along with some that are typically found in forests and rely on the presence of trees[131].

The lower diversity of moths observed in vineyards is not surprising as this was the pattern for all insects. However, there was also lower diversity of Lepidoptera in riparian zones which previously presented high values for all insects. This suggests that most species found in these habitats belong to other orders and that result was confirmed by the frequency of occurrence analysis with the most frequent species always being Diptera.

Seasonal effect was more marked on Lepidoptera communities where all habitats except riparian zones presented higher diversity towards the end of the summer, including vineyards that had lower overall insect richness in this season. Moreover, whereas in July only olive groves presented significantly higher richness than vineyards and riparian zones, in September all habitat combinations revealed significant richness differences, except vineyards and riparian zones that presented the lowest values. However, this pattern of higher richness of Lepidoptera in late summer does not match what has been found in other studies in central Europe, where the highest number of species is found during the warmest nights, usually in July[113]. This could either be a local pattern of the region, or a result of an atypical year with warmer nights later in the summer and/or other unmeasured variables, as well as stochastic events. Temperature in particular is well known to play a role in insect metabolism with higher temperatures leading to higher insect activity. This has already been described for Lepidoptera, with positive correlations between temperature and species richness[78,132]. Temperatures were abnormally high during all summer, which was also particularly dry. Furthermore, many other weather-related factors such as wind strength and direction are likely to affect low mass flying species, potentially transported by wind currents, even across long distances. However, these are also more unpredictable and harder to model. Still, it would be important to repeat the sampling in other years and check if this pattern is a result of year effect or not.

4.3.2. Taxonomic distinctiveness

Looking at diversity in different ways may provide different ecological insights. Unlike OTU richness, the taxonomic distinctiveness index was higher in July when considering all taxa. Therefore, those samples contain species taxonomically further apart, which also differ in morphology and other characteristics. Taxonomic distinctiveness differed according to habitat, with riparian zones presenting the highest values for this metric. These areas had the highest taxonomic distances within samples, suggesting about their high natural value. The trapping method is also specific for moths and they do not seem to occur in such high proportions in riparian zones according to the richness and frequency of occurrence results. Thus, many of the specimens gathered in theses zones were other arthropods, likely separated by larger taxonomic distances.

Calculating the taxonomic distinctiveness index among samples considering only OTUs that belong to the same order might explain the lack of significant differences in Lepidoptera assemblages. As all species are taxonomically related the index is likely to produce values that are more similar among samples. Lepidoptera taxonomic distinctiveness differed very little between habitats (low significance) and contrarily to insect richness it was similar across seasons. This probably means that in September the communities become dominated by Lepidoptera, thus raising their average OTU richness but decreasing the overall taxonomic distinctiveness.

4.3.3. OTU composition

Clear separation in species composition according to season was expected as most insects, including moths, emerge as adults at a strict period of the year. When looking at each month separately we could see some habitat structuring. However, habitats that differed from others also change differently with season, which indicates that the singularity of a certain habitat in terms of species assemblages relies on the sampling period. As above-mentioned, we should consider that different weather conditions are likely to produce different results as it is known that temperature for instance plays a role in controlling most insects' activity. This should be accounted for in all invertebrate studies as sampling period may influence the results, thus affecting its translation to management actions such as defining protected areas or protection periods. Likewise, Lepidoptera communities appear to have followed similar composition patterns to other insect communities, presenting strong compositional differences between seasons.

When looking at the interaction of the variables (habitat and season) we can see that the most distinct assemblages are those occurring in riparian zones in July. It was possible to verify that the effect of habitat is dependent on season for all insect OTUs and Lepidoptera. We could also analyze which species are more likely to be responsible for the difference between habitats on both seasons and relate it with their frequency of occurrence. Usually, these are species that tend to appear very often in one or two habitats and barely show up in the others. There were almost no unique species to a certain habitat, although in the only case it happened was in riparian zones. It would be interesting to compare these results with other years as this can be a way of identifying habitat preferences for certain species. Moreover, by looking at the frequency of occurrence we can infer that species that tend to occur in a broader range of habitats are more likely to have no preference (generalists), namely those that occur in all habitats in this study. For example, two of the top three most frequent species in vineyards and olive groves are the same, pointing out the similarity between these areas again. On the other hand, the three most frequent species in cork oak woodlands never appear on the top three of the other habitats and are all Lepidoptera species, whereas in riparian zones the most frequent species are all Diptera. The presence of a different Lepidoptera species when we look at all habitats together highlights the importance of

scale and how patterns may differ at local and landscape level. Nevertheless, it is important to keep in mind that "pure" control plots are hard to define for any of these habitats in such a fragmented landscape, as border effects will enable species characteristic of certain habitats to wander around in the surrounding ones (Figure 4-1). This can be a reason why among the OTUs that are shaping the differences between samples revealed by the SIMPER analysis, we can find few that are unique of only one habitat and most of them have only differences in frequency of occurrence. Using such a combination of tools can be helpful for when trying to research the biology and ecology of a species of interest, as knowing habitat preferences is a key to keep track of species distributions.

Figure 4-1: The fragmented landscape of PNRVT which resembles most of the Trás-os-Montes area. Border of a vineyard patch with an olive grove patch, surrounded by scrub-like land.

4.4. Into functional ecology through trait analysis

For a more complete community ecology analysis, it becomes evermore interesting to investigate functional traits rather than just looking at the species occurrences in certain area. Understanding ecosystem functioning goes beyond knowing local species pools and integrative studies may be a key to better understand species relations such as those of co-occurrence, competition, and other ecological processes. Unfortunately, because most insect communities are understudied, and moths are not an exception, the information on their traits is dispersed in several sources. In Portugal there were a total of 2657 species of described moth species in 2016, belonging to 76 different families[87]. In other words, insect taxonomy is a nightmare when it comes to gathering all the desired information in a single document. Each family is often entitled to a book of their own and these books are often expensive and only available in libraries many times far apart from ones working place. This reflects the lack of open, updated and centralized access to taxonomic information on moths, which extends to insects in general, and highlights the need to make this information accessible.

At a micro scale, insect communities can be quite different in short geographical distances in such way that on the same sampling evening, different species can be gathered in traps that are less than 1km away[133]. For example, the weather differs between valleys which is enough to find or not a certain species. Slight changes in temperature and humidity, together with soil, and other environmental variables boost the appearance of different plant, insect and other species whose ocurrence is dynamic[132]. This results in patchily, hard to assess and ever-changing distribution ranges that require continuous studies. In the Iberian Peninsula this information is particularly scarce. Traits, as the maximum number of generations per year, vary a lot with the availability of host plants that can be different between different geographic areas. Simultaneously, insects affect larger scale ecosystem processes as they participate in decomposition, pollination or they may even act as devastating plagues[23]. It becomes obvious why it is so challenging to study such communities.

Still, by merging information from over twenty different online open sources, it was possible to collect several traits of the occurring species to perform the fourthth corner analysis. The occurrence of larger moths in both riparian and cork oak areas indicates that these naturalized habitats hold the conditions to host larger species, which usually require more food. Leaf miners are known to cause severe damage to many crop species and their occurrence was related to vineyard habitats. Detritivores species tend to show up more in September, which is the end of the summer and most of the vegetation is dry. Species with more generations/year occurred more frequently in September. It is known that July is the flight peak of many species, probably including many of the single generation species, thus species appearing in September are more likely to be those with several generations per year. Migratory behavior related negatively with cork oak and riparian zones. The negative relation between presence of migratory moths and cork oak habitats and riparian zones is a pattern that deserves further research. Cork oak trees are mostly grown in the Mediterranean basin[128], thus species from elsewhere are unlikely to rely on them. These habitats might be richer in other local species that by occupying most ecological niches reduce the availability of natural resources as food or habitat. The presence of plant species that are uncommon in other areas can perhaps prevent colonization by migratory wanderers. This can also be a reason for why specialist feeding habits only related to cork oak woodland which also suggests good quality habitat, as these species are usually more demanding. By feeding on less host plants, they need to assure they exist in enough amounts to ensure survival and successful reproduction. Combining insect data with local plant diversity variables could be a way to better understand these patterns as insect and plant diversity are deeply connected.

5. Conclusions

The results of this study indicate that it is possible to detect most of the species present in bulk samples using DNA metabarcoding. We successfully used this method to identify arthropod communities, focusing part of the study on Lepidoptera for practical reasons but also due to their important ecological functions. Although the method does not provide abundance data and it was designed to target Lepidoptera, its high taxonomic resolution allowed to identify differences in richness and composition of species assemblages between different habitats, whether for all arthropods and only the target order. Although sample size was not enough to fully describe the diversity in each habitat, we can still assume that most common species were detected. This would easily be overcome by collecting more samples in the following studies and would almost certainly provide a more accurate snapshot of local species diversity.

Additionally, the sampling effort was unequal between habitats 22 points of each habitat type were sampled, except for riparian zones that are represented in a maximum of 13 sampling points; and seasons and the points in cork oak forest are distributed in only 8 localities, having thus a more clustered distribution within the landscape. Still, it was possible to verify that species occurrence relates to species traits that as their migratory behavior or morphological, phenological and dietary traits. These results demonstrate that such kind of monitoring study has potential to reveal some deeper insights into the relationship between species occurrence and their adaptations. Although it was possible to get a clue on trait selection under different habitat conditions. As species traits relate to certain ecological functions, the method has potential be used to predict which species are more likely to prosper in different habitats, enabling for example to track potential plagues or invasive species. Additionally, this kind of data has potential to be used in sampling in large geographical areas and to be analyzed from a geospatial perspective.

