Developing and validating the efficacy of DNA metabarcoding for nationwide insect biomonitoring schemes

Dimitrios Petsopoulos

Doctor of Philosophy

Newcastle University

School of Natural and Environmental Sciences

December 2022

Abstract

Next-generation sequencing (NGS) approaches are revolutionizing the study of biodiversity and ecosystem functioning by providing scalable, economic and efficient methods, with a range of immediate applications for national insect biomonitoring schemes. However, such approaches are still in their infancy and are yet to be applied to established schemes in the UK. With insect biodiversity being under threat and the recent evidence on global insect declines, there is a clear need for reliable and fast methods for describing insect communities that overcome the pitfalls of traditional monitoring. This thesis tests and validates the application of DNA metabarcoding to identify insect communities in bulk samples of insects collected as part of two major biomonitoring schemes in the UK. First, I give an overview of insect monitoring, and the molecular approaches used for insect monitoring particularly in an agricultural context, highlighting limitations and potential future directions. Second, I assess the efficacy of DNA-metabarcoding to identify aphid archival samples of a 16-year time series from the Rothamsted Insect Survey (RIS). I show how DNA-metabarcoding can reliably identify samples that are as old as 18 years with high congruence (~80%), but without destroying their morphological integrity. I apply these methods to the corresponding bycatch (i.e non-target taxa captured) samples from the same trap and reveal over 800 insect taxa caught over the same period, many of which can be regarded as ecosystem service providers, pests and/or newly discovered species in the UK. I highlight the potential of bycatch samples to construct time-series for hundreds of insect taxa, enabling previously intractable questions to be addressed. I then apply these methods to analyse bycatch samples from the Fera's Yellow Water Pan Trap Network and show the utility of DNA metabarcoding for describing and comparing biodiversity without relying on conventional taxonomy. Finally, I show how and why combining molecular approaches and long-term schemes can help to understand and mitigate insect declines. I discuss the potential biases and limitations of these methods, whilst identifying direct applications for existing and future insect biomonitoring schemes.

Dedicated to my parents Μανώλη and Αθηνά.

I could have never made it where I am now without their support and to my grandma Ευτυχία which will forever call all insects, μικρόβια (microbes)

Acknowledgments

I would first like to acknowledge many of my friends, family and colleagues who have provided insight and support over many years of academic journeys. Especially those who have assisted me in the lab and for computer work: previous PhD fellows that are now Doctors: Justin Byrne and Johanna Weston, current and past members of my group: Dr. Kirsten Miller, Dr. Jordan Cuff, Kyle Miller (who might have spiked all my samples with Carcelia), Dr. Pilar Suarez for her positive vibes in the lab, Dr. Fredrick Windsor and Douglas Boyes. Thanks to many of my friends up in Newcastle and particularly my flatmates and ex-flatmates for the moral support throughout the years and for making life in cold Newcastle fun: Thanks to Elena, Ferran (or rattata), Gaby, Mimi and Selene and all of other fellow PhD students at Newcastle which we've spent time together. Finally, many thanks to my family: Manolis, Nana, Pantelis and my grandmother (Eutuchia) who have been supporting me all these years.

I'd like to especially thanks Dr. James Kitson member of my group and supervisor for his support in the lab and bioinformatics who provided many hours of advice, assistance and troubleshooting. Thanks to my supervisor Dr. James Bell for his advice and feedback and for assisting enormously during the sample collection at Rothamsted. Thanks to Dr. Larissa Collins, for her help throughout the project and for managing the collection process of samples at Fera. I would also like to thank all of my other supervisors: Dr. Ramiro Morales-Hojas, Dr. Neil Boonham and Dr. Dave Lunt for all their help during this project. Finally, many thanks to all the taxonomists and researchers at Rothamsted (thanks Chris Shortall) and Fera (thanks to Jackie Dunn) who have been collecting and identifying thousands of insects all of these years, which without it wouldn't be possible for my project to take place.

Finally, my sincere thanks to Professor Darren Evans my main supervisor for all of his help throughout all aspects and problems that came up during my PhD, his feedback helped me greatly in improving many aspects of my PhD project. He has provided enormous support both professionally and personally on many occasions.

Table of Contents

Contents

List of Figures

List of Tables

Chapter 1: General introduction

1.1 Insects and insect monitoring in the UK

Insects represent one of the most diverse groups of animals on Earth with estimates ranging from 1.5-5.5 million species, but many taxa remain poorly studied, which is concerning given their importance for many ecological processes such as pollination, pest regulation, decomposition, and cultural services. In the United States the estimated value of such services from 'wild' insects alone reaches $57 billion (Zhang *et al.*, 2007; Ekström and Ekbom, 2011; Fijen *et al.*, 2018). At the same time, they represent one of the most 'endangered' taxa with their number plummeting likely due to increased anthropogenic disturbances such as habitat change (Hanski *et al.*, 2007) and climate change (Lister and Garcia, 2018). Moreover, fewer than 1% of the 1.4 million described invertebrate species have been assessed by the IUCN, but of those that have ~40% are considered threatened (Dirzo *et al.*, 2014). Indeed, insects are embedded in complex networks of ecological interactions (Pocock, Evans and Memmott, 2012) of which we know even less. Case studies highlighting substantial insect declines around the world have raised alarm, with a recent meta-analysis showing an average decline of terrestrial insect abundance of ~9% per decade (though an apparent increase in freshwater insect abundances (van Klink *et al.*, 2020). Yet it is clear that the state of insect diversity remains unknown for many parts of the world, as the geographical distribution of the studies, but also datasets available for insect populations (e.g. Global Biodiversity Information Facility - GBIF) are highly skewed in their distribution (Rocha-Ortega, Rodriguez and Córdoba-Aguilar, 2021).

Understanding how insect diversity and populations change throughout time requires long-term data. Such long-term data typically originate from long-term monitoring schemes themselves (Thomas, 2005). Insect monitoring can serve a variety of purposes as the aim of a monitoring scheme might be from pest monitoring and surveillance (including invasive species) to biodiversity assessment that is driven by national and international policies and frameworks; for example the recently established pollinator monitoring scheme in the UK (POMS- https://ukpoms.org.uk/). The original purposes of insect biomonitoring schemes are

varied and often target particular groups (e.g. agricultural pests) and at certain stages of their life-cycle (e.g. migration) or group of taxa characterised as biodiversity indicators or surrogates. Identification is typically done via traditional morphological means by trained taxonomists. In the UK insect monitoring has a long history and has been characterised as the most comprehensive and represents some of the longest-established efforts to monitor insect populations when compared to other countries. Many of the current schemes date back to the start of the 1960s (Bratton, 1991). There are numerous schemes at a national level focused on invertebrate monitoring (Morecroft *et al.*, 2009) many of which are also driven by volunteers. Briefly, some of the most important monitoring schemes in the UK which use passive sampling are: The National Moth Recording Scheme, launched in 2007 yet with similar schemes running from 1967-1982 (Butterfly Conservation), The Rothamsted Insect Survey which compromise of two networks: fifteen 12.2-meter high suction traps and a light-trap network recording univoltine moths, both started around 1964, with 15 traps already working by the early 70s. Another example is the UK Environmental Change Network (ECN) where measurements are aimed at moths, ground beetles and spiders by using light traps and pitfall traps respectively (Morecroft *et al.*, 2009). Some of the best datasets of insect time-series come from the Rothamsted Insect Survey (RIS) and Butterfly Monitoring Scheme (UKMBS), that have been crucial for understanding long-term insect population trends (Conrad *et al.*, 2006; Bell *et al.*, 2015), as well as correlatively identifying the major threats to insect biodiversity (Wagner *et al.*, 2021). However, as the focus of the schemes is applied (for example for pest monitoring) the breadth of taxa for which such time series are available is rather small. For example, in RIS only moths and aphids are identified to the lowest possible taxonomic level as they are the main focus taxa of the scheme. Yet, for many such schemes and particularly the ones that use passive sampling techniques, considerable non-target insect 'bycatch' is routinely collected and stored (see Hribar, 2020) but excluded from analyses due to lack of resources and/or expertise or discarded outright. Therefore, a lot of potential insect bycatch data remain unexamined or even thrown away, which is a loss of much-needed potential data given the lack of understanding on insect declines (Spears and Ramirez, 2015)

There can be many barriers in the development of more comprehensive insect monitoring schemes, particularly ones that could include bycatch diversity. Most of which can be

narrowed down to taxonomical, logistical and financial reasons (Montgomery *et al.*, 2021), 2021). Taxonomical impediments derive partially from the diversity of insects, with only 1 million species described to date and with most still undiscovered given that estimated numbers range from 5 to 30 million (Stork, 2017). Indeed, the sheer diversity along with the numerous niches they occupy makes them difficult to monitor, something known as the 'Linnean' and 'Hutchinsonian' shortfalls (Cardoso *et al.*, 2011). Insects pests in agriculture, despite being less diverse, can also be difficult to identify. In the case of aphids for example, one of the most important pests in the temperate world, taxonomic identification is impossible for some species without knowledge about the host (Bell *et al.*, 2015). Pest predators and beneficial insects are hard to distinguish as many of them are small and inconspicuous and usually require rigorous taxonomic knowledge. Particularly difficult are hymenopteran parasitoids, with their cryptic diversity and their obscure life histories (Smith *et al.*, 2008). Such diverse groups with high intra-specific complexes and subspecies make such taxonomic hindrances impossible to overcome without the use of molecular techniques (Blackman and Eastop, 2000). Finally, practical impediments such as the breadth of taxonomic expertise needed, the financial cost of doing it, the space requirements for archiving samples and the logistics of identifying and processing thousands of insect specimens can make it impractical for schemes to do this for all the samples, especially, when the aim of the scheme is to monitor a particular group of insects. However, recent advances in Next Generation Sequencing (NGS) can overcome many of such obstacles by providing scalable and cost-efficient methods that do not require taxonomic knowledge, enabling the processing of thousands of samples and species and thus revolutionizing the ways in which biodiversity can be monitored (van Klink *et al.*, 2022).

This thesis focuses on how NGS can significantly advance insect monitoring schemes by scaling up the diversity of insects being monitored with the inclusion of bycatch, as well as creating time series of general insect biodiversity data by examining archival samples. It also considers how merging NGS and the large spatial scales of insect samples as part of monitoring schemes can help to identify patterns of insect diversity throughout the UK. This introduction will present how DNA based approaches have been used to date in the context of insect monitoring, present technical considerations of DNA-metabarcoding that are

relevant to insect monitoring, and finally highlight uses of such tools to date as well as potential future directions.

1.2 From DNA barcoding to metabarcoding: DNA based tools for monitoring

The use of genetic approaches to identify species has a long history, yet most of it can be traced back as 'DNA barcoding' to the landmark study by Hebert *et al.*, (2003) on using DNA barcodes for species identification and delimitation. DNA barcodes are 'short, standardized genetic markers used for the taxonomic identification of isolated specimens'. With the advent of barcoding approaches, new insights on species diversity were unveiled such as the cryptic diversity of insect species (Hebert *et al.*, 2004). Despite overcoming many taxonomic challenges using DNA barcoding methods, a large challenge was scaling up the number of individuals and species one could process. Coupled with significant costs of Sanger Sequencing platforms, the routine use of barcoding approaches was not possible. It was the advent of Next-Generation Sequencing approaches, the sudden drop in sequencing costs led by market competition and the diminishing of specialist knowledge required to identify species *en masse* that made this possible. The term DNA metabarcoding was first coined in 2012 (Taberlet *et al.*, 2012) however the first study can be dated back to 1990 (Giovannoni *et al.*, 1990). Metabarcoding refers to 'simultaneous DNA-based identification of many taxa found in an environmental sample'; the environmental sample can derive from water, soil, sediments, faeces or bulk samples from traps (Taberlet *et al.*, 2018). It has gained popularity as a tool to monitor biological communities ranging from bacteria to megafaunas (Cristescu, 2014). Metabarcoding of terrestrial arthropods was pioneered by Yu *et al.*, (2012) where they showed that it is possible to DNA metabarcode complete samples from malaise traps, in this instance with more than 300 species and recovering enough taxonomic information to give α and β-diversity estimates. Although the species recovered from metabarcoding had lower resolution than the ones from sanger sequencing (16 and 35 species respectively) the study established the possibilities that arise from such approaches for arthropod biodiversity screening. Thus, metabarcoding can be used to reliably identify pest complexes (e.g. aphids)

which are otherwise impossible to identify without the use of DNA identification. In a larger study, Liu *et al.*, (2013) compared the applicability of using metabarcoding to address policy and management questions while also comparing efforts and costs for the taxonomic identification of species based on standard approaches and metabarcoding. Cost effectiveness was much higher for metabarcoding, although associated costs for standard approaches were not presented. Recent efforts suggest that it is now possible to increase the comprehensiveness of arthropod assessments and for policy and decision-making not to be strictly limited on 'surrogates' of biodiversity such as biodiversity indicators like beetles, ants and bees (Barsoum *et al.*, 2019).

1.2.1 DNA-metabarcoding and technical considerations

The effectiveness of metabarcoding can be broken down to three main steps: 1) study design and sampling, 2) sample processing in the lab and 3) bioinformatics analysis. Here I briefly give some considerations of the metabarcoding workflow with a particular focus on insects. For specific technical details and general considerations see (Liu *et al.*, 2020) and (Deiner *et al.*, 2017) respectively. In the study design and sampling the study goal needs to be established as well as the target taxa. The type of sample in pest monitoring schemes can be either individual samples, bulk samples, or trap media. It is important for these to be defined *a priori*, and a suitable preservation method to be chosen that would minimize contamination and DNA degradation as the longer the samples stay in the field the higher the degradation rates (Krehenwinkel *et al.*, 2018). Processing in the laboratory also needs careful development as it can have a direct impact on the downstream steps of the molecular workflow. Care is needed for DNA extraction methods depending on the purpose of the study as the DNA yield and the subsequent community comparisons can be affected by the method of choice (Majaneva *et al.*, 2018). Extraction methods can be destructive or non-destructive (Carew, Coleman and Hoffmann, 2018), something of particular importance for museums and pest monitoring schemes which archive their samples and wish to retain them for further use (e.g. RIS samples). Non-destructive methods are also appropriate to enrich reference databases such as BOLD (Ratnasingham and Hebert, 2007). The preservative or trap media has also been used as an alternative to non-destructive tissue extraction, the idea being that taxa within a

trap media can leach DNA into the trap media (Shokralla, Singer and Hajibabaei, 2010). As this method requires less processing time it is particularly interesting for monitoring schemes where fast identification times are needed (Zizka *et al.*, 2018; Kirse *et al.*, 2023). However, results to date show that recovered communities from preservatives like Ethanol differ significantly from the ones recovered from tissue based DNA extraction and they can be more prone to contamination (Chimeno *et al.*, 2022). An alternative is non-destructive tissue based extraction where samples are digested in a media for a short period of time (Martoni *et al.*, 2022; Kirse *et al.*, 2023). This again can have biases towards certain taxa as smaller and softer organisms tend to leach out their DNA faster when compared with taxa that have high sclerotization, but this approach is much less prone to contamination and gives similar results to tissue homogenization based DNA extraction (Kirse *et al.*, 2023). All DNA-metabarcoding studies rely on enriching or amplifying a mixture of DNA-templates from the DNA extracts. This is done to create multiple copies of the targeted DNA templates. It is typically done by the polymerase chain reaction (PCR) where through cycles of raising and lowering the temperature in a machine multiple copies of target template can be created chemically using reagents such as polymerase enzymes, primers and nucleotides. Primers are short fragments of DNA typically between 18-25 base pairs long that attach to a DNA molecule where the polymerase enzyme starts “copying” the DNA strand creating a new one. This process is repeated many times, which results in an exponential increase of the target template (Liu *et al.*, 2020). Another approach is hybridisation enrichment, here fragmented DNA is hybridised to baits that are complementary to regions of interests (Mariac *et al.*, 2018). Both of these approaches require a target region within genes to be amplified which is achieved by the primers. In DNA-barcoding and metabarcoding of insects, the most common targeted genes are mitochondrial such as: Cytochrome c oxydase I (COI) and Cytochrome c oxydase II (COII) due to their high inter and intra species divergence, or ribosomal such as 16s rRNA. Primers have been developed to be able to amplify DNA from as many taxa as possible by increasing the degeneracy of the primer (Elbrecht, Hebert and Steinke, 2018).

The primer and loci of choice can generate biases on the amplification and sequencing step which can be irreversible. Some primers and loci will amplify certain taxa while others can have very low PCR success rate e.g. for hymenoptera by using COI primers (Yu *et al.*, 2012). Variation in amplification efficiency will lead to some taxa being observed disproportionately.

For primers the specificity vs. coverage also needs to be considered, degenerate primers will give a more comprehensive coverage of the total community, while primers developed for certain taxa will increase higher level taxonomic resolution (i.e. to the species level) (Bálint *et al.*, 2018). From the two most dominant loci (Cytochrome oxydase 1 (COI) and 16s rRNA (16s)) there are many advantages and disadvantages with debates on what is more suitable for metabarcoding insects (Clarke *et al.*, 2014). 16s is considered better if the aim is wide taxa coverage. There are also taxa that are known to have problems amplification with COI like hymenopteran beneficials such as parasitoids and wild bee pollinators (Piper *et al.*, 2021). Yet, there are certain characteristics of COI that make it the dominant marker to date. One is higher intraspecific variation thus, if species-level identification is needed then COI should be the marker of choice. Additionally, COI has more complete reference databases (Deagle *et al.*, 2019). As no perfect primer or marker exists (certain primers amplify specific taxa better than others) it is important to consider marker's choice at the initial steps of the study. Equally important is to optimize and test different primers in silico with software such as PrimerMiner (Elbrecht and Leese, 2017) or EcoPrimer (Riaz *et al.*, 2011). For example, if a particular group of taxa is the target of the study (like aphids). Primers can be evaluated by checking in silico if DNA sequences from the target group can be amplified by that primer. Additionally, the number of DNA sequences (corresponding to species) for the loci of choice is an important factor determining successful primer evaluation . It is also important to include positive and negative controls and try to minimize contaminations as much as possible e.g. by having pre-PCR and post-PCR laboratory spaces. By using a mineral oil vapour barrier in the PCR mastermix, sequencing negative and positive controls can increase quality control of the study (Kitson *et al.*, 2019). The sequencing platform to be used should be defined in the study design, comparisons of sequencing experiments between Illumina MiSeq, Ion Torrent PGM and Ion Torrent S5, yielded similar results but currently Illumina MiSeq can give higher quality reads (Braukmann *et al.*, 2019). Additionally high output platforms like the Illumina NovaSeq can uncover many more taxa as the sequencing depth achieved can be many times higher than the most common platform used to date: the Illumina MiSeq (Singer *et al.,* 2019). There are also emerging sequencing technologies like the Oxford Nanopore Sequencers offering real time sequencing of long-reads (Krehewinkel *et al*., 2019). Taxonomic identification of DNA sequences consists of comparing known or unknown sequences from the experiment against a reference database containing annotated taxonomic information for sequences (Hleap *et*

al., 2021). Reference databases can contain thousands of taxonomically annotated sequences for organisms coming from sequencing experiments and in some cases such sequences can correspond to single voucher specimens (such as the BOLD database) (Ratnasingham and Hebert, 2007). Species or sequences that are closely related are characterized as Molecular Operational Taxonomic Units (MOTUs), these are identified by using cluster algorithms and a predefined percentage sequence similarity threshold (Ryberg, 2015). A further build-up on MOTUs is the Barcode Index Numbers system defined as: 'a species-level taxonomic registry based on the analysis of patterns of nucleotide variation in the barcode region of the COI gene' (Ratnasingham and Hebert, 2013). Finally, there are also other methods where clustering is not applied and Amplicon Sequence Variants or Exact Sequence Variants are inferred from the dataset, which can capture diversity at the haplotypic level (Callahan *et al.*, 2016). There are four main categories for taxonomic assignment methods: 1) methods based on sequence similarity (like Blast (Camacho *et al.*, 2009); 2) methods based on sequence composition (like RDP classifiers which is a classification algorithm based on a Naive Bayesian model (Wang *et al.*, 2007)); 3) phylogenetic methods (based on evolutionary placement algorithms (Czech *et al.*, 2022)) and 4) probabilistic methods (Hleap *et al.*, 2021). Many comparisons for the methods have been made to date (Hleap *et al.*, 2021) and on many occasions sequence similarity and composition methods fare better than other approaches and have been the most commonly used in DNA-metabarcoding studies. Finally, one of the major issues with metabarcoding is gaining quantitative and accurate information about the species abundances in the samples (Lamb *et al.*, 2019), something that is of particular importance for monitoring programs. To date, NGS methods that estimate species abundance are based on the assumption that sequenced reads correlate with the initial input of DNA, thus if biomass of each species in the bulk mixture were known in advance then an estimate of the number of each specimen per taxonomic unit could be inferred (Gueuning *et al.*, 2019). However, differential species detection rates from biases in PCR amplification (e.g. primer biases), starting material (e.g. some species will have more mitochondrial DNA than others, or the DNA itself can be more degenerate) and sequencing errors, all obscure the accuracy of such estimates (Deagle *et al.*, 2019). It is suggested that mitogenomics and by using PCR-free methods, reliable estimation of abundance in bulk samples is possible (Gomez-Rodriguez *et al.,* 2015; Bista *et al.*, 2018), but such methods are still more costly than metabarcoding for routine use, can have low coverage of target sequences and high false positives leading to

different community compositions when compared with morphological datasets (Gueuning et al., 2019). For now, metabarcoding can be a very robust tool for presence/absence data with bulk samples which can depict similar ecological patterns to traditional datasets. However, there are ways of making such datasets quantitative by the use of correction factors (Thomas *et al.*, 2016) or by spiking in known amounts of DNA (Deagle *et al.*, 2019; Ji *et al.*, 2020). The short barcodes from Illumina reads can also be a limiting factor to the accuracy of the taxonomic assignment, particularly at low taxonomic levels and for complex communities such as insects. With the ongoing improvement of third-generation long read sequencing platforms such as PacBio SMRT-seq and Oxford Nanopore (MinION) this will lead to long-read scalable DNA-metabarcoding which can help to identify otherwise difficult taxonomic complexes (Srivathsan *et al.*, 2021) and develop tools for monitoring in the field.

1.3 Monitoring insects with DNA-metabarcoding

Insect monitoring is still largely conducted using conventional morphological identification, and the adoption of DNA-metabarcoding has yet to be implemented within national insect monitoring schemes. In comparison, for example in the case of freshwater systems, environmental DNA (eDNA) approaches have been approved for use in monitoring (for example: the Great Lakes eDNA Monitoring Program) (Jerde *et al.,* 2013). Yet the complexity of terrestrial systems and the difficulty of implementing changes required for DNA-based monitoring within already established schemes complicate the adoption of such methods. Despite DNA-metabarcoding being used and accepted widely as a tool for scalable biodiversity identification, its usability within insect monitoring has been mainly discussed in review articles (van Klink *et al.*, 2022). Below I provide evidence to date where DNA-metabarcoding has been applied for invertebrate biodiversity assessment or monitoring within agricultural or natural systems, highlighting its potential as a tool for monitoring but also for ecological research more generally.

1.3.1 Pollinators

Pollination has a major effect on global food systems and is directly related to food production. Pollination can account up to 8% of the global food production and it has an estimate worth that ranges from 235$ to 577$ billion globally each year (Potts *et al.*, 2016). Additionally, non-

bee pollinators can also have a big impact on local food production but the value of many non-bee pollinator species is not well documented (Panziera *et al.*, 2022). Monitoring pollinators is an urgent task given the recent reports on pollinator losses and declines both globally and within Britain (Powney *et al.*, 2019). To date, most studies on pollinators focused on metabarcoding the pollen or deciphering mutualistic interactions within agro-ecosystems and not the insects per se (see Pornon *et al.*, 2016; Jones *et al.,* 2022). DNA-metabarcoding can be a good tool to understand species-interactions between insects and plants, which are time-consuming to do so by traditional morphology, and it can lead to the uncovering of interactions between taxa that have been previously unknown even for very common plants (Evans and Kitson, 2020; Lowe *et al.,* 2022). Gueuning et al. (2019) assess three different methods of Next-Generation Sequencing from a monitoring program of wild bees. The methods were: metabarcoding, metagenomics and NGS barcoding where individual information is retained. Metabarcoding did yield similar ecological patterns when based on presence/absence data, more importantly in this context it was presented as the most scalable tool particularly at high sample sizes, which supports that samples from monitoring schemes can be reliably identified with metabarcoding. Yet, metabarcoding is not limited to species identification. DNA-metabarcoding can also be used to identify the effect of anthropogenic disturbances such as habitat fragmentation on pollinator diversity and interactions (Tommasi *et al.*, 2022). Additionally, biological events such as pollinator migrations can also be monitored with metabarcoding, giving insights into phenomena such as transcontinental migrations from Africa to Europe, which might be more regular than previously thought (Suchan *et al.*, 2019).

1.3.2 Parasitoids

Another group of beneficial species particularly important for pest control are parasitoids. Parasitoids have a direct influence in decision-making (for example on timings of pesticide spraying) and constitute an integral part of Integrated Pest Management aimed to minimise the use of pesticides by maximising effective pest control (Zhang and Swinton, 2009). To date there is an 'ecological' need for developing tools that monitor pest predator populations and establishing population thresholds under which pesticide applications can be avoided, especially when current thresholds in the UK have been criticised as 'outdated' (Leather and

Atanasova, 2017). The use of DNA barcoding for the identification of parasitoids has long been appreciated (Hrček and Godfray, 2015), however as with all such methods the amount of samples that can be processed is minimal. DNA metabarcoding helps in reliably scaling up the number of individuals processed (Sigut *et al.*, 2017). It is also possible to estimate parasitism rates of crop pests by linking sequence reads to abundances in the bulk DNA mixture, although it is very likely that these over-estimate parasitism rates (Sow *et al.*, 2019). Furthermore, metabarcoding has the potential to uncover cryptic parasitoid species and multi-parasitism, which is difficult to achieve with rearing or even standard DNA barcoding (Sow *et al.,* 2019). Despite all such recent efforts focused on beneficials, the application of metabarcoding with a focus on pest monitoring en masse has being scarce (Batovska *et al.*, 2018) (see also below). But, as with new and developing methods there is continuous effort to validate and create tools for pest monitoring (Miller, Polaszek and Evans, 2021).

1.3.3 Rapid diagnostics of invading species

Invasive insect species and quarantine pests are monitored routinely throughout Europe, mainly using specific trapping techniques (e.g pheromone traps). Then, they are identified by appropriate personnel, including occasionally by using DNA barcoding as they can be indistinguishable by morphological characteristics from closely related species (Augustin *et al.*, 2012). Invasive species can have dramatic impact on the environment and native fauna and flora, however their effects can remain elusive before real damage has already occurred. The time lag between invasion, establishment and identification makes it particularly difficult to predict their effects (Simberloff *et al.*, 2013). Thus, rapid detection and control is very important to avoid the associated ecological and economic costs (Simberloff *et al.,* 2013). In freshwater ecosystems eDNA approaches have been used for the surveillance of aquatic invasions (e.g. Jerde and Mahon, 2015) yet until recently the potential of eDNA surveillance for terrestrial invaders have only been discussed. Valentin *et al.*, (2018) are the first to present the use of eDNA for the rapid diagnostics of the non-native pest *Halymorpha halys* within an agricultural ecosystem. The crops/fruits eaten by this insect are collected all together and rinsed the sampled rinsed water is thenused as the source for the eDNA extraction. Based on a qPCR approachminimal amounts of trace DNA could be picked up for the target organism. Hardulak *et al.*, (2020) show how DNA-metabarcoding can be used to monitor both pests and

invasive species within a national park in Germany by identifying invasive species of pests for the first time in their sampling region. DNA-based approaches can be more sensitive when it comes to very small populations such as the starting colony of an invasive species and can lead to rapid diagnostics of pests or migratory events which help asses expansion risks (Comtet *et al.*, 2015). Finally, surveillance programs for invasive pests that are vectors of human diseases such as mosquitos require accurate and rapid diagnostics of both vector and virus. Metabarcoding can also be used in such cases as a sensitive tool to evaluate the potential risks from virus outbreaks (Batovska *et al.*, 2018).

1.3.4 Bycatch diversity

Bycatch refers to non-target taxa trapped or collected unintentionally along with the target taxa. It is not a well-established concept within entomology and this might derive from a non-universal use of the term 'bycatch' and the alternative versions that exists e.g. 'non-target organisms' (Spears and Ramirez, 2015). There are important scientific and ethical reasons for bycatch to be highlighted in the context of monitoring. Within insect monitoring schemes, there is a lot of bycatch although direct comparisons are scarce. Spears and Ramirez (2015) assessed the number of bycatch species within the Cooperative Agricultural Pest Survey (U.S.) aimed at surveying three invasive moth pests. They found for one sampling period at least 82 non-target species such as pollinators, moths, and natural enemies. The RIS bycatch has also been used by Pérez-Rodríguez, Shortall and Bell, (2015) by linking the Rothamsted metadata on the grain aphid *Sitobion avenae* to catches of its parasitoids from the suction traps. What usually refers to by-catch could be potential target organisms within monitored ecosystems such as natural enemies and pollinators, but they are rarely analysed as it is prohibitively expensive or not the focus of the study or monitoring scheme. Metabarcoding could potentially complement and lead to synchronous screening of pest and beneficial insects altogether within pest monitoring schemes (Kitson *et al.*, 2019; Sow *et al.*, 2019). Another issue are the ethical considerations of bycatch. Malaise traps have a lot of bycatch and could potentially have an unknown effect on species that are trapped (e.g. rare species) yet the sampling effect *per se* is not usually considered (Fischer and Larson, 2019). As a start, rapid species inventories could be established for monitoring schemes via bulk sample metabarcoding (Yu *et al.,* 2012), and with a bycatch species list, a more targeted and

comprehensive examination of the effects on particular taxa (rare species) could be assessed. RIS represents a rare example where all its bycatch is archived (Harrington, 2013) so it could be possible to look for biodiversity trends by creating time-series of the entire bycatch and at the same time look for anthropogenic effects such as pesticide applications. Such DNA based time-series are considered to have enormous potential for agro-ecological research and will help understand our footprint on such systems (Bálint *et al.*, 2018). Currently I have identified only two examples of monitoring scheme that use all of their trapped sampled taxa: The Global Malaise Network (http://biodiversitygenomics.net/projects/gmp/) represents a unique monitoring program aimed at identifying all arthropod diversity trapped within Mailaise traps and the creation of species lists and barcode references globally. It represents an example of how species not known to science can be incorporated into monitoring programs via BINs and OTUs (Morinière *et al.*, 2016). Second the DINA (Diversity of Insects in Nature protected areas where DNA-metabarcoding has been used to monitor insect communities in nature reserves in Germany (Lehmann *et al.*, 2021). There are specific guidelines for the consideration of bycatch and some authors are suggesting database creation and storing the samples so other researchers can make use of such samples (Buchholz *et al.*, 2011). As the application of DNA-metabarcoding and adoption of NGS for routine monitoring continues to grow. These tools will become vital for scaling up insect monitoring. However, as previously discussed, the validity of DNA-metabarcoding within already established monitoring schemes has been lacking and needs to be assessed.

1.4 Creating species-interaction networks using DNA-metabarcoding

I have so far discussed and presented how DNA-metabarcoding has been applied as a tool for scalable biodiversity assessment. Yet, DNA-metabarcoding can also uncover other layers of ecological information: ecological networks and their species interactions (Kitson *et al*., 2019). Coupling long-term archival samples from monitoring schemes recovered by DNA-metabarcoding with advances in network science can help move towards a more mechanistic understanding of insect declines and the cascading effects on biodiversity and ecosystem function. Ecological networks describe the interactions between species, the underlying structure of communities and the function and stability of ecosystems (Montoya, Pimm and Solé, 2006). They are particularly well suited to the study of species loss and how this can

cause extinction cascades across ecosystems (Pocock, Evans and Memmott, 2012; Kehoe, Frago and Sanders, 2021). Whilst new network construction methods are gaining traction (Evans *et al.*, 2016; Staniczenko *et al.*, 2017) there is a dearth of understanding of the complex ways in which insects interact (Miller, Polaszek and Evans, 2021), and long-term species-interaction datasets are scant, which currently limits our ability to use networks predictively (Raimundo, Guimarães and Evans, 2018). Insect monitoring schemes, especially those with preserved sample archives, hold considerable potential for the construction of highly-resolved, long-term ecological networks, which can then be uniquely used to examine the impacts of environmental change on network structure, complexity and robustness (a measure of the tolerance to species extinctions) across scales, the impacts on ecosystem functioning (especially pollination, pest regulation) and new restoration methods.

1.5 Summary, thesis aims and structure

NGS approaches are revolutionising the ways insect communities can be monitored. They offer scalable and cost-efficient approaches to identify ecological communities and their patterns. To date, insect monitoring schemes mainly focus on certain target taxa, yet they capture a diverse range of insect taxa which remain unknown: the bycatch. Applications of NGS approaches on samples from insect monitoring schemes are scarce and it is unclear to managers of schemes what the added value is. Here the overarching hypothesis is that DNA metabarcoding can accurately describe insect diversity, both from temporally stored archive samples and from spatially distributed nationwide schemes. The thesis focuses on the 'added value' that DNA-metabarcoding can offer to on-going UK insect monitoring schemes: the Rothamsted Insect Survey (RIS) and the Fera Yellow Water Pan Trap network (YWP), both of which focus on monitoring aphids but have considerable bycatch. I do this by first assessing the validity of DNA-metabarcoding to recover already identified archival samples of insects (both target taxa and bycatch) from a pest monitoring scheme. I then focus on the scalability of the diversity of insects being monitored with the inclusion of bycatch. Finally, I consider how merging NGS and the large spatial scales of pest monitoring can help to identify patterns of bycatch insect diversity throughout the UK.

Evidence of the use of DNA-metabarcoding for identifying archival samples is scarce, particularly in the context of insect monitoring. Therefore, there is a need for validation of such tools. In Chapter 2, I assess the validity of DNA-metabarcoding to uncover already identified archived samples from RIS. I focus on recovering a 16-year time-series of aphid samples that has been identified by RIS and compare the datasets coming from metabarcoding and morphology to assess the congruence between the two. As the morphological integrity of archival samples is important for the future usability and accessibility of such samples, I evaluate non-destructive and destructive methods of DNA-extraction. I identify and discuss potential influences of different factors for the congruence of morphology and metabarcoding. Overall, the aim of the chapter is to validate DNA-metabarcoding as a potential tool for accessing archival samples which would open up thousands of archival samples of insects within RIS, allowing researchers to retrospectively create time-series of insects and understand how their populations change with time.

After validating DNA-metabarcoding as a tool for recovering target (aphid) archival communities, in Chapter 3 I focus on the bycatch diversity within RIS. The scheme has been archiving all their samples, bycatch or not, for decades. Identifying bycatch via morphological means can be very difficult. The breadth of taxa recovered and the sheer number of insect specimens within the samples adds complexity to its identification as sorting of specimens can be very time consuming and the taxonomic expertise needed to identify complex samples at low taxonomic levels such as genera or species might not be available. DNA-metabarcoding has the potential to overcome such challenges. I use DNA-metabarcoding to uncover the bycatch diversity within RIS by using a 16-year time series of bycatch samples. I evaluate a non-destructive method and provide for the first time a taxa list for bycatch taxa in RIS at a species level and identify potential taxa of interest in agriculture, such as pollinators, pests and predators, highlighting the importance of bycatch species for increasing the breadth of taxa being monitored. I show the value of bycatch taxa as ecological data as it is possible to use them to create time series for hundreds of insect taxa for which their insect populations remain unknown. Finally, I show how bycatch taxa not only include organisms of agricultural interest, but also show how it can include diversity previously unrecorded in the geographical regions sampled, highlighting its use for an early warning system for potential invasive species.

In Chapter 4 the focus remains on bycatch diversity. After validating DNA-metabarcoding as a tool and highlighting bycatch diversity and its potential uses, I further build up on this by showing how DNA-metabarcoding can be used at large spatial scales to uncover diversity patterns nationwide. I use the Fera Yellow Water Pan trap network (YWP), a network of yellow pan traps across the UK that is used to monitor aphids within farms. Here, bycatch is typically discarded, but I show how it is possible to retain a digital archive of it by using DNA-metabarcoding. I show how DNA-metabarcoding can uncover bycatch diversity from sampling sites across the UK with more than 83 locations. A further major limitation for NGS based studies is the sampling effort per sample, typically measured in sequencing depth. In this chapter I also assess the differences between sequencing platforms with very different data generating capabilities. This way Iindirectly assessthe influence of different sequencing depths for the recovery of insect communities at large spatial scales. As money can be an important factor when choosing the sequencing platform, particularly in the context of insect monitoring appropriate sequencing depths need to be achieved for the recovered community to be a close reflection of the actual communities. Finally, I show how DNA-metabarcoding can be used to analyse patterns of diversity both spatially and temporally nationwide within the UK and highlight the wide diversity of bycatch found despite sampling in an agricultural monoculture.

Lastly, in Chapter 5 I summarise the work and discuss further implications of this thesis and future directions for the use of DNA-metabarcoding within insect monitoring schemes. I discuss how merging the sampling from large spatial and temporal scales that takes place in monitoring schemes can be coupled with NGS approaches and other approaches based on network ecology to understand not only how insect communities change over time but also changes in their interactions which can hopefully give us a more comprehensive understanding of insect declines.

1.6 References

Augustin, S. *et al.* (2012) 'A list of methods to detect arthropod quarantine pests in Europe*', *EPPO Bulletin*. John Wiley & Sons, Ltd, 42(1), pp. 93–94. doi: 10.1111/J.1365-2338.2011.02523.X.

Bálint, M. *et al.* (2018) *Environmental DNA Time Series in Ecology*, *Trends in Ecology and Evolution*. doi: 10.1016/j.tree.2018.09.003.

Barsoum, N. *et al.* (2019) 'The devil is in the detail: Metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity', *Ecological Indicators*, 101, pp. 313–323. doi: 10.1016/j.ecolind.2019.01.023.

Batovska, J. *et al.* (2018) 'Effective mosquito and arbovirus surveillance using metabarcoding', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 18(1), pp. 32–40. doi: 10.1111/1755-0998.12682.

Bell, J. R. *et al.* (2015) 'Long-term phenological trends, species accumulation rates, aphid traits and climate: five decades of change in migrating aphids', *Journal of Animal Ecology*. Edited by K. Wilson. John Wiley & Sons, Ltd (10.1111), 84(1), pp. 21–34. doi: 10.1111/1365-2656.12282.

Bista, I. *et al.* (2018) 'Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples', *Molecular Ecology Resources*, pp. 0–2. doi: 10.1111/1755-0998.12888.

Blackman, R. L. and Eastop, V. F. (no date) 'Taxonomic issues.', in *Aphids as crop pests*. Wallingford: CABI, pp. 1–36. doi: 10.1079/9781780647098.0001.

Bratton, J. . (1991) 'British red data books: part 3: invertebrates other than insects', *British red data books*.

Braukmann, T. W. A. *et al.* (2019) 'Metabarcoding a Diverse Arthropod Mock Community', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111). doi: 10.1111/1755-0998.13008.

Buchholz, S. *et al.* (2011) 'Bycatches of ecological field studies: Bothersome or valuable?', *Methods in Ecology and Evolution*, 2(1), pp. 99–102. doi: 10.1111/j.2041-210X.2010.00051.x.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/1471-2105-10-421/.

Cardoso, P. *et al.* (2011) 'The seven impediments in invertebrate conservation and how to overcome them', *Biological Conservation*, 144(11), pp. 2647–2655. doi: 10.1016/j.biocon.2011.07.024.

Carew, M. E., Coleman, R. A. and Hoffmann, A. A. (2018) 'Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding?', *PeerJ*, 6, p. e4980. doi: 10.7717/peerj.4980.

Chimeno, C. *et al.* (2022) 'Depicting environmental gradients from Malaise trap samples: Is ethanol-based DNA metabarcoding enough?', *Insect Conservation and Diversity*. John Wiley & Sons, Ltd. doi: 10.1111/ICAD.12609.

Clarke, L. J. *et al.* (2014) 'Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 14(6), pp. 1160–1170. doi: 10.1111/1755-0998.12265.

Comtet, T. *et al.* (2015) 'DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens', *Biological Invasions 2015 17:3*. Springer, 17(3), pp. 905–922. doi: 10.1007/S10530-015-0854-Y.

Conrad, K. F. *et al.* (2006) 'Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis', *Biological Conservation*. doi: 10.1016/j.biocon.2006.04.020.

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

Czech, L. *et al.* (2022) 'Metagenomic Analysis Using Phylogenetic Placement—A Review of the First Decade', *Frontiers in Bioinformatics*. Frontiers Media SA, 2, p. 44. doi: 10.3389/FBINF.2022.871393/BIBTEX.

Deagle, B. E. *et al.* (2019) 'Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?', *Molecular Ecology*, 28(2), pp. 391–406. doi:

10.1111/mec.14734.

Deiner, K. *et al.* (2017) 'Environmental DNA metabarcoding: Transforming how we survey animal and plant communities', *Molecular Ecology*. John Wiley & Sons, Ltd, 26(21), pp. 5872–5895. doi: 10.1111/MEC.14350.

Dirzo, R. *et al.* (2014) 'Defaunation in the Anthropocene', *Science*, 345(6195), pp. 401–406. doi: 10.1126/science.1251817.

Ekström, G. and Ekbom, B. (2011) 'Pest Control in Agro-ecosystems: An Ecological Approach', *https://doi.org/10.1080/07352689.2011.554354*. Taylor & Francis Group , 30(1–2), pp. 74–94. doi: 10.1080/07352689.2011.554354.

Elbrecht, V., Hebert, P. D. N. and Steinke, D. (2018) 'Slippage of degenerate primers can cause variation in amplicon length', *Scientific Reports*. Nature Publishing Group, 8(1), p. 10999. doi: 10.1038/S41598-018-29364-Z.

Elbrecht, V. and Leese, F. (2017) 'PrimerMiner: an r package for development and in silico validation of DNA metabarcoding primers', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 8(5), pp. 622–626. doi: 10.1111/2041-210X.12687.

Evans, D. M. *et al.* (2016) 'Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems', *Functional Ecology*. Edited by T. Poisot, 30(12), pp. 1904–1916. doi: 10.1111/1365-2435.12659.

Evans, D. M. and Kitson, J. J. (2020) 'Molecular ecology as a tool for understanding pollination and other plant–insect interactions', *Current Opinion in Insect Science*. Elsevier, 38, pp. 26–33. doi: 10.1016/J.COIS.2020.01.005.

Fijen, T. P. M. *et al.* (2018) 'Insect pollination is at least as important for marketable crop yield as plant quality in a seed crop', *Ecology Letters*. John Wiley & Sons, Ltd, 21(11), pp. 1704–1713. doi: 10.1111/ELE.13150.

Fischer, B. and Larson, B. M. H. (2019) 'Collecting insects to conserve them: a call for ethical caution', *Insect Conservation and Diversity*. John Wiley & Sons, Ltd, 12(3), pp. 173–182. doi: 10.1111/ICAD.12344.

Giovannoni, S. J. *et al.* (1990) 'Genetic diversity in Sargasso Sea bacterioplankton', *Nature*

1990 345:6270. Nature Publishing Group, 345(6270), pp. 60–63. doi: 10.1038/345060a0.

Gueuning, M. *et al.* (2019) 'Evaluating next-generation sequencing (NGS) methods for routine monitoring of wild bees: Metabarcoding, mitogenomics or NGS barcoding', *Molecular Ecology Resources*. Blackwell Publishing Ltd, 19(4), pp. 847–862. doi: 10.1111/1755-0998.13013.

Hanski, I. *et al.* (2007) 'Deforestation and apparent extinctions of endemic forest beetles in Madagascar', *Biology letters*. Biol Lett, 3(3), pp. 344–347. doi: 10.1098/RSBL.2007.0043.

Hardulak, L. A. *et al.* (2020) 'DNA metabarcoding for biodiversity monitoring in a national park: Screening for invasive and pest species', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 20(6), pp. 1542–1557. doi: 10.1111/1755-0998.13212.

Harrington, R. (2013) 'The Rothamsted Insect Survey strikes gold', *Antenna*, 38(3), pp. 158–166. Available at: www.rothamsted.ac.uk/examine.

Hebert, P. D. N. *et al.* (2003) 'Biological identifications through DNA barcodes', *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), pp. 313–321. doi: 10.1098/rspb.2002.2218.

Hebert, P. D. N. *et al.* (2004) 'Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 101(41), pp. 14812–14817. doi: 10.1073/PNAS.0406166101/.

Hleap, J. S. *et al.* (2021) 'Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 21(7), pp. 2190–2203. doi: 10.1111/1755-0998.13407.

Hrček, J. and Godfray, H. C. J. (2015) 'What do molecular methods bring to host-parasitoid food webs?', *Trends in Parasitology*, 31(1), pp. 30–35. doi: 10.1016/j.pt.2014.10.008.

Hribar, L. J. (2020) 'Expanding Basic Entomological Knowledge by Using Mosquito Surveillance Bycatch', *Annals of the Entomological Society of America*, 113(6), pp. 439–446. doi: 10.1093/aesa/saaa014.

Jerde, C.L, *et al.* (2013) 'Detection of Asian carp DNA as part of a Great Lakes basin-wide

surveillance program, *Canadian Journal of Fisheries and Aquatic Sciences*, 70(4). doi: 10.1139/cjfas-2012-0478.

Jerde, C. L. and Mahon, A. R. (2015) 'Improving confidence in environmental DNA species detection', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 15(3), pp. 461–463. doi: 10.1111/1755-0998.12377.

Ji, Y. *et al.* (2020) 'SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes', *Molecular Ecology Resources*, 20(1), pp. 256–267. doi: 10.1111/1755-0998.13057.

Jones, L. *et al.* (2022) 'Temporal Patterns of Honeybee Foraging in a Diverse Floral Landscape Revealed Using Pollen DNA Metabarcoding of Honey, *Integrative and Comparative Biology*, 62(2), pp. 199-210. doi: 10.1093/icac029

Kehoe, R., Frago, E. and Sanders, D. (2021) 'Cascading extinctions as a hidden driver of insect decline', *Ecological Entomology*. John Wiley & Sons, Ltd, 46(4), pp. 743–756. doi: 10.1111/EEN.12985.

Kirse, A. *et al.* (2023) 'Comparison of destructive and nondestructive DNA extraction methods for the metabarcoding of arthropod bulk samples', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 23(1), pp. 92–105. doi: 10.1111/1755-0998.13694.

Kitson, J. J. N. *et al.* (2019) 'Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding', *Molecular Ecology*, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

van Klink, R. *et al.* (2020) 'Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances', *Science*, 368(6489), pp. 417–420. doi: 10.1126/science.aax9931.

van Klink, R. *et al.* (2022) 'Emerging technologies revolutionise insect ecology and monitoring', *Trends in Ecology & Evolution*. Elsevier, 37(10), pp. 872–885. doi: 10.1016/J.TREE.2022.06.001.

Krehenwinkel, H. *et al.* (2018) 'The effect of DNA degradation bias in passive sampling

devices on metabarcoding studies of arthropod communities and their associated microbiota', *PLoS ONE*, 13(1), pp. 1–14. doi: 10.1371/journal.pone.0189188.

Krehenwinkel, H. *et al.* (2019) 'Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale, *GigaScience,* 8(5), doi: 10/1093/gigascience/giz006

Lamb, P. D. *et al.* (2019) 'How quantitative is metabarcoding: A meta-analytical approach', *Molecular Ecology*. John Wiley & Sons, Ltd, 28(2), pp. 420–430. doi: 10.1111/MEC.14920.

Leather, S. R. and Atanasova, D. (2017) 'Without up-to-date pest thresholds sustainable agriculture is nothing but a pipe-dream', *Agricultural and Forest Entomology*. John Wiley & Sons, Ltd, 19(4), pp. 341–343. doi: 10.1111/AFE.12244.

Lehmann, G. U. C. *et al.* (2021) 'Diversity of Insects in Nature protected Areas (DINA): an interdisciplinary German research project', *Biodiversity and Conservation*. Springer Science and Business Media B.V., 30(8–9), pp. 2605–2614. doi: 10.1007/S10531-021-02209-4.

Lister, B. C. and Garcia, A. (2018) 'Climate-driven declines in arthropod abundance restructure a rainforest food web', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(44), pp. E10397–E10406. doi: 10.1073/PNAS.1722477115/SUPPL_FILE/PNAS.1722477115.SD04.CSV.

Liu, M. *et al.* (2020) 'A practical guide to DNA metabarcoding for entomological ecologists', *Ecological Entomology*. John Wiley & Sons, Ltd, 45(3), pp. 373–385. doi: 10.1111/EEN.12831.

Liu, S. *et al.* (2013) 'SOAPBarcode: Revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons', *Methods in Ecology and Evolution*. Edited by D. Orme, 4(12), pp. 1142–1150. doi: 10.1111/2041-210X.12120.

Majaneva, M. *et al.* (2018) 'Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples', *Metabarcoding and Metagenomics*. doi: 10.3897/mbmg.2.26664.

Mariac, C. *et al.* (2018) 'Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms', *PLOS ONE*. Public Library of Science, 13(9), p. e0202976. doi: 10.1371/JOURNAL.PONE.0202976.

Martoni, F. *et al.* (2022) 'Disentangling bias for non-destructive insect metabarcoding', *PeerJ*. PeerJ Inc., 10, p. e12981. doi: 10.7717/PEERJ.12981.

Miller, K. E., Polaszek, A. and Evans, D. M. (2021) 'A dearth of data: fitting parasitoids into ecological networks', *Trends in Parasitology*. Elsevier Current Trends, 37(10), pp. 863–874. doi: 10.1016/J.PT.2021.04.012.

Montgomery, G. A. *et al.* (2021) 'Standards and Best Practices for Monitoring and Benchmarking Insects', *Frontiers in Ecology and Evolution*. Frontiers Media S.A., 8, p. 513. doi: 10.3389/FEVO.2020.579193/BIBTEX.

Montoya, J. M., Pimm, S. L. and Solé, R. V. (2006) 'Ecological networks and their fragility', *Nature 2006 442:7100*. Nature Publishing Group, 442(7100), pp. 259–264. doi: 10.1038/nature04927.

Morecroft, M. D. *et al.* (2009) 'The UK Environmental Change Network: Emerging trends in the composition of plant and animal communities and the physical environment', *Biological Conservation*. Elsevier, 142(12), pp. 2814–2832. doi: 10.1016/J.BIOCON.2009.07.004.

Morinière, J. *et al.* (2016) 'Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix', *PLoS ONE*. doi: 10.1371/journal.pone.0155497.

Panziera, D. *et al.* (2022) 'The Diversity Decline in Wild and Managed Honey Bee Populations Urges for an Integrated Conservation Approach', *Frontiers in Ecology and Evolution*. Frontiers Media S.A., 10, p. 93. doi: 10.3389/FEVO.2022.767950/BIBTEX.

Pérez-Rodríguez, J., Shortall, C. R. and Bell, J. R. (2015) 'Large-scale migration synchrony between parasitoids and their host', *Ecological Entomology*. John Wiley & Sons, Ltd (10.1111), 40(5), pp. 654–659. doi: 10.1111/een.12241.

Piper, A. M. *et al.* (2021) 'Computational Evaluation of DNA Metabarcoding for Universal Diagnostics of Invasive Insect Pests', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2021.03.16.435710. doi: 10.1101/2021.03.16.435710.

Pocock, M. J. O., Evans, D. M. and Memmott, J. (2012) 'The robustness and restoration of a network of ecological networks', *Science*, 335(6071), pp. 973–977. doi:

10.1126/science.1214915.

Pornon, A. *et al.* (2016) 'Using metabarcoding to reveal and quantify plant-pollinator interactions', *Scientific Reports*, 6. doi: 10.1038/srep27282.

Potts, S. G. *et al.* (2016) 'Safeguarding pollinators and their values to human well-being', *Nature 2016 540:7632*. Nature Publishing Group, 540(7632), pp. 220–229. doi: 10.1038/nature20588.

Powney, G. D. *et al.* (2019) 'Widespread losses of pollinating insects in Britain', *Nature Communications 2019 10:1*. Nature Publishing Group, 10(1), pp. 1–6. doi: 10.1038/s41467-019-08974-9.

Raimundo, R. L. G., Guimarães, P. R. and Evans, D. M. (2018) 'Adaptive Networks for Restoration Ecology', *Trends in ecology & evolution*. Trends Ecol Evol, 33(9), pp. 664–675. doi: 10.1016/J.TREE.2018.06.002.

