Development of workflows for metabarcoding of mass-samples

A case study on Diptera

Dissertation

Vorgelegt von:
Jan-Frederic Struwe
aus Meschede

Bonn, August 2018

Zur Erlangung des Doktorgrades (Dr. rer. Nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Die Dissertation wurde am Zoologischen Forschungsmuseum Alexander Koenig (ZFMK) in Bonn durchgeführt.

Erstgutachter:	Prof. Dr. Johann Wolfgang Wägele
Zweitgutachter:	Prof. Dr. Thomas Bartolomaeus
Kommissionsmitglied (fachnah):	Prof. Dr. Bernhard Misof
Kommissionsmitglied (fachfremd):	apl. Prof. Dr. Ullrich Wüllner

Tag der Promotion: 25.06.2019

Erscheinungsjahr: 2020

Publication: Searching for the Optimal Sampling Solution; PLOS ONE

Gossner MM, Struwe J-F, Sturm S, Max S, McCutcheon M, Weisser WW, Zytynska SE (2016) Searching for the Optimal Sampling Solution: Variation in Invertebrate Communities, Sample Condition and DNA Quality. PLoS ONE 11(2): e0148247. https://doi.org/10.1371/journal.pone.0148247

„Krautsalat?“

- Felice Kremer -

Content

Publication: Searching for the Optimal Sampling Solution; PLOS ONE 3
1 Introduction 1
1.1 Background 1
1.2 Diptera and their relevance in ecosystems 2
1.2.1 The diversity and influence of Diptera 2
1.3.2 The challenges 4
1.2.2 The benefits of monitoring Diptera 7
1.3 DNA barcoding for biodiversity assessments 8
1.3.1 The concept 8
1.4 Hybridisation based target enrichment 9
1.4.1 The exclusive selection of a genomic region 9
1.4.2 Designing specific oligonucleotide probes 10
1.5 The necessity to build up a sequence reference data pool 11
1.6 The advantage of automated sampling 11
1.7 The Biodiversity Exploratories 12
1.8 Outlining the project goals 13
2 A Diptera sequence reference list and COI discriminability 15
2.1 Introduction 15
2.2 Material and Methods 15
2.2.1 Material acquisition and processing 15
2.2.2 Testing universality and discriminative properties of COI 17
2.3 Results 20
2.3.1 Material acquisition and library growth 20
2.3.2 The universality and discriminative properties of COI 21
2.4 Discussion 25
2.4.1 Building up a sequence reference list for German Diptera 25
2.4.2 Universal application of COI 25
2.4.3 The capability to discriminate species 26
2.4.4 Conclusion 30
3. Test of conservation liquids for traps 31
3.1 Introduction 31
3.2 Material and Methods 31
3.2.1 Experimental setup 31
3.2.2 Arthropod identification and classification 32

3.2.3 Measure of quality for morphological species determination 32
3.2.4 Species identification through DNA barcoding 32
3.3 Results 35
3.3.1 Condition of the samples 36
3.3.2 Order level 37
3.3.3 Species level analyses 38
3.4 Discussion 41
3.4.1 The usage of different conservation liquids for traps 41
3.4.2 Conclusion and recommendations 44
4 Mixed species DNA-samples 46
4.1 Introduction 46
4.2 Material and Methods 46
4.2.1 Material acquisition and sample preparation 46
4.2.2 Next Generation Sequencing 49
4.2.3 Data processing 50
4.2.4 The influence of sample treatment, species number and species quantity 51
4.3 Results 51
4.3.1 Differences in the chimera identification 52
4.3.2 Identification of species 54
4.3.3 Analyzing DNA input ratios 55
4.4 Discussion 56
4.4.1 Chimeras and sequencing errors 57
4.4.2. Factors inducing the forming of chimeras 58
4.4.3 DNA ratios and species abundances 59
4.4.4 Recommendations 59
5 Empirical biodiversity assessment 61
5.1 Introduction 61
5.2 Material and Methods 61
5.2.1 The automated interval sampler (AIS) 61
5.2.2 Study area 64
5.2.3 Preparations for a genetic biodiversity assessment 69
5.2.4 Data Processing 71
5.3 Results 75
5.3.1 MOTU analysis - alpha diversity 76
5.3.2 MOTU analysis - beta diversity 77
5.3.3 Population level analysis 80

5.3.4 Comparing the morphological and genetical determinations 83
5.4 Discussion 89
5.4.1 The automated interval sampler (AIS) 89
5.4.2 How to efficiently sequence bulk samples 91
5.4.3 MOTU based assessment 94
5.4.4 Morphological versus sequence based species determinations 98
5.4.5 Conclusion 100
6 General conclusion and future prospects 102
7 Bibliography 105
8 Supplement 134
S1 134
S2 136
S3 141
9 APPENDIX 163
Zusammenfassung 164
Erklärung 166

1 Introduction

1.1 Background

The loss of biodiversity the world is facing today has already reached the status of a global biodiversity crisis; and the decline does not appear to be slowing down (Singh 2002; Clausnitzer et al. 2009; Butchart et al. 2010; Brodeur and Candioti 2017; Hallmann et al. 2017). This disastrous development is reflected not only by a decline in local species abundancies but also in a massive decline of their biomass (Dirzo et al. 2014; Hallmann et al. 2017). Habitat loss, climate change, invasive species, land use and overexploitation are drivers of these changes (Lowe et al. 2000; Bradshaw 2009; Knapp 2017). Great dismay and consternation dominate scientists, politicians, and the general public as cascading effects on ecosystems, economy and humanity must be expected (Morse 1971; McIntyre et al. 2006; Ollerton et al. 2011; Cardinale et al. 2012; Hallmann et al. 2017). Recent research efforts and intergovernmental organizational formations regarding biodiversity conservation assessments and policies demonstrate the general demand for a better understanding of the relationships between ecological drivers and the affected species communities (Newbold et al. 2015; Schulp et al. 2016; Teixeira et al. 2016; Hevia et al. 2017; Mori et al. 2017; Intergovernmental Science–Policy Platform on Biodiversity and Ecosystem Services, IPBES, www.ipbes.net).

The inventory and differentiation of species is the fundamental basis of many of these ecological studies. Yet datasets often seem to be spatially and temporally insufficient and new efforts in monitoring and contributing technologies are being suggested (Gonzales et al. 2016; Hillebrand et al. 2017). The required spatial distribution and the demand for regular updates of datasets strongly suggest the automation of these processes including the process of species identification.

Considering that batches of insect samples currently are even being weighted to enable their evaluation (Hallmann et al. 2017) implies that the amount of data that is already being produced by current traps consumes an almost unreasonable effort of time and money whenever species differentiation is based on traditional morphological expertise. Additionally a morphological approach often needs several different taxonomic experts for every taxonomic order found in a sample. It can be expected that the required material and conversational exchange only leads to further delayment. In contrast to this a genetically based assessment of species offers the chance to make knowledge usually restricted to taxonomists widely applicable (Hebert et al. 2003a, 2003b; Geiger et al. 2016). Another great advantage of the genetic approach is that while the phenotype might change (e.g. egg vs larvae vs adult, male vs female) the genotype stays the same. This will not only accelerate and facilitate the identification process but also support taxonomy by helping to identify synonyms and reveal cryptic species (Smith et al. 2006; Whitman and Agrawal 2009; Scheffers et al. 2012; Janzen et al. 2017). Genetic approaches like DNA barcoding therefore help to maintain and extend our taxonomic knowledge.

DNA barcoding can even deal with yet unknown species. Publications already demonstrated how sequence data can be handled by defining them as operational taxonomic units or molecular taxonomic units (OTUs/MOTUs) which then function as placeholders until species information is available (Floyd et al. 2002; Schloss and

Handelsmann 2005; Blaxter 2004; Quince et al 2009). The development of a Barcode Indexing Number (BIN) system in the Barcode of Life Database (BOLD) supports this process, adapting to what some people might call a post-Linnaean taxonomic system (Ratnasingham et al. 2013; Blaxter 2016). These MOTUs, OTUs or BINs approved to be capable entities that can successfully be used in ecological surveys (Gibson et al. 2015; Blaxter 2016). And as most of the global biodiversity still remains unknown, these entities allow assessing diversity and documenting changes in community structures (Mora et al. 2011).

1.2 Diptera and their relevance in ecosystems

1.2.1 The diversity and influence of Diptera

Within the diverse arthropod fauna insects are the most speciose class. This is impressive considering that until now only 7-10% of insect species are known to science. The ecosystem impact of the known fraction of insects already has to be termed significant. Considering all existing species their functional significance must be enormous (Samways 1993). Even a single taxon can already cover a wide range of ecological functions. In the order of Diptera there are pollinators, predators, parasites and decomposer; and finally they also act as a food resource themselves (Morse 1971; Ziegler 2003; Borkent and Harder 2007; Ollerton et al. 2011). This ecologically diverse group has successfully colonized all continents and almost every habitat except the open sea (Teskey 1976; Ferrar 1987; Hovemeyer 2000; Courtney and Merritt 2008; Courtney et al. 2009). Estimates range from 120,000 to 150,000 species that are spread over the whole world (Colless and McAlpine 1991; Schumann 1992; Schuhmann et al. 1999, 2002; Brown 2001; Merritt et al. 2003). Current ratings put them on 3rd position, with 152,244 species, after Lepidoptera, with 156,793 species and Coleoptera, with 359,891 species (Evenhuis et al. 2007; Adler and Foottit 2009). Considering this richness it seems not surprising that the group of Diptera belongs to the major insect groups in the Palearctic. Less apparent might be that at the same time their species are highly endemic to it (Konstantinov et al. 2009). At higher latitudes Diptera even start to dominate insect communities, leaving species counts for beetles far behind (Khruleva 1987; Danks and Smith 2009).

From Germany over 9,000 species have been reported in more than 1900 genera (Schuhmann 1992; Schumann et al. 1999, 2002). These species interact with their environment at nearly all trophic levels (Ziegler 2003). Their ecological demands and ability to colonize different habitats while taking significant roles designate them valuable bio-indicators (Kühner 1992; Frouz 1999). Several ecological functions can be covered by a single species as the demands of the specimens often change through different life stages and can differ also according to their sex. Species of mosquitos (Culicidae; Figure 1.1) are valuable decomposer during their larval stage as filter feeders (Walker et al. 1988). They also can pollinate plants by nectar-feeding on them as adults (Thien 1969; Jhumur 2006; Borkent and Harder 2007; Peach and Gries 2016). And the female consumption of blood makes them commonly known as ectoparasites. Mosquitos even change from aquatic to terrestrial habitats during their life cycle. The significance and involved mechanisms that are influenced by certain single species in an ecosystem can hardly be described comprehensively. And although various species have been characterized for particular ecological conditions or habitats (Strenzke 1950;

Rozkošný 1986; Kühner 1992) still a lot more needs to be learned (Frouz 1999; Kenis et al. 2016). This holds true particularly for species of no economic interest. Besides that many investigations rather focus on groups that are considered more ‘‘attractive’’ for the public and the researcher, e.g. butterflies and ladybirds (Thurber et al. 1994; Kenis et al. 2009; Kenis et al. 2016; Rondoni et al. 2017) whereas flies and mosquitos are not very much appreciated by the public. But despite their common reputation and public ignorance Diptera also take important functions that would be highly appreciated if people would be aware of them. While usually bees are highly valued for their pollination service, Diptera often contribute significantly to plant reproductive success. For some plants they even are the main pollinators (Eberling and Olersen 1999; Larson et al. 2001; Borkent and Harder 2007). It might seem unfair that bees get credit for pollination even when the observed specimen was actually a fly. This deception is often caused by specimen of the family of Syrphidae (Figure 1.1). This group contains several species that successfully mimic the appearance of bees and wasps and occasionally even behave like them when visiting flowers (Golding et al. 2000). However rather inconspicuous is the appearance of small midges (Ceratopogonidae; Figure 1.1) of whom some are the exclusive pollinators of the highly specialized flowers of cacao (Young 1986, 1994). Although their service is probably very much appreciated by most of us, people still do not particularly know them and the way they benefit from them (Mursu et al 2004; Messerli 2012).

Impelling for a broader perception is always when species are perceived as a threat to human health or economic interests. For Diptera this affects for example species of the genus Aedes (Culicidae; Figure 1.1). *A. japonicus* and *A. albopictus* are considered invasive species and known vectors of severe infections. This includes West Nile fever, yellow fever, dengue fever and St Louis encephalitis just to mention a few (Turel et al. 2001; Kutz et al. 2003; Caminade et al. 2012; Huber K et al. 2012). Observations of these two species indicate that they have established stable populations in Germany. While *A. japonicus* was already documented in 2008 (Schaffner et al. 2009), *A. albopictus* has recently been documented for its successfully overwintering from 2015 to 2016 in southwest Germany (Pluskota et al. 2016). Its habitat expansion and colonization of Europe has been observed with concern since it was detected in the 1970s in Albania (Becker et al. 2017). A certain public interest for the presence or absence of these species is relatable but scientifically the will to gain further ecological insights should be essential. Species invasion almost always affects native animal populations and it can be assumed that local species will be affected by these invaders as they have been affected before on other occasions (Juliano and Lounibos 2005). Studies have shown that species replacements can have dramatic cascading effects on the environment. But again the research focus lies not on Diptera. The best documented records of these effects and the ecological damage on the native fauna that followed from it can be found for invasive ants (Hill M et al. 2003; O’Dowd et al. 2003; Kenis et al. 2009). Information on Diptera is often less comprehensive and rather concentrates on them as biocontrol agents, agricultural pests, or again when they interact with “attractive” or “exotic” species (McPheron and Steck 1996; Munro et al. 2002; Kellogg et al. 2003; Fessl et al. 2006; Koop et al. 2016; Knutie et al. 2017; Oberhauser et al. 2017). Nevertheless these studies do give an insight into the species interactions and into the extensive ecosystem effects that Diptera species can have. Considered beneficial are for example snail-killing flies (Sciomyzidae; Figure 1.1) that are used as biological control

agents for populations of the intermediate hosts of trematodes causing bilharzia (Berg and Knutson 1978, Maharaj et al. 1992). Parasitic Tachinidae (Figure 1.1) are used against pestiferous Lepidoptera (DeBach and Rosen 1991; Munro et al. 2002). From agricultural interest are fruit flies (Tephritidae; Figure 1.1) which are the cause of considerable economic damage in fruits and vegetables (McPheron and Steck 1996). The families of Leaf miners (Agromyzidae; Figure 1.1) and Gall midges (Cecidomyiidae; Figure 1.1) also contain several species known as severe plant pests impairing the cultivation of crop (Spencer 1973, 1990; Pollard 2000; Gagné et al. 2000). Black fungus gnats (Sciaridae; Figure 1.1) can cause considerable damage on seedlings by feeding on roots and carrying plant pathogens on their bodies and in their feces (Harris et al. 1996). And acknowledged as a threat to human health mosquitos (Culicidae; Figure 1.1) have been claimed to be an indirect cause of more morbidity and mortality among humans than any other group of organisms (Harbach 2007; Courtney et al. 2009). We therefore must admit that Diptera are ubiquitous and have influenced human life directly or indirectly whether we have known about them or not.

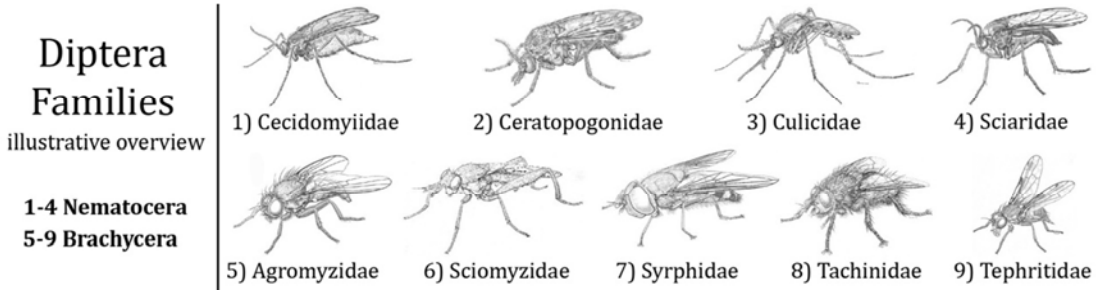


Figure 1.1: Exemplary individuals of the 9 above mentioned German Diptera families for an illustrative overview (modified after: McAlpine et al. 1981 and 1987). 1) *Cecidomyia resinicola*; 2) *Culicoides variipennis*; 3) *Aedes sticticus*; 4) *Sciara sp*.; 5) *Agromyza albipennis*; 6) *Sepedon fuscipennis*; 7) *Syrphus torvus*; 8) *Tachina florum*; 9) *Tephritis angustipennis*. Alltogether about 120 families of Diptera with more than 9000 species are known to occur in Germany (Schumann et al. 1999, 2002).

1.3.2 The challenges

1.3.2.1 Missing DNA references

In an idealized system a reference based approach would refer to a database covering all existing species associated with unique DNA sequences. Unfortunately we do have two major problems. First is: we can never be sure to have all species covered in a database, and second is: gene history does not always resemble species history (Maddison 1997). First means we do not always have a reference and second means, even if we do, we might be misled. Still the query sequences need to be differentiated in a biological meaningful way. Although it has to be kept in mind, that DNA barcoding does not define species, but discriminates molecular entities, a maximum of coherence to taxonomy should be striven for. This would make the results more seizable in the context of the typological background in the history of taxonomy. The advantage of these molecular entities is that the discrimination can be standardized across different datasets, making even yet unknown and unidentified biological life comparable across these datasets (Ratnasingham et al. 2013). Molecular entities have already been given several names in the history of genetics (Eldredge et al. 1980; Oliver et al. 1993; Moreira et al. 2002, Ratnasingham et al. 2013). Molecular Operational Taxonomic Unit (Floyd et al. 2002), also short: MOTU is the one that will continuously be used in this text when

dealing with different molecular entities. After assigning sequences to MOTUs, each MOTU can be classified according to constantly growing reference databases (Ratnasingham & Hebert 2007; O'Leary et al. 2016) and regarding current taxonomy. By this a Linnaean species name can be given to a MOTU sequence if it is found in a database. This allows further insights into the ecological composition of the samples.

1.3.2.2 Samples containing multiple DNA sources

A problem with samples containing multiple DNA sources is that these are susceptible for biases due to the often inevitable PCR steps in some of the processes. A PCR of a multiple template sample tends to amplify different sequences with different effectiveness and to introduce artefacts which are also referred to as chimeras (Suzuki and Giovannoni 1996; Cronn et al. 2002; Kanagawa 2003; Bellemain et al. 2010). These chimeras are sequence mixtures from different templates and are usually initiated through incomplete strand synthesis during PCR amplification (Figure 1.2).

Samples containing multiple species templates could influence the formation of chimeras in two different ways: by the number of species present in this sample (as the number of different sequence templates in a sample) and by the relation of the species to each other (as the similarity of the sequence templates to each other). With a higher number of different templates chances increase for an incomplete synthesized sequence to bind to a different template for completion than from which it was initially generated from. The second scenario could influence the affinity of the incomplete sequences for a different template as high template similarity can increase the affinity between the incomplete synthesized sequence and the "wrong" template for completion. To examine which of the mentioned factors play a major role, a simulated sample approach of known species compositions were therefore tested and evaluated in a pre-study to limit their influence whenever possible in later studies.

Figure 1.2: 1. Chimeras form during PCR by incomplete strand synthesis of the template (template A). 2. The incomplete complementary sequence (copy A) binds to a different template (template B) in the following cycle, where the strand synthesis is being completed. The resulting chimeric sequence acts as a template (new template AB) within the next cycles itself.

1.3.2.3 Inaccuracies in multiplex sequencing

While new sequencing technologies are making economical use of available sequencing capacities, they also introduce new obstacles. The necessary "library constructions" are preparations of the sequence targets to bring them into a form, the so called "library",

that is compatible with the sequencing system to be used (Head et al. 2014). To be able to sequence multiple samples in parallel, sample-specific indexes are embedded in one of the adapters that are added to the sequences in a library. Subsequently, multiple libraries can be pooled and sequenced together. This process is known as multiplex sequencing. After sequencing the samples can be separated again computationally based on their index sequence. This greatly increases experimental scalability, but also introduces the danger of falsely assigning sequences to their original samples. Introducing indexes in both of the adapters significantly reduces the risk of false assignments and helps identifying PCR artefacts (Kircher et al. 2012). Samples of identical origin were therefore analyzed using a single and a double indexing approach for their performance.

1.3.2.4 Avoiding errors caused by multi template fragment assemblages

Short DNA fragments of single specimens are often assembled from full length sequences up to whole genomes (Bonfield et al. 1995; Simpson and Pop 2015; Yoon et al. 2016). But despite the continuous improvements these assemblies can be inaccurate and introduce various errors into the final product. Large parts of a sequence can be omitted; others might be rearranged or otherwise deformed (Cheung et al. 2003; Schmutz et al. 2004; Salzberg and Yorke 2005; Phillippy et al. 2008; Alkan et al. 2011; Zhang and Backström 2014). Trying to simultaneously assemble short fragments from the same DNA segment (COI) from different species to again species specific full sequence barcodes only increases the chance for errors. Some regions within COI are highly conserved over different species due to their essential functions. Most of its discriminatory information resides in the non-amino acid specific bases (Ward and Holmes 2007; Forsdyke 2017). Sequence fragments need to overlap at these regions containing the specific information to correctly build species specific full length sequences. Otherwise artefact sequences are possibly being formed.

To avoid errors and artefacts sequence fragments were not assembled but sorted into two subsets according to their position in reference to a full length COI sequence (Figure 1.3). MOTU overestimation caused by separate non-overlapping sequence fragments originating from the same species will also be reduced by this as well as the computational effort will be reduced through smaller datasets. A parallel evaluation of these subsets will further allow a cross validation of the species identifications so that no species information will be lost during the process. Identification success for shorter than full length (658bp) sequences should not be a problem as species information can be high for even very short fragments (Meusnier 2008; Derocles et al. 2015; Lee et al. 2015; Villa et al. 2016).

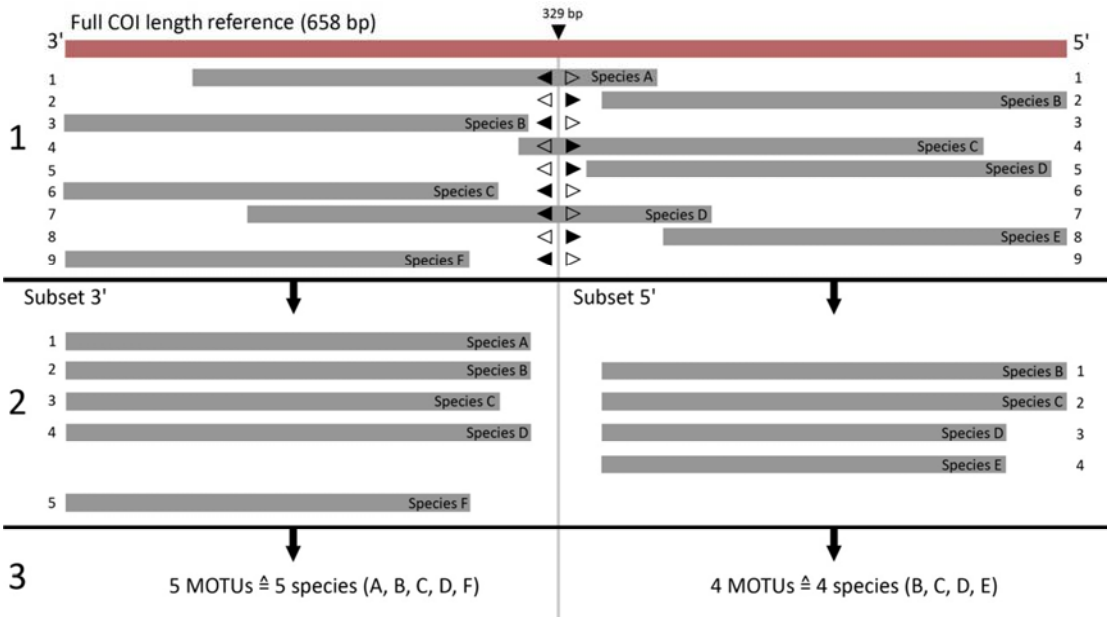


Figure 1.3: COI fragment sorting after alignment avoids erroneous fragment assemblages, reduces computational effort due to the formation of smaller subsets, and is further expected to improve the MOTU vs species ratio. Without sorting the ratio would be 9 MOTUs ≙ 6 species in this example. After sorting the rations improved to 5 MOTUs ≙ 5 species for 3'- sorted sequences and 4MOTUs ≙ 4 species (for 5'). 1: COI fragments alignment to COI reference, 2: fragment sorting according alignment position, 3: separate MOTU classification and identification.

1.3.2.5 Further problems

Nuclear mitochondrial pseudogenes (numts) can cause problems when mitochondrial markers are used. They are non-functional copies of mitochondrial DNA (mtDNA) that have been translocated into the nuclear genome (Bensasson et al. 2001; Lopez et al. 1994). Their unintended amplification have affected analyses in various taxa (Bensasson et al. 2001; Buhay 2009; Richly and Leister 2004; Song et al. 2008).

Mitochondrial heteroplasmy is the mixture of more than one type of mitochondrial genome within a single individual. The coamplification of different heteroplasmic copies of mtDNA, can lead to artificial clades and bias species numbers (Hebert et al. 2004; Hulcr et al. 2007; Rubinoff et al. 2006).

Further problems that come with a single-gene approach like DNA barcoding are introgression and hybridization as well as incomplete lineage sorting (Chase et al. 2005; Funk and Omland 2003; Meyer and Paulay 2005; Rosenberg and Tao 2008).

1.2.2 The benefits of monitoring Diptera

The observation of species with known ecosystem functions can be used as an indication for ecosystem condition and diversity (Kremen et al. 1993). A single species can be the origin of a chain reaction that alters its environment. Its introduction or removal can affect an ecosystem in different ways: it can change the availability and quality of nutrient resources, trophic resources and physic resources, as space, light, water, etc. (Vitousek 1990). This framework of resources on which other species depend helps to clarify how the presence or absence of certain species is linked to other biota (Crooks 2002). Their presence or absence can also indicate the status and the change of habitats over time. With their combination of rapidly migrating adults and almost sessile larvae, Diptera are perfect indicators to quickly notice habitat changes

due to the mobile spectrum of individuals and the impact length and intensity of the disturbance due to the lesser mobile individuals (Delettre et al. 1992; Frouz 1999).

Large scale ecosystem monitoring could meet the challenge to provide expanded spatial scale data and geographical coverage, while uncovering complex relationships through cross-taxon, multitrophic sampling that would help to understand the responses of biodiversity to drivers of ecological change and the effects of biodiversity on ecosystem properties. The problem with evaluating taxa of Diptera has been the difficult taxonomic determination and the little knowledge about the ecology of a large proportion of the Dipteran groups. However, even the identification of the lesser known species is still valuable when analyzing community structures (Timms et al. 2013). The diversity and distribution of species frequencies in a community can for example indicate the level of disturbance (Frouz 1999). And the problem of taxonomic determination gets new impulses from the development of DNA barcoding (Hebert et al. 2003b). Recent advances in sequencing technologies and processing software allowed establishing workflows that are able to deal with the high data in- and output of large scale monitoring (Hajibabae et al 2011; Porter et al. 2014). This progress is encouraging, as it will become possible to detect shifts in species composition in actual response to habitat disturbance for the first time (Geiger et al. 2016).

1.3 DNA barcoding for biodiversity assessments

1.3.1 The concept

The benefits and shortcomings of DNA barcoding have been discussed vividly. Some scientists deemed it as the end of taxonomy others envisioned a bright future (Will and Rubikoff 2004, Hebert and Gregory. 2005; Bucklin et al. 2011). Basically DNA barcoding is an identification system that relies on genomic variance. The concept of DNA barcoding is based on the assumption that a chosen DNA segment can discriminate different species as they exhibit characteristic nucleotide differences. Usually barcoding refers to the segment of cytochrome c oxidase subunit 1 (COI, CO1 or also Cox1) as it is among the most conservative protein-coding genes in the mitochondrial genome of animals, making it the preferred candidate for an identification barcode (Brown 1985; Folmer 1994; Hebert et al. 2003b). The usage is similar to the Universal Product Code (UPC) or the European Article Number (EAN) that is used to identify retail products. Alternate numerals from 0-9 at a defined number of sites (12-UPC, 13-EAN) generate billions of unique identifiers. Genomic barcodes have four alternate nucleotides at each position. The number of inspectable sites is the length of the chosen genetic marker. COI is one of three mitochondrial DNA (mtDNA) encoded subunits (MT-CO1, MT-CO2, MT-CO3) of the respiratory complex IV. Its coding sequence with a length of around 658 bp (base pairs) has proven to be a capable universal discriminator (Folmer et al. 1994; Hebert et al. 2003a; Hebert et al. 2005) and is already widely used (Ratnasingham & Hebert 2007). This is because the mitochondrial DNA mutation rate of 0.02 substitutions per base pair per million years exceeds the evolution of nuclear DNA by a factor of 10 (Brown et al. 1979). A substitution rate this high is expected to be sufficient to discriminate species as it means that two close species that diverged over one million years ago probably gathered around 13 genetic substitutions in their COI sequence. Because of these species specific differences in this DNA fragment, the COI segment of previously identified specimen can function as molecular markers for these species and

can be stored in a reference database. Species name and COI sequence are then associated under the assumption that every species has its own characteristic DNA sequence. Any time a documented DNA sequence is found again, it can be assumed that it originated from the associated species, no matter what life stage, or gender the DNA originated from (Hebert et al. 2003b).

1.4 Hybridisation based target enrichment

1.4.1 The exclusive selection of a genomic region

It is very probable that in the near future whole-genome sequencing of organisms will become routine. While gaining a deeper understanding of the full spectrum of genetic variation at the same time many of the above mentioned problems will be solved. However, financial and temporal inefficiency does not yet allow sequencing large numbers of complex genomes in their entirety. The extent of sequencing data being produced by this would also overstrain current software applications and data storage capabilities. Therefore current efforts concentrate on selectively capturing genomic regions from DNA samples before sequencing (Mamanova et al. 2010; Hedges et al. 2011). In the past important steps before sequencing required labor-intensive marker development together with single-locus polymerase chain reaction to amplify a region of interest. However, a PCR simultaneously using multiple different DNA templates is susceptible to various biasing factors as the above mentioned forming of chimeras and the possible exclusive amplification of only parts of the initially used DNA template. If in contrast the experimental focus is not the amplification of a single gene from different sources but the amplification of multiple genes from a single source, the simultaneous use of many primer pairs can generate a high level of nonspecific amplification, caused by interaction between the primers, and moreover amplicons can also fail to amplify (Suzuki and Giovannoni 1996; Mamanova et al. 2010). Opposing to the process of amplification, allows the hybridization based target enrichment (also known as hybrid capture) to eliminate all genomic DNA regions that are of no experimental interest (Figure 1.4). This enables researchers to specifically enrich only those genes, exons or other genomic regions that are relevant for a survey (Mamanova et al. 2010; Lemmon et al. 2012; Kozarewa et al. 2015).

Hybrid capture comprises different techniques that accumulate the targeted DNA before the sequencing process. Oligonucleotide probes (also referred to as baits), designed to be specifically similar to the target, are used to hybridize to the targeted sequence sections in a DNA sample. The hybridization of target regions can occur either on a solid surface (microarray) or in solution (Mamanova et al. 2010; Kozarewa et al. 2015; Mayer et al. 2016). In this study the focus was put on the solution-based enrichment as it has several advantages over the array-based approach.

1.4.1.1 Array-based target enrichment

In an array-based hybrid capture, genomic DNA is first sheared into fragments of a desired size range and a sequencing library is being constructed. An immobilized probe is then used to capture the targets in the fragmented library. Nonspecific hybrids are then washed away and hybridized probes are eluted. While array-based capture methods are efficient in enriching targets of interest, they have several limitations, such as expensive hardware, limitations to the number of samples and the relatively large

amount of DNA needed (around 10–15 µg) for the library preparation (Mamanova et al. 2010).

1.4.1.2 Solution-based target enrichment

Solution-based target enrichment is designed to overcome the limitations imposed by array-based capture methods. In contrast to the later, the solution capture has an excess of probes over the target allowing lower DNA quantities for a library construction (around 3 µg). A pool of probes targeting the desired region is added in solution to the readily prepared library. The hybridized probes are then captured and purified by magnetic beads and subsequently amplified and sequenced. The solution-based target enrichment can be performed in a common 96-well plate and is more readily scalable than the array-based target enrichment as the later relies on a time consuming usage of microarray slides that only function with a special purchasable hybridization station (Mamanova et al. 2010; Ernani and LeProust 2016).

