Dissertation submitted in partial fulfillment

of the requirements for the doctoral degree

(Dr. rer. nat.)

Metabarcoding Invertebrate Communities in a Natural Park: From Methods to Diversity Patterns

by

Ameli Kim Kirse

from Siegburg

Bonn, April 2020

Faculty of Mathematics and Natural Sciences,

Rheinische Friedrich-Wilhelms-Universität Bonn

prepared at the Zoological Research Museum A. Koenig, Bonn

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn.

Angefertigt am Zoologischen Forschungsmuseum Alexander Koenig, Bonn.

Finanziert durch das BMBF geförderte Projekt ‘German Barcode of Life‘

Betreut von: Dr. Vera G. Fonseca

Erstgutachter: Prof. Dr. Bernhard Misof

Zweitgutachterin: apl. Prof. Dr. Birgit Gemeinholzer

Fachnahes Kommissionsmitglied: Prof. Dr. Dietmar Quandt

Fachfremdes Kommissionsmitglied: Prof. Dr. Thomas Döring

Tag der Promotion: 14. Dezember 2020

Erscheinungsjahr: 2021

Für meine Eltern

Summary

Biodiversity is worldwide suffering from a dramatic decline, whose underlying causes are still fairly unknown. The scientific investigation and identification of potential drivers of biodiversity loss are strongly hampered by the taxonomic impediment, which describes the shortage of taxonomic experts. As a result, studies dealing with changes in species communities usually only target a small set of taxa whereas the complex reciprocal relationships between taxa but also between the abiotic and biotic environment remain largely unresolved. By using Next Generation Sequencing (NGS) platforms it is now possible to assess unprecedented levels of biodiversity, allowing for the documentation of changes in species composition in a timely and cost-efficient manner.

The Caucasus region has been described to be one of the world's biodiversity hotspots, although the full magnitude of existing biodiversity is still unknown, due to the poorly developed local scientific infrastructure. Here we show that metabarcoding of soil samples allows for the timely assessment of the enormous degree of local existing biodiversity, what enables the detection of changes in species communities triggered by changes of the environmental parameters, e.g. height. To our knowledge, this is the first metabarcoding study aiming to assess invertebrate diversity of the Georgian Caucasus region. As eDNA metabarcoding of soils is still to mature, little is known about best practice procedures. Here, technical variables including choice of primer and *in-silico* filter are evaluated for their suitability and efficiency to monitor species communities along an elevational gradient (Chapter I).

With ongoing scientific research, taxonomic resolution and sensitivity of molecular identification methods are continuously improving. In order to monitor the progress of ongoing renaturation measures in the Eifel National Park, Germany, changes in invertebrate community composition along a forest conversion gradient from non-native Norway spruce (*Picea abies*) to European beech (*Fagus sylvatica*) are documented. To capture a picture as complete as possible, species inhabiting the above ground habitat were sampled with Malaise traps (Chapter II) while ground dwelling organisms were captured by eDNA metabarcoding of soil samples (Chapter III). For both source materials important methodological considerations are discussed, comprising the evaluation of a new non-destructive DNA extraction method for bulk samples. Here it is shown that with a well-considered choice of marker, primer and source material, metabarcoding is a powerful tool for nature conservation purposes as it allows to monitor changes in biodiversity in a timely manner, enabling the identification of underlying causes of biodiversity loss.

Invertebrate communities are defined and influenced by a complex network of interrelations and mutual influences between local occurring species. By metabarcoding the two types of samples, soil and Malaise trap bulk samples a complex picture of species communities encompassing both, above and below ground biodiversity is achieved. Here, a time lagged overlap of species occurrence between the two strata is observed, highlighting the power of metabarcoding to uncover species biology and thereby reciprocal relationships between species and how species community dynamics are affected by conversation measures influencing the abiotic and biotic environment (Chapter IV).

This work contributes significantly to the development and refinement of metabarcoding approaches for the assessment of invertebrate diversity. In the future these methods will gain growing importance as the ongoing biodiversity loss remains strong making the identification of its driver to one of the most important tasks of current nature conversation research.

Content

Chapter II - Monitoring Biodiversity Shifts – How Arthropod Communities Respond to Human Managed Ecological Reconstructions

Chapter III - Unearthing the Potential of eDNA Metabarcoding of Soil – Towards Best Practice Advises for Invertebrate Assessments

Chapter IV - Arthropod Diversity Assessment From the Ground Up: Detection and Seasonal Overlap between Soil and Malaise Traps

Introduction

Biodiversity

In the origin of species (Darwin 1859) it says

"(…) whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved".

Darwin was not the first to recognize the immense number of lifeforms on earth, but his hypothesis concerning the evolution of species was groundbreaking. Today when referring to the entity of these "endless forms of life" we often use the word "biodiversity". The term was coined by Walter Rosen in 1986 (Takacs 1996) during the National Forum of BioDiversity in Washington, USA. A group around Rosen and the second initiator of the event Edward Wilson wanted to create a new buzzword to promote awareness of nature conservation (Erasga 2012). The term 'biodiversity' was formerly used to simply encompass the variety of life on earth but today it has become one of the most important concepts in modern science. Indeed it has become a buzzword although many other definitions are existing and there are even more measures to describe it. A large number of people equalize biodiversity with the total number of occurring species, also referred to as species richness which can be described by several measures like alpha-, beta- and gamma diversity (Whittaker 1960). But there is a wide consensus in the scientific world that biodiversity cannot be reduced to a single number, although this is often claimed by decision-makers like politicians (Purvis and Hector 2000). For scientific purposes a broader definition of biodiversity is needed which includes genetic diversity, diversity on the ecosystem level and one that includes spatial and compositional attributes to cover important features such as function and resilience (Mace et al. 2012). The 'Convention About Biological Diversity', which came into force on the 29th December 1993 and is nowadays signed by 193 countries, defines biological diversity as

"the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems."

This definition is outstanding as it highlights the variability that arises from species being part of ecological complexes. Furthermore, it outpoints that the resulting interactions are both cause and consequence of biodiversity (Mace et al. 2012).

Biodiversity is characterizing and forming habitats all over the world (Jones et al. 1994). Next to abiotic factors, the existing biodiversity is having major influence on the development and conservation of habitats. Many studies already pointed out that ecosystem functions are highly influenced by the functional traits of organisms (Heemsbergen et al. 2004; Thompson and Townsend 2006) and that the existence of specific organisms is actually crucial for the maintenance and function of the corresponding ecosystem (Jones et al. 1994; Simberloff 1998; Ellison et al. 2005; Siddig et al. 2016). It is well-known, that a high biodiversity loss rate is often associated with changes in ecosystem functions, like a reduction in productivity (Hooper et al. 2012). Highly diverse ecosystems have a higher productivity, recycling and decomposing rate, also they capture more biologically essential resources (Balvanera et al. 2006; Cardinale et al. 2011; Hooper et al. 2012). Moreover, biodiversity is of crucial importance for ecosystem functioning, sustainability and resilience, but the magnitude and relative diversity of global species richness remains unknown (Fonseca et al. 2010). Approximations for the total number of species living on earth may vary between 3 to 100 million (May 2010). These wide disparities on species richness highlight a large gap in our basic knowledge about life on earth (Mora et al. 2011), which furthermore underlines the importance to study biodiversity. Nevertheless, many scientists agree that the total number of species on earth ranges around 8.7 million species worldwide (Mora et al. 2011), but biodiversity is not evenly distributed across the globe (Mittermeier et al. 1999; Myers et al. 2000; Grosberg et al. 2012). Life can be found in terrestrial, aquatic and marine environments, but species composition differs dramatically within these habitats (Grosberg et al. 2012). Transition of species between these three realms is limited by at least three factors – productivity, habitat complexity and the physical properties of the surrounding media (Grosberg et al. 2012). These key factors directly affect the kind of adaption and thereby to what extent speciation is likely to occur, which is moreover influencing the interactions between species (Grosberg et al. 2012). Around 80% of all macro-

scopic organism species are adapted to terrestrial habitats, while only 15% and 5% are inhabiting marine and aquatic habitats, respectively (Grosberg et al. 2012). This is mainly attributed to historical reasons, but also to the terrestrial physical setting which is more fragmented, opening up more possibilities to a higher diversity and the development of new dispersal strategies (Vermeij and Grosberg 2010). Terrestrial ecosystems are influenced by a wide variety of abiotic and biotic factors. Their interactions lead to the formation of numerous unique habitats hosting thousands of species including ecosystem engineers (Jones et al. 1994), foundation species (Ellison et al. 2005) and keystone species (Simberloff 1998). Species belonging to either one of the two latter mentioned groups are mandatory for the existence and health of the corresponding ecosystem (Siddig et al. 2016). Ecosystem engineers are creating, modifying and maintaining the habitat for example by modulating the availability of resources to other species (Siddig et al. 2016), often paving the road for the introduction of new species by establishing micro habitats leading to higher diversity levels (Vieira and Romero 2013). The existence of foundation- and keystone species underlines that every single species can be crucial for keeping an ecosystem in balance. If this eco-balance gets disturbed the consequences can be disastrous. In 2005 a study showed, that the local environment highly depends on the existence of a foundation tree species (Ellison et al. 2005). Removing such trees from their habitat had a major impact on many local species, for example the disruption of fundamental ecosystem processes like nutrient fluxes, carbon sequestration and rates of decomposition and thereby even dramatically alter the dynamic of associated aquatic ecosystems (Ellison et al. 2005). In addition, the local extinction of animal species can lead to a misbalance. The term “mesopredator-release” refers to the cascade effect, which can be triggered by the disappearance of a top predator from a habitat (Soulé et al. 1988). As a consequence the population of smaller carnivores can grow explosively, which leads to a dramatic decline or even the extinction of prey species (Soulé et al. 1988; Crooks and Soulé 1999). This example describes a well-understood relationship between species loss and the resulting consequences for the habitat. However, the interconnections of causes and consequences often remain unclarified. Ecosystems are sometimes still affected by events that took place several years or decades ago or

the consequences of an event are realized heavily delayed (Tilman et al. 1994; Essl et al. 2015a, 2015b). The latter is the case for the ongoing event of biodiversity loss.

Biodiversity loss

Only recently we realized that a silent loss of species in the order of magnitudes was and is still taking place all over the globe. In 2017 entomologists evidenced that Germanys insect biomass has declined by 70% over the last 27 years (Hallmann et al. 2017). On a yearly basis, this is a loss of approximately 2.8%. In 2018 a study revealed a 78%-98% decline of ground foraging and canopy dwelling arthropods in a tropical forest over a 36 years period, which is a mean decline between 2.7 and 2.2% per year (Lister and Garcia 2018). If both of the above mentioned studies were encompassing a shorter period, the loss rates would probably have been regarded as statistically non-significant (Sánchez-Bayo and Wyckhuys 2019). Therefore, longtime monitoring studies are indispensable. Fluctuations in insect populations can be a short-term response to stochastic environmental events or just the statistical 'noise' of natural population cycles (Woiwod and Harrington 1994; Conrad et al. 2004). Only with long time monitoring series the identification of significant directional trends is possible (Fox 2013). Unfortunately, long-term data is only available for a handful of taxa, that are mostly well-studied but species poor, like bumblebees and butterflies (Biesmeijer et al. 2006; Conrad et al. 2006; Franzén and Johannesson 2007; Frankie et al. 2009; Dupont et al. 2011; Fox 2013; Melero et al. 2016). Although this data is very revealing, the question remains on how applicable these trends are for further insect taxa (Tilman 2000; Hambler and Speight 2004; Fox 2013). The above mentioned examples are especially shocking viewed in the context of estimations that account approximately 80% of all terrestrial arthropod species to insects (Stork et al. 2015). Insect loss will probably have a cascade effect on birds, fish, mammals and amphibians as with declining insect biomass, food shortage for the above mentioned groups is just a question of time. In 2004 Thomas et al. showed that butterfly declines are exceeding similar changes as

described for birds and vascular plants in Great Britain (Thomas et al. 2004b; Fox 2013). Another study conducted in the Netherlands and Britain links local extinction of plant species with the decline of the associated pollinator populations (Biesmeijer et al. 2006). The observation of these cascade effects as well as the already observed enormous dimension of insect decline led scientist to the statement that the 6th big mass extinction is already on its way (Thomas et al. 2004b). Next to dramatic changes in ecosystems function, insect decline is influencing the well-being of humankind. Insects are accounting to the pollination of 80% of wild plant species (Potts et al. 2010) and 75% of cultivated species (Klein Alexandra-Maria et al. 2007). As a result insect decline is directly influencing the human population, as a shortage in food resources is likely to occur (Vanbergen and the Insect Pollinators Initiative 2013). The research field dealing with the influence of ecosystems and its drivers on humanity is called ecosystem services.

The silent disappearance of insects is not a new fact, but has been ignored for a long time (Nilsson et al. 2008; Potts et al. 2010; van Swaay et al. 2013) although it directly affects ecosystems functions and services (Vanbergen and the Insect Pollinators Initiative 2013). On a global scale biodiversity loss and population declines have been recognized since the late 1970th and as a result, scientific efforts escalated focusing on biodiversity assessment, conservation studies and the direct impacts of biodiversity loss on ecosystem functions and services (Perrings et al. 1992; Daily 1997; Cardinale et al. 2012). Thereby it was shown, that some insect groups directly impact the delivery of key ecosystem services like pollination and pest control (Balmford and Bond 2005; Fox 2013). The Rothamsted Insect Survey (RIS) (Woiwod and Harrington 1994) was established in the early 1960s in order to provide and store information on the spatial variation of insect abundances. Based on the collected data it became clear that the macro-moth populations of England decreased around 12% during 10 year intervals. In numbers: from among 337 moth species, 222 populations declined between 1968 and 2003 (Conrad et al. 2006). Similar dramatic declines were reported from Sweden. Macro-lepidopteran population at the Kullaberg Nature Reserve were monitored over a time period of 50 years (Franzén and Johannesson 2007). From former 597 species, 159 were no longer found in 2004, while only 22 new species had newly colonized the area. In total macro-lepidopteran

population were suffering from a loss of 45% (Franzén and Johannesson 2007). In Spain 66 butterfly species were monitored over a period of 20 years (1994-2014). Only 5 out of 66 populations remained stable, for 15 populations an increase were documented but with 46, the majority of the populations were declining (Melero et al. 2016).

Lepidopterans can be found in a broad range of habitats. They often express a high degree of host-plant specialization and are therefore often used as indicator species to assess the health of a habitat (Erhardt 1985; Summerville et al. 2004). Furthermore, they have a direct impact on the delivery of key ecosystem services like pollination (Johnson and Bond 1994). Additionally, lepidopterans serve as prey for numerous insectivorous animals. Therefore, the loss of butterflies and moths can have unforeseeable consequences for ecosystem services, function and stability. Next to lepidopterans also further insect groups are suffering from populations loss. One of the most famous groups of insects are the hymenopterans, which includes bees and wasps. Wild bees are accountable for approximately 23% of crop flower pollination (Rader et al. 2016), making them one of the main pillars of animal pollination and thereby for ecosystem services. Studies comparing historical records of bumblebee population with actual data draw a dark picture for the future of hymenopterans. In Denmark five out of 12 long-tongued bumblebee species disappeared over the last eighty years (Dupont et al. 2011). For central Europe 49 out of 60 Bombini species and subspecies were suffering from decline over the last 136 years, while four species even went extinct (Kosior et al. 2007). In addition, other wild bee population are declining. In Britain a decline in 52% of wild bee species were recorded since 1980. In the Netherland 67% of the species are suffering from decline (Biesmeijer et al. 2006). Additionally, a decline of wild bee populations outside of Europe has been observed. Between 1972 and 1996 a 60% decline of wild bee species visiting dry forests along a highway in Costa Rica were documented (Frankie et al. 2009). Next to monitoring and documenting species decline, several studies were conducted for the purpose of pinpointing the causes for biodiversity loss (Sánchez-Bayo and Wyckhuys 2019). Within these studies direct human impact on the habitat e.g. changing habitats and pollution, was next to climate change and invasive species identified as a major trigger (Loreau et al. 2001; Thomas et al. 2004a; Pimm 2008; Butchart et al. 2010; Sánchez-Bayo and Wyckhuys 2019).

Over the last centuries and decades, human activities were increasingly shaping the landscape all over the world. Urbanization, agriculture and industrialization resulted in the destruction of natural habitat and finally in landscape fragmentation (Sánchez-Bayo and Wyckhuys 2019). It is an undisputable fact, that especially the intensification of agriculture frontier, has contributed to the progress of insect decline. Since 1700 croplands as well as pasturelands have multiplied five times (Ramankutty et al. 2018). Especially during the times when agriculture took major steps from traditional, low-input farming to industrial characterized intensive productions (Ollerton et al. 2014; Sánchez-Bayo and Wyckhuys 2019) insect populations were suffering. The cultivation of genetically-uniform monocultures, the invention of synthetic fertilizers and pesticides and the removal of local biodiversity havens (e.g hedgerows, wild meadows and small forests) withdrawals insects basis for life (Williams and Osborne 2009; Brooks et al. 2012; Kennedy et al. 2013; Morandin et al. 2014; Sánchez-Bayo and Wyckhuys 2019). Many studies account the use of synthetic fertilizers, pesticides and herbicides to the second biggest drivers for biodiversity decline (Sánchez-Bayo and Wyckhuys 2019). Due to its high degree of toxicity the application of insecticides has a direct negative influence on affected local insect communities (Mulé et al. 2017). But also herbicides can trigger the decline of insect populations. Herbicides have a negative effect on the diversity of the local vegetation and as a result on certain plants specialized arthropod species are vanishing from the habitat together with these plants (Marshall et al. 2003). To maximize crop yield many farmers are applying large amounts of fertilizers to their crops. The excessive use of fertilizers have led to the eutrophication of many aquatic systems, which has recently been linked to a decline in populations of insects with a partly aquatic lifestyle like dragonflies, Chironomidae, Trichoptera and Ephemeroptera (Kalkman 2010; Jenderedjian et al. 2012). Next to human activities also biological factors can play a role in insect decline. A good example is the arrival of new species within an intact ecosystem. Invasive species can cause food networks to collapse (Kenta et al. 2007). However, the drivers of Insect decline cannot be considered individually as they are often strongly interwoven with each other. In the curse of globalization species are often transported as stowaways to new habitats. Depending on the environmental conditions of the arrival location they can easily harm the local populations. This is sometimes intensified by the

effects of climate change, which now allows many species to immigrate to new habitats (Thomas 2010), while local species are forced to narrow their distribution area (Freeman et al. 2018). Species decline is often a result of an interplay of many factors. Therefore, monitoring biodiversity shifts is of uppermost importance, so that first indicators for a starting diversity loss won't be overlooked

Assessment Methods

In former days species identification studies were mainly based on the morphology of the organisms under investigation. The term "morphology" is derived from the ancient greek μορφή, morphé, meaning "form", and λόγος, lógos, meaning "word, study, research". Thus, morphology is the study of the form and structure of organisms. Because most taxa differ in their outer appearance from other taxa, species description and identification has been based on the external morphology (outer appearance characterized by shape, color, size, patterns and structure) for a long time, but later also on the internal morphology (bones, organs). Based on their shared morphological characters, types of organisms are grouped together in a meaningful ways. These groups are called taxa and are assigned to a taxonomic rank. In modern taxonomy the eight major ranks are 'Domain', 'Kingdom', 'Phylum', 'Class', 'Order', 'Family', 'Genus' and 'Species' The system of classifying organisms in a nested hierarchy and naming them after a binominal nomenclature has been developed by the Swedish botanist Carl Linnaeus. Today, the classification groupings have dramatically changed but it's still their major aim to reflect evolutionary relationships.

Generally speaking, the more distant two taxa are from each other, the more morphologically different they are. But there's no rule without an exception - not every species is unique in its outer appearance. Cryptic species can express an identical exterior but are evolutionary separated from each other and are therefore unrelated. On the other hand, some species can express a broad range of phenotypic plasticity, meaning that outer appearance can vary largely within a species. A good example is the strawberry poison frog (*Oophaga pumilio*),

which can occur in more than 15 different color morphs, ranging from plain red to multi colored spotted specimens (Summers et al. 2003; Hagemann and Pröhl 2007). A taxonomist is able to identify between 1000 and 2000 species. To successfully conduct a study aiming to investigate total biodiversity of a habitat several experts have to co-operate. This is often challenging as experts are rare and inundated with enquiries. Considering the fact that worldwide 7 million species are still waiting to be described, 8700 experts would be needed to do the job (https://www.bolgermany.de). Unfortunately, the worldwide number of experts lies between 4000 and 6000. This shortage is usually referred to as the 'taxonomic impediment' (Wheeler et al. 2004). For the sake of completeness, it must be mentioned that not only the insufficient number of experts is described by the taxonomic impediment but also the incomplete knowledge of existing biodiversity as described before, experts unbalanced distribution across the globe and a thereby attendant underdeveloped worldwide taxonomic infrastructure (Coleman 2015). The taxonomic impediment has a direct influence on research. Due to the shortage of experts many studies are only focusing on certain sets of taxa instead of assessing total existing diversity. These narrowed studies prohibit conclusions about interactions, complex networks and causes and consequences of biodiversity shifts.

Invertebrates often serve as a proxy for habitat biomonitoring since they are widespread, very abundant, sensitive to environmental change and highly diverse (Rosenberg and Resh 1993). To assess the actual health status of a habitat on the basis of invertebrate diversity two possible methods can be applied (Lindenmayer and Likens 2011): First, the proof of the existence of an indicator species. An indicator species can be an individual taxon or a group of species, which is taken as a proxy for particular environmental conditions or changes (Lindenmayer and Likens 2011). Secondly, approaches which focus on the assessment of the entirety of the existing biodiversity (Lindenmayer and Likens 2011). Assessing the actual health status of an ecosystem based on indicator species has the great advantage that scientists can exclusively focus on sampling these target organisms, while assessing the total existing biodiversity underlies the assumption that every species inhabiting the habitat has been captured and identified beyond doubt. Nevertheless, the use of indicator species is controversially discussed as

the definition of an indicator species is not always clear and its concept can be highly idiosyncratic (Lehmkuhl et al. 2008; Roth and Weber 2008; Lindenmayer and Likens 2011). Additionally, despite the fact that several species are strongly associated with certain environmental conditions or habitat types (Lindenmayer and Likens 2011; Morrison et al. 2012), it can be premature to assume that this species is a robust indicator for it (Lindenmayer and Likens 2011). The bivalve *Velesunio ambiguous* was long time thought to be an indicator of the presence of heavy metals and therefore of the degree of pollution of the corresponding habitat (Walker 1981). However, subsequent studies proved that there is no correlation between the degree of pollution and the uptake of heavy metals by the bivalve (Millington and Walker 1983; Lindenmayer and Likens 2011). Indicator species should therefore be chosen with caution. Additionally, populations of indicator species could be affected by unforeseeable changes in abiotic and biotic factors, which measurements are not subject to the corresponding study. Because of these drawbacks, it can be dangerous to claim that monitoring population dynamics or the protection of only certain taxa can be as informative or efficient as collecting data of the overall biodiversity. To clarify the differences Purvis and Hector (2000) used the following metaphor: “Conserving one population of every species is rather like having one of each note in the Mozart concerto” (Purvis and Hector 2000). Overall, the assessment of ecosystem health could be more accurate when studies are based on the total existing diversity shaping the habitat (Lindenmayer and Likens 2011), which has longtime been hampered by the taxonomic impediment.

Barcoding

The development of sequencing tools opened the door for new monitoring methods on the basis of molecular characteristics. These tools partly circumvent the taxonomic impediment. Similar to most commercial products which can be easily identified on the basis of their associated unique barcodes, species are clearly demarcated by unique DNA Barcodes (Hebert et

al. 2003). A DNA Barcode is defined as a short gene sequence, taken from a standardized portion of a gene (Hebert et al. 2003). The international barcode of life (iBOL) initiative, which has the overarching objective to assess the magnitude and relative diversity of global species richness, uses a 648bp base-pair region of the mitochondrial cytochrome c oxidase subunit I (COI) for the identification of metazoans on species level.

To create a new reliable barcode for a taxon, individuals of the target taxon must be collected and identified. This still requires taxonomic expertise, whereas molecular tools are not entirely independent from the taxonomic impediment as well. However, once individuals are collected and identified all metadata must be carefully collected and stored in a database together with a high resolution picture of the corresponding specimens. After DNA extraction, each specimen is deposited in a collection to serve as voucher specimen. The extracted DNA is the basis for the creation of the barcode. The standard molecular tools for generating DNA barcode libraries are PCR amplification of the target region on behalf of a well-executed design and selection of primers, followed by Sanger sequencing (Shokralla et al. 2014; Kress et al. 2015). The basics for sequencing is the process of DNA synthesis and replication, which is the result of a complex interplay of DNA helicase, DNA polymerase, deoxynucleotides and finally the template DNA. At the beginning of every DNA replication process DNA helicase unwinds the twisted double helix structure and splits it into two single-stranded DNA strands. Afterwards RNA primase binds a primer to the single stranded DNA, which is used as a starting point for the DNA polymerase to create an “extension product” by attaching nucleotides to the 3’-hydroxyl end of the RNA primer. The growth of the extension product is performed in 5’ to 3’ direction by attaching new nucleotides with phosphodiester bridges between the 3’-hydroxyl group at the extension product and the 5’ phosphate group of the incoming deoxynucleotides (Clark and Pazdernik 2013). Because of their restricted ability to form hydrogen bonds with each other, only nucleotides complementary to the actual base of the template DNA are incorporated by the Polymerase. Sanger presented a sequencing approach in 1977 (Sanger et al. 1977, 1978), which has been improved over the last three decades. For replication, Sanger uses next to deoxynucleotides (dNTPs), dideoxynucleotides (ddNTPs) which are lacking the 3’-

hydroxyl group. This is hampering the formation of phosphodiester bonds. For Sanger sequencing four separate reactions are set up, each of them are containing radioactively labeled nucleotides and either ddA, ddT, ddC or ddG. Depending on concentration of dNTPs and ddNTPs, during elongation of the extension product DNA polymerase incorporates now a deoxynucleotides or the corresponding dideoxynucleotide. In contrast to the dNTPs the incorporation of ddNTPs ends the chain extension and the extension product is no longer elongated. This results in the formation of several extension products differing in size. At the beginning of Sanger Sequencing the resulting DNA fragments were separated from each other by electrophoresis using a polyacrylamide gel (Tipu and Shabbir 2015). For separation, an electrical field is applied forcing the negatively charged fragments to move through the gel towards the positive electrode. The higher the molecular weight of a fragment is, the slower it moves through the gel. With electrophoresis it is even possible to separate fragments from each other, which are only differing by one nucleotide in size (Clark and Pazdernik 2013). Today, sanger Sequencing is able to achieve read length of up to 1000bp (Kircher and Kelso 2010; Shokralla et al. 2014). This can be very helpful, when sequencing unknown genomes (de novo) or rearrange genomes segments (Shendure and Ji 2008; Mardis 2013). But the traditional Sanger DNA-sequencing method is not suitable for large-scale studies (Shokralla et al. 2012) as it is lacking the ability to read DNA from multiple templates in parallel, making it impossible to sequence the DNA of several hundreds of specimen at the same time (Shokralla et al. 2012).

Metabarcoding

During the last years tremendous steps have been taken towards a new technology which is able to produce an enormous amount of data when sequencing the DNA of several thousands of specimen in parallel (Barba et al. 2014). In 2000 the first “Next-Generation-Sequencing” was launched, which was capable of sequencing up to 25 million bases with an accuracy of 99% or more in one to four hour run with de novo sequencing (Margulies et al. 2005). Since 2006 this has increased to 400 million bases per run providing the possibility to sequence the

genome of single individuals (Noonan et al. 2006; Nowrousian et al. 2012). The first High Throughput Sequencing (HTS) which was applied to large scale biodiversity analysis was Roche's 454 pyrosequencing platform in bacteria and later in eukaryotes (Sogin et al. 2006; Fonseca et al. 2010; Hajibabaei 2012; Lallias et al. 2015). Since then, NGS platforms were continuously further developed. Today, the Illumina MiSeq,HiSeq (Caporaso et al. 2012) and NextSeq (Elbrecht et al. 2016) platforms are increasingly used for environmental studies.

Prior to sequencing specific Illumina adapters are ligated to the library fragments. Illumina platforms work with bridge PCR. Single molecule amplicons are pumped through the channels of a flow cell (Kircher and Kelso 2010), on which surface high-density forward and reverse primers, complementary to the ligated Illumina adapters, are attached (Shokralla et al. 2012). By active heating and cooling steps (Shokralla et al. 2012), the primers hybridize to the library fragments and a new strand is formed by reverse strand synthesis. When the new strand bends over, it hybridizes to one of the immobilized primers, forming a bridge between two primers (Kircher and Kelso 2010). In this way, a second strand is formed and finally several 1000 copies are generated (Kircher and Kelso 2010). All amplicons arising from a single template molecule during amplification remain clustered on a single location (Shendure and Ji 2008). The density of the amplified clusters is thereby defined by the ratio of the primers. For sequencing a primer is hybridized to a universal sequence, which is flanking the strand. Illumina works with sequencing by synthesis concept, meaning that in each cycle a mixture of four modified deoxynucleotides are added. Each of it contains one of four fluorescent labels together with a reversibly terminating moiety (Shendure and Ji 2008). The moiety ensures, that the incorporation reaction is stopped after each cycle. With fluorescent dyes, the label of the incorporated base is analyzed. The next cycle is started after removing the fluorescent label and the moiety with the help of chemical cleavage (Shendure and Ji 2008). The most common errors in Illumina are substitutions, which can be often traced back to an incomplete cleavage of the labels (Shendure and Ji 2008). Furthermore, A/C and G/T bases are respectively similar in their emission spectra and the limited separation from each other can only be observed using optical filters. As a consequence, an increase in the number of substitution errors,

between A/C and G/T, is found (Kircher and Kelso 2010). Chemistry crystals, dust and lint particles are sometimes identified as clusters. This equals the so called 'ghost wells' in the Roche platform (Kircher and Kelso 2010). Furthermore, likely as a consequence of the growing amount background noises, error rate increases with ongoing sequencing process (Kircher and Kelso 2010).

On the basis of NGS new research fields have emerged making it difficult to use the right terminology as this is often still in flux (Douglas et al. 2012). In 2010 Fonseca et al. published the first NGS study dealing with the assessment of eukaryotic community composition. The analysis of homologous genes in order to perform a large-scale analysis of taxon richness was termed metagenetics (Creer et al. 2010). Since Fonseca et al (2010) published their pioneer work further marker gene studies were published using similar methods but giving them various nomenclature. Next to 'metasystematics' (Hajibabaei 2012) and 'amplicon-sequencing' (Creer et al. 2010, 2016), Taberlet et al. introduced the term metabarcoding (Taberlet et al. 2012a). Today the term 'metabarcoding' is widely established to describe the method of using gene-specific PCR primers to amplify DNA from complex DNA mixtures (Creer et al. 2016; Fonseca 2018).

A crucial step for assessing large-scale molecular biodiversity is the choice of the marker gene. So far, no marker has been identified to be capable to identify all taxa across the taxonomic tree. However, it was found that some barcode regions can provide a high taxonomic resolution and coverage for certain taxonomic groups of interest e.g. Cytochrome oxidase subunit I gene (COI) for arthropods (Andújar et al. 2018; Elbrecht et al. 2019), the small subunit 18S rRNA gen (18S) for Metazoa (Fonseca et al. 2010), the internal transcribed spacer (ITS) for fungi (Schoch et al. 2012), the maturase K gene (matK) and the large sub-unit of ribulose 1,5-biphosphate carboxylase gene (rbcL) for plants (CBOL Plant Working Group et al. 2009) and the small subunit 16S rRNA gen (16S) for bacteria (Chakravorty et al. 2007).

As the power of metabarcoding is directly linked to complexity and completeness of available databases (Somervuo et al. 2017) a lot of effort has been taken to establish well-maintained

reference databases populated with millions of taxonomically verified DNA reference sequences (e.g. BOLD, GenBank, Greengenes, SILVA). Despites the increasingly large number of references and the often pronounced high taxonomic resolution of the above outlined marker it has been noticed that several metabarcoding studies still imply a high number of false negative results. For instance, it has been observed that the COI marker widely fails to amplify for hymenopterans when they are part of a bulk sample (Brandon-Mong et al. 2015). The COI marker is a very polymorphic, fast evolving gene and thus highly suitable for the identification of taxa on species level (Creer et al. 2010; Deagle et al. 2014). However, due to its protein coding character the marker is less constraint by selection whereas it is showing an increasingly high number of mutation at the third codon of the primer binding sites leading to primer mismatches when targeting genetically diverse taxonomic groups (Piñol et al. 2015; Clarke et al. 2017). As a consequence the use of universal primers increasingly results in falsely estimated alpha and beta diversities since some taxonomic groups will be missed due to primer bias (Ficetola et al. 2010; Deagle et al. 2014). Therefore, the use of several primers targeting the same gen region is highly recommended to enlarge taxonomic resolution (Deagle et al. 2014). In contrast to the COI marker is the 18S marker characterized by highly conserved primer sites, enabling the identification of several taxa on the basis of a single primer (Fonseca et al. 2011). Combined with the fact that more than 150 copies of the marker gene can be found within each cell of the organism, facilitating the amplification of barcodes from microscopic organism, the nuclear 18S marker offers a comparatively large taxonomic coverage (Creer et al. 2010). However, in contrast to the COI marker, evolves the 18S gene much slower, making it difficult and sometimes even impossible to identify taxa on species level. As a result studies based on the 18S marker often tend to underestimate truly existing biodiversity (Tang et al. 2012). It is therefore increasingly recommended to use several markers as this practice leads to a better taxonomic resolution and a wider taxonomic coverage (Zhang et al. 2018; Marquina et al. 2019a).

A great advantage of metabarcoding is that it reduces the sampling effort to a minimum, while increasing the probability to sample every species inhabiting the target habitat, including microbial organisms and bacteria. Furthermore, it enables the possibility to sequence the genetic

information of whole-populations (Lazarevic et al. 2009) by analyzing DNA directly recovered from bulk (Braukmann et al. 2018; Watts et al. 2019) or environmental samples (Watts et al. 2019), which can for example consist of soil, sediment, water or even air. (Taberlet et al. 2012a; Barnes and Turner 2016; Deiner et al. 2017; Ruppert et al. 2019). Environmental DNA (eDNA) is defined as all DNA captured from environmental samples, without first isolating any target organisms (Taberlet et al. 2012a). eDNA can inter alia originate from skin, mucous, saliva, sperm, secretions, eggs, feces, urine, blood, roots, leaves, fruits, pollen, rotting bodies or microorganisms. eDNA comprises several types of DNA. First, intracellular DNA, originating from living cells and living multicellular organisms, and secondly extracellular DNA (Levy-Booth et al. 2007; Pietramellara et al. 2009; Taberlet et al. 2012a, 2012b; Creer et al. 2016), which results from cell lysis after cell death (Pietramellara et al. 2009; Taberlet et al. 2012a, 2012b). While intracellular DNA is often intact, extracellular DNA can be highly degraded (Pietramellara et al. 2009). The degree of degradation can significantly influence species detection rate as is has a direct influence on PCR bias (Krehenwinkel et al. 2018). However, the fade of eDNA in the environment has been studied only rudimentarily. This is especially true for terrestrial habitats. While several studies have focused on the investigation of eDNA degradation processes in aquatic habitat (Strickler et al. 2015; Turner et al. 2015), little is known about environmental influences affecting the persistence of eDNA in soil. However it is known that once DNA is in the environment it is vulnerable to degradation by enzymatic nucleases, while other fractions of the negatively charged DNA adsorb to negatively charged silica particles, clay and organic matter on behalf of cation and phosphate bridging (England et al. 2004; Taberlet et al. 2012b). This makes the DNA much more resistant to digestion by enzymatic nuclease. Depending on source material and the local abiotic and biotic factors (Barnes et al. 2014) the timeframe of DNA persistence can vary from hours to hundreds and thousands of years (Jørgensen et al. 2012; Thomsen et al. 2012; Kucherenko et al. 2018). Furthermore, the degradation dynamics of eDNA are influenced by intrinsic properties like DNA source (Pietramellara et al. 2009), G+C content (Hofreiter et al. 2001; Vuillemin et al. 2017), molecular purity (Nielsen et al. 2000) and molecular weight (Ogram et al. 1994; Pietramellara et al. 2009). But

as outpointed already, extrinsic conditions like soil mineralogy (Levy-Booth et al. 2007; Pietramellara et al. 2009; Gardner and Gunsch 2017), organic components (Ogram et al. 1988; Saeki et al. 2011), pH (Levy-Booth et al. 2007), temperature (Widmer et al. 1996; Gulden et al. 2005) and moisture (Widmer et al. 1996; Sirois and Buckley 2019) are influencing the fade of eDNA as well. Nevertheless, our understanding of the degradation processes and its influences on DNA persistence, especially for terrestrial habitats is still limited as only a few studies provide quantitative understanding of the underlying dynamics (Sirois and Buckley 2019). However, the importance to study eDNA fade in the environment is of uppermost importance as it can significantly influence the accuracy of conducted biodiversity assessments. Next to the assessment of recent habitat communities, eDNA can also be used to reconstruct ancient ones e.g. by using ice or permafrost as source material (Haile et al. 2009; Jørgensen et al. 2012; Thomsen and Willerslev 2015). Accordingly, long time persistence of eDNA could distort conclusions made about current existing biodiversity. Additionally, the origin of extracted eDNA often remains unclear as it could have been transported into the habitat by birds, large predators, stream currents or wind (Creer et al. 2016; Deiner et al. 2017). To circumvent these biases, community analysis are often based on DNA extracted from bulk samples, which has been coined ‘community DNA’ (Deiner et al. 2017). There is a thin line between eDNA and community DNA. Next to the DNA of the collected organisms, bulk samples can also contain DNA from organisms not existing at the study site. This DNA can originate from the gut content of the collected specimens or is cutaneous intracellular and extracellular DNA (e.g. parasites) (Deiner et al. 2017). Depending on the purpose of the study and chosen target specimen, an appropriate source material must be chosen (Shokralla et al. 2010; Hajibabaei 2012; Taberlet et al. 2012b; Lallias et al. 2015, 2015; Valentini et al. 2016). As already outpointed, the source material has a direct influence on the interpretation of the study concerning time and space. While community DNA provides relatively exact information of the current composition of a local community, the risk that extracted eDNA is a relic of a past community or contains DNA from organisms no longer existing in the study area is significantly increased. Hence, the interpretation of eDNA data must be performed very carefully to avoid false positive inferences (Deiner et al. 2017).

Depending on the study organisms, bulk samples for community DNA are collected in various ways. In terrestrial habitats, flying insects are usually collected with Malaise traps (Hallmann et al. 2017), while ground dwelling beetles are caught with pit falls traps (Greenslade 1964). If researchers are interested in nematode or annelid community compositions Baerman funnels are an appropriate method to circumvent sorting and sampling of soil samples manually (Walker and Wilson 1960; Van Vliet et al. 2004). In aquatic habitats, kick sampling is an appropriate method to sample macroinvertebrates (Lenat 1988). Until further processing, collected specimens are frozen or preserved e.g. in ethanol. A huge advantage of bulk samples is that newly discovered species are morphologically unharmed and it is still possible to describe them. In 2010 it was shown, that it is possible to extract DNA of the collected specimen directly from the preservative ethanol (Shokralla et al. 2010). Nevertheless, there is still no guarantee that the DNA of every specimen is contained within ethanol, as the mechanics of DNA release are not fully understood yet. Specimens contained in one bulk sample are differing in size, shape and morphology. While some beetles have a thick cuticle, small and fragile dipterans are much more vulnerable to disruptions. Furthermore, the size of the specimens has an influence on the amount of retrieved DNA. Due to the fact, that the amount of DNA of a large specimen in the resulting DNA template will exceed the one of a smaller organism, especially when grinding has been performed prior to extraction, an increasing number of metabarcoding studies is now based of presorted bulk samples, which only contain specimens of similar sizes. Without size sorting a few large specimens would dominate the resulting dataset while smaller specimens would probably remain undetected (Elbrecht et al. 2017). Elbrecht et al. (2017) showed that sorting by size and balancing the amount of tissue used per size fraction prior to grinding can increase the number of recovered taxa by 30%. After sorting the specimens by size they are grinded to ensure that DNA of every specimen is present in the resulting DNA soup from which DNA will be extracted.

When working with eDNA the extraction methods are highly depending on the type of target DNA. To extract intracellular DNA, living cells have to be lysed by adding a lysing agent to the sample and/or by applying mechanical disruption methods to it. However these approaches

are thought of to negatively affect quality and quantity of extracted extracellular DNA (England et al. 2004). Next to intracellular DNA extracellular DNA can be used for the reconstruction of species communities as well. Former studies have indicated, that there must be a competition between DNA and phosphate concerning the adsorption of particles (Taberlet et al. 2012b), therefore extracellular DNA can be selectively extracted from soil samples by using a saturated phosphate buffer (England et al. 2004).

For downstream analysis several bioinformatic tools are available. These tools keep developing in a phenomenal pace and new tools are made publicly accessible on a regular basis. The large number of available algorithms, programs and scripts is making the decision complicated. But as its application can have a strong influence on the outcome of the analysis the choice of the *in-silico* tools should be carefully considered in order to achieve reliable results (Kopylova et al. 2016). To narrow down the choice of tools the usage of pipelines like QIIME (Caporaso et al. 2010) and Vsearch (Rognes et al. 2016) can provide a good service. Pipelines can be programs, interfaces or web tools guiding the user through the whole analysis. Each pipeline is implementing a broad range of tools the user can choose from while performing the analysis. The choice of the pipeline should be closely coordinated with the choice of marker gene. Most pipelines have in common that they rely on a similar workflow comprising eight steps, illustrated in Figure 1. During preprocessing, raw sequences are trimmed to remove Illumina adapters as well as low quality bases (Martin 2011). When working within Illumina datasets, the forward and reverse reads need to be joined together as Illumina allows the sequencing of both ends of a fragment (Quail et al. 2012). Subsequently a quality filtering step ensures that the merged paired end reads have a certain size and that the phred score, which indicates a predicted error probability for

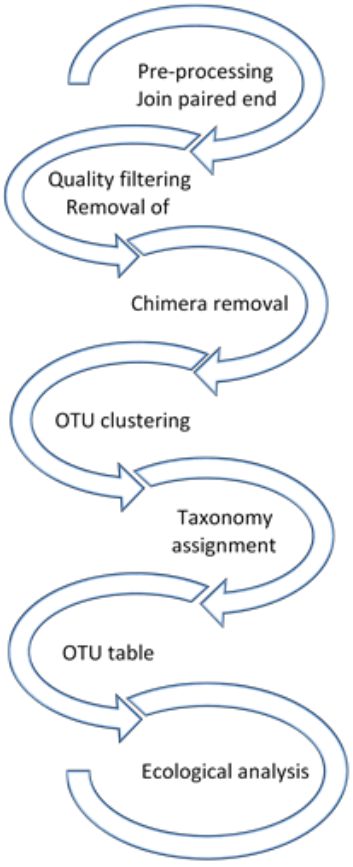


Figure 1 Bioinformatic workflow as found in several metabarcoding pipelines

each base introduced during sequencing, is of acceptable level (Caporaso et al. 2010). During quality filtering, PCR artifacts and sequencing errors are removed. Special emphasis must be paid to Chimera detection and removal (Edgar et al. 2011). Chimeras are known to be hybrid products which are formed during PCR, when an incomplete extension of a read is leading to a foreign strand, which is copied to completion in the next cycle (Fonseca et al. 2012). After chimera removal, sequences are clustered into Operational Taxonomic Units (OTUs) (Kopylova et al. 2016). An OTU is a cluster of sequences falling into a fixed similarity threshold, which is commonly 97% sequence similarity level, but can vary within studies. There are several algorithms available to cluster sequences into OTUs. The applied OTU clustering algorithm can have a huge impact on the number and quality of the resulting clusters (Kopylova et al. 2016), therefore choice should be made with care. The resulting OTU list should also undergo revision. Many rare OTUs, mainly OTUs with a low number of assigned sequences to it, are a result of PCR and sequencing errors (Huse et al. 2010). Approximately 38% of all OTUs containing only one sequence are erroneous (Brown et al. 2015). To exclude false positives from the dataset usually all singletons are discarded from the dataset (Huse et al. 2010; Kunin et al. 2010; Brown et al. 2015). Unfortunately, this practice means, that sometimes rare but real OTUs are lost as well. Depending on the studied ecosystem, genuinely infrequent species can be expected to be the largest proportion, but the exclusion of rare OTUs from the dataset distorts the overall picture of the community, as it only retains dominant species (Frøslev et al. 2017). Nowadays many post-clustering filter algorithms are available to filter the resulting OTU table for erroneous OTUs like LULU (Frøslev et al. 2017) and dbotu3 (Olesen et al. 2017). The core task of LULU is to identify and merge ‘daughter’ OTUs with consistently co-occuring sequences that are similar but more abundant ‘parent’ OTUs across a multi-sample data set, under the assumption that the ‘daughter’ OTUs are artefacts (Frøslev et al. 2017). These new tools are now increasingly implanted into studies to draw a more exact picture of species diversity (Drinkwater et al. 2019; Fløjgaard et al. 2019; Marquina et al. 2019b).

For taxonomy assignment curated OTUs are aligned against a reference database. Several well-maintained databases are publicly available and can be downloaded to local workstations. First and foremost the choice of the reference database depends on the chosen gene

marker. The SILVA database and Genebank can be used for the analysis of 16S and 18S datasets, while the BOLD database is a good choice when working with COI sequences. Although most databases try to stick to a “gold-standard” erroneous sequence assignments can occur. In such cases DNA reference libraries lack information on certain taxonomic groups. Therefore, the use of more than one reference database is expected to maximizing taxonomic coverage and reliability of results (Macher et al. 2017). The resulting OTU table is a good basis for the following ecological analysis, which can easily be performed using Gnu R. Today several packages are available for conducting ecological analysis on the basis of OTU tables e.g. the package vegan (Dixon 2003; Oksanen et al. 2007) but also other bioinformatic tools, which are often open source provide goof results (e.g. EstimateS).

Metabarcoding is a powerful tool to create species lists and allow insights into shifts of community composition. Therefore, it is a suitable method for long time monitor programs, which are urgently needed to document and understand population dynamics in relation to environmental conditions. A growing understanding of the dynamics will enable a better protection of the native fauna and flora and will likely provide hints how to stop ongoing insect decline.

Literature

Andújar C., Arribas P., Yu D.W., Vogler A.P., Emerson B.C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology. 27:3968–3975.

Balmford A., Bond W. 2005. Trends in the state of nature and their implications for human well-being. Ecology Letters. 8:1218–1234.

Balvanera P., Pfisterer A.B., Buchmann N., He J., Nakashizuka T., Raffaelli D., Schmid B. 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. Ecology Letters. 9:1146–1156.

Barba M., Czosnek H., Hadidi A. 2014. Historical perspective, development and applications of next-generation sequencing in plant virology. Viruses. 6:106–136.

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

Barnes M.A., Turner C.R., Jerde C.L., Renshaw M.A., Chadderton W.L., Lodge D.M. 2014. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. Environmental Science & Technology. 48:1819–1827.

Biesmeijer J.C., Roberts S.P., Reemer M., Ohlemüller R., Edwards M., Peeters T., Schaffers A., Potts S.G., Kleukers R., Thomas C. 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science. 313:351–354.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

Braukmann T.W., Ivanova N.V., Prosser S.W., Elbrecht V., Steinke D., Ratnasingham S., Jeremy R. deWaard, Sones J.E., Zakhariv E.V., Hebert P.D. 2018. Revealing the Complexities of Metabarcoding with a Diverse Arthropod Mock Community. bioRxiv.:433607.

Brooks D.R., Bater J.E., Clark S.J., Monteith D.T., Andrews C., Corbett S.J., Beaumont D.A., Chapman J.W. 2012. Large carabid beetle declines in a United Kingdom monitoring network increases evidence for a widespread loss in insect biodiversity. Journal of Applied Ecology. 49:1009–1019.

Brown S.P., Veach A.M., Rigdon-Huss A.R., Grond K., Lickteig S.K., Lothamer K., Oliver A.K., Jumpponen A. 2015. Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecology. 13:221–225.

Butchart S.H.M., Walpole M., Collen B., Strien A. van, Scharlemann J.P.W., Almond R.E.A., Baillie J.E.M., Bomhard B., Brown C., Bruno J., Carpenter K.E., Carr G.M., Chanson J., Chenery A.M., Csirke J., Davidson N.C., Dentener F., Foster M., Galli A., Galloway J.N., Genovesi P., Gregory R.D., Hockings M., Kapos V., Lamarque J.-F., Leverington F., Loh J., McGeoch M.A., McRae L., Minasyan A., Morcillo M.H., Oldfield T.E.E., Pauly D., Quader S., Revenga C., Sauer J.R., Skolnik B., Spear D., Stanwell-Smith D., Stuart S.N., Symes A., Tierney M., Tyrrell T.D., Vié J.-C., Watson R. 2010. Global Biodiversity: Indicators of Recent Declines. Science. 328:1164–1168.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7:335–336.

Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Huntley J., Fierer N., Owens S.M., Betley J., Fraser L., Bauer M., Gormley N., Gilbert J.A., Smith G., Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal. 6:1621–1624.

Cardinale B.J., Duffy J.E., Gonzalez A., Hooper D.U., Perrings C., Venail P., Narwani A., Mace G.M., Tilman D., Wardle D.A. 2012. Biodiversity loss and its impact on humanity. Nature. 486:59.

Cardinale B.J., Matulich K.L., Hooper D.U., Byrnes J.E., Duffy E., Gamfeldt L., Balvanera P., O'connor M.I., Gonzalez A. 2011. The functional role of producer diversity in ecosystems. American Journal of Botany. 98:572–592.

CBOL Plant Working Group, Hollingsworth P.M., Forrest L.L., Spouge J.L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M.W., Cowan R.S., Erickson D.L. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences. 106:12794–12797.

Chakravorty S., Helb D., Burday M., Connell N., Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods. 69:330–339.

Clark D.P., Pazdernik N.J. 2013. Molecular Biology. Academic Press Elsevier Inc.

Clarke L.J., Beard J.M., Swadling K.M., Deagle B.E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 7:873–883.

Coleman C.O. 2015. Taxonomy in times of the taxonomic impediment–examples from the community of experts on amphipod crustaceans. Journal of Crustacean Biology. 35:729–740.

Conrad K.F., Warren M.S., Fox R., Parsons M.S., Woiwod I.P. 2006. Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis. Biological conservation. 132:279–291.

Conrad K.F., Woiwod I.P., Parsons M., Fox R., Warren M.S. 2004. Long-term population trends in widespread British moths. Journal of Insect Conservation. 8:119–136.

Creer S., Deiner K., Frey S., Porazinska D., Taberlet P., Thomas W.K., Potter C., Bik H.M. 2016. The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution. 7:1008–1018.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W., Power D.M., Packer M., Carvalho G.R., Blaxter M.L., Lambshead P.J.D., Thomas W.K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology. 19:4–20.

Crooks K.R., Soulé M.E. 1999. Mesopredator release and avifaunal extinctions in a fragmented system. Nature. 400:563.

Daily G.C. 1997. Nature's services. Island Press, Washington, DC.

Darwin C. 1809-1882. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London: John Murray, 1859.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F., Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology Letters. 10:20140562.

Deiner K., Bik H.M., Mächler E., Seymour M., Lacoursière-Roussel A., Altermatt F., Creer S., Bista I., Lodge D.M., De Vere N. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology. 26:5872–5895.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 14:927–930.

Douglas W.Y., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Drinkwater R., Schnell I.B., Bohmann K., Bernard H., Veron G., Clare E., Gilbert M.T.P., Rossiter S.J. 2019. Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. Molecular Ecology Resources. 19:105–117.

Dupont Y.L., Damgaard C., Simonsen V. 2011. Quantitative historical change in bumblebee (Bombus spp.) assemblages of red clover fields. PloS one. 6:e25172.

Edgar R.C., Haas B.J., Clemente J.C., Quince C., Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 27:2194–2200.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Elbrecht V., Peinert B., Leese F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7:6918–6926.

Elbrecht V., Taberlet P., Dejean T., Valentini A., Usseglio-Polatera P., Beisel J.-N., Coissac E., Boyer F., Leese F. 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ. 4:e1966.

Ellison A.M., Bank M.S., Clinton B.D., Colburn E.A., Elliott K., Ford C.R., Foster D.R., Kloeppel B.D., Knoepp J.D., Lovett G.M. 2005. Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Frontiers in Ecology and the Environment. 3:479–486.

England L., Vincent M., Trevors J., Holmes S. 2004. Extraction, detection and persistence of extracellular DNA in forest litter microcosms. Molecular and cellular probes. 18:313–319.

Erasga D. 2012. Sociological Landscape: Theories, Realities and Trends. BoD–Books on Demand.

Erhardt A. 1985. Diurnal Lepidoptera: sensitive indicators of cultivated and abandoned grassland. Journal of Applied Ecology.:849–861.

Essl F., Dullinger S., Rabitsch W., Hulme P.E., Pyšek P., Wilson J.R., Richardson D.M. 2015a. Delayed biodiversity change: no time to waste. Trends in Ecology & Evolution. 30:375–378.

Essl F., Dullinger S., Rabitsch W., Hulme P.E., Pyšek P., Wilson J.R., Richardson D.M. 2015b. Historical legacies accumulate to shape future biodiversity in an era of rapid global change. Diversity and Distributions. 21:534–547.

Ficetola G.F., Coissac E., Zundel S., Riaz T., Shehzad W., Bessière J., Taberlet P., Pompanon F. 2010. An in silico approach for the evaluation of DNA barcodes. BMC Genomics. 11:434.

Fløjgaard C., Frøslev T.G., Brunbjerg A.K., Bruun H.H., Moeslund J., Hansen A.J., Ejrnæs R. 2019. Predicting provenance of forensic soil samples: Linking soil to ecological habitats by metabarcoding and supervised classification. PloS one. 14:e0202844.

Fonseca V., Packer M., Carvalho G., Power D., Lambshead J., Creer S. 2011. Isolation of marine meiofauna from sandy sediments From decanting to DNA extraction. Protocol Exchange.

Fonseca V.G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources. 18:923–926.

Fonseca V.G., Carvalho G.R., Sung W., Johnson H.F., Power D.M., Neill S.P., Packer M., Blaxter M.L., Lambshead P.J.D., Thomas W.K., Creer S. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications. 1.

Fonseca V.G., Nichols B., Lallias D., Quince C., Carvalho G.R., Power D.M., Creer S. 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. Nucleic Acids Research. 40:e66–e66.

Fox R. 2013. The decline of moths in Great Britain: a review of possible causes. Insect Conservation and Diversity. 6:5–19.

Frankie G.W., Rizzardi M., Vinson S.B., Griswold T.L. 2009. Decline in bee diversity and abundance from 1972-2004 on a flowering leguminous tree, Andira inermis in Costa Rica at the interface of disturbed dry forest and the urban environment. Journal of the Kansas Entomological Society. 82:1–21.

Franzén M., Johannesson M. 2007. Predicting extinction risk of butterflies and moths (Macrolepidoptera) from distribution patterns and species characteristics. Journal of Insect Conservation. 11:367–390.

Freeman B.G., Scholer M.N., Ruiz-Gutierrez V., Fitzpatrick J.W. 2018. Climate change causes upslope shifts and mountaintop extirpations in a tropical bird community. Proceedings of the National Academy of Sciences. 115:11982–11987.

Frøslev T.G., Kjøller R., Bruun H.H., Ejrnæs R., Brunbjerg A.K., Pietroni C., Hansen A.J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8:1–11.

Gardner C.M., Gunsch C.K. 2017. Adsorption capacity of multiple DNA sources to clay minerals and environmental soil matrices less than previously estimated. Chemosphere. 175:45–51.

Greenslade P. 1964. Pitfall trapping as a method for studying populations of Carabidae (Coleoptera). The Journal of Animal Ecology.:301–310.

Grosberg R.K., Vermeij G.J., Wainwright P.C. 2012. Biodiversity in water and on land. Current Biology. 22:R900–R903.

Gulden R.H., Lerat S., Hart M.M., Powell J.R., Trevors J.T., Pauls K.P., Klironomos J.N., Swanton C.J. 2005. Quantitation of transgenic plant DNA in leachate water: real-time polymerase chain reaction analysis. Journal of Agricultural and Food Chemistry. 53:5858–5865.

Hagemann S., Pröhl H. 2007. Mitochondrial paraphyly in a polymorphic poison frog species (Dendrobatidae; D. pumilio). Molecular Phylogenetics and Evolution. 45:740–747.

Haile J., Froese D.G., MacPhee R.D., Roberts R.G., Arnold L.J., Reyes A.V., Rasmussen M., Nielsen R., Brook B.W., Robinson S. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. Proceedings of the National Academy of Sciences. 106:22352–22357.

Hajibabaei M. 2012. The golden age of DNA metasystematics. Trends in Genetics. 28:535–537.

Hallmann C.A., Sorg M., Jongejans E., Siepel H., Hofland N., Schwan H., Stenmans W., Müller A., Sumser H., Hörren T., Goulson D., Kroon H. de. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLOS ONE. 12:e0185809.

Hambler C., Speight M.R. 2004. Extinction rates and butterflies. Science. 305:1563–1565.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:313–321.

Heemsbergen D., Berg M., Loreau M., Van Hal J., Faber J., Verhoef H. 2004. Biodiversity effects on soil processes explained by interspecific functional dissimilarity. Science. 306:1019–1020.

Hofreiter M., Serre D., Poinar H.N., Kuch M., Pääbo S. 2001. ancient DNA. Nature Reviews Genetics. 2:353.

Hooper D.U., Adair E.C., Cardinale B.J., Byrnes J.E., Hungate B.A., Matulich K.L., Gonzalez A., Duffy J.E., Gamfeldt L., O'Connor M.I. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature. 486:105.

Huse S.M., Welch D.M., Morrison H.G., Sogin M.L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology. 12:1889–1898.

Jenderedjian K., Hakobyan S., Stapanian M.A. 2012. Trends in benthic macroinvertebrate community biomass and energy budgets in Lake Sevan, 1928–2004. Environmental Monitoring and Assessment. 184:6647–6671.

Johnson S., Bond W. 1994. Red flowers and butterfly pollination in the fynbos of South Africa. Plant-animal interactions in Mediterranean-type ecosystems. Springer. p. 137–148.

Jones C.G., Lawton J.H., Shachak M. 1994. Organisms as ecosystem engineers. Ecosystem management. Springer. p. 130–147.

Jørgensen T., Kjaer K.H., Haile J., Rasmussen M., Boessenkool S., Andersen K., Coissac E., Taberlet P., Brochmann C., Orlando L. 2012. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Metabarcoding. Molecular Ecology. 21:1980–1988.

Kalkman V.J. 2010. European red list of dragonflies. IUCN (International Union for Conservation of Nature).

Kennedy C.M., Lonsdorf E., Neel M.C., Williams N.M., Ricketts T.H., Winfree R., Bommarco R., Brittain C., Burley A.L., Cariveau D. 2013. A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. Ecology Letters. 16:584–599.

Kenta T., Inari N., Nagamitsu T., Goka K., Hiura T. 2007. Commercialized European bumblebee can cause pollination disturbance: an experiment on seven native plant species in Japan. Biological Conservation. 134:298–309.

Kircher M., Kelso J. 2010. High-throughput DNA sequencing–concepts and limitations. Bioessays. 32:524–536.

Klein Alexandra-Maria, Vaissière Bernard E, Cane James H, Steffan-Dewenter Ingolf, Cunningham Saul A, Kremen Claire, Tscharntke Teja. 2007. Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B: Biological Sciences. 274:303–313.

Kopylova E., Navas-Molina J.A., Mercier C., Xu Z.Z., Mahé F., He Y., Zhou H.-W., Rognes T., Caporaso J.G., Knight R. 2016. Open-Source Sequence Clustering Methods Improve the State Of the Art. mSystems. 1:e00003-15.

Kosior A., Celary W., Olejniczak P., Fijał J., Król W., Solarz W., Płonka P. 2007. The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe. Oryx. 41:79–88.

Krehenwinkel H., Fong M., Kennedy S., Huang E.G., Noriyuki S., Cayetano L., Gillespie R. 2018. The effect of DNA degradation bias in passive sampling devices on metabarcoding studies of arthropod communities and their associated microbiota. PloS one. 13.

Kress W.J., García-Robledo C., Uriarte M., Erickson D.L. 2015. DNA barcodes for ecology, evolution, and conservation. Trends in Ecology & Evolution. 30:25–35.

Kucherenko A., Herman J.E., Iii E.M.E., Urakawa H. 2018. Terrestrial Snake Environmental DNA Accumulation and Degradation Dynamics and its Environmental Application. Herpetologica. 74:38–49.

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

Lallias D., Hiddink J.G., Fonseca V.G., Gaspar J.M., Sung W., Neill S.P., Barnes N., Ferrero T., Hall N., Lambshead P.J.D., Packer M., Thomas W.K., Creer S. 2015. Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. The ISME Journal. 9:1208–1221.

Lazarevic V., Whiteson K., Huse S., Hernandez D., Farinelli L., Østerås M., Schrenzel J., François P. 2009. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. Journal of Microbiological Methods. 79:266–271.

Lehmkuhl J.F., Peffer R.D., O'Connell M.A. 2008. Riparian and Upland Small Mammals on the East Slope of the Cascade Range, Washington. nwsc. 82:94–107.

Lenat D.R. 1988. Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. Journal of the North American Benthological Society. 7:222–233.

Levy-Booth D.J., Campbell R.G., Gulden R.H., Hart M.M., Powell J.R., Klironomos J.N., Pauls K.P., Swanton C.J., Trevors J.T., Dunfield K.E. 2007. Cycling of extracellular DNA in the soil environment. Soil Biology and Biochemistry. 39:2977–2991.

Lindenmayer D.B., Likens G.E. 2011. Direct Measurement Versus Surrogate Indicator Species for Evaluating Environmental Change and Biodiversity Loss. Ecosystems. 14:47–59.

Lister B.C., Garcia A. 2018. Climate-driven declines in arthropod abundance restructure a rainforest food web. Proceedings of the National Academy of Sciences. 115:E10397–E10406.

Loreau M., Naeem S., Inchausti P., Bengtsson J., Grime J.P., Hector A., Hooper D.U., Huston M.A., Raffaelli D., Schmid B., Tilman D., Wardle D.A. 2001. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. Science. 294:804–808.

Mace G.M., Norris K., Fitter A.H. 2012. Biodiversity and ecosystem services: a multilayered relationship. Trends in Ecology & Evolution. 27:19–26.

Macher J.-N., Macher T.-H., Leese F. 2017. Combining NCBI and BOLD databases for OTU assignment in metabarcoding and metagenomic datasets: The BOLD_NCBI _Merger. Metabarcoding and Metagenomics. 1:e22262.

Mardis E.R. 2013. Next-Generation Sequencing Platforms. Annual Review of Analytical Chemistry. 6:287–303.

Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.-J., Chen Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437:376.

Marquina D., Andersson A.F., Ronquist F. 2019a. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources. 19:90–104.

Marquina D., Esparza-Salas R., Roslin T., Ronquist F. 2019b. Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular Ecology Resources.

Marshall E., Brown V., Boatman N., Lutman P., Squire G., Ward L. 2003. The role of weeds in supporting biological diversity within crop fields. Weed Research. 43:77–89.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.

May R.M. 2010. Tropical Arthropod Species, More or Less? Science. 329:41–42.

Melero Y., Stefanescu C., Pino J. 2016. General declines in Mediterranean butterflies over the last two decades are modulated by species traits. Biological Conservation. 201:336–342.

Millington P.J., Walker K.F. 1983. Australian freashwater mussel Velesunio ambiguus (Philippi) as a biological monitor for zinc, iron and manganese. Marine Freshwater Research 34:873–892.

Mittermeier R.A., Myers N., Mittermeier C.G., Robles Gil P. 1999. Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions. CEMAX, S.A., Mexico City.

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., Worm B. 2011. How Many Species Are There on Earth and in the Ocean? PLOS Biology. 9:e1001127.

Morandin L.A., Long R.F., Kremen C. 2014. Hedgerows enhance beneficial insects on adjacent tomato fields in an intensive agricultural landscape. Agriculture, Ecosystems & Environment. 189:164–170.

Morrison M.L., Marcot B., Mannan W. 2012. Wildlife-Habitat Relationships: Concepts and Applications. Island Press.

Mulé R., Sabella G., Robba L., Manachini B. 2017. Systematic review of the effects of chemical insecticides on four common butterfly families. Frontiers in Environmental Science. 5:32.

Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403:853–858.

Nielsen K.M., Smalla K., van Elsas J.D. 2000. Natural Transformation of Acinetobactersp. Strain BD413 with Cell Lysates of Acinetobacter sp., Pseudomonas fluorescens, and Burkholderia cepaciain Soil Microcosms. Applied and Environmental Microbiology. 66:206–212.

Nilsson S.G., Franzen M., Jönsson E. 2008. Long-term land-use changes and extinction of specialised butterflies. Insect Conservation and Diversity. 1:197–207.

Noonan J.P., Coop G., Kudaravalli S., Smith D., Krause J., Alessi J., Chen F., Platt D., Pääbo S., Pritchard J.K. 2006. Sequencing and analysis of Neanderthal genomic DNA. Science. 314:1113–1118.

Nowrousian M., Teichert I., Masloff S., Kück U. 2012. Whole-genome sequencing of Sordaria macrospora mutants identifies developmental genes. G3: Genes, Genomes, Genetics. 2:261–270.

Ogram A., Sayler G.S., Gustin D., Lewis R.J. 1988. DNA adsorption to soils and sediments. Environmental Science & Technology. 22:982–984.

Ogram A.V., Mathot M.L., Harsh J.B., Boyle J., Pettigrew C.A. 1994. Effects of DNA polymer length on its adsorption to soils. Applied and Environmental Microbiology. 60:393–396.

Oksanen J., Kindt R., Legendre P., O'Hara B., Stevens M.H.H., Oksanen M.J., Suggests M. 2007. The vegan package. Community ecology package. 10:631–637.

Olesen S.W., Duvallet C., Alm E.J. 2017. dbOTU3: A new implementation of distribution-based OTU calling. PloS one. 12:e0176335.

Ollerton J., Erenler H., Edwards M., Crockett R. 2014. Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. Science. 346:1360–1362.

Perrings C., Folke C., Mäler K.-G. 1992. The ecology and economics of biodiversity loss: the research agenda. Ambio.:201–211.

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

Pimm S.L. 2008. Biodiversity: Climate Change or Habitat Loss — Which Will Kill More Species? Current Biology. 18:R117–R119.

Piñol J., Mir G., Gomez-Polo P., Agustí N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources. 15:819–830.

Potts S.G., Biesmeijer J.C., Kremen C., Neumann P., Schweiger O., Kunin W.E. 2010. Global pollinator declines: trends, impacts and drivers. Trends in Ecology & Evolution. 25:345–353.

Purvis A., Hector A. 2000. Getting the measure of biodiversity. Nature. 405:212.

Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H.P., Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 13:341.

Rader R., Bartomeus I., Garibaldi L.A., Garratt M.P., Howlett B.G., Winfree R., Cunningham S.A., Mayfield M.M., Arthur A.D., Andersson G.K. 2016. Non-bee insects are important contributors to global crop pollination. Proceedings of the National Academy of Sciences. 113:146–151.

Ramankutty N., Mehrabi Z., Waha K., Jarvis L., Kremen C., Herrero M., Rieseberg L.H. 2018. Trends in global agricultural land use: implications for environmental health and food security. Annual Review of Plant Biology. 69:789–815.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Rosenberg D., Resh V. 1993. Freshwater biomonitoring and macroinvertebrates. Springer US .

Roth T., Weber D. 2008. Top predators as indicators for species richness? Prey species are just as useful. Journal of Applied Ecology. 45:987–991.

Ruppert K.M., Kline R.J., Rahman M.S. 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation.:e00547.

Saeki K., Ihyo Y., Sakai M., Kunito T. 2011. Strong adsorption of DNA molecules on humic acids. Environmental Chemistry Letters. 9:505–509.

Sánchez-Bayo F., Wyckhuys K.A. 2019. Worldwide decline of the entomofauna: A review of its drivers. Biological Conservation. 232:8–27.

Sanger F., Coulson A.R., Friedmann T., Air G.M., Barrell B.G., Brown N.L., Fiddes J.C., Hutchison C.A., Slocombe P.M., Smith M. 1978. The nucleotide sequence of bacteriophage ϕX174. Journal of Molecular Biology. 125:225–246.

Sanger F., Nicklen S., Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74:5463.

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

Shendure J., Ji H. 2008. Next-generation DNA sequencing. Nature Biotechnology. 26:1135–1145.

Shokralla S., Gibson J.F., Nikbakht H., Janzen D.H., Hallwachs W., Hajibabaei M. 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources. 14:892–901.

Shokralla S., Singer G.A.C., Hajibabaei M. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. BioTechniques. 48:305–306.

Shokralla S., Spall J.L., Gibson J.F., Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology. 21:1794–1805.

Siddig A.A., Ellison A.M., Ochs A., Villar-Leeman C., Lau M.K. 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. Ecological Indicators. 60:223–230.

Simberloff D. 1998. Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? Biological Conservation. 83:247–257.

Sirois S.H., Buckley D.H. 2019. Factors governing extracellular DNA degradation dynamics in soil. Environmental Microbiology Reports. 11:173–184.

Sogin M.L., Morrison H.G., Huber J.A., Welch D.M., Huse S.M., Neal P.R., Arrieta J.M., Herndl G.J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." PNAS. 103:12115–12120.

Somervuo P., Yu D.W., Xu C.C., Ji Y., Hultman J., Wirta H., Ovaskainen O. 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution. 8:398–407.

Soulé M.E., Bolger D.T., Alberts A.C., Wrights J., Sorice M., Hill S. 1988. Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. Conservation Biology. 2:75–92.

Stork N.E., McBroom J., Gely C., Hamilton A.J. 2015. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proceedings of the National Academy of Sciences. 112:7519–7523.

Strickler K.M., Fremier A.K., Goldberg C.S. 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biological Conservation. 183:85–92.

Summers K., Cronin T.W., Kennedy T. 2003. Variation in spectral reflectance among populations of Dendrobates pumilio, the strawberry poison frog, in the Bocas del Toro Archipelago, Panama. Journal of Biogeography. 30:35–53.

Summerville K.S., Ritter L.M., Crist T.O. 2004. Forest moth taxa as indicators of lepidopteran richness and habitat disturbance: a preliminary assessment. Biological Conservation. 116:9–18.

van Swaay C., van Strien A., Harpke A., Fontaine B., Stefanescu C., Roy D., Kühn E., Õnuao E., Regan E., Švitra G. 2013. The European grassland butterfly indicator: 1990–2011. EEA Technical Reports. 11.

Taberlet P., Coissac E., Hajibabaei M., Rieseberg L.H. 2012a. Environmental DNA. Molecular ecology. 21:1789–1793.

Taberlet P., PRUD'HOMME S.M., Campione E., Roy J., Miquel C., Shehzad W., Gielly L., Rioux D., Choler P., CLÉMENT J. 2012b. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology. 21:1816–1820.

Takacs D. 1996. The Idea of Biodiversity: Philosophies of Paradise. Johns Hopkins University Press.

Tang C.Q., Leasi F., Obertegger U., Kieneke A., Barraclough T.G., Fontaneto D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences. 109:16208–16212.

Thomas C.D. 2010. Climate, climate change and range boundaries. Diversity and Distributions. 16:488–495.

Thomas C.D., Cameron A., Green R.E., Bakkenes M., Beaumont L.J., Collingham Y.C., Erasmus B.F.N., Siqueira M.F. de, Grainger A., Hannah L., Hughes L., Huntley B., Jaarsveld A.S. van, Midgley G.F., Miles L., Ortega-Huerta M.A., Peterson A.T., Phillips O.L., Williams S.E. 2004a. Extinction risk from climate change. Nature. 427:145–148.

Thomas J.A., Telfer M.G., Roy D.B., Preston C.D., Greenwood J., Asher J., Fox R., Clarke R.T., Lawton J.H. 2004b. Comparative losses of British butterflies, birds, and plants and the global extinction crisis. Science. 303:1879–1881.

Thompson R., Townsend C. 2006. A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. Journal of Animal Ecology. 75:476–484.

Thomsen P.F., Kielgast J., Iversen L.L., Møller P.R., Rasmussen M., Willerslev E. 2012. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. PLOS ONE. 7:e41732.

Thomsen P.F., Willerslev E. 2015. Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation. 183:4–18.

Tilman D. 2000. Causes, consequences and ethics of biodiversity. Nature. 405:208.

Tilman D., May R.M., Lehman C.L., Nowak M.A. 1994. Habitat destruction and the extinction debt. Nature. 371:65–66.

Tipu H.N., Shabbir A. 2015. Evolution of DNA sequencing. Journal of the College of Physicians and Surgeons - Pakistan. 25:210–5.

Turner C.R., Uy K.L., Everhart R.C. 2015. Fish environmental DNA is more concentrated in aquatic sediments than surface water. Biological Conservation. 183:93–102.

Valentini A., Taberlet P., Miaud C., Civade R., Herder J., Thomsen P.F., Bellemain E., Besnard A., Coissac E., Boyer F., Gaboriaud C., Jean P., Poulet N., Roset N., Copp G.H., Geniez P., Pont D., Argillier C., Baudoin J.-M., Peroux T., Crivelli A.J., Olivier A., Acqueberge M., Le Brun M., Møller P.R., Willerslev E., Dejean T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology. 25:929–942.

Van Vliet P., Beare M., Coleman D., Hendrix P. 2004. Effects of enchytraeids (Annelida: Oligochaeta) on soil carbon and nitrogen dynamics in laboratory incubations. Applied Soil Ecology. 25:147–160.

Vanbergen A.J., the Insect Pollinators Initiative. 2013. Threats to an ecosystem service: pressures on pollinators. Frontiers in Ecology and the Environment. 11:251–259.

Vermeij G.J., Grosberg R.K. 2010. The Great Divergence: When Did Diversity on Land Exceed That in the Sea? Integrative and Comparative Biology. 50:675–682.

Vieira C., Romero G.Q. 2013. Ecosystem engineers on plants: indirect facilitation of arthropod communities by leaf-rollers at different scales. Ecology. 94:1510–1518.

Vuillemin A., Horn F., Alawi M., Henny C., Wagner D., Crowe S.A., Kallmeyer J. 2017. Preservation and significance of extracellular DNA in ferruginous sediments from Lake Towuti, Indonesia. Frontiers in Microbiology. 8:1440.

Walker J., Wilson J. 1960. The separation of nematodes from soil by a modified Baermann funnel technique. Plant Disease Reporter. 44:94–97.

Walker K. 1981. Ecology of freshwater mussels in the River Murray. Australian Water Resources Council technical paper 63. Ecology of freshwater mussels in the River Murray. Australian Water Resources Council Technical Paper. 63:0.

Watts C., Dopheide A., Holdaway R., Davis C., Wood J., Thornburrow D., Dickie I.A. 2019. DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques. Austral Entomology.

Wheeler Q.D., Raven P.H., Wilson E.O. 2004. Taxonomy: Impediment or Expedient? Science. 303:285–285.

Whittaker R.H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. Ecological Monographs. 30:279–338.

Widmer F., Seidler R., Watrud L. 1996. Sensitive detection of transgenic plant marker gene persistence in soil microcosms. Molecular Ecology. 5:603–613.

Williams P.H., Osborne J.L. 2009. Bumblebee vulnerability and conservation world-wide. Apidologie. 40:367–387.

Woiwod I., Harrington R. 1994. Flying in the face of change: the Rothamsted Insect Survey. Long-term experiments in agricultural and ecological sciences.:321–342.

Zhang G.K., Chain F.J., Abbott C.L., Cristescu M.E. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications. 11:1901–1914.

Chapter I

Assessing Soil Biodiversity along an Elevation Gradient in the Caucasus Region using Metabarcoding – Evaluation of Methods

Abstract

The Caucasus region is known for hosting a high degree of biodiversity but its full magnitude is still unknown as biodiversity assessment studies were long time hampered by the limited local scientific infrastructure. The Georgian–German Biodiversity Center (GGBC) Initiative is aiming to close this large gap of knowledge by improving the local infrastructure e.g. through the establishment of modern molecular biodiversity assessment methods including barcoding and metabarcoding. This preliminary study highlights the power of metabarcoding to uncover levels of soil diversity in the Caucasus region, which has never been described before. We were able to prove that metabarcoding of soil samples is able to unravel changes in the degree of biodiversity along an elevational gradient. Furthermore, we outline the current challenges of metabarcoding to assess soil diversity and show that the aim of GGBC to establish a well curated and complete barcode reference database for the region is of uttermost importance to allow for a timely, cost-efficient and reliable assessment of local biodiversity. However, more work has to been done to enfold the complete power of metabarcoding for the assessment of soil invertebrate biodiversity e.g. by developing suitable primer-pairs and best-practice guidelines. To our knowledge, this is the first metabarcoding study based on eDNA from soil which compares the suitability of *in-silico* post-clustering filtering strategies for the assessment of invertebrate community composition. We show that choice of methods significantly influences assessed community patterns, whereas it would be desirable to develop best-practice guidelines to ensure

comparability of studies allowing for the detection of general ecological patterns (e.g. changes in degree of biodiversity along an elevational gradient).

Introduction

It is alarming that we are witnessing a dramatic decline in biodiversity (Butchart et al. 2010; Hallmann et al. 2017), without even knowing yet the full magnitude of biodiversity on earth (Mora et al. 2011). Estimates of total existing biodiversity range between 3 and 100 million species (May 2010) but up to today only about 1.5 million species have been described. To fill this large gap of knowledge several studies with the overreaching aim to assess national biodiversity and if necessary to describe new species, were launched all over the globe. While barcoding the German fauna and flora the 'German Barcode of Life'- Initiative (GBOL) detected several new species like the Sciaridae (Insecta: Diptera) species *Ctenosciara alexanderkoenigi* (Heller and Rulik 2016), which was found in the center of a crowded city. This is highlighting that even in countries which have long time been considered to be well studied, many species are still waiting to be described. By using Next Generation Sequencing (NGS) platforms it is now possible to assess unprecedented levels of biodiversity in a timely and cost-efficient manner, partly circumventing the shortage of taxonomic experts (Douglas et al. 2012; Beng et al. 2016; Barsoum et al. 2019; Elbrecht et al. 2019) also referred to as the taxonomic impediment (Wheeler et al. 2004). However, molecular biodiversity assessment methods heavily rely on the completeness and complexity of available reference databases. One of the major aims of the GBOL initiative was the establishment of a large publicly available reference database, storing the barcodes of the German fauna and flora. This database provides the basis for the incorporation of molecular methods for biodiversity assessment tasks within Germany also

outside of academia e.g. assessing the saprobic index for evaluation of water quality. The well-established scientific infrastructure of Germany has likely significantly contributed to the successful completion of this large project. Nevertheless, similar successful projects were also conducted in other parts of the world, but most of them had in common that the countries under investigation had already a well-established scientific infrastructure. However, biodiversity is not evenly distributed across the globe and some areas are outstanding in the number of occurring species. These areas are called biodiversity hotspots (Myers 1988, 1990; Mittermeier et al. 1999; Myers et al. 2000). Myers et al. (2000) defined 25 hotspots across the globe, which only comprise around 1.4 percent of total land surface but host around 44% of all vascular plants and 25% of all species in four vertebrate groups (Myers et al. 2000). Many theories have tried to define the reasons for the accumulation of species in these regions (Brown 2014). The majority suggests, that it is not just a single factor that triggers for the tremendously fast development of new species. Moreover, they postulate that it is a combination of several factors like the historical background, the benign character of the physical environment (Karr & Freemark 1983) and the existing biological heterogeneity (Brown 2014, Karr & Freemark 1983) that sustain the development of an enormous range of lifeforms. One of those hotspots is the Caucasus region. The Caucasus region is located between the Black Sea and the Caspian Sea and stretches over several countries including Georgia, which scientific infrastructure is still in the developmental phase. Therefore, the local fauna and flora has only been rudimentary studied and it is likely that many species are still waiting to be described. To close these gaps of knowledge the Georgian-German Biodiversity Center (GGBC) Initiative has embraced the goal to explore the magnitude and diversity of the local fauna and flora in the Caucasus area. Within the scope of GGBC a first joint project of the Ilia-State University in Tbilisi, Georgia and the Zoological Research Museum Alexander Koenig Bonn, Germany was launched in April 2018 with the aim to investigate the local biodiversity in the Kintrishi Protected Area. The park is located in southwestern Georgia (N41.75, E42.03) and covers an area of approximately 13.000

hectares, varying in altitude from 300 to almost 2500m above sea level. As a plio-pleistocene refugium (Tarkhnishvili et al. 2012) many local lifeforms are endemic, which is further featured by the extraordinary climate of the region. With 3000mm annual precipitation, Kintrishi is together with the Mtirala National Park the most humid area in the Caucasus region, supporting pristine mountain humid forests. Furthermore, subalpine and alpine belts are forming unique ecosystems at the steep slopes of the Lesser Caucasus, from dense forests to snow-capped stone deserts. The scientific investigations of the influence of height on species diversity has a long history. Darwin and Humboldt have noticed dramatic changes along an elevation gradient (Lomolino 2001). In the course of time, two main but contradicting hypotheses were framed to describe these changes. The discontinuity hypothesis states that groups of species have similar distribution areas along an environmental gradient. The transition to another group is thereby more or less clearly defined (Grytnes and McCain 2007). In contrast to that, the continuum hypothesis proposes that every species in a habitat is independently distributed along an environmental gradient and that the distribution area of the existing species are not limited by a common boundary (Grytnes and McCain 2007). During the 50ies and 60ies Whittaker conducted several studies to test for the correctness of these hypotheses (Whittaker 1952, 1960, 1967; Whittaker and Niering 1965). He ended up with the conclusion that species form independent range boundaries along an environmental gradient, supporting the continuum hypothesis. Moreover, Whittaker found that some organism groups were decreasing in diversity along an elevational gradient, while others were shaping a humped curve with maximum species richness at mid-elevation levels. Furthermore, Whittaker investigated the degree of change in species composition which he coined “beta-diversity” (Whittaker 1960). Although some findings of Whittaker had an enormous influence on the future of the research field of ecology by underlining the importance of the use of ordination techniques, his finding of the humped shape distribution of species richness along an elevational gradient, had in the course of time largely been forgotten. Thus, the assumption

that biodiversity decreases with height had widely been accepted (Grytnes and McCain 2007). It was only many years later in 1995 when Rahbek reviewed 97 papers on changes in biodiversity patterns along an elevational gradient to verify the at the time widely held assumption of biodiversity declines with increasing elevation level. Surprisingly he found that most studies describe an increase in biodiversity until a peak is reached at mid-elevation levels, followed by a decline at higher elevations (Rahbek 1995). Nevertheless, most of the reviewed studies were only focusing on certain taxa, which is also a result of the taxonomic impediment (Wheeler et al. 2004) which describes the challenges of biodiversity assessment studies arising from the shortage of taxonomic experts.

Although, the existing biodiversity in the Caucasus region is still to wide parts unknown and barcode reference databases are therefore extremely incomplete, the concept of Operational Taxonomic Units (OTUs) allows an insight to the magnitude of the existing local biodiversity and is furthermore able to mirror changes in species composition over environmental gradients (Beng et al. 2016). To assess total existing biodiversity as complete as possible the study substrate must be appropriately chosen. Commonly used sampling devices usually only target a narrowed set of taxa e.g. Malaise traps will largely miss to assess ground dwelling invertebrates like annelids. However, for aquatic habitats it has been observed that sediments serve as a sink for environmental DNA originating from all kind of organisms including Nematoda, Arthropoda and Chordata (Fonseca et al. 2017). Similar to these findings the extraction and analysis of eDNA from soil is thought to provide meaningful insights into the local terrestrial biodiversity. Soil is teeming with life and the existing local biodiversity can vary strongly already on small geographical scale as it highly depends on moisture, acidity, temperature, nutrient content, organic matter and last but not least on soils physical appearance described by texture and structure (McCredie et al. 1992; Baker et al. 1998; Curry 2004). In contrast to aquatic samples has soil been described to be a rather poor integrator of the whole local biodiversity (Taberlet et al. 2018). While for aquatic habitats many studies have been conducted with the aim to establish a best-practice guideline

(Leese et al. 2016), depending on target organisms this is still lacking for soil. Next to source material (Koziol et al. 2019), marker (Yang et al. 2013) and primer choice (Elbrecht et al. 2019), DNA extraction methods (Dopheide et al. 2019), completeness of reference library (Somervuo et al. 2017) are also *in-silico* steps (e.g. choice of filtering and clustering algorithms) known to directly influence taxonomic composition and diversity estimates (Frøslev et al. 2017). Some species show a high degree of intraspecific variations and as a result sequences belonging to the same species are clustered into more than one Operational Taxonomic Unit (OTU), leading to an artificial increase of calculated α-diversity (Frøslev et al. 2017). Furthermore, PCR and sequencing errors are known to contribute significantly to an artificial increase in number of recovered OTUs (Quince et al. 2009; Goodwin et al. 2016). To circumvent these sources of errors several algorithms are now available, which were developed to remove artifactual sequences from large-scale datasets. Most of them targeting low-quality reads (Huse et al. 2010) and chimeric sequences (Edgar et al. 2011; Schloss et al. 2011). To further decrease error rate most studies apply a second filtering step post OTU-clustering. The majority of these post-clustering filtering strategies are built on the assumption that the highest number of infrequent OTUs result from erroneous sequences (Huse et al. 2010; Kunin et al. 2010; Brown et al. 2015). As a result, OTUs with a low sequence abundance are often excluded from the dataset, although it can be expected that especially in very diverse habitats, rare but mostly infrequent species are accounting for a high proportion of the existing diversity (Nemergut et al. 2011; Frøslev et al. 2017). While some studies are following a “light” strategy by only excluding singletons, meaning OTUs that encompass only a single sequence, some studies cut off all OTUs which do not account for a random chosen proportion of total sequence count per sample. Both methods have in common that mostly OTUs of dominant species are retained, while rare species often remain undetected (Frøslev et al. 2017). Recently a post-clustering algorithm has been developed which is following a different strategy, LULU. The LULU algorithm identifies ‘daughter-OTUs’ with consistently co-occuring sequence similar, but more abundant ‘parent’ OTUs (Frøslev et al. 2017). The algorithm

assumes that the ‘daughter’ OTUs are artefacts and should therefore be merged with the ‘parent’ OTUs.

Aims of the Study

To our knowledge, up to today no metabarcoding study has been conducted in the Georgian Caucasus area. Species lists and biodiversity estimates for that region are solely based on traditionally generated data and computer models. Here, we are presenting the first metabarcoding data investigating the hypothesis that biodiversity is shaping a humped curve over an elevation gradient in a biodiversity hotspot. We hypothesize that most of the discovered OTUs won't be assigned to any taxon, as many species and even genera and families existing in that region have not yet been described and have therefore not been barcoded. Furthermore, we investigate the influence of *in-silico* filtering methods on taxa detection rate.

Material and Methods

Sampling Strategy

26 soil samples were collected in Kintrishi Protected Area in south-western Georgia (N41.75, E42.03) (figure I.1). The sample sites were distributed over a height gradient encompassing six elevation levels (EL) ranging from 325m to 2450m above sea level (figure I.2, table I.1).

Figure I.1 Location of Kintrishi Protected Area in south-western Georgia (N41.75, E42.03)

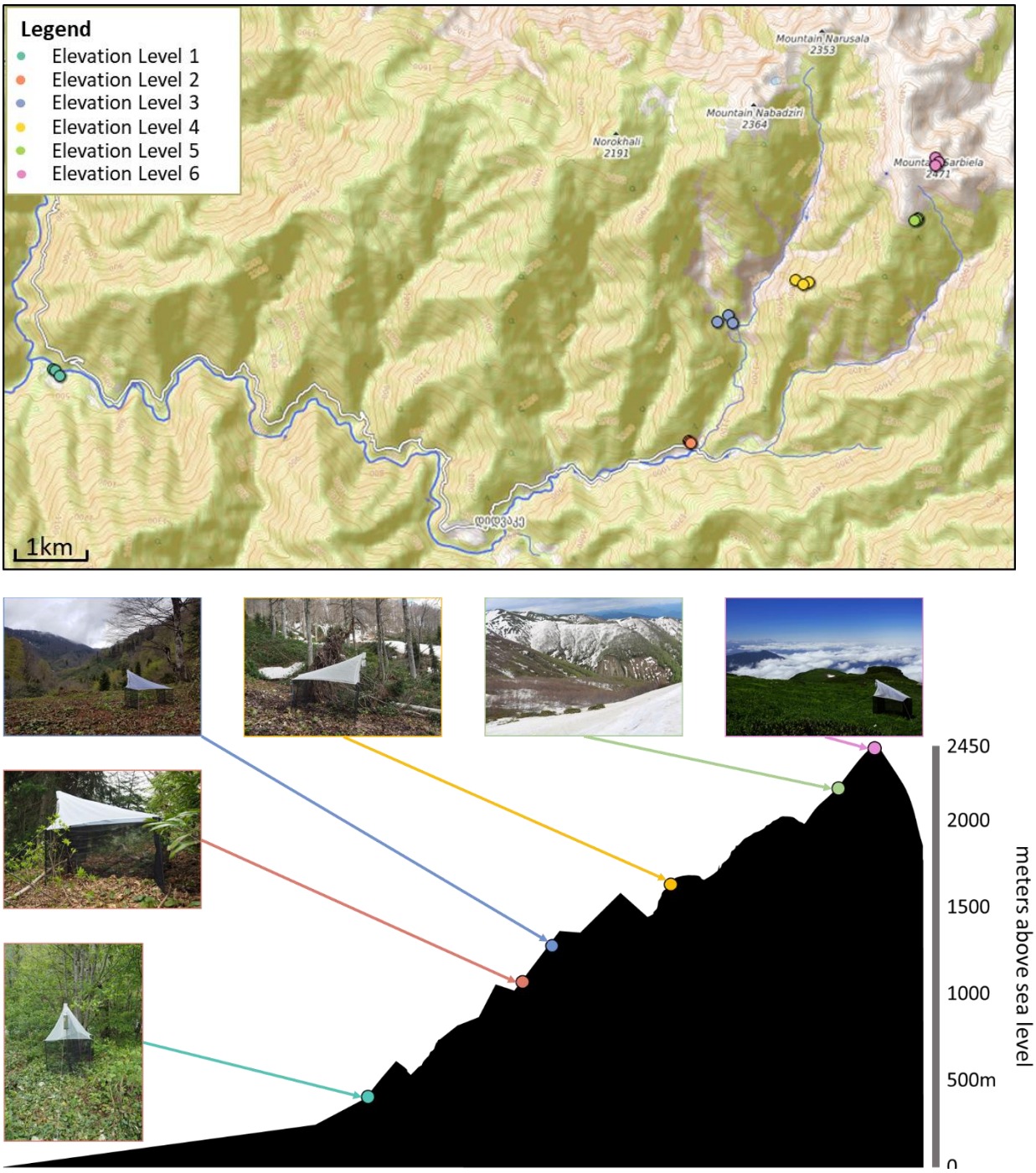


Figure I.2 Location of sampling sites in Kintrishi protected Area in south-western Georgia. In total six elevation levels were sampled, here highlighted in coloration of sample site.

Table I.1 Geographical characteristics of the 18 sampling sites. Depicted for each sampling site is the height (vertical distance from sea level in meters), elevation level (EL1-EL6), coordinates (altitude N and latitude E) and sampling dates (day.month.year).

Sample Site	Height [m]	Elevation Level	Coordinates [GMS]	Sample Date
Sample Site 07	404	EL1	41°44'13.7"N 41°58'45.2"E	21.04.2018
Sample Site 08	403	EL1	41°44'16.0"N 41°58'42.9"E	21.04.2018
Sample Site 09	401	EL1	41°44'16.2"N 41°58'41.6"E	21.04.2018
Sample Site 01	1035	EL2	41°43'45.7"N 42°04'39.3"E	20.04.2018
Sample Site 02	1020	EL2	41°43'46.0"N 42°04'38.9"E	20.04.2018
Sample Site 03	1031	EL2	41°43'46.4"N 42°04'38.2"E	20.04.2018
Sample Site 04	1264	EL3	41°44'38.9"N 42°05'00.3"E	20.04.2018
Sample Site 05	1252	EL3	41°44'36.3"N 42°04'54.4"E	20.04.2018
Sample Site 06	1235	EL3	41°44'36.0"N 42°05'02.7"E	20.04.2018
Sample Site 10	1697	EL4	41°44'52.8"N 42°05'45.3"E	24.04.2018
Sample Site 11	1637	EL4	41°44'51.7"N 42°05'42.5"E	24.04.2018
Sample Site 12	1634	EL4	41°44'53.5"N 42°05'38.4"E	24.04.2018
Sample Site 13	2268	EL5	41°45'18.6"N 42°06'45.0"E	02.06.2018
Sample Site 14	2280	EL5	41°45'19.1"N 42°06'46.2"E	02.06.2018
Sample Site 15	2280	EL5	41°45'19.4"N 42°06'46.7"E	02.06.2018
Sample Site 16	2465	EL6	41°45'44.9"N 42°06'56.6"E	16.06.2018
Sample Site 17	2462	EL6	41°45'43.0"N 42°06'58.4"E	16.06.2018
Sample Site 18	2450	EL6	41°45'41.8"N 42°06'56.9"E	16.06.2018

At each sampling site three soil samples (biological replicates) were taken. The three soil samples were collected between approximately 20m and 50m apart from each other. For sampling a 44mm id stainless steel sampler was used. Each biological sample comprised four pooled 44mm

diameter x 100mm soil samples, taken two meters apart from each other. After sampling soil cores were stored at -20°C until further processing. With the exception of elevation level 5, climate stations were installed at each height level. Every hour the mean soil and air temperature [°C] as well as the mean relative air humidity [%] of the last 60 minutes were locked. Next to this, climate station 3 located at elevation level 3 collected data of the local solar radiation [W/m²]. Unfortunately, climate station 3 failed 4 hours after set-up, leading to insufficient data.

DNA Extraction

DNA extraction was performed using the Macherey Nagel NucleoSpin® Soil kit, following manufacture's protocol.

Choice of Primers and Library Preparation

In order to increase taxonomic coverage and resolution a nuclear and a mitochondrial marker were chosen. The nuclear 18S rRNA marker which can detect a high percentage of the existing invertebrate groups as it has a high taxonomic coverage together with the more variable mitochondrial cytochrome oxidase subunit I (COI) marker, which enables a better taxonomic resolution but at the cost that several groups are missed (Tang et al. 2012; Cowart et al. 2015). To amplify the 313bp of the COI barcoding region, the primers mlCOIintF *(5'*-ACACTCTTTCCCTACACGACGCTCTTCCGATCT**GGWACWGGWTGAACWGTWTAYCCYCC** *-3')* (Leray et al. 2013) and dgHCO2198 *(5'*-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

TAAACTTCAGGGTGACCAAARAAYCA-*3')* (Leray et al. 2013) were used. To amplify the 380bp V4 region of the nuclear 18S rDNA, the primers TAReuk454FWD1 *(5'-*ACACTCTTTCCCTACACGACGCTCTTCCGATCT **CCAGCASCYGCGGTAATTCC**-*3')* and TAReukREV3r *(5'-*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **ACTTTCGTTCTTGATYRA**-*3'*) were chosen (Stoeck et al. 2010). For library preparation a two-step PCR amplification procedure was done (figure I.3).

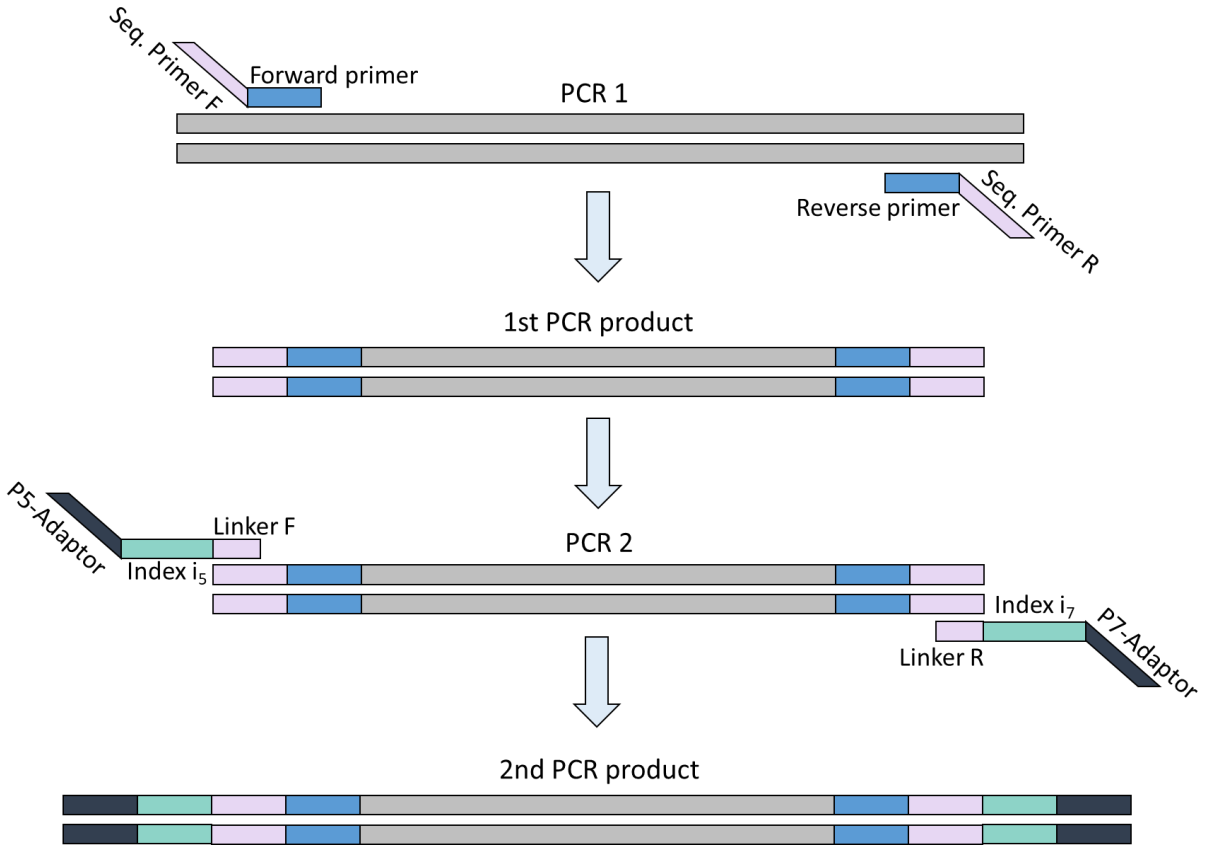


Figure I.3 Illustration of the 2-step PCR approach

Approximately 10ng of template DNA was used for all PCR reactions. For PCR1 the mastermix consisted of 7.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1µl Sigma H2O, 0.5ul of forward primer, 0.5µl of reverse primer, 0.5µl Bovine Serum Albumin (thermoscientific) and 1µl template DNA, making up a total of 15µl. The first PCR (PCR1) conditions consisted of an initial denaturation step of 2 min at 98°C, followed by 20 cycles with 40 s at 98°C, 40 s at 45°C, 30 s at 72°C (COI) or 20 cycles with 40 s at 98 °C, 40 s at 55 °C, 30 s at 72 °C (18S), and a final extension of 3 min at 72°C. Following PCR1 the PCR products were purified with HT ExoSAP-IT^{TM} (appliedbiosystems) by adding 4µl of HT ExoSAP-IT^{TM} to each sample. Samples were first heated up for 15 min at 37°C, followed by 15 min at 80°C and subsequently cooled down for 5 min at 4°C. To add the Illumina index tag adaptors a second PCR (PCR2) using 8µl of purified PCR1 products was performed. For PCR2 the purified PCR products were split onto two PCR tubes. Each tube contained 12.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 3ul Sigma H2O, 1.2µl of forward primer, 1.2µl of reverse primer and 8µl purified PCR1 product. The PCR2 conditions consisted of an initial denaturation of 2 min at 98°C, followed by 20 cycles with 40 s at 98°C, 30 s at 55°C, 30 s at 72°C and a final extension of 3 min at 72°C. PCR2 products were visualised by gel electrophoresis and purified using the QIAquick Gel Extraction Kit (Qiagen), according to manufacturer's instructions. All final purified amplicons (PCR2) were quantified using the Quantus Fluorometer (Promega) and diluted to the same concentration (3 ng/µl) before being pooled together to create two amplicon libraries (18S and COI). The resulting purified amplicon pools were sequenced on two runs of Illumina Miseq (2x 300bp) sequencing platform at Liverpool University's Centre for Genomic Research (Liverpool, UK).

Bioinformatics and Data Analysis

Preparation of OTU Table

Data sequenced at the Centre of Genomic Research (Liverpool, UK) had already undergone first quality check: The raw fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1. Furthermore, sequences were trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 20bp after trimming were removed.

The remaining fastq sequences were checked for the presence of the COI and 18S primers with Cutadapt version 1.18 (Martin 2011) using the following settings: maximum error rate (-e): 0.1, minimum Overlap (-O): 20, minimum sequence length (-m): 50). Sequences lacking either forward or reverse primer were removed. From the retained sequences detected primers were trimmed before paired-end reads were merged with vsearch version 2.7.0 (Rognes et al. 2016). Merged sequences with a length of 360-400bp for the 18S and 293-333bp for the COI dataset respectively were retained for further analysis and filtered with a maxEE threshold of 1.0 using vsearch (version 2.7.0) (Rognes et al. 2016). Subsequently fastq-sequences were demultiplexed using the script split_libraries_fastq.py implemented in QIIME1 (Caporaso et al. 2010). A phred quality threshold of 19 was chosen. Dereplication, size sorting, denovo chimera detection as well as OTU clustering with a 97% cutoff was conducted with vsearch 2.7.0 (Rognes et al. 2016). Finally, an OTU table was build by using the --usearch_global function in vsearch 2.7.0 (Rognes et al. 2016) followed by the python script “uc2otutab.py” written by Robert Edgar (https://drive5.com/python/uc2otutab_py.html). Taxonomy assignment was performed with blastn (version 2.9.0) (Altschul et al. 1990). For the COI Dataset next to the BOLD database (including barcodes of Annelida, Arthropoda, Chordata, Mollusca, Nematoda and Tardigrades), the complete GBOL database was used as reference database. For the 18S dataset

the sequences and taxonomy files were downloaded according to these criteria: ((18S) OR V4 AND ((animals[filter] OR fungi[filter] OR plants[filter]))). To investigate whether the post-clustering filter algorithm had an influence on the community composition, two filter protocols were applied.

In-silico Filter: 0.01-Filter

The following R script written by Vasco Elbrecht was applied to cut off the tail of the OTU rank-abundance curve (i.e. low alpha-diversity) at 0.01% abundance:

```
data <- read.csv("5_OTU_table_0.01_taxonomy.csv", sep = ";", header = T)
ncol(data)
data <- data[,-1]
data <- data[-c(2:10)]
ncol(data)
data <- data[-c(107)]
mysum <- colSums(data[,-1])
ncol(mysum)
data2 <- data[1:106]
for (i in 1:105){data2[,i+1] <- (data[,i+1])/(mysum[i])*100}
colSums(data2[,-1])
meep <- data2[,-1]
ncol(meep)
meep[,-1][meep[,-1]<0.01] <- 0
temp <- meep[, 1:105]
write.table(temp, sep=",", row.names=FALSE, file="OTU_Table_001_percentages.txt")
data_perc <- read.csv("OTU_Table_001_percentages.txt", stringsAsFactors=F)
```

```
ncol(data_perc)
data3 <- data[,2:106] * (data_perc [,1:105] > 0)
data4<-cbind(data$ID, data3)
write.table(data4, sep=",", row.names=FALSE, file="OTU_Table_001_filtered.csv")
```

To be considered a valid OTU for each sample tested, the number of sequences assigned to each OTU had to account for at least 0.01% of the total number of sequences found within the sample. If the number of sequences was lower the corresponding OTU was considered invalid and was removed from the sample. If the OTU was considered invalid for all tested samples, it was entirely removed from the OTU table.

In-silico Filter: Lulu curation

The second post-clustering filter approach tested was the Lulu-algorithm (Frøslev et al. 2017) Curation was started with an initial blasting of OTU representative sequences against each other using blastn (version 2.9.0). The following parameter settings were chosen for both datasets (COI and 18S): 'query coverage high-scoring sequence pair percent' (-qcov_hsp_perc) was set to 80, meaning that a sequence was reported as match when 80% of the query formed an alignment with an entry of the reference file. Secondly 'minimum percent identity' (-perc_identity) was set to 84, requiring the reference and query sequence to match at least to 84% to be reported as a match. The format of the output file was customized using the –outfmt settings '6 qseqid sseqid pident'. The resulting output file included the names of the query sequences and the names of the reference sequences next to percentage of identical match. Subsequently, the resulting list was uploaded into R (version 3.5) (R CoreTeam 2013) in order to apply the R-package 'lulu' (version 0.1.0) (Frøslev et al. 2017) for post-clustering filtering using standard settings. The LULU algorithm filters the dataset for suspicious OTUs. Afterwards, suspicious OTUs are either

classified as “daughter OTU” and merged with the corresponding “parent OTU” or are discarded from that dataset.