Ratnasingham, S. and Hebert, P. D. N. (2007) 'bold: The Barcode of Life Data System (http://www.barcodinglife.org)', *Molecular Ecology Notes*. John Wiley & Sons, Ltd, 7(3), pp. 355–364. doi: 10.1111/J.1471-8286.2007.01678.X.

Ratnasingham, S. and Hebert, P. D. N. (2013) 'A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System', *PLoS ONE*. Edited by D. Fontaneto. Public Library of Science, 8(7), p. e66213. doi: 10.1371/journal.pone.0066213.

Riaz, T. *et al.* (2011) 'ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis', *Nucleic Acids Research*. Oxford University Press, 39(21), p. e145. doi: 10.1093/NAR/GKR732.

Rocha-Ortega, M., Rodriguez, P. and Córdoba-Aguilar, A. (2021) 'Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction', *Ecological Entomology*. John Wiley & Sons, Ltd, 46(4), pp. 718–728. doi: 10.1111/EEN.13027.

Ryberg, M. (2015) 'Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi', *Molecular Ecology*. John Wiley & Sons, Ltd, 24(23), pp. 5770–5777. doi: 10.1111/MEC.13444.

Shokralla, S., Singer, G. A. C. and Hajibabaei, M. (2010) 'Direct PCR amplification and

sequencing of specimens' DNA from preservative ethanol', *BioTechniques*. Biotechniques, 48(3), pp. 305–306. doi: 10.2144/000113362.

Sigut, M. *et al.* (2017) 'Performance of DNA metabarcoding, standard barcoding, and morphological approach in the identification of hostparasitoid interactions', *PLoS ONE*. Edited by M. Hajibabaei, 12(12), pp. 1–18. doi: 10.1371/journal.pone.0187803.

Simberloff, D. *et al.* (2013) 'Impacts of biological invasions: what's what and the way forward', *Trends in Ecology & Evolution*. Elsevier, 28(1), pp. 58–66. doi: 10.1016/J.TREE.2012.07.013.

Smith, M. A. *et al.* (2008) 'Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 105(34), pp. 12359–12364. doi: 10.1073/pnas.0805319105.

Sow, A. *et al.* (2019) 'Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding', *Scientific Reports 2019 9:1*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41598-019-40243-z.

Spears, L. R. and Ramirez, R. A. (2015) 'Learning to Love Leftovers', *American Entomologist*. Narnia, 61(3), pp. 168–173. doi: 10.1093/ae/tmv046.

Srivathsan, A. *et al.* (2021) 'MinION barcodes: biodiversity discovery and identification by everyone, for everyone', *bioRxiv*. BMC Biology, p. 2021.03.09.434692. Available at: https://doi.org/10.1101/2021.03.09.434692.

Staniczenko, P. P. A. *et al.* (2017) 'Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks', *Ecology Letters*, 20(6), pp. 693–707. doi: 10.1111/ele.12770.

Stork, N. E. (2017) 'How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth?', *Annual Review of Entomology*. Annual Reviews, 63(1), pp. 31–45. doi: 10.1146/annurev-ento-020117-043348.

Suchan, T. *et al.* (2019) 'Pollen metabarcoding as a tool for tracking long-distance insect migrations', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 19(1), pp. 149–

162. doi: 10.1111/1755-0998.12948.

Taberlet, P. *et al.* (2018) 'Environmental DNA: For Biodiversity Research and Monitoring', *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford University Press, pp. 1–253. doi: 10.1093/OSO/9780198767220.001.0001.

Thomas, A. C. *et al.* (2016) 'Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material', *Molecular Ecology Resources*, 16(3), pp. 714–726. doi: 10.1111/1755-0998.12490.

Thomas, J. A. (2005) 'Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 360(1454), p. 339. doi: 10.1098/RSTB.2004.1585.

Tommasi, N. *et al.* (2022) 'DNA metabarcoding unveils the effects of habitat fragmentation on pollinator diversity, plant-pollinator interactions, and pollination efficiency in Maldive islands', *Molecular Ecology*. John Wiley & Sons, Ltd, 00, pp. 1–11. doi: 10.1111/MEC.16537.

Valentin, R. E. *et al.* (2018) 'Early detection of invasive exotic insect infestations using eDNA from crop surfaces', *Frontiers in Ecology and the Environment*, 16(5), pp. 265–270. doi: 10.1002/fee.1811.

Wagner, D. L. *et al.* (2021) 'Insect decline in the Anthropocene: Death by a thousand cuts', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 118(2), p. e2023989118. doi: 10.1073/PNAS.2023989118.

Wang, Q. *et al.* (2007) 'Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 73(16), p. 5261. doi: 10.1128/AEM.00062-07.

Yu, D. W. *et al.* (2012) 'Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring', *Methods in Ecology and Evolution*, 3(4), pp. 613–623. doi: 10.1111/j.2041-210X.2012.00198.x.

Zhang, W. *et al.* (2007) 'Ecosystem services and dis-services to agriculture', *Ecological Economics*. Elsevier, 64(2), pp. 253–260. doi: 10.1016/J.ECOLECON.2007.02.024.

Zhang, W. and Swinton, S. M. (2009) 'Incorporating natural enemies in an economic

threshold for dynamically optimal pest management', *Ecological Modelling*, 220(9–10), pp. 1315–1324. doi: 10.1016/j.ecolmodel.2009.01.027.

Zizka, V. M. A. *et al.* (2018) 'DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method', *Genome*, pp. 1–15. doi: 10.1139/gen-2018-0048.

Chapter 2: Assessing the efficacy of DNA-metabarcoding to identify archived insect bulk samples: a case study using the Rothamsted Insect Survey

2.1 Abstract:

Insect populations are declining in many parts of the world yet, there is a lot of uncertainty regarding the state of insect biodiversity and population sizes remain unknown. Lack of long-term monitoring data is a major impediment to understanding how insects respond to environmental changes. Next-Generation Sequencing (NGS) approaches, such as metabarcoding, have the potential to revolutionize insect biomonitoring schemes by i) rapid identification of bulk samples; ii) scaling-up monitoring in both time and space to detect rare or invasive species iii) uncovering species which are difficult to identify via taxonomy and/or are of economic interest (e.g. pest and invasive species) or conservation concern iv) provide important additional data on species-interactions (e.g. parasitism/natural pest control) that can help us better understand ecosystem provision within agriculture. However, it is unclear to what extent NGS can be applied to long-term stored insect samples. This is important as it could provide valuable information regarding past changes to biodiversity and/or potentially be used for forecasting. Here I assess the efficacy of DNA-metabarcoding to process and identify archived samples from the longest passive monitoring scheme in the UK: The Rothamsted Insect Survey (RIS). I focus on aphids as the target taxa of RIS suction-traps and analyse a time-series of 16 years going back to 2003 using metabarcoding in a non-destructive way. I show that a non-destructive DNA extraction can be a good alternative to destructive DNA extraction, ensuring the integrity of archival samples. I then compare the historical taxonomically identified dataset with the metabarcoding data I generate and show that DNA-metabarcoding can identify most of the samples with varying success (mean of over 80%). I show year was not an important factor in determining congruence between the two datasets. Finally, I highlight the wider potential that NGS approaches have for insect monitoring schemes that store samples.

2.2 Introduction

Recent studies have highlighted that insect populations are declining in many parts of the world (Dirzo *et al.*, 2014; van Klink *et al.*, 2020). However, the global status of insects is more nuanced but what is clear is the lack of data on diversity and population trends in many parts of the world (Cardoso *et al.*, 2011), although Outhwaite *et al*., (2022) showed in the largest worldwide study of its kind using species abundance, presence/absence and richness for a wide range of taxa across the globe that a 50% reduction in the abundance and 27% in the number of species has occurred (Outhwaite, McCann and Newbold, 2022). A lack of data largely stems from the difficulties associated with monitoring insects coupled with a declining taxonomic experts (Drew, 2011). Lately, DNA-based biodiversity monitoring have shown promise as a way to scale-up biodiversity monitoring (Ji *et al.*, 2013; Bush *et al.*, 2017) coining the name "Next-Generation Biomonitoring" (Baird and Hajibabaei, 2012). Given the rapid effects of human influence on the environment, there is a clear need for faster, more efficient and comprehensive techniques for biodiversity monitoring (Makiola *et al.*, 2020). Despite the recent success of such approaches there is still no large-scale adoption for terrestrial systems, as validation is needed to assess how accurate these methods are for routine monitoring. Next-Generation Sequencing (NGS) based biomonitoring approaches like metabarcoding, mitogenomics, metagenomics can help address the difficulties associated with surveillance in the context of insect declines (Piper *et al.*, 2019).

NGS approaches have been extensively used for insect identifications (Yu *et al.*, 2012; Zhou *et al.*, 2013; Sigut *et al.*, 2017). NGS based DNA barcoding is an approach where DNA-barcoding is coupled with NGS: a PCR (Polymerase Chain Reaction) step is used to amplify a region of interest for a single target taxon and thousands of insect specimens can be loaded on an NGS platform (Srivathsan *et al.*, 2021). This approach is well established and has increased dramatically the number of available barcodes in sequence databases (Shokralla *et al.*, 2014). DNA-metabarcoding refers to a similar approach where DNA barcodes are used to identify the whole community of a sample, like multiple taxa from bulk-insect samples (such as a Malaise, pan or suction-trap). DNA metabarcoding is still a developing field yet it is routinely used for bulk species identification on large-scale studies as it scales better when compared with NGS barcoding or metagenomics (Gueuning *et al.*, 2019). The most used

region of interest for insect identification is Cytochrome Oxydase I or COI for both NGS barcoding and metabarcoding. Although, other markers are also gaining popularity (Marquina, Andersson and Ronquist, 2019). Metabarcoding has been shown to be more comprehensive for species identification than traditional taxonomic approaches, but more importantly it scales much more efficiently both in terms of costs and time (Ji *et al.*, 2013). Lab costs are continuing to decline and sample sorting is typically not required, although it can have advantages (Majaneva *et al.*, 2018). The scalability of the approach makes it an ideal tool for rapid-biodiversity assessment but also a tool for rapid diagnostics of pests or non-native species of economic importance (Kitson *et al.*, 2019; Piper *et al.*, 2019).

DNA metabarcoding of insect samples is not without its problems. It is affected by contamination and biases induced by PCR that can cause: a) mis-identifications; b) amplification of non-target taxa; and c) primer-template mismatches that minimize the potential for metabarcoding results to be quantitative (i.e. infer abundance information from reads) (Krehenwinkel *et al.*, 2017). An approach that avoids the PCR bias is 'metagenomics' where the PCR amplification step is skipped, and multi-taxa identification is based on whole-genome sequencing. In a landmark study, Ji *et al.,* (2020) constructed a metagenomic pipeline to process historical samples from an insect monitoring scheme in the artic that retains abundance information. Whilst metagenomic approaches are state-of-the-art, they require high-quality DNA, which is difficult to obtain from highly degraded archive samples. Moreover, the sequencing depth needed for such approaches can make the use of metagenomics for many and complex samples prohibitively expensive (Gueuning *et al.*, 2019). Nonetheless, the study highlights the overall potential of using NGS approaches on insect archival samples, with considerable applications for understanding insect responses to environmental change (Petsopoulos *et al.*, 2021).

Currently, one of the biggest impediments for using metabarcoding for biomonitoring is the destructiveness in the DNA extraction protocols in order to yield high quality DNA. Advances in non-destructive sample processing, however, show that such approaches can be comparable to homogenized tissue methods (Carew, Coleman and Hoffmann, 2018). Methods can vary considerably from quick tissue digestions to extraction of the preservative within the sample, but most of them depend on tissue digestion for a minimal amount of time, where lysis of tissue can happen between minutes or hours (Zizka *et al.*, 2018; Batovska *et al.*,

2021). Depending on the chemical or the time of the digestion morphological damage can occur, particularly to soft bodied insects. However, a range of other biases still exist: for example it was shown that with a non-destructive DNA extraction method, morphological characteristics such as sclerotization and molecular traits like primer-template mismatch can all affect the taxonomic composition of the samples (Martoni *et al.*, 2022). Despite all the biases that exist, the benefits of such approaches can overcome the drawbacks where preservation and accessibility (future or not) of specimens is considered important i.e. when the aim is establishment of a species inventory like in many cases of long-term monitoring schemes.

Here, I develop and evaluate the application of DNA-metabarcoding to a historic insect monitoring scheme in the UK: The Rothamsted Insect Survey (hereafter as RIS (Harrington, 2013)). RIS has been monitoring aphids and moths since the 1960s by using networks of suction and light traps respectively. A key aim of the network is to provide farmers with information on the timing and size of aphid migrations to prevent heavy prophylactic use of insecticides. I focus on the aphid fraction over a 16-year time-series from 2003-2018: aphids are taxonomically identified to species level and subsequently archived. My aims are: i) to establish a non-destructive metabarcoding approach to process historical stored samples; ii) assess congruence between the taxonomically identified dataset and the metabarcoding one; iii) identify potential limiting factors (age of sample, sequencing depth, biases associated with non-destructive protocols) when processing such samples, iv) determine the added value that such approaches offer to insect monitoring schemes by unlocking previously untapped resources of insect specimens.

2.3 Materials and Methods

2.3.1 The Rothamsted Insect Survey suction-trap samples:

The suction-trap network currently comprises 16 traps (12 in England, 4 in Scotland, see Figure 2. 1), each 12.2 m tall that continuously measure the aerial density of flying aphids and provide daily records during the main aphid flying season (April–November) and weekly records at other times (https://insectsurvey.com/). The network has been operational from

1964 till present. Just over 400 of the 600 aphid species on the British list aphid have been recorded to date. Samples, both of aphids and 'bycatch', are stored and are available for further research. A unique aspect of RIS is that all the samples have been archived. However, mainly the aphid fraction has been identified to species level as they compromise the main target taxa. Aphids from 1968-2002 have been cleared for identification purposes in a formalin solution that removed internal tissues, therefore these samples cannot be determined by DNA analysis. From 2003 onwards samples have been well preserved in 100% ethanol:glycerol solution at a ratio of 95:5 which slows DNA degradation more efficiently than other solutions (Kagzi *et al.*, 2022). For this reason, I focused on a subsample of a 16-year time-series (2003-2018) from a single suction-trap (Newcastle suction-trap) where aphids have been preserved at room temperature.

Figure 2. 1 The Rothamsted Insect Survey (RIS) suction trap network with 16 suction traps across the U.K.

2.3.2 Sample collection

I collected two monthly samples from the archive between May-October 2003-2018 imposing use-case criteria for the number of aphids within those samples. Specifically, sampled dates

must have a total number of aphids within ± one standard deviation from the overall aphid mean count of the corresponding month. This was mainly done as a rule of a thumb for logistic reasons (i.e., fitting all samples in a single sequencing run) but also to avoid samples with extreme number of aphids within tubes (>300). Note that the majority of samples within the Newcastle trap were below 100 individuals in every sample, but exceptions did exist (See section A.1; Table A. 1). This way, I standardized consumable volumes and sizes for the whole experiment. The resulting timeseries includes more than 66% genera (67 genera, 122 species) found in the complete daily time series of the Newcastle suction-trap, which included over 2500 samples (105 genera) between 2003-2018. This is a good representative sample (in terms of species coverage) from the series. The samples totalled 183, aiming for 12 samples per year (split in two datasets, see below).

2.3.3 Non-destructive extraction

As RIS wishes to retain insect samples for future researchers, I aimed to extract the DNA non-destructively by using a short-digestion time lysis step. The damage that can be done to the tissue is dependent on the time that the tissue is digested and the lysis buffer used (Carew *et al.*, 2018 Piper et al., 2022). I used a bead based protocol[protocol #6.3] (Oberacker *et al.*, 2019) with slight modifications on the amount of lysis volume used to adjust it for different sample sizes (in terms of numbers of aphids) (see Table A. 1; Appendix A). To establish the minimum amount of time required for the lysis digestion before damage to the tissue became visible, I conducted a preliminary experiment: 5 extra samples were chosen for testing on 3 different lysis digestion times: 1h, 2h, 6h (see Table A. 2 Appendix A), with amplification success assessed on a gel. Morphological damage was assessed on a microscope in accordance with taxonomists at the RIS. From those samples, the 1h digestion did not amplify, but samples with 2h showed almost no morphological damage and amplification success was high. Finally, samples for 6h showed slight tissue digestion damage and high amplification rates (see Figure A. 1; Appendix A). For the remaining 183 samples of the time series, I split them randomly in two datasets by using the *"sample"* function in R (v. 4.0.1, R Core Team 2021): 91 samples were extracted with a 2h treatment and 92 with the 6h treatment. This was done to assess any influence of digestion time on overall results. Note that samples in each treatment compromised 6 dates per year, one for every month of the sampled period

(12 per year in total for both datasets, except for missing dates due to trap inoperation). The initial digestions were carried out in 1.5 mL tubes, after which 62 µL of lysate was transferred to 96 well plates (irrespective of initial lysis volume). This was done to standardize volumes and use the same volumes and steps as the protocol. Finally for each plate, I included a DNA extraction positive and a DNA extraction negative. The DNA extraction positive was tissue from an ichneumonid pollinator wasp belonging to the genus *Acrolyta*. The DNA extraction negative included all reagents used for the DNA extraction and molecular grade water. All DNA extractions where quantified on a Qubit 4 (Thermo Fisher Scientific) with the 1x High Sensitivity assay.

2.3.4 PCR amplification and library preparation

I followed the nested-tagging method by Kitson *et al*., (2019) that uses a combinatorial indexing approach to multiplex samples. I targeted a 313 bp fragment of the cytochrome C oxidase subunit I barcode regions with the primers mLCOintF and jgHCO2198. The reason I decided the COI loci is because for aphids but also all other terrestrial arthropods there are very well covered and curated reference databases (Coeur d'acier *et al*., 2014) the same is not true for most other markers regions. This specific primer pair has been one of the most successful and widely used degenerate primers to characterize metazoan communities (Leray *et al*., 2013). It has also been used widely for arthropod community discovery (Geller *et al*., 2013; Porter and Hajibabaei, 2020). The size of the amplicon (313 bp) makes it also suitable for degraded samples such in our case. Although other primers do exist for the COI region (e.g. BF1R2 see Elbrecht and Leese, 2017) these have not been widely used for terrestrial insects and in the largest primer evaluation study so far the primer pair of choice comes amongst the top primers for insect identification particularly for Hemiptera (Piper *et al*., 2023). For all of the above reasons, I believe that our choice of primer is justified, and we proceeded with these primers for all of the remaining chapters except otherwise stated. Note that these are modified from Leray *et al.*, (2013) and they include the standard Illumina molecular identification tags, bridge sequences and heterogeneity spacers (see Kitson et al., (2019) for details). PCR's were carried out over 40 cycles (95^{o}C for 45 seconds (s), 51 ^{o}C for 15 s and 72 ^{o}C for 45 s in 20µL reactions using a high fidelity Taq mastermix (MyFi Mix Bioline), 2 µL of template DNA and each primer (final concentration at 0.5 µM). Our reasoning for more PCR

cycles here is that DNA from these archival samples is expected to be limited, with minimal amounts of DNA left, studies have shown that cycles above the routinely used ones (20-30 cycles) are often preferred for old specimens (Vierna *et al*., 2017). To further prevent cross contamination the wells were sealed using mineral oil, before all the other reagents and template DNA were added. Two PCR controls were used per plate, a PCR positive which was already extracted DNA from a moth belonging to the genus *Operophtera* and a PCR negative which included all PCR reagents but had no template DNA.

PCR success was checked using 5 µL of PCR products on 1.5% agarose gels. PCR negatives and DNA extraction negatives did not show any bands. I then conducted a bead-based normalisation by using 0.6:1 ratio of Solid Phase Reversible Immobilizations beads (SPRI) (9 µL) and 15 µL of PCR template for each sample. After clean-ups and prior to library preparation I pooled the samples in groups of 16, from which 4 µL from each sample was taken to create each pre-library. This process generated 12 libraries, 6 for every plate. This was done to further increase sequencing diversity during the initial cycles of the sequencing run this was suggested by the sequencing centre (Genomics Core Facility at Newcastle University) and results from a previous trial that included only one library and 40 samples for validation purposes (to see whether sequencing was successful before committing to a full MiSeq run) failed to produce any passing reads. To create each of these libraries I used a second PCR (PCR2) with 12 cycles (95^{o}C for 45 s, 51 ^{o}C for 15 s and 72 ^{o}C) and a final extension step of 5 min at 72 ^{o}C in 20 µL reactions using 5 µL of each pooled library, the same Taq (MyFi Mix Bioline) and each of the respective Illumina N5 and N7 adapters (at a concentration of 1 µM). For each library a PCR2 negative was also included. All libraries and PCR2 negatives were checked on gel. No bands were visible for any of the negatives. All of the controls pre and post PCR were sequenced along with the samples. I then performed a PCR2 clean-up to remove fragments smaller than the target amplicon by using 0.6:1 ratio of SPRI beads to template (9µL and 15µL respectively). After cleaning the libraries, they were checked on an Agilent TapeStation and were pooled equimolarly at approximately 7.6 ng/µL. The pooled final library was then sequenced on an Illumina MiSeq using a V3 (2x300) kit and 500 cycles (2x250) at the Genomics Core Facility at Newcastle University.

2.3.5 Bioinformatic analysis

Sample demultiplexing within individual libraries was done on the software *MetaBEAT* (https://github.com/HullUni-bioinformatics/metaBEAT). Only reads with the used combination of tags were kept. All other analyses were conducted in R (v. 4.0.1, R Core Team 2021) except if stated otherwise. The demultiplexed data were processed using package *DADA2* (Callahan *et al.*, 2016), removing primers using *cutadapt* v1.18 (Martin, 2011). *DADA2* filtered and trimmed sequences based upon read quality removing any reads with ambiguous "N" bases with the "filterandtrim" function. I then merged pair-end reads, removed chimeras with the "removeBimeraDenovo" function and finally inferred Amplicon Sequence Variants within DADA2. All of the functions were used with the default arguments. Taxonomy was assigned using two methods in order to assess their performance: first I used a trained RDP classifier with the *Insect* package (Wilkinson *et al.*, 2018) with a database and the classifier included within the package. Secondly, I used "blastn" function on the command line (Camacho *et al.*, 2009) with a curated database for all Metazoa downloaded from the MIDORI database 2 (Leray, Knowlton and Machida, 2022) to assign taxonomy and kept only the top hit for each ASV. More specifically, I kept only hits that had more than 99% query cover and 100% percent identity. For ASVs that assignment was not possible at the species level, only genus level information was kept. The remaining analysis was based on the packages *phyloseq* (Mcmurdie & Holmes, 2012) and *vegan* (Oskanen *et al.,* 2020) see below.

2.3.6 Statistical analyses

To assess congruence between the morphological dataset (hereafter described as MOTA) and the metabarcoding (hereafter described as META) one, I used base set functions in R ("*intersect*", "*setdiff*") to identify percentages of common and different taxa for two taxonomic levels: genus and species. Both of these functions use character strings, therefore I first standardized taxonomic annotations between the two datasets in R using the *tidyverse* package (Wickham *et al.*, 2019). Additionally, I've identified false positives and false negatives based on which taxa were found on META but not on MOTA. Taxa that were only identified at the family level on MOTA were dropped. An ANOVA was performed between the number of sequencing reads for each sample from the two treatments after the reads have been log transformed using the "*log*" function in R, to investigate if there was statistically significant difference in the output between the treatments. To see whether sequencing reads were

correlated with the abundance captured in the morphological dataset for the main taxa captured I performed a linear regression between these two in R after log-transforming the counts and reads. A further ANOVA was performed to examine whether the two treatments had an effect on congruence (measured as percentage of common species after transforming with the "log" function in R). I also applied sequencing thresholds for each sample to filter potential sample-to-sample contamination that might have resulted from RIS sample handling. The RIS handles samples daily by separating the aphid fraction from the bycatch, as multiple people work with the same samples, cross-sample contamination by using the same sorting equipment is possible. We therefore wanted to minimize this uncertainty of cross-contamination by using sample thresholds which are known to control for this (Drake *et al*., 2022). The thresholds were 1% and 0.5% of sample counts for each genus or species meaning that taxa with reads less than 1 or 0.5% of the sample were discarded. The same thresholds were used at the species level. To identify the level of false positives or false negatives within the pre-filtered dataset and filtered ones I've counted species mismatches between MOTA and META. Additional taxa within each sample that were identified by META but not MOTA were treated as false positives, whereas false negatives were taxa not found by META. Finally, to understand which of the factors influenced congruence I used a binomial generalized linear model (GLM) with congruence as a response variable and year, sequencing depth (log transformed reads) and sample evenness as predictor variables using the "glm" function in R accounting for possible interactions between the year and sequencing depth (model formula: Congruence ~ year+ sequence depth + sample evenness + year:sequencing depth, family = "binomial").

2.4. Results

2.4.1 PCR success and sequencing results

Overall, PCR success was 94% meaning that 173/183 samples produced a visible band on a gel. The run produced 23,000,000 reads (including PhiX), after demultiplexing 16 million reads were retained. After filtering, denoising, merging and chimera removal a total of 9,073,750 were retained. Reads per sample ranged from 484 to 256,137, there was very high variability between samples with a median 32,970 (1st quartile: 16,443, 3rd quartile: 71,883) whilst the

mean was 48,523 (± 44,978). Unsurprisingly, the samples with the lowest number of reads were the ones where amplification was not evident on an agarose gel (10 samples with reads less than <3700). Additionally, no reads from the DNA extraction or PCR negative passed any initial filters and DNA extraction, and PCR positives were not found in any of the other samples, suggesting minimal sample to sample contamination. There were no significant differences in the number of total reads for both treatments of digestions (ANOVA; $F_{1,182}$=0.77, p=0.37).

2.42 Morphology vs metabarcoding

Out of the 8,670,304 reads assigned to Hemiptera, 8,008,230 (92%) were assigned to genus level whilst 2,775,306 (36.1%) were assigned to species level when using the RDP classifier. However, of the 8,421,887 (97.1%) reads assigned to Hemiptera with Blast: 8,288,369 (95.5%) were assigned to species level and only 133,518 (0.01%) reads where assigned only to genus level or higher (family, order or class). I therefore present Blast results for all analyses except if mentioned otherwise. META consisted of 61 unique genera and 120 unique species. In comparison, MOTA compromised of 69 unique genera and 99 unique species (71% congruence for genera and 80% for species, see Figure 2. 3A). Out of the 15 genera not identified by META only two genera had more than 3 individuals across the time series: *Rhopalosiphonius* and *Mindarus* with 6 and 4 individuals. All the rest of the non-identified genera had less than 3 individuals across the time series with half of them having <2 individuals. The two dominant genera in the metabarcoding dataset were *Drepanosiphum* and *Rhopalosiphum* with 59.62% and 14.35% of total reads assigned to them respectively. A similar pattern was seen in the morphological dataset. However, *Drepanosiphum* (13.5% of total counts) were most abundant in the metabarcoding dataset whilst *Rhopalosiphum* (43% of total counts) were most abundant in the morphological dataset (see Figure A. 2; Table A. 3; Table A. 4). Overall, there was significant positive correlation between the number of individuals in MOTA with the reads from META (R^2 = 0.56, p-value= 1.079e-09, see Table 2. 1 Figure 2. 2)

Coefficients	**Estimate**	**Standard error**	**t-value**	**Pr (>t)**

Intercept	-1.96	0.71	-2.76	0.008 **
Counts metabarcoding	0.57	0.07	7.70	1.08e-09 ***

Table 2. 1 Results for the linear regression between morphological counts and counts from metabarcoding (both after log-transformed). Two asterisks (**) show significance at the 0.001 level while three asterisks (***) show significance at the 0.0001 level.

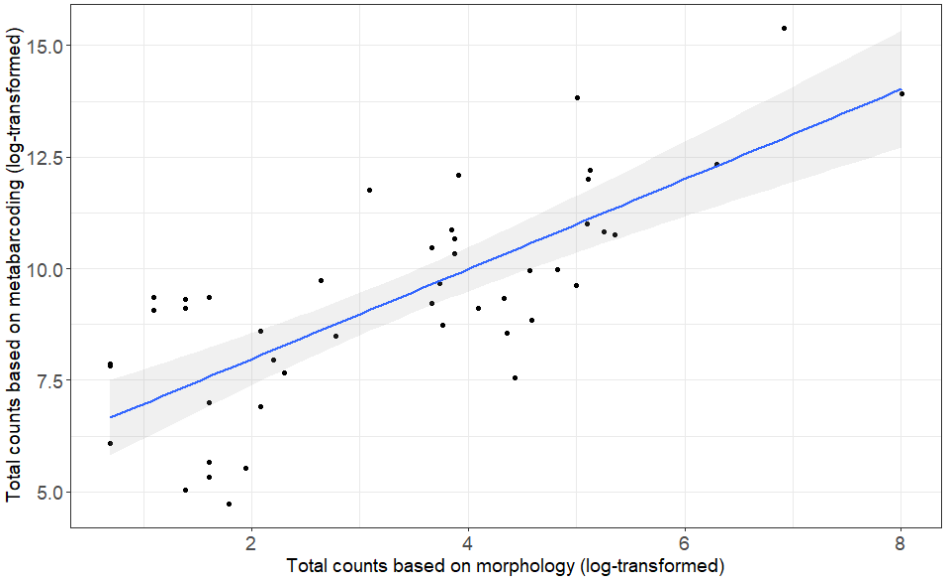


Figure 2. 2 Plotted best fit for the linear regression between counts from the morphological dataset and counts from the metabarcoding one (R^2 = 0.57).

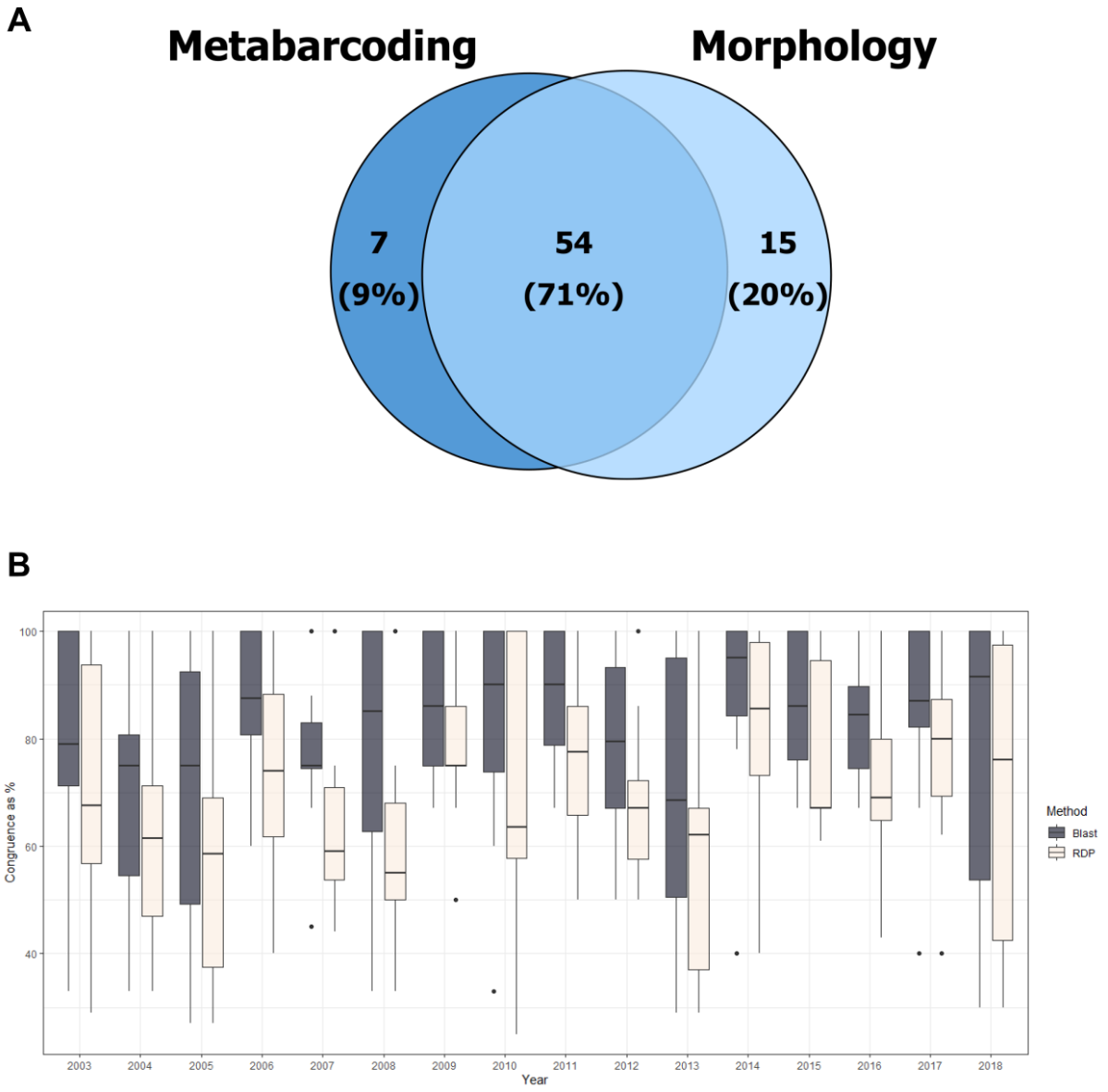


Figure 2. 3 A) Venn diagram for the two datasets (META and MOTA) at the genus level. B) Congruence across the two taxonomic assignment methods: BLAST (light grey) and RDP (light yellow)

There were no significant differences between the two treatments in the percentages of congruence (ANOVA; $F_{1,182} = 0.161$, p= 0.68). The percentage of congruence between both datasets showed high variability with a mean of 71.37 % (±20.39%) for genera for RDP and 80.75 % (±19.21) for BLAST. Between years the average percent of congruence varied from 69.80 % for year 2013 to 87.90% for year 2014 (see Figure 2. 3B for percentages throughout the years; Table 2. 3). For species level analyses the mean was found much lower with a mean of 49% for all the years (see Figure 2. 5 A;). The average number of taxa found as false positive

in META was 4 whilst false negatives had an average of 1.21. By using the sequencing thresholds of 0.1 and 0.5, the average number of false positives fell between 1.07-1.25 and the number of false negatives increased to 3.55 and 4.60 depending on the threshold. At the species level the false positives had a mean of 4.6 whilst false negatives had a mean of 2.34, the mean of false negatives dropped to 0.80 and false positives increased to 5.5 after applying the criteria. Unsurprisingly, congruence also fell to 50 and 40 % (see Figure 2. 4B and 2. 4C) for genera whilst for species it fell to 23% (see Figure 2. 5B). Year was not found to significantly affect congruence, neither did sequencing depth or any of the predictor variables was found significant in explaining congruence (see Table 2. 2)

Coefficients	**Estimates**	**Standard error**	**z-value**	**Pr(>z)**
Intercept	207.164	886.016	0.239	0.811
Year	-0.103	0.430	-0.241	0.809
Eveness	2.280	1.595	1.429	0.153
Sequence depth	-24.378	84.364	-0.289	0.773
Year:Sequence depth	0.011	0.041	0.292	0.770

Table 2. 2 Results from the binomial generalized linear model for factors explaining congruence as similarity

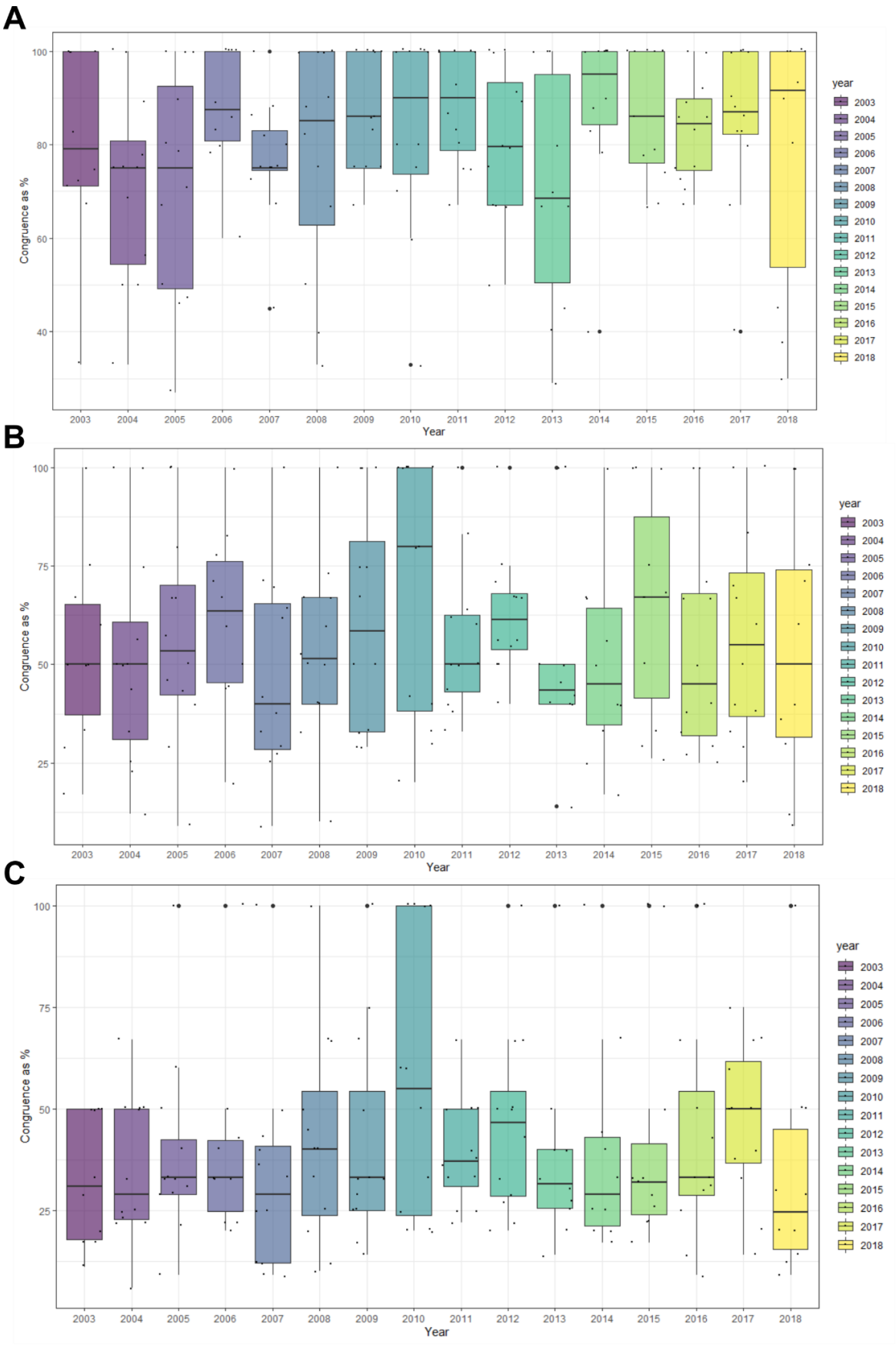


Figure 2. 4 Congruence across time: A) No sample filtering % threshold applied, B) 0.5 % Sample threshold applied, C) 1 % Sample threshold applied

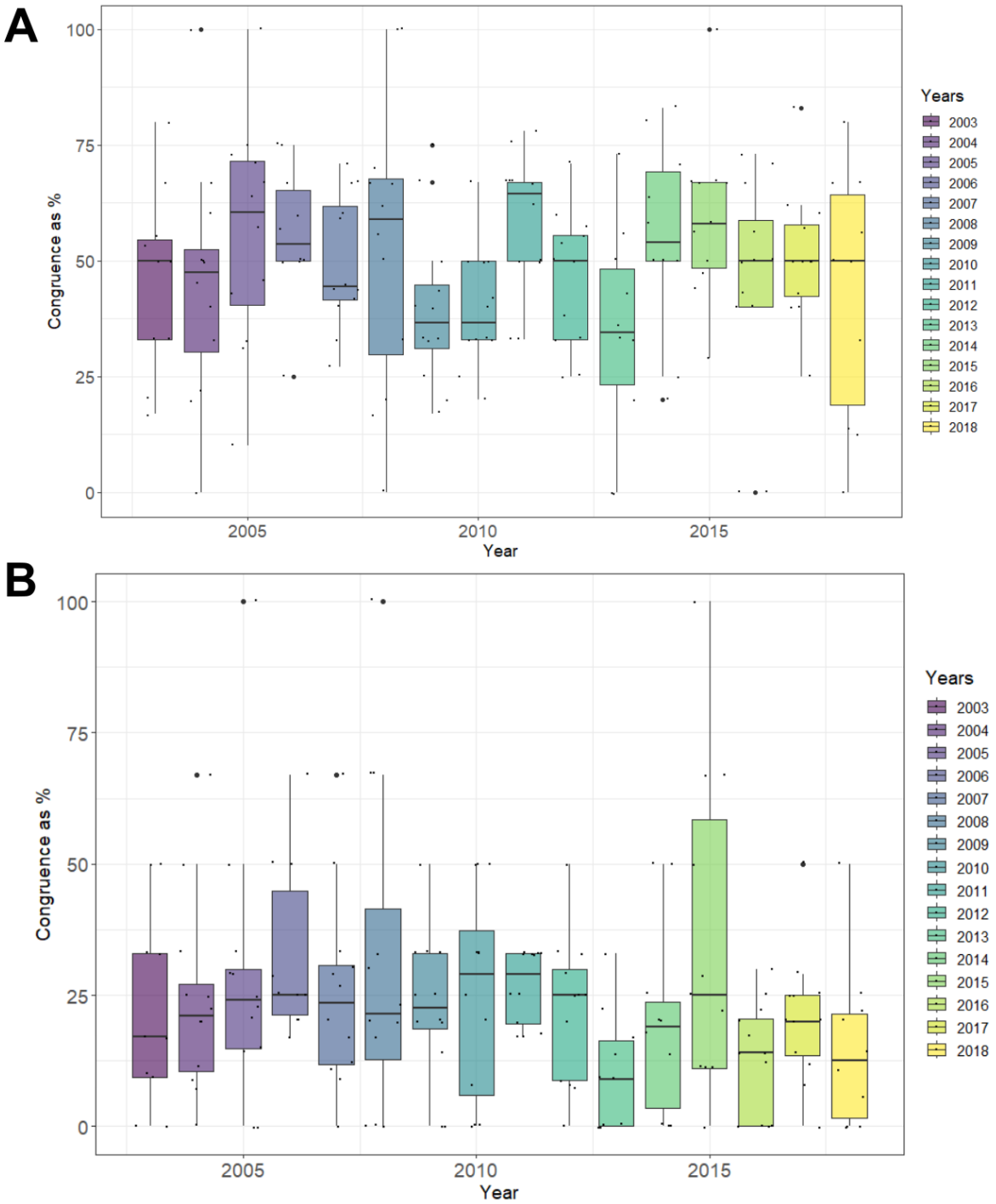


Figure 2. 5 Congruence across time for species level analysis: A) No sample % threshold applied, B) 1 % threshold applied

2.4.3 Non-aphid taxa in sequences

BLAST assigned 8,670,304 (92.1% of total reads) reads to Hemiptera (aphids and adelgids, which compromised the target taxa), 158,431 (1.74%) reads were assigned to other Arthropoda taxa, 1,304 (0.014%) reads were assigned to Chordata and 242,699 reads (2.64%)

were either assigned as root (i.e no identification was made) or were assigned to lower than Phylum. Total reads information entering and exiting the *DADA2* pipeline can be found in the Appendix A; Table A. 1. Reads that were assigned taxa other than the target taxa included common possible contaminants of RIS (Human and bird DNA) and other arthropod taxa commonly found within the samples before aphids are separated from the rest of the catch (see Figure 2. 6). The most abundant non-aphid orders were Diptera (54 samples), Diplostraca (9) and Hymenoptera (63). In the case of Hymenoptera over 50% of our samples had reads of aphid parasitoids. Certain samples were inspected for presence of non-aphid taxa such as Diplostraca (*Daphnia magna*) due to the high number of reads and because I believe it was unlikely to occur within RIS suction-traps. These taxa were not found and therefore I am uncertain if this is pre or post PCR contaminant. From the nine samples that had *Daphnia manga* reads one sample had 97.2% (24334) of those reads. However, certain samples did have other arthropods which included mainly Diptera, Thysanoptera and Psocoptera (See Figure 2. 6 for reads assigned to non-target taxa). As I was mainly interested in the target (identified) taxa and the congruence between morphology all analyses included only reads assigned to aphids (excluding the genus *Adelges*) except if otherwise mentioned.

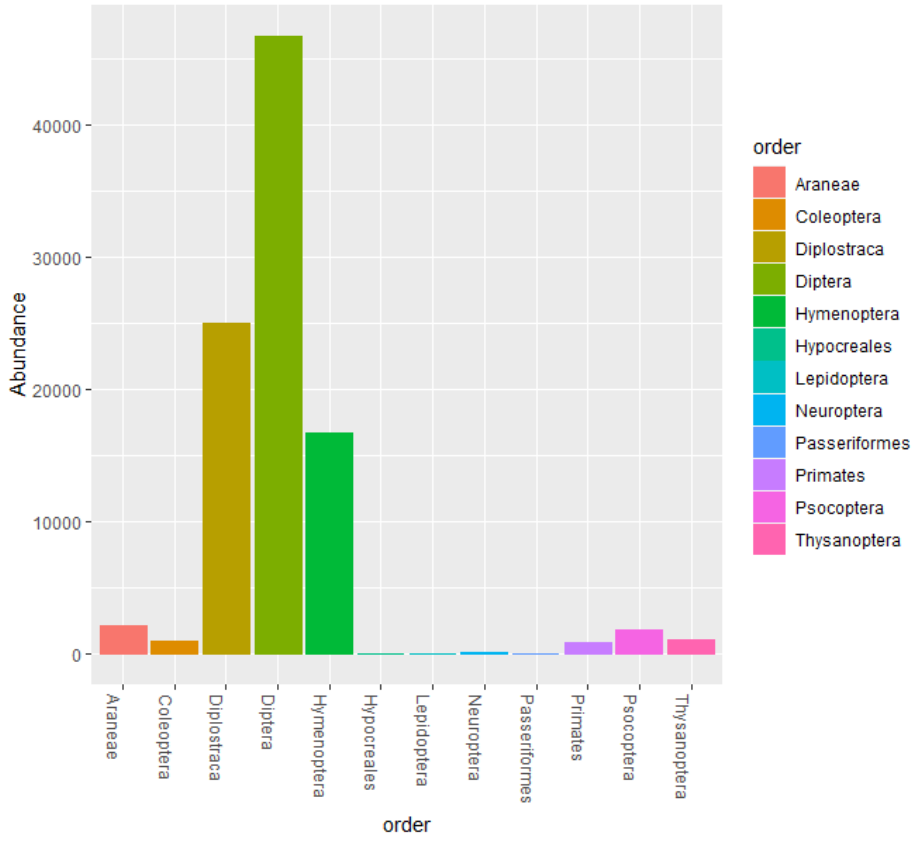


Figure 2. 6 Order level information on the read abundance for non-target taxa across the whole dataset

Year	Mean (s.d) congruence RDP	Mean (s.d) congruence BLAST	Number of unique genera morphology	Number of unique genera metabarcoding
2003	72.3 (± 22.2)	80.1 (± 21.5)	31	24
2004	61.8 (± 19.1)	70.8 (± 20.6)	32	22
2005	58.1 (± 22.9)	71.4 (± 24.5)	34	21
2006	74.4 (± 21.2)	87.6 (± 13.1)	22	18
2007	66.7 (± 14.2)	76.7 (± 13.2)	34	26
2008	62.5 (± 18.2)	77.0 (± 24.4)	28	24
2009	82.4 (± 12.2)	87.2 (± 12.4)	19	19
2010	77.6 (± 23.2)	83.1 (± 21.3)	23	16
2011	73.4 (± 15.1)	88.3 (± 12.1)	32	27
2012	67.5 (± 15.4)	80.4 (± 16)	30	19
2013	59.5 (± 25.8)	69.8 (± 25.9)	28	19
2014	83.3 (± 18.5)	87.9 (± 18.7)	20	16
2015	78.2 (± 16.2)	86.4 (± 14)	33	21
2016	73.3 (± 17.5)	83.0 (± 11.1)	29	25
2017	82.7 (± 17.7)	84.7 (± 17.4)	25	22
2018	69.3 (± 29.8)	77.6 (± 28.5)	31	28

Table 2. 3 Congruence measured as percentage for each year along with the number of unique genera identified by morphology and metabarcoding

2.5 Discussion

In this study I have demonstrated that DNA-metabarcoding can successfully identify insect species from long-term monitoring archive samples, and that this can be achieved non-destructively. I identified aphids (and other species) that have been archived for more than 18 years (albeit with varying DNA recovery success between years) despite the non-optimal conditions of the RIS collection (see below). Taxonomic identification fared slightly better than DNA-metabarcoding when assessed by congruence, which highlights the need to combine the two approaches (Keck *et al.*, 2022) . However, I show that it is possible to recover over 70% of genera within our time-series with DNA-metabarcoding that requires little taxonomic knowledge. Our study further highlights the added value of non-destructive DNA-based approaches for archival samples of insect collections, the importance of such collections but also their limitations.

2.5.1 Looking back in time: taxonomy vs metabarcoding

The congruence between taxonomy and metabarcoding varied by year. However, there is no clear linear relationship between them (see Figure 2. 3). This is important for collection-based research and particularly for RIS which has been archiving their samples since the 1970s. I here successfully analysed samples from one trap across a 16-year period. RIS compromises of 16 traps across the UK and daily catches of aerial insects are all collected and archived from all traps, most of the insects trapped remain unidentified as it would require a monumental effort to do this using taxonomy alone. In an era where long-term data are lacking making those samples accessible via DNA-metabarcoding can open new avenues for insect decline research (Petsopoulos *et al.*, 2021) and fill gaps on insect species populations that are still unknown in the UK. The taxonomical dataset compromised 15 genera unidentified by metabarcoding. This could be for a number of reasons, such as PCR-bias and primer-template mismatches (Alberdi *et al.*, 2018). The latter is known to be a problem for aphids in particular (Batovska *et al.*, 2021), yet whether different primers could fair better was beyond the scope of this study. Unidentified genera in the metabarcoding dataset mostly compromised of rare taxa (in abundance of 1-3 individuals) which can be difficult to identify when the samples are

dominated by other species. Despite these few genera, read counts overall were well correlated with morphological counts as seen by the linear regression (Fig. 2.2). As monitoring schemes need quantitative information on abundance to inform farmers on pest risk level, which indicates the quantitative nature of DNA-metabarcoding despite it being biased by multiple factors (Martoni *et al.,* 2022). Additionally, *Drepanosiphum* was the most dominant taxon in META whilst *Rhopalosiphum* was in MOTA. As *Drepanosiphim* is much larger than *Rhopalosiphum* this might have altered the relationship of read counts/morphological counts between the two here as larger insects can be overrepresented in the resultant community (Elbrecht *et al*., 2017). Surprisingly, evenness of the sample, sequencing depth or year were not found to significantly affect congruence which means that other factors are more influential. Our approach could not identify factors that affected congruence. Another option would be to try and partition bias influence throughout the protocol steps (Martoni *et al.*, 2022) could potentially help identify which factor was most influential however our study was not designed for that. Congruence was higher on average when using Blast for the taxonomic assignment compared to the RDP classifier, which further validates the need for more careful application of "newer" methods for taxonomic assignment (Hleap *et al.*, 2021). Overall, however, our study demonstrates that metabarcoding archived bulked samples shows considerable potential for unlocking insect time-series data.

2.5.2 A non-destructive approach for collections

To obtain high-quality DNA, destructive methods are usually applied, which is one of the reasons why collection samples are typically not processed (Raxworthy and Smith, 2021). This is especially true for old specimens where the DNA has been degraded after many years of collections even under optimal preservation conditions. RIS samples are stored in 100% ethanol:glycerol solution at a ratio of 95:5 at room temperature, which while cost-effective, is not ideal for DNA-preservation. However, I show that a non-destructive DNA extraction approach can accurately be used for sample identification irrespective of sample age as there was no difference between the destructive and non-destructive approach presented here. This is in line with other research suggesting non-destructive methods as an alternative for

DNA-metabarcoding (Martoni *et al.*, 2022). The approach used here relies on "quick" digestion of samples without external morphological damage. The quantity of the DNA extracted typically is influenced by the time of digestion and sclerotization of the species themselves (Carew, Coleman and Hoffmann, 2018). Aphids are soft-bodied insects and I found that a 2-h digestion was sufficient for species recovery. This also helps on bringing down DNA-extraction time which is important for DNA-based monitoring where the time from collection to identification is vital as for example in the case of invasive species (Piper *et al.*, 2019). But most importantly it ensures the future usability of such samples, like creating voucher specimens. There are approaches that are even faster (see Batovska *et al.*, 2021) during the DNA extraction step than the one presented here but it is uncertain how such approaches would fair with degraded samples further research would be needed to validate this.