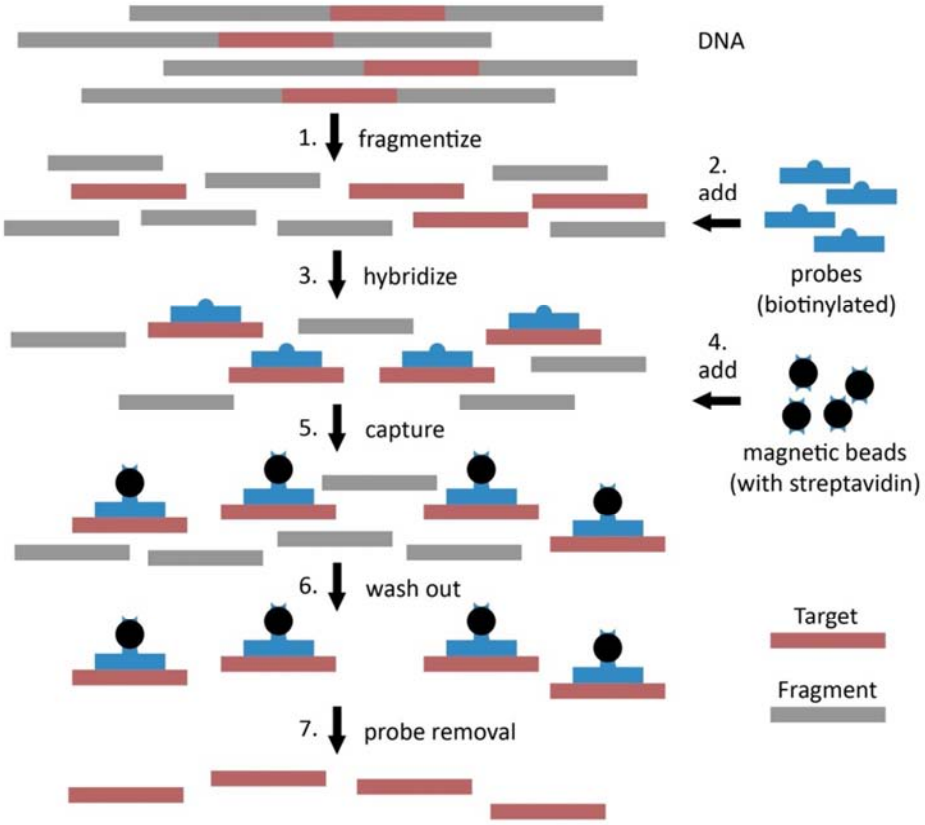


Figure 1.4: 1. During in-solution hybridization based target enrichment genomic DNA is mechanical or enzymatically fragmentized. 2.-3. RNA probes designed to identify and hybridize with the fragments of interest (the targets) are added to the sample. 4.-5. Magnetic beads covered with streptavidin are added to the sample and bind to the biotin of the probes. 6. While the targets are held back with a magnet the remaining fragments are washed out. 7. The targets are then extracted by removing beads and probes through RNA digestion.

1.4.2 Designing specific oligonucleotide probes

All companies have an assortment of already prefabricated oligonucleotide probes for various standard applications. Although the companies products can vary in length and their nucleic acid structure (e.g. Roche: 60–90-mer DNA; Agilent: 150-mer RNA), all oligonucleotides are biotinylated. This process generates a covalent bond between a nucleic acid and biotin. This bond is very specific and unlikely to perturb the function

due to the small size of biotin. It binds to streptavidin with an extremely high affinity and is therefore used for easy capture onto streptavidin-labeled magnetic beads (Green 1975; Mamanova et al. 2010; Ernani and LeProust 2016).

With increasing probe-to-target DNA sequence distance the enrichment effectivity of the probes decreases. Companies therefore also provide the option to customize the probes. However, designing probes covering a range of distantly related species can still pose a challenge. A high probe-to-target nucleotide sequence similarity in one species lineage can lead to a low similarity to another lineage. As a consequence these lineages can significantly differ in their enrichment quantity or even lead to a total enrichment failure in one lineage. In such cases more than one probe needs to be designed. Until recently no software was available that optimizes the number of probes for enriching target loci across a diverse group of species by dynamically adjusting the number of probes to the known taxonomic ingroup target locus divergence. Inspired by an approach that uses sections of infered ancestral nucleotide sequences as probes a new software (BaitFisher) was published that reduces redundancy and probe-to-target distances even further with a special sectional clustering approach (Hugall et al. 2015; Mayer et al. 2016). From this progress a more comprehensive enrichment of the targets can be expected. BaitFisher was therefore used in this study to design Diptera specific probes for the hybrid enrichment from a large set of Diptera sequences.

1.5 The necessity to build up a sequence reference data pool

A serious limitation to the utility of DNA barcoding is human error and uncertainty in creating and curating reference libraries (Collins et al. 2012). Of course identifications can only be as good as the quality of the references used. But large databases with several entries from different scientists do not only gather valuable data but also errors, contaminations and conflicting data (Becker et al. 2011; Collins 2012). The difficulty of maintaining their integrity grows with its size (Learn GH Jr et al. 1996; Yao Y-G et al. 2009). And although genetic databases are being generated all over the world (Ratnasingham S & Hebert PDN 2007; Haszprunar G 2009; O'Leary NA et al. 2016) it can be considered necessary to have a high quality reference of reasonable size when the goal is establishing and processing and identification workflow. A well maintained reference data stock allows more precise interpretation of the query data, because pitfalls (like species sharing identical barcodes) are known. Building up a high quality reference list for German Diptera sequences was therefore an essential part of the project. However, covering species diversity in Germany was neither necessary nor a realistic short term goal due to the large number of species (>9,000 species; checklist in Schumann et al. 1999, 2002). Every year new discoveries are added. And still there are more undescribed species in the hands of specialists awaiting species description. However, estimations are that there are about 300 species that are frequent in samples taken in Germany. All frequently caught species can be sequenced for a sequence reference list. By doing this the majority of the abundant species that would be compared to this reference list would be covered by it.

1.6 The advantage of automated sampling

Different techniques for catching Insects have been developed and improved over the years. While the common picture of a biologist is still more Linnaean oriented including a sweeping net to chase butterflies, various capturing methods and tools have extended

the repertoire for broader and differentiated applications since then (Abraham 1991). A widely used and very effective way for collecting flying insects is the Malaise trap (Malaise 1937; Ronquist 2010; Geiger et al. 2016; Aagaard et al. 2017). Although the tent-like construction is especially effective in capturing flying insects, also arthropods, including wingless species, can be caught accidentally, since they crawl into the net (Geiger et al. 2016). Continuous sampling over a longer period of time could therefore reflect a considerable amount of the extant fauna at the chosen localities. Additionally it is possible to monitor seasonal and annual changes. And as the tools for objective species differentiation are improving, it only seems consequent to also keep up with the automation of methods for collecting (Hebert et al. 2003; Hebert et al. 2009; Miller et al. 2016).

An essential necessary improvement appears to be obvious: Automation. With a single proximity Malaise trap a weekly change of the collecting jar seems to be a feasible effort, but with every additional trap that is set up the effort becomes more laborious. If these traps are then distributed over a larger geographical area or even set up in different countries, a weekly maintenance becomes very costly and at the end unfeasible. Individual advances have already been made (Simon et al. 2001; Selby et al. 2014) Published data show that automated traps can save 80% and more of the costs incurred, while efficiency even increases over time (Selby et al. 2014). The ideal trap would therefore be self-maintaining over an adequate period of time and by this provide a basis for extensive monitoring and standardized biodiversity assessments. Another benefit would be the avoidance of habitat disturbances as regular maintenance visits inevitably influence the study. And in times of climate change, traveling and its impact on the environment could be kept to a minimum.

1.7 The Biodiversity Exploratories

This study was part of a DFG priority program. Namely the “Biodiversity Exploratories”, a project for large-scale and long-term functional biodiversity research with contributors from all kinds of different science fields. Within this DFG project three larger areas in Germany were chosen beforehand for researchers to focus on (Figure 1.5). These areas, the “Exploratories” were planned as representatives for the variation in grassland and forest habitats and areas of typical land use in Germany. In its entirety it consists of (1) the UNESCO Biosphere Reserve Schorfheide-Chorin, which is situated in the lowlands of North-eastern Germany, a young glacial landscape with many wetlands, (2) the National Park Hainich and its surrounding areas, situated in the hilly lands of Central Germany, and (3) the UNESCO Biosphere Reserve Schwäbische Alb (Swabian Jura), which is situated in the low mountain ranges of South-western Germany (Fischer et al. 2010).

These three exploratories now serve as an open research platform for the various biodiversity and ecosystem research groups. The gathering of data from different research disciplines and its interlinkage has the potential to overcome disciplinary restrains and to achieve comprehensive insights in biodiversity and ecosystem functioning. It bears the chance for researchers to newly combine and integrate data from different fields. A major goal for al researchers in the Exploratories is therefore to overcome disciplinary isolation and to form productive data synthesis. The fundamental study objectives in the different research groups are:

- the understanding of the relationship between biodiversity of different taxa and levels
- the role of land use and management for biodiversity and
- the role of biodiversity for ecosystem processes

The chosen approach to reach these goals is a combination of biodiversity monitoring and ecosystem measures combined with manipulative experiments in a disciplinary overarching study design. A challenge for all researchers is not only the work and its results in a single project but the interdisciplinary integration of data for a comprehensive understanding of drivers and functional consequences of all facets of biodiversity.

An automation of data gathering should facilitate this process and allow applying more resources into necessary analyses. The development and testing of an automated sampler can therefore also be considered as valuable for all researchers. The samples acquired by this were planned as a realistic mass-sample supply for the in situ test of a high throughput, next generation sequencing (NGS) workflow. The chosen plots to set up prototypes of the automated insect sampler were located in the Schorfheide-Chorin.

Schorfheide-Chorin	
Location	NE Germany
Size	Ca. 1300 km²
Geology	Young glacial landscape
Human population density	23 km⁻¹
Altitude a.s.l.	3 –140 m
Annual mean temperature	8 – 8.5 °C
Annual mean precipitation	500 – 600 mm

Hainich-Dün	
Location	Central Germany
Size	Ca. 1300 km²
Geology	Calcareous bedrock
Human population density	116 km⁻¹
Altitude a.s.l.	285 – 550 m
Annual mean temperature	6.5 – 8 °C
Annual mean precipitation	500 – 800 mm

Schwäbische Alb	
Location	SW Germany
Size	Ca. 422 km²
Geology	Calcareous bedrock with karst
Human population density	258 km⁻¹
Altitude a.s.l.	460 – 860 m
Annual mean temperature	6 – 7 °C
Annual mean precipitation	700 – 1000 mm

Figure 1.5: Location and characteristics of the different Biodiversity Exploratories in Germany, representing a model for German landscapes and landuse characteristics. Part of this project focused on forests in the area of Schorfheide-Chorin.

1.8 Outlining the project goals

Considering the outlined challenges a basic question was the suitability of DNA barcoding for large scale biodiversity monitoring and the feasibility of the establishment of a Diptera focused workflow for rapid biodiversity assessment. This included i) the collection of DNA barcode data as a reference with a focus on German Diptera, ii) the testing of the efficacy of the genetic marker, iii) capture and evaluate

realistic mass-samples iv) the work with MOTUs for an evaluation aside Linnaean-references, iv) establishing a workflow oriented on this task, including a) a reliable capture device to provide sufficient suitable material and b) using parallelized high-throughput sequencing technologies, to initiate cost-effective and comprehensive studies of biodiversity. An exemplary and basic insight on the difference of diversity between two forest types was therefore planned.

2 A Diptera sequence reference list and COI discriminability

2.1 Introduction

Testing a genetic marker for its suitability to discriminate different species is elemental for a system aiming for species identifications based on their DNA. Cytochrome c oxidase subunit I (COI) was chosen for its already documented successful usage in previous researches (Folmer et al. 1994; Hebert et al. 2003a; Hebert et al. 2005.) Still it had to be confirmed that the chosen DNA section is a) amplifiable across different taxa of Diptera for PCR usage and b) discriminative to be able to differentiate between the species of those taxa.

A range of methods have been proposed, that apply different criteria for the purposes of assigning taxonomic identity based on DNA sequence data (Birch et al. 2017). One of these is the application of distance based methods to test the barcode efficacy. Combined with a distance threshold it can distinguish between inter- and intra-specific comparisons and by this prevent misidentifications that for example originate from the absence of conspecifics in the reference (Meier et al. 2006). Similarly monophyly has been used as a valuable description of data in terms of NJtrees. Still, unless nested directly within a cluster, the tree alone yields not very much information to discern if an "unknown" belongs to the closest topological species or not (Collins et al. 2012). With an incomplete reference library interpretations can therefore be misleading. However, with the specimens' identity previously known and a dataset with sufficient conspecifics tree evaluations still are valuable to validate specimen identifications. Criteria for such an approach have been given in Meier et al. 2006 and Birch et al. 2017. And alternative approach and especially promising for future high-throughput workflows is the use of a Bayesian classifier due to favorable trade-offs among automation, speed and accuracy (Newton and Roeselers 2012; Werner et al. 2012; Porter et al. 2014). Before continuing these distance-, tree- and Bayesian-based methods were used to assess barcode efficacy.

Diptera material came from identified samples taken in the exploratories. It was initially provided by the Department of Ecology and Ecosystem-management at the TU Munich. Amplification testing went along the gathering of the Diptera reference data. Common primer combinations were tested throughout the process, starting with LCO-1490 and HCO-2198 (Folmer et al. 1994), continuing with alternative primer combinations (Simon et a. 1994).

2.2 Material and Methods

2.2.1 Material acquisition and processing

Diptera material was provided by the TU Munich and consisted of Diptera samples from previous captures from the Exploratories. The provided material was supplemented through collections of the University of Bielefeld and personal collections of the taxonomist Michael von Tschirnhaus.

2.2.1.1 DNA extraction protocol for the Diptera database

For DNA extraction the following steps were conducted using the Qiagen DNeasy Blood and Tissue Kit. For large specimen (like Muscidae or Asilidae) a single leg was directly

incubated in 200 µl lysis buffer (180 µl ATL + 20 µl Proteinase K) in a 1.7 ml tube. Small individuals (e.g. Agromyzidae and Chloropidae) were put into the tube as a whole. Lysis lasts overnight in a thermo shaker set at 55°C. After that the individual is removed and stored in 99.7 % ethanol. 200 µl of AL buffer are added to the tube and incubated for 10 min at 70° C. After short vortexing 200 µl of pure ethanol are added and vortexed again to stop the lysis. The whole content is then pipetted into a DNeasy mini spin column and centrifuged 1 min at 8,000 rpm.

The mini spin column is then put in a 2ml collection tube, whereas the flow-through is discarded. After adding 500 µl AW1 buffer the mini spin column is centrifuged again 1 min at 8000 rpm. The mini spin column is again placed in a 2ml collection tube and the flow-through is discarded. Now 500 µl AW2 buffer are added and the column is centrifuged 3 min at 13,000 rpm. Finally the column is placed in a 1.7 ml tubed. 50 µl AE buffer are added and incubated 1 min at room temperature. The DNA can then be eluted centrifuging the column 1 min at 8,000 rpm. The last step is then repeated with another 50 µl AE buffer. The extraction can then be measured for its DNA concentration and prepared for PCR.

2.2.1.2 Measuring the DNA concentration

DNA concentration was measured with a Promega Quantus Fluorometer. First 98 µl of diluted TE (1:20 dilution of original 20xTE) were pipetted in a 0,5 ml tube. Then 2 µl of the sample extract was added. After this 100 µl of diluted dsDNA Dye (1:200 dilution of original 200xDye) was added and mixed thoroughly. The sample is then incubated for 5 min in the dark. After this the sample can be measured.

2.2.1.3 PCR amplification

For DNA amplification the Qiagen Multiplex PCR Kit was used on a thermocycler (Applied Biosystems 2720 Thermal Cycler). The reaction mixture of 20 µl for each sample consisted of the following constituents: 2,3 µl of H2O, 2 µl Q-Solution, 10 µl Qiagen Multiplex Mix, 1,6 µl of the forward primer (10 pmol/µl), 1,6 µl of the reverse primer (10 pmol/µl) and 2,5 µl from the DNA extraction. Primers for amplification were chosen from a list of commonly used standard primers (Table 2.1). LCO-1490 (3'GGTCAACAAATCATAAAGATATTGG'5; Folmer et al. 1994) was used as the standard forward primer. HCO-2198 (5'TAAACTTCAGGGTGACCAAAAAATCA'3; Folmer et al. 1994) was chosen as the reverse primer. After mixing all constituents the samples were placed in the thermocycler for amplification (Figure 2.1). The program started with an initial 15 min at 95°C, followed by 15 cycles of the following three steps: denaturation at 94°C for 0:35 min, annealing at 55°C for 1:30 min and an elongation step at 72°C for 1:30 min. At each cycle the annealing temperature drops 1°C. After the first 15 cycles the program continued with another 25 cycles repeating the following three steps: denaturation at 94°C for 0:35 min, annealing at 50°C for 1:30 min and an elongation step at 72°C for 1:30 min. The last elongation step is then prolonged for another 10 min. At the end the temperature is cooled down to 10°C until the samples are removed from the thermocycler. PCR success was checked via agarose gel electrophoresis. In case of a negative PCR result the reverse primer was substituted by the primer C1-N-2191 (5'CCCGGTAAAATTAAAATATAAACTTC'3) and the PCR amplification was repeated. When success still failed primers where exchanged with its JJ-variants (Table 2.1). After

successful amplification the products were sent to Macrogen Europe (Amsterdam, Netherlands; http://www.macrogen.com) for bidirectional sequencing.

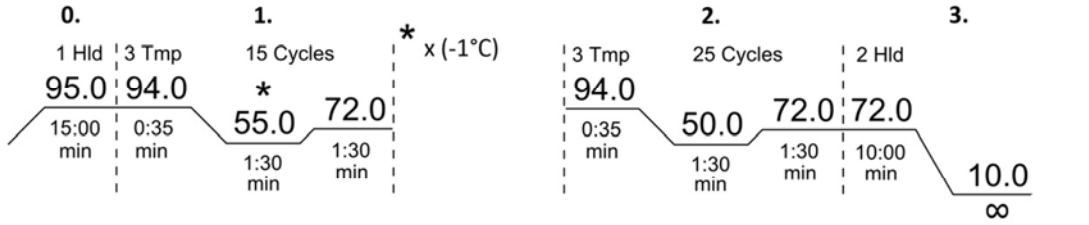


Figure 2.1: Graphic of the thermocycler program used during amplification of the reference sequences.

Name	Sequences	direction	References
LCO-1490	3'GGTCAACAAATCATAAAGATATTGG'5	forward	Folmer et al. 1994
HCO-2198	5'TAAACTTCAGGGTGACCAAAAAATCA'3	reverse	Folmer et al. 1994
C1-N-2191 (Nancy)	5'CCCGGTAAAATTAAAATATAAACTTC'3	reverse	Simon et al. 1994
LCO1490-JJ	3' CCRATATCTTTATGRTTWGTDG'5	forward	Astrin and Stüben 2008
HCO2198-JJ	5'AWACTTCVGGRTGVCCAAARAATCA3'	reverse	Astrin and Stüben 2008

Table 2.1:. List of primers that were used during PCR amplification.

2.2.1.4 Software processing of the sequence reads

After sequencing the sequences were assembled and edited with Geneious, using the versions 5.4.4 – 7.1.9 (Kearse et al. 2012; Biomatters Ltd.; http://www.geneious.com). The assembler was used on "Highest Sensitivity / Slow" mode. The maximum number of gaps that was allowed to be inserted into each read was set to 20% of the size of the overlap between the two reads. The maximum gap size was set to 5 bases. The minimum number of bases to match between two reads was set to 10 bases. The number of allowed ambiguities within this match was set to 16. Sequences were generated from the contiguous sequences using a 100% threshold meaning that only bases matching all sequences are used. The assigned quality score was calculated as the total of the contributing quality scores from both reads. Sequence reads and all additional information like species taxonomy, locality, GPS data, date, capturing method and ID number, were then connected using BioCode LIMS, an integrated laboratory information management system for DNA barcoding.

2.2.2 Testing universality and discriminative properties of COI

The universality of the genetic marker depends on its feasibility to be easily obtained from different Diptera taxa. The amplification success within the different families of Diptera was already checked when the first Diptera sequences were processed for the reference database. The PCR success was checked via agarose gel electrophoresis. For details see the previous chapter 2.2.1.3.

To test the discriminative properties of the COI sequence up to five different specimens per species from 12 of the most abundant Diptera families sampled in the exploratories in 2008 were chosen to be analyzed (List of abundant Diptera families in 2008, Supplement S1.1). Sequences were obtained from 166 different specimens (Table 2.2).

Family	Chosen Species	N	Family	Chosen Species	N
Calliphoridae	*Calliphora rohendorfi*	2	Hybotidae	*Hybos culciformes*	5
	Calliphora vicina	5		*Hybos femoratus*	5
	Calliphora vomitoria	5		*Hybos grossipes*	2
Cecidomyiidae	*Bryomyia apsectra*	5	Muscidae	*Coenosia agromyzina*	5
	Bryomyia bergrothi	3		*Coenosia albicornis*	5
	Bryomyia gibbosa	3		*Coenosia mollicula*	5
				Helina depuncta	5
Ceratopogonidae	*Atrichopogon levis*	4		*Helina evectra*	5
	Culcoides tauricus	5	Mycetophilidae	*Sciophila lutea*	5
	Stilobezzia ochracea	5		*Sciopihila nigronitida*	3
Chironomidae	*Tvetenia discoloripes*	4		*Tarnania fenestralis*	5
	Tvetenia veralli	5		*Tarnania nemoralis*	6
Chloropidae	*Platycephala planifrons*	5	Phoridae	*Megaselia angusta*	5
	Tricimba cincta	3		*Megaselia manicata*	5
	Oscinella frit	2		*Megaselia melanocephala*	5
Empididae	*Empis chioptera*	5	Sciaridae	*Leptosciarella fuscipalpa*	5
	Empis tesselata	5		*Leptosciarella wiebke*	5
	Empis univittata	3	OUTGROUP		
Fanniidae	*Fannia canicularis*	5	Ceratophyllidae	*Amalaraeus sp*	1
	Fannia lustrator	4	Hystrichopsyllidae	*Hystrichopsylla sp*	2
	Fannia serena	4			

Table 2.2: List of 39 species chosen to test the discriminative properties of COI. Species were chosen from the 12 most abundant Diptera families of the exploratories in 2008 (Supplement S1.1). Three specimens from two species of the order Siphonaptera were included into the dataset as outgroup species. The number of sequences for each species is given by N. In total 166 Sequences were examined.

2.2.2.1 Pairwise distance analysis

Species Identifier 1.8 was used for a distance based analyses as suggested in Meier et al. 2006. Each sequence was queried against the rest of the dataset of the chosen 39 species using a pairwise distance analysis. In a "best match" (BM) approach the query was considered a success when the sequence with the smallest distance to the query sequence was a conspecific. For a "best close match" (BCM) success a sequence had to be also within the 3% threshold of all intraspecific distances. Sequences were considered ambiguous when the closest sequences were a mixture of other and conspecific sequences (BM) and within the 3% threshold of all infraspecific distances (BCM) or when there was no match within the 3% threshold of all infraspecific distances (BCM). The query was considered misidentified when the closest sequence was not conspecific (BM) or when none of the matches within the 3% threshold are conspecific (Table 2.3). *Amalaraeus sp*. was not taken into account whenever the rating criteria required a conspecific in the dataset.

		Best Match (BM)	Best Close Match (BCM)
1	**identified**	Sequence(s) with smallest distance to query all conspecific	Sequence(s) with smallest distance to query conspecific and within the 3% threshold of all intraspecific distances
2	**ambiguous**	Sequence(s) with smallest distance to query a mixture of conspecific and other sequences	Sequence(s) with smallest distance to query a mixture of conspecific and other sequences and within the 3% threshold of all infraspecific distances
3	**misidentified**	Sequence(s) with smallest distance to query not conspecific	None of the matches within the 3% threshold are conspecific

Table 2.3: Identification criteria for a direct sequence comparison modified after Meier et al. 2006.

2.2.2.2 Clustering analysis

For the clustering analysis, the sequences were clustered at a 3% level using the Species Identifier 1.8 (Meier et al. 2006). A cluster was considered successful and the query

sequences as identified, when the cluster contained only sequences of a single species. It was considered ambiguous, when the cluster contained sequences of more than one species. The cluster was considered unidentified, when it contained only a single sequence.

1	**identified**	Query clusters only with conspecifics
2	**ambiguous**	Query clusters with more than one species
3	**unidentified**	Query does not cluster with other sequences “singleton”

Table 2.4: Identification criteria for a cluster analysis modified after Hebert et al. 2003b and Meier et al. 2006.

2.2.2.3 Tree based evaluation

A MUSCLE alignment with a maximum of 1,000 iterations was generated from the dataset. The maximum number of trees to build was limited to 5. The settings for the distance measure were kmer4_6 for the first iteration and pctid_kimura for all subsequent iterations. The clustering method for all iterations was UPGMB for free rooting. From this alignment a NJtree and a NJnetwork were calculated. For the NJtree the Geneious tree builder was used with the Tamura-Nei genetic distance model and *Amalaraeus sp.* as an outgroup sequence. The NJnetwork was calculated using SplitsTree4 v4.14.5 (Huson and Bryant 2006). Identification success was evaluated on species level following the criteria found in Birch et al 2017 (Table 2.5; Figure 2.2) and referring to Meier et al. 2006. *Amalaraeus sp.* was again not taken into account in this analysis as it was only represented by a single sequence in the dataset.

1	**identified**	a) Sequence at least one node into clade consisting of only conspecifics b) Sequence is a sister to a clade with only conspecifics c) Sequence in a polytomy with only conspecifics
2	**ambiguous**	a) Sequence in a polytomy with at least one conspecific and one different species b) Sequence is a sister to a clade with conspecifics and different species
3	**misidentified**	a) Sequence at least one node into clade consisting of only different species b) Sequence is a sister to a clade with different species c) Sequence in a polytomy with only different species

Table 2.5: Criteria for tree based sequence validation modified after Meier et al. 2006 and Birch et al. 2017.

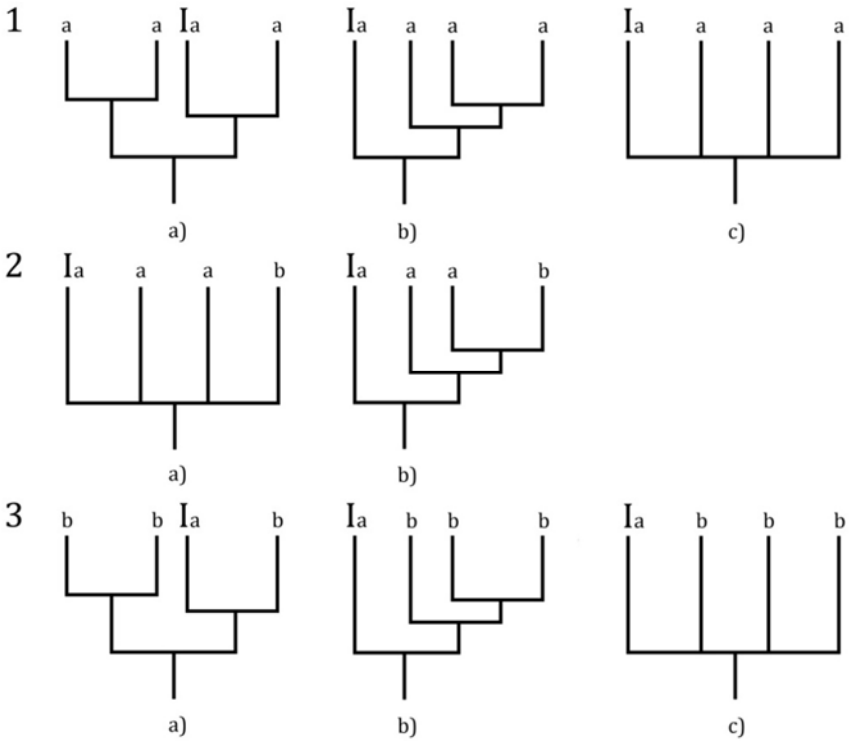


Figure 2.2: Criteria illustration for tree based sequence validation. 1-identified, 2-ambigous, 3-misidentified. "I" stands for the sequence that needs to be validated, "a" and "b" for the sequence variants of two different species (Modified after Meier et al. 2006 and Birch et al. 2017).

2.2.2.4 The Bayesian classifier

A Bayesian classifier implemented in the Mothur software (v.1.39.5) was used for species identifications (Wang et al. 2007; Schloss et al. 2009). The naïve Bayesian classifier processes the query sequence kmer by kmer, calculated the probability the query sequence would be in the given template taxonomy based on the kmers it contains. The assignment was then checked by a bootstrapping algorithm to find the confidence limit of the assignment. The confidence threshold was set to 80% (Wang et al. 2007; Cole et al. 2014; Porter et al. 2014; Vinje et al. 2015). The reference template contained about 48,000 sequences from over 10,000 species covering several different classes. Most sequences belong to the class of Insecta. It was created from the GBOL database comprising also the initial Diptera sequence list and was also applied in later experiments. It is referred to as the "GBOL-reference". During the classification process a copy of the query sequences remained in the reference. A species was considered identified when the assigned species matched the species of the query sequence. It was considered ambiguous when a different species was assigned to the query sequence sharing the identical COI sequence. The assignment was considered unidentified whenever the Bayesian classifier was not able to identify the sequence down to species level.

2.3 Results

2.3.1 Material acquisition and library growth

In the first three years the Diptera database had severe problems with the acquisition of suitable material for the database. The main reasons were:

1. Supplied specimens were often not in a suitable condition to extract utilizable DNA from it.
2. Supply with specimens from TU Munich was sparse due to lacking taxonomists for the necessary species identifications.
3. Prejudiced taxonomists refused to provide material for DNA barcoding because of two major reasons:
    a) The believe barcoding could threaten their future employment and
    b) lacking trust in the method

With the funding of the financially bigger scaled GBOL (German Barcode of Life, Pietsch & Rulik 2014) project at the same institute collaboration was the logical decision. The Diptera database was integrated into the GBOL database. For this all generated sequences and additional information were transferred to the GBOL database whereas all future emerging data are shared. The GBOL database is growing rapidly since then (Figure 2.3).

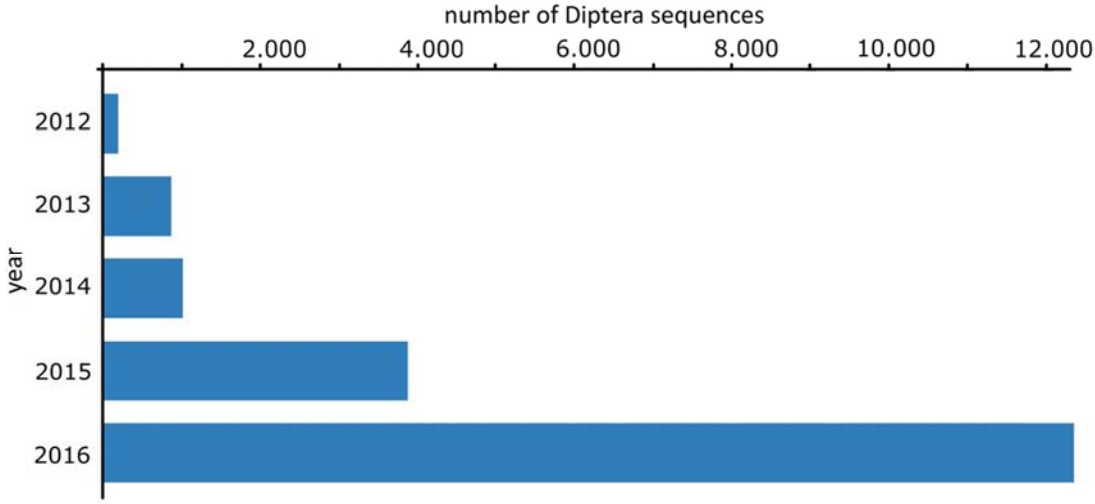


Figure 2.3: Growing number of the sequence references for Diptera in the past years in the project.

2.3.2 The universality and discriminative properties of COI

2.3.2.1 The universal application of COI

As the PCR amplification process was essential from the early beginning it soon became apparent that the common standard primers worked well throughout the tested Diptera families (Figure 2.4). Irregularities or amplification failures did not exceed normal laboratory routine. However, if at any time an alternative primer combination was needed literature offered several substitutes (Folmer et al. 1994; Simon et al. 1994; Astrin and Stüben 2008). Although GBOL has its workflow adapted to the "JJ-variants" of the primers HCO and LCO after laboratory work was committed to the GBOL employees, the previously used common standard primers as described in chapter 2.2.1.3 worked well during the first material acquisition (Figure 2.4).

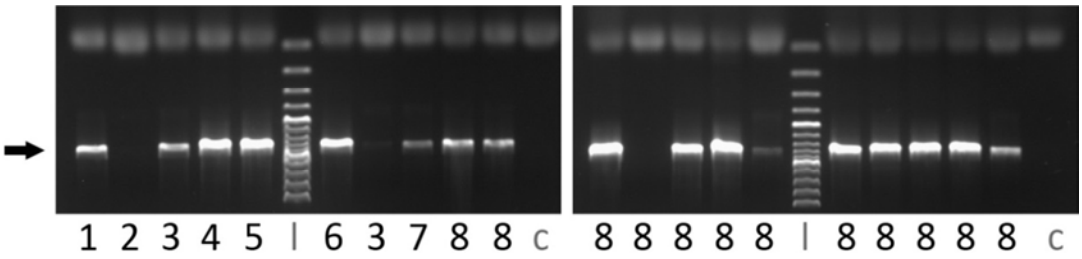


Figure 2.4: Positive gel control example of successful PCR using the primer combination HCO-LCO (January 25th 2012). Arrow points at amplified product. The PCR run contained the following families: 1-Anthomyiidae, 2-Hybotidae, 3-Muscidae, 4-Heleomyzidae, 5-Syrphidae, 6-Empidae, 7-Calliphoridae, 8-Chloropida (l-ladder, c-control). The DNA of number 2 was successfully amplified in a second run.

2.3.2.2Pairwise distance analysis

After using a pairwise distance approach to test the discriminative poperties of COI (as suggested in Meier et al. 2006) it showed that after testing each of the specimens sequences against the remaining sequences of all 39 species, identification success was very high (99.4%; Table 2.6) when applying the criteria for "best match" (Table 2.3). This means that in the pairwise distance analysis of 165 sequences of 39 species from the 12 most common Diptera families in the Schorfheide, 164 query sequences were identified by matching with a conspecific sequence. Only in a single query did the closest match not result in a conspecific match. The misidentification was a congeneric match within the family of Cecidomyiidae (Query sequence: *Bryomyia bergrothi*, incorrect match: *Bryomyia apsectra*, difference to query sequence: 51 bp / 7.8 %, differences to next conspecific: 57 bp / 8.7 %). According to "best close match" (Table 2.3) this match is not a misidentification as these sequence distances violate the 3% threshold. The query must therefore be considered as unidentified. Another eight sequences (5.4%) were also classified as unidentified according to the best close match criteria. Still the identification success for best close match is high with 94.6% identifications. The number of misidentifications is 0 (0.0%; Table 2.6)).