Many parts of the metabarcoding process can be improved, and most potential shortcomings of this relatively new method should be considered as opportunities to invest and improve the methodology. First, it is necessary to develop better universal PCR primers, to test them and to compile this information in an accessible way to all users of this technique. Even more desirable would be to cut dependency on PCR by directly sequencing DNA extraction products which could allow to overcome the fact that some orders are not detected: For example, Hymenoptera tend to pass undetected when using fragments of the COI gene. Although there has been a growing effort to improve the quality and assessibility of reference data bases, there is still a large effort to be made to complete them and to link databases as BOLD to taxonomic assignment programs. This can be boosted by the development of better barcoding pipelines, namely for chimera detection and taxonomy assignment protocols as well as with the rapid development of new software and hardware. Ultimately, one of the key goals would be to find a way to obtain relevant species abundance data from sequencing output.

Furthermore, many steps are to take in combining DNA metabarcoding with ecological research. Our results point out that in this context it would be interesting to further investigate the importance of the agroforestry mosaic for biodiversity in Mediterranean habitats. Monitoring for longer periods may enable to check diversity patterns throughout time (e.g. from year to year) and to identify areas where diversity is increasing or decreasing. It would also allow to chek if new comers/detections and the species that have been registered before share the same traits or if they differ. Metabarcoding can be used to measure and compare conservation measures in several kinds of situations, in this case as a dam was recently built and the area was protected, it can be used to understand the evolution of species assemblages. Moreover, metabarcoding produces data that is suitable to be combined with Earth observation data and produce geospatial analysis, and protocols on the matter are starting to appear[134].

For example, it would be interesting to use geographic information systems (GIS) to relate landscape variables, such as vegetaion cover or land use with species composition, instead of characterizing each sampling point as a simple habitat category. As the technique is used to identify whole communities it can be used as any other species identification method, the possibilities are endless for ecological research and the conditions are now set for metabarcoding to thrive in this field.

References

1. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294,** 804–808 (2001).
2. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75,** 3–35 (2005).
3. Rapport, D. J., Costanza, R. & McMichael, A. J. Assessing ecosystem health. *Trends Ecol. Evol.* **13,** 397–402 (1998).
4. Peterson, G., Allen, C. R. & Holling, C. S. Ecological Resilience, Biodiversity, and Scale. *Ecosystems* **1,** 6–18 (1998).
5. Bullock, J. M., Aronson, J., Newton, A. C., Pywell, R. F. & Rey-Benayas, J. M. Restoration of ecosystem services and biodiversity: Conflicts and opportunities. *Trends Ecol. Evol.* **26,** 541–549 (2011).
6. Pereira, H. M. *et al.* Scenarios for Global Biodiversity in the 21st Century. *Science* **330,** 1496–1501 (2010).
7. Hooper, D. U. *et al.* A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486,** 105–108 (2012).
8. Millennium Ecosystem Assessment Board. *Ecosystems and Human well-being: Health Synthesis*. *World Health Organization* (2005). doi:10.1196/annals.1439.003
9. Noonan-Mooney, K. & Gibb, C. How are people affecting biodiversity? in *The Youth Guide to Biodiversity* (eds. Christine, G., Neil, P. & Reuben, S.) 13–21 (FAO, 2013).
10. Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344,** 1246752 (2014).
11. Crutzen, P. & Stoermer, E. The 'anthropocene'. in *Earth System Science in the Anthropocene* **41,** 13–18 (Springer, Berlin, Heidelberg, 2006).
12. Dirzo, R. *et al.* Defaunation in the Antrhopocene. *Science* **345,** 401–406 (2014).
13. Pievani, T. The sixth mass extinction: Anthropocene and the human impact on biodiversity. *Rend. Lincei* **25,** 85–93 (2014).
14. Laurance, W. F. & Edwards, D. P. The search for unknown biodiversity. *Proc. Natl. Acad. Sci.* **108,** 12971–12972 (2011).
15. Tedesco, P. A. *et al.* Estimating how many undescribed species have gone extinct. *Conserv. Biol.* **28,** 1360–1370 (2014).
16. Barnosky, A. D. *et al.* Has the Earth's sixth mass extinction already arrived? *Nature* **471,** 51–57 (2011).
17. Stork, N. E. How many species are there? *Biodivers. Conserv.* **2,** 215–232 (1993).
18. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How many species are there on earth and in the ocean? *PLoS Biol.* **9,** e1001127 (2011).
19. Collen, B., Böhm, M., Kemp, R. & Baillie, J. E. M. *Status and trends of the world's invertebrates*. (Zoological Society of London, 2010).
20. UNEP/CBD/COP/6/20. *Decisions adopted by the conferene of the parties to the Convention on Biological Diversity at its sixth meeting*. 70–347 (2002).
21. Butchart, S. H. M. *et al.* Global biodiversity: Indicators of recent declines. *Science* **328,** 1164–1168 (2010).
22. Secretariat of the Convention on Biological Diversity. Global Biodiversity Outlook 3. 1–94 (2010). doi:10.7547/0003-0538-104.3.A1
23. Foottit, R. G. & Adler, P. H. *Insect biodiversity: science and society*. (Wiley-blackwell, 2009). doi:10.1002/9781444308211
24. Losey, J. E. & Vaughan, M. The economic value of ecological services provided by insects. *Bioscience* **56,** 311–323 (2006).
25. IUCN. The IUCN Red List of Threatened Species. (2018). Available at: http://www.iucnredlist.org.
26. Hoffmann, M. *et al.* The impact of conservation on the status of the world's vertebrates. *Science* **330,** 1503–1509 (2010).
27. Groom, M. J. Threats to Biodiversity. in *Principles of Conservation Biology* **3,** 63–109 (1989).

28. Tilman, D. *et al.* Forecasting Agriculturally Driven Environmental Change. *Am. Assoc. fo rthe Adv. Sci.* **292,** 281–284 (2001).
29. Foley, J. A. *et al.* Global Consequences of Land Use. *Science* **309,** 570–574 (2005).
30. Millennium Ecosystem Assessment. *Ecosystems and human well-being: Synthesis*. *World Resources Institute* **5,** (Island Press, 2005).
31. Bengtsson, J., Ahnström, J. & Weibull, A. C. The effects of organic agriculture on biodiversity and abundance: A meta-analysis. *J. Appl. Ecol.* **42,** 261–269 (2005).
32. Kotschi, J. & Müller-sämann, K. *The Role of Organic Agriculture in Mitigating Climate Change - a Scoping Study*. *International Federation of Organic Agriculture Movements* (2004).
33. Altieri, M. A. The ecological role of biodiversity in agroecosystems. *Agric. Ecosyst. Environ.* **74,** 19–31 (1999).
34. Powlson, D. S. *et al.* Limited potential of no-till agriculture for climate change mitigation. *Nat. Clim. Chang.* **4,** 678–683 (2014).
35. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418,** 671–677 (2002).
36. Benton, T. G., Vickery, J. A. & Wilson, J. D. Farmland biodiversity: Is habitat heterogeneity the key? *Trends Ecol. Evol.* **18,** 182–188 (2003).
37. Perfecto, I., Vandermeer, J., Hanson, P. & Cartín, V. Arthropod biodiversity loss and the transformation of a tropical agro-ecosystem. *Biodivers. Conserv.* **6,** 935–945 (1997).
38. Luoto, M., Rekolainen, S., Aakkula, J. & Pykälä, J. Loss of Plant Species Richness and Habitat Connectivity in Grasslands Associated with Agricultural Change in Finland. *AMBIO A J. Hum. Environ.* **32,** 447–452 (2003).
39. Chamberlain, D. E. & Fuller, R. J. Local extinctions and changes in species richness of lowland farmland birds in England and Wales in relation to recent changes in agricultural land-use. *Agric. Ecosyst. Environ.* **78,** 1–17 (2000).
40. McLaughlin, A. & Mineau, P. The impact of agricultural practices on biodiversity. *Agric. Ecosyst. Environ.* **55,** 201–212 (1995).
41. Haniotakis, G. E. *Olive pest control: Present status and prospects*. *IOBC/WPRS Bulletin* **28,** (2005).
42. Subbarao, G. *et al.* Scope and strategies for regulation of nitrification in agricultural systems - Challenges and opportunities. *CRC. Crit. Rev. Plant Sci.* **25,** 303–335 (2006).
43. Grasty, S. Agriculture and Climate Change. *TDRI Q. Rev.* **14,** 12–16 (1999).
44. Angelsen, a & Kaimowitz, D. Rethinking the causes of deforestation: lessons from economic models. *World Bank Res. Obs.* **14,** 73–98 (1999).
45. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320,** 1444–1449 (2008).
46. Lal, R. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* **304,** 1623–1627 (2004).
47. Lomba, A. *et al.* Mapping and monitoring High Nature Value farmlands: Challenges in European landscapes. *J. Environ. Manage.* **143,** 140–150 (2014).
48. Bignal, E. M. & McCracken, D. I. Low-Intensity Farming Systems in the Conservation of the Countryside. *J. Appl. Ecol.* **33,** 413–424 (1996).
49. Pienkowski, M. The nature conservation value of low-intensity farming systems. *EFNCP Occas. Publ. Number* **17,** 82–87 (1998).
50. Bignal, E. M. & McCracken, D. I. The nature conservation value of European traditional farming systems. *Environ. Rev.* **8,** 149–171 (2000).
51. Gallai, N., Salles, J. M., Settele, J. & Vaissière, B. E. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol. Econ.* **68,** 810–821 (2009).
52. Sharma, S., Kooner, R. & Arora, R. Insect Pests and Crop Losses. in *Breeding Insect Resistant Crops for Sustainable Agriculture* 45–66 (Springer, 2017). doi:10.1007/978-981-10-6056-4_2
53. Coissac, E., Riaz, T. & Puillandre, N. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* **21,** 1834–1847 (2012).
54. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270,** 313–321 (2003).
55. Fišer, C., Robinson, C. T. & Malard, F. Cryptic species as a window into the paradigm shift of