2.5.3 Contamination issues

The way samples are handled by RIS, which pre-dates advances in molecular ecology, means that some sample-to-sample contamination is inevitable. Metabarcoding is usually very prone to this type of contamination because of the PCR amplification step. Reads where assigned to other arthropod taxa and other common contaminants (like human DNA), but no reads were found in our negative controls therefore I believe this contamination comes pre DNA-extraction and can be attributed to sample handling in RIS. Some taxa found in our study besides aphids include commonly trapped insects in RIS and in rare cases, after re-examination of the samples under a microscope, certain insects like chiromonids and thrips were found in the aphid samples. Of particular interest were braconid parasitoids, 5 species of which all are aphid parasitoids had reads in more than 50% of the samples. This could either be contamination from the "by-catch" fraction of the samples before the aphids get separated into different tubes or represent real parasitism of flying aphids which is known to occur (Walton, Loxdale and Allen-Williams, 2011). If parasitized aphids are present within the tubes then there is a unique opportunity for constructing long-term host-parasitoid interaction networks (Petsopoulos *et al.*, 2021). However, this would require a different approach to the one presented here, with single aphid individuals processed through high-throughput DNA barcoding to validate whether it is parasitism or simple contamination. Perhaps, the most

limiting factor in this study is sample-to-sample contamination from aphid species themselves. The DNA metabarcoding dataset did in some cases reveal more aphid species than the morphology or identified completely different species. For example, in 30 samples DNA-metabarcoding had more unique genera than the morphological dataset. Example of genera not found in the morphological dataset include: *Pachypappa*, *Pineus*, *Hyalopteroides*, *Ericaphis.* In our case, where the samples have been already identified I could be informed in our decision as to which species are truly there or not. However, by applying sequence thresholds (here applied as a percentage of reads within a sample) this source of contamination can be minimized (Drake *et al.*, 2022). This process did drop the taxa identified only by metabarcoding (false positives) but also caused a significant drop in overall congruence between the datasets as rare taxa truly present would be dropped (false negatives, see Figure 2. 3). Further research could focus on validating whether this is actual contamination or represents true incongruence between taxonomy and metabarcoding by processing individual aphids within samples. Although, there are other approaches that can minimize even more contamination for example, increasing technical replicates like PCR replicates (Yang *et al.*, 2021), because contamination in RIS has been introduced during sample processing it would be better to avoid the PCR step altogether. Ji *et al.*, (2020) present such a mitogenomics approach for another insect monitoring scheme that also "suffers" from the same type of contamination where they successfully revisit archival samples. However, the cost of mitogenomics can still be comperatively hight when compared with DNA-metabarcoding and the cost doesn't scale well with increasing sample size (Guening *et al.,* 2019). RIS represents an archive of tens of thousands of daily bulk insect samples therefore an unprecedented potential to construct time-series for thousands of insect species. Thankfully, with the drop of sequencing costs these approaches could ultimately become the alternative.

2.6 Conclusions

Our study is the first attempt to assess the efficacy of DNA-metabarcoding for determining species identity of long-term stored aerial suction-trapped insects. I showed high congruence across years and managed to do so in a non-destructive way, meaning that metabarcoding

shows great potential for better understanding long-term insect trends using RIS. The biggest limitation of this approach are contamination issues that are likely due to the way insect samples are handled and processed in RIS, which is important to know as suggestions can be made on how to ensure future samples are contamination free by applying best practices. The archival collection of RIS includes thousands of un-identified insect bulk samples (bycatch) that could next be processed using metabarcoding. The temporal (50+ years of daily samples) and spatial (16 locations across the UK) characteristics of it make RIS a treasure vault for insect research. Perhaps, RIS is unique in that sense, but other insect monitoring schemes exist in other countries, our study highlighted how samples from such schemes can be accessed via DNA-metabarcoding in a non-destructive way.

2.7. References

Alberdi, A. *et al.* (2018) 'Scrutinizing key steps for reliable metabarcoding of environmental samples', *Methods in Ecology and Evolution*. Edited by A. Mahon. John Wiley & Sons, Ltd (10.1111), 9(1), pp. 134–147. doi: 10.1111/2041-210X.12849.

Baird, J.D. and Hajibabaei M. (2012) 'Biomonitoring 2.0: a new paradigm in ecosystem assesment made possible by next-generation DNA sequencing', *Molecular Ecology,* pp. 2039-2044. doi: 10.111/j.1365-294x.2012.05519x.

Batovska, J. *et al.* (2021) 'Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches', *Scientific Reports*. Nature Publishing Group UK, 11(1), pp. 1–14. doi: 10.1038/s41598-021-85855-6.

Bush, A. *et al.* (2017) 'Connecting Earth observation to high-throughput biodiversity data', *Nature Ecology and Evolution*. Macmillan Publishers Limited, 1(7), pp. 1–9. doi: 10.1038/s41559-017-0176.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*.

BioMed Central, 10(1), pp. 1–9. doi: 10.1186/1471-2105-10-421.

Cardoso, P. *et al.* (2011) 'The seven impediments in invertebrate conservation and how to overcome them', *Biological Conservation*, 144(11), pp. 2647–2655. doi: 10.1016/j.biocon.2011.07.024.

Carew, M. E., Coleman, R. A. and Hoffmann, A. A. (2018) 'Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding?', *PeerJ*, 6, p. e4980. doi: 10.7717/peerj.4980.

Coeur d' Acier A., (2014) 'DNA barcoding and the Associated PhylAphidB@se Website for the identification of European Aphids (Insecta: Hemiptera: Aphididae)' *Plos One*, doi: 10.1371/journal.pone.0097620

Dirzo, R. *et al.* (2014) 'Defaunation in the Anthropocene', *Science*, 345(6195), pp. 401–406. doi: 10.1126/science.1251817.

Drake, L. E. *et al.* (2022) 'An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 13(3), pp. 694–710. doi: 10.1111/2041-210X.13780.

Drew, L. W. (2011) 'Are We Losing the Science of Taxonomy?As need grows, numbers and training are failing to keep up.', *BioScience*. Oxford Academic, 61(12), pp. 942–946. doi: 10.1525/BIO.2011.61.12.4.

Elbrecht, V., and Leese, F. (2017) 'Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment', *Frontiers in Environmental Science*, doi: 10.3389/fenvs.2017.00011

Elbrecht, V., *et al*. (2017) 'Sorting things out: Assesing effects of unequal specimen biomass on DNA metabarcoding, *Ecology and Evolution.* doi: 10.1002/ece3.3192

Geller, J. *et al.* (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys, *Molecular Ecology Resources,* pp. 851-861, doi: 10.1111/1755-0998.12138

Gueuning, M. *et al.* (2019) 'Evaluating next-generation sequencing (NGS) methods for

routine monitoring of wild bees: Metabarcoding, mitogenomics or NGS barcoding', *Molecular Ecology Resources*. Blackwell Publishing Ltd, 19(4), pp. 847–862. doi: 10.1111/1755-0998.13013.

Harrington, R. (2013) 'The Rothamsted Insect Survey strikes gold', *Antenna*, 38(3), pp. 158–166. Available at: www.rothamsted.ac.uk/examine.

Hleap, J. S. *et al.* (2021) 'Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 21(7), pp. 2190–2203. doi: 10.1111/1755-0998.13407.

Ji, Y. *et al.* (2013) 'Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding', *Ecology Letters*, 16(10), pp. 1245–1257. doi: 10.1111/ele.12162.

Ji, Y. *et al.* (2020) 'SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes', *Molecular Ecology Resources*, 20(1), pp. 256–267. doi: 10.1111/1755-0998.13057.

Kagzi, K. *et al.* (2022) 'Environmental RNA degrades more rapidly than environmental DNA across a broad range of pH conditions', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 22(7), pp. 2640–2650. doi: 10.1111/1755-0998.13655.

Keck, F. *et al.* (2022) 'Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment', *Molecular Ecology*. John Wiley & Sons, Ltd, 31(6), pp. 1820–1835. doi: 10.1111/MEC.16364.

Kemp, B. M. and Smith, D. G. (2005) 'Use of bleach to eliminate contaminating DNA from the surface of bones and teeth', *Forensic science international*. Forensic Sci Int, 154(1), pp. 53–61. doi: 10.1016/J.FORSCIINT.2004.11.017.

Kitson, J. J. N. *et al.* (2019) 'Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding', *Molecular Ecology*, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

van Klink, R. *et al.* (2020) 'Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances', *Science*, 368(6489), pp. 417–420. doi:

10.1126/science.aax9931.

Krehenwinkel, H. *et al.* (2017) 'Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding', *Scientific Reports*, 7(1), pp. 1–12. doi: 10.1038/s41598-017-17333-x.

Leray, M., Knowlton, N. and Machida, R. J. (2022) 'MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences', *Environmental DNA*, 4(4), pp. 894–907. doi: 10.1002/edn3.303.

Majaneva, M. *et al.* (2018) 'Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples', *Metabarcoding and Metagenomics*. doi: 10.3897/mbmg.2.26664.

Makiola, A. *et al.* (2020) 'Key Questions for Next-Generation Biomonitoring', *Frontiers in Environmental Science*, 7, p. 197. doi: 10.3389/fenvs.2019.00197.

Marquina, D., Andersson, A. F. and Ronquist, F. (2019) 'New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 19(1), pp. 90–104. doi: 10.1111/1755-0998.12942.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), pp. 10–12. doi: 10.14806/EJ.17.1.200.

Martoni, F. *et al.* (2022) 'Disentangling bias for non-destructive insect metabarcoding', *PeerJ*, 10, p. e12981. doi: 10.7717/peerj.12981.

Mcmurdie, P. J. and Holmes, S. (no date) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data'. doi: 10.1371/journal.pone.0061217.

Oberacker, P. *et al.* (2019) 'Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation', *PLoS Biology*, 17(1), pp. 1–16. doi: 10.1371/journal.pbio.3000107.

Outhwaite, C. L., McCann, P. and Newbold, T. (2022) 'Agriculture and climate change are

reshaping insect biodiversity worldwide', *Nature 2022 605:7908*. Nature Publishing Group, 605(7908), pp. 97–102. doi: 10.1038/s41586-022-04644-x.

Petsopoulos, D. *et al.* (2021) 'Using network ecology to understand and mitigate long-term insect declines', *Ecological Entomology*. doi: 10.1111/een.13035.

Piper, A. M. *et al.* (2019) 'Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance', *GigaScience*. Oxford University Press, 8(8), pp. 1–22. doi: 10.1093/gigascience/giz092.

Piper, A.M. *et al.* (2021) 'Computational Evaluation of DNA Metabarcoding for Universal Diagnostics of Invasive Insect Pests', BioRxiv. doi: 10.1101/2021.03.16.435710

Porter, T.M, and Hajibabaei, M. (2020) 'Putting COI Metabarcoding in Context: The Utility of Exact Sequence Variants (ESVs) in Biodiversity Analysis', *Frontiers in Ecology and Evolution*, doi: 10.3389/fevo.2020.00248

Raxworthy, C. J. and Smith, B. T. (2021) 'Mining museums for historical DNA: advances and challenges in museomics', *Trends in Ecology & Evolution*. Elsevier, 36(11), pp. 1049–1060. doi: 10.1016/J.TREE.2021.07.009.

Shokralla, S. *et al.* (2014) 'Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens'. doi: 10.1111/1755-0998.12236.

Sigut, M. *et al.* (2017) 'Performance of DNA metabarcoding, standard barcoding, and morphological approach in the identification of hostparasitoid interactions', *PLoS ONE*. Edited by M. Hajibabaei, 12(12), pp. 1–18. doi: 10.1371/journal.pone.0187803.

Srivathsan, A. *et al.* (2021) 'MinION barcodes: biodiversity discovery and identification by everyone, for everyone', *bioRxiv*. BMC Biology, p. 2021.03.09.434692. Available at: https://doi.org/10.1101/2021.03.09.434692.

Vierna, J., *et al.* (2017) 'PCR cycles above routine numbers do not compromise high-throughput DNA barcoding results', *Genome*. doi: 10.1139/gen-2017-0081.

Walton, M. P. P., Loxdale, H. D. D. and Allen-Williams, L. J. J. (2011) 'Flying with a "death sentence" on board: Electrophoretic detection of braconid parasitoid larvae in migrating

winged grain aphids, Sitobion avenae (F.)', *Bulletin of Entomological Research*. Cambridge University Press, 101(4), pp. 443–449. doi: 10.1017/S000748531000074X.

Wickham, H. *et al.* (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. doi: 10.21105/joss.01686.

Wilkinson, S. P. *et al.* (2018) 'Taxonomic identification of environmental DNA with informatic sequence classification trees.' PeerJ Inc. doi: 10.7287/PEERJ.PREPRINTS.26812V1.

Yang, C. *et al.* (2021) 'Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 12(7), pp. 1252–1264. doi: 10.1111/2041-210X.13602.

Yu, D. W. *et al.* (2012) 'Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring', *Methods in Ecology and Evolution*, 3(4), pp. 613–623. doi: 10.1111/j.2041-210X.2012.00198.x.

Zhou, X. *et al.* (2013) 'Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification', *GigaScience*, 2(1), p. 4. doi: 10.1186/2047-217X-2-4.

Zizka, V. M. A. *et al.* (2018) 'DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method', *Genome*, pp. 1–15. doi: 10.1139/gen-2018-0048.

Chapter 3: Unlocking the potential of insect 'bycatch' using metabarcoding for biodiversity research

3.1 Abstract:

Insect biomonitoring schemes can have a lot of "unwanted" insect taxa trapped within their networks. These taxa, referred to as bycatch, can be seen as important biodiversity information. However, bycatch taxa rarely get identified or stored, which results in information loss that is unfortunate given the lack of data on general insect populations and particularly long-term data. There are many reasons why bycatch is not analysed, many of which have to do with financial constraints, or the lack of experienced personnel to identify diverse assemblages of insect taxa. Future biomonitoring tools, some of which are based on Next-Generation Sequencing, can scale up the way biodiversity is monitored by providing an efficient way to identify taxa in bulk, but is not dependent on taxonomic expertise. However, bycatch diversity still remains largely unknown, and it is unclear whether DNA-metabarcoding can describe bycatch diversity as this has not been tested, particularly from long-term stored insect samples. Here I focus on the bycatch fraction of a long-term monitoring scheme in the UK: The Rothamsted Insect Survey (RIS) suction trap network which monitors aphids but has also been storing bycatch taxa since the 1970s. I use DNA-metabarcoding to identify bycatch diversity from a 16-year time series from 2003 to 2018. I identified more than 800 arthropod species throughout the time series with the majority being Diptera. I showcase the potential of building time series from multiple taxa assemblages with DNA-metabarcoding highlighting bycatch as a valuable source for insect time series data. I found more than 25 aphid parasitoids and over 70 pollinators demonstrating the importance of bycatch to identify beneficial insects or potential pests. Finally, I highlight the potential of bycatch to scale-up the scope and breadth of biodiversity being monitored using DNA-metabarcoding in traditional biomonitoring surveys.

3.2 Introduction:

The term 'bycatch' which refers to non-target taxa captured within a sample, is mostly associated with the fishing industry, as non-target taxa that are captured can be of high conservation value and significant research and investment has gone into mitigating this (Komoroske and Lewison, 2015). Likewise, insect bycatch consists of non-target taxa trapped along with target organisms. Most of which is often discarded by researchers although there are notable exceptions (Petsopoulos *et al.*, 2021). Research efforts have focused on the effect of physical, visual and even chemical characteristics of the trapping methods for non-target taxa and how to minimize non-target taxa capture (Spears *et al.*, 2016; McCravy, 2018). For example, by changing the colour of water pan traps to attract less taxa or by narrowing the size of pitfall traps and even changing the chemical composition of pheromones used for trapping. However, bycatch can provide valuable biodiversity information such as species discovery, detection of pests or invasive species and it can also be used to provide more holistic ecological data for local insect populations (Hribar, 2020). In the context of insect declines bycatch can help increase the breadth of taxa assessed for long-term population changes, as currently the focus has been on a few set target taxa or charismatic groups (Rocha-Ortega, Rodriguez and Córdoba-Aguilar, 2021) where data are readily available. Bycatch also consists of a sample not only of taxa per se but of many known and unknown species interactions which could be uncovered with the aid of next-generation sequencing, ultimately leading to information beyond taxonomy that can be used to understand how ecological networks of species change through time (Petsopoulos *et al.*, 2021). Yet, despite its potential use, there are limited examples to date where bycatch species are identified or used for analyses and this is particularly true for mass trapping programs where bycatch species can often outnumber the target taxa (Skvarla and Holland, 2011; Hribar, 2020). One of the reasons why bycatch is often overlooked is because the lack of taxonomic personnel and knowledge to identify organisms beyond the target species and the financial support that this would need, another is space. As with many monitoring schemes, a diverse array of thousands of insects and other arthropods are trapped, often subsampling techniques are required to ease processing effort leading to additional information loss (Spears and Ramirez, 2015). However, novel approaches for identifying taxa based on short DNA fragments and Next-Generation Sequencing techniques can overcome these pitfalls (see DNA-

metabarcoding below). Furthermore, monitoring schemes typically have specific target taxa, for example to rapidly provide farmers with information on the best time to apply pesticides, or to assess current population levels (Harrington, 2013). Hribar (2020) presents examples of how using bycatch from a mosquito surveillance program in the USA can provide valuable information for biodiversity research by providing distribution information and new records of species. There are other examples of schemes that are used for monitoring a specific taxon but the breadth of other taxa getting trapped is high (Skvarla and Holland, 2011). There remain, however, limited examples of monitoring schemes where bycatch is archived and, despite calls, no monitoring scheme databases with bycatch information currently exist (Buchholz *et al.*, 2011).

One example where the bycatch has been archived for all the years that monitoring has taken place is the Rothamsted Insect Survey (RIS) suction trap network. RIS mainly target aphids which are identified to the species level. Although bycatch is not routinely identified, it is retained and stored in ethanol in room temperature (research on certain groups has been conducted in separate studies over the years) (Bell *et al.*, 2015; Pérez-Rodríguez, Shortall and Bell, 2015) and kept as a resource for researchers wishing to study the archive. Insect bycatch data such as this provides an exceptional but as yet untapped resource for better understanding and monitoring insect population trends, especially in the context of recently reported global declines (Petsopoulos *et al.*, 2021).

Next generation sequencing has been used extensively for identifying complex multi-taxa insect samples (Yu *et al.*, 2012) and can overcome some of the current limitations of processing bycatch data. DNA-metabarcoding is based on amplifying a region of genes usually Cytochrome Oxydase I (COI) with universal primers and then the amplicons are sequenced on NGS platforms like the Illumina MiSeq, HiSeq or NovaSeq series (Yang *et al.*, 2021). It is considered a cost-effective and scalable tool to identify biodiversity as it doesn't require taxonomic expertise (Ji *et al.*, 2013). Therefore, DNA-metabarcoding can be seen as an ideal tool to identify bycatch taxa within monitoring schemes in a cost-effective way, but to my knowledge this technology has not been validated and tested on archival samples (but see Krehewinkel *et al*., 2023).

The overall aim of this chapter is to test and validate DNA-metabarcoding to identify bycatch taxa from a time series of RIS archive samples over a 16-year period. My objectives are: i) to

examine the feasibility of creating a species inventory of archived bycatch taxa from a single suction trap between 2003 and 2018; ii) identify insects that provide ecosystem services/disservices in agriculture; and iii) discuss the potential of spatio-temporal bycatch samples, derived from DNA metabarcoding, as a novel source of invertebrate biodiversity time-series data.

3.3 Materials and methods

3.3.1 The Rothamsted Insect Survey suction trap samples:

The suction-trap network currently comprises 16 traps (12 in England, 4 in Scotland) (see Chapter 2, Figure 2. 1), each 12.2 metres tall that were primarily established to continuously measure the aerial density of flying aphids. The network provides daily records during the main aphid flying season (April–November) and weekly records at other times. The network has been operational from 1964 till present. Samples, both of aphids and bycatch, are separated and stored and are available as a resource for further research. A unique aspect of RIS is that all the samples have been archived at room temperature. The aphids are stored in 100% ethanol:glycerol solution at a ratio of 95:5 from 2003 onwards, all the previous years samples are stored in a formalin solution that removed internal tissues. The bycatch fraction has also been archived in a 100% ethanol:glycerol throughout the years with the same ratio, however it is uncertain whether the solution has been refreshed at any point. For example, in some samples I handled the solution had completely evaporated. Here I focused on bycatch, I subsampled a 16-year time-series (2003-2018) from a single suction trap based at Cockle Park Farm, Newcastle University, Newcastle Upon Tyne, UK. The primary reason for choosing this trap is practical as the amount of arthropods from a single trap can reach many thousands. Another reason is that previously I focused on the aphid fraction of this particular trap where I processed and compared already identified samples with DNA-metabarcoding (see Chapter 2). Therefore as I wanted to uncover biodiversity of bycatch taxa for the same dates, but also assess whether it is possible to find species that might be interacting with aphid pests (like aphid parasitoids or predators found in the bycatch), I focused on the same dates when it was possible (note samples from certain dates were not available: see Appendix B; section B.1; Table B. 1).

3.3.2 Sample collection and DNA extraction

I collected two monthly samples from the archive between May-October 2003-2018 (Total = 184), an additional sample was included for the bycatch (see Table B. 1). Briefly, the sampling was previously based on selecting samples chosen randomly from within +/- one standard deviation from the mean of the counts of aphids for that month. This was done to avoid excessive number of aphids and to standardise lab work throughout the experiment. For a more detailed explanation of the sampling see section 2.3 in Chapter 2 (Appendix section A.1). In the case where samples for the corresponding dates were unavailable, I picked the next day available in the time series for that month. For example, if a certain date was missing (because it was used for different projects), I sampled the next following date (If 26/06/2015 was unavailable, I would pick either the 25/06/2015 or 27/06/2015) see the Table B.1 for details on sampling and for the dates that have been changed. I extracted the DNA using a non-destructive method following the same procedures as Chapter 2 section 2.2 except otherwise stated. Briefly, I used a bead based protocol[protocol #6.3] (Oberacker *et al.*, 2019) with slight modifications on the amount of lysis volume used to adjust it for different sample sizes (See Table B. 2). 100 µl to 5 mL of Lysis solution were used. Whilst proteinase K ranged from 3ul to 30 µl per sample. Digestions were carried either in 1.5 mL tubes, 5 mL or 50 mL tubes if sample volume was too big for 3 hours. After that, 62 µL of lysate was transferred to 96 well plates (irrespective of initial lysis volume). This was done to standardise volumes and use the same volumes and steps as the protocol. Finally for each plate, I included a DNA extraction positive and a DNA extraction negative. The DNA extraction positive was a single individual from the species *Asellus aquaticus* which was selected for its low chance of being found in suction traps. The DNA extraction negative included all DNA-extraction reagents without and DNA which was substituted by molecular grade water.

3.3.3 PCR amplification and library preparation

I followed the nested-tagging method by Kitson et al. (2019) that uses a combinatorial indexing approach to multiplex samples. I targeted a 313 bp fragment of the cytochrome C oxidase subunit I barcode regions with the primers mLCOintF and jgHCO2198 (see section 2.3.4 in previous chapter for our reasoning on the primer choice). Note that these are modified from (Leray *et al.*, 2013) and they include the standard Illumina molecular

identification tags, bridge sequences and heterogeneity spacers (see Kitson *et al.*, 2019 for details). PCR's were carried out over 40 cycles (95^{o}C for 45 seconds (s), 51 ^{o}C for 15 s and 72 ^{o}C for 45 s in 20µL reactions using a high fidelity Taq mastermix (MyFi Mix Bioline), 2 µL of template DNA and each primer (final concentration at 0.5 µM). To further prevent cross contamination the wells were sealed using mineral oil, before all the other reagents and template DNA were added. PCR controls included a PCR positive and a PCR negative for each plate. For the PCR positive control, I selected the crayfish species *Homarus gammarus* while the PCR negative control included all reagents except any DNA template that was substituted by molecular grade water.

PCR success was checked by using 5 µL of PCR products on 1.5% agarose gels. PCR negatives and DNA extraction negatives did not show any bands. I then conducted a bead-based normalization by using 0.6:1 ratio of Solid Phase Reversible Immobilizations beads (SPRI) (9 µL) and 15 µL of PCR template for each sample. After clean-ups and prior to library preparation I pooled the samples in groups of 8 (plus 4 controls), 4 µL from each sample to create each pre-library. Totalling 23 libraries, 11 for the first plate and 12 for the second. This was done to further increase sequencing diversity during the initial cycles of the sequencing run. To create each of these libraries I used a second PCR (PCR2) with 12 cycles (95^{o}C for 45 s, 51 ^{o}C for 15 s and 72 ^{o}C) and a final extension step of 5 min at 72 ^{o}C in 20 µL reactions using 5 µL of each pooled library, the same Taq (MyFi Mix Bioline) and each of the respective Illumina N5 and N7 adapters (at a concentration of 1 µM). For each library a PCR2 negative was also included. All libraries and PCR2 negatives were checked on gel. No bands were visible for any of the negatives. I then did a PCR2 clean-up to remove fragments smaller than the target by using 0.6:1 ratio of SPRI to template (9µL and 15µL respectively). After cleaning the libraries, they were checked on an Agilent TapeStation and were pooled equimolarly at approximately 15 ng/µL. The pooled final library was then sequenced on an Illumina NovaSeq using a SP+XP (2x250) kit at the Genomics Core Facility at Newcastle University.

3.3.4 Bioinformatic analysis and processing

Sample demultiplexing within individual libraries was conducted using the *cutadapt* v1.18 software (Martin, 2011). All other analyses were conducted in *R* (v. 4.0.1, R Core Team 2020) except if stated otherwise. To check the quality of the reads for each sample we used *FastQC*

(Andrews, 2010) on individual samples which were then compared with *MultiQC* (Ewels *et al.*, 2016). The demultiplexed data were processed using package *DADA2* (Callahan *et al.*, 2016), removing primers using *cutadapt* v1.18 (Martin, 2011). *DADA2* filtered and trimmed sequences based upon read quality removing any reads with ambiguous "N" bases with the "filterAndTrim" function. I then merged pair-end reads with the "mergePairs" function, removed chimeras with the "removeBimeraDenovo" function and finally inferred Amplicon Sequence Variants (ASV's) within *DADA2* with the "dada" function. Taxonomy was assigned with Blastn (Camacho *et al.*, 2009) on the command line against a curated database for all Metazoa which was download from the MIDORI 2 database (Leray, Knowlton and Machida, 2022). More specifically, results from the taxonomic assignment were filtered based on the percentage of identical matches and query cover which we set at 99% for both. Furthermore, to curate the taxonomic community found, we further removed any ambiguous assignments where a species level identification was not possible.

3.3.5 Statistical analysis

I compared whether different quality filtering criteria influenced the number of ASV's and assigned taxonomy. This was done after checking the quality profiles of the reads due to the low overall quality for some samples within the run (see Figure 3. 1). To do so I used the "filterAndTrim" function in *DADA2* with three different options (see section B. 3; Appendix B). This resulted in three different datasets with different sequencing depths and different number of ASVs. I then compared whether there were significant differences in the number of ASV's and taxa discovery between the datasets, within the level of genera by performing an ANOVA with the function "aov" in R after checking of normality of the data. As the initial results showed no differences between the three datasets all the rest of the analyses were done using the most stringent filtering criteria from *DADA2* (see section 3.3.1)

Due to the nature of the dataset and the high probability of contamination from the way samples are handled in RIS, I also performed further filtering of the dataset by using sample percentage read thresholds (i.e., filtering all taxa per sample that had reads lower than a percentage of the total number of reads for the sample). I applied two thresholds: 0.5% and 1% and calculated taxa loss among these for two levels: genera and species. I also calculated the ratio of bycatch species to target species. To assess whether sampling effort was

representative of the bycatch within Newcastle trap I performed species accumulation curves using the *iNEXT* package (Hsieh, Ma and Chao, 2016) with the function "iNEXT". Species accumulation curves were calculated based on presence absence data. I also calculated two common diversity metrics: species richness and species turnover with the *codyn* package in R (Hallett *et al.*, 2016) to see how these change through time for both pre-filtered and filtered taxonomy after rarefying at the lowest sequencing depth (12918 reads; excluding samples were PCR amplification was not evident). Finally, to test if year and sequencing depth had an effect on species discovery, I fitted a Poisson Generalized Linear Model (GLM) using the "glm" function in base R (v. 4.0.1, R Core Team, 2022) with species richness as a response variable and year and sequencing depth as the predictor variables by reads were log transformed for the model.

3.3.6 Finding invertebrate taxa of agricultural interest

To assess whether certain taxa could be of agricultural interest: whether they are a pest or a beneficial insect, I matched the assigned taxonomy of the species found in my study to four databases: one of the largest pest-beneficials insect databases downloaded from Karp *et al.*, (2018), the EPPO-Q-bank Global Arthropod database (https://gd.eppo.int/), the database of Pollinator interactions (Balfour *et al.*, 2022) (https://www.sussex.ac.uk/lifesci/ebe/dopi/about), the database of Insects and their foodplants (Padovani *et al.*, 2020) and DoPI: The Database of Pollinator Interactions (Balfour *et al.,* 2022). Matched taxa received a tag of four possible categories: pest, predator, parasitoid and pollinator. The search in the databases was conducted using custom functions in R after standardizing the binomial names. For the pollinator database and the EPPO database we used the "*intersect"* function in R to identify shared taxa. Finally, to see whether any genera found in our datasets were recorded for the first time in the UK we matched taxa in our dataset with the UK species inventory database (Raper, 2014) (available at: https://data.nhm.ac.uk/dataset/uk-species-inventory-simplified-copy). I also cross-validated the results by checking the regions of the taxa found through GBIF (https://www.gbif.org/species/) this was done for all possible synonyms (which are already embedded within GBIF).

3.4 Results

3.4.1 Sequencing results

Figure 3. 1 Output from *MultiQC*, green lines represent the median of Phred quality scores per base for each sample. Forward reads (top), Reverse reads (bottom)

PCR success was high with 99% (182/184) samples producing a visible band on a gel. The run produced 320,628,559 reads (excluding PhiX), and after demultiplexing 302,333,205 were retained. After filtering, denoising, merging and chimera removal a total of 196,493,188 reads (23460 ASVs) were found for the first filtering criteria used in *DADA2*, 179,452,994 (22110 ASVs) for the second and 149,140,646 (21400 ASV's) for the third. There were no significant differences in either the number of ASV's or number of taxa assigned for the three datasets despite the difference in total number of reads ($F_{1,366}$=0.85, p=0.35; see Table B. 3) as the taxa

found in all datasets corresponded to the same genera. Therefore, I proceeded with the dataset that had stringent (and most conservative) filtering criteria that resulted in 149,140,646 reads. Reads per sample excluding the controls and the two samples that did not show a band ranged from 12,918 to 2,161,538 (see Table B. 3 in the Appendix B for reads per sample throughout the *DADA2* pipeline). The median was at 810,547 reads (1st quartile: 487,977, 3rd quartile: 1,053,536 whilst the mean was 810,547 (+- 449,377). Additionally, only two extraction negatives had reads passing the initial filters with 1 and 10 reads respectively. Surprisingly, for certain positives (including DNA extraction and PCR positives) in libraries no reads passed the filtering criteria and overall, of the ones that did there was a low number of reads passing (mean of 1000 reads). No reads from the positive controls were found in any of the samples suggesting that there was minimal to no sample-to-sample contamination throughout the lab work.

3.4.2 Assigned taxonomy

From the 149,140,646 reads: 98,353,488 reads could be assigned to species and there were 47,711,113 reads that could not be assigned to any level of taxonomic hierarchy and 3,076,045 were assignment could be done at the superkindom level. From the 98,353,488 reads that have been identified at least a genus/species level 97,989,753 (~99.6%) were assigned to Arthropoda which was the main target group of this study. Non-Arthropod taxa belonged to belonged to the following phyla: Annelida, Ascomycota, Mollusca, Basidiomycota, Chordata and Nematoda. Most of the reads within Arthropoda were assigned to the class Insecta (95,891,644 reads ~97% of total reads), some were assigned to Arachnida (0.01%) and very few to Collembola (0.001%). There were 12 insect orders assigned with the most abundant order being Diptera and with more than 84% reads assigned to it. This was followed by Psocoptera (3%), Hemiptera (1.5%) and Coleoptera (1.3%). For Arachnida, the most abundant order was Araneae followed by Trombidiformes.

3.4.3 Sample sequencing thresholds and taxa loss

I found 447 genera and 856 species for Insects whilst for Arachnids I found 13 genera and 15 species before applying any filtering criteria based on sample thresholds. After applying sample sequence thresholds, the number of species fell to 445 and 373 for the second and

third respectively. For genera level the numbers were 248 and 209 respectively (Table 3. 1). The taxonomic composition per year also changed after applying the filtering criteria with 8 of the taxonomic orders in the dataset completely disappearing (see Figure 3. 2B). However, Diptera dominated the dataset in every year of the time series irrespective of the filtering criteria applied.

Filtering threshold	Number of genera	Number of species
No filter	447	856
0.5% threshold	248	445
1% threshold	209	375

Table 3. 1 Number of unique genera and species found within the dataset before and after the application of sample filtering thresholds.

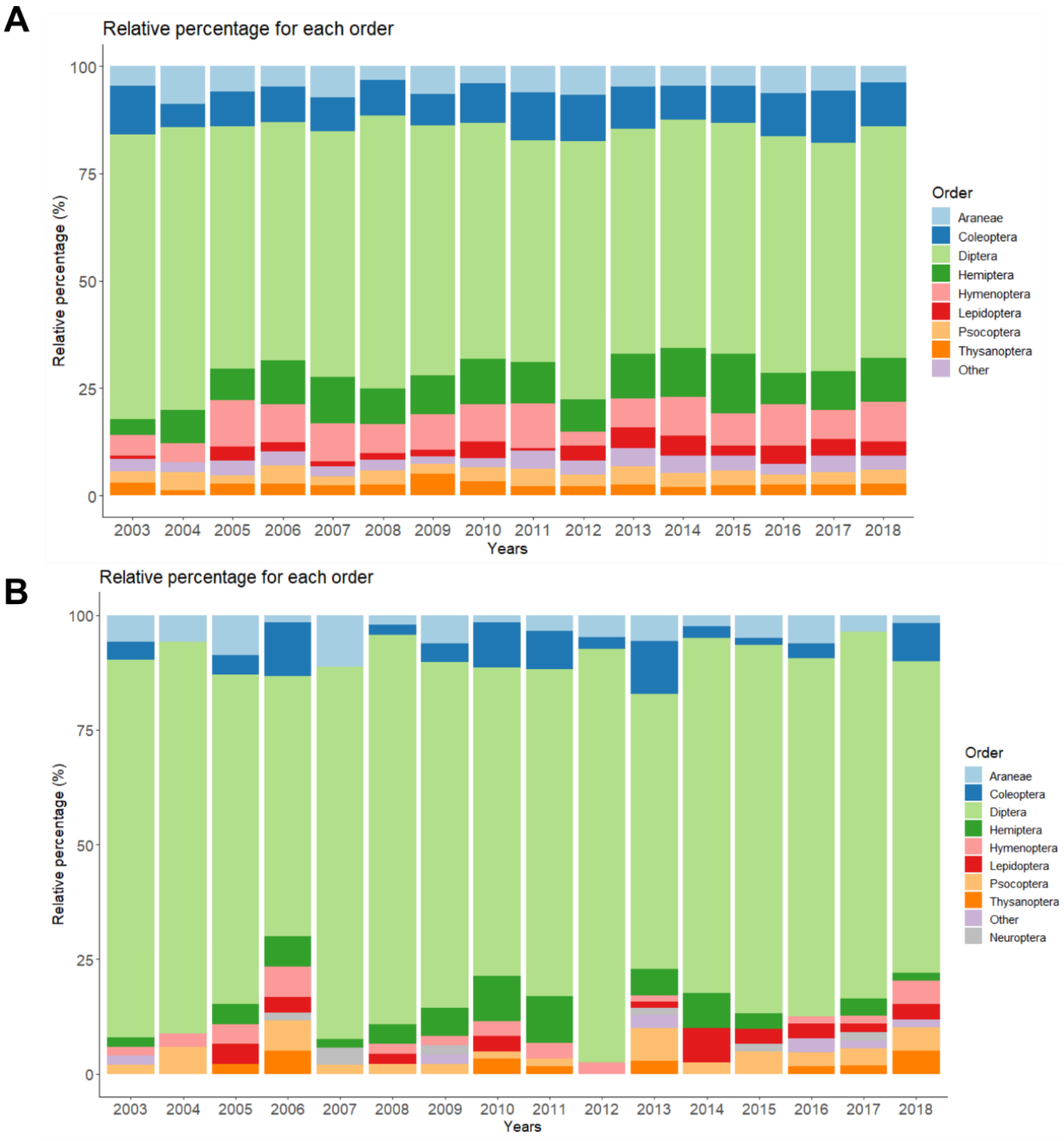


Figure 3. 2 A) Percentage of order present in every year before applying any filtering criteria. B) Percentage of order present in every year after applying filtering at the 1% threshold.

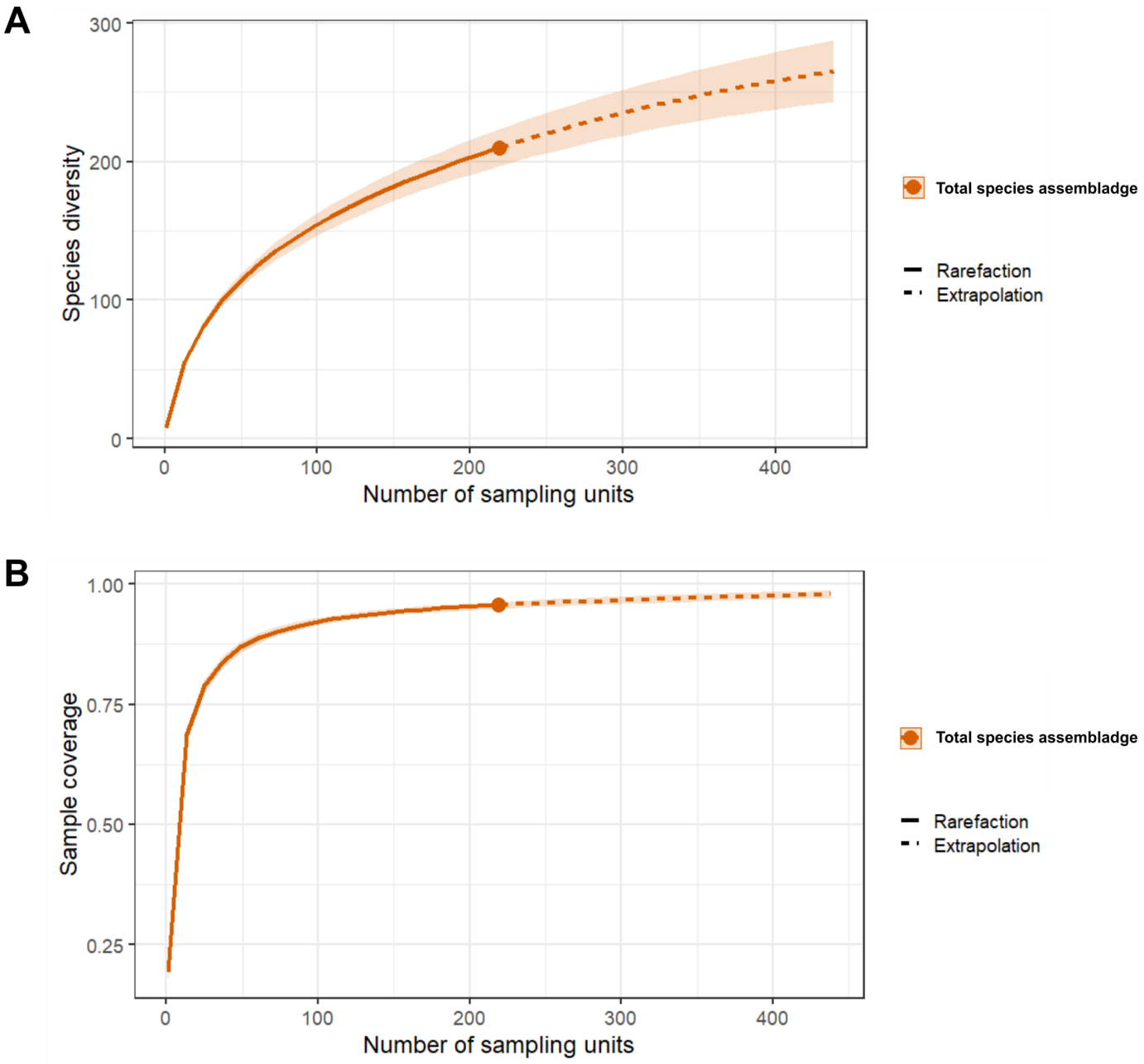


Figure 3. 3 A) Diversity as Hill-species richness of genera detected during the experiment after filtering. B) Associated sample coverage across the sampling units. Solid lines represent observed diversity and dashed lines extrapolated diversity.

3.4.4 Effect of year and sequencing depth on richness

Richness varied across and within years, with a similar pattern across the months for every year (see Figure B. 1). There is an increase in richness from May to June or July followed by a decrease then a second peak in September for some years, except for three years 2009,2017,2015. For all the major orders richness shows an increasing trend (see Figure 3. 4 A,C) when looking at the non-filtered dataset. However, with the application of sample

threshold filters, the trend changes for some orders as there very few species now presented at each order level, with the exception of Diptera (see Figure B. 2). Diptera richness showed an increasing trend with time that was captured for either filtered or unfiltered datasets. Turnover was high for most orders (see Figure 3. 4B). Estimated richness was calculated at 550 (+- 30) genera for the filtered dataset (Figure 3. 3A). Year had a significant positive effect on richness, but sequence depth did not, the interaction between sequencing depth and year was also not significant (*p*= 2e-16, see Table 3. 2).

Coefficients	Estimates	Standard error	Z value	Pr(>z)
Intercept	-5.512e+02	2.009e+02	-2.743	0.00670
Sequencing depth	2.886e-04	3.126e-04	0.906	0.35
Year	2.794e-01	9.996e-02	2.795	0.00575**
Sequencing depth: Year	-1.445e-07	1.554e-07	-0.930	0.35

Table 3. 2 Coefficients and significance of effects on species Richness. Residual deviance calculated 256 on 180 degrees of freedom, AIC= 1003, Number of Fisher Scoring iterations: 4

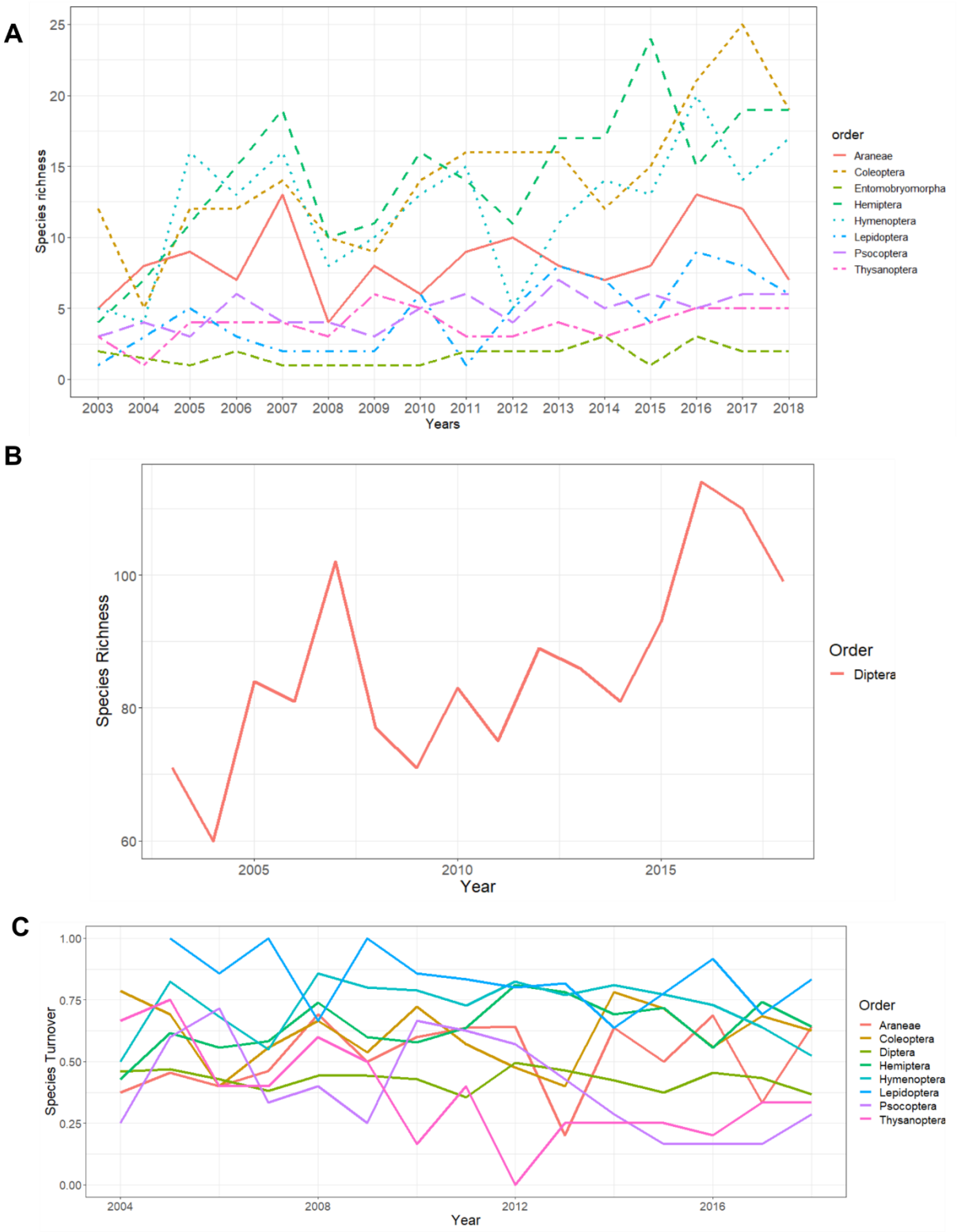


Figure 3. 4 Diversity as species richness over the years for the major orders that were presented throughout all the years after filtering. A) Richness for major orders identified throughout the time series. B) Richness of Diptera over time C) Species turnover over time for the major orders identified.

3.4.5 Bycatch species of agricultural interest

The mean of the ratio of bycatch species to target taxa was 8:1 (min 0.5, max 35) for genus level and 10:1 (min 0.5, max 54) for species level. In total 96 pests and beneficials (predators and parasitoids) species could be matched with a specific category. For comparisons with the EPPO global database 228 species of our dataset were found to be within it. However, the latter database did not include any tags therefore it was not possible to find specific categories for the taxa. A large issue was that the majority of taxa found for pest, predators and parasitoids did not have a tag (taxa were not found in the databases with a specific tag for each category) even though some are fairly known taxa. For example, in the pre-filtered dataset there were both aphid parasitoids and predators but only 5 parasitoids received a tag. I identified 29 species of Braconid wasps belonging to 17 unique genera of which all are known to parasitize aphids. Predators totalled 21 unique genera (1 Neuroptera, 19 Coleoptera, 9 Araneae). For other common taxa found the majority could not be classified for specific categories. However, most of the pollinators belong to Diptera, Hymenoptera and Lepidoptera. In total 74 genera of pollinators found with the majority belonging to the Diptera order (40 genera) followed by Hymenoptera (14), Lepidoptera (8), Coleoptera (7), Hemiptera (3) and Thysanoptera (2). Many identified taxa could also be identified as common pests like thrips, aphids, shield bugs, leafhoppers and planthoppers (see Figure 3. 5; Table B. 4 for the list of taxa where a category was found). Finally, I've found 13 taxa on the filtered dataset for which no record could match the UK species inventory list. To validate these results, we crossed check for synonyms of the species and used Blastn on the corresponding ASVs to see whether these could be misidentifications within the databases. One synonym was found for *Brassicogethes aeneus*, which is the common pollen beetle widely distributed in the UK. Additionally, I've found the species *Calopteryx maculata* a common damselfly of North America, however this could be attributed to a reference database error. As my reference database included mainly Metazoa, a miss annotation of *Calopteryx malculata* was found, as all the ASV's for this particular taxon could be identified as bacteria from the Blastn search. A total of 42 species were not found within the species inventory list of the UK for the pre-filtered dataset, after filtering 13 taxa only remained as mismatches. From these 13 taxa, one was found to be native in North America and Australia the barklouse *Ectopsocus californicus*, the rest were found to be native in mainland Europe but did not have a single record for the

UK in GBIF. Most belonged to the Diptera order (11) and one only belonged to Hymenopteran parasitoids: *Protapanteles fulvipes* (See Table B. 5 for all species).

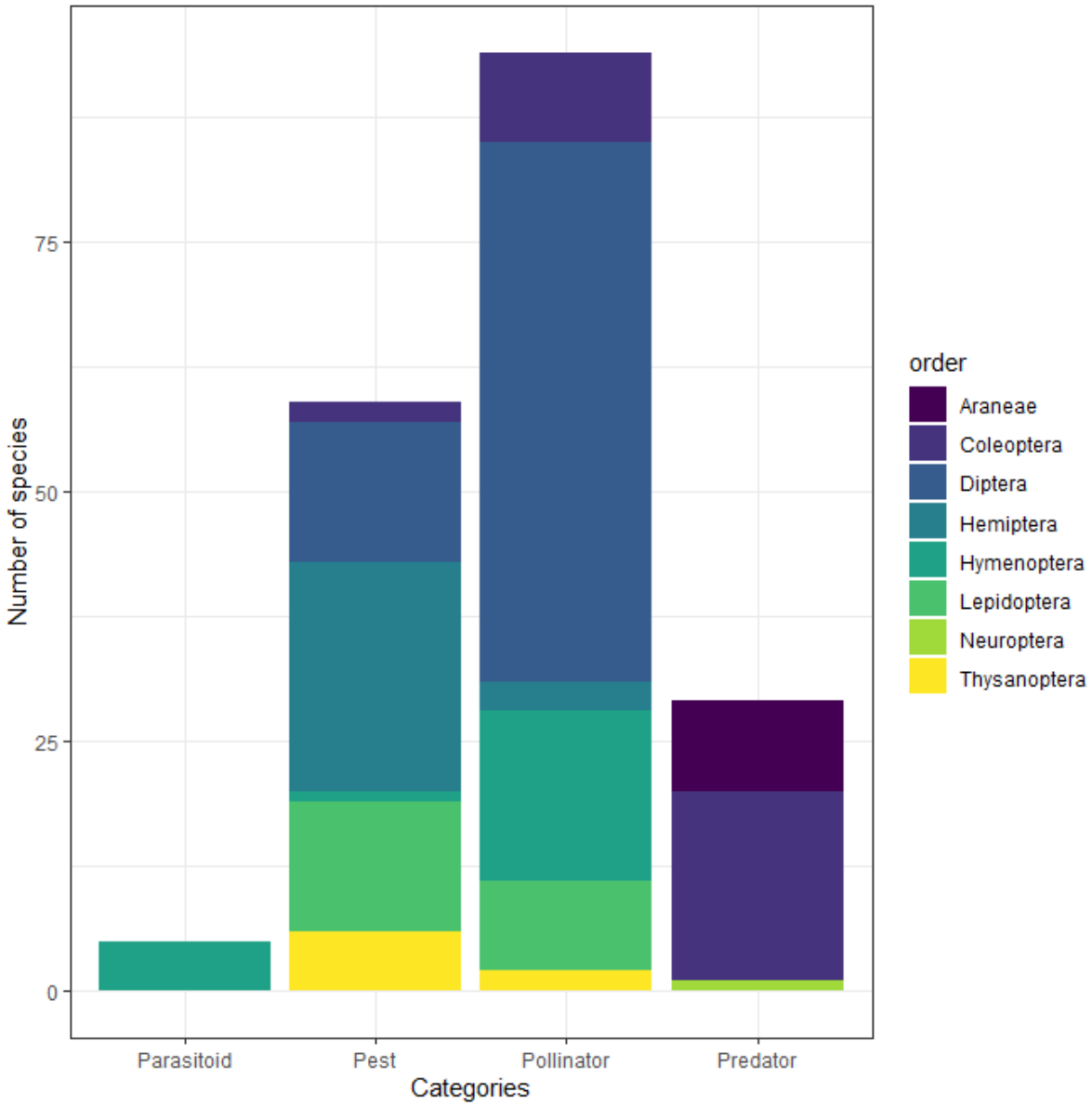


Figure 3. 5 Number of species in each category (Parasitoid, Pest, Pollinator and Predator) as identified from the databases. Colours correspond to the different orders.