Diversity and Community Analysis

The resulting curated OTU tables were loaded into Excel where data got cleaned up and formatted for upload into R (R CoreTeam 2013). For statistical analysis several R packages were used. We used R studio running R version 3.5. Rarefaction curves were calculated in R using the packages vegan (version 2.5-6) (Dixon 2003) and ggplot2 (version 3.2.1.) (Wickham 2016). Venn diagrams showing number and proportion of unique and shared OTUs between technical replicates (figure I.5a,b) and biological replicates (figure I.6a,b) respectively were prepared using the R package dplyr (version 0.8.3) (Wickham et al. 2015) and the R package Venn.Diagram (version: 1.6.20) (Chen and Boutros 2011). The resulting plots were modified using Microsoft PowerPoint. Similar to that Venn diagrams showing number of shared and unique OTUs assigned to Arthropoda between post-clustering filter strategy depending on marker and reference database used (figure I.12) were prepared. Barplot showing number of detected OTUs depending on elevation level and *in-silico* filter (figure I.7) as well as barplots visualizing differences in taxonomy assignment rate depending on marker, *in-silico filter* and database (figure I.13-I.16) were prepared with ggplot2 (Wickham 2016) and later modified using PowerPoint (Microsoft). UpsetR-Plots showing detailed number of unique and shared OTUs between elevation level (figure I.8a,b) were prepared using the R package UpSetR (version 1.4.0) (Conway et al. 2017). Heatmaps visualizing calculated jaccard similarity indices between elevation level (figure I.9) were prepared using the R package ggplot2 (Wickham 2016). Underlying calculation were done in Microsoft Excel based on the formula: $J(X,Y) = |X \cap Y| / |X \cup Y|$. PCoA plots indicating differences in assessed community composition for the COI and 18S dataset (figure I.10) were prepared using the R package betapart (version 1.5.1)

(Baselga and Orme 2012) and vegan (version 2.5-6) (Dixon 2003). Calculated Marioko plots visualizing number of assignments and corresponding BlastID (figure I.11) were prepared using the R package ggplot2 (Wickham 2016).

Results

A total of 10,895,629 sequences were generated using the Illumina Miseq sequencing platform which contained both COI primers and matched the required length of 293-333bp. After paired-end merging, further quality checking and chimera removal, the COI dataset comprised a total of 4,706,004 sequences. For the 18S dataset a total of 11,115,493 sequences were found containing the 18S rRNA region of interest indicated by the presence of both primers and a query length of 360 – 400bp. After pair-end read merging, chimera removal and further quality filtering a total of 8,965,469 sequences were retained in the dataset.

In the next step, retained sequences of both datasets were clustered into Operational Taxonomic Units (OTUs) with a cutoff of 97%. The 4,706,004 sequences of the COI dataset were clustered into 23414 OTUs, while the 8,965,469 18S sequences were split onto 131399 OTUs. Both post-clustering filter algorithm strongly reduced the number of recovered OTUs. In detail, LULU curation reduced the number of OTUs contained in the COI dataset by approximately 20% from 23414 OTUs to 18785. For the 18S dataset, LULU curation resulted in an even stronger reduction by approximately 70%. Out of the formerly 131399 OTUs, 39066 OTUs were retained in the 18S dataset. The application of the 0.01-Filter lead to a slightly stronger reductions in number of OTUs within both datasets. Number of COI OTUs fell from 23414 to 11457 OTUs after post cluster-filtering curation. This accounts for a reduction of almost 51%. Number of OTUs of the 18S were reduced by 79% from formerly 131399 OTUs to 27122 OTUs.

Suitability and Efficiency of Sampling Strategy

The conducted rarefaction curves did not reach an asymptote, which is indicating that sampling effort was not suitable for assessing total existing biodiversity. This is consistent for both markers and *in-silico* filters (figure I.4)

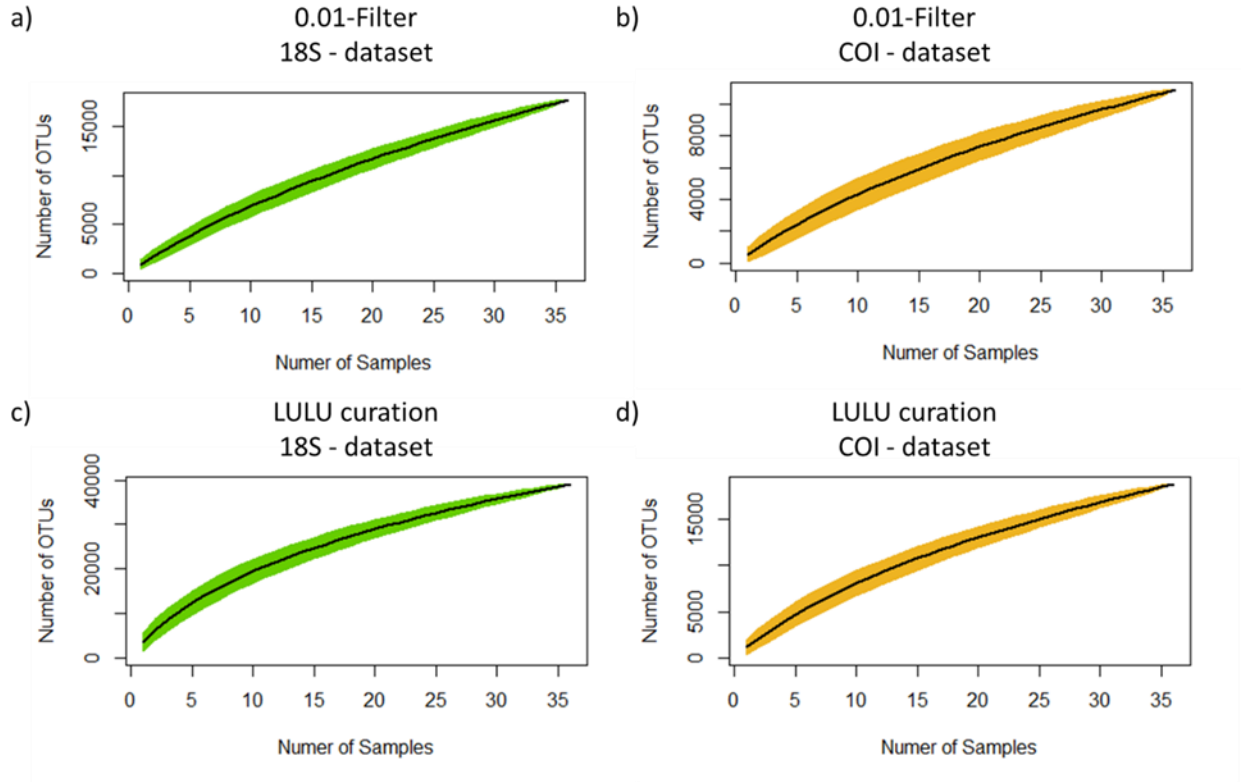


Figure I.4 Rarefaction curves calculated for the a) 18S dataset curated with the 0.01-Filter, b) COI dataset curated with the 0.01-Filter, c) 18S dataset curated with the LULU-algorithm, d) COI dataset curated with the LULU-algorithm

The two technical replicates taken of each sampling site were strongly differing in number of OTUs. This was especially true for the 18S dataset: It was observed that two technical replicates were differing in numbers by up to 691 OTUs. Additionally it was found, that only a small proportion of OTUs was present in both technical replicates (from here on referred to as 'shared OTUs) (figure I.5a and I.5b). A significant influence of filtering strategy on relative number of shared OTUs between technical replicates was observed. For the 18S dataset it was found that LULU curation increased average percentage of shared OTUs between technical replicates in comparison to the 0.01 Filter. Only at sample site 10, where technical replicates of the LULU curated dataset shared approximately 22.5% of OTUs a higher percentage was found when the same dataset was curated with the 0.01 Filter (35.5%). At the remaining 17 sampling sites, on average 13.79% of OTUs were shared between technical replicates when the 0.01-Filter was used for curation, while after LULU curation approximately 22.64% of OTUs were present in both replicates (figure I.5a).

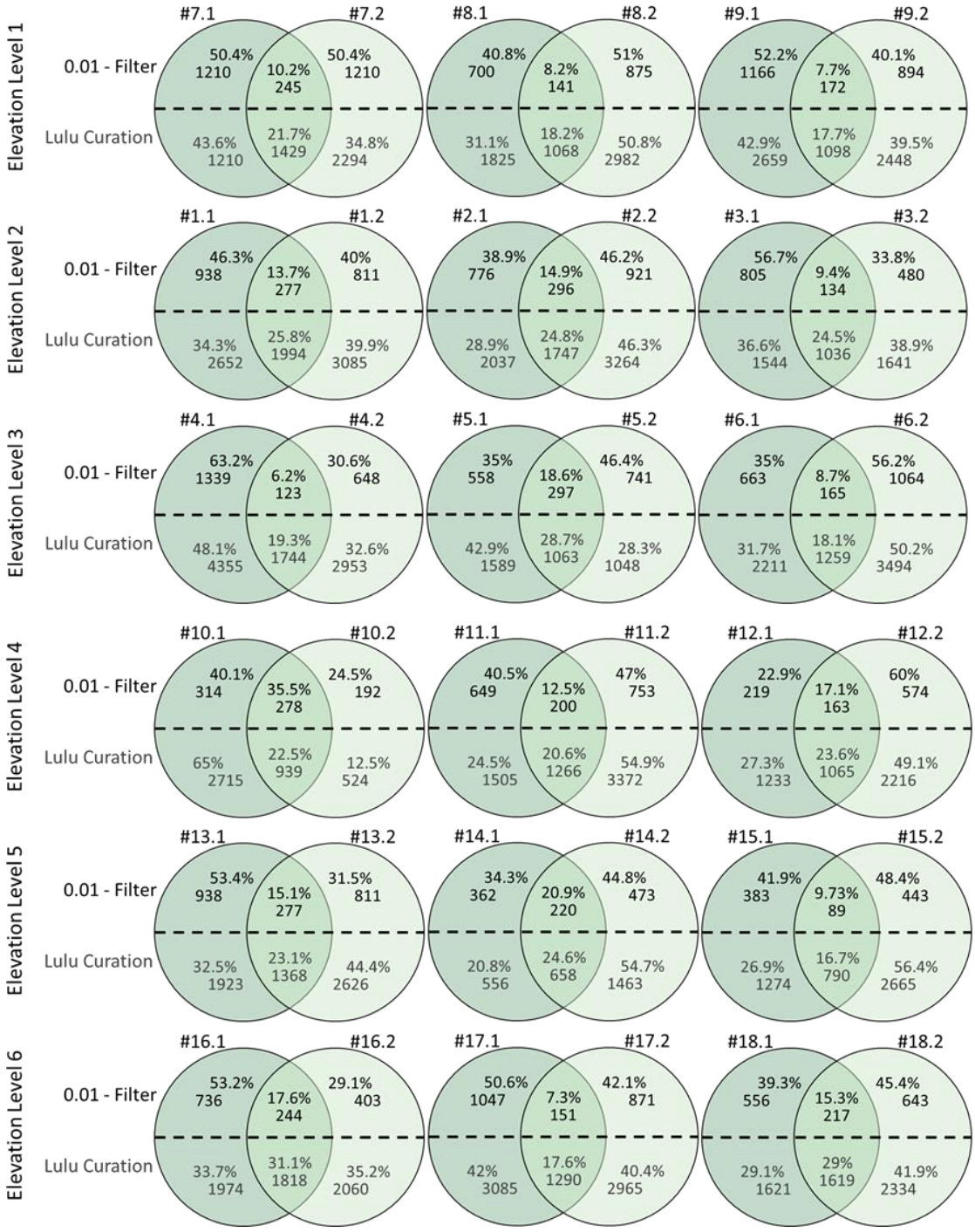


Figure I.5a Number and proportion of unique and shared OTUs between technical replicates of the 18S dataset, depending on curation method used. At each of the six elevation levels three biological replicates were taken (shown in columns, indicated by cardinals).

Figure I.5a (Continued.) Each biological replicate consisted of two technical replicates (indicated as ordinals) here shown as overlapping circles. The graph shows number and relative proportion of OTUs with which each technical replicate accounts to total number of OTUs per biological replicate. It further indicates number of OTUs detected with both technical replicates per biological replicate of each elevation level. The dashed line separate the two types of *in silico* analysis (0.01-Filter and LULU curation)

Similar results were found for the COI dataset. The two technical replicates amplified with the COI maker usually shared more OTUs when the LULU algorithm was used for post-clustering curation (LULU curation: 12.7% of OTUs; 0.01-Filtering: 11.53% of OTUs). Only for sample sites 5 (Lulu: 11.8%; 0.01: 12.2%), 10 (Lulu: 15%; 0.01: 24.8%), 14 (Lulu: 11.4%; 0.01: 14.3%) and 18 (Lulu: 19.4%; 0.01: 21.4%) a higher proportion of OTUs was shared between replicates when the 0.01-Filter was applied (figure I.5b). Regardless of applied post-clustering filter algorithm, proportion of OTUs shared between technical replicates of each sample site was higher for the 18s compared to the COI dataset. In detail, replicates amplified with the 18S primer pair and filtered with the 0.01-Filter shared on average 13.82% of OTUs, while replicates amplified with the COI primer pair shared on average 11.54% of all OTUs. When the Lulu algorithm was applied replicates of the 18S dataset shared on average 22.6% of OTUs, while on average 12.71% of the COI OTUS were detected in both replicates

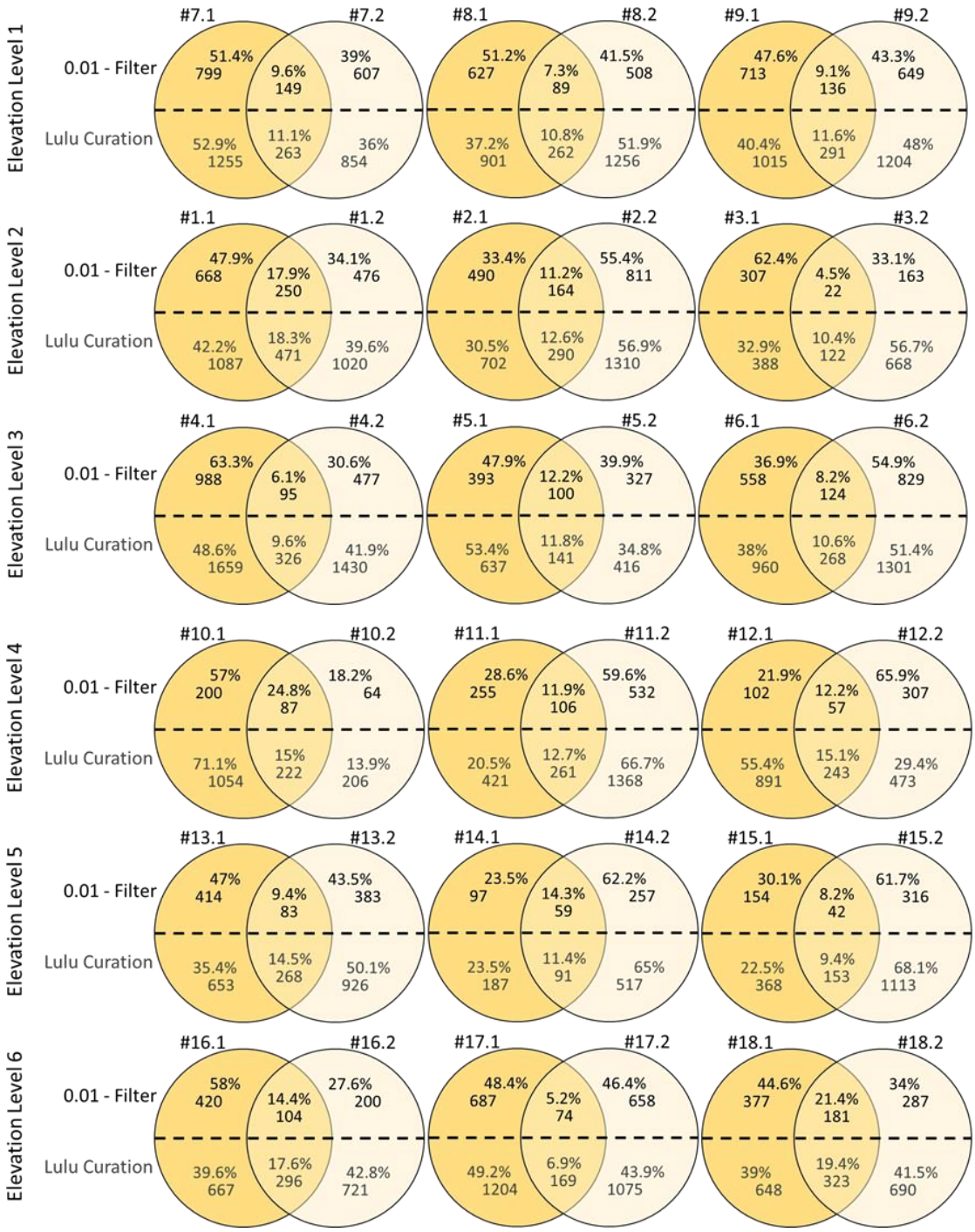


Figure I.5b Number and proportion of unique and shared OTUs between technical replicates of the COI dataset, depending on curation method used. At each of the six elevation levels

Figure I.5b (Continued.) three biological replicates were taken (shown in columns, indicated by cardinals). Each biological replicate consisted of two technical replicates (indicated as ordinals) here shown as overlapping circles. The graph shows number and relative proportion of OTUs with which each technical replicate accounts to total number of OTUs per biological replicate. It further indicates number of OTUs detected with both technical replicates per biological replicate of each elevation level. The dashed line separate the two types of *in-silico* analysis (0.01-Filter and LULU curation)

For the COI dataset the number of unique and shared OTUs between biological replicates of each of the six elevation level showed, that regardless of the applied filter algorithm, less than 15% of all recovered OTUs were shared between all three biological replicate (figure I.6a) taken at each elevation level. The relative number of shared OTUs between biological replicates was significantly higher when using the Lulu algorithm for post-clustering filtering (paired-sampled t-test: df=5, p=0.013). On average 6.91% of OTUs of the COI dataset filtered with the LULU algorithm were shared between all three sampling sites of each elevation level, while only 4.94% of OTUs were shared between biological replicates when the 0.01-Filter was used for post-clustering curation.

Within the 18S dataset up to 20.4% of OTUs were shared between the three biological replicates of each elevation level (figure I.6b). However, it was observed that relative number of shared OTUs varied strongly between applied filter algorithm depending on elevation level. *In-silico* filter had a significant influence on proportion of shared OTUs between biological replicates of each elevation level (paired-sampled t-test: df=5, p<0.001). For elevation level 2, 3 and 6 percentage of shared OTUs was more than three times higher when the dataset was curated with the LULU algorithm. For the remaining three elevation level percentage of shared OTUs at least doubled with LULU curation. On average biological replicates of each elevation

level shared 5.57% of all OTUs retained with the 0.01- Filter, while 13.88% of LULU OTUs were shared between biological replicates of each elevation level.

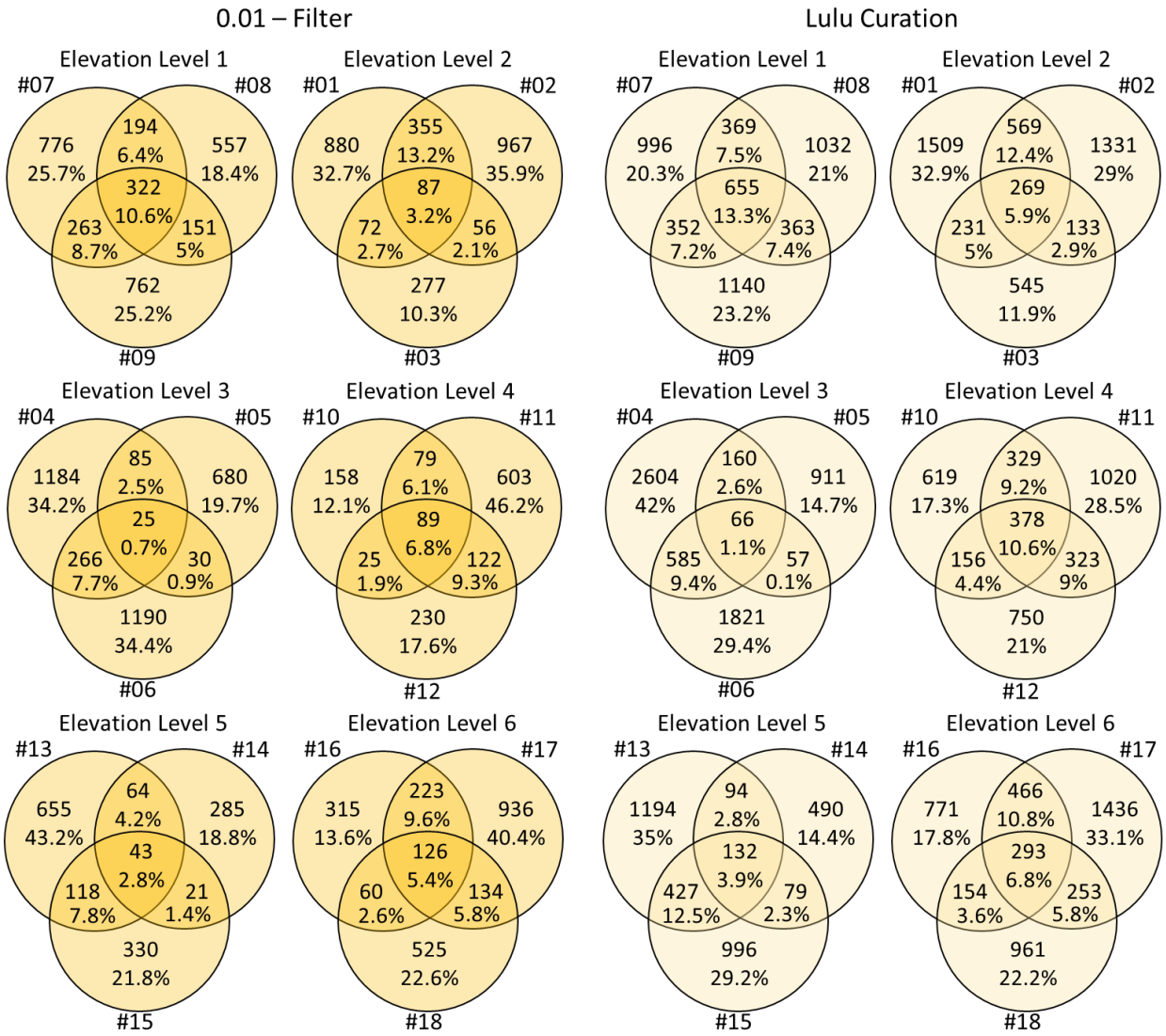


Figure I.6a Venn diagram showing number of unique and shared OTUs of the COI – dataset between biological replicates of each elevation level depending on applied *in-silico* filter algorithm.

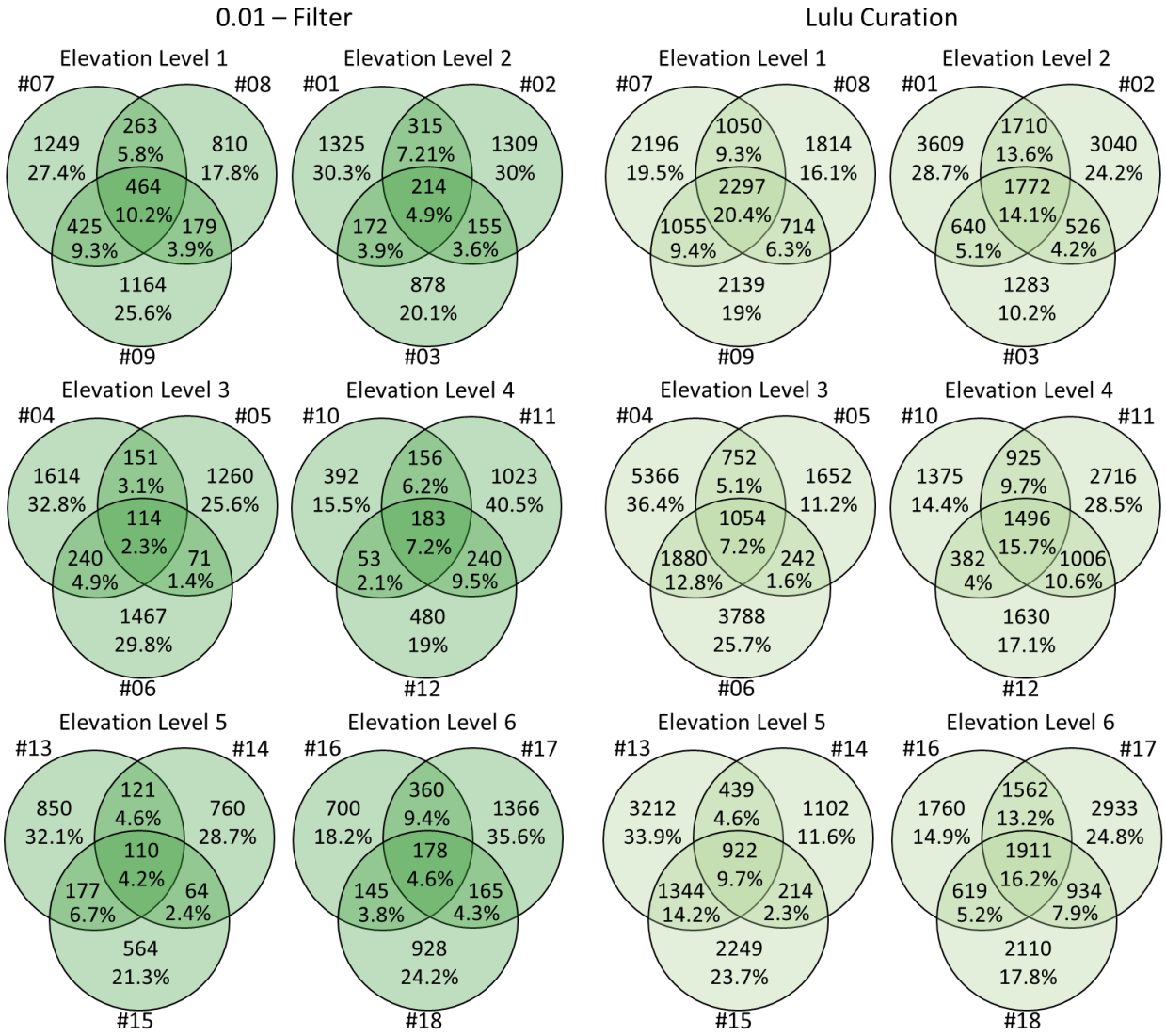


Figure I.6b Venn diagram showing number of unique and shared OTUs of the 18S – dataset between biological replicates of each elevation level depending on applied *in-silico* filter algorithm.

Biodiversity Patterns and Shifts along an Elevational Gradient

Within both datasets, number of detected OTUs were strongly varying between elevation levels (figure I.7). Regardless of curation method, the highest diversity was found at elevation level 3 (1030m a.s.l.) (figure I.7). Based on the dataset filtered with the 0.01-Filter lowest degree of diversity was found at elevation level 4 (1630m a.s.l.), while with Lulu-curation elevation level 5 was identified to host the lowest degree of biodiversity. This finding was consistent over both datasets (18S & COI). Furthermore, both datasets had in common, that degree of diversity at elevation level six was exceeding the one detected at elevation level 4 and 5.

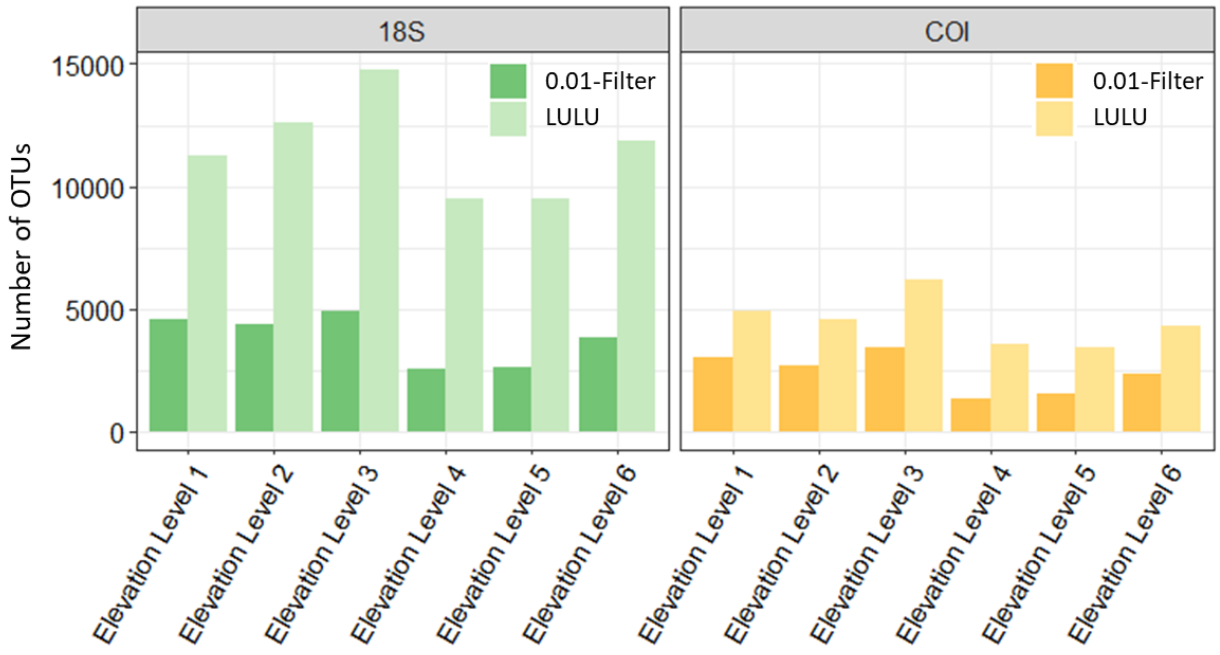


Figure I.7 Number of OTUs recovered from each elevation level depending on marker and applied post-clustering filter algorithm (0.01-Filter and Lulu curation). Color codes apply to marker-*in-silico* filter combinations

When curating the two datasets with the 0.01-Filter between 53.6% and 74.9% of OTUs were exclusively recovered from a single elevation level (from here on referred to as unique OTU) (figure I.8.1). Within the 18S dataset the highest number of unique OTUs was recovered from elevation level 1 (3285 OTUs), followed by elevation level 3 (3194 OTUs), elevation level 2 (2692 OTUs) and finally elevation level 6 (2418 OTUs). With only 1473 unique OTUs the lowest number was found at elevation level 4 and 5 (figure I.8.1a). Only 80 OTUs were shared between all six elevation levels. For the COI dataset a slightly differing pattern was observed. Elevation level 3 showed a slightly higher number of unique OTUs (2174 OTUs) compared to elevation level 1 (2154 OTUs). The next highest number of unique OTUs was found at elevation level 2 (1475 OTUs), followed by elevation level 6 (1357 OTUs), elevation level 5 (741 OTUs) and finally elevation level 4 (678 OTUs) (figure I.8a). The comparatively small number of 19 OTUs were detected at all six elevation levels (figure I.8.1b).

For the two marker datasets curated with the LULU algorithm, a similar pattern as found. The highest number of unique OTUs was found at elevation Level 3 (18S: 5542; COI: 3354 OTUs) while the lowest was found at elevation Level 5 (18S: 2621 OTUs; COI: 3354) (figure I.8.2). From all identified OTUs per dataset (18S: 39066 OTUs; COI: 18785 OTUs) a high percentage was exclusively found at a single elevation level. This applied for 62.01% of OTUs of the 18S dataset (figure I.8.2a) and 72.95% of all OTUs of the COI dataset (figure I.8.2b).Number of OTUs detected at all six elevation level varied strongly between the two marker dataset. With 1110 OTUs (2.8%) significantly more OTUs of the 18S were present at all six elevation level, while only 74 OTUs of the COI dataset (0.4%) were independently of elevation level found.

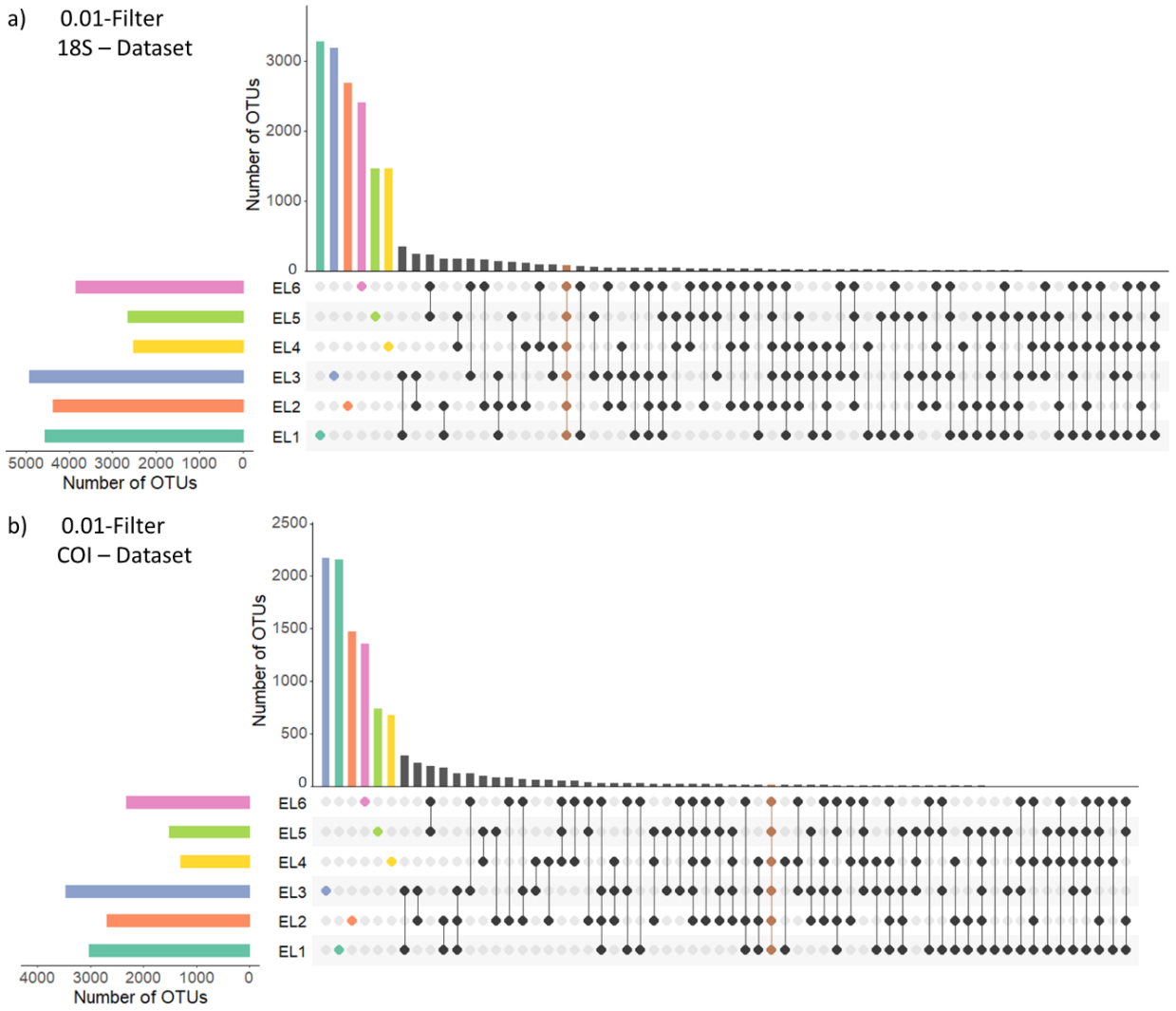


Figure I.8.1 Upset R-plots showing number of unique and shared OTUs between elevation levels for the a) 18S and b) COI dataset which were curated with the 0.01-Filter. Each row of the Upset R-plot corresponds to one elevation level (EL=Elevation level), while each column indicate a set consisting of a unique combinations of elevation levels. Light grey circles indicate that the elevation level was not part of the set, while black and colored points indicate a participation of the elevation level. Bars on top of each column indicate number of OTUs per set. Colored bars and dots indicate that the corresponding number of OTUs was exclusively recovered from one elevation level. Number of OTUs detected at all six elevation levels are highlighted in brown as well as the corresponding set.

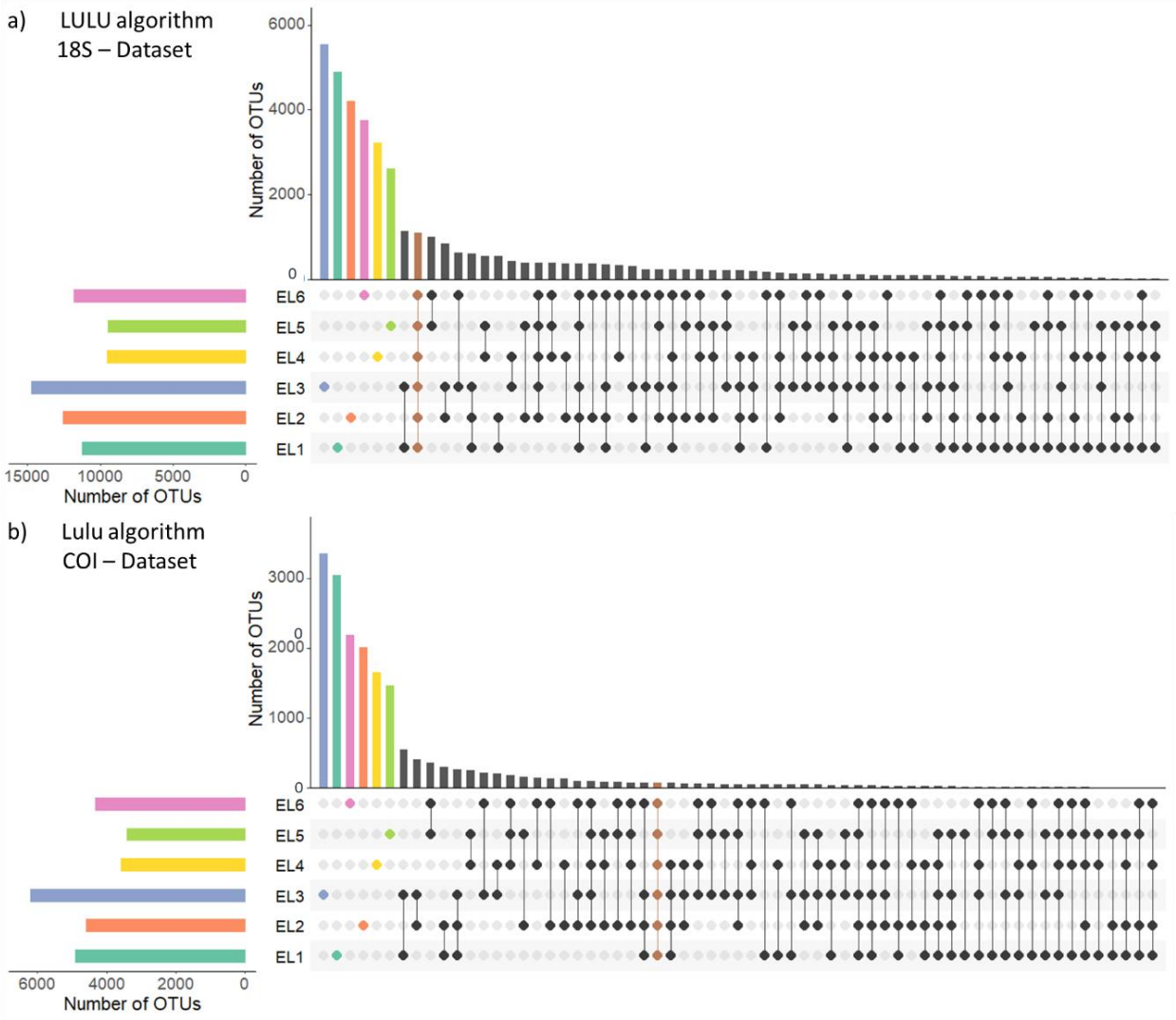


Figure I.8.2 Upset R-plots showing number of unique and shared OTUs between elevation levels for the a) 18S and b) COI dataset which were curated with the LULU algorithm. Each row of the Upset R-plot corresponds to one elevation level (EL = Elevation level), while each column indicate a set consisting of a unique combinations of elevation levels. Light grey circles indicating that the elevation level was not part of the set, while black and colored points indicate a participation of the elevation level. Bars on top of each column indicate number of OTUs per set. Colored bars and dots indicate that the corresponding number of OTUs was exclusively recovered from one elevation level. Number of OTUs detected at all six elevation levels are highlighted in brown as well as the corresponding set.

In order to analyze similarity levels of beta-diversity between the different elevation levels the Jaccard-similarity index was calculated. Jaccard-similarity indices were higher for the 18S dataset (0.04 up to 0.29) compared to the ones calculated for the COI dataset (0.02 up to 0.17). The on average calculated Jaccard-similarity indices for the COI dataset was depending on post clustering filter algorithm with 0.07 and 0.11 slightly lower compared to mean calculated Jaccard-similarity indices calculated for the 18S dataset (0.08 and 0.20). Comparing differences in Jaccard-similarity indices between curation algorithm, it becomes apparent that LULU curation results in a more homogenous diversity as Jaccard-similarity index doubled in comparison to the 0.01-Filter.

With an average Jaccard-similarity index of 0.15 and 0.20, which means that on average there is 15% and 20% similarity in terms of community composition between Level 5 and Level 6 using the COI or 18S marker, respectively highest Jaccard-similarity index was found between these two height levels. This was followed by elevation level 4 in comparison to elevation level 5 for which average Jaccard-similarities of 0.15 and 0.17 were calculated. The third highest similarity in terms of community composition between two height levels was found between elevation level 2 and 3 with average jaccard similarities between 0.14 and 0.18) (figure I.9).

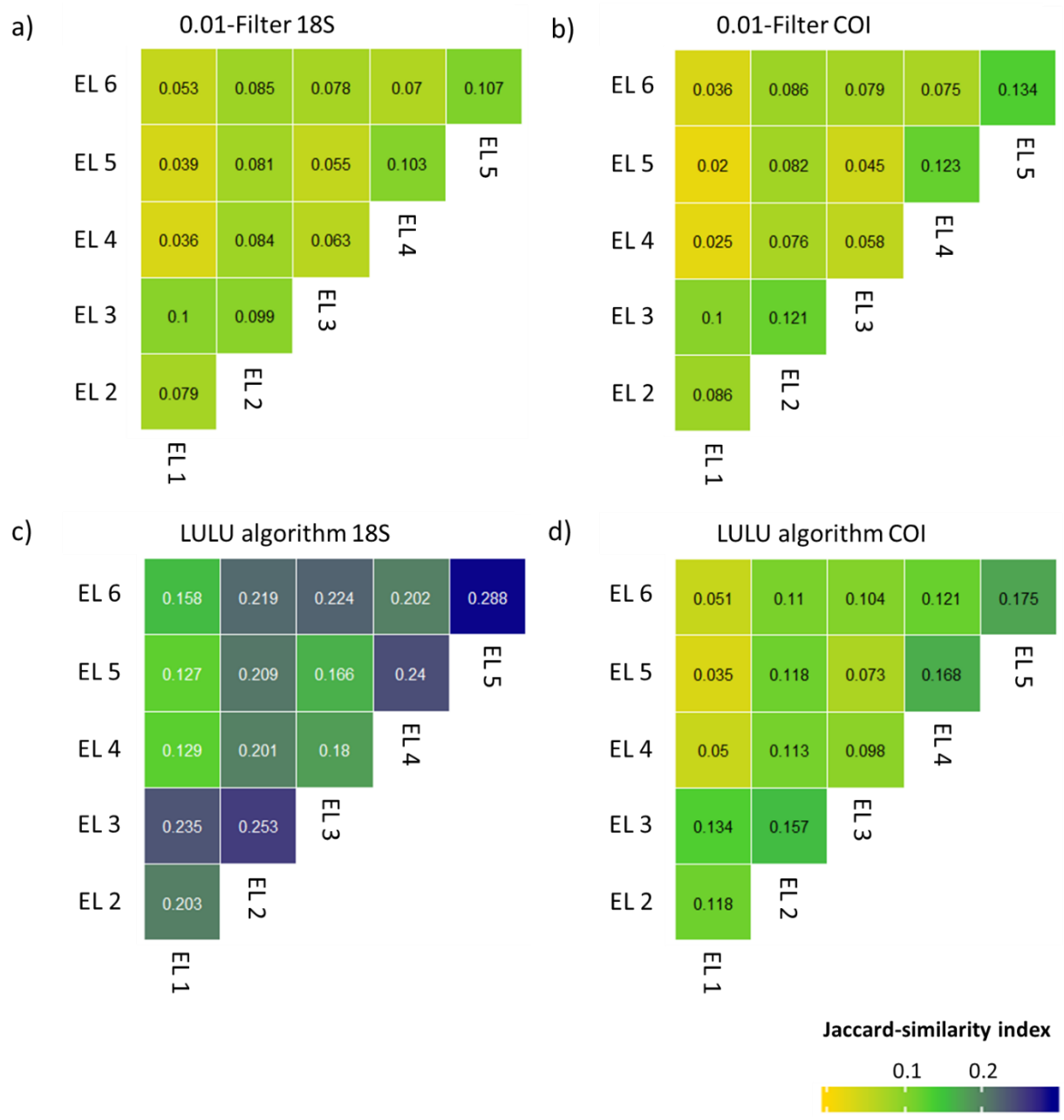


Figure I.9 Heatmaps based on Jaccard-similarity index indicating levels of beta-diversity between the different elevation levels (EL) depending on choice of marker and *in-silico* filter algorithm. a) 0.01-Filter + 18S dataset, b) 0.01 – Filter + COI dataset, c) 18S dataset + LULU algorithm, d) COI dataset + LULU algorithm. Increase in Jaccard-similarity indices is highlighted by fade of coloration from yellow over green to blue.

Regardless of choice of post-clustering filter strategy, principal component analysis (PCoA) for the COI and for the 18S dataset based on a prepared presence absence matrix, indicate that assessed species communities were significantly different between elevation levels (figure I.10). This was further confirmed with Adonis test (18S- 0.001-Filter 18S: F_5= 1.816, p=0.001; Lulu 18S: F_5=2.160, p=0.001; 0.001-Filter COI: F_5=1.839, p=0.001; Lulu COI: F_5=1.961, p=0.001). Regardless of post-clustering filter algorithm applied, we found that for the 18S dataset each elevation level formed a distinct cluster (figure I.10a,c). Similar finding were obtained for the COI dataset. However, a slightly overlap of clusters calculated for community composition found at elevation level 1 and 2 was detected (figure I.10b,d).

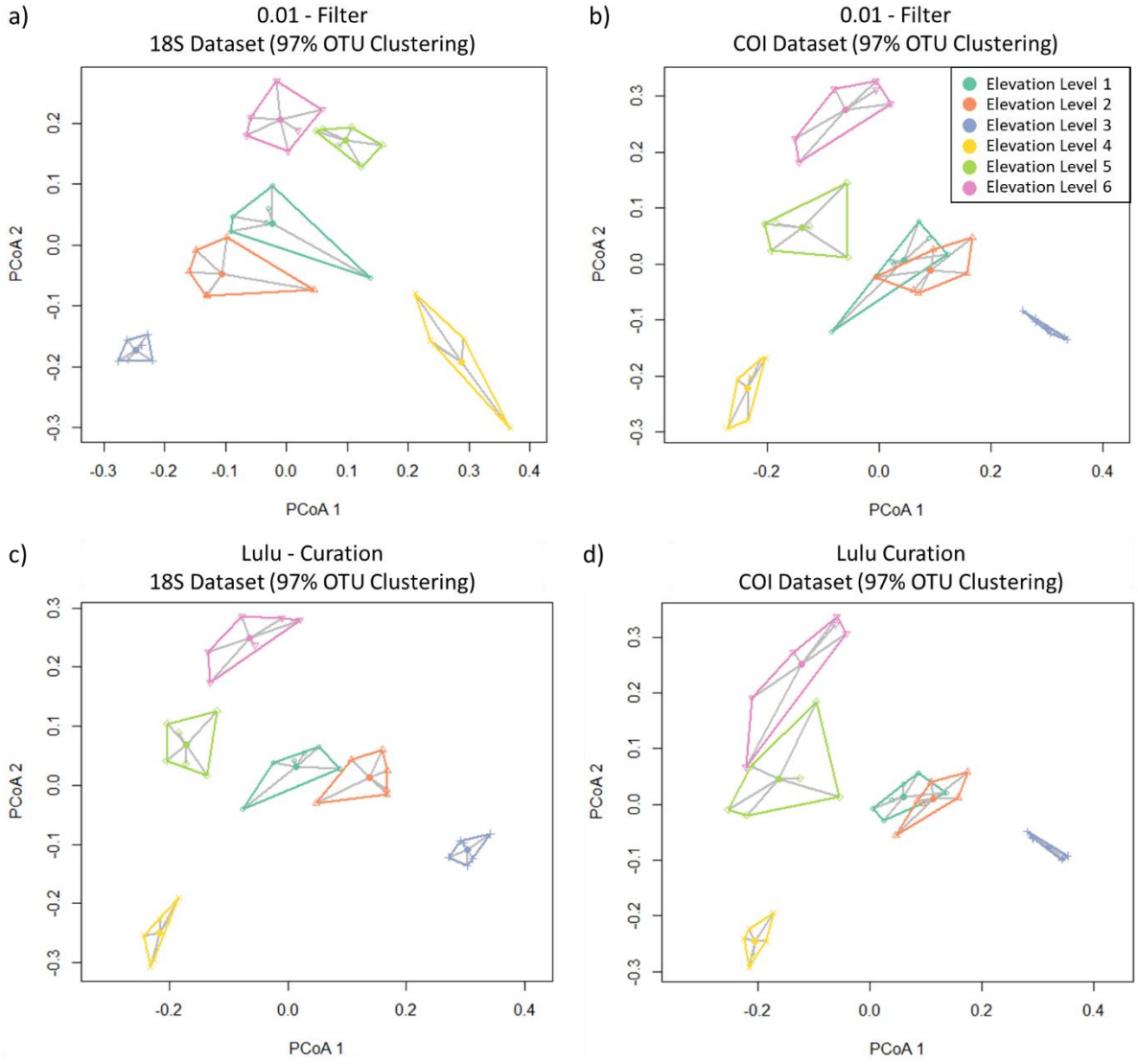


Figure I.10 Principal component analysis on OTU presence-absence matrix. a) PcOA for the 18S dataset curated with 0.01-Filter: PCoA1 explained for 6.88% of variation, PCoA2 explained for 5.78% of variation; b) PcOA for the COI dataset curated with 0.01-Filter: PCoA1 explained for 7.14% of variation, PCoA2 explained for 5.78% of variation; c) PcOA for the 18S dataset curated with 0 LULU-algorithm: PCoA1 explained for 9.10% of variation, PCoA2 explained for 6.75% of variation; d) PcOA for the COI dataset curated with the LULU algorithm: PCoA1 explained for 8.19% of variation, PCoA2 explained for 5.61% of variation.

Shifts in Metazoan Communities along an Elevational Gradient

To show differences in number and quality of taxonomy assignments between markers and dataset curated with different post-clustering algorithms a Marioko plot was prepared (figure I.11). Post-clustering filter strategy strongly influenced number of retrieved arthropod OTUs. For the 18S dataset almost twice as many OTUs assigned to Arthropoda with a blast hit of at least 90% were found for the 18S dataset curated with the 0.01-Filter (0.01-Filter: 669 OTUs; Lulu: 367 OTUs). For both COI datasets a contradicting picture was found. Almost twice as many OTUs with an assignment to Arthropoda were detected when the COI dataset was curated with the LULU algorithm (BOLD + 0.01 Filter: 57 OTUs; Bold + Lulu: 104 OTUs; GBOL + 0.01 Filter: 42 OTUs; GBOL + Lulu: 82 OTUs).

The average blastID of assignments to Arthropoda did not differ significantly between *in-silico* filters. On average OTUs of the 18S dataset curated with the 0.01-Filter were assigned to Arthropoda with a blast hit of 95.72%, while OTUs of the 18S dataset curated with the LULU algorithm were on average assigned to Arthropoda with blastID of 95.26%. Regardless of reference database used for taxonomy assignment of the COI dataset average blastID of assignments to Arthropoda was slightly higher for the COI dataset in comparison to the 18S dataset, but did not differ significantly between filter algorithms (BOLD + 0.01-Filter: 97.36; BOLD + Lulu: 97.05; GBOL + 0.01-Filter: 97.61; GBOL + Lulu: 97.31) (Figure I.11).

For the COI dataset number of assignments with a blastID of 99% or more to Arthropoda almost doubled when the LULU algorithm was used for post-clustering curation (BOLD + 0.01-Filter: 32 OTUs; BOLD + Lulu: 54 OTUs; GBOL + 0.01-Filter:29 OTUs; GBOL + Lulu: 50 OTUs). In contrast to that, no differences in number of OTUs with an high quality assignment to Arthropoda was found between *in-silico* filter methods for the 18S dataset. (0.01-Filter: 45 OTUs; Lulu 44 OTUs)

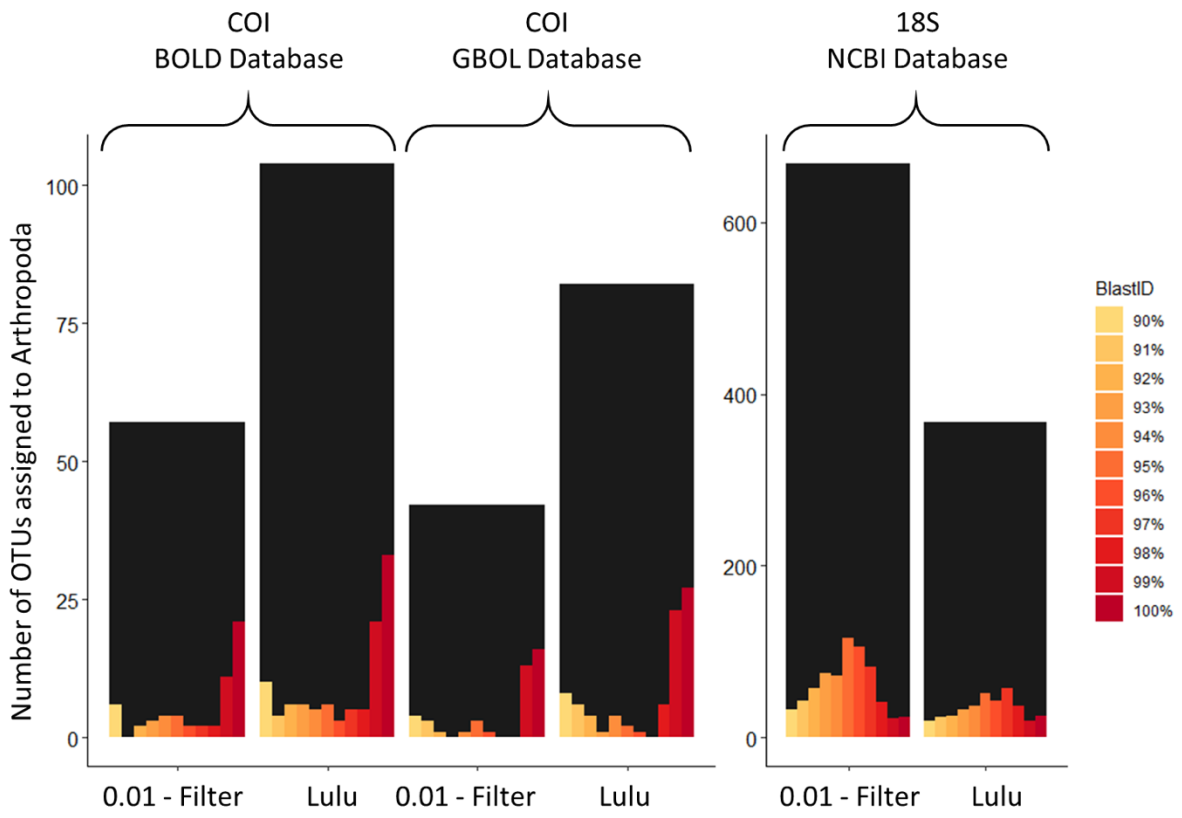


Figure I.11 Number of Arthropod OTUs retrieved with the two markers depending on post-clustering filter algorithm applied, including information about number assignments per BlastID ranging from 90-100%

Depending on reference database, between 43.2% and 45% of detected OTUs of the COI dataset assigned to Arthropoda were detected with both post-clustering filter methods (figure I.12). LULU curation resulted in the detection of 44 additional OTUs when the GBOL database was used as reference, while 54 additional arthropod OTUs were found with LULU curation when using the BOLD database. The 0.01-Filter resulted in the assignment of six and respectively seven additional OTUs to Arthropoda, which were not present in the list of assigned OTUs when curation was based on the LULU algorithm (figure I.12). Within the 18S dataset, 247 OTUs were assigned to Arthropoda regardless of post-clustering filter algorithm

applied. Furthermore the LULU algorithm retrieved 120 additional OTUs with an assignment to Arthropoda which were not present in the dataset filtered with the 0.01-Filter. Respectively 422 OTUs assigned to Arthropoda were contained in the 18S dataset filtered with the 0.01-Filter (figure I.22).

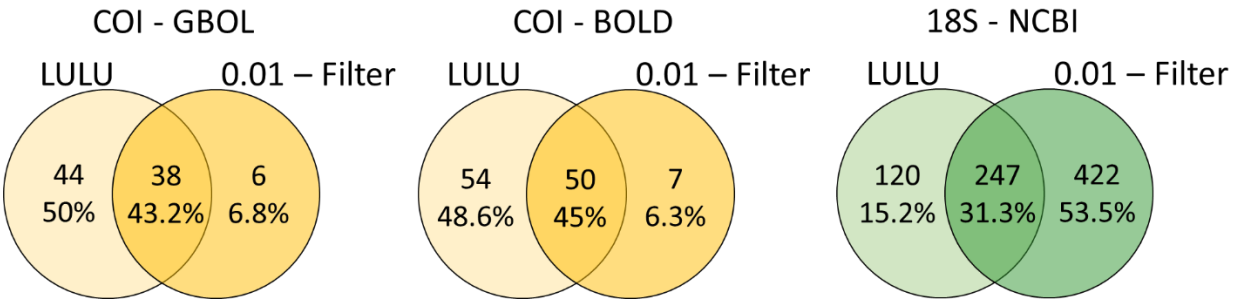


Figure I.12 Number of shared and unique OTUs assigned to Arthropoda between post-clustering filter strategy depending on marker and reference database used

For the 18S dataset a total of 10 Metazoa phyla were identified. Regardless of *in-silico* filter applied the highest number of OTUs was assigned to the four phyla Annelida, Arthropoda, Chordata and Nematoda. For all six elevation levels it was found that number of assigned OTUs increased when LULU curation was used for post-clustering filtering (figure I.13).

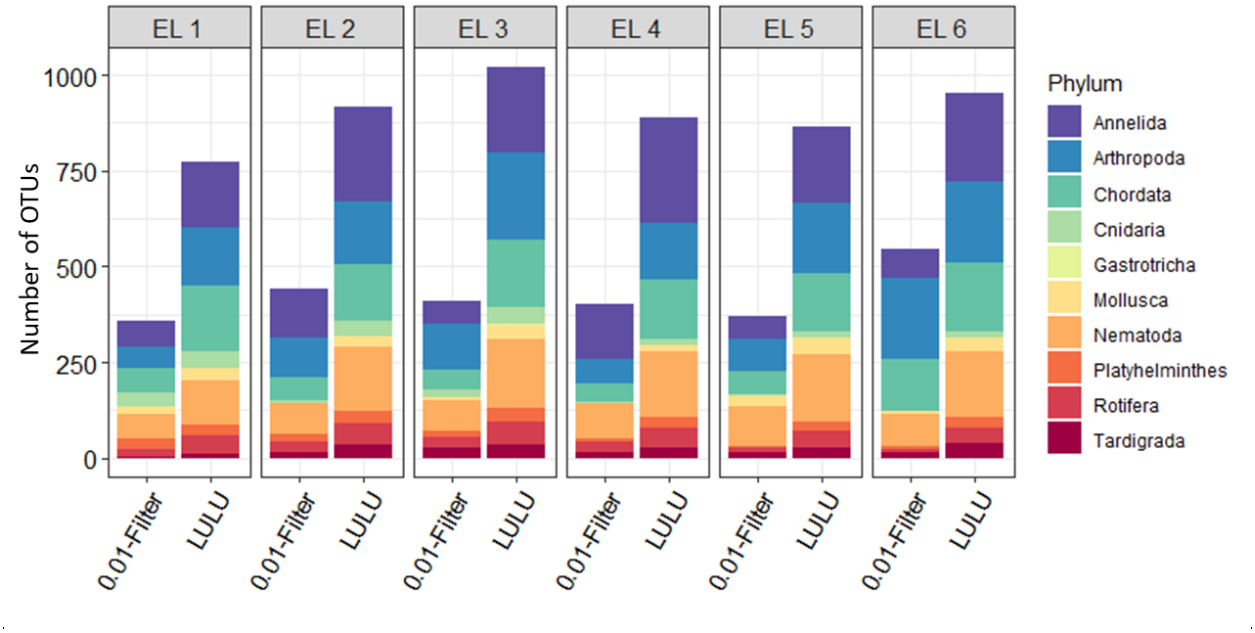


Figure I.13 Number of OTUs assigned on phylum level per elevation level for the two post-clustering curation algorithms when using the GenBank as reference database for taxonomy assignment of the 18S dataset

Taxonomy assignment of the COI dataset based on the BOLD database resulted in the detection of the four metazoan phyla Annelida, Arthropoda, Chordata and Nematoda (figure I.14). The post-clustering filter algorithm had a strong influence on number of detected OTUs, but its degree was found to be strongly depending on phylum. On average eight additional OTUs assigned to Arthropoda were found per elevation level when the dataset was curated with the LULU algorithm. However, approximately 7.5 additional OTUs assigned to Annelida were detected per elevation level when applying the 0.01- Filter algorithm. While for the Arthropods this finding was consistent over all six elevation level, the degree of influence of *in-silico* filter differed between elevation level for the phylum Annelida. While at elevation level 2 (+15 OTUs), elevation level 3 (+9 OTUs) and elevation level 4 (+26 OTUs) more OTUs assigned to Annelida were retrieved with the 0.01-Filter the opposite was the case for

elevation level 1 and elevation level 6 for which two and 3 additional OTUs were found when the dataset was curated with the LULU algorithm (figure I.14). For the phyla Chordata and Nematoda only low variations in number of detected OTUs depending on curation methods were found.

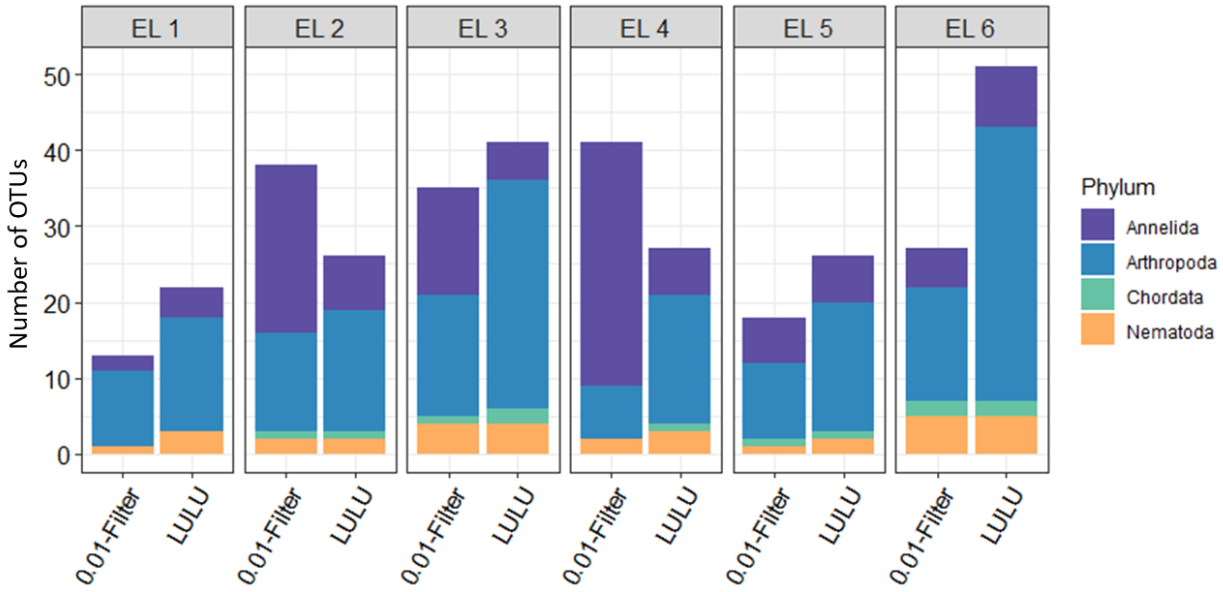


Figure I.14 Number of OTUs assigned on phylum level per elevation level for two post-clustering curation algorithms when using the BOLD database as reference database for taxonomy assignment of the COI dataset

For the COI dataset blasted against the GBOL database the four phyla Annelida, Arthropoda, Chordata and Streptophyta were identified (figure I.15). Number of assigned OTUs per elevation level was highest for the Arthropoda. The remaining three phyla were exclusively detected at a single elevation level. Furthermore, all three groups had in common that its detection was based on a single OTU. *In-silico* filter algorithm influenced number of OTUs

assigned to Arthropoda, but had no influence on number of OTUs assigned to Streptophyta and Chordata. The 0.01-Filter and LULU curation resulted in the detection of a single OTU at elevation level 3 which was assigned to Streptophyta. Additionally a single OTU assigned to Chordata was found at elevation level 6 with both curation methods. Based on LULU curation no OTU was assigned to the phylum Annelida. In contrast to that resulted the application of the 0.01-Filter in the detection of a single annelid OTU at elevation level 3. For the Arthropoda an increase of approximately 8.8 OTUs per elevation level was found when the LULU filter was used for post-clustering curation. However, the difference in number of OTUs between the two evaluated curation algorithms were strongly varying between elevation levels (figure I.15). While at elevation level 2 only one additional OTU was found when applying LULU curation instead of the 0.01-Filter at elevation level 6 18 additional OTUs were found.

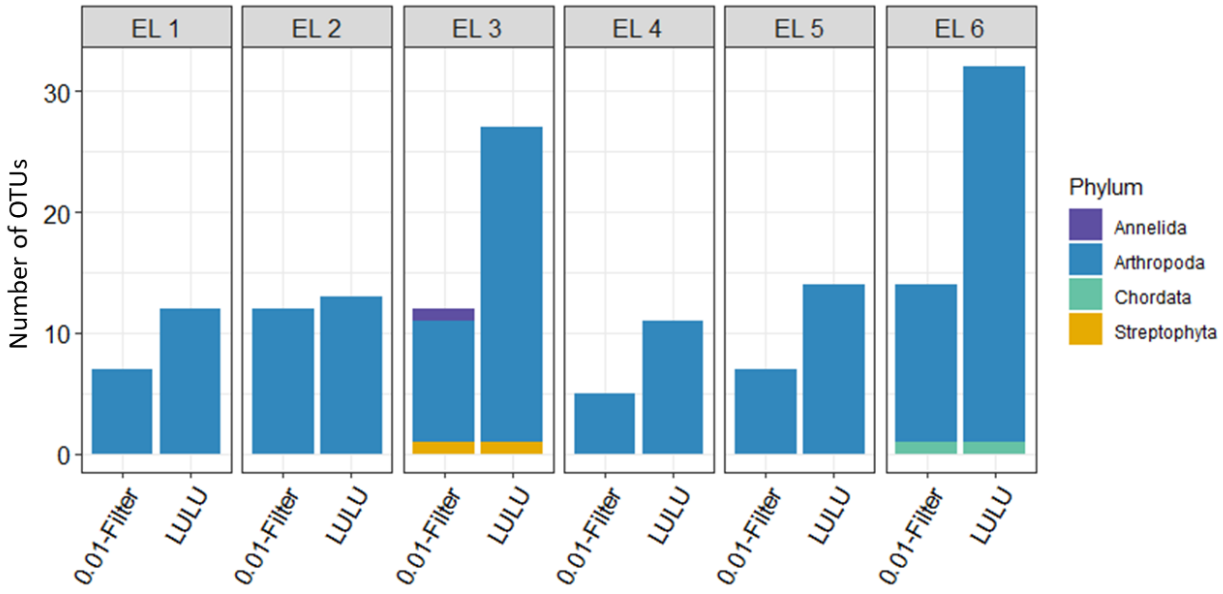


Figure I.15 Number of OTUs assigned on phylum level per elevation level depending on filter-strategy based on the GBOL database

The choice of marker has a significant influence on insect order detection rate. The order Ephemeroptera and Megaloptera were exclusively found with the 18S marker, while the order Thysanoptera and Neuroptera were exclusively detected with the COI marker (figure I.16). The remaining seven orders Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Orthroptera and Pscocoptera were detected with both markers. However, applied filter algorithm and in the case of the COI marker choice of reference database had an influence on taxa detection rate (figure I.16). With the 18S marker eight orders were identified regardless of filter algorithm applied. However, the ninth order Pscoptera was only detected when using the LULU-algorithm for post-clustering filtering. LULU curation also had an positive effect on number of detected orders with the COI maker. When blasting the COI sequences against the BOLD database six additional orders were identified. When using the GBOL database as reference LULU curation resulted in the detection of two additional orders (figure I.16).

Next to number of orders, number of detected insect families per order were depending on choice of reference database and marker significantly higher when using the LULU algorithm for post-clustering filtering (paired-sampled t-test: df = 16, p-value = 0.011). For the 18S dataset a slight increase in number of detected families of the orders Diptera and Psocoptera were found. However, the 0.01-Filter resulted in a slight increase in number of detected families of the order Megaloptera (figure I.16). For the COI dataset blasted against the BOLD database two coleopteran families, two hemipteran, one hymenopteran, two lepidopteran, one neuropteran, two psocopteran and one orthopteran families more were recovered when the dataset was curated with LULU. Taxonomy assignment on basis of the GBOL database led to similar results. An increase in number of families of the orders Coleoptera (+3), Diptera (+3), Hemiptera (+1), Hymenoptera (+3), Lepidoptera (+3), Neuroptera (+1), Orthoptera (+1) and Psocoptera (+2) were found.

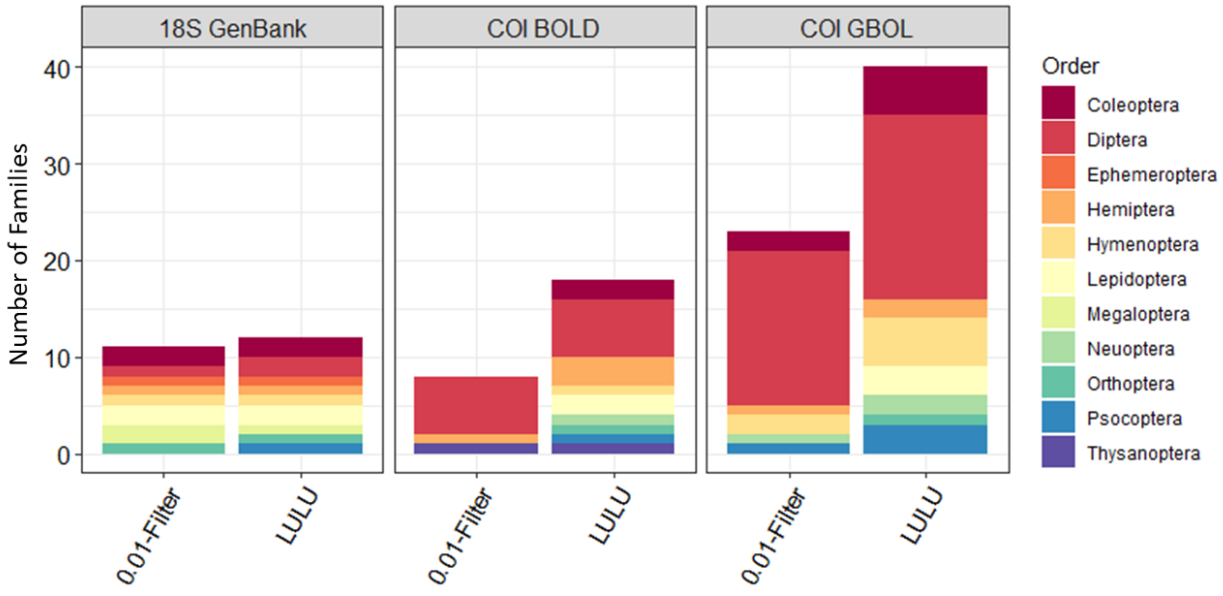


Figure I.16 Total number of insect families identified depending on applied *in-silico* filter strategy in combination with the 18S and COI marker respectively. Three different databases were used for taxonomy assignment (GenBank, BOLD and GBOL).

Measured Abiotic Parameters

Mean soil temperature measured ranged depending on elevation level between 3.06 °C and 10.54 °C (table I.2). Over a 24h period soil temperatures slightly oscillated at the most between ± 0.79°C (figure I.17), while air temperature showed considerably more variation over a one day trial period (± 5.44°C) (figure I.18). Relative air humidity varied broadly over 24 hours at all elevation levels (table I.2)

Table I.2 Measured abiotic parameters per elevation level. Table 6 gives information about mean soil temperature [°C], mean air temperature [°C] and mean relative humidity [%] at each elevation level locked for the day of sampling

	Elevation Level 1	Elevation Level 2	Elevation Level 3	Elevation Level 4	Elevation Level 6
mean soil temperature [°C]	10.54 (±0.47)	9.75 (±0.35)	7.70 (±0.13)	9.32 (±0.79)	3.06 (±0.28)
mean air temperature [°C]	12.03 (±5.74)	6.92 (±0.88)	8.44 (±1.85)	9.22 (±4.40)	9.24 (±4.88)
mean relative humidity [%]	75.8 (±11.7)	86.0 (±8.7)	66.4 (±10.7)	90.1 (±10.9)	50.8 (±12.7)

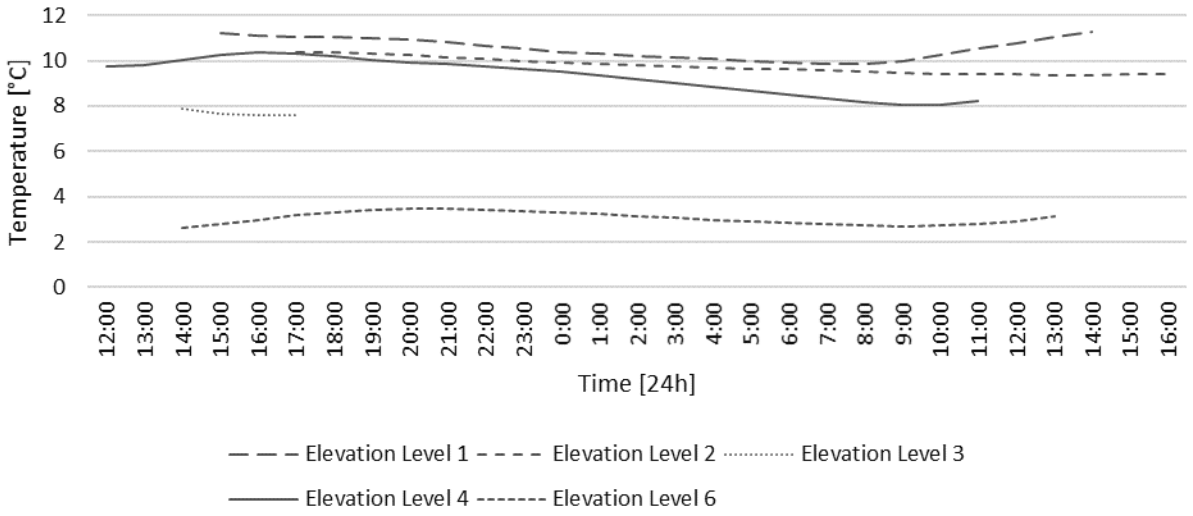


Figure I.17 Measured soil temperature at each elevation level for the first 24hours after sampling

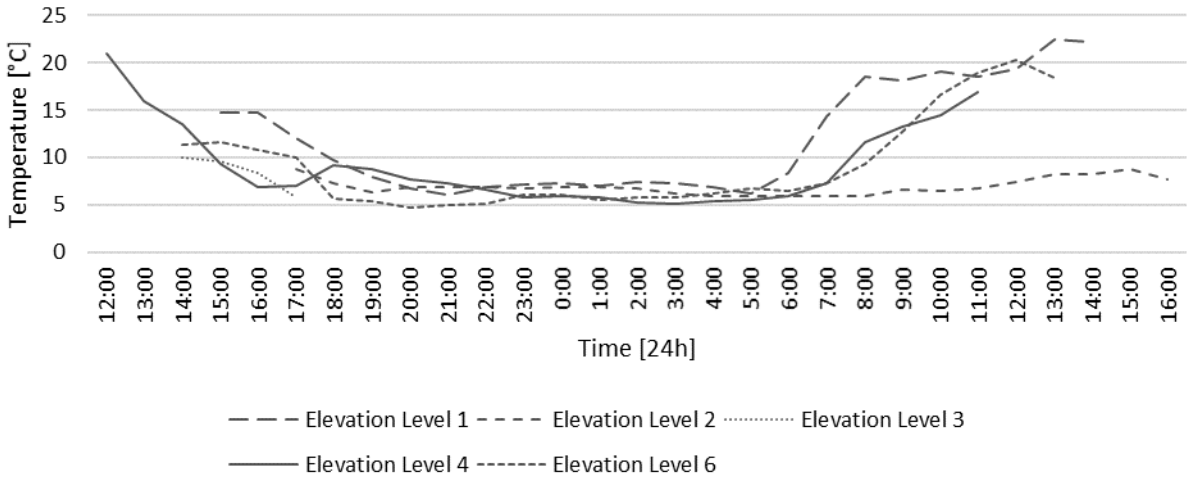


Figure I.18 Measured air temperature at each elevation level for the first 24hours after sampling

Discussion

This study agrees with previous studies showing that species community composition changes along an elevation gradient (Whittaker 1952; Vetaas and Grytnes 2002; Hodkinson 2005; McCain 2007). It also shows the applicability of metabarcoding to identify species and soil community structure for overall diversity assessments and the investigation of species distribution patterns. Using an amplicon-based metabarcoding approach it is possible to assess differences in species composition between samples on local scale (e.g. biological replicates) but also on larger scale by using Operational Taxonomic Units (OTUs) as a proxy for species. Nevertheless, the study underlines the importance of methodological considerations as choice of methods can directly influences study outcome.

Technical Variables: Gene Markers, *in-silico* Filters and Choice of Databases

Magnitude of Biodiversity Uncovered by Markers

Both marker revealed similar community patterns across an elevational gradient. However, the amplification of the 18S marker resulted in the detection of a significantly higher number of OTUs

compared to the COI marker. This is consistent with former studies (Macheriotou et al. 2019). It is known that differences in number of recovered OTUs are mainly contributed to differences in degree of variability within primer binding sites of the two marker (Tang et al. 2012; Borrell et al. 2017; Clarke et al. 2017). Because of its protein coding character is the COI marker less constrained by selection as mutations at many nucleotide positions do not change the coded protein (Deagle et al. 2014). As a result amplification success of the COI marker is highly differing between taxa whereas several groups are frequently missed (Yu et al. 2012; Zhou et al. 2013). This bias is further enhanced when metabarcoding highly diverse DNA mixtures as species with higher primer affinities will likely outcompete species with lower affinities (Morinière et al. 2016). Furthermore, the failed amplification of the taxa with lower affinities is often masked by an increase in sequences derived from taxa with higher primer affinities (Deagle et al. 2014). In contrast to the COI marker are the primer binding sites of the 18S marker gene highly conserved, what is enabling the amplification from a broader range of taxa (Lindeque et al. 2013; Hadziavdic et al. 2014). Especially, when working with eDNA extracted from soils, primer binding capacity can significantly influence study outcome as degree of biodiversity which can be found in one square meter of soil is outstanding (Schaefer and Schauermann 1990; Decaëns 2010). The more species are contributing to DNA mixture the higher the number of species which will be overseen when markers with highly variable binding sites are targeted (Morinière et al. 2016). We conclude, that due to differences in primer binding site the 18S marker results in the amplification of a broader range of taxa leading to an increase in number of recovered OTUs. This is also reflected in number of detected phyla. While the 18S marker detected 10 phyla, the COI marker identified only four.

Taxonomic Coverage and Resolution of Marker

While highly conserved primer binding sites ensures for a more even amplification of DNA from a broad range taxa, also referred to as “taxonomic coverage”, a higher variability within the barcode allows for a higher taxonomic resolution (Hebert et al. 2003a; Andújar et al. 2018b). Taberlet et al. (2018) defined the perfect barcode as a short DNA fragment displaying a highly variable sequence which is flanked by two highly conserved regions. The central variable region is discriminative for all species of the target group, that is, it’s sequence is uniquely associated to a given species and not shared with others. Although, the COI dataset contained less OTUs compared to the 18S dataset and furthermore even less OTUs matched an entry of the here used reference databases, number of recovered insect families exceeded the one detected with the 18S marker. Depending on insect order up to 17 families more were detected using the COI marker. This is contributed to the fact that the high taxonomic coverage of the 18S marker goes along with a decrease in taxonomic resolution (Clarke et al. 2017). Because of the low variability within the barcode region several species can possibly even share one barcode (Tang et al. 2012) leading to an increase in false negative results. In direct comparison has the COI barcode a significantly higher taxonomic resolution, what is enabling species delimitation (Hebert et al. 2003a; Andújar et al. 2018b). Especially when targeting arthropods it has been shown that the COI marker should be given priority over the 18S marker (Cowart et al. 2015). However, the COI marker is not an all-purpose answer. Our results indicate that the COI marker seems to underestimate truly existing nematode diversity. Nematodes has been described to be highly diverse in soils and can reach abundances of up to several hundred specimens per gram (Hoorman 2011). Former studies have shown, that for nematodes species level assignments are usually higher for the 18S marker (Andújar et al. 2018a; Macheriotou et al. 2019) compared to the COI marker. On the one hand this is contributed to a higher degree of variability within the COI primer binding sites (De Ley et al. 2005; Creer et al. 2010) inhabiting amplification of the barcode region, but could on the other hand also be a result of the absence of a definite barcode

gap (Derycke et al. 2010). The barcode gap is defined by the differences in genetic difference between the highest intraspecific variation and lowest interspecific divergence (Hebert et al. 2003a, 2003b; Wiemers and Fiedler 2007; Meier et al. 2008; Kvist 2014). A lack of a barcode gap has also been observed within the phylum Annelida. Especially for the earthworms it has been described that several species show 0% interspecific divergence (Kvist 2014) usually resulting in a strong underestimation of local annelid diversity. When comparing number of OTUs assigned to Annelida between the two marker datasets, significantly more annelid OTUs were detected on basis of the 18S marker. However, the COI dataset was blasted against two different databases: the GBOL and the BOLD database. On basis of the BOLD database the phylum Nematoda was detected at all six elevation levels while, when using the GBOL database as reference, not a single OTU was assigned to that phylum regardless of sampled elevation level. This is highlighting the importance of choice of database which is already known to have a direct influence on species detection rate (Somervuo et al. 2017; Andújar et al. 2018b). The here used GBOL database did not contain the barcode of any nematode species, prohibiting the detection of this phylum. In contrast to that contained the BOLD database barcodes of 64 nematode species. Similar observation were also made for the phylum Annelida. On basis of the BOLD database significantly more OTUs were detected compared to analysis using the GBOL database as reference. The BOLD database contained the barcodes of 233 annelid species, while number of species represented with barcodes in the GBOL database was with 27 species significantly lower. These findings are consistent with former studies. Somervuo et al. (2017) have found that probability of false positive and negative assignments is negative correlated with size and completeness of the reference database. The use of reference database with a high taxonomic coverage and depth usually leads to an increase in species detection rate (Andújar et al. 2018b).

However, In terms of completeness are COI databases unparalleled. This is mainly due to the fact that because of its high taxonomic resolution the Consortium for the Barcode of Life (CBOL) has

chosen the mitochondrial COI gene for standard DNA barcoding of single animal specimens (Hebert et al. 2003b; Ratnasingham and Hebert 2007). The first major program of the CBOL has been the BARCODE 500K project, which had the aim to establish a reference database containing the barcodes of 0.5 million species. This goal was met in August 2015, but the CBOL initiative kept going and even expanded. Today, more than 502.700 public barcode clusters are stored in the BOLD database. The barcode of life initiative encompasses now research organizations from 25 nations. Today, it can be said that no database for any other genetic region covers as many taxa as the Barcode of Life Database (BOLD) does for the COI (Deagle et al. 2014). For the 18S marker, no comparable efforts to establish a comparatively big and complete database have been undertaken, which might also result from the low taxonomic resolution of the 18S marker. Here it was shown that blast identity of COI sequences assigned to Arthropoda were higher compared to 18S sequences. This is likely an artefact of incomplete reference databases as the absence from species from the reference database increases, next to the risk of encompassing false negative results, also the risk of introducing false positive assignments as the next highest hit will be chosen. Despite the fact that the COI marker had a higher taxonomic resolution significantly less OTUs had an assignment. While for the COI dataset between 0.4% and 0.8% of total OTU count were assigned on phylum level, between 4.2% and 7.6% of OTUS of the 18S marker had a hit within the 18S reference database. This is either attributed to composition of extracted DNA mixture for the benefit of groups which increasingly fail to amplify using the COI marker, or secondly to incomplete reference databases. As already stated are COI databases rather complex. However, some areas are highly underexplored and several species are still waiting to be described. For the Caucasus region, new species are described on a regular basis (Kaplin 2019; Marin and Palatov 2019; Martynov and Žiak 2019; Teslenko et al. 2019). This is mainly due to two reasons. On the one hand, the Georgian Caucasus region is a biodiversity hot spot hosting an immense biodiversity. On the other hand Georgian biodiversity research was long time hampered by poorly developed scientific infrastructure. However, the largest proportion of newly described

species from that area are associated with above ground habitats including mammals and reptiles. Given the fact that soils host an immense biodiversity (Stork 1988; May 1990; André et al. 1994; Hågvar 1998; Curtis et al. 2002; Decaëns 2010; Nielsen et al. 2011; Andújar et al. 2015) it can be expected that numbers of species waiting to be described are exceeding number of above ground species several fold. Even soil from countries with a well established scientific infrastructure are known to be highly underexplored (Wall et al. 2005). In views to our study, the high number of undescribed species not being present in the databases has possibly lead to a high number of unassigned OTUs. Additionally it must be mentioned, that because of homoplasy variations at primer binding sites for the COI marker becomes saturated between distantly related taxa. As a result it is not possible to target for certain groups (e.g. arthropods) (Deagle et al. 2014). Furthermore, sample size, DNA extraction method as well as the number of PCR cycles can influence resulting biodiversity estimates for different taxonomic groups within soil (Dopheide et al. 2019). It is known that some regions of the bacteria genome are matching the commonly used COI primer binding sites, whereas next to metazoan DNA also bacteria DNA is amplified (Siddall et al. 2009; Yang et al. 2013, 2014; Horton et al. 2017; Mioduchowska et al. 2018). Especially when working with soil samples the undesired amplification of bacterial DNA can significantly influence study results and species identification rates as the extreme dominance of bacteria within the study substrate is leading to a strong reduction of amplification success of metazoan DNA for the benefit of the amplification of bacteria DNA (Yang et al. 2013; Horton et al. 2017). As the here used reference databases did not contain complete genomes of bacteria we cannot proof the influence of the above described observations on our dataset but the high number of unassigned reads is supporting this hypothesis.

Concluding we recommend to use both markers for the assessment of soil metazoan community composition. Annelida and Nematoda are important destruents highly influencing soil environmental parameter like texture, structure and nutrient composition (Coleman et al. 2004;

Lavelle et al. 2006) which in turn is directly influencing local fauna and flora. However, several arthropod species are ecosystems engineers and are indispensable for the maintenance of habitat health and function.

In-silico Filter Algorithm

Without taxonomic assignment it is difficult to make reliable statements about the degree of existing local diversity, especially as no best-practice advises are available for monitoring metazoan diversity on the basis of eDNA extracted from soil. Today it is well known that assessments of species communities via metabarcoding is significantly affected by applied methods including choice of marker, reference database (Andújar et al. 2018b), sampling strategy (Grey et al. 2018), extraction method (Deiner et al. 2015; Dopheide et al. 2019) and pipeline (Yang et al. 2013; Brandon-Mong et al. 2015; Anslan et al. 2018). In contrast to that, little known about the influence of post-clustering filter strategies on study outcome, which is why *in-silico* filter strategy is often randomly chosen. This study showed that post-clustering filter algorithm is depending on marker choice significantly affecting study outcome.

Influence of *In-silico* Filter on the COI-Dataset

Within the COI dataset the use of the 0.01-Filter resulted in the detection of a significantly lower number of OTUs assigned to Arthropoda when being compared to the LULU algorithm. In detail, regardless of reference database used almost twice as many OTUs were assigned to Arthropoda when the LULU algorithm was applied. Furthermore, although mean blastIDs did not differ significantly between filter algorithms, more OTUs with a high quality assignment (≥ 90% blastID) to Arthropoda were retained in the dataset when curating the dataset with LULU. This is

supporting the findings of Frøslev et al. (2017) who stated, that filter approaches cutting off the tail of a OTU rank-abundance curve (i.e. low alpha-diversity) are more prone to exclude real OTUs from the dataset. Our results contradicts the widely held assumption that that the highest number of infrequent OTUs result from erroneous sequences (Huse et al. 2010; Kunin et al. 2010; Brown et al. 2015). The fact that with the 0.01-Filter almost 50% less OTUs assigned to Arthropoda were recovered allows for the conclusion that many arthropod OTUs only contain a low number of sequences. This is mainly contributed to the high degree of diversity found in DNA mixture. It is likely that complexity of the dataset is determining the degree of influence of filter algorithm. A highly diverse DNA mixture results in the detection of a higher number of OTUs compared to a very homogenous mixture. In turn, a lower proportion of total read count is assigned to each OTU. Cutting of the tail of a rank abundance curve increases the risk of excluding low abundance OTUs derived from rare, infrequent species from a highly diverse dataset. Rare infrequent species are often making up the lion share of existing diversity (Nemergut et al. 2011; Frøslev et al. 2017). Especially when working with soil samples, the undesired amplification of non-target DNA can significantly influence study results. This is mainly due to the fact that non-target microorganisms (e.g. bacteria) are dominating the soil habitat (Whitman et al. 1998; Fierer et al. 2007). As already mentioned, is the COI marker known to co-amplify certain regions of the bacteria genome (Siddall et al. 2009; Yang et al. 2013; Horton et al. 2017; Mioduchowska et al. 2018) leading to a strong reduction of amplification success of metazoan DNA for the benefit of the amplification of non-target DNA (Yang et al. 2013; Horton et al. 2017; Marquina et al. 2019). It can be expected, that concentration of DNA of bacterial origin is surpassing amplified metazoan DNA by several orders of magnitudes (Yang et al. 2013; Horton et al. 2017). Additionally, distribution patterns and population density of macro-organisms are strongly differing from patterns found for microorganisms (Taberlet et al. 2012; Dopheide et al. 2019). In general are microorganisms more evenly distributed within a habitat, whereas the distribution patterns of macroorganisms are patchier as they are less abundant and bigger in size (Taberlet et al.

2012; Dopheide et al. 2019). As a result, several prokaryotic organisms of the same species contribute to DNA mixture, while in the DNA mixture contained DNA originated from a certain arthropod species is more likely to have derived from a single individual. As a result, OTUs resulting from metazoan DNA are more infrequent und usually contain less sequences, whereas they are more prone to be excluded from the dataset by the 0.01-Filter.

Influence of *In-silico* Filter on the 18S-Dataset

Contrarily, the number of OTUs of the 18S dataset assigned to Arthropoda was significantly higher when using the 0.01-Filter. Overall, LULU curation resulted in a much stronger reduction in number of arthropod OTUs compared to the 0.01-Filter. However, almost no differences in the number of detected insect families was found. Because of the low taxonomic resolution of the 18S marker the LULU algorithm has likely merged a high number of 'daughter' OTUs with consistently co-occuring sequence similarity, but more abundant 'parent' OTUs (Frøslev et al. 2017). In detail variation between taxa at lower taxonomic levels are strongly reduced which can justify the presence of false positive 'daughter' OTUs. As a result, number of 18S OTUs decreased much stronger in comparison to numbers of COI OTUs. The 0.01-Filter retained 670 arthropod OTUs in the dataset, while the LULU algorithm retained a lower number of 368 arthropod OTUs. However, within both filter datasets several OTUs were assigned to the same species/accession number. While the 0.01-Filter detected 137 different arthropod species, LULU curation resulted in the detection of 127 arthropod species. This is supporting our hypothesis that LULU curation merged similar OTUs, whereas number of species were only slightly affected. In total 114 arthropod species were recovered with both *in-silico* filter algorithms.

The fact that similar number of species were recovered with both *in-silico* filters leads to the conclusion that arthropod OTUs detected with the 18S marker likely contain a higher percentage

of total sequence count per sample compared to the COI dataset, and that moreover, 18S OTUs differ less dramatic in number of assigned sequences. This could either be due to the fact that a) the 18S dataset is less diverse supporting amplification of rare sequences; b) that sequences originating from several species were more likely to be assigned to the same OTU, increasing sequence count; c) because of the highly conserved primer binding sites, the 18S marker is supporting an equal amplification of sequences due to less dramatic differences in primer affinities.

Biodiversity Distribution Patterns along an Elevational Gradient

As already discussed is the detection rate of metazoan species based on the 18S marker less influenced by *in-silico* filter algorithm compared to the COI marker. Nevertheless, for the COI dataset is was observed that the application of the LULU algorithm resulted in a significant higher number of detected metazoan taxa compared to the 0.01-Filter. As LULU curation was not negatively influencing species detection rate for the 18S marker, it should be given priority over the 0.01-Filter. Therefore is the following ecological discussion based on the 18S and COI dataset curated with the LULU-algorithm.

The here presented metabarcoding study conducted in the Georgian Caucasus region found that each elevation level is hosting a unique biodiversity. However, specific local diversity (alpha diversity) niches present within a few meters apart, independently of height were identified. It has long been discussed that habitat zonation and height could drive species diversity (Lomolino 2001). The biodiversity patterns found here are supporting these findings.

Species turnover rate indicated by Jaccard-similarity index was not continuously highest between adjoining elevation level. Moreover, species composition associated with elevation level 1 was more similar to elevation level 6 than to elevation level 4 and 5. Therefore, the hypothesis that height is the main factor influencing species turnover rate has to be rejected. Nevertheless, height has undeniably a direct influence on local fauna and flora (Rahbek 1995; Sanders and Rahbek 2012), but the underlying causes are manifold. Next to climate (Grytnes 2003; Fu et al. 2006), productivity (Wang et al. 2004; Malhi et al. 2017), source-sink dynamics (Kessler et al. 2011), area (Rosenzweig 1995; Lomolino 2001; Sanders 2002), disturbances (Fu et al. 2006; Escobar et al. 2007; Rowe 2009), geometric constraints (Sanders 2002; Fu et al. 2006; Rowe 2009), evolutionary history (Machac et al. 2011), soil composition (Grieve et al. 1990) is scale (Rahbek 1995) directly influencing detected community composition. Rahbek (1995) underlined the importance of scale when assessing differences in species community composition over a height gradient. Studies based on a single transect were more likely to form a humped curve for species diversity over elevation level, than the ones performed over a whole mountain range. The classical species-area relationship (SAR) hypothesis supports these findings as it asserts that the bigger the area, the more species it will host (Rosenzweig 1995). For this study, scale was very narrowed as samples were taken along a transect, without taking into account that each elevation level consists of subalpine or alpine belts along the slopes of the mountain. Wind, light intensity and duration of direct solar radiation per day as well as steepness of slopes are only three factors that can already differ strongly on small local scales (Rahbek 1995) and thereby affecting local biodiversity. As a result, bare mountain sites as wells as dense coniferous forests can be part of the same elevation level but due to very localized sampling, biodiversity associated with only one area is assessed resulting in a strong underestimation of biodiversity which is associated with the corresponding elevation level.

However, diversity distribution patterns are highly dependent on the target organisms as well as on the local climate (Araújo et al. 2005). Local climate, in particular temperature is to some extent

directly influenced by height (Grytnes and McCain 2007; McCain 2007). While temperature at the lower height levels usually starts to increase earlier in the year, higher elevation levels often remain in a state of winter dormancy for a longer period of time. This has a direct influence on the start of the growing season. As a result, the lower elevation levels are already full of life, while at the upper elevation levels temperatures are still preventing melting processes (McCain and Grytnes 2001; Rolland 2003) whereas growing season is inhabited (Chmielewski and Rötzer 2002). The here presented study was conducted between April 2018 and July 2018. While the lower elevation levels 1-4 were sampled in April, the higher elevation levels were inaccessible at that time due to the high amount of snow covering the mountain. Nonetheless, elevation level 4 was still covered in snow at the time of sampling, which may have led to the low number of OTUs identified, as activity of many organisms groups is known to be limited by these harsh environmental conditions (Williams and Osman 1960). Although samples at elevation level 6 were taken approximately one month after sampling of the lower elevation levels, temperatures had still not reached a similar high level. Taken into account that many species use day length as main factor for the onset and offset of diapause (Gullan and Cranston 2014), it can be assumed that species inhabiting higher elevation levels must have developed certain adaptations to be able to survive in the more harsh environment including behavioral (Whitman 1988) and physiological strategies (Block 1990; Prange 1990; Heinrich 2013), eventually leading to a higher degree of local endemism. However, here we observed that the highest degree of endemism (number of unique OTUs per elevation level) was found at elevation level 1. This is contradicting to former studies (Stotz et al. 1996; Vetaas and Grytnes 2002): A study conducted in the Himalayan mountains showed that number of endemic vascular plants increases linearly from the lowland to 6000m asl. (Vetaas and Grytnes 2002). Contradicting to that, a second study showed that the highest number of endemic bird species was found at intermediate elevations (Stotz et al. 1996). We conclude that organisms groups show different patterns of endemism along an elevational gradient. The underlying causes for these

contradicting patterns cover a broad spectrum, which can roughly be grouped into four categories influencing organism groups at varying degrees: (1) Climatic factors, which encompasses most abiotic influences. (2) Spatial influences, including area and spatial constraints. (3) The historical background of a certain area and the (4) biotic factors like community overlap or interactions between fauna and flora (Grytnes and McCain 2007). Although, the driving causes for endemism are not yet fully understood, the two above cited studies (Stotz et al. 1996; Vetaas and Grytnes 2002) indicate that isolation played a major role for speciation, as the highest degree of endemism was found above the height gradient were species richness reached a peak. Here we found that biodiversity reached a peak at elevation level three. In combination with the fact that the highest degree of endemism was found at elevation level 1 our results are contradicting this hypothesis. However, the high percentage of unique OTUs at each elevation level suggests that the here assessed biodiversity is strongly niche-driven. Depending on dataset only a low proportion was found at all 6 elevation level indicating a small degree of cosmopolitanism (Fonseca et al. 2014). Nevertheless, we assume that the here observed diversity patterns are likely highly influenced by methodological issues as already discussed above. Additionally, former studies have shown that amount of soil used for extraction is influencing species detection rate including macro- and microorganisms (Ranjard et al. 2003; Dopheide et al. 2019).

It is common knowledge that when working with eDNA replication including biological and technical replication (e.g. PCR replicates) is of uttermost importance (Ranjard et al. 2003; Ficetola et al. 2015; Fonseca 2018; Nichols et al. 2018; Taberlet et al. 2018; Beentjes et al. 2019) due to the heterogenous nature of eDNA. Here we took three biological replicates per elevation level. Taking biological samples is indispensable as replication allows for the estimation of variance of results due to local heterogeneity among others (Searle et al. 2016; Taberlet et al. 2018). Additionally, it has been shown that technical replicates (including PCR replicates) can further increase species identification rate and limit false negative results

(Ficetola et al. 2015; Fonseca 2018; Nichols et al. 2018). Therefore, we took two technical replicates from each biological replicate. Nevertheless, conducted rarefaction curves as well as the fact that on average only 6.9% of all recovered OTUs were shared between all three biological replicates of each elevation level is indicating that number of taken replicates was inadequately scaled to assess total existing diversity on local scale (per elevation level) as well as on larger spatial scale (between elevation level).

On the one hand, this is contributed to the fact that both datasets were dominated by a high percentage of non-target DNA, indicated by a high number of non-assignments, likely originating from microorganisms including prokaryotes and fungi (Siddall et al. 2009; Yang et al. 2013; Horton et al. 2017; Mioduchowska et al. 2018). As the diversity and abundance of most prokaryotes and fungi are directly linked to local microhabitats it can be expected that high proportion of differences between assessed differences in species composition are attributed to these non-target species. On the other hand, sample size for assessing total existing macroorganisms diversity was not sufficient. Former studies have shown that the extraction of DNA from a larger amount of soil, results in a more accurate picture of local diversity (Dopheide et al. 2019). Additionally, the limited scale of the study has resulted in a narrowed picture of existing diversity. If a complete mountain belt would have been sampled it is likely that assessed degree of diversity would have increased by several orders of magnitude (Rahbek 1995), resulting in a more comprehensive picture of degree of diversity associated with each elevation level.

In contrast to studies following traditional sampling methods with later identification of collected specimens on basis of morphological characters, this study presented here is based on molecular methods. As already discussed can the use of eDNA be challenging due to methodological issues including primer bias, reference, sampling strategy and scale of sampled habitat. Additionally, it must be taken into account that, while morphological based identification allows for determining a time frame of species occurrence within the habitat, eDNA

degradation and persistence is prohibiting definite statements about species time of occurrence. Abiotic and biotic factors which can strongly differ between soil habitats are directly influencing the speed and pace of eDNA degradation (Ogram et al. 1988; Paget et al. 1998; Demaneche et al. 2001; Levy-Booth et al. 2007). Although it is well known that extracellular DNA is generally rapidly degrading in soils (Sirois and Buckley 2019), time of persistence can be extended due to a complex interplay of chemical, biological and physical factors (Taberlet et al. 2018) distorting assessments of present biodiversity. After being released into the environment, extracellular DNA is prone to degradation by enzymatic nucleases. By binding to inorganic and organic surface-reactive particles such as clay, sand, silt and humic substances, phosphate and cation bonds (Taberlet et al. 2018) are formed, enhancing the time DNA can persist in the soil (Ogram et al. 1988; Paget et al. 1998; Demaneche et al. 2001; Levy-Booth et al. 2007). Furthermore, next to the binding capacity is soil composition directly influencing microbial diversity and abundance, which in turn is directly influencing decay processes of eDNA (Eichmiller et al. 2016). Furthermore is temperature directly influencing the fate of eDNA in soil (Strickler et al. 2015; Eichmiller et al. 2016). For aquatic systems it has been shown, that lower temperatures increase the time of persistence of eDNA. An increase in time of persistence will likely increase probability to recover rare species as eDNA gets accumulated over time. In views to this study it is likely that due to lower temperatures at the higher elevation level, the extracted eDNA contained a higher percentage of DNA derived from organisms not any more present at the habitat.

Conclusion

The data presented here is the result of a preliminary study aiming to uncover levels of soil diversity in the Caucasus region, never described before, by using NGS. This study proofed the applicability of metabarcoding to assess differences in the degree of diversity over an height gradient. Despite the fact that this study showed that metabarcoding of soil sampling is a promising tool for future monitoring studies, it must be admitted that more work has to be done to enfold the complete power of it. Next to the completion of existing barcode reference databases further studies are needed to develop suitable primers for the assessment of soil metazoan communities which are limiting the co-amplification of non-target DNA to a minimum.

To our knowledge, this is the first metabarcoding study based on eDNA from soil which compares the suitability of *in-silico* post-clustering filtering strategies for the assessment of metazoan community composition. The study showed that choice of filtering strategy can significantly influences assessed community patterns. Therefore, it would be desirable to develop best-practice guidelines to ensure comparability of studies allowing for the detection of general ecological patterns e.g. the development of endemism depending on an elevational gradient.

Literature

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403–410.

André H.M., Noti M.-I., Lebrun P. 1994. The soil fauna: the other last biotic frontier. Biodiversity & Conservation. 3:45–56.

Andújar C., Arribas P., Gray C., Bruce C., Woodward G., Yu D.W., Vogler A.P. 2018a. Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. Molecular Ecology. 27:146–166.

Andújar C., Arribas P., Ruzicka F., Crampton-Platt A., Timmermans M.J., Vogler A.P. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. Molecular Ecology. 24:3603–3617.

Andújar C., Arribas P., Yu D.W., Vogler A.P., Emerson B.C. 2018b. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology. 27:3968–3975.

Anslan S., Nilsson R.H., Wurzbacher C., Baldrian P., Tedersoo L., Bahram M. 2018. Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. MycoKeys. 39:29-40.

Araújo M.B., Pearson R.G., Rahbek C. 2005. Equilibrium of Species' Distributions with Climate. Ecography. 28:693–695.

Baker G., Carter P., Curry J., Cultreri O., Beck A. 1998. Clay content of soil and its influence on the abundance of Aporrectodea trapezoides Duges (Lumbricidae). Applied Soil Ecology. 9:333–337.

Barsoum N., Bruce C., Forster J., Ji Y.-Q., Douglas W.Y. 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. 101:313–323.

Baselga A., Orme C.D.L. 2012. betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution. 3:808–812.