Pairwise distance analysis	**N**	**%**
Number of sequences in the dataset:	166	
Sequences with at least one matching conspecific:	165	100%
Sequences with a closest match at 0%:	81	49.1%
Allospecific matches at 0%:	0	0.0%
Correct identifications according to BM:	**164**	**99.4%**
Ambiguous according to BM:	0	0.0%
Misidentified according to BM:	1	0.6%
Correct identifications according to BCM:	**156**	**94.6%**
Ambiguous according to BCM:	0	0.0%
Misidentified according to BCM:	0	0.0%
Unidentified according to BCM (Sequences without any match closer than 3.0%):	9	5.4%

Table 2.6: Pairwise distance evaluation of the COI sequence of the most common Diptera families in the Schorfheide to test COI discriminative properties. *Amalarus sp*. was removed from the analysis as it had no conspecific sequence which was required for the analysis. N: Number of sequences. %: Percent of Sequences. Criteria for "best match" (BM) and "best close match" (BCM) have been suggested by Meier et al. 2006 and can be found in table 2.3.

2.3.2.3 Clustering evaluation

After a 3% clustering the 165 sequences of 38 species were distributed over 46 clusters. One cluster (2.2%) consisted of two species (species lumping: *Calliphora rohdendorfi* and *C. vicina*) and was therefore considered ambiguous according the criteria in table 2.4. Three clusters (6.5%) consisted of only a single sequence and were therefore considered unidentified. 42 clusters (91.3%) and its sequences were considered

successfully identified as the query sequences clustered only with conspecific sequences. The initial number of Linnaean species was 38 (without *Amalarus sp*. as it has no conspecific sequence). This results in a species overestimation of 21%, if the 46 clusters were considered as species-like entities. This overestimation is due to 17 species splits taking place in eight species which violated the selected 3% distance threshold. Tables can be found in the supplementary.

Clustering analysis (clustering at 3%)	N*	N
Original number of species:	(39)*	38
Number of clusters:	(47)*	46
Clusters with only one species:	(46)*	45
Clusters containing only one sequence:	(4)*	3
Clusters corresponding to traditional taxonomy:	(29)*	28
Species splits	(17)*	17
Species lumped	(1)*	1
Largest number of species in a cluster:	(2)*	2

Table 2.7: Summary of clustering analysis of 165 sequences clustering at 3% level. *: without removal of *Amalaraeus sp*. N: Number of clusters.

Species	**clusters**	**With other species**	**Species**	**clusters**	**With other species**
Amalaraeus sp	1	0	*Hybos culiciformis*	1	0
Atrichopogon levis	1	0	*Hybos femoratus*	1	0
Bryomyia apsectra	2	0	*Hybos grossipes*	1	0
Bryomyia bergrothi	2	0	*Hystrichopsylla sp*	1	0
Bryomyia gibbosa	1	0	*Leptosciarella fuscipalpa*	1	0
Calliphora rohdendorfi	1	1 (*C. vicina*)	*Leptosciarella wiebke*	1	0
Calliphora vicina	1	1 (*C. rohdendorfi*)	*Megaselia angusta*	1	0
Calliphora vomitoria	1	0	*Megaselia manicata*	1	0
Coenosia agromyzina	1	0	*Megaselia melanocephala*	1	0
Coenosia albicornis	1	0	*Oscinella frit*	2	0
Coenosia mollicula	1	0	*Platycephala planifrons*	1	0
Culicoides tauricus	2	0	*Sciophila lutea*	1	0
Empis chioptera	2	0	*Sciophila nigronitida*	1	0
Empis tessellata	1	0	*Stilobezzia ochracea*	1	0
Empis univittata	2	0	*Tarnania fenestralis*	1	0
Fannia canicularis	1	0	*Tarnania nemoralis*	1	0
Fannia lustrator	1	0	*Tricimba cincta*	3	0
Fannia serena	1	0	*Tvetenia discoloripes*	1	0
Helina depuncta	2	0	*Tvetenia verralli*	1	0
Helina evecta	1	0			

Table 2.8: List of the analyzed species. Number of clusters their sequences were placed in, and the number of other species (not conspecific) within these clusters at 3% clustering. Splitting events in: *B. apsectra, B. bergrothi, C. tauricus, E.chioptera, E. univittata, H. depuncta, O. frit*, and *T. cincta*. Lumping events in: *C. rohdendorfi* and *C. vicina*.

2.3.2.4 Tree-based evaluation

The tree analysis based on the stated criteria in table 2.5 and figure 2.2 resulted in 100% correct identifications. All query sequences were identified according their species membership. No ambiguities and no misidentifications were found in the tree topology. It also revealed the good distinction of the different genera. In addition the NJnetwork showed the conflicting information in the different clades. The conflicts become especially extensive above genus level. While the network confirmed the distinct separation of the different genera it also illustrates the unreadable signal due to the large amount of conflicts beginning on family level.

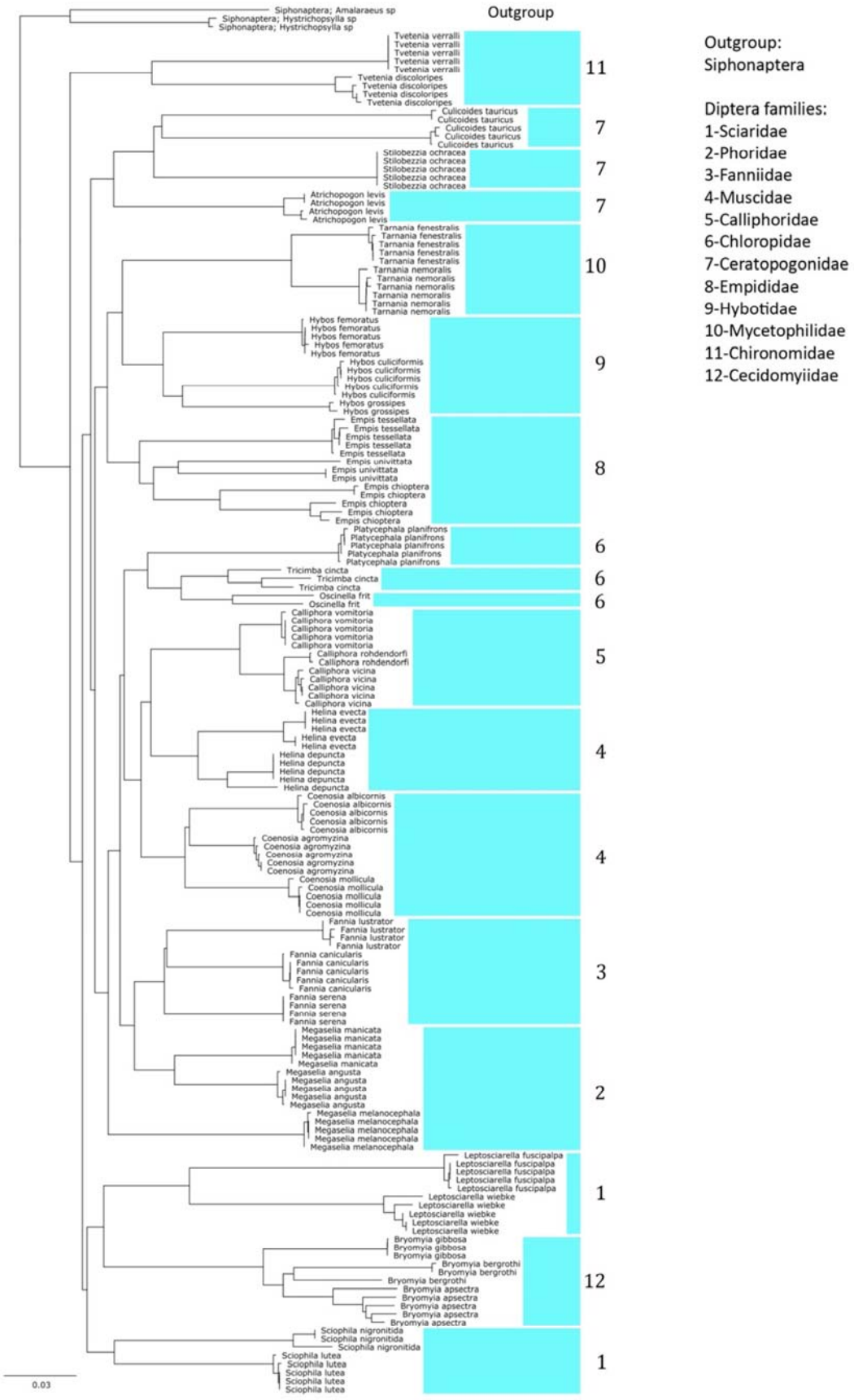


Figure 2.5: Neighbor Joining tree (NJtree) with 166 sequences from 39 species from the 12 most abundant Diptera families in the Exploratories. Two outgroup species were chosen from the order Siphonaptera. Coherent genera are marked as blue blocks whereas families are labelled by numbers. It must be kept in mind that the tree gives no information about deeper phylogenetic relationships. (Alternative NJ network in the supplement - S1.2)

2.3.2.5 Bayesian classifier

The Bayesian classifier assigned 95.2% of the 166 query sequences to the correct species names using the GBOL-reference list of over 48,000 sequences. 4.2% of the queries were ambiguous identifications. Of these ambiguous identifications five sequences belonged to the species *Tvetenia verralli* (Chironomidae) but were identified as *T. bavarica* as these two species share identical COI barcodes. Also two sequences of the species *Empis univittata* (Empididae) were ambiguously identified as this species also shares its barcode with *E. stercorea*. One sequence belonging to *E. univittata* was misidentified as *E. digramma* of whom 99.2% of its barcode are identical to *E. univittata*. This single misidentification made up to 0.6% of all identifications from 166 queries.

2.4 Discussion

2.4.1 Building up a sequence reference list for German Diptera

The gathering of material proved to be laborious as the continuous supply was not guaranteed. This had several reasons ranging from receiving molded specimen that were not suitable for a DNA extraction to the actual lack of material because specimens awaiting treatment were not identified due to missing capacities in taxonomy. Efforts to gain material from external taxonomists that were not involved in the project proved also to be difficult. Until recently collections were rarely stored in high-proof ethanol, as it causes great costs and a fire rist that should not be underestimated. Also taxonomic handling usually concentrated on dried material which is mostly unsuited for DNA work. Beside the actual scarcity of material some taxonomists also expressed their doubts in the method of DNA barcoding while others feared the new method would replace them, making expert knowledge expendable in the future. Especially the last argument seems to be a preconception that still is widespread amongst taxonomists. The truth is that taxonomists are already not able to manage all material that would need identification. And adding to the daily identification work museum collections are filled with innumerable drawers of new species that await species description. DNA barcoding can therefore be seen as a method that complements current research. It will canalize the mass of incoming material to a point that only samples and specimen of interest will need to be manually identified or described.

The incorporation of this project into the German Barcode of Life (GBOL) at the ZFMK proved to be very productive. In several conferences and events that promoted and explained techniques and benefits of DNA Barcoding it was possible to convince taxonomists to participate and contribute to the specimen collection. The supply with storage vials and high-proof ethanol as the appropriate storage liquid also helped to lower costs for the contributors. The database is rapidly growing since then.

2.4.2 Universal application of COI

Already in the first phase of building up the sequence reference it became apparent that the techniques for COI amplification are advanced and universally applicable. Standard primer worked well in every Diptera family they were applied to. Irregularities or amplification failures did not exceed normal laboratory routine. However, if an alternative primer combination is needed several substitutes can be found in the literature (Folmer et al. 1994; Simon et al. 1994; Astrin and Stüben 2008). In cases where the standard primer approach failed C1-N-2191 (Simon et al 1994) and

LCO1490-JJ, HCO2198-JJ (Astrin and Stüben 2008) were regularly used as primer alternatives. The GBOL protocol has now adapted to the permanent usage of the primer combination LCO1490-JJ and HCO2198-JJ as the GBOL project has a more universal approach, covering a wider range of taxa than Diptera alone. Cases for PCR failures while working on Diptera were rather caused by advanced age and adverse storage conditions of the material than the chosen COI primer combination or the individuals' membership to a certain Diptera family.

2.4.3 The capability to discriminate species

166 specimens from 39 previously identified species from 12 of the most abundant Diptera families in the Exploratories were tested for the discriminativity of the taxonomic signal present in their COI barcode sequence. For succesfull identification the genetic species assignment had to be identical with the previous morphological identification. Different methods after Hebert et al. 2003b, Meier et al. 2006, Wang et al 2007, Schloss et al. 2009 and Birch et al. 2017 were applied and evaluated. One specimen of the Genus Amalaraeus was left out from most of the evaluations as the criteria for succesfull identification were not appliable on a singleton species sequence. An exeption was the use of the Bayesian classifier as it used a different larger reference than the dataset itself (see chapter 2.3.2.5). The identification success within all the different methods ranged from 91.3% up to 100% (Figure 2.6).

Applying the criteria for best match (BM) on the single pairwise distance analyses the identification success reached over 99% from 165 tested sequences. Only a single sequence of *Bryomyia bergrothi* was paired incorrectly with *B. apsectra* as its next most similar sequence. However, the genetic distance between these two sequences is with 51 bp differences (7.8%) relatively high. The distinct taxonomic status described in literature also leaves no doubt about their individual species status as it is based on distinct morphological and ecological pronounced characters (Mamaev and Økland 1998, Edwards 1938). Best close match (BCM) threshold criteria account for such incidents and mark sequences not having a match within a 3% distance as unidentified. This largely avoids false positives while still yielding a high identification success. In this case almost 95% of the 165 sequences were correctly identified. However, a fixed 3% threshold can be and has been considered as somewhat problematic as different rates of evolution in different taxa (and climate zones) might result in different genetic species boundaries. Yet it can not at all be called an arbitrary boundary. In literature it is known and discussed as the "barcode gap" (Herbert et al. 2003b; Meyer & Paulay 2005; Meier et al. 2008). An automatic procedure that sorts sequences according a recursively calculated barcode gap has been proposed and proved to yield reliable species predictions (Puillandre et al. 2012; Ratnasingham et al 2013). But the more overlap there is between intra- and interspecific variation the less effective becomes a definite genetic species threshold (Meyer & Paulay 2005).

Clustering suffers from this aspect as species happen to be merged into a single cluster (lumping) or being split into two or more clusters based on the chosen threshold. The clustering evaluation indeed merged two close species within a single cluster as their interspecific distance was below 3% (*Calliphora vicina* and *C. rohdendorfi*: 2.2%). This effect of species merging is known as species "lumping". In contrast almost half of the analyzed species were at the same time also assigned into more than one cluster per species. All these separate clusters distinct from each other

more than 3%. This effect is known as species "splitting". Of these splittings three specimen were clustered as singletons as they did not cluster with a conspecific (*Bryomyia apsectra*, *B. bergrothi* and *Empis univittata*). However, these findings confirm the diverging intraspecific and interspecific distances from the clustering threshold depending on the analyzed genus respectively. Still the identification success for the cluster evaluation (CE) was above 91% for 165 analyzed specimens. Lumping and splitting events can also indicate taxonomic specifics that might have been overlooked until then. Different scenarios can explain the occurrence of lumping or splitting of species.

Sexual dimorphism has lead to the description of species based morphological differences that can cause the genetic lumping of these species later on. Sexual dimorphism occurs in many taxa and happened to account for species synonymy in the past and will probably also still do so in the future (Schlinger 1956; Reiskind 1965; Kraushaar and Blanckenhorn 2002; Poissant et al. 2010). Gender differences can be subtle but can also appear to us as a pure exaggeration of a character. These differences can make it hard to believe that the two specimens might belong to the same species. An impressive example can be found in the group of stalk eyed flies where males with sometimes an extreme large eye span are especially attractive to females (Chapman et al. 2017). These traits might appear arbitrary to us but can in fact be linked to genetic and fertility benefits for females and play a role in male rivalry (Small et al. 2009; Bellamy et al. 2013; Harley et al. 2013). Several of these sexual differences like eye stalks or male body elongation have experienced convergent evolution in Diptera but are not necessarily restricted to only a single sex (Kraushaar and Blanckenhorn 2002; Bonduriansky 2006). Body size is very well investigated and often linked to reproductive success (Darwin 1871; Clutton-Brock and Parker 1992). But increased body size not only leads to increased pairing success concerning male competition or female choice it for example can also lead to increased fecundity due to larger egg size when found in females (Honek 1993). However, taxonomic confusion and species revisions can already be caused by more subtle differences like the structure and insertion of the antennae or slight differences in wing venation (Schlinger 1956). Especially specimens that are quite rare in collections and not often brought together for study are susceptible for these errors. And this is not a problem restricted to specimen from the tropics. Until today Palearctic species descriptions need to be continuously revised as many species have been described from very few specimens and from one sex only (Disney 2006).

Aside from sexual dimorphism is the variance of appearance that can generally be found within a single species (be it morphological, behavioral, physiological or based on its life cycle) a fundamental factor causing confusion and species synonymy across all taxa (Schlichting and Pigliucci 1998; Whitman and Agrawal 2009). Some of the specimens' differences are mere species polymorphisms others are caused by phenotypic plasticity. Polymorphism describes differences of phenotypes of species or populations as a whole whereas phenotypic plasticity describes the different phenotypic responses of one individual genotype to different environmental conditions (West-Eberhard 1989). As a consequence from the fact that environments vary the ability to adapt ensures the survival of a species. Accordingly this form of variance is universal among all living things and can therefore also be found within the taxon of Diptera (Holloway 1993; Leclaire and Brandl 1994; Krüger and Garms 1999;

Chippindale et al. 2004; Kehlmaier et al. 2014). Inaccurate species identification and the failure to recognize polymorphism, phenotypic plasticity and sexual dimorphism can have severe consequences. It can hamper basic research, disease diagnosis and medical and agricultural pest control (Whitman and Agrawal 2009). The lumping of species in genetic analyses can be the first indication for these inaccuracies and can sometimes reveal unexpected species synonymy where two species are actually one (Krüger and Garms 1999; Kehlmaier et al. 2014).

On the opposite side there is "cryptic diversity". Cryptic diversity describes the incidence of two or more distinct species being initially classified as a single species due to similarities in their appearance (Bickford et al. 2007; Pfenninger and Schwenk 2007; Trontelj and Fišer 2009). Hence these species are often being split later on in clustering analyses due to the genetic distance between the species. Cryptic species is neither rare nor should it be trivialized. It is assumed to constitute at least 8–25% of all insect species (Smith et al. 2006). Of course this also affects many Diptera species (Cornel et al. 1996; Scheffer 2000; Selivon et al. 2005; Smith et al 2007; Hernández-Ortiz et al. 2012; Krosch et al. 2013; Thapa et al. 2017). A major focus is put on the unraveling of these species-complexes with several novel methods that have recently been introduced (Hausdorf et al. 2010; O'Meara 2010; Yang et al. 2010; Ence et al. 2011; Espíndola et al. 2016). Delimiting and identifying independent lineages brings valuable insights beyond the field of taxonomy which can affect agricultural, economical or ecological interests. When for example species assumed to be generalists reveal not to be generalists at all but in contrast highly specific cryptic species, species awareness and conservations efforts need to be revised (Smith et al. 2006; Costello et al. 2013; Canal et al. 2015). However, the identification of cryptic species remains challenging. The most promising are methods that integrate genetical, morphological and ecological data to clarify the taxonomic status of species in doubt (Selivon et al. 2005; Bickford et al. 2007; Hajibabaei et al. 2007; Rajaei Sh et al. 2013; Krosch et al. 2013; Contreras Gutiérrez et al. 2014). DNA barcoding plays a fundamental part in this process as it can complement research by providing background information that will help in the selection of species complexes for further analyses.

The tree based evaluation (TBE) yielded 100% of correct identifications of the 165 specimens. The computational demanding ability to simultaneously evaluate all sequence distance information for the specimens clustering instead of using the single pairwise distances of sequence pairs might explain the slightly better performance of the tree building evaluation. However, studies using simulated data also indicated that distance- and tree-based barcoding methods usually achieve similar levels of identification success (Ross et al. 2008; Austerlitz et al. 2009; Birch et al. 2017). More complex models of nucleotide evolution for the correct placement of individuals in a tree can be incorporated when analyzing more complex data although this again comes at a cost of being more time intensive (Astrin et al. 2012). Still the little additional resolution usually achieved by this does not reflect the additional computation time typically required when choosing tree-based analyses over distance based approaches (Birch et al. 2017). Whenever computing power is limited it is suggested to refrain from its application especially in research fields where massive parallel sequencing and high throughput workflows concentrate on rapid biodiversity assessments.

However, this is exactly the field where the Bayesian classifier can demonstrate its strengths (Wang et al. 2007, Schloss et al. 2009). Studies have shown that it is both

fast and accurate in providing taxonomic assignments for larger datasets (Liu et al. 2008, 2012; Porter and Golding 2012; Porter et al. 2014). In contrast to the previous methods the identification based on the GBOL-reference, a large dataset containing over 48,000 sequences from over 10,000 species created from the GBOL database comprising at that time also the initial Diptera sequence list. Although the dataset as a whole was multiple times increased by this, the identification completed in only 45 seconds using a standard desktop computer. At the sime time it also performed significantly well with over 95% of correct identifications. Furthermore 4% of the miss-assignments were actually species with an identical sequence to the query sequence. This circumstance of indistinguishability can therefore not be rated as a software error. Considering this the Bayesian classifier performed exceedingly well but for species determination this barcode overlap creates a problem. Although the amount of species sharing identical barcodes is seen as rather moderate it remains problematic whenever it occurs (Meier et al. 2006). In such cases identification can only be narrowed down to a complex of species sharing the same barcode when species synonymy can be excluded. To refine resolution in these complexes further data needs to be accessed. These could be morphological characters if the specimens are available but also a further genetic marker (Dowton et al. 2001; Quan et al. 2001; Elbrecht et al. 2016).

Unfortunately mitochondrial markers are especially susceptible for infections that lead to genetically indistinguishable species. Infections with maternally inherited symbionts can have direct influence on reducing the diversity of mtDNA and lead to identical mtDNA sequences among different species (Hurst and Jiggins, 2005). Especially the intracellular bacterium Wolbachia seems geographically and phylogenetically broadly distributed and causes problems also within different insect taxa (Werren and Windsor, 2000; Hurst and Jiggins, 2005). Despite its universal distribution, an examination on the Barcode of Life Datasystem (BOLD) found in only 0.16% of the cases an evidence for the presence of Wolbachia in COI data (Smith et al. 2012). It is therefore considered unlikely to compromise the accuracy of the DNA barcode library. However its presence has to be noted and considered. If species identification remains unclear or is in doubt further data must be consulted if available. Sometimes ecological information can lead to the exclusion of the unlikely species. And often are the morphological differences distinctly pronounced and can clarify the situation more easily. Other approaches use a second genetic marker when the species assignment remains uncertain. However, erroneous identifications can also always hint to taxonomic discrepancies that need further attention.

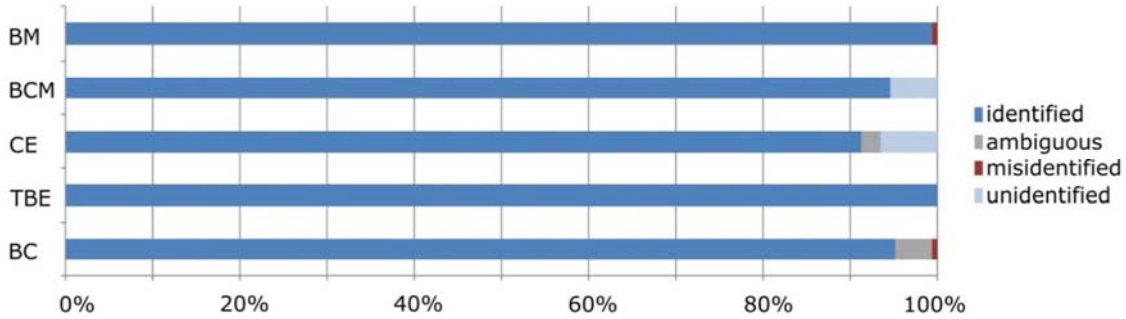


Figure 2.6: Identification success within 38 species based COI for: Best match (BM), 99.4% of 165 specimens; best close match (BCM), 94.6% of 165 specimens; cluster evaluation (CE), 91.3% of 165 specimens; tree based evaluation (TBE), 100% of 165 specimens; Bayesian classifier (BC), 95.2% of 166 specimens.

2.4.4 Conclusion

Despite the discussed discrepancies in particular species assignments the results for testing 166 specimens covering 38 species from 12 Diptera families are encouraging for the further usage of COI barcodes as species identifiers. Considering that only a single genetic marker was used an average identification of 96% is probably "as good as it gets" even under perfect conditions. However, barcoding is not meant to replace traditional taxonomy. Experts are overwhelmed by the material they have to examine. Studies taking samples have always produced masses of material from which a lot is never being evaluated. But until today large batches of insect samples are being weighted to enable their evaluation (Hallmann et al. 2017). Single identification is not realistic anymore as it too expensive and inefficient. Future tests for barcode efficiency and identification methods should therefore concentrate on these new sampling sizes, where hundreds of species need to be differentiated and identified. Still it must be admitted that whole sample barcoding already can definetly yield more detailed results than any weighting scale. It can further help to structure sample evaluation and give priority inidications for taxonomists. It complements them particularly in speed and sample coverage. Barcoding can be extremely efficient when workflows and automations are appropriately adjusted. New sequencing techniques allow evaluating multiple samples at once. Identification methods have to keep up with the vast amount of data being produced. The Bayesian classifier does not only compete with current methods, it also proved to be highly valuable in context of future high troughput approaches as its rapid identifications are at the same time very reliable. The identification algorithm is even significantly faster than current blast-based methods that are commonly used in environmental sequence surveys (Porter et al. 2014). Valuable results for high-throughput samples and biomonitoring studies can be expected using a combination of COI for DNA barcoding, a comprehensive and well maintained species reference and the Bayesian classifier.

3. Test of conservation liquids for traps

3.1 Introduction

In a cooperation study with the TU Munich sampling liquids used during the capturing process were tested for their influence on amplification success and sequence quality. The different liquids were considered as possible alternatives to the very effective but expensive and volatile high proof ethanol (Baird et al. 2011, Stein et al. 2013). The research article was published 2016 in PLOS ONE and is being summarized here. It can be found in the appendix for more detailed information (Gossner et al. 2016).

3.2 Material and Methods

3.2.1 Experimental setup

Material for the liquid testing was collected from the cooperation partner in the Wippenhauser Forest, North Freising, Germany (48.414°– 48.421°N / 11.714° - 11.732° E. Sampling areas were dominated by either beech (Fagus sylvatica) or spruce (*Picea abies*). The minimum distance of the plots to each other was 50 m. The maximum distance was 1400 m. A randomized block design was applied, with two different tree species, three sampling solutions, two trap types (canopy and understory) and two collection jars per trap (top and bottom jar) (Figure 3.1). Ten repeats lead to a total of 60 trees (30 of each species) with 120 traps and 240 collection jars. Each experimental block consisted of one forest plot with one tree for each of the three sampling solutions while only choosing trees of similar size. The trees were located at least 5 m, but no more than 10 m, from each other.

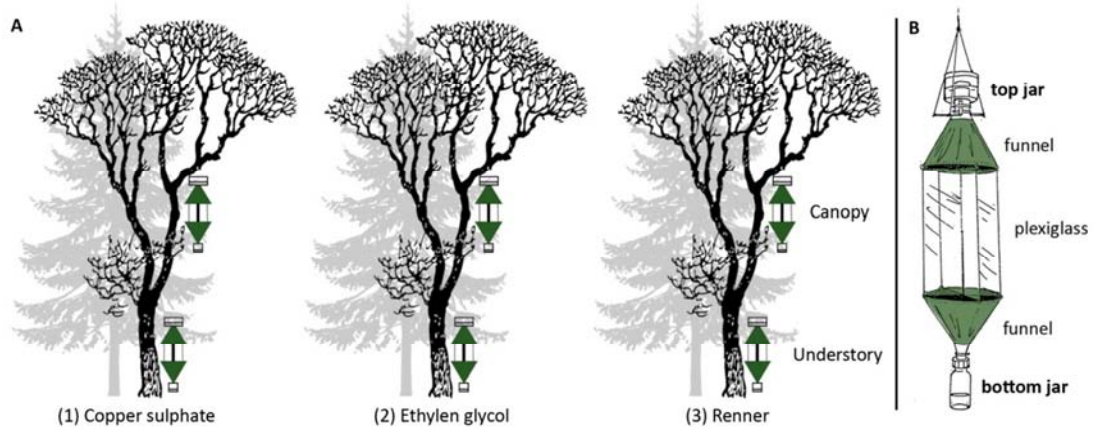


Figure 3.1: Graphic of the experimental setup. A: In each of the plots one tree was used for each sample solution. Two traps were installed in each tree (canopy and understory), each with two jars (bottom and top). Distance between trees within each plot was five to ten meters. Ten plots were used per tree species, beech and spruce. B: Construction of the flight interception trap used. (modified after: Gossner et al. 2014 and 2016)

The traps were flight interception traps made of crossed pairs of 40 cm x 60 cm sized transparent plastic shields with funnels and sampling jars attached at the top and the bottom of the trap (Figure 3.1). Three common sampling liquids (1) copper sulphate (CuSO4 3%), (2) ethylene glycol (50%) and (3) Renner solution: ethanol (40%)/glycerin (25%)/water (35%) excl. acetic acid, were chosen due to their effective

and persisting use in long-term, largescale biodiversity experiments (CuSO4: Röder et al. 2010, Gossner et al. 2014; ethylene glycol: Kovivula et al. 2003, Schmidt et al. 2006; Renner: Renner 1982, Dorow et al 2010; Sereda et al. 2014). Canopy traps were installed hanging them on suitable branches in the center of the tree crown (height: beech: 15.1±1.0; spruce: 18.5±0.7). The understory trap was installed next to the tree trunk (height 1.5 m). The traps were installed in early May and emptied twice, once in late-May (three weeks after installation) and again at the end of June (seven weeks after installation). The sampling solution was replaced after the first collection in all jars and all samples were immediately transferred to 70% ethanol in the field.

3.2.2 Arthropod identification and classification

All samples were sorted to arthropod order level in the laboratory in Munich. Subsequently, all Coleoptera and Hemiptera were identified to species level by specialists. Beetles were classified based on their feeding ecology and habitat requirements. All sampled species were classified into different feeding guilds (herbivores excl. xylophages, carnivores, mycetophages-fungi, mycetophages-mould, decomposers-wood, decomposers excl. wood), habitat guilds (ground dweller, eurytopic, vegetation, rotten substrate/nests/fungi-excl. wood) according to Köhler 2000, and more specific dead wood substrate guilds: old dead wood (od-dweller), fresh dead wood (fd-dweller), wood mould and specific dead wood structures (rh- and sdweller), wood fungi (fu-dweller) according to Schmidl and Bußler 2004. Among saproxylics the feeding guilds mycetophages, xylophages and carnivores were distinguished. Body size was measured as a functionally meaningful trait (Fountain-Jones et al. 2014).

3.2.3 Measure of quality for morphological species determination

During sorting, all samples from the June collection were classified according to the conditions of the insects with respect to mould and completeness of the insects (Table 3.1). This was used as a measure of quality for morphological species determination. Values ranged from 0.75 (excellent condition, no mold and insects all complete) to 3.25 (totally moldy and insects largely fragmented) in steps of 0.25.

Value	Mold	Insect completeness
0.75	no mould at all and all insects complete	
1	small traces of mould, but specimens without mould	single body parts missing in less than 5% of the specimens
1.5	medium traces of mould, but specimens without mould	body parts are missing from more than 10% of the specimens
2	bigger traces of mould, but specimens without mould	body parts are missing from more than 50% of the specimens
2.5	bigger traces of mould and specimens slightly covered by mould	body parts are missing from more than 90% of the specimens
3	bigger traces of mould and specimens conspicuously covered by mould	extremities are missing from all specimens
3.25	insects in a plug of mould and largely fragmented	

Table 3.1: Overview on the criteria that were used to classify the samples in terms of quality for subsequent morphological species identification.

3.2.4 Species identification through DNA barcoding

Ten species were chosen across the different sampling solutions (Table 3.2). Their COI sequences was analyzed and evaluated at the ZFMK in Bonn.

Order	Family	Genus	Species	Det.	N
Coleoptera	Leiodidae	*Agathidium*	*seminulum*	(Linnaeus 1758)	15
Coleoptera	Elateridae	*Athous*	*subfuscus*	(O. F. Muller 1764)	15
Coleoptera	Latridiidae	*Corticaria*	*abietorum*	(Herbst 1783)	15
Coleoptera	Latridiidae	*Corticarina*	*lambiana*	Sharp 1910	15
Coleoptera	Nitidulidae	*Cychramu*	*variegatus*	(Herbst 1792)	15
Coleoptera,	Cantharidae	*Metacantharis*	*discoidea*	(Ahrens 1812)	15
Coleoptera	Staphylinidae	*Plectophloeus*	*fischeri*	(Aube 1833)	15
Coleoptera	Curculionidae	*Rhynchaenus*	*fagi*	(Linnaeus 1758)	15
Coleoptera	Throscidae	*Trixagus*	*dermestoides*	(Linnaeus 1766)	15
Hemiptera	Miridae	*Psallus*	*varians*	(Herrich-Schäffer 1841)	15

Table 3.2: List of ten species found in all sampling solutions, to analyze their barcoding feasibility.

3.2.4.1 Sample preparation

For each of the three sampling solutions five individuals from each species were sampled. To minimize microclimatic bias all 15 specimens of one species were taken from one forest type and one vertical stratum. Effects of microclimatic differences on DNA quantity and quality could not be tested because of an insufficient number of specimens covering stratum and forest type. Individuals were washed with 70% ethanol with the remaining ethanol allowed to evaporate. The whole insect was used for the DNA extraction, except for *Athous subfuscus*, *Cychramus variegatus* and *Metacantharis discoidea* where leg material was used because of the specimens' size. Sample material was weighted and recorded. DNA was extracted using the Salting-Out method (Sunnucks and Hales 1996).