3.5 Discussion

In this study I demonstrate that it is possible to identify bycatch taxa, a fraction of biodiversity that often remains untouched in insect monitoring schemes, by using DNA-metabarcoding. I successfully identified archival samples from the historical Rothamsted Insect Survey suction traps over a period of 16 years, doing so non-destructively. My study highlights the potential to identify and construct time-series for hundreds of insect taxa, many of which have not been studied in the context of insect declines. I further showed that bycatch species can include many agriculturally important pests and beneficial taxa that are not the target of the insect monitoring schemes, which shows promise for scaling up the breadth of taxa being monitored. Overall, this study validates the use of DNA-metabarcoding for scaling up biodiversity monitoring and particularly insect monitoring by combining it with an already established monitoring scheme in the UK that focuses mainly on aphids.

3.5.1 Objective 1) Illuminating archived bycatch diversity within RIS

By using DNA-metabarcoding I recovered 856 insect species belonging to 414 genera in the full dataset. This was the first attempt to my knowledge to identify complete bycatch samples of RIS suction traps by using DNA metabarcoding. The estimated species richness was around 550 genera, close to the number of genera found in this study. This however, is probably an underestimate of real bycatch diversity. As the complete time series for the Newcastle trap has over 2500 samples for the period of 2003-2018 there are probably many more species to be recovered. Bycatch unsurprisingly, consisted of a wide diversity of taxa (more than 10 insect orders) which is why it also makes it is difficult to identify via traditional taxonomic means due to the breadth of taxonomic expertise required. Additionally the complexity increases even more as passive monitoring techniques can capture thousands of individuals which makes sorting those samples very time consuming. Molecular techniques can overcome such limitations as taxonomic expertise is not required (Wang *et al.*, 2018). The complete dataset comprised 12 insect orders belonging to 126 different families. Most insects belonged to the Diptera order with more than 90% of recovered taxa. This was expected as Diptera are some of the most common catches within suction traps (Blandenier, 1998). Besides insects, we have also found 43 species of spiders, many of which were expected as "ballooning" spiders (e.g. *Erigone atra*) are commonly found catches in suction traps

(Blandenier, 1998) and pseudoscorpions which were probably “hitchhiking” on other taxa. However, the number of species recovered drastically changed based on the filtering criteria applied on this study. Insect monitoring schemes and particularly RIS process their samples in a way that is prone to sample-to-sample contamination. For example, in RIS, it is uncertain whether sorting trays are cleaned after each sample identification. Here I focus only on Arthropoda and particularly insects, however reads were assigned to other common contaminants as well such as cattle or birds or non-eukaryotic taxa. In my study bird DNA was expected as feathers were found within the samples but also because birds can rest at the suction traps, other contaminants like cattle could be found from carrion flies (Rodgers *et al.*, 2017). Non-Arthropod reads however were less than 0.0001% of the total reads. As our study focuses on insects and arthropods, we could filter all non-Arthropoda taxa and our control samples didn’t show any cross contamination. However, contamination from arthropods from other samples is possible at RIS. I previously showed that this type of contamination can be controlled to an extent by using sample read thresholds decreasing false positive rates but increasing false negatives (see Chapter 2.3.2). Sample read thresholds can be applied in many ways (see Drake *et al.*, 2022) here I applied sample-based thresholds which resulted in a drop of overall taxa from 800 to 400 (see Table 3. 1). Because taxonomic information is not available for the bycatch the extent of which false positives/negatives were minimised is uncertain. For aphids, I’ve found that a 1% threshold would drop the average false positive rate from an average of 4 to 0.1 (see Chapter 2.3.2). Similarly, other studies have found that thresholds between 0.5 -1% offer a good true/positive rate (Drake *et al.*, 2022). However, it’s very likely that true positives disappear as commonly found insects within RIS traps like ladybirds could not be found after the application of filtering techniques. Future research efforts should focus on validating this by combining DNA-metabarcoding with morphological identifications or avoiding contamination issues that are amplified by PCR by using metagenomic approaches (Ji *et al.*, 2020). As many monitoring schemes were not designed for DNA based applications like DNA-metabarcoding this is an important caveat for such samples. Therefore, taxonomic information can add value to the reliability of DNA-metabarcoding and other molecular techniques like metagenomics. Here, I processed the samples in a non-destructive way therefore it would be possible for future research efforts to revisit these samples and assess the reliability of DNA-metabarcoding. Overall, I show that bycatch taxa can be identified via DNA-metabarcoding and in the case of RIS the bycatch

diversity is high with an average 10 bycatch species for every target taxon (here aphids) within a sample meaning that bycatch species dominate samples at RIS.

3.5.2 Objective 2) Highlighting beneficial insects and potential pests

Unsurprisingly, and in line with previous work, bycatch species include many taxa that can have an economic importance for agriculture (Hribar, 2020). A major limitation is that there exists, no large database with information on specific functional groups for each taxon. The dataset used here first from Karp *et al.,* (2018), was the largest of its kind yet many insects known to provide ecosystem services or disservices such as pest control and pests were not present. Here, I focused on aphids where I found more than 29 species of parasitoids (with the most abundant genera being *Aphidius* and *Praon)* and 15 species of predators belonging to 4 orders Coleoptera, Neuroptera, Hemiptera and Diptera. However, I also found many pollinators including bees, wasps and syrphid flies. Some include cosmopolitan species like the drone fly *Eristalis tenax* or hymenoptera including *Bombus bombus*. I also identified agricultural pests beyond the target taxa most of which belonged to the Hemiptera and Diptera orders. Additionally, I found insects that have not been recorded in the UK before, such as *Ectopsocus californicus* a species of bark lice that is typically found in North America and Australia and more than another 9 arthropods (common to Europe, or America). In my case, these should be validated with an expert taxonomist retrospectively, as DNA-metabarcoding can have many biases and taxonomic expertise is required to validate potential non-native invasive taxa. But this shows how bycatch catches can be used for expanding knowledge of the geographic distribution of insects (Hribar, 2020). Although it would be very expensive to process all the daily samples from RIS, metabarcoding is scalable and cheap when compared to other methods such as taxonomy or metagenomics (Bista *et al.*, 2018). Therefore, such taxa could be incorporated in future monitoring efforts based on metabarcoding. This can also lead to the early detection of newly invading insects pests (Piper *et al.*, 2019) or even description of newly found species (Spears and Ramirez, 2015). The breadth of biodiversity information that is lost by not analysing or even throwing away bycatch species is disheartening particularly because there is a lot of uncertainty regarding the state of insect diversity in the world (van Klink *et al.*, 2020).

3.5.3 Objective 3) Bycatch as a novel source for biodiversity time series

Long-term data are important to understand biodiversity changes throughout the years, yet insect data for many groups and regions are scarce. Previously, efforts have been made to process bycatch species taxonomically to help understand changes in phenology or abundance within RIS but most have used subsamples of specific groups like Diptera (Grabener *et al.*, 2020) or subsampled dates. Here by using a 16-year time series of unknown bycatch species. I show that it is possible to create time series for hundreds of insect taxa with DNA-metabarcoding. Additionally, I show that it is possible to estimate common diversity metrics like richness or turnover to understand biodiversity changes for multiple species assemblages (see Figure 3. 4). Species accumulation curves show that the sample size covers a significant amount of the estimated species richness (Figure 3. 3.B). However, as the focus was on a subsample of the dates within each month, these should not be seen as real patterns of biodiversity but simply showcase the potential of understanding biodiversity patterns for groups of insects largely unstudied (like Hymenoptera and Hemiptera). Additionally, the presence of beneficial insects like aphid parasitoids throughout the years could make it possible to construct time series of species interactions for aphids and parasitoids by using co-occurrence networks. RIS could be seen as a unique opportunity to not only construct multi species time series but capture changes of species interactions through decades (Petsopoulos et al., 2021). However, metabarcoding has its drawbacks too as it does not yet give quantitative information (Ji et al., 2020). But long-term presence-absence data can still be relevant, for example to study migration patterns and phenological shifts (Grabener *et al.*, 2020). Year and sequencing depth had a positive effect on species richness with species richness increasing over the years for most orders. For aphids at RIS I haven't found significant effect on year for species richness which might be an artefact of real differences of richness for the dates sampled and bias from sequencing depth as it can drastically alter the resultant community (Shirazi, Meyer and Shapiro, 2021). Further validation with more samples would enable us to understand how diversity patterns change throughout the years. Finally, RIS has a network of suction traps throughout the UK with archival bycatch samples for each location. This means that it could be possible to construct time series that are spatially and temporally replicated throughout the UK for hundreds of insect taxa, providing insect data to better understand how insect populations have been changing since the 60s.

3.6 Conclusions

My study highlights the potential of bycatch catches within insect monitoring schemes as a unique source of insect data. This is the first attempt to identify archival bycatch samples from RIS with DNA-metabarcoding, we do so by analysing samples spanning a 16-year period from 2003 to 2018. I show that a high diversity of insect taxa is being "missed" from monitoring schemes as bycatch remains untouched or not processed due to taxonomic, financial, or logistic impediments. But this can be done by DNA-metabarcoding opening a treasure vault of insect data, particularly for taxa that might be important for agriculture like beneficials and pests. However, contamination is still one of the major limitations when it comes to processing samples from insect monitoring schemes. Therefore, coupling molecular data with morphological can increase the reliability of such methods. Overall, bycatch can be seen as invaluable biodiversity information for insect research even more for insect decline research and DNA-metabarcoding a promising tool for accessing such biodiversity information.

3.7 References

Andrews, S. (2010) 'FastQC: A Quality Control Tool for High Throughput Sequence Data'. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Balfour, N. J. *et al.* (2022) 'DoPI: The Database of Pollinator Interactions', *Ecology*. John Wiley & Sons, Ltd, 103(11), p. e3801. doi: 10.1002/ECY.3801.

Bell, J. R. *et al.* (2015) 'Long-term phenological trends, species accumulation rates, aphid traits and climate: five decades of change in migrating aphids', *Journal of Animal Ecology*. Edited by K. Wilson. John Wiley & Sons, Ltd (10.1111), 84(1), pp. 21–34. doi: 10.1111/1365-2656.12282.

Bista, I. *et al.* (2018) 'Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples', *Molecular Ecology Resources*, pp. 0–2. doi: 10.1111/1755-0998.12888.

Blandenier, G. (1998) 'Ballooning spiders caught by a suction trap in an agricultural landscape in Switzerland'.

Buchholz, S. *et al.* (2011) 'Bycatches of ecological field studies: Bothersome or valuable?',

Methods in Ecology and Evolution, 2(1), pp. 99–102. doi: 10.1111/j.2041-210X.2010.00051.x.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/1471-2105-10-421/FIGURES/4.

Drake, L. E. *et al.* (2022) 'An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 13(3), pp. 694–710. doi: 10.1111/2041-210X.13780.

Ewels, P. *et al.* (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*. Oxford Academic, 32(19), pp. 3047–3048. doi: 10.1093/BIOINFORMATICS/BTW354.

Grabener, S. *et al.* (2020) 'Changes in phenology and abundance of suction-trapped Diptera from a farmland site in the UK over four decades', *Ecological Entomology*. Blackwell Publishing Ltd, 45(5), pp. 1215–1219. doi: 10.1111/EEN.12873.

Hallett, L. M. *et al.* (2016) 'codyn: An r package of community dynamics metrics', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 7(10), pp. 1146–1151. doi: 10.1111/2041-210X.12569.

Harrington, R. (2013) 'The Rothamsted Insect Survey strikes gold', *Antenna*, 38(3), pp. 158–166. Available at: www.rothamsted.ac.uk/examine.

Hribar, L. J. (2020) 'Expanding Basic Entomological Knowledge by Using Mosquito Surveillance Bycatch', *Annals of the Entomological Society of America*, 113(6), pp. 439–446. doi: 10.1093/aesa/saaa014.

Hsieh, T. C., Ma, K. H. and Chao, A. (2016) 'iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers)', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 7(12), pp. 1451–1456. doi: 10.1111/2041-210X.12613.

Ji, Y. *et al.* (2013) 'Reliable, verifiable and efficient monitoring of biodiversity via

metabarcoding', *Ecology Letters*, 16(10), pp. 1245–1257. doi: 10.1111/ele.12162.

Ji, Y. *et al.* (2020) 'SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes', *Molecular Ecology Resources*, 20(1), pp. 256–267. doi: 10.1111/1755-0998.13057.

Karp, D. S. *et al.* (2018) 'Crop pests and predators exhibit inconsistent responses to surrounding landscape composition', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(33), pp. E7863–E7870. doi: 10.1073/PNAS.1800042115.

Kitson, J. J. N. *et al.* (2019) 'Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding', *Molecular Ecology*, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

van Klink, R. *et al.* (2020) 'Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances', *Science*, 368(6489), pp. 417–420. doi: 10.1126/science.aax9931.

Komoroske, L. M. and Lewison, R. L. (2015) 'Addressing fisheries bycatch in a changing world', *Frontiers in Marine Science*, 2(OCT), pp. 1–11. doi: 10.3389/fmars.2015.00083.

Leray, M. *et al.* (2013) 'A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents', *Frontiers in Zoology*, 10(1), pp. 1–14. doi: 10.1186/1742-9994-10-34.

Leray, M., Knowlton, N. and Machida, R. J. (2022) 'MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences', *Environmental DNA*, 4(4), pp. 894–907. doi: 10.1002/edn3.303.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), pp. 10–12. doi: 10.14806/EJ.17.1.200.

McCravy, K. (2018) 'A Review of Sampling and Monitoring Methods for Beneficial

Arthropods in Agroecosystems', *Insects*, 9(4), p. 170. doi: 10.3390/insects9040170.

Oberacker, P. *et al.* (2019) 'Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation', *PLoS Biology*, 17(1), pp. 1–16. doi: 10.1371/journal.pbio.3000107.

Padovani, R. *et al.* (2020) 'Insect species richness for each plant species and insect-plant interactions from the Database of Insects and their Food Plants [DBIF] version 2'. NERC Environmental Information Data Centre. Available at: https://doi.org/10.5285/33a825f3-27cb-4b39-b59c-0f8182e8e2e4.

Pérez-Rodríguez, J., Shortall, C. R. and Bell, J. R. (2015) 'Large-scale migration synchrony between parasitoids and their host', *Ecological Entomology*. John Wiley & Sons, Ltd (10.1111), 40(5), pp. 654–659. doi: 10.1111/een.12241.

Petsopoulos, D. *et al.* (2021) 'Using network ecology to understand and mitigate long-term insect declines', *Ecological Entomology*. doi: 10.1111/een.13035.

Piper, A. M. *et al.* (2019) 'Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance', *GigaScience*. Oxford University Press, 8(8), pp. 1–22. doi: 10.1093/gigascience/giz092.

Raper, C. (no date) 'UK Species Inventory - Simplified copy [Data set]', *Natural History Museum*. doi: https://doi.org/10.5519/0070624.

Rocha-Ortega, M., Rodriguez, P. and Córdoba-Aguilar, A. (2021) 'Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction', *Ecological Entomology*. John Wiley & Sons, Ltd, 46(4), pp. 718–728. doi: 10.1111/EEN.13027.

Rodgers, T. W. *et al.* (2017) 'Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 17(6), pp. e133–e145. doi: 10.1111/1755-0998.12701.

Shirazi, S., Meyer, R. S. and Shapiro, B. (2021) 'Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding', *Ecology and Evolution*. John Wiley & Sons, Ltd, 11(22), pp. 15766–15779. doi: 10.1002/ECE3.8239.

Skvarla, M. J. and Holland, J. D. (2011) 'Nontarget insects caught on emerald ash borer

purple monitoring traps in Western Pennsylvania', *Northern Journal of Applied Forestry*, 28(4), pp. 219–221. doi: 10.1093/njaf/28.4.219.

Spears, L. R. *et al.* (2016) 'Pheromone Lure and Trap Color Affects Bycatch in Agricultural Landscapes of Utah', *Environmental Entomology*, 45(4), pp. 1009–1016. doi: 10.1093/ee/nvw085.

Spears, L. R. and Ramirez, R. A. (2015) 'Learning to Love Leftovers', *American Entomologist*. Narnia, 61(3), pp. 168–173. doi: 10.1093/ae/tmv046.

Team, R. C. (2022) 'R: A language and environment for statistical computing', *R foundation for Statistical Computing, Vienna, Austria.* Available at: https://www.r-project.org/.

Wang, W. Y. *et al.* (2018) 'Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing', *Molecular Ecology Resources*. John Wiley & Sons, Ltd (10.1111), 18(3), pp. 490–501. doi: 10.1111/1755-0998.12751.

Yang, C. *et al.* (2021) 'Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 12(7), pp. 1252–1264. doi: 10.1111/2041-210X.13602.

Yu, D. W. *et al.* (2012) 'Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring', *Methods in Ecology and Evolution*, 3(4), pp. 613–623. doi: 10.1111/j.2041-210X.2012.00198.x.

Chapter 4: Metabarcoding bycatch insect species from a nationwide monitoring scheme: The Fera Yellow Water Pan Trap Network (YWP)

Declaration: Samples in this chapter come from the monitoring scheme of the Fera Yellow Water Pan Trap Network. The sampling was mainly managed by Fera. Therefore, I would like to acknowledge all the staff responsible for collecting such samples and the farmers providing such samples. However, sample choice, sample processing and data analysis were all done by me.

4.1 Abstract

Bycatch is ubiquitous among insect monitoring schemes. Yet, despite its wider biodiversity research purposes, bycatch is usually not identified or analysed and, in some cases, even thrown away. As insect monitoring schemes serve a very specific purpose, for example monitoring specific pests or forecasting potential viral risk transmission, bycatch is usually not identified. Other reasons include the logistics, financial and the taxonomic expertise required to identify such samples. Next Generation Sequencing approaches can overcome many of the obstacles for simultaneously processing thousands of insect specimens that arise from nationwide biomonitoring schemes. In this study, I focus on bycatch diversity among the Fera Yellow Water Pan Trap Network (YWP) that focuses on monitoring aphids (and the viruses they carry) in potato fields at a farm level. To examine spatial and temporal variation in insect diversity, I use DNA-metabarcoding to identify bycatch taxa among the nationwide monitoring scheme with more than 80 locations across the UK. I found high bycatch diversity with more than 900 taxa. Furthermore, I compared the effect of different sequencing efforts on the recovery of bycatch diversity by comparing two Illumina platforms with a 20-fold difference in sequencing depth. I found high similarity with over 90% of the total taxa shared between the platforms. Finally, I show complex patterns of biodiversity with high temporal and spatial turnover (mean of 0.90 Jaccard index of dissimilarity) among the field samples despite these being a monoculture. The turnover itself is driven mainly by specific taxa that were the most abundant in the datasets. This study shows the potential of DNA-metabarcoding for analysing and identifying bycatch diversity and its patterns at the nationwide level.

4.2 Introduction

Recent studies have shown that insects are declining in many parts of the world. Yet, for many regions there is still a lot of uncertainty regarding the state of insect populations (van Klink *et al.*, 2020). A major reason for this is that there is a lack of data, particularly long-term data for many regions. Typically, sources for understanding insect declines come from long-term monitoring schemes; examples in the UK include the Butterfly Monitoring Scheme and the Rothamsted Insect Survey (RIS) (Conrad, Fox and Woiwod, 2007; Bell, Blumgart and Shortall, 2020). But there are many other insect biomonitoring schemes that usually serve a very particular purpose (i.e. pest monitoring) that can be used as more general sources of insect data, and this can also be done retrospectively by looking at archival specimens from monitoring schemes (Zizka *et al.*, 2022). Such invertebrate monitoring schemes usually capture thousands of insect species and individuals yet, as the focus can be pest monitoring or forecasting, only a fraction of those caught go on to be identified. The remaining sample, usually called "bycatch", often remains untouched or even thrown away, with very few exceptions where bycatch is archived and stored. In the UK, one exception is RIS, which has been storing all of their daily catches since the 1960s (Harrington, 2013). These can be viewed as unique biobanks of insect data that can be used to understand insect declines for numerous insect taxa (Petsopoulos *et al.*, 2021). Bycatch from monitoring schemes comprises of important biodiversity information that has been used to identify and describe new species, detection of newly discovered pests and invasive species (Spears and Ramirez, 2015; Hribar, 2020). Despite its potential use as a novel source for insect data, to date such samples have largely been untouched as it can be very difficult to identify all species due to the wide taxonomic expertise needed. Financial, logistic and storage limitations also apply as it is typically expensive to process thousands of specimens via taxonomic means and the space and personnel needed to do so further limits the accessibility of bycatch.

New emerging developments in Next Generation Sequencing (NGS) approaches for biodiversity assessment can alleviate existing problems in insect biomonitoring by offering alternative cost-effective methods that are not reliant on taxonomic expertise for mass identification of insect taxa (Zhou *et al.*, 2013). NGS approaches are revolutionising the way

we can assess biodiversity. By coupling short-fragments of DNA with NGS, it is possible to process thousands of specimens (Ji *et al.*, 2013). Three main approaches have been used so far: DNA barcoding (Shokralla *et al.*, 2014), DNA-metabarcoding and metagenomics (Derocles *et al.*, 2018). DNA barcoding relies on amplifying short fragments of DNA from one specimen with thousands of samples being able to fit within one sequencing experiment (Srivathsan *et al.*, 2021). Its use varies from simply identifying species to relatedness analysis between difficult to delimit species via traditional means and phylogenetic analyses (Hebert *et al.*, 2003). DNA-metabarcoding similarly makes use of short fragments but the sample unit can compromise thousands of specimens, for example a sample from an entomological pan trap. In particular, it can be used for the rapid assessment of thousands of specimens and samples (Piper *et al.*, 2019). Yet, DNA-metabarcoding has its limitations, the PCR amplification step induces biases that can make the data non-quantitative i.e. abundance information is very difficult to obtain from DNA-metabarcoding experiments, something that can hinder its use for routine monitoring where abundance information is regarded important (Martoni *et al.*, 2022). But there are methods to overcome such obstacles, for example by using correction factors (Krehenwinkel *et al.*, 2017) or spiking with known amounts of DNA (Ji *et al.*, 2020). Lastly, metagenomics is a PCR-free approach where abundance information can be retained. Here, the extracted DNA from a sample of multiple taxa is sequenced directly, avoiding any PCR biases. Yet, there can be still biases from other factors such as unknown copies of mitochondrial DNA (mtDNA) for the organisms within the sample, but most importantly the cost of metagenomics can be prohibitively expensive for routine monitoring purposes or for very degraded samples like archival collections (but see (Ji *et al.*, 2020)). All such approaches offer great promise for biodiversity assessment particularly in the context of insect declines, where fast and scalable methods are needed (Zizka *et al.*, 2022). Due to its scalability, ease of use and cost-efficiency, DNA-metabarcoding can been seen as one of the most promising tools for making bycatch samples accessible for further research. Yet, evidence of its use within the context of insect monitoring of bycatch species is lacking.

Here, I evaluate the potential of DNA-metabarcoding for analysing insect biodiversity from bycatch as part of a nationwide monitoring scheme in the UK: The Fera Yellow Water Pan Trap Network (YWP). First, I evaluate the potential of DNA-metabarcoding to identify bycatch species within YWP and examine the biodiversity information generated. I then compare the

effect of sequencing depth by analysing datasets obtained from two different sequencing technologies: the Illumina MiSeq and NovaSeq with output varying 20-fold. Finally, I show the utility of the approach by comparing UK biodiversity patterns across the different sampled regions throughout the sampling year.

4.3 Materials and Methods

4.3.1 The Fera Yellow Water Pan Trap Network (YWP)

YWP consists of a network of pan traps within potato fields that are used to monitor aphids and inform farmers about aphid populations density and assess viral transmission risk (https://aphmon.fera.co.uk/). Fera sends pan traps which farmers put within their potato fields and keep for a week without emptying. After a week has passed the farmers send the samples inside the tubes provided back to Fera for identification. All aphids are identified to the lowest possible taxonomic level, in some cases ladybirds are also identified. However, the bycatch is usually thrown away. Sorting of the samples might happen in the same trays but each of these trays is bleached after processing each sample, this might lead to some contamination during the sample sorting phase. Every year a different number of traps has been set up and monitored as the scheme depends on voluntary uptake by farmers. There can be up to 100 sample localities and sampling takes place in May and can continue up to October. It stops when farmers stop sending samples. Therefore, each field can have a different number of samples throughout the period. The scheme launched in 2003 and continued up to 2020 from which point it was discontinued.

4.3.2 Sample collection and DNA extraction

In 2019, Fera collected and stored 376 samples in 99% ethanol from 93 unique potato fields. From these I subsampled 192 samples based on which localities had repeated samples throughout the period of June 2019 to August 2019 (83 unique fields). Some samples had been sampled for all the months, but some were not. Therefore all sites where monthly samples were taken were used. I aimed for samples that were at least present two months or more (see Appendix C; Table C. 1 for more details and Figure 4. 1 for a graphical distribution of the locations of the fields sampled).

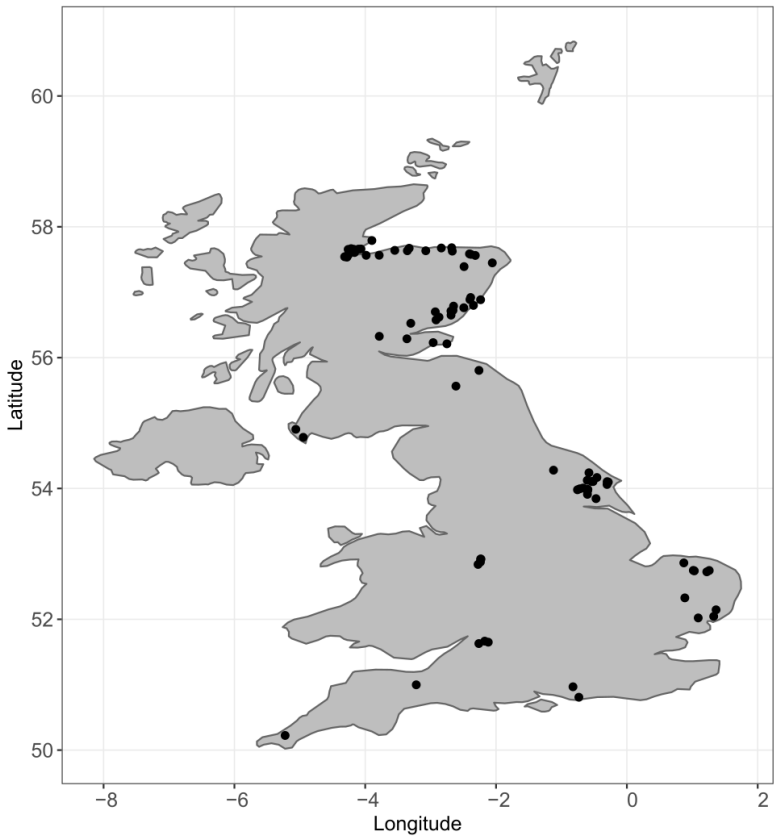


Figure 4. 1 Distribution of 83 unique fields included in this study. Each field is represented by a black dot.

I extracted the DNA using a non-destructive method following the same methods as described in Chapter 2.3.1, except where otherwise stated. Briefly, I used a bead-based protocol [protocol #6.3] (Oberacker *et al.*, 2019) with slight modifications on the amount of lysis volume used to adjust it for different sample sizes (see Table C. 2). Digestions were carried either in 50 mL tubes or 100 mL tubes if sample volume was large for 3 hours. After that, 62 µL of lysate was transferred to 96 well plates (irrespective of initial lysis volume). This was done to standardise volumes and use the same volumes and steps as the protocol. Finally for each plate, I included a DNA extraction positive and a DNA extraction negative. The DNA extraction positive was tissue from Salmon (*Salmo salar*), the DNA extraction negative included all reagents used for the DNA extraction without any tissue, and simply added molecular grade water.

4.3.3 PCR amplification

I followed the nested-tagging method by Kitson *et al.*, (2019) that uses a combinatorial indexing approach to multiplex samples. I targeted a 313 bp fragment of the cytochrome C oxidase subunit I barcode regions with the primers mLCOintF and jgHCO2198 (see section 2.3.4 for the reasoning behind the primer choice). Note that these are modified from Leray *et al.*, (2013) and they include the standard Illumina molecular identification tags, bridge sequences and heterogeneity spacers (see Kitson *et al.,* (2019) for details). For every PCR, a PCR positive and PCR negative was included, the PCR positive was DNA from a fish species (*Astatotilapia calliptera*) for the MiSeq dataset, but for the NovaSeq a crayfish species was selected (*Hommarus gammarus*) while the negatives included all PCR reagents without any template DNA. Four PCR for four samples were repeated and were included for both datasets. PCRs were carried out over 40 cycles (95^{o}C for 45 seconds (s), 51 ^{o}C for 15 s and 72 ^{o}C for 45 s in 20µL reactions using a high fidelity Taq mastermix (MyFi Mix Bioline), 2 µL of template DNA and each primer (final concentration at 0.5 µM). To further prevent cross contamination the wells were sealed using mineral oil before all the other reagents and template DNA were added. PCR success was checked by using 5 µL of PCR products on 1.5% agarose gels. PCR negatives and DNA extraction negatives did not show any bands. I then conducted a bead-based normalization by using 0.6:1 ratio of Solid Phase Reversible Immobilizations beads (SPRI) (9 µL) and 15 µL of PCR template for each sample.

4.3.4 Library preparation and sequencing

After clean-ups and prior to library preparation different pooling strategies were used for the two different sequencing technologies (MiSeq and NovaSeq). For the MiSeq I pooled the samples in groups of 16 (plus 4 controls), 4 µL from each sample to create each pre-library. Totalling 12 libraries, 6 for every plate. For the NovaSeq I pooled the samples in groups of 8 (plus 4 controls), 4 µL from each sample to create each pre-library which totalled 24 libraries, 12 for every plate. This was done to increase sequencing diversity through the early sequencing cycles because the different sequencers required different sequencing diversity inputs, as suggested by the sequencing centre. All of the following procedures to create each of the library used the same amounts as followed: To create each of these libraries I used a second PCR (PCR2) with 12 cycles (95 ^{o}C for 45 s, 51 ^{o}C for 15 s and 72 ^{o}C) and a final extension step of 5 min at 72 ^{o}C in 20 µL reactions using 5 µL of each pooled library, the same Taq (MyFi

Mix Bioline) and each of the respective Illumina N5 and N7 adapters (at a concentration of 1 µM) 12 adapters were used for the MiSeq run and 24 for the NovaSeq. For each library a PCR2 negative was also included. All libraries + PCR2 negatives were checked on gel. No bands were visible for any of the negatives. I then did a PCR2 clean-up to remove fragments smaller than the target by using 0.6:1 ratio of SPRI to template (9µL and 15µL respectively). After cleaning the libraries, they were checked on an Agilent TapeStation and were pooled equimolarly at approximately 15 ng/µL. The pooled final libraries were then sequenced both on an Illumina MiSeq using the 2x250 kit and a NovaSeq using a SP+XP (2x250) kit at the Genomics Core Facility at Newcastle University.

4.3.5 Bioinformatic analysis

Sample demultiplexing within individual libraries was conducted using the software *cutadapt* v1.18 software (Martin, 2011). All other analyses were conducted in *R* (v. 4.0.1, R Core Team, 2021) except if stated otherwise. The demultiplexed data were processed using package *DADA2* (Callahan *et al.*, 2016), removing primers using *cutadapt* v1.18. Due to the large amount of data from the NovaSeq dataset (120GB), the dataset was processed on a High-Performance Computing cluster at Newcastle University (Rocket). *DADA2* filtered and trimmed sequences based upon read quality removing any reads with ambiguous "N" bases with the "filterAndTrim" function. I then used the learn "learnErrors" function to estimate error rates for the datasets and then inferred amplicon sequence variants (ASV's) with the dada function. Finally, I then merged pair-end reads with the "mergePairs" and removed chimeras with the "removeBimeraDenovo" functions. Default function parameters were used for most functions, except if stated otherwise. Taxonomy was assigned with Blastn (Camacho *et al.*, 2009) for both datasets with a curated database downloaded from MIDORI2 reference (Leray, Knowlton and Machida, 2022). Our main target were insects and other arthropods, we thus filtered the datasets to include only classes of Arthropoda (see section 4.6 for further details on the other taxa). As we were uncertain if samples were contaminated throughout the handling at Fera, additional filtering criteria were applied at the sample level. Here, I applied a one percent sample threshold for both datasets to minimize false negatives. More specifically, if a taxon within a sample had less reads than 1% of the total amount of reads for that sample, then it was removed. The one percent was chosen based on previous work (see

Chapter 2 section 2.3 for further details and it is considered stringent as typical thresholds are between 0.2 to 0.6 (Drake *et al.*, 2022). Finally, to see whether sequencing depth was correlated between the two datasets I performed a linear model between the two.

4.3.6 Statistical analyses

To compare the datasets from MiSeq and NovaSeq for both pre-filtered and filtered taxonomy I first compared total taxonomic composition between the datasets with the "settdiff" and "intersect" functions in *R* (v. 4.0.1, R Core Team 2021). For sample-to-sample differences I computed the Jaccard similarity index in R implemented with a custom function. Jaccard's similarity index can be defined as:

$$Jaccard's\ index = \frac{a}{b + c - a}$$

Where a= The number of shared species between each sample, b= the number of species in the sample from dataset b (here MiSeq), c= the number of species in the sample from dataset c (here NovaSeq). Finally, we compared whether the number of ASVs differed between datasets with the "aov" function in R (v. 4.0.1, R Core Team 2021). To visualize the taxa found across the samples we produced heatmaps using the *Metacoder* package in R (Foster, Sharpton and Grünwald, 2017).To identify potential pests and beneficials within bycatch I used the same four databases as in chapter 3 (section 3.3). First to find additional pests relevant to potato fields I used the database of plant to insect interactions (Padovani *et al.*, 2020) and matched my taxa with taxa that are interacting with potato plants (*Solanum tuberosum*). Second, I use the pollinator and beneficial database from chapter 3 to find additional beneficial insects that were either pollinators, predators or parasitoids. Species richness was calculated by using the diversity function in *Vegan* (Oksanen *et al.,* 2022) . Richness for Diptera was very high for this reason Diptera were plotted separately (see Figure 4. 3). To account for differences in sequencing depth I subsampled using the "rrarefy" function in *Vegan* (Oksanen *et al.,* 2022) at 1500 reads for MiSeq and 10000 reads in the NovaSeq dataset for all diversity comparison analyses. These values where chosen to minimize sample drop and were based on the minimum read depths (for samples were an amplification band was visible) for each dataset. To analyse patterns of species richness between the fields and months I first performed a multivariate GLM with mvabund (Wang *et al.*, 2012) assuming a negative binomial distribution, to see whether there where significantly different

communities between the fields and months sampled. Assumptions were checked by visually plotting the residuals of the model (see Figure C. 1). To see whether richness differences could be explained by environmental variables we fitted a Generalized Linear Model (GLM) with the "glm" function in R with a Poisson family distribution, with richness as the response variable and latitude and landscape diversity as the predictor variables. To calculate landscape diversity at the field level, I downloaded the UKCEH land cover map for 2019 (Morton *et al.*, 2020). Landscape diversity for each field was calculated within a 3km circular buffer zone based on the categories that were either natural, grasslands or farmland (excluding categories such as: urban or saltwater) with the package *landscapemetrics* (Hesselbarth *et al.*, 2019). Beta diversity was calculated with "avgdist" function in Vegan (Oksanen *et al.,* 2022), with subsampling based on the minimum read depth of samples which was 1500 for the MiSeq and 10000 for the NovaSeq. To assess whether compositional dissimilarity (here as beta diversity) was correlated with physical distance of the samples themselves we checked the correlation between them using a mantel test. Beta diversity was calculated with "avgdist" function with presence-absence data. Physical distance between sampling points (i.e different fields) was computed with the "geo_dist" function in the *geodist* package (Padgham, 2021). Finally, to examine temporal changes between samples I partitioned the diversity into loss and gain components (Tatsumi, Iritani and Cadotte, 2022) to assess total changes in beta diversity and how beta diversity was influenced by specific taxa. Gain components referred to species added between the months and locations (referred as Colonization) and loss components where species that disappeared for the months and samples (referred as Extinction). The species comparison was made at the month level between: June-July and July-August. This analysis was done only on a subsample of locations, where samples for all the months were present. The subsample compromised of 28 fields. All the analyses were performed for both the MiSeq and NovaSeq datasets. However, within the main text figures and results for the models for diversity patterns, these include only the NovaSeq dataset -due to higher sequencing depth for each sample overall which led to less samples being dropped when subsampling. If an analysis or figure includes the MiSeq dataset this is made explicit in the legends. Furthermore, the models were additionally run only on the subset of locations (27) for which all months were present. This was done to see whether differences in sampling completeness affected the overall results of the models (see Appendix C; Table C. 4; C. 5)

4.4 Results

4.4.1 Sequencing results

PCR success was high with more than 184 out of 188 samples producing a visible band on a gel for the MiSeq and 185 out of 188 samples for the NovaSeq. The MiSeq run produced 11,962,584 reads (excluding PhiX), after demultiplexing, filtering, denoising, merging and chimera removal a total of 8,824,533 reads were retained (without accounting 399,993 reads dropped as they were assigned to non-arthropod taxa). The NovaSeq run produced a total of 300,000,000 reads of which after following the same steps as the MiSeq dataset a total of 91,083,100 reads were retained (without accounting reads that were dropped as non-arthropod taxa). Reads per sample excluding controls ranged from 127 to 122,777 for the MiSeq with a median 35,972(1^{st} quartile: 16,545, 3^{rd} quartile: 48645). For the NovaSeq dataset reads ranged from 953 to 1,416,083 with a median 515,423 (1^{st} quartile: 292,249, 3^{rd} quartile:647,891) whilst the mean was 487,150 (± 283,980). No reads from the positive controls were found in any of the samples for both datasets. In the MiSeq dataset 2 extraction negatives had 30 and 25 reads respectively. However, for the NovaSeq dataset 3 negatives (2 extraction and 1 PCR negative) had between 7-10 reads. Similar to a previous chapter (see Chapter 3.3), certain positive controls did not have any passing reads, suggesting a probable issue with the tag system used or very low concentrations of the samples, as the positive controls chosen were single individuals whereas the bulk samples comprised of hundreds of individuals. Samples with low read output n= 4 (<1,500) or no reads for MiSeq corresponded to samples where there was no visible band on a gel. However, for the NovaSeq 10 samples with low read output (<10,000 reads) did not correspond to failed PCR's. From those only 3 samples did. Despite this, sequencing depth between the samples was significantly correlated (R^2= 0.40, p= 2e-16).

4.4.2 Assigned taxonomy: MiSeq and NovaSeq similarities

For both datasets around 71% (6 million for MiSeq) and 75% (91 million for NovaSeq) reads could be assigned at the genus and species level in the Arthropoda phylum. Total ASVs found were 4,379 and 6,060 respectively corresponding to 553 and 546 unique taxa at the genus level. In both datasets the class Insecta dominated the reads with 98% and 99% respectively, around 0.3% belonged to Arachnida and few reads (<0.01%) to Collembola. There were a total of 12 insect orders found with the most abundant being Diptera with over 60% reads and 104 taxa, followed by Hemiptera, Coleoptera and Hymenoptera for both the MiSeq dataset and the NovaSeq. However, the remaining orders did vary in total reads for each of the datasets (see Figure 4. 2A). Within the three months sampled, Diptera were always the most abundant taxa for both datasets and the following major orders reflected the total read distribution of species among the orders (see Figure 4. 2B). The most abundant genera found for both datasets were *Delia*, *Scaptomyza*, *Brassicognethes*, *Bombus* and *Eupeodes* (see Figure 4. 3). There were significant differences when the number of ASV's were compared between the two sequencing methods (F_{value}=42.41, p=2.3e-10). However, compositional similarity between the datasets was high with more than 490 (81%) taxa shared at the genus level, for species level the similarity was 69% (764 species shared). Within samples the similarity was high with a mean Jaccard index of 0.80 (s.d ±0.10) (see Figure C. 3). Samples with low similarity typically had lower number of total reads when compared with the mean for each dataset. Surprisingly, after applying filtering criteria the MiSeq dataset had more taxa remaining. After using a one percent threshold both datasets lost taxa, Miseq totalled 244 taxa whilst NovaSeq retained 218, despite the 100 more taxa discovered. The similarity between the taxonomies remained high after applying the filtering (see Table 4. 1). Agricultural taxa of interest included 7 parasitoids, 15 pests, 136 pollinators and 36 predators. Most pollinators belonged to the order Diptera (89) and Hymenoptera (26). Predators were mostly found within the orders of Coleoptera (20), and Araneae (10). Finally, most pests belonged to Hemiptera (13) and parasitoids to Hymenoptera (7) (see Figure C. 3 for all the taxa that a category could be found see Table C. 4).

	Total number of taxa at species level (before filtering)	Total number of taxa at genus level (before filtering)	Shared taxa between MiSeq and NovaSeq (before filtering)	Total number of taxa at genus level (after filtering)	Shared taxa between MiSeq and NovaSeq (after filtering)	Median: Jaccard's similarity index
MiSeq	874	553	764/499 (species/genera)	244	209	0.80
NovaSeq	988	546		218		

Table 4. 1 Number of taxa found between the two platforms at the species and genus level for both pre-filtered and filtered datasets along with summary statistics for Jaccard's similarity index

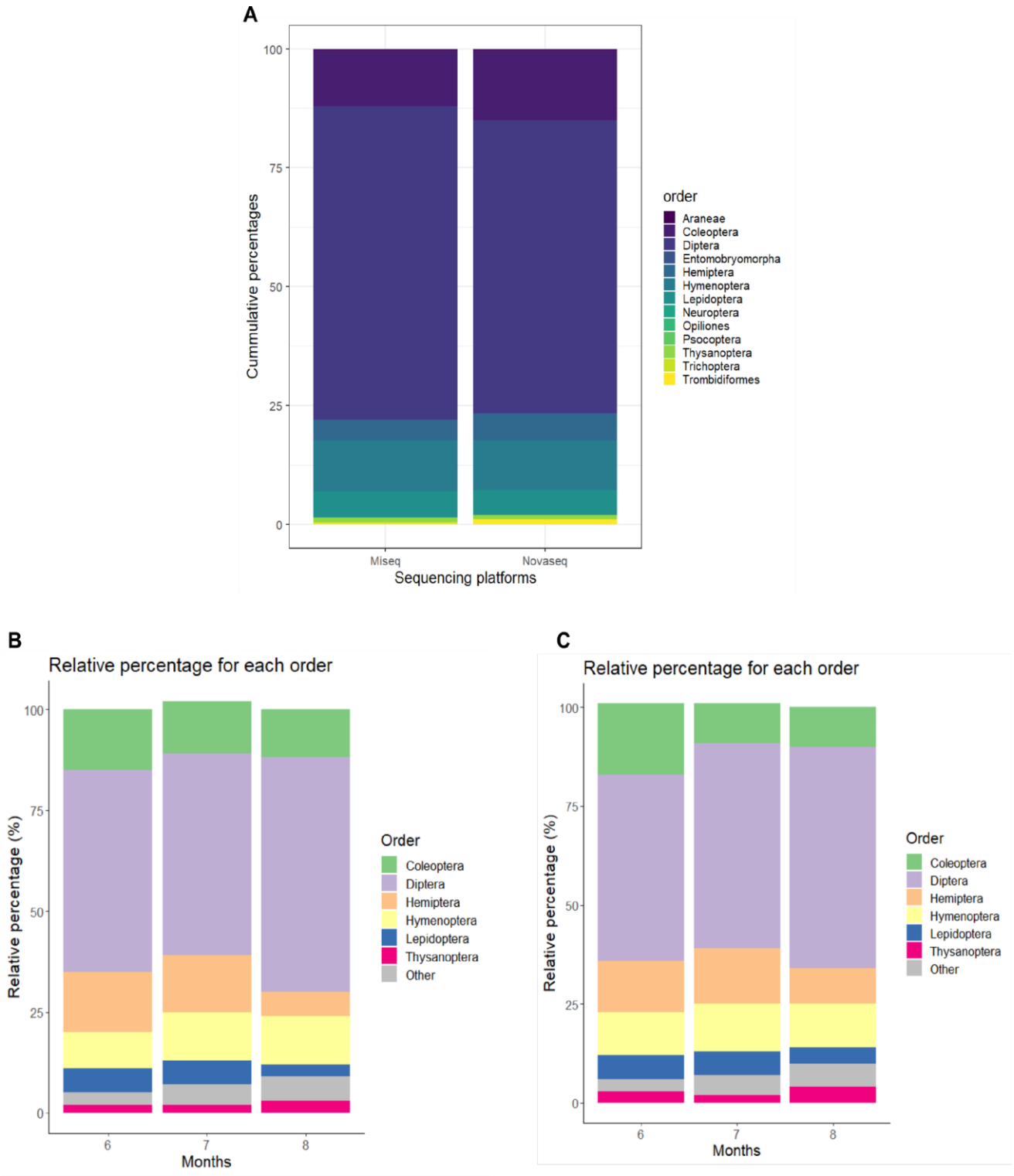


Figure 4. 2 A) Taxonomic composition of the two sequencing platforms according to the relative percentage of each species in the complete dataset. B) Relative percentage of taxa within the major orders for the three months sampled in the MiSeq dataset. C) Relative percentage of taxa within the major orders for three months sampled in the NovaSeq dataset (percentages have been rounded, therefore in B and C some of the bars look slightly above or below 100%)

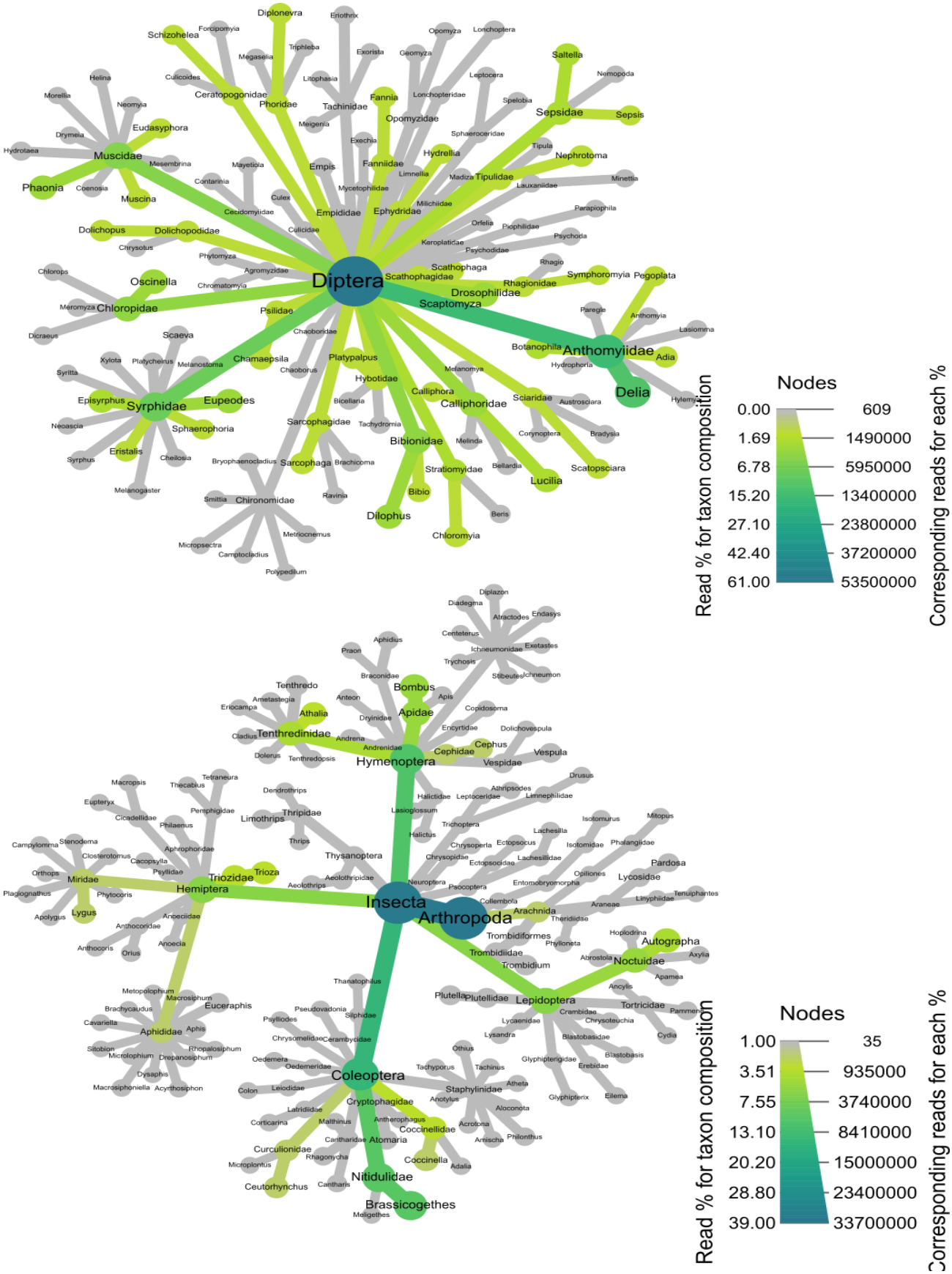


Figure 4. 3 Heat trees showing the total diversity captured within the filtered dataset. Each node represents a taxonomic level, lowest node presented here corresponds to the genus

level. Coloured edges and nodes correspond to the relative percentage of reads for each taxon and the number of reads. (Top) Only Diptera are plotted with the highest diversity, (bottom) all the rest major Arthropod orders are plotted together

4.4.3 Diversity patterns

Species richness within samples ranged between 1 and 21 for the MiSeq dataset and 1 to 19 for the NovaSeq dataset (filtered datasets). Beta diversity was found to be high between most of the samples with a median of 0.90. The GLM fitted with *mvabund* revealed significant differences between the communities for fields sampled but not for months (see Table 4. 2).

Coefficient	Res.df	Df.Diff	Dev	Pr(>dev)
Intercept	1329			
Field	1328	1	205.3	0.001*
Month	1327	1	264.4	0.199

Table 4. 2 Coefficients and results for the multivariate GLM, only Field was found significant with (Likelihood ratio test: 205)

Richness was found to be negatively affected by increasing latitude, but the relationship was weak (see Table 4. 3; Figure C. 2; Appendix C). Compositional dissimilarity was not correlated with physical distance (mantel R statistic= 0.046, significance= 0,031 at the 90% upper quantile; see Figure 4. 4;AB) for either of the datasets. Partitioning the diversity into gain and loss components (colonization and extinction) between the three months sampled showed low total changes in the components between June/July but high between July/August (see Table 4. 4). Colonization and extinction from many taxa had a significant effect on beta diversity (see Figures 4. 5;4. 6), highlighting that species turnover is the major driver of beta diversity. The taxa-level comparison showed that the genera that had the highest influence in beta diversity were: *Brassicogethes*, *Scaptomyza, Oscinella* and *Delia* for both timepoint comparisons with *Episyrphus* having a significant influence for the July-August comparison but not June-July (see Figure 4. 6). These taxa were also the most abundant ones in the dataset.