the species concept. *Mol. Ecol.* **27,** 613–635 (2018).
56. Ratnasingham, S. & Hebert, P. D. N. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Notes* **7,** 355–364 (2007).
57. Ji, Y. *et al.* Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* **16,** 1245–1257 (2013).
58. Chacón, I. A., Janzen, D. H., Hallwachs, W., Sullivan, J. B. & Hajibabaei, M. Cryptic species within cryptic moths: New species of Dunama Schaus (Notodontidae, Nystaleinae) in Costa Rica. *Zookeys* **264,** 11–45 (2013).
59. Janzen, D. H. *et al.* Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol. Ecol. Resour.* **9,** 1–26 (2009).
60. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. & Hebert, P. D. N. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci.* **103,** 968–971 (2006).
61. Janzen, D. H. *et al.* Reading the complex skipper butterfly fauna of one tropical place. *PLoS One* **6,** 1–15 (2011).
62. Evans, D. M., Kitson, J. J. N., Lunt, D. H., Straw, N. A. & Pocock, M. J. O. Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Funct. Ecol.* **30,** 1904–1916 (2016).
63. Yu, D. W. *et al.* Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* **3,** 613–623 (2012).
64. Gibson, J. *et al.* Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proc. Natl. Acad. Sci.* **111,** 8007–8012 (2014).
65. Shokralla, S., Spall, J. L., Gibson, J. F. & Hajibabaei, M. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* **21,** 1794–1805 (2012).
66. Andersen, K. *et al.* Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* **21,** 1966–1979 (2012).
67. Schmidt, P.-A. *et al.* Illumina metabarcoding of a soil fungal community. *Soil Biol. Biochem.* **65,** 128–132 (2013).
68. Jerde, C. L., Mahon, A. R., Chadderton, W. L. & Lodge, D. M. 'Sight-unseen' detection of rare aquatic species using environmental DNA. *Conserv. Lett.* **4,** 150–157 (2011).
69. Ficetola, G. F., Miaud, C., Pompanon, F. & Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **4,** 423–425 (2008).
70. Mata, V. A. *et al.* Female dietary bias towards large migratory moths in the European free-tailed bat (Tadarida teniotis). *Biol. Lett.* **12,** 20150988 (2016).
71. Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (Prionailurus bengalensis) in Pakistan. *Mol. Ecol.* **21,** 1951–1965 (2012).
72. Macías-Hernández, N. *et al.* Molecular gut content analysis of different spider body parts. *PLoS One* **13,** e0196589 (2018).
73. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21,** 2045–2050 (2012).
74. Agrawal, A. A. *et al.* Filling key gaps in population and community ecology. *Front. Ecol. Environ.* **5,** 145–152 (2007).
75. Calabrese, J. M., Certain, G., Kraan, C. & Dormann, C. F. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* **23,** 99–112 (2014).
76. Guisan, A. & Rahbek, C. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J. Biogeogr.* **38,** 1433–1444 (2011).
77. Logue, J. B., Mouquet, N., Peter, H. & Hillebrand, H. Empirical approaches to metacommunities: A review and comparison with theory. *Trends Ecol. Evol.* **26,** 482–491 (2011).
78. Wisz, M. S. *et al.* The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.* **88,** 15–30 (2013).
79. Gonzalez, A. & Loreau, M. The Causes and Consequences of Compensatory Dynamics in Ecological Communities. *Annu. Rev. Ecol. Evol. Syst.* **40,** 393–414 (2009).

80. Cardillo, M. & Meijaard, E. Are comparative studies of extinction risk useful for conservation? *Trends Ecol. Evol.* **27,** 167–171 (2012).
81. McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21,** 178–185 (2006).
82. Stuart-Smith, R. D. *et al.* Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501,** 539–542 (2013).
83. Corley, M. F. V. & Buchner, P. Depressaria villosae sp. nov., a new species from Portugal, Spain and Greece (Depressariidae). *Entomol. Rec. J. Var.* **130,** 105–111 (2018).
84. Corley, M. F. V., Merckx, T., Marabuto, E. M., Arnscheid, W. & Maravalhas, E. New and interesting Portuguese Lepidoptera records from 2012 (Insecta : Lepidoptera). *Shil. Rev. Lepidopterol.* **41,** 449–477 (2013).
85. Corley, M. F. V., Marabuto, E., Maravalhas, E., Pires, P. & Cardoso, J. P. New and interesting Portuguese Lepidoptera records from 2009 (Insecta: Lepidoptera). *Shil. Rev. Lepidopterol.* **39,** 15–35 (2011).
86. Corley, M. F. V. *et al.* New and interesting Portuguese Lepidoptera records from 2015 (Insecta: Lepidoptera). *Shil. Rev. Lepidopterol.* **44,** 615–643 (2016).
87. Corley, M. F. V. *et al.* New and interesting Portuguese Lepidoptera records from 2016 (Insecta: Lepidoptera). *Shil. Rev. Lepidopterol.* **46,** 33–56 (2018).
88. Corley, M. F. V. *et al.* New and interesting Portuguese Lepidoptera records from 2014 (Insecta: Lepidoptera). *Shil. Rev. Lepidopterol.* **43,** 583–613 (2015).
89. New, T. R. Moths (Insecta: Lepidoptera) and conservation: Background and perspective. *J. Insect Conserv.* **8,** 79–94 (2004).
90. Oerke, E. C. Crop losses to pests. *J. Agric. Sci.* **144,** 31–43 (2006).
91. UNESCO World Heritage Center. UNESCO World Heritage List. (2001). Available at: http://whc.unesco.org/en/list/1046.
92. Diário da República. *Regulamento do Parque Natural Regional do Vale do Tua*. 2–4 (2013).
93. Promoção dos serviços de ecossistemas no Parque Natural Regional do Vale do Tua: Controlo de Pragas Agrícolas e Florestais por Morcegos. (2016).
94. Elbrecht, V., Peinert, B. & Leese, F. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecol. Evol.* **7,** 6918–6926 (2017).
95. Elbrecht, V. & Leese, F. Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Front. Environ. Sci.* **5,** 1–11 (2017).
96. Folmer, O., BLACK, M., HOEH, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3,** 294–299 (1994).
97. Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C. & Jones, G. Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Mol. Ecol. Resour.* **11,** 236–244 (2011).
98. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* **5,** 9687 (2015).
99. Boyer, F. *et al.* OBITOOLS: a unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16,** 176–182 (2016).
100. Kimber, I. UKmoths. (2018). Available at: http://www.ukmoths.org.uk.
101. Prichard, T. The Moths of Suffolk. (2018). Available at: http://www.suffolkmoths.org.uk/mos/index.shtml.
102. Pitkin, B., Ellis, W., Plant, C. & Edmunds, R. The leaf and stem mines of British flies and other insects. (2018). Available at: http://www.ukflymines.co.uk/.
103. Mazzei, P., Morel, D. & Panfili, R. Moths and Butterflies of Europe and North Africa. (2018). Available at: http://www.leps.it/.
104. Hoddé, C. & Mothiron, P. Lépi'net. (2018). Available at: http://www.lepinet.fr/.
105. Unger, M. Moths and butterflies of Sweden. (2018). Available at: http://www.lepidoptera.se/.
106. Flemish Entomological Society. Phegea. (2018). Available at: http://www.phegea.org/.
107. de Jong, Y. *et al.* Fauna Europaea. (2018). Available at: https://fauna-eu.org/.
108. Ratnasingham, S. & Hebert, P. D. N. BOLD - Barcode of Life Data System. (2018). Available at: http://www.boldsystems.org/.
109. Guyonne, A. Papillon de Poitou-Charentes. (2018).