Insect tissue was homogenized in 300 µl TNES buffer (50 mM Tris-HCl pH8, 20 mM EDTA, 400 mM NaCl, 0.5% SDS) with 5 µl of Proteinase K (20 mg/µl) and incubated overnight in a water bath at 37°C. Then 85 µl of 5MNaCl was added, vortexed and the sample centrifuged at 13,000 rpm for 10 min. The supernatant was placed into a new tube and 400 µl of 100% ethanol were added. The sample was kept at -20°C for a minimum of one hour for DNA precipitation. After centrifugation at 13,000 rpm for 20 min the DNA pellet was washed twice in 70% ethanol. Before the resuspension in TE buffer (10mM Tris- HCl (pH 8.0), 1 mM EDTA) the remaining ethanol was allowed to evaporate. Samples were stored at -20°C. The DNA concentration (ng/µl) was measured using a Quantus Fluorometer (Promega) to calculate DNA yield (ng/mg).

DNA integrity was determined at the ZFMK in Bonn, using a Fragment Analyzer (Advanced Analytical Technologies, Inc., Ames, IA) with its current software version 1.0.2. Samples were prepared following the manufacturers specifications for the use of the high sensitivity kit (DNF-488 High Sensitivity Genomic DNA Analysis Kit). Whenever needed the DNA concentration was diluted to a maximum of 5 ng/µl to level sample concentrations. Smear analyses were conducted using the provided software PROSize 2.0 Software Version 1.3 to quantify the proportion of higher genomic DNA (1,000 bp – 20,000 bp) to the rest of the sample and to identify the average size of the measured sequence lengths. The 1,000 bp cutoff was chosen to include intact DNA as well as partially degraded DNA into the measurement (AATI gDNA Anlysis, 2012). Partially degraded DNA still maintains potential to lead to successful identifications due to the limited length of COI and the abundant occurrence of mtDNA in the cells. Furthermore 1,000 bp is also the maximum threshold to which irrelevant RNA might be present in the extraction (User Guide DNF-488 High Sensivity Genomic DNA Analysis Kit 2014MAR11).

A 658 bp fragment of the COI gene was amplified using the universal primers LCO1490 (Folmer et al. 1994) and C1-N-2191 (Simon et al. 1994). PCR amplification was performed again in Munich, using 20 µl volume, with 1.5 µl template DNA, 1 U Bioline MyTaq DNA polymerase, 4 µl 5X Bioline MyTaq Reaction Buffer, 0.5 µMprimer F, 0.5 µMprimer R. A touchdown PCR (53–48°C) was performed on a thermocycler (BIOER Lifetouch ™): 95°C 5 min, followed by 10 cycles of 94°C 15 s, 53–48°C 30 s (-0.5°C per cycle), 72°C 30 s, then 25 cycles of 94°C 15 s, 48°C 30 s, 72°C 30 s, finally ending with 72°C for 6 min and then cooled to 4°C. The results were visualised on a 1.2% agarose gel using DNA Stain G (SERVA, Germany) and visualised using a gel documentation system (Intas Gel-Stick, Royal Biotech, Germany). An 8 µl aliquot of the PCR product was then cleaned using ExoSap (0.1 U FastAP, 0.4 U Exonuclease I), and incubated at 37°C for 30 min followed by 80°C for 15min in a thermocycler. Cleaned products were sequenced by Macrogen. Samples that failed to produce a PCR band or sequencing product were repeated.

3.2.4.2 Sequence processing, quality evaluation and identification

Sequencing results were processed at the ZFMK, Bonn. All sequence reads were fully processed in Geneious v7.1.9 (Kearse et al. 2012). Raw reads were assembled (de novo assemble) using the geneious assembler on highest sensitivity default mode (Allow Gaps: true; Word length: 10; Index word length: 10; Maximum mismatches per reads: 50%; Maximum ambiguity: 16; Maximum gap size: 5). Read directions (forward and reverse) were checked and corrected if necessary. Primer sections were trimmed. From each of the assembled reads consensus sequences were calculated, setting the threshold for matching bases at 100% identical, allowing ambiguities according to the IUPAC Ambiguity Code to encode for ambiguous positions (IUPAC 1997). The resulting sequences were assigned into the categories High, Medium and Low for an overall sequence quality rating. These categories resulted from the evaluation of further quality thresholds. A sequence of the category *High* is allowed no ambiguities and a minimum length of 300 bases. 90% of the bases need to have a phred score of 40. Only 10% of the bases are allowed to have a phred score below 20. For *Medium* classification a maximum of 5 ambiguities is accepted. The minimum sequence length is again 300 bases. 80% of the bases need to have a phred score 40 and 15% are allowed to have a phred score below 20. Any sequence that did not fit these criteria was assigned *Low*.

Sequences were then cross-checked for matches in BOLD and NCBI databases. In BOLD the offered BOLD Identification System (IDS) for animal identification was used (Ratnasingham and Hebert 2007). In NCBI the query sequences were analyzed using BLASTN 2.2.31+ (Zhang et al. 2000). The data (available matches and their corresponding values: similarity (BOLD) and Max score and Identity (NCBI)) were then transferred into a table.

3.2.4.3 Data analysis

After sequence identification and quality evaluation data was send to TU Munich for statistical analysis All analyses were done in R (version 3.2.0; R Core Development Team 2015) using R studio (version 0.98.977). Eight response variables were analyzed: (1) sample condition, (2) order richness, (3) order diversity, (4) order abundance, (5) Coleoptera species richness, (6) Coleoptera species diversity, (7) Hemiptera: Heteroptera species richness, and (8) Hemiptera: Heteroptera species diversity. For the

order abundance data, the nine orders with greater than 500 individuals collected across all treatments (99.0% of the data) were analyzed. The diversity was calculated as the exponential Shannon diversity using the vegan package (Oksanen et al. 2013). Linear mixed effects models, using R package nlme (Pinheiro et al. 2015) were used to determine the effect of the fixed factors: tree species, sampling solution, trap type (canopy or understory) and collection jar (top or bottom), and the random effects used were plot, sampling solution, trap type and jar to account for the hierarchical structure of the data set. For the order abundance model, an additional fixed effect of 'order' was used in the model. Full models were fit first, including all interactions, and then each model was simplified by removing the most non-significant term first, using a backwards selection procedure. Post-hoc contrasts from the models in R are presented to show differences among levels within a factor. The pH of copper sulphate varied between the top and bottom jars, thus a further model to determine if the effects on sample condition were dependent on the changing pH was run on the copper sulphate data. Reduced (minimum adequate models) are presented in the results. For a more detailed analysis of changes in community composition a RLQ analysis was used (Doledec et al. 1996) for an ordination of species, species traits and sites on the main environmental gradients (package ade4 in R; Dray and Dufour 2007). A fourth-corner analysis was applied (Dray et al. 2014; Dray and Legendre 2008) as a statistical test of the relation of the biological traits and the environmental variables through the link of a community data table. In the RLQ analyses the relationships between species traits (Q) and environmental variability (R) are revealed by maximizing the congruency between three data tables: Beetle abundance data (L-matrix), traits data (Q-matrix), and environmental data (R-matrix). The genetic barcoding data were analyzed using general linear models. The number of successful identifications was analyzed using a generalized linear model with quasibinomial error distribution. The response variable was identification success (0,1) and the explanatory variables were species and sampling solution (and the interaction). To analyze DNA yield and DNA quality (the average fragment length (bp) and the concentration of DNA above 1000bp (ng/µl)) variation as response variables a linear models with normal error distributions was applied. The DNA yield data was log-transformed to achieve normal errors, and the explanatory variables were species and sampling solution (and the interaction) for each response variable. A second model, using the available pH and sample condition data (N = 109, from 150 total), was analyzed for each response variable. The variables of DNA yield and quality were correlated to assess the influence of these variables on the barcoding success. For this generalized linear mixed effects models (binomial error) with species and solution as a random factor were applied. Full models were fit first, including all interactions and then each model was simplified by removing the most non-significant term first, using a backwards selection procedure.

3.3 Results

The traps collected 76,588 individuals from 29 orders; 5,938 additional holometabolic larvae were excluded from further analyses. Nine orders contained more than 500 individuals (Acari, Araneae, Coleoptera, Collembola, Diptera, Hemiptera, Hymenoptera, Psocoptera, Thysanoptera; Table A in S4 File in: Gossner et al. 2016).

	Order richness			Order diversity			Order abundance		
	df	**F**	**P**	**df**	**F**	**P**	**df**	**F**	**P**
Source		na			na		8.19	12.08	**<0.001**
tree species	1.18	0.22	0.647	1.18	3.01	0.100	1.18	0.99	0.333
sampling solution	2.38	0.93	0.404	2.38	2.38	0.107	2.38	0.94	0.401
trap type	1.58	16.86	**<0.001**	1.58	18.00	**<0.001**	1.59	0.20	0.658
jar	1.12	88.43	**<0.001**	1.12	33.55	**<0.001**	1.12	18.91	**<0.001**
tree species x trap type	1.58	4.50	**0.038**	1.58	4.20	**0.045**	-	-	-
trap type x jar	1.12	-	-	1.12	10.18	**0.002**	-	-	-
order x tree species		na			na		8.19	2.92	**0.003**
order x trap type		na			na		8.19	3.47	**<0.001**
order x jar		na			na		8.19	4.15	**<0.001**

Table 3.3: Summary of the statidtical results for order richness, diversity and abundance. df-Value: df stands for degrees of freedom, the number of values which are free to vary. F-Value: Values close to 1 show, that the source does not affect the residual variance (null hypothesis is true). Large values indicate a reduction in residual variance that can be attributed to the source (rejection of the null hypothesis). P-Value: It is based on the sampling distribution of F-values under the null hypothesis (source is uncorrelated to the residual variance) and indicates the probabilities of getting F-values larger than the one obtained. —: the source was not retained in the minimum adequate model. All interactions (including the 4-way between all main effects) were tested; the ones not shown were not retained in the minimum adequate model. Significant (P<0.05) sources are highlighted in bold. Order abundance tests the nine orders with at least 500 individuals (99.0% of total). na: the source was not included in the model. (Residual variance is the variance that is not explainable by the experimental variation. It is the variance that needs to be explained)

3.3.1 Condition of the samples

The condition of the samples was found to be influenced by a number of treatment factors, including a 3-way interaction between tree species, sampling solution and trap type ($F_{2.54}$ = 4.69, P = 0.013). Here, the copper sulphate sampling solution samples were observed to be in a worse condition than samples from the other solutions, and this is most apparent in the top jars of traps–particularly in canopy traps on beech and understory traps on spruce (Fig 3.2, Table A in S5 File in: Gossner et al. 2016). The pH of the copper sulphate samples was lower in the top than the bottom jars of the traps ($F_{1.39}$ = 926.4, P<0.001), likely due to dilution of the bottom jars from rain (Fig A in S5 File in: Gossner et al. 2016); the pH of the other solutions was constant across the jars. However, neither the pH ($F_{1.38}$ = 1.75, P = 0.193) nor the jar ($F_{1.38}$ = 1.24, P = 0.272) affected the condition of the samples and as such the effect of sampling solution on sample condition was not linked to the pH.

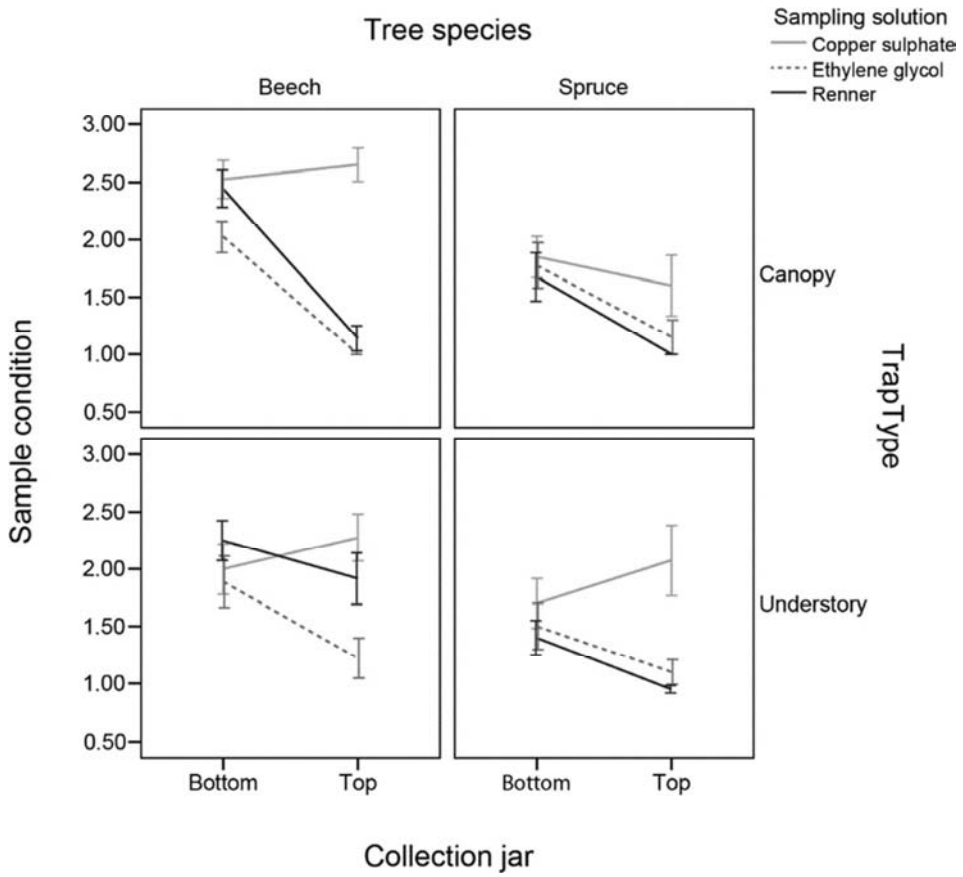


Figure 3.2: The condition of the samples across the different treatments. Higher values indicate worse condition. Error bars represent ±1SE (Gossner et al. 2016).

3.3.2 Order level

There was no effect of sampling solution on richness, diversity or community composition at the order level (richness: $F_{2.38} = 0.93$, $P = 0.404$; diversity: $F_{2.38} = 2.38$, $P = 0.107$; composition: $R = -0.001$, $P > 0.9$). However, the understory traps had higher order richness and diversity than the canopy traps (Table 3.3). Order composition also differed between trap type ($R = 0.110$, $P < 0.001$), as understory traps collected more Acari, Isopoda and Orthoptera, whereas the canopy traps collected more Diptera, Mecoptera and Megaloptera. The effect of trap type also differed dependent on the tree species (Tree species x trap type interaction; Table 3.3); there was a greater difference in order richness and diversity between the understory and canopy traps for spruce than for beech. In general, the bottom jar collected more individuals than the top (Table 3.3) but a significant interaction between trap type and jar (Table 3.3) showed that the difference was larger in the understory than the canopy traps. The abundance of nine orders, in which more than 500 individuals were collected in total, was analyzed with respect to the treatments (Table 3.3). The sampling solution did not alter the abundance of these orders. All other fixed effects were significant as interaction effects (Table 3.3). The abundance of the different orders changed dependent on: the tree species, with spruce having more Diptera (post-hoc test: $t = 2.49$, $P = 0.017$) and beech more Thysanoptera ($t = 2.36$, $P = 0.018$); the trap type, with canopy traps collecting more Diptera ($t = 4.74$, $P < 0.001$); and, the jar, with the bottom jar collecting more individuals overall, but specifically more Diptera ($t = 2.49$, $P = 0.017$).

3.3.3 Species level analyses

3.3.3.1 Species richness and diversity

5,432 individuals of Coleoptera were collected (Table A in S4 File in: Gossner et al. 2016), of which 5,278 were identified to 326 species. The most abundant species were *Xyleborus germanus* (651 individuals), *Athous subfuscus* (611 individuals), *Eusphalerum sorbi* (505 individuals) and *Rhynchaenus fagi* (368 individuals). The samples from spruce were more species rich, with 243 species identified compared to 182 from beech. The species richness of the samples differed among the treatment factors with a three-way interaction between tree species, sampling solution and jar (Table 3.4); this was primarily driven by an increase in species richness in the bottom jars of understory traps containing the Renner sampling solution, to a greater extent in spruce than beech (Fig A in S6 File in: Gossner et al. 2016). However, this 3-way interaction term was not significant for the species diversity, where sampling solution was not present in any significant terms within the final model (Table 3.4).

3,468 individuals from the order Hemiptera were collected, 164 individuals from the sub-order Heteroptera were identified into 32 species. Due to the lack of Heteroptera in the top jars, the data was combined from the bottom and top jars of each trap. The sampling solution influenced Heteroptera species richness ($F_{2,27}$ = 3.33, P = 0.051) and diversity ($F_{2,27}$ = 3.66, P = 0.039), with the Renner solution yielding fewer species and a reduced diversity than the other two solutions. The species richness and diversity of the samples did not differ among the tree species (species richness $F_{1,17}$ = 0.019, P = 0.892; diversity $F_{1,17}$ = 0.09, P = 0.771), or trap type (species richness: $F_{1,25}$ = 0.56, P = 0.462; species diversity $F_{1,25}$ = 1.21, P = 0.281). We found no interactions between the treatment factors.

	Species richness			Species diversity		
Source	**df**	**F**	**P**	**df**	**F**	**P**
tree species	1.18	12.29	**0.003**	1.18	17.59	**<0.001**
sampling solution	2.36	1.31	0.284	2.38	1.23	0.303
trap type	1.56	15.65	**<0.001**	1.58	6.22	**0.016**
Jar	1.86	753.67	**<0.001**	1.91	460.56	**<0.001**
tree species x solution	2.36	0.71	**0.498**	-	-	-
tree species x trap type	1.56	9.14	**0.004**	1.58	8.61	**0.005**
solution x trap type	2.56	0.43	0.650	-	-	-
tree species x jar	1.86	6.83	**0.011**	1.91	8.19	**0.005**
solution x jar	2.86	3.07	0.051	-	-	-
trap type x jar	1.86	0.47	0.495	-	-	-
tree species x solution x jar	2.86	3.11	**0.050**	-	-	-

Table 3.4: Summary of species richness and Shannon diversity results for Coleoptera species. df-Value: df stands for degrees of freedom, the number of values which are free to vary. F-Value: Values close to 1 show, that the source does not affect the residual variance (null hypothesis is true). Large values indicate a reduction in residual variance that can be attributed to the source (rejection of the null hypothesis). P-Value: It is based on the sampling distribution of F-values under the null hypothesis (source is uncorrelated to the residual variance) and indicates the probabilities of getting F-values larger than the one obtained. —: the source was not retained in the minimum adequate model. All interactions (including the 4-way between all main effects) were tested; the ones not shown were not retained in the minimum adequate model. Significant (P<0.05) sources are highlighted in bold. Order abundance tests the nine orders with at least 500 individuals (99.0% of total). na: the source was not included in the model. (Residual variance is the variance that is not explainable by the experimental variation. It is the variance that needs to be explained)

3.3.3.2 Community composition

Due to the low number of adult Hemiptera in the samples the analyses of community composition concentrated on Coleoptera. The changes in the community composition are illustrated by RLQ analysis (Figures in: Gossner et al. 2016). The first axis of the RLQ

analysis explained 79.4% and the second axis 16.3% of the total variation. The RLQ analysis captured 92.7% and 81.5% of the total inertia of the R–L and the Q–L analyses on the first RLQ axis indicating that the environment–species relationship and the trait–species relationships are both very close in the dataset. Sampling solution had a major influence on the composition of the beetle assemblages along the first RLQ axis. The high explanatory value of the second RLQ axis on the other hand is due to the high significance of the factor "forest type".

The correlation between beetle traits and the RLQ axes showed that feeding guild (Correlation Ratios CR; axis 1: 0.33, axis 2: 0.16) and habitat preference (CR axis 1: 0.54, axis 2: 0.06) but not body size (CR axis 1: 0.06, axis 2: 0.06) were well represented along the gradient of the two axes. Fresh dead wood dwellers, among those mainly mycetophagous ambrosia beetles, were pronounced in traps with Renner solution (Table A in S6 File in: Gossner et al. 2016). While wood-decomposers were more pronounced in spruce forests, other decomposers and herbivores were more important in beech forests (Fig 4). In a fourth corner analyses, however, only the relationship between habitat guild and sampling solution was significant (Table A in S6 File in: Gossner et al. 2016). A complete list of all sampled species is given in Table A in S7 File in: Gossner et al. 2016.

3.3.3.3 Species identification through DNA barcoding

65 individuals (of 150) were successfully identified using genetic barcoding. Three species did not produce any successful barcode sequences (*A. subfuscus*, *C. lambiana* and *P. fischeri*), whereas almost all individuals in another three species were correctly identified from the barcoding sequencing (*C. abietorum*, *C. variegatus* and *R. fagi*) (Figure 3.4). The number of successful identifications was therefore strongly species-specific (GLM binomial: $F_{9.15} = 23.65$, $P<0.001$) and was also significantly influenced by sampling solution ($F_{2.15} = 9.41$, df = 2, $P<0.001$), with copper sulphate samples producing fewer successful identification than ethylene glycol ($t = 2.87$, $P = 0.004$) or Renner ($t = 3.88$, $P<0.001$) (Figure 3.4). There was no significant interaction between species and solution on barcoding success.

DNA yield varied across the species (average 15.30–634.99 ng/mg; $F_{9.12} = 15.29$, $P<0.001$) and was again dependent on the sampling solution (interaction: $F_{18.12} = 1.71$, $P = 0.045$; Table 3.5). DNA fragmentation was found to also explain the variation in obtaining a successful sequence across the different species and solutions. The average length of the DNA and the concentration of DNA fragments above 1000bp correlated ($r = 0.666$, $P<0.001$). Both were influenced by a significant species and solution interaction (fragment length: $F_{18.12} = 3.12$, $P<0.001$; concentration above 1000bp: $F_{18.12} = 2.23$, $P = 0.005$) (Fig 3.4). There was no evidence that higher DNA yield, in general, produced more successful sequence identifications (GLM binomial: $X^2 = 2.46$, df = 1, $P = 0.117$), but samples with larger average fragment length and higher concentration of good DNA (above 1000bp) did lead to higher sequencing success (GLM binomial: length $X^2 = 6.48$, df = 1, $P = 0.011$; concentration above 1000bp $X^2 = 14.12$, df = 1, $P<0.001$). In particular, *P. fischeri* samples produced very low DNA yield (15.30 ± 4.5 ng/mg; Table 3.5) (posthoc: $t = -2.27$, $P = 0.025$) and, while the average length was around 1000bp the extremely low concentration of DNA led to no successful barcoding sequences being obtained from this species. However, many species-solution combinations had a lower average fragment length than the required 658bp, yet sequences for their identification

were successfully obtained. This indicates that although the success rate increased with DNA quality not all variation could be explained by this. It showed that samples with at least 3.1 ng/µl of DNA above 1000bp would lead to an 80% success rate of sequence identification (Fig 3.3).

There was little association between sample condition and DNA yield ($F_{1.71}$ = 0.90, P = 0.345) or average fragment length ($F_{1.71}$ = 2.05, P = 0.157). But, lower DNA yield was to some extent associated with a lower pH of the solution, i.e. in copper sulphate samples ($F_{1.80}$ = 3.19, P = 0.078), although the average fragment length was not affected ($F_{1.80}$ = 0.67, P = 0.414). When copper sulphate was used the obtained reads were also shorter on average (630 bp) than when ethylene glycol (650 bp) or Renner (646 bp) was used. The quality of the DNA sequences obtained was categorized as 'high' for 77%, 75%, and 75% of samples from copper sulphate, ethylene glycol and Renner, respectively. Thus, while the read length and number of successful identifications was lower for copper sulphate samples, for those sequences that were successfully obtained sequence quality was similar to the other sampling solutions.

	DNA yield (ng/mg)		
Species	**Copper sulphate**	**Ethylene glycole**	**Renner**
Agathidium seminulum	137.82 ± 66.91	937.89 ± 373.45	170.68 ± 77.25
Athous subfuscus	595.96 ± 67.49	375.45 ± 108.94	155.59 ± 48.70
Corticaria abietorum	113.72 ± 66.38	223.00 ± 114.23	25.48 ± 8.20
Corticarina lambiana	124.68 ± 32.73	212.78 ± 84.94	171.48 ± 97.11
Cychramus variegatus	263.77 ± 71.33	241.45 ± 16.36	280.17 ± 50.13
Metacantharis discoidea	453.40 ± 110.21	565.18 ± 213.51	541.51 ± 167.62
Plectophloeus fischeri	18.43 ± 9.29	10.50 ± 0.55	16.97 ± 10.78
Rhynchaenus fagi	221.02 ± 88.55	921.31 ± 514.25	113.21 ± 43.08
Trixagus dermestoides	225.68 ± 93.38	855.65 ± 203.19	823.64 ± 164.54
Psallus varians	749.45 ± 289.83	747.95 ± 114.68	722.40 ± 292.20

Table 3.5: DNA yield recovered from all species across the sampling solutions. DNA yield was measured by using a Quantus™Fluorometer (Promega).

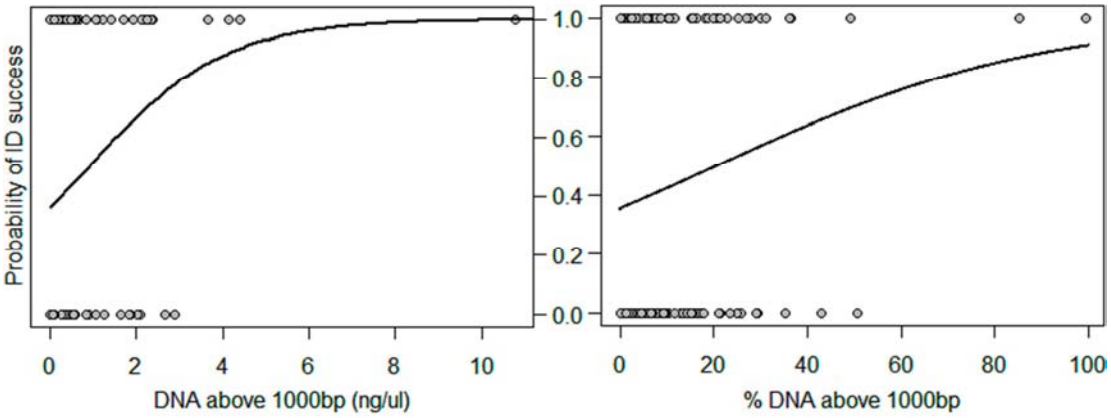


Figure 3.3: The probability of successful identification (ID) of a sample as a function of the concentration (left diagram) and % (right diagram) of good DNA in the sample (Gossner et al. 2016).

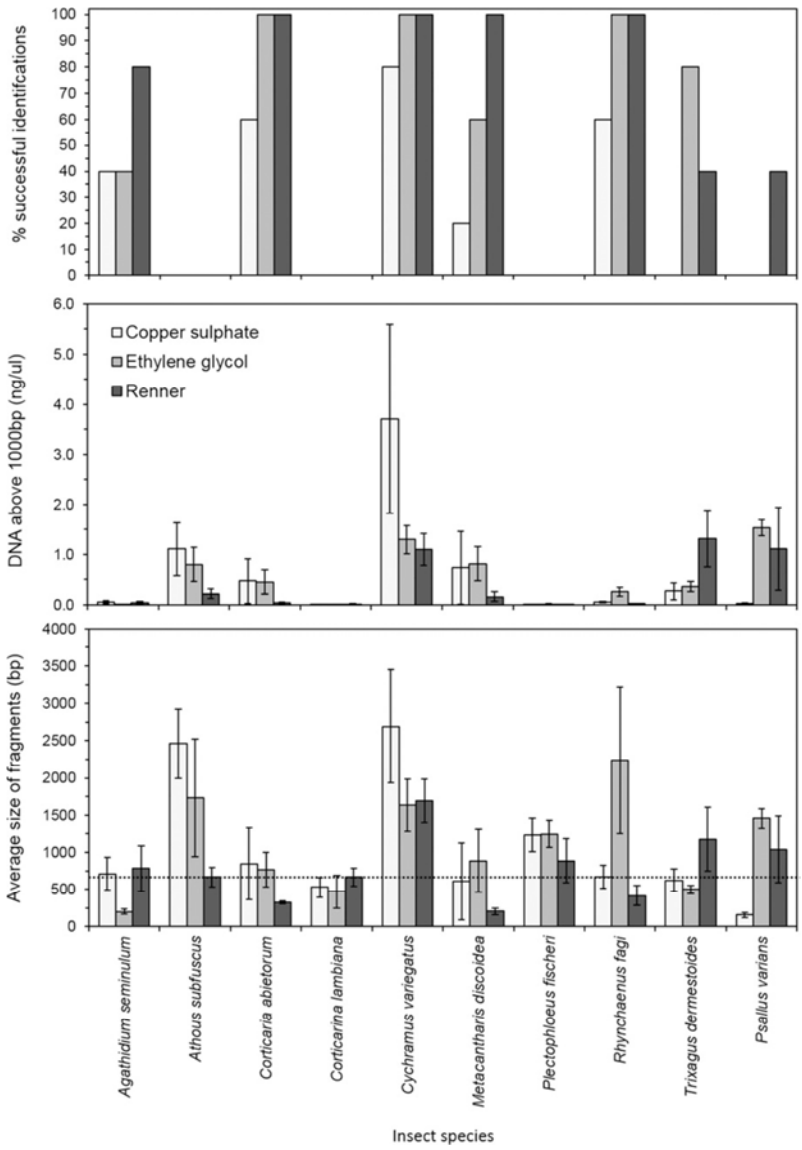


Figure 3.4: DNA yield and quality of studied species (Gossner et al. 2016). The percentage of samples with successful identification of the genetic barcode (top), the concentration of DNA above 1000bp (ng/µl) and average length of recovered DNA fragments (bp), with the dotted line showing the 658bp fragment length required (bottom), across the solutions and species tested. DNA yield and fragment length was measured by a Fragment Analyzer. Error bars ±1SE.

3.4 Discussion

3.4.1 The usage of different conservation liquids for traps

Testing of less expensive alternatives to the usage of ethanol as sampling and conservation liquid showed that condition of samples as well as the composition of sampled arthropod communities clearly depends on the sampling solution. The effects of sampling solution also depended on the forest type, the vertical stratum, and whether top or bottom jars of the flight interception trap were used. Although vertical stratum and jar positioning were not a primary focus in this study some findings are discussed here. However a more detailed analysis and discussion can be found in the publication: Gossner et al. 2016.

The condition of the samples caught in a flight interception trap (Figure 3.1) was more moldy and fragmented in top jars filled with copper sulphate solution and in bottom jars in general when compared to top jars filled with ethylene glycol and Renner

solution. The community parameters, richness, diversity, and abundance, were highly affected by forest type and vertical stratum, but less so by sampling solution on the order level (Figure 3.4).

Species richness and community composition differed greatly. Renner solution had either a highly attractant (beetles) or repellent (true bugs) effect on species when comparing to the other solutions. The change in community composition by sampling solution was mainly related to ambrosia beetles which colonize fresh dead wood and were most likely attracted by ethanol containing Renner solution.

DNA barcoding was highly successful for three species (*C. abietorum*, *C. variegatus* and *R. fagi*; Figure 3.4*)*, very unsuccessful for three species (*A. subfuscus*, *C. lambiana* and *P. fischeri*; Figure 3.4) and the other four species tested produced variable results. Overall, samples collected in copper sulphate showed lower barcoding success than for the other two sampling solutions, which was not directly related to sample condition or reduced DNA yield although the low pH of copper sulphate did influence DNA yield to some degree. Best results were obtained from ethylene glycol and Renner solution, both liquids containing derivates of alcohol.

3.4.1.1 Sample condition

The difference in observed sample conditions was not affected by pH, although pH differed between solutions and changed under field conditions. The generally more moldy and fragmented insects of samples from the bottom jars might be caused by a rainfall-related dilution effect. Samples from ethylene glycol were generally less moldy and fragmented than the other solutions, particularly in beech forests. This indicates that ethylene glycol serves as the best alternative under extreme rainfall-caused sample dilution events.

The sample conditions in the top jars, where the solution was not diluted, was generally better for ethylene glycol or Renner solution than in the bottom jar, but not so for copper sulphate. The better sample preservation in ethylene glycol and Renner solution compared to copper sulphate is in line with studies using pitfall traps with roofs as rainwater protection (Stoeckle et al. 2010; Engel et al. 2003). The incidence of mold on species sampled in top jars filled with copper sulphate is also in line with the observations of samples collected in similar jars attached to stem eclectors in spruce forests (Engel et al. 2003). This is, however, somehow surprising as copper-containing fungicides including copper sulphate are widely used in agriculture (Teviotdaleo et al. 1989; Darriet et al. 2001; Hardy et al. 2007). High humidity in the top jars resulting from water evaporation possibly allowed the molding at the surface of the copper-sulphate solution, where the fungicidal properties did not have an effect on the material. In contrast to previous findings that showed e.g. an attraction of Diptera by decay-induced volatiles the attractiveness of the sampling solution seemed not to change by the level of decay (Schmidt et al. 2006).

However, the suitability for subsequent morphological or genetic analyses is influenced by difference in sample conditions (Schmidt et al 2006; Jud et al. 2008). The impact on the material can depend on the taxonomic group as different cuticle consistency of e.g. soft bodied vs. hard bodied specimens might influence molding and DNA fragmentation (Dillon et al. 1996; Bisanti et al. 2009; Stoeckle et al. 2010). A comprehensive study analyzing DNA quality of species sampled by different sampling

solutions in a multi-taxa approach would certainly be valuable for further and more specific insights.

3.4.1.2 Effects of sampling solutions on insect communities

Analyses on order level did not reveal differences among sampling liquids. However, studies based on pitfall traps found significant differences in the sampled number of individuals among sampling solutions (Engel et al. 2003; Schmidt et al 2006). In literature a decay-induced attraction of Diptera to Renner solution due to ethanol evaporation is mentioned, as well as an attraction of true-flies (Brachycera), snails and slugs (Gastropoda), and wasps (Hymenoptera: Vespidae) to ethylene glycol (Engel et al. 2003; Schmidt et al 2006). Copper sulphate is described as being least attractant (Engel et al. 2003). But then some studies are being discordant in the description of the effects (Holopainen 1990, 1992; Engel et al. 2003; McCravy et al. 2007). Thus the comparison of studies based on different sampling liquids needs to be evaluated critically. However, it has been shown that ethanol is released in decaying wood, probably by microorganisms, and that this attracts bark beetle species (Graham 1968; Byers 1992, 2007; Bouget et al. 2009). This would explain the attraction of fresh dead-wood colonizer to samples filled with Renner. Due to a lack of supporting evidence in literature, differences found in the distribution of Heteroptera can either be a repellent effect of Renner solution or the attraction of the other solutions. Comparisons between communities sampled by copper sulphate and ethylene glycol seemed to be less biased.