Beentjes K.K., Speksnijder A.G., Schilthuizen M., Hoogeveen M., van der Hoorn B.B. 2019. The effects of spatial and temporal replicate sampling on eDNA metabarcoding. PeerJ. 7:e7335.

Beng K.C., Tomlinson K.W., Shen X.H., Surget-Groba Y., Hughes A.C., Corlett R.T., Slik J.W.F. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. Scientific Reports. 6:24965.

Block W. 1990. Cold tolerance of insects and other arthropods. Philosophical Transactions of the Royal Society of London. B, Biological Sciences. 326:613–633.

Borrell Y.J., Miralles L., Do Huu H., Mohammed-Geba K., Garcia-Vazquez E. 2017. DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. PloS one. 12:e0183347.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

Brown S.P., Veach A.M., Rigdon-Huss A.R., Grond K., Lickteig S.K., Lothamer K., Oliver A.K., Jumpponen A. 2015. Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecology. 13:221–225.

Butchart S.H.M., Walpole M., Collen B., Strien A. van, Scharlemann J.P.W., Almond R.E.A., Baillie J.E.M., Bomhard B., Brown C., Bruno J., Carpenter K.E., Carr G.M., Chanson J., Chenery A.M., Csirke J., Davidson N.C., Dentener F., Foster M., Galli A., Galloway J.N., Genovesi P., Gregory R.D., Hockings M., Kapos V., Lamarque J.-F., Leverington F., Loh J., McGeoch M.A., McRae L., Minasyan A., Morcillo M.H., Oldfield T.E.E., Pauly D., Quader S., Revenga C., Sauer J.R., Skolnik B., Spear D., Stanwell-Smith D., Stuart S.N., Symes A., Tierney M., Tyrrell T.D., Vié J.-C., Watson R. 2010. Global Biodiversity: Indicators of Recent Declines. Science. 328:1164–1168.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 7:335–336.

Chen H., Boutros P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics. 12:35.

Chmielewski F.-M., Rötzer T. 2002. Annual and spatial variability of the beginning of growing season in Europe in relation to air temperature changes. Climate Research. 19:257–264.

Clarke L.J., Beard J.M., Swadling K.M., Deagle B.E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 7:873–883.

Coleman D.C., Crossley Jr D.A., Hendrix P.F. 2004. Fundamentals of Soil Ecology. Academic Press.

Conway J.R., Lex A., Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940.

Cowart D.A., Pinheiro M., Mouchel O., Maguer M., Grall J., Miné J., Arnaud-Haond S. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one. 10:e0117562.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W., Power D.M., Packer M., Carvalho G.R., Blaxter M.L., Lambshead P.J.D., Thomas W.K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology. 19:4–20.

Curry J.P. 2004. Factors affecting the abundance of earthworms in soils. Earthworm Ecology. 9:91-114.

Curtis T.P., Sloan W.T., Scannell J.W. 2002. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences. 99:10494–10499.

De Ley P., De Ley I.T., Morris K., Abebe E., Mundo-Ocampo M., Yoder M., Heras J., Waumann D., Rocha-Olivares A., Jay Burr A. 2005. An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences. 360:1945–1958.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F., Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology letters. 10:20140562.

Decaëns T. 2010. Macroecological patterns in soil communities. Global Ecology and Biogeography. 19:287–302.

Deiner K., Walser J.-C., Mächler E., Altermatt F. 2015. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation. 183:53–63.

Demaneche S., Jocteur-Monrozier L., Quiquampoix H., Simonet P. 2001. Evaluation of biological and physical protection against nuclease degradation of clay-bound plasmid DNA. Applied and Environmental Microbiology 67:293–299.

Derycke S., Vanaverbeke J., Rigaux A., Backeljau T., Moens T. 2010. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. PLoS One. 5: e13716..

Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 14:927–930.

Dopheide A., Xie D., Buckley T.R., Drummond A.J., Newcomb R.D. 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution. 10:120–133.

Douglas W.Y., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Edgar R.C., Haas B.J., Clemente J.C., Quince C., Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 27:2194–2200.

Eichmiller J.J., Best S.E., Sorensen P.W. 2016. Effects of Temperature and Trophic State on Degradation of Environmental DNA in Lake Water. Environmental Science & Technology. 50:1859–1867.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Escobar F., Halffter G., Arellano L. 2007. From forest to pasture: an evaluation of the influence of environment and biogeography on the structure of beetle (Scarabaeinae) assemblages along three altitudinal gradients in the Neotropical region. Ecography. 30:193–208.

Ficetola G.F., Pansu J., Bonin A., Coissac E., Giguet-Covex C., De Barba M., Gielly L., Lopes C.M., Boyer F., Pompanon F. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources. 15:543–556.

Fierer N., Breitbart M., Nulton J., Salamon P., Lozupone C., Jones R., Robeson M., Edwards R.A., Felts B., Rayhawk S. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Applied and Environmental Microbiology. 73:7059–7066.

Fonseca V., Sinniger F., Gaspar J., Quince C., Creer S., Power D.M., Peck L.S., Clark M.S. 2017. Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. Scientific Reports. 7:6094.

Fonseca V.G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources. 18:923–926.

Fonseca V.G., Carvalho G.R., Nichols B., Quince C., Johnson H.F., Neill S.P., Lambshead J.D., Thomas W.K., Power D.M., Creer S. 2014. Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. Global Ecology and Biogeography. 23:1293–1302.

Frøslev T.G., Kjøller R., Bruun H.H., Ejrnæs R., Brunbjerg A.K., Pietroni C., Hansen A.J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications. 8:1–11.

Fu C., Hua X., Li J., Chang Z., Pu Z., Chen J. 2006. Elevational patterns of frog species richness and endemic richness in the Hengduan Mountains, China: geometric constraints, area and climate effects. Ecography. 29:919–927.

Goodwin S., McPherson J.D., McCombie W.R. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 17:333-351.

Grey E.K., Bernatchez L., Cassey P., Deiner K., Deveney M., Howland K.L., Lacoursière-Roussel A., Leong S.C.Y., Li Y., Olds B. 2018. Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. Scientific Reports. 8:1–10.

Grieve I., Proctor J., Cousins S. 1990. Soil variation with altitude on volcan Barva, Costa Rica. Catena. 17:525–534.

Grytnes J.A. 2003. Species-richness patterns of vascular plants along seven altitudinal transects in Norway. Ecography. 26:291–300.

Grytnes J.-A., McCain C.M. 2007. Elevational trends in biodiversity. Encyclopedia of Biodiversity. 2:1–8.

Gullan P.J., Cranston P.S. 2014. The insects: an outline of entomology. John Wiley & Sons.

Hadziavdic K., Lekang K., Lanzen A., Jonassen I., Thompson E.M., Troedsson C. 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PloS one. 9: e87624.

Hågvar S. 1998. The relevance of the Rio-Convention on biodiversity to conserving the biodiversity of soils. Applied Soil Ecology. 9:1–7.

Hallmann C.A., Sorg M., Jongejans E., Siepel H., Hofland N., Schwan H., Stenmans W., Müller A., Sumser H., Hörren T., Goulson D., Kroon H. de. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PloS one. 12:e0185809.

Hebert P.D., Ratnasingham S., De Waard J.R. 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:S96–S99.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003b. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:313–321.

Heinrich B. 2013. The hot-blooded insects: strategies and mechanisms of thermoregulation. Springer Science & Business Media.

Heller K., Rulik B. 2016. Ctenosciara alexanderkoenigi sp. n.(Diptera: Sciaridae), an exotic invader in Germany? Biodiversity Data Journal. 4: e6460-1

Hodkinson I.D. 2005. Terrestrial insects along elevation gradients: species and community responses to altitude. Biological Reviews. 80:489–513.

Hoorman J.J. 2011. The role of soil protozoa and nematodes. Fact Sheet: Agriculture and Natural Resources.(Smith KL), The Ohio State University Extension, Colombus, Ohio.

Horton D.J., Kershner M.W., Blackwood C.B. 2017. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. European Journal of Soil Biology. 80:43–48.

Huse S.M., Welch D.M., Morrison H.G., Sogin M.L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology. 12:1889–1898.

Kaplin V. 2019. New Species of the Genus Allopsontus Silv.(Archaeognatha, Machilidae) from the Caucasus and Tajikistan. Entomological Review. 99:393–402.

Kessler M., Hofmann S., Krömer T., Cicuzza D., Kluge J. 2011. The impact of sterile populations on the perception of elevational richness patterns in ferns. Ecography. 34:123–131.

Koziol A., Stat M., Simpson T., Jarman S., DiBattista J.D., Harvey E.S., Marnane M., McDonald J., Bunce M. 2019. Environmental DNA metabarcoding studies are critically affected by substrate selection. Molecular Ecology Resources. 19:366–376.

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

Kvist S. 2014. Does a global DNA barcoding gap exist in Annelida? Mitochondrial DNA Part A. 27:2241–2252.

Lavelle P., Decaëns T., Aubert M., Barot S., Blouin M., Bureau F., Margerie P., Mora P., Rossi J.-P. 2006. Soil invertebrates and ecosystem services. European Journal of Soil Biology. 42:S3–S15.

Leese F., Altermatt F., Bouchez A., Ekrem T., Hering D., Meissner K., Mergen P., Pawlowski J., Piggott J., Rimet F. 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes. 2:e11321.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T., Machida R.J. 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. 10:34.

Levy-Booth D.J., Campbell R.G., Gulden R.H., Hart M.M., Powell J.R., Klironomos J.N., Pauls K.P., Swanton C.J., Trevors J.T., Dunfield K.E. 2007. Cycling of extracellular DNA in the soil environment. Soil Biology and Biochemistry. 39:2977–2991.

Lindeque P.K., Parry H.E., Harmer R.A., Somerfield P.J., Atkinson A. 2013. Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PloS one. 8: e81327..

Lomolino M.V. 2001. Elevation gradients of species-density: historical and prospective views. Global Ecology and Biogeography. 10:3–13.

Machac A., Janda M., Dunn R.R., Sanders N.J. 2011. Elevational gradients in phylogenetic structure of ant communities reveal the interplay of biotic and abiotic constraints on diversity. Ecography. 34:364–371.

Macheriotou L., Guilini K., Bezerra T.N., Tytgat B., Nguyen D.T., Phuong Nguyen T.X., Noppe F., Armenteros M., Boufahja F., Rigaux A. 2019. Metabarcoding free-living marine nematodes using curated 18S and CO1 reference sequence databases for species-level taxonomic assignments. Ecology and Evolution. 9:1211–1226.

Malhi Y., Girardin C.A., Goldsmith G.R., Doughty C.E., Salinas N., Metcalfe D.B., Huaraca Huasco W., Silva-Espejo J.E., del Aguilla-Pasquell J., Farfán Amézquita F. 2017. The variation of productivity and its allocation along a tropical elevation gradient: a whole carbon budget perspective. New Phytologist. 214:1019–1032.

Marin I., Palatov D. 2019. A new species of the genus Niphargus (Crustacea: Amphipoda: Niphargidae) from the south-western part of the North Caucasus. Zoology in the Middle East. 4:1–11.

Marquina D., Esparza-Salas R., Roslin T., Ronquist F. 2019. Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular Ecology Resources. 19:1516-1530

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.

Martynov A.V., Žiak M. 2019. A new micropterous species of Leuctra (Plecoptera: Leuctridae) from the Lesser Caucasus (Georgia). Zootaxa.4:581-588.

May R.M. 1990. How many species? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 330:293–304.

May R.M. 2010. Tropical Arthropod Species, More or Less? Science. 329:41–42.

McCain C.M. 2007. Area and mammalian elevational diversity. Ecology. 88:76–86.

McCain C.M., Grytnes J. 2001. Elevational gradients in species richness. John Wiley & Sons Chichester. pp. 1-10.

McCredie T., Parker C., Abbott I. 1992. Population dynamics of the earthworm Aporrectodea trapezoides (Annelida: Lumbricidae) in a Western Australian pasture soil. Biology and fertility of soils. 12:285–289.

Meier R., Zhang G., Ali F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Systematic Biology. 57:809–813.

Mioduchowska M., Czyż M.J., Gołdyn B., Kur J., Sell J. 2018. Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too "universal"? PloS one. 13: e0199609..

Mittermeier R.A., Myers N., Mittermeier C.G., Robles Gil P. 1999. Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions. CEMAX, S.A., Mexico City.

Mora C., Tittensor D.P., Adl S., Simpson A.G.B., Worm B. 2011. How Many Species Are There on Earth and in the Ocean? PLOS Biology. 9:e1001127.

Morinière J., de Araujo B.C., Lam A.W., Hausmann A., Balke M., Schmidt S., Hendrich L., Doczkal D., Fartmann B., Arvidsson S. 2016. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PloS one. 11:e0155497.

Myers N. 1988. Threatened biotas: "Hot spots" in tropical forests. Environmentalist. 8:187–208.

Myers N. 1990. The biodiversity challenge: Expanded hot-spots analysis. Environmentalist. 10:243–256.

Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403:853–858.

Nemergut D.R., Costello E.K., Hamady M., Lozupone C., Jiang L., Schmidt S.K., Fierer N., Townsend A.R., Cleveland C.C., Stanish L. 2011. Global patterns in the biogeography of bacterial taxa. Environmental Microbiology. 13:135–144.

Nichols R.V., Vollmers C., Newsom L.A., Wang Y., Heintzman P.D., Leighton M., Green R.E., Shapiro B. 2018. Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources. 18:927–939.

Nielsen U.N., Ayres E., Wall D.H., Bardgett R.D. 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. European Journal of Soil Science. 62:105–116.

Ogram A., Sayler G.S., Gustin D., Lewis R.J. 1988. DNA adsorption to soils and sediments. Environmental Science & Technology. 22:982–984.

Paget E., Lebrun M., Freyssinet G., Simonet P. 1998. The fate of recombinant plant DNA in soil. European Journal of Soil Biology. 34:81–88.

Prange H.D. 1990. Temperature regulation by respiratory evaporation in grasshoppers. Journal of Experimental Biology. 154:463–474.

Quince C., Lanzén A., Curtis T.P., Davenport R.J., Hall N., Head I.M., Read L.F., Sloan W.T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods. 6:639-641.

R CoreTeam. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahbek C. 1995. The elevational gradient of species richness: a uniform pattern? Ecography. 18:200–205.

Ranjard L., Lejon D.P., Mougel C., Schehrer L., Merdinoglu D., Chaussod R. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environmental Microbiology. 5:1111–1120.

Ratnasingham S., Hebert P.D.N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes. 7:355–364.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Rolland C. 2003. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. Journal of Climate. 16:1032–1046.

Rosenzweig M.L. 1995. Species diversity in space and time. Cambridge University Press.

Rowe R.J. 2009. Environmental and geometric drivers of small mammal diversity along elevational gradients in Utah. Ecography. 32:411–422.

Sanders N.J. 2002. Elevational gradients in ant species richness: area, geometry, and Rapoport's rule. Ecography. 25:25–32.

Sanders N.J., Rahbek C. 2012. The patterns and causes of elevational diversity gradients. Ecography. 35:1–3.

Schaefer M., Schauermann J. 1990. The soil fauna of beech forests: comparison between a mull and a moder soil. Pedobiologia. 34:299–314.

Schloss P.D., Gevers D., Westcott S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one. 6:e27310.

Searle D., Sible E., Cooper A., Putonti C. 2016. 18S rDNA dataset profiling microeukaryotic populations within Chicago area nearshore waters. Data in brief. 6:526–529.

Siddall M.E., Fontanella F.M., Watson S.C., Kvist S., Erséus C. 2009. Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. Systematic Biology. 58:445–451.

Sirois S.H., Buckley D.H. 2019. Factors governing extracellular DNA degradation dynamics in soil. Environmental microbiology reports. 11:173-184.

Somervuo P., Yu D.W., Xu C.C., Ji Y., Hultman J., Wirta H., Ovaskainen O. 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution. 8:398–407.

Stoeck T., Bass D., Nebel M., Christen R., Jones M.D., BREINER H., Richards T.A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology. 19:21–31.

Stork N.E. 1988. Insect diversity: facts, fiction and speculation. Biological journal of the Linnean Society. 35:321–337.

Stotz D.F., Fitzpatrick J.W., Parker III T.A., Moskovits D.K. 1996. Neotropical birds: ecology and conservation. University of Chicago Press.

Strickler K.M., Fremier A.K., Goldberg C.S. 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biological Conservation. 183:85–92.

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

Taberlet P., PRUD'HOMME S.M., Campione E., Roy J., Miquel C., Shehzad W., Gielly L., Rioux D., Choler P., CLÉMENT J. 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology. 21:1816–1820.

Tang C.Q., Leasi F., Obertegger U., Kieneke A., Barraclough T.G., Fontaneto D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences. 109:16208–16212.

Tarkhnishvili D., Gavashelishvili A., Mumladze L. 2012. Palaeoclimatic models help to understand current distribution of Caucasian forest species. Biological Journal of the Linnean Society. 105:231–248.

Teslenko V.A., Palatov D.M., Semenchenko A.A. 2019. Description of new apterous winter species of Leuctra (Plecoptera: Leuctridae) based morphology and DNA barcoding and further records to stonefly fauna of the Caucasus, Georgia. Zootaxa. 4585:547–560.

Vetaas O.R., Grytnes J. 2002. Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. Global Ecology and Biogeography. 11:291–301.

Wall D.H., Fitter A.H., Paul E.A. 2005. Developing new perspectives from advances in soil biodiversity research. Biological Diversity and Function in Soils. Cambridge University Press, Cambridge. pp. 31–43.

Wang C., Wang Q., Long R., Jing Z., Shi H. 2004. Changes in plant species diversity and productivity along an elevation gradient in an alpine meadow. Acta Phytoecological Sinica. 28:240–245.

Wheeler Q.D., Raven P.H., Wilson E.O. 2004. Taxonomy: Impediment or Expedient? Science. 303:285–285.

Whitman D.W. 1988. Function and evolution of thermoregulation in the desert grasshopper Taeniopoda eques. The Journal of Animal Ecology. 57:369–383.

Whitman W.B., Coleman D.C., Wiebe W.J. 1998. Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences. 95:6578–6583.

Whittaker R.H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. Ecological Monographs. 22:1–44.

Whittaker R.H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. Ecological Monographs. 30:279–338.

Whittaker R.H. 1967. Gradient analysis of vegetation. Biological Reviews. 42:207–264.

Whittaker R.H., Niering W.A. 1965. Vegetation of the Santa Catalina Mountains, Arizona: a gradient analysis of the south slope. Ecology. 46:429–452.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.

Wickham H., Francois R., Henry L., Müller K. 2015. dplyr: A grammar of data manipulation. R package version 0.8.3. 3.

Wiemers M., Fiedler K. 2007. Does the DNA barcoding gap exist?–a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology. 4:8.

Williams C., Osman M. 1960. A new approach to the problem of the optimum temperature for insect activity. Journal of Animal Ecology. 29:187–189.

Yang C., Ji Y., Wang X., Yang C., Douglas W.Y. 2013. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. Science China Life Sciences. 56:73–81.

Yang C., Wang X., Miller J.A., de Blécourt M., Ji Y., Yang C., Harrison R.D., Douglas W.Y. 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators. 46:379–389.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Zhou X., Li Y., Liu S., Yang Q., Su X., Zhou L., Tang M., Fu R., Li J., Huang Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience. 2:4.

Chapter II

Monitoring Biodiversity Shifts – How Arthropod Communities Respond to Human Managed Ecological Reconstructions

Abstract

The timely assessment of biodiversity is indispensable for the protection of local species communities. With metabarcoding it is possible to monitor biodiversity in a timely and cost-efficient manner, allowing the evaluation of the current condition of a habitat based on total existing biodiversity, rather than on indicator species. Nevertheless, there are still no best practice recommendations for the assessment of insect community composition via metabarcoding. Here we show that the choice of primer, marker and reference database directly influences species detection rate, but that a careful selection of the above-mentioned parameters allows for the reliable and timely assessment of local insect diversity. Using a 'one-locus-several-primers'-strategy we were able to monitor changes in insect species community composition along a renaturation gradient from Norway spruce to European beech. Our study underlines the importance of renaturation measures for Germany's insect diversity, as many native insect species rely on deciduous forests, which have over the last centuries increasingly been replaced with non-native boreal forests. We found that the dominating tree species directly influences insect community composition, but that with ongoing renaturation measures species communities found at spruce monocultures are in terms of composition more and more shifted towards species communities associated with beech monocultures. This study shows the applicability of metabarcoding for conservation purposes as it allows for the timely and cost efficient monitoring of species communities.

Introduction

Germanys Biodiversity at Risk

In 2015 the Bundesamt für Naturschutz determined the number of animal species described for Germany to 48.000. Although Germanys fauna is thought to be well studied, it can be expected that several species are still waiting to be discovered. Only recently a new species of Sciaridae (Insecta: Diptera) was found in the garden of a zoological research museum (Heller and Rulik 2016), a museum which is specialized on the assessment of new species and located in the center of a lively city. Despite the lack of knowledge about how many species exist in Germany, several studies have already shown that Germany is suffering from a dramatic insect decline (Hallmann et al. 2017; Homburg et al. 2019). In May 2013 the 'IUCN Red List of Threatened Species' released a call of action entitled "Germany's biodiversity at risk" (Sánchez 2013). The report is based on the results the IUCN gained from the analysis of data collected for 1383 species. Six percent of these species were assigned to the status "threatened", an additional 8% to "near threatened" and another three species already went extinct in the last few years. Unfortunately, the data are incomplete and information for large important taxonomic groups like Hymenoptera, and Diptera are missing (Sánchez 2013). Further studies draw an even darker picture of the future for Germany's biodiversity. In 2015 the 'Bundesamt für Naturschutz' (BfN) released a report about species conservation in Germany, based on the red lists conducted for Germany by several experts (Ludwig et al. 1996; Haupt et al. 2009; Binot-Hafke et al. 2011; Gruttke et al. 2016). In

2011 Binot-Hafke and colleagues assigned 2704 out of 6057 invertebrate species to the category ‘threatened’, ‘extremely rare’ or ‘extinct’, making up a total of 45,8%. Even more dramatic is the fact that it was not possible to stop the population decline of 1234 Invertebrate species of which scientists were aware of at the very latest since 1998 (Binot-Hafke et al. 2011). Finally, in 2018 a study brought worldwide attention to the ongoing insect decline, by stating that insect biomass of protected sites in Germany has declined by 75% over the last 27 years (Hallmann et al. 2017).

Conservation Measures

In order to stop the ongoing biodiversity loss, efforts have been made to reduce scientifically proven negative influences on species diversity to a minimum e.g. by reduction of emissions (Phelps et al. 2012), the establishment of protected areas on a large scale but also by the foundation of national parks on a local scale (Naughton-Treves et al. 2005). Since the first recognition of diversity loss, 14 areas distributed across Germany were declared as national parks making up a total of 16 German national parks. The Eifel National Park, which was founded in 2004, is spread over an area of 10.880 hectares of which 8.190 hectares are covered with forest. A large biotope mapping project conducted between 2003 and 2005 assigned 50% of all trees in the Eifel National Park as invasive alien species e.g. spruce (figure II.1).

One of the main tasks of national parks is the protection of the native fauna and flora. In order to meet these requirements measures were undertaken to restore the previous status of the Eifel National Park area, which was mainly characterized by deciduous mixed forests. Next to intensive large-scale deforestation activities in some areas, spruces were reduced by actions with a low intervention intensity e.g. by rejuvenation measures like underplanting. Measures of the first category are very effective but usually go along with the complete destruction of the affected

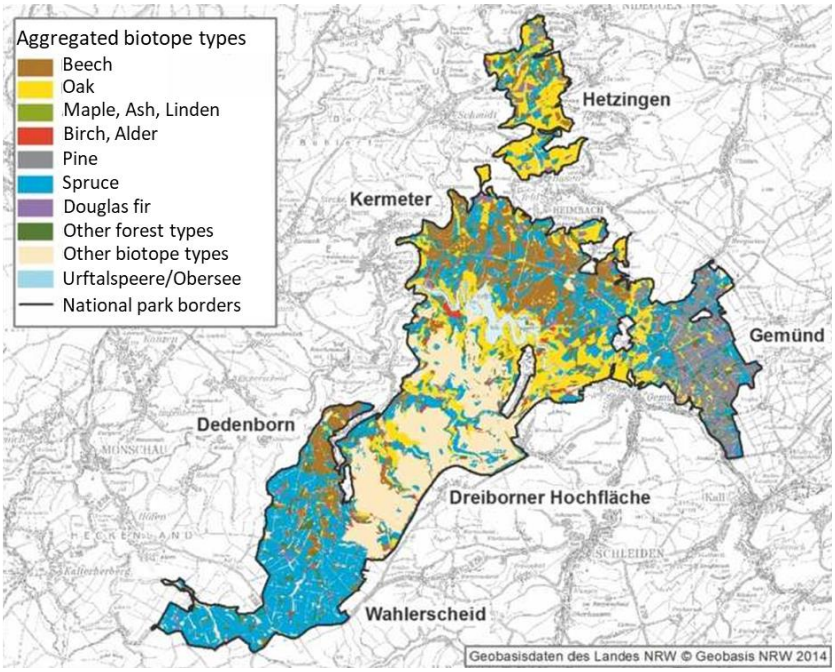


Figure II.1 Existing forest types and its distribution in the Eifel National Park as monitored between 2003 and 2005. Especially in the southern parts of the National Park are forests dominated by spruces. (© Nationalparkverwaltung Eifel)

ecosystem. In contrast to that, actions with a low intervention intensity transform the habitat over a longer period of time allowing for the slow transition of a spruce forest to the desired beech monoculture. This process is supposed to go along with changes in faunal species communities whose composition is directly linked to the current status of the ecosystems. Former studies have already shown that the dominating tree species has a major influence on abiotic (e.g. light, humidity, ph-value) (Barkman 1992; Leuchner et al. 2007, 2011) and biotic factors (e.g. symbionts) (Magura et al. 2002; Konôpka et al. 2013) and thereby on local fauna and flora (Schmidt and Weckesser 2001; Barbier et al. 2008). In the literature, several examples can be found linking species presence and absence with certain environmental conditions like solar radiation (Gossner 2009), microclimate (Seibold et al. 2016) and food resources (Feeny 1970).

The Forest Habitat

Temperate deciduous forests like beech forests are an extreme habitat. No other forest type undergoes more dramatic changes throughout a one year period (Röhrig and Ulrich 1991; Ulyshen 2011). Life cycles in deciduous forests are highly driven by seasonal variations in light, precipitation and temperature. Depending on age of the forest and thickness of the stands, beech forests form monolayer canopies during summer season which have a major influence on the amount of light reaching the forest floor (Knapp and Jeschke 1991; Scherzinger 1996) and thereby on seasonal gradients in temperature (Frady et al. 2007). Later in the year, when temperature decreases, beech trees start to shed their leaves in order to get ready for winter dormancy. With the beginning of winter, a large amount of foliage is accumulated on the forest floor in layers of various thicknesses. These sheet-like layers of beech leaves have a toxic effect on many tree seedlings and are hardly promoting understory growth (Augusto et al. 2003). In spring, when temperatures increase and higher light levels reach the forest floor, the growing season is induced. Over the next weeks, new leaves and shoots emerge which will finally form a new canopy shading the forest floor. Compared to deciduous forests, evergreen boreal forests are at first sight less constrained by the seasons. Throughout the whole year, the thickness of the canopy does not undergo dramatic changes. Nevertheless, abiotic factors still have a major influence on the onset, course and offset of the growing season. During the growing season, coniferous trees form new needle sets. Unlike deciduous trees, lost or damaged leaves or rather needles are not replaced. Usually a branch of the Norwegian spruce consists of 5 to 8 needle sets (Sander and Eckstein 2001), one formed each year. After the last needle set has been formed, the branch will stop growing. Within a boreal forest, single dead trees usually leave a long-lasting gap. The prevailing all year round cold and shady climate in the boreal forest hampers tree growth. Nevertheless, spruce saplings can hold out in the understory for up to 250 years until thinning out of the canopy allows increased growth (Leibundgut 1984; Scherzinger 1996). Within deciduous forests the death of a single tree affects the ecosystem for a much shorter period of

time. The resulting clearing is flooded with light, which enables the growth of light demanding species next to beech regenerations. Nevertheless, when the latter finally reaches the height of the canopy the forest floor will soon be shaded again and the forest floor will again be covered in beech foliage (Scherzinger 1996). These two extremes already indicate that next to the dominating tree species, also the age of the forest has a direct influence on species communities associated with it.

In general, it can be said that regardless of the dominating tree species, each established forest has undergone several forest-development stages (Leibundgut 1982). The first trees settling on a deforested area are usually light demanding species like members of the genus Betula (Fagales: Betulaceae). Later, more and more shade tree species occupy the space underneath the light demanding trees. Together with the light-demanding trees they form a transitional forest. After some years the shade tree species will over tower the light demanding trees and as a result the latter will disappear from the habitat and a circuit forest is formed (Leibundgut 1982). In the next decades the circuit forest will undergo several phases. During the optimal phase, a monolayer canopy characterizes the forest. During the aging phase single trees will die, which goes along with the formation of canopy gaps. The forest is now either entering the rejuvenation phase during which new emerging, shade tolerant trees are rapidly replacing dead trees. Alternatively, the circuit forests enters in the decay phase. During the decay phase several single gaps are formed which with time will merge, until finally large deforested areas are formed. The deforested areas are then colonized again by light-demanding trees (Leibundgut 1982).

Not only the dominating trees undergo succession. Moreover, total existing flora, including grasses and mosses is subject to change over the years. A forest is a stratified habitat consisting roughly of five layers (figure II.2). The ground layer encompasses plants like mosses and lichens. This layer is highly influenced by all strata over-towering it (Barbier et al. 2008; Budde et al. 2011). Not only microclimate conditions are mainly influenced by the conditions formed by the higher strata (Barbier et al. 2008), but also plant debris originating from the latter are accumulated here

Figure II.2 Schematic Visualization of the typical stratification characterizing European forests

(Augusto et al. 2003). The next layer is the herb layer, which is characterized by herbaceous plants like grasses, herbs, wildflowers and ferns. The herb layer is over-towered by the shrub layer, which is characterized by a woody vegetation like bushes and brambles or young immature trees. Subsequently, the next layer, the understory, encompasses immature trees of the habitat dominating trees. As soon as a gap opens in the canopy, trees of the undershrub will enhance growth and refill the gap (Leibundgut 1982). Finally, the uppermost stratum is the canopy layer.

As already mentioned, depending on forest type seasonal variation in thickness of the canopy can occur. The canopy has a major influence on light intensity reaching the understory, the shrub- and the herb layer (Barkman 1992; Leuchner et al. 2007). Species composition of the herb and shrub vegetation is mainly driven by hummus accumulation, pH-value of the soil and light, which differs significantly between deciduous and boreal forests. Budde et al (2011) showed that relative light irradiance was lowest in pure beech stands, while coniferous stands where significantly lighter. Several authors have already pointed out the relevance of light for the development of a diverse herb and under shrub layer (Hill 1979; Jennings et al. 1999; Barbier et al. 2008). Indeed, Budde et al. (2011) showed that the dominating tree species of a forest have major influence on the diversity of the herb and shrub layer. The increase in light as well as the absence of a thick, toxic layer of foliage increases herb diversity within coniferous stands. While in pure beech stands nine herbs and five mosses were found, in pure spruce stands 19 herbs and

11 mosses were sampled (Budde et al. 2011). Additionally, in mixed stands of beech and spruce the number of herbs and mosses was marginally higher (20 herbs, 13 mosses) compared to the number of species found at pure spruce stands. The results of Budde et al. (2011) are in accordance with further studies showing that floral diversity of the undershrub and herb layer changes slowly from pure spruce stand over mixed stands to beech stands (Ammer et al. 2002; Fritz 2006). All studies have in common that species diversity reached a peak at the mixed stands. It can be expected that the more diverse the shrub and herb layers are, the more niches are provided for insects, favoring an increase in insect diversity. We therefore hypothesize that the replacement of spruces by beeches will lead to changes in the existing Insect species community. Furthermore, we argue that monitoring the changes in insect community composition will allow the control of the efficacy of ongoing renaturation measures.

Methods for Assessing Arthropod Diversity

Over the last decades several methods for monitoring community compositions have been developed. One of the most efficient systems for catching flying insects are Malaise traps. Malaise traps are named after the Swedish entomologist René Malaise (1892-1978) who first used this new style of trap for a biodiversity assessment study in Burma (Malaise 1937). The Malaise trap has a tent like structure, which is raised at one end. The construction of the traps is based on a passive sampling strategy. Close to the ground flying organisms enter the tent where a black net forces them to change direction. Most of the insects will try to escape by flying upwards into the direction of light, which is reinforced by a white net forming the roof of the trap. The highest point of the trap is equipped with a bottle of preservative ethanol in which the organisms get finally accumulated and preserved. As the Malaise trap follows a passive sampling strategy its effectiveness highly depends on the structure of the sampled habitat, the setup and

overall on the lifestyle of target organisms. Some organisms can escape from the Malaise traps (Campbell and Hanula 2007), while other insects such as many beetles accidentally escape by dropping to the ground when striking an obstacle like the net of the malaise trap (Juillet 1963). Nevertheless, catches with Malaise traps are very efficient and one bottle can contain up to several thousands of specimens (Geiger et al. 2016).

Due to the direct preservation of caught organisms in ethanol, the collected specimens are usually well preserved. Therefore, Identification can be based either on morphological characteristics or on molecular data. Morphological identification suffers from the taxonomic impediment, which describes the shortage of experts, making identification a time consuming and costly process (Wheeler et al. 2004). In contrast to that, molecular species identification is to a certain extent independent of taxonomic experts, although taxonomists are still needed for the establishment of reference databases and the description of new species. Today, a broad range of species can already be identified with molecular tools (Elbrecht et al. 2019), without having to contact a taxonomist. Molecular species identification encompasses barcoding and metabarcoding. For identifying species on the basis of barcoding, each specimen must be individually treated which goes along with a high temporal expenditure. In contrast to that, metabarcoding allows for the simultaneous identification of thousands of specimens in a timely and cost-efficient manner. All three methods have their advantages and drawbacks. While morphological identification and barcoding provide information about abundances as each specimen is treated individually, metabarcoding is still lacking this feature. Nevertheless, abundance data can provide important information about changes in species communities (Alpha-, beta-, gamma diversity) (Whittaker 1960). Changes in species composition are often early indicators that an ecosystem is losing its balance. However, species presence-absence data is also a valuable source of information. The trade-off between required information density, workload, budget and last but not least project term must be taken into account when choosing an appropriate identification method.

Metabarcoding Arthropod Diversity

Metabarcoding is a straightforward method which has already been implemented into ecological landmark studies (Fonseca et al. 2010). Because of the wide range of applicability of metabarcoding e.g. water samples, soil samples, bulk samples and even feces, no standardized protocol has so far been developed. Even within sample groups, no consensus has so far been found. It has already been shown that some methods e.g. choice of primers and extraction method can perform well for several groups of organisms but will perform poorly for other groups of interest (Drummond et al. 2015; Dopheide et al. 2019). For identification of species accumulated in a bulk sample several extraction methods have so far been tested. In 2010 a study showed that it is possible to extract DNA directly from the preservative ethanol (Shokralla et al. 2010). Two years later it was shown that up to 87% of all taxa included in a bulk sample can be detected by using that noninvasive extraction method (Hajibabaei et al. 2012). Nevertheless, the findings of some later conducted studies contradict these results. Marquina et al. (2019b) showed that depending on extraction method, the insects traits like size or degree of sclerotization play a major role in determining if a species will be detected or not. Concluding, it can be assumed that DNA extraction directly from the preservative ethanol is unreliable. Therefore several studies now use grinding of the specimens (Gibson et al. 2014; Elbrecht et al. 2019). However, the development of a non-destructive extraction method is still desirable as the grinding of species deprives the possibility of a morphological identification at a later stage e.g. when OTUs indicate that a rare, not yet barcoded or even a non-described species is present in the sample.

Next to the extraction method, marker choice is a main factor influencing taxon detection rate as well as the accuracy of species identification against marker-specific reference databases (Andújar et al. 2018). Because of the rapidly expanding reference databases (Ratnasingham and Hebert 2007; Porter and Hajibabaei 2018) and its comparatively good taxonomic resolution

(Meusnier et al. 2008) the 658 bp long region of the Cytochrome c oxidase I subunit (COI) (Folmer et al. 1994) has become the marker of choice for assessing arthropod diversity. Making the COI marker to the standard metazoan marker is not without criticism. Especially when working with eDNA, microbial DNA is often coamplified (Yang et al. 2014). Furthermore, the COI marker is still struggling with taxonomic bias for PCR amplification within certain groups (Yu et al. 2012; Zhou et al. 2013; Clarke et al. 2014; Deagle et al. 2014; Brandon-Mong et al. 2015; Andújar et al. 2018). Today, several primer combinations are available targeting various parts of the Folmer region. The consideration of which primer pair to use is critical for the outcome of the study (Elbrecht et al. 2019). Depending on target organisms some primers can perform poorly as they do not match the template sequence, resulting in substantial bias in taxon detection (Elbrecht and Leese 2015; Piñol et al. 2015). PCR biases mainly result from the fact that the primer binding regions of the COI marker are not highly conserved. By definition, the perfect barcode is a short DNA fragment displaying a highly variable sequence which is flanked by two highly conserved regions. The central variable region is discriminative for all species of the target group, that is, its sequence is uniquely associated to a given species and not shared with others (Taberlet et al. 2018). Primers amplifying the chosen region will perfectly bind to the conserved regions and will ensure that DNA originating from the target species will be amplified without bias while amplification of non-target taxa is prevented (Taberlet et al. 2018). The COI marker is a protein coding gene, which often goes along with third base wobbles, leading to taxonomic bias for PCR amplification within certain groups (Deagle et al. 2014). To avoid these biases, some studies have chosen makers displaying highly conserved primer binding sites like the nuclear ribosomal gene 18S (Creer et al. 2010; Fonseca et al. 2010). Indeed, the 18S marker is more conserved than the COI marker, which is a curse and savior at the same time. On the one hand, the primer binding sites are more strongly conserved, which allows for a broader taxonomic coverage across the eukaryotic domain of life (Clarke et al. 2017). But on the other hand a broad taxonomic coverage is only achieved at the cost of a lower taxonomic resolution, as the central variable region of the target DNA fragment is often not uniquely associated with a given species (Tang et al. 2012). The four most

frequently used arguments for focusing on the COI marker are (a) the availability of an incomparably large reference database, (b) the good taxonomic resolution, (c) advances in the development of robust and reliable bioinformatic tools, assisting with quality filtering and the removal of spurious sequences and (d) the rapid progress in primer design (Andújar et al. 2018).

For the analyses of bulk samples some authors recommend the use of multiple primer sets or multiple marker genes (Alberdi et al. 2018; Zhang et al. 2018) to ensure the possible best taxonomic coverage, without reducing taxonomic resolution. In contrast to that, some authors state that there is no need to employ several primer sets targeting the same maker (Elbrecht et al. 2019). This already indicates that choice of primer plays a major role for the accuracy and completeness of metabarcoding studies. Especially the primer design has taken tremendous steps forward in recent years. Facing the challenge of third base wobbles, a broad palette of degenerate primers has been developed over the years. Many of these primers have been developed to target certain groups of organisms like certain Insect orders e.g. Hemiptera (Park et al. 2011), Hymenoptera (Mitrović and Tomanović 2018) or Coleoptera (Astrin and Stüben 2008), while others were developed to target a broader range of organisms like all members of the phylum Arthropoda (Gibson et al. 2014). Although some of the available COI primers target different regions within the COI gene, many primers target the same region. These primers usually differ in their degree of degeneracy. A degenerate primer is a mixture of DNA oligunucleotides that differ in base composition for one or several nucleotide positions. The higher the proportion of degenerate positions, the higher the degree of degeneracy (Andújar et al. 2018). The use of highly degenerated primers is not without its critics. A higher degree of degeneracy usually goes along with a decrease in specifity, which may result in non-specific amplification, dimerization and primer slippages (Elbrecht et al. 2018, 2019).

Aims of this Study

Part 1 – Evaluation of Methods

Within this study, we tested three primer pairs amplifying the barcode region of two makers for their taxonomic coverage and resolution. We examine the influence of primer and marker choice on resulting beta diversity and calculated community composition patterns. Furthermore, we tested a new extraction method which is based on the application of a lysis buffer followed by salt precipitation.

For taxonomy assignment two different databases were used in order to distinguish if a geographically limited database perform better than a bigger, but less specialized database.

Part II – Ecological analysis

We hypothesize that the replacement of spruces by beeches will lead to changes in the existing Insect species community. Here we use metabarcoding to answer the question if monitoring of changes in insect community composition allows for conclusions about ongoing renaturation measures and if metabarcoding is yet capable to assess community patterns associated with the dominating tree species.

Material and Methods

Sampling Strategy

The here sampled 14 locations were located in the Eifel Nationalpark, which is situated in the south-western part of Germany close to the Belgian border (figure II.3).

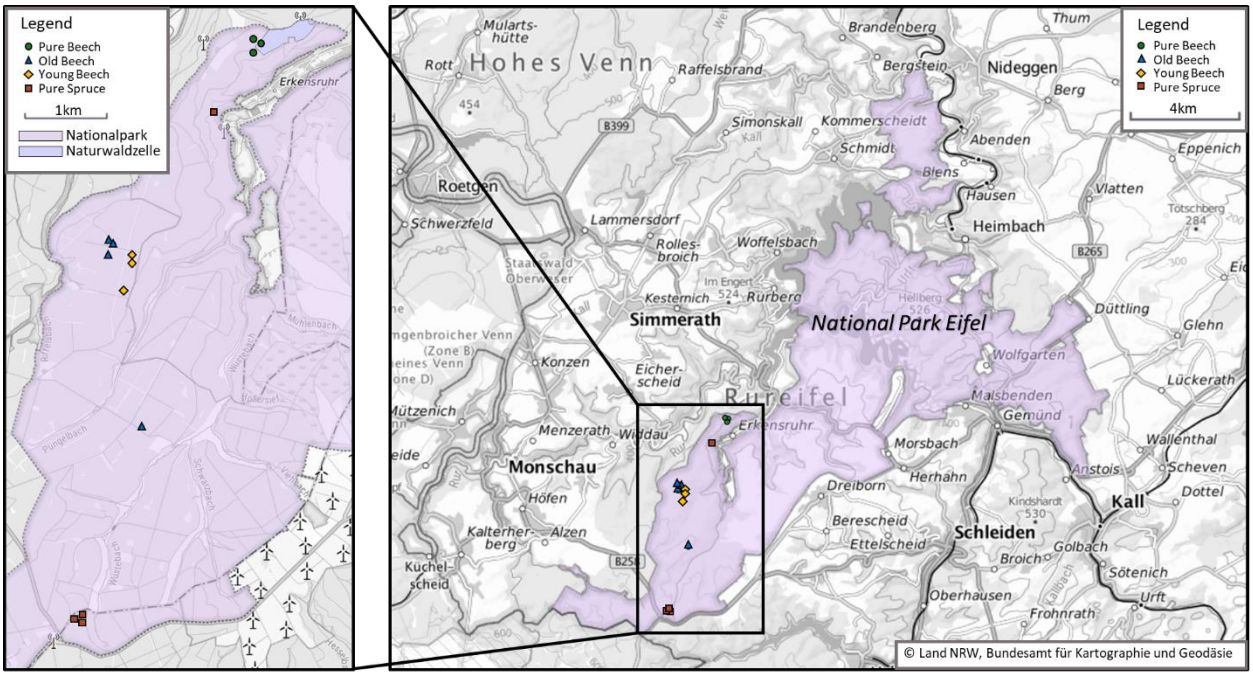


Figure II.3 Map of sampling sites. Map of all 14 sites sampled for this study. Area highlighted in purple corresponds to the current area of the Eifel national park

For this study a forest conversion gradient from Norway Spruce (*Picea abies*) to European Beech (*Fagus sylvatica*) was sampled. To reflect the different stages of conversion from spruce to beech, four forest types were defined: pure beech (PB), old beech (OB), young beech (YB) and pure spruce (PS) (table II.1).

Table II.1 Geographical location and ecological characteristics of the 14 sampling sites. Depicted for each sampling site are the Coordinates (altitude N and latitude E) and the associated forest type

Sample Sites	Coordinates	Forest Type
Sample Site 01	50° 34'11.7984''N 6°21'32.1012''E	Pure Beech
Sample Site 02	50° 34'07.7016''N 6°21'27.3996''E	Pure Beech
Sample Site 03	50° 34'12.9000''N 6°21'27.3996''E	Pure Beech
Sample Site 04	50° 32'44.5992''N 6°20'15.2988''E	Young Beech
Sample Site 05	50° 32'41.3016''N 6°20'15.6984''E	Young Beech
Sample Site 06	50° 32'29.7996''N 6°20'11.1012''E	Young Beech
Sample Site 07	50° 32'29.7996''N 6°20'11.1012''E	Old Beech
Sample Site 08	50° 31'35.1984''N 6°20'25.2996'E	Old Beech
Sample Site 09	50° 32'48.3000''N 6°20'03.4008''E	Old Beech
Sample Site 10	50° 30'17.2008''N 6°19'48.1008''E	Pure Spruce
Sample Site 11	50° 30'18.2988''N 6°19'51.4020''E	Pure Spruce
Sample Site 12	50° 33'15.8004''N 6°21'07.3008''E	Pure Spruce
Sample Site 13	50° 30'16.0056''N 6°19'51.4704''E	Pure Spruce
Sample Site 14	50° 32'49.9632''N 6°20'00.7296''E	Old Beech

The four forest types differed in tree species composition as well as in approximate age of trees. The pure beech and pure spruce sampling sites were located in monoculture stands which were either dominated by beeches or spruces respectively. The pure beech monoculture stands were approximately 180 years old and partly under special protection through North-Rhine Westfalia (Naturwaldzelle) (Sample Site 01). With a mean age of 60 years, the pure spruce stands were substantially younger. Spruces of the same age dominated the young beech sampling sites at which young beeches had been planted only recently. The beeches had therefore not yet reached three meters in size at the time of sampling. At the old beech sampling sites, beeches had already reached a height of more than 3 meters and actions to remove spruces from the forest had already been undertaken.

Figure II.4 Forest conversion gradient: From spruce monocultures over with beeches underplanted spruce forests to beech monocultures. Photos were taken in summer 2016.

In July 2016, 12 Malaise traps were set up in the Eifel National Park, North-Rhine Westfalia, Germany. At the beginning of the study three Malaise traps were placed in each forest type (table II.2). To ensure that the orientation of the Malaise traps did not affect sampling success, the highest point of each Malaise traps was set up pointing south. The traps were left in the field for the full duration of the experiment until April 2017 to ensure that insects were collected from

exactly the same locations. In October 2016 two additional traps (Malaise Trap 13 and Malaise Trap 14) were installed – one in the pure spruce stands (Sample Site 13) and a second in the old beech forests (Sample Site 14). All traps were equipped with a bottle filled with approximately 1 litre of 99,96% pure ethanol over a two week period in July 2016 (13.07-27.07), October 2016 (13.10-27.10), January 2017 (11.01-25.01) and April 2017 (12.04-26.04). The ethanol bottles were replaced after one week in the field. After each weekly collection, the ethanol was replaced to ensure that concentration of the preservative ethanol was stable. After final collection, the traps were left unequipped in the field until the start of the next sampling period. Between October 2016 and January 2017 nine malaise traps were destroyed by heavy snow fall. The damaged traps were replaced at the start of the new sampling season in January 2017.

Table II.2 Malaise trap collection periods. For each sampling season the time of the year, number of traps and time period of collection is depicted

Season	Time of the year	Number of Traps	Sampling Dates
Season 1	Summer	12	13.07.2016 – 27.07.2016
Season 2	Autumn	14	13.10.2016 – 27.10.2016
Season 3	Winter	14	11.01.2017 – 25.01.2017
Season 4	Spring	14	12.04.2017 – 26.04.2017

DNA Extraction

DNA extraction was performed after overnight incubation in lysis buffer using non-destructive methods. We followed a modified protocol of Aljanabi & Martinez (1997) adjusted for our purposes. Organisms were first sieved from the collecting ethanol. To ensure that no specimen was overseen or lost, we poured the content of each bottle through a mesh filter (MICROFIL®V

Filter White Gridded 0.45µm-diameter 47mm & 100ml Funnel Sterilized). To reduce the risk of accidentally loosing small specimens, we processed the filter with the specimens. The insects were dried for 10 minutes at room temperature. Depending on biomass between 15 and 25ml of extraction buffer (0.4M NaCl, 10mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0) and 2% Sodium dodecyl sulphate (SDS) was added to each bulk sample. Finally, 400 µg Proteinase K per ml was added to lysis buffer solution. The samples were left to digest at 52°C at 30rpm in the orbital shaker overnight. The next day, the lysate was poured out of the bottles using the MICROFIL®V Filter (White Gridded 0.45um-Dia 47mm & 100ml Funnel Sterilized – Q :24). The solution was equally split into three 15ml falcon tubes. A 6M NaCl solution was added to the falcon tubes to a concentration of 4mmol. After vortexing for 30 seconds, the tubes were centrifuged at 4700rpm for 30 seconds. The supernatant was transferred to a new falcon tube to which an equal amount of isopropanol was added. After carefully mixing by turning the tube a few times upside down, the tubes were left at -20°C for one hour and subsequently centrifuged at 4700 rpm for 60 minutes. The supernatant was removed from the tube before submerging the resulting pellet with 20ml of ice cold 70% ethanol. The ethanol containing tubes were centrifuged for 15 minutes at 4700 rpm. Afterwards, the supernatant was discarded and the pellet was left to dry at 20°C overnight. The next day, 1ml of sterile H_2O was added to the dry pellet. The resulting DNA solution was stored at -20°C until further processing.

Choice of Primers and Library Preparation

For amplicon library preparation of Malaise trap samples, three primer pairs targeting two different markers were used, using a two step PCR approach (figure II.4). In the first step, the fragment of interest was amplified using gene specific primers including an illumina adaptor overhang (referred to here as PCR 1), and in the second step (referred to here as PCR 2), Illumina

index adaptors were added (Bourlat et al. 2016; Fonseca and Lallias 2016).

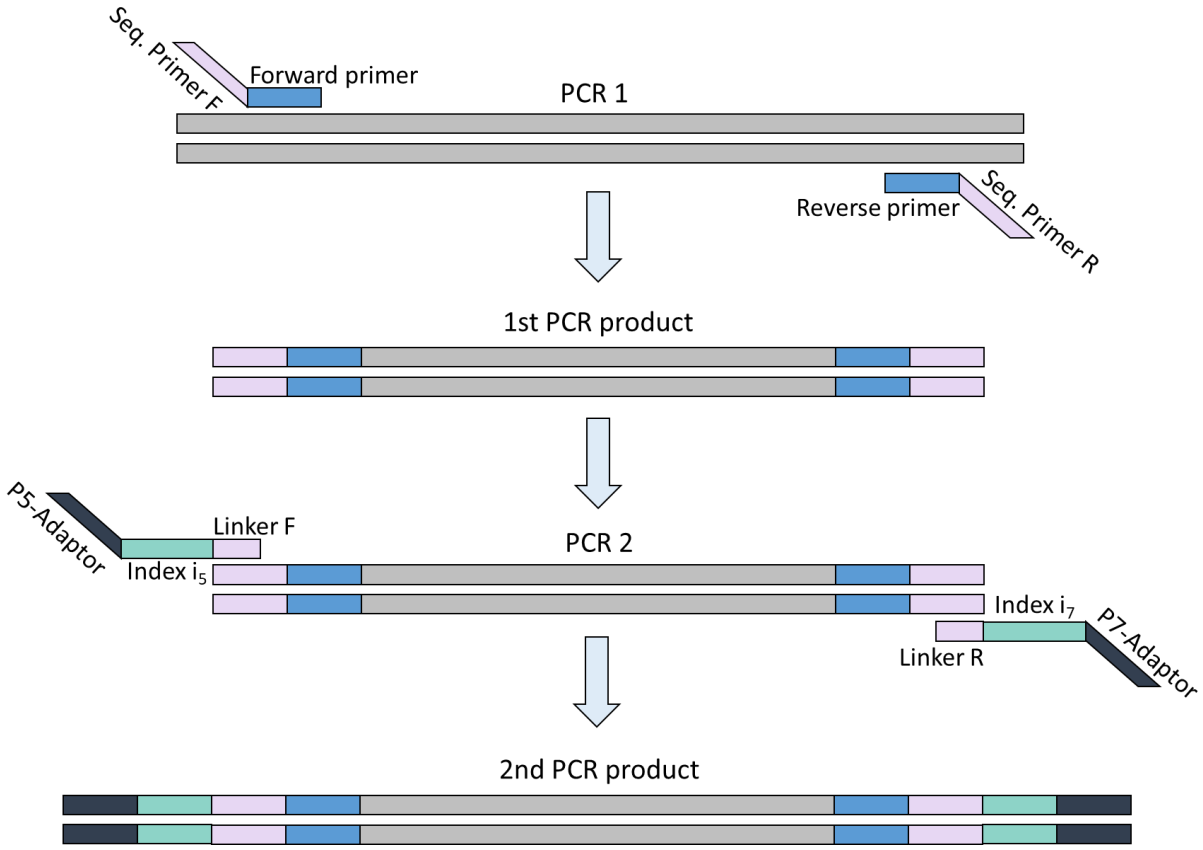


Figure II.5 Illustration of the 2-step PCR approach

For the amplicon PCR we decided to use the nuclear 18S rDNA (18S) marker which is more conserved but can detect a wide range of taxonomic groups, together with the more variable mitochondrial Cytochrome C Oxidase Subunit I (COI) marker, which enables a better taxonomic resolution within the Metazoa but at the cost that several other groups are only incompletely assessed e.g. Chordata and Echinoderms (Tang et al. 2012; Cowart et al. 2015).

For 18S, the primers TAReuk454FWD1 *(5'-CCAGCASCYGCGGTAATTCC-3')* and TAReukREV3r *(5'-*

ACTTTCGTTCTTGATYRA-3') were chosen, amplifying approximately 380bp of the V4 region of the nuclear 18S rDNA (Stoeck et al. 2010). Furthermore, two different primer pairs targeting the same 313bp of the 658bp long barcoding region of the mitochondrial Cytochrome C Oxidase subunit I gene (COI). Firstly, the forward primer mlCOIintF *(5'-GGWACWGGWTGAACWGTWTAYCCYCC-3')* (Leray et al. 2013) combined with the reverse primer Fol-degen-rev *(5'-TANACYTCNGGRTGNCCRAARAAYCA-3')* (Yu et al. 2012) and secondly mlCOIintF combined with the reverse primer dgHCO2198 *(5'-TAAACTTCAGGGTGACCAAAAAATCA-3')*, a less degenerate version of Fol-degen-rev (Leray et al. 2013).

mlCOIintF/Fol-degen-rev

10ng of template DNA was used for PCR1 using the forward primer mlCOIintF *(5'-*ACACTCTTTCCCTACACGACGCTCTTCCGATCT **GGWACWGGWTGAACWGTWTAYCCYCC***-3')* (Leray et al. 2013) and the reverse primer Fol-degen-rev *(5'-*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TANACYTCNGGRTGNCCRAARAAYCA***-3')* (Yu et al. 2012). The amplicon PCR (PCR 1) consisted for each sample of 7.5µ Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1 µl Sigma H2O, 0.5ul forward primer (10µM), 0.5ul reverse primer (10µM), 0.5ul Bovine Serum Albumin (thermoscientific) and 1ul template DNA making up a total of 15ul. The following PCR Program was applied: 1 cycle 98°C for 2 minutes, 25 cycles of 98°C for 1 minute, 50°C for 45 seconds and 72°C for 1 minute. Final elongation 72°C for 5 minutes.

After PCR1 the resulting product was checked for size by gel electrophoresis. Successfully amplified samples were quantified using the Quantus Fluorometer (Promega). Samples with a DNA concentration lower than 3ng/µl as well as samples that failed to amplify were repeated. After a cleanup step using Ampure XP magnetic beads (Beckman Coulter) at a ratio of 0.8:1 to remove primer dimers, PCR1 amplicon products were sent for further processing (indexing, pooling and sequencing) to MacroGen Inc. Seoul. Here PCR2 was conducted using the following

PCR program: 1 cycle 95°C for 3 minutes, 10 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds. Final elongation 72°C for 5 minutes.

The resulting purified amplicon pool was sequenced on an Illumina Miseq (2x 300bp) sequencing platform (MacroGen Inc. Seoul).

miCOIintF/dgHCO2198

10ng of template DNA was used for PCR1 using the primers mlCOIintF *(5'-*ACACTCTTTCCCTACACGACGCTCTTCCGATCT **GGWACWGGWTGAACWGTWTAYCCYCC** *-3')* (Leray et al. 2013) and dgHCO2198 *(5'-*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TAAACTTCAGGGTGACCAAARAAYCA***-3')* (Leray et al. 2013). The PCR1 mix for each sample consisted of 7.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1ul Sigma H2O, 0.5µl of forward Primer (10µM), 0.5µl of reverse primer (10µM), 0.5ul Bovine Serum Albumin (thermoscientific) and 1µl template DNA, making up a total of 15µl. The following PCR Program was applied: 1 cycle 98°C for 2 minutes, 20 cycles 98°C for 40 seconds followed by 45°C for 40 seconds and 72°C for 30 seconds with a final elongation at 72°C for 3 minutes. Following PCR1 the PCR products were purified by adding 4µl of HT ExoSAP-IT^{TM} (Applied Biosystems) to each sample. Samples were first heated at 37°C for 15 minutes removing excess primers and dNTPs, then at 80°C for 15 minutes to inactivate the enzyme and subsequently cooled at 4°C for 5 minutes.

For PCR2 (index PCR) the purified PCR products were split into two PCR tubes. While with PCR1 the gene region of interest was amplified, in PCR2 molecular identification (MID) tags in combination with NGS platform specific primers were incorporated. For each sample a unique combination of MID tags targeting both amplicon ends were chosen.

Each tube contained 12.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 3µl

Sigma H2O, 1.2µl of Index forward primer (10µM) (AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN ACACTCTTTCCCTACACGACGCTC), 1.2µl of index reverse primer (10µM) CAAGCAGAAGACGGCATACGAGAT NNNNNNNN GTGACTGGAGTTCAGACGTGTGCTC) and 8µl of purified PCR1 product. The PCR2 program was run as follows: 1 cycle at 98°C for 2 minutes, followed by 20 cycles of 40 seconds at 98°C, 30 seconds at 55°C and 30 seconds at 72°C. Final elongation at 72°C for 3 minutes. The PCR Products were visualised by gel electrophoresis and bands of the expected size were cut out. Cut out gel pieces of the same sample were merged before being purified using the QIAquick Gel Extraction Kit (Qiagen). DNA quantification was conducted with the Quantus Fluorometer (Promega). All purified and quantified tagged amplicons were pooled equimolarly. The resulting purified amplicon pool with a concentration of 3ng/µl was sequenced on a Illumina Miseq (2x 300bp) sequencing platform at Liverpool University's Centre for Genomic Research (Liverpool, UK).

TAReuk454FWD1/TAReukREV3

10ng of template DNA was used for PCR1 using the primers TAReuk454FWD1 *(5'-* ACACTCTTTCCCTACACGACGCTCTTCCGATCT **CCAGCASCYGCGGTAATTCC**-*3')* and TAReukREV3r *(5'-* GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **ACTTTCGTTCTTGATYRA**-*3'*). The PCR1 (amplicon PCR) mix for each sample consisted of 7.5ul Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1ul Sigma H2O, 0.5ul of forward primer (10µM), 0.5ul of reverse primer (10µM), 0.5µl Bovine Serum Albumin (thermoscientific) and 1µl template DNA making up a total of 15µl. The following PCR Program was applied: 1 cycle at 98°C for 2 minutes, 20 cycles at 98°C for 40 seconds, 55°C for 40 seconds, 72°C for 30 seconds. final elongation at 72°C for 3 minutes. Following PCR1 the PCR products were purified with HT ExoSAP-IT^{TM} (appliedbiosystems) by adding 4µl of HT ExoSAP-IT^{TM} to each sample. Samples were first heated up to 37°C for 15minutes then to 80°C for 15 minutes and subsequently cooled down at 4°C for 5 minutes.

For PCR2 (index PCR) the purified PCR products were split into two PCR tubes. Each tube contained 12.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 3µl Sigma H2O, 1.2µl of index forward primer (10µM) (AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN ACACTCTTTCCCTACACGACGCTC), 1.2µl of index reverse primer (10µM) CAAGCAGAAGACGGCATACGAGAT NNNNNNNN GTGACTGGAGTTCAGACGTGTGCTC) and 8µl purified PCR 1 product. The PCR2 program started with 1 cycle at 98°C for 2 minutes, followed by 20 cycles at 98°C for 40 seconds, 30 seconds at 55°C and 30 seconds at 72°C. Final elongation at 72°C for 3 minutes. The PCR Products were visualised by gel electrophoresis and bands of the expected size were cut out. Cut out gel pieces of the same sample were merged before being purified using the QIAquick Gel Extraction Kit (Qiagen). DNA quantification was conducted with the NanoDrop (XXX). All purified and quantified tagged amplicons were pooled equimolarly. The resulting purified amplicon pool was sequenced on a Illumina Miseq (2x 300bp) sequencing platform at Liverpool University's Centre for Genomic Research (Liverpool, UK).

Bioinformatics and Data Analysis

Data sequenced at the Centre of Genomic Research (Liverpool, UK) had already undergone first quality check: The raw fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1. Furthermore, sequences were trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 20bp after trimming were removed. Sequences were received from MacroGen Inc. Seoul in Casava 1.8 paired-end demultiplexed fastq format.

The fastq sequences were checked for the presence of the COI and 18S primers with Cutadapt version 1.18 (Martin 2011) using the following settings: maximum error rate (-e): 0.1, minimum

Overlap (-O): 20, minimum sequence length (-m): 50). Sequences lacking either forward or reverse primer were remove. Detected primers were trimmed from the remaining sequences. Paired-end reads were merged with vsearch version 2.7.0 (Rognes et al. 2016). Merged sequences with a length of 360-400bp for the 18S and 293-333bp for the COI dataset respectively were retained for further analysis and filtered with a maxEE threshold of 1.0 using vsearch (version 2.7.0) (Rognes et al. 2016). Subsequently fastq-sequences were demultiplexed using the script split_libraries_fastq.py implemented in QIIME1 (Caporaso et al. 2010). A phred quality threshold of 19 was chosen. Dereplicating, size sorting, denovo chimera detection as well as OTU clustering with a 97% cutoff was conducted with vsearch 2.7.0 (Rognes et al. 2016). Finally, an OTU table was build by using the --usearch_global function in vsearch 2.7.0 (Rognes et al. 2016) (Rognes et al. 2016) followed by the python script “uc2otutab.py” written by Robert Edgar (https://drive5.com/python/uc2otutab_py.html). For taxonomy assignment sequences were blasted against the GBOL database using blastn 2.9.0+ (Altschul et al. 1990).

The resulting OTU table was loaded into Excel were data got cleaned up and formatted for upload into R (R CoreTeam 2013). Statistical analysis was conducted with R studio running R version 3.5 using several R packages. Visualizations comparing number of OTUs/Species depending on choice of marker, primer and database were prepared using the R package ggplot2 (Wickham 2016). Data used for the preparation of Venn diagrams visualizing the number and percentage of shared and unique insect species with a BlastID of at least 99% between primer-pairs depending on organisms group (figure 14) were processed with the R package dplyr (version 0.8.3) (Wickham et al. 2015) before final diagrams were prepared with the R package Venn.Diagram (version: 1.6.20) (Chen and Boutros 2011). Bar plots showing number of recovered OTUs/Specimens depending on method, marker and primer were prepared with the R package ggplot2 (Wickham 2016). PCoA Plot, showing differences in recovered insect (BlastID ≥ 90%) communities depending on combination primer, marker and database choice were prepared using the R package betapart (version 1.5.1) (Baselga and Orme

2012) and vegan (version 2.5-6) (Dixon 2003). Calculated heatmaps visualizing calculated jaccard similarity indices for insect communities were prepared using the R package ggplot2 (Wickham 2016). Underlying calculation were done in Microsoft Excel based on the formula: $J(X,Y) = |X \cap Y| / |X \cup Y|$. With the R package ggplot2 (Wickham 2016) stacked bar plots were created showing number of OTUs assigned to species level within the suborder Nematocera, depending on blastID. Calculated Marioko plots visualizing number of species depending on sampling season and forest type were prepared using the R package ggplot2 (Wickham 2016). The same R package was used for preparing stacked barplot showing number of species per insect order depending on forest type and season. PCOA plots showing differences in species communities depending on forest type and season were prepared using the R package betapart (version 1.5.1) (Baselga and Orme 2012) and vegan (version 2.5-6). The same two packages were used for the visualization of the distance to centroid indicating homogeneity between samples within forest types depending on season (figure 20B). Homogeneity between samples and forest types was further visualized via Venn Diagrams. Venn Diagrams were prepared on basis of the R package Venn.Diagram (version: 1.6.20) (Chen and Boutros 2011) and modified using Microsoft PowerPoint. Stacked barplots showing number of permanent resident and visitor species were prepared using the R package ggplot2 (Wickham 2016). Finally, UpsetR-Plot showing detailed number of unique and shared permanent resident species (BlastID ≥ 99%) between forest types depending on season was prepared using the R package UpSetR (version 1.4.0) (Conway et al. 2017).

Results & Discussion

Part 1 – Evaluation of Methods

Choice of Marker

Many studies have already shown that metabarcoding is a powerful tool, enabling the assessment of biodiversity in a reliable, cost efficient and timely manner (Gibson et al. 2014; Braukmann et al. 2018; Elbrecht et al. 2019). Over the last years, several metabarcoding studies were conducted in order to assess diversity of various organism groups. Depending on target group different genetic markers were chosen to draw a picture as complete as possible of existing diversity (Yu et al. 2012; Elbrecht et al. 2016, 2019; Marquina et al. 2019a). For the assessment of arthropod diversity the COI marker have increasingly been used (Elbrecht et al. 2019; Thomsen and Sigsgaard 2019). However, due to its protein coding character several authors recommend the use of more conserved markers e.g. 18S (Deagle et al. 2014; Horton et al. 2017). Unquestionable the choice of marker has a significant influence on study outcome. This is also mirrored in the here presented study. The two markers tested here, resulted in the detection of significant different numbers of OTUs (figure II.6). The two COI datasets contained, depending on primer, between 3077 (Fol_degen_rev) and 2565 OTUs (dgHCO2198), while the 18S dataset

contained the lower slightly lower number of 1558 OTUs (figure II.6). The fact that with the COI marker almost up to twice as many OTUs were found, is likely contributed to the different degrees of variability of the two genes of interest. In direct comparison is the 18S marker much stronger conserved. This is a curse and savior at the same time. On the one hand the primer binding sites are less variable, which is supporting binding of primers whereas a broader taxonomic coverage across the eukaryotic domain of life is achieved (Clarke et al. 2017). On the other hand this goes for the costs of a lower taxonomic resolution. Due to its lower degree of variability within the discriminative region of the gene, reads originating from distinct species are more likely to be clustered together into the same OTU. As a result, especially on lower taxonomic level existing biodiversity is often strongly underestimated (Wangensteen et al. 2018). As a results the gap between number of assigned OTUs widens between the two markers with decreasing taxonomic level.

Out of the 1558 OTUs of the 18S dataset, 44.5% (694 OTUs) were assigned to Arthropoda, while depending on choice of primer and reference database, between 43% and 62% of the OTUs of the COI dataset were assigned to this phylum (GBOL: dgHCO2198: 1578 OTUs, Fol_degen_rev: 1599 OTUs; BOLD database: dgHCO2198: 1605 OTUs, Fol_degen_rev: 1323 OTUs) (figure II.6a). On average, the COI marker detected slightly more than twice as many Arthropod OTUs. On class-level this gap widened. Approximately more than four times as many OTUs were assigned to Insecta using the COI marker. Between 91% and 94% of the recovered COI OTUs, with an assignment to Arthropoda were assigned to Insecta (1206 OTUs -1505 OTUs). Within the 18S dataset only 50% of the 649 recovered Arthropod OTUs (361 OTUs) accounted for this class (figure II.6b).

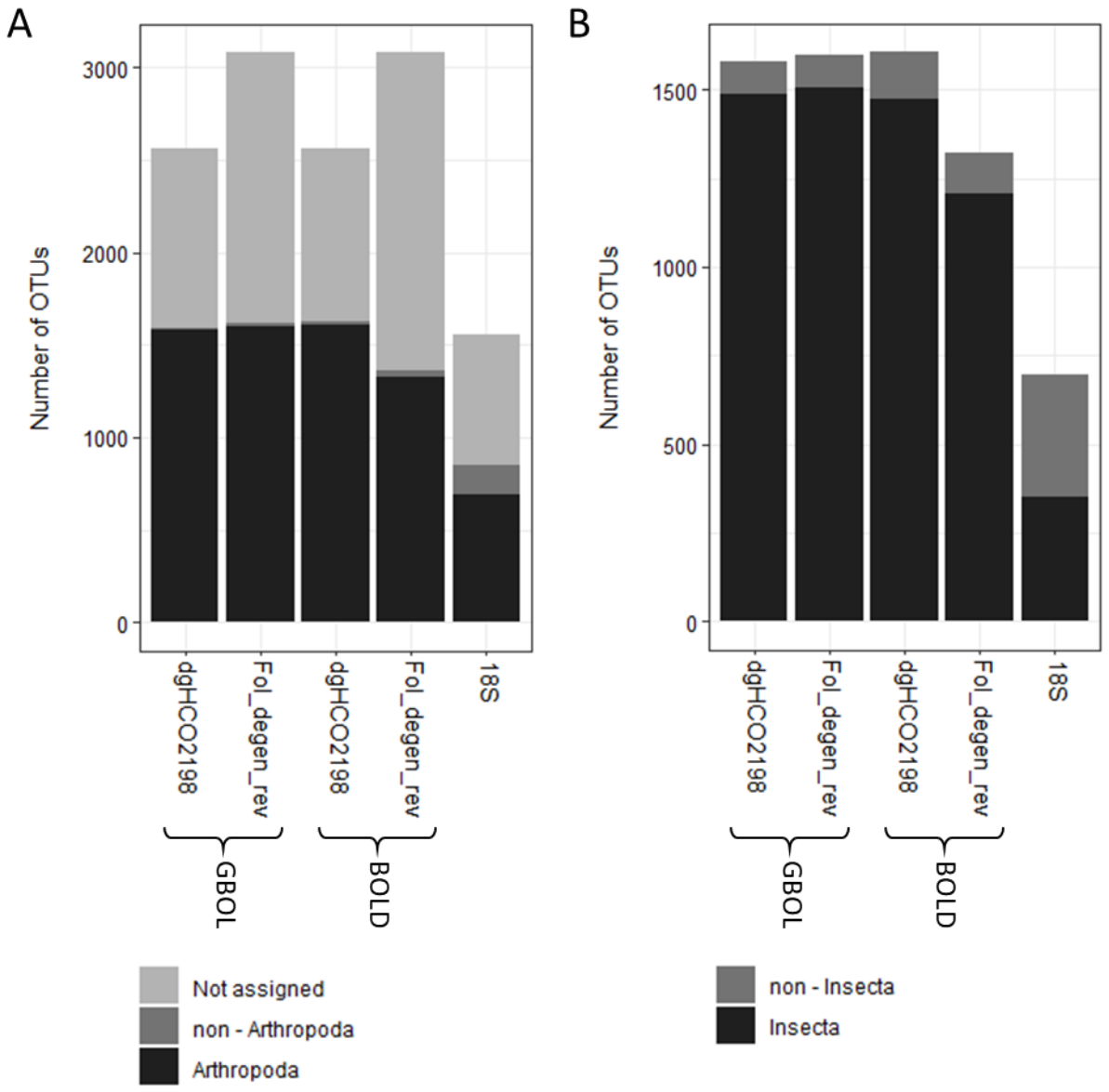


Figure II.6 Number of assigned OTUs depending on choice of marker, primer and used reference database. A) Total number of detected OTUs, including number of 'Not assigned' OTUs and highlight the number of detected OTUs assigned to Arthropoda. B) number of OTUs assigned to Insecta out of total number of OTUs assigned to Arthropoda

As a result of differences in taxonomic resolution, the degree of insect diversity assessed by the two markers showed major variations. While the COI marker identified depending on primer pair between 17 and 18 insect orders, only 12 orders were recovered with the 18S marker (figure II.7). Out of them, five orders encompass regardless of primer and marker choice at least 86% of all recovered Insect OTUs. The five orders were Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera (figure II.7A), from here on referred to as highly abundant insect orders. Out of them, the order Diptera accounted for the highest share of OTUs (GBOL Database: Fol-degen-rev: 52.01%, dgHCO2198: 68.40%; BOLD Database: Fol_degen_rev: 48.1%, dgHCO2198: 69.10%; 18S: 56.5%). The remaining four highly abundant insect orders Coleoptera (GBOL: Fol-degen-rev: 6.93%, dgHCO2198: 5.47%; BOLD: Fol-degen-rev: 9.22%, dgHCO2198: 5.20%; 18S: 3.99%), Hymenoptera (GBOL: Fol-degen-rev: 20.04%, dgHCO2198: 7.40%; BOLD: Fol-degen-rev: 20.53%, dgHCO2198: 6.96%; 18S: 8.75%), Hemiptera (GBOL: Fol-degen-rev: 9.21%, dgHCO2198: 6.44%; BOLD: Fol-degen-rev: 8.65%, dgHCO2198: 6.25%; 18S: 5.04%) and Lepidoptera (GBOL: Fol-degen-rev: 4.15%, dgHCO2198: 3.67%; BOLD: Fol-degen-rev: 5.5.8%, dgHCO2198: 3.77%; 18S: 12.2%) were represented by significantly less OTUs. Next to the five highly diverse orders, several less diverse orders were identified. Its number was strongly depending on choice of marker (figure II.7B). The COI marker resulted in the detection of at least 11 less diverse orders while with the 18S marker only seven were identified. Furthermore, numbers of OTUs assigned to less diverse orders were strongly differing between marker datasets. While with the COI marker, depending on primer between 49 and 67 OTUs were assigned to less diverse orders, only 26 OTUs of the 18S dataset had an assignment to one of these taxa (figure II.7B).

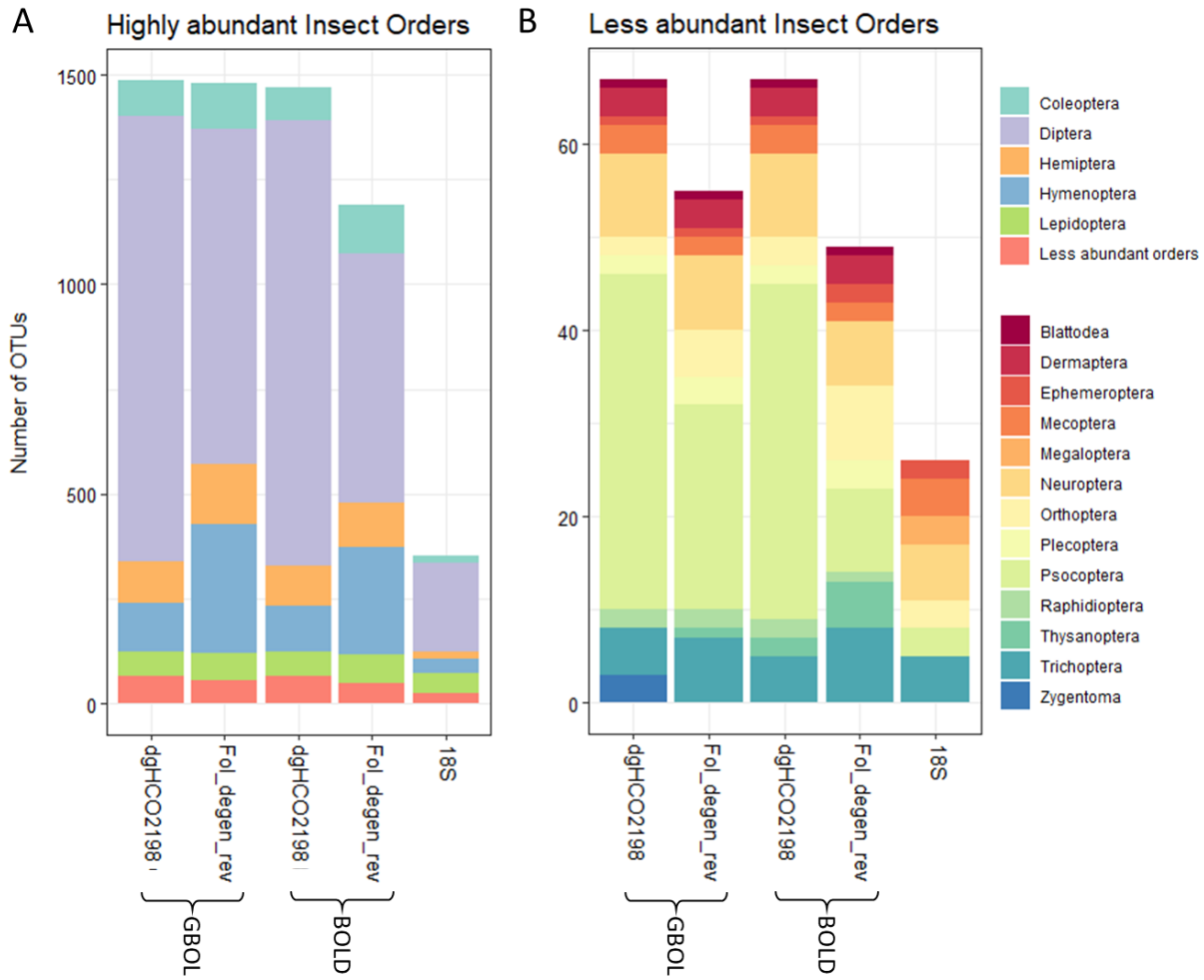


Figure II.7 Number of OTUs assigned to orders of the class Insecta with a BlastID of at least 90%, depending on choice of primer, marker and reference database. Figure II.6A shows the highly diverse groups, while Figure II.6B includes all orders represented by less than 20 OTUs.

We conclude that the lower the taxonomic level, the more dramatic the influence of marker choice. Therefore and as expected led the lower taxonomic resolution of the 18S marker to a significantly lower number of identified species. To reliable assign OTUs on species level a blast hit of at least 99% was required. This applied to OTUs assigned to Insecta of the 18S dataset. In contrast to that, depending on primer choice between 478 and 810 Insect OTUs of the two COI

datasets had an high-quality assignment on species level to Insecta. However, the total number of identified species was lower as several species had double assignments, what means that they were represented by more than one OTU. In terms of total number of identified insect species, the COI marker was clearly superior over the 18S maker. While with the 18S marker 30 insect species were reliable identified the COI marker detected up to 671 species (figure II.8).

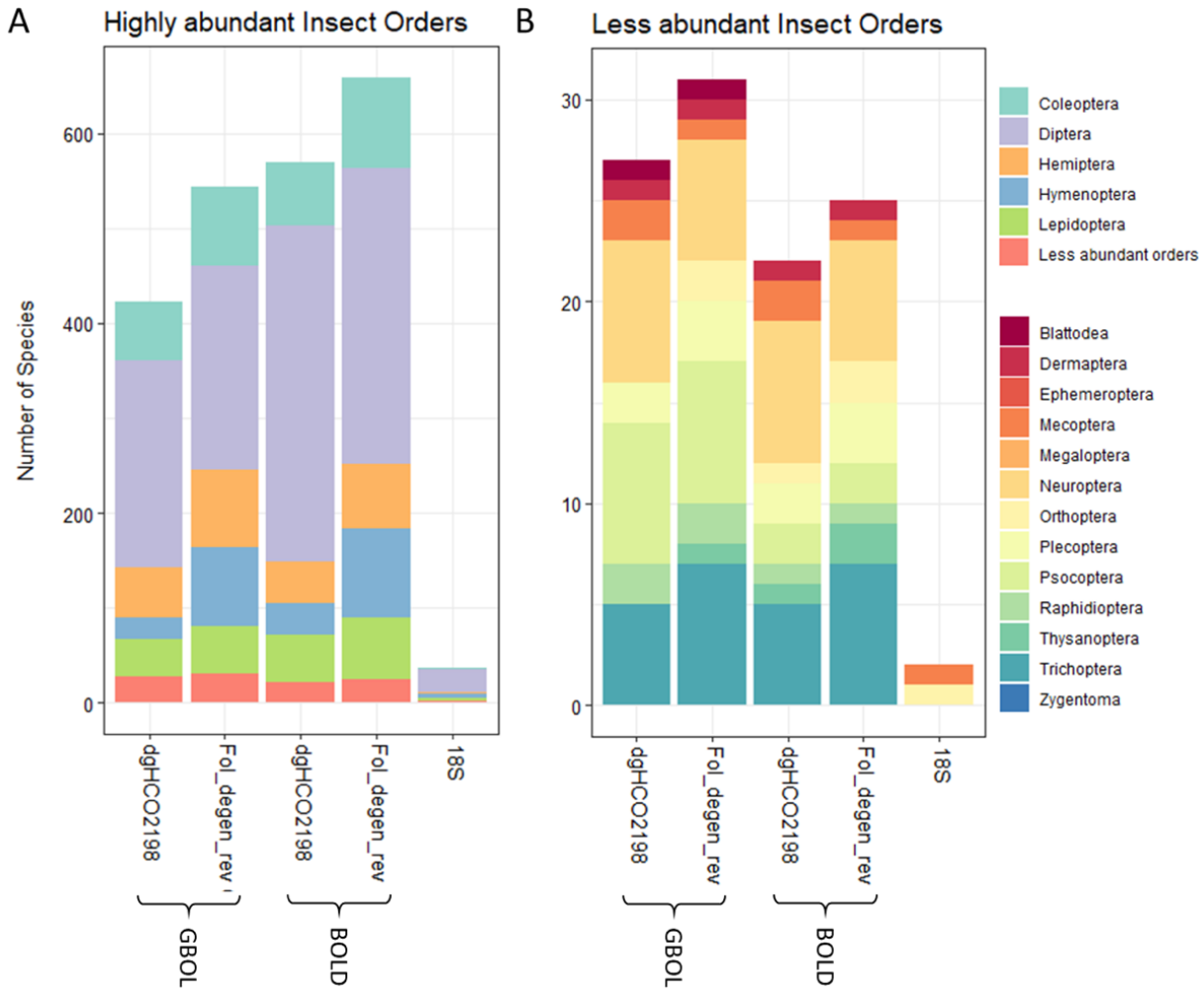


Figure II.8 Number of recovered insect species (blastID ≥ 99%) per order, depending on choice of marker, primer and reference database. Figure II.7A includes all species of the highly

Figure II.8 (Continued.) diverse orders Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera, while Figure II.7B shows number of detected species of remaining, usually less diverse insect orders

Choice of Reference Database

The higher number of species recovered with the COI marker is not least attributable to the extremely large publicly available reference database for this marker. It is known that study outcome and completeness of resulting species list is closely linked to complexity and completeness of available reference databases (Gibson et al. 2014; Corse et al. 2019). A study investigating fecal samples from bats for their preferred diet, has shown that samples collected in the Paleartic displayed a much higher taxonomic resolution compared to samples taken in Africa (Corse et al. 2019). The authors of the study hypothesized that the lack of precise species discrimination is contributed to the fact that megadiverse environments like equatorial areas are suffering from relatively incomplete public databases. A fact that has also been mentioned by further scientists (Cowart et al. 2015; Beng et al. 2016; Lopes et al. 2017). Because of its high taxonomic resolution, the Consortium for the Barcode of Life (CBOL) has chosen the mitochondrial COI gene for standard DNA barcoding of single animal specimens (Hebert et al. 2003; Ratnasingham and Hebert 2007). This decision contributed significantly to the establishment of the now unparalleled large publicly available reference databases. In August 2015 the international barcode of life initiative completed their first major program BARCODE 500K, which had the overreaching goal of barcoding 0.5 million species. Today, more than 502.700 public barcodes clusters are stored in the BOLD database. The barcode of life initiative encompasses now research organizations from 25 nations. All barcodes retrieved in the scope of the barcode of life initiative are supposed to be stored in the international BOLD Database in order to make them publicity available. Today, it can be said that no database for any other

genetic region covers as many taxa as the Barcode of Life Database (BOLD) does for the COI (Deagle et al. 2014). However, many of the national initiative which were conducted under the auspices of the CBOL have established own, geographically limited barcode reference databases. The German Barcode of Life project (GBOL) was launched in 2012 with the overreaching goal to barcode the German fauna and flora. In the scope of the GBOL project the GBOL database was established. The GBOL database now comprises the barcodes of 16906 arthropod species, what accounts for approximately 45% of the for Germany described arthropod diversity (German Barcode of Life Consortium et al. 2011). Although it is the major aim of the Barcode of Life initiative to store all produced barcode retrieved under the auspices of the CBOL in a single reference database – the BOLD database this goal has yet not been met. Out of the 16906 barcodes contained in the GBOL database, many are still waiting to be uploaded to the BOLD database.

Here it was found that the usage of the BOLD database as reference database resulted in a higher number of detected species of the four highly diverse orders Coleoptera, Diptera, Hymenoptera and Lepidoptera. When the GBOL database was used as reference database a higher number of species of less abundant insect orders (figure II.8B) were identified. Furthermore, number of detected species of the highly diverse order Hemiptera was significantly higher. While incomplete databases are usually leading to wrong negative results, very complex databases can increase the risk of false positive assignments as many taxonomy assignment algorithms are only choosing the best hit, without taking the probability of species existence within the habitat into account. When blasting our dataset against the BOLD database the dipteran species *Atypophthalmus inustus* was found detected in a single bulk sample with a blastID of 99.02%. A closer look in the reference database showed that the species is represented by the small number of three sequences retrieved from specimens found in Finland. Although, the species has been described for Germany there are to our knowledge no recent records. Sequences originating from undescribed

species are more likely to be assigned to a closely related species the more complex the database becomes.

However, the comparatively higher number of barcodes contained in the BOLD database is mirrored in number of OTUs being assigned on species level (figure II.8). Especially within the dipterans, a better species coverage was achieved with the BOLD database as reference. In detail, up to 145 species more were identified compared to the GBOL database. However, many species of the order Diptera are highly active and are capable to travel long distances in a rather short time. This is making the order Diptera only conditionally suitable for assessment of the current status of a habitat. Furthermore, for many dipteran species little is known about their biology and their associated habitat. Often the less diverse groups e.g. Raphidioptera and Thysanoptera and much better studied, making them to more reliable indicator species for the current status of the habitat. For these less diverse groups a better taxonomic resolution was achieved with the GBOL database (figure II.8B). The GBOL database is well curated and there were no entries on species level, without a definite species name. This was not the case for the BOLD database in which several entries were found with an unclear entry on species level, either only containing an accession number or the hint, that the species had not yet been described. This made the analysis to a time consuming task, as every entry had to be checked twice. Another advantage of the GBOL database is the fact that it only contains species for which a proof of existence for the geographical area Germany exists. Although this may prohibits for the detection of possible invasive species, this is a great advantage as it limits the risk of encompassing false positive assignments. Therefore, the use of a less complex but for the area more accurate databases is recommended. Additionally, it should be noted that all barcodes obtained with the GBOL project are supposed to be placed into the BOLD database. Unfortunately, this goal has not yet been met whereas it is likely that several species which main distribution area is located within Germany are not yet present in the BOLD database, whereas the species would likely be missed when using the BOLD database as reference.

Choice of COI Primer Pair

Because of the above discussed differences in performence of databases, the following analysis is based on the GBOL database. In contrast to the choice of marker had the choice of primer less influence on number of recovered taxa on higher taxonomic level. On class level, the use of the dgHCO2198 primer pair resulted in the detection of one insect order less in comparison to the Fol_degen_rev primer pair (figure II.7). Nevertheless, on lower taxonomic level significant differences in number of recovered taxa were observed (figure II.9). Depending on insect order the Fol_degen_rev and HCO2198 primer pair resulted in the detection of two distinct insect species communities (figure II.14). Only 53% of all recovered species were detected with both used primer pair (figure II.9). With 120 species more, a significantly higher number of species was detected with the Fol_degen_rev primer pair. As both primers are targeting for the same region of the mitochondrial COI region, the large discrepancy in number of detected species is not attributable to incomplete reference databases. Moreover, variations in primer affinities are likely to be the main reason for these findings.

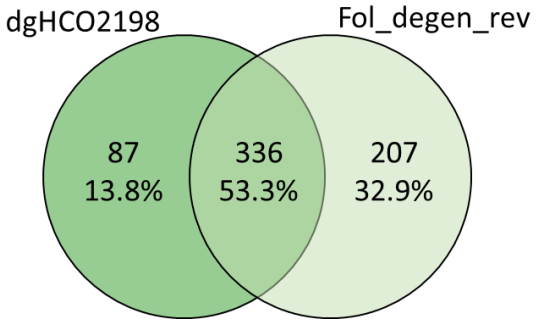


Figure II.9 Venn diagram showing number of unique and shared detected insect species (BlastID ≥ 99%) between COI primer pairs.

Both primer pairs shared the forward primer mlCOIintF. The mlCOIintF primer is only partly degenerated. It contains seven wobbles bases, each of them allowing two different bases to match (figure II.10). The calculated degeneracy value for mlCOIintF is 128, resulting in 128 different primer combinations.

mlCOIintF 5'-GGWACWGGWTGAACWGTWTAYCCYCC-3'

Figure II.10 Forward primer mICOIintF. Wobble bases are highlighted in red

While both primer pairs consisted of the same forward primer, the chosen reverse primers were strongly differing in their degree of degeneracy (figure II.11). While the Fol-degen-rev is highly degenerated, the dgHCO2198 does only allow for two base wobbles. The first wobble position allows the binding of Adenine and Guanine, while the second allows for Thymine and Cytosine to bind. The Fol-degen-rev primer has three position at which Adenine and Guanine are allowed to bind and two positions at which Thymine can be replaced by Cytosine and vice versa. Furthermore, three base positions allow any nucleotide to bind. The calculated amount of degeneracy for the Fol-degen-rev is 2048, what dramatically increases the amount of number of sequences which can be amplified.

dgHCO2198 5'-TAAACTTCAGGGTGACCAAARAAYCA-3'
Fol-degen-rev 5'-TANACYTCNGGRTGNCCRAARAAYCA-3'

Figure II.11 Reverse primers dgHCO2198 and Fol-degen-rev. Wobble bases are highlighted in blue and red respectively

A high degree of degeneracy enables the amplification of a broader taxonomic range as PCR biases are mitigated (Yu et al. 2012; Krehenwinkel et al. 2017). The number of mismatches between the primers and the template DNA defines the affinity of the primer for the different templates (Marquina et al. 2019a). The higher the affinity the more easily the template DNA will be amplified. As a result several organism groups are subsequently overrepresented (Clarke et

al. 2014), while other groups will largely be missed (Yu et al. 2012). Some studies have already shown, that the amplification success for different organism groups therefore even depends on the degree of degeneracy by describing a positive effect of primer degeneracy degree and amplification success (Morinière et al. 2016). In general, species with higher affinities will likely capture more primer molecules during PCR while species with lower affinities will yield lower level amplicons and fewer reads (Hajibabaei et al. 2011; Brandon-Mong et al. 2015). When degenerated primers are used this bias can to some extend be reduced (Clarke et al. 2014; Elbrecht and Leese 2017; Elbrecht et al. 2019).

The study presented here is partly supporting these findings. Based on the direct comparison of retrieved insect communities from six Malaise trap bulk samples, collected at the pure beech sites in summer and autumn season the performance of the two COI primer pairs, differing in its degree of degeneracy were compared. It was found that depending on target insect order the two primer pairs recovered different insect communities in terms of composition and diversity.

The number of identified hymeopteran families varied widely between chosen primer pairs. On basis of the six malaise traps samples taken at the pure beech sites in summer and autumn season we found that a higher number of insect families were retrieved with the Fol_degen_rev primer pair. While the less degenerated dgHCO2198 primer detected 12 hymenopteran families, the Fol_degen_rev primer retrieved the same 12 hymenopteran families but also detected six additional families (figure II.12), which were not found by the less degenerated dgHCO2198 primer pair.

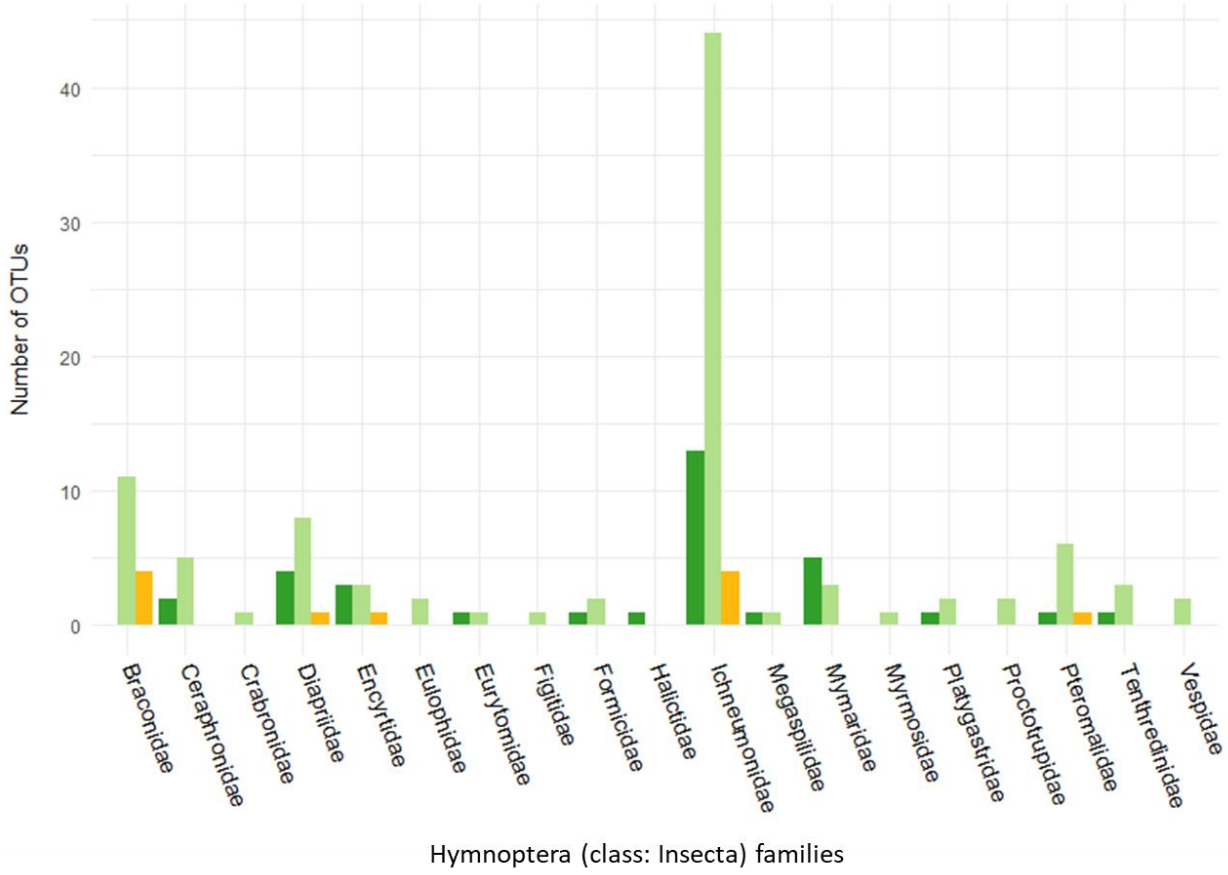


Figure II.12 Number of OTUs assigned to families of the insect order Hymenoptera. Coloration of bars are indicating choice of marker, primer and reference database.

Similar results were obtained for the dipteran suborder Nematocera. Amplification with the Fol_degen_rev primer resulted in the detection of 16 nematoceran families out of which only 15 were also recovered with the dgHCO2198 primer pair (figure II.13). These results indicate that the amplification success of the order Hymenopteran and the suborder Nematocera is positively correlated with the degree of degeneracy of chosen primer pair. Nevertheless, using primers with a high degree of degeneracy has also its drawbacks. A recent study found that highly degenerated primers performed poorly on a malaise trap sample (Elbrecht et al. 2019). Often, a high degree of degeneracy goes along with a decrease in specifity, which may result in non-specific

amplification, dimerization and primer slippages (Elbrecht et al. 2018, 2019). As a result, a higher degree of degeneracy may promote the detection of certain organism groups, while hampering the identification of other.

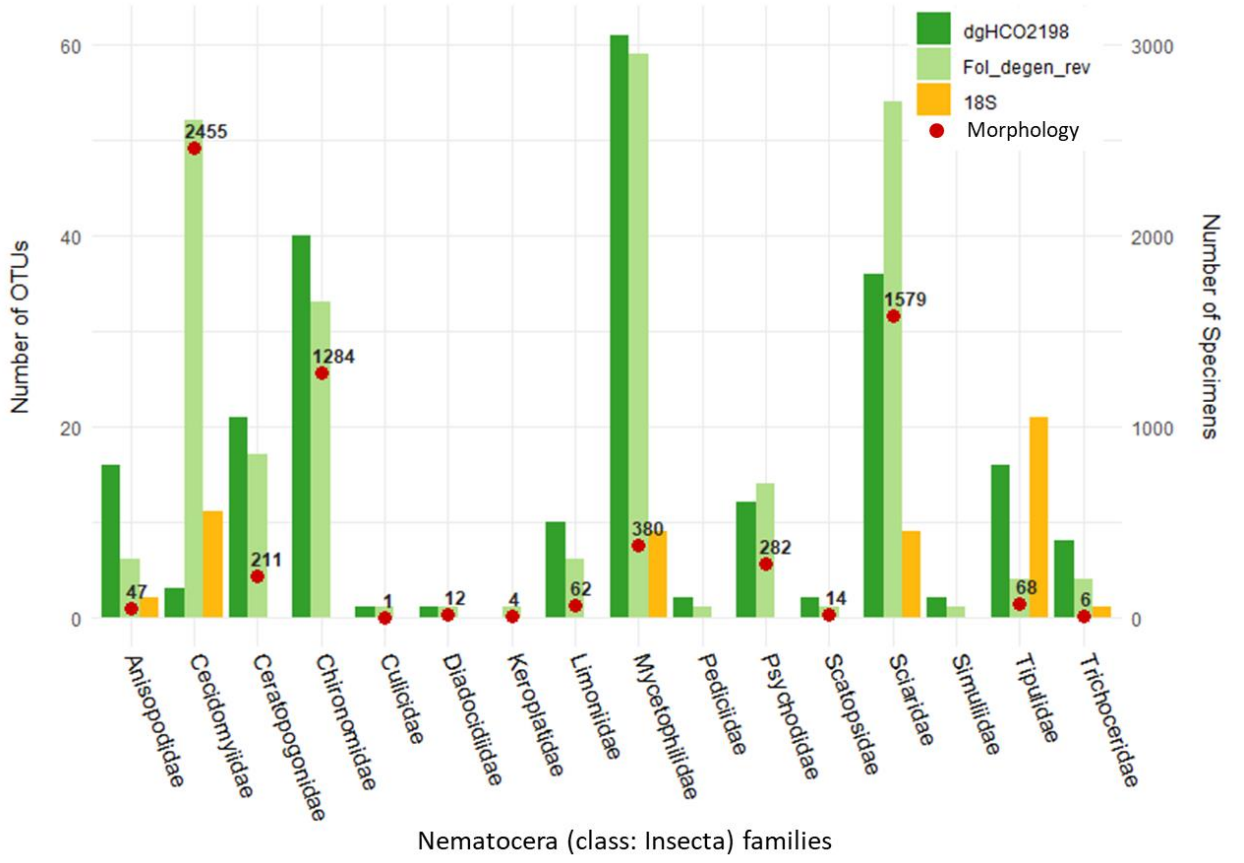


Figure II.13 Number of Individuals and OTUs assigned to families of the Insect suborder Nematocera. Bars are representing number of detected OTUs with molecular methods differing in choice of primer, marker and reference database, while morphological identified number of specimens are shown as red points.

Assessed Coleoptera community composition revealed even more significantly differences in primer performance depending on degree of primer degeneracy. While the Fol_degen_rev

primer pair detected 21 families, the dgHCO2198 primer pair identified only 18 (figure II.14). Although the number of detected species retrieved with the Fol_degen_rev primer pair surpassed the number found with the dgHCO2198, it cannot be stated that the highly degenerated primer pair outperformed the latter. This is because the two primer pairs did not result in the detection of the same 18 families. Moreover, the five families Coccinellidae, Eucnemidae, Monotomidae, Oedemeridae and Ptinidae were exclusively found with the Fol_degen_rev primer pair, while the two families Leiodidae and Sphindidae were exclusively recovered with the less degenerated dgHCO2198 primer pair.

However, when taking the full dataset into account and not only the here investigated six bulk samples, the Fol_degen_rev primer pair was also capable to detect the two families Leiodidae and Sphindidae. Thus, the highly degenerated primer pair is generally working for these families, which is indicating that bulk sample composition might influence primer performance. This is in alignment with previous studies (Brandon-Mong et al. 2015). It has been shown, that the COI region of several hymenopteran specimens can be successfully amplified using the Leray primer pair mICOIintF/HCO2198 for single-specimen PCR, but the same species will be missed when being part of a bulk sample (Brandon-Mong et al. 2015). A possible explanation for these findings is that the DNA mixture extracted from the unsorted Malaise traps is highly diverse. DNA with a lower affinity to the primer will more likely be outcompeted by DNA with a higher affinity, leading to false negative results (Morinière et al. 2016). Morinière et al. (2016) showed that metabarcoding of presorted Malaise traps on order level recovered a higher number of high score BINs compared to metabarcoding of unsorted traps. It can be expected that taxonomic coverage of a primer pair is directly influenced by number and taxonomic classification of the organisms contained in the sample as it directly influences competition for primer molecules resulting from different affinities of species in their primer binding sites (Hajibabaei et al. 2011; Brandon-Mong et al. 2015).

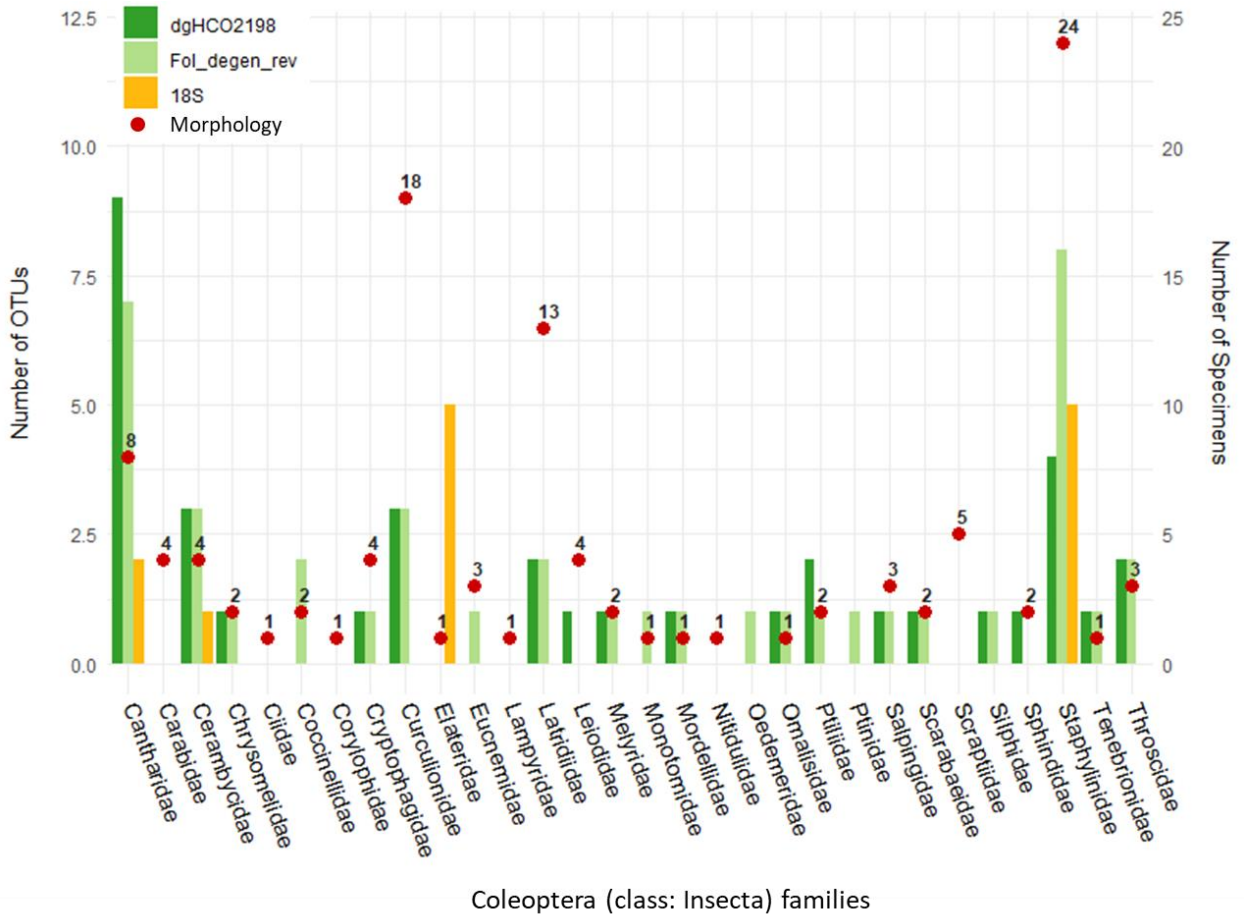


Figure II.14 Number of OTUs (90% BlastID) and Individuals per detected Coleoptera family depending on choice of primer, marker, reference database and identification method. Number of recovered OTUs are represented by bars, while number of morphological identified specimens are represented by red points.