Coefficients	Estimate	Standard error	Z value	Pr(>z)
Intercept	4.50	0.67	6.71	<1.8e-11
Latitude	-0.03	0.01247	-2.407	0.0151
Landscape diversity	-0.054	0.08	-0.647	0.5174

Table 4. 3 Coefficients and estimates for the GLM for the different effect of factors for species richness. AIC: 676.46, residual deviance 291 on 81 degrees of freedom

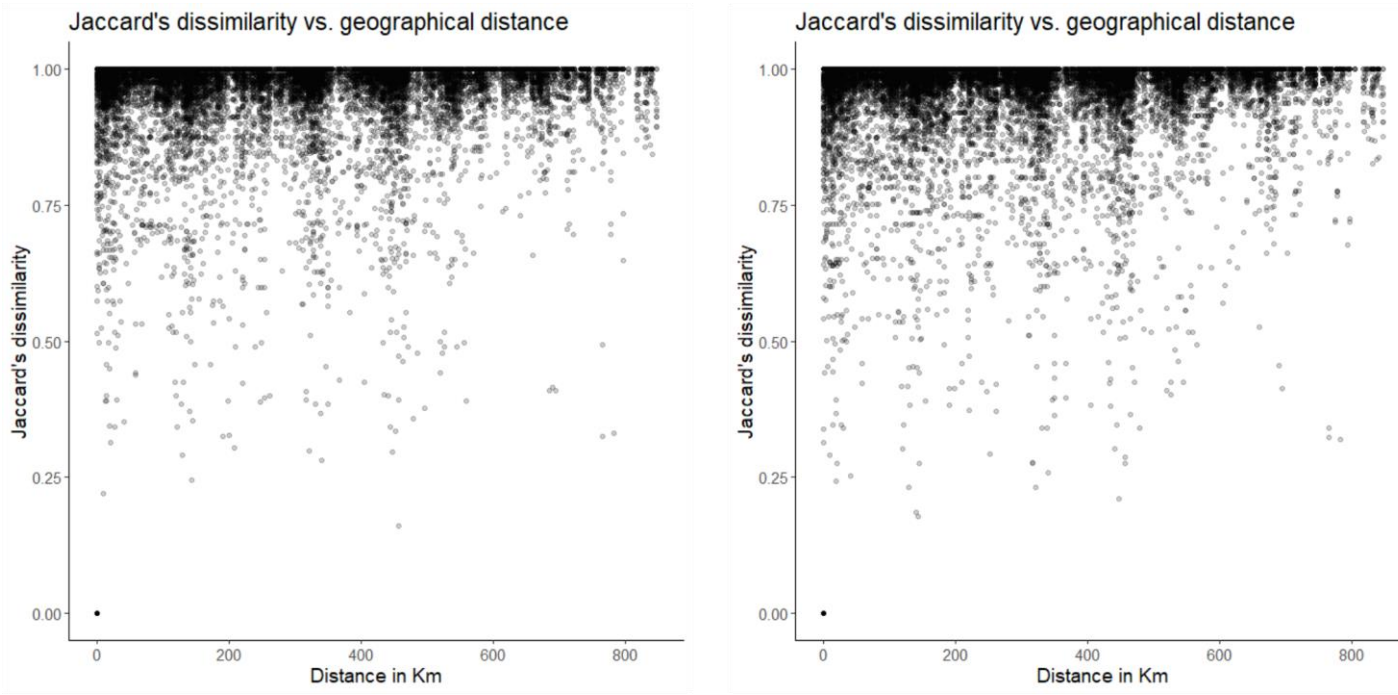


Figure 4. 4 Jaccard's dissimilarity between all sample comparisons against physical distance of sampling points for the MiSeq (Left) and NovaSeq (right) datasets.

Extinction and Gain components	Comparison for June-July (Temporal change in beta diversity)	Comparison for July-August (Temporal change in beta diversity)
Extinction component effect (Loss of species)	0.084	0.29
Colonization component effect (Gain of species)	0.06	0.50

Table 4. 4 Temporal component (gain and extinction) effects on beta diversity based on extinction (loss) and colonisation (gain) of species.

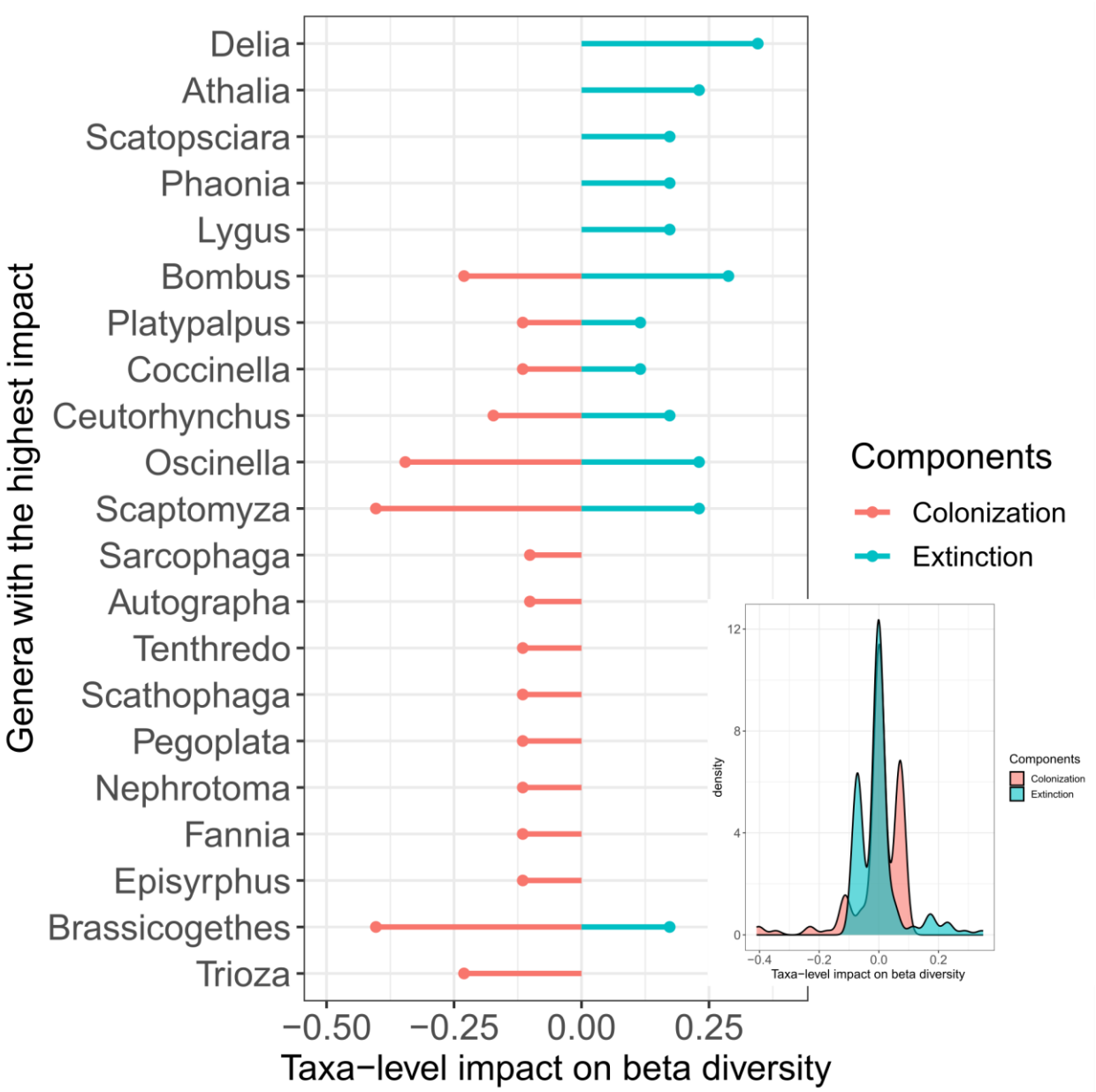


Figure 4. 5 Taxa level impact on beta diversity for genera that increased or decreased by a-+ 0.10 change the overall beta diversity, comparison between June-July. The density plot shows distribution of values for all taxa for the main components (colonaziation and extinction).

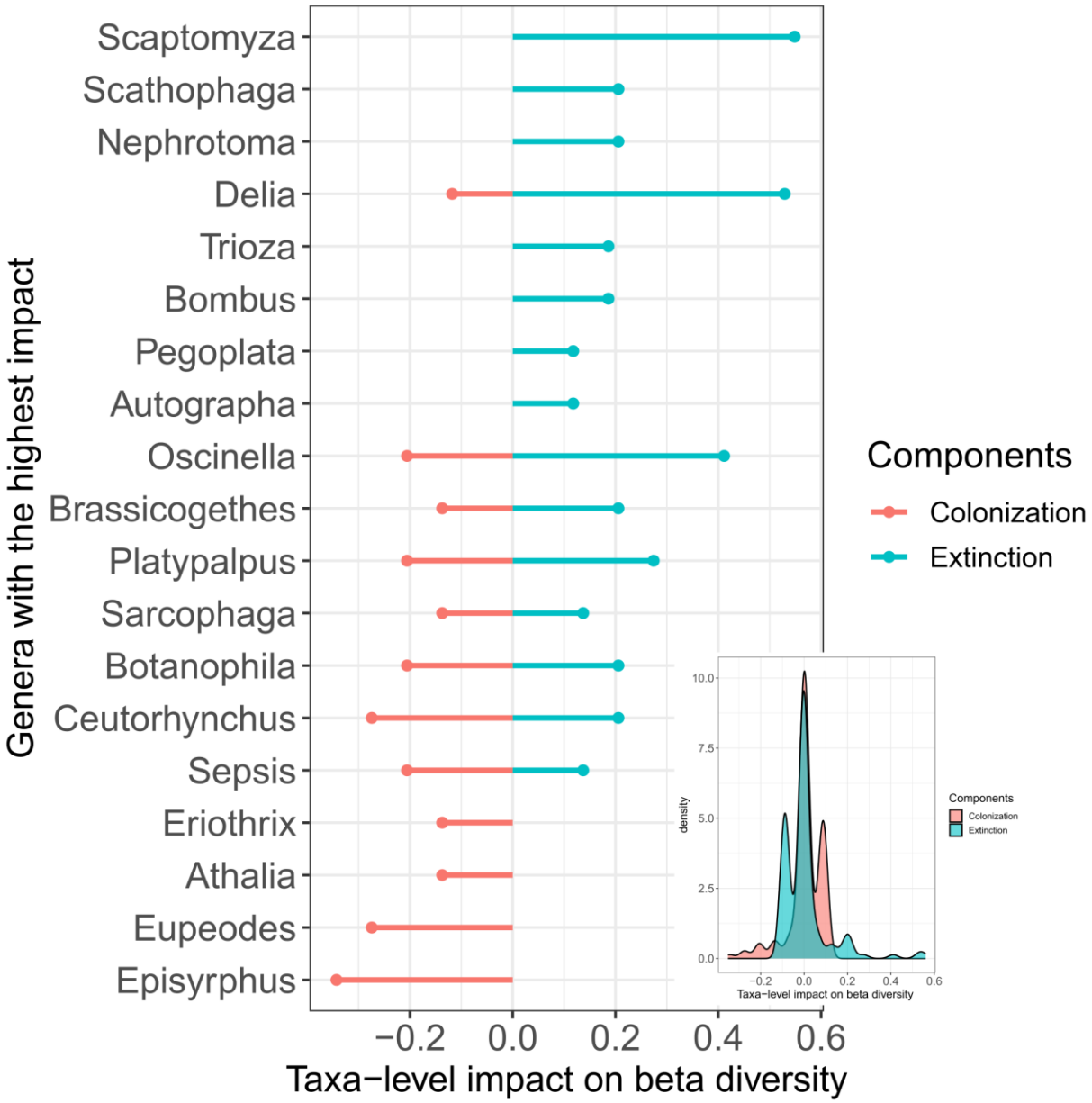


Figure 4. 6 Taxa level impact on beta diversity for genera that increased or decreased by a -+ 0.10 change the overall beta diversity, comparison between July August. The density plot shows distribution of values for all taxa for the main components (colonaziation and extinction).

4.5 Discussion

In this study, I have demonstrated that DNA-metabarcoding can successfully identify bycatch diversity in active nationwide insect monitoring schemes, with more than 800 species identified, as well as describing community-level spatio-temporal differences, highlighting the value of bycatch diversity as a source for insect data. Furthermore, by comparing two different sequencing platforms (MiSeq and NovaSeq), I show overall very high similarity between the two with more than 90% of taxa shared but also within-sample similarity despite the 20-fold difference in sequencing output. Finally, I show that bycatch diversity patterns are complex with a mean beta-diversity of 0.90 despite study sites being agricultural monocultures (potato fields) which was found to be mainly driven by taxa turnover. This study further validates the added value of DNA-metabarcoding for insect monitoring schemes by making bycatch accessible at a nationwide scale.

4.5.1 Bycatch diversity within potato fields

Overall, bycatch diversity within the YWP network was high with more than 800 taxa for the pre-filtered datasets and around 500 for the filtered at the species level (see Table 4. 1). Bycatch is typically not analysed within monitoring schemes as it can be regarded as prohibitively expensive, requires more effort and taxonomic expertise. But Next-Generation approaches can offer a tool for biodiversity monitoring with no taxonomic expertise (Yu *et al.*, 2012). It can be seen as valuable taxonomic information as it (a) can inform early pest detection, (b) be used for broadening knowledge regarding distributions of taxa and (c) even identify new species (Hribar, 2020). I showed that DNA-metabarcoding can identify such samples with more than 12 insect orders within the fields sampled. Among the taxa identified, many could be characterized as beneficial insects such as pollinators: *Bombus, Apis* and *Syrphus* but also aphid parasitoids and predators which included: *Aphidius*, *Praon* and *Coccinella*. Additionally, 6 additional taxa were also found to be potato plant pests, such as: *Delia platura*, *Dilophus febrilis*, *Closterotomus norwegicus*, *Eupteryx atropunctata*, *Campylomma verbasci* and *Noctua pronuba*. YWP focuses on informing farmers about aphid populations and risk of viral transmission, therefore this could be additional information

provided to farmers that could aid management choices. DNA-metabarcoding does not provide quantitative information, and this is a limiting factor for its use in monitoring, however there are ways for extracting such information (Bista *et al.*, 2018; Deagle *et al.*, 2019). Diptera was the most abundant taxa both in relative reads and total number of species which was expected given that Diptera can be some of the most abundant flying insects (de Souza Amorim *et al.*, 2022). Within Diptera there can be many species that provide ecosystem services for farms like predators, or pollinators. In my data, at least 4 very common taxa can be highlighted that can provide such services, the genera: *Episyrphus*, *Eupeodes*, *Syrphus* and *Sphaerophoria*. Yet overall, we identified more than 170 species that can provide such services. The ratio of target species to bycatch species was found on average 1:10, which means that bycatch species dominate the catches in the YWP network. Here, I showed that despite the sampling that took place within an agricultural monoculture (potato fields) diversity was high (556 genera) and, overall, very different in its taxonomic composition from field to field as shown by the high beta-diversity. As bycatch is typically thrown away within this monitoring scheme and perhaps many more, the loss of insect data needs addressing and DNA-metabarcoding can be seen as a tool that can address this problem, even if the lack of space within a monitoring scheme is an issue. Then DNA-metabarcoding can provide an alternative digital archive of the communities captured (Petsopoulos *et al.,* 2021).

4.5.2 Sequencing technologies differences

Next-generation sequencing approaches can vary greatly in their output number of reads and sequencing depth can be a limiting factor when trying to identify complex multi-taxa samples (Alberdi *et al.*, 2018). Here we compared two Illumina sequencing platforms the MiSeq and NovaSeq with a 20-fold difference in total output and 16-fold difference within the sample read averages between the two datasets. Despite the differences in output, I found high compositional similarity (see Table 4. 1) overall, at both species and genus level. However, the NovaSeq dataset did include more than 100 additional species than the MiSeq which was reflected to the total difference in numbers of ASVs between the datasets (4379 for MiSeq and >6000 for NovaSeq). At the genus level the difference was minimal with NovaSeq capturing only 7 more additional genera. NovaSeq datasets have been previously shown to

capture more ASV's even when compared at similar sequencing outputs as the MiSeq (Singer *et al.*, 2019). In this study, we observed a similar pattern however, when ASVs were collapsed into genus level there were no differences in the taxonomic composition. Similarity remained high (90%) after filtering thresholds were applied but the overall taxa within the datasets fall to 227 and 245 for NovaSeq and MiSeq respectively. Surprisingly, NovaSeq lost more taxa than the MiSeq which might reflect a limitation on the filtering criteria applied here. As NovaSeq had many more reads for each sample rare taxa which had less reads got filtered out due to the high threshold in each sample. Each sample within the NovaSeq had more rare taxa (114 more species) than the MiSeq the high sequence depth which would equal higher read thresholds resulted in many more taxa being dropped in the NovaSeq dataset because of the higher number of rare taxa per sample. Whilst such taxa remained in the MiSeq dataset as the limit for dropping out taxa was lower overall. Our stringent criteria most likely removed false positives, but as previously shown (see Chapter 2) it is uncertain as to which extent this was done but also to which extent false positives were also dropped as is the case when applying such filtering methods (Drake *et al.*, 2022). As morphological information is not available for these samples it's uncertain to which degree false positives and negatives are removed. Our approach for extracting the DNA of these samples was not destructive, therefore this could be checked by revisiting the samples themselves. Finally, the high similarity between the two datasets showcases that despite the high number of samples (188) the MiSeq dataset managed to capture the bycatch diversity with low read outputs which means that costs can be kept low which is a major factor for deciding sampling efforts in monitoring schemes.

4.5.3 Bycatch diversity patterns within YWP nationwide monitoring scheme

Insect biodiversity patterns can be very complex, particularly over large spatial scales (Grabener *et al.*, 2020; Outhwaite, McCann and Newbold, 2022). Yet, diversity is not a homogenous and static measure, changes over space and time lead to turnover of that diversity and alterations to structural properties of the diversity like nestedness, ultimately driving broader ecological differences. Despite sampling within an agricultural monoculture, we find overall high diversity and very different communities across the fields sampled and

the months (see Figures 4. 4;AB). Although, we found species richness decreasing with latitude in our dataset, which is observable in other groups of taxa as well (Hillebrand, 2004). The relationship was weak and there was no apparent spatial grouping between the communities found, despite certain fields being in very close proximity with each other (see Figure 4. 4). This is most likely due to factors influencing communities at a local level, even different management practices within the fields, as arthropods can be very sensitive to environmental variations (Nunes *et al.*, 2020). Unfortunately, management information was not available for the fields sampled but further studies could incorporate other environmental factors such as temperature, rain or wind to further understand if these could explain differences in species richness across the dataset. Throughout the dataset, a similar temporal pattern with Diptera dominating can be seen for most of the months sampled with July having the overall highest bycatch diversity with more than 180 taxa throughout the network. I found that beta diversity changes are driven mainly by species turnover. Species turnover can be a major component of beta diversity in arthropods (Nunes *et al.*, 2020) influenced by a number of factors such as environmental dissimilarity and geographic distance (Buckley *et al.*, 2008). Turnover was found to be mainly driven by the disappearance and appearance of certain taxa (Figures 4. 5;4. 6). These taxa were also the most abundant taxa in the dataset, this could imply that when these taxa are present and dominate the reads, rare taxa are not picked up leading to higher beta diversity measures. There are several reasons why this might be, including uneven sample abundances or PCR biases (Martoni *et al.*, 2022). However, as I did not identify such samples morphologically or did not estimate any abundance proxies, such as biomass, my study was not able to attribute such differences to such biases. For example, in Chapter 2, when the dominant taxa were present in a sample congruence was found lower. Sampling completeness could be one of the reasons for such high beta diversity estimates, as many of the fields sampled did not have samples throughout the time period. However, an analysis with all the samples that had replicated samples throughout the months still showed high beta diversity (see Table C. 3) and similar diversity patterns (i.e. decreasing richness with increasing latitude). Further explorations of beta diversity which would include many more environmental factors such as: temperature and rainfall could help better illuminate what is driving the high species turnover in this study. As geographical distance was not correlated with beta diversity and even very closely located sites showed high beta diversity, we expect local factors to be the main driver of these differences. NMDS plots of the community across

environmental gradients (temperature or rainfall) could be used to highlight any groupings associated with those gradients. Insects represent one of the most biodiverse groups of taxa on Earth, yet there is still a lot of uncertainty regarding insect population particularly in the context of insect declines. Bycatch that results from unwanted catches in monitoring schemes typically doesn't get analysed as it is very expensive and the labour costs in both time and money further complicate its use. I have shown here that with DNA-metabarcoding we can analyse patterns of bycatch biodiversity within a nationwide scale even at very low sequence outputs.. While relying on relatively affordable sequencing technology of the MiSeq we found minimal differences between the sequencing platforms in the resultant communities,. As sequencing technologies are getting cheaper and cheaper (Preston, VanZeeland and Peiffer, 2021) this shows great promise for DNA-based monitoring of insect communities.

4.6 Limitations of this study and conclusions

In this study I have showed that DNA-metabarcoding can be a valuable tool for characterising the diversity of bycatch taxa. Focusing on a nationwide monitoring scheme at Fera (YWP), we additionally showed how DNA-metabarcoding can help us explain global patterns across the scheme whilst using two different sequencing platforms. The biggest impediment in this study is the lack of morphological data to assess the accuracy of DNA-metabarcoding against morphological identifications. My DNA-extraction approach was not destructive therefore these samples could be revisited for this purpose. However, at the scale of this study this includes a serious workload as thousands of insect's specimens were trapped with overall high species complexity (12 orders, 545 genera). Additionally, insect monitoring needs fast and scalable methods. Morphological identification can act as an impediment to the efficiency of overall monitoring efforts as it is generally more time-consuming. A better approach would be to address the issue at the source, by controlling cross-sample contamination that might be introduced during sample handling. We used sequencing threshold filtering approaches to account for this which led to 230 taxa being dropped from the dataset. Yet, it would be beneficial for such schemes to establish multiple layers of controls. For example, field negative controls are needed to address contamination from nearby sources. As these traps are left for weeks in agricultural fields many organisms can contaminate the water in the pan traps (i.e. farm animals drinking water). Establishing field controls would also help minimize

contamination from flying insect DNA giving a more accurate representation of the community being sampled. In this study, we focused on arthropods as the main aim was insect bycatch diversity. However, non-arthropod taxa were recovered, some included common contaminants (like human DNA, or cow DNA which was not unexpected as this study takes place within agricultural fields). Others included birds, or gastropods some of which I've found tissue within the samples. This means that contaminants were present at the stage of sample collection some of which can be biologically explained (non-arthropod taxa in fields) but others point to contamination at the sample handling level (human DNA).

4.7 References

Bell, J. R., Blumgart, D. and Shortall, C. R. (2020) 'Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain', *Insect Conservation and Diversity*, 13(2), pp. 115–126. doi: 10.1111/icad.12412.

Bista, I. *et al.* (2018) 'Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples', *Molecular Ecology Resources*, pp. 0–2. doi: 10.1111/1755-0998.12888.

Buckley, L.B., and Jetz, W. (2008) 'Linking global turnover of species and environments', *Proceedings of the National Academy of Science, 105(46),* pp. 17836-17841. doi: 10.1073/pnas.0803524105

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/1471-2105-10-421.

Conrad, K. F., Fox, R. and Woiwod, I. P. (2007) 'Monitoring biodiversity: Measuring long-term changes in insect abundance', *Insect Conservation Biology: Proceedings of the Royal Entomological Society's 23rd Symposium*. CABI Publishing, pp. 203–225. doi: 10.1079/9781845932541.0203.

Deagle, B. E. *et al.* (2019) 'Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?', *Molecular Ecology*, 28(2), pp. 391–406. doi:

10.1111/mec.14734.

Derocles, S. A. P. *et al.* (2018) 'Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into Ecological Network Analysis', *Advances in Ecological Research*, 58, pp. 1–62. doi: 10.1016/bs.aecr.2017.12.001.

Drake, L. E. *et al.* (2022) 'An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 13(3), pp. 694–710. doi: 10.1111/2041-210X.13780.

Foster, Z. S. L., Sharpton, T. J. and Grünwald, N. J. (2017) 'Metacoder: An R package for visualization and manipulation of community taxonomic diversity data', *PLOS Computational Biology*. Public Library of Science, 13(2), p. e1005404. doi: 10.1371/JOURNAL.PCBI.1005404.

Grabener, S. *et al.* (2020) 'Changes in phenology and abundance of suction-trapped Diptera from a farmland site in the UK over four decades', *Ecological Entomology*. Blackwell Publishing Ltd, 45(5), pp. 1215–1219. doi: 10.1111/EEN.12873.

Harrington, R. (2013) 'The Rothamsted Insect Survey strikes gold', *Antenna*, 38(3), pp. 158–166. Available at: www.rothamsted.ac.uk/examine.

Hebert, P. D. N. *et al.* (2003) 'Biological identifications through DNA barcodes', *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), pp. 313–321. doi: 10.1098/rspb.2002.2218.

Hesselbarth, M. H. K. *et al.* (2019) 'landscapemetrics: an open-source R tool to calculate landscape metrics', *Ecography*. Blackwell Publishing Ltd, 42(10), pp. 1648–1657. doi: 10.1111/ECOG.04617.

Hillebrand, H. (2004) 'On the generality of the latitudinal diversity gradient', *American Naturalist*, 163(2), pp. 192–211. doi: 10.1086/381004.

Hribar, L. J. (2020) 'Expanding Basic Entomological Knowledge by Using Mosquito Surveillance Bycatch', *Annals of the Entomological Society of America*, 113(6), pp. 439–446. doi: 10.1093/aesa/saaa014.

Ji, Y. *et al.* (2013) 'Reliable, verifiable and efficient monitoring of biodiversity via

metabarcoding', *Ecology Letters*, 16(10), pp. 1245–1257. doi: 10.1111/ele.12162.

Ji, Y. *et al.* (2020) 'SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes', *Molecular Ecology Resources*, 20(1), pp. 256–267. doi: 10.1111/1755-0998.13057.

Kitson, J. J. N. *et al.* (2019) 'Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding', *Molecular Ecology*, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

van Klink, R. *et al.* (2020) 'Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances', *Science*, 368(6489), pp. 417–420. doi: 10.1126/science.aax9931.

Krehenwinkel, H. *et al.* (2017) 'Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding', *Scientific Reports*, 7(1), pp. 1–12. doi: 10.1038/s41598-017-17333-x.

Leray, M. *et al.* (2013) 'A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents', *Frontiers in Zoology*, 10(1), pp. 1–14. doi: 10.1186/1742-9994-10-34.

Leray, M., Knowlton, N. and Machida, R. J. (2022) 'MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences', *Environmental DNA*, 4(4), pp. 894–907. doi: 10.1002/edn3.303.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*. EMBnet Stichting, 17(1), pp. 10–12. doi: 10.14806/EJ.17.1.200.

Martoni, F. *et al.* (2022) 'Disentangling bias for non-destructive insect metabarcoding', *PeerJ*, 10, p. e12981. doi: 10.7717/peerj.12981.

Morton, R. D. *et al.* (2020) 'Land Cover Map 2019 (25m rasterised land parcels, GB)'. NERC Environmental Information Data Centre. Available at: https://doi.org/10.5285/f15289da-

6424-4a5e-bd92-48c4d9c830cc.

Nunes, C. A. *et al.* (2020) 'High Temporal Beta Diversity in an Ant Metacommunity, With Increasing Temporal Functional Replacement Along the Elevational Gradient', *Frontiers in Ecology and Evolution*, 8(November), pp. 1–12. doi: 10.3389/fevo.2020.571439.

Oberacker, P. *et al.* (2019) 'Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation', *PLoS Biology*, 17(1), pp. 1–16. doi: 10.1371/journal.pbio.3000107.

Oksanen, J.A.R.I., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P. and Stevens, M.H.H. (2020) 'vegan: Community Ecology Package. R package version 2.5–7. 2020.'

Outhwaite, C. L., McCann, P. and Newbold, T. (2022) 'Agriculture and climate change are reshaping insect biodiversity worldwide', *Nature 2022 605:7908*. Nature Publishing Group, 605(7908), pp. 97–102. doi: 10.1038/s41586-022-04644-x.

Padgham, M. (2021) 'geodist: Fast, Dependency-Free Geodesic Distance Calculations', R package. Available at: https://github.com/hypertidy/geodist.

Padovani, R. *et al.* (2020) 'Insect species richness for each plant species and insect-plant interactions from the Database of Insects and their Food Plants [DBIF] version 2'. NERC Environmental Information Data Centre. Available at: https://doi.org/10.5285/33a825f3-27cb-4b39-b59c-0f8182e8e2e4.

Petsopoulos, D. *et al.* (2021) 'Using network ecology to understand and mitigate long-term insect declines', *Ecological Entomology*. doi: 10.1111/een.13035.

Piper, A. M. *et al.* (2019) 'Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance', *GigaScience*. Oxford University Press, 8(8), pp. 1–22. doi: 10.1093/gigascience/giz092.

Preston, J., VanZeeland, A. and Peiffer, D. A. (2021) 'Innovation at Illumina: The road to the $600 human genome', *Nature Portfolio*.

Shokralla, S. *et al.* (2014) 'Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens'. doi:

10.1111/1755-0998.12236.

Singer, G. A. C. *et al.* (2019) 'Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater', *Scientific Reports*. Springer US, 9(1), pp. 1–12. doi: 10.1038/s41598-019-42455-9.

de Souza Amorim, D. *et al.* (2022) 'Vertical stratification of insect abundance and species richness in an Amazonian tropical forest', *Scientific Reports 2022 12:1*. Nature Publishing Group, 12(1), pp. 1–10. doi: 10.1038/s41598-022-05677-y.

Spears, L. R. and Ramirez, R. A. (2015) 'Learning to Love Leftovers', *American Entomologist*. Narnia, 61(3), pp. 168–173. doi: 10.1093/ae/tmv046.

Srivathsan, A. *et al.* (2021) 'MinION barcodes: biodiversity discovery and identification by everyone, for everyone', *bioRxiv*. BMC Biology, p. 2021.03.09.434692. Available at: https://doi.org/10.1101/2021.03.09.434692.

Tatsumi, S., Iritani, R. and Cadotte, M. W. (2022) 'Partitioning the temporal changes in abundance-based beta diversity into loss and gain components', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 13(9), pp. 2042–2048. doi: 10.1111/2041-210X.13921.

Wang, Y. *et al.* (2012) 'mvabund– an R package for model-based analysis of multivariate abundance data', *Methods in Ecology and Evolution*. John Wiley & Sons, Ltd, 3(3), pp. 471–474. doi: 10.1111/J.2041-210X.2012.00190.X.

Yu, D. W. *et al.* (2012) 'Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring', *Methods in Ecology and Evolution*, 3(4), pp. 613–623. doi: 10.1111/j.2041-210X.2012.00198.x.

Zhou, X. *et al.* (2013) 'Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification', *GigaScience*, 2(1), p. 4. doi: 10.1186/2047-217X-2-4.

Zizka, V. M. A. *et al.* (2022) 'Long-term archival of environmental samples empowers biodiversity monitoring and ecological research', *Environmental Sciences Europe*. Springer Berlin Heidelberg, 34(1). doi: 10.1186/s12302-022-00618-y.

Chapter 5: Discussion

Declaration: An abridged version of the discussion has already been published: The published piece can be found at:
https://resjournals.onlinelibrary.wiley.com/doi/full/10.1111/een.13035

5.1 Thesis objectives and summary of main findings

Global biodiversity is currently under threat and rigorous and scalable monitoring is needed to understand and mitigate the impacts. Insects are one of the most diverse terrestrial groups of biodiversity that are facing numerous threats, but despite the claims of global insect declines (van Klink *et al.*, 2020), the picture is blurred with a lot of uncertainty for insect populations (Wagner *et al.*, 2021). One of the main reasons for this uncertainty is the lack of long-term data for insect populations for many regions of the world. There has also been a taxonomic bias towards more charismatic groups for biodiversity assessment (Rocha-Ortega, Rodriguez and Córdoba-Aguilar, 2021), resulting in less than 1% of insect taxa assessed for IUCN status. Insects are of course one of the most diverse groups with millions of species discovered and described so far (Stork, 2017) which further complicates our ability to assess the state of insect populations broadly. Therefore, there is a clear need for more comprehensive monitoring efforts, which could be facilitated by advances in molecular ecology. A great source for long-term data has been a small number of insect monitoring schemes, even though these have been developed with different aims (such as pest surveillance). A good example of this is the Rothamsted Insect Survey (RIS) suction trap network, which has been monitoring aphids daily in the UK since the 1960s, providing long-term data for decades and helping to understand ecological patterns but also highlighting potential threats to insect populations (Bell, Blumgart and Shortall, 2020; Grabener *et al.*, 2020). As in the case of RIS the focus of insect monitoring schemes tends to be for particular taxa depending on the aim of the scheme whether if it is for surveillance or biodiversity assessment. Yet, they capture hundreds of other taxa as well, which is typically called the bycatch. Bycatch usually remains untouched (or is discarded) as it is not the principal aim of such monitoring schemes and there are multiple barriers to retain it, especially the taxonomic expertise needed, financial and logistical reasons (Hribar, 2020). However, new tools can help us overcome such challenges. Next Generation Sequencing (NGS) approaches can provide us with the necessary tools that can scale up insect biodiversity monitoring (especially bulk

samples caught in a range of entomological traps) and effectively revolutionising the way we can monitor insects (van Klink *et al.*, 2022).

The overall aim of this thesis is to test and evaluate the added value that NGS approaches, and more specifically DNA-metabarcoding, can offer to insect monitoring schemes. Although the use of DNA-metabarcoding has been largely used for a range of other biodiversity assessments, its use with application to insect monitoring schemes that serve as bio-surveillance, has been scarce. In this thesis I aimed to develop and evaluate the reliability and use of DNA-metabarcoding for insect monitoring schemes by focusing on three major aims that correspond to individual chapters.

Aim 1 (Chapter 2): Develop and evaluate a DNA-metabarcoding approach to uncover archival (target aphid) samples from insect monitoring schemes.

Aim 2 (Chapter 3): Determine bycatch diversity from insect monitoring schemes, highlight its value for scaling up monitoring efforts and present archival bycatch samples as a source for constructing multi taxa time series

Aim 3 (Chapter 4): Use DNA-metabarcoding as a tool to uncover bycatch diversity within a nationwide insect monitoring scheme, showcasing its use for insect monitoring at large spatial scales.

5.2 Summary of findings

In Chapter 2, I assessed the reliability of DNA-metabarcoding for retrieving insect sample DNA from archives. Molecular methods have previously been used successfully to identify a variety of old specimens including insects (Gilbert *et al.*, 2009). Despite the limitations that archival samples have, especially DNA degradation, methods have been developed for extracting DNA from insect specimens up to hundreds of years old (Gilbert *et al.*, 2007). Insect monitoring schemes, particularly ones with archival samples, do not necessarily preserve their samples in ideal conditions for DNA-research. RIS for example, although archiving all of their samples in a suitable solution for long-term preservation of DNA, the samples are archived in room temperature which is not ideal for extracting high quality DNA (Gray, Pratte and Kellogg, 2013). I focused on a subsample of RIS aphid samples that included more than 66% of all genera identified within RIS for the suction trap at Newcastle and spanned 16 years from 2003-2018

with the oldest samples being 18 years old at the point of the DNA-extraction. I found high congruence for the taxonomic composition of the DNA-metabarcoding dataset compared with taxonomically identified data (80% at the genus level, 51% at the species level), but with a lot of variation between the samples. Sample age did not affect congruence at either taxonomic level and the relationship was not linear, this means that it is possible to use this method to go back in time irrespective of sample age. This is good news for researchers wishing to use archival samples at RIS as it means that thousands of insect samples could potentially be identified for the first time with the use of DNA-metabarcoding. Doing this using taxonomy would be prohibitively expensive and laborious. Indeed, previous efforts to do so within RIS typically focused on high taxonomic levels such as family or insect biomass (Grabener *et al.*, 2020). Although I was not able to model a high amount of congruence variability to a specific factor, such as sample evenness or sample age, taxonomic assignment did have an impact on congruence with BLAST having a mean congruence of 80% and RDP of 69%, which supports growing evidence that newer approaches are not always the best and there is a need for comparison/validation between taxonomic assignment methods (Hleap *et al.*, 2021). Sequencing depth did not influence congruence in this study despite the well documented importance of achieving sufficient depth in DNA-metabarcoding studies (Shirazi, Meyer and Shapiro, 2021). There can be other factors influencing congruence such as primer bias and choice of loci (van der Loos and Nijland, 2021), which were not considered in this study as the main aim was to see if sample age had an influence. A limitation with my approach was that DNA-metabarcoding had on average 3.6 false positives; I show that this could be controlled by using the taxonomic information or by using sample sequence thresholds. However, congruence fell drastically to 50% when using sequence threshold filtering and false negatives increased from 1 to 4 on average which means a lot of real information was dropped from the dataset. The scale of false positives and false negatives should be further assessed, for example by picking up samples and using single specimen DNA-barcoding to highlight any potential misidentifications by the taxonomists or by avoiding the PCR step altogether and processing the samples with DNA-metagenomics. It is important to note however, that the source of such contamination within RIS that might lead to false positives or false negatives is more likely to be from the way samples are handled in RIS prior to archiving. Therefore, addressing the issue at the source by following processing procedures that minimize sample to sample contamination which comes from potentially using the same

sorting trays could solve the problem of contamination. Furthermore I showed that a non-destructive DNA extraction yielded similar results with no effect on congruence. This is important for insect monitoring schemes that can be hesitant to provide samples for DNA analysis if destructive methods are used. This ensured that the samples processed can be accessible in the future or even revisited to create specimen-based reference databases.

Chapter 3 focused on bycatch diversity within RIS (using the accompanying samples caught with the aphids in Chapter 2). Bycatch within RIS does not typically get analysed and therefore can be seen as an untouched but valuable resource for long-term insect data as it has been preserved for decades. Here, DNA-metabarcoding can be seen as a potential key to unlocking this resource. After validating DNA-metabarcoding and understanding the limitations of it for RIS (which were the need of taxonomic information for filtering out false positives or the use of filtering methods with the drawback of losing information), I used DNA-metabarcoding to recover the bycatch diversity from a time series of 16 years providing for the first time a species list for bycatch within the RIS 'Newcastle' trap. Unsurprisingly, bycatch diversity was high with more than 800 insect taxa identified to a species level. Within the bycatch, many taxa could be identified as insect beneficials (e.g. pollinators) and even additional pests, highlighting the value that bycatch has to offer for insect monitoring more broadly. For example, around 50 taxa of aphid parasitoids and predators were identified. As RIS currently focuses on informing farmers regarding aphid migration for the prophylactic use of insecticides, information regarding beneficial presence could be used as well, although with caution as presence of the predators does not necessarily equal higher biological control. Additionally, I identified more than 9 species which have no record within the UK, some of which were found only in North America. Validation for these would be necessary by expert taxonomists, but this shows how bycatch can be used, for example, for early detection of new invasive species or to expand knowledge on current geographic species ranges. Similarly, to Chapter 2, I controlled for false positives by using sample sequence thresholds with stringent filtering criteria of 1%, resulting in more than half the taxa identified being dropped. As I do not have morphological information it is difficult to assess exactly how false positives or negatives are affected by this, something that should be confirmed in future studies. However, identifying morphologically diverse communities such as bycatch with more than 14 arthropod orders here is a monumental task. Especially given the high number of insect

numbers within samples, that can add up to thousands of individuals. As I purposely selected one trap from RIS and a subset of the samples going back to 2003, the number of species recovered here is probably an underestimate of real bycatch diversity. Nevertheless I have shown how it is possible to create time series for hundreds of insect taxa, and calculate diversity metrics such as richness and species turnover. Richness increased over time for the majority of orders, yet as I only subsampled the whole time series from one trap it is uncertain whether this is a real pattern or an artefact of the sampling, despite choosing sampling days at random. Finally, sequencing depth did not show an increasing positive effect on richness, which was unexpected. Although in this study sequencing depth was high as I used the Illumina NovaSeq platform, and the suggested species accumulation curves show a coverage of more than 95% for my sampling, I believe that bycatch samples can be highly complex and may harbour increased richness requiring even deeper sequencing. Interestingly, sample age had a slight impact on species richness as suggested by the model fitted, suggesting that degradation of DNA might be an important consideration for the bycatch samples in RIS.

Finally, in chapter 4 I apply DNA-metabarcoding to a live, nationwide scheme where archiving samples is not possible. I use DNA-metabarcoding to identify and analyse patterns of bycatch diversity within the nationwide pest monitoring scheme at Fera: the Fera Yellow Water Pan Trap Network (YWP), highlighting the potential of the tool at a large spatial scale. Here, throughout the UK (with some geographical biases as the sampling at Fera is based on farmers interest and uptake) I identified more than 500 insect taxa, despite samples taken in agricultural monocultures (i.e potato fields). Additionally, more than 100 taxa were identified as beneficials, such as pollinators, predators, and parasitoids. A further 7 potato pests were also identified, which might be important for the farmers using YWP. As part of this work I also assessed potential differences in biodiversity descriptors between sequence platforms of varying output: the Illumina MiSeq and NovaSeq with a 20-fold output difference. Surprisingly, both datasets were highly similar in their taxonomic composition sharing more than 90% of taxa captured and similar biodiversity patterns. This is important for monitoring schemes as the cost can be prohibitively expensive when it comes to large datasets such as the one presented in this chapter (i.e. with more than 188 bulk arthropod samples and 83 locations across the UK), as it would mean that for the bycatch within YWP a single MiSeq run could suffice to provide a list of species. YWP methodology currently discards all samples that are

captured as there is limited space for archiving. DNA-metabarcoding can provide a digital record of bycatch diversity for many monitoring schemes missing the capabilities of archiving samples, and there are suitable databases for such information to be stored (such as GBIF see (Andersson *et al.*, 2022). Overall, I found complex patterns of biodiversity with high species turnover across the sites and months sampled, with a median beta diversity of 0.90 based on the Jaccard index of dissimilarity. The most dominant insect order from these pan traps (indeed, as in the case of samples from the RIS suction traps) were Diptera, as they can be some of the most abundant flying insects. Two of the main components of beta diversity are nestedness and turnover both driven by environmental variation and geographical distances. Species turnover was the main driver of beta diversity here and I found no correlation between geographical distance and dissimilarity, despite many samples collected being very close to each other. Highlighting the importance of local in field factors despite the locations being agricultural monocultures. Partitioning the dataset into gain and loss components revealed that beta diversity is mainly driven by taxa themselves. Meaning that the most dominant (in terms of relative read percentage and frequency) taxa also drove the differences in beta diversity, meaning that appearance or disappearance of such taxa led to the high differences observed. Finally, by using metadata in the dataset, and by computing indices of landscape complexity, I found that richness decreased by increasing latitude but no effect of landscape complexity was found. This suggests that other factors particular at the local farm level, such as climate or even management practices can be at play. A further build-up on this dataset could include many more factors such as in field pest-management strategies and local environmental variables such as rainfall, temperature and wind. Which could help explain why species turnover is so high among these monocultures.

5.3 Filling up insect data gaps by merging NGS and archival samples

This thesis adds to growing evidence that NGS approaches such as DNA-metabarcoding can help researchers to overcome many of the obstacles inherent in traditional monitoring schemes that are dependent on morphological identification by skilled taxonomists (van Klink *et al.*, 2022). I have shown how DNA-metabarcoding can add value to two nationwide monitoring schemes in the UK. However, this is not limited to these two schemes; it could be applied to any insect monitoring scheme that uses passive sampling techniques or not

irrespective of their capabilities to archive samples. Although it is difficult to identify such schemes, databases of long-term insect data such as Insect Change (van Klink *et al.*, 2021). EntoGEM (Grames *et al.*, 2022) (found in: https://entogem.github.io/) can act as a first source for finding them by looking at studies that make use of long-term data. Other sources for identifying schemes with collections or insect archival samples can be found on the Entomological Collection Network (https://ecnweb.net/). Despite numerous databases it can still be difficult to identify such schemes. For example, the Rothamsted Insect Survey is not included in some of the above databases. Development of databases where insect monitoring schemes along with the information of their samples can be found is a crucial step for streamlining applications of NGS at larger scales. There are currently few published studies that make use of DNA-metabarcoding or NGS tools on long-term datasets from insect monitoring schemes (Ji *et al.*, 2020; Basset *et al.*, 2022). Perhaps the best example of NGS applications for global insect monitoring is the BIOSCAN program (https://ibol.org/programs/bioscan/), monumental task for capturing and describing arthropod diversity, patterns and their interactions at the global scale by using malaise traps with more than 2000 sampling locations to date. RIS is a unique monitoring scheme in the sense that they have been archiving all of their bycatch samples since the 1960s. With a network of 16 suction traps across agricultural fields and daily archived samples it offers a unique opportunity to construct time series at a nationwide level for hundreds or even thousands of insect taxa. I have shown that this is possible for samples dating back to 2003 (~18 years old). Indeed, given that DNA degradation does not appear to be an issue it could be possible, given sufficient funding, to metabarcode all RIS samples unlocking some of the longest multi-taxa time series in the world. The effort of doing this via traditional means like morphological identification would be monumental, but NGS can offer the appropriate tools for such efforts (van Klink *et al.*, 2022). However, a major disadvantage for DNA-metabarcoding for insect monitoring schemes is the lack of quantitative data (Lamb *et al.*, 2019). Insect monitoring schemes and long-term datasets to understand insect declines require abundance information on the species captured. Yet, DNA-metabarcoding and other approaches such as mitogenomics can bias the quantitative nature of such data. Metabarcoding requires amplification of the template DNA via PCR, primer-template mismatches, unsuitable primer choice or loci can all alter the relationship of input DNA/biomass of species and the reads assigned to them (Deagle *et al.*, 2019). Mitogenomics,

although avoiding the PCR step biases, can still come from the different number of mitochondrial copies available in each taxa (Ji *et al.,* 2021), something that is unknown for many organisms. Despite all such biases and limitations current efforts are changing this and it is now possible to retrieve quantitative information from both DNA-metabarcoding and mitogenomics (Shelton *et al.*, 2022). Obtaining quantitative information will very likely increase the adoption of such methods and insect monitoring schemes like RIS offer ideal datasets to validate the quantitative nature of NGS approaches.

5.4 Scaling up insect monitoring with information on bycatch

Insect monitoring schemes with archival samples such as RIS are scarce, most monitoring schemes perhaps do not have the capability to archive their samples, such as the Fera Yellow Water Pan Trap Network. I have shown that even when that is the case, DNA-metabarcoding can still add value to such schemes by analysing the bycatch. Bycatch, although ubiquitous in monitoring schemes, is not well documented for insects (Hribar et al., 2021). A literature search for the terms “by-catch OR bycatch” and insects shows that less than 40 published articles can be found in Web of Science. Although this can partially be attributed to non-standardized use of the term ‘bycatch’ in studies, results show that when it comes to insect bycatch it is much less known (Spears and Ramirez, 2015). Many studies focus on finding ways of minimising bycatch (Spears *et al.*, 2016; Sétamou *et al.*, 2019) yet very few on what actually bycatch diversity consists of. Of those that do, many have highlighted the importance of bycatch for entomological and ecological research (Hribar, 2020). As shown in both chapters 3 and 4, bycatch diversity within monitoring schemes can be very high and complex. It can include beneficial species within agricultural systems such as predators, parasitoids and pollinators. All such ‘beneficials’ can provide ecosystem services that have impact at the farm level by pest suppression or even yield increase by pollination (Stein *et al.*, 2017). Therefore, it could be possible for pest monitoring schemes that provide farmers information regarding pest populations to give data on pest predators, parasitoids and pollinators too, so farmers can have a more informed decision regarding their management practices. Admittedly, this would require a better link between the presence of beneficials and actual provision of services as the degree that, for example, pest suppression can be linked to abundance of beneficial insects is limited. However, newer approaches that are based on species

interactions uncovered by NGS can help us better understand such interactions (see below section 5.5.6) and link it to ecosystem functioning (Windsor *et al.*, 2022). I also found numerous additional pests within the bycatch for both schemes and potential insects that have previously never been recorded in the UK, one of which is native only to North America and Australia. Therefore, bycatch could be used as an early warning system for invasive species aided by the rapid species identification offered by DNA-metabarcoding, but it can also be used to increase our understanding about geographical distribution of species. Finally, there is also an ethical reason for the inclusion of bycatch taxa. Particularly in the context of insect declines, numerous insects and arthropods species are captured within pest monitoring schemes most of which are bycatch, many of which are never analysed or worse still thrown away (Fischer and Larson, 2019). As I have shown, the mean ratio of bycatch species to target pests was around 10 to 1 for RIS and YWP. The degree to which such passive sampling methods influence local populations of bycatch is uncertain and requires further investigation. Additionally, it is saddening that so much information is potentially lost as pest monitoring schemes that do not have the resources to collect or analyse bycatch throw these away, especially when there is a clear lack of insect data for many parts of the world (Wagner *et al.*, 2021). Finally, it is clear that bycatch within monitoring schemes is rarely documented. There is an urgent need for databases where information on schemes that have bycatch is available to researchers so that collaboration and research efforts on bycatch are promoted. Thankfully, guidelines for doing so have been published (Montgomery *et al.*, 2021) and there are many biodiversity databases that could already incorporate such datasets such as GBIF (https://www.gbif.org/dna).

5.5 Beyond simply identifying species: future approaches

NGS approaches are of course a big part of the tools that can help researchers to scale up insect monitoring, but not the only ones. Advances in computer vision, acoustic monitoring and radar technologies all offer up immense potential for scaling up insect monitoring and biomonitoring in general (Van Klink *et al.,* 2022). The combination of such tools could potentially lead to the much anticipated automated monitoring systems (Bohan *et al.*, 2017; Derocles *et al.*, 2018). Of course, it must be noted that such technologies will still never replace the need for taxonomists and specialist taxonomic knowledge, but should be taken

as complementary. For example, bycatch species could be initially screened by DNA-metabarcoding and then picked up by expert taxonomists that are interested in particular taxonomic groups for further morphological analyses. The non-destructive method for DNA-extraction, developed in my thesis, would also make it possible for experts to create well-presented voucher specimens which can populate reference databases such as BOLD (Ratnasingham and Hebert, 2007) as currently complete and well curated reference databases for insects are an issue in DNA based studies (Magoga *et al.*, 2022). This could be a win-win situation for both molecular ecologists and taxonomists. Furthermore, I have shown that without the taxonomic information, DNA-metabarcoding, although reliable, there is loss of information by using filtering criteria. There can also be uncertainty regarding the degree that false positives or negatives can impact the recovered community which furthers highlight the need for morphological data. Such issues need to be addressed by combining metabarcoding and taxonomical identifications, as done in the second chapter of this thesis. Additionally, pest surveillance requires a rapid turnover from sampling to identification. This is usually done within a day in RIS, whilst metabarcoding can take a couple of hours to days depending on the methods used for DNA extraction, amplification and sequencing, which is still a bottleneck for rapid DNA-based identification. Emerging technologies such as portable sequence devices, like the Oxford Nanopore sequencer, can help alleviate such issues by providing real-time and in-field sequencing of samples (Kipp *et al.*, 2021). A unique aspect of NGS when compared to other tools for insect monitoring is the eco-evolutionary nature of the data. As multiple organisms are sequenced both ecological and phylogenetic signals can be captured with such datasets which enable us to answer questions beyond just simply identifying species (Tedersoo *et al.*, 2021). Additionally, long-read sequencing technologies can overcome many of the challenges of short-read based technologies which DNA-metabarcoding is currently based on. Giving us a way to increase the signal from such datasets but also provide more accurate species identifications (Porter and Hajibabaei, 2020). Yet, perhaps one of the most important layers of information that can be uncovered via NGS approaches is that of species interactions, as species interactions can not only provide a mechanistic framework to understand insect declines (Petsopoulos *et al.*, 2021) but also quantify ecosystem functioning such as pest regulation and pollination, which would be important information for management practices. Recent advances in molecular ecology, machine learning, big data and network theory (e.g. Bohan *et al.*, 2013; Makiola *et al.*, 2020))

provide new opportunities for unlocking a more holistic understanding of the mechanisms driving insect declines.

5.5.6 Constructing ecological networks using insect survey data

Before describing how insect surveys can be used to construct networks, I distinguish between two forms of insect survey datasets (see Figure 5. 1A, for graphical representation): 'physical' where biological samples are retained, taxonomically identified or not, and 'digital' where time series of insect samples are identified to some taxonomic level and stored electronically. The forms can affect the way ecological networks can be constructed. There are three main, non-exclusive methods for constructing networks (Delmas *et al.*, 2019): i) literature searches where trophic and/or other interactions are described; ii) observation based on empirical study; and iii) predicting species interactions from community data (Figure 5. 1B). The rapid growth and interest in network ecology in recent years has resulted in the proliferation of datasets around the world (e.g. Mangal https://mangal.io; Poisot *et al.*, (2016)), often with a focus on bipartite interactions. Empirically derived networks are the result of painstaking observations of species-interactions in the field, but are increasingly being augmented using DNA-metabarcoding, especially to determine difficult-to-observe interactions (Kitson *et al.*, 2019). Alongside these developments, network inference approaches are being applied to insect community data (e.g. to species lists generated using environmental DNA) whereby species-interactions are predicted based on co-occurrence (presence-absence) conditional probabilities (but see Blanchet *et al.*, (2020) for a critique).

Recent work by Vanbergen *et al.*, (2018) provided a novel way of showing how multiple sources of biological recording data, that included citizen science records, were used to build nationwide plant-pollinator networks, and found positive relationships between agricultural land cover and both pollinator generality (one of many network metrics) and robustness under several extinction scenarios. Building on this, combining network construction methods using long-term target and non-target insect biomonitoring scheme data represents a significant opportunity to understand the extinction dynamics of more holistic insect interaction networks.