110. Lepiforum e. V. Lepiforum. (2018). Available at: http://www.lepiforum.de.
111. Jonko, C. Lepidoptera Mundi. (2018). Available at: http://www.lepidoptera.eu.
112. Jonko, C. Lepidoptera EU. (2018). Available at: http://www.old.lepidoptera.eu.
113. Museum, T. N. H. HOSTS - a Database of the World's Lepidopteran Hostplants. (2018). Available at: http://www.nhm.ac.uk/our-science/data/hostplants/.
114. Ziegler, H. Euroleps. (2018). Available at: http://www.euroleps.ch/.
115. Ellis, W. N. Plant Parasites of Europe - leafminer, galls and fungi. (2018). Available at: https://bladmineerders.nl/.
116. Muus, T. S. T. & Corver, S. C. Microlepidoptera.nl, atlas of the small butterflies in the Netherlands. (2018). Available at: http://www.microvlinders.nl/.
117. Clarke, K. R. & Warwick, R. M. A taxonomic distinctness index and its statistical properties. *J. Appl. Ecol.* **35,** 523–531 (1998).
118. Warton, D. I. *et al.* So Many Variables: Joint Modeling in Community Ecology. *Trends Ecol. Evol.* **30,** 766–779 (2015).
119. Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J. & Elbrecht, V. Revealing the Complexities of Metabarcoding with a Diverse Arthropod Mock Community. *bioRxiv* 433607 (2018). doi:https://doi.org/10.1101/433607
120. Bhagwat, S. A., Willis, K. J., Birks, H. J. B. & Whittaker, R. J. Agroforestry: a refuge for tropical biodiversity? *Trends Ecol. Evol.* **23,** 261–267 (2008).
121. Borron, S. Building resilience for an unpredictable future: how organic agriculture can help farmers adapt to climate change. *Food and Agriculture Organization of the United Nations* (2006).
122. Wilson, J. D., Morris, A. J., Arroyo, B. E., Clark, S. C. & Bradbury, R. B. A review of the abundance and diversity of invertebrate and plant foods of granivorous birds in northern Europe in relation to agricultural change. *Agric. Ecosyst. Environ.* **75,** 13–30 (1999).
123. Matson, P. A., Parton, W. J., Power, A. G. & Swift, M. J. Agricultural intensification and ecosystem properties. *Science* **277,** 504–509 (1997).
124. Thorbek, P. & Bilde, T. Reduced numbers of generalist arthropod predators after crop management. *J. Appl. Ecol.* **41,** 526–538 (2004).
125. Gregory, S. V., Swanson, F. J., McKee, W. A. & Cummins, K. W. An Ecosystem Perspective of Riparian Zones. *Bioscience* **41,** 540–551 (1991).
126. Naiman, R. J. & Decamps, H. The Ecology of Interfaces: Riparian Zones. *Annu. Rev. Ecol. Syst.* **28,** 621–658 (1997).
127. Jose, S., Holzmueller, E. & Gillespie, A. R. Tree-Crop Interactions in Temperate Agroforestry. *North Am. Agrofor. An Integr. Sci. Pract. Agrofor. An Integr. Sci. Pract.* 57–74 (2009). doi:10.2134/2009.northamericanagroforestry.2ed.c4
128. Carrión, J. S., Parra, I., Navarro, C. & Munuera, M. Past distribution and ecology of cork oak (Quercus suber) in the Iberian Peninsula: a pollen-analytical approach. *Divers. Distrib.* **6,** 29–44 (2000).
129. Gerson, U. Lichen-Arthropod Associations. *Lichenol.* **5,** 434–443 (1973).
130. Bugalho, M. N., Caldeira, M. C., Pereira, J. S., Aronson, J. & Pausas, J. G. Mediterranean cork oak savannas require human use to sustain biodivsersity and ecosystem services. *Front. Ecol. Environ.* **9,** 278–286 (2011).
131. Jose, S. Agroforestry for ecosystem services and environmental benefits: An overview. in *Agroforestry Systems* **76,** 1–10 (ed. HE Garrett, 2009).
132. Jonason, D., Franzén, M. & Ranius, T. Surveying Moths Using Light Traps: Effects of Weather and Time of Year. *PLoS One* **9,** e92453 (2014).
133. Hodkinson, I. D. Terrestrial insects along elevation gradients: Species and community responses to altitude. *Biol. Rev. Camb. Philos. Soc.* **80,** 489–513 (2005).
134. Bush, A. *et al.* Connecting Earth observation to high-throughput biodiversity data. *Nat. Ecol. Evol.* **1,** 0176 (2017).

Appendix I-A

Sampling localities

LOCALITY	LATITUDE	LONGITUDE	SAMPLE	SEASON	HABITAT
01	41.21646119	-7.439015468	IT0010	July	Riparian_zone
01	41.21646119	-7.439015468	IT0070	September	Riparian_zone
01	41.21692668	-7.439548828	IT0009	July	Riparian_zone
01	41.21692668	-7.439548828	IT0069	September	Riparian_zone
02	41.21533833	-7.440673254	IT0072	September	Vineyard
02	41.21566345	-7.441023769	IT0071	September	Olive_grove
03	41.21188769	-7.427407078	IT0007	July	Vineyard
03	41.21188769	-7.427407078	IT0073	September	Vineyard
03	41.21309655	-7.427678187	IT0008	July	Olive_grove
03	41.21309655	-7.427678187	IT0074	September	Olive_grove
04	41.20554765	-7.403392632	IT0001	July	Cork_oak_woodland
04	41.20554765	-7.403392632	IT0075	September	Cork_oak_woodland
04	41.20583082	-7.403351744	IT0002	July	Cork_oak_woodland
04	41.20583082	-7.403351744	IT0076	September	Cork_oak_woodland
05	41.22826848	-7.386477951	IT0003	July	Vineyard
05	41.228342	-7.385596	IT0077	September	Vineyard
05	41.2288898	-7.3847196	IT0004	July	Olive_grove
06	41.22667391	-7.406635946	IT0005	July	Vineyard
06	41.22667391	-7.406635946	IT0079	September	Vineyard
06	41.22681147	-7.406273896	IT0006	July	Olive_grove
06	41.22681147	-7.406273896	IT0080	September	Olive_grove
07	41.3483723	-7.305618995	IT0013	July	Vineyard
07	41.3483723	-7.305618995	IT0081	September	Vineyard
07	41.34848616	-7.305225367	IT0014	July	Vineyard
07	41.34848616	-7.305225367	IT0082	September	Vineyard
08	41.3549874	-7.296756058	IT0012	July	Olive_grove
08	41.3549874	-7.296756058	IT0084	September	Olive_grove
08	41.35568522	-7.295892738	IT0011	July	Olive_grove
08	41.35568522	-7.295892738	IT0083	September	Olive_grove
09	41.34936573	-7.289928225	IT0016	July	Cork_oak_woodland
09	41.34936573	-7.289928225	IT0086	September	Cork_oak_woodland
09	41.34994197	-7.28945322	IT0015	July	Cork_oak_woodland
09	41.34994197	-7.28945322	IT0085	September	Cork_oak_woodland
10	41.34633227	-7.279207426	IT0017	July	Riparian_zone
10	41.34633227	-7.279207426	IT0087	September	Riparian_zone

10	41.34704294	-7.277694907	IT0018	July	Riparian_zone
10	41.34704294	-7.277694907	IT0088	September	Riparian_zone
11	41.33188949	-7.255711574	IT0022	July	Olive_grove
11	41.33188949	-7.255711574	IT0090	September	Olive_grove
11	41.33248321	-7.255118792	IT0021	July	Vineyard
11	41.33248321	-7.255118792	IT0089	September	Vineyard
12	41.32967405	-7.240786207	IT0020	July	Olive_grove
12	41.32967405	-7.240786207	IT0092	September	Olive_grove
12	41.33064472	-7.240130774	IT0019	July	Vineyard
12	41.33064472	-7.240130774	IT0091	September	Vineyard
13	41.3135814	-7.2212788	IT0146	September	Cork_oak_woodland
13	41.313774	-7.2205704	IT0144	September	Cork_oak_woodland
13	41.3139458	-7.220831	IT0145	September	Cork_oak_woodland
13	41.314726	-7.2201841	IT0143	September	Cork_oak_woodland
13	41.315647	-7.219312	IT0024	July	Cork_oak_woodland
13	41.315647	-7.219312	IT0094	September	Cork_oak_woodland
13	41.31569963	-7.219785207	IT0023	July	Cork_oak_woodland
13	41.31569963	-7.219785207	IT0093	September	Cork_oak_woodland
14	41.32739252	-7.368112612	IT0025	July	Riparian_zone
14	41.32739252	-7.368112612	IT0095	September	Riparian_zone
14	41.32742887	-7.367688494	IT0026	July	Riparian_zone
14	41.32742887	-7.367688494	IT0096	September	Riparian_zone
15	41.33313008	-7.337704078	IT0027	July	Vineyard
15	41.33313008	-7.337704078	IT0097	September	Vineyard
15	41.33347649	-7.337398151	IT0028	July	Vineyard
15	41.33347649	-7.337398151	IT0098	September	Vineyard
16	41.3225093	-7.359209652	IT0029	July	Olive_grove
16	41.3225093	-7.359209652	IT0099	September	Olive_grove
16	41.32325647	-7.358178826	IT0030	July	Olive_grove
16	41.32325647	-7.358178826	IT0100	September	Olive_grove
17	41.28070644	-7.364140192	IT0032	July	Vineyard
17	41.28070644	-7.364140192	IT0102	September	Vineyard
17	41.28119974	-7.364093046	IT0031	July	Olive_grove
17	41.28119974	-7.364093046	IT0101	September	Olive_grove
18	41.27580092	-7.377442141	IT0034	July	Vineyard
18	41.27580092	-7.377442141	IT0104	September	Vineyard
18	41.27601447	-7.376900279	IT0035	July	Vineyard
18	41.27601447	-7.376900279	IT0105	September	Vineyard
18	41.27623074	-7.37590176	IT0033	July	Olive_grove

18	41.27623074	-7.37590176	IT0103	September	Olive_grove
19	41.27872606	-7.390492147	IT0036	July	Cork_oak_woodland
19	41.27909507	-7.390965451	IT0037	July	Cork_oak_woodland
19	41.284849	-7.3871808	IT0106	September	Cork_oak_woodland
19	41.2848652	-7.3864131	IT0107	September	Cork_oak_woodland
19	41.2849521	-7.3860158	IT0139	September	Cork_oak_woodland
19	41.2851007	-7.3855968	IT0140	September	Cork_oak_woodland
20	41.4039292	-7.424184445	IT0039	July	Olive_grove
20	41.4039292	-7.424184445	IT0109	September	Olive_grove
20	41.40445381	-7.423626477	IT0038	July	Vineyard
20	41.40445381	-7.423626477	IT0108	September	Vineyard
21	41.38821113	-7.411579205	IT0040	July	Vineyard
21	41.38821113	-7.411579205	IT0110	September	Vineyard
21	41.38869375	-7.411817043	IT0041	July	Olive_grove
21	41.38869375	-7.411817043	IT0111	September	Olive_grove
22	41.35888039	-7.402107677	IT0043	July	Riparian_zone
22	41.35888039	-7.402107677	IT0113	September	Riparian_zone
22	41.35900283	-7.401794928	IT0042	July	Riparian_zone
22	41.35900283	-7.401794928	IT0112	September	Riparian_zone
23	41.34839786	-7.403592236	IT0045	July	Olive_grove
23	41.34839786	-7.403592236	IT0115	September	Olive_grove
23	41.34868012	-7.402782951	IT0044	July	Olive_grove
23	41.34868012	-7.402782951	IT0114	September	Olive_grove
24	41.34642385	-7.405332368	IT0046	July	Vineyard
24	41.34642385	-7.405332368	IT0116	September	Vineyard
24	41.34682212	-7.404242954	IT0047	July	Vineyard
24	41.34682212	-7.404242954	IT0117	September	Vineyard
25	41.28312538	-7.395625131	IT0049	July	Olive_grove
25	41.28328282	-7.395395149	IT0048	July	Olive_grove
25	41.28328282	-7.395395149	IT0118	September	Olive_grove
25	41.2838939	-7.395576842	IT0050	July	Vineyard
25	41.2838939	-7.395576842	IT0120	September	Vineyard
26	41.27951026	-7.39976681	IT0052	July	Cork_oak_woodland
26	41.27951026	-7.39976681	IT0122	September	Cork_oak_woodland
26	41.27957209	-7.399581073	IT0051	July	Cork_oak_woodland
26	41.27957209	-7.399581073	IT0121	September	Cork_oak_woodland
27	41.40645943	-7.162168639	IT0054	July	Riparian_zone
27	41.40645943	-7.162168639	IT0124	September	Riparian_zone
27	41.40692542	-7.162469497	IT0053	July	Riparian_zone