However, next to primer affinities also the amount of DNA with which a species contribute to extracted DNA mixture is influencing species detection rate. For the insect order Coleoptera and the dipteran suborder Nematocera all families being present in the six bulk samples were morphologically identified. When the morphologically obtained data were compared to list of detected families with metabarcoding we found that all formerly morphological identified nematoceran families were detected using the highly degenerated Fol_degen_rev primer pair

(100% detection rate). The dgHCO2198 primer pair found one family less (93.75% detection rate)(figure II.11). Out of the 27 morphological identified coleopteran families, 18 were identified with the Fol_degen_rev primer pair (66.67% detection rate), while the dgHCO2198 primer pair recovered 17 (62.96% detection rate). The six families Carabidae, Ciidae, Corylophidae, Lampyridae, Nitidulidae and Scraptiidae were not identified with metabarcoding targeting the COI marker. The reasons for the incomplete assessment are manifold.

The family Scraptiidae had no record in the GBOL database, whereas it was not possible to retrieve this family with metabarcoding using this database for taxonomy assignment. The remaining five families were represented in the database by several species and barcodes. Nevertheless, a closer look on number of recovered individuals per family showed, that for four out of the five families left undetected only a single individual was found in the six bulk samples (figure II.14). Former studies have already discussed the effect of unbalanced biomass of species detection rate. It has been shown, that taxa are more likely to be detected when they are part of a bulk sample which only contains specimen of similar sizes (Elbrecht et al. 2017). The larger a specimen, with the more DNA it contributes to DNA mixture used for library preparation. Reads with higher abundances will most likely capture more primer molecules leading to a high amplicon yield, while lower abundance reads may not yield any amplicons (Hajibabaei et al. 2011). Thus, highly abundant species or species of larger body size are more likely to be captured with metabarcoding, while smaller or less abundant species are more likely to be missed. Therefore it is increasingly recommended to presort samples prior to DNA extraction (Elbrecht et al. 2017). However, here a lysis buffer based DNA extraction method was followed, for which the influence of differences specimen size and abundance in combination with complexity of the bulk sample has so far not been tested. However, here it was observed, that that an increase in degree of degeneracy does not guarantee for a more precise biodiversity assessment which allows for the assumption that the above mentioned parameter also plays an important role for study outcome when using non-destructive extraction methods.

Choice of Extraction Method

Former studies have already successfully applied a modified protocol of the non-destructive DNA-extraction method of Aljanabi and Martinez for the extraction of DNA from feces (Vesterinen et al. 2016; Kaunisto et al. 2017). Additionally, a second study showed the suitability of the method for the extraction of Arthropod DNA from bulk samples (Ritter et al. 2019). Ritter et al. (2019) tested the efficiency of five DNA extraction protocols on five insect bulk samples. As they did not find any significant differences in number of captured sequences between the protocols, they concluded that the success rates of the protocols are comparable. Kaunisto et al. (2017) compared the salt extraction method modified from Aljanabi and Martinez with two commercial kits. The salt extraction method produced the highest number of reads, but the number of detected taxa was only marginal higher than the one found with the commercial Macherey Nagel Nucleospin Kit. To our knowledge, the salt-precipitation protocol has not yet been tested with a mock community, whereas the suitability of this method might vary between different groups of organisms. Specimens morphology could potentially influence DNA extraction success rates. Especially within the coleopterans a high number of families were not recovered (figure II.14). Next to the above discussed primer bias, the extraction methods can likely strongly contributed to the here observed poor detection rate. Beetles are often characterized by a thick and hard cuticle, which could possibly hamper the release of DNA into the lysis buffer, while soft tissue specimens like members of many dipteran families are more prone to be destructed by the lysis buffer, leading to a higher read yield. As already discussed, the proportion with which a species contributes to the DNA mixture is influencing species detection rate. Reads with higher abundances will most likely capture more primer molecules leading to a high amplicon yield, while lower abundance reads may not yield any amplicons (Hajibabaei et al. 2011). To circumvent these bias deeper sequencing can be performed which goes along with a possible tradeoff in terms of non-specific binding and increased costs (Hajibabaei et al. 2011; Brandon-Mong et al.

2015). To prove if the success rate of the salt-extraction method is indeed biased by the morphology of specimens, tests with a mock community should be performed.

Within the suborder Nematocera two families were detected by molecular methods which were not morphologically indentified. For the coleopterans three families were exclusively recovered with metabarcoding. Given the fact that these organisms were not overseen by morphological identification, the absence of whole specimens of these families in the bulk sample does not necessarily point to false positive results. DNA of these species could be brought into the Malaise trap bulk samples in several ways such as from the gut contents of arthropods or in form of eDNA attached to insect being present in the sample. Regarding the gut contents of arthropods in a Malaise trap, it has been documented that many species tend to regurgitate their gut contents while being preserved (Zizka et al. 2018). Although preservative ethanol in each Malaise traps was exchanged at collection and finally removed before extraction, it cannot be excluded that parts of the gut contents remained in the sample. In addition, the organisms found in a bulk sample have been interacting with the environment before entering the malaise trap, introducing other potential sources of eDNA to the sample. A recent study has shown that eDNA can be extracted from the surface flowers, revealing the various pollinator species that have visited the flowers (Thomsen and Sigsgaard 2019). Flower pollinators may serve as potential vectors for the distribution of exogenous DNA into the Malaise traps. When grinding specimens prior to extraction it can be assumed that exogenous DNA represents only DNA traces in relation to DNA originating from whole specimen in the sample and will therefore likely be highly underrepresented and as a result outcompeted by more abundant sequences. However, the amount of DNA released by each specimen when performing DNA extraction with the above described non-destructive extraction method remains unknown. Therefore, it cannot be excluded that the here used non-destructive extraction method is more prone to introduce false positive results into analysis in comparison to methods based on grinding of insects.

Influence of Choice of Methods on Detected Community Patterns

Marker Choice

The two markers revealed different insect community patterns (figure II.15). PCoA plots indicate that across all seasons insect communities assessed with the 18S marker were not significantly different between the three by spruces dominated forest types (figure II.15). However, over the full year it was observed that insect communities associated with the pure beech stands were different from the one found at the remaining three forests. This is indicated by the formation of a distinct cluster. In contrast to that, insect species community composition detected with the COI marker were depending on season slightly different at all four forest stands. Regardless of choice of primer pair and reference database each forest type formed a distinct cluster (figure II.15). (figure II.15).

It can be assumed that these contradicting findings are a direct result of differences in taxonomic resolution of the two marker. Because of the lower taxonomic resolution of the 18S marker, less insect OTUs were detected. On higher taxonomic level, many taxa do not show a high degree of specialization, but on lower taxonomic level many species are highly adapted to the colonized habitat. Furthermore and as already discussed, the higher degree of conservation of the 18S marker may lead to the fact sequences originating from several species are clustered into a single OTU. As a result differences in species communities between the four habitats become blurred, hampering a precise assessment of the current status of the ecosystem based on total existing biodiversity.

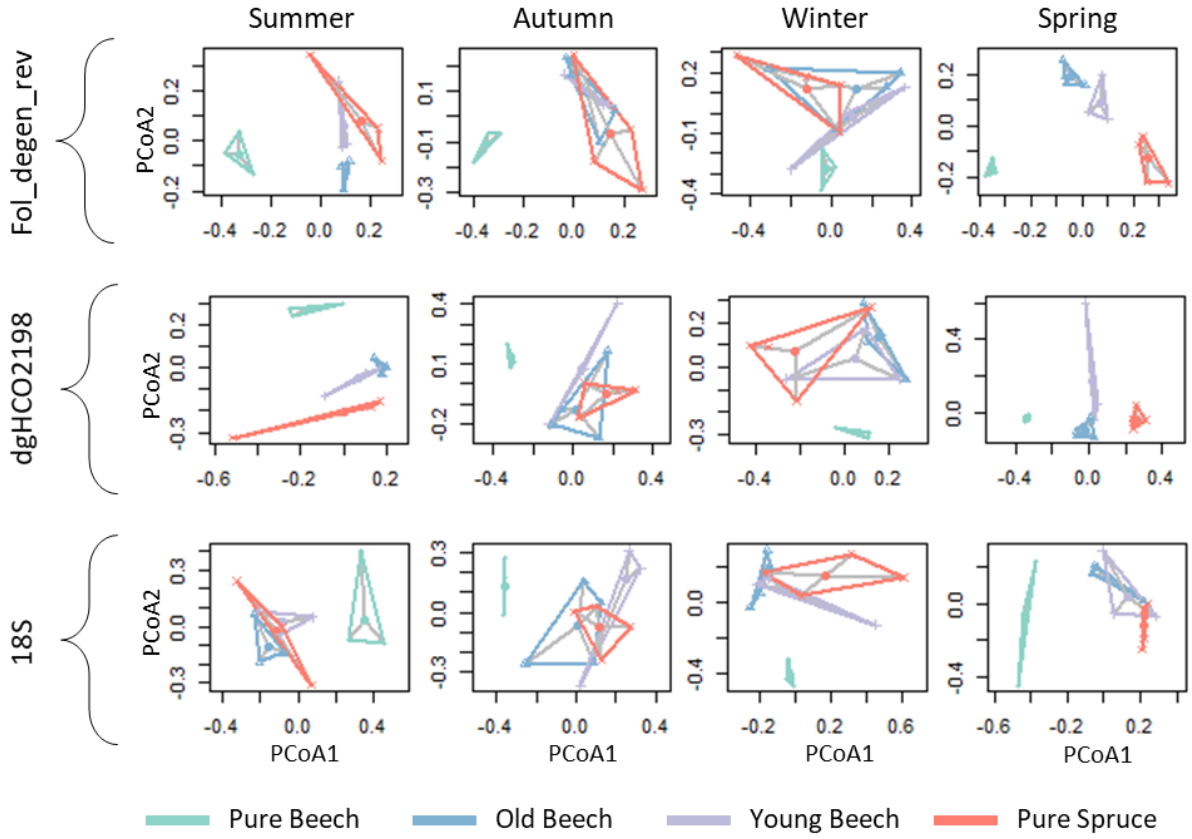


Figure II.15 PCOA plots indicating differences in Insect communities between forest types depending on sampling season, choice of marker, primer and reference database. PCoA plots were calculated on the basis of OTUs with an assignment to Insecta based on a blastID of at least 90%.

The resulting lower variation in insect diversity within the four forest stands is also mirrored in the calculated beta diversity. The Jaccard similiarty-index was calculated on the basis of a presence absence matrix based on all OTUs assigned to the class Insecta. Overall, it was found that Jaccard-similarity indices calculated for the 18S dataset were significantly lower compared to the ones calculated for the COI dataset, regardless of choice of primer (figure II.16). Regardless of primer and marker used, the lowest index was calculated for the comparison of the two monocultures, indicating strong differences in local insect communities of the two forest types.

The highest Jaccard-similarity indices were calculated for the two mixed stands, which is highlighting that insects communities of these forests were more similar compared to the one found at the two monocultures. When comparing insect community composition found in the pure beech stands with insect communities of pure spruce and mixed spruce stands, beta-diversity dropped from old beech over young beech to pure spruce stands. This finding was consistent for both marker datasets. For the pure spruce stands a less definite result was obtained. While the 18S dataset calculated an increase of the Jaccard-similarity index from pure beech over old beech to young beech stands, the COI datasets showed a higher Jaccard similarity index for the old beech than for the young beech stands (figure II.16).

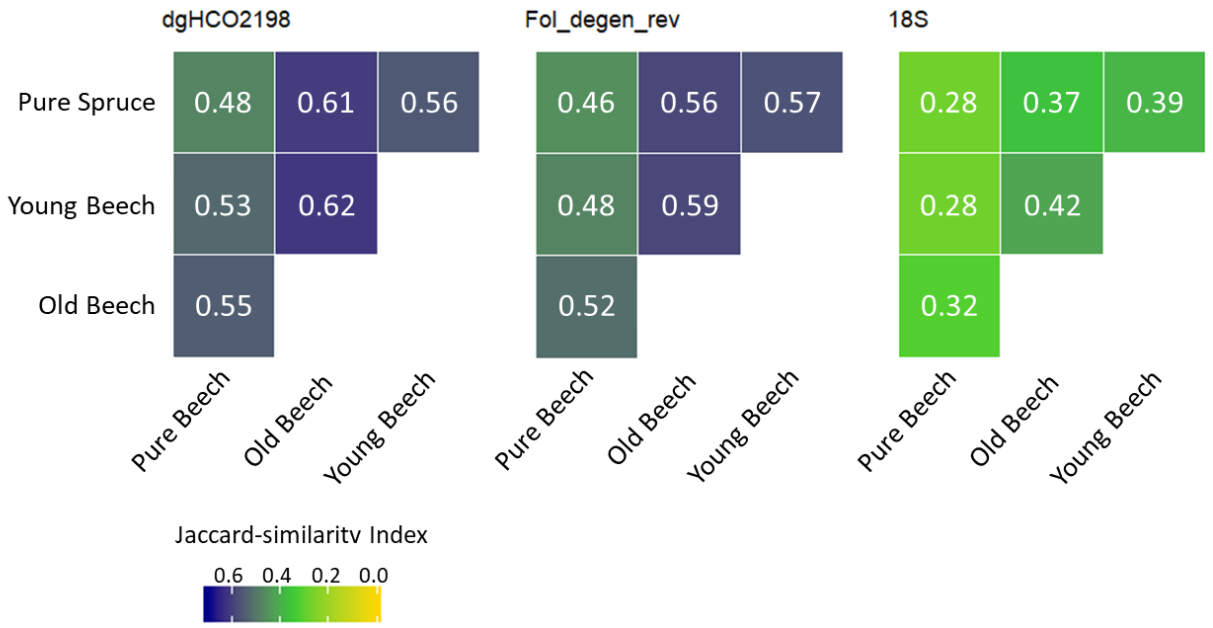


Figure II.16 Heatmaps showing jaccard similarity indices between forest types, depending on choice of primer and reference database. Analsis is exclusively based on OTUs assigned to Insecta with a blastID of at least 90%

Here it was shown, that marker choice significantly influences detected biodiversity patterns. However, a higher taxonomic resolution leads to a more accurate calculation of beta diversity and observed species turnover. In case of ecological studies, which are dealing with Insect diversity patterns, the use of a marker with a high taxonomic resolution should therefore be given priority over a marker with a broader taxonomic coverage.

Primer Choice

Although, calculated insect species communities were strongly depending on COI primer used (figure II.14), the detected species distribution patterns show only slight variations (figure II.15, figure II.16). PCoA plot based on assessed species communities calculated for the summer season were slightly differing between primers. While PCoAs based on the dgHCO2198 primer pair indicate that species community of the young beech and old beech sites were more similar to each other, the PCoAs of the Fol_degen_rev primer pair dataset indicate a higher overlap of species communities of the pure spruce and young beech sites (figure II.15). Similar patterns were observed in autumn season. Based on the Fol_degen_rev dataset, the pure spruce, young beech and old beech sites were clustering togetherr, PCoA plot calculated for the dgHCO2198 dataset showed overlapping clusters for the old beech and young beech sites as well as for the old beech and pure spruce sites, but not for the young beech and pure spruce stands. In winter and spring the two primer pairs revealed similar community patterns (figure II.15). As already discussed has the choice of primers a significant influence on species identification rate. The lower the taxonomic level the more significant the observed differences between the two primer pairs in terms of assessed community composition. On family level nematoceran communities were only slightly differing between the two primer datasets (figure II.13). However, on species major differences in communities are observed (figure II.17). Surprisingly found the less

degenerated dgHCO2198 primer pair more OTUs with an assignment on species level (figure II.17A). However, several of these OTUs were assigned to the same accession number and thereby to the same species. After merging these double assignments, number of detected OTUs with assignment on species level was slightly higher for the Fol_degen_rev dataset (figure II.17B). Next to that, number of high-quality species assignments, meaning assignments with a blastID of at least 99% were slightly higher with Fol_degen_rev primer pair (figure II.17C,D).

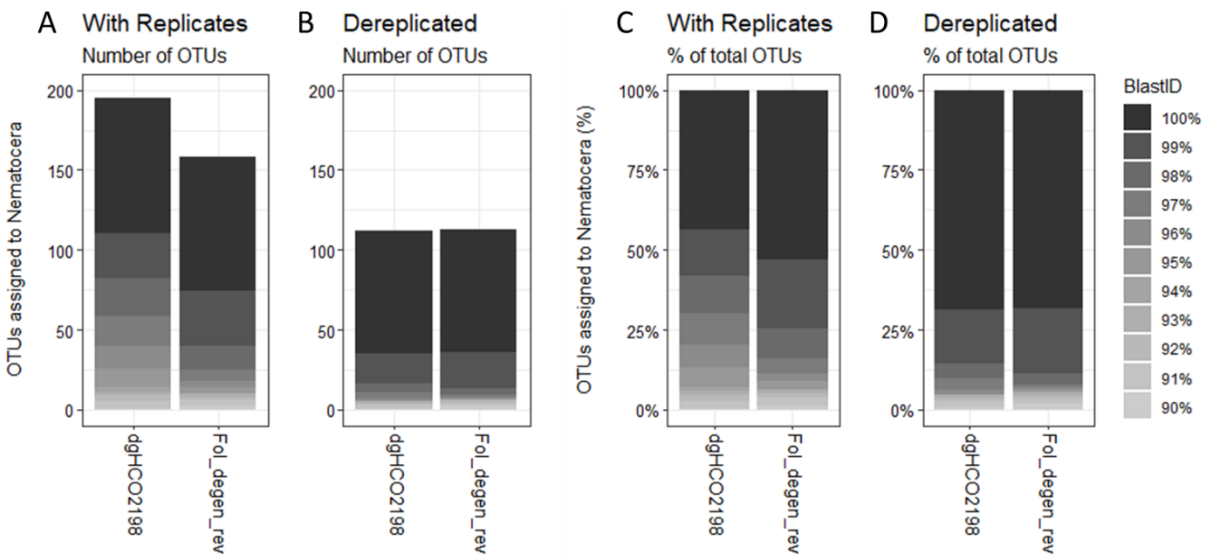


Figure II.17 Total and relative number of OTUs assigned on species level to the dipteran suborder Nematocera. Shade of coloration indicates blastID of assignment. While figure A and C are including all OTUs assigned to Nematocera on family level, figure B and D are based on a dereplicated dataset, meaning that OTUs assigned to the same accession number were counted as one.

Despite that similar number of OTUs were assigned on species level to an unique entry of the reference database, resulting species list were strongly differing between primer pairs (figure II.18A). For Nematocera slightly more species were detected using the Fol_degen_rev primer pair

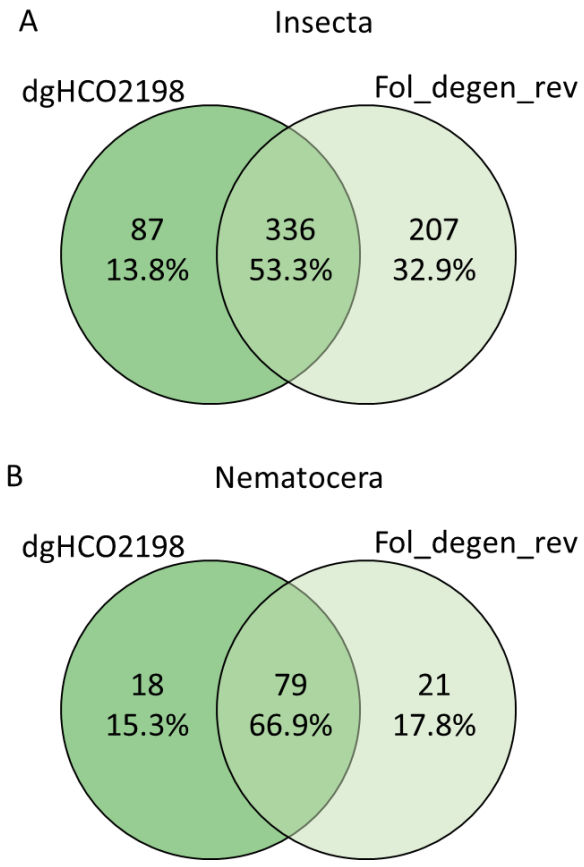


Figure II.18 Venn diagrams showing Number of detected insect (A) and nematoceran (B) species (blastID ≥ 99%) depending on choice of primer pair.

(Fol_degen_rev: 100 species, dgHCO2198: 97 species). Out of them 79 species were detected by both primer pairs, making up 66.9% of total recovered nematoceran species diversity (figure II.18B). On broader taxonomic scale, encompassing all recovered species of the class Insecta only half of all species were recovered by both primer pairs (figure II.18A). These results clearly show that a lack of knowledge about species composition within a single Malaise trap bulk sample makes general predictions on primer performance based on their degree of degeneracy premature. Nevertheless, it clearly demonstrates that the choice of primer plays a important for the detection probability of several species. As already discussed is composition and number of species contained in a bulk sample influencing amplification and thereby detection rate of species (Morinière et al. 2016). Former studies which rely on the exact identification of specimens on species level are therefore increasingly using a “one-locus-several-primer-sets” strategy (Corse et al. 2019). This is attributed to the fact, that some taxa are recovered by one primer set but not by another and vice versa. Studies on species level rely on a taxonomic resolution as precise as possible. False negative results can lead to wrong conclusion about species behavior, species occurrence, species preference and even protection status assigned to the corresponding species and finally to wrong conclusions about the effect of abiotic on biotic factors on species occurrence.

Conclusion – Part I

Here we were able to show that choice of marker, primer and reference database can significantly influence species detection rate. Because of its higher taxonomic resolution and the higher completeness of available reference database should priority rather be given to the COI over the 18S marker. Although, both marker revealed similar species community patterns, with the COI marker retrieved species lists were more detailed. In views to conservation studies, species lists should be as complete as possible as not the total number of recovered species provide information about conservational value of the forest, rather the number of endangered, specialists and endemic species should be considered.

Our study has shown that choice of primer can dramatically influence resulting species lists. Although, the degree of degeneracy of primers is positively influencing taxa detection rate, degeneracy is not universal tool. Here several species were exclusively found with the less degenerated dgHCO2198 primer pair. Former studies which rely on the exact identification of specimens on species level are increasingly using a “one-locus-several-primer-sets” strategy (Corse et al. 2019). The here presented results are encouraging this trend as the combination of a highly degenerated primer with a second primer pair resulted in the detection of up to 16 percent more species.

Results & Discussion

Part II – Ecological Analysis

Because of the above discussed findings, the following ecological analysis is based on a “on one-locus-several-primer-sets” strategy. After combining the COI datasets all sequences were blasted against the GBOL reference database. OTUs with an assignment on species level to Insecta were checked for double assignments and if necessary subsequently merged. In total 717 insect species with a BlastID of at least 90% were identified. Out of them, 630 species were identified with a blastID of at least 99%. Only those were taken into account for ecological analysis.

Influence of Biotic and Abiotic Conditions

Depending on season, we found distinct insect community compositions among the four sampled forest types. Overall species richness was highest at the old beech stands (425 species), followed by the pure spruce (390 species), pure beech (377 species) and finally young beech sites (347 species). Hoewver, depending on season these proportions shifted. We observed that degree of fluctuation in species diversity between seasons strongly depended on composition of dominating tree species (figure II.19).

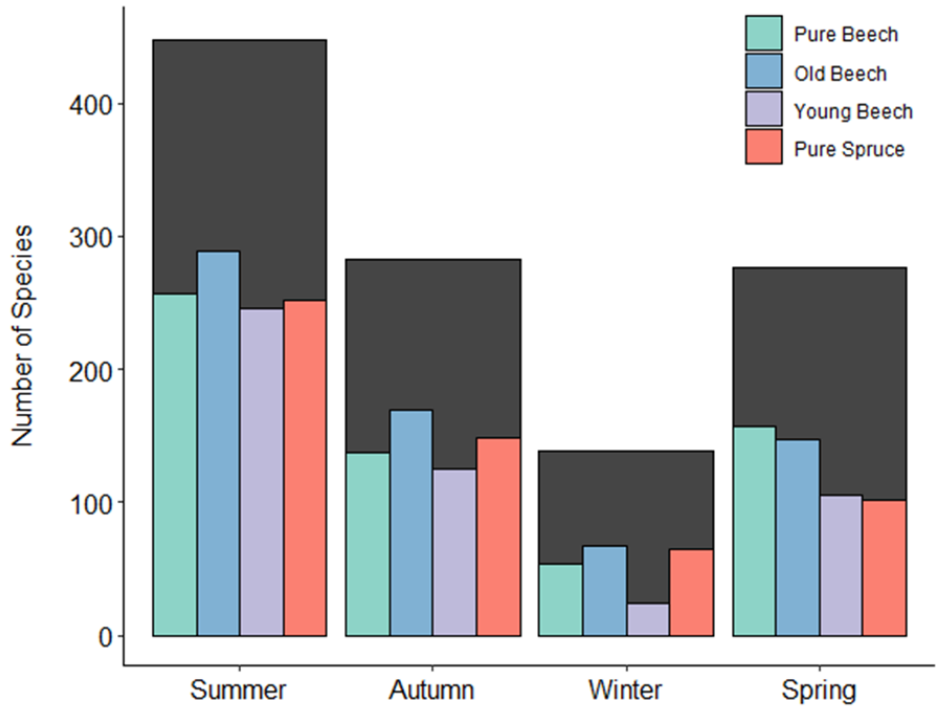


Figure II.19 Number of species (blastID ≥ 99%) found per forest type and season. Black bars are indicating total number of recovered species per season. Green bars are indicating number of species recovered from the pure beech sample sites, blue bars are number of species found at the old beech sample sites, purple bars show number of species detected at the young beech forest stands and red is indicating number of species recovered from the pure spruce stands

Seasonal Fluctuations within Habitats

Former studies have already shown that seasonal and annual variation in environmental conditions lead to changes in population size and thereby to changes in community structures (Schowalter 2016). This study is confirming these results. Within all four sampled forest types, species diversity was shaping a reverse humped curve over the one-year sampling period. A peak

was reached in summer season, before species diversity dropped again until a low was reached in winter with approximately more than 200 species less (figure II.19).

Insects are poikilothermic organisms, which body temperature is closely linked to prevailing air temperature (Szujecki 1986). As a result, degree of activity of insects is mainly driven by air temperature. In detail, it has been shown that insect activity increases until temperature reaches an optimum but past the optimum activity will decrease again (Williams and Osman 1960). As Malaise traps are flight interception traps, its efficiency is directly linked to species degree of activity. Because of the harsh abiotic conditions in wintertime, insect activity decreases to a minimum resulting in a strongly reduced number of caught insects. Nevertheless, the widely held assumption that insects are unable to regulate their body temperature is not true. Several species show adaptions to the prevailing abiotic conditions in winter allowing them to stay active. Next to behavioral strategies, which includes avoidance or active seeking of warm or cold places (Whitman 1988), also physiological strategies were invented e.g. by changing rates of evaporation and respiration (Prange 1990; Heinrich 2013), control of body temperature due to an increase in muscular activity (Heinrich 2013) and the accumulation of polyhydric alcohols and antifreeze proteins can be found within the cells of many insect species (Block 1990). Finally, body shape and body color are often adapted to the local climatic conditions (Heinrich 2013). Here one OUT was found, which had an 93.7% assignment to the Mecoptera family Boreidae (figure II.20). The family Boreidea was exclusively detected in winter season, which is the typical time of appearance of the adult form of this family (Glime 2006) earning it the English name 'snow scorpionflies'. Snow scorpionflies are dark

Figure II.20 Photography of voucher specimen of *Boreus westwoodi* (Mecoptera: Boreidea) (© BY-SA 4.0: GBOL / Museum Koenig)

colored, allowing them to adsorb long-wave and short-wave radiation. This enables the regulation of body temperature to a certain extent and unlinks species body temperature from the prevailing air temperature (Courtin et al. 1984).

Influence of Day Length on Insect Activity

Changes in seasonal activity of the Mecoptera family Boreidae shows that light plays an important role in the life of most insects. Only insects living in caves or deeper soil layers are not directly affected by light, while all other insects are at least in one life stage highly driven by it (Szujecki 1986). On a seasonal basis is light the main abiotic factor timing reproduction, activity-patterns, development, body structure and even sex-determination in insects. In temperate zones, day length is usually closely linked to changes in temperature and humidity, which makes it to a reliable and good predictor for insects as it ensures a perfect timing for mating and the development of the different life stages in views to seasonal conditions e.g. food availability (Gullan and Cranston 2014). The seasonal development cycle of most Insects is therefore to wide parts determined by day length (Gullan and Cranston 2014). Day length is acting as a signaling factor stimulating a complex physiological process during which endocrinal glands release hormones, which in turn activates the sexual and moulting glands (Szujecki 1986). Day length, light intensity, incoming wave length as well as temperature plays an important role for the onset and offset of diapause in the life of insects during winter season (Gullan and Cranston 2014). Time of onset and offset of diapause can directly influence species occurrence within the different seasons.

For most species only the imagoes are targeted with Malaise traps. Therefore, the life stage in which an organism hibernates is often linked to time at which species can be captured with Malaise traps in the next year. While species hibernating as imagoes usually occur earlier in the year, species which overwinter as larvae or in the egg stage usually first complete development

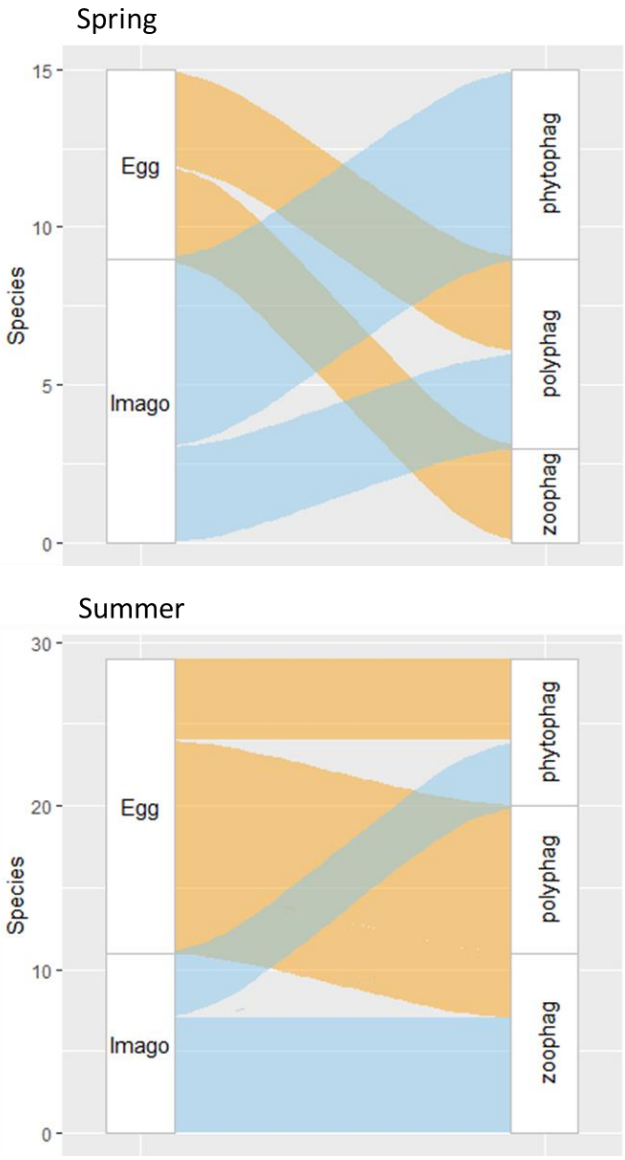


Figure II.21 Number of detected Heteropta species in summer and spring season depending on hibernation strategy and preferred diet

before emerging. They are usually appear a bit delayed in the Malaise trap catchments. Depending on species, both hibernation strategies can be observed within the suborder Heteroptera (Insecta: Hemiptera). Here, 15 heteropteran species were detected in the Malaise trap catchments collected in April. Later in the year, in summer 29 species were found (table II.3). Out of the 15 species found in spring, six are known to hibernate in the egg stage and nine as Imagoes (figure II.21). In summer season, this proportion shifted. Out of the 27 detected species, 18 have been described to overwinter in the egg stage and only eleven as Imagoes. We conclude that species hibernating in the egg stage are more likely to occur later in the year. A closer look at the diet of the discovered species revealed that none of the six species in the egg stage hibernating species which were found in spring season, was having an exclusively phytophagous diet, while out of the 18 in summer detected egg-hibernating-species 5 are known to be exclusively phytophagous (figure II.21). Some insects are adapted to feeding on early leaves, as leaves can become toxic or too tough for a certain species to feed on (Feeny 1970), while other species can occur later in the year, as they require other food resources. It can be concluded that hibernation

and foraging strategy are closely linked to time of species appearance within a habitat and thereby to detection probability with Malaise traps.

Influence of Forest Types on Seasonal Dynamics of Insect Communities

As discussed is day length one of the main drivers for the onset and offset of several activity patterns as it is fairly independent from any prevailing condition whereas further abiotic factors like humidity, direct solar radiation and temperature can quickly change on small geographical scale, characterizing several temporal ecological niches, which are occasionally occupied by various insect groups. As a result, it might be that some groups of insects are more diverse in habitats characterized by certain abiotic and biotic conditions than others. A strong influence on prevailing abiotic and biotic conditions is having the dominating tree species (Leuchner et al. 2011).

Within all four sampled forest types, the highest number of species found were members of the order Diptera. At least 50% of all detected insect species per forest type and season were dipterans (figure II.22). Next to Diptera, the orders Coleoptera, Hymenoptera, Lepidoptera and Hemiptera were showing a high degree of diversity across all seasons within all forest types. The remaining ten here detected order were not represented by more than 10 species within each season strongly limiting its informative value about the current status of the ecosystem. Therefore focus was set on the five highly abundant orders.

All of them show strong variations in diversity between forest types and seasons. Additionally, they have in common that a peak in diversity was found in summer season within all four forest types (figure II.22). In spring, the highest diversity of Diptera, Lepiodptera and Hemiptera showed was observed at the pure beech sample sites, while hymenopterans were most diverse at the old beech sampling sites (figure II.22). These differences in species diversity between forest types

depending on insect order is the result of a complex interplay of abiotic and biotic factors in combination with species lifestyle and adaptations.

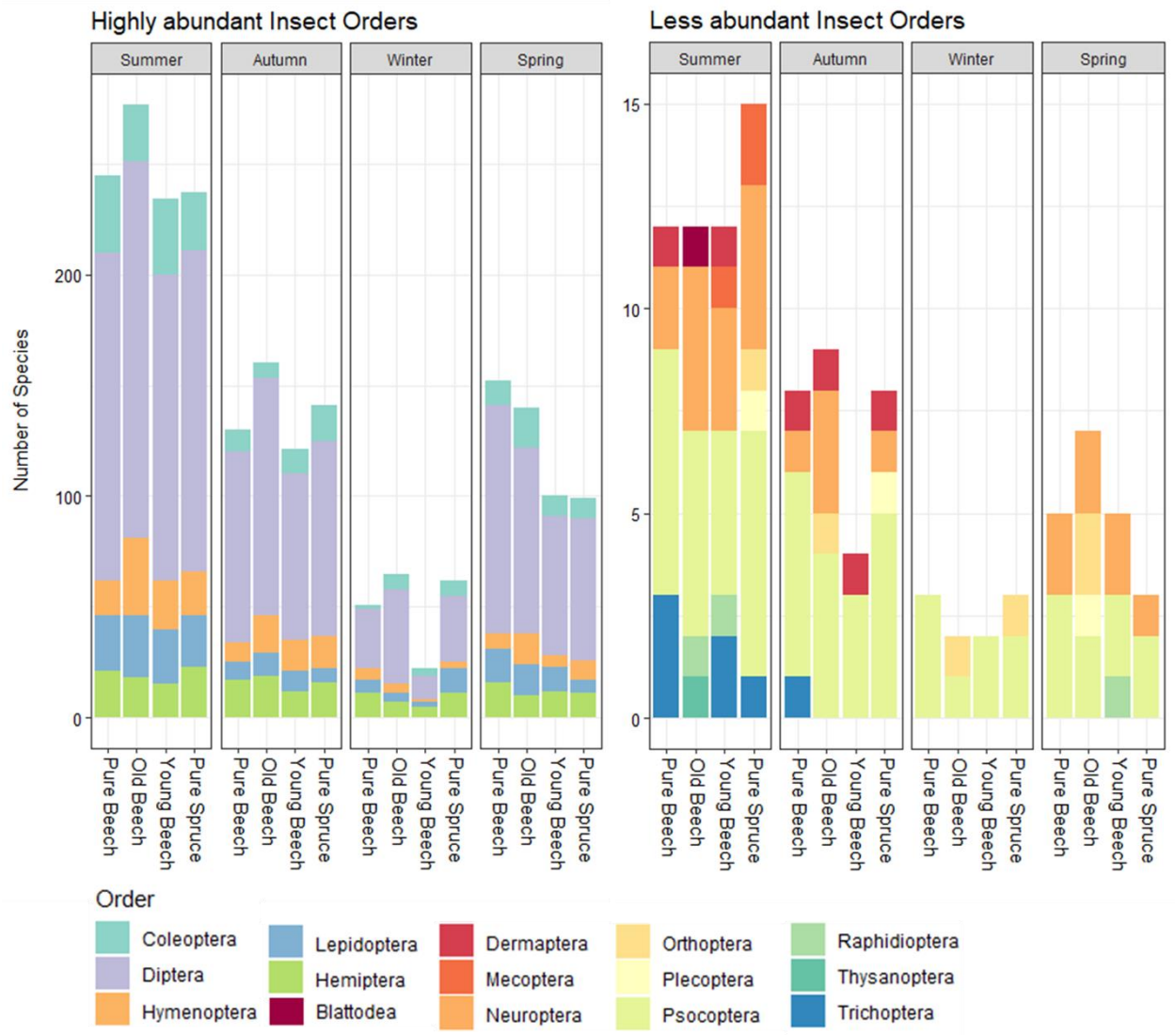


Figure II.22 Number of species per insect order recovered from each forest type depending on sampling season.

At the beginning of the year, the pure beech sample sites are flooded with light. The leafless canopy is not shading the forest floor and the incoming sunlight, which intensity and duration increases during spring will warm up air temperature and thereby also soil and foliage. Mean soil temperature at the pure beech sample sites in spring had already reached 12.3 °C while mean soil temperature at the pure spruce sample sites was 6.8 °C. The higher temperature at the pure beech sample sites, together with an increased amount of light had possibly already induced spring growth while the colder spruce stands were still in the state of dormancy. Phytophagous Insects and their corresponding host plants are often stimulated by similar temperature and photoperiodic cues to initiate spring growth and activity (Chapin III et al. 2011). Depending on species, diapause is terminated by increasing length of photoperiod or increasing temperatures (Koštál 2006; Gill et al. 2017). As the light intensity and temperatures are increasing more sharply in the leafless beech forests compared to the shady spruce sample sites, a reactivation of insect activity will probably occur earlier in the year at the pure beech sites than in the shady pure and mixed spruce stands. This would explain for the higher degree of diversity found in spring at the pure beech sampling sites. More insects had already been triggered to emerge from the hibernation grounds, than at any other location. This is also indicated by calculated jaccard-similarity index between seasons within each forest type (figure II.23).

In all forest stands, the calculated jaccard indices were lowest for the winter season compared to any other season, indicating that a comparatively small number of insects were active in winter. Nevertheless, only at the pure beech stands a comparatively high beta diversity was calculated for comparison of insect diversity of spring and summer season (0.29), while at the remaining three forest types the calculated jaccard similarity was rather low (Old Beech:0.19; Young Beech:0.17 Pure Spruce:0.15). Especially in comparison to the one calculated for the same forest types for comparison of local diversity of summer with the one of autumn season (Pure Beech: 0.29; Old Beech: 0.28; Young Beech: 0.22 Pure Spruce: 0.21) (figure 19). This is indicating that already in spring conditions at the pure beech sites met the requirements of species

occurring in summer season, which was only to a lesser extent the case at the spruce dominated forests.

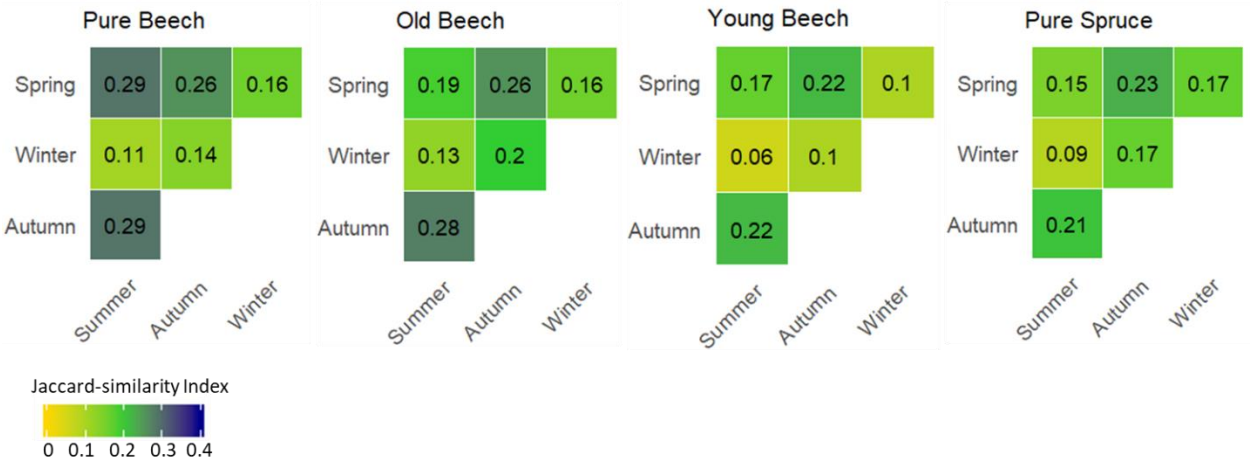


Figure II.23 Jaccard similarity indices of insect communities between seasons depending on forest type

Like Insect communities are also plant communities of the undershrub and herb layer highly influenced by the dominating trees (Budde et al. 2011). It has been shown that the herb layers of mixed forest stands are known to be more diverse compared to the one of monocultures. Furthermore, former studies found that understory species composition of spruce-beech mixed stands are more closely related to the one found at pure spruce stands than to understory species composition associated with pure beech stands (Budde et al. 2011). This is mainly due the fact that thick layers of foliage are usually covering the forest floor of beech monocultures. The foliage has a toxic effect on many tree seedlings resulting in a strongly reduced understory growth (Budde et al. 2011) (Figure 28). A reduction in understory growth has a direct influence on faunal diversity as an increase in herb layer and undershrub diversity goes along with an increase in faunal species diversity (Ekschmitt et al. 2003; Bos et al. 2007). Many Insects like pollinators and phytophagous species are feeding on the plants of the herb and undershrub layer (Proesmans et

al. 2019). While spruces and beeches are anemophilic (wind pollination) species which depend on anemochory (wind dispersal), most plants of the herb layer have showy floral displays, indicating that insects and other animals are potentially important pollinators (Barrett and Helenurm 1987). Looked at the opposite way, this means that a broader range of possible food resources are provided for insects the more diverse the undershrub layer is.

Additionally, it has been observed that in temperate deciduous forests the understory shrub is generally developing leaves much earlier than the dominating trees (Walter 1984). Herbs and trees are following different life strategies. While herbs are starting to grow and photosynthesis is onset, as soon as that conditions are favorable, larger trees usually start a bit delayed into growing season (Walter 1984). This is mainly contributed to their physiology. Herbs have only a very narrow root system, not providing any access to deeper soil structures. That makes them very susceptible to short term droughts (Walter 1984). In contrast to trees, herbs won't reduce their photosynthetic activity. Moreover, photosynthetic activity is performed, until the complete aerial shoot system and leaves are dead (Walter 1984). In contrast to that is the photosynthetic activity of trees induced but also inhibited by changes of the abiotic environment (Walter 1984). These major differences in plant life strategies can result in a much longer photosynthetic active phase of herbs. The complexity and diversity of the herb layer and undershrub of a forest can therefore drastically influence arthropod species composition of a habitat as it provides several food resources.

The fact that the hymenopterans were more diverse at the old beech and pure spruce stands, can likely be a result of different lifestyles. While at the pure beech sites seven hymenopteran species were recovered in spring, 14 hymenopteran species were found at the old beech sites (figure II.22). Next to an increase in number of species associated with a parasitic life style, especially the family Tenthredinidae was more diverse at the old beech stand (table II.7). A high proportion of Imagoes of the family Tenthredinidae are known to be common flower visitors, while the larvae are usually feeding on plant parts. Former studies have already shown, that

diversity of the herb layer is directly influencing pollinator abundances (Proesmans et al. 2019). We therefore conclude that the higher number of possible food resources of forest stands characterized by a highly diverse undershrub and herb layer is likely to be the driver for the increase in number of pollinator species at the old beech sites.

Influence of Forest Type on Insect Community Composition

Highest total number of Insect species was found at the old beech stands (425 species) (figure II.11). The young beech stands, which were differing from the old beech stands in age of the underplanted beeches, harbored 347 species which was the lowest number of species detected within all four forest types. With 390 species 30 more species were recovered from the pure spruce stands in comparison to the pure beech forests (Figure II.19). Next to total number of recovered species, seasonal species community composition were strongly differing between forest types (figure II.24A)

The calculated PCoA Plots based on species composition recovered from the three and respectively four sample sites of each forest type showed depending on season different patterns (figure II.24A). In summer and autumn season a distinct cluster was calculated for the pure beech sites, while the remaining three forest types were clustering together. In winter season a similar picture was found, with the exception that the clusters of the pure beech and old beech sites were also slightly overlapping. In spring a distinct cluster was formed for each forest type (figure II.24A). When calculating the distance of each datapoint to the centroid of the corresponding cluster we observed that especially samples of the autumn and spring season taken within one forest type showed a high degree of heterogeneity (figure II.24B).

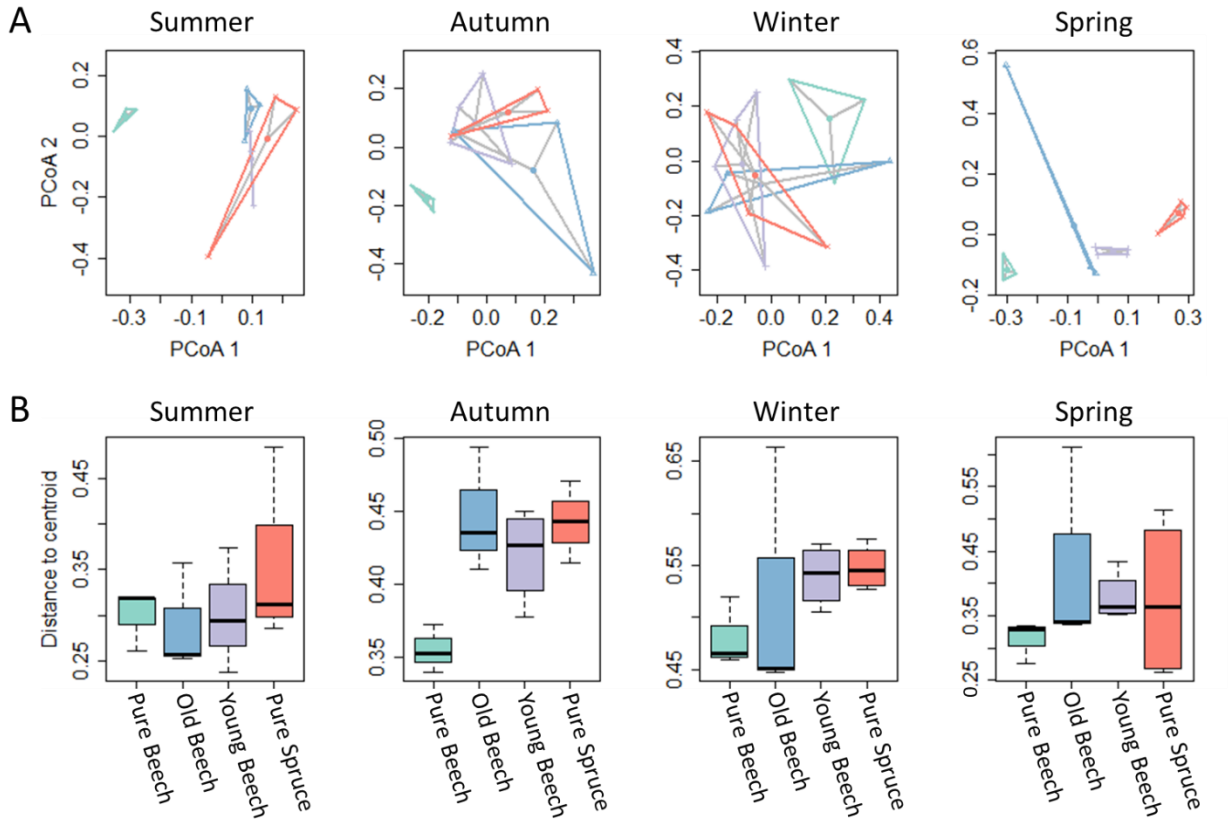


Figure II.24 Calculated PCoA per forest type and season (A). The more distinct a formed cluster, the less similar the associated insect species community composition to the remaining three forest types. Figure II.21B shows the degree of homogeneity between samples per forest type and season. The lower the distance to centroid value, the more homogenous are the replicates.

A closer look at the number of shared and unique species within biological replicates per forest type and season clearly showed that detected insect diversity is strongly varying even within forest types (figure II.25), confirming results indicated by calculation of distance to centroid for each forest type per season (figure 24B). Within each season, only a small proportion ranging from 1.96% to 34.6% of all OTUs were found in all three and respectively four biological replicates

of each forest type per season. Moreover, up to 50% of total species count retrieved from a single forest type within one season was recovered from only biological replicate.

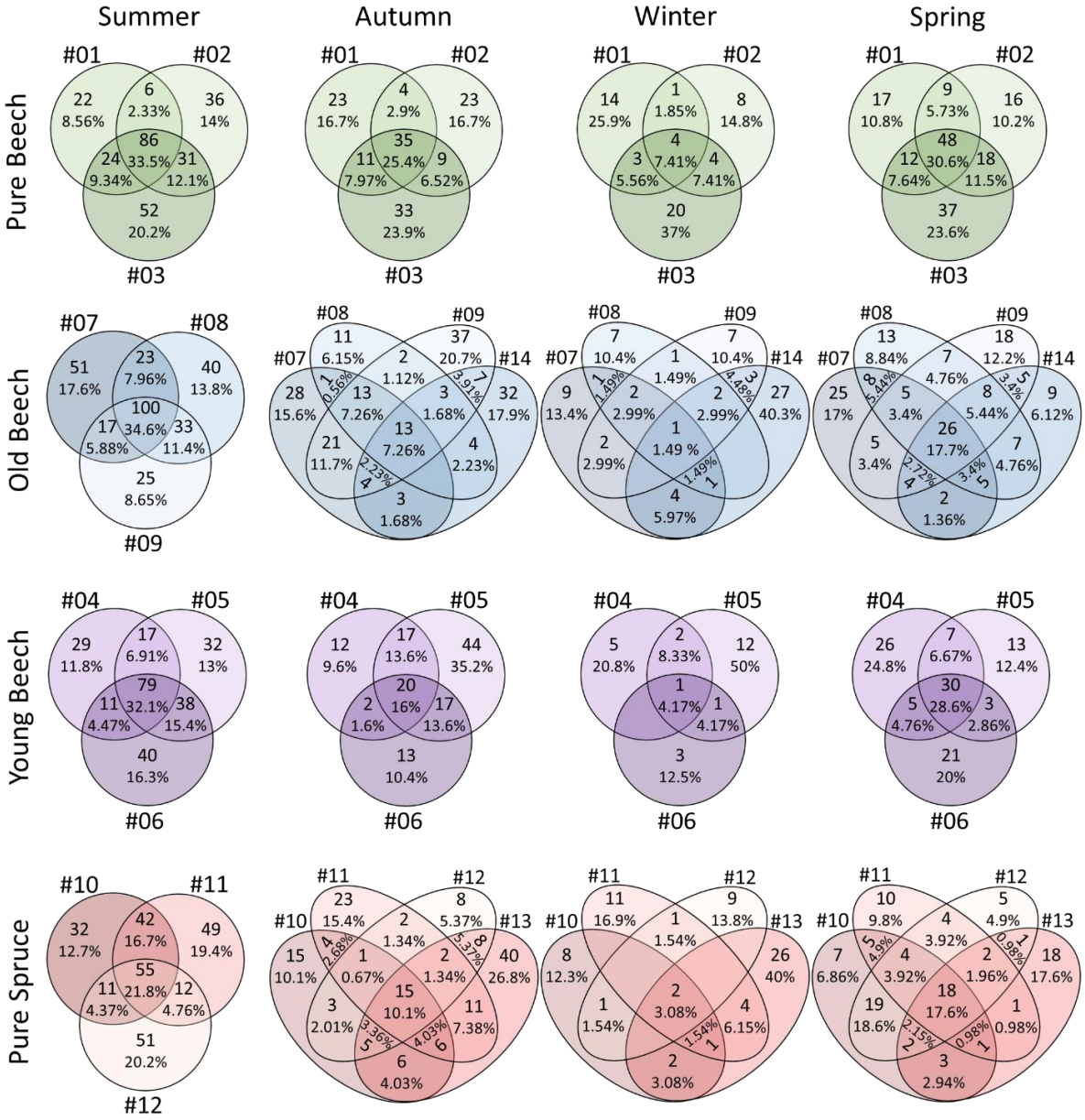


Figure II.25 Number and relative percentage of unique and shared species between sample sites of each forest type depending on season

Although, relative number of unique species were differing between seasons, relative number of shared species between all sample sites of one habitat did not exceeded 34.6% at any point. A high percentage of unique species can indicate that numbers of samples taken were not sufficient to assess total existing diversity. Nevertheless, it should be kept in mind that many insects are highly active and are capable to travel long distances. Therefore, it cannot be excluded that several of the caught insects are possibly just passing through a habitat without being adapted or strongly interacting with the local fauna and flora.

Therefore, here it is differentiated for further analysis between **visitor species**, which were detected in only one biological replicate per forest type and season and **permanent residents,** which were found in more than one biological replicate per forest type and season. Taking only permanent residents into consideration the number of detected insect species per forest type dropped significantly (Figure II.26) by up to 60%. Out of the formerly 630 detected species, 367 were identified as permanent residents of at least one of the four habitats. The remaining 263 species were classified as visitor species (Figure II.26). While in summer and spring the greater part of detected species were indeed permanent resident species, in autumn and winter more than half of all detected species were potential visitors. Especially in winter the proportion of permanent resident species were lower than 30% (figure II.26).

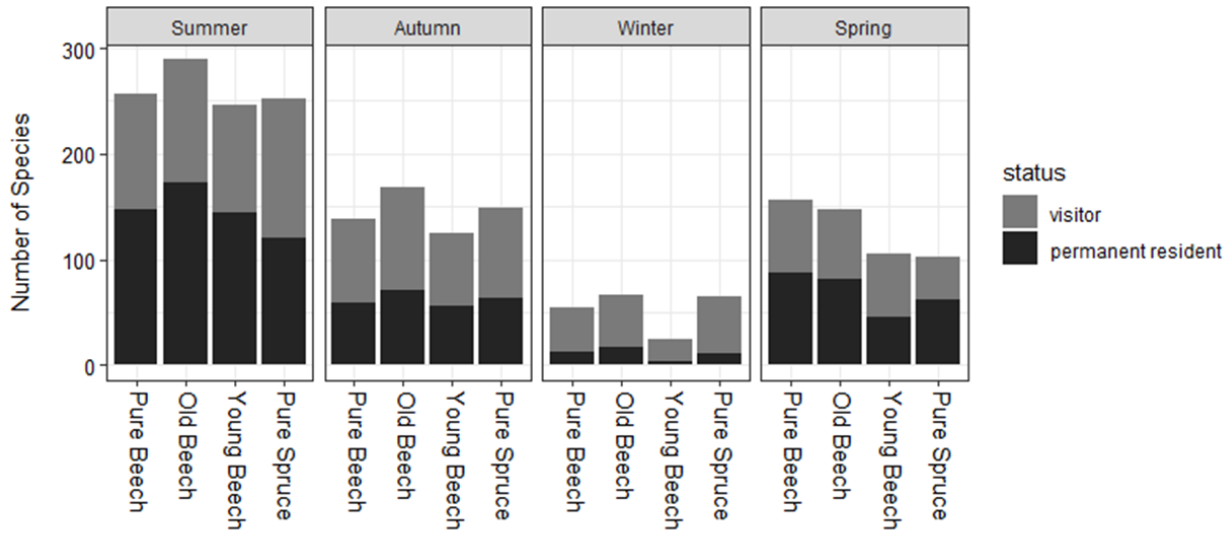


Figure II.26 Number of permanent resident and visitor species associated with each forest type depending on season

After excluding all visitor species from the dataset, we found that depending on season between 3.23% and 21% of all recovered species were present in all four forest types (figure II.27). Nevertheless, in winter season only 31 species were left in the dataset out of which 70.9% occurred only within a single forest type. In the remaining three seasons between 36.98% and 53.49% of the detected species were unique to one of the four forest types (figure II.27). The highest relative number of unique species were found in spring season. In detail, the highest number of species exclusively found in one habitat was found at the pure beech sites (34 species) which is further underlining the above discussed variation in degree of influence of dominating tree species on abiotic and biotic factors influencing species diversity in spring.

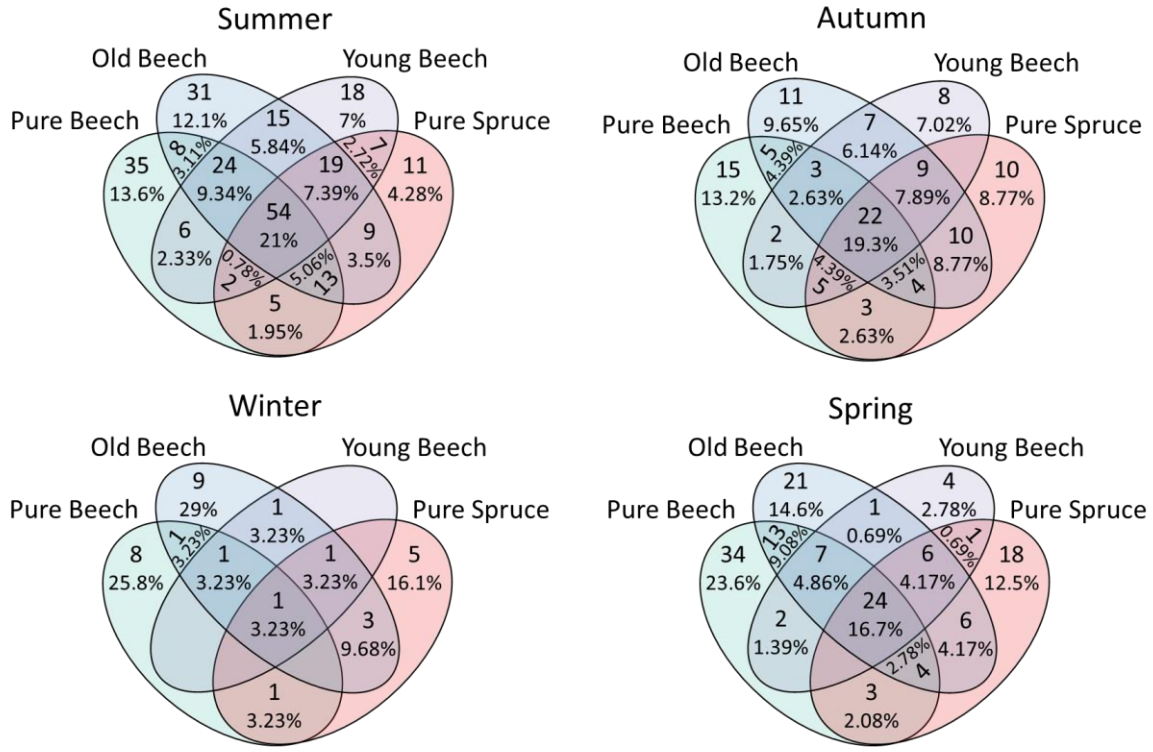


Figure II.27 Venn Diagram showing number and relative percentage of unique and shared permanent resident species between forest types depending on season

While a comparatively high number of permanent resident species occurred within several habitats the calculated Upset-R plot showed that a high number of permanent resident species were exclusively occurring in one of the four sampled seasons and were not swapping between habitats between the seasons (figure II.28). Furthermore, especially at the pure beech and old beech stands many permanent resident species were found in summer and spring, but not at any other forest type or season. The upset R plot only includes intersection encompassing more than two species. Out of the 367 detected permanent resident species, 230 were exclusively found in one season. However, number of shared species between summer and autumn season was higher compared to number of species occurring in summer and spring season, again indicating that several species had not yet emerged from hibernating sites in spring

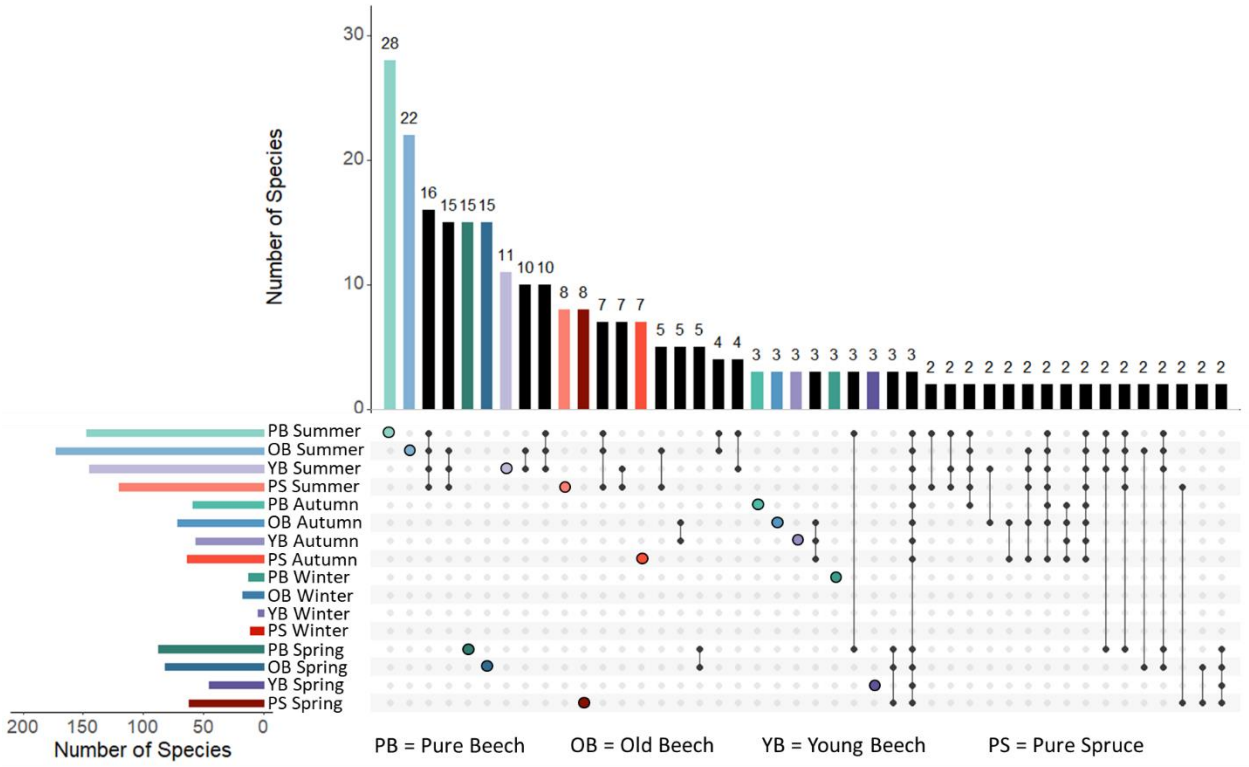


Figure II.28 Upset-R Diagram showing number of unique and shared permanent resident species between forest types, depending on season. The Diagram is only showing intersections encompassing more than one species.

Detected Specialists

Species presence and absence is directly linked to available food resources (Proesmans et al. 2019). Especially within phytophagous insects, food preferences play a major role for species occurrence in certain habitats. Within the beech forest as well as within the spruce forests we found a range of species which occurrence is in the literature often associated with the respective dominating tree species of the two monocultures. These species are usually specialists of which

at least one development stage depends on the dominating tree species as food source. If focus is set on potentially permanent resident species we were able to identify several specialists for the two dominating tree species.

Specialists were found within all major Insects order. A closer look at the resulted species lists (Table II.3-II.8), clearly showed that especially the less diverse groups e.g. Neuroptera were more likely to be recovered from the pure beech sites, compared to the pure spruce sites (table II.3). In contrast to that, hymenopteran species were more likely to be permanent residents at the spruce than at the pure beech stands (table II.7). Overall, the highest number of permanent resident hymenopteran species were recovered from the old beech stands in summer season. The reasons for these findings have already been discussed.

The species *Hemerobius pini* and *Coniopteryxs pygmaea* (figure II.29), both members of the order Neuroptera has been described to be strictly bound to coniferous trees (Paoletti 2012). Indeed, both species were solely found at sampling sites where spruces were present (table II.3). Within the groups of Lepidoptera the four species *Chionodes electella, Eupithecia tantillaria, Epinotia pygmaeana and Epinotia tedella* (figure II.29) were identified. Although they belong to three different families they have in common that the larvae are specialized on feeding on the needles of spruces. Out of them, none species was found at the pure beech sites, while two were detected in all of the three forest stands characterized by the presence of spruces (table II.8).

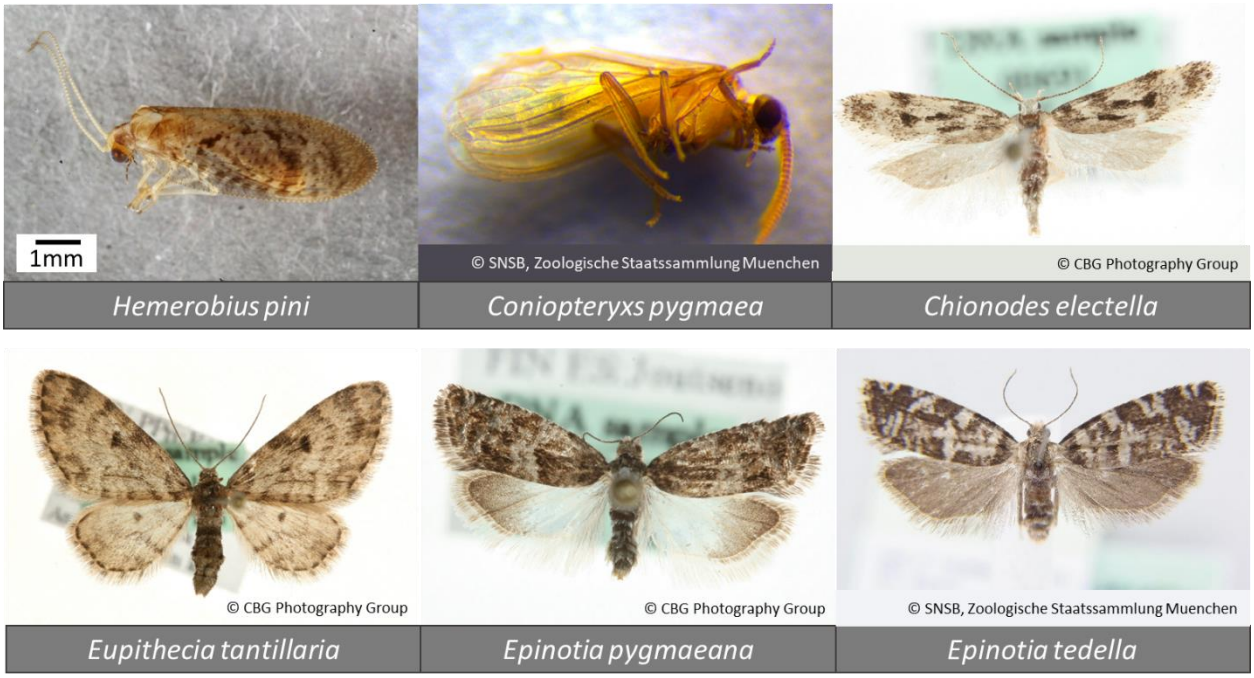


Figure II.29 Photos of voucher specimens of species which occurrence is associated with spruces: *Hemerobius pini* (Insecta: Neuroptera), *Coniopteryxs pygmaea* (Insecta: Neuroptera), *Chionodes electella* (Insecta: Lepidoptera), *Eupithecia tantillaria* (Insecta: Lepidoptera), *Epinotia pygmaeana* (Insecta: Lepidoptera) and *Epinotia tedella* (Insecta: Lepidoptera) (if not other stated: © BY-SA 4.0: GBOL / Museum Koenig)

For the lepidopterans specialized on foraging on beeches a similar pattern was found. Six species were found which have been described to exclusively feed on beeches. Out of them the species *Nematopogon swammerdamellus, Cyclophora linearia, Operophtera fagata* and *Stigmella tityrella* (figure II.30) were exclusively detected in the pure beech stands (table II.8). The two species *Phyllonoryceter maestingella* and *Stigmella hemargyrella* (figure II.30) are also known to have larvae which are specialized on feeding on beech. In contrast to the larvae, are adult lepidopterans lacking a foraging preference. Furthermore, their ability to fly enables them to travel long distances. This could be an explanation why both above mentioned species were next to the pure beech and mixed beech stands also found at the pure spruce stands (table II.8).

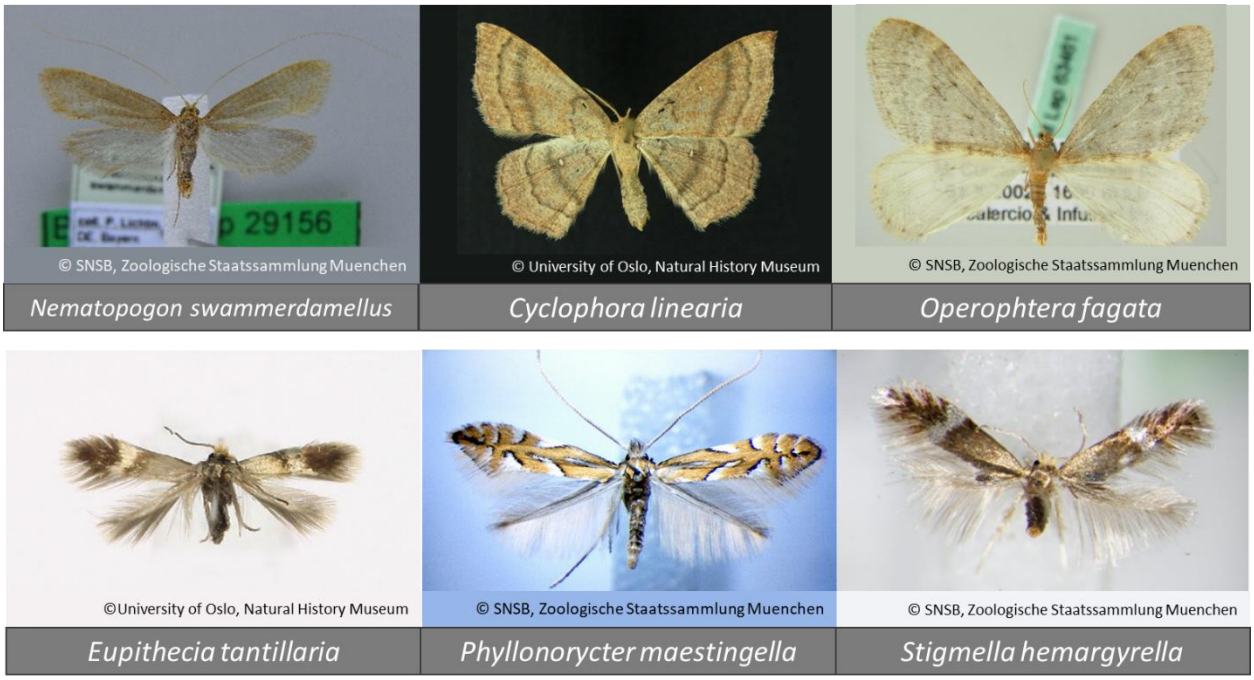


Figure II.30 Photos of voucher specimens of species which occurrence is associated with beeches: *Nematopogon swammerdamellus* (Insecta: Lepidoptera), *Cyclophora linearia* (Insecta: Lepidoptera), *Operophtera fagata* (Insecta: Lepidoptera), *Stigmella tityrella* (Insecta: Lepidoptera), *Phyllonoryceter maestingella* (Insecta: Lepidoptera) and *Stigmella hemargyrella* (Insecta: Lepidoptera)

Specialists can not only be found within the order Lepidoptera. Out of the Insecta order Coleoptera three species belonging to the genus Ernobius (Coleoptera: Anobiidae) were identified. *Ernobius abietinus, Ernobius abietis and Ernobius angusticollis* (figure II.31). Their occurrence has been described to be associated with the presence of coniferous trees. Indeed, all three species were not found at the pure beech sample sites, but occurred in at least one of three forest stands being characterized by the presence of spruces (table II.4). *Rhynchaenus fagi* (figure II.31), a coleopteran species specialized on feeding on beeches was not found at the pure spruce sites but within all three beech forest types. The larvae of the spruce web-spinning sawfly *Cephalcia avensis* (Hymenoptera:Pamphiliidae) (figure II.31) are exclusively feeding on the

needles of spruces. Between May and July active flying adults can be found in the habitat. Within our dataset the species was detected at the old beech, young beech and pure spruce sites during summer season but was absent from the pure beech sampling sites (table II.7).

Our definition of permanent residents and potential visitors requires the sampled species to occur comparatively frequently within one habitat. Unfortunately, many species are rare. They are not absent from a habitat because of a lack of specialization but rather because of a comparatively small population size. The beech-sawfly *Nematus fagi* (Hymenoptera: Tenthredinidae) (figure II.31) has only recently been described to be a character species of old beech forests (Kraus and Floren 2002). Long time no correlation between dominating tree species and occurence of *N. fagi* was observed, as the sawfly seemed to be not very frequently in forests. In contrast to that Kraus and Floren (2002) stated that *Nematus fagi* is by no means rare. The species has rather been overlooked because of its hidden life. They described *Nematus fagi* as a species which spend large parts of their life in the canopies of beech forests, where they are easily overseen. Our data underlines these findings as members of the species were exclusively found at the pure beech sampling sites, although they did not meet our definition of a permanent resident, which is possibly a result of the insufficient sampling strategy not encompassing the canopy of the forest. This result show that collected insects are primarily reflecting species community associated with the undershrub layer.

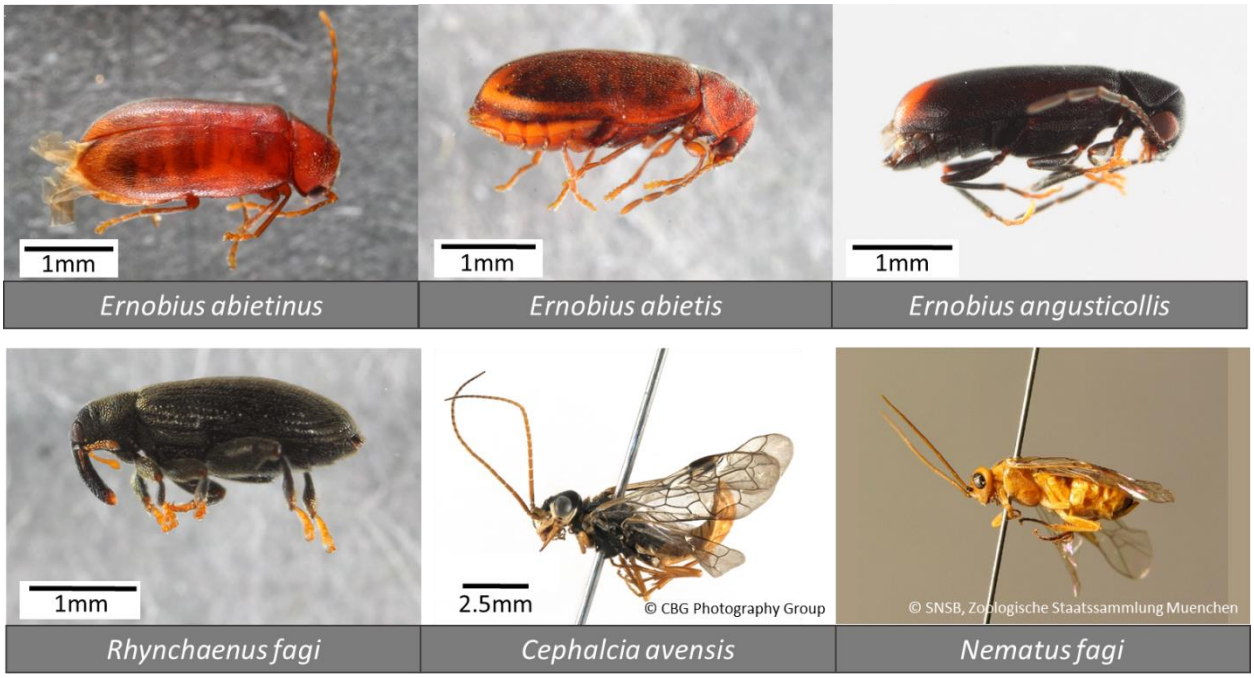


Figure II.31 Photos of voucher specimens: Occurrence of the three coleopeteran species *Ernobius abietinus, Ernobius abietis* and *Ernobius angusticollis* as well as the hymenopteran species *Cephalcia avensis* has been described to be associated with spruces, while *Rhynchaenus fagi* and *Nematus fagi* are associated with beech trees (if not other stated: © BY-SA 4.0: GBOL / Museum Koenig)

The overall number of permanent resident species was highest at the old beech stands (Pure beech: 194 species; Old beech: 233 species; Young beech: 183 species; Pure spruce 176 species). This is consistent with finding of former studies (Engel and Ammer 2001; Elmer et al. 2004). The mixed-species stands are more stratified and provide therefore more niches, which goes along with an increase in food resources, foraging niches and nest sites (Ammer et al. 2008). Pictures taken at the four sampling areas are showing changes in habitat complexity (Figure II.32).The stratification of the habitat is unquestionable an important factor influencing species diversity. As already discussed, the more diverse the different layer are the more ecological niches and food resources are provided.

Figure II.32 Photos of the four sampled forest types in summer season

Pollinators

As already discussed, especially several hymenopteran species are known to require a diverse herb and undershrub layer (Proesmans et al. 2019). Within the group of pollinators, bees (Hymenoptera: Apoidea) are the most specialized pollinators as almost all species feed exclusively on floral resources as both larvae and Imago (Kevan and Baker 1983; Proctor et al. 2012). Next to bees, hoverflies (Diptera: Syrphidae) are often mentioned as one of most

important groups of flower visitors (Larson et al. 2001). Within our dataset we found four bee species and 12 species belonging to the hoverflies. Out of the four Apidae species only one was assigned to the group of permanent residents, while the remaining three families seem to occur only occasionally within a habitat. Within the Syrphidae, nine species were permanent residents of one or more habitats. Like all other insect groups is the occurrence of pollinators directly linked to available food resources, which is highly driven by season and associated abiotic factors like light intensity and temperature.

Within the bees only the species *Bombus terrestris* was active during spring season. The species was found at the pure beech and young beech sampling sites but at both sites they occur as visitors. Out of the hoverflies, seven species already occurred in spring. Three of them were found at the pure beech sampling sites, four at the old beech and young beech sites respectively and two at the pure spruce sites. Out of them, the species *Syrphus torvus* occurred in all four above mentioned habitats. *S. torvus* is known to be able to travel long distances and has even been described to be a migratory species (Schmid 1999; Jensen 2001). Nevertheless, it was marked as permanent species for all four habitats in spring season.

In summer season six Syrphidae species were found at the pure beech sampling sites. Out of them, only two were permanent resident of one of at least one of the four sampled forest types. Between spring and summer season hoverfly diversity strongly increased in the two mixed forest stands. In both stands, six species were found. All of them were permanent residents of the old beech stands and four of the young beech stand. This is likely due to the fact that with increasing temperatures, growing season was induced. At the pure spruce sampling sites five species were found out of which only one was marked as permanent resident species. A single bees species was found in summer season. The visitor species *Bombus hortorum* was detected at the old beech and young beech sites. Only four species of the above mentioned pollinator groups were marked as permanent resident for the pure beech stand within the summer season. The low number of pollinators belonging to these two big groups can likely be a results of the reduced herb layer.

This is once more highlighting the importance of dominating tree species for the herb and understory layer and thereby on pollinator species composition.

Parasitoids

Some authors argue that plant diversity does not necessarily increase insect diversity (Haddad et al. 2009; Scherber et al. 2010). With an increase in trophic level and degree of omnivory, effect of plant diversity on insect richness weakens (Haddad et al. 2009; Scherber et al. 2010). Therefore, it is also interesting to have a closer look at co-occurrence of parasitoids and their hosts or of possible mutual relationships and predator-prey interactions. The major parasitoid groups identified in this study are the parasitoid wasps including the two hymenopteran families Ichneumonidae and Braconidae and the dipteran family Tachinidae. Five permanent resident species were found within the family Tachinidae. Within our dataset the species *Cyzenis jucunda* (figure II.33), belonging to the genus Cyzenis was found at the young beech, old beech and pure beech sampling sites in spring. The genus Cyzenis is known to parasite on lepidopteran larvae. Especially the species *Cyzenis albicans* (figure II.33) is known to have a host preference on the winter moth *Operophtera brumata* (Hassell 1968) (Lepidoptera: Geometridae), but has also been observed to parasite other members of the family Geometridae. Although the preferred host species of *C. jucunda* is still unknown (Herting 2017), former studies have shown that *C. jucunda* and *C. albicans* are closesly related (O'Hara and Cooper 1992), what possibly could point to a similar host preference. *C. jucunda* appeared as permanent resident at the pure beech stands in spring 2017 due to the time when the Geometridae member *Opheroptera fagata* (figure II.32) was present in the same habitat. Similar correlation were also be observed for further parasitoids. Nevertheless, the identification of parasitoid-host relationship is difficult. Firstly, within the group of parasitoids many different strategies has been evolved. Next to idiobiont parasitoids, which usually paralyze the host species, some parasitoids have evolved koinobiont

strategies, which still allows the host species to keep on moving, feeding and developing (Gullan and Cranston 2014). Furthermore, parasitoids can be ecto- and endoparasitoids. While the latter develop inside the host, ectoparasitoids are feeding on the host from the outside. While idiobiont parasitoids can be endo- and ectoparasitoids, koinbiont strategies are usually only observed within the group of endoprasitoids (Gullan and Cranston 2014). Hosts of koinbiont parasitoids are usually feeding in comparatively less sheltered locations, where they can easily be attacked by the parasitoids (Quicke 2015). Many of these host species usually pupate in more concealed and safe locations, which is also an advantage for the parasitoids as its transformation from larvae to imago is therefore happening in a sheltered location (Quicke 2015).

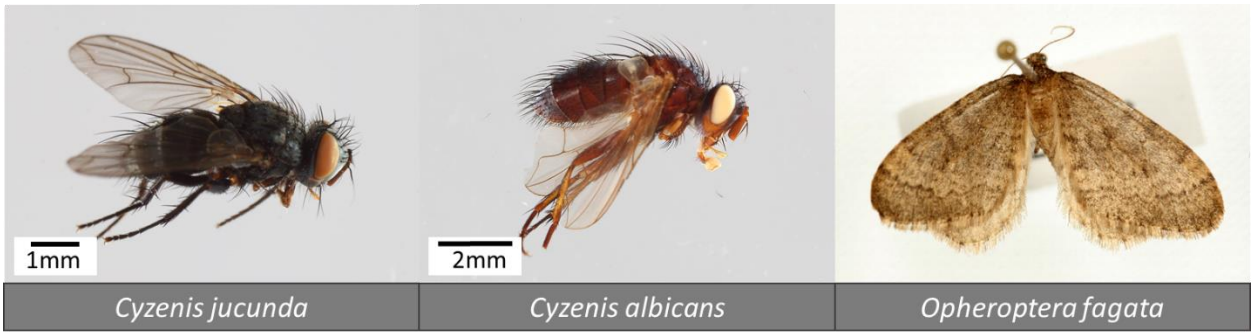


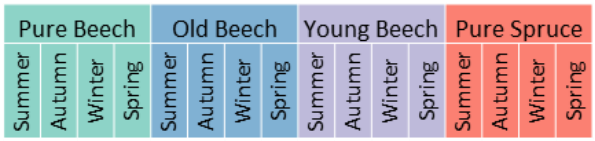
Figure II.33 Photos of voucher specimens of *Cyzenis jucunda* (Insecta: Diptera), *Cyzenis albicans* (Insecta: Diptera) and *Opheroptera fagata* (Insecta: Lepidoptera) (© BY-SA 4.0: GBOL / Museum Koenig)

These wide variety in parasitoid lifestyle makes the simultaneous assessment of host species and their parasitoids often very complicated. Time constraints are hampering the simultaneous assessment of parasitoids and their hosts. Parasitoids exploiting the larvae of their host, usually occur a bit later in the year after their host has already laid their eggs. Some lepidopterans even hibernate in the egg stage, which means that their Imagoes are usually active in autumn, while flying specimens of the associated parasitoid is usually found in spring. This is leading us to

another challenge when it comes to investigating parasitoid-host relationships. Not all parasitoids are attacking flying insects. Some species are specialized on exploiting spiders and ground living coleopterans, which won't be captured by Malaise traps.

Table II.3 Detected permanent resident species of less diverse orders

Less diverse orders



Mecoptera		
Family	Genus	Species
Panorpidae	Panorpa	Panorpa germanica
Neuroptera		
Family	Genus	Species
Coniopterygidae	Coniopteryx	Coniopteryx pygmaea
Coniopterygidae	Coniopteryx	Coniopteryx tineiformis
Hemerobiidae	Hemerobius	Hemerobius micans
Hemerobiidae	Hemerobius	Hemerobius pini
Psocoptera		
Family	Genus	Species
Caeciliusidae	Valenzuela	Valenzuela flavidus
Ectopsocidae	Ectopsocus	Ectopsocus briggsi
Elipsocidae	Elipsocus	Elipsocus moebiusi
Peripsocidae	Peripsocus	Peripsocus subfasciatus
Psocidae	Loensia	Loensia fasciata
Stenopsocidae	Graphopsocus	Graphopsocus cruciatus
Trichoptera		
Family	Genus	Species
Hydropsychidae	Hydropsyche	Hydropsyche siltalai

Table II.4 Detected permanent resident species of the order Coleoptera

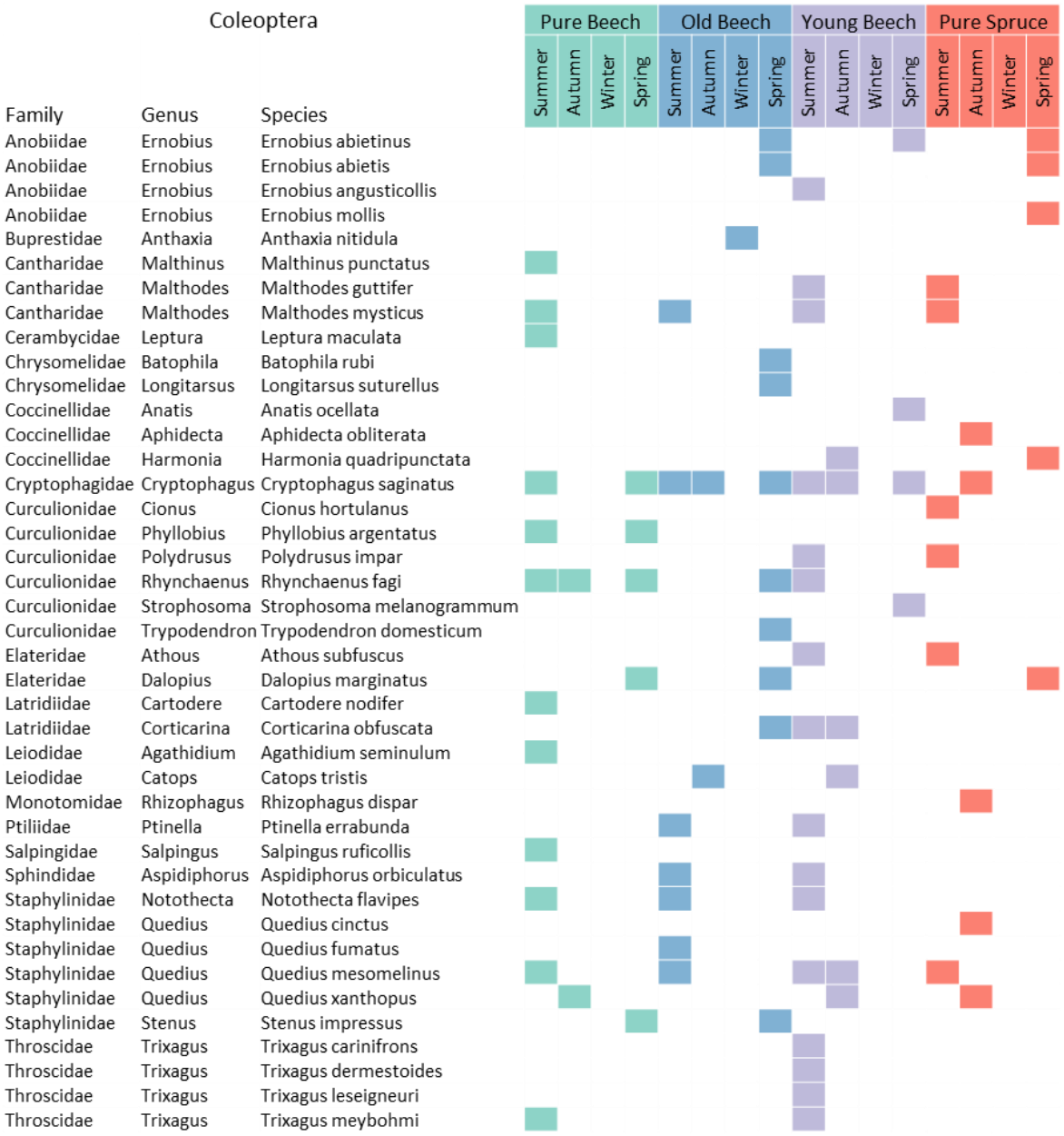


Table II.5a Detected permanent resident species of the order Diptera

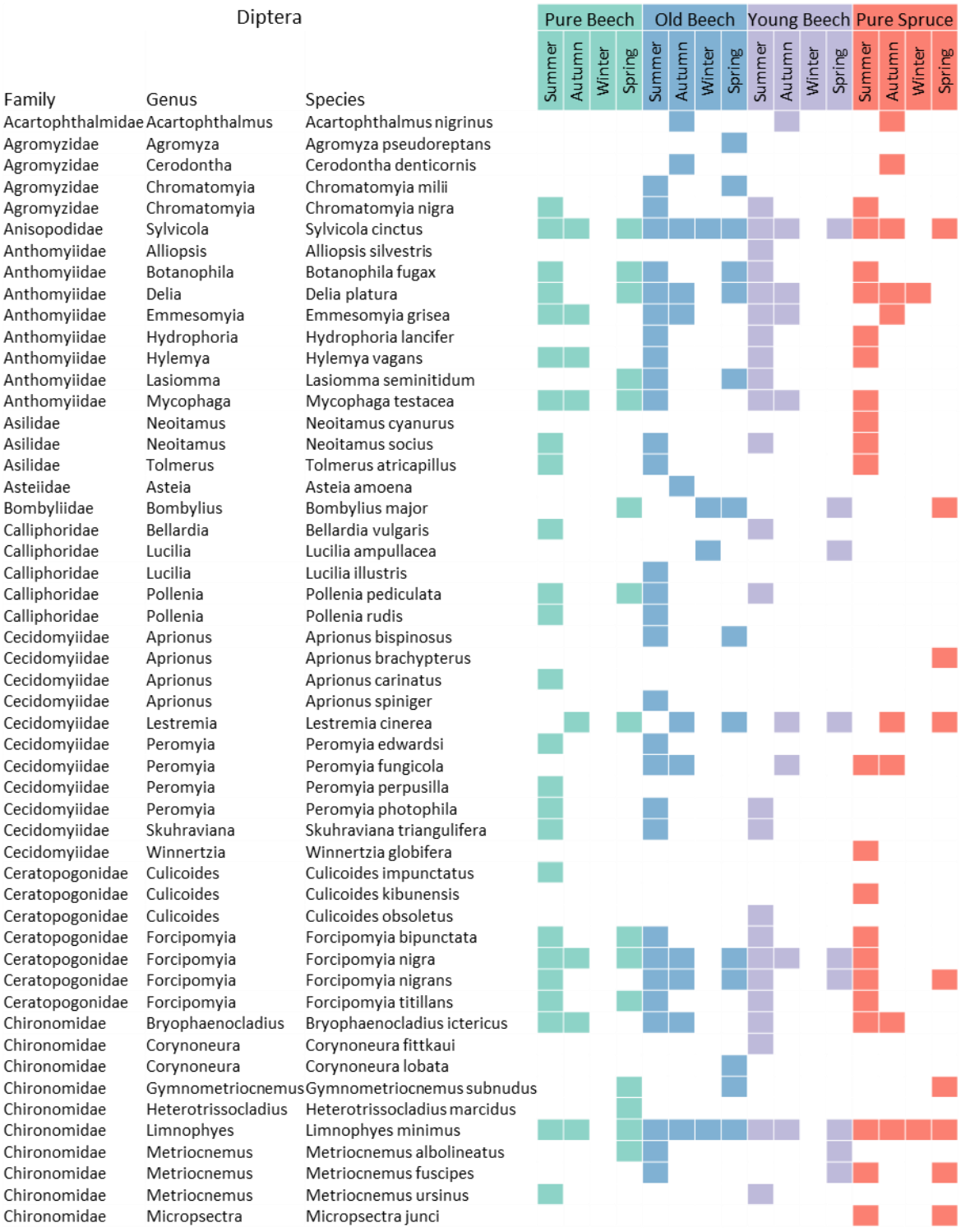


Table II.5b Detected permanent resident species of the order Diptera

Table II.5c Detected permanent resident species of the order Diptera

Table II.5d Detected permanent resident species of the order Diptera

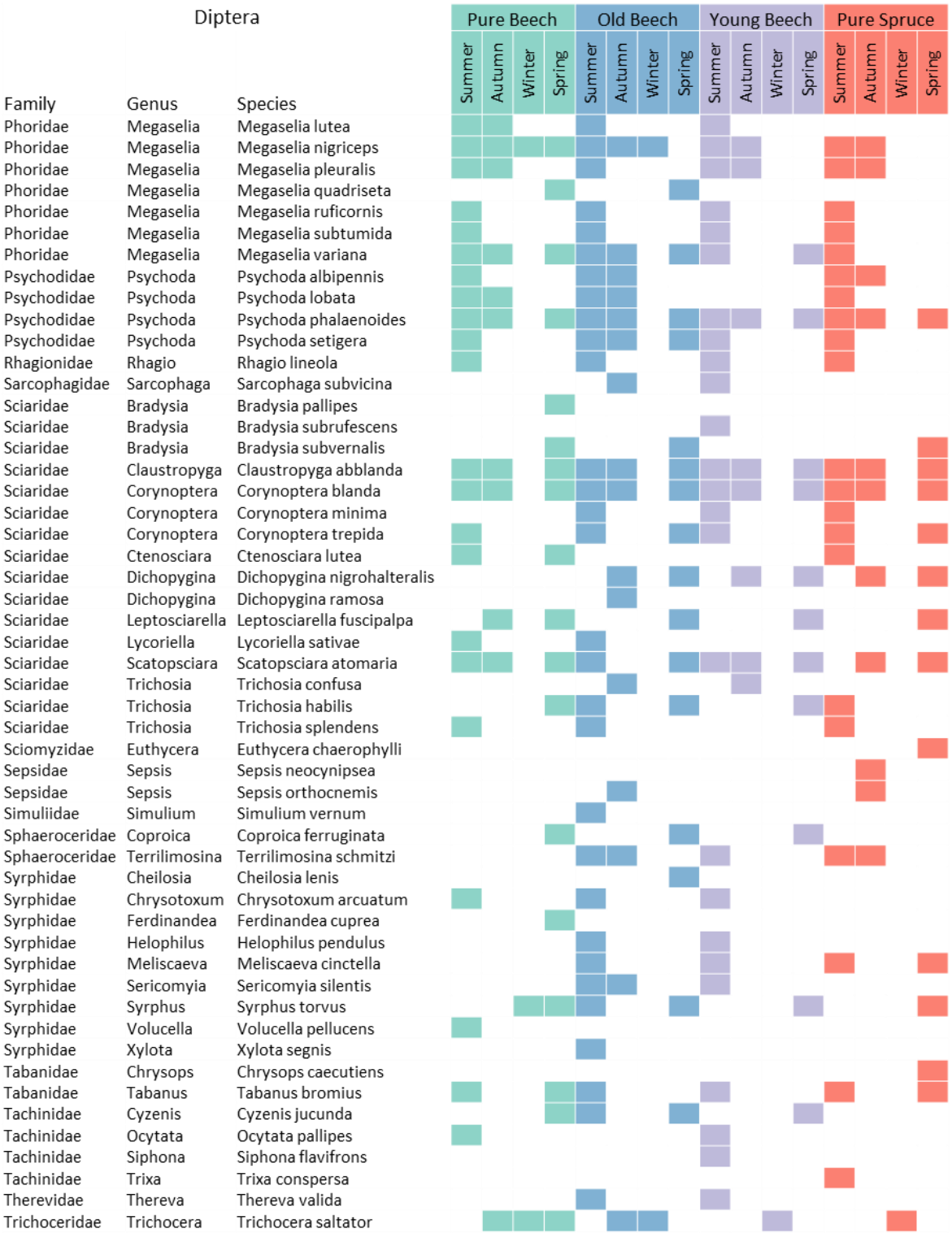


Table II.6 Detected permanent resident species of the order Hemiptera

Hemiptera

Family	Genus	Species
Acanthosomatidae	Elasmostethus	Elasmostethus minor
Anthocoridae	Anthocoris	Anthocoris confusus
Aphididae	Aphis	Aphis idaei
Aphididae	Drepanosiphum	Drepanosiphum platanoidis
Aphididae	Euceraphis	Euceraphis betulae
Aphididae	Euceraphis	Euceraphis punctipennis
Cicadellidae	Alnetoidia	Alnetoidia alneti
Cicadellidae	Balclutha	Balclutha punctata
Cicadellidae	Empoasca	Empoasca vitis
Cicadellidae	Errhomenus	Errhomenus brachypterus
Cicadellidae	Eupteryx	Eupteryx atropunctata
Cicadellidae	Typhlocyba	Typhlocyba quercus
Delphacidae	Javesella	Javesella pellucida
Issidae	Issus	Issus coleoptratus
Microphysidae	Loricula	Loricula elegantula
Miridae	Atractotomus	Atractotomus magnicornis
Miridae	Deraeocoris	Deraeocoris lutescens
Miridae	Monalocoris	Monalocoris filicis
Miridae	Pachytomella	Pachytomella parallela
Miridae	Parapsallus	Parapsallus vitellinus
Miridae	Phytocoris	Phytocoris longipennis
Miridae	Psallus	Psallus varians
Nabidae	Nabis	Nabis limbatus
Pentatomidae	Troilus	Troilus luridus
Psyllidae	Arytaina	Arytaina genistae
Pyrrhocoridae	Pyrrhocoris	Pyrrhocoris apterus
Rhyparochromidae	Gastrodes	Gastrodes abietum
Scutelleridae	Eurygaster	Eurygaster testudinaria
Triozidae	Trioza	Trioza remota
Triozidae	Trioza	Trioza urticae

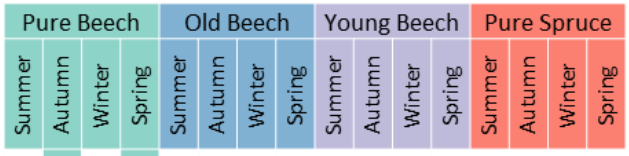


Table II.7 Detected permanent resident species of the order Hymenoptera

Hymenoptera

Family	Genus	Species
Apidae	Bombus	Bombus terrestris
Braconidae	Aphidius	Aphidius ervi
Braconidae	Aphidius	Aphidius rhopalosiphi
Braconidae	Microplitis	Microplitis mediator
Crabronidae	Trypoxylon	Trypoxylon clavicerum
Ichneumonidae	Allomacrus	Allomacrus arcticus
Ichneumonidae	Apechthis	Apechthis rufata
Ichneumonidae	Bathythrix	Bathythrix pellucidator
Ichneumonidae	Cratichneumon	Cratichneumon viator
Ichneumonidae	Cylloceria	Cylloceria melancholica
Ichneumonidae	Eusterinx	Eusterinx argutula
Ichneumonidae	Eusterinx	Eusterinx bispinosa
Ichneumonidae	Gelis	Gelis rufogaster
Ichneumonidae	Gnathochorisis	Gnathochorisis crassulus
Ichneumonidae	Ichneumon	Ichneumon gracilentus
Ichneumonidae	Misetus	Misetus oculatus
Ichneumonidae	Orthocentrus	Orthocentrus asper
Ichneumonidae	Orthocentrus	Orthocentrus spurius
Ichneumonidae	Picrostigeus	Picrostigeus recticauda
Ichneumonidae	Plectiscus	Plectiscus callidulus
Ichneumonidae	Plectiscus	Plectiscus impurator
Ichneumonidae	Polyaulon	Polyaulon paradoxus
Pamphiliidae	Cephalcia	Cephalcia arvensis
Tenthredinidae	Athalia	Athalia cordata
Tenthredinidae	Cladius	Cladius compressicornis
Tenthredinidae	Dolerus	Dolerus varispinus
Tenthredinidae	Empria	Empria tridens
Tenthredinidae	Euura	Euura vaga
Tenthredinidae	Pristiphora	Pristiphora cincta
Tenthredinidae	Pristiphora	Pristiphora nigriceps
Tenthredinidae	Pristiphora	Pristiphora pallidiventris
Vespidae	Dolichovespula	Dolichovespula saxonica
Vespidae	Symmorphus	Symmorphus gracilis
Vespidae	Vespa	Vespa crabro
Vespidae	Vespula	Vespula vulgaris

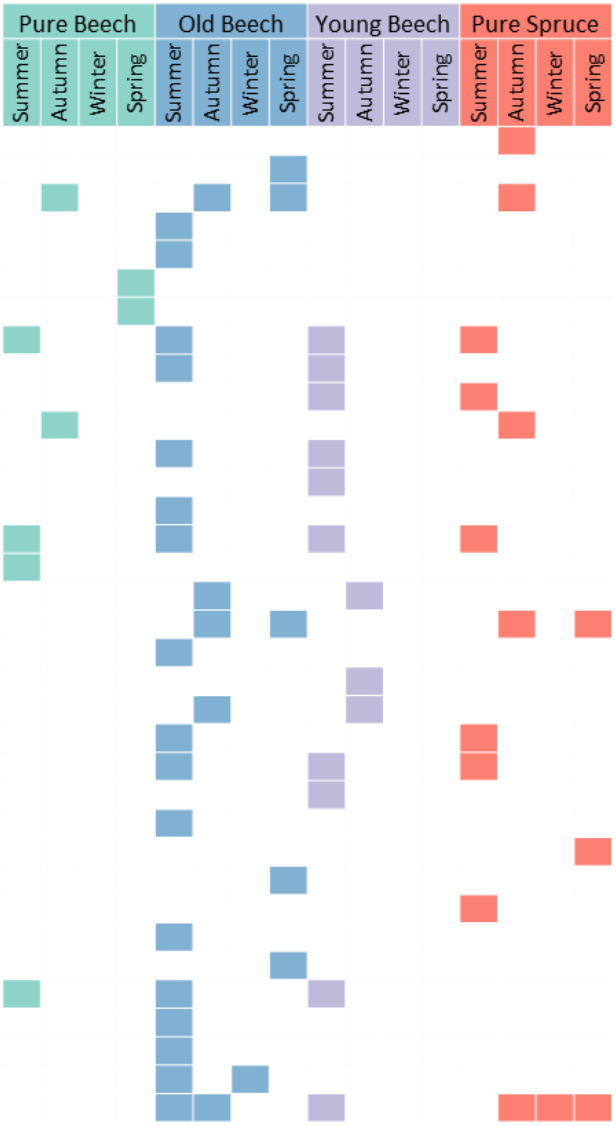
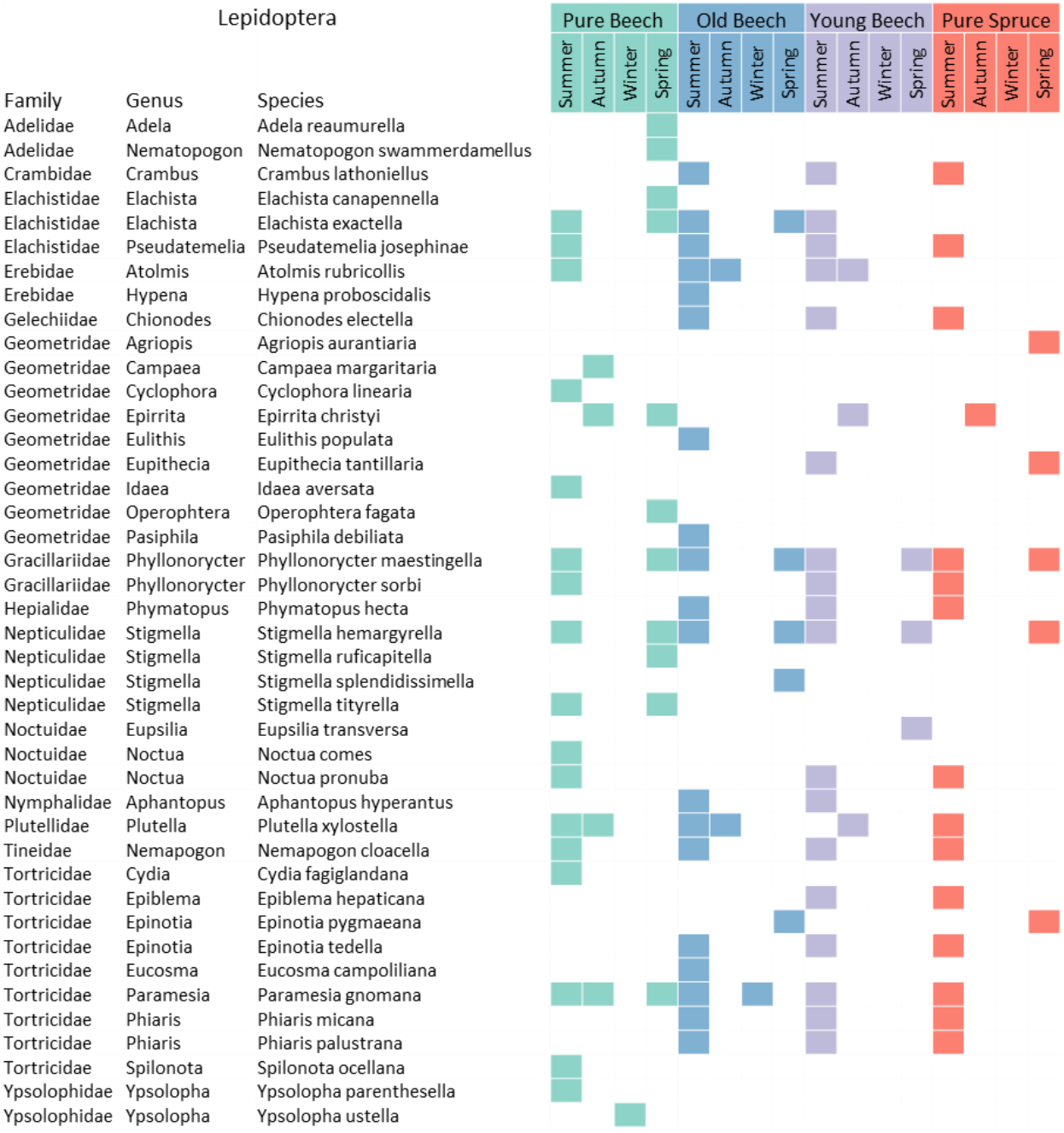


Table II.8 Detected permanent resident species of the order Lepidoptera



Influence of Methods on Study Outcome

Sampling Strategy

Regardless of targeted ecosystems, ecologists are facing several difficulties when assessing insect community composition. Because of insects small size, various life cylces which often goes along with seasonal variations in lifeforms, activity level, food preferences and the exceptional high number of species, the sampling strategy must be chosen with care. Additionally, next to the above mentioned obstacles, geographic circumstances and the local flora characterizing the study area can highly influence insect sampling success. While insect activity in grasslands and other treeless landscapes is to wide extent restricted to an area which can be easily accessed by ecologists, ecologists working in forests are overtowered by their study substrate (Leather 2008). Overall, assessing insect diversity associated with forests can be challenging as many more factors have to be taken into account. Former studies have shown that older forests potentially host a greater diversity. With increasing age, the canopies become structurally more complex, resulting in a greater variety of niches and microhabitats (Ishii et al. 2004). This often goes along with an increase in arthropod diversity (Jeffries et al. 2006) as the more niches are present the more requirements of species on a habitat are potentially matched. Like most sampling devices, Malaise traps are not equally efficient for all groups of insects. Malaise traps are passive sampling advices, which success rate is directly linked to species morphology, habitat type, and species activity level. Malaise traps were invented as flight interception traps, which main purpose is to capture insects moving with low distance to the ground. Therefore, the traps are particularly effective for capturing insects associated with the understory environment, on condition that they are actively flying through the environment (Leather 2008). Within our study, Malaise traps were exclusively placed on the forest floor. As a result, arthropod species associated with the canopies were not assessed, although it can be expected that insect diversity associated with the

forest canopy is exceptional high (Erwin 1983, 2001; Novotny and Basset 2005; Dial et al. 2006) and could provide valuable information about current status of the habitat. Former studies have shown that diversity and community composition of arachnids (Larrivée and Buddle 2009), hymenopterans (Sobek et al. 2009), coleopterans (Vance et al. 2003; Maguire et al. 2014), dipterans (Maguire et al. 2014) and lepidopterans (Schulze et al. 2001) are significantly differing between forest strata. It is likely that sampling of the canopies would had significantly influenced number of recovered species and thereby calculated degree of diversity of each of the four sampled forest types.