3.4.1.3 Interaction between sampling solution and forest type/stratum

The abundance of Diptera was higher in the canopy than in the understory. This pattern can only be found in part in previous studies and can highly depend on the structure and heterogeneity of the forest (Maguire et al. 2014; Scherber et al 2014). Also Diptera were more abundant in spruce compared to beech forests. This might be explained by tree species specificities of these Diptera. The higher availability of dead wood resources for saproxylic species in spruce compared to beech forests in this study might also explain this pattern. The dead wood distribution is a general pattern found in commercial forests of Europe (Ammer et al. 2008; Hessenmöller et al. 2011; Gossner et al. 2014). The sampling solution effects highly depended on the forest type. The attractive effect of Renner solution was highly significant and more pronounced in the understory of spruce forests. This is mainly due to the higher abundance and species richness of fresh wood dwellers in spruce forests compared to beech forests (Gossner et al. 2013). Under these different prerequisites in species occurrences comparisons among sampling solutions (Renner vs. ethylene glycol / copper sulphate) must be considered somehow biased and should be done with caution.

3.4.1.4Species identification through DNA barcoding

Samples generally had lower barcoding success when they were collected in copper sulphate solution, which is consistent with the effects on sampling condition previously discussed. However, while copper sulphate solution reduced the number of sequencing reads that could successfully be assembled, it produced similar yields of DNA as from the Renner solution. This suggests that copper sulphate with its high amount of water may lead to an increased rate of DNA degradation. Indeed shorter sequence lengths were obtained from copper sulphate samples than from the other sampling solutions. However, the overall average length of fragments in samples from copper sulphate was

not substantially lower than from the other solutions, but much more variation in sequence lenghts both among and within the different species was verified for copper sulphate collected samples. Oxidative DNA damage and DNA breaks mediated by copper ions have been verified before (Lloyd et al. 1999; Cervantes-Cervantes et al. 2005; Jose et al. 2011). It is also suspected that copper negatively affects DNA synthesis and leads to single base substitutions, which potentially already influences the amplification of sequences during PCR (Tkeshelashvili et al 1991). There is also evidence that the low pH of copper sulphate might have led to the reduced DNA yield. It is known that a neutral or alkaline pH is limiting DNA degradation (Lindahl 1993). However, pH did not have a significant effect on identification success of the samples.

Although A*. subfuscus* yielded seemingly sufficient DNA it produced no sequence results. A possible explanation is a mutational change in the primer binding site. While the primers used are considered universal or standard primer individual variance can occur (Folmer et al. 1994; Simon et al. 1994). A very low DNA concentration can be assumed to have caused amplification failure in the two other species samples. Extrapolations from the data showed that a concentration of 3.1 ng/µl of DNA was required to achieve an overall success rate of 80%.

Following from this the choice of sampling solution and species of interest will therefore have a large impact on the results if the experiment is based on methods depending on amplification processed like PCR. The most adequate primer pair should then be used that best fits the taxonomic group of interest. This should be kept in mind when evaluating results to avoid biasing the scientific research towards those species that consistently produce good DNA for analyses.

Only three alternatives to ethanol that are commonly used in current biodiversity studies were tested in this study. Recently, 2% SDS and 100mMEDTA were recommended as a cheap, stable and easily transportable alternative to ethanol for preserving specimens and their DNA collected in the field (Pokluda et al. 2014). Its attracting effect has, however, not been tested and while it might be a good choice for DNA preservation it is unknown if community biases (as we found for Renner) occur.

3.4.2 Conclusion and recommendations

In biodiversity studies many different properties of sampling solutions have to be considered; costs, toxicity, evaporation, attractiveness to selected species, and good preservation properties for subsequent morphological and genetic analyses. Whenever high-proof ethanol (96%) is ineligible and an alternative must be used the following recommendations can be made considering the results and previous studies:

1. To obtain optimally preserved insects ethylene glycol instead of Renner solution or copper sulphate is suggested as this solution has better preservation properties in all tested microclimatic situations. Propylene glycol might be used as a less toxic but even more expensive alternative as it showed no different attraction compared to ethylene glycol in previous studies (Weeks et al. 1997).

2. When decisions on sampling solutions are financially restricted and morphological identification is targeted, copper sulphate can be a suitable alternative because it costs only 7% of the price of ethylene glycol. By reducing sampling intervals, molding of copper sulphate samples most likely could be minimized.

3. Meta-analyses of data sampled with flight-interception traps using ethylene glycol or copper sulphate are assumed to be insignificantly biased, because –in contrast to pitfall trap studies- no differences in abundance, species richness or community composition between samples were found. Comparisons with samples caught with a Renner solution should, however, be critically questioned.

4. When aiming at subsequent DNA analyses ethanol has mostly been used in the past. But because it is very susceptible to evaporation from open receptacles, ethylene or propylene glycol might be an alternative (Vink et al. 2005; Höfer et al. 2006; McCravy et al 2007). Samples from ethylene glycol and Renner performed similarly in producing applicable sequences for species identification. For copper sulphate the amount was generally lower and strongly dependent on the species used. Due to the potential bias towards or against certain species by the Renner sampling solution ethylene glycol can be suggested as an alternative sampling solution when genetic analyses tools are intended and high proof ethanol cannot be used.

4 Mixed species DNA-samples

4.1 Introduction

Most sequencing techniques are at some point still based on the application of PCR for sequence amplification. As effective this method has proven to be in the amplification of a DNA section of a single template as delicate is its application when multiple templates are involved. A source of such multiple templates can be found in the DNA of different species. The attempt to amplify the same genetic region of different species within one sample usually results in a set of sequences that are not in all aspects identical to its origin. Of course some of the original sequences can be recovered but other might be missing. And again some sequences are different in a way that they are mixture of the original input sequences (Suzuki and Giovannoni 1996; Cronn et al. 2002; Kanagawa 2003; Bellemain et al. 2010). These artefacts or chimeras are especially hard to identify if the original input is unkown. On the other hand is the genetic species identification especially desirable in diversity assessments of ecosystems with unknown composition. To estimate to which extend the result is biased by the species amount and species composition samples with a known and defined species input were artificially created and processed. It was assumed that an increasing number of different species (different templates) leads to the increased formation of chimeric artefacts. It was also assumed that close relatives (similar templates) also fascilitate chimera formation as sequence fragments tend to easily bind to a different template if the different temaplate is similar to the original template. It was further tested if different species ratios in a sample can be recovered by the number of sequences in the output.

4.2 Material and Methods

4.2.1 Material acquisition and sample preparation

All species where selected from the material that initiated the barcoding reference dataset. Limiting the choice to these specimen it was guaranteed, that all chosen specimen already fulfilled a minimum quality standard, as they had already been checked when introduced to the reference set. This means species had already correctly been identified morphologically and were then successfully processed via single Sanger sequencing, yielding high quality species specific COI sequences. To ensure optimal quality for the following steps only material matching the following criteria were considered further. (1) Previous Sanger sequencing of the material had to result in the full segment length of 658 bp. (2) No ambiguities were allowed in these sequences. (3) The phred score for good quality bases had to be ≥40 for at least 95% of the bases and finally (4) not more than 1% of the bases were allowed to have a phred score of ≤20. Following these criteria 33 species samples were selected to cover the range from sibling species up to different infraorders, aiming at testing the influence of a) the number of different species, b) the patristic distance and c) the species proportions on the recovery of species and the formation of sequence chimera in the barcoding process (Species and their taxonomy in figure 4.2). a), b), and c) can be considered as natural or external influencing factors. As a methodical or internal influencing factor the effect of d) PCR processing was analyzed. Addressing the later factor two approaches were

chosen. In the first treatment (T1) single species DNA extractions were mixed and afterwards the COI sequences were amplified via PCR. In the second treatment (T2) the COI sequence of each species sample was amplified first and then the amplified COI sequences were mixed according to the different species combinations already used in the first approach.

	Identical source material for both treatments (T1 and T2)	
-	DNA extraction of 33 identified species	
-	Measurement and equalization of the extracted DNA concentration (dilution)	
	Mixed amplification (T1)	**Single amplification (T2)**
-	*Mixing of the DNA **extracts** (!)*	- COI sequence amplification (single PCR)
-	COI sequence amplification (mixed PCR)	- *Mixing of the **amplicons** (!)*
-	Illumina MiSeq sequencing	- Illumina MiSeq sequencing

Table 4.1: Two different treatments (T1 and T2) of identical species compositions sequences were carried out to test in which way PCR influences the result of a sample with different mixed DNA templates (species). The first is the amplification of mixed DNA, T1 the DNA mixture. The second is a single amplification different templates that are mixed afterwards, T2 the amplicon mixture.

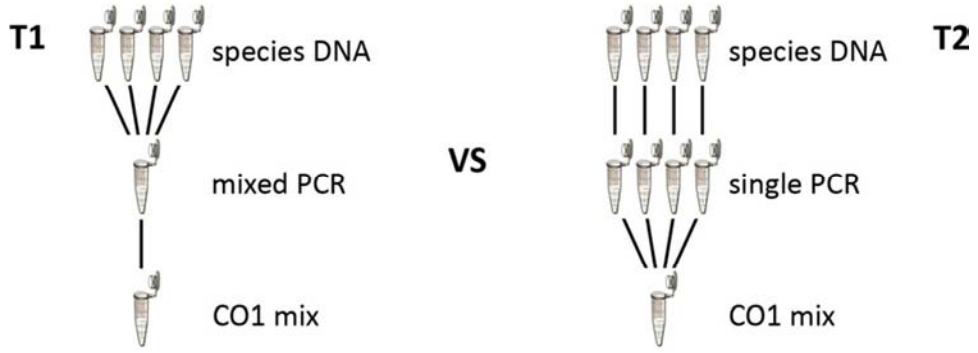


Figure 4.1: Illustration of the different treatments to obtain the different samples. On the left T1: a PCR with mixed DNA of different species. On the right T2: multiple single PCRs of the different species with mixing the amplified sequences afterwards. T1 simulates a common step of PCR based sequencing approaches; T2 simulates PCR free methods in sequencing approaches.

Accordingly 2 x 95 samples were prepared in different sets, containing either different species combinations or species differing in their contributing DNA proportions (Table 4.2; detailed species list of mixtures in supplement S2.1). Thus the DNA quantity was standardized over all source samples to generate quantifiable entities (0.45 ng/µl). These entities were then combined according the experiment. Samples analyzing the influence of species numbers and their relationship/patristic distance contained 1 µl per species. Samples with growing proportions had different DNA quantities per contributing species (1 µl to 5 µl to 10 µl). Each setup had a ten times repetition. The first set combined randomly chosen species in different quantities of 5, 10, 20 and 30 species. This set with a mean patristic distance of 0.1545 comprising 40 samples was named “arbitrary” (A). The second set consisted only of species with low patristic distances to each other. This set was named “close” (C) for choosing only closely related species. Species of this set were combined to quantities of 5 and 10 species and had a mean patristic distance of 0.1365. The third set consisted only of species with a high patristic distance. This set was named “distant” (D) for choosing distantly related species. The species were again combined to quantities of 5 and 10 species and had a mean patristic distance of 0.1630. In a fourth set for analyzing proportions (P) the set always contained 3 randomly chosen different species each with different DNA

quantities (0.45 ng/µl – 2.25 ng/µl – 4.50 ng/µl). All samples where then sent to LGC Genomics GmbH, in Berlin, Germany for further processing.

No. 01	**No. 02**	**No. 03**	**No. 04**	**No. 05**	**No. 06**	**No. 07**	**No. 08**	**No. 09**	**No. 10**	**No. 11**
5 G	5 A	5 A	5 A	5 A	5 A	5 A	5 A	5 A	5 A	5 A
No. 12	**No. 13**	**No. 14**	**No. 15**	**No. 16**	**No. 17**	**No. 18**	**No. 19**	**No. 20**	**No. 21**	**No. 22**
5 C	5 C	5 C	5 C	5 C	5 C	5 C	5 C	5 C	5 C	5 D
No. 23	**No. 24**	**No. 25**	**No. 26**	**No. 27**	**No. 28**	**No. 29**	**No. 30**	**No. 31**	**No. 32**	**No. 33**
5 D	5 D	5 D	5 D	5 D	5 D	5 D	5 D	5 D	10 G	10 A
No. 34	**No. 35**	**No. 36**	**No. 37**	**No. 38**	**No. 39**	**No. 40**	**No. 41**	**No. 42**	**No. 43**	**No. 44**
10 A	10 A	10 A	10 A	10 A	10 A	10 A	10 A	10 A	10 C	10 C
No. 45	**No. 46**	**No. 47**	**No. 48**	**No. 49**	**No. 50**	**No. 51**	**No. 52**	**No. 53**	**No. 54**	**No. 55**
5 C	5 C	5 C	10 C	10 C	10 C	10 C	10 C	10 D	10 D	10 D
No. 56	**No. 57**	**No. 58**	**No. 59**	**No. 60**	**No. 61**	**No. 62**	**No. 63**	**No. 64**	**No. 65**	**No. 66**
10 D	10 D	10 D	10 D	10 D	10 D	10 D	20 G	20 A	20 A	20 A
No. 67	**No. 68**	**No. 69**	**No. 70**	**No. 71**	**No. 72**	**No. 73**	**No. 74**	**No. 75**	**No. 76**	**No. 77**
20 A	20 A	20 A	20 A	20 A	20 A	20 A	30 G	30 A	30 A	30 A
No. 78	**No. 79**	**No. 80**	**No. 81**	**No. 82**	**No. 83**	**No. 84**	**No. 85**	**No. 86**	**No. 87**	**No. 88**
30 A	30 A	30 A	30 A	30 A	30 A	30 A	33 G	3 P	3 P	3 P
No. 89	**No. 90**	**No. 91**	**No. 92**	**No. 93**	**No. 94**	**No. 95**				
3 P	3 P	3 P	3 P	3 P	3 P	3 P				

Table 4.2: Samples with the number of species used and the type of mixture: A – for arbitrary chosen species, C – for closely related species, D – for distantly related species, P – for different DNA proportions per species. (Detailed list of species mixtures in supplement S2.1).

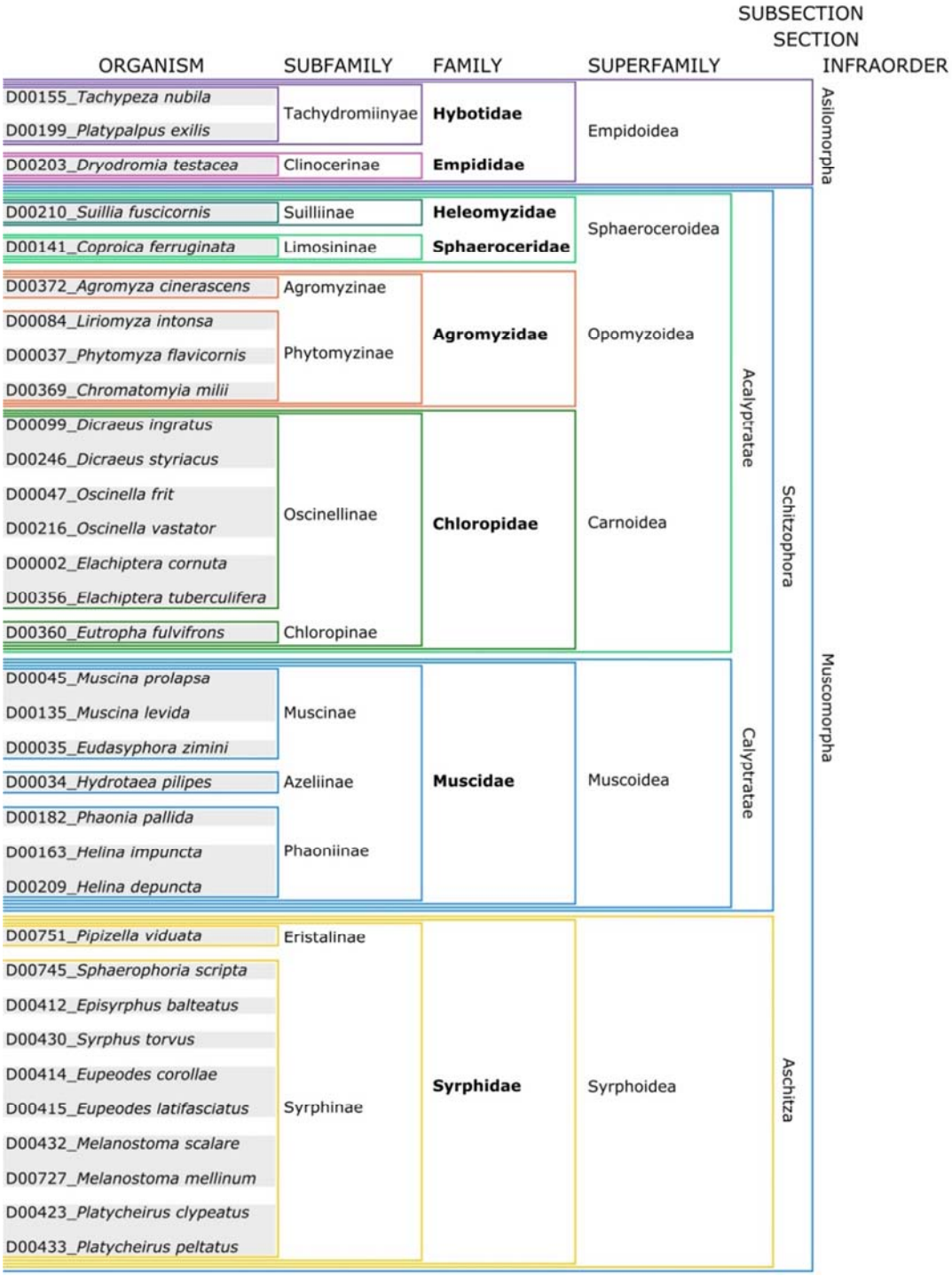


Figure 4.2: Taxonomic classification of the 33 different specimens that were used in the different species mixtures. Initial numbers are unique identifiers of the internal sequence reference list.

4.2.2 Next Generation Sequencing

At LGC (LGC Genomics GmbH, in Berlin, Germany) all 2x 95 sample mixtures were tagged using Illumina TruSeq adapters with an individually indexed identifier (Table 4.3; Figure 4.3).

TruSeq Universal Adapter
5’AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT’3
TruSeq Indexed Adapter
5’GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-NNNNNN-ATCTCGTATGCCGTCTTCTGCTTG’3

Table 4.3: LGC True Seq Adapter for sample tagging, where “N” stands for a combination of six nucleotides to form are a unique index-sequence which can readily be identified as unique to one library.

Samples were sequenced using 300 bp paired-end reads on the Illumina MiSeq V3 (LGC Genomics GmbH, Berlin, Germany). CASAVA data analysis software (Illumina, San Diego, U.S.A.) was used for demultiplexing. Data were sorted according to sample origin.

Figure 4.3: Sequence strand after adapter merging through PCR.

4.2.3 Data processing

Read pairs (R1, R2) were treated as individual datasets to compare findings of each set. MiSeq Sequencing also does not produce sufficient overlap for pairing when templates with a length of 658 bp or longer are used. All raw-sequences were filtered for homopolymers with a maximum of 10 base repeats as this was the longest repeat found in a sequence of *Epidapus microthorax* (10 x G) present in the GBOL databse. For this and the following steps Mothur v.1.39.5 was used (Schloss et al. 2009). Duplicate sequences were merged for more efficient computation. A count table was generated containing the names of the unique sequences, the groups and the quantity of unique sequence in each group. The query sequences were then aligned to a reference alignment based on the used specimen. The template sequence for the query sequence was searched for using a kmer search with 8mers to find the best match. A pairwise alignment is then created using the Needleman-Wunsch method with a reward of +1 for a match and penalties of -1 for a mismatch and -2 for a gap. Columns that only contain gap characters (i.e. '-' or '.') were being ignored to accelerate the calculation of distances. If processing generated new duplicate sequences through removing characters the duplicates were merged again. Single-linkage preclustering was then applied to remove sequencing errors (Huse 2010). Potentially chimeric sequences were identified using uchime and a reference template (Edgar et al. 2011). The reference template contained the 33 chosen species sequences. Chimeric Sequences were then removed. For a “blind” chimera analysis with no reference the mothur based rewrite of Chris Quince's chimera detection program, Perseus was used (Quince et al. 2011). After this sequences were then classified using a naïve Bayesian classifier looking at the query sequence kmer by kmer, calculating the probability the query sequence would be in the given reference template taxonomy based on the kmers it contains. The assignment was then checked by a bootstrapping algorithm to find the confidence limit of the assignment (Wang et al. 2007; Cole et al. 2014; Porter et al. 2014; Vinje et al. 2015). Vsearch clustering assigned the sequences to MOTUs (Rognes et al. 2016). Abundance based greedy clustering (agc) was chosen as the clustering algorithm as it is very robust and generates the most stable MOTUs (Westcott and Schloss 2015). MOTU taxonomy was assigned using the consenus taxonomy of the given MOTU. For a taxon to be included at least 51% of the sequences had be identically classified at that level in the given MOTU.

4.2.4 The influence of sample treatment, species number and species quantity

The percentage of identified chimeras per sample was calculated. It was then evaluated how the percentage of chimeras is distributed over all samples in regard to the different approaches T1 and T2 and in each of both reads (R1 and R2).

A Boxplot analysis conducted in R (Version 3.2.3) was used to visualize the influence of a) the number of different species used in the sample and b) the averaged patristic distances of, (A) arbitrary chosen species, (C) species that are taxonomically close, (D) and species that are taxonomically distant to each other for the approaches T1 the mixed amplification and T2 the single amplification for the first and second sequencing run R1 and R2.

The a) the number of species used in the sample and b) the averaged patristic distances of (A), (C), and (D) were tested for its significant influence on the formation of chimeras using a Kruskal-Wallis rank sum test.

All sequence lengths of the first treatment T1 were then manually shortened for all R1 and R2 samples involved in testing the influence of growing species numbers within samples on the formation of chimeras. The edited samples where then checked again with Perseus (Quince et al. 2011). The number of chimeric sequences found was then compared again with the number of chimeras found in samples with full length sequences using Boxplots.

A generalized linear model was used to predict the development of chimeras for growing species numbers within a sample for both treatments T1 and T2. It was also used to analyze the influence of the treatments and different DNA proportions on the amount of identified sequence clusters after sequencing. A Kruskal-Wallis test was used to test whether the influence was based on the T1 or T2 data.

To test wether the number of sequence reads identified for a species depended on the different treatments, the sequence reads per species were first tested for normal distribution in T1 and T2 using the Anderson darling test. A Welch test for unpaired samples with different variances was then applied to test if T1 and T2 data was significantly different from each other.

4.3 Results

Sequencing was ordered as “paired end sequencing”. This means products are sequenced from both directions resulting in two data sets, R1 for the first set of sequencing reads and R2 for the second set. Instead of pairing the sets the second set (R2) was planned to be treated as a repetition of the first (R1) as no overlap between the two read directions could achieved due to the product length of 658 bp. Comparing the first and second sequencing run (R1 and R2) directly after sequencing a significantly higher amount of unique sequence reads was found in R2 (Kruskal-Wallis chi-squared = 102.89, df = 1, p-value < 2.2e-16) (Figure 4.4). The two different treatments T1 and T2 however, did not differ much in the amount of unique sequences.

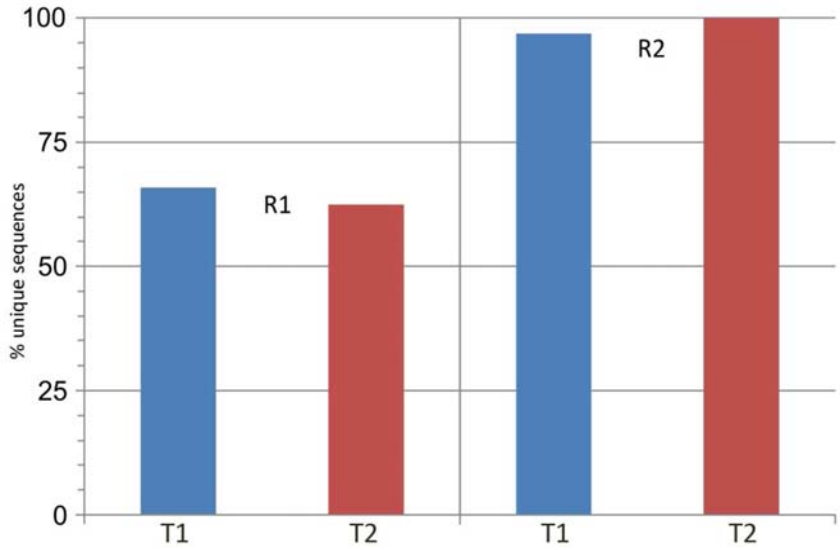


Figure 4.4: Percentage of unique sequences in the treatments T1 for the mixed amplification and T2 for the single amplification of the two sequencing sets R1 and R2. R2 has significantly more unique sequence reads than R1 ($p<0.001$).

4.3.1 Differences in the chimera identification

The amount of identified chimeras was normally distributed for both treatments T1 and T2 in R1 but not so for R2 following the Shapiro-Wilk normality test (Table 4.4). With its extreme amount of unique sequences data must be considered to be strongly biased. R2 data was therefore excluded from most of all further analysis from this point.

Run	treatment	W	p-value
R1	T1	0.96391	0.2274
	T2	0.98275	0.7899
R2	T1	0.92559	**0.01159***
	T2	0.73693	**4.156e-07***

Table 4.4: Shapiro-Wilk normality test of chimera distribution for the two treatmens in both runs. For R1 data shows to be normally distributed. R2 reveals to be not normally distributed. Significant values are highlited.

A clear trend of increasing numbers of chimeras in reference to increasing numbers of total sequences within a sample was found (Figure 4.5). However, in the second treatment, amplicon mix T2, the slope is not as steep as in T1 the mixed amplification treatment. Only for samples with an extreme low amount of sequences a reversal of the trend can be observed. T1 data is also more widely scattered on the x-axis than T2 revealing a larger variance in chimera detection for T1 reads (Figure 4.5).

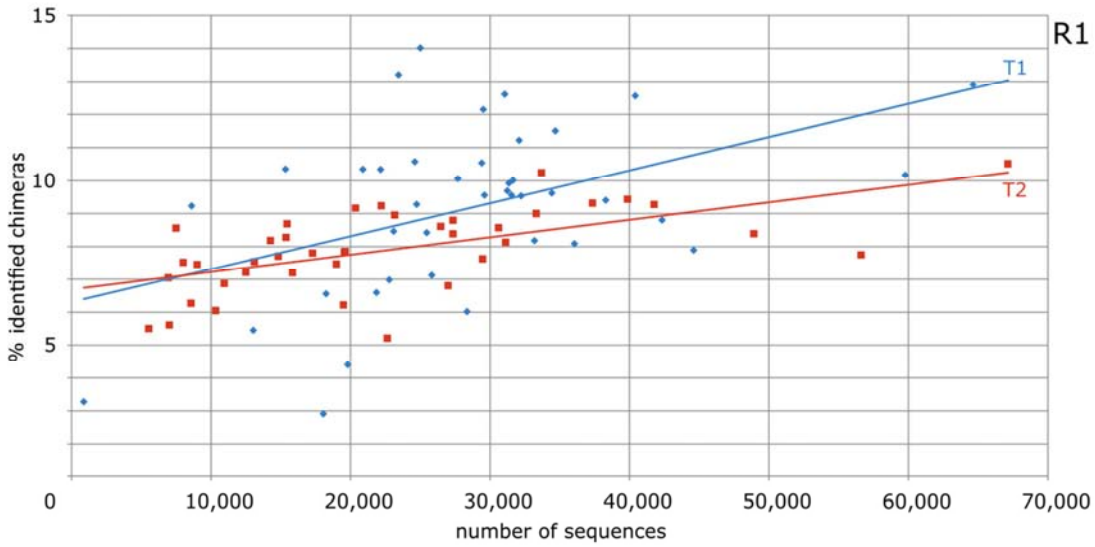


Figure 4.5: Trend line about the emergence of chimeras over all samples with randomly chosen species combinations (Table 4.2, A) for the first sequencing run R1 (R2 in supplement S2.2). Higher numbers of total sequences within the sample lead to the identification of more chimeras in T1, the mixed amplification than in T2, the single amplification.

Boxplots show that the number of species used in a sample influenced the amount of chimeric sequences in R1 significantly in treatment T1 (mixed amplification) but not in T2 (single amplification) (Figure 4.6, left). This result was confirmed by a Kruskal-Wallis rank sum test. It showed that the influence of species number on the amount of identified chimeric sequences is highly signific (p=0.001294*) (Table 4.5). The different averaged patristic distances, nameningly (A), (C), and (D) of the used species within a sample was not found to affect the number of chimeric sequences in T1 or in T2 (Figure 4.6, right). This result is also validated by the Kruskal-Wallis rank sum test revealing no significant influence (Table 4.5).

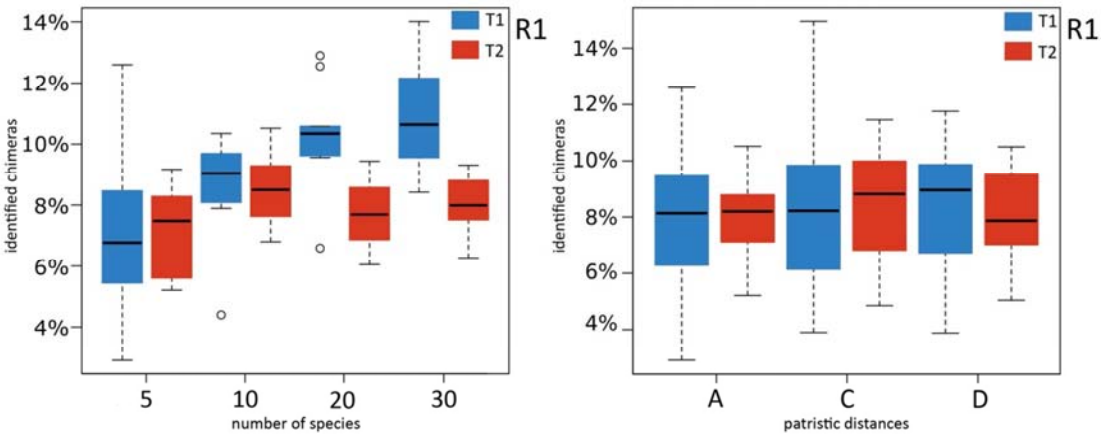


Figure 4.6: Boxplot diagram for R1 sequencing data. Left: Identified chimeric sequences in relation to the number of different species (5, 10, 20, and 30) within a sample of treatment T1 (mixed amplification) and T2 (single amplification). Right: Identified chimeric sequences in relation to the averaged patristic distances of, (A) arbitrary chosen species, (C) species that are taxonomically close, (D) and species that are taxonomically distant to each other. (R2 in supplement S2.3)

R1	test of	chi-squared	df	p-value
T1	species number	15.72	3	**0.001294***
T2	species number	5.0239	3	0.1701
T1	patristic distances	0.80248	2	0.6695
T2	patristic distances	1.6751	2	0.4328

Table 4.5: Kruskal-Wallis rank sum test to test the influence of species numbers and patristic distances on the amount of chimeric sequences in the first sequencing run R1. Significant value is highlited. (Table for R2 data in supplement S2.4)

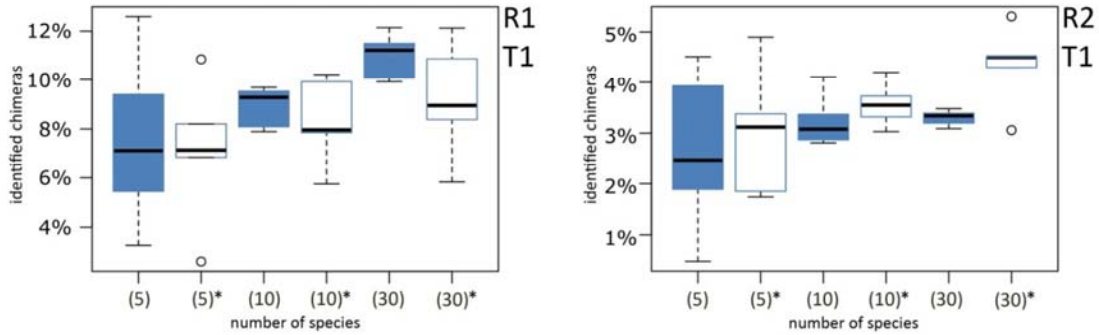


Figure 4.7: Comparison of the amount of chimeric sequences (in relation to the total amount of sequences in a sample) found in T1 samples with different amounts of species, where the sequences where shortened to a length of 190 bp (species number*) and un-shortened sequences (species number). The shortening of the sequences showed to have contrary effects for R1 (left) and R2 (right). For R1 the number of chimeras decreased in comparison to the original samples while it increased for R2. However, R2 samples still had lesser chimera identifications compared to R1.

After inspecting the reads of the two data sets R1 and R2 it was found that the reads started to gather multiple N's in the data set from the second sequencing run R2 especially after the 190's base position. The sequence reads from R1 this problem was not found. LGC Genomics confirmed that R2 data can suffer from increased sequencing errirs due to chemical and sequence "exhaustion". To test the influence of these errors a large erroneous fragment of the reads was cut off. For R1 data fewer chimeras were identified after the original reads were shortened from the 5' end (Figure 4.7, left). The opposite effect was found for R2 data after shortening. Here the number of identified chimera was higher after the sequences cut after that position (Figure 4.7, right).