5.5.7 Network applications for insect biomonitoring schemes

Using the RIS 12.2m suction-trap network as an exemplar, I show how biomonitoring schemes can be used to generate different types of network data, and, here I use the bycatch from the suction traps as a representative sample of the aerial community. First, it is possible to use collected insect samples to look for direct interactions between species. For example, molecular methods can be used to screen ladybird gut contents for aphids (predator-prey interactions), and/or aphids (and non-target species) for symbionts and parasitoids (Figure 5. 2B; ii), with interaction data retained in bioinformatic pipelines that can then be used to construct networks (see Kitson *et al.,* 2019). Network data can then be used to examine long-term changes in species-interactions, in this case regarding questions of disease transmission and pest regulation. Second, the bycatch can be identified by metabarcoding the bulk sample, representing a more holistic community of interacting species when trapping allows (acknowledging in this instance the focus on aerial insects using a particular trapping method). Testing a range of co-occurrence algorithms on insect community data derived from metabarcoding is necessary (Figure 5. 2B; iii), but for RIS validation is easier as the interactions between agricultural insects are generally well documented. Third, networks can be constructed in space and time by scaling up the molecular processing of catches using automation, resulting in daily, weekly and monthly insect networks at each suction trap across the country (acknowledging the need for validation (Piper *et al.*, 2019) and appropriate methods for obtaining insect abundance data (Ji *et al.*, 2020). Importantly, there is an opportunity to construct historic networks by metabarcoding stored insect sample archives, potentially non-destructively and assuming DNA degradation is not a hindrance. Long-term changes in network structure, complexity, interaction turnover and robustness can then be examined in relation to environmental change, allowing new insights into the drivers of insect declines and the consequences for ecosystem functioning.

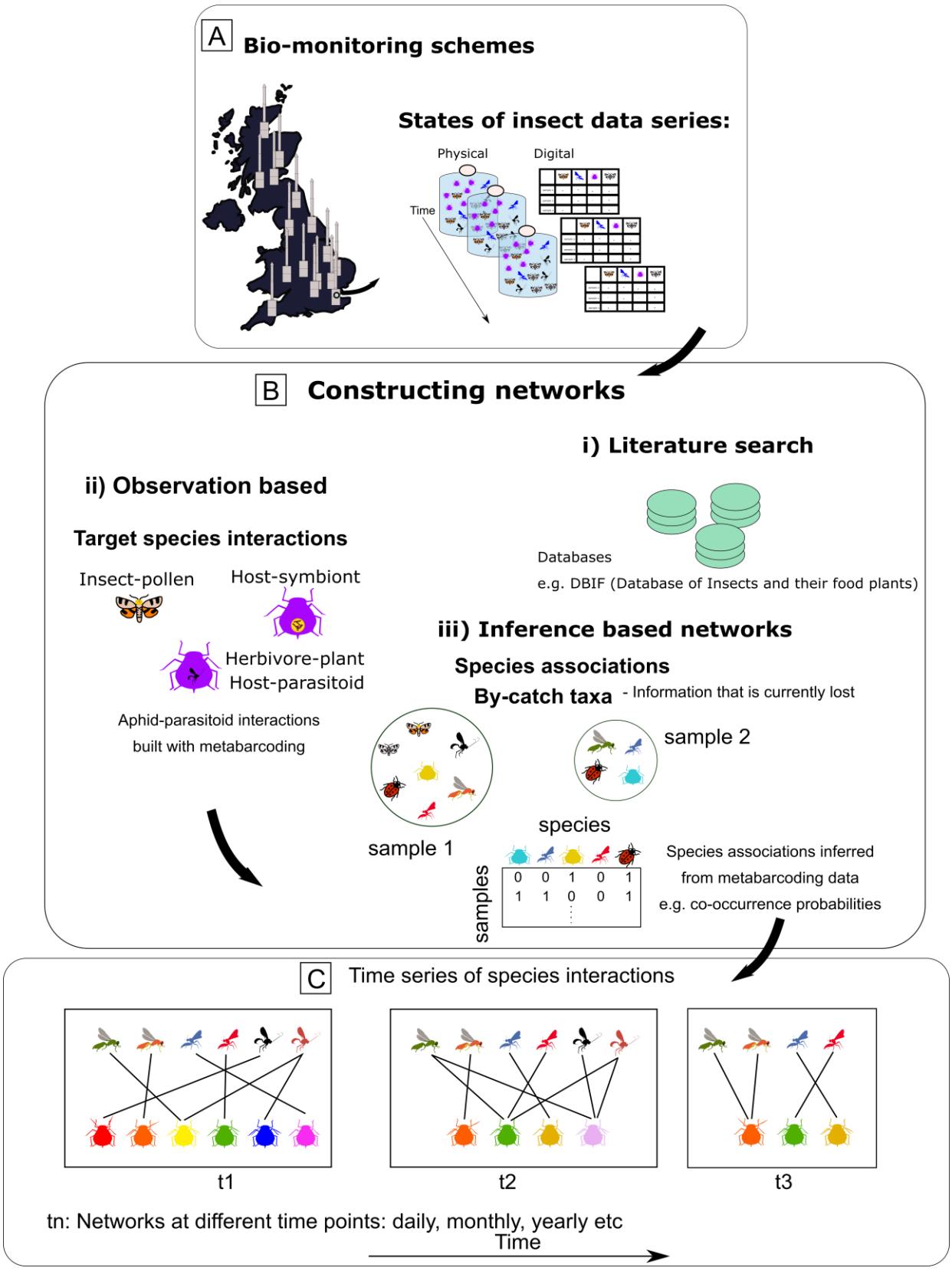


Figure 5. 1 A) Insect biomonitoring schemes can have collections of insect data series which can be physical (archived samples) or digital (in databases). B) Constructing networks: Constructing networks from such samples can be achieved from: i) literature searches databases etc.; ii) observation, based on targeted interactions via metabarcoding e.g. host-

parasitoids; iii) From inference, where species associations are inferred based on co-occurrence. C) Using different sources can result in time series of species interactions at different time scales (e.g. daily, weekly, yearly).

5.5.8 Understanding insect declines: an ecological network approach

General reviews on the analysis, applications and limitations of networks already exist (Delmas *et al.,* 2019). In the context of insect declines, however, I suggest the following directions for further research (see Figure 5. 2). First, the construction of insect networks using 'physical' and 'digital' long-term datasets can be applied to multiple types of insect monitoring schemes around the world (and well beyond those highlighted here). Some additional examples include The Global Malaise Trap Program (http://biodiversitygenomics.net/projects/gmp/) and global volunteer insect monitoring initiatives (see Bried *et al.*, 2020). This can provide new opportunities to examine whether network structure, complexity and interaction turnover has changed over time and across large spatial scales, whilst identifying the key drivers. Bohan *et al.,* (2017) show how next-generation sequencing data combined with machine-learning could be combined in future global biomonitoring schemes, through autonomous samplers deployed over large geographical areas. This could construct highly replicated networks of ecological interactions, allowing potential changes in ecosystem function to be observed for the first time. Second, examining the robustness of networks to species extinctions shows promise (see Kehoe, Frago and Sanders, 2021) regarding extinction cascades as a driver of insect decline). For mitigation purposes it not only has the potential to identify 'fragile' insect groups, but also species which are disproportionately important for the integrity of the network which could be targeted for conservation management. For example, Pocock, Evans and Memmott, (2012) identified insect pollinator networks as most vulnerable to species loss on farmland compared to the other animal groups studied. At the farm scale, their analysis identified common agricultural plants such as clovers (*Trifolium* spp.), thistles (*Cirsium* spp.) and buttercups (*Ranunculus* spp.) that theoretically could be managed to increase robustness and improve overall agroecosystem resilience. Similarly, Evans et al. (2013) showed how habitat robustness analyses could be used to identify key agricultural habitats for targeted management to increase resilience, in this case hedgerows and waste ground, which together comprised <5%

of the total farm area. Scaling up further, Redhead *et al.*, (2018) showed how network robustness analyses can also be used to identify key species traits that enable persistence in highly perturbed landscapes. Thus, robustness measures could be used in ecosystem restoration to boost the resilience of insect communities, although this is yet to be tested empirically. Third, the use of DNA-metabarcoding (and in the future metagenomics (Cordier *et al.*, 2020) to construct phylogenetically-structured networks is a research priority (Raimundo, Guimarães and Evans, 2018). Currently, the use of adaptive network models for predicting ecological restoration outcomes shows considerable promise but are severely hampered by the lack of long-term species-interaction data. We contend that this could quickly be overcome by making use of long-term biomonitoring insect archive samples, such as RIS. Finally, recent advances have started to pull together different network types into multilayer networks (Pilosof *et al.*, 2017). This presents a new way of examining the implications of insect declines on a large array of other taxa that interact with them (e.g. birds and bats), providing new ways to examine how the loss of some insect groups leads to further insect extinctions.

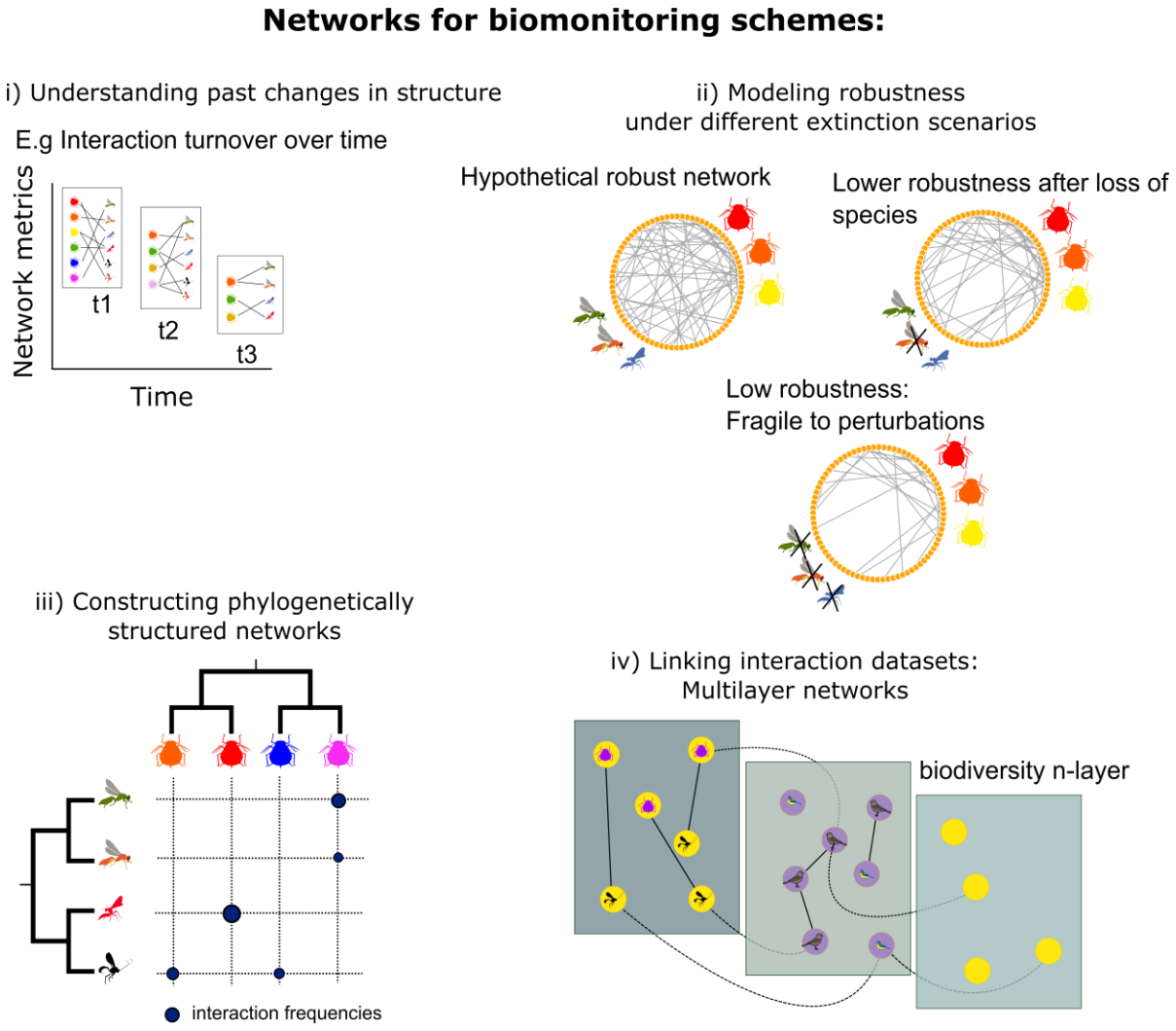


Figure 5. 2 Networks for biomonitoring schemes: Having constructed such networks we highlight four possible uses: i) Understanding past changes in structure by analyzing network metrics through time; ii) Modelling robustness under extinction scenarios; iii) Construct phylogenetically structured networks to examine eco-evolutionary dynamics; iv) Linking interaction datasets with multilayer networks e.g. insect interaction with birds or bats

5.6 Conclusions

Insect monitoring schemes have some of the best long-term datasets to date that could potentially help us understand how insect populations have been changing across decades. Highlighting some of the major factors that might be affecting them such as climate change and other human driven disturbances. Yet, there is still a lot of uncertainty regarding the state of insect diversity in many parts of the world. In this thesis, I have shown how DNA-

metabarcoding can add value to two monitoring schemes in the UK that focus on pest monitoring and surveillance. I first assessed the reliability of DNA-metabarcoding for identifying archival insect samples that come from RIS, highlighting the advantages and limitations of such an approach but most importantly the potential of using archival samples for constructing time series with the aid of DNA-metabarcoding. I then applied this approach to the bycatch collection of RIS. I have shown that DNA-metabarcoding can be one approach to unlock such collections, scaling up the potential to construct time series for hundreds of taxa for an unknown fraction of biodiversity within insect monitoring schemes: the bycatch. Additionally, I've highlighted the applied importance of bycatch diversity within a biomonitoring context as bycatch includes potential beneficial taxa that provide many ecosystem services such as pest regulation and pollination, but it can also include many additional pests. Finally, I show how by merging DNA-metabarcoding at large spatial and temporal scales that the Fera Yellow Water Pan Trap Network can be used to understand diversity patterns at a nationwide level, something that was possible because of the scalability of Next Generation Sequencing approaches. As growing evidence on the applicability and reliability of NGS approaches continues to grow, adoption of these newly emerging approaches and concepts within insect monitoring schemes that rely on traditional morphological identification will increase, leading to scalable and cost-efficient ways to monitor insect communities and understand how their patterns have changed through time.

5.7 References

Anders F. Andersson *et al.* (2022) 'Publishing DNA-derived data through biodiversity data platforms', Version 30. Available at: https://docs.gbif.org/publishing-dna-derived-data/1.0/en/.

Basset, Y. *et al.* (2022) 'Comparison of traditional and DNA metabarcoding samples for monitoring tropical soil arthropods (Formicidae, Collembola and Isoptera)', *Scientific Reports 2022 12:1*. Nature Publishing Group, 12(1), pp. 1–16. doi: 10.1038/s41598-022-14915-2.

Bell, J. R., Blumgart, D. and Shortall, C. R. (2020) 'Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain', *Insect Conservation and Diversity*, 13(2), pp. 115–126. doi: 10.1111/icad.12412.

Blanchet, F. G., Cazelles, K. and Gravel, D. (2020) 'Co-occurrence is not evidence of ecological interactions', *Ecology Letters*, 23(7), pp. 1050–1063. doi: 10.1111/ele.13525.

Bohan, D. A. *et al.* (2013) *Networking agroecology. integrating the diversity of agroecosystem interactions*, *Advances in Ecological Research*. doi: 10.1016/B978-0-12-420002-9.00001-9.

Bohan, D. A. *et al.* (2017) 'Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks', *Trends in Ecology and Evolution*, 32(7), pp. 477–487. doi: 10.1016/j.tree.2017.03.001.

Bried, J. *et al.* (2020) 'Towards global volunteer monitoring of odonate abundance', *BioScience*, 70(10), pp. 914–923. doi: 10.1093/biosci/biaa092.

Cordier, T. *et al.* (2020) 'Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap', *Molecular Ecology*, (April), pp. 1–22. doi: 10.1111/mec.15472.

Deagle, B. E. *et al.* (2019) 'Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data?', *Molecular Ecology*, 28(2), pp. 391–406. doi: 10.1111/mec.14734.

Delmas, E. *et al.* (2019) 'Analysing ecological networks of species interactions', *Biological Reviews*. John Wiley & Sons, Ltd (10.1111), 94(1), pp. 16–36. doi: 10.1111/brv.12433.

Derocles, S. A. P. *et al.* (2018) 'Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into Ecological Network Analysis', *Advances in Ecological Research*, 58, pp. 1–62. doi: 10.1016/bs.aecr.2017.12.001.

EJ, K. *et al.* (2021) 'Nanopore adaptive sampling for mitogenome sequencing and bloodmeal identification in hematophagous insects'. doi: 10.1101/2021.11.11.468279.

Fischer, B. and Larson, B. M. H. (2019) 'Collecting insects to conserve them: a call for ethical caution', *Insect Conservation and Diversity*. John Wiley & Sons, Ltd, 12(3), pp. 173–182. doi: 10.1111/ICAD.12344.

Gilbert, M. T. P. *et al.* (2007) 'DNA extraction from dry museum beetles without conferring external morphological damage', *PLoS ONE*. doi: 10.1371/journal.pone.0000272.

Gilbert, M. T. P. *et al.* (2009) 'Non-Destructive Sampling of Ancient Insect DNA', *PLoS ONE*, 4(4), p. e5048. doi: 10.1371/journal.pone.0005048.

Grabener, S. *et al.* (2020) 'Changes in phenology and abundance of suction-trapped Diptera from a farmland site in the UK over four decades', *Ecological Entomology*. Blackwell Publishing Ltd, 45(5), pp. 1215–1219. doi: 10.1111/EEN.12873.

Grames, E. M. *et al.* (2022) 'A framework and case study to systematically identify long-term insect abundance and diversity datasets', *Conservation Science and Practice*. John Wiley & Sons, Ltd, 4(6), p. e12687. doi: 10.1111/CSP2.12687.

Gray, M. A., Pratte, Z. A. and Kellogg, C. A. (2013) 'Comparison of DNA preservation methods for environmental bacterial community samples', *FEMS Microbiology Ecology*. Oxford Academic, 83(2), pp. 468–477. doi: 10.1111/1574-6941.12008.

Hleap, J. S. *et al.* (2021) 'Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 21(7), pp. 2190–2203. doi: 10.1111/1755-0998.13407.

Hribar, L. J. (2020) 'Expanding Basic Entomological Knowledge by Using Mosquito Surveillance Bycatch', *Annals of the Entomological Society of America*, 113(6), pp. 439–446. doi: 10.1093/aesa/saaa014.

Ji, Y. *et al.* (2020) 'SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes', *Molecular Ecology Resources*, 20(1), pp. 256–267. doi: 10.1111/1755-0998.13057.

Kehoe, R., Frago, E. and Sanders, D. (2021) 'Cascading extinctions as a hidden driver of insect decline', *Ecological Entomology*. John Wiley & Sons, Ltd, 46(4), pp. 743–756. doi: 10.1111/EEN.12985.

Kitson, J. J. N. *et al.* (2019) 'Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding', *Molecular Ecology*, 28(2), pp. 471–483. doi: 10.1111/mec.14518.

van Klink, R. *et al.* (2020) 'Meta-analysis reveals declines in terrestrial but increases in

freshwater insect abundances', *Science*, 368(6489), pp. 417–420. doi: 10.1126/science.aax9931.

van Klink, R. *et al.* (2021) 'InsectChange: a global database of temporal changes in insect and arachnid assemblages', *Ecology*. Ecological Society of America, 102(6), p. e03354. doi: 10.1002/ECY.3354.

van Klink, R. *et al.* (2022) 'Emerging technologies revolutionise insect ecology and monitoring', *Trends in Ecology & Evolution*. Elsevier, 37(10), pp. 872–885. doi: 10.1016/J.TREE.2022.06.001.

Lamb, P. D. *et al.* (2019) 'How quantitative is metabarcoding: A meta-analytical approach', *Molecular Ecology*. John Wiley & Sons, Ltd, 28(2), pp. 420–430. doi: 10.1111/MEC.14920.

van der Loos, L. M. and Nijland, R. (2021) 'Biases in bulk: DNA metabarcoding of marine communities and the methodology involved', *Molecular Ecology*. John Wiley & Sons, Ltd, 30(13), pp. 3270–3288. doi: 10.1111/MEC.15592.

Magoga, G. *et al.* (2022) 'Curation of a reference database of COI sequences for insect identification through DNA metabarcoding: COins', *Database*. Oxford Academic, 2022, p. 55. doi: 10.1093/DATABASE/BAAC055.

Makiola, A. *et al.* (2020) 'Key Questions for Next-Generation Biomonitoring', *Frontiers in Environmental Science*, 7, p. 197. doi: 10.3389/fenvs.2019.00197.

Montgomery, G. A. *et al.* (2021) 'Standards and Best Practices for Monitoring and Benchmarking Insects', *Frontiers in Ecology and Evolution*. Frontiers Media S.A., 8, p. 513. doi: 10.3389/FEVO.2020.579193/BIBTEX.

Petsopoulos, D. *et al.* (2021) 'Using network ecology to understand and mitigate long-term insect declines', *Ecological Entomology*. doi: 10.1111/een.13035.

Pilosof, S. *et al.* (2017) 'The multilayer nature of ecological networks', *Nature Ecology and Evolution*. Nature Publishing Group, p. 0101. doi: 10.1038/s41559-017-0101.

Piper, A. M. *et al.* (2019) 'Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance', *GigaScience*. Oxford University Press, 8(8), pp. 1–22. doi: 10.1093/gigascience/giz092.

Pocock, M. J. O., Evans, D. M. and Memmott, J. (2012) 'The robustness and restoration of a network of ecological networks', *Science*, 335(6071), pp. 973–977. doi: 10.1126/science.1214915.

Poisot, T. *et al.* (2016) 'mangal – making ecological network analysis simple', *Ecography*. John Wiley & Sons, Ltd, 39(4), pp. 384–390. doi: 10.1111/ECOG.00976.

Porter, T. M. and Hajibabaei, M. (2020) 'Putting COI Metabarcoding in Context: The Utility of Exact Sequence Variants (ESVs) in Biodiversity Analysis', *Frontiers in Ecology and Evolution*. Frontiers Media S.A., 8, p. 248. doi: 10.3389/FEVO.2020.00248/BIBTEX.

Raimundo, R. L. G., Guimarães, P. R. and Evans, D. M. (2018) 'Adaptive Networks for Restoration Ecology', *Trends in ecology & evolution*. Trends Ecol Evol, 33(9), pp. 664–675. doi: 10.1016/J.TREE.2018.06.002.

Ratnasingham, S. and Hebert, P. D. N. (2007) 'bold: The Barcode of Life Data System (http://www.barcodinglife.org)', *Molecular Ecology Notes*. John Wiley & Sons, Ltd, 7(3), pp. 355–364. doi: 10.1111/J.1471-8286.2007.01678.X.

Redhead, J. W. *et al.* (2018) 'Potential landscape-scale pollinator networks across Great Britain: structure, stability and influence of agricultural land cover', *Ecology Letters*, 21(12), pp. 1821–1832. doi: 10.1111/ele.13157.

Rocha-Ortega, M., Rodriguez, P. and Córdoba-Aguilar, A. (2021) 'Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction', *Ecological Entomology*. John Wiley & Sons, Ltd, 46(4), pp. 718–728. doi: 10.1111/EEN.13027.

Sétamou, M. *et al.* (2019) 'Screening Sticky Cards as a Simple Method for Improving Efficiency of Diaphorina citri (Hemiptera: Liviidae) Monitoring and Reducing Nontarget Organisms', *Journal of Economic Entomology*. Oxford Academic, 112(3), pp. 1167–1174. doi: 10.1093/JEE/TOZ045.

Shelton, A. O. *et al.* (2022) 'Toward quantitative metabarcoding', *Ecology*. John Wiley & Sons, Ltd, p. e3906. doi: 10.1002/ECY.3906.

Shirazi, S., Meyer, R. S. and Shapiro, B. (2021) 'Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding', *Ecology*

and Evolution. John Wiley & Sons, Ltd, 11(22), pp. 15766–15779. doi: 10.1002/ECE3.8239.

Spears, L. R. *et al.* (2016) 'Pheromone Lure and Trap Color Affects Bycatch in Agricultural Landscapes of Utah', *Environmental Entomology*, 45(4), pp. 1009–1016. doi: 10.1093/ee/nvw085.

Spears, L. R. and Ramirez, R. A. (2015) 'Learning to Love Leftovers', *American Entomologist*. Narnia, 61(3), pp. 168–173. doi: 10.1093/ae/tmv046.

Stein, K. *et al.* (2017) 'Bee pollination increases yield quantity and quality of cash crops in Burkina Faso, West Africa', *Scientific Reports 2017 7:1*. Nature Publishing Group, 7(1), pp. 1–10. doi: 10.1038/s41598-017-17970-2.

Stork, N. E. (2017) 'How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth?', *Annual Review of Entomology*. Annual Reviews, 63(1), pp. 31–45. doi: 10.1146/annurev-ento-020117-043348.

Tedersoo, L. *et al.* (2021) 'Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 87(17), pp. 1–19. doi: 10.1128/AEM.00626-21.

Vanbergen, A. J. *et al.* (2018) *Potential landscape-scale pollinator networks across Great Britain: structure, stability and influence of agricultural land cover*, *Ecology Letters*. doi: 10.1111/ele.13157.

Wagner, D. L. *et al.* (2021) 'Insect decline in the Anthropocene: Death by a thousand cuts', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 118(2), p. e2023989118. doi: 10.1073/PNAS.2023989118.

Windsor, F. M. *et al.* (2022) 'Network science: Applications for sustainable agroecosystems and food security', *Perspectives in Ecology and Conservation*. Elsevier, 20(2), pp. 79–90. doi: 10.1016/J.PECON.2022.03.001.

Appendix A: Supplementary information for Chapter 2

A.1 DNA extraction details:

As RIS wishes to retain insect samples for future researchers, I aimed to extract the DNA non-destructively by using a short-digestion time lysis step. I used a bead based protocol[protocol #6.3] (Oberacker *et al.*, 2019) with slight modifications on the amount of lysis volume used to adjust it for different sample sizes (in terms of numbers of aphids). See Table 2 column 4 for the amount of lysis and proteinase K used for each sample. In general 62ul to 620 ul of Lysis solution were used. Whilst proteinase K ranged from 3ul to 30ul per sample. A hyperlink is attached to the protocol used in this thesis. With the exception of the lysis step where different volumes were used, all the following steps were followed as in 6.3 tissue protocol: https://bomb.bio/protocols/

Unique.id	Aphid_number	input	filtered	denoisedF	denoisedR	merged	nonchim	Lysis solution (ul)	Proteinase K
00110R	22	139716	87373	87282	87253	84841	82678	240	12
00114R	35	116056	75421	75290	75329	74937	74471	300	15
00116R	41	43203	25956	25885	25825	25046	23941	360	18
00119R	64	90301	52236	51860	52063	49402	45938	600	30
0011R	11	116801	76572	76503	76432	74607	73749	120	6
00121R	56	141794	96914	96802	96820	96162	95375	600	30
00124R	40	108399	73586	73499	73460	73172	72914	600	30
00125R	16	37605	19196	19155	19112	18683	18344	240	12
0015R	74	92838	57113	56952	56981	54272	52649	600	30
0018R	19	70018	41642	41592	41592	40367	39701	240	12
00211R	11	115886	69730	69643	69626	68259	65930	240	12
00213R	13	5843	3238	3156	3161	2903	2898	240	12
00214R	29	25692	12355	12270	12280	12011	11975	300	15
00219R	30	18188	9714	9687	9685	9470	9398	300	15
0021R	17	253025	161543	161397	161428	159969	159116	240	12
00220R	14	45790	30382	30319	30293	29968	29538	240	12
00222R	35	34427	16014	15947	15975	15633	15510	300	15
00225R	19	6176	3310	3268	3302	3204	3204	240	12
0022R	22	216275	140912	140718	140706	138671	138585	240	12
0025R	72	160470	93034	92739	92672	87026	77170	600	30
0028R	18	104568	42210	42045	42151	40106	38829	240	12
0029R	64	105951	33814	33710	33688	32998	32962	600	30
00311R	15	3785	2246	2180	2226	2040	2040	240	12
00312R	48	102655	25607	25391	25504	23716	23253	600	30
00315R	50	90495	35647	35503	35538	33935	30828	600	30
00317R	83	21879	8268	8230	8216	8216	8216	600	30
00318R	78	90190	35423	33001	30160	30160	30160	600	30
00319R	96	50988	28874	28775	28735	28029	27696	600	30

0031R	19	52757	22390	22230	22307	21482	20581	240	12
00323R	11	60714	36753	36647	36685	35663	34369	240	12
00326R	18	80578	52069	51888	52028	50317	49834	240	12
0034R	28	84754	27331	27288	27293	27038	26998	300	15
0036R	24	20473	12621	12575	12564	12325	12203	300	15
0038R	28	19761	13551	13321	10133	10133	10133	300	15
00411R	20	91644	60403	60237	60227	59085	58796	300	15
00413R	32	77763	49106	49053	49069	48376	48138	300	15
00415R	23	205911	133016	132762	132896	130413	128239	300	15
00417R	74	88526	52775	52669	52694	52287	51485	600	30
00419R	74	12576	8026	7961	7997	7631	7399	600	30
0041R	6	324041	195256	195001	195029	190425	171704	60	3
00421R	82	178914	12348	12305	12306	12028	11866	600	30
0043R	13	29615	18407	18361	18314	18018	17971	120	6
0048R	56	114773	74933	74837	74858	73238	71956	600	30
0049R	52	25622	16476	16414	16428	15953	15883	600	30
00511R	14	139555	83446	83122	83202	77113	76469	240	12
00512R	14	72387	47985	47940	47951	44965	44760	240	12
00515R	41	166046	112963	112900	112888	106337	105907	600	30
00518R	81	187747	117866	117544	117612	107905	103763	600	30
00519R	92	93408	50305	50164	50202	48445	47147	600	30
00521R	64	109771	72516	72367	72387	69602	64921	600	30
00523R	102	59508	31754	31631	31678	30149	29790	600	30
00525R	7	8294	4154	4115	4124	3927	3786	60	3
0053R	14	24406	7879	7837	7847	7701	7471	120	6
0054R	62	82228	36464	36345	36381	34111	33516	600	30
0056R	41	202174	131757	131633	131637	122053	120528	600	30
0058R	70	72290	31783	31732	31709	30732	30633	600	30
00610R	65	248445	133528	132885	133215	126395	121308	600	30
00614R	27	43946	16504	16451	16432	16177	16158	300	15
00616R	18	6708	3425	3402	3404	3353	3353	240	12

00617R	16	36982	23583	23446	23521	22911	22581	240	12
00621R	50	16597	10503	10468	10464	10325	10325	600	30
00622R	108	157830	100048	99849	99778	92669	85781	600	30
00623R	86	78841	46988	46904	46888	45920	45062	600	30
00626R	8	14963	10331	10245	10290	10044	9967	60	3
0063R	15	47206	29066	28991	29020	28574	28520	120	6
0065R	24	140362	70104	70013	70018	67722	64413	240	12
0066R	36	36976	23263	23215	23230	22653	22364	300	15
0069R	36	83647	40670	40530	40512	38572	36673	300	15
00710R	12	156571	104960	104775	104839	104223	103955	120	6
00711R	38	133994	87189	87108	87082	85588	85048	300	15
00714R	64	53762	21451	21415	21413	21031	20893	600	30
00716R	96	402313	264844	264616	264695	260301	254130	600	30
00717R	70	230080	155914	155802	155716	152857	143295	600	30
00718R	58	30027	20068	20019	19940	18896	18314	600	30
00719R	50	32273	22150	22137	22142	22107	22076	600	30
0072R	8	33248	16998	16969	16961	16635	16273	60	3
0073R	39	299710	205527	205344	205310	204077	200689	300	15
0074R	20	61704	24559	24484	24497	24161	24131	240	12
0075R	18	11164	7375	7315	7343	7106	7055	240	12
0078R	32	17424	8397	8374	8367	8235	8235	300	15
00810R	16	45583	28695	28659	28634	28212	27904	240	12
00812R	54	33506	20624	20567	20568	20159	19957	600	30
00815R	16	65398	22640	22589	22598	22337	22337	240	12
00817R	14	11181	4899	4871	4878	4576	4576	120	6
0081R	26	47765	32300	32212	32112	30175	26833	300	15
00820R	24	29798	14487	14447	14464	14183	13948	300	15
00822R	28	33288	19963	19907	19897	19452	19311	300	15
00825R	10	46046	28577	28304	28514	27437	26741	120	6
0082R	39	95190	57993	57945	57950	57788	57475	600	30
0085R	90	124932	82523	82131	82478	81639	80863	600	30

0086R	27	60589	41476	41325	41334	39982	38422	300	15
0089R	24	2540	1199	1180	1180	1140	1099	300	15
00910R	76	169521	106785	106513	106226	92634	75875	600	30
00912R	64	124226	72777	72455	72400	60175	49552	600	30
00914R	14	45534	26755	26677	26707	26100	25894	120	6
00916R	26	95135	59556	59489	59502	58821	58246	300	15
00919R	22	37504	20412	20340	20375	20006	19822	300	15
00921R	72	31670	16028	15966	15978	15614	15335	600	30
00923R	78	89219	25486	25453	25451	24945	24830	600	30
00924R	100	67101	40631	40491	40528	38232	36463	600	30
0093R	6	28483	9986	9942	9944	9648	9578	60	3
0095R	86	262304	175772	175601	175592	171889	169697	600	30
0097R	24	156860	104200	104049	104103	100951	95696	300	15
0099R	92	116112	68758	68569	68579	64388	61857	600	30
01011R	37	87168	2759	2759	2755	2695	2695	300	15
01014R	20	32522	17586	11130	11111	10873	10839	240	12
01018R	48	25549	14999	14752	14736	14681	14681	600	30
01019R	20	92496	11145	11120	11093	11093	11093	240	12
0101R	79	455753	278349	276121	275531	271536	270743	600	30
01020R	34	84846	24416	24647	24647	24647	24647	300	15
01022R	59	23702	1866	1659	1659	1659	1659	600	30
0102R	49	96600	52925	47828	47828	47828	47828	600	30
0105R	61	232761	139218	139218	139218	139218	139218	600	30
0106R	39	177196	90401	90401	90401	90401	90401	300	15
0107R	44	47051	15281	15281	15281	15281	15281	600	30
0109R	8	113238	56978	54388	54388	54388	54388	60	3
01110R	12	48575	33104	32329	32329	32329	32329	120	6
01112R	56	98101	10617	10617	10617	10617	10617	600	30
01117R	12	82338	47732	46768	46768	46768	46768	120	6
01118R	20	116008	79714	78937	78937	78937	78937	240	12
01119R	7	201367	132011	130175	130175	130175	130175	60	3

01123R	10	8500	5375	5133	5133	5133	5133	120	6
0112R	16	135107	564	587	587	587	587	120	6
0114R	30	37078	18939	17516	17516	17516	17516	300	15
0118R	54	65362	7802	7802	7802	7802	7802	600	30
0119R	44	31183	1787	1459	1459	1459	1459	420	21
01211R	48	44776	22595	26861	26861	26861	26861	420	21
01215R	19	120860	65677	65677	65677	65677	65677	240	12
01216R	8	140017	90037	86043	86043	86043	86043	60	3
01218R	11	154370	88165	88425	88425	88425	88425	120	6
01220R	42	86602	49149	47554	47554	47554	47554	420	21
0122R	53	145448	92365	92365	92365	92365	92365	600	30
0123R	48	213624	143163	138039	138113	137038	136219	600	30
0126R	52	149345	99374	83342	83326	79365	74227	600	30
0127R	56	206389	138270	137741	137741	137741	137741	600	30
0129R	21	125948	83610	74489	74489	74489	74489	240	12
01312R	11	113044	75104	72111	72111	72111	72111	120	6
01317R	64	51991	34904	30653	30653	30653	30653	600	30
01318R	40	27933	17265	20759	20781	19930	19599	420	21
0131R	108	133484	84978	78484	78484	78484	78484	600	30
01322R	13	32803	20877	20411	20411	20411	20411	120	6
01323R	45	21803	12788	12522	12522	12522	12522	420	21
0132R	38	97684	57123	45615	45626	42860	42072	420	21
0135R	42	77653	52087	52174	52209	50914	49832	420	21
0136R	55	80263	45920	44775	44775	44775	44775	600	30
0138R	77	85973	52274	52622	52724	51841	50648	600	30
0139R	18	121120	78669	79599	79559	78321	77321	240	12
01410R	10	76028	52767	50706	50706	50706	50706	120	6
01415R	15	113403	79731	77358	77358	77358	77358	240	12
01416R	4	18006	12388	12087	12087	12087	12087	60	3
01418R	42	85158	57749	49467	49467	49467	49467	600	30
01419R	41	33568	22171	21549	21549	21549	21549	600	30

01421R	44	34046	21846	22020	22020	22020	22020	600	30
01423R	37	46993	32739	31082	31082	31082	31082	420	21
0142R	9	54874	33376	33322	33322	33322	33322	120	6
0143R	13	104028	69512	67378	67378	67378	67378	120	6
0145R	38	30777	18811	17787	17787	17787	17787	420	21
0146R	22	144157	100092	85250	85264	82882	76606	300	15
0149R	22	118682	74340	74340	74340	74340	74340	300	15
01511R	18	125706	85415	51052	51127	50039	47366	240	12
01512R	9	153539	104716	100589	100589	100589	100589	120	6
01514R	6	73128	51181	47418	47418	47418	47418	60	3
01516R	13	85917	55335	53226	53226	53226	53226	120	6
01518R	51	40487	27481	24011	23998	23683	22797	600	30
01520R	63	181486	125960	101057	101057	101057	101057	600	30
01522R	83	35318	24088	22869	22869	22869	22869	600	30
01523R	59	70048	48981	45161	45161	45161	45161	600	30
0152R	8	50950	32171	32171	32171	32171	32171	60	3
0153R	53	115929	78003	72457	72457	72457	72457	600	30
0155R	22	55731	37444	37003	37003	37003	37003	300	15
0158R	7	11957	6086	5823	5823	5823	5823	60	3
01610R	46	65060	44225	35301	35279	32656	30209	600	30
01611R	21	48413	32423	30800	30800	30800	30800	300	15
01612R	6	54240	35428	31119	31119	31119	31119	60	3
01614R	45	86165	56219	49084	49084	49084	49084	600	30
01615R	56	34784	11954	11842	11842	11842	11842	600	30
0161R	12	116647	79238	77454	77454	77454	77454	120	6
0162R	79	104505	72124	60445	60445	60445	60445	600	30
0164R	7	96230	66689	64772	64772	64772	64772	60	3
0165R	39	44825	31061	28074	28074	28074	28074	420	21
0169R	13	55951	38191	33297	33297	33297	33297	120	6
00EE8R	45	23059	13094	13049	13031	12724	12724	600	30
0EE127R	142	243129	160570	75290	75329	74937	74471	600	30

00ExPos_1	1	10773	5736	5731	5726	5726	5726		60	3
00ExPos_2	1	24090	13996	13979	13967	13947	13947		60	3
00PCRNeg_1	0	6	0	0	0	0	0		62	3
00PCRNeg_2	0	9	0	0	0	0	0		62	3
00PCRpos_1	1	11334	5020	2692	2677	2269	2269	na	na	
00PCRpos_2	1	7934	3797	3788	3652	3652	3652	na	na	

Table A. 1 Unique codes for each sample, number of aphids within individual samples, reads per sample for each filtering step in the DADA2 pipeline, Lysis solution amount and proteinase K used for the samples

SAMPLE	1 HOUR	2 HOURS	6 HOURS
00EEER1	No	Yes	Yes
00EEER2	No	Yes	Yes
00EEER3	No	No	Yes
00EEER4	No	Yes	Yes
00EEER5	No	Yes	Yes

Table A. 2 Samples for assessing PCR amplification success and tissue damage during extraction

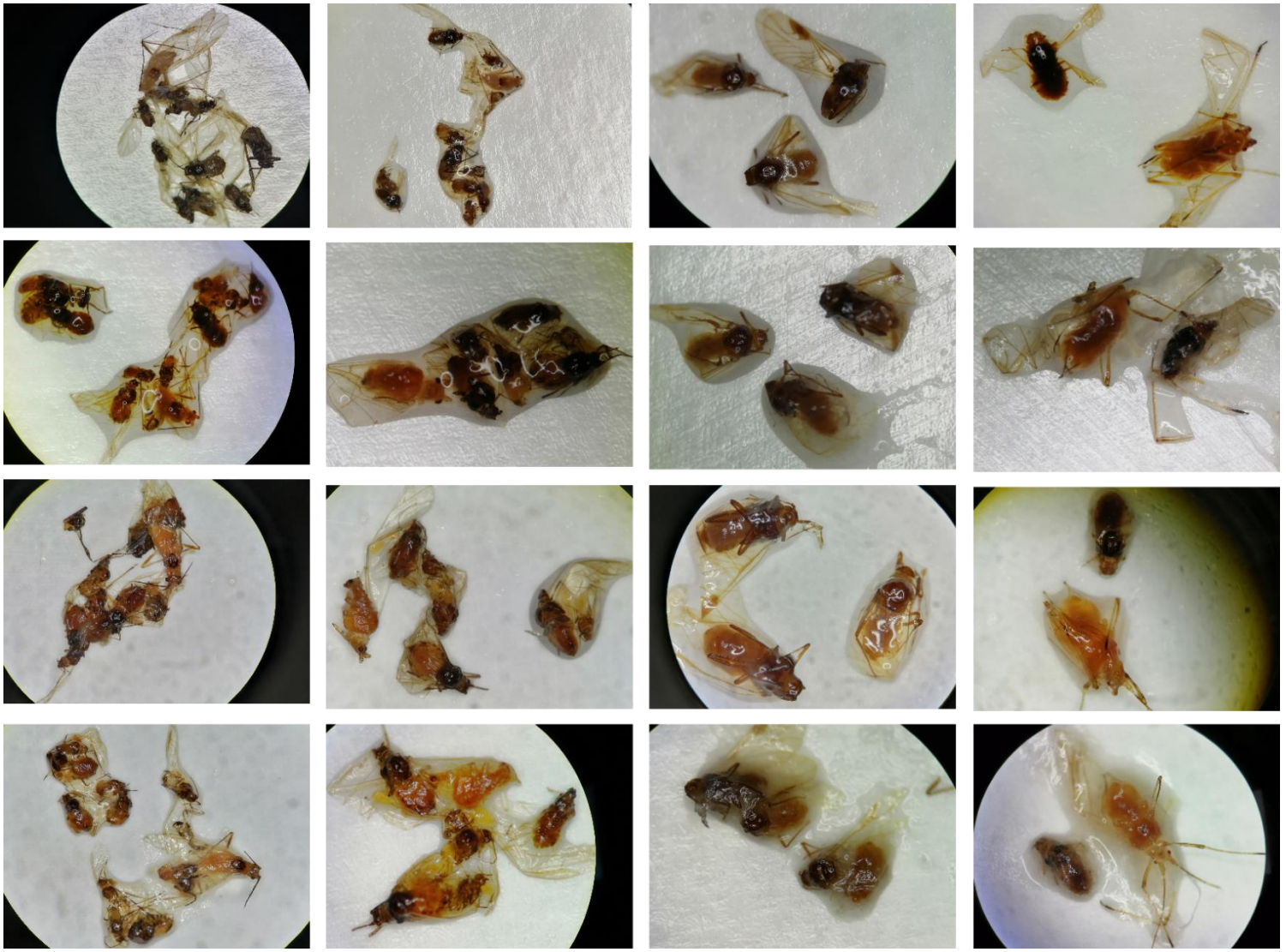
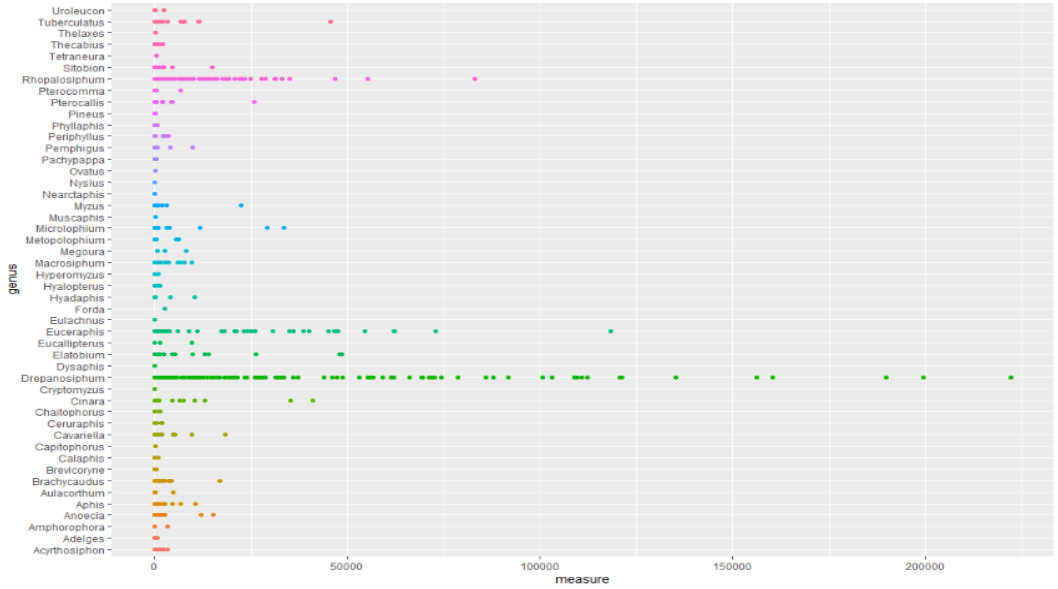


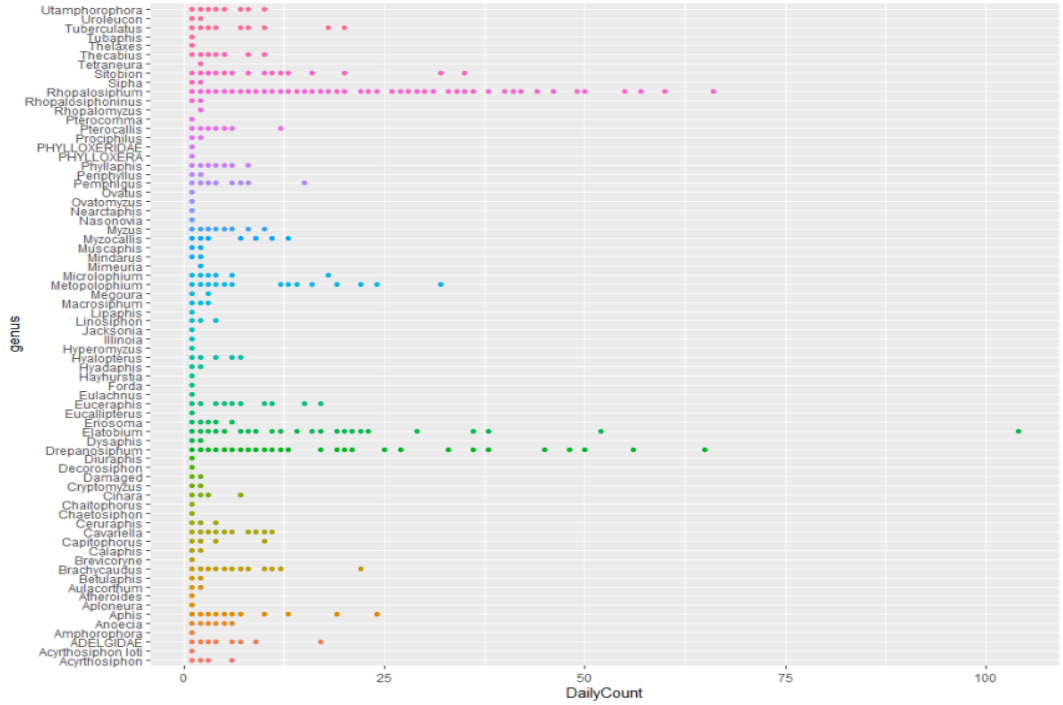
Figure A. 1 Photographs of four samples during the DNA-extraction, Samples before extraction first row, 1h after the lysis second row, 2h after the digestion third row, 6 h after the digestion.

A


genus
Uroleucon
Tuberculatus
Thelaxes
Thecabius
Tetraneura
Sitobion
Rhopalosiphum
Pterocomma
Pterocallis
Pineus
Phyllaphis
Periphyllus
Pemphigus
Pachypappa
Ovatus
Nysius
Nearctaphis
Myzus
Muscaphis
Microlophium
Metopolophium
Megoura
Macrosiphum
Hyperomyzus
Hyalopterus
Hyadaphis
Forda
Eulachnus
Euceraphis
Eucallipterus
Elatobium
Dysaphis
Drepanosiphum
Cryptomyzus
Cinara
Chaitophorus
Ceruraphis
Cavariella
Capitophorus
Calaphis
Brevicoryne
Brachycaudus
Aulacorthum
Aphis
Anoecia
Amphorophora
Adelges
Acyrthosiphon
0
50000
100000
150000
200000
measure


genus
Acyrthosiphon
Adelges
Amphorophora
Anoecia
Aphis
Aulacorthum
Brachycaudus
Brevicoryne
Calaphis
Capitophorus
Cavariella
Ceruraphis
Chaitophorus
Cinara
Cryptomyzus
Drepanosiphum
Dysaphis
Elatobium
Eucallipterus
Euceraphis
Eulachnus
Forda
Hyadaphis
Hyalopterus
Hyperomyzus
Macrosiphum
Megoura
Metopolophium
Microlophium
Muscaphis
Myzus
Nearctaphis
Nysius
Ovatus
Pachypappa
Pemphigus
Periphyllus
Phyllaphis
Pineus
Pterocallis
Pterocomma
Rhopalosiphum
Sitobion
Tetraneura
Thecabius
Thelaxes
Tuberculatus
Uroleucon

B


genus
Utamphorophora
Uroleucon
Tuberculatus
Tubaphis
Thelaxes
Thecabius
Tetraneura
Sitobion
Sipha
Rhopalosiphum
Rhopalosiphoninus
Rhopalomyzus
Pterocomma
Pterocallis
Prociphilus
PHYLLOXERIDAE
PHYLLOXERA
Phyllaphis
Periphyllus
Pemphigus
Ovatus
Ovatomyzus
Nearctaphis
Nasonovia
Myzus
Myzocallis
Muscaphis
Mindarus
Mimeuria
Microlophium
Metopolophium
Megoura
Macrosiphum
Lipaphis
Linosiphon
Jacksonia
Illinoia
Hyperomyzus
Hyalopterus
Hyadaphis
Hayhurstia
Forda
Eulachnus
Euceraphis
Eucallipterus
Eriosoma
Elatobium
Dysaphis
Drepanosiphum
Diuraphis
Decorosiphon
Damaged
Cryptomyzus
Cinara
Chaitophorus
Chaetosiphon
Ceruraphis
Cavariella
Capitophorus
Calaphis
Brevicoryne
Brachycaudus
Betulaphis
Aulacorthum
Atheroides
Aploneura
Aphis
Anoecia
Amphorophora
ADELGIDAE
Acyrthosiphon loti
Acyrthosiphon
0
25
50
75
100
DailyCount


genus
Acyrthosiphon
Acyrthosiphon loti
ADELGIDAE
Amphorophora
Anoecia
Aphis
Aploneura
Atheroides
Aulacorthum
Betulaphis
Brachycaudus
Brevicoryne
Calaphis
Capitophorus
Cavariella
Ceruraphis
Chaetosiphon
Chaitophorus
Cinara
Cryptomyzus
Damaged
Decorosiphon
Diuraphis
Drepanosiphum
Dysaphis
Elatobium
Eriosoma
Eucallipterus
Euceraphis
Eulachnus
Forda
Hayhurstia
Hyadaphis
Hyalopterus
Hyperomyzus
Illinoia
Jacksonia
Linosiphon
Lipaphis
Macrosiphum
Megoura
Metopolophium
Microlophium
Mimeuria
Mindarus
Muscaphis
Myzocallis
Myzus
Nasonovia
Nearctaphis
Ovatomyzus
Ovatus
Pemphigus
Periphyllus
Phyllaphis
PHYLLOXERA
PHYLLOXERIDAE
Prociphilus
Pterocallis
Pterocomma
Rhopalomyzus
Rhopalosiphoninus
Rhopalosiphum
Sipha
Sitobion
Tetraneura
Thecabius
Thelaxes
Tubaphis
Tuberculatus
Uroleucon
Utamphorophora

Figure A. 2 Sequence reads (A) and counts (B) for the major species identified by metabarcoding and morphology respectively.