As flight interception traps are Malaise traps especially efficient when targeting flying insects like dipterans and hymenopterans, whereas non-flying insects are often missed. However, insects not being capable to flight might even be better indicator species for the health and current status of an ecosystem as they are more limited in their travel distances. Furthermore, our study set up does not allow for general statements about species presence and absence depending on season. Two weeks of sampling allow only for a short glimpse on existing local diversity within each season. Many highly adapted species occur only for a short time period like many species of mayflies and stoneflies (Dobrin and Giberson 2003). Even highly abundant species, which are capable to fly are often missed by Malaise traps because of insects complex life cycles and accompanying time constraints.

Time Constraints

Insects can roughly be divided into two groups: the Hemimetabola and the Holometabola. While the Hemimetabola are undergoing a gradual change from egg over nymph to imago, holometabolous insects are undergoing a complete metamorphosis including a pupal stage (Dettner and Peters 2011). After hibernating as larvae the sawfly *Athalia cordata* pupate and finally emerge as Imago in late April. They soon lay they eggs underneath the leaves of various

plants belonging to the genus Ribes. During May and June the eggs undergo a full development until the second generation of imagos hatch in late June. Although the species is present in the habitat all around the year, the forms capable to fly are usually observed from late April till September. Within this study *A. cordata* was detected during summer season at the old beech and young beech sample sites, but did only match the requirements for being a permanent resident in the latter (table II.7). It is likely that sampling in spring (12.04 – 26.04.2017) took place shortly before the emergence of the adult forms *of A. cordata*, while at the time of sampling in autumn (13.10 – 27.10.2016) the adult forms had already passed away. These time constraints can influence study outcome even more dramatically, when only some members of a species are capable to fly in a short time frame. A good example is the hymenoptera family Formicidae. As social insects, each ant of an ant colony has a certain task, which is important for the maintenance of the colony. To be best adapted to their tasks several morphological adaptation can be found within one ant species. This can make the proof of existence of a certain ant species within a habitat challenging, especially if only Malaise traps are used as sampling devices. In our study only three species of the family Formicidae were detected. Formicidae are known to be highly abundant and as ecosystem engineers their presence is even crucial for the maintenance of the health of the habitat. While in summer, the species *Myrmica ruginodis* was found at the Pure beech and Old beech sampling sites, in autumn it was only recovered from the pure beech sites. *M. ruginodis* is known to have a short flight season between mid and late summer (Noordijk et al. 2008). For the genus Myrmica is has been described, that mating takes place at aggregation sites, meaning that male dominated swarms occur in high abundances at places (Noordijk et al. 2008) which are often characterized by certain structures or abiotic parameters like a high light intensity (Kannonski 1963). Members of the genus Myrmica are considered to travel only short distances (Elmes 1991; Hicks 2012). Therefore, is can be assumed that *Mrymica ruginodis* is present at the old beech and pure beech stands throughout the year but did not appear at any time as permanent resident species. We conclude, that because of the short time frame of flight activity the absence of flying forms during winter and spring season proves that a Malaise trap is

not a suitable sampling device for the assessments of organisms which are only actively flying for a short period of time when sampling is limited to a short period of time.

Furthermore, some insect orders encompass flightless as well as flightworthy forms. The highly diverse order Coleoptera consists of 211 families (Bouchard et al. 2011). One of the feature which distinguish members of the order Coleoptera from most other insects are the development of elytra, which are hard wing-cases formed out of the front wing pair (Honomichl et al. 1998). Depending on family and genus the degree of hardness of the elytra is differing. In general, to be able to fly elytra must be raised enabling movement of hind wings. Some species has lost the ability to fly like some members of the family Carabidae (Honomichl et al. 1998) what makes a proof of existence of these species on the basis of Malaise traps catchments unreliable. For middle Europe a total of 800 species belonging to the family Carabidae has been described (de Jong et al. 2014). The carabids are a highly diverse group, which belong to the most important insect predators (Lövei and Sunderland 1996; Riley and Browne 2011) within the forest ecosystem. In our study, only one species of the family Carabidae - *Harpalus rufipes* was identified. In contrast to many other carabid species is *Harpalus rufipes* still able to fly, although this behavior is only rarely observed (Alford 2014). Nevertheless, *H. rufipes* was detected at the pure beech, young beech and pure spruce sites at different times of the year. However, the species does not match the requirements of being a permanent resident of a certain forest type at any time of the year. Nevertheless, *H. rufipes* is known to be a cultural successor and not very exacting. It can be expected that the species is present within all three forest types thoughout the year, but is because of their limited fly activity only occasionally caught.

Next to seasonal variations, also daily variations in number of caught insects have been observed (Matthews and Matthews 1983). Abiotic factors are having a major influence on species activity status and thereby on the effectiveness of Malaise traps. It has already been described that during long lasting periods which are characterized by a climate hampering insect activity, the effectiveness of Malaise traps is strongly reduced (Matthews and Matthews 1983).

Metabarcoding for Assessing Arthropod Communities

Metabarcoding is a powerful tool to circumvent the taxonomic impediment. Although, metabarcoding has unquestionable many advances over the morphological identification of species, like cost efficiency, timely results and the lower risk of misidentifying cryptic species or species expressing a broad range of phenotypic plasticity, it also has some drawbacks. When working with Malaise traps, species are accumulated in so called bulk samples. DNA isolated from bulk samples is called community DNA (cDNA) (Deiner et al. 2017). cDNA is a mixture of the DNA of all organisms contained in the sample. Many authors assume that the analysis of cDNA mirrors the accurate species composition which was present at the specified time at the sample site (Deiner et al. 2017). However, this is only partly true. Many insects are known to travel long distances to avoid unfavourable abiotic conditions. Within the order Lepidoptera the monarch butterfly *Danaus plexippus* migrates from the west coast of California to his overwintering sites in central Mexico and back (Brower 1995, 1996; Knight et al. 1999; Rogg et al. 1999). Within Europe the butterfly *Vanessa cardui* migrates each year from Scandinavia to West Africa (Stefanescu et al. 2013). Next to the butterflies many other insects has been described to travel long distances like hoverflies (Syrphidae) (Schmid 1999; Jensen 2001) and beetles (Hodek et al. 1993). On their way from one site to another they will get in contact with the biotic environment. As a results they can easily be infested with parasites like mites (Arachnida: Acari). Next to DNA of the host animal, also the DNA of the parasite is now part of the community DNA extracted from the bulk sample, although the parasite does not naturally occur at the sample site. Furthermore, community DNA can also be infested with eDNA originating from a total different habitat. While migrating from one habitat to another highly mobile insects are easily becoming transport systems for eDNA. eDNA can originate from faeces, mucus and many other sources, which may serve as food for the migrating insects. A recent study has shown that eDNA from pollinators get accumulated in the flower buds (Thomsen and Sigsgaard 2019). Migrating

hoverflies, feeding on such a flower will potentially uptake spurs of eDNA originating from prior flower visitor species.

Like the DNA originating from parasites, also the eDNA attached to the captured insect will become part of the community DNA. Although we did not destroy the specimens, we cannot exclude the possibility that DNA originating from the gut content of the caught insects were released into the bulk sample. Therefore, the above mentioned definition about community DNA of Deiner et. al (2017) should be revised. In views to our study the introduction of eDNA into the bulk sample could have let to wrong positive results concerning species activity level. Many insects are hibernating on the ground underneath the foliage, or under the bark of trees where they are vulnerable to attacks from predators being active during winter season. Furthermore, depending on climate and abiotic conditions eDNA can persist a long time in the environment (Harrison et al. 2019). After getting in contact with it, winter active species like the winter snow scorpion fly could become potentially vectors transporting eDNA into the Malaise traps, what will lead to false positive detection of species.

Furthermore, metabarcoding of bulk samples is still suffering from primer bias. Although many universal primer pairs has been developed a recent study have shown that none of the 21 most frequently used primer pairs had a 100% detection rate within a mock community (Elbrecht et al. 2019) . Nevertheless, Elbrecht et al (2019) showed that average detection rate for taxa was high with 91%. Depending on sequencing depth the primer pair miCOIIntF/Fol-degen-rev on which this study is partly based, had a recovery rate of up to 95.7 %.

Another crucial drawback of metabarcoding is the missing ability to quantify species relative abundances (Fonseca 2018). Template secondary structure (Fonseca et al. 2012), annealing temperature (Elbrecht et al. 2019), primer-template mismatches, copy number variation of the target loci, amplicon size and taxon-specific amplification bias (Elbrecht and Leese 2015; Krehenwinkel et al. 2017) can alter abundance estimates. Species relative abundances are crucial for nature protection strategies. Next, to scientific interest in numbers, numbers are a powerful

tool to raise awareness for insect decline. The study of Hallmann et al (2017), which stated that in the last 27 years insect biomass were suffering from a decrease of 75% arose world wide interest and politics have become aware of the urgent need of action. Since than, several research initiatives have been funded.

Although this study cannot claim to have assessed the full existing diversity within four forest types, a good overview over species diversity being associated with the undershrub along a forest conversion gradient have been acquired. This study has proven several hypothesis concerning changes of faunal diversity over a forest conversion gradient which were long time based on a limited number of insect groups. We showed that insect diversity patterns of the undershrub are changing with seasons. The age of the mixed forests play thereby an important role. The old beech forests were more stratified, which goes along with an increase in number of niches. Nevertheless, the conservational value of a forest is not measured by the number of species which it harbors. Moreover, the occurrence of native, endemic and endangered species should be taken into account. For conservation purposes a less diverse natural forests can be more valuable for conservation purposes than an artificial diverse monoculture stand (Ammer et al. 2008). Overall, non endangered species was recovered with this study, nor did we find an impact of forest type on species abundance as metabarcoding did not allow for this sample design. Nevertheless, the maintenance and protection of the native insect fauna is only possible when restoring the native flora. This study clearly demonstrate that the fauna is directly linked to the local flora. Moreover, we observed that calculated jaccard similarity indices were increasing with ongoing process of renaturation indicating that species communities are slowly changing from the ones adapted to pure spruce stands to pure beech stands (Figure 6). When monitoring the process of ongoing renaturation measurements sampling should be performed over a longer period of time, otherwise, many species which may serve as indicator species are likely missed.

Conclusion Part II

As hypothesized current status of renaturation measures was mirrored in changes in insect species composition. However, the complex interplay of the abiotic and biotic environment consisting of flora and fauna is making it hard to identify the main drivers of these changes. Although, many experiments and studies have investigated reciprocal relationships between all kinds of influential factors, it would still be presumptuous to make predictions about community fade on the basis of a single factor e.g. dominating tree species.

While a decrease in biodiversity was recorded for only recently underplanted spruce forests, we showed that with increasing age the establishment of mixed forests led to an increase in species diversity. However, it should be kept in mind, that the conservational value of a forest is not measured by the number of species which it harbors. Moreover, the occurrence of native, endemic and endangered species should be taken into account. For conservation purposes a less diverse natural forests can be more valuable for conservation purposes than an artificial diverse monoculture stand (Ammer et al. 2008). Out of the 367 recovered permanent resident species, which were discovered within this study (Table 1-6) the low number of 57 had an entry in the register of the 'Rote Liste Deutschland'. None of these 57 species were listed as endangered or threatened by extinction. Therefore, we were not able to show the value of the renaturation measures for the protection of endangered species.

Nevertheless, we were able to show that number of beech specialists were more likely to occur in beech forests, while species adapted to spruces were only present within these habitats. Therefore, we conclude that the transformation of the non-native spruce forests to deciduous

forests, will likely go along with the disappearance of several species, for the benefit of the establishment of insect species community which is more similar to the one formerly being associated with the deciduous forests characterizing the Eifel before the establishment of the commercial wood industry.

Literature

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

Alford D.V. 2014. Pests of fruit crops: a colour handbook. CRC press.

Aljanabi S.M., Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research. 25:4692–4693.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of molecular biology. 215:403–410.

Ammer C., Bickel E., Kölling C. 2008. Converting Norway spruce stands with beech-a review of arguments and techniques. Austrian Journal of Forest Science. 125:3–26.

Ammer U., Engel K., Förster B., Goßner M., Kölbel M., Leitl R., Simon U., Simon U., Utschick H. 2002. Vergleichende waldökologische Untersuchungen in Naturwaldreservaten (ungenutzten Wäldern) und Wirtschaftswäldern unterschiedlicher Naturnähe (unter Einbeziehung der Douglasie) in Mittelschwaben. Abschlussbericht des BMBF-Forschungsvorhabens A. 339735.

Andújar C., Arribas P., Yu D.W., Vogler A.P., Emerson B.C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular ecology. 27:3968–3975.

Astrin J.J., Stüben P.E. 2008. Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). Invertebrate Systematics. 22:503–522.

Augusto L., Dupouey J.-L., Ranger J. 2003. Effects of tree species on understory vegetation and environmental conditions in temperate forests. Annals of Forest Science. 60:823–831.

Barbier S., Gosselin F., Balandier P. 2008. Influence of tree species on understory vegetation diversity and mechanisms involved—a critical review for temperate and boreal forests. Forest ecology and management. 254:1–15.

Barkman J. 1992. Canopies and microclimate of tree species mixtures. Special publications... of the British Ecological Society.

Barrett S.C., Helenurm K. 1987. The reproductive biology of boreal forest herbs. I. Breeding systems and pollination. Canadian Journal of Botany. 65:2036–2046.

Baselga A., Orme C.D.L. 2012. betapart: an R package for the study of beta diversity. Methods in ecology and evolution. 3:808–812.

Beng K.C., Tomlinson K.W., Shen X.H., Surget-Groba Y., Hughes A.C., Corlett R.T., Slik J.W.F. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. Scientific Reports. 6:24965.

Binot-Hafke M., Haupt H., Balzer S., Becker N., Gruttke H., Hofbauer N., Ludwig G., Matzke-Hajek G., Strauch M. 2011. Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands - Band 3: Wirbellose Tiere: Teil 1. Bundesamt für Naturschutz.

Block W. 1990. Cold tolerance of insects and other arthropods. Philosophical Transactions of the Royal Society of London. B, Biological Sciences. 326:613–633.

Bos M.M., Höhn P., Saleh S., Büche B., Buchori D., Steffan-Dewenter I., Tscharntke T. 2007. Insect diversity responses to forest conversion and agroforestry management. Stability of Tropical Rainforest Margins. Springer. p. 277–294.

Bouchard P., Bousquet Y., Davies A.E., Alonso-Zarazaga M.A., Lawrence J.F., Lyal C.H., Newton A.F., Reid C.A., Schmitt M., Ślipiński S.A. 2011. Family-group names in Coleoptera (Insecta). ZooKeys. 88:1.

Bourlat S.J., Haenel Q., Finnman J., Leray M. 2016. Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method. Marine Genomics. Springer. p. 197–207.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

Braukmann T.W., Ivanova N.V., Prosser S.W., Elbrecht V., Steinke D., Ratnasingham S., Jeremy R. deWaard, Sones J.E., Zakhariv E.V., Hebert P.D. 2018. Revealing the Complexities of Metabarcoding with a Diverse Arthropod Mock Community. bioRxiv.:433607.

Brower L. 1996. Monarch butterfly orientation: missing pieces of a magnificent puzzle. Journal of Experimental Biology. 199:93–103.

Brower L.P. 1995. Understanding and misunderstanding the migration of the monarch butterfly(Nymphalidae) in North America: 1857-1995. Journal of the Lepidopterists Society. 49:304–385.

Budde S., Schmidt W., Weckesser M. 2011. Impact of the admixture of European beech (Fagus sylvatica L.) on plant species diversity and naturalness of conifer stands in Lower Saxony. Waldökologie, Landschaftsforschung und Naturschutz. 11:49–61.

Campbell J.W., Hanula J. 2007. Efficiency of Malaise traps and colored pan traps for collecting flower visiting insects from three forested ecosystems. Journal of Insect Conservation. 11:399–408.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7:335–336.

Chapin III F.S., Matson P.A., Vitousek P. 2011. Principles of terrestrial ecosystem ecology. Springer Science & Business Media.

Chen H., Boutros P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics. 12:35.

Clarke L.J., Beard J.M., Swadling K.M., Deagle B.E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 7:873–883.

Clarke L.J., Soubrier J., Weyrich L.S., Cooper A. 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular ecology resources. 14:1160–1170.

Conway J.R., Lex A., Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940.

Corse E., Tougard C., Archambaud-Suard G., Agnèse J., Messu Mandeng F.D., Bilong Bilong C.F., Duneau D., Zinger L., Chappaz R., Xu C.C. 2019. One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. Ecology and evolution. 9:4603–4620.

Courtin G., Shorthouse J., West R. 1984. Energy relations of the snow scorpionfly Boreus brumalis (Mecoptera) on the surface of the snow. Oikos. 43:241–245.

Cowart D.A., Pinheiro M., Mouchel O., Maguer M., Grall J., Miné J., Arnaud-Haond S. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one. 10:e0117562.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W., Power D.M., Packer M., Carvalho G.R., Blaxter M.L., Lambshead P.J.D., Thomas W.K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology. 19:4–20.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F., Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology letters. 10:20140562.

Deiner K., Bik H.M., Mächler E., Seymour M., Lacoursière-Roussel A., Altermatt F., Creer S., Bista I., Lodge D.M., De Vere N. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular ecology. 26:5872–5895.

Dettner K., Peters W. 2011. Lehrbuch der Entomologie. Springer-Verlag.

Dial R.J., Ellwood M.D., Turner E.C., Foster W.A. 2006. Arthropod abundance, canopy structure, and microclimate in a Bornean lowland tropical Rain forest 1. Biotropica. 38:643–652.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 14:927–930.

Dobrin M., Giberson D.J. 2003. Life history and production of mayflies, stoneflies, and caddisflies (Ephemeroptera, Plecoptera, and Trichoptera) in a spring-fed stream in Prince Edward Island, Canada: evidence for population asynchrony in spring habitats? Canadian Journal of Zoology. 81:1083–1095.

Dopheide A., Xie D., Buckley T.R., Drummond A.J., Newcomb R.D. 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution. 10:120–133.

Drummond A.J., Newcomb R.D., Buckley T.R., Xie D., Dopheide A., Potter B.C., Heled J., Ross H.A., Tooman L., Grosser S. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. GigaScience. 4:s13742-015.

Ekschmitt K., Stierhof T., Dauber J., Kreimes K., Wolters V. 2003. On the quality of soil biodiversity indicators: abiotic and biotic parameters as predictors of soil faunal richness at different spatial scales. Agriculture, ecosystems & environment. 98:273–283.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Elbrecht V., Hebert P.D., Steinke D. 2018. Slippage of degenerate primers can cause variation in amplicon length. Scientific reports. 8:10999.

Elbrecht V., Leese F. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PloS one. 10:e0130324.

Elbrecht V., Leese F. 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science. 5:11.

Elbrecht V., Peinert B., Leese F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7:6918–6926.

Elbrecht V., Taberlet P., Dejean T., Valentini A., Usseglio-Polatera P., Beisel J.-N., Coissac E., Boyer F., Leese F. 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ. 4:e1966.

Elmer M., La France M., Förster G., Roth M. 2004. Changes in the decomposer community when converting spruce monocultures to mixed spruce/beech stands. Plant and soil. 264:97–109.

Elmes G. 1991. Mating strategy and isolation between the two forms, macrogyna and microgyna, of Myrmica ruginodis (Hym. Formicidae). Ecological Entomology. 16:411–423.

Engel K., Ammer U. 2001. Analyse und Bewertung von Umbaumaßnahmen in Fichtenreinbeständen anhand ökologischer Gilden der Wirbellosen-Fauna. Forstwissenschaftliches Centralblatt vereinigt mit Tharandter forstliches Jahrbuch. 120:242–255.

Erwin T.L. 1983. Tropical forest canopies: the last biotic frontier. Bulletin of the ESA. 29:14–20.

Erwin T.L. 2001. Forest canopies, animal diversity. Encyclopedia of Biodiversity 3.

Feeny P. 1970. Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. Ecology. 51:565–581.

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 3:294–299.

Fonseca V.G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular ecology resources. 18:923–926.

Fonseca V.G., Carvalho G.R., Sung W., Johnson H.F., Power D.M., Neill S.P., Packer M., Blaxter M.L., Lambshead P.J.D., Thomas W.K., Creer S. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications. 1:98.

Fonseca V.G., Lallias D. 2016. Metabarcoding marine sediments: preparation of amplicon libraries. Marine Genomics. Springer. p. 183–196.

Fonseca V.G., Nichols B., Lallias D., Quince C., Carvalho G.R., Power D.M., Creer S. 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. Nucleic Acids Research. 40:e66–e66.

Frady C., Johnson S., Li J. 2007. Stream macroinvertebrate community responses as legacies of forest harvest at the HJ Andrews Experimental Forest, Oregon. Forest Science. 53:281–293.

Fritz P. 2006. Ökologischer Waldumbau in Deutschland–Fragen, Antworten, Perspektiven. oekom, München. 153.

Geiger M.F., Moriniere J., Hausmann A., Haszprunar G., Wägele W., Hebert P.D., Rulik B. 2016. Testing the Global Malaise Trap Program–How well does the current barcode reference library identify flying insects in Germany? Biodiversity data journal. 4: e10671

German Barcode of Life Consortium, Wägele W., Haszprunar G., Eder J., Xylander W., Borsch T., Quandt D., Grobe P., Pietsch S., Geiger M.F., Astrin J.J., Rulik B., Hausmann A., Morinière J., Holstein J., Krogmann L., Monje C., Traunsprunger W., Hohberg K., Lehmitz R., Müller K., Nebel M., Hand R. 2011. GBOL Webportal. .

Gibson J., Shokralla S., Porter T.M., King I., van Konynenburg S., Janzen D.H., Hallwachs W., Hajibabaei M. 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proceedings of the National Academy of Sciences. 111:8007–8012.

Gill H.K., Goyal G., Chahil G. 2017. Insect diapause: a review. J. Agric. Sci. Technol. 7:454–473.

Glime J.M. 2006. Terrestrial Insects: Holometabola - Mecoptera. Bryophyte ecology. Michigan Technological University. p 1-26

Gossner M.M. 2009. Light intensity affects spatial distribution of Heteroptera in deciduous forests. European Journal of Entomology. 106:241–252.

Gruttke H., Haupt H., Balzer S., Binot-Hafke M., Ludwig G., Matzke-Hajek G., Ries M. 2016. Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands - Band 4: Wirbellose Tiere: Teil 2. Bundesamt für Naturschutz.

Gullan P.J., Cranston P.S. 2014. The insects: an outline of entomology. John Wiley & Sons.

Haddad N.M., Crutsinger G.M., Gross K., Haarstad J., Knops J.M., Tilman D. 2009. Plant species loss decreases arthropod diversity and shifts trophic structure. Ecology Letters. 12:1029–1039.

Hajibabaei M., Shokralla S., Zhou X., Singer G.A., Baird D.J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one. 6:e17497.

Hajibabaei M., Spall J.L., Shokralla S., van Konynenburg S. 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC ecology. 12:28.

Hallmann C.A., Sorg M., Jongejans E., Siepel H., Hofland N., Schwan H., Stenmans W., Müller A., Sumser H., Hörren T., Goulson D., Kroon H. de. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLOS ONE. 12:e0185809.

Harrison J.B., Sunday J.M., Rogers S.M. 2019. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B. 286:20191409.

Hassell M. 1968. The behavioural response of a tachinid fly (Cyzenis albicans (Fall.)) to its host, the winter moth (Operophtera brumata (L.)). The Journal of Animal Ecology. 37:627–639.

Haupt H., Ludwig G., Gruttke H., Binot-Hafke M., Otto C., Pauly A. 2009. Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands. Bundesamt für Naturschutz.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:313–321.

Heinrich B. 2013. The hot-blooded insects: strategies and mechanisms of thermoregulation. Springer Science & Business Media.

Heller K., Rulik B. 2016. Ctenosciara alexanderkoenigi sp. n.(Diptera: Sciaridae), an exotic invader in Germany? Biodiversity Data Journal.

Herting B. 2017. A critical revision of host records of Palearctic Tachinidae (Diptera) until 19371. Integrative Systematics: Stuttgart Contributions to Natural History. 10:41–174.

Hicks B.J. 2012. How does Myrmica rubra (Hymenoptera: Formicidae) disperse in its invasive range? Record of male-only swarming flights from Newfoundland. Myrmecological News. 16:31–34.

Hill M. 1979. The development of a flora in even-aged plantations. The Ecology of Even-Aged Forest Plantations. Institute of Terrestrial Ecology, Cambridge. p. 175–192.

Hodek I., Iperti G., Hodkova M. 1993. Long-distance flights in Coccinellidae (Coleoptera). European Journal of Entomology. 90:403–414.

Homburg K., Drees C., Boutaud E., Nolte D., Schuett W., Zumstein P., von Ruschkowski E., Assmann T. 2019. Where have all the beetles gone? Long-term study reveals carabid species decline in a nature reserve in Northern Germany. Insect Conservation and Diversity. 12:268–277.

Honomichl K., Bellmann H., Jacobs W., Renner M. 1998. Biologie und Ökologie der Insekten. Fischer, Stuttgart, Jena, Lübeck.

Horton D.J., Kershner M.W., Blackwood C.B. 2017. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. European journal of soil biology. 80:43–48.

Ishii H.T., Tanabe S., Hiura T. 2004. Exploring the relationships among canopy structure, stand productivity, and biodiversity of temperate forest ecosystems. Forest Science. 50:342–355.

Jeffries J.M., Marquis R.J., Forkner R.E. 2006. Forest age influences oak insect herbivore community structure, richness, and density. Ecological Applications. 16:901–912.

Jennings S., Brown N., Sheil D. 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry: An International Journal of Forest Research. 72:59–74.

Jensen J.-K. 2001. An invasion of migrating insects (Syrphidae and Lepidoptera) on the Faroe Islands in September 2000. Norwegian Journal of Entomology. 48:263–268.

de Jong Y., Verbeek M., Michelsen V., de Place Bjørn P., Los W., Steeman F., Bailly N., Basire C., Chylarecki P., Stloukal E. 2014. Fauna Europaea–all European animal species on the web. Biodiversity data journal.

Juillet J. 1963. A comparison of four types of traps used for capturing flying insects. Canadian Journal of Zoology. 41:219–223.

Kannonski P.H. 1963. The flight activities of formicine ants. Symposia Genetica et Biologica Italica. 12:74–102.

Kaunisto K.M., Roslin T., Sääksjärvi I.E., Vesterinen E.J. 2017. Pellets of proof: First glimpse of the dietary composition of adult odonates as revealed by metabarcoding of feces. Ecology and evolution. 7:8588–8598.

Kevan P., Baker H. 1983. Insects as flower visitors and pollinators. Annual review of entomology. 28:407–453.

Knapp H., Jeschke L. 1991. Naturwaldreservate und Naturwaldforschung in den ostdeutschen Bundesländern. Schriftenreihe für Vegetationskunde. 21:21–54.

Knight A., BROWER L.P., WILLIAMS E.H. 1999. Spring remigration of the monarch butterfly, Danaus plexippus (Lepidoptera: Nymphalidae) in north-central Florida: estimating population parameters using mark-recapture. Biological Journal of the Linnean Society. 68:531–556.

Konôpka B., Pajtík J., Noguchi K., Lukac M. 2013. Replacing Norway spruce with European beech: A comparison of biomass and net primary production patterns in young stands. Forest Ecology and Management. 302:185–192.

Koštál V. 2006. Eco-physiological phases of insect diapause. Journal of insect physiology. 52:113–127.

Kraus M., Floren A. 2002. Pflanzenwespen (Hymenoptera, Symphyta) und Stechimmen (Chrysididae, Pompilidae, Sphecidae) aus Baumkronenbeneblungen (Fogging) von Eichen und Rotbuchen in Bayern (Unterfranken), Thüringen (Hainich), Slovenien und Rumänien. Galathea Supplement. 11:93–102.

Krehenwinkel H., Wolf M., Lim J.Y., Rominger A.J., Simison W.B., Gillespie R.G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports. 7:17668.

Larrivée M., Buddle C.M. 2009. Diversity of canopy and understorey spiders in north-temperate hardwood forests. Agricultural and Forest Entomology. 11:225–237.

Larson B., Kevan P., Inouye D.W. 2001. Flies and flowers: taxonomic diversity of anthophiles and pollinators. The Canadian Entomologist. 133:439–465.

Leather S.R. 2008. Insect sampling in forest ecosystems. John Wiley & Sons.

Leibundgut H. 1982. Europäische Urwälder der Bergstufe: dargestellt für Forstleute, Naturwissenschafter und Freunde des Waldes. Haupt Verlag Bern/Stuttgart.

Leibundgut H. 1984. Die Urwaldreste der Schweiz. Urwälder der Alpen.-München. p. 95–103.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T., Machida R.J. 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. 10:34.

Leuchner M., Hertel C., Menzel A. 2011. Spatial variability of photosynthetically active radiation in European beech and Norway spruce. Agricultural and forest meteorology. 151:1226–1232.

Leuchner M., Menzel A., Werner H. 2007. Quantifying the relationship between light quality and light availability at different phenological stages within a mature mixed forest. Agricultural and forest meteorology. 142:35–44.

Lopes C.M., Sasso T., Valentini A., Dejean T., Martins M., Zamudio K.R., Haddad C.F. 2017. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Molecular ecology resources. 17:904–914.

Lövei G.L., Sunderland K.D. 1996. Ecology and behavior of ground beetles (Coleoptera: Carabidae). Annual review of entomology. 41:231–256.

Ludwig G., Schnittler M., Vollmer I. 1996. Rote liste gefährdeter pflanzen Deutschlands. Bundesamt für Naturschutz.

Maguire D.Y., Robert K., Brochu K., Larrivée M., Buddle C.M., Wheeler T.A. 2014. Vertical stratification of beetles (Coleoptera) and flies (Diptera) in temperate forest canopies. Environmental entomology. 43:9–17.

Magura T., Elek Z., Tóthmérész B. 2002. Impacts of non-native spruce reforestation on ground beetles. European Journal of Soil Biology. 38:291–295.

Malaise R. 1937. A new insect-trap. Entomologisk tidskrift. 58:148–160.

Marquina D., Andersson A.F., Ronquist F. 2019a. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular ecology resources. 19:90–104.

Marquina D., Esparza-Salas R., Roslin T., Ronquist F. 2019b. Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular ecology resources. 19: 1516-1530

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.

Matthews R.W., Matthews J.R. 1983. Malaise traps: The Townes model catches more insects. Contributions of the American Entomological Institute.

Meusnier I., Singer G.A., Landry J.-F., Hickey D.A., Hebert P.D., Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. BMC genomics. 9:214.

Mitrović M., Tomanović Ž. 2018. New internal primers targeting short fragments of the mitochondrial COI region for archival specimens from the subfamily Aphidiinae (Hymenoptera, Braconidae). Journal of Hymenoptera Research. 64:191–210.

Morinière J., de Araujo B.C., Lam A.W., Hausmann A., Balke M., Schmidt S., Hendrich L., Doczkal D., Fartmann B., Arvidsson S. 2016. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PloS one. 11:e0155497.

Naughton-Treves L., Holland M.B., Brandon K. 2005. The role of protected areas in conserving biodiversity and sustaining local livelihoods. Annu. Rev. Environ. Resour. 30:219–252.

Noordijk J., Morssinkhof R., Boer P., Schaffers A.P., Heijerman T., Sýkora K.V. 2008. How ants find each other; temporal and spatial patterns in nuptial flights. Insectes sociaux. 55:266–273.

Novotny V., Basset Y. 2005. Host specificity of insect herbivores in tropical forests. Proceedings of the Royal Society B: Biological Sciences. 272:1083–1090.

O'Hara J.E., Cooper B.E. 1992. Revision of the Nearctic species of Cyzenis Robineau-Desvoidy (Diptera: Tachinidae). The Canadian Entomologist. 124:785–813.

Paoletti M.G. 2012. Invertebrate biodiversity as bioindicators of sustainable landscapes: Practical use of invertebrates to assess sustainable land use. Elsevier.

Park D.-S., Foottit R., Maw E., Hebert P.D. 2011. Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). Plos one. 6:e18749.

Phelps J., Webb E.L., Adams W.M. 2012. Biodiversity co-benefits of policies to reduce forest-carbon emissions. Nature Climate Change. 2:497.

Piñol J., Mir G., Gomez-Polo P., Agustí N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular ecology resources. 15:819–830.

Porter T.M., Hajibabaei M. 2018. Over 2.5 million COI sequences in GenBank and growing. PloS one. 13:e0200177.

Prange H.D. 1990. Temperature regulation by respiratory evaporation in grasshoppers. Journal of experimental biology. 154:463–474.

Proctor E., Nol E., Burke D., Crins W.J. 2012. Responses of insect pollinators and understory plants to silviculture in northern hardwood forests. Biodiversity and Conservation. 21:1703–1740.

Proesmans W., Bonte D., Smagghe G., Meeus I., Verheyen K. 2019. Importance of forest fragments as pollinator habitat varies with season and guild. Basic and Applied Ecology. 34:95–107.

Quicke D.L. 2015. The braconid and ichneumonid parasitoid wasps: biology, systematics, evolution and ecology. John Wiley & Sons.

R CoreTeam. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratnasingham S., Hebert P.D.N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes. 7:355–364.

Riley K.N., Browne R.A. 2011. Changes in ground beetle diversity and community composition in age structured forests (Coleoptera, Carabidae). ZooKeys.:601.

Ritter C.D., Häggqvist S., Karlsson D., Sääksjärvi I.E., Muasya A.M., Nilsson R.H., Antonelli A. 2019. Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. Genome. 62:147–159.

Rogg K.A., Taylor O.R., Gibo D.L. 1999. Mark and recapture during the monarch migration: A preliminary analysis. :133.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Röhrig E., Ulrich B. 1991. Temperate deciduous forests. Elsevier Amsterdam.

Sánchez S. 2013. Germany’s biodiversity at risk - call for action. International Union for Conservation of Nature

Sander C., Eckstein D. 2001. Foliation of spruce in the Giant Mts. and its coherence with growth and climate over the last 100 years. Annals of forest science. 58:155–164.

Scherber C., Eisenhauer N., Weisser W.W., Schmid B., Voigt W., Fischer M., Schulze E.-D., Roscher C., Weigelt A., Allan E. 2010. Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. Nature. 468:553-556.

Scherzinger W. 1996. Naturschutz im Wald: Qualitätsziele einer dynamischen Waldentwicklung. Ulmer.

Schmid U. 1999. Schwebfliegen auf dem Mittelmeer (Diptera, Syrphidae). Volucella. 4:167–170.

Schmidt W., Weckesser M. 2001. Struktur und Diversitat der Waldvegetation als Indikator fur eine nachhaltige Waldnutzung. Forst und Holz. 56:493–498.

Schulze C.H., Linsenmair K.E., Fiedler K. 2001. Understorey versus canopy: patterns of vertical stratification and diversity among Lepidoptera in a Bornean rain forest. Tropical forest canopies: Ecology and management. Springer. p. 133–152.

Seibold S., Bässler C., Brandl R., Büche B., Szallies A., Thorn S., Ulyshen M.D., Müller J. 2016. Microclimate and habitat heterogeneity as the major drivers of beetle diversity in dead wood. Journal of Applied Ecology. 53:934–943.

Shokralla S., Singer G.A.C., Hajibabaei M. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. BioTechniques. 48:305–306.

Sobek S., Tscharntke T., Scherber C., Schiele S., Steffan-Dewenter I. 2009. Canopy vs. understory: Does tree diversity affect bee and wasp communities and their natural enemies across forest strata? Forest Ecology and Management. 258:609–615.

Stefanescu C., Páramo F., Åkesson S., Alarcón M., Ávila A., Brereton T., Carnicer J., Cassar L.F., Fox R., Heliölä J. 2013. Multi-generational long-distance migration of insects: studying the painted lady butterfly in the Western Palaearctic. Ecography. 36:474–486.

Stoeck T., Bass D., Nebel M., Christen R., Jones M.D., BREINER H., Richards T.A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular ecology. 19:21–31.

Szujecki A. 1986. Ecology of forest insects. Springer Netherlands.

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

Tang C.Q., Leasi F., Obertegger U., Kieneke A., Barraclough T.G., Fontaneto D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences. 109:16208–16212.

Thomsen P.F., Sigsgaard E.E. 2019. Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. Ecology and Evolution. 9:1665–1679.

Ulyshen M.D. 2011. Arthropod vertical stratification in temperate deciduous forests: implications for conservation-oriented management. Forest Ecology and Management. 261:1479–1489.

Vance C., Kirby K., Malcolm J., Smith S. 2003. Community composition of longhorned beetles (Coleoptera: Cerambycidae) in the canopy and understorey of sugar maple and white pine stands in south-central Ontario. Environmental Entomology. 32:1066–1074.

Vesterinen E.J., Ruokolainen L., Wahlberg N., Peña C., Roslin T., Laine V.N., Vasko V., Sääksjärvi I.E., Norrdahl K., Lilley T.M. 2016. What you need is what you eat? Prey selection by the bat Myotis daubentonii. Molecular Ecology. 25:1581–1594.

Walter H. 1984. Vegetation of the earth and ecological systems of the geo-biosphere. Springer Verlag Berlin Heidelberg New York Tokyo.

Wangensteen O.S., Palacín C., Guardiola M., Turon X. 2018. DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. PeerJ. 6:e4705.

Wheeler Q.D., Raven P.H., Wilson E.O. 2004. Taxonomy: Impediment or Expedient? Science. 303:285–285.

Whitman D.W. 1988. Function and evolution of thermoregulation in the desert grasshopper Taeniopoda eques. The Journal of Animal Ecology. 57:369–383.

Whittaker R.H. 1960. Vegetation of the Siskiyou mountains, Oregon and California. Ecological monographs. 30:279–338.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.

Wickham H., Francois R., Henry L., Müller K. 2015. dplyr: A grammar of data manipulation. R package version 0.8.3. 3.

Williams C., Osman M. 1960. A new approach to the problem of the optimum temperature for insect activity. Journal of animal ecology. 29:187–189.

Yang C., Wang X., Miller J.A., de Blécourt M., Ji Y., Yang C., Harrison R.D., Douglas W.Y. 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators. 46:379–389.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Zhang G.K., Chain F.J., Abbott C.L., Cristescu M.E. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications. 11:1901–1914.

Zhou X., Li Y., Liu S., Yang Q., Su X., Zhou L., Tang M., Fu R., Li J., Huang Q. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. Gigascience. 2:4.

Zizka V.M., Leese F., Peinert B., Geiger M.F. 2018. DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. Genome. 62:122–136.

Chapter III

Unearthing the Potential of eDNA Metabarcoding of Soil – Towards Best Practice Advises for Invertebrate Assessments

Abstract

Metabarcoding has proven to be a powerful tool for the assessment of species communities from all kinds of habitats. When aiming to assess terrestrial invertebrate diversity, studies are frequently based on bulk samples which collection requires a comparatively high sampling effort. eDNA metabarcoding has been shown to be capable to limit sampling effort to a minimum while uncovering diversity of organism groups frequently being missed when collecting bulk samples (e.g. Nematodes). However, several authors have stated that eDNA from soil is a rather poor integrator for total biodiversity. Here the potential of eDNA metabarcoding of soil samples for assessing invertebrate biodiversity was evaluated. It is shown that choice of marker and extraction method is significantly influencing species detection rates, observed distribution patterns and thereby assessed invertebrate community composition. Next to an increase in amount of starting material, the application of a lysis step led to an increase in number of detected invertebrate species. While DNA extraction from a larger amount of soil allows for the detection of species with lower population densities, lysis enabled the assessment of transient organisms. Furthermore, choice of marker significantly influenced detected community composition. While the 18S marker resulted in the detection of a higher number of annelid and nematode species, the COI marker was more suitable for monitoring changes in arthropod community structure. This is highlighting the advantage of multiplexed studies for monitoring complex invertebrate communities. The here presented study significantly contributes to the development of best

practice guidelines for metabarcoding of soil samples by highlighting advantages and disadvantages of several important methodological considerations.

Introduction

"*It's not the soil itself, it's the soil life that is the most important element*." – Geoff Lawton

Soil is teeming with life. The degree of biodiversity which can be found in one square meter of soil is outstanding (Schaefer and Schauermann 1990; Decaëns 2010), making soil to the largest reservoirs of biodiversity on Earth (Lukac 2017). On a global scale, 25% of all described multicellular species reside in the soil (Curtis et al. 2002; Nielsen et al. 2011; Andújar et al. 2015), which is making up more than five times the known biodiversity of forest canopy (Stork 1988; May 1990; Decaëns 2010). Already in 1985 soil communities were described as 'the poor man's rainforest', referring to the enormous degree of biodiversity of which only a small portion has already been described (Usher 1985). A statement of Wall et al. (2005) best describes the large gap of knowledge about existing metazoan community composition in soils: There is probably no soil in the world, for which it would be possible to identify all residing invertebrates and microorganisms.

The enormous dimension of biodiversity associated with the soil habitat (André et al. 1994; Decaëns 2010) is further highlighted by the fact that it can vary strongly already on a small geographical scale. This is because biodiversity in soil is highly driven by moisture, acidity, temperature, nutrient content, organic matter and last but not least by soils physical appearance described by texture and structure. Both, texture and structure depend on geological factors which led to the formation of the soil parent material (Orgiazzi et al. 2016). Texture describes the

type and composition of mineral particles, which are classified according to their diameter into the four major groups gravel (> 2 mm), sand (2.0 - 0.063 mm), silt (0.63 - 0.02 mm) and clay (< 0.002 mm) (Orgiazzi et al. 2016). Structure refers to the arrangement of these soil particles into larger aggregates of different sizes and shapes and the resulting pore spaces formed between them (Orgiazzi et al. 2016). With increasing sampling depth structure and texture can change several times in a single soil core, resulting in the formation of several strata.

Most soils can be structured into five strata- which are also referred to as horizons (Orgiazzi et al. 2016). The uppermost layer is the O-horizon. This horizon is characterized by a high degree of organic content consisting of undecomposed or partly compost plant remains. Depending on vegetation cover, diversity and composition the thickness of the O-horizon can vary between a few millimeters and several centimeters. Underneath the O-horizon lays the A-horizon (Orgiazzi et al. 2016), which is characterized by a mixture of organic and mineral material. The combinations of these minerals is supporting biological and chemical activities whereas the A-horizons is often referred to as “engine room” of the soil (Orgiazzi et al. 2016). Together with the O-horizon forms the A-horizon the top soil. Underneath the A-horizon is the B-horizon located. This horizon is dominated by clay and iron minerals (Bardgett 2005). However, due to abiotic transport and bioturbation also smaller amounts of organic material can frequently be found in the B-horizon. The next lower level is the C-horizon which is located directly on top of the bedrock or parent material which is called the R-horizon (Orgiazzi et al. 2016).

Ever since the invention of agriculture, humanity was aware of the importance of a wide range of abiotic factors influencing fertility of arable land, but the importance of the biosphere was thereby widely neglected (Decaëns et al. 2006; Decaëns 2010). This led to the implementation of various practices in forestry and agriculture, like the application of fertilizers and pesticides whose influences on the existing biodiversity was and still is to wide parts unknown (Decaëns 2010). It was not until the beginning of the 21st century that politicians from all over the world became aware of the urgent need to protect soil biodiversity particularly as soil degradation can

lead to annual costs of several billion dollars (Kuhlman et al. 2010; Telles et al. 2011). With the 'Message from Malahide' in 2004 the EU outlined the priority objectives and targets for 2010 in order to halter biodiversity decline. In 2006 the EU released the "Thematic Strategy on Soil Protection" with the objective to define a common and comprehensive approach, focusing on the preservation of soil functions which includes prevention and restoration measurements (https://eur-lex.europa.eu/homepage.html references: COM(2006)231, COM(2006)232, SEC(2006)1165, SEC(2006)620). In 2012 the EU published a report of the implementation of the 'Soil Thematic Strategy' which clearly outpoints that after five years the formerly defined goals were still not met, mainly because a EU-wide monitoring and protection program had not yet been designed (https://eur-lex.europa.eu/homepage.html references: COM(2012)46). In the meantime, the ongoing decline of biodiversity is still strong thus making the development of timely and cost-efficient assessment strategies even more important than ever before.

New molecular tools such as metabarcoding and new sequencing platforms are a promising tool for this kind of tasks. It has already been shown that these methods are capable to accelerate and refine species community assessment studies (Elbrecht et al. 2019). Especially eDNA metabarcoding studies are expected to contribute significantly to the gain of knowledge about existing biodiversity in the near future. In contrast to studies based on community DNA, meaning DNA extracted from bulk samples, eDNA metabarcoding studies require only a very limited sampling effort. Environmental DNA (eDNA, i.e genomic DNA from different organisms and/or cellular material (Creer et al. 2016)) can be extracted from soil without separating target organisms prior to DNA extraction. Several studies have already shown the applicability of eDNA metabarcoding for the assessment of invertebrate diversity associated with the soil habitat (Bienert et al. 2012; Yang et al. 2013; Horton et al. 2017; Oliverio et al. 2018). However, a proof of concept for many invertebrate taxa like arthropods is still missing, precluding the development of best practice advices. However, it is well known that methodological decisions often have a direct influence on the recovered community composition (Alberdi et al. 2018; Dopheide et al.

2019). One crucial step that influences the outcome of diversity assessments is the choice of the DNA extraction method (Delmont et al. 2011; Fonseca 2018). While for microbial communities, studies have already investigated the influence of extraction methods (Delmont et al. 2011) and sample size (Kang and Mills 2006) on assessed community composition, similar detailed studies for many groups of invertebrates are still missing. It is likely, that for invertebrates specific sampling strategies could be more appropriate than the ones identified as most suitable for the assessment of prokaryote communities, because invertebrate morphology, size and abundance is more heterogeneous (Taberlet et al. 2012; Dopheide et al. 2019).

Today, many commercial kits are available, but most of them are optimized for assessing microbial diversity (Zinger et al. 2016). This is also reflected in the low amount of required starting material. However, even for microbial communities it has been shown that a low amount of source material results in less consistent and representative pictures of the local microbial community (Ranjard et al. 2003). Presumably, this bias is further enhanced when targeting invertebrates because of the above outlined differences in abundances, size and distribution patterns. A recent study has shown that extraction methods using larger amounts of soil resulted in the detection of a significantly higher number of invertebrate species (Dopheide et al. 2019). Nevertheless, Dopheide et al. (2019) did not take soil composition into account, although it has been shown that soil composition is having major influence on extracellular DNA yield (Taberlet et al. 2018). For microbial communities it has been observed that soil composition is determining suitability and thereby effectiveness of extraction methods and that under certain condition some methods should be preferably used although they provide poor results when conditions are not met (Sagova-Mareckova et al. 2008). This is because available commercial kits can strongly differ in workflow and composition of chemicals.

Another important consideration which is required when aiming to assess invertebrate communities composition from eDNA samples is the choice of marker (Fonseca 2018). Marker choice can significantly influence the composition of the recovered community as it has a direct

influence on taxa detection rate as well as on the accuracy of species identification against marker-specific reference databases (Yu et al. 2012; Yang et al. 2013; Andújar et al. 2018). This is inter alia attributed to the fact that markers are strongly differing is the degree of variability of primer binding sites (Deagle et al. 2014) and barcode region (Yu et al. 2012). By definition, "the perfect barcode is a short DNA fragment displaying a highly variable sequence which is flanked by two highly conserved regions. The central variable region is discriminative for all species of the target group, that is, it's sequence is uniquely associated to a given species and not shared with others" (Taberlet et al. 2018). Markers with highly conserved binding sites are usually more effectively amplified. However, decreasing variability usually goes along with a decrease in taxonomic resolution. Furthermore, complexity of reference databases plays a major role for completeness of the resulting species lists. For assessing invertebrates, the 18S and COI gene have the most appropriate reference databases. Both genes have already been tested for targeting soil community composition (Yang et al. 2013; Horton et al. 2017). While the COI marker seems to miss a wide range of taxa (Yang et al. 2013), the taxonomic resolution of the 18S database is hampering taxonomic identification at the species level (Tang et al. 2012; Yang et al. 2013). Several studies have now proven that a multiplexed study design (combination of several markers) can significantly contribute to a more complete picture of assessed biodiversity (Cowart et al. 2015; Drummond et al. 2015; Zhang et al. 2018; Marquina et al. 2019a). However, for soil invertebrates it remains unclear which marker are most suitable. Here we test the COI marker, known for its high taxonomic resolution for arthropods (Andújar et al. 2018), with the 18S marker which has already successfully been amplify for a broad range of metazoans (Horton et al. 2017).

Aims of the Study

1) Evaluation of suitability of different extraction methods for the assessment of soil invertebrate communities
2) Validation of the COI and 18S gene marker to examine soil invertebrate communities and test for the hypothesis that a multiplex approach is more powerful compared to single marker assessment studies.

Material and Methods

Sampling Strategy

All sample sites were located in the Eifel Nationalpark, which is situated in the south-western part of Germany close to the Belgian border (figure III.1).

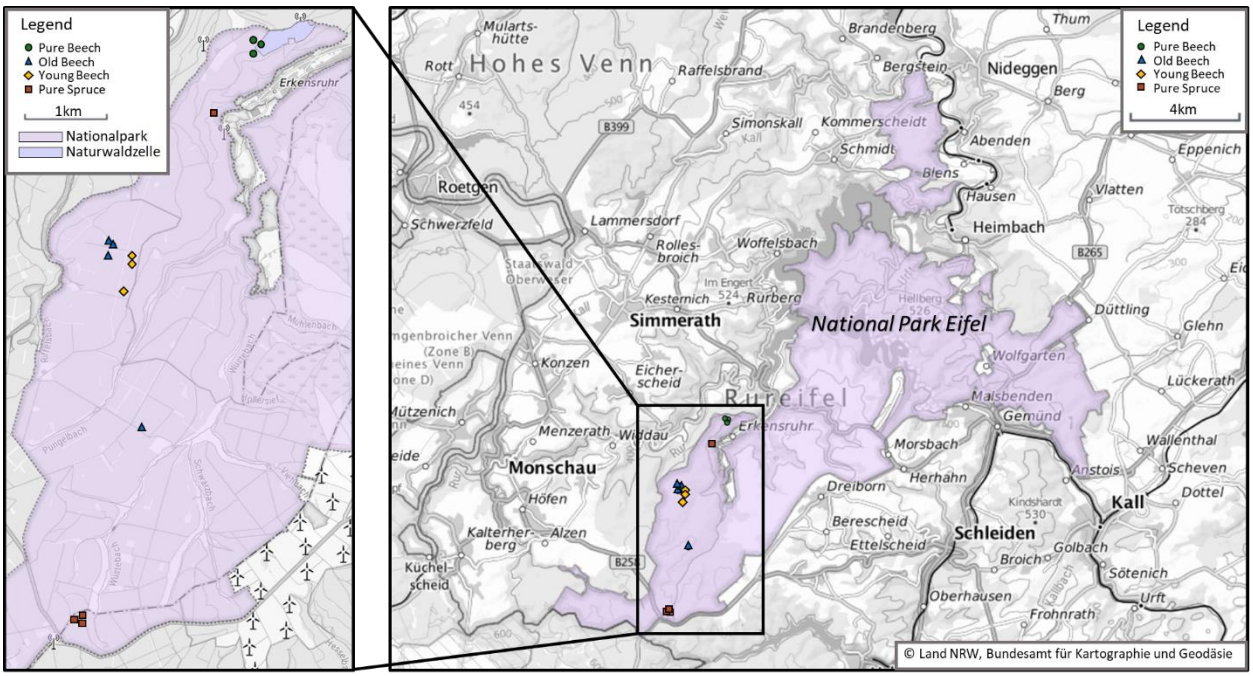


Figure III.1 Location of the 14 sampling sites located at the Eifel National Park, Germany

For this study a forest conversion gradient from Norway Spruce (*Picea abies*) to European Beech (*Fagus sylvatica*) was sampled. To reflect the different stages of conversion from spruce to beech, four forest types were defined: pure beech (PB), old beech (OB), young beech (YB) and pure spruce (PS) (table 1).

Table III.1 Geographical location and ecological characteristics of the 14 sampling sites. Depicted for each sampling site are the Coordinates (altitude N and latitude E) and the associated forest type

Sample Sites	Coordinates	Forest Type
Sample Site 01	50° 34'11.7984''N 6°21'32.1012''E	Pure Beech
Sample Site 02	50° 34'07.7016''N 6°21'27.3996''E	Pure Beech
Sample Site 03	50° 34'12.9000''N 6°21'27.3996''E	Pure Beech
Sample Site 04	50° 32'44.5992''N 6°20'15.2988''E	Young Beech
Sample Site 05	50° 32'41.3016''N 6°20'15.6984''E	Young Beech
Sample Site 06	50° 32'29.7996''N 6°20'11.1012''E	Young Beech
Sample Site 07	50° 32'29.7996''N 6°20'11.1012''E	Old Beech
Sample Site 08	50° 31'35.1984''N 6°20'25.2996'E	Old Beech
Sample Site 09	50° 32'48.3000''N 6°20'03.4008''E	Old Beech
Sample Site 10	50° 30'17.2008''N 6°19'48.1008''E	Pure Spruce
Sample Site 11	50° 30'18.2988''N 6°19'51.4020''E	Pure Spruce
Sample Site 12	50° 33'15.8004''N 6°21'07.3008''E	Pure Spruce
Sample Site 13	50° 30'16.0056''N 6°19'51.4704''E	Pure Spruce
Sample Site 14	50° 32'49.9632''N 6°20'00.7296''E	Old Beech

The forest types differed in tree species composition as well as in approximate age of the trees. The pure beech and pure spruce sampling sites were located in monoculture stands which were dominated either by beeches or spruces respectively. The pure beech monoculture stands were approximately 180 years old and partly under special protection through North-Rhine Westfalia (Naturwaldzelle) (Sample Site 01). With a mean age of 60 years, the pure spruce stands were substantially younger. Spruces of the same age dominated the young beech sampling sites. At the young beech sampling sites, young beeches had been planted only recently and had not yet reached three meters in size at the time of sampling. At the old beech sampling sites, beeches had already reached a height of more than 3 meters and actions to remove spruces from the forest had already been undertaken (figure III.2).

Figure III.2 Forest conversion gradient from spruce monocultures over with beeches underplanted spruce forests to beech monocultures. Photos were taken during summer season (July 2016).

Each soil sample consisted of approximately twenty 44mm diameter x 100mm cores, taken 5 cm apart from each other. As already mentioned, is especially the top soil teeming with life (Coleman et al. 2004; Orgiazzi et al. 2016), which is why here only the O- and A-horizons (top 10cm) were sampled. A total of 162 soil samples were collected and kept in individual 250 ml containers.

Shortly after sampling, soil samples were transported to the laboratory and stored at -20°C until further processing. To monitor changes in local existing biodiversity over a time period of one year a sampling trial was conducted at each season (Summer, Autumn, Winter Spring) (table III.2).

Table III.2 Time of sampling. Depicted for each sampling period is the name of the season, associated time of the year, number of samples taken and sampling dates (d,m,y)

Season	Time of the year	Number of Samples	Sampling Dates
Season 1	Summer	36	27.07.2016
Season 2	Autumn	48	27.10.2016
Season 3	Winter	48	25.01.2017
Season 4	Spring	48	26.04.2017

DNA Extraction

Two different extraction strategies were applied for the extraction of eDNA from soil samples. Firstly, a commercial DNA extraction kit and secondly the Taberlet et al. (2012) phosphate buffer protocol. The first method uses 0.5 g of soil per sample followed by DNA extraction using a commercial silica membrane based kit. Briefly, after thoroughly mixing the sample, DNA was extracted from 0.5g soil per sample using the Macherey-Nagel NucleoSpin® Soil kit following the manufacturer's protocol.

The second method allows the extraction of DNA from larger amounts of starting material. With a phosphate buffer based solution (Na2HPO4; 0.12M; pH 8) (Taberlet et al. 2012) eDNA from 50 -100g of soil per sample were extracted following the protocol of Taberlet et al. (2012). Briefly,

soil samples were removed from the -20°C chamber approximately 12 hours before DNA extraction and stored at +4°C overnight. The next morning, each sample was thoroughly mixed before a saturated phosphate buffer solution of equal weight was added. All samples were processed in duplicates and placed in an orbital shaker at 120 rpm for 15 minutes. Two 2ml Eppendorf safe lock tubes were filled with 1.7 ml of the resulting mixture and centrifuged for 10 min at 10000g. Afterwards, 400µl of the resulting supernatant were transferred to a new 2ml collection tube to which 200µl of SB binding buffer, included in the Macherey-Nagel NucleoSpin® Soil kit, was added. Lysate duplicates were loaded onto the same NucleoSpin® Soil Column and centrifuged at 10000 g for one minute. For the rest of the extraction, the standard manufacturer's protocol of the Macherey-Nagel 'Genomic DNA from soil' kit was followed starting from step 8. Finally, DNA was eluted with 50µl of SE Buffer (Macherey-Nagel). 10ul of the resulting elution step was combined with 90µl pure H_2O (Sigma), followed by DNA purification using the PowerClean® Pro DNA Clean-Up Kit (MO Bio Laboratories, Inc.) following the manufacturer's protocol.

Choice of Primers and Library Preparation

For amplicon library preparation of soil samples, two primer pairs targeting two different markers were used. Library preparation was conducted following a two step PCR approach (figure III.3). In the first step, the fragment of interest was amplified using primers targeting specific gene regions together with annealing an Illumina adaptor overhang (referred to as PCR 1). In the second step (referred to as PCR 2) Illumina adaptor and index-tag primers are added (Bourlat et al. 2016; Fonseca and Lallias 2016).

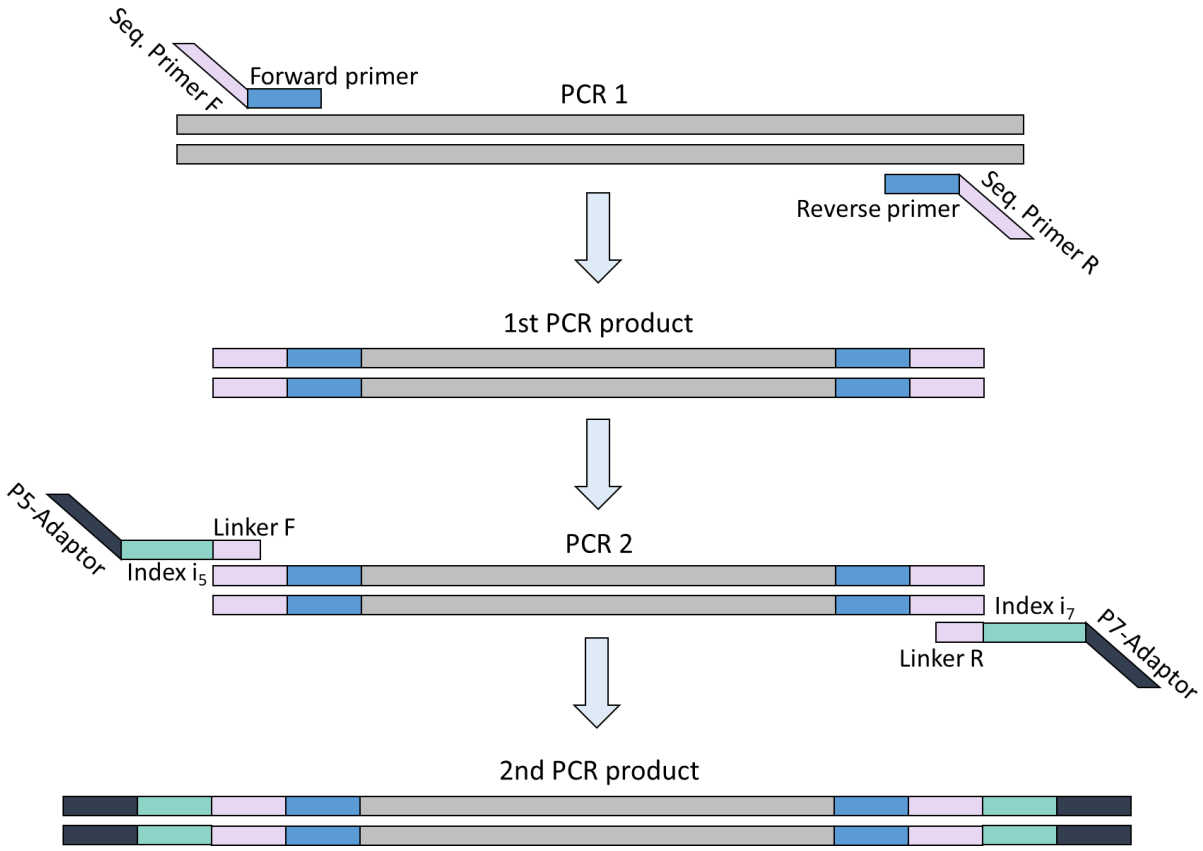


Figure III.3 Illustration of the 2-step PCR approach

Here two different markers were targeted: Firstly, the 18S rRNA (18S) nuclear marker which is more conserved but can detect a wide range of taxonomic groups and secondly, the more variable Cytochrome C Oxidase Subunit I (COI) mitochondrial marker which enables a better taxonomic resolution within the Metazoa, but at the cost of an incomplete assessment of other groups e.g. Chordata (Tang et al. 2012; Cowart et al. 2015). To amplifying approximately 380bp of the V4 region of the nuclear 18S rRNA the forward primer TAReuk454FWD1 *(5'-CCAGCASCYGCGGTAATTCC-3')* was combined with the reverse primer TAReukREV3r *(5'-ACTTTCGTTCTTGATYRA-3'*) (Stoeck et al. 2010). The mitochondrial COI primer pair consisted of the forward primer mlCOIintF *(5'-GGWACWGGWTGAACWGTWTAYCCYCC-3')* (Leray et al. 2013)

and the reverse primer dgHCO2198 *(5'-TAAACTTCAGGGTGACCAAAAAATCA-3')* (Leray et al. 2013), together targeting a 313bp region of the 658bp long barcoding COI gene.

Approximately 10ng of template DNA was used for all PCR reactions. For PCR1 the chosen forward and reverse primers were preceded by the universal Illumina adaptors, *5'-* ACACTCTTTCCCTACACGACGCTCTTCCGATCT *primer- 3' and 5'-* GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT *primer* -3', respectively. For PCR1 the mastermix consisted of 7.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1µl Sigma H2O, 0.5µl of forward Primer, 0.5µl of reverse primer, 0.5µl Bovine Serum Albumin (thermoscientific) and 1µl template DNA, making up a total of 15µl. The first PCR (PCR1) conditions consisted of an initial denaturation of 2 min at 98°C, followed by 20 cycles with 40 s at 98°C, 40 s at 45°C, 30 s at 72°C (COI) or 20 cycles with 40 s at 98 °C, 40 s at 55 °C, 30 s at 72 °C (18S), and a final extension of 3 min at 72°C. Afterwards, PCR1 products were purified with HT ExoSAP-IT^{TM} (appliedbiosystems) by adding 4µl of HT ExoSAP-IT^{TM} to each sample. Subsequently, samples were first heated up for 15 min at 37°C, followed by 15 min at 80°C and finally cooled down for 5 min at 4°C. To add the Illumina index tag adaptors a second PCR (PCR2) using 8µl of purified PCR1 products was performed. For the index PCR2 the forward and reverse primers included the P5 and P7 index Illumina adaptors, followed by an 8-bp Illumina Nextera barcode and the universal Illumina adaptors, respectively. For PCR2 the purified PCR products were split onto two PCR tubes. Each tube contained 12.5µl Q5 Hot Start High-Fidelity 2X MasterMix (New England BioLabs), 3µl Sigma H2O, 1.2µl of forward primer, 1.2µl of reverse primer and 8ul purified PCR1 product. The PCR2 conditions consisted of an initial denaturation of 2 min at 98°C, followed by 20 cycles with 40 s at 98°C, 30 s at 55°C, 30 s at 72°C and a final extension of 3 min at 72°C. PCR2 products were visualized by gel electrophoresis and purified using the QIAquick Gel Extraction Kit (Qiagen), according to manufacturer's instructions. All final purified amplicons (PCR2) were quantified using the Quantus Fluorometer (Promega) and diluted to the same concentration (3ng/µl) before being pooled together to create two amplicon libraries (18s and

COI). The resulting purified amplicon pools were sequenced on two runs of Illumina Miseq (2x 300bp) sequencing platform at Liverpool University's Centre for Genomic Research (UCGR, Liverpool).

Bioinformatics and Data Analysis

Raw data was initially quality checked at UCGR, where fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1. Sequences were then trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 20bp after trimming were removed. Sequences were received from MacroGen Inc. Seoul in Casava 1.8 paired-end demultiplexed fastq format.

The fastq sequences were checked for the presence of the COI and 18S primers with Cutadapt version 1.18 (Martin 2011) using the following settings: maximum error rate (-e): 0.1, minimum Overlap (-O): 20, minimum sequence length (-m): 50). Sequences lacking either forward or reverse primer were removed from the dataset. Subsequently paired-end reads were merged with vsearch version 2.7.0 (Rognes et al. 2016). Merged sequences with a length of 360-400bp for the 18S and 293-333bp for the COI dataset respectively were retained for further analysis and filtered with a maxEE threshold of 1.0 using vsearch (version 2.7.0) (Rognes et al. 2016). Afterwards fastq-sequences were demultiplexed using the script split_libraries_fastq.py implemented in QIIME1 (Caporaso et al. 2010). Thereby A phred quality threshold of 19 was chosen. Dereplicating, size sorting, denovo chimera detection as well as OTU clustering with a 97% cutoff was conducted with vsearch 2.7.0 (Rognes et al. 2016). Finally, an OTU table was built by using the --usearch_global function in vsearch 2.7.0 (Rognes et al. 2016) followed by the python script "uc2otutab.py" written by Robert Edgar

(https://drive5.com/python/uc2otutab_py.html). Resulting OTU tables of both markers were further curated with LULU (Frøslev et al. 2017). Curation was started with an initial blasting of OTU representative sequences against each other using blastn (version 2.9.0). The following parameter settings were chosen: 'query coverage high-scoring sequence pair percent' (-qcov_hsp_perc) was set to 80, meaning that a sequence was reported as match when 80% of the query formed an alignment with an entry of the reference file. Secondly minimum percent identity (-perc_identity) was set to 84, requiring the reference and query sequence to match at least to 84% to be reported as a match. The format of the output file was customized using the –outfmt settings '6 qseqid sseqid pident'. The resulting output file included the name of the query sequence and the name of the reference sequence next to percentage of identical match. Subsequently, the resulting filtered OTU match list was uploaded into R (version 3.5) (R CoreTeam 2013). Subsequently the R-package 'lulu' (version 0.1.0) (Frøslev et al. 2017) was used to perform post clustering curation using standard settings. The LULU algorithm filters the dataset for suspicious OTUs. Afterwards, suspicious OTUs are either classified as "daughter OTU" and merged with the corresponding "parent OTU", or are discarded from that dataset.

For taxonomy assignment COI sequences were blasted against the BOLD database using blastn 2.9.0+ (Altschul et al. 1990). As the BOLD database is strongly limited in number of bacterial sequences and barcodes of many eukaryotic species outside Metazoa, a second database was downloaded from GenBank using the following search critera: (COI[All Fields] OR COX1[All Fields]) OR CO1[All Fields] AND (fungi[filter] OR protists[filter] OR bacteria[filter] OR archaea[filter]). All sequences not assigned to Metazoa when blasted against the downloaded BOLD database, where in a second step compared to the GenBank reference database. For taxonomy assignment of the 18S sequences a similar approach was used. First, all sequences were blasted against a customized reference database downloaded from GenBank according to the following criteria: ((18S) OR V4 AND ((animals[filter] OR fungi[filter] OR plants[filter]))).

Sequences without assignment were in a second step blasted against the newly released SILVA132 release.

The resulting OTU table was loaded into Excel and formatted for upload into R (R CoreTeam 2013). For statistical analysis several R studio v3.5 packages were used. To compare the number of OTUs and proportion of taxonomic assignments between samples and per gene marker the R package ggplot2 was used (Wickham 2016). The proportion of species/ OTUs recovered per gene marker were calculated using the R package dplyr (version 0.8.3) (Wickham et al. 2015) and visualized using R package Venn. Diagram (version: 1.6.20) (Chen and Boutros 2011). The resulting plots were modified using Microsoft PowerPoint. Similar analysis were done to visualize the number of shared and unique species (including arthropods) between extraction methods depending on choice of marker and the number of unique and shared arthropod species between sampling sites per forest type and season. Taxonomic composition for each extraction method was prepared using Microsoft PowerPoint after calculating the overlap and number of unique species using the R package Venn.Diagram (version: 1.6.20) (Chen and Boutros 2011). To identify differences in community composition for the COI and 18S dataset a Principal Coordinate Analysis (PCoAs) was calculated using the R package betapart (version 1.5.1) (Baselga and Orme 2012) and vegan (version 2.5-6) (Dixon 2003), thereby “jaccard” was chosen as family of dissimilarity indices. To identify ecological trends of community composition and species distribution along the different forests and seasons Marioko plots using the R package ggplot2 (Wickham 2016) were calculated. To further assess spatial patterns of beta diversity the R package betapart (version 1.5.1) (Baselga and Orme 2012) and vegan (version 2.5-6) were used. This allowed multiple-site dissimilarity measures to account for compositional heterogeneity between samples within forest types and depending on season. UpsetR-Plot showing detailed number of unique and shared species (BlastID ≥ 99%) between forest types was prepared using the R package UpSetR (version 1.4.0) (Conway et al. 2017). To visualize differences in chemical composition between soil samples

the R package 'yarrr' (version 0.1.5) (Phillips 2017) was used. Finally, sediment classification after Plot and Pye were visualized using GRADISTAT (Blott and Pye 2001).

Granulometry

Granulometry tests were performed at Instituto Português do Mar e da Atmosfera (IPMA) located in Portugal. For soil granulometry analysis all samples went through several steps: pre-treatment, sieving, grain size by laser diffraction, sieve merging, Malvern data, sample classification and elemental analysis of carbon, nitrogen and hydrogen.

Pre-treatment

All soil samples were pre-treated due to high organic matter content. Samples were subjected to hydrogen peroxide (H2O2) attack, intended to destroy the organic matter. On day one, 25 ml of hydrogen peroxide (130 volumes) was added to the sub-samples, stirred and allowed to react. Depending on the reaction (noted by effervescence), a further 50 ml of hydrogen peroxide with the same concentration was added if necessary. The next day, 50 ml of the same compound was added, and depending on the reaction repeated with 75 ml of the compound. The process was repeated until no reaction was any more noticeable. Afterwards, the sediment was placed in an oven for 24h, at a temperature of 60°C, here the soil samples were covered by an aluminium foil in order to ensure that no trace of effervescence could contaminate further samples. This step was repeated as often as necessary to allow evaporation of the excess H2O2. During this stage, the subsamples were stirred several times a day. On a regular basis distilled water was added to replace evaporated water. Finally,

samples were let to evaporate in order to eliminate any vestiges of hydrogen peroxide and gases. This process was carried out 4-5 times, until the sediment was clean and the water was clear. From this moment on the samples were prepared for the next phase.

Sieving

The grain size analysis included the separation of two main fractions by wet sieving diameters larger than 500µm and less than 500µm. Each fraction was dried at the oven (60°C), where the maximum total weight was around 19g. The >500 µm fraction accounted for a minimum of 0.26g and on average of 3g of total weight. The fraction above 500µm was sieved out using mesh screens of different sizes (8mm, 4mm, 2mm, 1mm, 500µm and bottom). The resulting fractions were weighed and the respective percentage calculated.

Grain Size Analysis by Laser Diffraction

The fraction below 500µm was analysed using the Mastersizer 2000 (Malvern Inc.) through a wet dispersion unit - the Hydro G. This equipment uses the laser diffraction as a measure of grain-size. It can detect very small particles of up to 0.01µm. The equipment measures the relative amount of volume, with which each size class contributes to the whole sample. This type of analysis follows specific “Standard Operating Procedures” (SOP) according to the sample type. This is of paramount importance in order to ensure that consistent results are obtained. The SOP started with the register of the sample name followed by a laser calibration and the measurement of laser background. A few µl of a sodium hexametaphosphate solution was added to the wet dispersion unit. The mixture was submitted to ultrasonic power for 3 minutes, where the sample deflocculate and disaggregate. The SOP used included 3 measurements and the respective average calculation. The measurements required

consistency. If case of major differences within the three measurements analysis was repeated.

Merging of Sieves and Malvern Data

In contrast to Malvern data were sieving results weight based. Malvern data are given in volume percentage of sediments of each size fraction. Therefore, the results are not directly comparable making it difficult to assimilate size data obtained by using several methods. In order to achieve results including grain-size of the whole samples, results achieved with both methods were merged. The respective percentages were recalculated with the assumption that the density of the materials is equal for both methods.

The calculated results (in percentages) were loaded into GRADISTAT program (version 8.0) (Blott & Pye, 2001). The following statistics were calculated: mean, mode(s), sorting (standard deviation), skewness, kurtosis, and a range of a cumulative percentile values (the grain size at which a specified percentage of the grains are coarser), namely D10, D50, D90 among others. The statistical parameters used in this work were the “Geometric method of moments” (Blott & Pye, 2001).

Sample Classification

The samples were classified according to the percentages of the main grain-sizes classes (gravel, sand and mud) (Blott and Pye 2012). Classification encompass five first-order size classes (boulder, gravel, sand, silt and clay), each of them has five second-order subdivisions with limits defined at one phi intervals.

Elemental Analysis (C, N)

Total Carbon (TC) was determined using an elemental analyzer TruSpec CHNS Micro (Version 2.72) from LECO. The elemental analyzer measured the relative proportion of elemental content of carbon, nitrogen and hydrogen of dried samples. The applied detection method for carbon is an optimized, low noise and non-dispe rsive Infrared (IR) absorption. On a first stage, the samples were grinded on a planetary micro mill from Fritsch (Pulverisette 7 Classic Line) over 5 minutes at 600 rpm. To remove remaining organic matter, the samples were submitted to a 3 hours combustion at 400°C, with 1 hour rising temperature levels in a Labotherm L3/S muffle from Navertherm. Using tin capsules (3,2 x 4 mm) with 1-2 mg of each sample, the elemental analyses were performed on three trials of each grinded sample and on two trials on the respective combusted sample. The relative content of total carbon (TC) was obtained from the first three trials, while relative total inorganic carbon content (TIC) was assessed from the last two trials. The total organic carbon content (TOC) was calculated as difference between TC and TIC. The relative precision range of repeated measurements of the samples and standards was 0.05%.

Results & Discussion

Influence of Marker Choice

Amplification of the COI marker resulted in the detection of 25,036,251 high quality-filtered reads, which were subsequently clustered into 31781 OTUs. When amplifying the V4 region of the 18S marker 22,036,784 quality filtered reads were obtained, which were clustered into 33953 OTUs. After Lulu curation 23004 OTUs remained in the COI dataset (72.4%), while 53.9% of detected 18S OTUs were excluded from the dataset leaving 15650 OTUs for further analysis.

Although the COI barcode is already on its way to become the standard metazoan community DNA metabarcode, it still stands in the firing line of critics when working with eDNA (Deagle et al. 2014). Several authors have already stated, that the applicability of the COI marker is strongly limited for eDNA studies, as it appears to amplify primarily bacteria DNA (Yang et al. 2013; Horton et al. 2017). This is leading to a high percentage of non-target OTUs which can often not be assigned to any taxon. The here presented study is supporting these findings. When comparing the two marker datasets, assignment rate of the COI dataset was significantly lower (figure III.4). While 64% of all retrieved 18S OTUs had an assignment, only 12.59% of the COI OTUs (figure III.4a) matched an entry in the reference databases based on a blast identity cutoff of at least 90%. Considering the very complex available COI databases for Metazoa the low assignment rate suggests that the COI marker is more constraint by non-

target amplification in comparison to the 18S marker. However, only 17 OTUs of the COI dataset had an assignment to Bacteria, which is contradicting the results of Yang et al. (2013) who stated that the highest percentage of OTUs retrieved from soil samples accounts for bacterial species. Nevertheless, we cannot entirely reject Yang et al. (2013) hypothesis as Horton et al. (2017) suggests, that primer binding regions of the metazoan COI gene are similar to DNA regions within the genomes of other taxonomic groups including Bacteria. To test for this hypothesis and to proof the findings of Yang et al. (2013) it would be necessary to blast all unassigned OTUs against a widely extended reference database containing the genomes of several bacteria. However here it was shown that next to bacteria are also further non-target groups are increasingly amplified by the COI marker e.g. fungi. While the COI marker is highly discriminative for several metazoan groups, it widely fails to distinguish species of these non-target groups (Dentinger et al. 2011). As a result, COI reference databases for eukaryotic taxa outside Metazoa are fairly incomplete, hampering taxa identification. Nevertheless, despite incomplete databases a high number of COI OTUs was assigned to Fungi. This shows that the use of the mICOIIntF primer in combination with the dgHCO2198 COI primer lead to the amplification of a high amount of non target DNA. It is likely that with growing complexity of databases, number of assignments to non-target taxa will further increase. However, here focus was set on Metazoa. While 21.4% of OTUs contained in the 18S dataset (3350 OTUs) were assigned to Metazoa, only 4.1% of the OTUs contained in the COI dataset had an assignment to this taxon (953 OTUs) (figure III.4b).

The kingdom “Metazoa” comprises a broad range of phyla encompassing several marine, aquatic and terrestrial life forms. With the 18S marker 10 phyla were identified: Annelida, Arthropoda, Chordata, Cnidaria, Gastrotricha, Mollusca, Nematoda, Platyhelminthes, Rotifera, Tardigrada (figure III.4b). With the COI maker only six phyla were identified. This is not least contributed to the fast that the here used customized COI-reference database did not contain reference sequences of the phyla Cnidaria, Gastrotricha, Platyhelmintes and Rotifera as these

taxa were not targeted. However, it cannot be excluded that some unassigned OTUs of the COI dataset derived from these four phyla as 2% of OTUs detected with the 18S dataset were assigned to them.

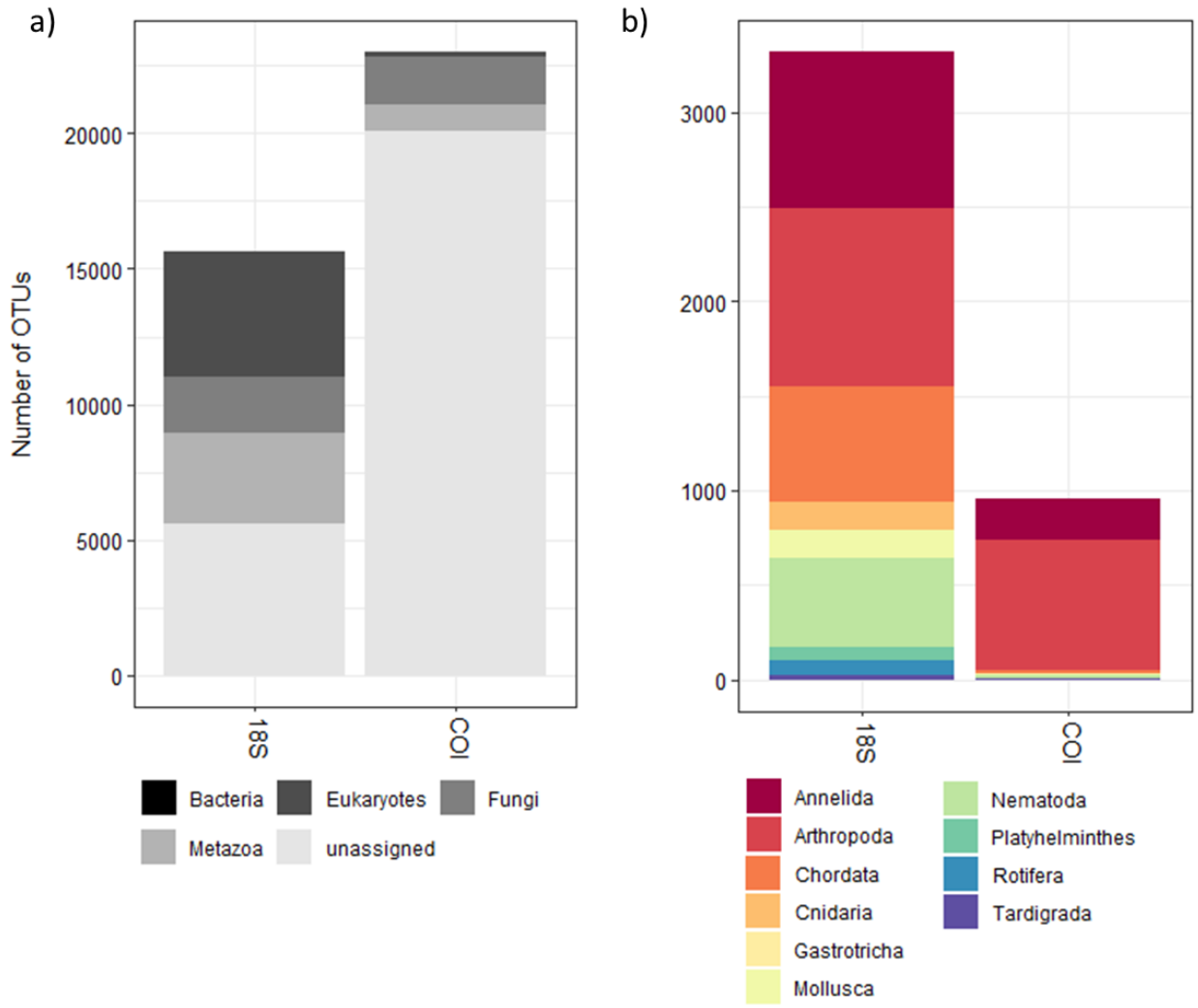


Figure III.4 Number of OTUs assigned to Bacteria, Fungi, Metazoa and remaining Eukaryotes (blastID ≥ 90%). Furthermore, number of unassigned OTUs is indicated in light grey (a). Number of OTUs assigned to Metazoa on phyla level (blastID ≥ 90%) (b)

Here the focus was set to the Metazoa phyla Annelida, Arthropoda and Nematoda. This choice was made as all three phla play a major role for the maintenance of the soil habitat (Coleman et al. 2004). For the three chosen phlya the numbers of detected species and therefore also the assessed taxonomic composition varied according to the gene marker used (figure III.5). For the annelids 12 different species were identified at 99% blast match sequence similarity with GenBank and BOLD database respectively. The 18S rRNA gene recovered the eight annelid species *Achaeta affinis*, *Achaeta unibulba*, *Bryodrilus ehlersi*, *Eisenia fetida*, *Mesenchytraeus armatus*, *Mesenchytraeus flavus, Mesenchytraeus pelicensis* and *Oconnorella cambrensis,* whereas the COI marker detected the four annelid species *Achaeta bifollicula*, *Achaeta brevivasa*, *Dendrobaena octaedra* and *Dendrodrilus rubidus*. Accordingly, there were no shared annelid species between the two markers (Figure III.5a), evidencing how complementary nuclear and mitochondrial markers can be (Drummond et al. 2015).

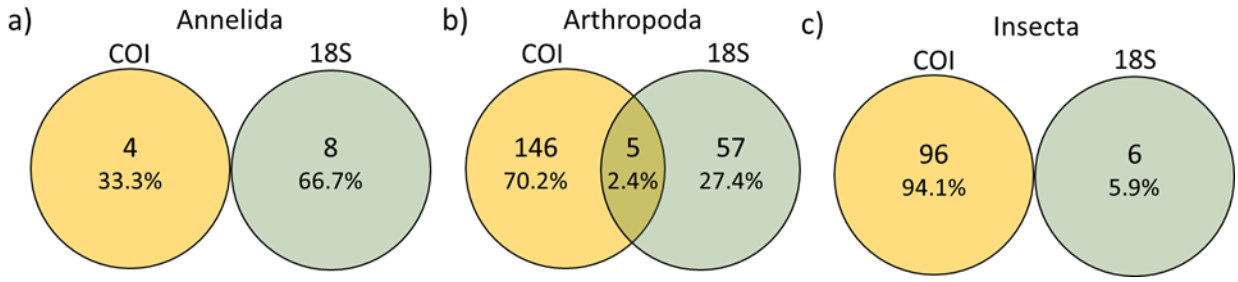


Figure III.5 Number of species (blastID ≥ 99%) recovered with either or both markers depending on organism group

To check for possible biases introduced by incomplete reference databases, the completeness of the two used reference database for the 12 detected annelid species were investigated. All of the eight species recovered with the 18S marker had next to an entry in GenBank also an entry in the BOLD database. In contrast to that, none of the four species detected with the

COI marker had an entry in the in-house customized 18S GenBank database. This is indicating that the COI maker may not be a perfect match when targeting annelids, while identification of annelids on basis of the 18S marker is likely to be more hampered by incomplete reference databases. The reasons why the COI marker failed to detected the eight species found with 18S marker can be manifold. The fact that the metazoan primer binding sites are often not highly conserved for the COI gene limits DNA amplification in favor for DNA of other groups. Deagle et al. (2014) showed that as a result of homoplasy, variation at the primer binding sites becomes saturated between distantly related taxa, what is hampering the development of primers, targeting only certain groups. A fact, which becomes increasingly important the higher the proportion of non-target DNA is in the DNA mixture. Additionally, it has been described that several annelid taxa are lacking an appropriate sized COI barcode gap (Kvist 2014). The barcode gap is the differences in genetic difference between the highest intraspecific variation and lowest interspecific divergence (Hebert et al. 2003a, 2003b; Wiemers and Fiedler 2007; Meier et al. 2008; Kvist 2014). A lack of this gap usually leads to false negative and positive results. In 2014 Kvist et al. showed that of the phylum Annelida especially earthworms are suffering from an insufficiently sized barcode gap. In fact, interspecific comparison of many earthworm species show 0% divergence. In views to this study the absence of a sufficiently sized barcode gap might have led to the assignment of sequences originating from different species to the same taxon, leading to false negative results and finally to the underestimation of local existing annelid diversity.

While the COI marker failed to recover eight species, the 18S marker left four species undetected. As all of them had no entry in the 18S database its detection was not possible. This highlights that the power of metabarcoding is directly linked to completeness of available reference databases. Andújar et al. (2018) stated that the utility of a reference sequence database is a function of (a) the inherent power of the marker for taxonomic assignment; (b) the taxonomic coverage (number of species and phylogenetic diversity represented in the database)

and depth (number of individuals sequenced per species) of reference sequences; and (c) the adequate formatting and curation of the database. To tell a long story short, the size and completeness of a reference database is negatively correlated with the probability of false positive and negative assignments (Somervuo et al. 2017). The more complex the reference database is, the more likely are reliable assignments on lower taxonomic ranks (Somervuo et al. 2017). In terms of completeness are COI databases unparalleled. Due to its high taxonomic resolution the Consortium for the Barcode of Life (CBOL) decided to use the COI marker for standard DNA barcoding of single animal specimens (Hebert et al. 2003b; Ratnasingham and Hebert 2007). The first major program of the CBOL has been the BARCODE 500K project, which had the overreaching aim to establish a reference database containing the barcodes of 0.5 million species. This goal was met in August 2015, but the CBOL kept going and even expanded. Today, more than 502.700 public barcodes clusters are stored in the BOLD database. No database for any other genetic region covers as many taxa as the Barcode of Life Database (BOLD) does for the COI (Deagle et al. 2014). With some exceptions, the 18S marker does not have species-level resolution and thus discriminating between close-related species (e.g. cryptic) could be daunting or impossible. Many closely related species do only show slight variation within the 18S barcode regions whereas it cannot be excluded, that two distinct species are showing the same barcode (Tang et al. 2012). Additionally, eventual low maintenance of existing nuclear databases can also contribute to 18S poor taxonomic assignments in particular on lower taxonomic ranks. All of this is increasing the risk of false positive assignments and a resulting distorting of the assessed degree of existing biodiversity. Nevertheless, it should be kept in mind that some groups are also largely underrepresented within the BOLD database e.g. Nematodes. Up to today around 30.000 nematode species have been described but estimates put the number of total nematode species as high as a million (Kiontke and Fitch 2013). Out of the 30.000 described species, the barcodes of 2141 species are stored in the BOLD database. In comparison to that, contained the in-house customized 18S GenBank database the barcodes of 6796 Nematode species. While 469 OTUs of

the 18S dataset were assigned to Nematoda, only 22 OTUs of the COI dataset had an assignment to this phylum. Against the background that one gram of soil can hold several hundreds of nematodes (Hoorman 2011) these results allow for the assumption that the COI marker largely underestimated local nematode diversity. Former studies have already shown that the Folmer region gives very poor amplification results for marine nematode species (De Ley et al. 2005; Bhadury et al. 2006; Derycke et al. 2010b). It has been observed that nematodes show a high nucleotide variability and indels at the COI primer sites (Blouin et al. 1998; Creer et al. 2010). Due to these exceptional high mutation rates the design of universal primer targeting this gene is impossible as its use will in any case result in poor amplification and consequently poor taxa recovery rate. Although amplification success of the 18S region was significantly higher it has been described that the low taxonomic resolution of this marker does often not allow for the discrimination of closely related species (De Ley et al. 2005; Derycke et al. 2010a). Therefore, it was not possible to fully assess nematode diversity on basis of the here used marker combination.

While the 18S marker detected more annelid and nematode species, assessed arthropod diversity with the COI marker was significantly higher. The COI gene exclusively uncovered 146 species, ca. 70% of all detected arthropod species, whereas the 18S retrieved only 57 species not found by the COI marker. Additional five species were detected by both markers (figure III.5b). This result accords with several former studies showing that the 18S marker is prone to underestimate the real diversity of several metazoan species (Tang et al. 2012), including Arthropoda (Cowart et al. 2015; Drummond et al. 2015). While the primer binding sites of the 18S marker are stronger conserved (Clarke et al. 2017), species-level resolution is strongly hampered for the 18S marker. Consequently, it agglutinates distinct species into the same OTU. Especially on lower taxonomic level the higher degree of conservation is therefore likely to lead to a strong underestimation of the total existing biodiversity. This is also reflected in number of recovered species of the Arthropoda class Insecta. While 96 insect species with a blastID of at least 99% were detected with the COI marker, solely the six insect species *Agriotes acuminatus*

(Insecta: Coleoptera), *Cooloola ziljan* (Insecta: Orthoptera), *Geostiba circellaris* (Insecta: Coleoptera), *Hadrotes crassus* (Insecta: Coleoptera), *Photophorus jansonii* (Insecta: Coleoptera) and *Stigmella lycii* (Insecta: Lepidoptera) were found when analysis was based on the 18S marker. Remarkably, none of the six species was detected with the COI marker. A closer look on species biology marks the results obtained with 18S marker as suspicious. While the reference database used for the COI dataset was geographically limited to middle Europe, the here used 18S reference sequences retrieved from GenBank were not geographically bound. The available literature for the six insect species, detected with the 18S maker, reveal that all of them are only described from areas outside of Europe, which is indicating an accumulation of false positive results. This is highlighting that the use of markers with low taxonomic resolution in combination with unsuitable reference databases increases the risk of encompassing false positive results. Overall, arthropod reference databases for ribosomal markers are still very limited for most taxonomic groups (Clarke et al. 2014), which supports the use of mitochondrial markers for these taxa.

This study indicates that a multiplex approach can significantly increase number of detected target invertebrate taxa (Annelida, Arthropoda, Nematoda). Despite the fact that the COI region is unique among at least 95% of studied species of diverse groups of organisms including birds, insects, fish etc. (Hajibabaei et al. 2007) and is thus capable of delineating most species of these groups, it is not an all-purpose answer as it's taxonomic resolution and coverage is limited for many invertebrate taxa (e.g. Annelida and Nematoda) (Kvist 2014; Creer et al. 2016). Combining the COI marker with a ribosomal marker can strongly increase detection rate, as despite its low taxonomic resolution for many arthropod taxa it has been shown that the 18S marker is capable of identifying several nematodes on species level (Bhadury et al. 2006). The use of a multiplexed study design is further supported by the fact that so far no primer or single gene region has been identified that will amplify all taxa on an eDNA sample, hampering assessments of complete biodiversity levels. The use of several genes/ regions will come closer to having the

full biodiversity picture of a given habitat (Cowart et al. 2015; Drummond et al. 2015; Zhang et al. 2018; Marquina et al. 2019a). A recent study found that the combination of at least two markers can improve the taxonomic resolution by up to 10% (Marquina et al. 2019a). However, marker performance and resulting number of detected species is strongly depending on references databases. Especially for arthropods databases for ribosomal markers are very limited (Clarke 2014). However, databases are fast growing. Today, more than 2.5 million COI sequences (Porter and Hajibabaei 2018) and 1.4 million 18S sequences are inter alia stored in GenBank, making it to one of the largest repository of genetic data for biodiversity (Strasser 2008; Porter and Hajibabaei 2018) and thus allowing for the reliable identification of a broad range of taxa (Leray et al. 2019). With ongoing effort to complete databases which is nowadays supported by the application of new methods like genome skimming, and the detection of markers with high taxonomic resolution, the power of multiplex studies will further enfold and open the door for biodiversity studies aiming to reliable identify a wide range of taxa.

Influence of Extraction Method

Former studies have already shown that eDNA extraction method commonly used for soil samples can significantly influence taxa detection rate (Dopheide et al. 2019). Dopheide et al. (2019) compared community composition recovered from 1.5g of soil extracted with the MoBio Powersoil RNA extraction kit (MoBio Laboratories, Carlsbad, CA) with the community composition assessed from 15g of soil extracted with a phosphate buffer following the Taberlet et al. (2012) protocol. The protocol includes the use of the Nucleospin Soil DNA extraction Kit (Macherey Nagel GMBH & CO. KG) for purifying and eluting DNA. Prior to this study, the effectiveness of the MoBio Powersoil RNA extraction kit with the Nucleospin Soil DNA extraction kit (Macherey-Nagel GMBH & CO. KG) was compared. It was observed that the use of the Macherey Nagel Kit resulted in higher DNA yields. These findings are likely attributable to two

main factors. Firstly, the Macherey Nagel Kit can to some extent be adjusted to soil composition of the sample, secondly, the two extraction kits are differing in number of washing and purification steps. Depending on soil composition (organic content, inorganic matter, humic substances etc.) DNA yield and DNA purity can be strongly influenced by the buffer utilized. While with the MoBio kit only a single extraction buffer system is available, the Macherey Nagel Kit allows for the choice between two lysis buffers and also an enhancer (called SX). For soil samples under investigation, the combination of lysis Buffer SL2 and the enhancer SX clearly increased DNA yields in comparison to any other combination. Furthermore, the two kits work with several purification steps, which likely increases DNA purification and concentration of remaining PCR inhibitors. As differences in the protocol workflow and chemical composition of the commercial DNA extraction kits may have resulted in different PCR biases due to variations in persisting amounts of PCR inhibitor substances, the results of Dopheide et al. (2019) are questionable. It cannot be stated without doubt that the extraction of a larger amount of soil leads to an increase in arthropod species detection rate. To exclude this bias we here compared communities assessed with the phosphate buffer after Taberlet et. al (2012) and the commercial Macherey-Nagel kit.

DNA extraction with the commercial Macherey Nagel Kit resulted in the detection of 18339 COI OTUs and 13164 18S OTUs. When using the phosphate buffer for DNA extraction the resulting 18S dataset contained 13034 OTUs, which were slightly less compared to the number of OTUs retrieved when extraction was conducted with the Macherey Nagel kit. Similar findings were obtained for the COI dataset. Approximately 1000 OTUs less were recovered when extraction was conducted with the phosphate buffer (17329 OTUs) (figure III.6) These results are surprising as according to Dopheide et al. (2019) amount of starting material used for extraction is positively correlated with number of retrieved OTUs. Besides to amount of starting material the two extraction methods were targeting different kinds of eDNA. While the phosphate buffer exclusively allows for the extraction of extracellular DNA, extraction with

the Macherey Nagel Kit includes a lysis step, which enables the extraction of intracellular DNA next to extracellular DNA. The highest amount of intracellular DNA in soil usually originates from microbial organisms (Taberlet et al. 2012). As shown here and also described in the literature both markers are suffering from non-target amplification (Yu et al. 2012; Yang et al. 2014), a bias which is likely to be increased with the application of a lysis step as more microbial DNA is made available for extraction. Indeed, for both marker datasets an increase in number of detected microbial eukaryotes was found when DNA extraction was performed with the Macherey Nagel kit (figure III.6). Next to an increasing amount of available microbial DNA, the amplification of intracellular microbial DNA is further supported by the fact that intracellular DNA is usually of a higher quality compared to extracellular DNA, which is supporting binding of primers leading to an higher amplification success. Nevertheless, for Bacteria an inconclusive result was obtained when comparing number of retrieved OTUS depending on extraction method. While for the 18S marker an increase in detected bacterial OTUs was found when extraction was conducted with the Macherey-Nagel kit, contradicting results were obtained with the COI marker (figure III.6). This is surprising but might be best explained by the here used reference database which were not suitable for the identification of bacterial species. As already discussed, number of detected bacteria fall short of expectations and the here identified number of taxa does not represent truly existing bacterial diversity. To test for the hypothesis if an increased amount of bacterial DNA is influencing invertebrate species detection rate more specialized markers, for which well curated databases are available, should be used. However, because of the high number of unassigned reads, we cannot exclude the possibility that number of OTUs derived from bacterial reads might be strongly differing between extraction methods, especially as number of OTUs detected after extraction with the Macherey-Nagel kit exceeded the one obtained with the phosphate buffer (figure III.6). Furthermore, an increase in amount of bacterial DNA might have influenced metazoan species detection rate as with an increase in number of species

contributing to eDNA mixture the primer affinities will concomitantly change (Morinière et al. 2016; Elbrecht et al. 2017). The higher diverse and rich the samples, the lower the taxonomic coverage (Fonseca et al. 2012) presumably resulting in the assessment of different community composition.

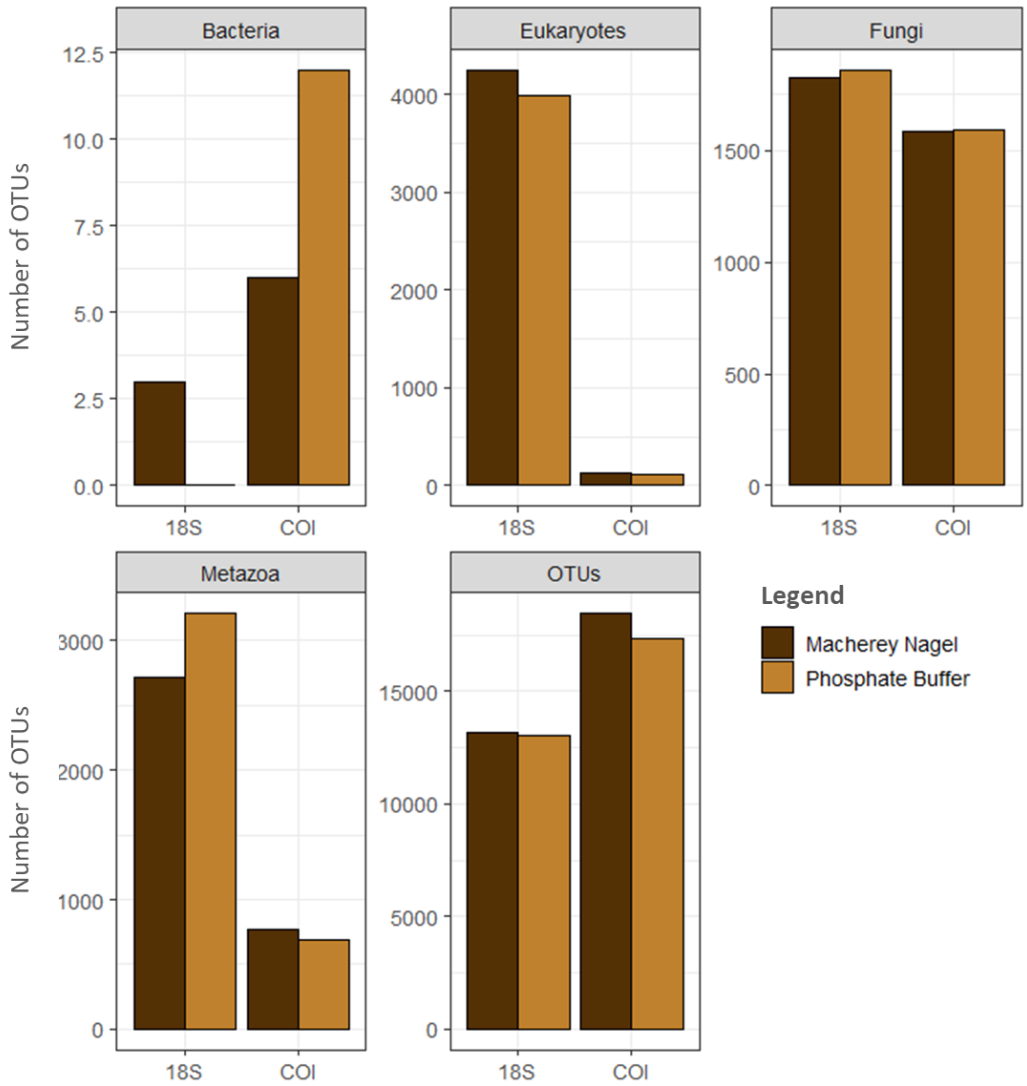


Figure III.6 Total number of OTUs and number of OTUs assigned (blastID ≥ 90%) to Bacteria, Eukaryotes (excluding Fungi and Metazoa), Fungi and Metazoa depending extraction methods and marker choice.