27	41.40692542	-7.162469497	IT0123	September	Riparian_zone
28	41.40286667	-7.165467122	IT0055	July	Olive_grove
28	41.40286667	-7.165467122	IT0125	September	Olive_grove
28	41.40287015	-7.166291163	IT0056	July	Olive_grove
28	41.40287015	-7.166291163	IT0126	September	Olive_grove
29	41.41120066	-7.17159127	IT0058	July	Cork_oak_woodland
29	41.41120066	-7.17159127	IT0128	September	Cork_oak_woodland
29	41.41158184	-7.17119856	IT0057	July	Cork_oak_woodland
29	41.41158184	-7.17119856	IT0127	September	Cork_oak_woodland
30	41.40108631	-7.160958973	IT0059	July	Vineyard
30	41.40108631	-7.160958973	IT0129	September	Vineyard
30	41.40125679	-7.160080301	IT0060	July	Vineyard
30	41.40125679	-7.160080301	IT0130	September	Vineyard
31	41.3792164	-7.207574053	IT0061	July	Olive_grove
31	41.3792164	-7.207574053	IT0131	September	Olive_grove
31	41.37927226	-7.208257517	IT0062	July	Vineyard
31	41.37927226	-7.208257517	IT0132	September	Vineyard
32	41.37348626	-7.209405207	IT0064	July	Cork_oak_woodland
32	41.37372397	-7.209455403	IT0063	July	Cork_oak_woodland
32	41.37372397	-7.209455403	IT0133	September	Cork_oak_woodland
32	41.3738223	-7.209726	IT0134	September	Cork_oak_woodland
33	41.37307806	-7.215178704	IT0066	July	Olive_grove
33	41.37307806	-7.215178704	IT0136	September	Olive_grove
33	41.37307867	-7.2145147	IT0065	July	Vineyard
33	41.37307867	-7.2145147	IT0135	September	Vineyard
34	41.38792758	-7.201763158	IT0067	July	Riparian_zone
34	41.3883003	-7.1987251	IT0137	September	Riparian_zone
34	41.38838635	-7.199995984	IT0068	July	Riparian_zone
34	41.38838635	-7.199995984	IT0138	September	Riparian_zone
35	41.2555563	-7.4149703	IT0142	September	Cork_oak_woodland
35	41.2557056	-7.4145513	IT0141	September	Cork_oak_woodland

Appendix I-B

Sample codes and geographical coordinates

SAMPLE	SEASON	HABITAT	LATITUDE	LONGITUDE
IT0001	July	Cork_oak_woodland	41.20554765	-7.403392632
IT0002	July	Cork_oak_woodland	41.20583082	-7.403351744
IT0003	July	Vineyard	41.22826848	-7.386477951
IT0004	July	Olive_grove	41.2288898	-7.3847196
IT0005	July	Vineyard	41.22667391	-7.406635946
IT0006	July	Olive_grove	41.22681147	-7.406273896
IT0007	July	Vineyard	41.21188769	-7.427407078
IT0008	July	Olive_grove	41.21309655	-7.427678187
IT0009	July	Riparian_zone	41.21692668	-7.439548828
IT0010	July	Riparian_zone	41.21646119	-7.439015468
IT0011	July	Olive_grove	41.35568522	-7.295892738
IT0012	July	Olive_grove	41.3549874	-7.296756058
IT0013	July	Vineyard	41.3483723	-7.305618995
IT0014	July	Vineyard	41.34848616	-7.305225367
IT0015	July	Cork_oak_woodland	41.34994197	-7.28945322
IT0016	July	Cork_oak_woodland	41.34936573	-7.289928225
IT0017	July	Riparian_zone	41.34633227	-7.279207426
IT0018	July	Riparian_zone	41.34704294	-7.277694907
IT0019	July	Vineyard	41.33064472	-7.240130774
IT0020	July	Olive_grove	41.32967405	-7.240786207
IT0021	July	Vineyard	41.33248321	-7.255118792
IT0022	July	Olive_grove	41.33188949	-7.255711574
IT0023	July	Cork_oak_woodland	41.31569963	-7.219785207
IT0024	July	Cork_oak_woodland	41.315647	-7.219312
IT0025	July	Riparian_zone	41.32739252	-7.368112612
IT0026	July	Riparian_zone	41.32742887	-7.367688494
IT0027	July	Vineyard	41.33313008	-7.337704078
IT0028	July	Vineyard	41.33347649	-7.337398151
IT0029	July	Olive_grove	41.3225093	-7.359209652
IT0030	July	Olive_grove	41.32325647	-7.358178826
IT0031	July	Olive_grove	41.28119974	-7.364093046
IT0032	July	Vineyard	41.28070644	-7.364140192
IT0033	July	Olive_grove	41.27623074	-7.37590176
IT0034	July	Vineyard	41.27580092	-7.377442141
IT0035	July	Vineyard	41.27601447	-7.376900279

IT0036	July	Cork_oak_woodland	41.27872606	-7.390492147
IT0037	July	Cork_oak_woodland	41.27909507	-7.390965451
IT0038	July	Vineyard	41.40445381	-7.423626477
IT0039	July	Olive_grove	41.4039292	-7.424184445
IT0040	July	Vineyard	41.38821113	-7.411579205
IT0041	July	Olive_grove	41.38869375	-7.411817043
IT0042	July	Riparian_zone	41.35900283	-7.401794928
IT0043	July	Riparian_zone	41.35888039	-7.402107677
IT0044	July	Olive_grove	41.34868012	-7.402782951
IT0045	July	Olive_grove	41.34839786	-7.403592236
IT0046	July	Vineyard	41.34642385	-7.405332368
IT0047	July	Vineyard	41.34682212	-7.404242954
IT0048	July	Olive_grove	41.28328282	-7.395395149
IT0049	July	Olive_grove	41.28312538	-7.395625131
IT0050	July	Vineyard	41.2838939	-7.395576842
IT0051	July	Cork_oak_woodland	41.27957209	-7.399581073
IT0052	July	Cork_oak_woodland	41.27951026	-7.39976681
IT0053	July	Riparian_zone	41.40692542	-7.162469497
IT0054	July	Riparian_zone	41.40645943	-7.162168639
IT0055	July	Olive_grove	41.40286667	-7.165467122
IT0056	July	Olive_grove	41.40287015	-7.166291163
IT0057	July	Cork_oak_woodland	41.41158184	-7.17119856
IT0058	July	Cork_oak_woodland	41.41120066	-7.17159127
IT0059	July	Vineyard	41.40108631	-7.160958973
IT0060	July	Vineyard	41.40125679	-7.160080301
IT0061	July	Olive_grove	41.3792164	-7.207574053
IT0062	July	Vineyard	41.37927226	-7.208257517
IT0063	July	Cork_oak_woodland	41.37372397	-7.209455403
IT0064	July	Cork_oak_woodland	41.37348626	-7.209405207
IT0065	July	Vineyard	41.37307867	-7.2145147
IT0066	July	Olive_grove	41.37307806	-7.215178704
IT0067	July	Riparian_zone	41.38792758	-7.201763158
IT0068	July	Riparian_zone	41.38838635	-7.199995984
IT0069	September	Riparian_zone	41.21692668	-7.439548828
IT0070	September	Riparian_zone	41.21646119	-7.439015468
IT0071	September	Olive_grove	41.21566345	-7.441023769
IT0072	September	Vineyard	41.21533833	-7.440673254
IT0073	September	Vineyard	41.21188769	-7.427407078
IT0074	September	Olive_grove	41.21309655	-7.427678187