4.3.2 Identification of species

The quantified DNA of the chosen 33 species (Figure 4.2) was used between 29 and 38 times to be distributed over 95 samples for T1 and T2 (190 samples total). The DNA of all individual species together was used 1078 times for T1 and T2 (2,156 times total). In 90.54% of these cases for T1 (mixed amplification) species coul be recovered. For T2 (single amplification) species were recovered in 99.54% of the cases. In other words for T2 detection only failed in 5 of 1078 cases. Six species were responsible for the occurring false negatives in T1. Of these six species only two could not be recovered in T2. Responsible for most of the false negatives was *Phaonia pallida*. The DNA of this species was integrated into 36 samples each for T1 and T2 (72 samples total). Its detection failed in 75% of the T1 mixed amplification samples it was applied to. In T2, where all species were individually amplified, detection failed in 11.11% for *P. pallida*. The other five species were *Dryodromia testacea* (T1: 75%, T2: 0%; of 32 samples each), *Eudasyphora zimini* (T1: 56.25%, T2: 0%; of 32 samples each), *Suillia fuscicornis* (T1: 51.52%, T2: 0%; of 33 samples each), *Platypalpus exilis* (T1: 33%, T2: 0%; of 30 samples each) and *Muscina prolapsa* (T1: 18.75%, T2: 3.13%; of 32 samples each).

The amount of false positives makes a cross-contamination of the samples very likely. In all samples small amounts of almost all the used species were found. However, their concentration was usually considerably lower than the concentration of species that were supposed to be in the samples.

4.3.3 Analyzing DNA input ratios

Analyzing the samples, always containing three randomly chosen species in differing combinations and increasing proportions of DNA, differences in the numbers of species associated sequence reads can be found. A generalized linear model described the difference between the proportions 1 and 10 to be significant (Estimate= 33.050, Std. Error= 11.975, t-value= 2.760, p=0.007802*). A Kruskal-Wallis test showed a significant influence of the proportions used for T2 (chi-squared= 9.7419, df = 2, p= 0.007666*) but not so for T1 (chi-squared= 2.7955, df= 2, p= 0.2472).

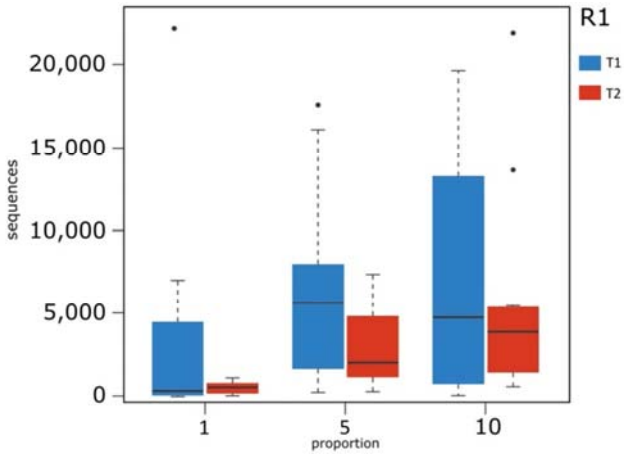


Figure 4.8: Increasing proportions (1: 0.45 ng/µl; 5: 2.25 ng/µl; 10: 4.50 ng/µl) of species DNA does generally influence the number of species read output. However, the individual results vary strongly from species to species making a general estimate about the DNA input quantity impossible.

Comparing the numbers of species specific sequence reads over all samples initially using identical amounts of DNA for each species, it becomes apparent that the number of sequence reads does not depend on the treatment.

T1 species sequence data was not normally distributed according to the Anderson darling test, whereas T2 species sequences did show a normal distribution (T1 p= <0.05*; T2 p= 0.228). It also showed that T1 data was more scattered than T2 data (Figure 4.9). However, their mean sequence read number per species did not differ very much. A Welch test that is commonly used to compare samples with different variances showed that T1 and T2 sequence outcome for species sequences was not significantly different (t= 1.0012; df=64; p= 0.3205). Instead read number depends on the Diptera species used (Figure 4.10).

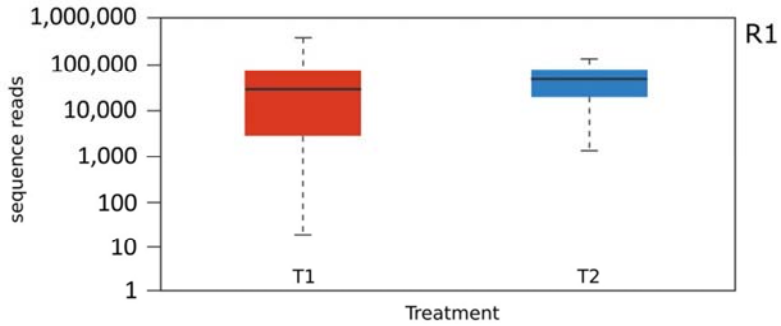


Figure 4.9: Distribution of sequence read numbers for all 33 species used in the experiment show that T1 data is more scattered than T2. However, their mean sequence read number per species does not differ very much between T1 and T2.

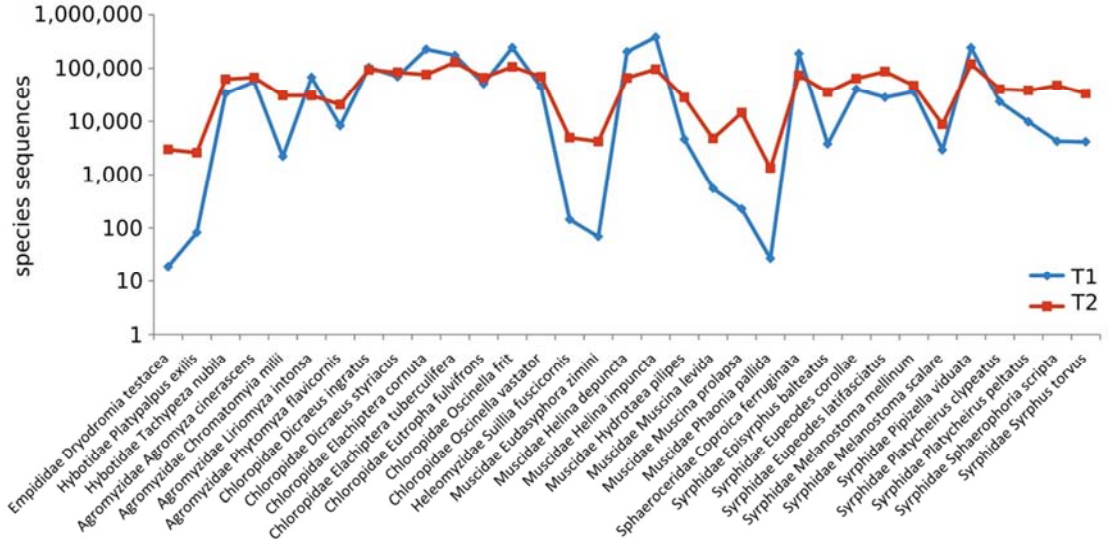


Figure 4.10: The number of identified species sequences in the different treatments T1 and T2 shows a higher variance in T1 than in T2. However, sequence numbers reveal a trend to correlate with the species used for amplification.

4.4 Discussion

As reliable as PCR has proven to be in amplifying distinct regions of single specimen as limited is it when it comes to amplifying identical regions in samples of mixed specimen. Studies have shown that these samples are susceptible for errors when PCR is applied to such mixed samples (Mullis et al. 1987; Thompson et al. 2002; Ashelford et al. 2005; Edgar RC et al. 2011). This becomes especially problematic when PCR is applied to samples with unkwon species content for diversity assessments. Results are strongly biased due to the selective amplification of different sequence templates and the forming of hybrid sequence copies due to disruptions and resumptions of the amplification process (Hansen et al. 1998; Wang and Wang 1996, 1997; Polz and Cavanaugh 1998; Acinas et al. 2005). The experiments showed that these errors are PCR induced and are linked to the number of different species within a sample (Figure 4.6, left). Surprisingly species relation and with this sequence similarity did not affect the forming of chimeras (Figure 4.6, right). At least not in the tested extend of differences found in the family of Diptera and the species used.

Analyzing the linkage between DNA concentration input and read output of a sample it can be said, that the amount of species DNA certainly affects the number of its

output reads, however clear estimates about the original input quantity cannot be made. The variations can be massive and are not only PCR based but also species specific (Figure 4.8; Figure 4.9).

4.4.1 Chimeras and sequencing errors

When the sequence reads of the first and second sequencing run of the identical samples (R1 and R2) were analyzed it initially seemed that data from R2 sequencing had considerably less chimeras. It soon turned out to be an artifact of extensive sequencing errors within R2. Checking for unique sequences in R2 data it revealed that over 90% of unique sequences were found in the first treatment and about 100% of unique sequences in the second treatment (Figure 4.3). This seems unlikely considering that only 33 different species were used in the 95 analyzed samples. In contrast to this the software detected 0% chimeras in some of the samples where chimeras should have been found according to the analyzis of R1 data. The non-normal distribution of chimeras in R2 then additionally indicated that the results were strongly biased (Table 4.4). After checking back with LGC Genomics in Berlin, the company admitted that an increased error type can be expected in the second reading process (R2) as the chemicals and COI strands can experience some sort of exhausten during processing. This effect has also been described as "phasing" and one of the main sources of sequencing error on the Illumina platform (Kircher et al. 2009; Kircher et al. 2012). After this confirmation R2 data was excluded from most of the following analyses.

However, it should not be underestimated how the ability to detect chimeras with current software can be inhibited due to sequencing errors. Most algorithms sort sequences after their frequency and then start to compare the less common sequences as potential chimeric sequences with the more frequent sequences as their potential sources. If the software does not recognize a chimeric sequence as similar enough to its sources (e.g. due to several sequencing errors) it will not be identified as a chimera. As most of these errors cumulated towards the end of the sequence, all R2 sequences were manually shortened to a length of 190 bp. A manually inspection revealed that after that position multiple "N's" (IUPAC: N for any Nucleotid) started to accumulate in the strand (IUPAC 1997). At the same time this length still guarantees reasonable chances for a sequence to lead to species identification (Meusnier et al. 2008). When the amount of chimeras was compared between the original and the shortened sequences, contrary results were found for R1 and R2 (Figure 4.7). While in R1 the number of identified chimeric sequences declined it increased in R2. The explanation for effect is that in R1 the detection of chimeric sequences worked well. It can be assumed that almost all chimeric sequences have identified as such. In R2 chimeric identification did not work so well. Although results should be identical the amount of chimeric sequences detected in R2 is far less than the amount detected in R1 (Figure 4.7). What happens now, when a part of a chimeric sequence is being cut off differs in R1 from R2. In R1 chances are high that the sequence was already identified chimeric and that the chimeric part is being cut off. As a result the number of chimeric sequences decreases. In R2 the chimeric sequence was not identified as such because of the high error rate within the sequence. When the erroneous part that has a disguising effect on the chimera is cut off, the software is able to identify the chimera as such again. This then leads to increasing numbers of chimeric sequences in R2 as most have not been identified before (Figure 4.7, right). This means that under certain conditions shortening a sequence can have a

positive effect on data quality. However, identifying chimeras and detecting sequencing errors goes hand in hand. Multiple tools have been and still are being invented to improve quality filtering and detect chimeric sequences (Huber et al. 2004; Ashelford et al. 2005; Haas et al. 2011; Edgar et al. 2011; Wright et al. 2012; Callahan et al. 2016; Edgar 2016).

4.4.2. Factors inducing the forming of chimeras

It is known that DNA of multiple species witin a single sample makes it prone for the forming of chimeras when PCR is used to amplify the regions of interest (Kanagawa 2003). It can therefore be supposed that increasing numbers of species within a sample would lead to an increased formation of chimeras during PCR in relation to the number of different species. From the process of chimera forming it could also be supposed that similar template sequences further promote the origination of chimeras as uncompleted sequence copies should have a higher affinity to bind to more similar DNA strands during PCR than sequence templates that are less similar (Figure 1.2). This would mean that samples containing closely related species could be especially susceptible for the origination of chimeras.

The experiments confirmed that increasing the number of species in a sample significantly increases the origination of chimeras (Figure 4.6, left). Apparently this effect is closely associated with the usage of PCR and it must be expected that larger numbers of different species increase this effect further (Figure 4.10). However, this effect can be reduced when the number of PCR-cycles with mixed sequences is being minimized as done in the second treatment T2 where the species DNA was amplified seperately (Table 4.1; Figure 4.1 and Figure 4.6, left). Although PCR could not fully be eliminated, due to the process of specific sample tagging for the sequencing process. Here PCR was used to mark sequences with individual tags to be able to assign them to their original sample (Figure 4.3). Still no significant influence was found between the number of identified chimeras and the number of species within a sample for samples with the treatment T2.

Surprisingly sequence similarity had no effect on the number of identified chimeras (Figure 4.6, right). Although it could be assumed that during the formation of chimeras incomplete complementary sequences bind more easily to a more similar sequence template no such effects were found. However, experiments included only Diptera sequences with a maximum distance of 19.3% (127 differing bases between the sequences of *Muscina prolapsa* and *Liriomyza intonsa*) and even more distant taxa could possibly lead to reduced numbers of chimeras.

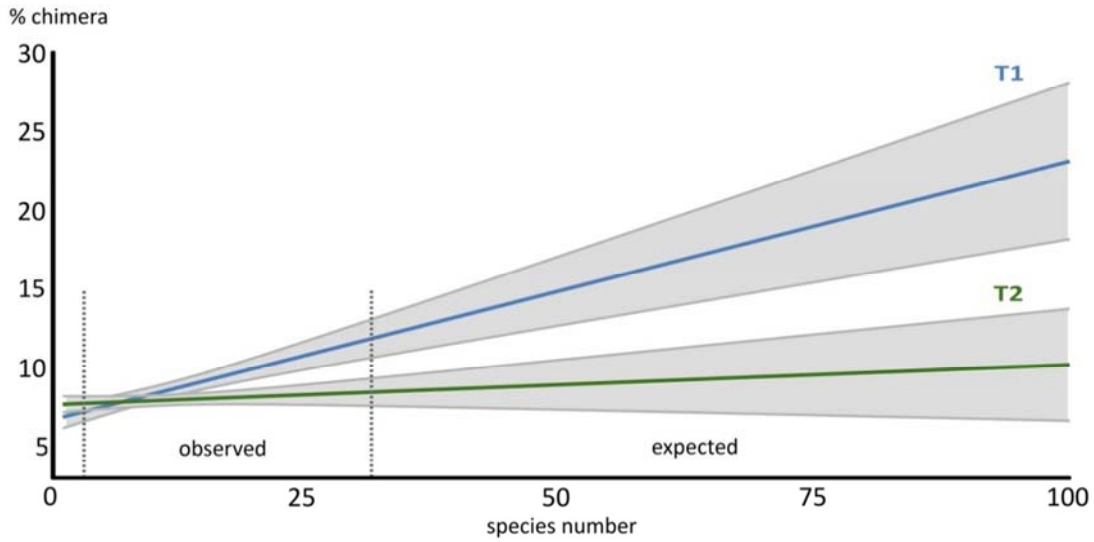


Figure 4.10: A prediction based on a generalized linear model assumes fewer chimeras for T2 than T1 when the number of species in a sample increases. In grey: 95% confidence interval.

4.4.3 DNA ratios and species abundances

In certain aspects DNA-based assessments have proved their superiority over morphological assessments in multiple taxa (Hebert et al. 2004; Smith et al. 2006, 2007; Stein et al. 2014; Janzen et al. 2017). This digital available comprehensive knowledge allows extensive evaluations of samples with these various taxa. However, in one aspect no sufficient progress was made in the last years. The measuring of species abundances based on genetic approaches is requested especially by ecologists. The modest results of the experiments analyzing the effect of DNA ratios in mock samples accords to the results found in literature (Elbrecht et al. 2015; Piñol et al. 2015). Although a significant influence can be confirmed, absolute numbers of reads are highly variable. The inconsistency found in the results does not allow a generalized assessment of the amount of species specific DNA used at the beginning. As the experiment based on quantified DNA, assessments under more realistic condictions, like using whole specimens, can only be assumed to vary even more due to large differences in the biomass of the different species. This means at least for PCR based methods that species abundances cannot be reliably assessed.

There is evidence that variation lowers when the amount of PCR cycles is reduced (Figure 4.8). But the evaluation of the non-ratio samples also showed that the number of species seuqences was very inconsistent and did not depend on the sample treatment but was rather species specific (Figure 4.9). However, both treatments T1 and T2 did make use of PCR although T2 only for applying sample indices to the sequences. Further experiments on the quantification of the abundance of species should therefore concentrate on the strict avoidance of PCR to make sure not to introduce any bias.

4.4.4 Recommendations

As the goal of this study is the development of workflows for high-throughput sample metabarcoding suggestions only concentrate on aspects relevant for such an approach. Of course it can be discussed that if the number of species influences the amount of chimeric sequences, samples could be previously sorted to reduce the number of species for each sample. But this would also decrease time and cost efficiency. Also it is

unclear what would be the appropriate number of species in a sample, and if the species within a sample should be as homogeneous as possible, as diverse as possible, and if they also should be sorted after size. For all these approaches reasonable arguments can be found in literature, either describing enhanced taxon recovery or the loss of data during Illumina sequencing (Krueger et al. 2011; Morinière et al. 2016; Elbrecht et al. 2017). Therefore the most convenient and certainly fastest way would be the evaluation of samples as a whole. This means without any previous sorting or splitting of the sample. As long as the DNA acquisition is done in a non-destructive way, samples can always be sorted afterwards or when DNA results imply the necessity for it. At least when species abundances are required samples should still be manually sorted.

Hybrid-enrichment is one of the most promising approaches to enrich target sequences like COI as it does not rely on the amplification through PCR. Original species abundances might be affected to a lesser extend as their sequence ratios are not biased through amplification due to the selective binding of primers. It can also be aspect that the number of chimeras should be on a constant low level due to lesser PCR cycles compared to common PCR based methods.

5 Empirical biodiversity assessment

5.1 Introduction

Biodiversity studies in local biotopes often collect thousands of specimens with Malaise traps even in temperate regions (Geiger et al 2016). The sorting and identification through morphological approaches usually occupies experts to capacity. With DNA barcoding as a method to objectively and rapidly identify species a new tool for assessing diversity has been developed (Hebert et al.2003a, 2003b; Hebert et al. 2009). Advancing and extending this approach to a workflow with (partly) automated steps will help to seize its great potential. Efforts to improve workflows should start at the very beginning, the capturing of what to be analyzed. The automation of a step this early has shown to save 80% and more of the costs incurred in sampling (Selby et al. 2014). Another benefit automated sampling is the standardization of the capturing process that helps to make future data originating from different studies more easily comparable. To make the most out of it the following processes should ideally also be standardized. This experiment therefore ecompasses a full workflow that starts with the devolpment of a capturing device with an integrated automated long-term interval-sampling device and ends with the evaluation of the collected data. The main aim is to demonstrate the general applicability of this workflow for future biodiversity research.

Collected samples were processed with latest hybridization based target enrichment strategies to significantly reduce biasing factors intrudued by using PCR based approaches documented in the previous chapter. Species identification was perfomed with the Bayesian classifier as it proved to be extremely fast and at the same time very reliable. It has shown that it is also significantly faster than current blast-based methods that are commonly used in environmental sequence surveys (Porter et al. 2014). Identifications that were obtained by this were then compared to morphological identifications by two taxonomists. But as as current sequence databases still lack large parts of the local diversity a MOTU (molecular taxonomic unit) based approach was also applied.

5.2 Material and Methods

5.2.1 The automated interval sampler (AIS)

5.2.1.1 Requirements and planning

The automated interval sampler was planned as a module, allowing combining it with different kinds of traps. This modular feature makes the interval trap very versatile as sampling criteria (like e.g temporal resolution) usually accordingly apply to many different trap types. Their further standardization by adding the interval sampling module can therefore improve the comparability of data.

The main focus during the construction was as put on the maintenance independence and reliability of the sampler while providing valuable data. To achieve this goal the design needed to fulfill the following criteria:

1. Capture in programmable intervals
2. Keep samples separated and distinguishable

3. Keep samples and DNA in good condition to be analyzed
4. Independent energy supply
5. Robustness for outdoor usage

Further desirable attributes taken into account were a certain compactness of construction, simplicity for the ease in setting it up and taking it down and a reasonable weight to allow transportation.

In this study the AIS was planned to be combined with a Malaise trap using ethanol in its sampling jars. This combination is particularly effective for collecting flying insects and their identification using molecular markers, since ethanol-preserved specimens are perfectly suited for DNA analysis (Malaise 1937; Hebert et al 2003b; Geiger et al. 2016 Aagaard et al. 2017).

5.2.1.2 Hardware solutions

Considering the criteria above, the basic design of the interval sampler resulted in a horizontal revolver mechanism for changing the storing-jars (Figure 5.2). Twelve storing-jars were chosen for collecting while an additional storing jar signals a halt position. This thirteenths bottle allows a defined beginning and ending of the collection for the twelve other jars. All thirteen bottles were filled with 96.3 % Ethanol. In the first its version (the prototype) were hung under a rotating plate with several passage holes fixed to a second plate on top of it leaving a single fixed passage to the collecting funnel-jar positioned on the top of the Malaise trap. The inital wooden plates were later exchanged for UV resistant acryl plates. The modification also changed from hanging the storing jars under a rotating plate to fixing them upon a rotating device to gain more distance between the rotating and the fixed plates. For controlling intervals and the motor driven rotation the programmable open-source platform Arduino Uno Rev3 was used. These micro controlled boards are able to read digital data inputs as from sensors, buttons or time clocks and turn them into outputs, like activating a motor, turning on an LED or saving data on a storage device. A solar charged battery provided the required energy. At the beginning the AIS was equipped with a 20W solar panel, combined with a charging regulator and a 12V 9Ah sealed lead-acid battery. A second panel was later added for forest usage. Each solar panel is attached with a 5 meter cable for variable installation capabilities in the surrounding. Three legs were fixed to the construction also allowing their leveling. The mechanisms, electronics and all bottles were put into an acrylic housing to restrict sight and secure it from jar removals (Figure 5.1).

Figure 5.1: From left to right: (1) Prototype version of the automated interval sampler (AIS) with a single hanging storing-jar and solar panel. (2) Inside of the acryl housing of the second version with storing jars with black sealing rings fixed upon on a rotating plate filled with ethanol. (3) AIS set up with Malaise trap in the Schorfheide-Chorin on plot SEW7 (Photos: Struwe 2015).

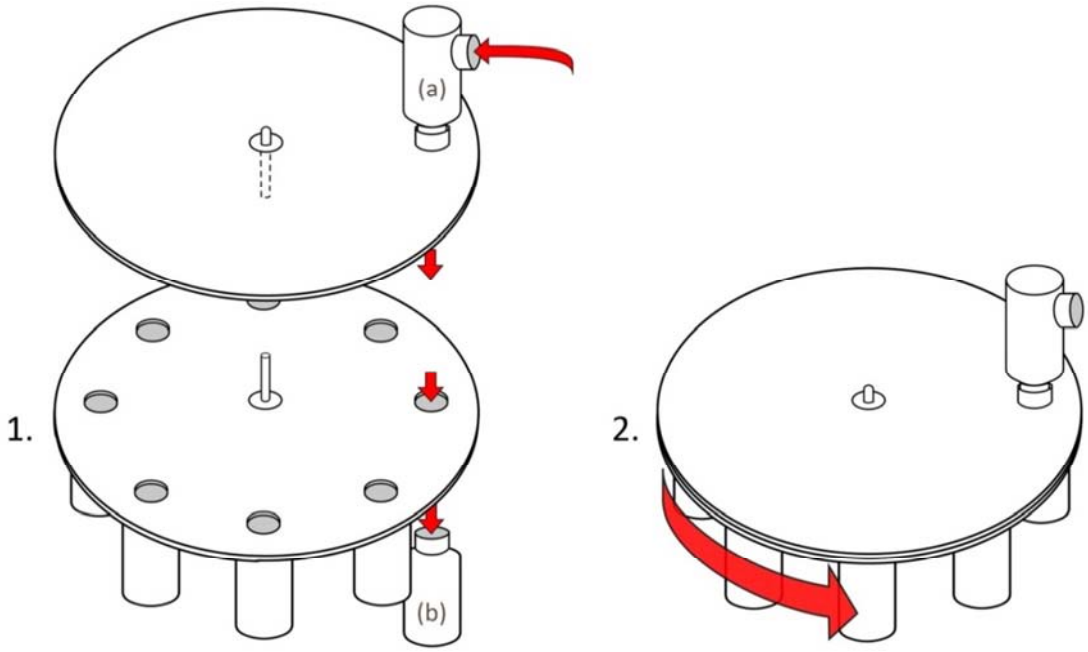


Figure 5.2: Concept of the revolver mechanism used in the automated interval sampler (AIS). 1. Insects are led by an adaptable trap into the capturing funnel-jar (a), from where they drop into the storing-jar (b) that is filled with high proof ethanol. 2. Storing-jars rotate in determined intervals for a temporal sample resolution.

5.2.1.3 Program functions

It is possible to program various capturing intervals. For the experiment the sampler was set to one week of capturing for every storing jar. Every time the holes of the top funnel-jar and the bottom storing-jar line up, the rotating mechanism stops for the chosen programmed interval, in this case seven days. After passing each of the 13 bottle-jars the last one signals a halt. Thus the seasonal transition of species can be documented. This basic modus can be alternated by e.g. changing the capturing time or the number of bottle-jar iterations. Shortening the capturing time to 12 hours and optionally skipping the thirteenth jar allows for example a separate capture of diurnal and nocturnal species over a defined time frame. As the Arduino Uno Rev3 has several possible data inputs a sensor extension is possible to react to temperature, light or for

example rain. By this the sampler could react to its surrounding and capture in flexible intervals.

5.2.2 Study area

5.2.2.1 Schorfheide-Chorin in general

The chosen region for a pilot study of automated biodiversity monitoring was the UNESCO Biosphere Reserve Schorfheide-Chorin (Figures 1.5 and 5.3). Its status originated in the national park program during the last days of the former German Democratic Republic in September 1990. It covers the districts of Uckermark, Barnim, Märkisch-Oderland and Oberhavel. In 2011 it became part of the UNESCO Biosphere Reserve program (Succow et al. 2012). Today about 35,000 people are living and working in the biosphere reserve. Populations dates back until the Stone Age. Germanic and Slavic tribes colonized the landscape. Since the 12th century progressive German colonization started to further shape the regional appearance. During the more recent history, through the times of the German Empire (1871-1918), the Third Reich (1933-1945) and the German Democratic Republic (1949-1990), the area was repeatedly used as hunting grounds for changing sovereigns. This circumstance protected large parts of the area from human influences and are a fundamental basis for the nature protection programs of the later years.

5.2.2.2 Geology and geological history

The regional landscape and different soil types (podzolised brown earth, lessivé, pararendzina, podzols, and bog soils) are the basis for diverse vegetation and were shaped during the last glacial maximum in Northern Europe (Figure 5.3). This period is known as the Weichselian glaciation and ranged from 115,000 to 10,000 years BP (before present). Its end about 10,000 years ago marked the beginning of the Holocene (Walker et al. 1994). During its late phase a sharp decline in temperature temporarily stopped the retreat of the glaciers in the Northern Hemisphere. The change was relatively sudden, taking place in decades, and led again to advances of the glacier. These late activities resulted in the characteristic geological and geographical elements found in this region (Heine et al. 2009). Ground and end moraines as glacial deposits are stretching in chains from North to South, flat sandurs and wide glacial valleys (as the thorn-eberswalder urstromtal) where the melting ice waters ran north are typical for this relatively young and structurally diverse landscape. Kames formed plateaus, while kettle holes formed by melting dead ice are adding up to the local landscape diversity (Weiße 1995). The melting of the remaining ice created the extraordinarily high number of lakes, bogs and water bodies which regulate the retention and filter capacity in the biosphere reserve. Nutrient rich alder bogs, reed beds and sedge-bogs but also nutrient poor, sour formations like peat moss-bogs and cotton grass-reed-vegetations cover approx. 10% of the total surface area. The 230 lakes have a total surface of over 90 km^2, not counting the more than 3,000 ponds and puddles. However almost no natural running waters can be found in the biosphere due to its water shed positioning between the North Sea and the Baltic Sea and the drainage of the large flood plain of the river Oder, involving avulsions of the meandering Oder and the construction of the Finow canal.

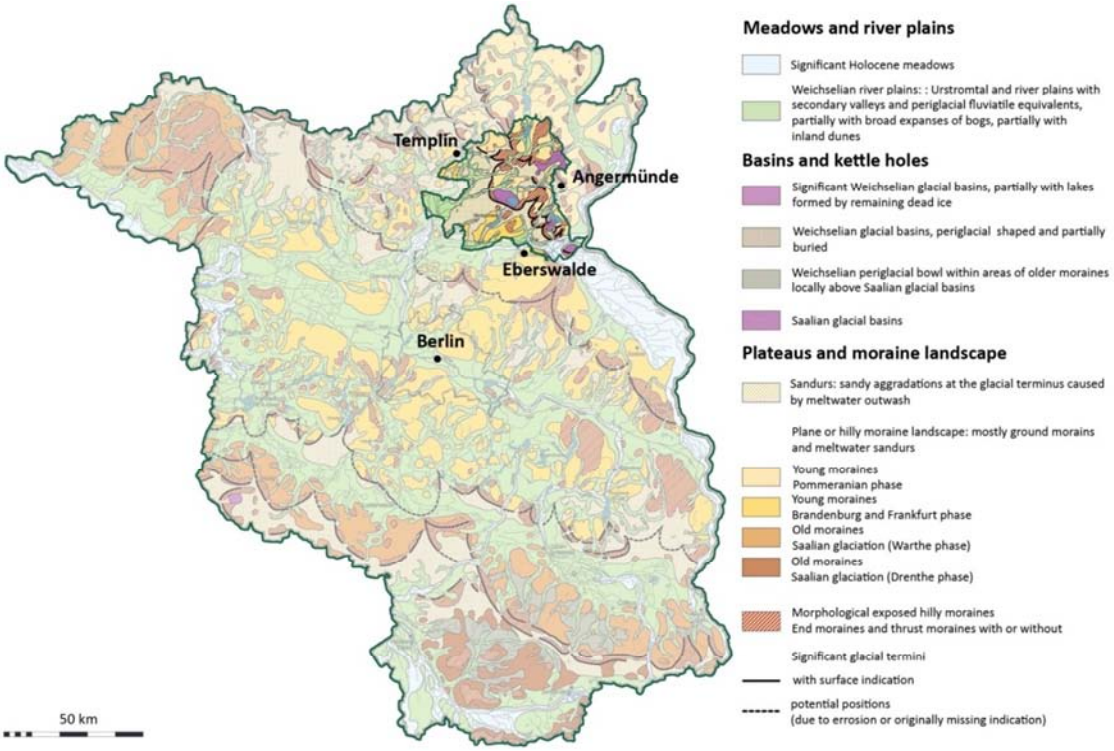


Figure 5.3: Location of the Schorfheide-Chorin Biosphere Reserve within the geological diverse landscape of Brandenburg in the north of Berlin (Modified after: Atlas zur Geologie von Brandenburg 4. Auflage 2010).

5.2.2.3 Climate

The eastern area of the Schorfheide-Chorin is situated in one of the driest regions of Germany and has a sub continental climate (Figure 5.4). The biosphere reserve usually experiences a rapid warming during spring followed by a relatively hot summer. Most of the annual precipitation accumulates in the central part of the biosphere reserve. Along the river Oder the sum of precipitation does not reach 500mm per year. The number of days with frost range between 75 and 110. The area around the Oderbruch usually has lesser days of freezing than the rest of the reserve. The number of days with a permanent temperature below 0°C range from 22 to 31 per year. Vegetation start is at the beginning of April, usually around the 4th and the 8th. On 30 to 40 days in the year the temperature can reach 25°C. On 3 to 8 days the temperature can go up to 30°C (Source: DWD – Deutscher Wetter Dienst, 1961-1990). However recent data over the past 30 years show a tendency to higher temperatures and less colder days (Source: DWD).

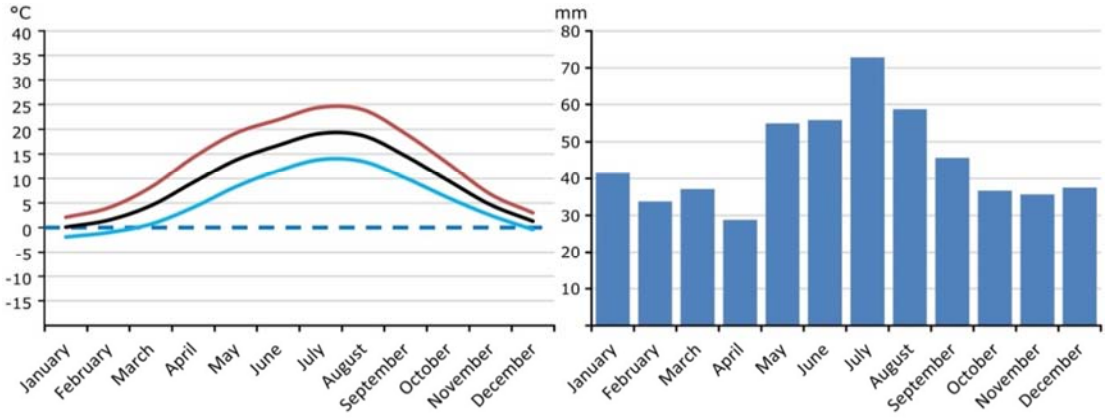


Figure 5.4: Regional weather statistic. On the left in red: Average max temperature per month within 24h. In blue: Average min temperature per month within 24h. In black: Average temperature. The graph on the right shows the precipitation amount during a month. Data measured in the period of 1992–2017 at the weather station in Angermünde. Source: WMO (World Meteorological Organization).

5.2.2.4 Forests

The biosphere reserve Schorfheide-Chorin contains 64.580 ha of woodland communities, from pine-monoculture (*Pinus sylvestris*) to natural alder marshlands (*Alnus glutinosa)*. The deciduous forests have been preserved almost in their entirety for several hundred years and now remained undisturbed for about 20 years. A significant element at the core of the Biosphere Reserve is the millet grass beech forest (*Fagus sylvatica*) in the territory of Grumsin. For the further preservation of its natural exceptional conditions it is part of the UNESCO World Heritage natural sites (BMUB 2015).