Due to differences in obtained number of OTUs depending on extraction method it is hypothesized that extraction method is significantly influencing assessed community composition. This was confirmed by analysis of variance (ANOVA) (COI: adonis: F_{323}=11.629, p<0.001; 18S: F_{323}=34.051, p<0.001) and subsequently visualized with a Principal Coordinate Analysis (PCoAs). Based on calculated presence-absence matrices for both marker datasets it was found that data obtained with each of the two extraction methods formed a distinct cluster (Figure III.7a,b). To further identify differences in terms of community composition between DNA extraction methods (beta diversity) the degree of sample dispersion homogeneity was tested (Figure III.7b). For the COI as well as for the 18S dataset betadisper-test indicated a very heterogeneous dispersion within samples of each extraction group (COI: F_1= 5.92, p=0.02; 18S: F_1=4.79, p<0.03). This suggests that even on small scales, large variation in species composition occur. This pattern has already been observed along and within marine coast transects of only 10 m apart using metabarcoding (Fonseca et al. 2014).

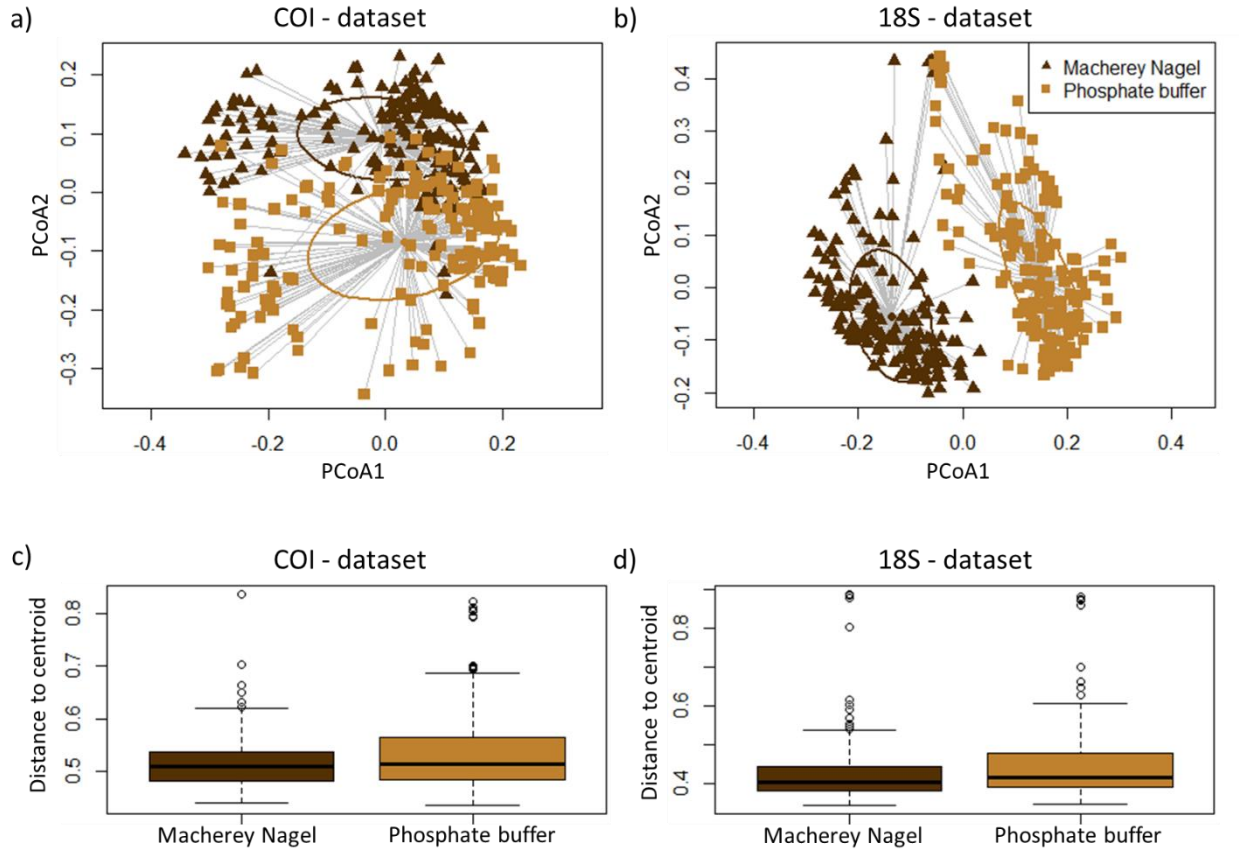


Figure III.7 PCoA plots indicating differences in assessed community composition for the (a) COI and (b) 18S dataset. Light brown coloration is indicating community composition assessed using the phosphate buffer, while community composition assessed with the Macherey Nagel kit is highlighted with a dark brown colored. The same coloration scheme was used for the group dispersion plots indicating homogeneity of assessed community composition within all soil samples extracted with the same extraction method for the (c) COI dataset and (d) 18S dataset.

When focusing on the two target groups Annelida and Arthropoda no definite trend was observed indicating that one extraction methods is superior over the other in terms of revealed degree of diversity. Moreover, number of assigned OTUs with either of the two extraction methods was closely linked to choice of marker. Amplification of the COI marker

resulted in a higher number of OTUs assigned to Arthropoda and Annelida respectively when the Macherey Nagel kit was used for DNA extraction (figure III.8a). The opposite was found with the 18S dataset for which number of OTUs assigned to Annelida and Arthropoda was higher when the phosphate buffer was used for DNA extraction (figure III.8b). As already outpointed is the taxonomic resolution of the COI marker limited for the phyla Annelida and Nematoda. Additionally, it has been shown that although the COI marker is often described as the currently best choice for the identification of Metazoa (Andújar et al. 2018) it still fails to amplify some groups of arthropods (Marquina et al. 2019a), especially when these arthropods are contained in a bulk samples (Brandon-Mong et al. 2015). A bias which is probably enhanced when working with eDNA. Due to the much higher complexity of processed DNA mixture shifts in primer affinities occur. The fact that marker choice was influencing number of OTUs assigned to Arthropoda and Annelida depending on choice of extraction method is supporting these findings. Due to its low variability in primer sites, amplification success of the 18S marker might be less influenced by complexity of DNA mixture as primer affinities are more similar for many invertebrate taxa. However, an increase in amount of starting material is likely to correlate with variability in contained DNA, hampering COI amplification success. Additionally, the lysis step implemented into the extraction with the Macherey Nagel kit could have significantly contributed to the above described findings. Lysis led to an increase in amount of DNA derived from transient species. This intracellular DNA is of much better quality, supporting amplification. Furthermore, it is likely that amount of DNA released from transients invertebrate species due lysis, clearly exceeded the one found as traces in the soil. The amount of DNA with which a species contributes to DNA mixture is significantly influencing its detection rate (Elbrecht et al. 2017). It is therefore likely, that the higher number of annelid and arthropod species recovered with the COI marker when extraction was based on the Macherey Nagel kit is directly contributed to the application of the lysis buffer and to changes in occurring primer and amplification bias.

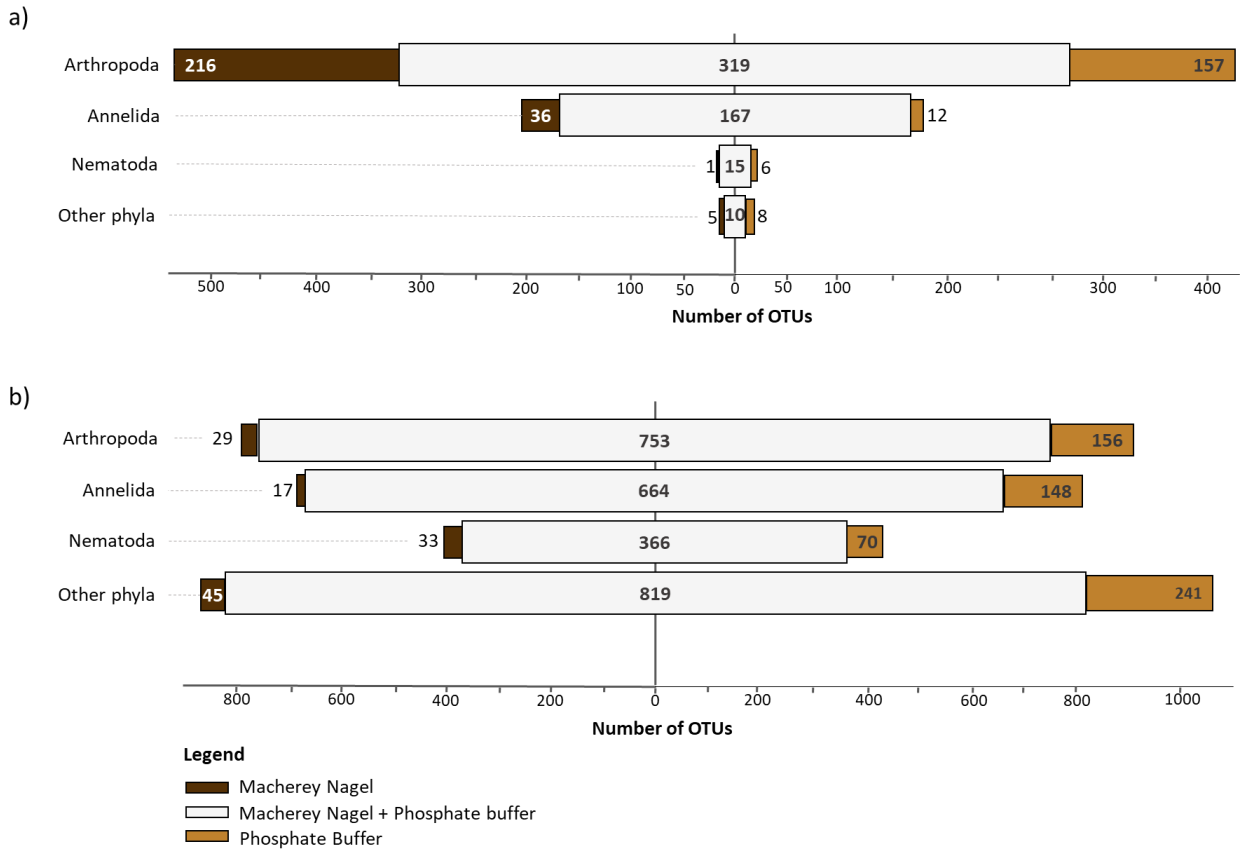


Figure III.8 Number of assigned OTUs (BlastID ≥ 90%) to the three target groups Annelida, Arthropoda and Nematoda depending in extraction method based on the a) COI marker and b) 18S marker.

However, while the 18S primer is known to have a rather poor taxonomic resolution whereas many OTUs are often not assigned on species level or are assigned to the same species, it is surprising that out of the 692 arthropod OTUs recovered with the COI marker only 177 were assigned on species level (figure III.8a). This either suggests that (a) a high percentage of ground dwelling arthropods has not yet been barcoded or (b) that several OTUs are artifacts which could be the result of a high number of sequencing artifacts or degradation.

Former studies have already stated, that most commercial kits are not suitable for the assessment of arthropod community composition (Dopheide et al. 2019). The majority of commercial kits are designed for the assessment of microbial communities which is thought to require the extraction of a much smaller amount of soil (Taberlet et al. 2012; Dopheide et al. 2019). It is often argued that macro-organism distribution patterns, size and population density is strongly differing from the ones of microorganisms (Taberlet et al. 2012; Dopheide et al. 2019), whereas an increase in amount of starting material could likely increase detection probability. Some studies have already proven this hypothesis by showing that number of detected macroorganisms increased with amount of source material used for DNA extraction (Dopheide et al. 2019). Our study is only partly supporting these findings. Indeed, on basis of the 18S marker higher number of arthropod and annelid species were retrieved when DNA was extracted from a larger amount of soil (figure III.9ab). However, this positive relationship was not found for the COI dataset (figure III.9ab). In detail, the two extraction methods recovered exactly the same four annelid species when the COI marker was amplified (figure III.9b). These findings might be strongly affected by the low taxonomic resolution of the COI marker for earthworms (Kvist 2014). Former studies have shown that annelids can reach abundances of up to 134 000 specimens per m^2 (Coleman et al. 2004). Additionally, detection probability of annelids is further increased by the fact that several Enchytraeids produce a high amount of fecal pellets. It has been shown that these pellets can make up to 29% of the volume of the higher A- horizon in a Scottish grassland soil (Davidson et al. 2002). Although only 0.5g of soil were used for the extraction with the Macherey Nagel Kit, the possibility of capturing Enchytraeid DNA was therfore exceptional high. Additionally, the absence of a COI barcoding gap may have led to the assignment of sequences originating from several species to the same OTU, resulting in a strong underestimation of total existing annelid diversity.

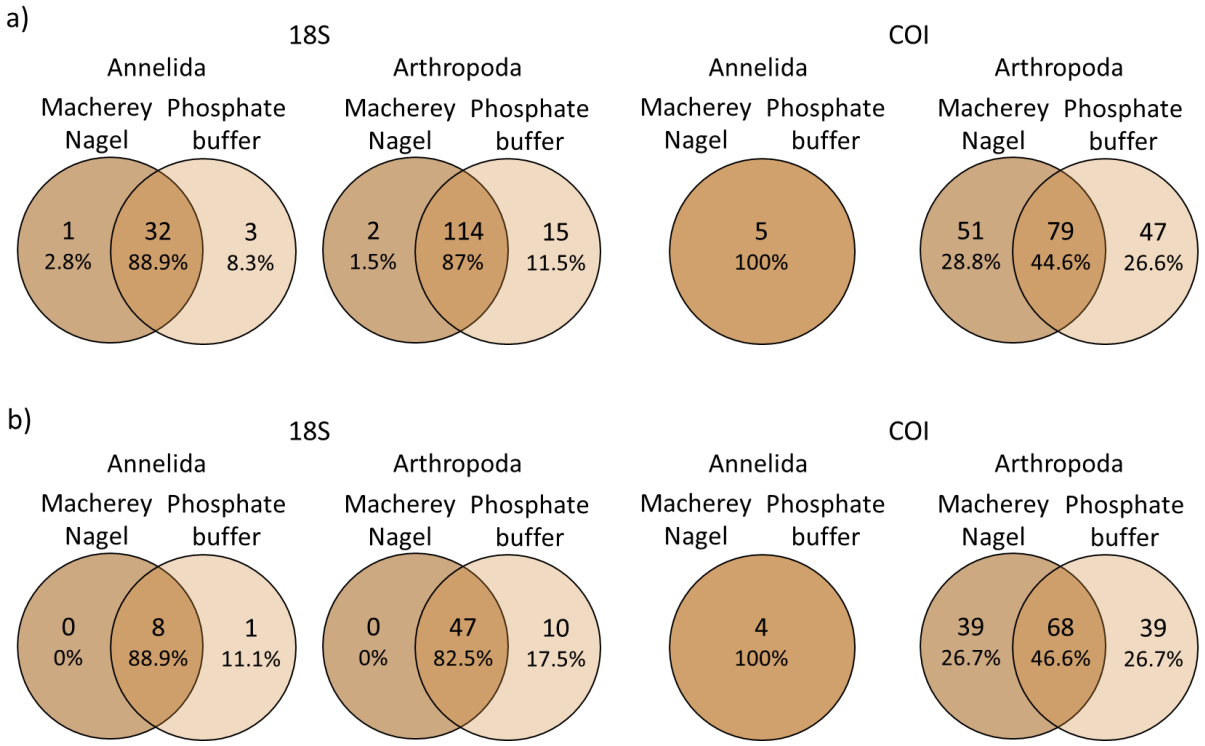


Figure III.9 Venn diagram showing number of shared and unique species between extraction methods depending on choice of marker. Figure III.9a is based on a BlastID of at least 90%, while figure III.9b) required a BlastID of at least 99%.

Using the COI marker both extraction methods resulted in the detection of similar numbers of arthropod species based on a blastID of at least 99% (Figure III.9b). Overall, 146 arthropod species were identified. Out of them, 39 species were exclusively found by each of the two extraction methods. The remaining 68 were recovered with both methods (figure III.9b). This is indicating that the two extraction methods are targeting different arthropod communities. It can be expected that several species were exclusively detected by the phosphate buffer due to the larger amount of soil used for extraction (Taberlet et al. 2012; Dopheide et al. 2019). Furthermore, differences in detected species composition are inherent to the specifies of the methods, one targeting extracellular and intracellular DNA, with cell lysis disruption and the

other only targeting extracellular eDNA. The application of the lysis step made DNA contained in living cells (e.g. eggs, larvae and other dormant lifeforms) available for DNA extraction. A closer look on order level of the arthropod class Insecta further supports this hypothesis. During the summer season when insect activity is higher, extraction with the Macherey Nagel indicated a peak in diversity of the order Diptera (figure III.10).

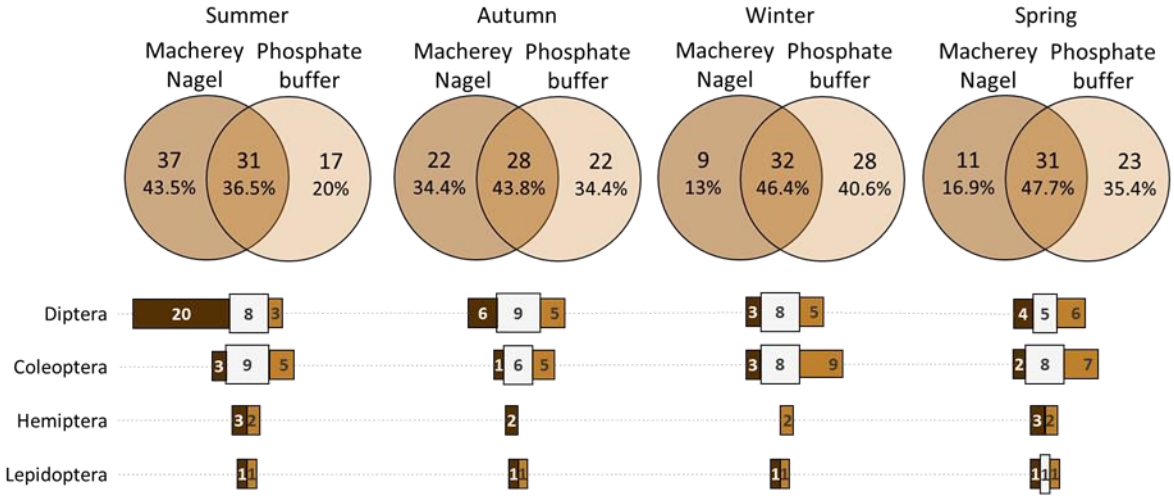


Figure III.10 Number of shared and unique detected arthropod species (blastID ≥ 99%) between extraction methods depending on season. Detailed number of species per insect order recovered with either one or both of the two extraction methods are shown.

In particular, the assessed community composition of the nematoceran families Sciaridae, Mycetophilidae and Limoniidae as well as the brachyceran family Phoridae showed a higher degree of diversity when DNA extraction was conducted with the Macherery Nagel kit (figure III.11). The families Sciaridae and Mycetophilidae both belong to the Sciaroidae (Barnard 2011), a superfamily which encompass a high number of species which larvae feeds on either fungal fruiting bodies or mycelia in dead wood and soil litter (Jakovlev 2012). Especially the

larvae of the family Sciaridae are known to be usually associated with soil litter, where they feed on plant roots (Binns 1981). The third nematoceran family for which an increase in diversity was observed when the Macherey Nagel kit was used for DNA extraction was the Liimonidae. Although the majority of species belonging to this family are known to develop in aquatic habitats, next to them this family also encompass species known to develop in drier soils (Reusch and Oosterbroek 1997). Indeed, this has been described for three of the here found four Liimonidae genera (Chilotrichia, Dicranomyia and Limonia). Next to the three above described nematoceran families, an increase in diversity was also found for the braychceran family Phoridae when extracting DNA with the Macherey Nagel kit. Similar to described nematoceran taxa is the larvae of Phoridae commonly found in forest soils (Disney 2012). This accumulation of in the soil residing species supports the hypothesis that the Macherey Nagel kit and the phosphate buffer are targeting different arthropod community. As already discussed enables the application of a lysis step the extraction of DNA from small, at the time of sampling living organisms including transient species (e.g. eggs and pupae). It is likely that lysis resulted in the release of a high amount of intracellular DNA originated from these living organisms. Due to this increase in amount of high quality DNA, amplification of these species is supported. While extraction with the Macherey Nagel kit is therefore supporting insect detection rate, the opposite is likely the case for the extraction with the phosphate buffer. As only DNA traces are recovered amplification probability is strongly decreased. Additionally, the richness and diversity of the eDNA mixture could have changed primer affinity. In fact, the amplification success of hymenopteran DNA depends on composition and complexity of bulk eDNA samples (Brandon-Mong et al. 2015). High abundant species with higher affinities (close related) will likely capture more primer molecules during PCR, while species with lower affinities will yield lower level amplicons and fewer reads (Hajibabaei et al. 2011; Brandon-Mong et al. 2015). When larger amounts of soil are extracted, the DNA mixture becomes more complex. As a result, changes in primer affinities may occur, whereas amplification of DNA of some species is now hampered as

other species have a better primer affinity. Furthermore, eDNA mixtures extracted from soil are known to be dominated by non-target species, likely occurring in higher abundances (e.g. bacteria and fungi) (primer-template mismatches that lead to taxonomic detection bias).

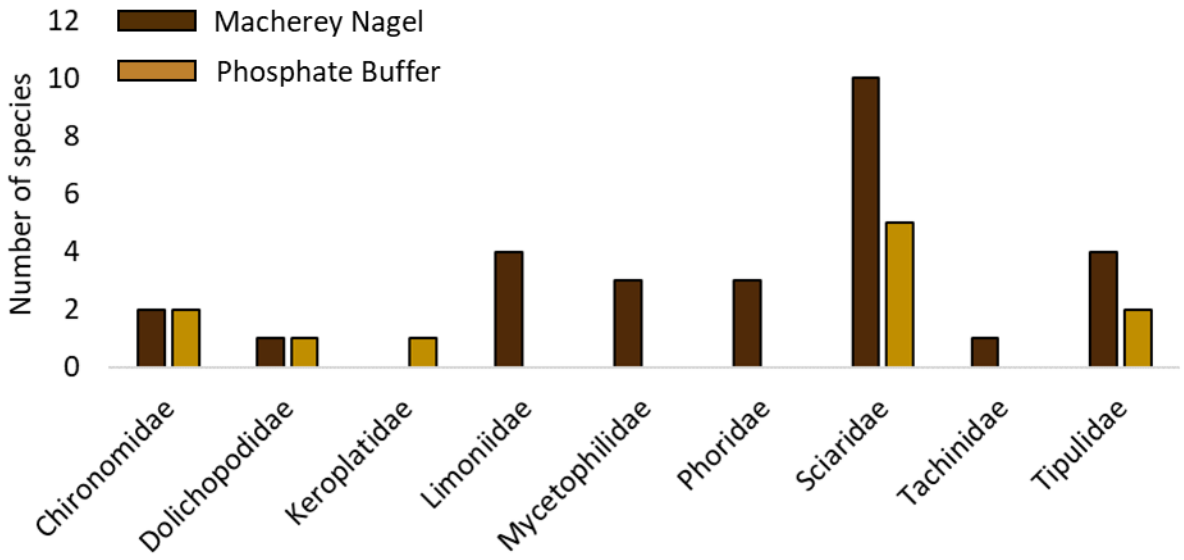


Figure III.11 Number of detected dipteran species (BlastID 99%) per family in summer season depending on extraction method

When comparing arthropod species composition recovered from each of the three biological replicates taken per forest type and season, the phosphate buffer dataset uncovered a more homogenous species distribution across taken biological replicates. Depending on season between 33.32% and 62.9% of arthropod species retrieved with the Macherey Nagel kit were exclusively detected in one of the three biological replicates (figure III.12a), while with the phosphate buffer between 30% and 45.92% of arthropod species were found in only one of the three soil samples taken at each sampling site (figure III.12b). This indicates that arthropod

diversity identified with the phosphate buffer is more homogenous across samples what corroborates the idea of when using larger amounts of soil for DNA extraction, it will increase the chances to assess a more complete picture of existing arthropods diversity.

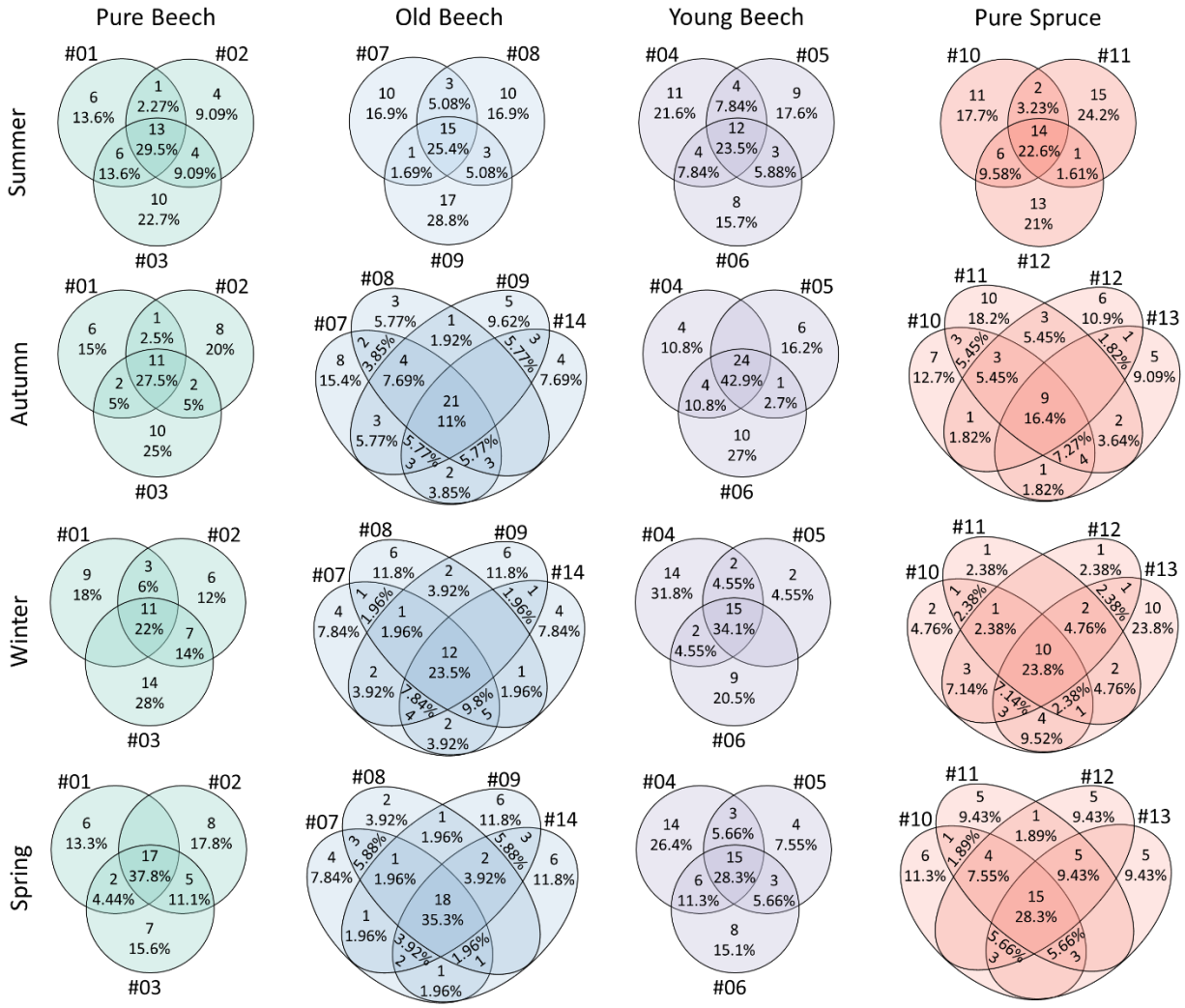


Figure III.12a Number of unique and shared species between sampling sites (cardinals) depending on forest type (columns) and season (rows) using the Macherey-Nagel kit. The data shown here comprises all arthropod species detected with either one or both of the two used

Figure III.12a (Continued.) markers (18S and COI). Only species detected with a BlastID of at least 99% to the reference databases are considered.

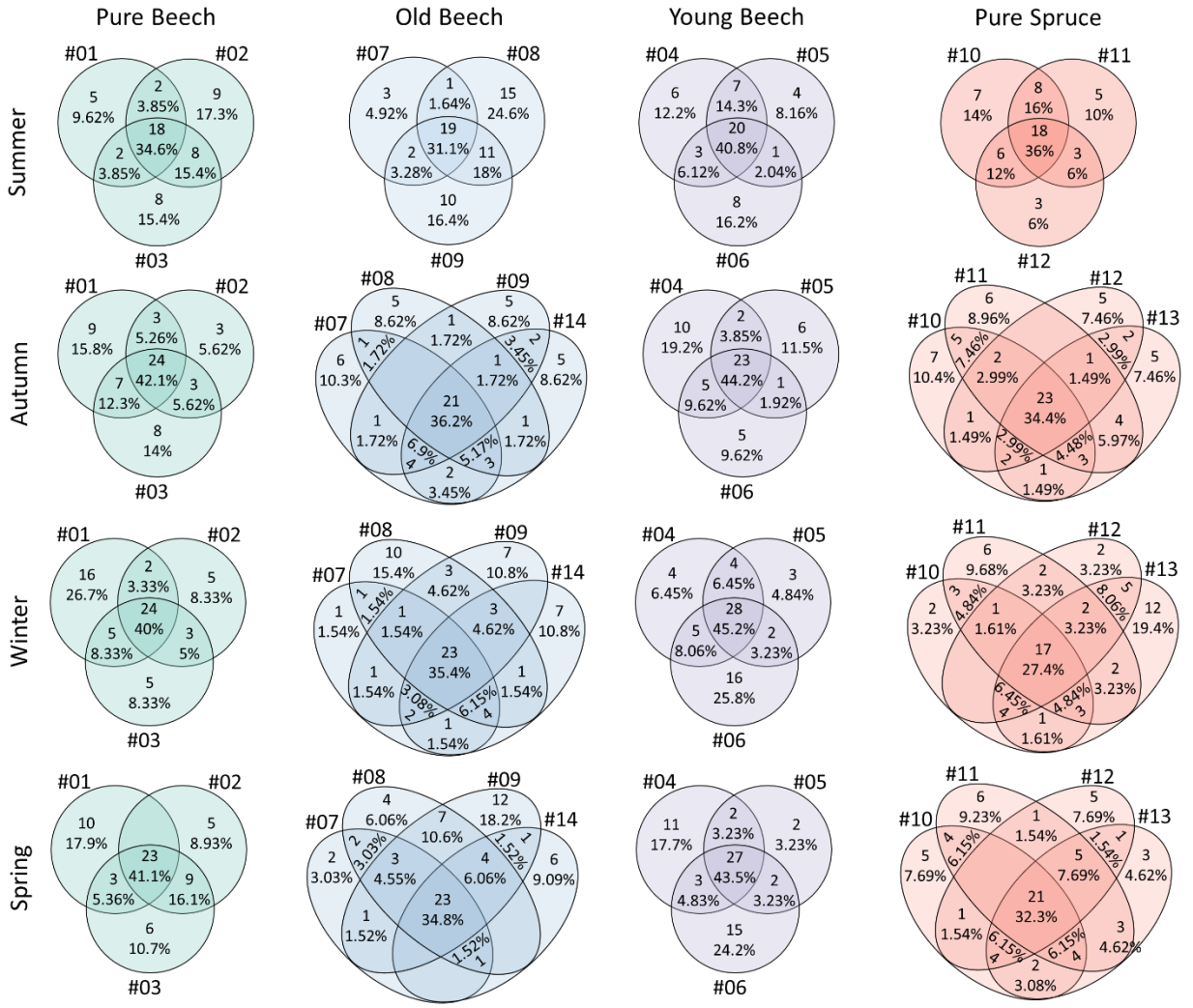


Figure III.12b Number of unique and shared species between sampling sites (cardinals) depending on forest type (columns) and season (rows) based on the extraction with the phosphate buffer. The data shown here comprises all arthropod species detected with either one or both of the two used markers (18S and COI). Only species detected with a blastID of at least 99% to the reference databases are considered.

However, the relative high percentages of species exclusively recovered from one sampling site across both extraction methods indicates that sampling strategy was not adequately scaled and that a higher number of samples is required to assess total exiting arthropod diversity. Rarefaction curves with extrapolation (figure III.13a) and sample completeness curves (figure III.13b) are supporting this hypothesis. Sample completeness curves show that the here taken 162 soil samples resulted in a slightly lower sample coverage when extracted with the Macherey Nagel kit. This is indicating that more samples are needed to be extracted with the Macherey Nagel kit compared to the phosphate buffer in order to reach a similar high coverage. Nonetheless, rarefaction curves clearly indicate that the use of a single extraction method is not suitable for the assessment of total existing arthropod diversity. While individual extrapolation of the Macherey Nagel kit dataset and the phosphate buffer dataset respectively calculated a maximum of approximately 215 arthropod species when 750 samples are taken, the combined dataset reached a significantly higher arthropod diversity of 270 species at the same sampling depth. This underlines our hypothesis that two distinct arthropod communities were assessed by the two extraction methods and furthermore allows for the statement that a combination of both methods could significantly increase species detection rate.

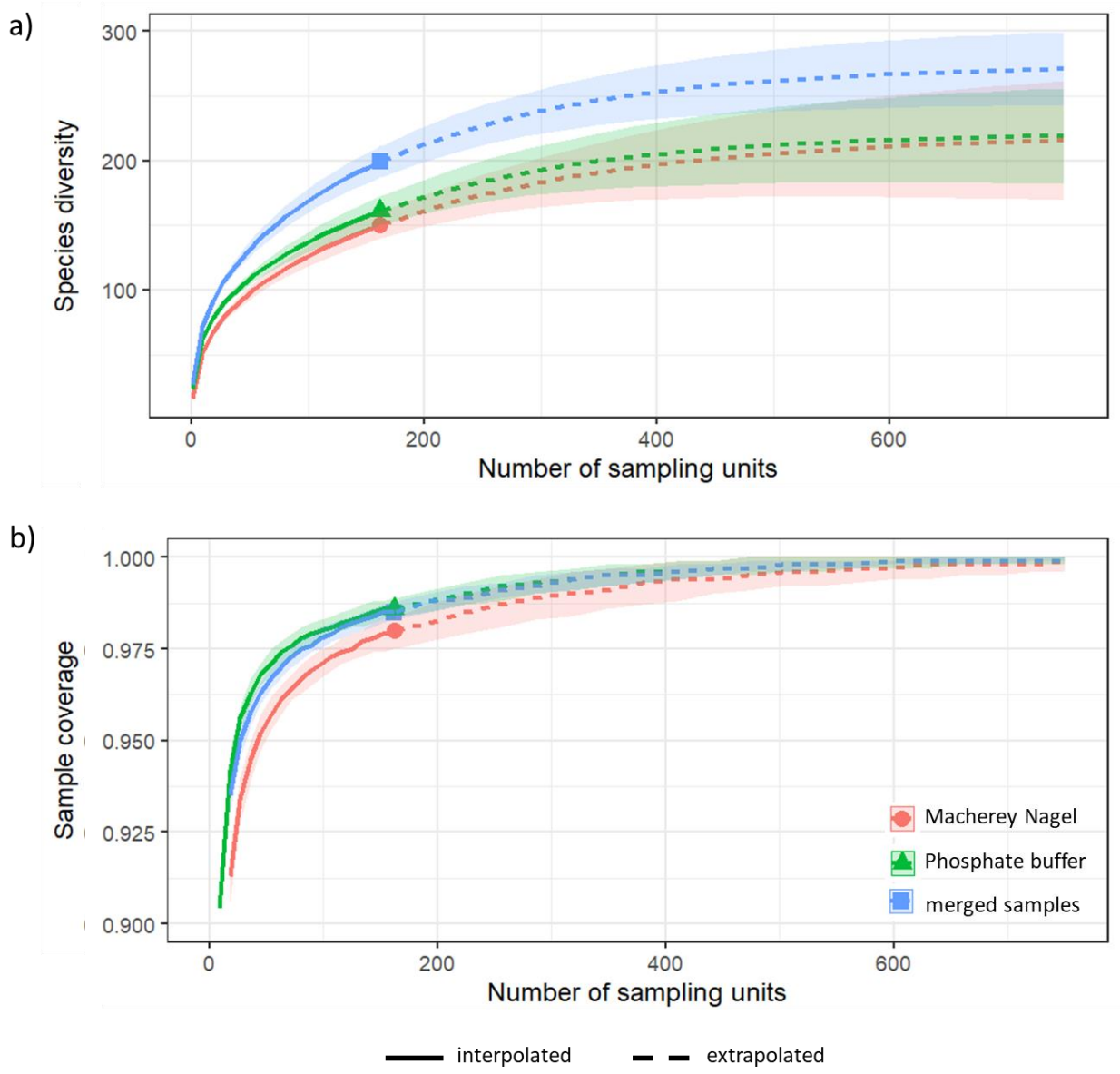


Figure III.13 Rarefaction curves (solid line) with extrapolation (dashed line) showing number of detected arthropod species depending on number of soil samples taken with each extraction method (a). Additionally sample completeness curves were calculated showing sample coverage depending on number of soil samples (b).

Here it was shown that the choice of marker had a significant influence on retrieved species composition (figure III.4). Additionally, it was observed that the two extraction methods revealed two distinct invertebrate communities. When aiming to assess complete species composition the use of the Macherey- Nagel kit can support the detection of dormant and non-active species. However, this study is also supporting the hypothesis that extracting DNA from larger amounts of soil is enhancing the probability to detect macro-organism with large distribution patterns but small population densities (Taberlet et al. 2012; Dopheide et al. 2019). It should be kept in mind, that the more complex the extracted DNA mixture becomes, the higher the resulting primer bias. To assess a picture of existing species diversity as complete as possible including species in dormant life stages the use of a commercial DNA kit, which includes a lysis step in combination with a sampling strategy of great extent, which comprises the sampling of several hundreds of samples, should be given priority over the use of a phosphate buffer. However, if a study is not interested in the assessment of transient species (e.g. eggs, pupae), the phosphate buffer is a less labor intensive and more cost efficient choice, which provides a good overview of existing species community.

Influence of Forest Type on Soil Composition

Forest stands covered with different tree species generate a divergence in soil properties and may thereby influence soil microbial communities (Chandra et al. 2016) which could concomitantly drive mesofauna diversity as well. In this study, it was observed that soil composition differed significantly between the four forest types (figure III.14 and figure III.15). This study compared the five soil components gravel, sand and mud, which comprises inter alia silt and clay between the four forest types. It was observed that the amount of mud was

significantly different between forest types (Anova: F_3=46.25, p<0.001) (figure III.14). Tukey post-hoc analysis revealed that the pure beech stands are characterized by a significant (p<0.001) lower amount of mud compared to the other three forest types (old beech – pure beech (16.68, 95%-CI[11.75, 21.63]); young beech – pure beech (22.34, 95%-CI[17.13, 27.54]) pure spruce – pure beech (15.82, 95%-CI[10.88, 20.76])). Furthermore, the forest types were characterized by a significantly different amount of gravel (Kruskal-Wallis test: Chi square=87.47, p<0.001, df=3) (figure III.14). Subsequent conducted post-hoc test (Tukey HSD) showed that the amount of gravel was significantly higher (p<0.001) at the pure beech sites compared to the other three stands (old beech – pure beech (-20.54, 95%-CI[-24.99, -16.08]) young beech – pure beech (-24.45, 95%-CI[-29.14, -19.75]) pure spruce – pure beech (-15.14, 95%-CI[-19.59, -10.69])). Additionally, soils of the pure spruce stands contained a higher amount of gravel than the two mixed stands (pure spruce – old beech (5.40, 95%-CI[1.20, 9.60], p=0.005), pure spruce – young beech (9.31, 95%-CI[4.85, 13.76], p<0.001)) (figure III.14). These results underline observations in the field that the O- and A-horizon of the pure beech sites were strongly reduced. The O-horizon consists of dead plant debris e.g. leafs and twigs. The A-horizon is located directly under the O-horizon and it is characterized by a mixture of organic matter and minerals (Orgiazzi et al. 2016). Underneath the A-horizon is the B-horizon which is dominated by clay and iron minerals (Orgiazzi et al. 2016). However, also smaller amounts of organic material can be found here which usually get washed into the layer by rainwater (Orgiazzi et al. 2016). It is known that thickness of the O-horizon can dramatically differ between forest types (Scheu et al. 2003). Generally, it is known that litter of spruce stands tend to accumulate at the forest floor in thick layers (Nihlgård 1971; Mardulyn et al. 1993; Albers et al. 2004; Berger and Berger 2012). This is inter alia attributed to morphological properties of needles and leafs which are directly influencing texture, compaction, density and the environmental characteristics of the litter layer (Paluch and Gruba 2012). Litter input in pure spruce stands is significantly lower compared to pure beech stands (Scheu et al. 2003).

Furthermore, spruce needles decompose more quickly than beech leaves (Albers et al. 2004). However, a major difference between spruce needles and beech leaves is the higher content in polyphenols (Gallet and Lebreton 1995) known to delay decomposition processes, by controlling N-dynamics which is the limiting factor for microbial activity (Albers et al. 2004). As a result, polyphenols have been described to be a driving force for the formation of thick organic layers, which further influences soil abiotic parameters such as pH-value. Abiotic conditions prevailing in different forest stands have a direct influence on local decomposition rates. Interestingly, the degree of influence of prevailing conditions on decomposition rate is differing between spruce needles and litter of deciduous tree species (Albers et al. 2004).

However, regardless of thickness of the O-horizon the first 10cm of the forest floor were sampled. The applied sampling was designed to sample the A and O-horizon, but variation in relative proportions of horizons contributing to sample might have occurred between samples. Because of different amounts of soil originating from each strata, ratio of minerals and organic content within samples might have been shifted and resulted in variations in composition of the mud fraction across forets types. Mud consists of silt (4-63µm) and clay (< 4µm). The amount of silt differed significantly between forest types (Kruskal-Wallis rank sum test: Chi square=80.70, $p<0.001$, df=3). Conducted post-hoc test showed significant differences between each of the mixed stands and the monoculturous stands (young beech-pure beech: paired-wilcoxon-test, $p<0.001$; young beech-pure spruce: $p=0.001$; old beech-pure beech: $p<0.001$; old beech-pure spruce: $p=0.024$, table 3). Furthermore, soils at the pure beech and pure spruce stands, differed significantly in amount of silt ($p<0.001$).

Table III.3 Paired-Wilcoxon-test results for comparison of amount of silt [%] between sampling sites.

	pure beech	old beech	young beech	pure spruce
pure beech	1	>0.001***	>0.001***	>0.001***
old beech	>0.001***	1	0.227	0.024*
young beech	>0.001***	0.227	1	0.001**
pure spruce	>0.001***	0.024*	0.001**	1

Additionally the relative percentage of clay found at each sampling site differed between forest types (Kruskal-Wallis rank sum test: Chi^2=24.851, p>0.001). Conducted post hoc-test showed significant differences between the two monocultures (paired-sampled Wilcoxon test: p<0.001), the two mixed stands (paired-samples Wilcoxon test: p<0.001) as well as between pure spruce and old beech stands (paired-sampled Wilcoxon test: p=0.017), and pure beech and young beech stands (paired-sampled Wilcoxon test: p<0.001). No differences were found between the young beech and pure spruce stands, nor between the old beech and pure beech sites. (table III.4, figure III.15).

Table III.4 Paired-Wilcoxon-test results for comparison of amount of clay [%] between sampling sites.

	pure beech	old beech	young beech	pure spruce
pure beech	1	0.838	>0.001***	>0.001***
old beech	0.839	1	>0.001***	0.017*
young beech	>0.001***	>0.001***	1	0073
pure spruce	>0.001***	0.017*	0.073	1

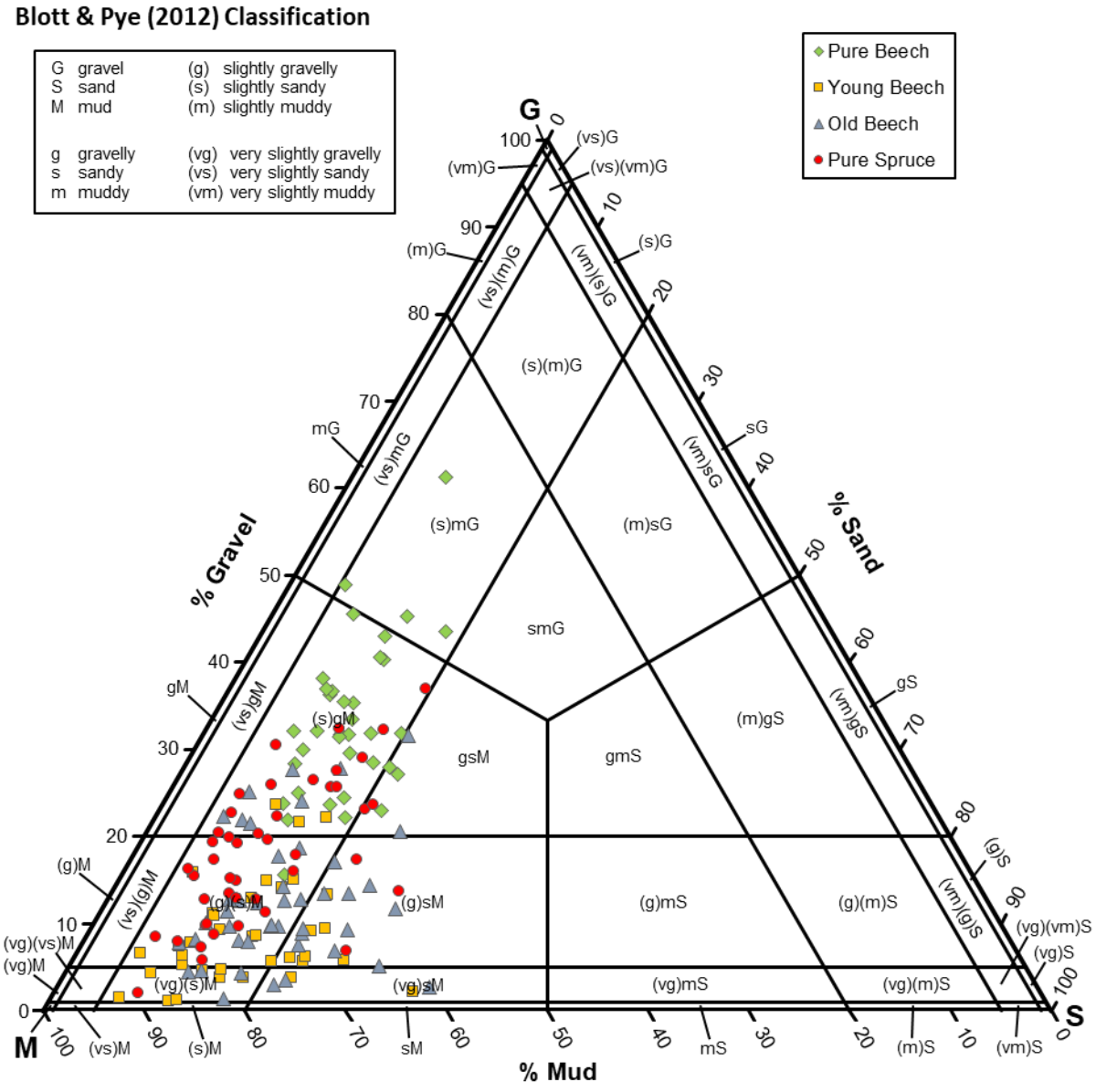


Figure III.14 Classification of soil texture after Blott and Pye (2012) depending on forest type. Considered parameters were proportion of gravel, sand and mud

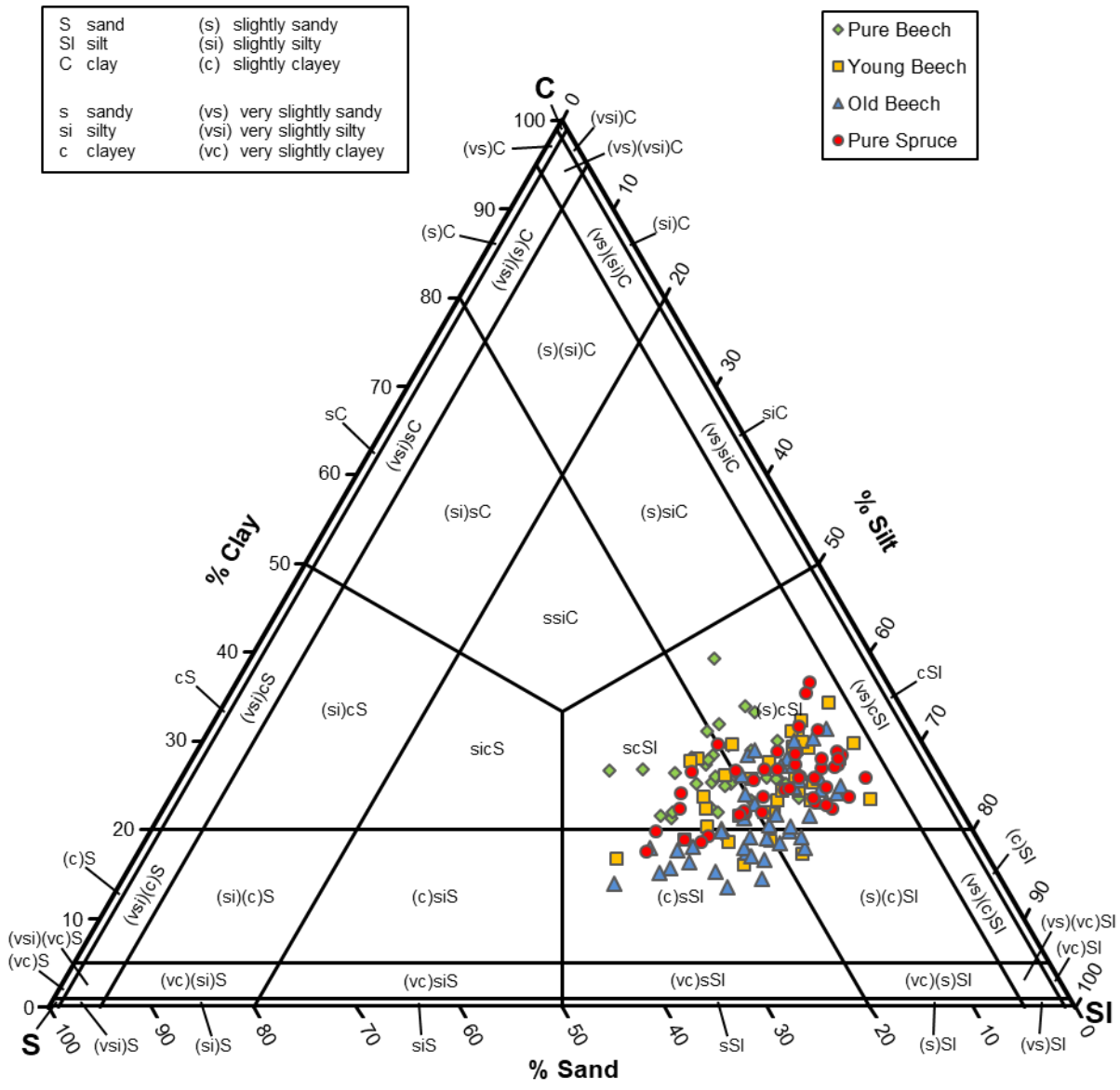


Figure III.15 Classification of soil texture after Blott and Pye (2012) depending on forest type. Considered parameters were proportion of sand and silt and clay.

Former studies showed that soil texture is often closely linked to local dominating tree species (Dutta and Agrawal 2002). Although usually soil texture is thought to be a limiting factor for faunal species distribution it has been observed that trees can alter soil texture (relative amount of silt and clay) to a certain extent. This study is partly supporting these findings. The old beech and pure beech stands showed similar amount of clay, so does the pure spruce and young beech stands. This indicates that with ongoing renaturation process, clay content slowly shifts. However, tree species are differing in their ability to modify the same (Dutta and Agrawal 2002). Dutta and Agrawal (2002) discussed that this is mainly attributed to the fact that the local flora is maintaining nutrient regeneration by addition of various sources of organic matter, altering abiotic and biotic conditions. An increase in organic matter enhances N-mineralization leading to an increase of plant available nutrients (Dutta and Agrawal 2002) supporting the establishment of a diverse herb and undershrub layer and thereby soil-nutrient cycling. Additionally, with increasing plant cover porosity of soils decreases. Porosity is a main driver for fragmentation, redistribution and aggregation of particles in soils (Dutta and Agrawal 2002). However, the pure beech and old beech stands were strongly differing in development of a herb and shrub layer. This is also reflected in total organic content (TOC)-Nitrogen (N) ratio (TOC:N). The TOC:N is a value giving information about the humus form as it is closely linked to variation in tree species composition (Lovett et al. 2002; Albers et al. 2004). Conducted Kruskal-Wallis rank sum test showed significant differences in TOC/N-ratio between forest types (Chi^2=49.81, $p<0.001$) (figure III.16). Subsequently conducted paired-sampled Wilcoxon test showed significant differences between all four samples sites.

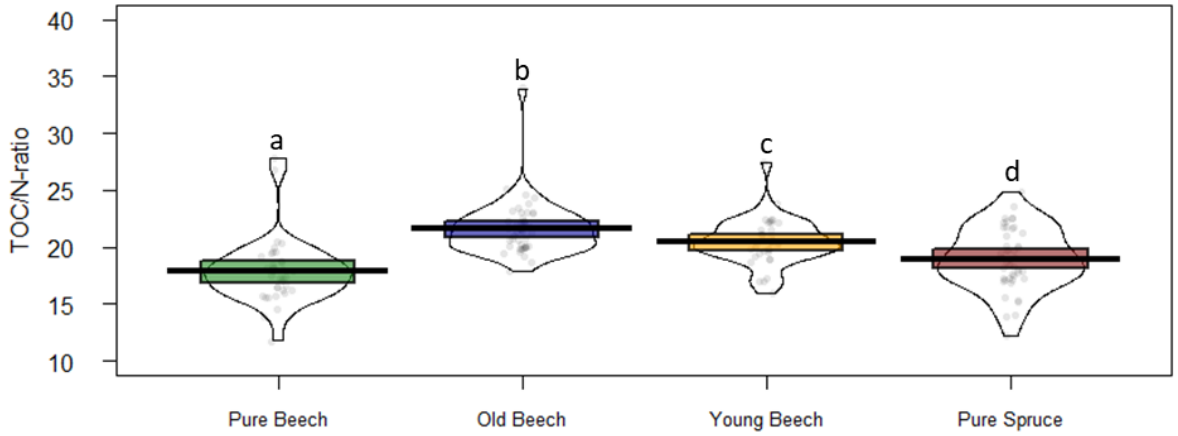


Figure III.16 Total organic content - Nitrogen ratio depending on forest type. The TOC:N ratio allows for conclusions about changes in humus form between forest types.

Followed by differences in soil physical appearance and humus type significant differences in chemical composition were found. Results of chemical soil composition analysis showed that content of nitrogen (Kruskal-Wallis-test: Chi square =43.71, $p<0.001$, df=3), carbon (Kruskal-Wallis-test: Chi square=51.89, $p<0.001$, df=3), hydrogen (Kruskal-Wallis-test: Chi square=43.71, $p=0.002$, df=3), organic nitrogen (Kruskal-Wallis-test: Chi square=39.04, $p<0.001$, df=3), carbon carbonate (Kruskal-Wallis-test: Chi square=37.07, $p<0.001$, df=3) and total organic carbon (Kruskal-Wallis-test: Chi square= 51.501, $p<0.001$, df=3) were significantly different between forest types (figure III.17). Post-hoc tests (Mann-Whitney-U-Test) showed that the old beech stands were differing from the three other forest types in views of carbon, nitrogen, hydrogen, total organic carbon and carbon carbonate content. Furthermore and as already mentioned organic nitrogen content of the two mixed stands was significantly different from the one found at the two monocultures but was not significantly differing between mixed stands. Additionally, the young beech stands were significantly different from the remaining three stands in terms of nitrogen-, carbon-, hydrogen- and total organic content. The two monocultures were not differing in views of chemical soil composition.

Several studies have shown that tree species have a significant influence on C and N content in forest soils (Nihlgård 1971; Vesterdal et al. 2013; Cremer et al. 2016). For beech and spruce monocultures usually a different amount of N content is documented (Nihlgård 1971; Vesterdal et al. 2013; Cremer et al. 2016). The data presented here are not supporting these findings. However, these result are in agreement with further studies (Hagen-Thorn et al. 2004; Trum et al. 2011). It has been shown, that these contradicting results arose from sampling depth. The classification of soil horizons is corresponding to successive stages of decomposition within the different layers (Wardle 1993). Vesterdal et al. (2008) observed that the influence on tree species highly depended on sampled soil layer. While species with a lesser content of Carbon (C) and Nitrogen (N) in the forest floor, had usually a higher content of C and N in the associated mineral soil than species with a higher C and N content in the forest floor. As a result, when combining forest floor and mineral soil this opposite trend may lead to an offset in differences in forest floor C and N resulting in similar contents for several tree species. Next to Vesterdal et al. further studies have observed similar trends (Finzi et al. 1998; Hagen-Thorn et al. 2004; Oostra et al. 2006).

However, a strong increase in carbon and nitrogen content at the two spruce-beech mixed stands (figure III.17) was observed. Former studies have already shown that spruce needles and beech leafs tend to decompose more quickly in beech or spruce-beech mixed stands (Albers et al. 2004) compared to pure spruce stands. Additionally, it has been shown that mixing the litter of the two tree species leads to an accelerated decomposition process (Sariyildiz et al. 2005). Both facts are mainly a result of different microbial activities. The accumulation of spruce needles lead to a retardation of decomposition processes as it goes along with a decreasing pH values. Humus formation encompasses the binding of base cations within organic residues which can then not be used for nutrient transformation (Albers et al. 2004). The more base cations are captured the higher the concentration of protons the rhizosphere plants exchange for base cation (Albers et al. 2004). The low pH value in

combination with a high polyphenol content are likely the main drivers for the retardation of decomposition processes which goes along with an decrease in microbial biomass (Albers et al. 2004). In the mixed stands decomposition is accelerated, although it does not meet the one at pure beech stands. This is mainly due to the fact that abiotic conditions are still differing between mixed and beech monoculture stands as well as the fact that microbial communities of spruce stands seem to be less responsive to environmental conditions (Albers et al. 2004). However, the continuous input of leaf litter due to the presence of spruce trees results in an increased amount of N and C. It can be assumed that the combination of both factors leads to a higher amount of C and N at the mixed stands compared to the two monocultures.

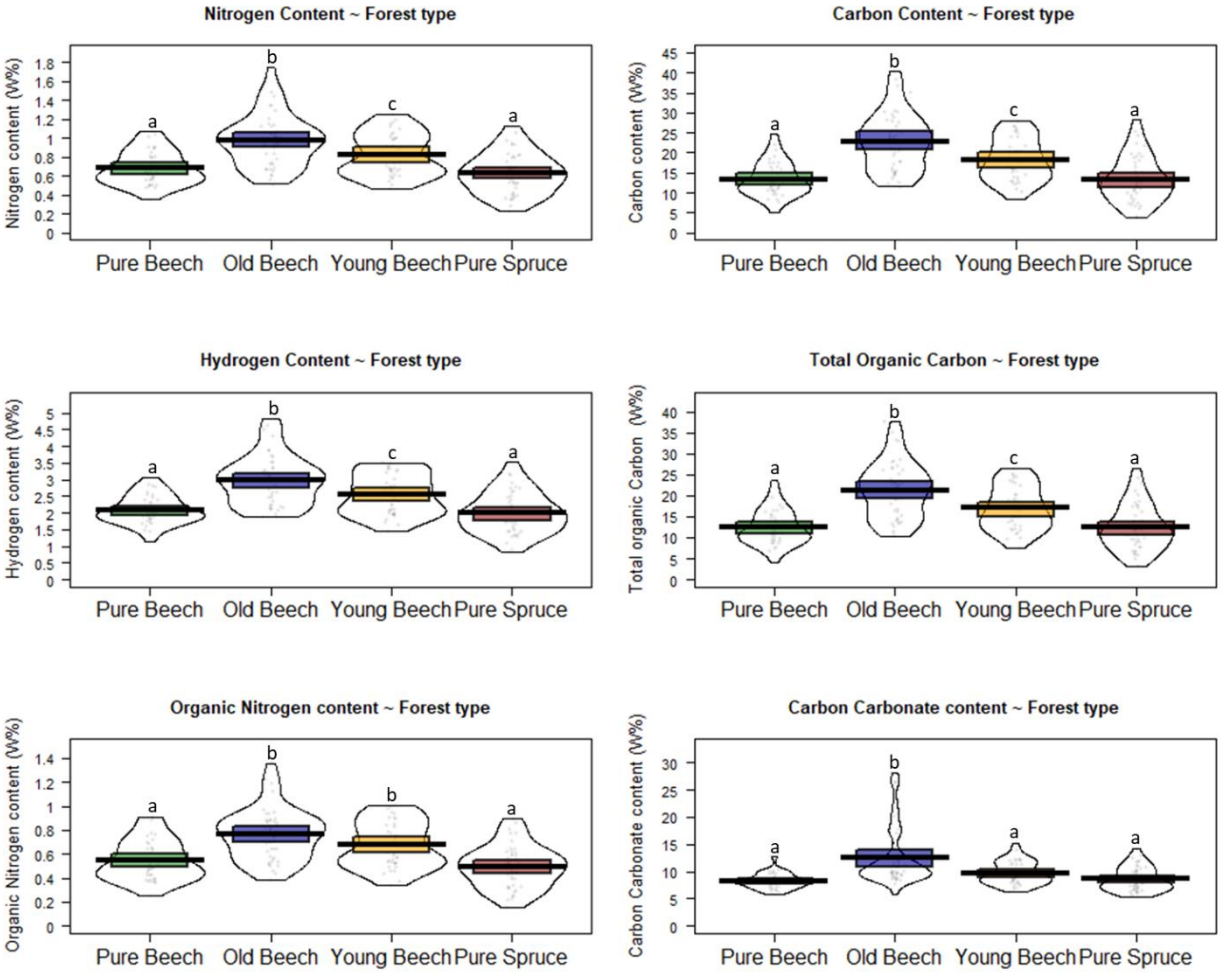


Figure III.17 Differences in soil chemical composition between the four tested forest types. While the two monoculture were not differing from each other in tested chemical composition. The mean calculated nitrogen, carbon, hydrogen, total organic carbon and organic nitrogen content was significantly higher at the two mixed forests, reaching a peak at the old beech stands.

Influence of Soil Composition on Assessed Species Communities

The fate of eDNA in soils may plays a role for the reconstruction of species communities. We have already discussed that the dominating tree species had a direct influence on soils physical and chemical composition. The fade of extracellular DNA in the environment depends on a complex interplay of chemical, biological and physical factors, which determines times of persistence and degree of degradation (Pietramellara et al. 2009). After being released into the environment, extracellular DNA is prone to degradation by enzymatic nucleases. However, by binding to inorganic and organic surface-reactive particles such as clay, sand, silt and humic substances, phosphate and cation bonds are formed (Taberlet et al. 2018). This is a curse and savior at the same time. Binding of extracellular DNA to environmental particles enhances the time DNA can persist in the soil (Ogram et al. 1988; Paget et al. 1998; Demaneche et al. 2001; Levy-Booth et al. 2007) but hampers DNA extraction (Zinger et al. 2016). However several studies suggested that extracellular DNA in soil is degrading rapidly (Sirois and Buckley 2019). Nevertheless, the number of binding sites are varying widely between soil particles. It has been shown, that clay particles are capable to bind more DNA than sand (Levy-Booth et al. 2007). Sandy loam soils consisting of 6.2% of clay adsorbed 40% more DNA compared to sandy soils with a relative amount of 0.1% of clay (Blum et al. 1997). In views to this study an increased amount of silt and clay may have resulted in incremented DNA binding, protecting DNA from nuclease mediated enzymatic hydrolisation (Crecchio et al. 2005; Levy-Booth et al. 2007) and by adsorbing nucleases and DNAses, reducing the potential for enzymatic DNA restriction (Blum et al. 1997; Levy-Booth et al. 2007). As a result, it cannot be excluded that the higher amount of clay at the pure beech and young beech sites may have resulted in a lower degree of DNA degradation leading to a more exact picture of the existing diversity. For the sake of completeness, it should be mentioned that not only soil composition is influencing the fade of extracellular DNA. Moreover, pH value and cation concentration influences DNA

binding capacity of soil. However, because of a defect-sampling device, measured pH-value could not be included into analysis.

Influence of Forest Type on Annelid Community

The dominating tree species has a major influence on soil chemical and physical properties (Dutta and Agrawal 2002), which again is directly driving annelid diversity and abundances. However, as each tree species is affecting local abiotic and biotic conditions and thereby soil properties to a different degree, investigations of its influence on existing annelid diversity are challenging.

Several studies have assessed annelid diversity of different forests and have emphasized that annelid diversity in deciduous forests is surpassing the one found in boreal stands (Reich et al. 2005; Schwarz et al. 2015). The results presented here are contradicting these findings. The Conducted PCoA on basis of all OTUs assigned to Annelida (presence-absence matrix of the multiplexed dataset) indicate no significant differences in annelid community composition between forest types (figure III.18). It cannot be excluded that these findings are strongly distorted by the already outlined disadvantages of the here used markers for assessing annelid diversity. As shown and confirmed by the literature (Kvist 2014; Lejzerowicz et al. 2015), none of the markers was capable of retrieving total annelid species composition. In detail, only 12 species were reliably detected with a blastID of at least 99% (figure III.5), what is indicating an underestimation of existing annelid diversity. Next to basing analysis on number of recovered species, several studies are trying to assess total biodiversity by using number of OTUs as proxy for species to avoid false negative detections due to incomplete databases. However, it is questionable if total existing annelid diversity is reflected by the number of the here observed OTUs. The total number of OTUs assigned to Annelida with a blastID of at least 90% was 768 OTUs, which is in terms of described annelid diversity for the here investigated forest types

exceptional high. On species level it was observed that up to 159 OTUs were assigned to the same species, resulting in the detection of the smaller number of 40 species. Out of the 159 OTUs which were assigned to *Achaeta unibulba*, six OTUs had a blast ID of at least 99% percent. It is hypothesized that the high number of OTUs per species is a result of (a) incomplete databases, leading to false negative results or (b) because of the high intraspecific variability within the group of Annelida (Kvist 2014) members of the same species were clustered into different OTUs, artificially increasing OTU count.

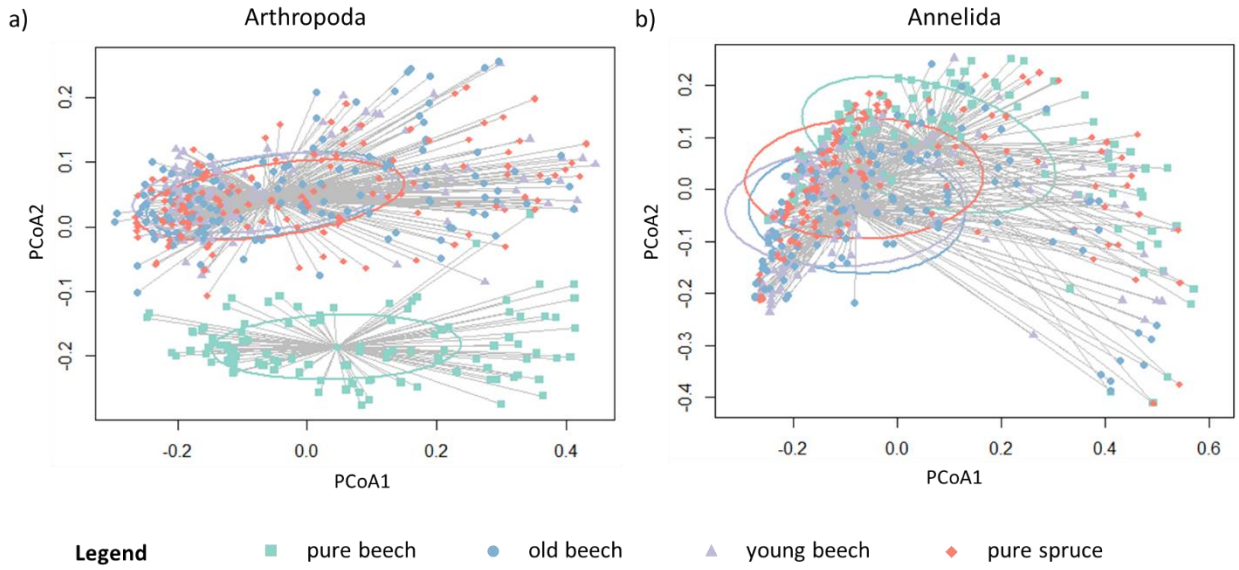


Figure III.18 PCoA plot indicating differences in assessed a) arthropod and b) annelid communities depending on forest type

Here, no influence of soil physical properties on annelid diversity was found. In detail no correlation between annelid OTU abundances or species richness and soil texture (particle size) was observed. This is surprising as further studies have shown, that annelid species differ

in their preferred habitat due to differences in soil texture (Baker et al. 1998; Ivask et al. 2008). While some species prefer sandy habitats, some prefer loamy soils (Beylich and Graefe 2009). It has been discussed that this is mainly a result of the influence of soil composition on water storage capacity and aeration by differences in texture (Beylich and Graefe 2009). While coarse textured soils have large pores, facilitating aeration, fine textures soils are characterized by small particles, which allow only for small pores, hampering aeration. After a heavy rain shower when pores are filled up with water, coarse textured soils like soils dominated by sands are rapidly draining, while drainage of fine soil e.g. soils dominated by clay can take up to several days. Temperature and soil moisture are the main driver of annelid biology and life processes (Gerard 1967; Phillipson et al. 1976). Many annelids living in temperate areas have a temperature preference of 10 to 20 degrees with relatively low upper lethal temperatures of 25 to 35 degrees (Curry 2004). In temperate regions, high temperatures are usually closely linked to moisture shortage. However, former studies have discussed that temperature is likely not the driving force of seasonal earthworm mortality but moisture stress (Gerard 1967; Phillipson et al. 1976). It has been suggested that population density of annelids is strongly driven by moisture (McCredie et al. 1992). The reasons for a lacking correlation between annelid diversity and soil texture can be manifold. Next to methodological issues affecting study outcome (primer bias, fade of eDNA in soil, incomplete reference databases, uneven soil sampling, low number of samples) further abiotic and biotic parameters e.g. food preferences and food availability could have influenced annelid species composition associated, limiting the driving effect of soil texture. It has been observed that litter input is correlating with an increase in earthworm abundances (Curry 2004; Eisenhauer et al. 2009). Indeed, here an increase in number of OTUs assigned to Annelida was found at the pure beech sites in autumn season, when litter input was highest. It is likely that the high organic input in form of litter is possibly limiting the role of clay as storage system for organic content and moreover, is the accumulation of leafs on the forest floor presumably limiting evaporation extent. However, an

increase in annelid diversity was not found at the two mixed sites in autumn season, although an increase of organic input occurred. Although soil composition has been described to affect annelid community composition, several more parameters must be taken into account, including biological factors like existing plant functional groups (Deleporte 2001; Campana et al. 2002; Eisenhauer et al. 2009). The effect of flora on annelid community composition and age structure has been described to strongly vary between seasons. Eisenhauer et al. (2009) observed that especially surface feeding aneric earthworm abundance are highly influenced by plant productivity and high quality inputs of plant residues. To some extend this is supported by this study. Annelid species diversity based on a blastID of at least 99% was significantly differing with changes in TOC:N ratio in summer (spearman correlation test: $p=0.02$) and spring seasons (spearman correlation test: $p=0.005$). Number of OTUs assigned to Annelida found at the two mixed stands, clearly exceeded the number of OTUs found at the pure beech monocultures (figure III.19) where an undershrub or herb layer was widely absent. Only in autumn, when litter input dramatically increased at the pure beech sites an increase in number of OTUs was found. The fact that the highest diversity of annelids was found at the pure spruce sites in autumn and spring is surprising as a former study has observed that earthworm abundances are lower at spruce stands compared to beech forests (Reich et al. 2005). Additionally thinning of spruce forests can lead to an increase in earthworm numbers (Castin-Buchet and André 1998). Overall, spruces have a negative effect on earthworms richness (Schwarz et al. 2015). Similar to number of OTUs (blastID ≥ 90%), the highest number of identified annelid species (BlastID ≥ 99%) was found at the pure spruce forests. While at the pure beech and old beech sites 9 and respectively 10 species were found, 11 species were detected at the pure spruce and young beech sites. Although some observations of changes in annelid diversity are consistent with recent literature focusing on annelid biology the already discussed methodological issues might have distort the here presented diversity patterns.

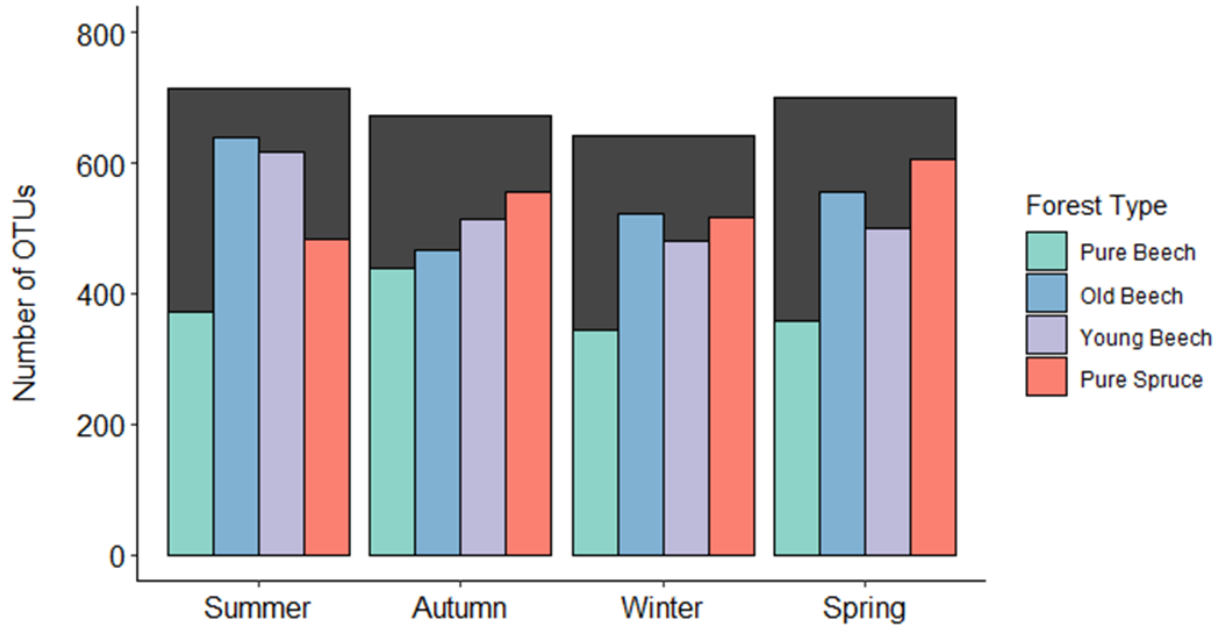


Figure III.19 Number of detected annelid OTUs (BlastID ≥ 90%) within the four sampled forest types depending on season

Influence of Forest type on Arthropod Community

Compared to extraction with the phosphate buffer resulted the use of the Macherey Nagel kit in the detection of a higher number of unique arthropod OTUs. This is likely contributed to an increase in number of detected transient species (e.g. eggs and pupae). Extraction with the phosphate buffer uncovered a more homogenous picture of existing biodiversity per forest type, here indicated by a reduced number of OTUs exclusively detected at one sample site per forest type and season as well as by a higher sample coverage reached with 162 soil samples

(figure III.13b). Nevertheless, the phosphate buffer has likely failed to detected a wide range of transient species, which were at the time of sampling not actively interacting with the soil habitat. Next to extraction method, also choice of marker significantly influenced number of recovered species. The 18S maker strongly underestimated the truly exiting insect diversity although several arthropod species were exclusively detected with this marker. This is highlighting the advantages of a multiplexed study. Therefore, the following ecological discussion on arthropod community patterns depending on landscape parameters is based on the species community data retrieved by both markers. Furthermore, only OTUs assigned on species level with a blast identity of at least 99% are taken into account.

In contrast to the assessed annelids community composition, differed arthropods community structure significantly between forest types (figure III.18b) (Adonis: F_3=14.419, p>0.001). For each forest type it was observed that a high number of arthropod species were exclusively recovered from one sampling site per season. In detail between 30.88% and 50% of all recovered species were unique to one sampling site (figure III.20). This suggests that the sampling strategy was not sufficient to assess total existing diversity.

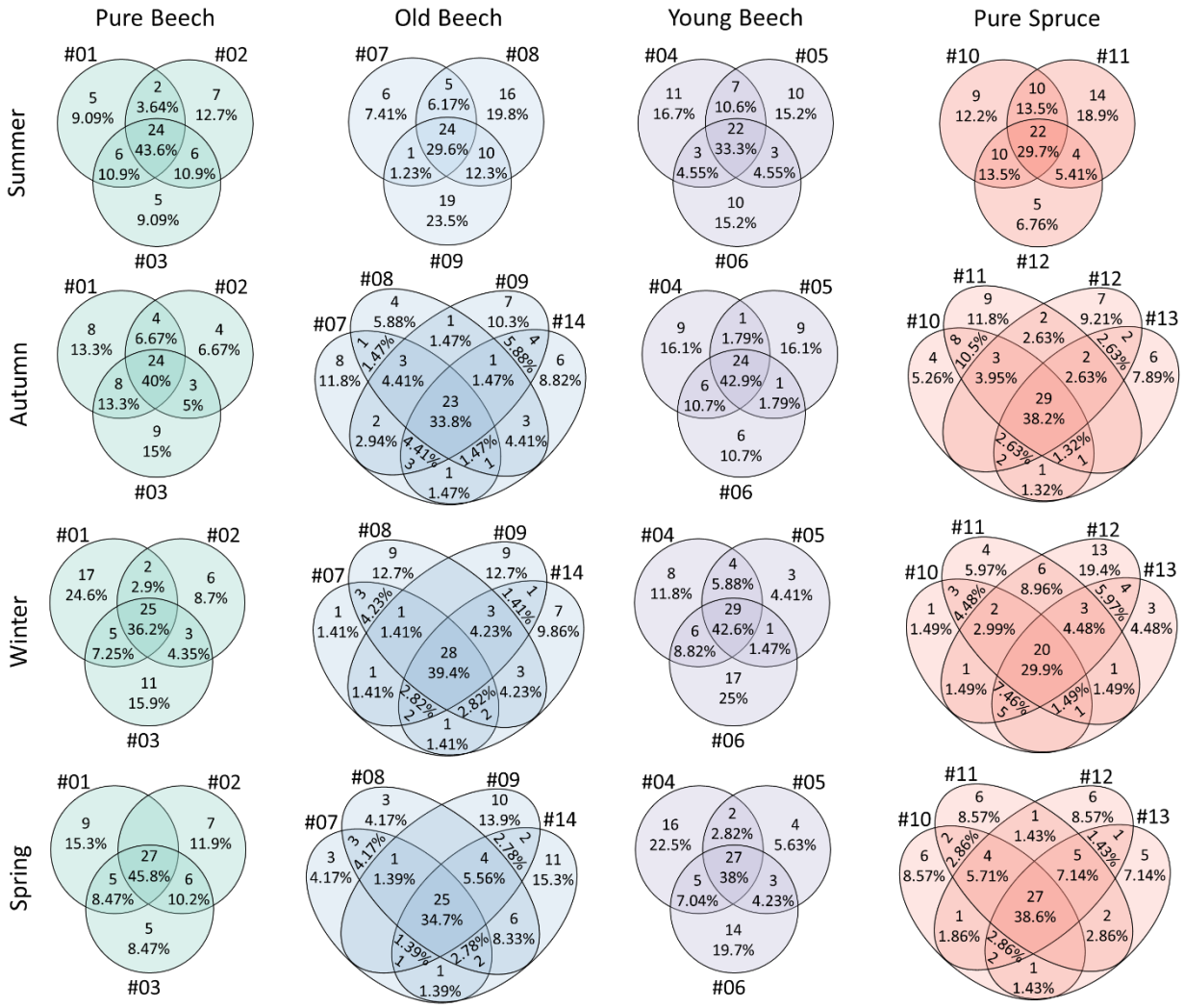


Figure III.20 Number of unique and shared arthropod species (blastID ≥ 99%) between sample sites of each forest type depending on season

Depending on species and the at the time of sampling expressed life stage, distribution patterns of arthropods can strongly vary. While many flying forms (e.g. Lepidoptera, Diptera or Hymenoptera) are capable to travel long distances, less mobile organisms (e.g. Chilopoda, Diplopoda or Malacostraca) are more likely to stay within a habitat. Additionally, species life

cycle and behavioral adaptations can directly influence detection rate. Overall is the soil matrix thought to be a rather poor integrator of the whole local biodiversity compared to aquatic ecosystems (Taberlet et al. 2018). Several studies have already shown that eDNA extracted from soil samples is poorly reflecting diversity of flying insects (Marquina et al. 2019b; Ritter et al. 2019). However, several flying organisms are also occasional inhabitants of the soil layer (Coleman et al. 2004). Depending on activity patterns and life cycle, species presence in the soil layer can be documented with molecular methods when sampling strategy is adequately scaled and timed. Because of the rapid degradation time of DNA in soils (Sirois and Buckley 2019) time frame for sampling is limited as many flying species are only occasionally interacting with the soil and time of occurrence is usually strongly correlated with species lifecycle. Additionally, extraction method must be appropriately chosen. The appearance of the hymenopteran species *Cephalcia lariciphila* is usually closely linked to the presence of coniferous forests as the larva feeds on the needles of the trees (Billany and Brown 1980). Four weeks after hatching the mature larva drops to the ground, where it quickly digs into the soil (EFSA Panel on Plant Health (PLH) et al. 2017). The larva builds an earth-walled chamber at a depth of 5-20 cm, where it finally pupates. During the dormant stage of the larvae, detection probability is enhanced when applied extraction method encompasses a lysis step. While *C. lariciphila* was captured with the Macherey Nagel kit at the pure spruce site in autumn and at the old beech sites in autumn and summer, extraction with the phosphate buffer resulted in the detection of the species in autumn at the pure spruce sites. The probability to detect *C. lariciphila* in soils sampled during autumn season is strongly enhanced, as it corresponds to the in the literature described time of the year when larvae drops to the ground in order to dig into the soil for pupation. The fact that the Macherey Nagel kit retrieved the species also in summer seasons underlines the already highlighted assumption that a lysis step enables the detection of transient species. However, it has been described that the larvae of *C. lariciphila* can remain in the soil for several years (EFSA Panel on Plant Health (PLH) et al. 2017), whereas its detection would have

been possible and likely for the full duration of the study. Again, methodological issues could explain for observed absence of the species during winter and spring season. As already mentioned, only a lysis step can enable the extraction of DNA from insects currently undergoing inactive life stages. This applies for the Macherey-Nagel kit. However the kit has been designed for the extraction of DNA from the low amount of 0.5g of soil. The likelihood of recapturing the same species again in a random sample of 0.5g of soil is rather low unless several biological replicates are taken or the species has a high population density.

Depending on arthropod class, number of species recovered from only a single or several forest stands were differing. While 42.6% of detected arachnid species were recovered from all four forest types, this applied for only 9.8% of observed insect species. Furthermore, almost 60% of all insect species were exclusively recovered from one forest type, but only 26% of the recovered arachnid species showed this distribution pattern. These differences are likely a direct effect of food preferences and life style of these two investigated arthropod classes.

The here detected arachnid species belong to a total of six orders. While the Mesostigmata, Ixodiden, Opiliones and Trombidiformes were represented by four, one, six and three species respectively, while significantly more species of the orders Araneae and Sarcoptiformes were found. Out of the 24 Sarcoptiformes species found, 23 were detected in at least two of the four sampled forest types. Only a single species was unique to one forest type. This applied for a much higher proportion of detected Araneae species. From 22 species, 9 were exclusively recovered from a single forest stands. Former studies showed that ground-dwelling spider communities can be significantly affected by the covering tree species (Schuldt et al. 2008; Ziesche and Roth 2008). To be exact, it's not the dominating tree species itself which is influencing spider assemblages, moreover the underneath the canopy established microhabitat mosaic defined by differences in litter type, ground vegetation, soil characteristics and microclimatic parameters like irradiation and humidity, which are not necessarily bound to the stand type itself is the major driver (Schuldt

et al. 2008; Ziesche and Roth 2008). Floral diversity is positively correlated with structural diversity, which in turn strongly influences spider community composition (Jocque 1973). The four sampled forest types were differing in its degree of stratification. While the two mixed stands were characterized by the presence of a strongly developed ground-, litter-, herb- and undershrub layer, the latter two layers were fairly absent at the pure beech sites. It is hypothesized that due to a limited amount of niches provided for possible prey species, less spiders were present in the habitat. Spiders are generalist predators which have already been observed to be limited in abundances by densities of prey organisms (Wise et al. 1999). However, the limited number of niches does not entirely explain the here presented findings. It is striking that only six Araneae species were found within the pure beech stands as spider assemblages of this habitat type have been described to be very diverse (Scharff et al. 2003). Overall, the fact that only one Araneae species was found at all four forest types indicate a high rate of false negative errors, likely resulting from methodological issues. Next to primer bias, the sampling method might not have been suitable. It has already been discussed that larger metazoans are more likely to be detected when a larger amount of soil is used for DNA extraction (Dopheide et al. 2019). This study showed that a high proportion of detected arthropods were exclusively recovered from one sampling site per forest type depending on season (figure III.21b). This leads to the hypothesis that sampling effort might has not been suitable to assess total exiting arthropod community. Suggesting that smaller organisms with a higher population density e.g. mites will have a higher detection rate. Former studies showed that number of mites can exceed number of spiders in the same habitat by hundred fold (Huhta and Koskenniemi 1975). Migge et al. (1998) showed that the average oribatid mite (Arachnida: Sarcoptiformes) density in European spruce and beech forests is as high as 160 000 ind/m^2 encompassing 68 species. Next to the higher population density are mites more actively interacting with the soil layer, which is likely increasing detection probability. Oribatid mites are known to be important components in the decomposer community (Huhta et al. 2012). Here a total of 28 mite species were found when

combining the data from all four forest types. Out of them, only two mite species were unique to one forest type. This supports the hypothesis that species population density is directly influencing species detection rate. However, in comparison to former studies is the total number of detected Acari species rather low, suggesting a high number of false negative result (Baulieu et al. 2019). However, in comparison to spiders are databases for mites fairly incomplete increasing the risk of incorporating false negative results. The here presented data suggests that Araneae distribution patterns are increasingly determined by local microhabitat mosaics, while mites are less influences by local abiotic and biotic conditions. nonetheless spiders are omnivore predators making this hypothesis highly unlikely, as it has been shown for arthropods that a higher degree of omnivory weakens the effects of plant diversity on species richness (Haddad et al. 2009; Scherber et al. 2010).

While spiders are omnivorous predators, insects are occupying several food niches and are thereby often showing a high degree of specialization (e.g. parasitoids, pollinators etc.). Within the arthropod class Insect, species from eight orders were identified, but number of species were not equally distributed between orders. Only one species each of the orders Hymenoptera, Orthoptera, Psocoptera and Thysanoptera were found. In contrast to that were number of detected species of the orders Lepidoptera, Hemiptera, Diptera and Coleoptera with 6, 9, 44 and 39 species significantly higher. The latter four orders, from here on referred to as highly diverse orders, had in common that a high number of species was exclusively recovered from one forest type (figure III.21a). Out of the 39 detected coleopteran species, 23 species were exclusively found in one forest stand. This accounts for 59% of all recovered coleopteran species. Similar but slightly lower proportions of unique species were observed for the remaining three highly diverse orders: 56% of the recovered dipteran species, 50% of lepidopteran species and 56% of all detected hemipteran species were unique to one habitat.

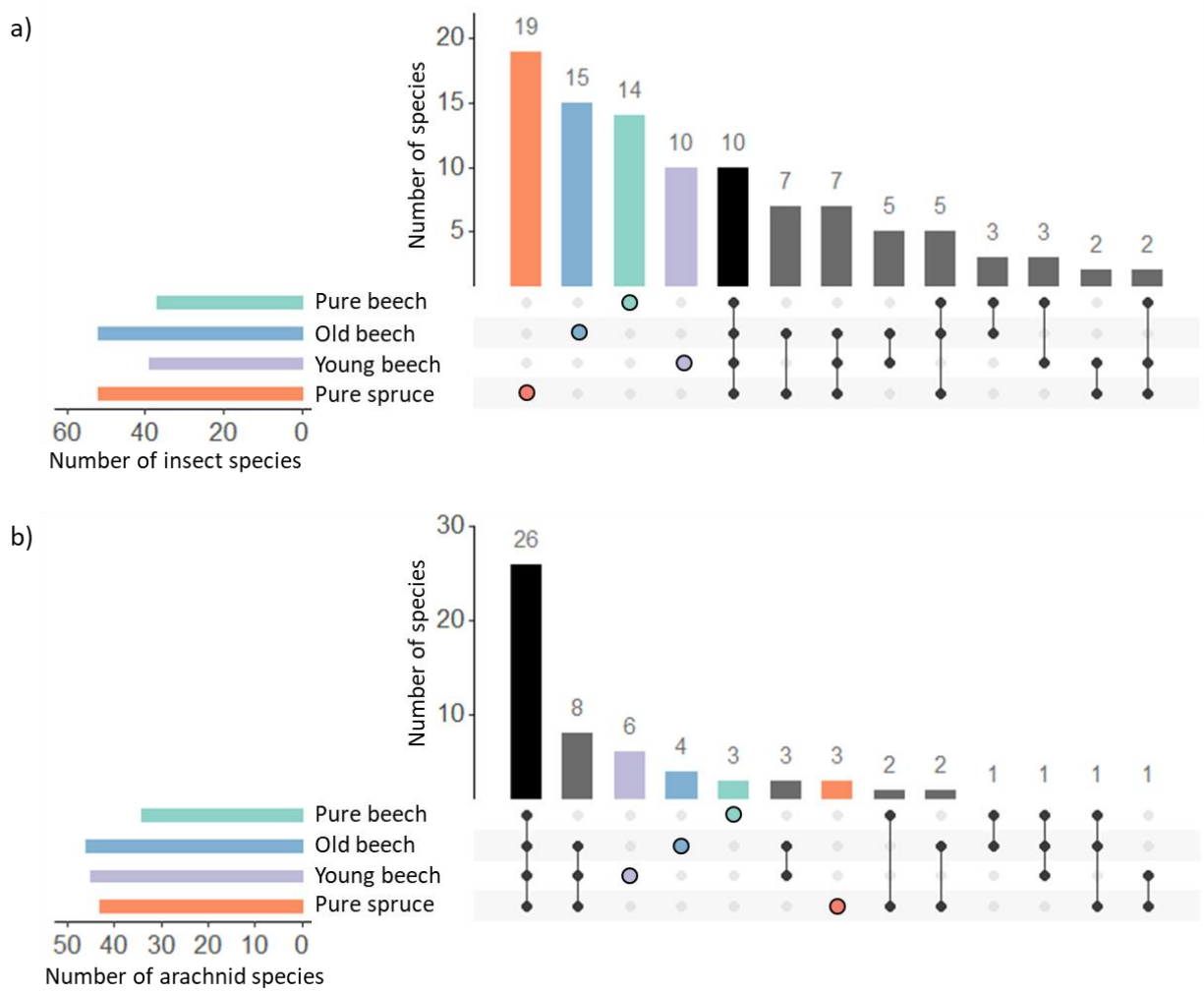


Figure III.21 Upset-R plot showing number of unique and shared a) Insect and b) arachnid species identified on basis of a blastID of at least 99% between forest types.

The high number of species only occurring within one forest type could suggest that either a high percentage of the here recovered insect species show a high degree of specialization or that the dataset is suffering from an accumulation of false negative results. The later hypothesis is supported by the fact, that species rarefaction curves indicate an insufficient sampling effort per forest type and season for assessing total existing biodiversity. (figure III.13).

As already discussed had the extraction method a significant influence on the number of detected species and thereby on assessed insect community composition (figure III.9, III.10, III.13). While for the beetles DNA extraction with the phosphate buffer resulted in a higher number of detected species, for the dipterans a higher number of species was found when using the Macherey-Nagel kit for DNA extraction (figure III.10). Comprising data from all seasons and forest types 17 coleopteran species were only detected when the phosphate buffer was used for DNA extraction. In contrast to that, revealed extraction with the Macherey Nagel kit the presence of five coleopteran species, which were left undetected by the phosphate buffer. For the dipterans, a contradicting picture was found. Extraction with the Macherey Nagel kit resulted in the detection of 18 species not found by the phosphate buffer, but the same method left seven species undetected. As discussed, is the phosphate buffer more suitable for the detection of larger metazoan species with low population densities, whereas the Macherey Nagel kit enables the detection of transient species but fails to detect organisms with a low population density. As a result, the number of dipteran species recovered with the Macherey-Nagel kit exceeded the one found with the phosphate buffer (figure III.22). A high percentage of them are members of families known to have larvae developing in the organic layer of the forest floor. Due to their small body size it is likely that the amount of DNA released by dipteran larvae into the soil is rather low compared to species with larger body sized (e.g. coleopterans). As a result, DNA originating from larger organisms is likely surpassing dipteran DNA in available extracellular DNA. For bulk samples, it has been shown that body size plays an important role for detection rate. Species with larger body sizes accounts for a significantly higher proportion of DNA in the DNA mixture. It has been shown that reads with higher abundances will most likely capture more primer molecules leading to a higher amplicon yields, while lower abundance reads may not yield any amplicons (Hajibabaei et al. 2011). Unfortunately, little is known about the release of eDNA into the environment and to what extent differences depending on body size, morphological characteristics but also species might occur. Therefore, it remains unclear to what extent the

above discussed concept is applicable for eDNA. However, when applying a lysis step proportion of DNA originated from transient species in extracted DNA mixture is likely to be increased. As a result a shift in proportions of DNA contributing to DNA mixture away from non-target microorganisms towards these bigger organisms occurs, subsequently enhancing detection probability of the latter.

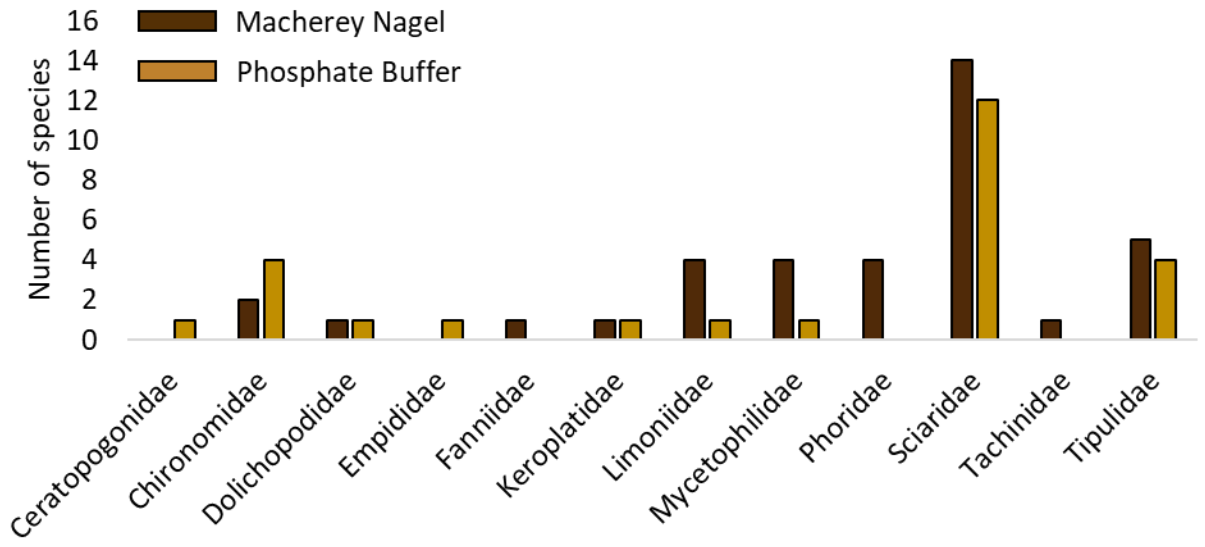


Figure III.22 Number of detected dipteran species (BlastID ≥ 99%) per family depending on extraction method.

A similar shift in proportion might be observed, when small species show a high population density as already described for Acari. Although number of dipteran species exclusively recovered with the Macherey-Nagel kit exceeded the one detected with the phosphate buffer, a high number of small sized Sciaridae species was recovered with both extraction methods. Sciaridae are known to dominate nematoceran larvae communities within forest soils (Schulz 1996; Elmer and Roth 2001). For several Sciaridae species it has been observed that up to 250 000 larvae can

accumulate on a very narrowed area (Menzel and Mohrig 2000; Menzel and Schulz 2007). The high population density of Sciaridae larvae is supporting detection probability with both extraction methods as it goes along with an increase of proportion with which the species contributes to extracted eDNA mixture.

In contrast to nematoceran larvae are population densities of coleopteran larvae usually lower. This is intera alia attributed to the fact that the larvae of several beetle families are predatory and even show cannibalism (Tschinkel 1981; Osawa 1989; Currie et al. 1996; Dodds et al. 2001). As a result detection probability was much lower with the Macherey-Nagel kit compared to the phosphate buffer. Furthermore, many dipteran species only occasionally inhabit the ground layer e.g. as larvae like the dipteran families Tipulidae and Limoniidae (Coleman et al. 2004). As a result species occurrences within these habitats are closely linked to time of the year. In contrast to that, many coleopteran families can be found in the soil layer all year round. Several members of the coleopteran family Staphylinidae and the ground beetle Carabidae occupy the ground strata throughout the year, resulting in an increase in the detection probability, especially when using the phosphate buffer. However, out of the 23 detected coleopteran species 21 were exclusively found in one season. This is once more indicating an insufficient sampling strategy.

Concluding it can be assumed that sampling was not adequately scaled leading to a high proportion of false negative results distorting invertebrate assessment. The differences found in species composition depending on forest type are therefore more likely to be an artefact of inappropriate sampling strategy.

Conclusion

Here it was shown that species detection rates, distribution patterns and finally resulting assessed community composition are significantly influenced by choice of DNA extraction method, marker and sampling strategy.

DNA extraction of larger amounts of soil does not necessarily lead to a more complete assessment of local community composition. Moreover, depending on target taxa DNA extraction from smaller amount of soil can obtain similar high numbers of recovered species. Further studies are required to identify the main drivers for these findings. The here presented results indicate that the application of a lysis step is strongly influencing assessed community composition. Among other things, this might be the result of changes in eDNA composition due to the release of intracellular DNA. However, little is known about natural and with a lysis step induced eDNA release processes and how it might vary between species. Specimen size and morphology might play next to behavioral aspects an important role. These large gaps of knowledge highlight the importance of further scientific research in order to develop, scale and refine eDNA metabarcoding approaches for the assessment of soil invertebrate diversity.

The enormous diversity of eDNA found in soil requires the development of target specific primer to limit the amplification of non-target DNA to a minimum. With increasing amount of soil used for extraction this tasks gains in importance as the higher the amount of starting

material, the more complex the resulting DNA mixture. This is curse and savior at the same time. On the one hand species with low population densities are more likely to be detected with increasing amount of starting material, on the other hand, the higher number of species contributing to eDNA mixture further increases primer bias

However here it was shown that eDNA is a promising tool for the assessment of invertebrate communities. We assume that with increasing completeness of databases and the development of more specific primer pairs metabarcoding will significantly increase knowledge about life in soil.

Literature

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

Albers D., Migge S., Schaefer M., Scheu S. 2004. Decomposition of beech leaves (Fagus sylvatica) and spruce needles (Picea abies) in pure and mixed stands of beech and spruce. Soil Biology and Biochemistry. 36:155–164.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403–410.

André H.M., Noti M.-I., Lebrun P. 1994. The soil fauna: the other last biotic frontier. Biodiversity & Conservation. 3:45–56.

Andújar C., Arribas P., Ruzicka F., Crampton-Platt A., Timmermans M.J., Vogler A.P. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. Molecular Ecology. 24:3603–3617.

Andújar C., Arribas P., Yu D.W., Vogler A.P., Emerson B.C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology. 27:3968–3975.

Baker G., Carter P., Curry J., Cultreri O., Beck A. 1998. Clay content of soil and its influence on the abundance of Aporrectodea trapezoides Duges (Lumbricidae). Applied Soil Ecology. 9:333–337.

Bardgett R. 2005. The biology of soil: a community and ecosystem approach. Oxford university press.

Barnard P.C. 2011. The royal entomological society book of British insects. John Wiley & Sons.

Baselga A., Orme C.D.L. 2012. betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution. 3:808–812.

Baulieu F., Knee W., Nowell V., Schwarzfeld M., Lindo Z., Behan-Pelletier V.M., Lumley L., Young M.R., Smith I., Proctor H.C. 2019. Acari of Canada. ZooKeys. 819:77-168.

Berger T.W., Berger P. 2012. Greater accumulation of litter in spruce (Picea abies) compared to beech (Fagus sylvatica) stands is not a consequence of the inherent recalcitrance of needles. Plant and Soil. 358:349–369.

Beylich A., Graefe U. 2009. Investigations of annelids at soil monitoring sites in Northern Germany: reference ranges and time-series data. Soil Organisms. 81:175–196.

Bhadury P., Austen M.C., Bilton D.T., Lambshead P.J.D., Rogers A.D., Smerdon G.R. 2006. Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. Marine Ecology Progress Series. 320:1–9.

Bienert F., De Danieli S., Miquel C., Coissac E., Poillot C., BRUN J., Taberlet P. 2012. Tracking earthworm communities from soil DNA. Molecular Ecology. 21:2017–2030.

Billany D., Brown R. 1980. The web-spinning larch sawfly, Cephalcia lariciphila Wachtl.(Hymenoptera: Pamphiliidae) a new pest of Larix in England and Wales. Forestry: An International Journal of Forest Research. 53:71–80.

Binns E. 1981. Fungus gnats (Diptera: Mycetophilidae/Sciaridae) and the role of mycophagy in soil: a review. Revue d'écologie et de biologie du sol. 18:77–90.

Blott S.J., Pye K. 2001. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. Earth Surface Processes and Landforms. 26:1237–1248.

Blott S.J., Pye K. 2012. Particle size scales and classification of sediment types based on particle size distributions: Review and recommended procedures. Sedimentology. 59:2071–2096.

Blouin M.S., Yowell C.A., Courtney C.H., Dame J.B. 1998. Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics. Molecular Biology and Evolution. 15:1719–1727.

Blum S.A., Lorenz M.G., Wackernagel W. 1997. Mechanism of retarded DNA degradation and prokaryotic origin of DNases in nonsterile soils. Systematic and Applied Microbiology. 20:513–521.

Bourlat S.J., Haenel Q., Finnman J., Leray M. 2016. Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method. Marine Genomics. Springer. p. 197–207.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

Campana C., Gauvin S., Ponge J.-F. 2002. Influence of ground cover on earthworm communities in an unmanaged beech forest: linear gradient studies. European Journal of Soil Biology. 38:213–224.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 7:335–336.

Castin-Buchet V., André P. 1998. The influence of intensive thinning on earthworm populations in the litters of Norway spruce and Douglas fir. Pedobiologia. 42:63.

Chandra L., Gupta S., Pande V., Singh N. 2016. Impact of forest vegetation on soil characteristics: a correlation between soil biological and physico-chemical properties. 3 Biotechnology. 6:188.

Chen H., Boutros P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 12:35.

Clarke L.J., Beard J.M., Swadling K.M., Deagle B.E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 7:873–883.

Clarke L.J., Soubrier J., Weyrich L.S., Cooper A. 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources. 14:1160–1170.

Coleman D.C., Crossley Jr D.A., Hendrix P.F. 2004. Fundamentals of Soil Ecology. Academic Press.

Conway J.R., Lex A., Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940.

Cowart D.A., Pinheiro M., Mouchel O., Maguer M., Grall J., Miné J., Arnaud-Haond S. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one. 10:e0117562.

Crecchio C., Ruggiero P., Curci M., Colombo C., Palumbo G., Stotzky G. 2005. Binding of DNA from Bacillus subtilis on montmorillonite–humic acids–aluminum or iron hydroxypolymers. Soil Science Society of America Journal. 69:834–841.