IT0075	September	Cork_oak_woodland	41.20554765	-7.403392632
IT0076	September	Cork_oak_woodland	41.20583082	-7.403351744
IT0077	September	Vineyard	41.228342	-7.385596
IT0079	September	Vineyard	41.22667391	-7.406635946
IT0080	September	Olive_grove	41.22681147	-7.406273896
IT0081	September	Vineyard	41.3483723	-7.305618995
IT0082	September	Vineyard	41.34848616	-7.305225367
IT0083	September	Olive_grove	41.35568522	-7.295892738
IT0084	September	Olive_grove	41.3549874	-7.296756058
IT0085	September	Cork_oak_woodland	41.34994197	-7.28945322
IT0086	September	Cork_oak_woodland	41.34936573	-7.289928225
IT0087	September	Riparian_zone	41.34633227	-7.279207426
IT0088	September	Riparian_zone	41.34704294	-7.277694907
IT0089	September	Vineyard	41.33248321	-7.255118792
IT0090	September	Olive_grove	41.33188949	-7.255711574
IT0091	September	Vineyard	41.33064472	-7.240130774
IT0092	September	Olive_grove	41.32967405	-7.240786207
IT0093	September	Cork_oak_woodland	41.31569963	-7.219785207
IT0094	September	Cork_oak_woodland	41.315647	-7.219312
IT0095	September	Riparian_zone	41.32739252	-7.368112612
IT0096	September	Riparian_zone	41.32742887	-7.367688494
IT0097	September	Vineyard	41.33313008	-7.337704078
IT0098	September	Vineyard	41.33347649	-7.337398151
IT0099	September	Olive_grove	41.3225093	-7.359209652
IT0100	September	Olive_grove	41.32325647	-7.358178826
IT0101	September	Olive_grove	41.28119974	-7.364093046
IT0102	September	Vineyard	41.28070644	-7.364140192
IT0103	September	Olive_grove	41.27623074	-7.37590176
IT0104	September	Vineyard	41.27580092	-7.377442141
IT0105	September	Vineyard	41.27601447	-7.376900279
IT0106	September	Cork_oak_woodland	41.284849	-7.3871808
IT0107	September	Cork_oak_woodland	41.2848652	-7.3864131
IT0108	September	Vineyard	41.40445381	-7.423626477
IT0109	September	Olive_grove	41.4039292	-7.424184445
IT0110	September	Vineyard	41.38821113	-7.411579205
IT0111	September	Olive_grove	41.38869375	-7.411817043
IT0112	September	Riparian_zone	41.35900283	-7.401794928
IT0113	September	Riparian_zone	41.35888039	-7.402107677
IT0114	September	Olive_grove	41.34868012	-7.402782951

IT0115	September	Olive_grove	41.34839786	-7.403592236
IT0116	September	Vineyard	41.34642385	-7.405332368
IT0117	September	Vineyard	41.34682212	-7.404242954
IT0118	September	Olive_grove	41.28328282	-7.395395149
IT0120	September	Vineyard	41.2838939	-7.395576842
IT0121	September	Cork_oak_woodland	41.27957209	-7.399581073
IT0122	September	Cork_oak_woodland	41.27951026	-7.39976681
IT0123	September	Riparian_zone	41.40692542	-7.162469497
IT0124	September	Riparian_zone	41.40645943	-7.162168639
IT0125	September	Olive_grove	41.40286667	-7.165467122
IT0126	September	Olive_grove	41.40287015	-7.166291163
IT0127	September	Cork_oak_woodland	41.41158184	-7.17119856
IT0128	September	Cork_oak_woodland	41.41120066	-7.17159127
IT0129	September	Vineyard	41.40108631	-7.160958973
IT0130	September	Vineyard	41.40125679	-7.160080301
IT0131	September	Olive_grove	41.3792164	-7.207574053
IT0132	September	Vineyard	41.37927226	-7.208257517
IT0133	September	Cork_oak_woodland	41.37372397	-7.209455403
IT0134	September	Cork_oak_woodland	41.3738223	-7.209726
IT0135	September	Vineyard	41.37307867	-7.2145147
IT0136	September	Olive_grove	41.37307806	-7.215178704
IT0137	September	Riparian_zone	41.3883003	-7.1987251
IT0138	September	Riparian_zone	41.38838635	-7.199995984
IT0139	September	Cork_oak_woodland	41.2849521	-7.3860158
IT0140	September	Cork_oak_woodland	41.2851007	-7.3855968
IT0141	September	Cork_oak_woodland	41.2557056	-7.4145513
IT0142	September	Cork_oak_woodland	41.2555563	-7.4149703
IT0143	September	Cork_oak_woodland	41.314726	-7.2201841
IT0144	September	Cork_oak_woodland	41.313774	-7.2205704
IT0145	September	Cork_oak_woodland	41.3139458	-7.220831
IT0146	September	Cork_oak_woodland	41.3135814	-7.2212788

Appendix II

Arthropod DNA extraction from bulk samples using the E.Z.N.A.® Tissue DNA Kit

Before starting:

- Set the oven at 56ºC and place the Gordon's and elution buffer in it;
- Clean the working space and material with bleach and water and leave it under the UV-light on for at least 15 mins;
- Use filter tips at all steps;
- Use a negative control for each batch (23 samples);
- If necessary, prepare Gordon's buffer by using an appropriate protocol as guidance.

Procedure:

1. Prepare the number of 2mL tubes needed for the extraction.
2. Distribute 1000 µL of Gordon's lysis buffer and 25 µL of OB Protease per tube. When distributing the lysis buffer dispense the volume slowly and aim to the tube not your sample or you may risk cross contamination by dust flying out of the tube. If using a pipette dispenser set the dispenser speed to the lowest setting.
3. Quickly vortex all tubes and leave it the oven for 30 minutes.
4. While the samples are in the oven clean the working space and material with bleach and ethanol and prepare a cutting space with paper cloth and aluminum foil.
5. Cut each Inhibitex tablet in 2 pieces.
6. Turn-on the dry bath and set it at 70ºC.
7. Per sample prepare:
    - one 2mL tube with Inhibitex;
    - one 2ml tube with 25 µL of OB Protease
    - two 1.5mL tube for both final elutions.
8. Short-spin samples and transfer up to 700 µL of the supernatant to the tube with Inhibitex.
9. Vortex for 1min and centrifuge for 30s at 14000rpm.
10. Transfer up to 500 µL of the supernatant (avoid to pipette sediments or the white membrane on the surface) to the 2mL tube containing OB Protease.
11. Add 200 µL of BL buffer. Vortex at maximum speed for 15 seconds and short-spin.
12. Place the samples in the dry-bath for 10 minutes. Use this time to label the columns and, if using, set the vacuum system.
13. Short-spin samples to remove condensation.
14. Add 400 µL of ethanol 100% and vortex at maximum speed for 20 seconds. Short-spin samples.

Safe point to stop! *If needed, you can put samples in the freezer (covered in aluminum foil to avoid contaminations). If stopping, when re-starting the protocol allow samples to come to room temperature for a few minutes.*

<u>If using the centrifuge see steps 16 to 23. If using QIAGEN Vacuum Pump see steps 24 to 30.</u>

When using the centrifuge:

15. Transfer up to 600 µl of supernatant to the column with a collector tube. Centrifuge at 1000rpm for 1 minute.
16. Place the column in a new collection tube.
17. Repeat step 16 and 17 if you still have volume left from step 15.
18. Place the column in a new collection tube. Add 500 µl of HB Buffer. Centrifuge at 1000rpm for 1 minute.
19. Place the column in a new collection tube. Add 700 µl of DNA Wash Buffer. Centrifuge at 1000rpm for 1 minute.
20. Repeat previous step.
21. Place the column in a new collection tube. Centrifuge ate 14000rpm for 2 minutes to completely dry the membrane.
22. Proceed to step 30.

When using the QIAvac:

Do not forget to place a VacConnector between the QIAvac and EZNA column.

23. Transfer up to 600 µl of supernatant to the column placed in the QIAvac 24. Repeat if necessary.
24. Turn on the Vacuum pump. The ideal vacuum pressure should be between -80 and -90 kPa.
25. Add 500 µl of HB Buffer. If one or more columns appears to be clogged turn off the pump when possible. Place the clogged column in a collector tube and centrifuge at 1000rpm for 1 minute. Return the column to its position in the QIAvac.
26. Add 700 µl of DNA Wash Buffer.
27. Repeat previous step.
28. Allow the columns membrane to completely dry for 5 to 10 mins (depends on the room temperature).
29. Proceed to step 30.

DNA Elution:

30. Transfer the column plate to 1.5mL labelled tube. Add 50 µL Elution Buffer and Incubate at room temperature for 5 minutes. Centrifuge at 10000 rpm for 1 minutes. This will be your 1st elution.
31. Transfer the column plate to 1.5mL labelled tube. Add 50 µL Elution Buffer and Incubate at room temperature for 5 minutes. Centrifuge at 10000 rpm for 1 minutes. This will be your 2nd elution.

**When enough samples have been extracted in tubes, transfer them into a plate and include this information in the database and in the lab book.*

**Always check the number of collector tubes available. The extraction kit only accounts for one change of the collector tube. However, we have a stock of those since most extractions are performed using the QIAGEN Vacuum Pump.*

*******After using VacConnecters placed them in a bag labelled as "Used" and when the bag is full placed them in a used tips box with bleach 60% overnight in Cristina's room. Replace the bleach with distilled water for at least 6 hours and they should be dried in the oven. NEVER AUTOCLAVE VacConnecters.*

Appendix III

Protocol testing

Table 0-1: Dark colors indicate detection of species by one of the primers with dark brown for species detected only when using B2 and dark green for LCO. Lighter colors indicate the identifications for which only one of the primers allowed for an accurate (non-ambiguous) species level identification.