GENUS	READS	PERCENTAGE
CAVARIELLA	59950	0.007
CINARA	140059	0.016
DREPANOSIPHUM	4722348	0.553
ELATOBIUM	220302	0.026
EUCERAPHIS	1021972	0.12
MACROSIPHUM	55590	0.007
MICROLOPHIUM	177019	0.021
RHOPALOSIPHUM	1158691	0.136
SITOBION	199147	0.023
TUBERCULATUS	200492	0.023

Table A. 3 Reads and total relative percentage for taxa with >0.02 (Metabarcoding dataset)

GENUS	TOTAL	PERCENTAGE
APHIS	211	0.031
BRACHYCAUDUS	191	0.028
CAVARIELLA	164	0.024
DREPANOSIPHUM	1015	0.147
ELATOBIUM	543	0.079
EUCERAPHIS	150	0.022
METOPOLOPHIUM	148	0.021
RHOPALOSIPHUM	3004	0.434
SITOBION	165	0.024
TUBERCULATUS	169	0.024

Table A. 4 Counts and total relative percentage for taxa with >0.02 (Morphological dataset)

Appendix B: Supplementary information for Chapter 3

B.1 Collection of samples:

Collected dates can be found on Table B.1. Certain samples were either used for a different project or were not found within the archive. Such dates were substituted by the dates found in Different_Date column in Table B.1.

Collection date	Collected	Different_date
06/05/2003	yes	
22/05/2003	yes	
02/06/2003	yes	
19/06/2003	yes	
02/07/2003	yes	
05/08/2003	yes	
16/08/2003	yes	
10/09/2003	yes	
24/09/2003	yes	
17/10/2003	yes	
26/10/2003	yes	
10/05/2004	yes	
18/05/2004	yes	
06/06/2004	yes	
29/06/2004	yes	
06/07/2004	no	09/07/2004
17/07/2004	yes	
03/08/2004	no?	02/08/2004
10/08/2004	yes	
12/09/2004	no	13/09/2004
20/09/2004	no	21/09/2004
03/10/2004	yes	
27/10/2004	yes	
12/05/2005	yes	
31/05/2005	yes	
11/06/2005	yes	
30/06/2005	yes	
22/07/2005	yes	
26/07/2005	yes	
18/08/2005	yes	
29/08/2005	yes	
08/09/2005	yes	
10/09/2005	yes	
08/10/2005	yes	
31/10/2005	yes	
12/05/2006	yes	
29/05/2006	yes	
16/07/2006	yes	29/06/2006

26/07/2006	yes	
08/08/2006	yes	
26/08/2006	yes	
06/09/2006	yes	
20/09/2006	yes	
17/10/2006	yes	
29/10/2006	yes	
18/05/2007	yes	
27/05/2007	yes	
06/06/2007	yes	
23/06/2007	yes	
14/07/2007	yes	
19/07/2007	yes	
12/08/2007	yes	
28/08/2007	yes	
05/09/2007	yes	
25/09/2007	yes	
21/10/2007	yes	
30/10/2007	yes	
19/05/2008	yes	
29/05/2008	yes	
09/06/2008	yes	
30/06/2008	yes	
03/07/2008	yes	
29/07/2008	yes	
15/08/2008	yes	
21/08/2008	yes	
17/09/2008	yes	
24/09/2008	yes	
06/10/2008	yes	
22/10/2008	yes	
16/05/2009	yes	
26/05/2009	yes	
16/06/2009	yes	
18/06/2009	yes	
20/07/2009	yes	
31/07/2009	yes	
11/08/2009	yes	
27/08/2009	yes	
15/09/2009	yes	
19/09/2009	yes	
19/10/2009	yes	
22/10/2009	yes	
15/05/2010	no	16/05/2010
27/05/2010	no	26/05/2010
16/06/2010	no	17/06/2010
23/06/2010	yes	
12/07/2010	yes	

16/07/2010	yes	
01/08/2010	yes	
23/08/2010	no	24/08/2010
09/09/2010	yes	
25/09/2010	yes	
08/10/2010	yes	
31/10/2010	yes	
15/05/2011	yes	
28/05/2011	yes	
16/06/2011	yes	
26/06/2011	yes	
14/07/2011	yes	
24/07/2011	yes	
12/08/2011	yes	
23/08/2011	yes	
16/09/2011	yes	
25/09/2011	yes	
14/10/2011	yes	
16/10/2011	yes	
24/05/2012	yes	
28/05/2012	yes	
18/06/2012	yes	
24/06/2012	no	25/06/2012
04/07/2012	yes	
18/07/2012	no	20/07/2012
03/08/2012	no	04/08/2012
24/08/2012	no?	05/08/2012
20/09/2012	no	26/09/2012
24/09/2012	no	25/09/2012
03/10/2012	yes	
16/10/2012	yes	
06/06/2013	yes	
19/06/2013	yes	
22/07/2013	yes	
28/07/2013	yes	
01/08/2013	yes	
14/08/2013	yes	
18/09/2013	yes	
29/09/2013	yes	
03/10/2013	yes	
29/10/2013	yes	
17/05/2014	yes	
23/05/2014	no	22/05/2014
16/06/2014	no	15/06/2014
18/06/2014	yes	
07/07/2014	yes	
22/07/2014	yes	
19/08/2014	yes	

20/08/2014	yes	
06/10/2014	yes	
31/10/2014	yes	
22/05/2015	yes	
31/05/2015	yes	
24/06/2015	yes	
26/06/2015	no?	25/06/2015
11/07/2015	yes	
18/07/2015	yes	
07/08/2015	yes	
23/08/2015	no	24/08/2015
16/09/2015	yes	
18/09/2015	yes	
26/10/2015	yes	
29/10/2015	yes	
23/05/2016	yes	
28/05/2016	yes	
15/06/2016	yes	
17/06/2016	yes	
09/07/2016	yes	
18/07/2016	yes	
19/08/2016	yes	
31/08/2016	yes	
12/09/2016	yes	
20/09/2016	yes	
04/10/2016	yes	
14/10/2016	yes	
15/05/2017	yes	
24/05/2017	yes	
04/06/2017	yes	
25/06/2017	yes	
21/07/2017	yes	
24/07/2017	yes	
09/08/2017	yes	
24/08/2017	yes	
09/09/2017	yes	
20/09/2017	yes	
12/10/2017	yes	
20/10/2017	yes	
16/05/2018	yes	
30/05/2018	yes	
13/06/2018	yes	
22/06/2018	yes	
28/07/2018	yes	
30/07/2018	yes	
13/08/2018	yes	
26/08/2018	yes	
08/09/2018	yes	

	15/09/2018	yes

Table B. 1 Dates collected for bycatch samples that correspond to the dates in Chapter 2. Additional dates for samples not found in the archive are included

B.2 DNA Extraction.

I used a bead-based protocol [protocol #6.3] (Oberacker *et al.*, 2019) with slight modifications on the amount of lysis volume used to adjust it for different sample sizes. In the protocol 62 ul of TNES solution is used with 3 ul of Proteinase K. Here to fully cover the tissues within the samples we had to increase the amount of TNES used, ranging from 0.6ml to 20 ml of TNES solution. Proteinase K was used at a targeted concentration of 100ug/mL for each sample except the controls (see Table B3 bellow).

B.3 Quality filtering in DADA2

As initial quality assessment with FastQC pointed to some libraries having lower quality (assessed by Phred scores). We used three filtering criteria within the DADA2 pipeline which resulted in 3 datasets. The following functions were used to create each dataset by changing the maxEE parameter to relax quality filtering:

```
Dataset_1 <- filterAndTrim(cutFs, filtFs, cutRs, filtRs, maxN = 0, maxEE = c(5, 5),
                truncQ = 2, minLen = 200, rm.phix = TRUE, compress = TRUE, multithread =
TRUE)

Dataset_2 <- filterAndTrim(cutFs, filtFs, cutRs, filtRs, maxN = 0, maxEE = c(3, 3),
                truncQ = 2, minLen = 200, rm.phix = TRUE, compress = TRUE, multithread =
TRUE)

Dataset_3 <- filterAndTrim(cutFs, filtFs, cutRs, filtRs, maxN = 0, maxEE = c(2, 2),
                truncQ = 2, minLen = 200, rm.phix = TRUE, compress = TRUE, multithread =
TRUE)
```

Coefficient	**Df**	**Sum of Squares**	**Mean Sq**	**F-value (Pr>F)**
DADA2_Group	2	25962	12846	1.006 (0.36)
Residuals	549	70081333	12765	

Table B. 2 Anova for the three datasets on the number of ASV's

Sample	input	filtered	denoisedF	denoisedR	merged	nonchim	Lysis solution(ml)	Proteinase K
00B001	1677676	733402	730421	730747	636622	609116	3	15
00B002	1248858	790079	786634	788104	685518	620945	3	15
00B003	1198534	774246	770088	771085	640445	624044	4	20
00B004	1131308	685628	683982	683476	593133	588511	3	15
00B005	861537	510039	508932	508472	443524	403495	3	15
00B006	1272279	856599	852698	854063	768377	599435	3	15
00B007	1287422	770669	764539	767790	692511	648231	4	20
00B008	1169921	768466	767552	766973	662877	591456	1	5
00B009	1502820	1043524	1037191	1038909	990586	835023	3	15
00B010	1519338	1033818	1030364	1031840	1012943	925700	3	15
00B011	1673998	1278760	1273532	1276471	1226440	1133758	4	20
00B012	1588375	1119691	1116423	1116111	1078291	895738	4	20
00B013	1328309	987721	983636	985281	949693	822128	3	15
00B014	914879	607274	603547	604490	581749	546258	3	15
00B015	1205201	723245	717439	719536	674598	622871	3	15
00B016	1299085	901644	897849	899235	861916	791622	3	15
00B017	2235205	778186	766924	775029	410778	353258	3	15
00B018	1837876	693242	686226	690470	358805	293950	3	15
00B019	1736334	706132	692155	702509	365423	296977	4	20
00B020	1997683	763812	753938	757862	380023	291579	3	15
00B021	2642281	971480	967854	957527	544900	447778	3	15
00B022	1911542	810812	808551	809384	534359	483007	3	15
00B023	1197940	368840	365839	366725	311073	239519	3	15
00B024	1560592	685876	680683	683979	560698	411699	3	15
00B025	3231207	2213678	2205831	2209017	2134808	1764026	4	20
00B026	3169867	2198670	2187338	2190482	2111146	1841125	3	15

00B027	3348317	2541964	2529685	2534843	2387989	2161538	20	100
00B028	3623297	2530898	2521296	2523486	2426690	2093856	4	20
00B029	3076449	2067973	2062466	2063068	1977982	1700477	3	15
00B030	2773608	1963919	1955556	1954057	1852355	1502334	4	20
00B031	3094850	2004110	1986750	1993966	1894386	1599178	4	20
00B032	3135651	2336724	2329741	2331685	2238405	1937942	4	20
00B033	3387890	2026775	2020770	2021091	1943835	1590341	4	20
00B034	2514313	1781181	1774737	1775763	1703709	1361609	4	20
00B035	2372812	1709993	1696738	1699436	1613004	1346741	20	100
00B036	2911575	1968914	1962611	1963240	1898900	1577689	20	100
00B037	2548463	1850950	1841920	1844947	1772982	1533184	4	20
00B038	2510330	1756208	1746671	1748880	1667652	1405916	4	20
00B039	2944665	1725068	1719499	1721748	1624850	1452545	4	20
00B040	2899079	2151801	2143952	2147029	2053739	1766232	3	15
00B041	2954924	1477633	1469855	1472099	1358986	1222377	4	20
00B042	1491105	963942	955495	959509	805092	679503	4	20
00B043	1690530	1217191	1197517	1205507	1062344	784719	20	100
00B044	2344566	1626133	1621186	1621131	1556775	1323375	4	20
00B045	1686526	1060958	1056428	1057588	1005404	948434	4	20
00B046	1066469	700936	697089	697389	659273	597250	4	20
00B047	2359285	1410837	1407784	1406364	1359346	1276759	4	20
00B048	1667937	930521	928052	928361	884888	805674	1	5
00B049	4085916	1775961	1767525	1768061	1675049	1379117	4	20
00B050	2267523	1676658	1671857	1673149	1615695	1272150	4	20
00B051	978243	658328	656519	656744	607290	586657	3	15
00B052	2655721	1866977	1864106	1862573	1773834	1634825	4	20
00B053	1960998	1282021	1277413	1277978	1203979	1076304	4	20
00B054	2839170	2102400	2097319	2095806	2019682	1716355	4	20
00B055	2733102	1912622	1907511	1908997	1846871	1626641	4	20
00B056	2488723	1816650	1809918	1812084	1738154	1554073	4	20
00B057	2823663	1693107	1688191	1688479	1620414	1414591	3	15

00B058	1602401	800209	798249	798291	766749	745836	3	15

00B059	2276097	1664512	1659209	1659142	1593879	1256615	4	20
00B060	2594782	1760756	1757522	1756202	1706325	1346465	4	20
00B061	2730209	1581771	1577036	1579137	1507039	1348247	3	15
00B062	2062832	1463025	1457327	1458646	1393072	1112663	4	20
00B063	2602116	1740708	1736599	1736462	1683012	1398638	4	20
00B064	2136020	1498875	1491213	1490879	1398784	1095908	10	50
00B065	1745935	1043928	1042133	1040709	1015736	889739	3	15
00B066	1394336	1003898	999238	1001158	956416	863339	4	20
00B067	1590868	1248443	1243307	1246704	1201456	1074522	4	20
00B068	1513448	1002671	999474	1000193	963688	878504	4	20
00B069	1534171	1131791	1128878	1129714	1086392	997286	4	20
00B070	1524886	1164222	1159526	1161356	1110151	1014587	3	15
00B071	1446661	894079	890696	889176	858848	822347	4	20
00B072	1358715	1037886	1033943	1035000	1007624	855090	4	20
00B073	1643942	964581	961198	962475	918980	847550	4	20
00B074	1276959	877587	874766	875249	816306	675770	4	20
00B075	1281739	946716	944010	945105	914662	811015	4	20
00B076	959951	656338	654608	654572	637350	633017	4	20
00B077	1680559	1237860	1234277	1228755	1192069	990087	4	20
00B078	1682910	1199248	1195694	1196360	1160379	1012032	4	20
00B079	1172168	802880	800571	799380	758811	668694	4	20
00B080	1304158	861625	858288	859498	777460	654563	4	20
00B081	2298209	1311533	1309787	1308157	1261625	1157747	3	15
00B082	1439643	1064894	1060863	1060510	1019900	828115	4	20
00B083	1366245	1004525	1000843	1001159	962742	843061	5	25
00B084	1105480	785400	781201	781963	746492	643741	5	25
00B085	899957	690850	686902	688550	660279	558557	5	25
00B086	1648956	1285889	1282339	1282993	1244586	1112266	4	20
00B087	1528449	968359	964300	966055	915693	856135	4	20
00B088	1372173	965682	959395	960828	898718	823223	4	20
00B089	571593	358583	358231	357949	329538	328162	1	5

00B090	349239	228539	228080	228198	224902	224664	1	5
00B091	424610	302218	299780	300557	290136	276312	4	20
00B092	307784	184975	183896	184322	180425	179357	4	20
00B093	2025375	1512127	1505949	1506266	1431461	1232234	4	20
00B094	1020898	751005	742691	745635	704005	612280	5	25
00B095	1449257	935636	929814	931465	865730	733918	5	25
00B096	867925	605185	599992	601812	557417	517645	4	20
00B097	1587731	800281	790550	796295	505847	470950	4	20
00B098	1582696	776135	772739	772686	456956	413181	3	15
00B099	1497168	857167	847490	853791	594362	476616	4	20
00B100	1552300	854027	848137	850995	631257	522173	4	20
00B101	1052197	446938	444957	445305	253941	250554	1	5
00B102	1388674	683326	679958	682131	435883	424637	1	5
00B103	1397457	261148	260336	260667	170330	164678	2	10
00B104	1473579	765164	761493	764133	496314	422886	4	20
00B105	2141096	1124247	1121350	1121848	1077907	1046540	4	20
00B106	1302974	870649	865946	867359	835773	778744	4	20
00B107	1442924	1056019	1053187	1053147	1024979	963722	4	20
00B108	1240888	795202	792321	792281	754427	700403	10	50
00B109	1502425	968590	963284	965096	922701	775311	4	20
00B110	1351049	961210	958889	958696	923983	900817	4	20
00B111	1561313	972556	966482	968024	932076	851550	4	20
00B112	1456492	1000391	994701	996963	951860	920047	4	20
00B113	2710119	1609959	1605478	1604716	1439114	1305882	4	20
00B114	1786936	1141936	1140844	1140654	1062431	989543	4	20
00B115	1428293	935396	931108	931802	835826	782000	4	20
00B116	1929314	1256081	1253800	1250782	1136863	1101963	5	25
00B117	1944772	1202866	1192448	1198134	1056083	916039	5	25
00B118	1611159	1047891	1043781	1045580	944734	888189	5	25
00B119	1729237	913060	907826	909008	861506	824952	4	20
00B120	1749618	1095557	1092455	1093250	1038710	973198	4	20

00B121	2099592	1188447	1185735	1186531	1155070	1128979	4	20
00B122	1408276	967437	961667	962935	913310	816161	4	20
00B123	1893508	1457565	1453223	1454888	1411800	1191829	4	20
00B124	1580102	1110607	1104618	1105292	1062326	939525	10	50
00B125	1318258	942092	937039	937152	896681	814991	4	20
00B126	1856241	1386331	1382096	1382891	1345468	1186380	4	20
00B127	2205255	1575499	1571344	1571093	1531837	1432191	4	20
00B128	2065201	1250716	1247918	1248382	1199664	1153705	4	20
00B129	1490563	1068520	1064820	1066267	1027021	907947	4	20
00B130	1259576	844488	841326	840865	812172	734034	4	20
00B131	1476544	1116719	1114197	1113925	1081120	956637	4	20
00B132	1521797	1013915	1010061	1011820	956757	937641	4	20
00B133	1542799	1113953	1111084	1113083	1033462	985863	3	15
00B134	1234189	816825	815404	815180	795160	781935	4	20
00B135	1224169	743205	740864	741218	713028	646235	4	20
00B136	1448041	905898	903688	903897	874038	835184	4	20
00B137	3040519	1674624	1667173	1667360	1562152	1365417	4	20
00B138	18406	1928	1729	1753	1627	1627	1	5
00B139	1488809	1110923	1107999	1109023	1059663	996578	4	20
00B140	1590959	1036161	1032156	1033453	977869	926897	5	25
00B141	1360912	938151	934752	936400	892045	837601	4	20
00B142	1788544	1305262	1297889	1300194	1223161	1014589	5	25
00B143	1496269	981233	972404	975900	906039	776983	5	25
00B144	1768025	1180807	1174675	1175872	1093260	974499	5	25
00B145	929041	118215	117924	118122	24303	22253	0.1	0.5
00B146	1410202	221248	219566	220709	97087	90784	4	20
00B147	1570976	257737	255609	256663	183890	156325	4	20
00B148	52170	2707	2377	2570	2336	2336	4	20
00B149	487087	28946	28552	28862	25442	22811	3	15
00B150	1315178	282964	282140	282265	200504	178126	1	5
00B151	1185736	222517	220548	219859	166455	131373	3	15

00B152	1297558	249920	246830	248997	172637	137689	4	20
00B153	1539394	938913	934787	935969	565726	522791	4	20
00B154	884747	451600	445947	449224	337712	292631	4	20
00B155	46844	14845	14490	14692	13360	12918	1	5
00B156	1538830	673856	662884	669542	599058	548878	10	50
00B157	1302200	804419	800232	801316	666538	572558	4	20
00B158	1136414	641182	636499	630218	545854	489633	4	20
00B159	1013461	561608	560325	560632	472975	403893	4	20
00B160	605946	170194	169187	169688	148869	144445	4	20
00B161	2523560	761328	751033	756426	520953	464560	4	20
00B162	1480325	499694	493857	497325	281978	251331	3	15
00B163	1352345	625601	622445	623783	411177	395626	4	20
00B164	1503550	336241	329914	334392	132788	115914	4	20
00B165	1862627	645671	634495	642796	406889	378518	4	20
00B166	1379698	559525	554850	557281	353854	328010	4	20
00B167	1745749	448912	445215	447216	309425	287176	4	20
00B168	503457	109912	109738	109751	55040	54201	0.1	0.5
00B169	1560577	527744	526685	526191	512757	509984	3	15
00B170	837074	551640	547467	548874	519758	459014	20	100
00B171	984931	683401	679733	680265	648089	590118	20	100
00B172	609787	403319	402437	402126	393614	392849	4	20
00B173	972251	673139	670505	670973	642872	614275	3	15
00B174	43695	21388	21245	21249	21091	21083	3	15
00B175	790026	503172	500320	500585	484066	468398	4	20
00B176	959012	614297	612316	612729	567957	500548	3	15
00B177	1802128	1060042	1054459	1057150	842497	807380	3	15
00B178	681326	487849	486946	486711	472402	471705	3	15
00B179	1340952	984741	978428	980698	897990	839342	4	20
00B180	1159070	844367	837242	839263	790803	702266	10	50
00B181	1358210	984444	980689	981246	940692	814759	3	15
00B182	1094392	838192	835594	834747	796669	749716	4	20

00B183	1482120	1082872	1079013	1080103	1019029	904493	4	20
00B184	977780	640846	637010	639175	595712	552367	4	20
extr_neg10	76	0	1	1	0	0	0.06	3
extr_neg11	31	0	2	2	0	0	0.06	3
extr_neg12	23	0	2	3	0	0	0.06	3
extr_neg13	13	0	1	1	1	1	0.06	3
extr_neg14	281	1	3	3	0	0	0.06	3
extr_neg15	251	0	1	1	0	0	0.06	3
extr_neg16	679	8	15	10	10	10	0.06	3
extr_neg17	78	5	4	4	0	0	0.06	3
extr_neg18	137	0	1	1	0	0	0.06	3
extr_neg19	22	0	1	1	0	0	0.06	3
extr_neg2	195	6	0	0	0	0	0.06	3
extr_neg20	129	3	0	0	0	0	0.06	3
extr_neg21	40	2	0	0	0	0	0.06	3
extr_neg22	15	0	0	0	0	0	0.06	3
extr_neg23	50	0	0	0	0	0	0.06	3
extr_neg3	25	0	0	0	0	0	0.06	3
extr_neg4	123	16	0	0	0	0	0.06	3
extr_neg5	65	4	0	0	0	0	0.06	3
extr_neg6	16	0	0	0	0	0	0.06	3
extr_neg7	56	0	0	0	0	0	0.06	3
extr_neg8	87	1	0	0	0	0	0.06	3
extr_neg9	9	1	0	0	0	0	0.06	3
extr_pos10	70534	5832	6	6	0	0	0.06	3
extr_pos11	52006	4755	340	336	336	336	0.06	3
extr_pos12	28562	2315	1296	1272	1271	1271	0.06	3
extr_pos13	5333	173	209	212	208	208	0.06	3
extr_pos14	2073	271	2416	2474	2409	2409	0.06	3
extr_pos15	76831	3018	7505	7554	5916	5916	0.06	3
extr_pos16	107088	8218	529	541	528	528	0.06	3

extr_pos17	27974	3400	596	598	596	596	0.06	3
extr_pos18	33250	2289	890	893	702	702	0.06	3
extr_pos19	36197	85	795	802	794	794	0.06	3
extr_pos2	18669	1784	1	1	0	0	0.06	3
extr_pos20	20851	999	1	1	0	0	0.06	3
extr_pos21	1517	6	404	401	232	232	0.06	3
extr_pos22	3844	345	108	110	107	107	0.06	3
extr_pos23	21923	1306	783	803	749	749	0.06	3
extr_pos3	41380	220	26	42	0	0	0.06	3
extr_pos4	35392	2496	62	71	0	0	0.06	3
extr_pos5	77796	7568	1717	1732	1587	1587	0.06	3
extr_pos6	6915	543	1994	1997	1636	1636	0.06	3
extr_pos7	6965	610	1353	1357	1148	1148	0.06	3
extr_pos8	15032	904	602	631	331	331	0.06	3
extr_pos9	7779	815	25	2	0	0	0.06	3
pcr_neg1	22	0	0	0	0	0	0.06	3
pcr_neg10	3623	0	0	0	0	0	0.06	3
pcr_neg11	111	0	0	0	0	0	0.06	3
pcr_neg12	185	0	0	0	0	0	0.06	3
pcr_neg13	4	0	0	0	0	0	0.06	3
pcr_neg14	102	0	0	0	0	0	0.06	3
pcr_neg15	1598	0	0	0	0	0	0.06	3
pcr_neg16	3349	0	0	0	0	0	0.06	3
pcr_neg17	87	0	0	0	0	0	0.06	3
pcr_neg18	275	0	0	0	0	0	0.06	3
pcr_neg19	56	1	0	0	0	0	0.06	3
pcr_neg2	20	0	0	0	0	0	0.06	3
pcr_neg20	154	1	0	0	0	0	0.06	3
pcr_neg21	35	0	0	0	0	0	0.06	3
pcr_neg22	70	0	0	0	0	0	0.06	3
pcr_neg23	764	0	0	0	0	0	0.06	3

pcr_neg3	21	0	0	0	0	0	0.06	3
pcr_neg4	315	0	0	0	0	0	0.06	3
pcr_neg5	86	0	0	0	0	0	0.06	3
pcr_neg6	29	0	0	0	0	0	0.06	3
pcr_neg7	40	0	0	0	0	0	0.06	3
pcr_neg8	234	0	0	0	0	0	0.06	3
pcr_neg9	12	0	0	0	0	0	0.06	3
pcr_pos10	8442	411	0	0	0	0	0.06	3
pcr_pos11	2041	114	0	0	0	0	0.06	3
pcr_pos12	10743	807	0	0	0	0	0.06	3
pcr_pos13	3681	46	0	0	0	0	0.06	3
pcr_pos14	808	75	0	0	0	0	0.06	3
pcr_pos15	71416	1750	1200	988	555	381	0.06	3
pcr_pos16	40636	2042	2000	1843	1672	1672	0.06	3
pcr_pos17	15065	1389	0	0	0	0	0.06	3
pcr_pos18	18123	636	0	0	0	0	0.06	3
pcr_pos19	22323	41	0	0	0	0	0.06	3
pcr_pos2	2878	200	0	0	0	0	0.06	3
pcr_pos20	3907	67	0	0	0	0	0.06	3
pcr_pos21	1074	4	0	0	0	0	0.06	3
pcr_pos22	3255	240	0	0	0	0	0.06	3
pcr_pos23	5013	155	0	0	0	0	0.06	3
pcr_pos4	27413	1427	0	0	0	0	0.06	3
pcr_pos5	9274	597	0	0	0	0	0.06	3
pcr_pos6	1958	93	0	0	0	0	0.06	3
pcr_pos7	13585	790	0	0	0	0	0.06	3
pcr_pos8	13365	493	0	0	0	0	0.06	3
pcr_pos9	1887	171	0	0	0	0	0.06	3

Table B. 3 Sample information with reads entering and exiting the DADA2 pipeline. Lysis solution volume used and corresponding Proteinase K amount

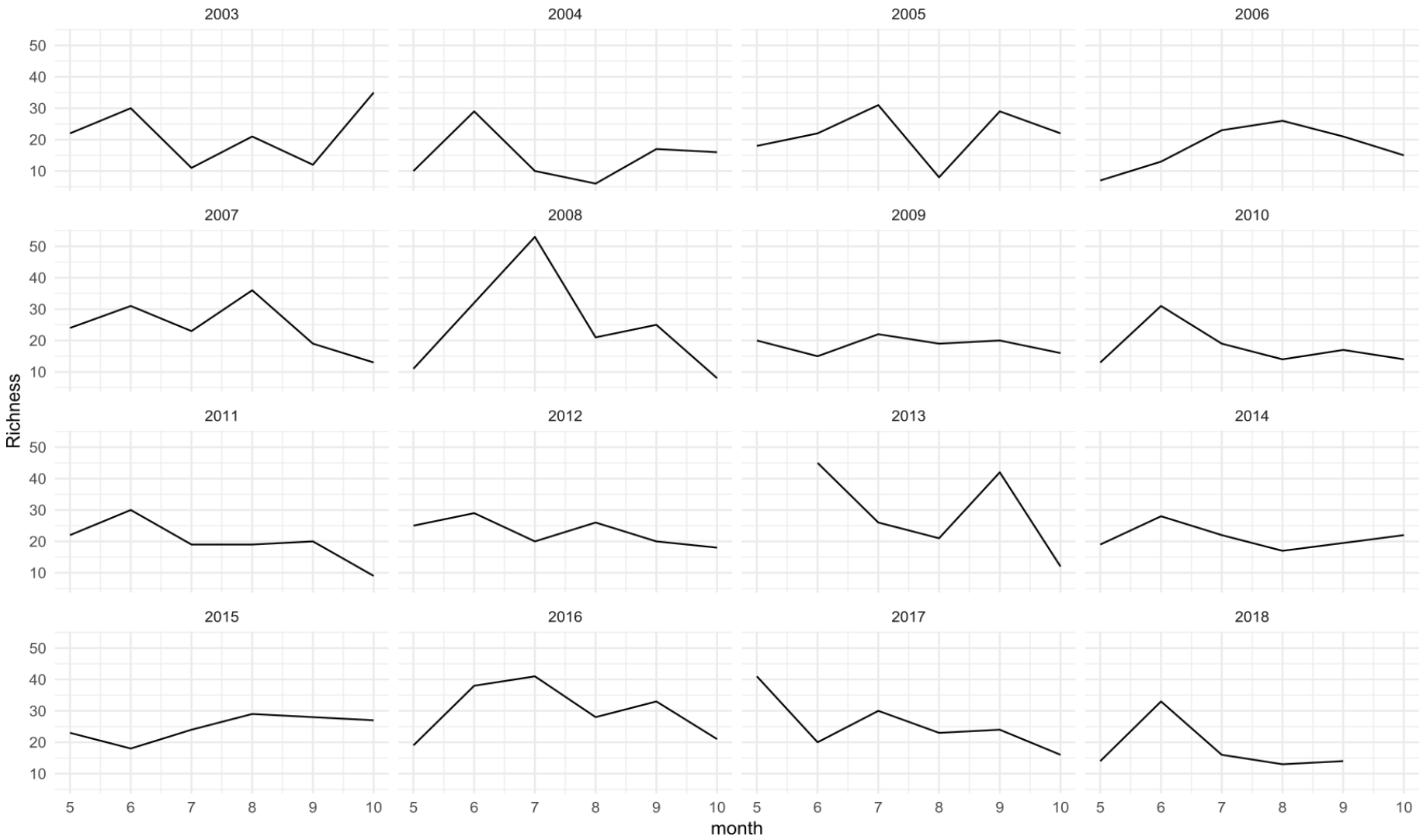


Figure B. 1 Species richness throughout the years of the time series for each month

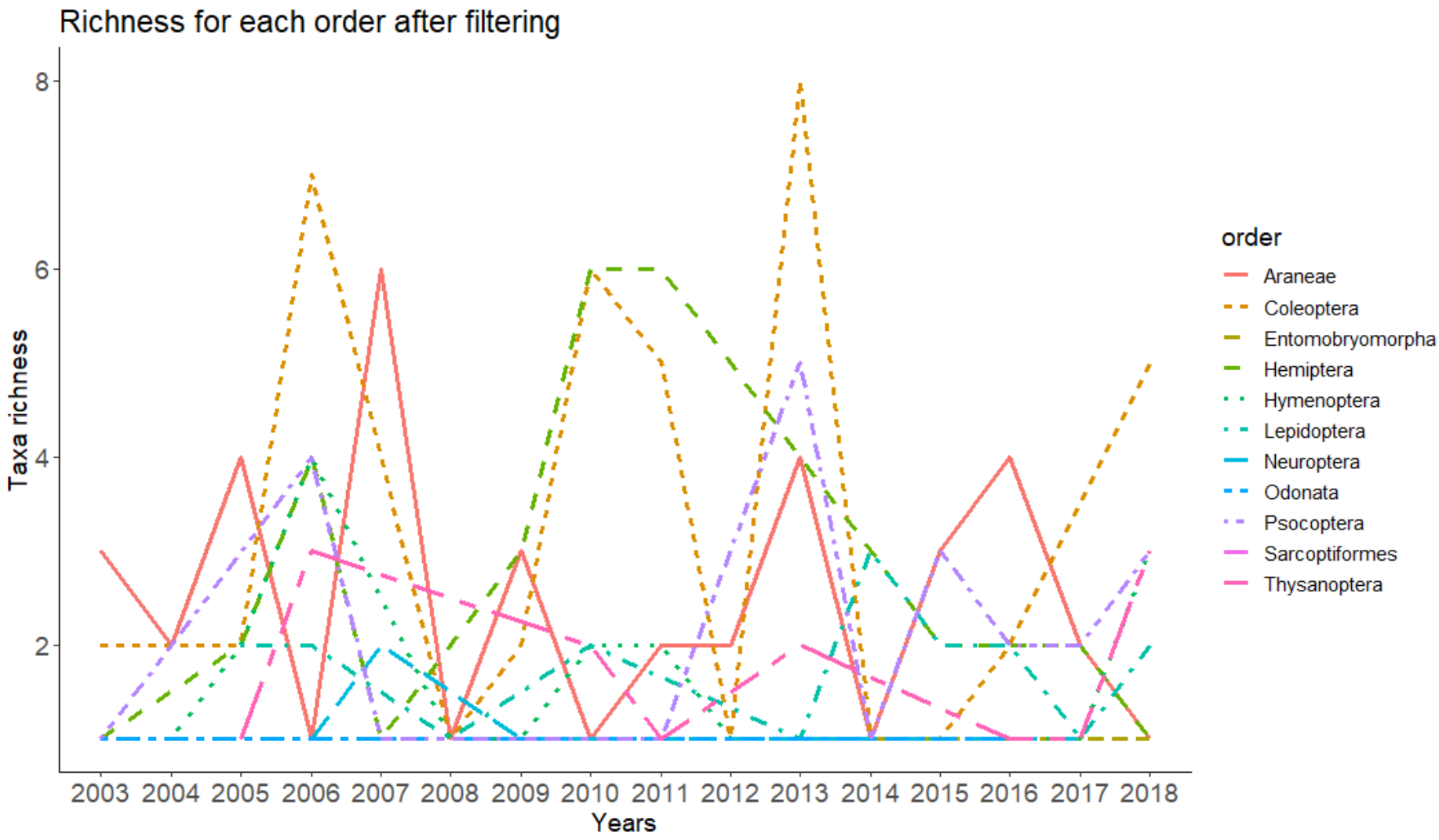


Figure B. 2 Richness for the major orders after using sample filtering thresholds at 1%

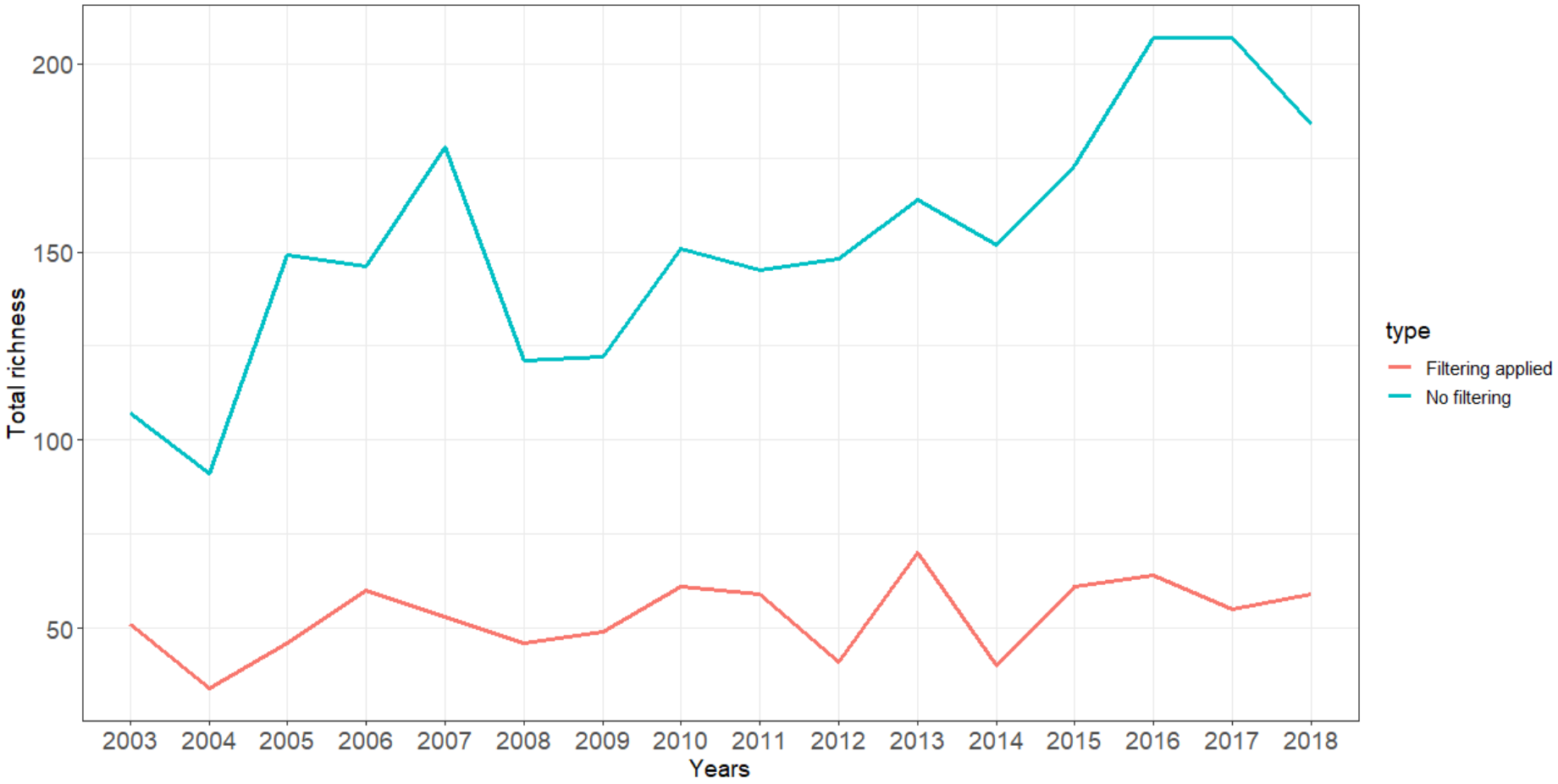


Figure B. 3 Richness over time before and after applying filtering with a 1% threshold the entire dataset.

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
species	Category	genus	family	order	class	phylum	kingdom	superkingdom
erigone atra	Predator	Erigone	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
tenuiphantes tenuis	Predator	Tenuiphantes	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
tachyporus hypnorum	Predator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
aphidius rhopalosiphi	Parasitoid	Aphidius	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
chrysoperla carnea	Predator	Chrysoperla	Chrysopidae	Neuroptera	Insecta	Arthropoda	Metazoa	Eukaryota
oedothorax apicatus	Predator	Oedothorax	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
anotylus rugosus	Predator	Anotylus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
rhopalosiphum padi	Pest	Rhopalosiphum	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
porrhomma pygmaeum	Predator	Porrhomma	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
aphidius avenae	Parasitoid	Aphidius	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
tachyporus obtusus	Predator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
trechus quadristriatus	Predator	Trechus	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
aphidius ervi	Parasitoid	Aphidius	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
aloconota gregaria	Predator	Aloconota	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
praon volucre	Parasitoid	Praon	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
erigone dentipalpis	Predator	Erigone	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
atheta celata	Predator	Atheta	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
anotylus tetracarinatus	Predator	Anotylus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
centromerita bicolor	Predator	Centromerita	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
tachyporus chrysomelinus	Predator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
microlinyphia pusilla	Predator	Microlinyphia	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
bembidion aeneum	Predator	Bembidion	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
amischa analis	Predator	Amischa	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota

porrhomma microphthalmum	Predator	Porrhomma	Linyphiidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
aphis fabae	Pest	Aphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
chromaphis juglandicola	Pest	Chromaphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
diaeretiella rapae	Parasitoid	Diaeretiella	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
macrosiphum euphorbiae	Pest	Macrosiphum	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
amischa nigrofusca	Predator	Amischa	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
bembidion guttula	Predator	Bembidion	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
amara similata	Predator	Amara	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
rhyzobius litura	Predator	Rhyzobius	Coccinellidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
anotylus sculpturatus	Predator	Anotylus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
pterostichus strenuus	Predator	Pterostichus	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
aphis gossypii	Pest	Aphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips tabaci	Pest	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
philonthus cognatus	Predator	Philonthus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
nebria brevicollis	Predator	Nebria	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
larinioides cornutus	Predator	Larinioides	Araneidae	Araneae	Arachnida	Arthropoda	Metazoa	Eukaryota
sitobion avenae	Pest	Sitobion	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
anchomenus dorsalis	Predator	Anchomenus	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
myzus cerasi	Pest	Myzus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
acyrthosiphon pisum	Pest	Acyrthosiphon	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
dilophus febrilis	Pest	Dilophus	Bibionidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
sitodiplosis mosellana	Pest	Sitodiplosis	Cecidomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
trioza urticae	Grass	Trioza	Triozidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
trypeta zoe	Urtica	Trypeta	Tephritidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
hydrellia maura	Grass	Hydrellia	Ephydridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips flavus	Polyphagus	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
agrotis exclamationis	Pest	Agrotis	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota

geomyza tripunctata	Pest	Geomyza	Opomyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
agrochola litura	Pest	Agrochola	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
amphipyra tragopoginis	Pest	Amphipyra	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaptomyza flava	Garden	Scaptomyza	Drosophilidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
trioza remota	Pest	Trioza	Triozidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
chlorops pumilionis	Pest	Chlorops	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
amphipyra pyramidea	Pest	Amphipyra	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
apamea lithoxylaea	Grass	Apamea	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips major	Pest	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaptomyza graminum	Garden	Scaptomyza	Drosophilidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
delia platura	Pest	Delia	Anthomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
mayetiola destructor	Pest	Mayetiola	Cecidomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
frankliniella tenuicornis	Pest	Frankliniella	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
hydrellia griseola	Pest	Hydrellia	Ephydridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
cerodontha denticornis	Pest	Cerodontha	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
chromatomyia milii	Grass	Chromatomyia	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
contarinia tritici	Pest	Contarinia	Cecidomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
bibio marci	Pest	Bibio	Bibionidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
autographa gamma	Pest	Autographa	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
blastobasis adustella	Pest	Blastobasis	Blastobasidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
noctua pronuba	Pest	Noctua	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
macrosteles sexnotatus	Garden	Macrosteles	Cicadellidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
eupteryx atropunctata	Pest	Eupteryx	Cicadellidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
chromatomyia horticola	Pest	Chromatomyia	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
muellerianella extrusa	Grass	Muellerianella	Delphacidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
conomelus anceps	Garden	Conomelus	Delphacidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
oxythrips ajugae	Pest	Oxythrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
pinalitus cervinus	Pest	Pinalitus	Miridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota

neuroterus quercusbaccarum	Pest	Neuroterus	Cynipidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
phyllonorycter maestingella	Pest	Phyllonorycter	Gracillariidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
oscinella maura	Garden	Oscinella	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
hyalopterus pruni	Pest	Hyalopterus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
apamea monoglypha	Garden	Apamea	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
phlogophora meticulosa	Pest	Phlogophora	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
athalia cordata	Garden	Athalia	Tenthredinidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
euceraphis betulae	Pest	Euceraphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
macrosteles ossiannilssoni	Grass	Macrosteles	Cicadellidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
alnetoidia alneti	Pest	Alnetoidia	Cicadellidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
pandemis heparana	Garden	Pandemis	Tortricidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
frankliniella intonsa	Garden	Frankliniella	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
dicraeus vagans	Garden	Dicraeus	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips minutissimus	Pest	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
elatobium abietinum	Pest	Elatobium	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
yponomeuta evonymella	Pest	Yponomeuta	Yponomeutidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
phyllotreta undulata	Pest	Phyllotreta	Chrysomelidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
macustus grisescens	Garden	Macustus	Cicadellidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
pinalitus viscicola	Garden	Pinalitus	Miridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
trioza galii	Garden	Trioza	Triozidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips angusticeps	Pest	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
pterocallis alni	Pest	Pterocallis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
agrotis ipsilon	Pest	Agrotis	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
phyllaphis fagi	Pest	Phyllaphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
euceraphis punctipennis	Pest	Euceraphis	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota

liriomyza flaveola	Pest	Liriomyza	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
coleophora caespititiella	Garden	Coleophora	Coleophoridae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
contarinia fagi	Pest	Contarinia	Cecidomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
xestia xanthographa	Pest	Xestia	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
cacopsylla pruni	Pest	Cacopsylla	Psyllidae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
philaenus spumarius	Grass	Philaenus	Aphrophoridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
blepharidopterus angulatus	Pest	Blepharidopterus	Miridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
phytomyza rufipes	Pest	Phytomyza	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
sitona lineatus	Pest	Sitona	Curculionidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
periphyllus testudinaceus	Pest	Periphyllus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
phyllonorycter messaniella	Pest	Phyllonorycter	Gracillariidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
eulachnus agilis	Pest	Eulachnus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
eucallipterus tiliae	Pest	Eucallipterus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
dilophus febrilis	Pollinator	Dilophus	Bibionidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaptomyza pallida	Pollinator	Scaptomyza	Drosophilidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
psychoda phalaenoides	Pollinator	Psychoda	Psychodidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
eudasyphora cyanella	Pollinator	Eudasyphora	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
platypalpus agilis	Pollinator	Platypalpus	Hybotidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
lotophila atra	Pollinator	Lotophila	Sphaeroceridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
copromyza equina	Pollinator	Copromyza	Sphaeroceridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
psychoda grisescens	Pollinator	Psychoda	Psychodidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
nephrotoma appendiculata	Pollinator	Nephrotoma	Tipulidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
bibio nigriventris	Pollinator	Bibio	Bibionidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
scathophaga stercoraria	Pollinator	Scathophaga	Scathophagidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
hydrellia maura	Pollinator	Hydrellia	Ephydridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota

tachyporus hypnorum	Pollinator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
agrotis exclamationis	Pollinator	Agrotis	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
bicellaria vana	Pollinator	Bicellaria	Hybotidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
bellardia vulgaris	Pollinator	Bellardia	Calliphoridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
bellardia vulgaris	Pollinator	Bellardia	Orobanchaceae	Lamiales	Magnoliopsida	Streptophyta	Viridiplantae	Eukaryota
geomyza tripunctata	Pollinator	Geomyza	Opomyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
lonchoptera lutea	Pollinator	Lonchoptera	Lonchopteridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
amphipyra tragopoginis	Pollinator	Amphipyra	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
botanophila fugax	Pollinator	Botanophila	Anthomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaptomyza flava	Pollinator	Scaptomyza	Drosophilidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
chlorops pumilionis	Pollinator	Chlorops	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
oscinella nigerrima	Pollinator	Oscinella	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
rhamphomyia sulcata	Pollinator	Rhamphomyia	Empididae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
sepsis cynipsea	Pollinator	Sepsis	Sepsidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
vespula vulgaris	Pollinator	Vespula	Vespidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
thrips major	Pollinator	Thrips	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaptomyza graminum	Pollinator	Scaptomyza	Drosophilidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
anthocoris nemorum	Pollinator	Anthocoris	Anthocoridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
coproica ferruginata	Pollinator	Coproica	Sphaeroceridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
delia platura	Pollinator	Delia	Anthomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
tachyporus obtusus	Pollinator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
chlorops hypostigma	Pollinator	Chlorops	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
hydrellia griseola	Pollinator	Hydrellia	Ephydridae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
syrphophilus tricinctorius	Pollinator	Syrphophilus	Ichneumonidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
cerodontha denticornis	Pollinator	Cerodontha	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
fannia rondanii	Pollinator	Fannia	Fanniidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota

azelia cilipes	Pollinator	Azelia	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
delia florilega	Pollinator	Delia	Anthomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
aphidius ervi	Pollinator	Aphidius	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
chromatomyia milii	Pollinator	Chromatomyia	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
platypalpus longicornis	Pollinator	Platypalpus	Hybotidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
praon volucre	Pollinator	Praon	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
bibio marci	Pollinator	Bibio	Bibionidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
autographa gamma	Pollinator	Autographa	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
noctua pronuba	Pollinator	Noctua	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
helina depuncta	Pollinator	Helina	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
ceutorhynchus obstrictus	Pollinator	Ceutorhynchus	Curculionidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
muscina prolapsa	Pollinator	Muscina	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
thaumatomyia notata	Pollinator	Thaumatomyia	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
tachyporus chrysomelinus	Pollinator	Tachyporus	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
sepsis punctum	Pollinator	Sepsis	Sepsidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
hyalopterus pruni	Pollinator	Hyalopterus	Aphididae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
apamea monoglypha	Pollinator	Apamea	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
eristalis tenax	Pollinator	Eristalis	Syrphidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
phlogophora meticulosa	Pollinator	Phlogophora	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
sepsis flavimana	Pollinator	Sepsis	Sepsidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
tetanocera elata	Pollinator	Tetanocera	Sciomyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
azelia nebulosa	Pollinator	Azelia	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
dacnusa maculipes	Pollinator	Dacnusa	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
anthomyia pluvialis	Pollinator	Anthomyia	Anthomyiidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
vespula germanica	Pollinator	Vespula	Vespidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
mesopolobus incultus	Pollinator	Mesopolobus	Pteromalidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota

frankliniella intonsa	Pollinator	Frankliniella	Thripidae	Thysanoptera	Insecta	Arthropoda	Metazoa	Eukaryota
philorinum sordidum	Pollinator	Philorinum	Staphylinidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
dicraeus vagans	Pollinator	Dicraeus	Chloropidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
diglyphus isaea	Pollinator	Diglyphus	Eulophidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
empis nigripes	Pollinator	Empis	Empididae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
aphidius matricariae	Pollinator	Aphidius	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
diaeretiella rapae	Pollinator	Diaeretiella	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
empis nuntia	Pollinator	Empis	Empididae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
scaeva pyrastri	Pollinator	Scaeva	Syrphidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
amara similata	Pollinator	Amara	Carabidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
cantharis flavilabris	Pollinator	Cantharis	Cantharidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
epuraea melanocephala	Pollinator	Epuraea	Nitidulidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
praon dorsale	Pollinator	Praon	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
agrotis ipsilon	Pollinator	Agrotis	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
bombus lucorum	Pollinator	Bombus	Apidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
chrysocharis pubicornis	Pollinator	Chrysocharis	Eulophidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
sympycnus pulicarius	Pollinator	Sympycnus	Dolichopodidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
colastes braconius	Pollinator	Colastes	Braconidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
liriomyza flaveola	Pollinator	Liriomyza	Agromyzidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
xestia xanthographa	Pollinator	Xestia	Noctuidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
philaenus spumarius	Pollinator	Philaenus	Aphrophoridae	Hemiptera	Insecta	Arthropoda	Metazoa	Eukaryota
themira annulipes	Pollinator	Themira	Sepsidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota

plutella xylostella	Pollinator	Plutella	Plutellidae	Lepidoptera	Insecta	Arthropoda	Metazoa	Eukaryota
sepsis violacea	Pollinator	Sepsis	Sepsidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
muscina levida	Pollinator	Muscina	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
lasius niger	Pollinator	Lasius	Formicidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
sitona lineatus	Pollinator	Sitona	Curculionidae	Coleoptera	Insecta	Arthropoda	Metazoa	Eukaryota
meliscaeva auricollis	Pollinator	Meliscaeva	Syrphidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota
dendrocerus carpenteri	Pollinator	Dendrocerus	Megaspilidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
apis mellifera	Pollinator	Apis	Apidae	Hymenoptera	Insecta	Arthropoda	Metazoa	Eukaryota
neomyia cornicina	Pollinator	Neomyia	Muscidae	Diptera	Insecta	Arthropoda	Metazoa	Eukaryota

Table B. 4 Table with species and their corresponding taxonomy and categories found in the databases

Species	Found in UK
Psychoda satchelli	No
Calopteryx maculata	Database error
Alnetoidia alneti	Yes
Ectopsocus californicus	No
Metriocnemus atriclava	
Diamesa hyperborea	No
Smittia stercoraria	No
Cricotopus relucens	No
Sylvicola stackelbergi	No
Brassicogethes aeneus	Yes (Synonym *Meligethes aeneus*)
Tipula mediterranea	No
Protapanteles fulvipes	No
Chironomus melanescens	No