Creer S., Deiner K., Frey S., Porazinska D., Taberlet P., Thomas W.K., Potter C., Bik H.M. 2016. The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution. 7:1008–1018.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W., Power D.M., Packer M., Carvalho G.R., Blaxter M.L., Lambshead P.J.D., Thomas W.K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology. 19:4–20.

Cremer M., Kern N.V., Prietzel J. 2016. Soil organic carbon and nitrogen stocks under pure and mixed stands of European beech, Douglas fir and Norway spruce. Forest Ecology and Management. 367:30–40.

Currie C.R., Spence J.R., Niemelä J. 1996. Competition, cannibalism and intraguild predation among ground beetles (Coleoptera: Carabidae): a laboratory study. The Coleopterists' Bulletin. 50:135–148.

Curry J.P. 2004. Factors affecting the abundance of earthworms in soils. Earthworm Ecology. CRC Press. p. 91-114

Curtis T.P., Sloan W.T., Scannell J.W. 2002. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences. 99:10494–10499.

Davidson D., Grieve I., Young I. 2002. Impacts of fauna on an upland grassland soil as determined by micromorphological analysis. Applied Soil Ecology. 20:133–143.

De Ley P., De Ley I.T., Morris K., Abebe E., Mundo-Ocampo M., Yoder M., Heras J., Waumann D., Rocha-Olivares A., Jay Burr A. 2005. An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences. 360:1945–1958.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F., Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology Letters. 10:20140562.

Decaëns T. 2010. Macroecological patterns in soil communities. Global Ecology and Biogeography. 19:287–302.

Decaëns T., Jiménez J.J., Gioia C., Measey G., Lavelle P. 2006. The values of soil animals for conservation biology. European Journal of Soil Biology. 42:S23–S38.

Deleporte S. 2001. Changes in the earthworm community of an acidophilous lowland beech forest during a stand rotation. European Journal of Soil Biology. 37:1–7.

Delmont T.O., Robe P., Cecillon S., Clark I.M., Constancias F., Simonet P., Hirsch P.R., Vogel T.M. 2011. Accessing the soil metagenome for studies of microbial diversity. Applied and Environmental Microbiology. 77:1315–1324.

Demaneche S., Jocteur-Monrozier L., Quiquampoix H., Simonet P. 2001. Evaluation of biological and physical protection against nuclease degradation of clay-bound plasmid DNA. Applied and Environmental Microbiology 67:293–299.

Dentinger B.T., Didukh M.Y., Moncalvo J.-M. 2011. Comparing COI and ITS as DNA barcode markers for mushrooms and allies (Agaricomycotina). PLoS One. 6.

Derycke S., De Ley P., Tandingan De Ley I., Holovachov O., Rigaux A., Moens T. 2010a. Linking DNA sequences to morphology: cryptic diversity and population genetic structure in the marine nematode Thoracostoma trachygaster (Nematoda, Leptosomatidae). Zoologica Scripta. 39:276–289.

Derycke S., Vanaverbeke J., Rigaux A., Backeljau T., Moens T. 2010b. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. PLoS One. 5.

Disney H. 2012. Scuttle flies: the Phoridae. Springer Science & Business Media.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 14:927–930.

Dodds K.J., Graber C., Stephen F.M. 2001. Facultative intraguild predation by larval Cerambycidae (Coleoptera) on bark beetle larvae (Coleoptera: Scolytidae). Environmental Entomology. 30:17–22.

Dopheide A., Xie D., Buckley T.R., Drummond A.J., Newcomb R.D. 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution. 10:120–133.

Drummond A.J., Newcomb R.D., Buckley T.R., Xie D., Dopheide A., Potter B.C., Heled J., Ross H.A., Tooman L., Grosser S. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. GigaScience. 4:s13742-015.

Dutta R.K., Agrawal M. 2002. Effect of tree plantations on the soil characteristics and microbial activity of coal mine spoil land. Tropical Ecology. 43:315–324.

EFSA Panel on Plant Health (PLH), Jeger M., Bragard C., Caffier D., Candresse T., Chatzivassiliou E., Dehnen-Schmutz K., Gilioli G., Jaques Miret J.A., MacLeod A. 2017. Pest categorisation of Cephalcia lariciphila. EFSA Journal. 15:e05106.

Eisenhauer N., Milcu A., Sabais A.C., Bessler H., Weigelt A., Engels C., Scheu S. 2009. Plant community impacts on the structure of earthworm communities depend on season and change with time. Soil Biology and Biochemistry. 41:2430–2443.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Elbrecht V., Peinert B., Leese F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7:6918–6926.

Elmer Mi., Roth M. 2001. Effekte des Umbaus von Fichtenmonokulturen in Buchenmischwälder auf die saprophagen Dipteren. Mitteilungen der Deutschen Gesellschaft für allgemeine und angewandte Entomologie. 13:319–324.

Finzi A.C., Van Breemen N., Canham C.D. 1998. Canopy tree–soil interactions within temperate forests: species effects on soil carbon and nitrogen. Ecological Applications. 8:440–446.

Fonseca V.G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources. 18:923–926.

Fonseca V.G., Carvalho G.R., Nichols B., Quince C., Johnson H.F., Neill S.P., Lambshead J.D., Thomas W.K., Power D.M., Creer S. 2014. Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. Global Ecology and Biogeography. 23:1293–1302.

Fonseca V.G., Lallias D. 2016. Metabarcoding marine sediments: preparation of amplicon libraries. Marine Genomics. Springer. p. 183–196.

Fonseca V.G., Nichols B., Lallias D., Quince C., Carvalho G.R., Power D.M., Creer S. 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. Nucleic Acids Research. 40:e66–e66.

Frøslev T.G., Kjøller R., Bruun H.H., Ejrnæs R., Brunbjerg A.K., Pietroni C., Hansen A.J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications. 8:1–11.

Gallet C., Lebreton P. 1995. Evolution of phenolic patterns in plants and associated litters and humus of a mountain forest ecosystem. Soil Biology and Biochemistry. 27:157–165.

Gerard B. 1967. Factors affecting earthworms in pastures. The Journal of Animal Ecology.:235–252.

Haddad N.M., Crutsinger G.M., Gross K., Haarstad J., Knops J.M., Tilman D. 2009. Plant species loss decreases arthropod diversity and shifts trophic structure. Ecology Letters. 12:1029–1039.

Hagen-Thorn A., Callesen I., Armolaitis K., Nihlgård B. 2004. The impact of six European tree species on the chemistry of mineral topsoil in forest plantations on former agricultural land. Forest Ecology and Management. 195:373–384.

Hajibabaei M., Shokralla S., Zhou X., Singer G.A., Baird D.J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One. 6:e17497.

Hebert P.D., Ratnasingham S., De Waard J.R. 2003a. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:S96–S99.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003b. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences. 270:313–321.

Hoorman J.J. 2011. The role of soil protozoa and nematodes. Fact Sheet: Agriculture and Natural Resources.(Smith KL), The Ohio State University Extension, Colombus, Ohio.

Horton D.J., Kershner M.W., Blackwood C.B. 2017. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. European Journal of Soil Biology. 80:43–48.

Huhta V., Koskenniemi A. 1975. Numbers, biomass and community respiration of soil invertebrates in spruce forests at two latitudes in Finland. Annales Zoologici Fennici. 12:164-182.

Huhta V., Siira-Pietikäinen A., Penttinen R. 2012. Importance of dead wood for soil mite (Acarina) communities in boreal old-growth forests. Soil Organisms. 84:499–512.

Ivask M., Kuu A., Meriste M., Truu J., Truu M., Vaater V. 2008. Invertebrate communities (Annelida and epigeic fauna) in three types of Estonian cultivated soils. European Journal of Soil Biology. 44:532–540.

Jakovlev J. 2012. Fungal hosts of mycetophilids (Diptera: Sciaroidea excluding Sciaridae): a review. Mycology. 3:11–23.

Jocque R. 1973. The spider fauna of adjacent woodland areas with different humus types. Biologisch Jaarboek. 41:153–179.

Kang S., Mills A.L. 2006. The effect of sample size in studies of soil microbial community structure. Journal of Microbiological Methods. 66:242–250.

Kiontke K., Fitch D.H. 2013. Nematodes. Current Biology. 23:R862–R864.

Kuhlman T., Reinhard S., Gaaff A. 2010. Estimating the costs and benefits of soil conservation in Europe. Land Use Policy. 27:22–32.

Kvist S. 2014. Does a global DNA barcoding gap exist in Annelida? Mitochondrial DNA Part A. 27:2241–2252.

Lejzerowicz F., Esling P., Pillet L., Wilding T.A., Black K.D., Pawlowski J. 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Reports. 5:13932.

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

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T., Machida R.J. 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. 10:34.

Levy-Booth D.J., Campbell R.G., Gulden R.H., Hart M.M., Powell J.R., Klironomos J.N., Pauls K.P., Swanton C.J., Trevors J.T., Dunfield K.E. 2007. Cycling of extracellular DNA in the soil environment. Soil Biology and Biochemistry. 39:2977–2991.

Lovett G.M., Weathers K.C., Arthur M.A. 2002. Control of Nitrogen Loss from Forested Watersheds by Soil Carbon: Nitrogen Ratio andTree Species Composition. Ecosystems. 5:0712–0718.

Lukac M. 2017. Soil biodiversity and environmental change in European forests. Central European Forestry Journal. 63:59–65.

Mardulyn P., Godden B., Echezarreta P.A., Penninckx M., Gruber W., Herbauts J. 1993. Changes in humus microbiological activity induced by the substitution of the natural beech forest by Norway spruce in the Belgian Ardennes. Forest Ecology and Management. 59:15–27.

Marquina D., Andersson A.F., Ronquist F. 2019a. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources. 19:90–104.

Marquina D., Esparza-Salas R., Roslin T., Ronquist F. 2019b. Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular Ecology Resources. 19: 1516-1530

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.

May R.M. 1990. How many species? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 330:293–304.

McCredie T., Parker C., Abbott I. 1992. Population dynamics of the earthworm Aporrectodea trapezoides (Annelida: Lumbricidae) in a Western Australian pasture soil. Biology and Fertility of Soils. 12:285–289.

Meier R., Zhang G., Ali F. 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. Systematic Biology. 57:809–813.

Menzel F., Mohrig W. 2000. Revision der paläarktischen Trauermücken (Diptera: Sciaridae). Ampyx.

Menzel F., Schulz U. 2007. Die Trauermücken in Deutschland-ökosystemare Bedeutung, zönozoologische Koinzidenzen und bioindikatorisches Potential (Diptera: Sciaridae). Beiträge zur Entomologie= Contributions to Entomology. 57:9–36.

Migge S., Maraun M., Scheu S., Schaefer M. 1998. The oribatid mite community (Acarina) of pure and mixed stands of beech (Fagus sylvatica) and spruce (Picea abies) of different age. Applied Soil Ecology. 9:115–121.

Morinière J., de Araujo B.C., Lam A.W., Hausmann A., Balke M., Schmidt S., Hendrich L., Doczkal D., Fartmann B., Arvidsson S. 2016. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PloS One. 11:e0155497.

Nielsen U.N., Ayres E., Wall D.H., Bardgett R.D. 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. European Journal of Soil Science. 62:105–116.

Nihlgård B. 1971. Pedological influence of spruce planted on former beech forest soils in Scania, South Sweden. Oikos. 22:302–314.

Ogram A., Sayler G.S., Gustin D., Lewis R.J. 1988. DNA adsorption to soils and sediments. Environmental Science & Technology. 22:982–984.

Oliverio A.M., Gan H., Wickings K., Fierer N. 2018. A DNA metabarcoding approach to characterize soil arthropod communities. Soil Biology and Biochemistry. 125:37–43.

Oostra S., Majdi H., Olsson M. 2006. Impact of tree species on soil carbon stocks and soil acidity in southern Sweden. Scandinavian Journal of Forest Research. 21:364–371.

Orgiazzi A., Bardgett R.D., Barrios E. 2016. Global soil biodiversity atlas. European Commission.

Osawa N. 1989. Sibling and non-sibling cannibalism by larvae of a lady beetle Harmonia axyridisPallas (Coleoptera: Coccinellidae) in the field. Population Ecology. 31:153–160.

Paget E., Lebrun M., Freyssinet G., Simonet P. 1998. The fate of recombinant plant DNA in soil. European Journal of Soil Biology. 34:81–88.

Paluch J.G., Gruba P. 2012. Effect of local species composition on topsoil properties in mixed stands with silver fir (Abies alba Mill.). Forestry: An International Journal of Forest Research. 85:413–426.

Phillips N.D. 2017. Yarrr! The pirate’s guide to R. APS Observer. 30.

Phillipson J., Abel R., Steel J., Woodell S. 1976. Earthworms and the factors governing their distribution in an English beechwood. Pedobiologia. 16:258-285

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

Porter T.M., Hajibabaei M. 2018. Over 2.5 million COI sequences in GenBank and growing. PloS One. 13:e0200177.

R CoreTeam. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranjard L., Lejon D.P., Mougel C., Schehrer L., Merdinoglu D., Chaussod R. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environmental Microbiology. 5:1111–1120.

Ratnasingham S., Hebert P.D.N. 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes. 7:355–364.

Reich P.B., Oleksyn J., Modrzynski J., Mrozinski P., Hobbie S.E., Eissenstat D.M., Chorover J., Chadwick O.A., Hale C.M., Tjoelker M.G. 2005. Linking litter calcium, earthworms and soil properties: a common garden test with 14 tree species. Ecology Letters. 8:811–818.

Reusch H., Oosterbroek P. 1997. Diptera Limoniidae and Pediciidae, short-palped crane flies. Aquatic Insects of North Europe-A Taxonomic Handbook. 2:105–132.

Ritter C.D., Häggqvist S., Karlsson D., Sääksjärvi I.E., Muasya A.M., Nilsson R.H., Antonelli A. 2019. Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. Genome. 62:147–159.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Sagova-Mareckova M., Cermak L., Novotna J., Plhackova K., Forstova J., Kopecky J. 2008. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Applied and Environmental Microbiology. 74:2902–2907.

Sariyildiz T., TÜFEKÇİOĞLU A., Küçük M. 2005. Comparison of decomposition rates of beech (Fagus orientalis Lipsky) and spruce (Picea orientalis (L.) Link) litter in pure and mixed stands of both species in Artvin, Turkey. Turkish Journal of Agriculture and Forestry. 29:429–438.

Schaefer M., Schauermann J. 1990. The soil fauna of beech forests: comparison between a mull and a moder soil. Pedobiologia. 34:299–314.

Scharff N., Coddington J.A., Griswold C.E., Hormiga G., de Place Bjørn P. 2003. When to quit? Estimating spider species richness in a northern European deciduous forest. The Journal of Arachnology. 31:246–273.

Scherber C., Eisenhauer N., Weisser W.W., Schmid B., Voigt W., Fischer M., Schulze E.-D., Roscher C., Weigelt A., Allan E. 2010. Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. Nature. 468:553-556.

Scheu S., Albers D., Alphei J., Buryn R., Klages U., Migge S., Platner C., Salamon J. 2003. The soil fauna community in pure and mixed stands of beech and spruce of different age: trophic structure and structuring forces. Oikos. 101:225–238.

Schuldt A., Fahrenholz N., Brauns M., Migge-Kleian S., Platner C., Schaefer M. 2008. Communities of ground-living spiders in deciduous forests: Does tree species diversity matter? Biodiversity and Conservation. 17:1267–1284.

Schulz U. 1996. Vorkommen und Habitatanforderungen von Bodenmakroarthropoden in Natur- und Wirtschaftswäldern: ein Vergleich. Cuvillier Verlag Göttingen. .

Schwarz B., Dietrich C., Cesarz S., Scherer-Lorenzen M., Auge H., Schulz E., Eisenhauer N. 2015. Non-significant tree diversity but significant identity effects on earthworm communities in three tree diversity experiments. European Journal of Soil Biology. 67:17–26.

Sirois S.H., Buckley D.H. 2019. Factors governing extracellular DNA degradation dynamics in soil. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12725.

Somervuo P., Yu D.W., Xu C.C., Ji Y., Hultman J., Wirta H., Ovaskainen O. 2017. Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. Methods in Ecology and Evolution. 8:398–407.

Stoeck T., Bass D., Nebel M., Christen R., Jones M.D., BREINER H., Richards T.A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology. 19:21–31.

Stork N.E. 1988. Insect diversity: facts, fiction and speculation. Biological journal of the Linnean Society. 35:321–337.

Strasser B.J. 2008. GenBank--Natural History in the 21st Century? Science. 322:537–538.

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

Taberlet P., PRUD'HOMME S.M., Campione E., Roy J., Miquel C., Shehzad W., Gielly L., Rioux D., Choler P., CLÉMENT J. 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular Ecology. 21:1816–1820.

Tang C.Q., Leasi F., Obertegger U., Kieneke A., Barraclough T.G., Fontaneto D. 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences. 109:16208–16212.

Telles T.S., Guimarães M. de F., Dechen S.C.F. 2011. The costs of soil erosion. Revista Brasileira de Ciência do Solo. 35:287–298.

Trum F., Titeux H., Ranger J., Delvaux B. 2011. Influence of tree species on carbon and nitrogen transformation patterns in forest floor profiles. Annals of Forest Science. 68:837–847.

Tschinkel W.R. 1981. Larval dispersal and cannibalism in a natural population of Zophobas atratus (Coleoptera: Tenebrionidae). Animal Behaviour. 29:990–996.

Usher M. 1985. Population and community dynamics in the soil ecosystem. Special publications series of the British Ecological Society 243-265.

Vesterdal L., Clarke N., Sigurdsson B.D., Gundersen P. 2013. Do tree species influence soil carbon stocks in temperate and boreal forests? Forest Ecology and Management. 309:4–18.

Vesterdal L., Schmidt I.K., Callesen I., Nilsson L.O., Gundersen P. 2008. Carbon and nitrogen in forest floor and mineral soil under six common European tree species. Forest Ecology and Management. 255:35–48.

Wall D.H., Fitter A.H., Paul E.A. 2005. Developing new perspectives from advances in soil biodiversity research. Biological Diversity and Function in Soils. Cambridge University Press, Cambridge.:31–43.

Wardle D. 1993. Changes in the microbial biomass and metabolic quotient during leaf litter succession in some New Zealand forest and scrubland ecosystems. Functional Ecology. 7:346–355.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.

Wickham H., Francois R., Henry L., Müller K. 2015. dplyr: A grammar of data manipulation. R package version 0.8.3. 3.

Wiemers M., Fiedler K. 2007. Does the DNA barcoding gap exist?–a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology. 4:8.

Wise D.H., Snyder W.E., Tuntibunpakul P., Halaj J. 1999. Spiders in decomposition food webs of agroecosystems: theory and evidence. Journal of Arachnology. 27:363–370.

Yang C., Ji Y., Wang X., Yang C., Douglas W.Y. 2013. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. Science China Life Sciences. 56:73–81.

Yang C., Wang X., Miller J.A., de Blécourt M., Ji Y., Yang C., Harrison R.D., Douglas W.Y. 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators. 46:379–389.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Zhang G.K., Chain F.J., Abbott C.L., Cristescu M.E. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications. 11:1901–1914.

Ziesche T.M., Roth M. 2008. Influence of environmental parameters on small-scale distribution of soil-dwelling spiders in forests: What makes the difference, tree species or microhabitat? Forest Ecology and Management. 255:738–752.

Zinger L., Chave J., Coissac E., Iribar A., Louisanna E., Manzi S., Schilling V., Schimann H., Sommeria-Klein G., Taberlet P. 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biology and Biochemistry. 96:16–19.

Chapter IV

Arthropod Diversity Assessment From the Ground Up:

Detection and Seasonal Overlap Between Soil and Malaise Traps

Abstract

In the light of recent evidence for dramatic insect declines worldwide, it has become increasingly important to deliver methods for the rapid, cost efficient and reliable assessment of insect communities. Metabarcoding matches these requirements and furthermore allows the assessment of biodiversity in a non-invasive way. This is because metabarcoding can be performed using different types of source material ranging from community DNA extracted from bulk tissue samples, to environmental DNA (eDNA) which encompasses intracellular and extracellular DNA captured from environmental samples such as water, soil, sediment or air. While many studies have already shown that eDNA originating from a broad spectrum of aquatic organisms gets accumulated in the sediments, it is still under discussion how well eDNA extracted from soil reflects the whole local biodiversity. Here we show that metabarcoding of eDNA from soil is a valuable complement to community DNA extracted from Malaise trap catches. Several species were identified in both types of sample at the same location but at different times of the year. Although soil seems to be a rather poor integrator for above ground arthropod biodiversity, it is capable of reflecting biodiversity associated with the belowground strata for at least some life stages of the species also found in Malaise traps. Furthermore, we showed that Malaise traps and soil samples respectively represented only a fraction of the total detected arthropod diversity, resulting in the characterization of two significantly different arthropod communities. We conclude that analysis based on a single substrate will likely underestimate total local

diversity but the use of more than one substrate source in the same location and a time series of samples allows to capture a more complete picture of local species diversity and turnover.

Introduction

In the last years, tremendous steps have been taken towards developing timely and cost- efficient methods for the monitoring of biodiversity. Since then, metabarcoding has increasingly been used to address ecological questions which have long remained unresolved due to methodological limitations (Kartzinel et al. 2015; Lopes et al. 2015; Purcell et al. 2017). In contrast to the traditional morphology based identification of species and barcoding, metabarcoding is capable of identifying several thousands of specimens in parallel. Nevertheless, metabarcoding still has some major drawbacks. While with barcoding and traditional morphological identification species abundances are easily assessed as each specimen is individually treated and identified, for terrestrial arthropods it is not yet possible to draw conclusions from the metabarcoding results as to number of individuals in the sample. However it is possible to infer relative abundances of the species in a sample. In terms of biomonitoring is metabarcoding clearly superior to conventional physical methods as it allows the assessment of local biodiversity in an non-invasive way, without causing destructive impacts to the biological community or the environment (Wheeler et al. 2004) hence metabarcoding can be performed using a wide range of source material. Depending on source material, the extracted DNA can be divided into two major groups: community DNA (cDNA) and environmental DNA (eDNA). cDNA encompasses all DNA which is isolated from bulk-extracted mixtures of organisms which were prior to DNA extraction separated from the substrate (Deiner et al. 2017). In contrast to that is eDNA all DNA captured from an environmental sample without first isolating any target organisms (Taberlet et

al. 2012a). Especially the latter has opened up new possibilities for the detection of local diversity and how it is connected to environmental parameters, in a non-invasive way. For aquatic habitats, several protocols have been invented to effectively monitor metazoan species on the basis of eDNA directly extracted from water (Valentini et al. 2016) and sediment samples (Fonseca et al. 2017). Based on these experiences efforts are now undertaken to develop gold-standard genomic tools and novel ecogenomic indices for routine application in biodiversity assessments of aquatic habitats (Leese et al. 2016). In contrast to that is metabarcoding from soil samples yet to mature.

A well-known problem of eDNA studies is the risk of false negative results, meaning that target species are present in a habitat, but DNA is not recovered (Guillera-Arroita et al. 2017). An important parameter which defines whether a species is detected or not is the concentration of eDNA from the targeted species in the environment (Garlapati et al. 2019). The concentration of eDNA is among other parameters, determined by the way it got deposited in the environment. Sources of genetic material organisms shed into the environment are manifold (Barnes and Turner 2016). For example, eDNA may result from fecal matter (Andersen et al. 2012), tissue from decaying carcasses (Kamoroff and Goldberg 2018) or shed epithelial cells (Sassoubre et al. 2016). Former studies have already shown, that species with a small body size, highly sclerotized species and species with a low population density such as rare and perhaps endangered species are more likely to be missed (Guillera-Arroita et al. 2017) as less genetic material is released into the environment resulting in a lower DNA concentration. Furthermore, sampling strategy and extraction method are known to directly influence species detection rate (Ranjard et al. 2003; Dopheide et al. 2019).

Former metabarcoding studies have already shown that in aquatic habitats, species detection rate is directly linked to the substrate type used for eDNA extraction (Koziol et al. 2019). It is likely that this is also true for terrestrial habitats, especially as the soil matrix seems to be a rather poor integrator of the whole local biodiversity compared to aquatic ecosystems (Taberlet et al. 2018).

While many studies have already dealt with the persistence of eDNA in aquatic habitats (Dejean et al. 2011; Barnes et al. 2014; Deiner and Altermatt 2014; Collins et al. 2018; Salter 2018), little is known about the fate of eDNA in soils. A recent study has shown that moisture, temperature, habitat and agricultural management are influencing eDNA degradation rate (Sirois and Buckley 2019), but the underlining causes are still to a wide extent unknown. Furthermore, Ogram et al. (1988) showed that soil characteristics have an major influence on absorption rate of eDNA to soil particles and thereby on eDNA degradation rate (Cai et al. 2006) but the complexity and wide range of possible binding mechanisms (Pietramellara et al. 2009) is not allowing for general statements pointing to accurate prediction of time of eDNA persistence. This is why the identification of soil dwelling organisms from eDNA can be particularly challenging. During DNA extraction, humic substances are often co-extracted along with the nucleic acids (Wang et al. 2017). Humic substances are known to inhibit PCR (Matheson et al. 2010) which results in major PCR biases leading to the underestimation of the local diversity due to false negative results. Although eDNA extracted from soil samples has already been proven to provide reliable information about the existing diversity of several organism groups like annelids (Bienert et al. 2012), plants (Yoccoz et al. 2012) and vertebrates (Andersen et al. 2012; Leempoel et al. 2019), a proof of concept is still missing for many invertebrate groups.

However, robust molecular techniques would be particularly desirable for assessing soil invertebrate communities as due to their high taxonomic and functional diversity, the role of many soil invertebrate taxa for ecosystem function still remains unclear (Wardle 2006; Cardoso et al. 2011). Among the invertebrates, arthropods are one of the most diverse groups whose members can be found across a wide range of terrestrial habitats. Arthropod diversity is especially outstanding in forests as these ecosystems are very complex providing a broad range of niches (Szujecki 1986). Forests are roughly divided into two strata: the ground and the above ground layer. While some arthropods such as Chilopoda are usually associated with the ground layer, some groups like many lepidopterans are usually associated with the above ground strata.

Both strata are closely linked to each other by mutual relationships of the associated abiotic and biotic environment (Wardle et al. 2004), but the line between the strata is blurred. While some species are present in one stratum all year round, the appearance of other taxa in one or both strata is seasonally driven. The latter is often observed for insects, which usually have a complex lifecycle (Danks 2007). Insects with ground dwelling larvae can, once developed into an imago, leave the below ground strata and find their way into above ground habitats. The time an insect persists in any life stage varies strongly between species and is often seasonally driven (Danks 2007). When monitoring both strata over a longer time period, it is likely that a time lagged overlap of species occurrences can be observed between strata. However, depending on target organisms several studies are still based on a single sampling method e.g. Malaise traps, pit fall traps, light traps, bait traps or soil samples. Flying insects are often sampled with Malaise traps but time and duration of sampling can thereby strongly influence whether a species is sampled, as several species only show flight activity for a short period of time (e.g. ants) (Noordijk et al. 2008). Species with a short flight period are therefore more prone to be missed when only a single sampling strategy is applied, leading to false negative results. Combining several sampling strategies will likely increase number of recovered species, even when only targeting flying species.

We expect that species associated with the ground strata at least in one life stage can be well monitored with eDNA extracted from soil. Nevertheless, it is still unclear to what extent soil serves as a sink for eDNA originating from species not actively interacting with it. However, it is likely that species detection rate is directly linked to the current life stage of the organism. This encompasses two cases: first, the organism is not present in the soil anymore and therefore not interacting with it (e.g. Imagoes which have already found their way out to above ground habitats) and secondly, the organism is present in the soil, but it is currently not interacting with it (e.g. organisms in the pupal or dormant stage). While organisms that are actively interacting with the habitat leave a track of DNA traces (e.g. originating from feces, excretions or epithelial

cells) through the habitat, inactive organisms might be harder to detect unless directly captured. Choice of method for DNA extraction can therefore directly influence species detection rate. For example, methods involving or not a lysis step can be used. eDNA encompasses intracellular and extracellular DNA. The first is defined as DNA which is still located within cell membranes, while the latter has already been released from it. Intracellular DNA can only be extracted from the sample when a lysis step is applied, cracking open the cell membranes resulting in the release of the DNA. When using methods that target exclusively extracellular DNA, it is likely that transient species in the form of eggs or pupae are missed. Applying a lysis step might therefore increase the chance that non-active organisms are also detected in the environmental sample.

Studies encompassing both soil eDNA and malaise trap samples are rare (Yang et al. 2014; Horton et al. 2017; Marquina et al. 2019) and to our knowledge only two studies have so far used both methods to address ecological questions (Yang et al. 2014; Marquina et al. 2019). Although several species are known to be only occasional inhabitants of either of the two habitats and time of occurrence is highly depending on species life stage, no study has so far addressed seasonal turnover rates between substrates and thereby habitats. It is therefore likely that detected differences in arthropod species composition (Marquina et al. 2019) might have resulted from insufficient sampling, whereas year-round monitoring could reveal less distinct arthropod communities between the two habitats.

Aims of the study

The aims of this study are to investigate the possibility of soil eDNA as a source of information for existing local (ground and above ground) arthropod diversity. We aim to answer the following questions:

1) Do soil samples and Malaise traps detect similar arthropod species communities?
2) To what extent does variation in Arthropod species communities exist between seasons and forest types depending on sampling method?
3) Is there a time-lagged overlap of species occurrences between the ground and above ground strata?

Material and Methods

Sampling strategy

All sample sites were located in the Eifel National park, which is situated in the south-western part of Germany close to the Belgian border (figure IV.1).

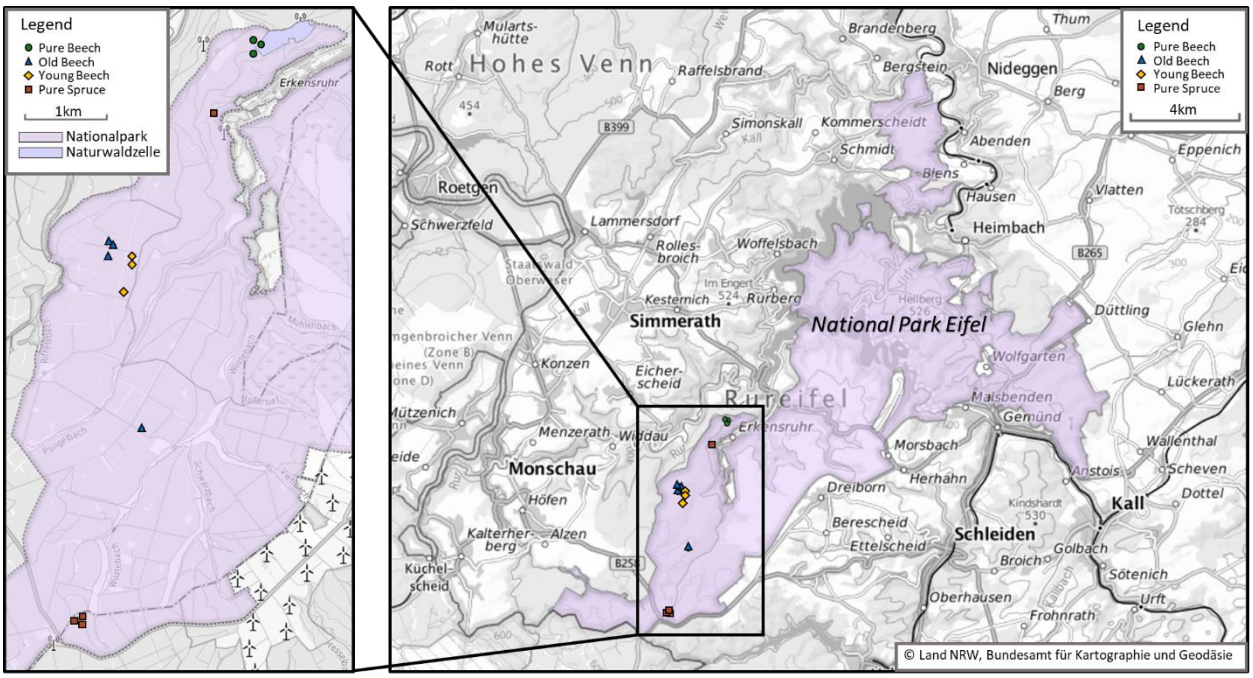


Figure IV.1 Location of sampling sites. Current area of the Eifel National Park is highlighted in purple.

For this study a forest conversion gradient from Norway Spruce (*Picea abies*) to European Beech (*Fagus sylvatica*) was sampled. To reflect the different stages of conversion from spruce to beech, four forest types were defined: pure beech (PB), old beech (OB), young beech (YB) and pure spruce (PS) (table IV.1).

Table IV.1 Geographical and ecological characteristics of the 14 sample sites. Depicted for each sample site are the Coordinates (altitude N and latitude E) and the associated local forest type

Sample Sites	Coordinates	Forest Type
Sample Site 01	50° 34'11.7984''N 6°21'32.1012''E	Pure Beech
Sample Site 02	50° 34'07.7016''N 6°21'27.3996''E	Pure Beech
Sample Site 03	50° 34'12.9000''N 6°21'27.3996''E	Pure Beech
Sample Site 04	50° 32'44.5992''N 6°20'15.2988''E	Young Beech
Sample Site 05	50° 32'41.3016''N 6°20'15.6984''E	Young Beech
Sample Site 06	50° 32'29.7996''N 6°20'11.1012''E	Young Beech
Sample Site 07	50° 32'29.7996''N 6°20'11.1012''E	Old Beech
Sample Site 08	50° 31'35.1984''N 6°20'25.2996'E	Old Beech
Sample Site 09	50° 32'48.3000''N 6°20'03.4008''E	Old Beech
Sample Site 10	50° 30'17.2008''N 6°19'48.1008''E	Pure Spruce
Sample Site 11	50° 30'18.2988''N 6°19'51.4020''E	Pure Spruce
Sample Site 12	50° 33'15.8004''N 6°21'07.3008''E	Pure Spruce
Sample Site 13	50° 30'16.0056''N 6°19'51.4704''E	Pure Spruce
Sample Site 14	50° 32'49.9632''N 6°20'00.7296''E	Old Beech

The forest types differed in tree species composition as well as in approximate age of the trees. The pure beech and pure spruce sampling sites were located in monoculture stands which were

either dominated by beeches or spruces respectively. The pure beech monoculture stands were approximately 180 years old and partly under special protection through North-Rhine Westfalia (Naturwaldzelle) (Sample Site 01). With a mean age of 60 years, the pure spruce stands were substantially younger. Spruces of the same age dominated the young beech sampling sites. At the young beech sampling sites, young beeches had been planted only recently which had not yet reached three meters in size at the time of sampling. At the old beech sampling sites, beeches had already reached a height of more than 3 meters and actions to remove spruces from the forest had already been undertaken (figure IV.2).

Figure IV.2 Forest conversion gradient: From spruce monocultures over with beeches underplanted spruce forests to beech monocultures. Photos were taken during summer season (July 2016).

In July 2016, 12 Malaise traps were set up in the Eifel National Park, North-Rhine Westfalia, Germany. Three malaise traps were installed per forest type, to capturing flying insects (table IV.2). To ensure that the orientation of the Malaise traps did not affect sampling success, the highest point of each traps was set up pointing south. The traps were left in the field for the full duration of the experiment until April 2017, in order to ensure that insects were collected from exactly the same locations. In October 2016, two additional traps (Malaise Trap 13 and Malaise Trap 14) were installed at two further sample sites (Sample Site 13 and Sample Site 14). All traps were equipped with a bottle filled with approximately 1 litre of 99,96% pure ethanol over a two

week period in July 2016 (13.07-27.07), October 2016 (13.10-27.10), January 2017 (11.01-25.01) and April 2017 (12.04-26.04) (table IV.2). The ethanol bottles were replaced after one week in the field. After each weekly collection, the ethanol was replaced to ensure that concentration of the preservative ethanol was stable. After final collection, the traps were left unequipped in the field until the start of the next sampling period. Between October 2016 and January 2017 nine malaise traps were destroyed by heavy snow fall. The damaged traps were replaced at the start of the new sampling season in January 2017.

Table IV.2 Malaise trap collection periods. For each sampling season the time of the year, number of traps and time period of collection is depicted

Season	Time of the year	Number of Traps	Sampling Dates
Season 1	Summer	12	13.07.2016 – 27.07.2016
Season 2	Autumn	14	13.10.2016 – 27.10.2016
Season 3	Winter	14	11.01.2017 – 25.01.2017
Season 4	Spring	14	12.04.2017 – 26.04.2017

Close to each Malaise trap, three soil samples were taken. Soil sampling sites were located 4m and respectively 5m apart from the trap, forming a triangle in the centre of which the malaise trap is located (figure IV.3). One corner of the sampling triangle was pointing south, while the both remaining corners were pointing north-west and northeast respectively.

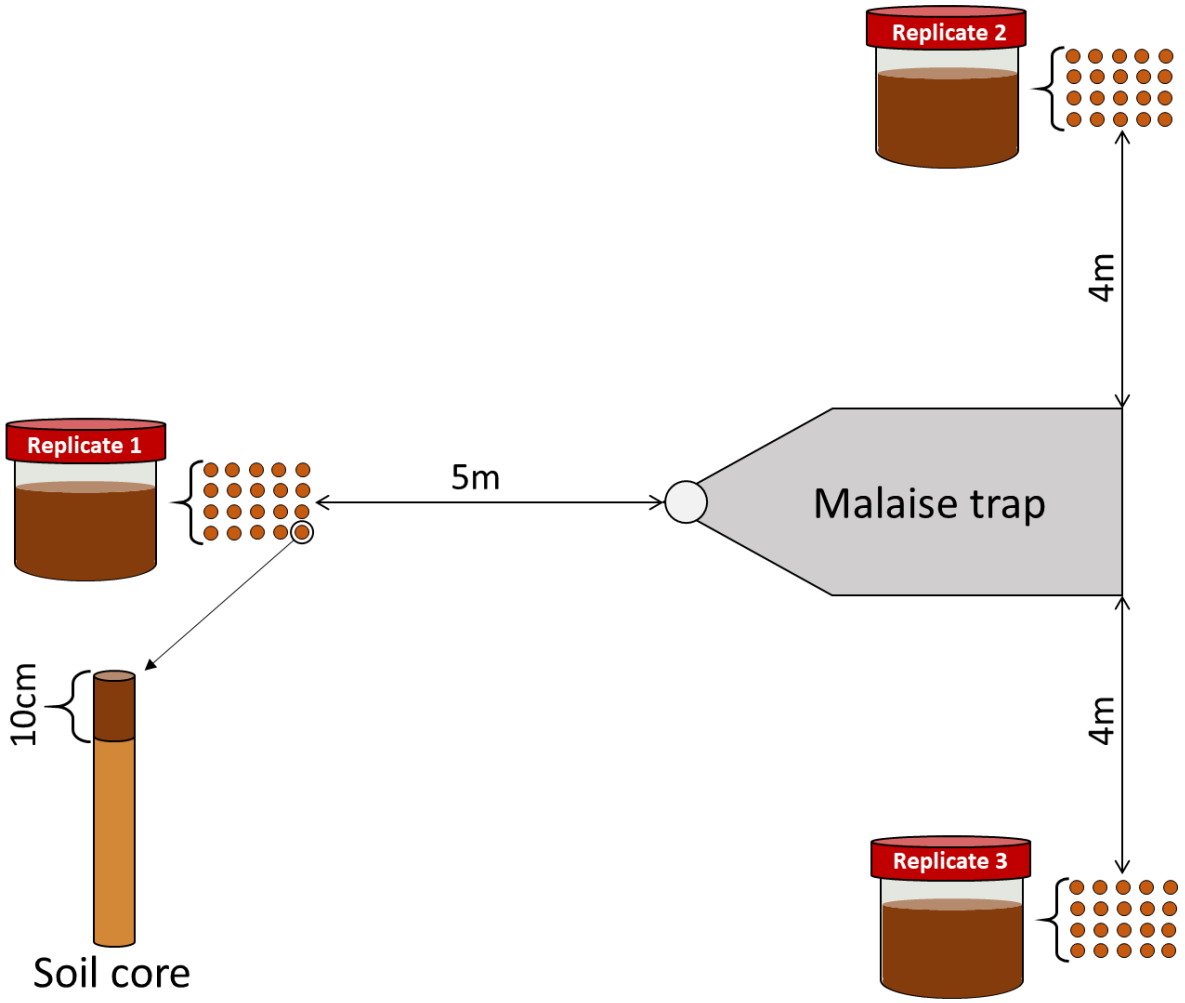


Figure IV.3 Overview of study set up. The Malaise trap is at the centre of a triangle formed by soil sampling locations. For soil sampling the upper 10cm of soil were taken. At each sampling site three biological replicates, each consisting of 20 soil cores were taken.

Each sampling site was sampled four times in the course of a one-year period. Soil samples and Malaise trap trials were synchronized. Soil sampling was scheduled on day 14 of each Malaise trap trial, when final bottles were collected (table IV.3). Each soil sample consisted of approximately twenty 44mm diameter x 100mm cores, taken 5 cm apart from each other. In total 162 soil samples were taken. Shortly after sampling soil samples were stored at -20°C until further processing.

Table IV.3 Soil sample collection periods. For each sampling season the time of the year, number of traps and date of collection is depicted

Season	Time of the year	Number of Samples	Sampling Dates
Season 1	Summer	36	27.07.2016
Season 2	Autumn	48	27.10.2016
Season 3	Winter	48	25.01.2017
Season 4	Spring	48	26.04.2017

DNA Extraction

Malaise Traps

DNA extraction was performed after overnight incubation in lysis buffer using non-destructive methods. We followed a modified protocol of Aljanabi & Martinez (1997) adjusted for our purposes. Organisms were first sieved from the collecting ethanol. To ensure that no specimen was overseen or lost, we poured the content of each bottle through a mesh filter (MICROFIL®V Filter White Gridded 0.45µm-diameter 47mm & 100ml Funnel Sterilized). To reduce the risk of accidentally loosing small specimens, we processed the filter with the specimens. The insects were dried for 10 minutes at room temperature. Depending on biomass between 15ml and 25ml of extraction buffer (0.4M NaCl, 10mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0) and 2% Sodium dodecyl sulphate (SDS) was added to each bulk sample. Finally, 400 µg Proteinase K per ml was added to lysis buffer solution. The samples were left to digest at 52°C at 30rpm in the orbital shaker overnight.

The next day the lysate was poured out of the bottles using the MICROFIL®V Filter (White Gridded

0.45um-Dia 47mm & 100ml Funnel Sterilized – Q :24). The solution was equally split into three 15ml falcon tubes. A 6M NaCl solution was added to the falcon tubes to a concentration of 4mmol. After vortexing for 30 seconds, the tubes were centrifuged at 4700rpm for 30 seconds. The supernatant was transferred to a new falcon tube to which an equal amount of isopropanol was added. After carefully mixing by turning the tube upside down a few times, the tubes were left at -20°C for one hour and subsequently centrifuged at 4700 rpm for 60 minutes. The supernatant was removed from the tube before submerging the resulting pellet with 20ml of ice cold 70% ethanol. The ethanol containing tubes were centrifuged for 15 minutes at 4700 rpm. The supernatant was then discarded and the pellet was left to dry at 20°C overnight. The next day, 1ml of sterile H_2O was added to the dry pellet. The resulting DNA solution was stored at -20°C until further processing.

Soil Samples

DNA extraction of the soil samples was conducted using two different extraction methods, a commercial DNA extraction Kit and the Taberlet et al. (2012b) phosphate buffer protocol. The first method uses 0.5 g of soil per sample followed by DNA extraction using a commercial silica membrane based kit. Briefly, after thoroughly mixing the sample, DNA was extracted from 0.5g soil per sample using the Macherey-Nagel NucleoSpin® Soil kit following the manufacturer's protocol.

The second DNA extraction method comprised using 50 - 150g of soil per sample using a phosphate buffer based solution (Taberlet et al. 2012b). This method allows DNA to be extracted from larger amounts of starting material. For DNA extraction using the phosphate buffer solution (Na2HPO4; 0.12M; pH 8) we followed the protocol of Taberlet et al. (2012b). Briefly, soil samples were removed from the -20°C chamber approximately 12 hours before DNA extraction and stored

at +4°C overnight. The next morning, each sample was thoroughly mixed and equal weight of saturated phosphate buffer solution was added. All samples were processed in duplicates and placed in an orbital shaker at 120 rpm for 15 minutes. Two 2ml Eppendorf safe lock tubes were filled with 1.7 ml of the resulting mixture and centrifuged for 10 min at 10000g. 400µl of the resulting supernatant were transferred to a new 2ml collection tube to which 200µl of SB binding buffer included in the Macherey-Nagel NucleoSpin® Soil kit was added. Each duplicate lysate were loaded onto a NucleoSpin® Soil Column and centrifuged at 10000g for one minute, respectively (merged duplicates from each sample). For the rest of the extraction, the standard manufacturer's protocol for the Macherey-Nagel NucleoSpin® kit was followed starting from step 8. Finally, DNA was eluted with 50µl of SE Buffer (Macherey-Nagel). 10ul of the resulting elution step was combined with 90µl pure H_2O (Sigma), followed by DNA purification using the PowerClean® Pro DNA Clean-Up Kit (MO Bio Laboratories, Inc.) following the manufacturer's protocol.

Choice of Primers and Library Preparation

For amplicon library preparation of soil samples, a COI primer pair targeting 313bp of the 658bp long barcoding region of the mitochondrial Cytochrome c Oxidase subunit I gene (COI) were used, using a two step PCR approach (figure IV.4). In the first step, the fragment of interest is amplified using primers targeting specific gene regions and also including an Illumina adaptor overhang (referred to as PCR 1), and in the second step (referred to as PCR 2) Illumina adaptor and index-tag primers are added (Bourlat et al. 2016; Fonseca and Lallias 2016).

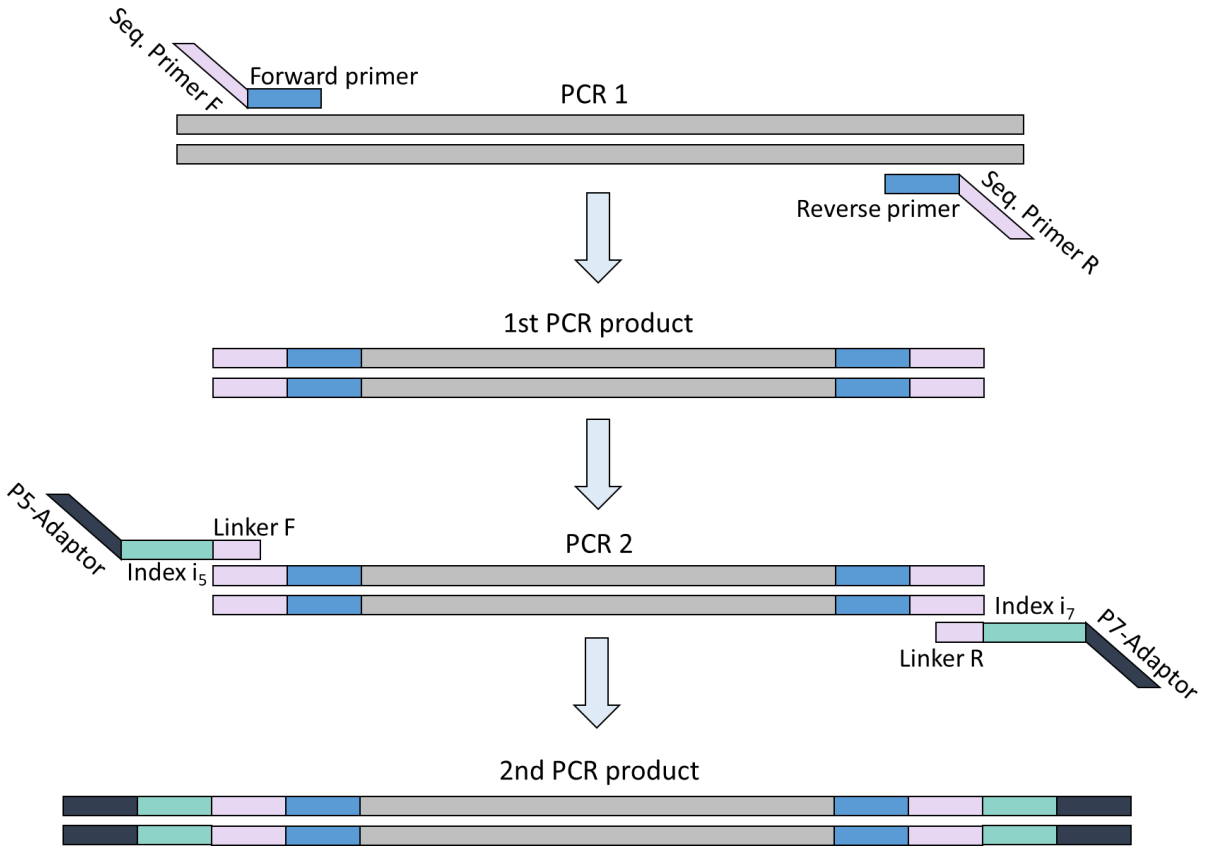


Figure IV.4 Illustration of the 2-step PCR approach

10ng of template DNA was used for PCR1 using the primers mlCOIintF *(5'*-ACACTCTTTCCCTACACGACGCTCTTCCGATCT **GGWACWGGWTGAACWGTWTAYCCYCC** *-3')* (Leray et al. 2013) and dgHCO2198 *(5'*-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT **TAAACTTCAGGGTGACCAAARAAYCA***-3')* (Leray et al. 2013). The PCR1 mix for each sample consisted of 7.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1ul Sigma H2O, 0.5µl of forward Primer (10µM), 0.5µl of reverse primer (10µM), 0.5ul Bovine Serum Albumin (thermoscientific) and 1µl template DNA, making up a total of 15µl.

The following PCR Program was applied: 1 cycle 98°C for 2 minutes, 20 cycles 98°C for 40 seconds

followed by 45°C for 40 seconds and 72°C for 30 seconds with a final elongation at 72°C for 3 minutes. Following PCR1 the PCR products were purified by adding 4µl of HT ExoSAP-IT^{TM} (Applied Biosystems) to each sample. Samples were first heated at 37°C for 15 minutes removing excess primers and dNTPs, then at 80°C for 15 minutes to inactivate the enzyme and subsequently cooled at 4°C for 5 minutes.

For PCR2 (index PCR) the purified PCR products were split into two PCR tubes. While with PCR1 the gene region of interest was amplified, in PCR2 molecular identification (MID) tags in combination with NGS platform specific primers were incorporated. For each sample a unique combination of MID tags targeting both amplicon ends were chosen.

Each tube contained 12.5µl Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 3µl Sigma H2O, 1.2µl of Index forward primer (10µM) (AATGATACGGCGACCACCGAGATCTACAC NNNNNNNN ACACTCTTTCCCTACACGACGCTC), 1.2µl of index reverse primer (10µM) CAAGCAGAAGACGGCATACGAGAT NNNNNNNN GTGACTGGAGTTCAGACGTGTGCTC) and 8µl of purified PCR1 product. The PCR2 program was run as follows: 1 cycle at 98°C for 2 minutes, followed by 20 cycles of 40 seconds at 98°C, 30 seconds at 55°C and 30 seconds at 72°C. Final elongation at 72°C for 3 minutes. The PCR Products were visualised by gel electrophoresis and bands of the expected size were cut out. Cut out gel pieces of the same sample were merged before being purified using the QIAquick Gel Extraction Kit (Qiagen). DNA quantification was conducted with the Quantus Fluorometer (Promega). All purified and quantified tagged amplicons were pooled equimolarly. The resulting purified amplicon pool with a concentration of 3ng/µl was sequenced on a Illumina Miseq (2x 300bp) sequencing platform at Liverpool University's Centre for Genomic Research (Liverpool, UK).

Bioinformatics and Data Analysis

Data sequenced at the Centre of Genomic Research (Liverpool, UK) had already undergone first quality check: The raw fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1. Furthermore, sequences were trimmed using Sickle version 1.200 with a minimum window quality score of 20. Reads shorter than 20bp after trimming were removed. Sequences were received from MacroGen Inc. Seoul in Casava 1.8 paired-end demultiplexed fastq format.

The fastq sequences were checked for the presence of the COI primers with Cutadapt version 1.18 (Martin 2011) using the following settings: maximum error rate (-e): 0.1, minimum Overlap (-O): 20, minimum sequence length (-m): 50). Sequences lacking either forward or reverse primer were remove. Detected primer pairs were trimmed off from the remaining sequences. Subsequently, paired-end reads were merged with vsearch version 2.7.0 (Rognes et al. 2016). Merged sequences with a length of 293-333bp were retained for further analysis and filtered with a maxEE threshold of 1.0 using vsearch (version 2.7.0) (Rognes et al. 2016) before fastq-sequences were demultiplexed using the script split_libraries_fastq.py implemented in QIIME1 (Caporaso et al. 2010). Thereby a phred quality threshold of 19 was chosen. Dereplicating, size sorting, denovo chimera detection as well as OTU clustering with a 97% cutoff was conducted with vsearch 2.7.0 (Rognes et al. 2016). Finally, an OTU table was build by using the --usearch_global function in vsearch 2.7.0 (Rognes et al. 2016) followed by the python script “uc2otutab.py” written by Robert Edgar (https://drive5.com/python/uc2otutab_py.html). For taxonomy assignment sequences were blasted against the GBOL database using blastn 2.9.0+ (Altschul et al. 1990).

The resulting OTU table was curated with LULU (Frøslev et al. 2017). Curation was started with an initial blasting of OTU representative sequences against each other using blastn (version

2.9.0). The following parameter settings were chosen: 'query coverage high-scoring sequence pair percent' (-qcov_hsp_perc) was set to 80, meaning that a sequence was reported as match when 80% of the query formed an alignment with an entry of the reference file. Secondly minimum percent identity (-perc_identity) was set to 84, requiring the reference and query sequence to match at least to 84% to be reported as a match. The format of the output file was customized using the –outfmt settings '6 qseqid sseqid pident'. The resulting output file included the name of the query sequence and the name of the reference sequence next to percentage of identical match. Subsequently, the resulting filtered OTU match list was uploaded into R (version 3.5) (R CoreTeam 2013). Subsequently the R-package 'lulu' (version 0.1.0) (Frøslev et al. 2017) was used to perform post clustering curation using standard settings. The LULU algorithm filters the dataset for suspicious OTUs. Afterwards, suspicious OTUs are either classified as "daughter OTU" and merged with the corresponding "parent OTU", or are discarded from that dataset.

The resulting curated OTU table was loaded into Excel where data were formatted for upload into R (R studio running R version 3.5). For statistical analysis, several R packages were used. Marioko plots (figure IV.2 and figure IV.12) were prepared using the R package ggplot2. Data used for the preparation of Venn diagrams visualizing the number and percentage of shared and unique arthropod species with a BlastID of at least 99% between soil samples treated with two different extraction methods (figure IV.4) were processed with the R package dplyr (version 0.8.3) before final diagrams were prepared with the R package Venn.Diagram (version: 1.6.20). PCoA Plots, showing differences in arthropod species (BlastID ≥ 99%) composition between sample types (figure IV.5) were prepared using the R packages betapart (version 1.5.1) and vegan (version 2.5-6). Additionally, PERMANOVA was calculated using the R package vegan. The bar chart showing number of unique and shared species (BlastID ≥ 99%) between soil samples and Malaise traps (figure IV.9) was prepared using Microsoft PowerPoint. Underlying calculations for each group of arthropods under investigation were

conducted using the R package Venn.Diagram. Heatmaps visualizing calculated Jaccard-similarity indices (figure IV.10, figure IV.11 and figure IV.15) for arthropod species communities ((BlastID ≥ 99%) were prepared using the R package ggplot2. Underlying calculations were done in Microsoft Excel based on the formula: $J(X,Y) = |X \cap Y| / |X \cup Y|$. UpsetR plots showing detailed number of unique and shared arthropod species (BlastID ≥ 99%) within and between sample type depending on forest type and season were prepared using the R package UpSetR (version 1.4.0).

Results and Discussion

Differences and Similarities between Arthropod Communities Recovered From Malaise Traps and Soil Samples

Number of recovered OTUs, depending on sample type and extraction method

The overall number of OTUs found with Malaise traps was significantly lower than number of OTUs retrieved from soil samples (figure IV.5). One reason for the large discrepancy in OTU numbers retrieved are the two different types of DNA extracted from soil and Malaise samples. DNA extracted from soil samples is called environmental DNA (eDNA). eDNA comprises DNA from all lifeforms present in a habitat, including Metazoa, fungi and also prokaryotes (Taberlet et al. 2018). Here two different methods for extracting eDNA from soil samples were used. While the phosphate buffer exclusively targets extracellular DNA, the Macherey Nagel kit incorporates a lysis step allowing for the extraction of intracellular and extracellular DNA. The broader spectrum of DNA targeted with the Macherey Nagel kit is also mirrored in the number of detected OTUs. A total of 12855 OTUs were found with the Macherey Nagel kit, out of which 1.5% (188 OTUs) were assigned to the phylum Arthropoda (figure IV.5). DNA extraction conducted with the phosphate buffer resulted in the detection of 11431 OTUs which were slightly less OTUs than being found with the Macherey Nagel kit. Nevertheless, the number of OTUs assigned to the phylum Arthropoda (1.6%, 185 OTUs) was similarly high for the two extractions methods. Although the

Macherey Nagel kit targets a broader range of eDNA, the phosphate buffer has been described to be the better choice when targeting DNA derived from arthropods (Dopheide et al. 2019). This is due to the much higher amount of soil which can be and was finally processed with the phosphate buffer. The Macherey Nagel Kit is optimized for the assessment of microbial diversity (Zinger et al. 2016), which is also reflected in the low amount of starting material required (this kit is designed to extract DNA from 0.5g of soil). In contrast to that, here the phosphate buffer was used to extract DNA from up to 150g of soil. Invertebrate morphology, size and abundance is more heterogeneous compared to prokaryote whereas the extraction of eDNA from a larger amount of soil is likely to increase the number of recovered species (Taberlet et al. 2012b; Dopheide et al. 2019). Nevertheless, even for microbial communities it has been shown that a low amount of source material results in a less consistent and representative overview of the local microbial community (Ranjard et al. 2003). Presumably, this bias is further enhanced when targeting invertebrates because of the above outlined differences to microbial groups. A recent study has shown, that methods using a higher amount of starting material were indeed capable of recovering a broader diversity of arthropod taxa (Dopheide et al. 2019). Nevertheless, here the number of OTUs assigned to Arthropoda was similarly high between the two extraction methods, which is indicating that a lysis step enables the extraction of DNA derived from 'whole' transient individuals, which are frequently missed when exclusively extracellular DNA traces are targeted.

In comparison to the number of OTUs retrieved from the soil samples, the number of OTUs detected with the Malaise traps was significantly lower, with a total of 9218 OTUs detected across all seasons and forest types. Out of them, 95.22% were assigned to Arthropoda (figure IV.5). As already described encompasses cDNA all DNA resulting from specimens contained in a bulk sample (Deiner et al. 2017). However, here we define DNA extracted from Malaise trap catches as lysis DNA (lysDNA). In addition to DNA originating from caught specimens, extracted lysDNA also contains DNA originating from other sources such as from the gut contents of arthropods in

the sample, pollen sacks on the legs of bees, or mammal blood derived from blood sucking insects. Regarding the gut contents of arthropods in a Malaise trap it has been documented that many species tend to regurgitate their gut contents while being preserved (Zizka et al. 2018). Although preservative ethanol in the Malaise traps was exchanged at collection and finally removed before extraction, it cannot be excluded that parts of the gut contents remained in the sample. In addition, organisms contained in one bulk sample have been interacting with the environment before entering the Malaise trap, introducing other potential sources of eDNA to the sample. A recent study has shown that eDNA can be extracted from the surface of flowers, revealing the various pollinator species that have visited the flowers (Thomsen and Sigsgaard 2019). Flower pollinators may serve as potential vectors for the distribution of exogenous DNA into the Malaise traps, although this would represent only DNA traces in relation to the cDNA.

The high number of OTUs found in soil samples reflects the extremely high biodiversity in soil, also described as the third biotic frontier (Wardle 2002; Wardle et al. 2004). One gram of soil can contain several hundreds of nematodes (Hoorman 2011) and up to 5×10^4 bacterial species (Roesch et al. 2007). As a result, only between 1.46% and 1.61% of all OTUs recovered from soil (depending on extraction method) were assigned to the phylum Arthropoda. On the contrary, the proportion of all OTUs assigned to Arthropoda from Malaise traps was as high as 95%. In addition, while a total of 400 Arthropod species were identified with the Malaise traps, only 119 were recovered from the soil samples. This comparatively low number is surprising, considering former studies which have described arthropods as particularly ubiquitous and diverse in soils (Usher and Parr 1977; Oliverio et al. 2018). It has been stated that one square meter of forest soil may host up to thousands of soil arthropod species (Schaefer and Schauermann 1990). The large discrepancy between these numbers indicates that the here presented metabarcoding study encompass a high number of false negative results. Indeed, other studies have shown that metabarcoding of soil samples tends to underestimate local arthropod diversity (Yang et al. 2013; Oliverio et al. 2018). Many studies have already tried to target the main drivers for an

accumulation of false negative results. Incomplete reference databases and primer biases have been frequently mentioned (Cowart et al. 2015; Beng et al. 2016; Lopes et al. 2017). The outstanding high degree of biodiversity which can be found in soil, is directly reflected in complexity of DNA extracts, whereas primer bias is strongly enhanced (Brandon-Mong et al. 2015; Morinière et al. 2016). This is due to differences in primer affinities. While the amplification of DNA originated from well-matched taxa is supported, species with low primer affinities often remain undetected. Several authors have tried to solve this issue. It is now increasingly recommended to sequence at a higher sequencing depth (Elbrecht et al. 2017; Alberdi et al. 2018) as it increases the probability of sequencing rare and infrequent sequences. The here used sequencing depth was lower than 100.000 sequences per sample, a value which is recommended for metabarcoding of macrozoobenthos bulk samples (Elbrecht and Steinke 2019). For eDNA studies an even higher sequencing depth would be required. A former study has shown that number of OTUs recovered from faecal sample was positive correlated with sequencing depth (Alberdi et al. 2018). A plateau in number of recovered OTUs was almost reached when sequencing depth was as high as 1.5 million sequences per sample. We therefore assume that sequencing depth chosen here was not deep enough leading to the observed accumulation of false negative results. Next to primer bias and insufficient sequencing depth incomplete databases are also known to be a limiting factor for the identification of arthropods associated with the soil habitat. Estimates indicate that less than 5% of below ground biodiversity has so far been described (Wall et al. 2005) and it can be expected that even fewer species are represented in barcode reference databases. Next to that, eDNA is more prone to degradation in comparison to DNA extracted from bulk samples. A recent study has shown that degradation bias plays only a minor role for metabarcoding studies working with Malaise traps (Krehenwinkel et al. 2017). This is also true for samples left in the field under unfavorable conditions whereas it is possible to monitor species occurrence over several weeks. For soil samples it has been shown that eDNA degrades quickly (Sirois and Buckley 2019). As a result, species occurrence can only be traced

back for a narrowed time period. The here detected low number of arthropod OTUs retrieved from soil samples might therefore also been attributed to a higher degradation rate.

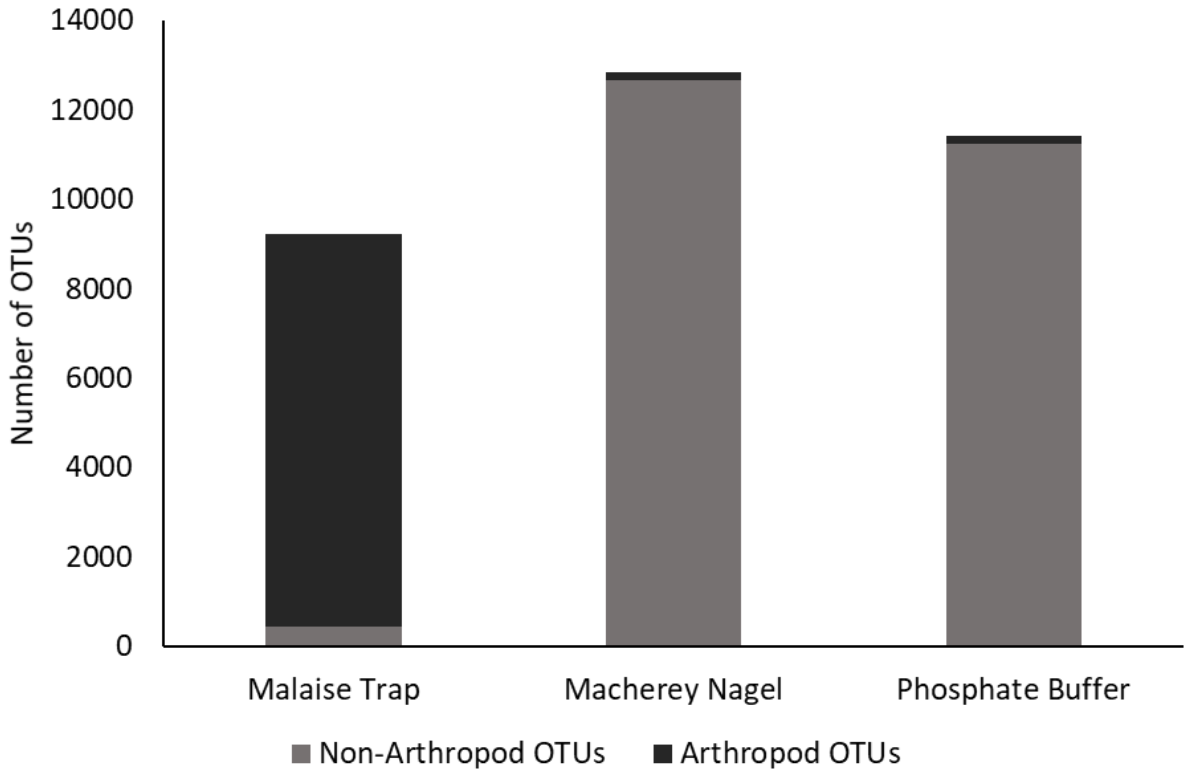


Figure IV.5 Number of retrieved OTUs (97% similarity) from the Malaise traps and soil samples depending on extraction method applied. For each dataset number of OTUs with an assignment to the phylum Arthropoda are highlighted in black.

Out of the 9218 OTUs recovered with the Malaise traps, 8777 OTUs were assigned to the Phylum Arthropoda with a blast identity of 90% or more. To reliably assign an OTU on species level, a blast identity of at least 99% was required. Out of all OTUs assigned to Arthropoda, 12.02% met these requirements resulting in the assignment of 1032 OTUs on species level. Out of the 185

arthropod OTUs recovered from soil with the phosphate buffer, 139 were assigned on species level (75%), while the Macherey Nagel Kit retrieved 135 Arthropod OTUs (71%) with an assignment to species level (figure IV.6). These results are indicating that reference databases are still suffering from a lack of completeness. Although the GBOL database is likely the most complete database for Germany's arthropods, the barcodes of several species are still missing (German Barcode of Life Consortium, 2011). In detail from the 37494 arthropod species which have been described for Germany only 45% (16906 species) have so far been barcoded. Additional 3% (1161 species) have already been collected but are still waiting to be barcoded. The remaining 52% have not been collected in the scope of the GBOL project (German Barcode of Life Consortium et al. 2011) which is highlighting that a significant proportion of arthropod taxa has so far been excluded from GBOL. It has been suggested that this is mainly contributed to insufficient expertise, which is highlighting that molecular biodiversity assessment approaches are only partly circumventing the taxonomic impediment. However, new research attempts are on its way to fill these large gaps of knowledge e.g. the GBOL III – Dark Taxa project.

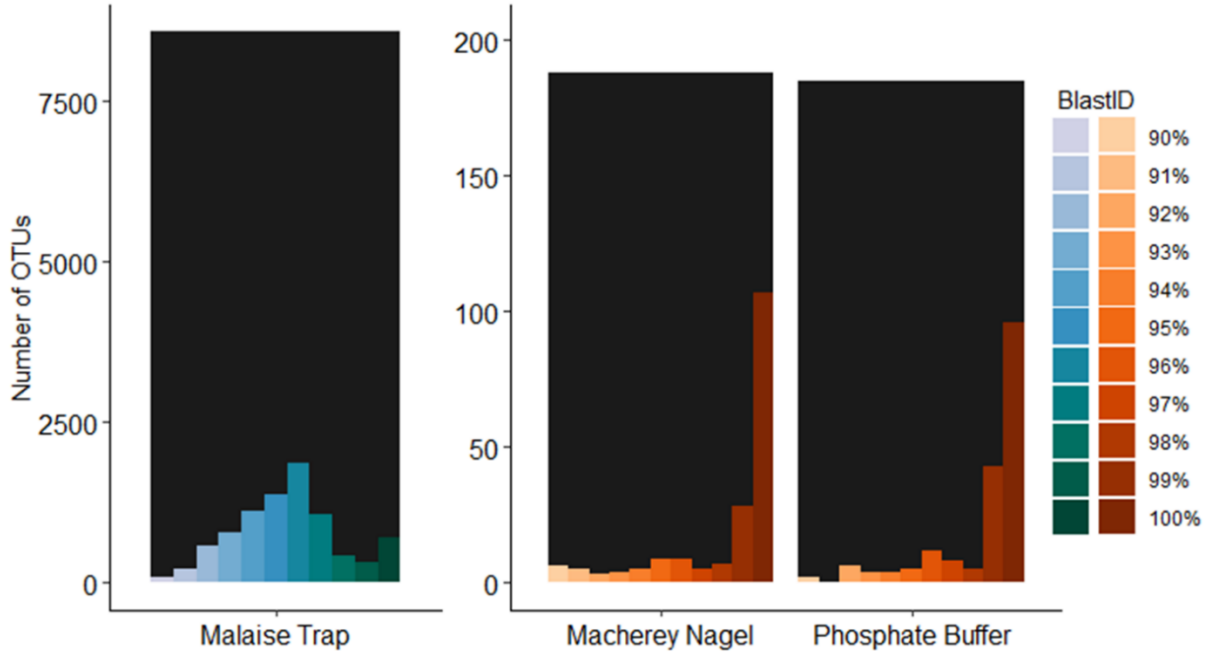


Figure IV.6 Number of OTUs assigned to Arthropoda retrieved from the Malaise traps and the two types of soil samples respectively. Next to overall number of detected arthropod OTUs (black bars) number of OTUs per blastID ranging from 90-100% are shown.

Influence of soil extraction method on detected species communities

As already mentioned, is choice of extraction method expected to directly affect number of detected species. To evaluate the effectiveness of the two soil extraction methods, the number of unique and shared OTUs assigned to Arthropoda at species level with a BlastID of at least 99% were compared. For the following analysis, all samples of each extraction type were combined (Phosphate Buffer: 162 Samples; Macherey Nagel: 162 Samples). It was found that 60 species, corresponding to half of all discovered arthropod species were detected in both types of soil samples. Additional 29 species were exclusively detected by the Macherey-Nagel kit, while phosphate buffer extraction resulted in the detection of 30 additional species (figure IV.8a). Overall, 128 species were identified from the soil samples, out of which 90 species are members

of the arthropod class Insecta. Out of them, 43 species were recovered by both extraction methods, while either one of the two extraction methods (figure IV.8b) exclusively recovered 24 and 23 species respectively.

This is highlighting the importance of choice of extraction method. As already mentioned, extraction with the Macherey Nagel Kit encompasses a lysis step which enables the extraction of intracellular DNA. In contrast to the phosphate buffer, the Macherey Nagel kit is therefore capable to target more organisms (such as hibernating individuals or individuals in the pupal stage). When comparing both extraction methods, the presence and absence of several species in the phosphate buffer samples and Macherey Nagel kit samples respectively underline this theory.

Within the Lepidoptera five species were recovered with the phosphate buffer, while eight were found with the Macherey Nagel kit (figure IV.8d). All species detected with the phosphate buffer were also retrieved with the Macherey Nagel kit. Additionally the species *Noctua pronuba*, *Phyllonorycter maestingella* and *Ypsolopha ustella* were found (figure IV.7).

Figure IV.7 Photos of voucher specimens of *Noctua pronuba, Phyllonorycter maestingella* and *Ypsolopha ustella* (if not other stated: © BY-SA 4.0: GBOL / Museum Koenig)

These three lepidopteran species have in common that they hibernate in the pupal stage, whereas time of occurrence in the soil strata is limited. Indeed, all three species were only detected in one season. Depending on life cycle stage, the three species occurred at different times of the year in the soil samples. *N. pronuba* was recovered from the soil samples in winter. The larvae of N. pronuba has been described to be present in the soil from September through March (Green et al. 2016). *P. maestingella* belongs to the group of the leafminers. The larva of *P. maestingella* mines the leaves of various *Fagus* species until it pupates on the lower side of the leaves where it also hibernates. In autumn, when the trees shed leaves, pupae of the species drop with the falling leafs to the ground (Miller 1973; Askew and Shaw 1979). Here *P. maestingella* was recovered from the soil samples taken in the autumn season, which aligns with the in the literature described occurrence of the species within the soil habitat. Within our study, *Y. ustella* was found in the soil samples taken during summer season. In contrast to the two formerly described species, is the imago of *Y. ustella* active during winter (Soszyńska-Maj and Buszko 2011). Pupae of this species are usually found in early summer between May and June .

While for the lepidopterans more species were recovered with the Macherey Nagel kit, for the coleopterans a higher number of species was found when the phosphate buffer was used for DNA extraction. In total, 35 coleopteran species were found, out of which 11 species were exclusively recovered from the phosphate buffer samples. The smaller number of five species was exclusively detected in soil samples extracted with the Macherey Nagel kit. The remaining 13 species were detected with both extraction methods (figure IV.8e). Overall, six more coleopteran species were found with the phosphate buffer. In contrast to the lepidopterans, are many coleopteran species are more closely and permanently associated with the ground layer, including active and inactive life stages. During the active phases, the species release DNA traces in the environment e.g. in form of feces. Depending on population size, the number of individuals occurring within a certain area can vary strongly. It can be assumed that the higher amount of

soil processed with the phosphate buffer resulted in a higher species detection rate as more extracellular DNA traces are contained.

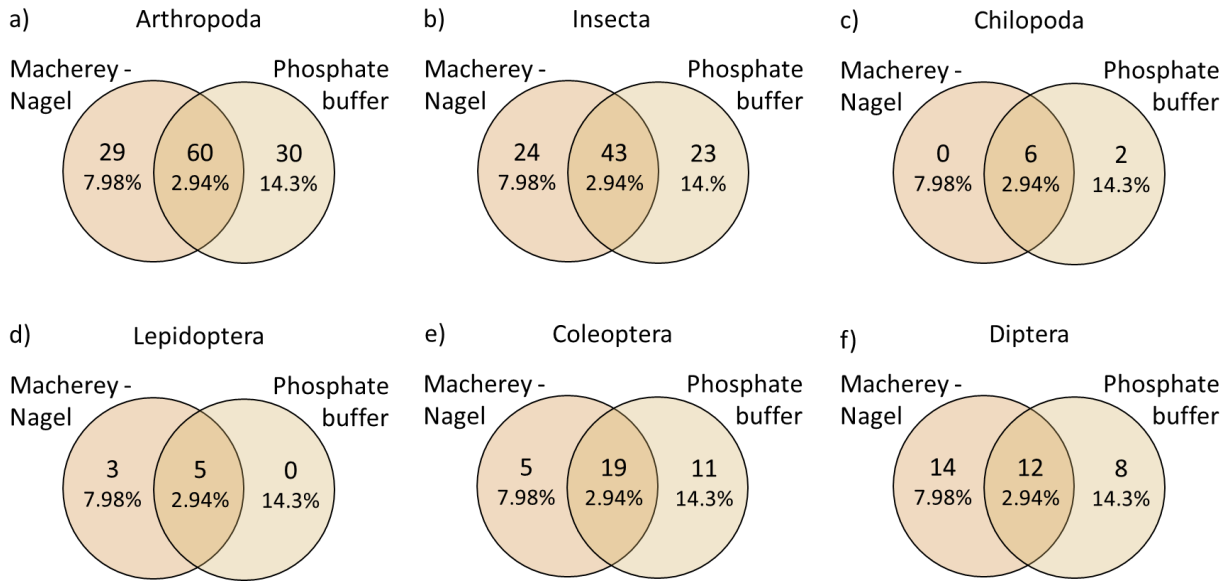


Figure IV.8 Venn Diagrams showing the number of unique and shared species (BlastID ≥ 99%) between the two types of soil samples (differing in extraction method). In total six Venn diagrams are shown, each based on a certain group a) Arthropoda b) arthropod class Insecta, c) arthropod class Chilopoda, d) insect order Lepidoptera e) insect order Coleoptera and f) insect order Diptera.

Despite these findings, the PCoA plot (figure IV.9a) did not indicate a significant difference in retrieved arthropod community composition between the two types of soil samples. When comparing arthropod species communities detected with the two types of soil samples and the Malaise traps on the basis of a PCoA plot, only two distinct clusters are formed. While the Phosphate Buffer and the Macherey Nagel Kit samples cluster together, the second cluster is formed by the Malaise trap samples. This indicates that arthropod species community composition recovered from Malaise traps is clearly different from species community

composition of soil samples, while, as expected, species communities recovered from the two soil sample extraction methods are more similar. When combining both types of soil samples, soil samples and Malaise traps form two distinct clusters (figure IV.9b). Major differences in arthropod species diversity between the two sample types is further confirmed with PERMANOVA (Pseudo-$F=12.15_1$, $p=0.001$).

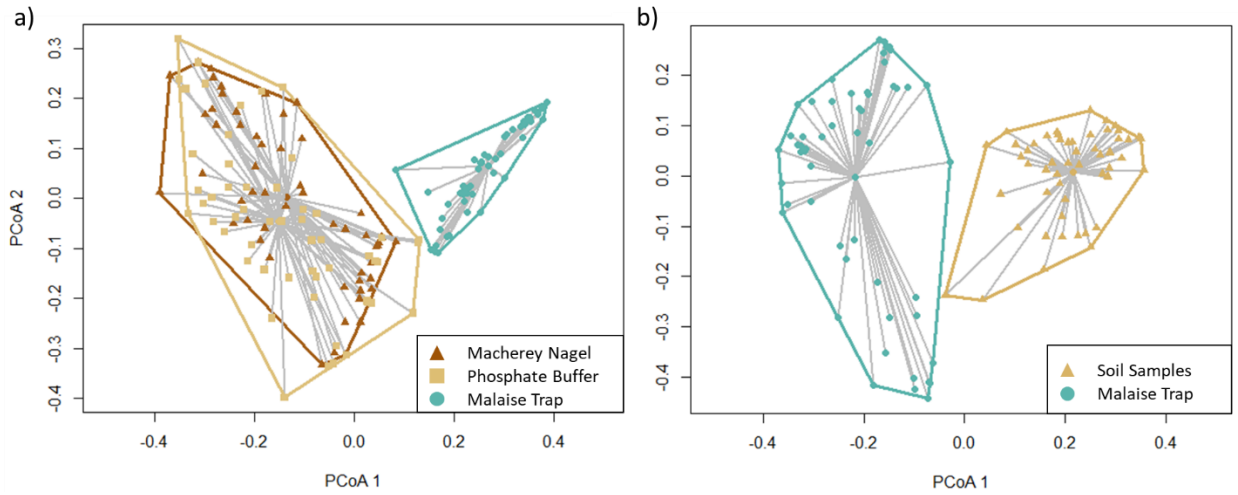


Figure IV.9 PCoA plots indicate significant differences in arthropod species communities assessed via soil and Malaise trap samples (blastID ≥ 99%). Figure IV.9a) shows that extraction method used for the extraction of eDNA from soil samples had no significant influence on recovered species community composition. For Figure IV.9b) species composition assessed with the two types of soil samples combined and compared to community composition assessed with Malaise traps. The resulting PCoA plot indicates that arthropod species community composition varies significantly between sample type which was further confirmed by PERMANOVA (Pseudo-$F=12.15_1$, $p=0.001$).

Although PCoA did not indicate significant differences between arthropod species communities assessed with either one of the two extraction methods, a high number of species only recovered with either one of the two extraction methods indicate that the two datasets complement each other (figure IV.8). In order to draw a picture as complete as possible for arthropod community

composition of the soil layer, the Macherey-Nagel kit samples and the phosphate buffer samples were combined for all further analyses. From here on, we refer to this combined dataset as 'soil samples'.

Influence of sample type (Malaise versus soil) on detected species communities

In total, 55 arthropod species were identified which were found in both types of samples - the soil and the Malaise trap samples. These 55 species accounted for 11.8% of all detected arthropod species. Overall, 74.3% of all detected species of the phylum Arthropoda were exclusively recovered from the Malaise traps (345 species), while 64 species (13.8%) were found in the soil samples but not in the Malaise traps. A closer look at the different arthropod classes reveal major differences in detected species diversity between soil samples and Malaise traps (figure IV.10).

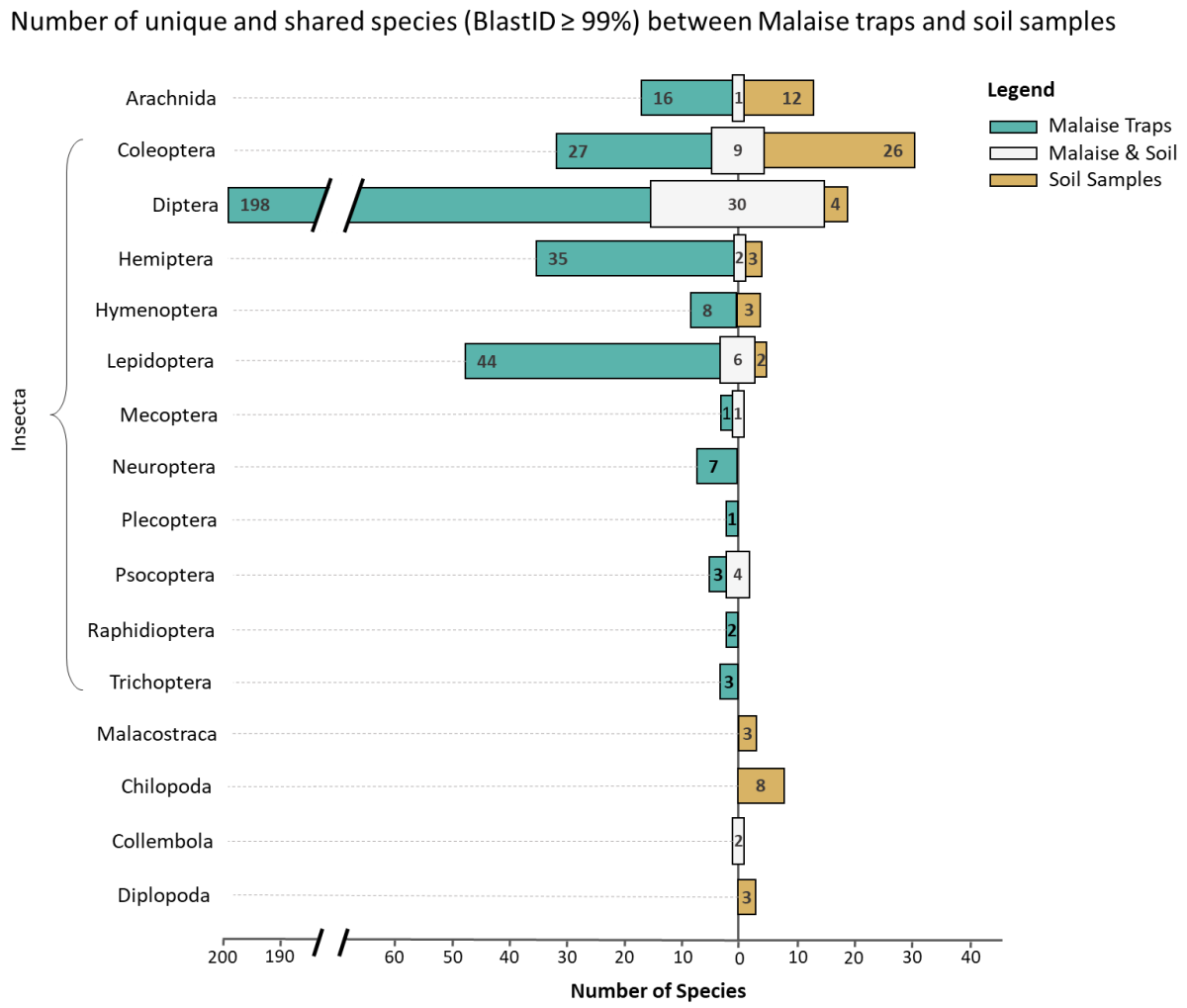


Figure IV.10 Number of unique and shared species (BlastID ≥ 99%) between Malaise traps and soil samples depending on class. For the class Insecta number of shared and unique species per order is visualized

Arachnida

A total of 29 Arachnida were detected, out of which 12 were only found in the soil samples, while 16 were exclusively recovered from the Malaise traps (figure IV.10). However, arachnids have been described to be extremely diverse in soil environments. When the conditions are suitable,

up to hundreds of thousands of individuals of the Arachnida order Acari can be found per square meter (Maraun and Scheu 2000). Surprisingly, we did not detect any species of the order in the soil samples, nor in the Malaise traps - all detected arachnid species were either members of the Arachnida order Araneae (23 species), Opiliones (six species) or Pseudscorpiones (one species) respectively. We assume that this is mainly attributed to incomplete databases, leading to false negative results. Especially Araneae are known to be a highly diverse group within forest ecosystems (Scharff et al. 2003). Spiders show a wide range of adaptions to the colonized forest habitat, including variation in foraging strategy and preferred habitat. While some species are adapted to the ground strata, some are more likely to be found in the above ground layers e.g. on tree trunks or in the canopy (Scharff et al. 2003). In this study, only a single arachnid species was present in the soil samples and the Malaise traps: *Tenuiphantes zimmermanni* (figure IV.11). The species has been described to colonize next to the leaf layer also the herb and shrub layer (Roberts 1996; Harvey et al. 2002). The detection of *T. zimmermanni* within both sample types is therefore aligned with the description of the biology if the species which can be found in the literature

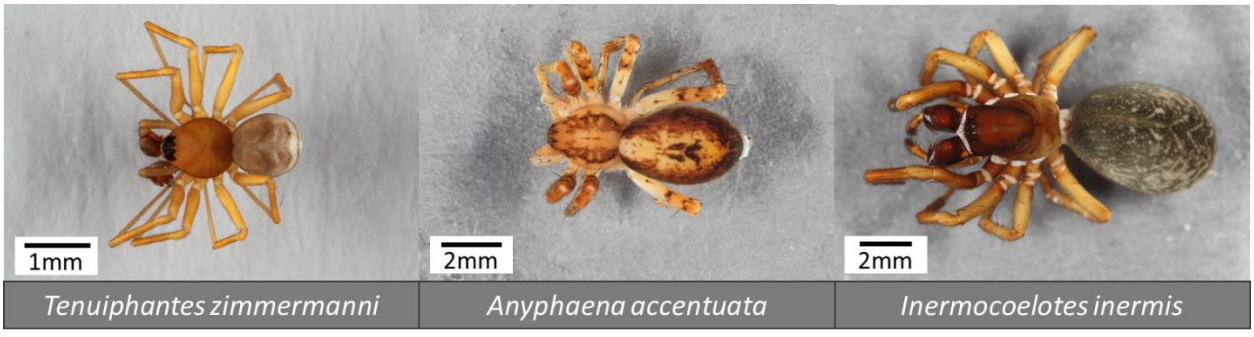


Figure IV.11 Photos of voucher specimes of the here detected Araneae species: *Tenuiphantes zimmermanni*, *Anyphaena accentuate* were recovered from the soil samples while *Inermocoelotes inermis* were found in the Malaise traps (© BY-SA 4.0: GBOL / Museum Koenig).

All further detected Araneae species were either detected within the soil samples or Malaise traps respectively. Remarkably, similar numbers of Araneae species were recovered from the soil samples (10 species) and Malaise traps (13 species).

As already mentioned, only one species was recovered from both sample types, indicating two different arachnid communities. The effectiveness of Malaise traps for assessing spider community composition has long been under discussion. General techniques such as sweeping, beating, suction samplers and pitfall traps (Churchill and Arthur 1999; Ludy and Lang 2004) were thought to be more efficient for assessing spider richness and diversity, although most of these techniques are dramatically more labor-intensive. Nevertheless, in the course of the last years several studies using Malaise traps have noticed that spiders are a common bycatch (Geiger et al. 2016; deWaard et al. 2019). This led to the question whether Malaise traps could also be useful for the assessment of local spider diversity as well as flying insect diversity. Oxbrough et al. (2010) finally tested the efficiency of Malaise traps for assessing local spider diversity in comparison to pitfall traps. They found that both methods were equally efficient in terms of numbers of retrieved species, but that the sampling devices target different species, resulting in the detection of two different spider communities. Given the fact that pitfall traps mirror ground diversity, the results presented here are supporting these findings. Nevertheless, Oxbrough et al. (2010) found that while the proportion of arboreal, web-building species was higher in Malaise traps, actively hunting species were more likely to be found in pitfall traps (Oxbrough et al. 2010). Further studies made similar observations (Jennings and Hilburn 1988). However, the study presented here is contradicting these findings. In detail, the only active hunting species detected was recovered from the Malaise traps (*Anyphaena accentuata* (figure IV.11)) (Kreuels and Buchholz 2006), while all species recovered from the soil were net-building predators. Instead of foraging strategy, association with habitat type seemed to be the major driver for determining whether a species was recovered with either of the two sampling methods. All species recovered from the soil samples are known to be typical inhabitants of the litter layer or the soil itself (e.g.

Inermocoelotes inermis (figure IV.11)). With the exception of *A. accentuata,* all species detected in the Malaise traps are net-building species which are usually found in the shrub layer. Although the description of recovered species in the literature matches our findings, the completeness of the assessed Araneae community is questionable. Several other studies have already shown the Arachnid diversity associated with temperate forests to be more diverse (Hsieh and Linsenmair 2012; Samu et al. 2014).

Malcostraca, Chilopoda & Diploda

As expected, species belonging to the three arthropod classes Malacostraca, Chilopoda and Diploda were found in the soil samples but not in the Malaise traps. This is due to the fact that all three classes are mainly associated with the ground strata as they lack the ability to fly and only occasionally climb up to higher forest strata.

Collembola

Two species of the class Collembola were found, which were both present in the soil samples as well as in the Malaise traps. Although Collembola are mainly described as colonizers of the soil and litter layer, many studies have already shown that collembols are also common inhabitants of the upper forest strata (Palacios-Vargas et al. 1998; Yoshida and Hijii 2006). Despite lacking the ability to fly, they are often found entering Malaise traps by climbing up the net (Fjellberg 1992). Despite the fact that the class Collembola has been described as one of the most diverse arthropod groups, the order was for a long time outside the focus of scientific interest. Therefore, the group is highly underexplored and several species are still waiting to be described (Rusek 1998). As a result, the reference databases are very incomplete, possibly leading to a large number of false negative results.

Insecta

Regardless of sample type, the class Insecta accounted for the highest proportion of recovered species. Insects were represented by 419 species out of which 329 were exclusively recovered from the Malaise traps (78.5%). In contrast, only 38 species (9.1%) were exclusively found in the soil samples, while the remaining 52 species (12.4%) were shared between both types of samples (figure IV.10). Because of the high functional and genetic diversity of the class Insecta, we decided to analyze this group at order level. In total 10 insect orders were detected, all of them had in common that the number of species recovered from the Malaise traps clearly exceeded the number of species found in the soil samples (figure IV.10). Moreover, all species of the orders Plecoptera, Neuroptera, Raphidioptera and Trichoptera were exclusively found in the Malaise trap. This already indicates that, as expected, flying species are more likely to be missed with soil sampling. This was further confirmed when analyzing data from highly diverse orders dominated by flying species such as the dipterans, hymenopterans and lepidopterans. All three orders were significantly less diverse in the soil samples. In detail, only four dipteran, three hymenopteran and two lepidopteran species were exclusively recovered from the soil samples, while 198, 8 and 44 species respectively were exclusively detected in the Malaise traps. For the hymenopterans no overlap between the two sample types were found (figure IV.10), although this group contains a wide range of species constantly occupying the ground as well as the above ground stratum, such as several ground nesting bees and wasps. Indeed, approximately 70% of all described bee and wasp species nest in the ground (Cope et al. 2019). However, the hymenopterans are one of the most difficult arthropod groups to target with metabarcoding using universal primers. Former studies have already shown that hymenopteran sequences tend to show a low affinity to primers compared to other arthropod group. With increasingly complexity of DNA mixture several hymenopteran species increasingly remain undetected (Brandon-Mong et al. 2015). The most diverse group of flying insects detected were the dipterans. In total 232 dipteran species were found out of which 198 species (85%) were exclusively detected in the Malaise traps, 30 species

were shared between soil samples and Malaise traps and only the four species *Corynoptera winnertzi*, *Pyratula zonata*, *Synapha fasciata*, *Tanytarsus ejuncidus* (figure IV.12) were exclusively recovered from the soil samples.

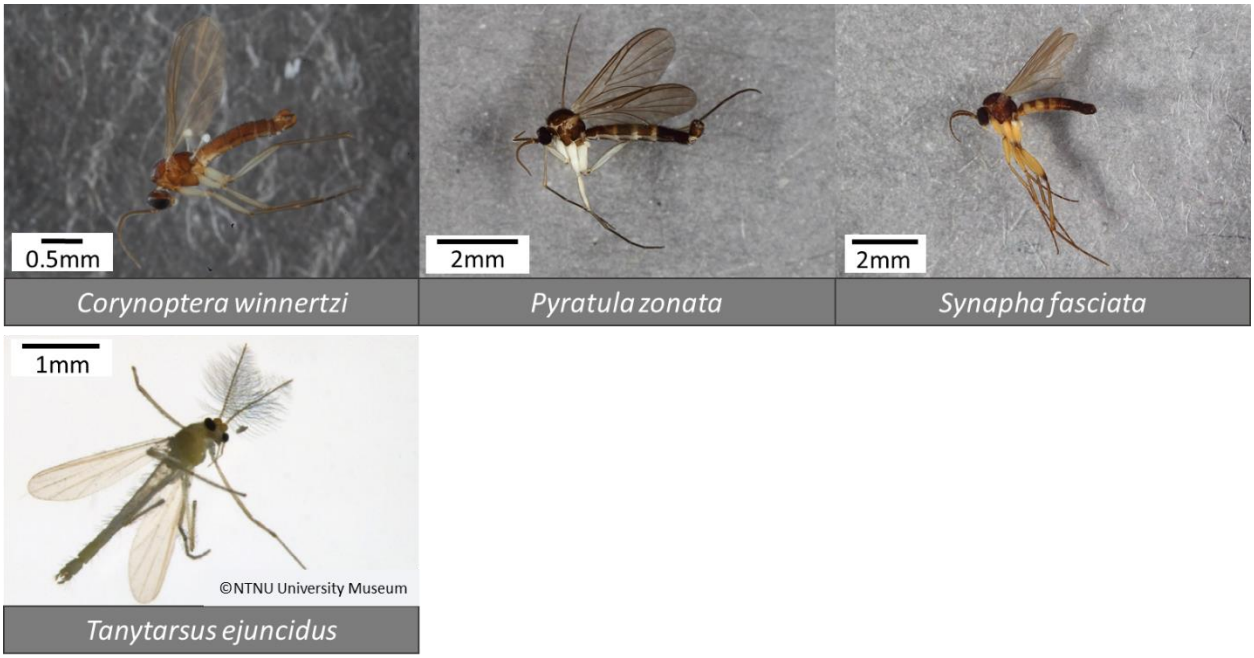


Figure IV.12 Photos of voucher specimens of the here exclusively recovered from the soil samples recovered dipteran species *Corynoptera winnertzi*, *Pyratula zonata*, *Synapha fasciata*, *Tanytarsus ejuncidus* (if not other stated: © BY-SA 4.0: GBOL / Museum Koenig)

The reasons why these four species were not part of the species community recovered from the Malaise traps can be manifold. Former studies have already shown that the detection of some species can be challenging with molecular methods when they are part of a bulk sample (Brandon-Mong et al. 2015). Because of different primer affinities (primer bias) (Hajibabaei et al. 2011) and variations in amount of DNA, especially less abundant and small species are often missed (biomass bias) (Elbrecht et al. 2017). Furthermore, sampling with Malaise traps was only

conducted over a rather short period of time. It is possible that no flying specimens were present in the habitat at the time of sampling, but that eggs and larvae of the species were present in the soil.

As already mentioned, the order Diptera was the most diverse group in the Malaise trap catchment. Out of the 400 arthropod species recovered, 228 were dipterans accounting for more than half of all recovered species. In the soil sample, a total of 34 dipteran species were found, making up 28.6%. This indicates that the soil is indeed a rather poor integrator for total existing biodiversity. Nevertheless, for the assessment of other groups, soil samples can be a valuable complement to Malaise traps catches, such as for coleopterans.

In total, 62 Coleoptera species were found out of which only nine were present in both sample types: the soil and the Malaise traps (figure IV.9). 27 coleopteran species were exclusively detected in the Malaise traps and 26 respectively in the soil samples. Nine further species were recovered from both sample types. The efficiency of Malaise traps to sample coleopterans is highly dependent on biology of the target species. Although former studies have shown that Malaise traps can efficiently capture a wide range of flying coleopteran species (Hosking 1979), this kind of trap is not suitable for the detection of ground dwelling species. Indeed, almost 43% of all coleopteran species were not recovered with the Malaise trap. A closer look at the species list reveals that especially species of the families Staphylinidae and Curculionidae were exclusively found in the soil samples. Within the Curculionidae a total of 11 species were detected of which six were not found in the Malaise traps. All six species (*Acalles ptinoides*, *Barypeithes araneiformis*, *Barypeithes pellucidus*, *Otiorhynchus scaber*, *Sciaphilus asperatus* and *Strophosoma capitatum* (figure IV.13)) are wingless weevils (Gratwick 1992; Cateau et al. 2016), which means they are unlikely to be detected by a flight interception trap.

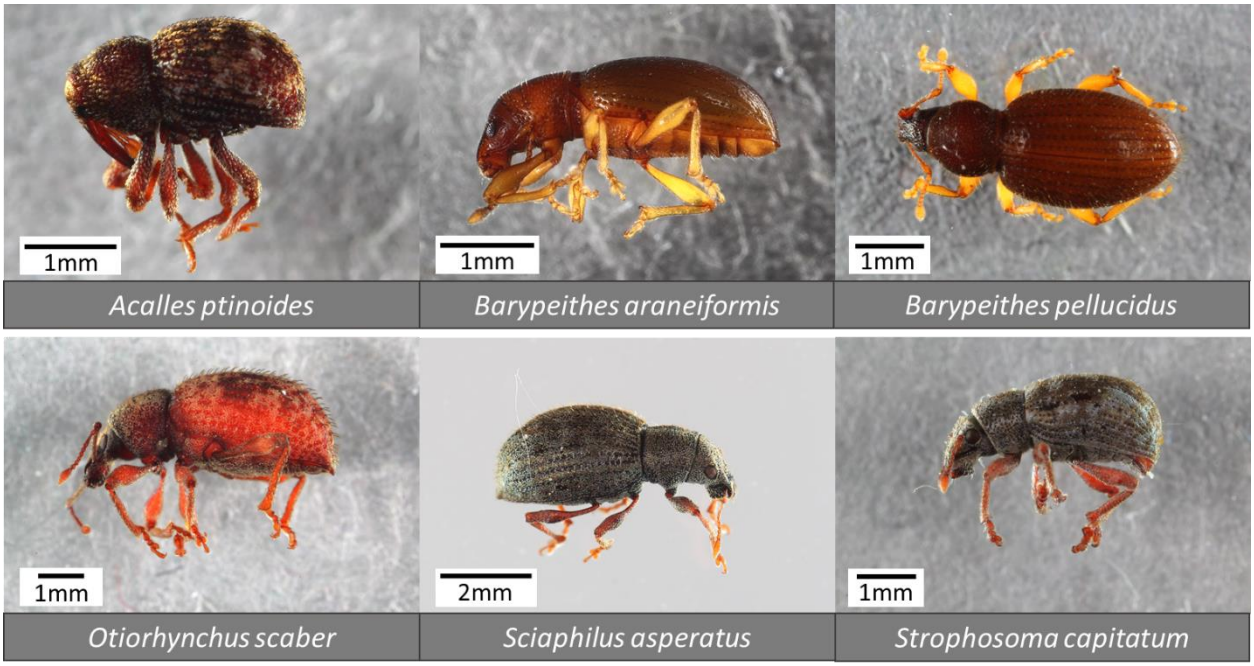


Figure IV.13 Wingless weevil species recovered from soil samples via metabarcoding: *Acalles ptinoides*, *Barypeithes araneiformis*, *Barypeithes pellucidus*, *Otiorhynchus scaber*, *Sciaphilus asperatus* and *Strophosoma capitatum* (© BY-SA 4.0: GBOL / Museum Koenig).

For the Staphylinidae, only four out of 14 species were detected in the Malaise traps. Although all 14 species have wings, the ten species exclusively recovered from the soil sample have in common that they are members of ground dwelling genera (Balog et al. 2003; Betz et al. 2018). We suggest that species lifestyle significantly influences whether a species is recovered from either of the two sample types. For the sake of completeness, it should be mentioned that the low overlap of species recovered from both types of sample may also result from methodological issues. In former experiments (see Chapter II) it was shown that out of 26 morphologically identified coleopteran families found in a total of six bulk samples, nine families were not recovered with metabarcoding. Similarly to this study DNA extraction was performed in a non-destructive manner, by overnight lysis in a lysis buffer. The effectiveness of lysis buffer on highly sclerotized insects has not been fully tested, but it has been shown in a recent study that highly

sclerotized insects release less DNA into the preservative ethanol than soft tissued ones (Marquina et al. 2019). A significantly lower amount of DNA from coleopterans could have led to false negative results as the percentage of DNA with which a species contributes to the DNA mixture determines whether a species is recovered (Elbrecht et al. 2017). Further experiments would need to be conducted using mock communities and testing non-destructive overnight lysis versus homogenization of the whole sample.

This study clearly demonstrates that Malaise traps and soil samples represented only a fraction of the total detected arthropod diversity, resulting in the characterization of two significantly different arthropod communities. This result is in accordance with prior conducted studies (Marquina et al. 2019). Additionally, similar results were also obtained from aquatic habitats. A former study investigated the effect of four commonly used eDNA substrates to explore taxonomic diversity of coastal ports (Koziol et al. 2019). As each of the four substrates drew a significantly different picture of the existing local biological diversity, the study concludes that an analysis based on a single substrate will likely underestimate total local diversity, which we can now also confirm for terrestrial habitats.

Seasonal Variations in Arthropod Species Communities Depending on Sample Type

When monitoring arthropod diversity over the duration of a full year, variations in the number of detected species and the composition of arthropod communities occur between seasons. This is mainly due to the fact that arthropods are poikilothermic organisms. Roughly said, their body temperature depends on the outside temperature (Szujecki 1986). The seasonal developmental cycle of many arthropods is therefore strongly determined by day length, as this is one of the most reliable predictors for upcoming abiotic conditions (Gullan and Cranston 2014). In

temperate zones, changes in day length usually predict changes in solar radiation, temperature and humidity. These factors highly influence seasonal conditions like food availability, whereas reliable predictions about upcoming changes allow for a perfect timing for mating and the development of the different life stages (Gullan and Cranston 2014). Day length thereby acts as a signaling factor stimulating a complex physiological process during which endocrine glands release hormones, which in turn activates the sexual and moulting glands (Szujecki 1986). As well as reproductive processes, the onset of diapause during winter is strongly driven by changes in day length, light intensity, incoming wavelength as well as temperature (Gullan and Cranston 2014). Although there is a wide consensus that arthropod and in particular insect activity is directly linked to time of year regardless of habitat preference, arthropod species communities recovered from the soil sample and from the Malaise traps showed different dynamics in the course of the year. In order to specify the main drivers for species variations within and between forest types depending on season we calculated the corresponding Jaccard-similarity indices for each of the two datasets. Within the soil samples the highest Jaccard-indices were found within the two monocultures between seasons (figure IV.14). In detail, arthropod communities of the two monocultures showed the highest Jaccard-similarity index between autumn season and the winter and spring season respectively (figure IV.14). This indicates that community composition associated with the ground layer of the two monoculture stands was more constrained by forest type than by seasonal variation.