Family	B2	LCO
Autostichidae	*Dysspastus fallax*	*Dysspastus fallax*
	Oegoconia quadripuncta	*Oegoconia quadripuncta*
	Symmoca nigromaculella	*Symmoca nigromaculella*
Blastobasidae	*Blastobasis phycidella*	
	sp.A	
Brachodidae	*Brachodes funebris*	*Brachodes funebris*
		Brachodes nanetta
Coleophoridae	*Coleophora bilineella*	*Coleophora bilineella*
	Coleophora strigosella	*Coleophora strigosella*
Crambidae	*Agriphila geniculea*	*Agriphila geniculea*
	Agriphila latistria	*Agriphila latistria*
	Catoptria staudingeri	*Catoptria staudingeri*
	Eudonia mercurella	*Eudonia mercurella*
	Evergestis dumerlei	*Evergestis dumerlei*
	Metasia ophialis	
	Nomophila noctuella	*Nomophila noctuella*
	Xanthocrambus delicatellus	
Drepanidae	*Watsonalla uncinula*	*Watsonalla uncinula*
		Ochromolopis staintonellus
Erebidae	*Catocala conjuncta*	*Catocala conjuncta*
	Catocala nymphagoga	*Catocala nymphagoga*
	Coscinia cribaria	*Coscinia cribaria*
	Cymbalophora pudica	*Cymbalophora pudica*
	Dysgonia algira	*Dysgonia algira*
	Eilema caniola	*Eilema caniola*
	Eilema uniola	*Eilema uniola*
	Eublemma parva	*Eublemma parva*
	Eublemma pura	*Eublemma pura*
	Lymantria dispar	*Lymantria dispar*
	Metachrostis velox	
	Ocneria rubea	
Gelechiidae	*Anacampsis scintillella*	*Anacampsis scintillella*
	Bryotropha dryadella	*Bryotropha dryadella*
	Epidola stigma	*Epidola stigma*
	Gelechia senticetella	
	Mirificarma lentiginosella	*Mirificarma lentiginosella*
	Neofriseria hitadoella	
Geometridae	*Adactylotis gesticularia*	*Adactylotis gesticularia*
	Camptogramma bilineata	*Camptogramma bilineata*
	Crocallis elinguaria/albarracina	*Crocallis elinguaria/albarracina*
	Cyclophora porata	*Cyclophora hyponoea/porata*
	Cyclophora punctaria	
	Cyclophora puppillaria	*Cyclophora puppillaria*
	Cyclophora ruficiliaria	
	Eupithecia centaureata	*Eupithecia centaureata*
	Eupithecia oxycedrata	*Eupithecia oxycedrata*
	Gymnoscelis rufifasciata	*Gymnoscelis rufifasciata*
	Idaea alyssumata	*Idaea alyssumata*

	Idaea calunetaria	
	Idaea degeneraria	*Idaea degeneraria*
	Idaea eugeniata	*Idaea eugeniata*
	Idaea infirmaria	
	Idaea joannisiata	*Idaea joannisiata*
	Idaea mustelata	*Idaea mustelata*
	Idaea obsoletaria	*Idaea obsoletaria*
	Idaea rhodogrammaria	
	Idaea subsericeata	*Idaea subsericeata*
	Nychiodes andalusiaria	*Nychiodes andalusiaria*
	Pachycnemia hippocastanaria	
	Peribatodes ilicaria	*Peribatodes ilicaria*
	Pseudoterpna coronillaria	*Pseudoterpna coronillaria*
	Rhodometra sacraria	
	Rhoptria asperaria	*Rhoptria asperaria*
	Scopula marginepunctata	
	Selidosema taeniolaria	*Selidosema taeniolaria*
	Tephronia sepiaria	*Tephronia sp.*
	Xenochlorodes olympiaria	
Gracillariidae	*Phyllonorycter roboris*	
Lasiocampidae	*Pachypasa limosa*	*Pachypasa limosa*
	Phyllodesma suberifolia	*Phyllodesma suberifolia*
	Psilogaster loti	*Psilogaster loti*
Lecithoceridae	*Eurodachtha canigella*	
	Homaloxestis briantiella	*Homaloxestis briantiella*
		Stigmella basiguttella
Noctuidae	*Acronicta rumicis*	*Acronicta rumicis*
	Agrotis bigramma	*Agrotis bigramma*
	Agrotis trux	
	Bryophila vandalusiae	*Bryophila vandalusiae*
	Callopistria latreillei	*Callopistria latrellei*
	Caradrina aspersa	*Caradrina aspersa*
	Caradrina flavirena	*Caradrina flavirena/noctivaga*
	Chloantha hyperici	*Chloantha hyperici*
	Craniophora pontica	
	Cryphia algae	*Cryphia algae/pallida*
	Cryphia pallida	*Cryphia algae/pallida*
	Epilecta linogrisea	*Epilecta linogrisea*
	Euxoa tritici/obelisca	*Euxoa tritici*
	Hoplodrina ambigua	*Hoplodrina ambigua*
	Hoplodrina hesperica	*Hoplodrina hesperica*
	Leucania putrescens	*Leucania putrescens*
	Luperina testacea	*Luperina testacea*
	Mesapamea didyma/secalis	
	Mesoligia furuncula	*Mesoligia furuncula*
	Mythimna albipuncta	
	Mythimna l-album	*Mythimna l-album*
	Mythimna riparia	*Mythimna riparia*
	Mythimna sicula	*Mythimna sicula*
	Mythimna vitellina	*Mythimna vitellina*
	Noctua janthe	*Noctua janthe*
	Noctua orbona	*Noctua orbona*
	Nyctobrya muralis	*Nyctobrya muralis*
	Peridroma saucia	
	Polymixis dubia	*Polymixis dubia*
	Spodoptera exigua	

	Tyta luctuosa	*Tyta luctuosa*
Nolidae	*Meganola strigula*	*Meganola strigula*
	Nola cucullatella	
Notodontidae	*Drymonia querna*	*Drymonia querna*
	Thaumetopoea herculeana	*Thaumetopoea herculeana*
Oecophoridae	*Batia lambdella*	*Batia lambdella*
	Pleurota honorella	*Pleurota sp.*
Psychidae	*Dissoctena albidella*	
Pterophoridae	*Stangeia siceliota*	*Stangeia siceliota*
Pyralidae	*Acrobasis bithynella*	*Acrobasis bithynella*
	Acrobasis consociella	*Acrobasis consociella*
	Acrobasis glaucella	*Acrobasis glaucella*
	Alophia combustella	*Alophila combustella*
	Ancylosis cinnamomella	*Ancylosis cinnamomella*
	Bostra obsoletalis	*Bostra obsoletalis*
	Elegia fallax	*Elegia fallax*
		Ephestia parasitella
	Etiella zinckenella	*Etiella zinckenella*
	Homoeosoma sinuella	*Homeosoma sinuella*
		Oxybia transversella
	Pempelia palumbella	*Pempelia palumbella*
	Phycita roborella	*Phycita roborella*
	Psorosa nucleolella	
	sp.A	*sp.A*
	Stemmatophora brunnealis	
	Stemmatophora rungsi	*Stemmatophora rungsi*
Tineidae	*Infurcitinea atrifasciella*	
Tortricidae	*Cydia amplana*	
	Cydia fagiglandana	*Cydia flagiglandana*
	Cydia splendana	*Cydia splendana*
	Notocelia incarnatana	
	Phalonidia contractana	*Phalonidia contractana*
		Lobesia botrana
Total	**131**	**105**

Appendix IV

Validation of DNA metabarcoding results

Table 0-1: Lepidoptera OTUs detected using either only DNA metabarcoding but not through visual identification and the opposite: families detected through visual identification but not when using DNA metabarcoding. Values refer to the percentage of occurences of species in the twelve samples identified exclusively by metabarcoding or visual ID as well as by both methods.

Family	Species	Metabarcoding ID	Visual ID	Both methods	Total
Autostichidae	*Apatema mediopallidum*	0	1	0	1
	Autostichidae sp. 1	1	0	0	1
	Symmoca signatella	0	0	3	3
	Symmoca sp. 1	1	0	0	1
Brachodidae	*Brachodes funebris*	6	0	0	6
	Brachodes gaditana	0	6	1	7
Crambidae	*Agrotera nemoralis*	0	0	1	1
	Anarpia incertalis	0	0	2	2
	Catoptria pinella	0	0	2	2
	Eudonia mercurella	0	1	10	11
	Metasia sp. 1	1	0	1	2
	Pleuroptya balteata	0	0	1	1
Depressaridae	*Agonopterix atomella*	1	0	0	1
	Agonopterix scopariella	1	1	2	4
	Depressaria adustatella	0	0	1	1
	Elegia cf fallaximima	0	1	0	1
Drepanidae	*Watsonalla uncinula*	0	1	5	6
Erebidae	*Catephia alchymista*	0	0	1	1
	Catocala nymphagoga	0	0	10	10
	Dysgonia algira	0	0	1	1
	Eublemma candidana	1	0	0	1
	Eublemma parva	0	1	5	6
	Eublemma pura	0	1	4	5
	Lymantria dispar	0	0	12	12
	Ocneria rubea	0	0	1	1
	Phragmatobia fuliginosa	0	0	1	1
	Zethes insularis	0	0	1	1
Gelechiidae	*Anacampsis scintillella*	2	0	0	2
	Aproaerema anthyllidella	0	0	1	1
	Aristotelia ericinella	0	1	1	2
	Epidola stigma	0	0	2	2
	Gelechiidae sp. 4	0	0	1	1
	Neofriseria hitadoella	0	1	6	7
	Neotelphusa cisti	1	0	1	2
	Pseudotelphusa occidentella	0	0	1	1
Geometridae	*Brachyglossina hispanaria*	0	0	1	1
	Camptogramma bilineata	0	0	3	3
	Charissa mucidaria	0	0	1	1
	Cyclophora puppillaria	0	0	4	4
	Gymnoscelis rufifasciata	0	0	2	2
	Idaea belemiata	0	12	0	12
	Idaea circuitaria	0	0	2	2
	Idaea consanguiberica	1	0	0	1
	Idaea eugeniata	0	0	1	1
	Idaea exilaria	0	0	2	2
	Idaea infirmaria	0	1	8	9
	Idaea mustelata	0	0	7	7
	Idaea nigrolineata	0	0	1	1
	Idaea obsoletaria	7	0	1	8
	Idaea rhodogrammaria	2	0	4	6
	Menophra abruptaria	0	0	3	3
	Menophra japygiaria	0	0	1	1
	Nychiodes andalusiaria	0	0	8	8
	Phaiogramma etruscaria	0	0	2	2
	Plagodis dolabraria	0	0	1	1
	Pseudoterpna coronillaria	1	0	4	5
	Rhoptria asperaria	0	0	5	5
	Scopula marginepunctata	1	0	0	1
	Scopula submutata	0	1	0	1
	Tephronia sepiaria	0	1	3	4