Table B. 5 Species that were not found in the UK species inventory list

Appendix C: Supplementary information for Chapter 4

	Unique_code	date	Field_ID	month.x	Region	easting	northing
1	000ABF	29/07/2019	3357	7	Scotland (Scottish Borders)	383735	656881
2	0010ABF	04/08/2019	3400	8	Staffordshire	384580	335810
3	0010AF	16/06/2019	3354	6	Scotland (Scottish Borders)	361412	630255
4	0011ABF	04/08/2019	3459	8	Staffordshire	384250	336060
5	0011AF	17/06/2019	3377	6	East Scotland (Aberdeenshire)	396648	839818
6	0012ABF	05/08/2019	3361	8	North Yorkshire	500755	475735
7	0012AF	17/06/2019	3376	6	East Scotland (Angus)	357902	751084
8	0013ABF	05/08/2019	3360	8	East Riding of Yorkshire	512325	468225
9	0013AF	17/06/2019	3324	6	East Riding of Yorkshire	491868	446904
10	0014ABF	05/08/2019	3362	8	North Yorkshire	492645	483625
11	0014AF	17/06/2019	3364	6	East Scotland (Aberdeenshire)	375775	777417
12	0015ABF	06/08/2019	3411	8	Highland	274557	865148
13	0015AF	17/06/2019	3363	6	East Scotland (Aberdeenshire)	385705	777094
14	0016ABF	06/08/2019	3412	8	Highland	276861	865305
15	0016AF	17/06/2019	3365	6	East Scotland (Aberdeenshire)	376458	780926
16	0017ABF	06/08/2019	3414	8	Highland	270245	865049
17	0018ABF	06/08/2019	3363	8	East Scotland (Aberdeenshire)	385705	777094
18	0019ABF	05/08/2019	3365	8	East Scotland (Aberdeenshire)	376458	780926
19	0019AF	17/06/2019	3437	6	East Riding of Yorkshire	487243	457500
20	001ABF	23/07/2019	3426	7	Suffolk	596749	273827
21	001AF	15/06/2019	3430	6	East Scotland (Grampian)	381129	852508
22	0020ABF	05/08/2019	3364	8	East Scotland (Aberdeenshire)	375775	777417

23	0020AF	17/06/2019	3353	6	East Scotland (Aberdeenshire)	370697	833533
24	0021ABF	06/08/2019	3346	8	Cornwall	170095	40955
25	0022ABF	05/08/2019	3372	8	East Scotland (Moray)	319119	861258
26	0022AF	22/06/2019	3402	6	Staffordshire	384010	330900
27	0023ABF	05/08/2019	3449	8	Somerset	303796	134493
28	0023AF	25/06/2019	3405	6	East Riding of Yorkshire	481720	454480
29	0024ABF	05/08/2019	3405	8	East Riding of Yorkshire	481720	454480
30	0024AF	21/06/2019	3378	6	Highland	263655	851851
31	0025ABF	05/08/2019	3406	8	East Riding of Yorkshire	492170	455320
32	0026ABF	02/08/2019	3378	8	Highland	263655	851851
33	0026AF	24/06/2019	3360	6	East Riding of Yorkshire	512325	468225
34	0027ABF	02/08/2019	3379	8	Highland	267943	866183
35	0027AF	24/06/2019	3335	6	East Scotland (Moray)	335952	860944
36	0028AF	24/06/2019	3392	6	Gloucestershire	391862	194593
37	0029ABF	05/08/2019	3421	8	East Scotland (Perth and Kinross)	315731	711439
38	0029AF	15/07/2019	3449	7	Somerset	303796	134493
39	002ABF	30/07/2019	3424	7	East Scotland (Angus)	346600	747800
40	002AF	17/06/2019	3373	6	East Scotland (Grampian)	307684	862149
41	0030ABF	06/08/2019	3416	8	Highland	320883	865502
42	0030AF	24/06/2019	3393	6	Gloucestershire	388103	196359
43	0031ABF	04/08/2019	3354	8	Scotland (Scottish Borders)	361412	630255
44	0031AF	25/06/2019	3323	6	East Scotland (Aberdeenshire)	379000	767500
45	0032AF	23/06/2019	3372	6	East Scotland (Moray)	319119	861258
46	0033ABF	06/08/2019	3394	8	Gloucestershire	381905	192262
47	0033AF	21/06/2019	3382	6	East Riding of Yorkshire	511130	468906
48	0034ABF	06/08/2019	3424	8	East Scotland (Angus)	346600	747800
49	0034AF	25/06/2019	3426	6	Suffolk	596749	273827
50	0035ABF	06/08/2019	3377	8	East Scotland (Aberdeenshire)	396648	839818

51	0035AF	24/06/2019	3441	6	East Scotland (Aberdeenshire)	375800	855200
52	0036ABF	05/08/2019	3380	8	East Scotland (Perth and Kinross)	289702	716266
53	0036AF	24/06/2019	3387	6	Somerset	314498	122795
54	0037ABF	06/08/2019	3430	8	East Scotland (Grampian)	381129	852508
55	0037AF	22/06/2019	3425	6	North Yorkshire	495650	468870
56	0038ABF	06/08/2019	3323	8	East Scotland (Aberdeenshire)	379000	767500
57	0038AF	24/06/2019	3337	6	East Scotland (Aberdeenshire)	359650	865777
58	0039ABF	07/08/2019	3423	8	East Scotland (Fife)	340500	704500
59	0039AF	28/06/2019	3326	6	West Sussex	482700	119400
60	003ABF	31/07/2019	3339	7	Norfolk	593352	333256
61	003AF	17/06/2019	3415	6	East Riding of Yorkshire	484035	456155
62	0040ABF	07/08/2019	3422	8	East Scotland (Fife)	353500	702300
63	0040AF	27/06/2019	3477	6	East Riding of Yorkshire	492073	453688
64	0041ABF	08/08/2019	3347	8	Highland	287251	879453
65	0041AF	27/06/2019	3442	6	East Scotland (Aberdeenshire)	377137	854287
66	0042ABF	06/08/2019	3431	8	East Scotland (Angus)	369950	763550
67	0042AF	29/06/2019	3459	6	Staffordshire	384250	336060
68	0043ABF	06/08/2019	3374	8	Highland	293212	854187
69	0043AF	29/06/2019	3401	6	Staffordshire	381560	326810
70	0044ABF	28/07/2019	3354	7	Scotland (Scottish Borders)	361412	630255
71	0044AF	29/06/2019	3400	6	Staffordshire	384580	335810
72	0045ABF	11/08/2019	3369	8	East Scotland (Angus)	357884	758135
73	0045AF	29/07/2019	3324	7	East Riding of Yorkshire	491868	446904
74	0046ABF	12/08/2019	3415	8	East Riding of Yorkshire	484035	456155
75	0046AF	28/06/2019	3379	6	Highland	267943	866183
76	0047ABF	12/08/2019	3376	8	East Scotland (Angus)	357902	751084
77	0047AF	28/06/2019	3352	6	North Yorkshire	497055	468755

78	0048ABF	12/08/2019	3417	8	Highland	281317	854420
79	0048AF	26/06/2019	3325	6	West Sussex	489300	101700
80	0049ABF	09/08/2019	3442	8	East Scotland (Aberdeenshire)	377137	854287
81	0049AF	28/06/2019	3349	6	East Riding of Yorkshire	500637	439795
82	004ABF	03/08/2019	3450	8	Highland	264052	854289
83	004AF	17/06/2019	3362	6	North Yorkshire	492645	483625
84	0050ABF	19/08/2019	3337	8	East Scotland (Aberdeenshire)	359650	865777
85	0050AF	28/06/2019	3350	6	North Yorkshire	491211	470723
86	0051ABF	19/08/2019	3335	8	East Scotland (Moray)	335952	860944
87	0051AF	29/06/2019	3429	6	Suffolk	612159	240335
88	0052ABF	19/08/2019	3373	8	East Scotland (Grampian)	307684	862149
89	0052AF	28/06/2019	3428	6	Suffolk	628271	243636
90	0053ABF	21/08/2019	3339	8	Norfolk	593352	333256
91	0053AF	28/06/2019	3427	6	Suffolk	630237	254970
92	0054ABF	21/08/2019	3351	8	Norfolk	603631	320925
93	0054AF	30/06/2019	3451	6	Highland	265184	864986
94	0055ABF	20/08/2019	3397	8	East Scotland (Angus)	343222	756893
95	0055AF	30/06/2019	3450	6	Highland	264052	854289
96	0056ABF	19/08/2019	3385	8	East Scotland (Perth and Kinross)	319895	737755
97	0056AF	01/07/2019	3394	7	Gloucestershire	381905	192262
98	0057ABF	22/08/2019	3410	8	East Scotland (Angus)	343750	743100
99	0057AF	01/07/2019	3364	7	East Scotland (Aberdeenshire)	375775	777417
100	0058ABF	30/07/2019	3412	7	Highland	276861	865305
101	0058AF	01/07/2019	3363	7	East Scotland (Aberdeenshire)	385705	777094
102	0059ABF	21/08/2019	3413	8	Highland	270835	859499
103	005ABF	03/08/2019	3451	8	Highland	265184	864986
104	005AF	17/06/2019	3389	6	West Scotland	203842	560905

105	0060ABF	01/09/2019	3376	9	East Scotland (Angus)	357902	751084
106	0060AF	01/07/2019	3374	7	Highland	293212	854187
107	0061ABF	02/09/2019	3373	9	East Scotland (Grampian)	307684	862149
108	0061AF	01/07/2019	3372	7	East Scotland (Moray)	319119	861258
109	0062AF	01/07/2019	3335	7	East Scotland (Moray)	335952	860944
110	0063AF	01/07/2019	3437	7	East Riding of Yorkshire	487243	457500
111	0064ABF	14/07/2019	3425	7	North Yorkshire	495650	468870
112	0064AF	30/06/2019	3404	6	Highland	265017	855669
113	0065ABF	16/07/2019	3414	7	Highland	270245	865049
114	0065AF	29/06/2019	3403	6	Highland	261573	852632
115	0066ABF	16/07/2019	3411	7	Highland	274557	865148
116	0066AF	01/07/2019	3415	7	East Riding of Yorkshire	484035	456155
117	0067ABF	16/07/2019	3359	7	East Riding of Yorkshire	511165	464085
118	0067AF	30/06/2019	3368	6	East Scotland (Angus)	359834	759248
119	0068ABF	16/07/2019	3346	7	Cornwall	170095	40955
120	0068AF	30/06/2019	3369	6	East Scotland (Angus)	357884	758135
121	0069ABF	16/07/2019	3323	7	East Scotland (Aberdeenshire)	379000	767500
122	0069AF	01/07/2019	3441	7	East Scotland (Aberdeenshire)	375800	855200
123	006ABF	30/07/2019	3350	7	North Yorkshire	491211	470723
124	006AF	17/06/2019	3390	6	West Scotland	210553	546951
125	0070ABF	17/07/2019	3338	7	Norfolk	604731	320165
126	0070AF	01/07/2019	3380	7	East Scotland (Perth and Kinross)	289702	716266
127	0071ABF	19/07/2019	3379	7	Highland	267943	866183
128	0071AF	06/07/2019	3400	7	Staffordshire	384580	335810
129	0072ABF	19/07/2019	3429	7	Suffolk	612159	240335
130	0072AF	05/07/2019	3378	7	Highland	263655	851851
131	0073ABF	19/07/2019	3352	7	North Yorkshire	497055	468755
132	0073AF	10/07/2019	3347	7	Highland	287251	879453
133	0074ABF	18/07/2019	3427	7	Suffolk	630237	254970

134	0074AF	10/07/2019	3382	7	East Riding of Yorkshire	511130	468906
135	0075ABF	27/07/2019	3450	7	Highland	264052	854289
136	0075AF	03/07/2019	3422	7	East Scotland (Fife)	353500	702300
137	0076ABF	22/07/2019	3410	7	East Scotland (Angus)	343750	743100
138	0077ABF	22/07/2019	3353	7	East Scotland (Aberdeenshire)	370697	833533
139	0077AF	03/07/2019	3423	7	East Scotland (Fife)	340500	704500
140	0078AF	08/07/2019	3430	7	East Scotland (Grampian)	381129	852508
141	0079ABF	22/07/2019	3368	7	East Scotland (Angus)	359834	759248
142	0079AF	10/07/2019	3351	7	Norfolk	603631	320925
143	007ABF	01/08/2019	3441	8	East Scotland (Aberdeenshire)	375800	855200
144	007AF	18/06/2019	3406	6	East Riding of Yorkshire	492170	455320
145	0080ABF	19/07/2019	3428	7	Suffolk	628271	243636
146	0080AF	11/07/2019	3387	7	Somerset	314498	122795
147	0081ABF	22/07/2019	3376	7	East Scotland (Angus)	357902	751084
148	0081AF	14/07/2019	3442	7	East Scotland (Aberdeenshire)	377137	854287
149	0082ABF	21/07/2019	3375	7	East Scotland (Aberdeenshire)	360142	859995
150	0082AF	12/07/2019	3326	7	West Sussex	482700	119400
151	0083ABF	22/07/2019	3361	7	North Yorkshire	500755	475735
152	0083AF	15/07/2019	3365	7	East Scotland (Aberdeenshire)	376458	780926
153	0084ABF	22/07/2019	3385	7	East Scotland (Perth and Kinross)	319895	737755
154	0084AF	15/07/2019	3420	7	Norfolk	620040	321470
155	0085ABF	22/07/2019	3370	7	East Scotland (Angus)	360406	766405
156	0085AF	15/07/2019	3418	7	Norfolk	617650	319080
157	0086ABF	22/07/2019	3377	7	East Scotland (Aberdeenshire)	396648	839818
158	0086AF	15/07/2019	3337	7	East Scotland (Aberdeenshire)	359650	865777

159	0087ABF	22/07/2019	3431	7	East Scotland (Angus)	369950	763550
160	0087AF	14/07/2019	3389	7	West Scotland	203842	560905
161	0088ABF	29/07/2019	3360	7	East Riding of Yorkshire	512325	468225
162	0088AF	15/07/2019	3393	7	Gloucestershire	388103	196359
163	0089ABF	30/07/2019	3413	7	Highland	270835	859499
164	0089AF	15/07/2019	3392	7	Gloucestershire	391862	194593
165	008ABF	01/08/2019	3428	8	Suffolk	628271	243636
166	008AF	17/06/2019	3336	6	East Scotland (Grampian)	350179	865448
167	0090ABF	22/07/2019	3416	7	Highland	320883	865502
168	0090AF	13/07/2019	3459	7	Staffordshire	384250	336060
169	0091ABF	22/07/2019	3373	7	East Scotland (Grampian)	307684	862149
170	0091AF	13/07/2019	3402	7	Staffordshire	384010	330900
171	0092ABF	29/07/2019	3421	7	East Scotland (Perth and Kinross)	315731	711439
172	0092AF	13/07/2019	3401	7	Staffordshire	381560	326810
173	0093ABF	27/07/2019	3404	7	Highland	265017	855669
174	0093AF	14/07/2019	3403	7	Highland	261573	852632
175	0094ABF	29/07/2019	3397	7	East Scotland (Angus)	343222	756893
176	0094AF	09/07/2019	3325	7	West Sussex	489300	101700
177	0095ABF	28/07/2019	3390	7	West Scotland	210553	546951
178	0095AF	15/07/2019	3362	7	North Yorkshire	492645	483625
179	0096AF	14/07/2019	3477	7	East Riding of Yorkshire	492073	453688
180	009ABF	04/08/2019	3401	8	Staffordshire	381560	326810
181	009AF	17/06/2019	3361	6	North Yorkshire	500755	475735
182	0017AF	16/06/2019	3386	6	North Yorkshire	457295	487465
183	0018AF	17/06/2019	3357	6	Scotland (Scottish Borders)	383735	656881
184	0021AF	17/06/2019	3394	6	Gloucestershire	381905	192262
185	0025AF	12/06/2019	3422	6	na	353500	702300
186	0028ABF	29/07/2019	3336	7	Scotland	350179	865448
187	0032ABF	12/06/2019	3422		Scotland	353500	702300
188	0059AF	27/07/2019	3369	7	Scotland	357884	758135

189	0062ABF	17/07/2019	3405	7	na	481720	454480
190	0063ABF	16/07/2019	3406	7	na	492170	455320
191	0076AF	31/07/2019	3422	7	na	350179	865448
192	0078ABF	24/07/2019	3349	7	na	500637	439795

Table C. 1 Metadata on the collected samples from Fera, sample name, date of collection, month and locality information.

Unique_code	*input*	*filtered*	*No chimera*	*Lysis solution (mL)*	*Proteinase k(uL)*
000ABF	48920	30027	27587	10	50
0010ABF	53664	25283	24775	5	25
0010AF	68097	60230	58038	15	75
0011ABF	48807	23564	23153	15	75
0011AF	54469	48051	41898	20	100
0012ABF	46812	23919	22787	10	50
0012AF	34760	29080	28630	10	50
0013ABF	24892	14932	14735	10	50
0013AF	81619	71212	65537	15	75
0014ABF	54473	32857	31733	15	75
0014AF	77774	68743	61120	5	25
0015ABF	40221	17475	16534	15	75
0015AF	84894	72300	69350	5	25
0016ABF	67706	59387	52971	10	50
0016AF	85483	74931	66441	10	50
0017ABF	66916	60166	57043	10	50
0018ABF	76527	68132	59999	10	50
0018AF	95	26	18	10	50
0019ABF	27580	24360	24045	10	50
0019AF	57372	48694	46828	10	50
001ABF	48284	25332	24761	10	50
001AF	78909	70162	67112	10	50
0020ABF	32321	27725	27360	15	75
0020AF	60769	50980	44678	15	75
0021ABF	71686	62894	58762	25	125
0021AF	203	2	0	10	50
0022ABF	77539	68403	65184	20	100
0022AF	59942	50844	49110	15	75

0023ABF	71182	62773	60771	10	50
0023AF	53034	45802	42489	10	50
0024ABF	77234	68845	50100	5	25
0024AF	58538	52137	48190	5	25
0025ABF	61216	50432	45512	30	150
0025AF	59796	51084	44855	15	75
0026ABF	71728	63504	48784	10	50
0026AF	52241	45844	40813	5	25
0027ABF	65413	57788	46453	10	50
0027AF	61195	52689	46565	15	75
0028AF	49636	41350	39204	10	50
0029ABF	63242	49535	44083	5	25
0029AF	55198	43005	39191	10	50
002ABF	49066	27927	27245	10	50
002AF	83687	73798	71095	10	50
0030ABF	76684	64729	51258	15	75
0030AF	56491	46156	43821	20	100
0031ABF	65779	54134	47894	5	25
0031AF	55333	46068	35945	10	50
0032ABF	3	1	1	25	125
0032AF	56738	47323	43610	15	75
0033ABF	66883	56979	56017	15	75
0033AF	85404	66900	64296	10	50
0034ABF	77676	66443	65157	10	50
0034AF	28621	24947	24491	5	25
0035ABF	19827	17049	16959	5	25
0035AF	88829	76775	74169	10	50
0036ABF	55131	43121	42785	10	50
0036AF	97652	84818	77159	15	75
0037ABF	81853	52189	49615	10	50
0037AF	265	204	127	5	25

0038ABF	69293	54543	53363	10	50
0038AF	91692	80418	76154	10	50
0039ABF	71867	59817	57730	15	75
0039AF	46195	37524	36502	25	125
003ABF	59244	36748	35670	5	25
003AF	87429	77929	58143	5	25
0040ABF	69377	57912	54113	10	50
0040AF	95964	82286	72630	5	25
0041ABF	76033	61378	59796	10	50
0041AF	81416	70881	67233	20	100
0042ABF	21203	16780	16612	20	100
0042AF	75192	63558	60833	5	25
0043ABF	70668	60255	59253	15	75
0043AF	61543	53687	52882	5	25
0044ABF	14848	12078	11791	5	25
0044AF	4181	3617	3279	5	25
0045ABF	75176	60124	59310	15	75
0045AF	31650	27681	27323	15	75
0046ABF	82807	68904	67441	10	50
0046AF	46847	40971	40358	10	50
0047ABF	81224	63543	62208	10	50
0047AF	85879	72769	70616	10	50
0048ABF	95203	72439	70859	5	25
0048AF	68119	54561	53369	5	25
0049ABF	70739	58684	49979	5	25
0049AF	64684	49613	44519	15	75
004ABF	38120	16103	15755	5	25
004AF	70696	61683	51895	10	50
0050ABF	66576	54795	54667	10	50
0050AF	64849	56685	47863	25	125
0051ABF	108511	89218	78484	15	75

0051AF	53297	43004	37820	25	125
0052ABF	51799	40433	37787	5	25
0052AF	50768	38920	37987	20	100
0053ABF	88309	72077	62987	10	50
0053AF	64592	46592	44328	15	75
0054ABF	108873	88703	57890	5	25
0054AF	61122	50235	44508	5	25
0055ABF	104105	80427	69682	5	25
0055AF	59516	48380	35775	5	25
0056ABF	70085	48110	47953	20	100
0056AF	59521	48153	40185	5	25
0057ABF	68557	54127	51522	15	75
0057AF	62097	46086	43885	5	25
0058ABF	3890	3089	2922	5	25
0058AF	67069	57497	55863	20	100
0059ABF	103230	83103	81583	15	75
005ABF	45972	23858	22050	5	25
005AF	77811	67747	63527	5	25
0060ABF	92401	69576	64783	5	25
0060AF	53921	43318	40595	100	500
0061ABF	50893	38071	36410	5	25
0061AF	59790	49985	45311	10	50
0062ABF	249	1	0	5	25
0062AF	47718	38824	37546	10	50
0063ABF	18319	32	0	5	25
0063AF	51162	40526	39768	5	25
0064ABF	77064	64751	62130	20	100
0064AF	51009	43524	41894	5	25
0065ABF	71026	60980	60347	10	50
0065AF	59157	51801	50507	5	25
0066ABF	65841	55828	54526	5	25

0066AF	66721	57858	50145	40	200
0067ABF	90600	78781	77657	5	25
0067AF	41765	35153	34710	5	25
0068ABF	86226	73571	64512	10	50
0068AF	58673	49674	48043	5	25
0069ABF	75196	66487	65958	10	50
0069AF	51037	43520	41563	20	100
006ABF	34297	14035	13601	10	50
006AF	86192	75106	67553	50	250
0070ABF	78254	69158	65474	30	150
0070AF	48528	41169	40238	5	25
0071ABF	79316	66755	64760	10	50
0071AF	61290	50717	49141	20	100
0072ABF	76219	63155	59857	20	100
0072AF	63675	55166	54878	20	100
0073ABF	83509	68597	59921	5	25
0073AF	72924	63538	48776	10	50
0074ABF	5618	4610	4472	10	50
0074AF	21466	17170	16775	10	50
0075ABF	79201	63460	61600	5	25
0075AF	28919	25499	20346	15	75
0076ABF	79719	69453	60706	5	25
0076AF	58657	47556	47082	5	25
0077ABF	73992	63055	55031	10	50
0077AF	56059	48757	35153	5	25
0078AF	45269	39826	32129	5	25
0079ABF	71904	60024	57469	20	100
0079AF	26390	21016	20066	20	100
007ABF	36046	13402	12889	15	75
007AF	66548	57529	54719	15	75
0080ABF	51942	42386	39596	10	50

0080AF	52249	42758	35584	10	50
0081ABF	94578	83184	73286	10	50
0081AF	52667	45570	44085	15	75
0082ABF	91527	78470	78155	10	50
0082AF	70790	60443	58803	10	50
0083ABF	79642	70189	69024	5	25
0083AF	49804	42960	40126	10	50
0084ABF	71817	59732	57645	20	100
0084AF	59490	52141	50848	10	50
0085ABF	80061	69648	66162	15	75
0085AF	52608	46818	45741	10	50
0086ABF	7170	5207	5203	10	50
0086AF	67127	57016	53724	10	50
0087ABF	90394	78548	72702	10	50
0087AF	66936	58313	55714	5	25
0088ABF	14809	12585	12567	15	75
0088AF	60545	53007	47519	5	25
0089ABF	79560	68238	66591	10	50
0089AF	61285	53577	46121	5	25
008ABF	49564	24473	24273	10	50
008AF	88888	76565	74332	5	25
0090ABF	11109	8845	8664	5	25
0090AF	24617	21730	20628	10	50
0091ABF	67116	56231	55677	5	25
0091AF	10135	8689	8156	5	25
0092ABF	68111	25414	24559	5	25
0092AF	62165	53805	48859	5	25
0093ABF	79970	66979	63253	5	25
0093AF	54231	48068	42370	5	25
0094ABF	80141	66762	65113	15	75
0094AF	69947	60426	59470	5	25

0095ABF	82396	70434	69672	5	25
0095AF	66875	56704	53671	10	50
0096AF	56626	48562	46898	10	50
009ABF	40552	15163	14262	5	25
009AF	88347	79354	74040	15	75
00ExNeg_1	1091	3	0	62	2
00ExNeg_11	53	36	36	62	2
00ExNeg_3	11	3	0	62	2
00ExNeg_5	1878	2	0	62	2
00ExNeg_8	41	31	31	62	2
00ExNeg_9	28	8	0	62	2
00ExPos_1	48169	23037	22942	62	2
00ExPos_10	57962	18622	18260	62	2
00ExPos_11	3917	1620	1559	62	2
00ExPos_12	37928	15837	15736	62	2
00ExPos_2	52170	17691	17455	62	2
00ExPos_3	31687	13827	13735	62	2
00ExPos_4	37435	12986	12875	62	2
00ExPos_5	77414	35465	35158	62	2
00ExPos_6	69849	30018	29794	62	2
00ExPos_7	9703	808	795	62	2
00ExPos_8	4239	1790	1765	62	2
00ExPos_9	57298	22229	22052	62	2
00PCRpos_1	3649	757	756	62	2
00PCRpos_10	700	95	93	62	2
00PCRpos_11	4428	861	858	62	2
00PCRpos_12	1295	193	192	62	2
00PCRpos_2	1127	157	155	62	2
00PCRpos_3	676	125	122	62	2
00PCRpos_4	29732	2231	2220	62	2
00PCRpos_5	504	63	0	62	2

00PCRpos_6	45219	7898	7850	62	2
00PCRpos_7	127	1	0	62	2
00PCRpos_8	502	98	97	62	2
00PCRpos_9	872	136	136	62	2

Table C. 2 Sample information with reads entering and exiting the DADA2 pipeline, Lysis solution (ml) volume for each sample and Proteinase K (ul)

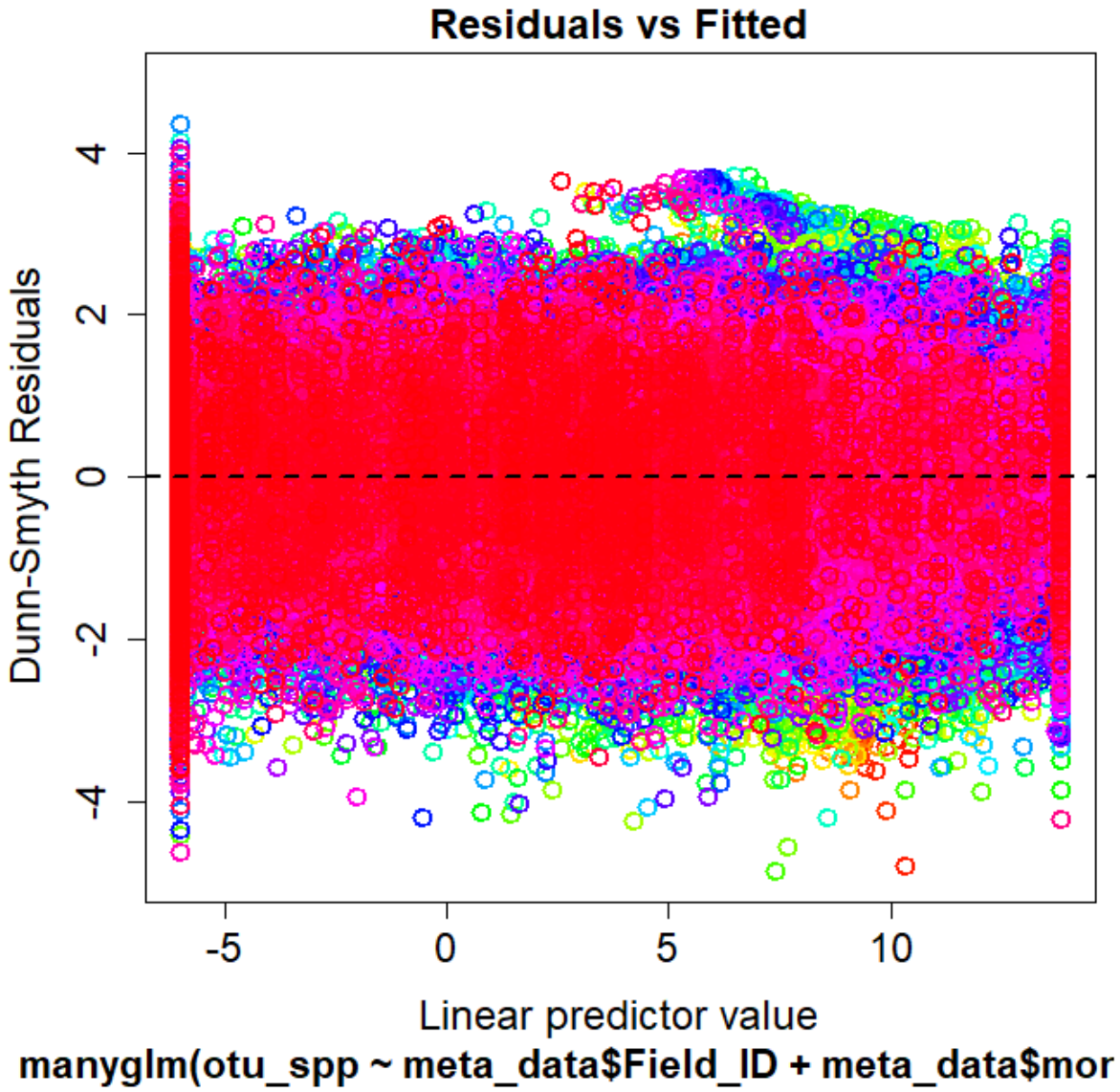


Figure C. 1 Residual plot for the GLM in mvabund for the two factors: Field and months.

Coefficients	*Estimate*	*Standard error*	*t-value*	*Pr(>t)*
Interecept	152.17	34.97	4.351	0.0001
Latitude	-2.44	0.64	-3.799	0.0007
Landscape_diversity	5.25	3.98	1.31	0.19

Table C. 3 Coefficients and estimates for the different effect of factors for species richness for the GLM. AIC: 204.62, residual deviance 1233 on 27 degrees of freedom

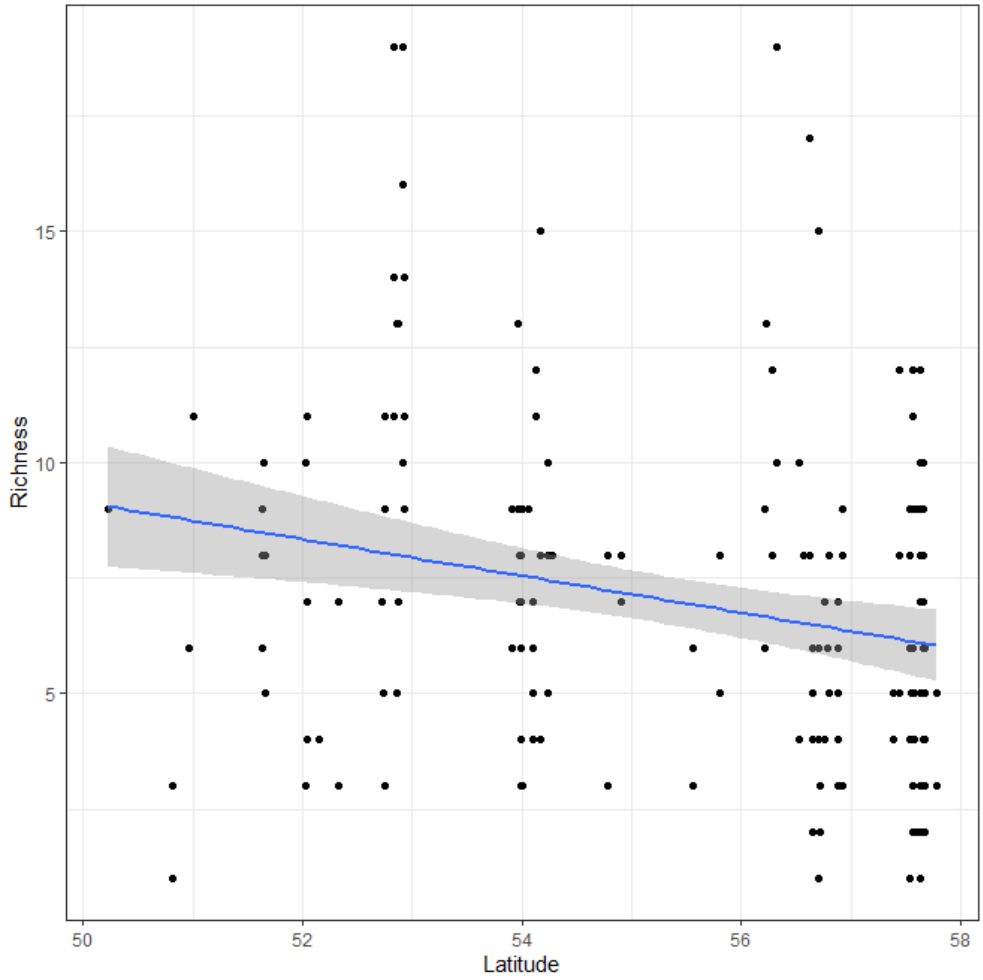


Figure C. 2 Plotted GLM blue line represent best fitted line, shaded areas are standard errors.

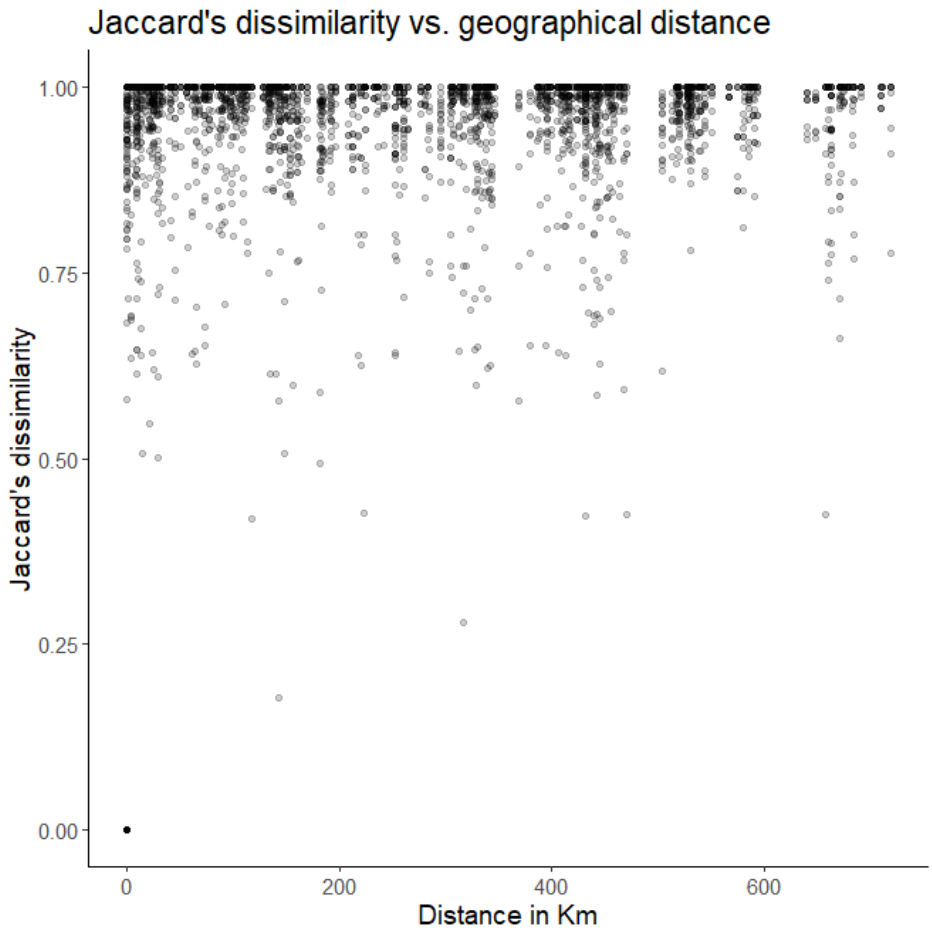


Figure C. 3 Beta diversity against physical distance for every sample pair (Includes only samples where all months have been sampled)

species	*Category*	*family*	*order*	*class*
bombus lucorum/terrestris	Pollinator	Apidae	Hymenoptera	Insecta
bombus terrestris	Pollinator	Apidae	Hymenoptera	Insecta
oscinella nigerrima	Pollinator	Chloropidae	Diptera	Insecta
ceutorhynchus obstrictus	Pollinator	Curculionidae	Coleoptera	Insecta
oedemera nobilis	Pollinator	Oedemeridae	Coleoptera	Insecta
rhagonycha fulva	Pollinator	Cantharidae	Coleoptera	Insecta
delia florilega	Pollinator	Anthomyiidae	Diptera	Insecta
adia cinerella	Pollinator	Anthomyiidae	Diptera	Insecta
athalia rosae	Pollinator	Tenthredinidae	Hymenoptera	Insecta
empis livida	Pollinator	Empididae	Diptera	Insecta
eriothrix rufomaculata	Pollinator	Tachinidae	Diptera	Insecta
eupeodes corollae	Pollinator	Syrphidae	Diptera	Insecta
eupeodes luniger	Pollinator	Syrphidae	Diptera	Insecta
orthops campestris	Pollinator	Miridae	Hemiptera	Insecta
saltella sphondylii	Pollinator	Sepsidae	Diptera	Insecta
syrphus ribesii	Pollinator	Syrphidae	Diptera	Insecta

bellardia vulgaris	Pollinator	Calliphoridae	Diptera	Insecta
calliphora vicina	Pollinator	Calliphoridae	Diptera	Insecta
delia platura	Pollinator	Anthomyiidae	Diptera	Insecta
pegoplata aestiva	Pollinator	Anthomyiidae	Diptera	Insecta
phaonia incana	Pollinator	Muscidae	Diptera	Insecta
scathophaga stercoraria	Pollinator	Scathophagidae	Diptera	Insecta
autographa gamma	Pollinator	Noctuidae	Lepidoptera	Insecta
sarcophaga subvicina	Pollinator	Sarcophagidae	Diptera	Insecta
lygus rugulipennis	Pollinator	Miridae	Hemiptera	Insecta
melinda viridicyanea	Pollinator	Calliphoridae	Diptera	Insecta
scaptomyza flava	Pollinator	Drosophilidae	Diptera	Insecta
thrips major	Pollinator	Thripidae	Thysanoptera	Insecta
botanophila fugax	Pollinator	Anthomyiidae	Diptera	Insecta
eudasyphora cyanella	Pollinator	Muscidae	Diptera	Insecta
lucilia caesar	Pollinator	Calliphoridae	Diptera	Insecta
cheilosia albitarsis	Pollinator	Syrphidae	Diptera	Insecta
bombus hortorum	Pollinator	Apidae	Hymenoptera	Insecta
scaptomyza pallida	Pollinator	Drosophilidae	Diptera	Insecta
coccinella septempunctata	Pollinator	Coccinellidae	Coleoptera	Insecta
plutella xylostella	Pollinator	Plutellidae	Lepidoptera	Insecta
diplazon laetatorius	Pollinator	Ichneumonidae	Hymenoptera	Insecta
glyphipterix simpliciella	Pollinator	Glyphipterigidae	Lepidoptera	Insecta
lonchoptera lutea	Pollinator	Lonchopteridae	Diptera	Insecta
pegoplata infirma	Pollinator	Anthomyiidae	Diptera	Insecta
melanogaster aerosa	Pollinator	Syrphidae	Diptera	Insecta
sarcophaga crassimargo	Pollinator	Sarcophagidae	Diptera	Insecta
andrena haemorrhoa	Pollinator	Andrenidae	Hymenoptera	Insecta
bombus lapidarius	Pollinator	Apidae	Hymenoptera	Insecta
fannia fuscula	Pollinator	Fanniidae	Diptera	Insecta
empis caudatula	Pollinator	Empididae	Diptera	Insecta
chloromyia formosa	Pollinator	Stratiomyidae	Diptera	Insecta
meigenia mutabilis	Pollinator	Tachinidae	Diptera	Insecta
hydrellia griseola	Pollinator	Ephydridae	Diptera	Insecta
delia coarctata	Pollinator	Anthomyiidae	Diptera	Insecta
sepsis punctum	Pollinator	Sepsidae	Diptera	Insecta
vespula vulgaris	Pollinator	Vespidae	Hymenoptera	Insecta
andrena cineraria	Pollinator	Andrenidae	Hymenoptera	Insecta
andrena dorsata	Pollinator	Andrenidae	Hymenoptera	Insecta
muscina levida	Pollinator	Muscidae	Diptera	Insecta
eristalis tenax	Pollinator	Syrphidae	Diptera	Insecta
hydrophoria lancifer	Pollinator	Anthomyiidae	Diptera	Insecta
bombus lucorum	Pollinator	Apidae	Hymenoptera	Insecta
dolichopus plumipes	Pollinator	Dolichopodidae	Diptera	Insecta
dolichopus simplex	Pollinator	Dolichopodidae	Diptera	Insecta
hydrotaea dentipes	Pollinator	Muscidae	Diptera	Insecta
drymeia hamata	Pollinator	Muscidae	Diptera	Insecta
muscina prolapsa	Pollinator	Muscidae	Diptera	Insecta

rhagio tringarius	Pollinator	Rhagionidae	Diptera	Insecta
coenosia tigrina	Pollinator	Muscidae	Diptera	Insecta
scaptomyza graminum	Pollinator	Drosophilidae	Diptera	Insecta
xylota segnis	Pollinator	Syrphidae	Diptera	Insecta
bellardia viarum	Pollinator	Calliphoridae	Diptera	Insecta
brachicoma devia	Pollinator	Sarcophagidae	Diptera	Insecta
dilophus febrilis	Pollinator	Bibionidae	Diptera	Insecta
eudasyphora cyanicolor	Pollinator	Muscidae	Diptera	Insecta
andrena nitida	Pollinator	Andrenidae	Hymenoptera	Insecta
dicraeus vagans	Pollinator	Chloropidae	Diptera	Insecta
tachyporus hypnorum	Pollinator	Staphylinidae	Coleoptera	Insecta
mesembrina meridiana	Pollinator	Muscidae	Diptera	Insecta
neomyia cornicina	Pollinator	Muscidae	Diptera	Insecta
sarcophaga nigriventris	Pollinator	Sarcophagidae	Diptera	Insecta
scaeva pyrastri	Pollinator	Syrphidae	Diptera	Insecta
geomyza tripunctata	Pollinator	Opomyzidae	Diptera	Insecta
sarcophaga variegata	Pollinator	Sarcophagidae	Diptera	Insecta
lucilia silvarum	Pollinator	Calliphoridae	Diptera	Insecta
cephus pygmeus	Pollinator	Cephidae	Hymenoptera	Insecta
oscinella frit	Pollinator	Chloropidae	Diptera	Insecta
apis mellifera	Pollinator	Apidae	Hymenoptera	Insecta
dolichopus popularis	Pollinator	Dolichopodidae	Diptera	Insecta
endasys plagiator	Pollinator	Ichneumonidae	Hymenoptera	Insecta
andrena nigroaenea	Pollinator	Andrenidae	Hymenoptera	Insecta
chlorops pumilionis	Pollinator	Chloropidae	Diptera	Insecta
bibio pomonae	Pollinator	Bibionidae	Diptera	Insecta
fannia armata	Pollinator	Fanniidae	Diptera	Insecta
bicellaria vana	Pollinator	Hybotidae	Diptera	Insecta
tenthredo arcuata	Pollinator	Tenthredinidae	Hymenoptera	Insecta
closterotomus norwegicus	Pollinator	Miridae	Hemiptera	Insecta
philaenus spumarius	Pollinator	Aphrophoridae	Hemiptera	Insecta
pseudovadonia livida	Pollinator	Cerambycidae	Coleoptera	Insecta
halictus rubicundus	Pollinator	Halictidae	Hymenoptera	Insecta
helina reversio	Pollinator	Muscidae	Diptera	Insecta
lucilia sericata	Pollinator	Calliphoridae	Diptera	Insecta
platycheirus manicatus	Pollinator	Syrphidae	Diptera	Insecta
cantharis flavilabris	Pollinator	Cantharidae	Coleoptera	Insecta
syrphus torvus	Pollinator	Syrphidae	Diptera	Insecta
plagiognathus arbustorum	Pollinator	Miridae	Hemiptera	Insecta
exorista rustica	Pollinator	Tachinidae	Diptera	Insecta
sarcophaga vagans	Pollinator	Sarcophagidae	Diptera	Insecta
helina evecta	Pollinator	Muscidae	Diptera	Insecta
neoascia podagrica	Pollinator	Syrphidae	Diptera	Insecta
sepsis cynipsea	Pollinator	Sepsidae	Diptera	Insecta
sepsis flavimana	Pollinator	Sepsidae	Diptera	Insecta
athalia bicolor	Pollinator	Tenthredinidae	Hymenoptera	Insecta

lucilia illustris	Pollinator	Calliphoridae	Diptera	Insecta
copidosoma floridanum	Pollinator	Encyrtidae	Hymenoptera	Insecta
ravinia pernix	Pollinator	Sarcophagidae	Diptera	Insecta
tachyporus obtusus	Pollinator	Staphylinidae	Coleoptera	Insecta
tachyporus chrysomelinus	Pollinator	Staphylinidae	Coleoptera	Insecta
aphidius matricariae	Pollinator	Braconidae	Hymenoptera	Insecta
bombus pascuorum	Pollinator	Apidae	Hymenoptera	Insecta
chlorops hypostigma	Pollinator	Chloropidae	Diptera	Insecta
schizohelea leucopeza	Pollinator	Ceratopogonidae	Diptera	Insecta
hylemya variata	Pollinator	Anthomyiidae	Diptera	Insecta
sphaerophoria scripta	Pollinator	Syrphidae	Diptera	Insecta
melanomya nana	Pollinator	Calliphoridae	Diptera	Insecta
anthocoris nemorum	Pollinator	Anthocoridae	Hemiptera	Insecta
vespula rufa	Pollinator	Vespidae	Hymenoptera	Insecta
bombus pratorum	Pollinator	Apidae	Hymenoptera	Insecta
dolerus aeneus	Pollinator	Tenthredinidae	Hymenoptera	Insecta
diplonevra funebris	Pollinator	Phoridae	Diptera	Insecta
apamea monoglypha	Pollinator	Noctuidae	Lepidoptera	Insecta
morellia simplex	Pollinator	Muscidae	Diptera	Insecta
paregle audacula	Pollinator	Anthomyiidae	Diptera	Insecta
abrostola tripartita	Pollinator	Noctuidae	Lepidoptera	Insecta
lasioglossum malachurum	Pollinator	Halictidae	Hymenoptera	Insecta
anthomyia liturata	Pollinator	Anthomyiidae	Diptera	Insecta
anthomyia confusanea	Pollinator	Anthomyiidae	Diptera	Insecta
phaonia tuguriorum	Pollinator	Muscidae	Diptera	Insecta
nephrotoma cornicina	Pollinator	Tipulidae	Diptera	Insecta
syritta pipiens	Pollinator	Syrphidae	Diptera	Insecta
eupeodes corollae	Predator	Syrphidae	Diptera	Insecta
coccinella septempunctata	Predator	Coccinellidae	Coleoptera	Insecta
sphaerophoria scripta	Predator	Syrphidae	Diptera	Insecta
pardosa amentata	Predator	Lycosidae	Araneae	Arachnida
rhagonycha fulva	Predator	Cantharidae	Coleoptera	Insecta
syrphus torvus	Predator	Syrphidae	Diptera	Insecta
aphidius avenae	Parasitoid	Braconidae	Hymenoptera	Insecta
tachyporus obtusus	Predator	Staphylinidae	Coleoptera	Insecta
aphis fabae	Pest	Aphididae	Hemiptera	Insecta
philonthus laminatus	Predator	Staphylinidae	Coleoptera	Insecta
aloconota gregaria	Predator	Staphylinidae	Coleoptera	Insecta
philonthus cognatus	Predator	Staphylinidae	Coleoptera	Insecta
acyrthosiphon pisum	Pest	Aphididae	Hemiptera	Insecta
metopolophium dirhodum	Pest	Aphididae	Hemiptera	Insecta
rhopalosiphum padi	Pest	Aphididae	Hemiptera	Insecta
erigone atra	Predator	Linyphiidae	Araneae	Arachnida
thrips tabaci	Pest	Thripidae	Thysanoptera	Insecta

tachyporus chrysomelinus	Predator	Staphylinidae	Coleoptera	Insecta
tenuiphantes tenuis	Predator	Linyphiidae	Araneae	Arachnida
aphidius rhopalosiphi	Parasitoid	Braconidae	Hymenoptera	Insecta
macrosiphum euphorbiae	Pest	Aphididae	Hemiptera	Insecta
chrysoperla carnea	Predator	Chrysopidae	Neuroptera	Insecta
sitobion avenae	Pest	Aphididae	Hemiptera	Insecta
tachyporus hypnorum	Predator	Staphylinidae	Coleoptera	Insecta
gyrohypnus angustatus	Predator	Staphylinidae	Coleoptera	Insecta
erigone dentipalpis	Predator	Linyphiidae	Araneae	Arachnida
atheta celata	Predator	Staphylinidae	Coleoptera	Insecta
trechus quadristriatus	Predator	Carabidae	Coleoptera	Insecta
aphidius ervi	Parasitoid	Braconidae	Hymenoptera	Insecta
brevicoryne brassicae	Pest	Aphididae	Hemiptera	Insecta
anotylus rugosus	Predator	Staphylinidae	Coleoptera	Insecta
anotylus tetracarinatus	Predator	Staphylinidae	Coleoptera	Insecta
harmonia axyridis	Predator	Coccinellidae	Coleoptera	Insecta
diaeretiella rapae	Parasitoid	Braconidae	Hymenoptera	Insecta
aleochara bipustulata	Predator	Staphylinidae	Coleoptera	Insecta
episyrphus balteatus	Predator	Syrphidae	Diptera	Insecta
dolycoris baccarum	Predator	Pentatomidae	Hemiptera	Insecta
myzus cerasi	Pest	Aphididae	Hemiptera	Insecta
trialeurodes vaporariorum	Pest	Aleyrodidae	Hemiptera	Insecta
porrhomma pygmaeum	Predator	Linyphiidae	Araneae	Arachnida
liriomyza huidobrensis	Pest	Agromyzidae	Diptera	Insecta
capitophorus elaeagni	Pest	Aphididae	Hemiptera	Insecta
pemphredon lugubris	Parasitoid	Crabronidae	Hymenoptera	Insecta
aleochara bilineata	Predator	Staphylinidae	Coleoptera	Insecta
passaloecus corniger	Parasitoid	Crabronidae	Hymenoptera	Insecta
bembidion lunulatum	Predator	Carabidae	Coleoptera	Insecta
amischa analis	Predator	Staphylinidae	Coleoptera	Insecta
hyadaphis foeniculi	Pest	Aphididae	Hemiptera	Insecta
myzus persicae	Pest	Aphididae	Hemiptera	Insecta
araeoncus humilis	Predator	Linyphiidae	Araneae	Arachnida
praon volucre	Parasitoid	Braconidae	Hymenoptera	Insecta
hippodamia variegata	Predator	Coccinellidae	Coleoptera	Insecta
microlinyphia pusilla	Predator	Linyphiidae	Araneae	Arachnida
pachygnatha degeeri	Predator	Tetragnathidae	Araneae	Arachnida
tetragnatha extensa	Predator	Tetragnathidae	Araneae	Arachnida
dinaraea angustula	Predator	Staphylinidae	Coleoptera	Insecta
pardosa lugubris	Predator	Lycosidae	Araneae	Arachnida
nasonovia ribisnigri	Pest	Aphididae	Hemiptera	Insecta

Table C. 4 Table of species of agricultural interest for which a category was found