The largest parts of the exploratory are covered by beech forests and pine-beech-forests. In these mixed forests the top tree layer is usually represented by pines, whereas beech trees built up a dense middle layer. Hornbeam (*Carpinus betulus*), oak (*Quercus petraea*, *Q. robur*), maple (*Acer platanoides, A. pseudoplatanoides*) and sometimes small-leaved lime (*Tilia cordata*) are other species that can also be found in the beech forests. Its herbal layer consists of *Anemone nemorosa*, *Galium odoratum*, *Lamium galeobdolon*, *Viola reichenbachiana* and *Viola riviniana* as well as *Melica uniflora* and at poorer habitats *Deschampsia flexuosa*, *Carex pilulifera* and *Maianthemum bifolium*.

The west contains large areas of dry bilberry-pine-forest; the south of the exploratory is dominated by extended pine plantations. Raspberry, blackberry and bush grass are the dominant species of the undergrowth layer. Although some unmanaged forest stands exist in the core area, most areas are managed as age-class forests. This structure consists mainly of trees of the same age or by proportions of trees in different age classes.

Further forest stands of minor economic interest are built up by birch (*Betula pendula*) and ash (*Fraxinus excelsior*). Allochthonous trees in the Exploratories are *Picea abies*, *Quercus rubra*, *Pinus strobus*, *Larix decidua* and *Pseudotsuga menziesii*. In between the different forest types several lakes, ponds, marhes and bogs are scattered harbouring diverse wetland communities (*Carex disticha C.acutiformis, C. riparia and C. acuta Caltha palustris, Iris pseudacorus*, *Juncus effusus, Lychnis flos-cuculi, Phragmites communis*). Within and around the many forest stands repeated patches of grassland with grazing cattle and agroecosystems harboring cereals, potatoes and tobacco are

situated (United Nations 2003, Biodiversity Exploratories 2017, Biosphärenreservat Schorfheide-Chorin 2017).

Figure 5.5: Forests on peat are characteristic for the exploratory Schorfheide-Chorin. The picture shows a birch stand with scattered alder and two single oaks in the distance; situated at the edge of an unmanaged beech forest in the territory of Arnimswalde near the experimental plot SEW7 (Struwe 2015).

5.2.2.5 Sampling plots

Each Exploratory has a large number of predetermined research sites. These sites are classified into a hierarchic system of study plots differing in their intensity of investigation. The basis is formed by about 1,000 grid plots (GP) from which a primary inventory was made containing ecological and geological data. Number and abundance of plant species, land use types and intensity, as well as soil data were recorded. Based on these data one hundred experimental plots (EP) were selected to represent broad gradients of land use intensity ranging from near-natural, protected sites to intensively used ecosystems in grasslands and forests. The diversity of further taxa is being assessed. Even the microbial community and the diversity of mycorrhizal fungi is being assessed. Researchers are therefore asked to do most of their research on VIPs followed by EPs to be able to connect and relate as much data as possible. The chosen sampling sites for the automated insect sampler lies therefore mainly in the very intensive plots. Although the main goal is using VIP plots, not all plots are accessible at all times as many of them are private property. Hunting can be a reason that sometimes prohibits the access to suitable VIP plots.

Besides these internal restrictions the study plots were chosen according to the following principles. A primary objective was that (1) the chosen study plots and plot types allow a comparison as they should have overlapping and opposing attributes. (2) The chosen types and their correlating plots should also have two resembling replicates. (3) The distance of all the study plots to each other ought to ensure that the sites do not interact or influence each other in a significant way. A further desirable characteristic was the avoidance of disturbance (trap removal, rambler, etc.). Grassland areas were excluded from the outset as they already have several unfavorable characteristics regarding the planned research activities. All taller installed instruments have a wide visibility. The installation would be exposed to disturbances by regular mowing and/or the grazing of cattle that could require taking down the installation. High temperature differences due to the exposure to the sun could cause the heating of the electronics and

sampling fluids. These differences in temperature could lead to DNA damage of the specimens caught and to an increased evaporation of the sampling fluid. The chosen study sites therefore focused on forest plots. In detail typically managed pine forest, and widely unmanaged beech forest were chosen (Figures 5.6 and 5.7; Table 5.1).

Figure 5.6: Typical pine (left) and beech plot (right) at the end of April with Malaise trap and automated interval sampler with installed solar panels (experimental plots SEW2 and SEW9; Struwe 2015).

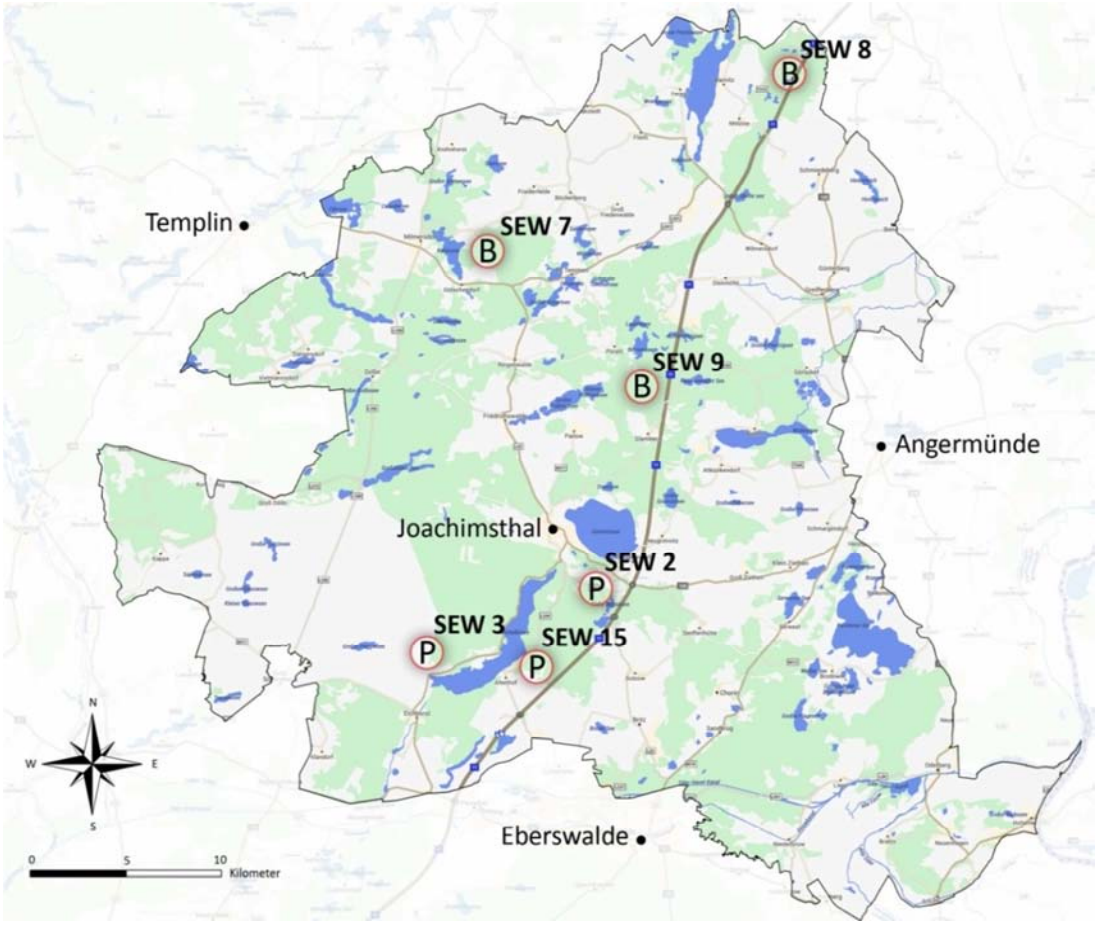


Figure 5.7: Sampled plots in the Schorfheide-Chorin Biosphere Reserve. P – managed pine forest, B – unmanaged beech forest.

Plot	Tree type	Land use	(WRB) Soil Type	Latitude	Longitude	(MASL) Height	Status	Territory
SEW2	Pine	Managed	Cambisol	52.952281	13.778288	62 m	FFH	Voigtwiese
SEW3	Pine	Managed	Cambisol	52.920645	13.642759	79 m	NSG2	Eichheide
SEW7	Beech	Unmanaged	Cambisol	53.107932	13.694059	104 m	NSG1	Arnimswalde
SEW8	Beech	Unmanaged	Albeluvisol	53.191725	13.929809	105 m	NSG1	Melzow
SEW9	Beech	Unmanaged	Cambisol	53.044157	13.810056	58 m	NSG1	Glambeck
SEW15	Pine	Managed	Cambisol	52.914054	13.737783	92 m	LSG	Altenhof

Table 5.1: Location and description of the sampled forest plot types. WRB: World Reference Base for Soil Resources. MASL: Meters above sea level. FFH: Nature reserve after FFH-policy (Flora-Fauna-Habitat). NSG: nature reserve, category: 1 & 2. LSG: protected landscape.

5.2.3 Preparations for a genetic biodiversity assessment

5.2.3.1 Sample preparation and extraction

From all samples taken on the study plots in the Schorfheide-Chorin four sampling bottles, two for each forest type, covering captures from the same sampling interval were chosen to be compared (Table 5.2). By this the samples covered the same sampling period regarding capturing date and capturing length. Also both plot types had to be covered by the same interval to allow comparisons. Although this setup is very basic it was chosen as a demonstration of the concept and to prove the informative content of samples being evaluated on a DNA basis.

Sampling Plot	Type	Trees	Start (dd.mm.yy)	End (dd.mm.yy)	Duration (d)	Indexing
SEW2	Managed	Pine	30.05.2015	06.06.2015	7	SI + DI
SEW3	Managed	Pine	28.05.2015	04.06.2015	7	SI + DI
SEW7	Unmanaged	Beech	27.05.2015	03.06.2015	7	SI + DI
SEW9	Unmanaged	Beech	27.05.2015	03.06.2015	7	SI + DI

Table 5.2: Analyzed samples from four of the sampled plots within the Schorfheide-Chorin. Two parallel sets were chosen to compare the results of a single- versus double-indexing approach. d: day; m: month; y: year; SI: single-indexing; DI: double-Indexing. The main tree type of the plot is given in the column "Trees", the duration of the sampling interval with its start end end dates is given in the corresponding collumns.

In a first step the sampling bottles from SEW2, SEW3, SEW7 and SEW9 were manually sorted for Nematocera and Syrphidae by taxonomic experts (Björn Rulik, ZFMK Bonn and Ximo Mengual, ZFMK Bonn). All identifications and the number of specimen were entered into a species list. All specimens were then transferred back into the bottles except for one male specimen of *Macrobrachius kowarzi* from sample 7. This specimen was used to extend the sequence reference list and was transferred into the GBOL workflow and database. After the identification the whole content of each bottle was dispersed using the Ultra Turrax T18 (IKA®-Werke GmbH & Co. KG, Staufen, www.ika.com). The rotational speed was 25.000 min-1 for 60 seconds followed by 90 seconds of 15.000 min-1. The dispersing tool was extensively cleaned after each sample preparation.

Cleaning: The disperser ran two times for 1 min at 4.000 min-1 in fresh H_2O each and one time at 12.000 min-1 in fresh H_2O. After that it was rinsed under running H_2O. It then ran for 30 sec. in C_2H_6O. It was then rinsed under C_2H_6O. It was again rinsed under H_2O after it had been taken apart. Its parts were then put under UV for 5 min, then autoclaved at 121°C and again put under UV for 5 min.

Because of the large quantity of caught specimen subsamples were taken from the bottles after homogenization to make sure to obtain an adequate representation of the initial sample composition. For this the dispersed samples were initially stirred up

and while the suspension sat down again 4 subsamples were taken. The procedure was then repeated. The first two subsamples of a bottle had a volume of 1,5 µl, whereas the following two hat a volume of 1,0 µl. By this 16 subsamples were created. The supernatant was then evaporated at 35°C in vacuo for about 1:45 h and 1:30 h, respectively. The remaining material was then measured using the Sartorius MC1 LC 6200 S (Sartorius Werkzeuge GmbH Co. KG, Ratingen, www.sartorius.de). Measurement table can be found in the supplement S3.1.

Extraction was conducted with the Qiagen DNeasy Blood & Tissue Kit (Qiagen N.V., Venlo, www.qiagen.com). The detailed laboratory protocol for the extraction can be found in the supplement S3.2. DNA concentration of the extractions were then measured using the Promega Quantus Fluorometer (Promega Corporation, Fitchburg, www.promega.com). Identical subsamples were combined for greater DNA quantity. Protocol and table in the supplement (S3.3 and S3.4). DNA integrity was checked using the AATI Fragment Analyzer (AATI Inc., Ankeny, www.aati-us.com).

5.2.3.2 Library preparation, hybrid enrichment and sequencing

The general preparation steps for the analysis were: (i) fragmenting/sizing the sequences to the desired length, (ii) converting into double-stranded DNA, (iii) attaching oligonucleotide adapters and (iv) sample indices to the fragments ends, (v) hybridizing and (vi)capturing of the targeted fragments, (vii) quantitating the final library product for sequencing.

Fragmentation (i)

To generate an appropriate fragment length for sequencing the sequences were cut enzymatically. Fragmentation was done with Fragmentase for dsDNA (NEB, Ipswich, www.neb.com). (Protocol in supplement S3.5)

Purification and concentration leveling (i)

Afterwards the samples were purified with Ampure XP Beads (Agilent Technologies, Santa Clara, www.agilent.com). (Protocol in supplement S3.6)

After measuring the DNA concentration the corresponding samples were homogenized and split to ensure equal DNA distributions in the both indexing approaches.

End repair (ii)

Subsequently the overhangs of the sequence fragments are converted into blunt ends by an end repair step. For this oligos are aligned to the open ends. (Protocol in supplement S3.7) After this DNA concentration was measured again. (Table in supplement S3.8)

A – tailing (ii)

A single A-nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the later adapter ligation reaction. A corresponding single T nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. (Protocol in supplement S3.9)

Adaptor ligation (iii)

The Agilent SureSelect platform for library preparation was used for adapter ligation of all single indexing samples. (Protocol in supplement S3.10)

All double indexing samples needed to be purified first. Here Ampure Beads were used again in a 1:1 concentration. All steps from the previous purification protocol (S3.6) were followed until step 10. Here only 30 µl dH2O were added. In step 13 only 28 µl of the eluate were taken out for the prepared tubes. For adapter ligation of the double indices NEBNext reagents were used. (Protocol in supplement S3.11)

Library preparation (iv)

The sample specific indices were then bound to the sequences for single- and double-indexing. (Protocols in supplement S3.12 and S3.13)

Baits hybridization (v)

All samples were pooled in Lo-Binding tubes according their indexing type and their forest plot. For single-indexing pool SI1 (pine samples) and SI2: (beech samples) and for double-indexing DI1 and DI2 accordingly. The samples were mixed thoroughly and vaporized to a volume of 3,5 µl. (Protocol in supplement S3.14)

Capturing the hybridized DNA (vi)

Streptavidin beads were used for capturing the hybridized DNA fragments. (Protocol in supplement S3.15)

Amplifying the captured libraries (vii)

Amplification proceeded with the minimum number of cycles possible. S1 and S2 needed 12 cycles to generate a sufficient library size, whereas D1 and D2 only needed 9 cycles. (Protocol in supplement S3.16) This resulted in 16 samples for the single indexing approach and 16 samples for the double indexing approach with sufficient DNA concentration for sequencing. (Table in supplement S3.17)

5.2.3.4 Sequencing

Sequencing took place at StarSeq GmbH in Mainz, Germany (www.starseq.com), on an Illumina MiSeq V3 using 1 flow cell. The requested sequencing type was 300 nt paired-end read. For Sequencing the pools SI1 and SI2 were combined, so were the pools DI1 and DI2. Data was pre-sorted by StarSEQ GmbH according the sample indices.

5.2.4 Data Processing

5.2.4.1 Initial quality filtering

Data was initially checked visually using fastqc v 0.11.15 (Andrews et al 2011). For adapter clipping and first quality filtering Trimmomatic v.0.36 was used (Bolger et al. 2014). This included the removal of the remaining Illumina TrueSeq adapters, leading and trailing low quality or N bases were also removed when the phred score was below 3. A 4-base sliding window then scanned the read and cut it when the averaged quality dropped below phred score 20. Reads shorter than 100 bases were removed. (Script in supplement S3.18)

5.2.4.2 Advanced data processing

Mothur v.1.39.5 was then used to generate contig sequences from the read pairs R1 and R2 (Schloss et al. 2009). For quality filtering all sequences containing ambiguities were removed, so were sequences containing more than 10 homopolymers (e.g. *Epidapus microthorax*: 10 x G) and sequences that did not have a minimum overlap of at least 50 bases. Duplicate sequences were merged for more efficient computation. For the same

reason a count table was generated containing the names of the unique sequences, the groups and the quantity of unique sequence in each group. The query sequences were then aligned to a reference alignment. The template sequence for a query sequence is searched for using a kmer search with 8mers to find the best match. A pairwise alignment is then created using the Needleman-Wunsch method with a reward of +1 for a match and penalties of -1 for a mismatch and -2 for a gap. Sequences that start at or before position 190 and end at or after position 468 were removed to make sure all sequences overlap in the same region. Columns that only contain gap characters (i.e. '-' or '.') were ignored to accelerate the calculation of distances. If processing generated new duplicate sequences through removing characters the duplicates were merged again. Single-linkage preclustering was then applied to remove sequencing errors (Huse 2010). Potentially chimeric sequences were identified using the uchime algorithm and a reference template (Edgar et al. 2011). The reference template contained about 48,000 sequences from about 10,000 species with taxonomic descriptions. It was created from GBOL database that at this time encompassed also the initial Diptera list. The reference is referred to as the "GBOL-reference". Identified chimeric sequences were removed from the data. After this the sequences were sorted into a 3'- or a 5'-subset according to their alignment position. For this the alignment data of the sequences were used to sort the sequences either into the 3' half or the 5' half of COI according to the region in which >50% of the COI fragment was positioned. By this the richness estimates were improved by not counting short sequences representing the same species, but located in contrary regions of the COI sequence, twice.

After this sequences were identified using a naïve Bayesian classifier looking at the query sequence kmer by kmer, calculating the probability the query sequence would be in the given template taxonomy based on the kmers it contains. The assignment was then checked by a bootstrapping algorithm to find the confidence limit of the assignment (Wang et al. 2007; Cole et al. 2014; Porter et al. 2014; Vinje et al. 2015).

Two clustering algorithms were chosen to assign the sequences to MOTUs (Rognes et al. 2016). Average neighbor (also: average link) clustering, was chosen as it consistently produces high quality MOTUs. Abundance based greedy clustering (agc) was chosen as it performs similar well concerning quality and stability by using less computational resources (Westcott and Schloss 2015; Westcott and Schloss 2017). Both algorithms were set to cluster at a 3% level. The average link clustering had to be aborted after 2 months because of its intensive memory use. The assigned 12 processors (2.67 GHz Intel® Xeon®) and a maximum of 128 GB showed to be insufficient for the task. For all further steps MOTUs from the agc clustering were used. Spurious MOTUs were removed following the recommondations for processing NGS data (Edgar 2010, http://drive5.com/usearch/manual/singletons.html; Edgar 2013; Edgar et al. 2015).

MOTU taxonomy was assigned using the consenus taxonomy of the given MOTU. For a taxon to be included at least 51% of the sequences had be identically classified at that level in the given MOTU. (Script in supplement S3.19)

After this a subset was generated containing only MOTUs that were classified to originate from Diptera taxa for further analyses.

5.2.4.3 MOTU alpha diversity analysis

Rarefaction curves were calculated to determine if the sampling effort was sufficient for an accurate assessment on MOTU richness and diversity. The rarefaction curves were generated for the 3'- and 5'-sorted subsets of the single- and double-indexed sequences to describe the number of MOTUs observed as a function of sampling effort for each of the described subsets.

As an objective measure of the degree of sample completeness the sample coverage was calculated. Originally it is the proportion of the total number of individuals in an assemblage that belong to the species represented in the sample. Here it was used to calculate the MOTU coverage instead of species coverage. The coverage **C** was calculated as the Good's coverage estimator which is defined as:

$$C = 1\text{-}\frac{F_1}{N}$$

where $\boldsymbol{F_1}$ is the number of singleton MOTUs (i.e., the number of MOTUs with only a single sequence) and $\boldsymbol{N}$ is the total number of all sequences.

For an estimation of MOTU richness the Chao1 estimator $\mathbf{M_{chao}}$ was calculated. Originally intended as a lower bound the Chao1 estimator has proven to yield good estimations for the total richness in an observed community (Chao 1984; Shen et al. 2002; Chao and Chen 2003). $\mathbf{M_{chao}}$ is defined as:

$$M_{chao} = M_{obs} + \frac{F_1^2}{2F_2}$$

where $\mathbf{M_{obs}}$ is the number of MOTUs in the sample, $\mathbf{F_1}$ is again the number of singletons (i.e., the number of MOTUs with only a single sequence) and $\mathbf{F_2}$ is the number of doubletons (the number of MOTUs with two sequences). To also use the information of higher frequencies (F_3 to F_{10}) the Abundance-based Coverage estimator (ACE) was also calculated and compared. It is used when the sample is expected to be more heterogeneous (Chao and Lee 1992; Chao and Chen 2003; Chao and Chiu 2016). $\mathbf{M_{ACE}}$ is defined as:

$$M_{ACE} = M_{abun} + \frac{M_{rare}}{C_{rare}} + \frac{F_1}{C_{rare}}\,\gamma^2_{rare}$$

where $\mathbf{M_{abun}}$ is the number of abundant MOTUs (i.e., the number of MOTUs with >10 sequences) and $\mathbf{M_{rare}}$ the number of rare MOTUs (i.e., the number of MOTUs with ≤10 sequences). $\mathbf{C_{rare}}$ measures the sample completeness of the subsample restricted to rare MOTUs where N is in this case defined as N_{rare}, the number of all rare MOTU sequences from M_{rare}. This means C_{rare} calculates the proportion of all sequences in M_{rare} that are not singletons. The value for $\boldsymbol{\gamma^2_{rare}}$ denotes the square of the estimated coefficient of variation (CV). The CV parameter is used to characterize the degree of heterogeneity among MOTU abundances. Abundant MOTUs carry almost no information about undetected MOTUs it is therefore more efficient to restrict the formula to the rare species group. The larger the value for CV becomes the greater will be the degree of

heterogeneity. The CV vanishes if and only if all MOTUs have the same abundances (i.e., the assemblage is homogeneous).

$$\gamma^2_{rare} = max\left\{\frac{M_{rare}}{C_{rare}}\frac{\sum_{i=1}^{10} i(i-1)F_i}{(N_{rare})(N_{rare}-1)} - 1{,}0\right\}$$

The formula for M_{ace} is undefined when all rare MOTUs are singletons ($F_1 = N_{rare}$, yielding $C_{rare} = 0$). In such a case literature advises to compute the bias-corrected form of Chao1 instead (Gotelli and Colwell 2011). (Software script in supplement S3.19)

5.2.4.4 MOTU beta diversity

To compare the shared richness among the samples 4-way Venn diagrams were analyzed for all possible different groupings of forest type, indexing approach and 3‘- 5‘-sorting.

Data was rarefied to a common number of sequences to calculate the Yue & Clayton index of dissimilarity of the different samples including the proportions of both the shared and non-shared MOTUs in each sample (Yue et al. 2005). Similarity heatmaps and a tree diagrams were used to describe the dissimilarity among the sample grouping. For the tree diagram the groups were clustered using the UPGMA algorithm using the distance between the samples as calculated from the Yue Clayton distance matrices describing the similarity of the samples.

To examine diversity further it was looked for general patterns across the samples of forest types. This means instead of viewing the sample as representing the community, it was viewed as having being generated by sampling from the community. A probabilistic modelling approach was used to cluster the different samples into communities. The data is modelled as an increasing number of communities and determines groups of communities with a similar composition based on Dirichlet multinomial mixtures as described by Holmes et al. 2012. A Laplace approximation estimates the evidence of the complete model. This means firstly, if a mixture of Dirichlets is more appropriate than a single Dirichlet prior for this data set and secondly, the number of components of the mixture. More specifically this means that a minimum Laplace value indicates the most probable number of community components. No rarefaction was needed as the method allows for clustering from samples with different numbers of sequences.

An analysis of molecular variance (AMOVA) was applied to test whether the genetic diversity within the pine and beech community is not significantly different from that which would result from pooling the two communities. The significance of the variance is tested using a permutational approach, eliminating the normality assumption that is conventional for analysis of variance but inappropriate for molecular data (Excoffier et al. 1992; Stewart and Excoffier 1996; Schloss 2008).

A test of homogeneity of molecular variance (HOMOVA) was performed to determine whether the amount of genetic variability in each forest type is significantly different from each other (Stewart and Excoffier 1996; Schloss 2008). (Software script in supplement S3.19)

5.2.4.5 Population-level analysis

An algorithm for the discovery of biomarkers was used to test whether there are any MOTUs that are differentially represented between the samples from pine forest and beech forest. The linear discriminant analysis (LDA) effect size (LEfSe) method supports high-dimensional comparisons with a focus on metagenomic analyses. LEfSe determines the MOTUs most likely to explain differences between the two forest types by coupling a Kruskal-Wallis sum-rank test for significant differential abundances (Kruskal and Wallis 1952) and a Wilcoxon rank-sum test as standard tests for statistical significance (Wilcoxon 1945; Mann and Whitney 1947) to test biological consistency with an additional LDA for effect relevance (Segata et al. 2011). (Software script in supplement S3.19)

5.3 Results

Single-indexing (SI) and double-indexing (DI) of the four Malaise trap samples had considerable differences in their read outcome after quality filtering. While starting with about the same amount of raw reads after sequencing, the number of DI sequences severely exceeded the number of SI sequences after processing (Table 5.3).

After sorting the sequences of each sample according to their positioning in relation to the 3' and 5' end of the COI sequence, it showed that the 3' and 5' subsets also had uneven sequence distributions (Table 5.4). The 3'-sorted subsets of a sample revealed to always contain more different sequences than the associated 5'-sorted subsets of the same sample. For SI the 5'-sorted subset contained 49% less sequences than the 3'-sorted subset from the same sample. For DI the 5'-sorted subset ocontained 37% less sequences than the 3'-sorted subset. From these fewer sequences also followed fewer MOTUs. The 5'-sorted subset had 31% less MOTUs than the 3'-sorted subset in the SI approach and 28% less MOTUs in DI approach. From the lesser MOTUs in the 5'-sorted subset also fewer species were identified; 13% less for SI and 5% less for DI than in the 3' sorted subsets of a sample. The observed differences in sequence and MOTU number between the 3'- and 5'-sorted subsets of a sample were generally higher in the SI approach than in the DI approach.

However, differences between the subsets and the sequencing approaches reduced towards the end of data processing when the MOTUs were filtered for Diptera sequences and assigned to species names. The mean length of the evaluated Diptera sequences was 338 bp with a standard deviation of 50 bp.

	SAMPLE	SEW2	SEW3	SEW7	SEW9	Total
(1)	SI raw reads	5,429,858	5,710,292	5,551,034	5,963,800	22,654,984
	DI raw reads	6,388,915	4,957,474	7,408,627	4,600,385	23,355,401
(2)	SI trim-paired	310,811	445,001	615,712	748,659	2,120183
	DI trim-paired	1,279,858	1,752,511	1,635,222	1,853,249	6,520,840
(3)	SI processed	33,293	44,212	93,146	123,600	294,251
	DI processed	242,832	445,568	420,499	299,548	1,250,063

Table 5.3: Total number of unfiltered (1) raw reads after sequencing, (2) sequences after initial filtering and read pairing, (3) sequences filtered for e.g. quality and chimeras before 3' 5' splitting and MOTU clustering.

SAMPLE		SEW2		SEW3		SEW7		SEW9	
Subset		3’	5’	3’	5’	3’	5’	3’	5’
SI	Seq	20,265	13,028	26,429	17,783	58,172	34,974	75,634	47,966
DI	Seq	142,384	100,448	287,184	168,384	247,812	172,687	174,594	124,954
SI	MOTU	898	610	1,100	741	996	685	1,482	1,047
DI	MOTU	1,772	1,305	2,164	1,508	1,495	1,035	1,940	1,388
SI	Species	137	115	152	121	124	105	175	153
DI	Species	173	161	183	172	125	116	194	180
SI	unclass	34	29	40	30	28	22	38	38
DI	uclass	56	41	48	41	35	31	48	47

Table 5.4: Total number of sequences, MOTUs and species before Diptera filtering sorted for their positioning on the 3’ and 5’ end on the COI. MOTUs based on agc clustering (abundance based greedy clustering) Species identifications based on Bayesian classifier. SI unclass and DI unclass are the number of MOTUs that could not be identified down to species level.

5.3.1 MOTU analysis - alpha diversity

5.3.1.1 Rarefaction curves

Rarefaction curves were calculated after filtering for Diptera sequences to determine if the sequencing effort was sufficient for covering Diptera MOTU diversity (Figure 5.8). The generally early terminations of the rarefaction curves for single-indexed samples (SI) show that the number of sequences obtained is relatively low to cover the diversity of Diptera MOTUs in both subset sortings (3’ and 5’). Double-indexed samples (DI) generally yielded more sequences and therefore also better approximations. Although the DI rarefaction curves are still on the edge of flattening out (and higher MOTU numbers can be expected with further sequences) sequencing output would have had to be increased severely to be able to detect a considerable number of further MOTUs.

The lowest number of sequences was found for the single-indexed 5’ subsample of SEW3 and the highest number for double-indexed 3’subsample of SEW7. The highest number of MOTUs per sequences was obtained by the double-indexed 3’ subsample of SEW3. A similar high number of MOTUs was reached by the double indexed 3’ and 5’ samples of SEW7 but with considerably more sequences. The lowest number of MOTUs was found in single-indexed 5’ subsample of SEW2 (Figure 5.8 and Table 5.5).

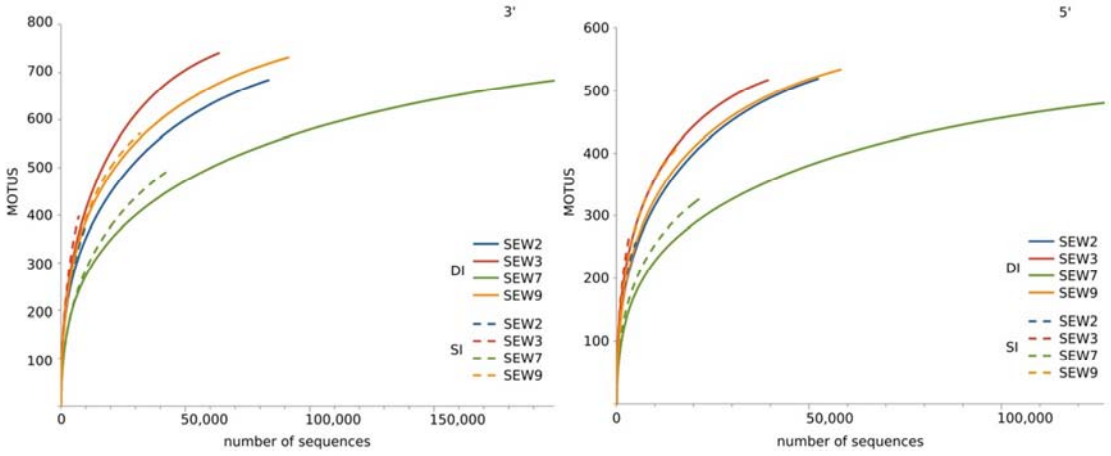


Figure 5.8: Rarefaction curves relating the sequencing effort to the number of MOTUs in the different SEW samples for single- and double-indexing approache and for the 3’ and 5’ sample subsets. X: Number sampled sequences Y: Number of MOTUs.

5.3.1.2 Coverage, richness and diversity

The Good’s coverage index calculates how much of the MOTUs in a sample are approximately covered during sequencing. MOTU coverage was calculated based on the

sequences that were identified to originate from Diptera. For all samples and subsets MOTU coverage can be considered high. Values generally ranged between 97% and 99% (Table 5.5).

The Chao1 richness estimates and Shannon diversity indices showed similar patterns between the single- (SI) and double- (DI) indexing approaches and the 3'- and 5'-sorted subsets. However, the 5' sorted subsets produced in general lower richness and diversity values. For the Chao1 estimator the highest MOTU richness was found in the double-indexed 3'-subset pine plot sample SEW3. A little lower than the SEW3 samples was the richness estimate in the beech plot samples from SEW9. Again a little lower was the richness found in the sample from the second pine plot SEW2. The second beech plot sample SEW7 had considerably lower richness values than all other plots with the lowest in the single-indexed 5'-subset. The results for the ACE richness estimator generally followed the pattern of the Chao1 estimator while having slightly higher values (Table 5.5).

The highest diversity was again found in the double-indexed 3'-subset pine plot sample from SEW3. The plot with the next highest diversity values was in contrast to the results of the richness estimate the second pine plot SEW2. Then the beech plots followed with SEW9 and SEW7. Thus the patterns found between the richness and diversity values slightly differed, whereas the plots with highest (SEW3) and lowest (SEW7) values stayed the same for the richness and diversity estimates (Table 5.5).