Lasiocampidae	*Phyllodesma suberifolia*	0	0	7	7
	Psilogaster loti	0	0	4	4
Lecithoceridae	*Eurodachtha canigella*	0	2	1	3
	Lecithoceridae sp. 1	1	0	0	1
Momphidae	*Mompha miscella*	0	1	0	1
Noctuidae	*Acontia lucida*	0	0	1	1
	Bryophila ravula	1	0	0	1
	Calophasia sp.	0	1	0	1
	Caradrina aspersa	0	0	8	8
	Chloantha hyperici	1	0	0	1
	Cryphia algae	1	1	5	7
	Cryphia pallida	1	0	0	1
	Epilecta linogrisea	0	0	6	6
	Euxoa tritici	0	0	1	1
	Hecatera dysodea	0	0	1	1
	Heliothis peltigera	0	1	0	1
	Lophoterges millierei	0	0	1	1
	Noctua comes	1	0	5	6
	Noctua fimbriata	0	0	1	1
	Noctua janthe	0	1	4	5
	Noctua orbona	0	1	0	1
	Noctua tirrenica	1	0	0	1
	Nyctobrya muralis	0	0	7	7
	Paucgraphia erythrina	0	0	1	1
	Pleurota honorella	0	0	1	1
	Polyphaenis sericata	0	2	1	3
	Tyta luctuosa	0	0	1	1
Nolidae	*Meganola strigula*	0	1	3	4
Notodontidae	*Cerura iberica*	1	0	0	1
	Harpyia milhauseri	0	0	3	3
	Spatalia argentina	0	0	1	1
	Thaumetopoea pityocampa	0	0	1	1
Nymphalidae	*Maniola jurtina*	0	0	1	1
Oecophoridae	*Epicallima mercedella*	0	1	4	5
	Goidanichiana jourdheuillella	0	0	1	1
Pterophoridae	*Crombrugghia laetus*	0	0	1	1
	Stangeia siceliota	0	0	1	1
Pyralidae	*Acrobasis glaucella*	2	0	4	6
	Acrobasis sodalella	0	2	0	2
	Acrobasis tumidana	0	0	1	1
	Ancylosis cinnamomella	0	1	1	2
	Bostra obsoletalis	0	2	5	7
	Endotricha flammealis	0	0	1	1
	Etiella zinckenella	0	0	6	6
	Galleria mellonella	0	0	1	1
	Homoeosoma sinuella	0	1	1	2
	Oxybia transversella	0	1	2	3
	Pempelia palumbella	0	0	7	7
	Phycita roborella	0	0	4	4
	Pyralidae sp. 2	1	0	0	1
	Synaphe punctalis	0	0	5	5
Scythridae	*Scythris dissimilella*	3	0	0	3
	Scythris parafuscoaenea	0	4	0	4
Tineidae	*Anomalotinea liguriella*	0	0	2	2
	Nemapogon agenjoi	0	0	2	2
Tortricidae	*Cydia fagiglandana*	0	0	12	12
Unknown	*Lepidoptera sp. 10*	1	0	0	1
	Lepidoptera sp. 12	2	0	0	2
Ypsolophidae	*Ypsolopha persicella*	0	1	0	1
Total		**44 (12.02%)**	**54 (14.75%)**	**268 (73.22%)**	**366 (100%)**

Appendix V

A) Multiple comparison test results for all insects OTU richness

Between habitats	Contrast			Z-ratio	P-value
July	Cork_oak_woodland	vs	Olive_grove	-5.264	**<0.0001**
	Cork_oak_woodland	vs	Riparian_zone	-3.39	**0.0196**
	Cork_oak_woodland	vs	Vineyard	-0.67	1
	Olive_grove	vs	Riparian_zone	1.406	1
	Olive_grove	vs	Vineyard	5.207	**<0.0001**
	Riparian_zone	vs	Vineyard	3.065	0.0609
September	Cork_oak_woodland	vs	Olive_grove	0.568	1
	Cork_oak_woodland	vs	Riparian_zone	1.042	1
	Cork_oak_woodland	vs	Vineyard	6.903	**<0.0001**
	Olive_grove	vs	Riparian_zone	0.547	1
	Olive_grove	vs	Vineyard	6.195	**<0.0001**
	Riparian_zone	vs	Vineyard	4.846	**<0.0001**
Between seasons					
Cork_oak_woodland	July	vs	September	-3.665	**0.0069**
Olive_grove	July	vs	September	2.405	0.4528
Riparian_zone	July	vs	September	1.08	1
Vineyard	July	vs	September	3.464	**0.0149**

B) Multiple comparison test results for Lepidoptera OTU richness

Between habitats	Contrast			Z-ratio	P-value
July	Cork_oak_woodland	vs	Olive_grove	-1.46	1
	Cork_oak_woodland	vs	Riparian_zone	2.524	0.3249
	Cork_oak_woodland	vs	Vineyard	2.406	0.4519
	Olive_grove	vs	Riparian_zone	4.06	**0.0014**
	Olive_grove	vs	Vineyard	4.295	**0.0005**
	Riparian_zone	vs	Vineyard	-0.48	1
September	Cork_oak_woodland	vs	Olive_grove	3.59	**0.0093**
	Cork_oak_woodland	vs	Riparian_zone	7.929	**<0.0001**
	Cork_oak_woodland	vs	Vineyard	8.048	**<0.0001**
	Olive_grove	vs	Riparian_zone	5.01	**<0.0001**
	Olive_grove	vs	Vineyard	4.33	**0.0004**
	Riparian_zone	vs	Vineyard	-1.49	1
Between seasons					
Cork_oak_woodland	July	vs	September	-8.463	**<0.0001**
Olive_grove	July	vs	September	-4.469	**0.0002**
Riparian_zone	July	vs	September	-2.509	0.3389
Vineyard	July	vs	September	-4.534	**0.0002**

Appendix VI

Species richness vs. taxonomic index (Delta+)

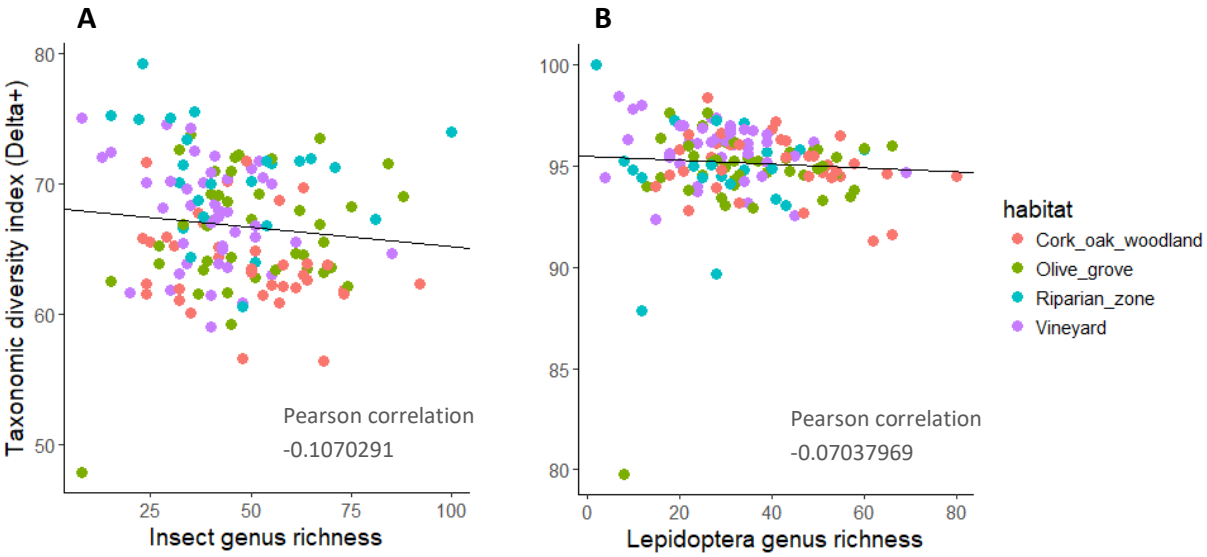


Figure A-VI-1-1: Comparison of species richness with the respective taxonomic index (Delta+) for each sample and the respective Pearson correlation values.

Appendix VII

A) Multiple comparison test habitat effect delta+ all insect OTUs

	Contrast			Z-ratio	P-value
Between habitats	Cork_oak_woodland	vs	Olive_grove	-2.132	0.1982
	Cork_oak_woodland	vs	Riparian_zone	-6.215	**<0.0001**
	Cork_oak_woodland	vs	Vineyard	-3.78	**0.0009**
	Olive_grove	vs	Riparian_zone	-4.517	**<0.0001**
	Olive_grove	vs	Vineyard	-1.706	0.5286
	Riparian_zone	vs	Vineyard	3.095	**0.0118**

B) Multiple comparison test habitat effect delta+ Lepidoptera OTUs

	Contrast			Z-ratio	P-value
Between habitats	Cork_oak_woodland	vs	Olive_grove	-0.857	1
	Cork_oak_woodland	vs	Riparian_zone	-4.348	**0.0004**
	Cork_oak_woodland	vs	Vineyard	-1.968	1
	Olive_grove	vs	Riparian_zone	-3.91	**0.0026**
	Olive_grove	vs	Vineyard	-1.243	1
	Riparian_zone	vs	Vineyard	2.85	0.1226