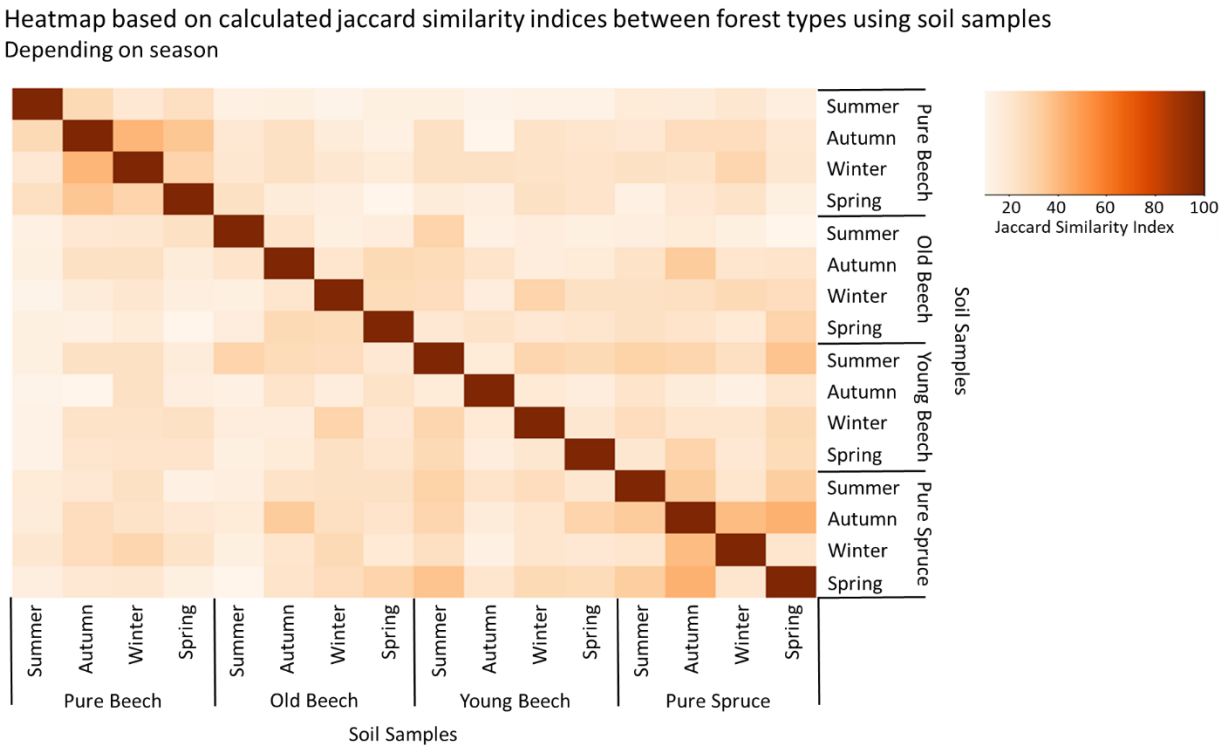


Figure IV.14 Heatmap showing variations in jaccard similarity indices between forest types depending on sampling season. The calculation was based on assessed arthropod species community composition assessed via soil samples

A different picture was found for arthropod species communities recovered from the Malaise traps. In contrast to the soil samples, Jaccard-similarity indices between forest types were higher when samples were taken at the same time of the year but at different locations (figure IV.15). This indicates that arthropods associated with the above ground habitat are more likely than ground dwelling species to be found in several habitats within one season.

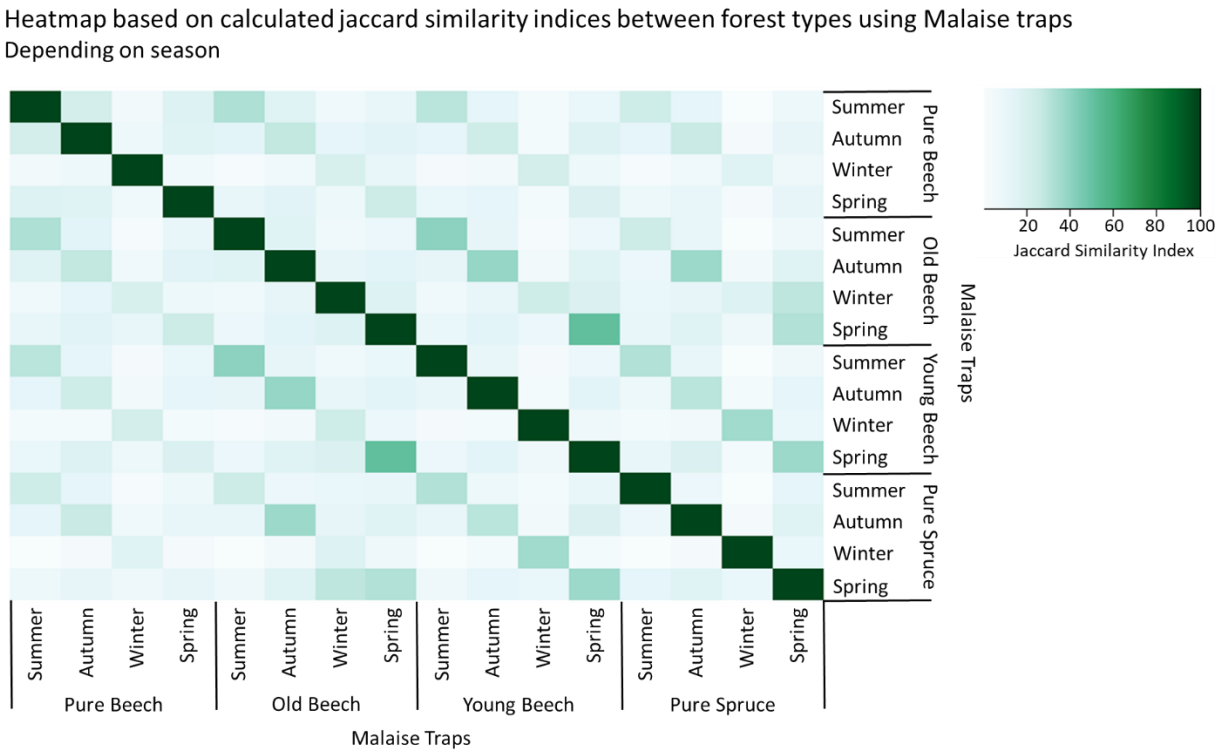


Figure IV.15 Calculated heatmap shows variations in jaccard similarity indices between forest types depending on sampling season. The calculation was based on arthropod species community composition assessed via Malaise traps.

These findings are not surprising. As already shown, several arthropod classes like Chilopoda, Malacostraca and Diplopoda were exclusively recovered from the soil samples, while the number of flying arthropod species recovered from the Malaise traps exceeded that found in the soil samples many times over. On the one hand, this clearly points to the fact that soil is indeed a poor integrator for flying arthropod diversity. On the assumption that eDNA extracted from the soil samples originated to a large degree from species directly interacting with the ground layer, instead of from traces coming from species inhabiting the forest canopies, the above described findings are not surprising. Flying arthropods can travel long distances and easily access several habitats. This is further supported by the fact that within each season and forest type several

species were only found in one of the three and respectively four Malaise traps. In a former study we were able to show that in summer season, when insects were most diverse, between 40% and 52.3% of all insect species detected within a certain forest type were only recovered from one of the three Malaise traps (Chapter II). This high portion of non-overlapping species between traps located in a single forest type may point to the fact that several species occur only in low abundances or secondly that several species are highly active and are just passing through a habitat. Species associated with the soil are less mobile and the highest portion of them won't travel long distances. Nevertheless, seasonal dynamics also occurred in arthropod species communities of the ground layer (figure IV.16).

As already mentioned, the overall number of arthropod species recovered from the Malaise traps exceeded the number of species found with the soil samples three fold (figure IV.16). Depending on sampling season, this proportion shifted towards the soil samples. For the Malaise trap samples, the highest total number of arthropod species was found in the summer season (251 species), followed by autumn (137 species) and spring (118 species). A low was reached in winter, when only 38 arthropod species were recovered (figure IV.16a). A different pattern was found for arthropod diversity recovered from the soil samples. In summer, autumn, winter and spring the number of recovered arthropod species differed only slightly. The lowest number of arthropod species was found in autumn with 45 species. Two more species were found in summer season (47 species) and an additional five species were detected in spring (52 species) (figure IV.16a). In contrast to the Malaise trap samples, the highest number of arthropod species recovered from the soil samples was found in winter (figure IV.16a). Similar findings were observed for the arthropod class Insecta (figure IV.16b). As already explained, Malaise traps are flight interception traps, the efficiency of which is directly linked to the degree of activity of existing flying insect communities. As a result of decreasing temperatures and low level of light in the winter months, many insects enter their hibernation state in later autumn. As a result, the number of insects caught in the winter months with Malaise traps decreases dramatically. In

contrast, an accumulation of arthropod species can be observed in soil samples during winter. This can be best explained by choice of hibernation site. While some insects prefer the bark of trees, dead wood, hedges and meadows, many insects actively seek leaf litter or dig soil chambers for hibernating. As a result, an accumulation of species in the soil can be observed in winter.

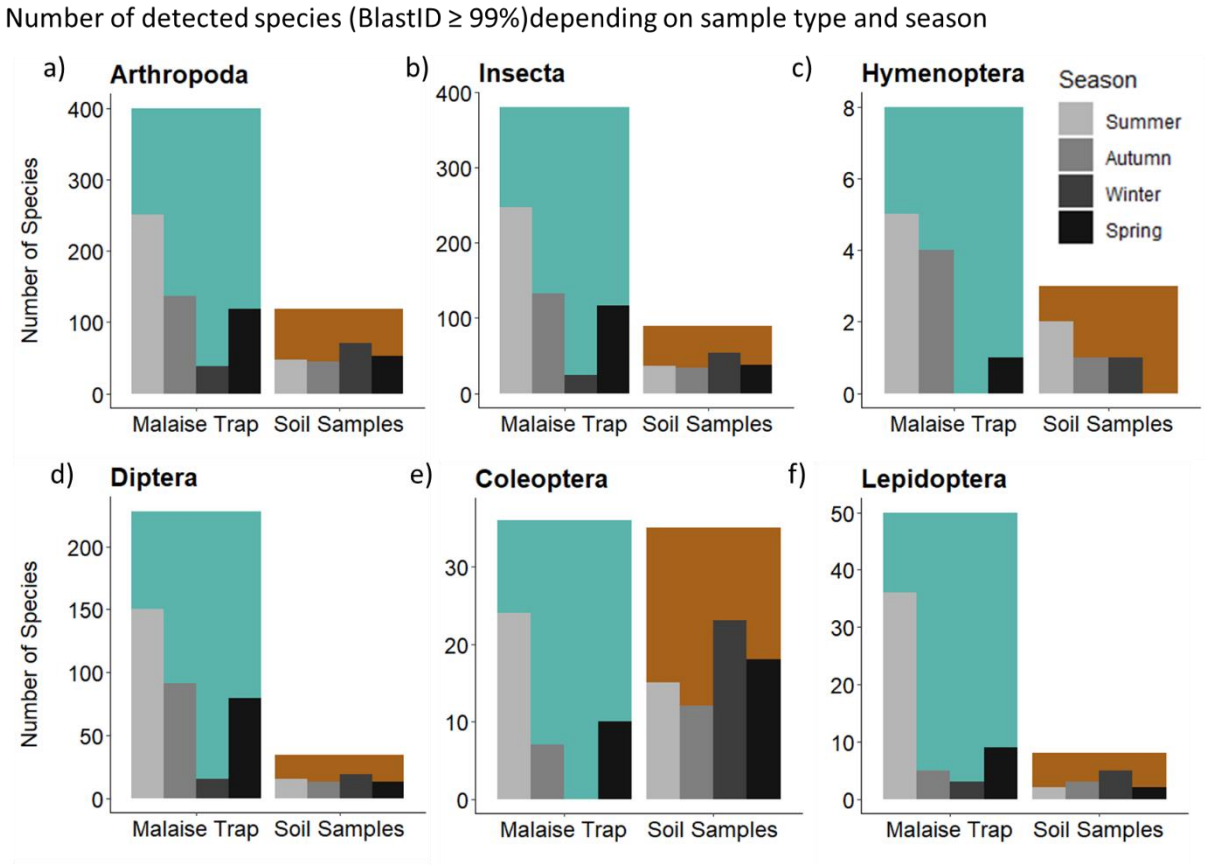


Figure IV.16 Number of detected arthropod species (blastID ≥ 99%) per sample type depending on sampling season. The colored bars indicate the total number of detected species per sample type and method, irrespective of forest type and season. The grey shaded bars represent numbers of detected species per season.

Hymenoptera

Eight species of the order Hymenoptera were found with the Malaise traps. Out of them five were detected in the summer season, four in autumn and one in spring, while none was detected in winter (figure IV.16c). From the soil samples, less than half as many species were recovered. Over the one year period, only three hymenopteran species were detected (figure IV.16c). Two of them belong to the order Formicidae (*Formica sanguinea* and *Myrmica ruginodis* (figure IV.17)), which are known to spend most of their lifetime on the forest floor. Additionally, we found the species *Cephalcia lariciphila* (figure IV.17), a member of the family Pamphiliidae.

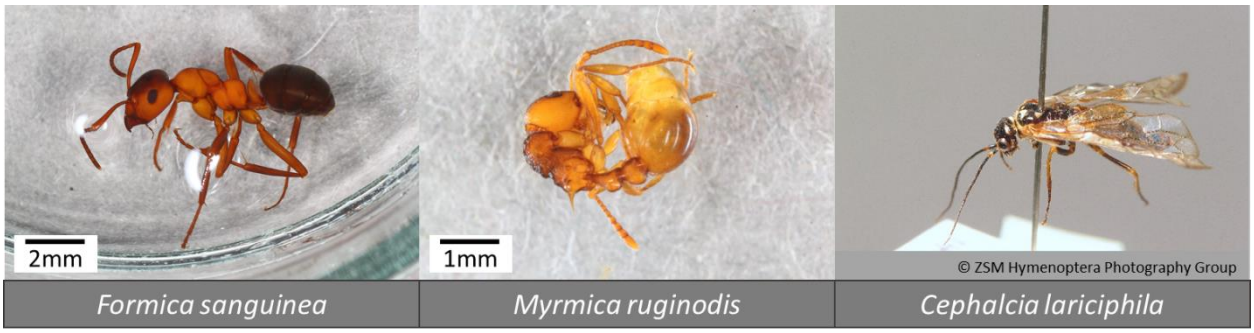


Figure IV.17 Photos of voucher specimens of the three Hymenoptera species *Formica sanguinea*, *Myrmica ruginodis* and *Cephalica lariciphila* (if not other stated: © BY-SA 4.0: GBOL / Museum Koenig)

C. lariciphila is usually found in coniferous forests where the larva feeds on the needles of the trees. Four weeks after hatching, the mature larva drops to the ground, where it quickly digs into the soil (EFSA Panel on Plant Health (PLH) et al. 2017). The larvae then builds an earth-walled chamber at a depth of 5-20 cm, where it finally pupates. The flight period of *C. lariciphila* is rather short. After emerging from the soil between May and June, adult individuals directly mate, before the females fly to the canopies for oviposition. Interestingly, females have also been observed

reaching the canopies by climbing up the tree trunk, rather than by flying (EFSA Panel on Plant Health (PLH) et al. 2017). The short flight period combined with a rather short sampling period can be expected to be the reason why no specimens of *C. lariciphila* were found in the Malaise traps. *C. lariciphila* was found in the soil samples taken at the pure spruce sites in autumn. Nevertheless, we would have expected to find the species also at other times of the year as it has been described that the larvae can remain in the soil for several years (EFSA Panel on Plant Health (PLH) et al. 2017). Again methodological issues could be made accountable for this. The larva remains inactive in its earth walled chamber until its emergence in May and June. During the inactive phase it could only be detected with the Macherey-Nagel kit, which is designed for the extraction of 0.5g of soil. The likelihood of recapturing the species again in a random sample of 0.5g of soil is rather low. Nevertheless, when interpreting these results it should be kept in mind that hymenopterans are one of the most difficult insect groups to target with metabarcoding using a universal primer (Brandon-Mong et al. 2015). The number of hymenopteran species recovered from Malaise traps (8 species) seems rather low, considering hymenopterans are among the most diverse insect orders. Studies have shown that several hundred parasitic hymenopteran species can be present in a single beech forest (Ulrich 1999a, 1999b). Universal COI primer pairs have already been shown to be unsuitable for the detection of hymenopterans contained in a bulk sample (Brandon-Mong et al. 2015), which is presumably even complicated further when working with eDNA. With further research, it should be possible to design a COI primer which fits better to the order Hymenoptera.

Lepidoptera

Within the Malaise traps a total of 50 lepidopteran species were identified. 36 species were present in summer, while only five, three and nine species were found in autumn, winter and

spring respectively. Within the soil samples, the highest number of lepidopteran species was found in the winter season (figure IV.16f).

Two Lepidopteran species were exclusively recovered from soil. *Nematopogon robertella* and *Nemophora congruella* (figure IV.18) have both been observed to have a rather short flying period from May until the end of June, an interval during which Malaise sampling was not conducted. The fact that both species were recovered from soil further confirms our hypothesis that soil samples can supplement sampling through Malaise traps when assessing arthropod diversity in a habitat.

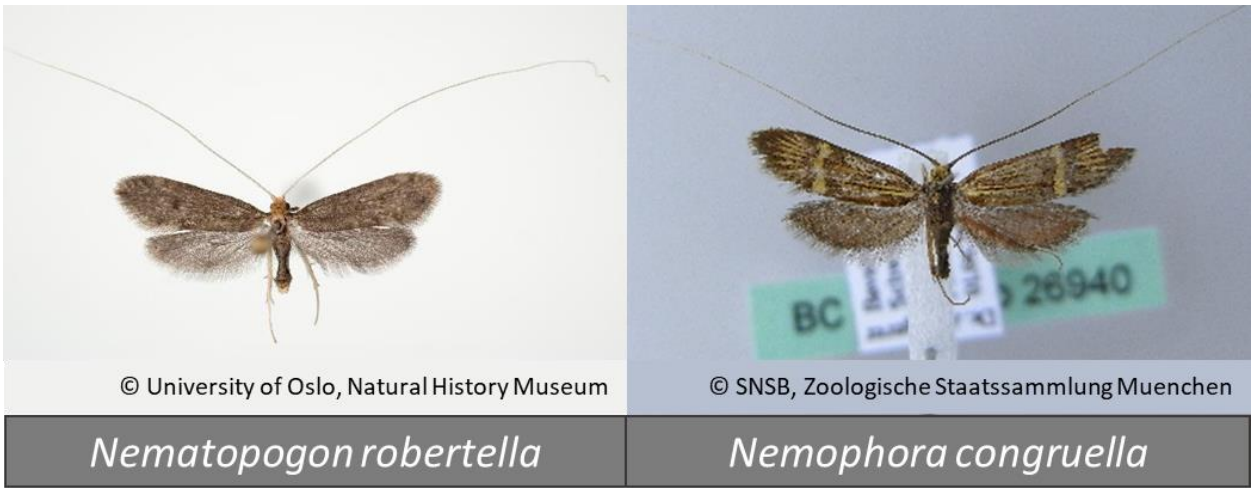


Figure IV.18 Photos of voucher specimens of the two Lepidoptera species *Nematopogon robertella* and *Nemophora congruella*

Nevertheless, it should be kept in mind that likelihood of detection is closely linked to species lifestyle. Our study shows that only species of which at least one life stage is associated with the ground layer are recovered from soil samples, suggesting that DNA is recovered from the dormant life stages (eggs, pupae, larvae) themselves.

Coleoptera

The total number of coleopteran species recovered from all samples of a single sample type was highest for the Malaise traps (36 species). A closer look at each season revealed that this pattern was only found in the summer season, while the number of detected species in autumn, winter and spring was higher in the soil samples (figure IV.16e). Interestingly, the highest number of species in the soil was found in the winter season, while with the Malaise traps no coleopteran species were found in the same season (figure IV.16e). In winter season, a total of 25 species was recovered from the soil samples. At the same time of the year no coleopteran species was captured with the Malaise traps. Overall, out of the 25 species detected in the soil samples in the winter season only eight were also present in the Malaise traps at some time of the year. As already mentioned, several coleopteran species are adapted to a ground dwelling lifestyle and flights of these species can only be occasionally observed. This dramatically lowers efficiency of Malaise traps when aiming to assess total existing coleopteran diversity. The reduced sampling success with Malaise traps directly leads to a limited species overlap between Malaise traps and soil samples.

Nevertheless, out of the 25 coleopteran species recovered from the soil samples in winter nine species were present in soils all year round while six species appeared in winter and in at least one additional season. The remaining ten species were recovered from soil in winter but had no record in any other season. Out of these ten species, three species (*Malthodes mysticus*, *Malthodes fuscus* and *Ernobius abientinus*) were also detected with Malaise traps but in summer and spring respectively. The remaining seven species (*Acalles ptinoides, Ampedus nigrinus, Otiorhynchus scaber, Propylea quatuordecimpunctata, Pteryx suturalis, Sciaphilus asperatus, Xantholinus laevigatus*) were exclusively recovered from the soil samples of the winter season. This underlines the necessity of combining several sampling strategies and seasons. Next to the assessments of species which cannot be detected with Malaise traps like the wingless weevils *A.*

ptinoides, *O. scaber, S. asperatus* (figure IV.13) and the ground dwelling Staphylinidae species *X. laevigatus* (figure IV.19), the use of several sampling strategies also increases the chance of sampling rather frequent species like *P. quatuordecimpunctata* (figure IV.19).

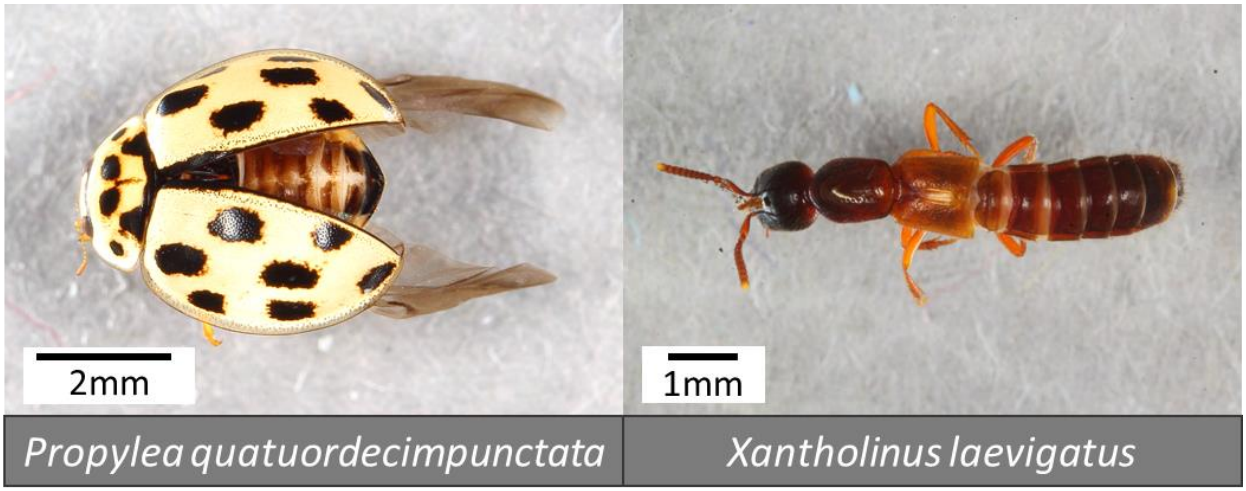


Figure IV.19 Photos of voucher specimens of the Coleoptera species *Propylea quatuoredecimpuctata* and *Xantholinus laevigatus* (© BY-SA 4.0: GBOL / Museum Koenig)

P. quatuordecimpunctata is a common beetle within several habitats (Pervez 2011). The species is specialized on feeding on aphids and other small arthropods, forcing it to travel through the habitat when searching for prey. In the warm summer months the species is highly active and flight can be observed on a regular basis. Nevertheless, *P. quatuordecimpunctata* was not found in the Malaise trap bulk samples. A possible explanation might be that coleopteran are known to be able to evade Malaise traps. Several species have been observed dropping to the ground after hitting the black middle mesh of the Malaise traps instead of flying upwards. Furthermore, like several other insects, coleopterans are only active during a certain time window. *P. quatuordecimpunctata* is known to have a rather long life span. The adult beetles have been observed to live up to 2 years, forcing them to hibernate twice (gbif.org). A typical hibernating

ground is the leaf layer close to the ground. As a result, probability increases that the species will be found in winter in soil samples.

Time-lagged Overlap of Arthropod Species Occurrences between Forest Strata

Overlap between soil bound and flying arthropod species across the seasons

Calculated jaccard similarity indices between soil samples and Malaise traps based on arthropod species identified with a blastID of at least 99% indicate that some species occupy different habitats at different times of the year (figure IV.18). Jaccard similarity index was calculated between the Malaise traps and soil samples depending on season and forest type (figure IV.18). Within all forest types, the highest similarity indices were found when soil samples were compared to Malaise traps of the summer season. Indeed, subsequently calculated UpsetR plot showed that malaise trap samples from summer had the highest overlap with soil samples (figure IV.20).

Four species were recovered from the Malaise traps in summer season, which were present in the soil throughout the year (*Athous subfuscus* (Insecta: Coleoptera), *Corynoptera minima* (Insecta: Diptera), *Ctenosciara lutea* (Insecta: Diptera), *Phymatopus hecta* (Insecta: Lepidoptera). Further six species were detected in the Malaise trap catchments of the summer season, which were also present in the soil samples of the winter season. Additionally, three species were detected in the soil samples taken in spring and winter and three species in the soil samples from autumn and winter season. Malaise traps of no other season had similarly high jaccard indices

when compared to the soil samples. Especially during the warm summer months, above ground arthropods are highly active (Szujecki 1986), which is also mirrored in the high number of arthropod species recovered in the summer season using Malaise traps (figure IV.16). Nevertheless, our results indicate that a large proportion of these species are seeking different habitats in the course of the year, mirroring the complex life cycle of arthropods (table IV.4, figure IV.20).

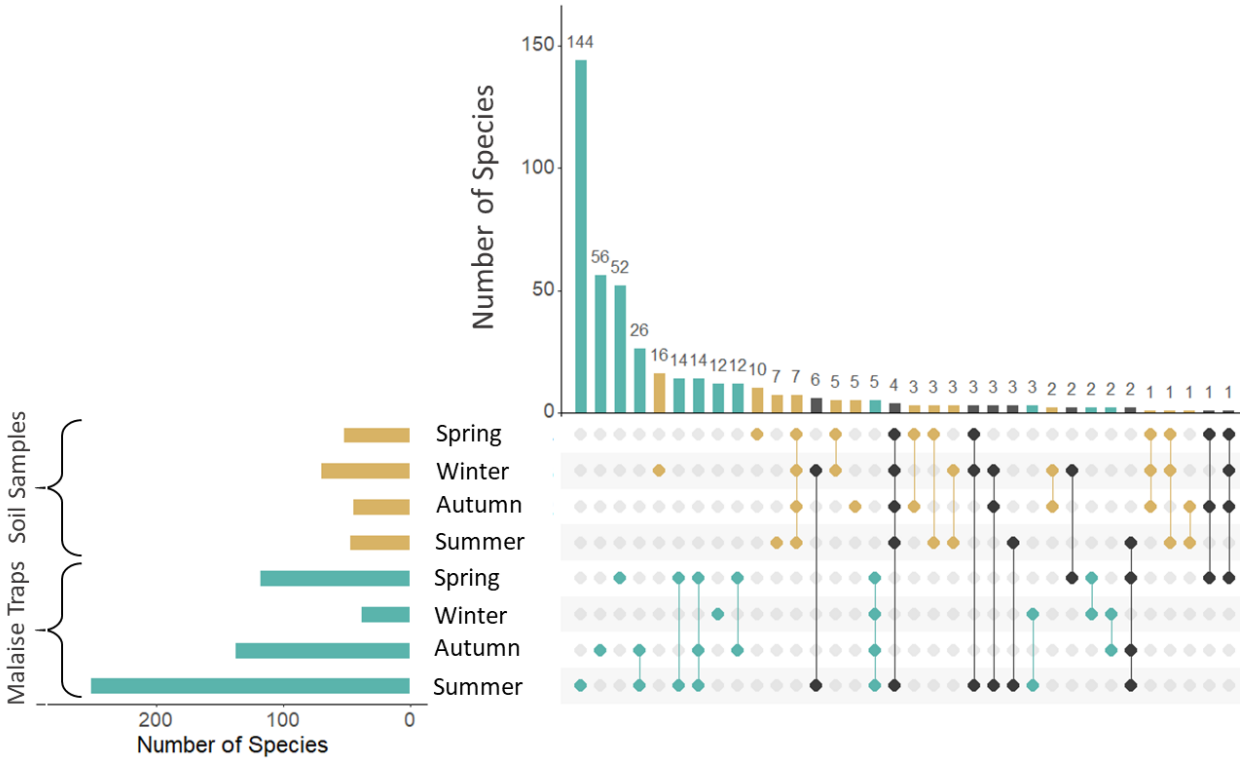
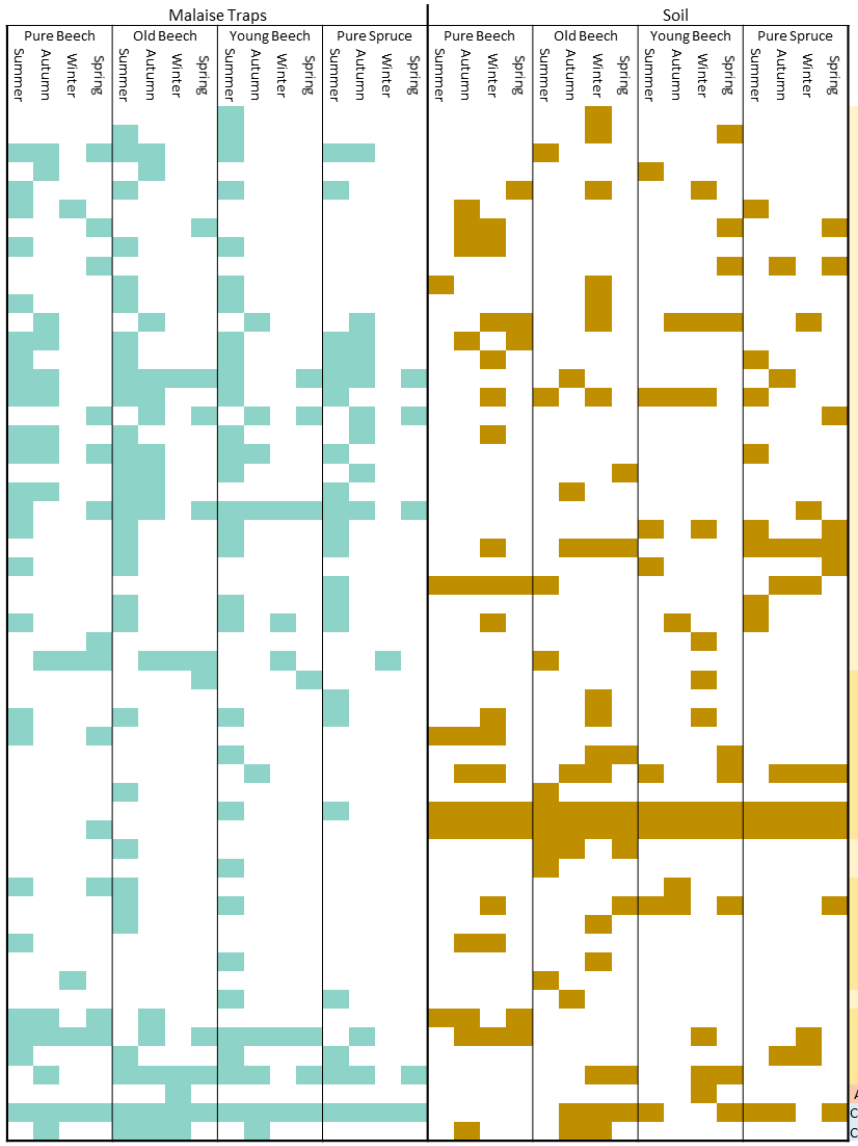


Figure IV.20 UpsetR plot showing number of unique and shared arthropod species between sample types depending on season.

Table IV.4 List of species found in soil samples and Malaise traps depending on forest type and sampling season



Class	Order	Family	Genus	Species
Insecta	Diptera	Asilidae	Neoitamus	Neoitamus cyanurus
Insecta	Diptera	Asilidae	Neoitamus	Neoitamus socius
Insecta	Diptera	Chironomidae	Bryophaenocladius	Bryophaenocladius ictericus
Insecta	Diptera	Chironomidae	Gymnometriocnemus	Gymnometriocnemus brumalis
Insecta	Diptera	Dolichopodidae	Neurigona	Neurigona quadrifasciata
Insecta	Diptera	Dolichopodidae	Sciapus	Sciapus platypterus
Insecta	Diptera	Empididae	Empis	Empis chioptera
Insecta	Diptera	Fanniidae	Fannia	Fannia polychaeta
Insecta	Diptera	Hybotidae	Bicellaria	Bicellaria nigra
Insecta	Diptera	Hybotidae	Platypalpus	Platypalpus luteicornis
Insecta	Diptera	Hybotidae	Platypalpus	Platypalpus nigritarsis
Insecta	Diptera	Hybotidae	Trichinomyia	Trichinomyia flavipes
Insecta	Diptera	Lauxaniidae	Lyciella	Lyciella platycephala
Insecta	Diptera	Muscidae	Phaonia	Phaonia pallida
Insecta	Diptera	Muscidae	Phaonia	Phaonia subventa
Insecta	Diptera	Muscidae	Polietes	Polietes lardarius
Insecta	Diptera	Mycetophilidae	Boletina	Boletina trivittata
Insecta	Diptera	Mycetophilidae	Synapha	Synapha vitripennis
Insecta	Diptera	Mycetophilidae	Tetragoneura	Tetragoneura sylvatica
Insecta	Diptera	Phoridae	Megaselia	Megaselia lata
Insecta	Diptera	Phoridae	Megaselia	Megaselia longicostalis
Insecta	Diptera	Psychodidae	Psychoda	Psychoda phalaenoides
Insecta	Diptera	Rhagionidae	Rhagio	Rhagio lineola
Insecta	Diptera	Sciaridae	Corynoptera	Corynoptera minima
Insecta	Diptera	Sciaridae	Corynoptera	Corynoptera trepida
Insecta	Diptera	Sciaridae	Ctenosciara	Ctenosciara lutea
Insecta	Diptera	Syrphidae	Sericomyia	Sericomyia silentis
Insecta	Diptera	Tabanidae	Tabanus	Tabanus bromius
Insecta	Diptera	Tipulidae	Tipula	Tipula submarmorata
Insecta	Diptera	Trichoceridae	Trichocera	Trichocera saltator
Insecta	Coleoptera	Anobiidae	Ernobius	Ernobius abietinus
Insecta	Coleoptera	Cantharidae	Malthodes	Malthodes fuscus
Insecta	Coleoptera	Cantharidae	Malthodes	Malthodes mysticus
Insecta	Coleoptera	Curculionidae	Phyllobius	Phyllobius argentatus
Insecta	Coleoptera	Curculionidae	Polydrusus	Polydrusus impar
Insecta	Coleoptera	Curculionidae	Strophosoma	Strophosoma melanogrammum
Insecta	Coleoptera	Elateridae	Ampedus	Ampedus balteatus
Insecta	Coleoptera	Elateridae	Athous	Athous subfuscus
Insecta	Coleoptera	Elateridae	Dalopius	Dalopius marginatus
Insecta	Hemiptera	Cicadellidae	Eupteryx	Eupteryx atropunctata
Insecta	Hemiptera	Miridae	Calocoris	Calocoris alpestris
Insecta	Lepidoptera	Gracillariidae	Phyllonorycter	Phyllonorycter maestingella
Insecta	Lepidoptera	Hepialidae	Phymatopus	Phymatopus hecta
Insecta	Lepidoptera	Noctuidae	Noctua	Noctua pronuba
Insecta	Lepidoptera	Tortricidae	Cydia	Cydia fagiglandana
Insecta	Lepidoptera	Tortricidae	Epinotia	Epinotia tedella
Insecta	Lepidoptera	Ypsolophidae	Ypsolopha	Ypsolopha ustella
Insecta	Mecoptera	Panorpidae	Panorpa	Panorpa germanica
Insecta	Psocoptera	Caeciliusidae	Valenzuela	Valenzuela flavidus
Insecta	Psocoptera	Ectopsocidae	Ectopsocus	Ectopsocus briggsi
Insecta	Psocoptera	Peripsocidae	Peripsocus	Peripsocus subfasciatus
Insecta	Psocoptera	Stenopsocidae	Graphopsocus	Graphopsocus cruciatus
Arachnida	Araneae	Linyphiidae	Tenuiphantes	Tenuiphantes zimmermanni
Collembola	Entomobryomorpha	Entomobryidae	Entomobrya	Entomobrya nivalis
Collembola	Entomobryomorpha	Entomobryidae	Orchesella	Orchesella flavescens

Lepidoptera

The lepidopteran species *Phymatopus hecta* (figure IV.16) was found in the soil in all four sampling seasons but was only present in the Malaise traps in summer season. Occurrence of *P. hecta* is associated with the herb layer. The larvae feed on the roots of the fern *Pteridium aquilinum* before they hibernate in the soil (Speidel 1994). The last larval instar starts feeding on young shoots at surface level, before they pupate in the soil. Unlike many other lepidopteran species, have females of *P. hecta* been observed to drops its eggs above possible food plants in flight (Turner 2015), which increases the likelihood that eggs will be found on the forest floor. The close association of *P. hecta* with the ground strata strongly increases the probability that the species is recovered from the soil. Other species such as *Epinotia tedell*a (figure IV.20) were only found in the soil samples during the winter season. In contrast to *P. hecta*, only one life stage of E. *tedella* is associated with the ground strata. *E. tedella* specializes its feeding on young spruce needles (Kosibowicz et al. 2014), forcing the species to spend most of its life in above the ground.

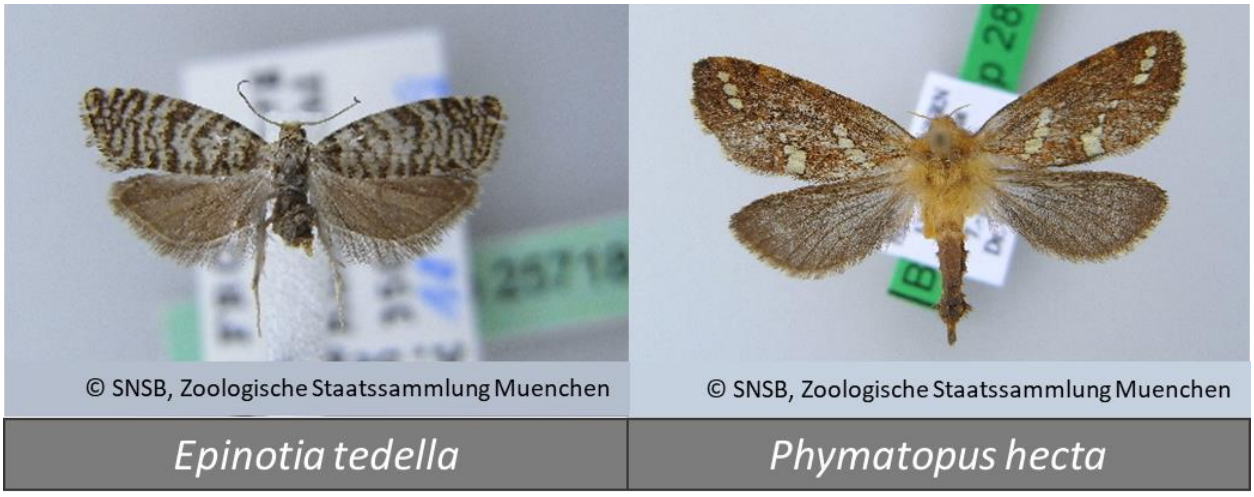


Figure IV.21 Photos of voucher specimens of the two lepidopteran species *Epinotia tedella* and *Phymatopus hecta*

Larvae of *E. tedella* drop to the forest floor for pupation where they hibernate. Field observations have even shown that some larvae remain in the tree canopy where they pupate and overwinter (Kosibowicz et al. 2014). Thus, some individuals of *E. tedella* do not interact with the ground strata at any life stage, but the probability of recovering *E. tedella* from soil samples is strongly increased in the winter season. *E. tedella* was caught in the Malaise traps in the summer season, which is the usual time of occurrence for the imagoes. The appearance of *E. tedella* in the two sample types at different times of the year mirrors the life cycle of the species. For all three above mentioned lepidopteran species, time of detection within the two strata matched developmental cycle described in the literature in the associated habitats, which underlines the power of metabarcoding to monitor species dynamics within habitats and seasons.

Coleoptera

Similar time shifts were also observed for species which are more challenging to target with metabarcoding. The Coleopteran species *Athous subfuscus* (figure IV.21) was present in the soil throughout the whole year, but was recovered from Malaise traps only in the summer season. *A. subfuscus* has a long developmental time of approximately six years, during which the larvae remain in the soil and undergo 12 different larval stages (Strey 1973). Imagoes emerge from the soil between April and July and are active until summer. The comparatively long developmental time of *A. subfuscus* from egg to imago is mirrored by the all year-round presence of *A. subfuscus* in soil samples, while the rather short lifespan of the imagoes limits the time during which the species can be detected with Malaise traps. Next to time constraints also methodological issues can hamper species detection rate with Malaise traps. The coleopteran species *Polydrusus impar* (figure IV.21) was recovered from soil samples of the winter and spring seasons, but was only found in Malaise traps in summer season.

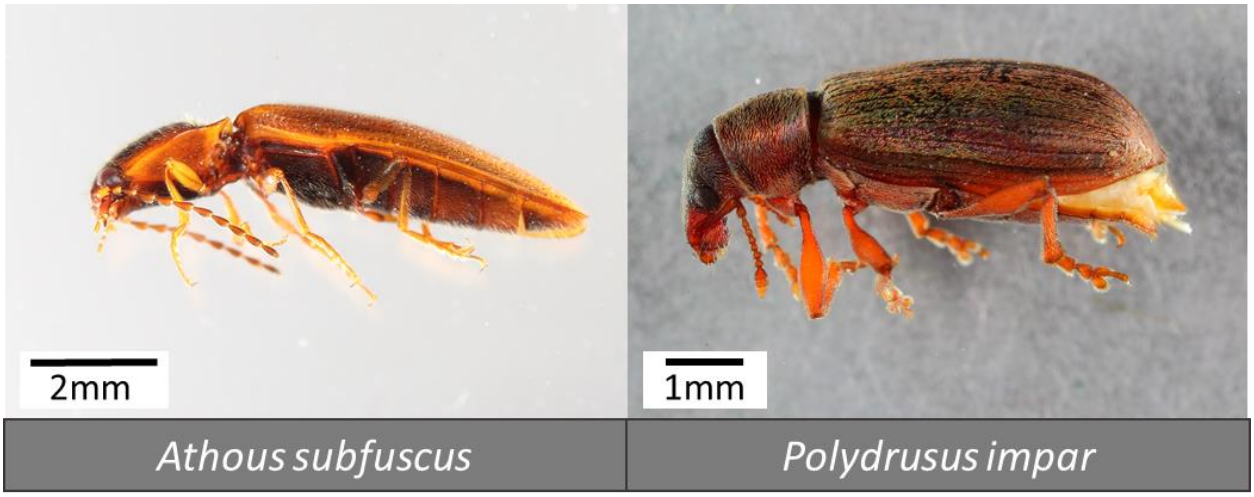


Figure IV.22 Photos of voucher specimens of the coleopteran species *Athous subfuscus* and *Polydrusus impar* (© BY-SA 4.0: GBOL / Museum Koenig)

While the larvae of *P. impar* feed on the roots of coniferous trees, the imagoes are usually found in the canopies during the summer season (Kula 2003). Species diversity associated with the canopies of forests is highly underexplored, not least because scientists aiming to investigate species diversity associated with the forest canopy are over towered by the study substrate. As a result, complex and costly methods like canopy fogging are required. As already described, several species like P. impar spend parts of their life associated with the ground strata, where their presence can now easily be confirmed with metabarcoding of soil samples. These examples highlight the power of metabarcoding to detect species diversity and turnover in a time series of samples collected in several substrates in the same location.

Diptera

For many dipteran species little is known about their biology, hampering direct comparison of our findings with existing descriptions of species biology. Overall, 30 species from 15 different

families were recovered from both sample types. The families widely differed in lifestyle, ranging from families mainly characterized by pollinator species (e.g. Syrphidae) to families with a high proportion of predatory species (e.g. Asilidae). 17 dipteran species were recovered from the soil samples in winter, which were not present in the Malaise traps at this time of the year. This allows for the assumption that like coleopterans and lepidopterans, dipterans are using the ground strata for hibernation. Nevertheless, it should be kept in mind that eDNA extracted from soil samples may also result from the decaying carcasses. Marquina et al. (2019) only found a deferred overlap between species occurrence in soil and Malaise trap samples for the insect order Diptera, which they assume was partly a result of the accumulation of dead specimens in the soil. The extended time lagged overlap was probably contributed by the fact that samples were taken late in the season when a majority of ground dwelling larvae had probably already developed into imagoes and had emerged from the ground habitat.

Influence of forest type on species recovery

Location of the Malaise traps (forest type) was not a factor influencing the recovery of a species next to Malaise traps although from the soil. Across all four forest types, highest jaccard-similarity indices calculated for the comparison of soil samples and Malaise traps were found for Malaise traps collected in summer season. A different picture was found when comparing arthropod species communities detected between soil samples taken within a specified forest type and Malaise traps. Soil samples taken at the pure beech sampling sites showed the highest jaccard-similarity index to Malaise traps when soil samples were taken in winter, while arthropod species communities of the soil samples taken at the old beech sites were more similar to Malaise traps when taken in autumn or spring respectively (figure IV.22). Finally, soil samples of the pure spruce sites showed the highest overlap with Malaise traps when soil sampling took place in summer. Overall, we observed an accumulation of species inhabiting the above ground strata during

summer at the ground layer of the pure beech sites in winter season. This may result from the different prevailing abiotic and biotic conditions at the four forest types favoring hibernation e.g. the accumulation of leaf litter.

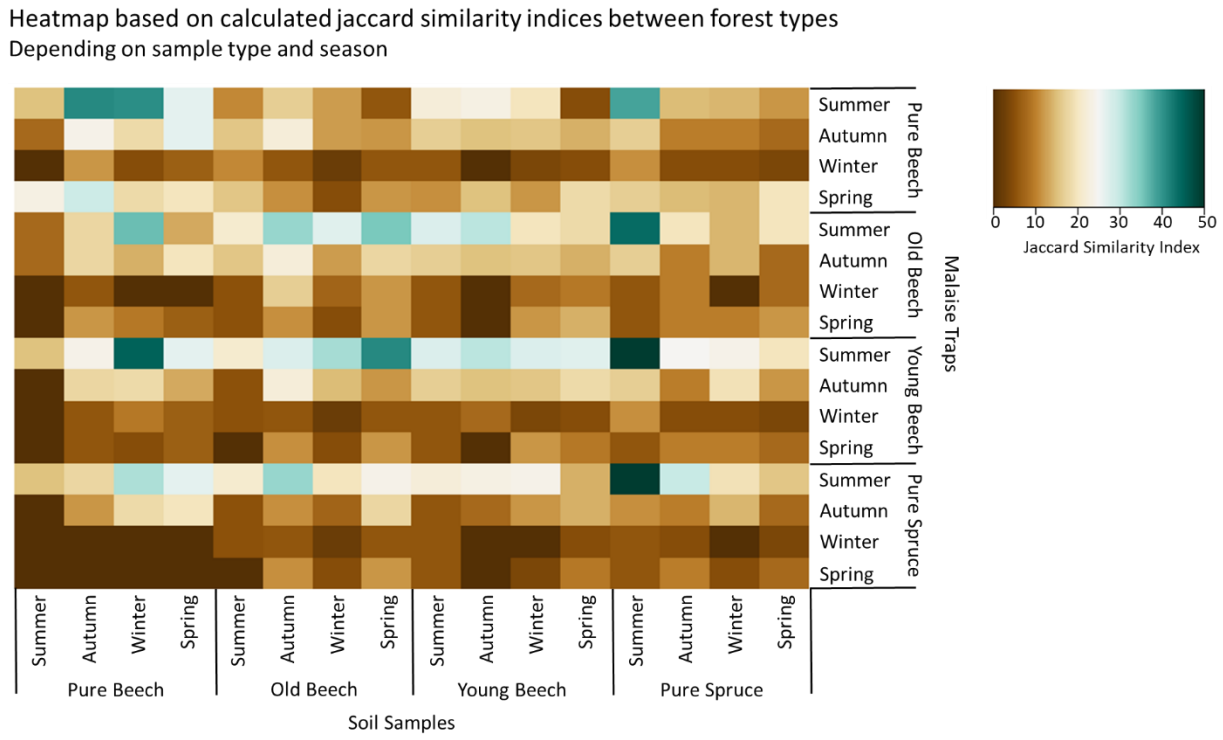


Figure IV.23 Heatmap showing jaccard similarity indices for the comparison of the assessed arthropod species community associated with each of the four forest types, depending on sample type and season

Conclusion

This study shows that choice of sampling method and substrate is critical for the outcome of DNA metabarcoding studies targeting Arthropoda. While the metabarcoding of bulk samples is becoming increasingly reliable (Elbrecht et al. 2019), assessing arthropod diversity on the basis of eDNA extracted from soil is still to mature. Nevertheless, this study has shown that despite the current limitations of soil eDNA metabarcoding, the method is already capable of supplementing results of species assessment studies based on Malaise trap catches. Our study confirms the findings of other studies which have shown that a limited sampling time window increases the risk of missing species, regardless of sampling method used (Marquina et al. 2019). All species described here were recovered from Malaise traps and soil samples respectively in concordance with species biology and lifecycle descriptions found in the literature. The complex life cycles and the time shifts between species occurrences in ground and above ground habitats required the combination of several sampling strategies to capture a picture as complete as possible of the occurring arthropod diversity. Our study underlines that depending on targeted taxa, the use of several sample types is highly recommended, when aiming to assess total species diversity. Furthermore, the use of several sample types allows for the scientific investigation of new research questions. It enables the monitoring of species occurrences over a longer period of time and can give deep insights into a species' ecology and life cycle. Furthermore, it also allows us to monitor seasonal species turnover rates between habitats e.g. when species prefer a certain habitat for hibernation. Based on this study, we recommend extending sampling with Malaise

traps over several seasons to lower the risk of missing species with a short flight period such as the hymenopteran family Formicidae and complementing it with metabarcoding of eDNA extracted from soil samples.

Literature

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

Aljanabi S.M., Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research. 25:4692–4693.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of molecular biology. 215:403–410.

Andersen K., Bird K.L., Rasmussen M., Haile J., Breuning-Madsen H., Kjaer K.H., Orlando L., Gilbert M.T.P., Willerslev E. 2012. Meta-barcoding of 'dirt'DNA from soil reflects vertebrate biodiversity. Molecular Ecology. 21:1966–1979.

Askew R., Shaw M. 1979. Mortality factors affecting the leaf-mining stages of Phyllonorycter (Lepidoptera: Gracillariidae) on oak and birch: 1. Analysis of the mortality factors. Zoological Journal of the Linnean Society. 67:31–49.

Barnes M.A., Turner C.R. 2016. The ecology of environmental DNA and implications for conservation genetics. Conservation genetics. 17:1–17.

Barnes M.A., Turner C.R., Jerde C.L., Renshaw M.A., Chadderton W.L., Lodge D.M. 2014. Environmental Conditions Influence eDNA Persistence in Aquatic Systems. Environ. Sci. Technol. 48:1819–1827.

Beng K.C., Tomlinson K.W., Shen X.H., Surget-Groba Y., Hughes A.C., Corlett R.T., Slik J.W.F. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. Scientific Reports. 6:24965.

Bienert F., De Danieli S., Miquel C., Coissac E., Poillot C., BRUN J., Taberlet P. 2012. Tracking earthworm communities from soil DNA. Molecular Ecology. 21:2017–2030.

Bourlat S.J., Haenel Q., Finnman J., Leray M. 2016. Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method. Marine Genomics. Springer. p. 197–207.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

Cai P., Huang Q.-Y., Zhang X.-W. 2006. Interactions of DNA with clay minerals and soil colloidal particles and protection against degradation by DNase. Environmental science & technology. 40:2971–2976.

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Peña A.G., Goodrich J.K., Gordon J.I., Huttley G.A., Kelley S.T., Knights D., Koenig J.E., Ley R.E., Lozupone C.A., McDonald D., Muegge B.D., Pirrung M., Reeder J., Sevinsky J.R., Turnbaugh P.J., Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 7:335–336.

Cardoso P., Erwin T.L., Borges P.A., New T.R. 2011. The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation. 144:2647–2655.

Churchill T.B., Arthur J.M. 1999. Measuring spider richness: effects of different sampling methods and spatial and temporal scales. Journal of Insect Conservation. 3:287–295.

Collins R.A., Wangensteen O.S., O'Gorman E.J., Mariani S., Sims D.W., Genner M.J. 2018. Persistence of environmental DNA in marine systems. Communications biology. 1:1–11.

Cope G.C., Campbell J.W., Grodsky S.M., Ellis J.D. 2019. Evaluation of nest-site selection of ground-nesting bees and wasps (Hymenoptera) using emergence traps. The Canadian Entomologist. 151:260–271.

Cowart D.A., Pinheiro M., Mouchel O., Maguer M., Grall J., Miné J., Arnaud-Haond S. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one. 10:e0117562.

Danks H. 2007. The elements of seasonal adaptations in insects. The Canadian Entomologist. 139:1–44.

Deiner K., Altermatt F. 2014. Transport distance of invertebrate environmental DNA in a natural river. PloS one. 9:e88786.

Deiner K., Bik H.M., Mächler E., Seymour M., Lacoursière-Roussel A., Altermatt F., Creer S., Bista I., Lodge D.M., De Vere N. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular ecology. 26:5872–5895.

Dejean T., Valentini A., Duparc A., Pellier-Cuit S., Pompanon F., Taberlet P., Miaud C. 2011. Persistence of environmental DNA in freshwater ecosystems. PloS one. 6: e23398..

deWaard J.R., Levesque-Beaudin V., deWaard S.L., Ivanova N.V., McKeown J.T., Miskie R., Naik S., Perez K.H., Ratnasingham S., Sobel C.N. 2019. Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. Genome. 62:85–95.

Dopheide A., Xie D., Buckley T.R., Drummond A.J., Newcomb R.D. 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution. 10:120–133.

EFSA Panel on Plant Health (PLH), Jeger M., Bragard C., Caffier D., Candresse T., Chatzivassiliou E., Dehnen-Schmutz K., Gilioli G., Jaques Miret J.A., MacLeod A. 2017. Pest categorisation of Cephalcia lariciphila. EFSA journal. 15:e05106.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Elbrecht V., Peinert B., Leese F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7:6918–6926.

Elbrecht V., Steinke D. 2019. Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology. 64:380–387.

Fjellberg A. 1992. Hypogastrura (Mucrella) arborea sp. nov., a tree-climbing species of Collembola (Hypogastruridae) from Vancouver Island, British Columbia. The Canadian Entomologist. 124:405–407.

Fonseca V., Sinniger F., Gaspar J., Quince C., Creer S., Power D.M., Peck L.S., Clark M.S. 2017. Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. Scientific reports. 7:6094.

Fonseca V.G., Lallias D. 2016. Metabarcoding marine sediments: preparation of amplicon libraries. Marine Genomics. Springer. p. 183–196.

Frøslev T.G., Kjøller R., Bruun H.H., Ejrnæs R., Brunbjerg A.K., Pietroni C., Hansen A.J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 8:1–11.

Garlapati D., Charankumar B., Ramu K., Madeswaran P., Murthy M.R. 2019. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Reviews in Environmental Science and Bio/Technology. 18:389–411.

Geiger M.F., Moriniere J., Hausmann A., Haszprunar G., Wägele W., Hebert P.D., Rulik B. 2016. Testing the Global Malaise Trap Program–How well does the current barcode reference library identify flying insects in Germany? Biodiversity data journal. 4: e10671.

German Barcode of Life Consortium, Wägele W., Haszprunar G., Eder J., Xylander W., Borsch T., Quandt D., Grobe P., Pietsch S., Geiger M.F., Astrin J.J., Rulik B., Hausmann A., Morinière J., Holstein J., Krogmann L., Monje C., Traunsprunger W., Hohberg K., Lehmitz R., Müller K., Nebel M., Hand R. 2011. GBOL Webportal.

Green J., Dreves A.J., McDonald B.W., Peachey R.E. 2016. Winter Cutworm: A New Pest Threat in Oregon. Oregon State University, Extension Service.

Guillera-Arroita G., Lahoz-Monfort J.J., van Rooyen A.R., Weeks A.R., Tingley R. 2017. Dealing with false-positive and false-negative errors about species occurrence at multiple levels. Methods in Ecology and Evolution. 8:1081–1091.

Gullan P.J., Cranston P.S. 2014. The insects: an outline of entomology. John Wiley & Sons.

Hajibabaei M., Shokralla S., Zhou X., Singer G.A., Baird D.J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one. 6:e17497.

Harvey P.R., Nellist D.R., Telfer M.G. 2002. Provisional atlas of British spiders (Arachnida, Araneae), Volume 1. Biological Records Centre, Centre for Ecology and Hydrology.

Hoorman J.J. 2011. The role of soil protozoa and nematodes. Fact Sheet: Agriculture and Natural Resources.(Smith KL), The Ohio State University Extension, Colombus, Ohio.

Horton D.J., Kershner M.W., Blackwood C.B. 2017. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. European journal of soil biology. 80:43–48.

Hosking G. 1979. Trap comparison in the capture of flying Coleoptera. New Zealand Entomologist. 7:87–92.

Hsieh Y., Linsenmair K.E. 2012. Seasonal dynamics of arboreal spider diversity in a temperate forest. Ecology and evolution. 2:768–777.

Jennings D.T., Hilburn D.J. 1988. Spiders (Araneae) captured in Malaise traps in spruce-fir forests of west-central Maine. Journal of Arachnology. 16:85–94.

Kamoroff C., Goldberg C.S. 2018. An issue of life or death: using eDNA to detect viable individuals in wilderness restoration. Freshwater Science. 37:685–696.

Kartzinel T.R., Chen P.A., Coverdale T.C., Erickson D.L., Kress W.J., Kuzmina M.L., Rubenstein D.I., Wang W., Pringle R.M. 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences. 112:8019–8024.

Kosibowicz M., Grodzki W., Jachym M. 2014. Local outbreak of the spruce needle tortricid Epinotia tedella Clerk (Lepidoptera, Tortricidae) in the Sudetes in Poland. Beskydy. 7:29–38.

Koziol A., Stat M., Simpson T., Jarman S., DiBattista J.D., Harvey E.S., Marnane M., McDonald J., Bunce M. 2019. Environmental DNA metabarcoding studies are critically affected by substrate selection. Molecular ecology resources. 19:366–376.

Krehenwinkel H., Wolf M., Lim J.Y., Rominger A.J., Simison W.B., Gillespie R.G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports. 7:17668.

Kreuels M., Buchholz S. 2006. Ökologie, Verbreitung und Gefährdungsstatus der Webspinnen Nordrhein-Westfalens: Erste überarbeitete Fassung der Roten Liste der Webspinnen (Arachnida: Araneae). Senden: Wolf & Kreuels.

Kula E. 2003. The population and seasonal dynamics of weevils developing in the soil of birch stands. Proceedings: Ecology, Survey and Management of Forest Insects Forest Service. USDA Forest Service. p. 33-38.

Leempoel K., Hebert T., Hadly E.A. 2019. A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. BioRxiv. 287:634022.

Leese F., Altermatt F., Bouchez A., Ekrem T., Hering D., Meissner K., Mergen P., Pawlowski J., Piggott J., Rimet F. 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes. 2:e11321.

Leray M., Yang J.Y., Meyer C.P., Mills S.C., Agudelo N., Ranwez V., Boehm J.T., Machida R.J. 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. 10:34.

Lopes C., De Barba M., Boyer F., Mercier C., da Silva Filho P., Heidtmann L., Galiano D., Kubiak B., Langone P., Garcias F. 2015. DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents. Heredity. 114:525.

Lopes C.M., Sasso T., Valentini A., Dejean T., Martins M., Zamudio K.R., Haddad C.F. 2017. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Molecular ecology resources. 17:904–914.

Ludy C., Lang A. 2004. How to catch foliage-dwelling spiders (Araneae) in maize fields and their margins: a comparison of two sampling methods. Journal of applied Entomology. 128:501–509.

Maraun M., Scheu S. 2000. The structure of oribatid mite communities (Acari, Oribatida): patterns, mechanisms and implications for future research. Ecography. 23:374–382.

Marquina D., Esparza-Salas R., Roslin T., Ronquist F. 2019. Establishing arthropod community composition using metabarcoding: surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. Molecular ecology resources. 19: 1516-1530

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10–12.

Matheson C.D., Gurney C., Esau N., Lehto R. 2010. Assessing PCR inhibition from humic substances. The Open Enzyme Inhibition Journal. 3:38-45.

Miller P.F. 1973. The biology of some Phyllonorycter species (Lepidoptera: Gracillariidae) mining leaves of oak and beech. Journal of Natural History. 7:391–409.

Morinière J., de Araujo B.C., Lam A.W., Hausmann A., Balke M., Schmidt S., Hendrich L., Doczkal D., Fartmann B., Arvidsson S. 2016. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PloS one. 11:e0155497.

Noordijk J., Morssinkhof R., Boer P., Schaffers A.P., Heijerman T., Sýkora K.V. 2008. How ants find each other; temporal and spatial patterns in nuptial flights. Insectes sociaux. 55:266–273.

Ogram A., Sayler G.S., Gustin D., Lewis R.J. 1988. DNA adsorption to soils and sediments. Environmental science & technology. 22:982–984.

Oliverio A.M., Gan H., Wickings K., Fierer N. 2018. A DNA metabarcoding approach to characterize soil arthropod communities. Soil Biology and Biochemistry. 125:37–43.

Oxbrough A., Gittings T., Kelly T.C., O'Halloran J. 2010. Can Malaise traps be used to sample spiders for biodiversity assessment? Journal of insect conservation. 14:169–179.

Palacios-Vargas J.G., Meneses G.C., Gómez-Anaya J. 1998. Collembola from the canopy of a Mexican tropical deciduous forest. Pan Pacific Entomologist. 74:47–54.

Pervez A. 2011. Ecology of aphidophagous ladybird Propylea species: A review. Journal of Asia-Pacific Entomology. 14:357–365.

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

Purcell R.V., Visnovska M., Biggs P.J., Schmeier S., Frizelle F.A. 2017. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Scientific reports. 7:11590.

R CoreTeam. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranjard L., Lejon D.P., Mougel C., Schehrer L., Merdinoglu D., Chaussod R. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environmental microbiology. 5:1111–1120.

Roberts M.J. 1996. Collins field guide. Spiders of Britain and northern Europe. HarperCollins Publishers.

Roesch L.F., Fulthorpe R.R., Riva A., Casella G., Hadwin A.K., Kent A.D., Daroub S.H., Camargo F.A., Farmerie W.G., Triplett E.W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME journal. 1:283-290.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584.

Rusek J. 1998. Biodiversity of Collembola and their functional role in the ecosystem. Biodiversity & Conservation. 7:1207–1219.

Salter I. 2018. Seasonal variability in the persistence of dissolved environmental DNA (eDNA) in a marine system: The role of microbial nutrient limitation. PloS one. 13: e0192409.

Samu F., Lengyel G., Szita É., Bidló A., Ódor P. 2014. The effect of forest stand characteristics on spider diversity and species composition in deciduous-coniferous mixed forests. The Journal of Arachnology. 42:135–141.

Sassoubre L.M., Yamahara K.M., Gardner L.D., Block B.A., Boehm A.B. 2016. Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. Environmental science & technology. 50:10456–10464.

Schaefer M., Schauermann J. 1990. The soil fauna of beech forests: comparison between a mull and a moder soil. Pedobiologia. 34:299–314.

Scharff N., Coddington J.A., Griswold C.E., Hormiga G., de Place Bjørn P. 2003. When to quit? Estimating spider species richness in a northern European deciduous forest. The journal of Arachnology. 31:246–273.

Sirois S.H., Buckley D.H. 2019. Factors governing extracellular DNA degradation dynamics in soil. Environmental microbiology reports. 11:173-184

Soszyńska-Maj A., Buszko J. 2011. Lepidoptera recorded on snow in Central Poland. Entomologica Fennica. 22:21–28.

Speidel W. 1994. Hepialidae. Die Schmetterlinge Baden-Württembergs Band 3 – Nachtfalter I. Stuttgart: Ulmer. p. 120–138.

Strey G. 1973. Ökoenergetische Untersuchungen an Athous subfuscus Müll. und Athous vittatus Fbr.(Elateridae, Coleoptera) in Buchenwäldern. Dissertation Göttingen.

Szujecki A. 1986. Ecology of forest insects. Springer Netherlands.

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

Taberlet P., Coissac E., Hajibabaei M., Rieseberg L.H. 2012a. Environmental DNA. Molecular ecology. 21:1789–1793.

Taberlet P., Prud'homme S.M., Campione E., Roy J., Miquel C., Shehzad W., Gielly L., Rioux D., Choler P., Clèment J. 2012b. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular ecology. 21:1816–1820.

Thomsen P.F., Sigsgaard E.E. 2019. Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. Ecology and Evolution. 9:1665–1679.

Turner J.R. 2015. The flexible lek: Phymatopus hecta the gold swift demonstrates the evolution of leking and male swarming via a hotspot (Lepidoptera: Hepialidae). Biological journal of the Linnean Society. 114:184–201.

Ulrich W. 1999a. The number of species of Hymenoptera in Europe and assessment of the total number of Hymenoptera in the world. Polskie Pismo Entomologiczne. 68:151–164.

Ulrich W. 1999b. Temporal stability of community structure of the parasitic Hymenoptera in a beech forest on limestone. Polish Journal of Ecology. 47:257–270.

Valentini A., Taberlet P., Miaud C., Civade R., Herder J., Thomsen P.F., Bellemain E., Besnard A., Coissac E., Boyer F., Gaboriaud C., Jean P., Poulet N., Roset N., Copp G.H., Geniez P., Pont D., Argillier C., Baudoin J.-M., Peroux T., Crivelli A.J., Olivier A., Acqueberge M., Le Brun M., Møller P.R., Willerslev E., Dejean T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology. 25:929–942.

Wall D.H., Fitter A.H., Paul E.A. 2005. Developing new perspectives from advances in soil biodiversity research. Biological Diversity and Function in Soils. Cambridge University Press, Cambridge. p. 31–43.

Wang H., Qi J., Xiao D., Wang Z., Tian K. 2017. A re-evaluation of dilution for eliminating PCR inhibition in soil DNA samples. Soil Biology and Biochemistry. 106:109–118.

Wardle D.A. 2002. Communities and ecosystems: linking the aboveground and belowground components. Princeton University Press.

Wardle D.A. 2006. The influence of biotic interactions on soil biodiversity. Ecology letters. 9:870–886.

Wardle D.A., Bardgett R.D., Klironomos J.N., Setälä H., Van Der Putten W.H., Wall D.H. 2004. Ecological linkages between aboveground and belowground biota. Science. 304:1629–1633.

Wheeler Q.D., Raven P.H., Wilson E.O. 2004. Taxonomy: Impediment or Expedient? Science. 303:285–285.

Yang C., Ji Y., Wang X., Yang C., Douglas W.Y. 2013. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. Science China Life Sciences. 56:73–81.

Yang C., Wang X., Miller J.A., de Blécourt M., Ji Y., Yang C., Harrison R.D., Douglas W.Y. 2014. Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators. 46:379–389.

Yoccoz N.G., Bråthen K.A., Gielly L., Haile J., Edwards M.E., Goslar T., von Stedingk H., Brysting A., Coissac E., Pompanon F. 2012. DNA from soil mirrors plant taxonomic and growth form diversity. Molecular Ecology. 21:3647–3655.

Yoshida T., Hijii N. 2006. Seasonal distribution of Xenylla brevispina (Collembola) in the canopy and soil habitat of a Cryptomeria japonica plantation. Pedobiologia. 50:235–242.

Zinger L., Chave J., Coissac E., Iribar A., Louisanna E., Manzi S., Schilling V., Schimann H., Sommeria-Klein G., Taberlet P. 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biology and Biochemistry. 96:16–19.

Zizka V.M., Leese F., Peinert B., Geiger M.F. 2018. DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. Genome. 62:122–136.

Discussion

Over the last years, metabarcoding has become a reliable and cost efficient tool, enabling the assessment of biodiversity in a timely manner. Despite enormous steps towards the implementation of metabarcoding (e.g. food security, wildlife protection, assessment of water quality) some aspects of metabarcoding still require more research to transition DNA-based biodiversity assessment to an accepted alternative to traditional morphology-based approaches.

Current Limitations

By definition metabarcoding describes 'the taxonomic identification of multiple species extracted from a mixed sample (community DNA or environmental DNA) which have been PCR-amplified and sequenced on a high-throughput platform' (Deiner et al. 2017). Several authors have stated that the dependency of metabarcoding on PCR is one of the major weaknesses of these methods since it can lead to primer taxonomic preferential amplification and diversity bias (Taberlet et al. 2012; Clarke et al. 2014; Cristescu 2014). Metabarcoding studies usually aim at amplifying DNA molecules derived from many species in a single PCR run. Therefore, it is of uttermost importance to use highly universal primers to ensure that sequences originating from different taxa are amplified with similar efficiencies and no species gets missed in diversity assessments (Taberlet et al. 2012). Consequently, several markers have been tested for their taxonomic resolution and coverage, but none is capable of providing a high-resolution picture of total existing biodiversity. However, several barcode regions were identified to have a high taxonomic resolution and coverage for certain taxonomic groups of interest. e.g. Cytochrome *c* oxidase subunit I gene (COI) for arthropods (Andújar et al. 2018; Elbrecht et al. 2019), the small subunit 18S rRNA gene (18S) for Metazoa (Fonseca et al. 2010), the internal transcribed spacer region (ITS) for fungi (Schoch et al. 2012; Blaalid et al. 2013), the maturase K gene (matK) and the large

subunit of ribulose 1,5-bisphosphate carboxylase gene (rbcL) for plants (CBOL Plant Working Group et al. 2009) and the small subunit 16S rRNA gen (16S) for bacteria (Chakravorty et al. 2007). Although all of them have proved capable of identifying a large number of taxa from groups of interest, an accumulation of false negative results is still frequently observed (CBOL Plant Working Group et al. 2009; Yu et al. 2012). This is mainly due to primer bias, inter alia introduced due to species-specific differences in primer affinities. For instance, the COI marker is a protein coding gene, therefore the third position of the codons is often highly variable as it is less constrained by selection (Piñol et al. 2015; Clarke et al. 2017). As a result, primer binding sites are not strictly conserved and primer affinities can significantly vary between species supporting the amplification of well-matched taxa, while species with low primer affinities often remain undetected. The degree of incorporated primer bias is highly unpredictable as it is determined by several factors (Morinière et al. 2016; Elbrecht et al. 2017). Together with the complexity of the DNA mixture (Morinière et al. 2016) and low sequencing depth, also differences in specimen biomass contained in bulk sample (Elbrecht et al. 2017) is known to increase PCR bias. In order to avoid or reduce primer bias, several strategies have been found and increasingly applied, e.g. by presorting samples (Morinière et al. 2016; Elbrecht et al. 2017), lowering the primer annealing temperature, or by using highly degenerate primers (Clarke et al. 2014; Elbrecht and Leese 2017). Nevertheless, an increase in degree of primer degeneracy as well as suboptimal annealing temperatures are known to introduce new potential biases like the amplification of non-target DNA (Siddall et al. 2009; Horton et al. 2017). While in bulk samples the lion's share of DNA accounts for trapped metazoan specimens, in eDNA mixtures extracted from soil samples this proportion is shifted towards microbial DNA resulting in a high amplification rate of non-target DNA (Yang et al. 2013; Macher et al. 2018). As a result, due to differences in primer affinities and the unwanted amplification of non-target DNA, several target species are missed as only a limited number of sequences of the target groups are amplified.

A recent study has shown that pursuing a *one locus several primers* strategy, meaning the use of several primer pairs targeting the same locus within the same group of taxa (Corse et al. 2019), can be key to efficiently describe species diversity (Esnaola et al. 2018). Although this approach seems to be a promising alternative to highly degraded primers, it doesn't come without drawbacks. The spectrum of target species can only be extended as far as it matches the taxonomic coverage and resolution of the chosen marker for the targeted group (Creer et al. 2010; Derycke et al. 2010; Kvist 2014; Andújar et al. 2018). Furthermore, even highly conserved primers do not entirely exclude the amplification of non-target DNA (Mioduchowska et al. 2018) and the use of a single marker is often problematic as intra- and interspecific variability can greatly vary within organism groups (Brown et al. 2015; Chain et al. 2016). Although the COI marker is often referred to as marker of choice for the identification of metazoans (Andújar et al. 2018), it fails to robustly identify annelids (Kvist 2014) and nematodes (De Ley et al. 2005; Bhadury et al. 2006; Derycke et al. 2010). Additional observations were made for matK for the identification of several plant taxa (CBOL Plant Working Group et al. 2009; Fahner et al. 2016). Furthermore, the use of a single marker only allows for the exploitation of a limited proportion of taxonomic information contained in the DNA mixture. To increase the proportion of valuable taxonomic information and thereby number of recovered taxa the combination of several markers is increasingly recommended (Cowart et al. 2015). By now several studies have shown that the combination of several markers can significantly increase taxonomic resolution and coverage (Zhang et al. 2018; Marquina et al. 2019). Nevertheless, when targeting several markers, library preparation and sequencing costs increase rapidly, together with duration of laboratory work and time of analysis. In order to reduce costs and effort, it is possible to multiplex several markers within the same PCR (De Barba et al. 2014). However, this method received criticism as well. Primer melting temperatures may differ considerably in sub-optimally compiled primer pairs, leading to strong variations in amplification success (Taberlet et al. 2018). Although this bias can be bypassed by adjusting melting temperatures and primer concentrations (Taberlet

et al. 2018) this method is highly susceptible to introduce a high number of false negative results as sub-optimal PCR conditions are limiting amplification success.

Outlook

Development of PCR-free Approaches

Such PCR-based approaches limitations contributed to a necessarily narrow primer and marker choice (Taberlet et al. 2018), attempts have been undertaken to develop PCR-free approaches like 'shotgun-based metabarcoding'. Shotgun sequencing, also often referred to as 'metagenomics', 'environmental genomics', 'ecogenomics' and 'community genomics' refers to a DNA-sequencing based method, aiming to profile mixed communities (Porter and Hajibabaei 2018) by sequencing all genomic material from many diverse taxa simultaneously (Venter et al. 2004). The genomic DNA is thereby fragmented to the size appropriate for the chosen sequencing platform (Piper et al. 2019). The resulting sequences represent a random subsample of the DNA extract and contain a complex set of taxonomical information of mitochondrial and nuclear origin (Tang et al. 2015; Crampton-Platt et al. 2016). Shotgun-based metabarcoding can be combined with a target enrichment step (Dowle et al. 2016; Shokralla et al. 2016). Although even without target enrichment, promising results can be obtained (Srivathsan et al. 2015, 2016), a target enrichment step is recommended as its application allows for the enrichment of a set of marker prior to sequencing (Porter and Hajibabaei 2018) leading to a strong reduction of sequencing cost and saving in sequencing time (Mamanova et al. 2010). For example, when targeting pathogens or parasites, it is possible to enrich samples with sequences from these organisms, while reducing the reads for taxa with highly abundant DNA, e.g. the host. Shotgun-based metabarcoding without enrichment by capture increases the cost at least 1000fold (Taberlet et al. 2018) and

furthermore leads to a high number of taxonomically non-informative reads (Crampton-Platt et al. 2016). As a result, the identification of some taxa is, depending on sequencing depth, frequently based on a very low number of informative reads (Pedersen et al. 2016), strongly increasing the risk of incorporating false negative results.

Today, several methods are pooled under the designation ‘target enrichment’, including commonly used PCR (Mamanova et al. 2010). These can be also found under the name of hybrid capture and molecular inversion probes (MIPs) as the most widely PCR-free used approaches (Mamanova et al. 2010; Porter and Hajibabaei 2018).

Molecular Inversion Probes (MIPs)

Target enrichment on the basis of MIPs uses single-stranded oligonucleotides consisting of universal sequences (common backbone), flanked by universal primers which are bordered by target-specific sequences (Stefan et al. 2016). After probes anneal on either site of the target region of polymerase fills the resulting sequence gap (Mamanova et al. 2010). Afterwards, the resulting loop is closed by ligation (Mamanova et al. 2010; Stefan et al. 2016). For PCR-amplification on the basis of universal primers directed at the common backbone (Mamanova et al. 2010) only circularized probes are kept, while exogenous host DNA and uncircularized probes are digested by exonucleases (Mamanova et al. 2010; Stefan et al. 2016; Porter and Hajibabaei 2018). Although this method has been proven to have a high specifity of MIP captures (Mamanova et al. 2010), the number of taxa which can be simultaneously targeted is rather low compared to other target enrichment approaches (Porter and Hajibabaei 2018). This is mainly due to the fact that the MIP approach is limited in number of target regions which can be captured under uniform conditions (Mamanova et al. 2010). In contrast to this method, stands target enrichment by hybrid capture, which is more suitable for biodiversity assessment studies as a complex set of taxa can be targeted.

Hybrid Capture

Similar to molecular inversion probes, hybrid enrichment uses large oligonucleotides for the capture of target DNA. A great advantage of this method is the possibility to encompass an extremely high number of oligonucleotides (Piper et al. 2019) designed complementarily to target specific target sequences (Jones and Good 2016), allowing to capture several loci (Piper et al. 2019) at once. For hybrid capture, oligonucleotides are bound to beads in solutions. After hybridization of target sequences to oligonucleotides, several washing steps are applied in order to remove non-target sequences from the solution. Hybrid capture favors the capture of short fragments with high specificity (Mamanova et al. 2010) as longer sequences tend to show a higher number of cross-hybridizations (Mamanova et al. 2010) and thereby an increasing proportion of off-target sequences (Gnirke et al. 2009). This preference towards the capture of short fragments is promoting the capture of degraded DNA, which is increasingly found in eDNA samples (Thomsen and Willerslev 2015). As a result, it can be anticipated that hybrid capture can significantly improve eDNA metabarcoding when being used for the replacement of the initial mixed template PCR step (Shokralla et al. 2016). Due to the high complexity of resulting DNA mixtures and the extreme dominance of microbial DNA, the amplification of rarer species of e.g. arthropods is strongly hampered (Yang et al. 2013; Shokralla et al. 2016; Horton et al. 2017). Even when aiming to amplify the Folmer region of the COI marker, which has been shown to be most suitable for the detection of arthropods, it has been observed that the vast majority of resulting amplified sequences account for bacterial DNA (Yang et al. 2013) hampering the identification of arthropods. Former studies have shown that hybrid capture is highly efficient for the capture of arthropod and insect DNA, leading to a higher detection rate of these taxa compared to traditional metabarcoding approaches which uses an initial mixed-template PCR step (Shokralla et al. 2016). This is making it to a promising tool for the development of monitoring approaches using eDNA. Furthermore, the absence of a PCR- step can likely lead to a more meaningful interpretation of read numbers, e.g. to try to infer species abundance estimates.

Metabarcoding as a Quantitative Tool

The development of molecular biodiversity assessment approaches are capable of retrieving relative abundance data but to obtain absolute abundances has not been possible and has already been referred to as the "holy grail" of metabarcoding (Clarke et al. 2017). Some studies have suggested that, depending on species, a modest but positive relationship between species abundance and sequencing read abundance exists (Evans et al. 2016). This relationship has been used to make semi-quantitative but valid statements about population growth trends for single species (Tillotson et al. 2018). Nevertheless, these studies are only approximations as it is well known that several methodological considerations can significantly influence number of reads (Elbrecht and Leese 2015; Piñol et al. 2015; Krehenwinkel et al. 2017). Especially the current dependency of metabarcoding on PCR is limiting its use for quantitative biodiversity survey studies (Elbrecht and Leese 2015; Piñol et al. 2015). Differences in primer affinities are leading to the introduction of amplification biases which degree is determined by complexity and bulk sample composition, making it highly unpredictable (Hajibabaei et al. 2011; Brandon-Mong et al. 2015; Elbrecht and Leese 2015; Morinière et al. 2016). However, although PCR bias can to some extent be overcome by replacing the initial mixed-template PCR step, e.g. by hybrid capture, it remains challenging to retrieve absolute abundance data (Krehenwinkel et al. 2017). The character of the target marker must be taken into account as depending on marker the number of gene copies varies strongly across cells, tissues, species and also individuals (Taberlet et al. 2018), making it indispensable to have a priori knowledge on the above outlined characteristics of the target marker for each species of interest. Databases containing the required data would allow for the development of bioinformatic algorithms capable of calculating a close estimation of species abundances. For the establishment of a database containing species specific data of marker characteristics for a broad range of taxa (e.g. all arthropods), an immense research effort is required. However due to the enormous gain of information, this effort would be reasonable

as the implementation of a correction factor would open the door for new field of application e.g. the development of theoretical models based on abundance data.

Development of Reference Databases

Up to now it was not possible to identify a single marker which is capable to identify all taxa across the tree of life. As this would be desirable due to cost and time constraints, new markers are emerging on a regular basis. Several of those markers have been shown to deliver comprehensive surveys, but are limited for the use of species identification due to incomplete databases (Elbrecht et al. 2016). The completeness of the associated reference databases is significantly influencing detection rate and furthermore determines how many reads are assigned to each taxon (Shokralla et al. 2016). Shokralla et al. (2016) showed that more sequences are assigned to orders which tend to be overrepresented in databases compared to less abundant orders. This highlights the importance of database complexity especially as metabarcoding is striving to provide quantitative results by using read number assigned to each taxon as a proxy for species abundance. However, it has been shown that the combination of two or more markers can increase species detection rate by up to 10%, a proportion which is likely to further increase with completeness of databases. Unfortunately, barcode reference databases are strongly differing in completeness between markers. While a lot of effort has been undertaken to establish well curated and comprehensive COI databases for arthropods, ribosomal marker reference databases are still very limited for most taxonomic groups (Clarke et al. 2014). This is inter alia due to the standard barcoding system (Coissac et al. 2016), which has long time relied on single species PCR-based approaches and Sanger sequencing for the establishment of references databases. The increasing number of multiplexed studies, circumventing the problematic recovery of groups for which universal primers have been shown to be poorly working (e.g., COI for hymenoptera and matK for various plant lineages), requires

an extension of the current standard barcoding systems (Coissac et al. 2016). In order to provide a database as comprehensive as possible, genome-skimming is now increasingly applied (see here: PhyloAlps project: phyloalps.org). Genome skimming is the low coverage shotgun sequencing of genomic DNA (Straub et al. 2012) resulting in comparatively deep sequencing of the multi-copy fraction of the genome (plastome, mitogenome, and repetitive elements) (Dodsworth 2015). As a result, all markers are obtained simultaneously and can be stored as potential barcodes in a single reference database. This allows for a strong increase in phylogenetic signal which increases taxonomic resolution and coverage (Ruhsam et al. 2015; Coissac et al. 2016). Furthermore, in contrast to PCR approaches, genome skimming is capable of retrieving the barcodes of museum specimens, including type specimens which are often suffering from a high degree of DNA degradation (Besnard et al. 2014; Coissac et al. 2016). Despite the fact that the costs for genome skimming are still exceeding the ones of traditional PCR-based approaches (Taberlet et al. 2018), the already outlined advantages of genome skimming justify its use for the establishment of reference databases in the future as with growing complexity of databases taxonomic resolution and coverage of multiplexed studies will be further increased (Zhang et al. 2018; Marquina et al. 2019).

Exact Sequence Variants

For a long time, metabarcoding was based on the concept of operational taxonomic units (OTUs) which used clusters of sequences at a given cutoff as proxy for number of present taxa. In the literature a cutoff *p*-distance of 97% - 99% is usually found to approximate species. However, the degree of sequence dissimilarity within species (intraspecific variability) (Chain et al. 2016) and between species (interspecific variability) (Brown et al. 2015) can widely vary. While intraspecific variation can exceed a threshold of 3% (Brown et al. 2015), resulting in the oversplitting of taxa into several OTUs, interspecific variation can be lower fusing different species into one OTU and

distorting biodiversity measurements (Chain et al. 2016). Recently, several studies were published emphasizing that new concepts might be more suitable and sensitive for the measurement of biodiversity (Callahan et al. 2016, 2017; Needham et al. 2017). So-called exact sequence variants (ESVs) (Callahan et al. 2017), also often referred to as amplicon sequence variants (ASVs) (Callahan et al. 2016) or zero-radius OTUs (zOTUs) (Edgar 2016), are roughly said clusters of sequences which differ by as little as one nucleotide from each other (Callahan et al. 2017) and thus represent true biological units (Porter and Hajibabaei 2018). The advantage of OTU over ASV approaches are the higher degree of robustness to an artificial increase of assessed biodiversity due to the incorporation of sequencing artefacts. However, with the development of new bioinformatic sequence denoising approaches this issue has been addressed. For example DADA2 generates a parametric error model, which takes the entire sequencing run into consideration and uses the data to correct and collapse sequencing errors. Several studies have now compared the use of ESVs and OTUs and found that both method detect similar community patterns (Glassman and Martiny 2018; Porter et al. 2019), although with exact sequence variants a slightly higher species diversity was found (Porter et al. 2019). However, the concept of exact sequence variants has an enormous advantage over the OTU concept as it is working with real biological units existing outside of the analyzed data (Callahan et al. 2017; Marshall and Stepien 2019). While the comparison of two datasets based on OTUs requires a complete new analysis, which, depending on number of sequences per dataset, can be very time-consuming and often requires a high amount of computational power, studies based on ASVs can easily be compared with each other due to consistent labeling as each ASV (Callahan et al. 2017) accounts for a precise haplotype (Marshall and Stepien 2019). Furthermore, the development of a large reference database based on ASVs would allow for a direct comparison of detected ASVs across studies allowing for the detection of invasive species which have so far not been described. However, the comparability of studies is highly dependent on study design and bioinformatics pipeline choice.

Standardization

One of the major concerns of metabarcoding is its current lack of standardization (Taberlet et al. 2018). Standardized metabarcoding protocols ensure that experiments are fully replicable in time, space and across laboratories allowing for the comparison of several independently conducted studies, thus enabling the documentation and investigation of biological phenomena on a global scale, e.g., the development of endemism over a height gradient or the influence of forest type on arthropod community composition (boreal vs. deciduous forests). Previous studies have shown that standardization facilitates the acceptance of new methods inside and outside the scientific community (Malorny et al. 2003). It is therefore likely that standardization will also lead to an increased acceptance of metabarcoding approaches for non-scientific purposes (e.g., food security, wildlife protection, assessment of water quality by assessing the saprobic index), but also inside the scientific world e.g. for calculation of theoretical models. Usually, species and corresponding abundance data are used for the development of theoretical models (Pereira et al. 2010; Kaschner et al. 2011), whereas metabarcoding has long time not been in the focus of the scientific-model community. However, today, metabarcoding retrieves valid biodiversity data, which provide a reliable basis for the development of theoretical semi-mechanistic community level models (Mokany and Ferrier 2011). While biodiversity assessment data based on traditional methods are rare and mostly limited to a small set of species, several thousand metabarcoding datasets are now publicly available containing information about current and past biodiversity patterns from across the world. If these datasets had been obtained under standardized conditions, an immense and highly valuable unified dataset would now be available for the development of biodiversity models which could better monitor, among other aspects, global shifts in biodiversity to better estimate correlation with climate change, land use, etc.

However, standardization also has a downside as it encompasses the risk of blocking further developments and thereby negatively influencing study results. While for metabarcoding of bulk samples several studies have already highlighted advantage and disadvantages of several

methodological considerations (e.g. marker choice, primer choice and bioinformatics tools) (Fonseca 2018), only a limited number of eDNA metabarcoding studies using soil samples to assess macroorganisms have so far been conducted whereas several methodological questions are still left unanswered. The development of a guideline or even best practice advises for eDNA soil studies is desirable but at this point not recommended as it could hamper future developments and lead to an accumulation of studies with improvable results. Furthermore, depending on the aim of the study the experimental design must be accordingly adjusted. Former studies have already shown that environmental parameters (e.g. soil composition) can significantly influence the effectiveness of eDNA extraction method (Sagova-Mareckova et al. 2008). For example DNA yield an purity extracted from soil samples can significantly vary between different types of soil depending on extraction methods. The assessed degree of biodiversity of soil samples are therefore only partially comparable with each other. This case also shows that prior to study, tests must be taken to ensure best possible results. Instead of formulating narrowly conceptualized guidelines and best practices, a first step towards standardization could be the use of a standardized eDNA template as positive control in each study (Taberlet et al. 2018). This would allow for subsequently conducted meta-analyses, and thereby for a direct control of effectiveness of study design. While increasingly more studies are now dealing with soil eDNA (Pietramellara et al. 2009; Sirois and Buckley 2019), little is known about taxon-dependent, differential release of eDNA into these environments. Therefore, it is of topmost importance to perform systematic laboratory experiments to establish a basis for the development of protocols and best practice instructions leading towards standardization and thereby to a promising use of metabarcoding for commercial purposes, namely biomonitoring and conservation studies.

Conclusion Remarks

Overall molecular methods combined with next generation sequencing platforms can provide reliable information about existing biodiversity. As these approaches keep developing at a tremendous pace, they are likely to contribute significantly to the identification and investigation of dimensions, drivers, and consequences of biodiversity loss, enabling the development of urgently needed biodiversity protection strategies in the near future.

Literature

Andújar C., Arribas P., Yu D.W., Vogler A.P., Emerson B.C. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology. 27:3968–3975.

Besnard G., Christin P.-A., Malé P.-J.G., Lhuillier E., Lauzeral C., Coissac E., Vorontsova M.S. 2014. From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. Journal of Experimental Botany. 65:6711–6721.

Bhadury P., Austen M.C., Bilton D.T., Lambshead P.J.D., Rogers A.D., Smerdon G.R. 2006. Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. Marine Ecology Progress Series. 320:1–9.

Blaalid R., Kumar S., Nilsson R.H., Abarenkov K., Kirk P., Kauserud H. 2013. ITS 1 versus ITS 2 as DNA metabarcodes for fungi. Molecular Ecology Resources. 13:218–224.

Brandon-Mong G.-J., Gan H.-M., Sing K.-W., Lee P.-S., Lim P.-E., Wilson J.-J. 2015. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. Bulletin of Entomological Research. 105:717–727.

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

Callahan B.J., McMurdie P.J., Holmes S.P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal. 11:2639–2643.

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A., Holmes S.P. 2016. DADA2: High resolution sample inference from Illumina amplicon data. Nature Methods. 13:581–583.

CBOL Plant Working Group, Hollingsworth P.M., Forrest L.L., Spouge J.L., Hajibabaei M., Ratnasingham S., van der Bank M., Chase M.W., Cowan R.S., Erickson D.L. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences. 106:12794–12797.

Chain F.J., Brown E.A., MacIsaac H.J., Cristescu M.E. 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Diversity and Distributions. 22:493–504.

Chakravorty S., Helb D., Burday M., Connell N., Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods. 69:330–339.

Clarke L.J., Beard J.M., Swadling K.M., Deagle B.E. 2017. Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution. 7:873–883.

Clarke L.J., Soubrier J., Weyrich L.S., Cooper A. 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources. 14:1160–1170.

Coissac E., Hollingsworth P.M., Lavergne S., Taberlet P. 2016. From barcodes to genomes: extending the concept of DNA barcoding. Molecular Ecology. 25:1423–1428.

Corse E., Tougard C., Archambaud-Suard G., Agnèse J., Messu Mandeng F.D., Bilong Bilong C.F., Duneau D., Zinger L., Chappaz R., Xu C.C. 2019. One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. Ecology and Evolution. 9:4603–4620.

Cowart D.A., Pinheiro M., Mouchel O., Maguer M., Grall J., Miné J., Arnaud-Haond S. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PloS one. 10:e0117562.

Crampton-Platt A., Yu D.W., Zhou X., Vogler A.P. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. GigaScience. 5:s13742-016.

Creer S., Fonseca V.G., Porazinska D.L., Giblin-Davis R.M., Sung W., Power D.M., Packer M., Carvalho G.R., Blaxter M.L., Lambshead P.J.D., Thomas W.K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology. 19:4–20.

Cristescu M.E. 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. Trends in Ecology & Evolution. 29:566–571.

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

De Ley P., De Ley I.T., Morris K., Abebe E., Mundo-Ocampo M., Yoder M., Heras J., Waumann D., Rocha-Olivares A., Jay Burr A. 2005. An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences. 360:1945–1958.

Deiner K., Bik H.M., Mächler E., Seymour M., Lacoursière-Roussel A., Altermatt F., Creer S., Bista I., Lodge D.M., De Vere N. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology. 26:5872–5895.

Derycke S., Vanaverbeke J., Rigaux A., Backeljau T., Moens T. 2010. Exploring the use of cytochrome oxidase c subunit 1 (COI) for DNA barcoding of free-living marine nematodes. PLoS One. 5.

Dodsworth S. 2015. Genome skimming for next-generation biodiversity analysis. Trends in Plant Science. 20:525–527.

Dowle E.J., Pochon X., C. Banks J., Shearer K., Wood S.A. 2016. Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. Molecular Ecology Resources. 16:1240–1254.

Edgar R.C. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv.:081257.

Elbrecht V., Braukmann T.W., Ivanova N.V., Prosser S.W., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D., Steinke D. 2019. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ Preprints. 7:e27801v1.

Elbrecht V., Leese F. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PloS One. 10:e0130324.

Elbrecht V., Leese F. 2017. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science. 5:11.

Elbrecht V., Peinert B., Leese F. 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution. 7:6918–6926.

Elbrecht V., Taberlet P., Dejean T., Valentini A., Usseglio-Polatera P., Beisel J.-N., Coissac E., Boyer F., Leese F. 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ. 4:e1966.

Esnaola A., Arrizabalaga-Escudero A., Gonzalez-Esteban J., Elosegi A., Aihartza J. 2018. Determining diet from faeces: Selection of metabarcoding primers for the insectivore Pyrenean desman (Galemys pyrenaicus). PloS One. 13.

Evans N.T., Olds B.P., Renshaw M.A., Turner C.R., Li Y., Jerde C.L., Mahon A.R., Pfrender M.E., Lamberti G.A., Lodge D.M. 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources. 16:29–41.

Fahner N.A., Shokralla S., Baird D.J., Hajibabaei M. 2016. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: recovery, resolution, and annotation of four DNA markers. PloS One. 11.

Fonseca V.G. 2018. Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources. 18:923–926.

Fonseca V.G., Carvalho G.R., Sung W., Johnson H.F., Power D.M., Neill S.P., Packer M., Blaxter M.L., Lambshead P.J.D., Thomas W.K., Creer S. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nature Communications. 1.

Glassman S.I., Martiny J.B. 2018. Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. MSphere. 3.

Gnirke A., Melnikov A., Maguire J., Rogov P., LeProust E.M., Brockman W., Fennell T., Giannoukos G., Fisher S., Russ C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology. 27:182.

Hajibabaei M., Shokralla S., Zhou X., Singer G.A., Baird D.J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS One. 6:e17497.

Horton D.J., Kershner M.W., Blackwood C.B. 2017. Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. European Journal of Soil Biology. 80:43–48.

Jones M.R., Good J.M. 2016. Targeted capture in evolutionary and ecological genomics. Molecular Ecology. 25:185–202.

Kaschner K., Tittensor D.P., Ready J., Gerrodette T., Worm B. 2011. Current and future patterns of global marine mammal biodiversity. PLoS One. 6.

Krehenwinkel H., Wolf M., Lim J.Y., Rominger A.J., Simison W.B., Gillespie R.G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports. 7:17668.

Kvist S. 2014. Does a global DNA barcoding gap exist in Annelida? Mitochondrial DNA Part A. 27:2241–2252.

Macher J., Vivancos A., Piggott J.J., Centeno F.C., Matthaei C.D., Leese F. 2018. Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. Molecular Ecology Resources. 18:1456–1468.

Malorny B., Tassios P.T., Rådström P., Cook N., Wagner M., Hoorfar J. 2003. Standardization of diagnostic PCR for the detection of foodborne pathogens. International Journal of Food Microbiology. 83:39–48.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., Turner D.J. 2010. Target-enrichment strategies for next-generation sequencing. Nature methods. 7:111.

Marquina D., Andersson A.F., Ronquist F. 2019. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources. 19:90–104.

Marshall N.T., Stepien C.A. 2019. Invasion genetics from eDNA and thousands of larvae: A targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. Ecology and evolution. 9:3515–3538.

Mioduchowska M., Czyż M.J., Gołdyn B., Kur J., Sell J. 2018. Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too “universal”? PloS One. 13.

Mokany K., Ferrier S. 2011. Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. Diversity and Distributions. 17:374–380.

Morinière J., de Araujo B.C., Lam A.W., Hausmann A., Balke M., Schmidt S., Hendrich L., Doczkal D., Fartmann B., Arvidsson S. 2016. Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. PloS One. 11:e0155497.

Needham D.M., Sachdeva R., Fuhrman J.A. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. The ISME Journal. 11:1614–1629.

Pedersen M.W., Ruter A., Schweger C., Friebe H., Staff R.A., Kjeldsen K.K., Mendoza M.L., Beaudoin A.B., Zutter C., Larsen N.K. 2016. Postglacial viability and colonization in North America's ice-free corridor. Nature. 537:45–49.

Pereira H.M., Leadley P.W., Proença V., Alkemade R., Scharlemann J.P., Fernandez-Manjarrés J.F., Araújo M.B., Balvanera P., Biggs R., Cheung W.W. 2010. Scenarios for global biodiversity in the 21st century. Science. 330:1496–1501.

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

Piñol J., Mir G., Gomez-Polo P., Agustí N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources. 15:819–830.

Piper A.M., Batovska J., Cogan N.O., Weiss J., Cunningham J.P., Rodoni B.C., Blacket M.J. 2019. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience. 8:giz092.

Porter T.M., Hajibabaei M. 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology. 27:313–338.

Porter T.M., Morris D.M., Basiliko N., Hajibabaei M., Doucet D., Bowman S., Emilson E.J., Emilson C.E., Chartrand D., Wainio-Keizer K. 2019. Variations in terrestrial arthropod DnA metabarcoding methods recovers robust beta diversity but variable richness and site indicators. Scientific Reports. 9:1–11.

Ruhsam M., Rai H.S., Mathews S., Ross T.G., Graham S.W., Raubeson L.A., Mei W., Thomas P.I., Gardner M.F., Ennos R.A. 2015. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in Araucaria? Molecular Ecology Resources. 15:1067–1078.

Sagova-Mareckova M., Cermak L., Novotna J., Plhackova K., Forstova J., Kopecky J. 2008. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Applied and Environmental Microbiology. 74:2902–2907.

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

Shokralla S., Gibson J., King I., Baird D., Janzen D., Hallwachs W., Hajibabaei M. 2016. Environmental DNA barcode sequence capture: targeted, PCR-free sequence capture for biodiversity analysis from bulk environmental samples. BioRxiv.:087437.

Siddall M.E., Fontanella F.M., Watson S.C., Kvist S., Erséus C. 2009. Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. Systematic Biology. 58:445–451.

Sirois S.H., Buckley D.H. 2019. Factors governing extracellular DNA degradation dynamics in soil. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12725.

Srivathsan A., Ang A., Vogler A.P., Meier R. 2016. Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. Frontiers in Zoology. 13:17.

Srivathsan A., Sha J.C., Vogler A.P., Meier R. 2015. Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (P ygathrix nemaeus). Molecular Ecology Resources. 15:250–261.

Stefan C.P., Koehler J.W., Minogue T.D. 2016. Targeted next-generation sequencing for the detection of ciprofloxacin resistance markers using molecular inversion probes. Scientific Reports. 6:1–12.

Straub S.C., Parks M., Weitemier K., Fishbein M., Cronn R.C., Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany. 99:349–364.

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

Taberlet P., Coissac E., Pompanon F., Brochmann C., Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology. 21:2045–2050.

Tang M., Hardman C.J., Ji Y., Meng G., Liu S., Tan M., Yang S., Moss E.D., Wang J., Yang C. 2015. High-throughput monitoring of wild bee diversity and abundance via mitogenomics. Methods in Ecology and Evolution. 6:1034–1043.

Thomsen P.F., Willerslev E. 2015. Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation. 183:4–18.

Tillotson M.D., Kelly R.P., Duda J.J., Hoy M., Kralj J., Quinn T.P. 2018. Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. Biological Conservation. 220:1–11.

Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A., Wu D., Paulsen I., Nelson K.E., Nelson W. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 304:66–74.

Yang C., Ji Y., Wang X., Yang C., Douglas W.Y. 2013. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. Science China Life Sciences. 56:73–81.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution. 3:613–623.

Zhang G.K., Chain F.J., Abbott C.L., Cristescu M.E. 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evolutionary Applications. 11:1901–1914.

Abbreviations

18S	small subunit 18S rRNA gene
16S	small subunit 16S rRNA gen
ASV	Amplicon Sequence Variant
C	Carbon
CBOL	Consortium for the Barcode of Life
cDNA	community DNA
COI	Cytochrome *c* oxidase subunit I gene
DNA	Deoxyribonucleic acid
eDNA	environmental DNA
EL	Elevation Level
ESV	Exact Sequence Variant
GBOL	German Barcode of Life
ITS	internal transcribed spacer region
lysDNA	Lyse DNA
matK	maturase K gene
N	Nitrogen

NGS	Next Generation Sequencing
OB	Old Beech
OTU	Operational Taxonomic Unit
PB	Pure Beech
PCR	Polymerase Chain Reaction
PS	Pure Spruce
rbcL	1,5-bisphosphate carboxylase gene
RNA	Ribonucleic acid
SAR	species-area relationship
SOP	Standard Operating Procedures
TC	Total carbon
TIC	Total inorganic carbon
TOC	Total organic carbon
YB	Young Beech

List of Tables

Chapter I

Chapter II

Chapter III

Chapter IV

List of Figures

Introduction

Chapter I

Chapter II

Chapter III

Chapter IV

Acknowledgments

Die Anfertigung meiner Dissertation wäre ohne die berufliche aber auch privater Unterstützung einer Vielzahl von Personen nicht möglich gewesen.

First of all, I would like to thank Dr. Vera G. Fonseca who gave me the opportunity to prepare this thesis in her research group. Without her input and her profound knowledge of metabarcoding, the preparation of this thesis wouldn't have been possible. She was always available when support and advice were needed and at all times willing to discuss emerging issues.

Gleich zu Beginn meiner Doktoranden Zeit hat sich Bernhard Misof bereit erklärt die Rolle des Erstgutachters zu übernehmen, wofür Ich ihm an dieser Stelle danken möchte. Weiterhin möchte Ich auch Birgit Gemeinholzer, Dietmar Quandt und Thomas Döring danken, dass sie sich die Zeit nehmen diese doch recht komplexe and interdisziplinäre Arbeit zu begutachten.

Besonderer Dank gilt weiterhin Kathrin Langen für ihre Einführung und Hilfestellungen im Labor, welche nicht besser hätte sein können. Auch Sarah Bourlat, Jonas Astrin und Björn Rulik gebührt besonderer Dank für Ihre tatkräftige Unterstützung, sei es im Feld oder während der Anfertigung dieser und anderer Arbeiten sowie für ihre stetige Bereitschaft mir in diversen Fragestellungen zur Seite zu stehen. Insgesamt möchte Ich auch dem Rest des Laborteams sowie aller aktuellen und auch ehemaligen Mitglieder der Sektion Metabarcoding dafür danken, dass sie immer bereit

waren bei Problemen zu helfen aber auch im Besonderen für das sehr gute Arbeitsklima. Auch möchte Ich hiermit allen Personen Danke sagen, welche mich während der Feldarbeit tatkräftig unterstützt haben und denen kein Schnee zu tief war.

Finanzielle Unterstützung erhielt ich von dem Bundesministerium für Bildung und Forschung sowie dem Forschungsmuseum Alexander Koenig, Bonn. Diesem möchte Ich im Besonderen für die Bereitstellung der Forschungsinfrastruktur, sowie auch für meinen Arbeitsplatz danken.

Neben dieser hervorragenden beruflichen Unterstützung, konnte Ich mich glücklich schätzen auch enormen Beistand auf privater Ebene erfahren zu dürfen. In erster Linie möchte Ich mich bei meiner Familie bedanken, welche mich bedingungslos unterstützt hat und zu jeder Zeit an mich geglaubt hat. Auch Tanja Salwiczek und Tillmann Eitelberg gebührt enormer Dank für Ihre Unterstützung, sei es auf privater aber auch auf fachlicher Ebene. Insbesondere möchte Ich mich dafür bedanken, dass sie mir stets den Rücken freigehalten haben und mir kleine Auszeiten ermöglichten. Gerade in der Endphase der Doktorandenzeit erhielt ich starken emotionalen Support von Juliane Glomm wofür Ich Ihr von Herzen danken möchte. Insgesamt schätze Ich mich überaus glücklich, dass Ich mich während dieser Zerreißprobe für das soziale Leben auf eine Vielzahl von Freunden verlassen konnte für deren Geduld und positive Bestärkung ich nicht dankbarer sein könnte. Ich danke euch.

Electronic Supplement

Chapter I

Please find Library Preparation strategy on the CD or at https://doi.org/10.22000/394 in the supplement file

Lib_preparation_Georgia.txt

Chapter II

Please find Library Preparation strategy on the CD or at https://doi.org/10.22000/394 in the supplement file

Lib_preparation_MalaiseTraps.txt

Chapter III

Please find Library Preparation strategy on the CD or at https://doi.org/10.22000/394 in the supplement files

Lib_preparation_PhosphatBuffer.txt

Lib_preparation_MachereyNagel.txt

Chapter IV

Please find Library Preparation strategy on the CD or at https://doi.org/10.22000/394 in the supplement files

Lib_preparation_MalaiseTraps.txt

Lib_preparation_PhosphatBuffer.txt

Lib_preparation_MachereyNagel.txt