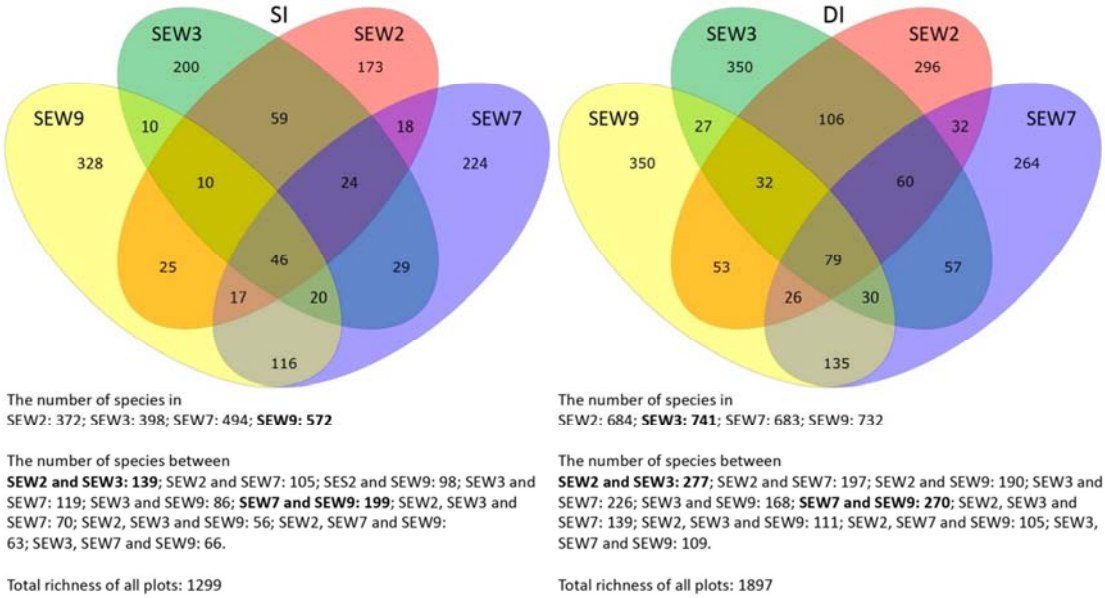
SAMPLE		SEW2		SEW3		SEW7		SEW9	
Subset		3'	5'	3'	5'	3'	5'	3'	5'
SI	Seq	9,549	4,955	7,148	3,550	43,145	22,301	31,779	16,137
DI	Seq	83,469	52,362	63,527	39,394	198,155	126,638	91,458	58,289
SI	MOTUs	372	258	398	277	494	330	572	412
DI	MOTUs	684	519	741	517	683	480	732	534
SI	Coverage	0.98	0.98	0.98	0.97	0.99	0.98	0.98	0.97
DI	Coverage	0.98	0.98	0.98	0.97	0.99	0.98	0.98	0.98
SI	Chao1	503.12	230.03	547.01	277.00	407.63	172.62	528.33	247.45
DI	Chao1	478.49	339.63	581.41	375.79	363.97	238.11	506.53	342.64
SI	ACE	629.16	403.07	698.61	518.88	509.93	316.67	630.50	436.70
DI	ACE	542.97	426.89	728.62	425.86	371.60	260.18	574.78	412.96
SI	Shannon	4.06	4.17	4.24	4.43	2.90	2.90	3.81	3.77
DI	Shannon	4.41	4.21	4.54	4.38	2.83	2.65	3.94	3.65

Table 5.5: Data overview after filtering for Diptera sequences. Results were sorted for the positioning of the original sequence reads in relation to the 3'- and 5'- end of COI. The table shows the sequence numbers, the corresponding numbers of Diptera MOTUs after agc clustering, the sample coverage and the Chao1 richness estimator, the ACE richness estimator, and the Shannon diversity index for single indexing approach (SI) and the double indexing approach (DI).

5.3.2 MOTU analysis - beta diversity

The Venn diagrams showed that each sample contains MOTUs that are unique to it and fractions of MOTUs that are shared with other samples (Figure 5.9 and supplement S3.20 - S3.23). The results of the samples differing only in the indexing-approach shared most of the presented MOTUs. Double-indexing samples revealed to encompass largest parts of the single-indexing MOTU diversity ($\tilde{x}$: 95.5 % for 3'- and 5'-sorting) while also containing further MOTUs that are exclusive to the double-indexing samples (Figures in supplement S3.20 - S3.23). Single- and double-indexing evaluation both showed that samples taken from plots sharing the same forest type (pine: SEW2-SEW3, beech: SEW7-SEW9) share larger proportions of MOTUs with each other than with samples that belong to a different forest type (Table 5.6). The median percentage of shared

MOTUs within the same forest type ranges between 24% for single- and 25% for double-indexing. Across forest types the median percentage ranges between 12% for single- and 17% for double-indexing. The percentage of shared MOTUs between three different plots lies between 6% and 9% with the most MOTUs shared between the plots SEW2 - SEW3 - SEW7.



The number of species in
SEW2: 372; SEW3: 398; SEW7: 494; **SEW9: 572**

The number of species between
SEW2 and SEW3: 139; SEW2 and SEW7: 105; SES2 and SEW9: 98; SEW3 and SEW7: 119; SEW3 and SEW9: 86; **SEW7 and SEW9: 199**; SEW2, SEW3 and SEW7: 70; SEW2, SEW3 and SEW9: 56; SEW2, SEW7 and SEW9: 63; SEW3, SEW7 and SEW9: 66.

Total richness of all plots: 1299

The number of species in
SEW2: 684; **SEW3: 741**; SEW7: 683; SEW9: 732

The number of species between
SEW2 and SEW3: 277; SEW2 and SEW7: 197; SEW2 and SEW9: 190; SEW3 and SEW7: 226; SEW3 and SEW9: 168; **SEW7 and SEW9: 270**; SEW2, SEW3 and SEW7: 139; SEW2, SEW3 and SEW9: 111; SEW2, SEW7 and SEW9: 105; SEW3, SEW7 and SEW9: 109.

Total richness of all plots: 1897

Figure 5.9: The Venn diagram for 3'-sorted MOTUs shows that the samples SEW2 – SEW3 and SEW7 – SEW9, which also share the forest type (pine – pine and beech – beech) also share more MOTUs than any other sample combination in both sequencing approaches (SI: single-indexing; DI: double-indexing). (Further diagrams in supplement S3.20 - S3.23)

	Shared MOTUs within forest type		**Shared MOTUs across forest type**			
Samples	**SEW2-SEW3**	**SEW7-SEW9**	**SEW2-SEW7**	**SEW2-SEW9**	**SEW3-SEW7**	**SEW3-SEW9**
3' SI	22 %	23 %	14 %	12 %	15 %	10 %
5' SI	26 %	25 %	12 %	12 %	14 %	12 %
Median SI		**24%**			**12 %**	
3' DI	24 %	24 %	16 %	13 %	18 %	16 %
5' DI	26 %	25 %	18 %	18 %	19 %	14 %
Median DI		**25%**			**17%**	

Table 5.6: Percentage of shared MOTUs between the different samples of single and double indexing and 3'- 5' sorting.

Values from the Yue and Clayton measure of dissimilarity also show that samples from identical plots, only differing in the indexing approach (SI vs DI), were again indicated most similar to each other (Yue et al. 2005). Also similar are the samples that originated from the same forest types (pine: SEW2-SEW3, beech: SEW7-SEW9). Least similar are the samples that originate from different forest types. These results were confirmed for both of the 3'- and 5'-sorted MOTU subsets of the sample and are visualized in the similarity heatmaps (Figure 5.10).

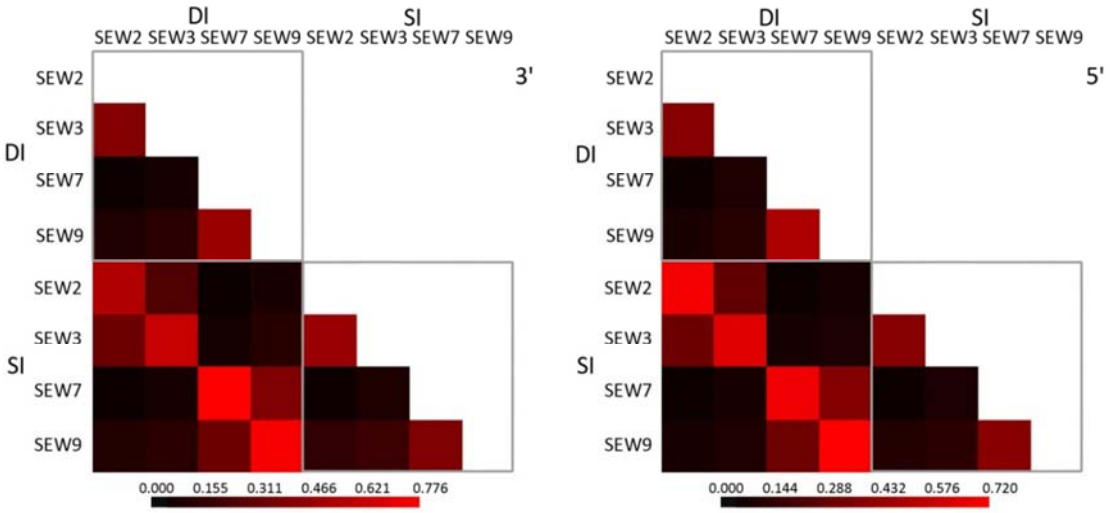


Figure 5.10: Similarity heatmaps for 3'-sorting (left) and 5'-sorting (right). Red colors indicate higher similarity of sequence communities. The Yue and Clayton θ index is a similarity index that includes species proportions of both the shared and non-shared species in each population. The index ranges from 0 to 1, with 1 = complete similarity and 0 = complete dissimilarity (Yue et al. 2005). Results from the same samples only differing in the indexing approach are most similar (e.g. SI SEW7 and DI SEW7) followed by samples sharing the same forest type (e.g. for pine: DI SEW3 and DI SEW2).

A UPGMA clustering of the Yue & Clayton values describes the sample groupings in a tree. Corresponding single- and double-indexing samples are closest and cluster together followed by the forest type clustering. The different forest types are distantly positioned to each other.

Dirichlet multinomial mixtures (DMM) were used to determine groups of communities (Holmes et al. 2012). The DMM community types affirm the separation of the forest type samples into separated communities and point to further subdivision of the beech plot community based on the 3'-sorted MOTUs. However the Laplace values for finding the correct community resolution are very close for K = 2 and K = 3 in both findings. The theta values for the 3' data (θ = ❶: 295.39, ❷: 1184.11, ❸: 1717.16) indicate a highly variable cluster ❶ and two more homogeneous clusters ❷ and ❸. Theta value for 5' data (θ = ❶: 247.55, ❷:199.98) indicate that here ❶ is the more homogeneous cluster and ❷ the variable cluster.

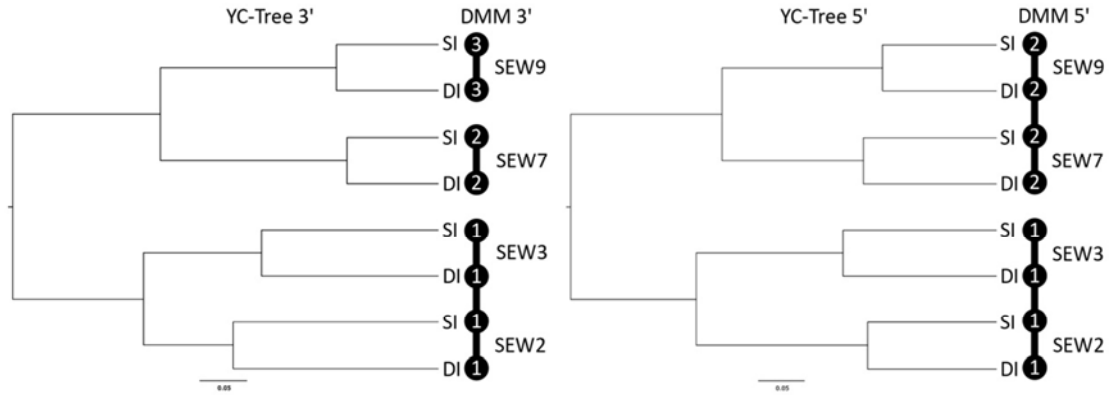


Figure 5.11: UPGMA clustered data of the Yue & Clayton values describe the sample groupings for 3' and 5' sorting. The DMM community types are based on the Dirichlet multinomial mixtures and affirm the individual grouping of the two forest types (pine: SEW2 – SEW3, beech: SEW7 – SEW9) and indicates a possible further subdivision of the beech plot community according to the 3'-sorting.

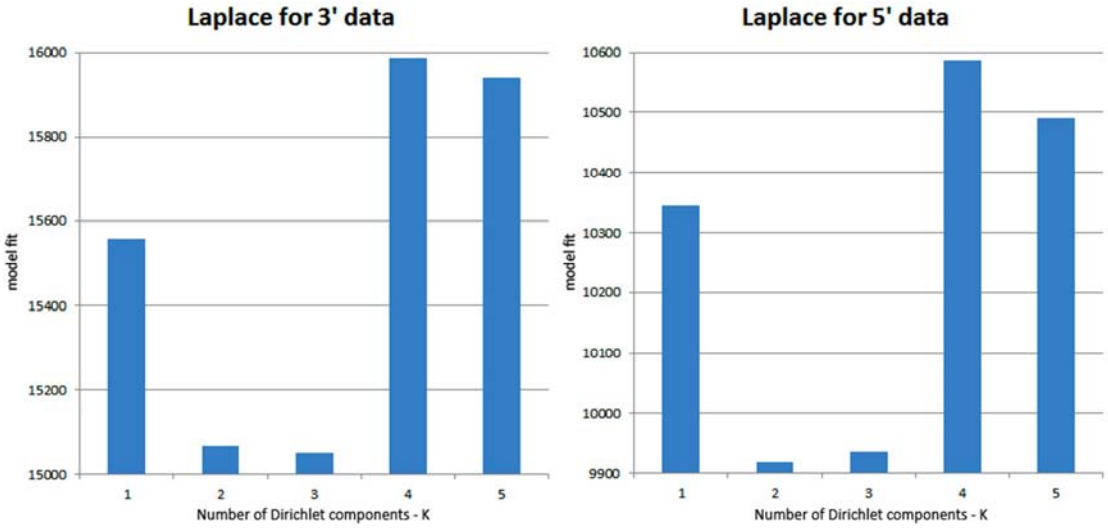


Figure 5.12: Model fit for mixture of Dirichlets prior to dataset. Evaluates model fit for increasing number of Dirichlet mixture components using the Laplace approximation to the negative log model evidence. For the 3' data the minimum lies at K=3 and for the 5' data at K=2.

The analysis of molecular variance (AMOVA) affirms the significantly different diversity structure of the pine and beech MOTU communities (3'p=0.0288, 5'p=0.0295). This means the genetic diversity within each community is significantly different from the average genetic diversity of both communities pooled together. However, the HOMOVA revealed that the amount of genetic variabilty is the same in the pine and beech communities. (3'p = 1, 5'p = 1).

5.3.3 Population level analysis

An algorithm for the discovery of biomarkers (LEfSe: Segata et al .2011) was used to test whether there are any MOTUs that are differentially represented between the samples from the pine and beech plots. The corresponding species classifications identified by the Bayesian classifier (Wang et al. 2007) were then assigned to the MOTUs found to be biomarkers. The discriminative species which have been identified for pine and beech plots are presented in the following tables (Table 5.7 and 5.8). It becomes apparent that most species share a xylobiont or saproxylic life stage. Most species also share a preference for some sort of moist condition ranging from damp habitats to habitats with bogs or other waterbodies. In the pine stands species with preferences for coniferous trees and deciduous trees can be found. In the beech stands species are rather restricted to deciduous trees. In both habitats are also species that are strongly associated with cattle indicating the presence of pastures in the vicinity of the stands. A full list of all identified species can be found in the supplement (S3.24).

Recognized possible biomarkers for the pine plots: SEW2, SEW3

Family	Species	Occurrence, habitat	Cited from
Agromyzidae	*Chromatomyia milii*	Milium is the eponymous main hostplant of the species that grows in **wet to mesic woodlands**; also associated with **flatwoods** in upland moraines	(Thompson 1980)
Anthomyiidae	*Botanophila fugax*	In various woodland areas all with **deadwood** including **spruce** and open and **damp habitat**	(Bartak 1998)
Anthomyiidae	*Emmesomyia grisea*	In **spruce plantations**	(Bartak 1998)
Anthomyiidae	*Hydrophoria linogrisea*	mixed areas with **deadwood** including **spruce** plantation and **damp habitat**	(Bartak 1998)

Cecidomyiidae	*Bryomyia apsectra*	Fungivorous. Larvae associated with **dead wood**. Known as a saproxylic insect	(Jaschhof 1998; Jaschhof et al. 2014)
Conopidae	*Myopa buccata*	Various **woods** and wood edges	(Stuke 2003)
Empididae	*Rhamphomyia longipes*	Known from **riverine forest** areas	(van der Weele et al. 2017)
Fanniidae	*Fannia sociella*	In **rotten wood** or beneath moss or bark on dead or dying **trees,** in **sandy** and **humus soil**	(Rozkošný et al. 1997)
Limoniidae	*Epiphragma ocellare*	Associated with streams, riverbanks and **wetland habitats** with the appearance of **dead wood** with known in occurrences in **alder, beech** and also **coniferous stands**	(Reusch et al. 2009; Reusch and Weber 2013)
Limoniidae	*Neolimonia dumetorum*	Larvae in **deadwood** of various **deciduous trees**	(Reusch and Weber 2013)
Muscidae	*Haematobosca stimulans*	Often found in the **presence of cattle**. Species is associated with **cow dung** for larval development. Bloodsucking, **parasiting cows** and mammals	(Ball 1984)
Muscidae	*Helina impuncta*	Associated with **cow dung** and **woodlands**	(Skidmore 1985)
Muscidae	*Mydaea nebulosa*	Associated with **forests,** also **cow dung** and **rotten elm wood**; larvae of several M. species develop primarily in fungi (e.g. *Suillus bovinus*, *Boletus pinophilus* in **coniferous** stands)	(Laux 1985; Skidmore 1985; Gminder et al.2000)
Muscidae	*Polietes lardarius*	P. species are found in **cow dung**	(Smith 1989)
Phaeomyiidae	*Pelidnoptera fuscipennis*	Larvae parasitic to Diplopoda. Associated with open **forests** and **watersides**.	(Merz and Kofler 2006)
Rhagionidae	*Rhagio annulatus*	Occurs at **forest edges** and is **water associated**	(Stuke 2010)
Sarcophagidae	*Sarcophaga subvicina*	Familiy well represented in **wetland habitats**; S. species can be associated with **bog-inhabiting plants,** where they consume dead insects trapped within the cups of the plants; Sarcophaga have larvae that that feed on the soupy mixture of insects and liquid that accumulates at the bottoms of the plant cups	(Dahlem 1991; Keiper et al. 2002)
Sarcophagidae	*Sarcophaga variegata*	Familiy well represented in **wetland habitats**; S. species can be associated with **bog-inhabiting plants,** consuming dead insects trapped within the cups of the plants	(Dahlem 1991; Keiper et al. 2002)
Scathophagidae	*Scathophaga furcata*	Ubiquitous with a tendency to nutritious and **wet places**	(Stuke and Schacht 2009)
Scathophagidae	*Scathophaga stercoraria*	Can be found anywhere at places associated with **cow dung**	(Stuke and Schacht 2009)
Sphaeroceridae	*Spelobia parapusio*	S. species are associated with **decaying plants** or **fungi,** in **wetlands** or close to **waterbodies**	(Bährmann et al. 2013)
Stratiomyidae	*Clitellaria ephippium*	Larvae in nests of *Lasius fuliginosus*. Ant nests primarily in standing **tree trunks** and **rotting stumps**.	(Wilson 1955; Haupt and Haupt 1995)
Syrphidae	*Blera fallax*	associated tree species **Pinus sylvestris** in microhabitats: tree hole, heartrot	Rotheray et al. (2001)
Syrphidae	*Chrysotoxum cautum*	In **deciduous forests** and **forest edges** and flowering Fabacea	(von der Dunk 2005)
Syrphidae	*Criorhina asilica*	Criorhina mainly select ovipositiosition sites in **dead wood** on exposed tree roots	(Warren et al. 1991)
Syrphidae	*Dasysyrphus tricinctus*	**Wet meadows** and **forest edges**	(von der Dunk 2005)
Syrphidae	*Didea fasciata*	Didea fasciata stronly associated with ancient **forest sites**	(Warren et al. 1991)
Syrphidae	*Epistrophe nitidicollis*	At **forest edges**	(von der Dunk 2005)
Syrphidae	*Rhingia campestris*	Common. Found at **pastures** and **waysides**	(von der Dunk 2005)
Syrphidae	*Syrphus ribesii*	At **meadows**	(von der Dunk 2005)
Syrphidae	*Temnostoma bombylans*	At **clearings** and **wet hardwood** (e.g. Salix)	(von der Dunk 2005)
Tabanidae	*Hybomitra bimaculata*	Especially attracted to **cows** (urine) and present in **woodlands**.	(Krcmar et al. 2006)
Tachinidae	*Peribaea fissicornis*	Parasitoid of winter moths and noctuids. Winter moths on **deciduous trees** and **shrubs** but also	(Glavendekić et al. 2006)

		found on **conifers** (Larch).	
Tipulidae	*Tanyptera atrata*	**Decaying sapwood**	(Rotheray et al. (2001)

Table 5.7: Excerpt of discriminative species recognized as biomarkers for the sampled pine plots from all genetically identified Diptera species (LEfSe: Segata et al. 2011). (Full species lists in supplement (S3.24)

Recognized possible biomarkers for the beech plots: SEW7, SEW9

Family	Species	Occurrence, habitat	Cited from
Cecidomyiidae	*Aprionus cardiophorus*	Larvae **mycetophag** on **deadwood** of **deciduous trees**	(Jaschhof 2009)
Ceratopogonidae	*Serromyia femorata*	S. is associated with **bogs, fens, wet meadows**, streams or small rivers. All species are **restricted to wooded regions.** Larvae of *S. femorata* in mosses at lake margins and mud associated with **marshlands**	(Borkent and Bissett 1990)
Dolichopodidae	*Dolichopus claviger*	*D. claviger* in **moist woodland** habitats, leaf dwelling. In general D. species prefer humid to moist conditions. Species exhibit a pronounced habitat affinity, the family as such serves well as bio-indicator or in site quality assessments, in particular of humid biotopes. Species can be found in large numbers especially in **humid forests**, **humid heathland**, **saltmarshes**, **dune slacks** and **waterbodies**.	(Pollet et al. 2003).
Dolichopodidae	*Dolichopus festivus*		
Dolichopodidae	*Dolichopus nigricornis*		
Dryomyzidae	*Dryomyza decrepita*	The adults are frequently found in **forests** in shaded **humid places** with low vegetation, searching for flowing sap	(Rozkošný 2006)
Empididae	*Empis aestiva*	Emergence often in dry **heathland** and less common near woods and ponds. Adults common at **ponds** and rare in heathland.	(Delettre et al. 1997)
Empididae	*Hilara interstincta*	Emergence often in dry **heathland** and less common near woods and ponds. Adults common at **ponds** and rare in heathland.	(Delettre et al. 1997)
Lauxaniidae	*Lyciella platycephala*	Most species associated with deciduous **forests**, **undergrowth** and **shrubs**. Larvae feed mainly on **decomposing leaves** of trees.	(Miller 1977; Keiper et al. 2002; Merz and Kofler 2008)
Limoniidae	*Austrolimnophila ochracea*	Species typical of **woodland**, **marsh fauna**. Feed on decayed **dead beech wood**.	(Service 1973; Hövemeyer and Schauermann 2003; Krivosheina 2009)
Mycetophilidae	*Leia crucigera*	Most L. species are mycetophagous, associated with fungi, either fruiting bodies or mycelia in **dead wood** and **soil litter** in **deciduous forests.**	(Økland et al. 2008; Jakovlev 2014)
Mycetophilidae	*Mycetophila fungorum*		
Mycetophilidae	*Phronia basalis*		
Phoridae	*Megaselia pleuralis*	Common. Associated with **marsh** and **riverine forests**. Megaselia can indicate disturbance or stress. Species benefit from habitat heterogeneity followed by various disturbance sources like cutting, wildfires or wind throw.	(Prescher et al. 1994; Brenner 2004; Durska et al. 2010)
Rhagionidae	*Ptiolina obscura*	Stem borer of moss. Associated with **moist habitats, deciduous forests** and moss.	(Hardy and McGuire 1947)
Rhagionidae	*Rhagio scolopaceus*	Common in **moist forest habitats**, adults often found on trunks.	(Haupt and Haupt 1995)
Rhagionidae	*Rhagio strigosus*	In **moist forest habitats.**	(Haupt and Haupt 1995)
Sciaridae	*Cratyna perplexa*	Strongly associated with **deciduous forests** with **beech** stands (Fagus sylvatica).	(Menzel and Schulz 2007)
Sphaeroceridae	*Spelobia palmata*	Sphaerocerids are known for **wetlands**, **peatlands** and **forest habitats**. Spelobia occur in wetland habitats where they feed on **dung**. Associated with pasturing **cattle**. Sometimes	(Hafez 1939; Coffey 1966; Buck 1994; Keiper et al. 2002; Papp 2002)

		associated with deer and fox feces.	
Syrphidae	*Temnostoma vespiforme*	Associated with **forests**. Often on **decaying birch** stems. Adults visiting flowers.	(Stammer 1933; Hauser et al. 1996)

Table 5.8: Excerpt of discriminative species recognized as biomarkers for the sampled beech plots from all genetically identified Diptera species (LEfSe: Segata et al. 2011). (Full species lists in supplement (S3.24)

5.3.4 Comparing the morphological and genetical determinations

5.3.4.1 Comparing total species numbers

From all samples 54 Diptera species were identified by two taxonomic experts; Björn Rulik from the ZFMK for fungus gnats and Ximo Mengual also from the ZFMK for hover flies (Table 5.9). From the identifications to species level based on morphological characters nine identifications could not be confirmed based on the sequencing approaches used (Table 5.10). However, all together 245 Diptera species could be discerned based on their DNA sequences. 17 of these identifications lead to unique database identifiers as the specimen the reference sequence originated from (voucher) was awaiting identification when its sequences was already integrated into the reference list. Linnean names were assigned to 228 of the DNA based identifications. Of these 130 belonged to the single-indexed samples (SI) and 227 to the double-indexed samples (Table 5.11).

For single-indexing (SI) on average 86 species were found per sample. For double-indexing (DI) on average 96 species were found per sample (including identifications with unique identifiers). Morphologically (MI) on average only 24 species were identified per sample (Table 5.12). The lower identification success based in morphological characters is partially based in the fact that several specimens could only be identified down to genus or even family level. For all those unresolved taxa sequencing revealed a much more detailed resolution by distinguishing between several species within these unresolved genera and families. For the family of Sciaridae that was not further resolved based on morphology 36 species from 12 genera were identified based on their DNA. The previous unresolved family of Cecidomyiidae was genetically resolved to 15 species from 10 genera. Within the family of Psychodidae 8 species from five genera were could be identified. Within the family of Limoniidae six species from six genera were identified. Within Chironomidae three species from three genera were identified. Within Ceratopogonidae two species from two genera were identified. Within Culicidae two species from two genera were identified. Moreover, for several further families additional species and genera could be identified based on DNA sequencing (Table 5.11).

In contrast to this DNA identification also assured previously uncertain species numbers from the morphological approach as for example in the family of Keroplatidae. Here it was confirmed that there are indeed no more than the four morphologically identified species in total, although several specimens of the family of Keroplatidae could not be identified further than to family or genus level based on morphological characters (Table 5.10 and 5.11).

DNA sequencing also identified further species belonging to Diptera families that were previously missed in the morphological approach because the expertise of the taxonomists was limited to the 16 identified families. Nevertheless 27 further Diptera families were identified using DNA sequencing yielding a total of 43 Diptera families based on DNA sequencing.

For 9 different species and one specimen from the genus Sciophila (Mycetophilidae) that were found by the taxonomists that could not be verified genetically, actually no or only sparse sequence data was present in the reference library used for sequence identification (Table 5.10). A problematic case is *Melanostoma scalare* that has 23 representative specimens in the reference list based on GBOL data. However, these speciemens all share an identical barcode reducing the actual reference to a single sequence for this species. Moreover it is known that large parts of *Melanostoma mellinum* populations share their COI barcode with their close relative *M. scalare*, making individuals affected by this condition genetically indistinguishable if only COI is used for their identification. As *M. mellinum* sequences were found in the samples, and the reference library also contains *M. mellinum* references identical to *M. scalare*, a distinction was not possible for these species. The same problem also accounts for *Acnemia amoena,* where an identical barcode for *A. nitidicollis* was discovered in the reference library. Finally a single female specimen was morphologically identified belonging to the genus Sciophila (Mycetophilidae). This was not verified genetically although 6 Sciophila species with differing sequences were present in the reference library (Table 5.9, 5.10, and 5.11).

The results of DNA identification can be considered a success despite the above mentioned missing species recoveries. Even when limiting the evaluation of species number identifications to a subset containing only families that had already been identified based on morphological traits, it becomes apparent that the sequencing approach still provides a higher resolution than the morphological approach (Table 5.12). (Full species list of all species identified based on their DNA in supplement S3.24)

5.3.4.2 The single- and double indexing performance and 3'-5'- subsets

Most species were identified from the 3'-subset of the double-indexed samples (DI, 3'-subset) followed by the 5'-subset (DI, 5'- subset). Comparing the species numbers in the Diptera families a difference of 1-4 species was found between the two sortings (Table 5.11). The single-indexing approach yielded generally fewer species numbers than the DI approach. Again more species were found within the 3'-subset (SI, 3'-subset) than in the 5'-subset (SI, 5'-subset). The differences in species number within the Diptera families between the two subsets were with 2-12 species more dissimilar than in the double-indexing approach. In general the congruence between 3'- and 5'- sorted subset identifications ranged between 75% - 80% for double-indexing and between 64% - 75% for single-indexing (Table 5.11). However, morphological identifications (MI) always yielded the lowest species numbers with less than half the species identified in the double-indexing approach.

Altogether the double-indexing approach performed best. It showed more species congruence between its two subsets (3' and 5') and yielded the highest species resolution for all samples through its higher species numbers.

Morpholgical determinations		**Forest type:**	**pine**		**pine**		**beech**		**beech**	
		Sample:	**SEW2**		**SEW3**		**SEW7**		**SEW9**	
Family	**Genus**	**Species**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
Anisopodidae	**unknown**	**unknown**	0	0	**1457**	**224**	0	0	0	0
Cecidomyiidae	**unknown**	**unknown**	**470**	**132**	**24**	**37**	**112**	**61**	**150**	**93**
Ceratopogonidae	**unknown**	**unknown**	**1**	**5**	**14**	**23**	**7**	**14**	**345**	**748**
Chaoboridae	*Chaoborus*	*obscuripes*	0	0	0	0	0	0	**5**	0
Chaoboridae	**unknown**	**unknown**	0	0	0	0	0	0	0	**16**
Chironomidae	**unknown**	**unknown**	**7**	**15**	**4**	**3**	**11**	**71**	**106**	**93**
Culicidae	**unknown**	**unknown**	0	0	0	0	**1**	0	0	**5**

Diadocidiidae	*Diadocidia*	*ferruginosa*	**2**	0	**3**	0	0	0	**4**	0
Diadocidiidae	*Diadocidia*	*spinosula*	**5**	0	**1**	0	0	0	0	0
Diadocidiidae	**unknown**	**unknown**	0	**3**	0	**3**	0	0	0	**2**
Ditomyiidae	*Ditomyia*	*fasciata*	0	0	0	0	0	0	**1**	0
Ditomyiidae	*Symmerus*	*annulatus*	0	0	0	0	0	0	**1**	0
Keroplatidae	*Keroplatus*	*testaceus*	**1**	0	0	0	0	0	0	0
Keroplatidae	*Orfelia*	*fasciata*	**1**	0	0	0	0	0	**4**	0
Keroplatidae	*Orfelia*	*nemoralis*	0	0	**4**	0	0	0	**1**	0
Keroplatidae	***Orfelia***	***sp.***	0	0	0	0	0	**1**	0	0
Keroplatidae	*Urytalpa*	*dorsalis*	0	0	**2**	0	0	0	0	0
Keroplatidae	**unknown**	**unknown**	0	0	0	**3**	0	0	0	**1**
Limoniidae	**unknown**	**unknown**	0	**14**	**15**	**4**	**1**	**2**	**7**	**6**
Mycetophilidae	*Acnemia*	*amoena*	0	0	**1**	0	0	0	0	0
Mycetophilidae	*Acnemia*	*nitidicollis*	0	0	0	0	**2**	0	**1**	0
Mycetophilidae	*Cordyla*	*insons*	0	0	**1**	0	0	0	0	0
Mycetophilidae	*Ectrepesthoneura*	*hirta*	0	0	**2**	0	0	0	0	0
Mycetophilidae	*Leia*	*subfasciata*	0	0	0	0	0	0	0	**2**
Mycetophilidae	*Macrobrachius*	*kowarzi +*	0	0	0	0	0	**1**	0	0
Mycetophilidae	*Monoclona*	*rufilatera*	0	0	0	0	0	0	**2**	0
Mycetophilidae	*Mycetophila*	*fungorum*	0	0	0	0	**1**	0	**2**	0
Mycetophilidae	*Mycetophila*	*ocellus*	0	0	0	0	0	0	**1**	0
Mycetophilidae	*Mycetophila*	*perpauca*	0	0	0	0	0	0	**1**	0
Mycetophilidae	*Mycomya*	*fimbriata*	0	0	**2**	0	0	0	0	0
Mycetophilidae	*Mycomya*	*parva*	0	0	0	0	0	0	**1**	0
Mycetophilidae	*Phronia*	*basalis*	0	0	0	0	**2**	0	**4**	0
Mycetophilidae	***Phronia***	***sp***	0	0	0	0	0	**2**	0	0
Mycetophilidae	***Sciophila***	***sp.***	0	0	0	0	0	**1**	0	0
Mycetophilidae	*Tetragoneura*	*sylvatica*	0	0	**27**	0	0	0	**1**	0
Mycetophilidae	***Tetragoneura***	***sp.***	0	0	0	0	0	**6**	0	0
Mycetophilidae	**unknown**	**unknown**	0	**9**	**15**	**25**	0	0	**1**	**12**
Psychodidae	**unknown**	**unknown**	**4**	**4**	**7**	**23**	**4**	**7**	**84**	**71**
Sciaridae	**unknown**	**unknown**	**13**	**16**	**7**	**14**	**36**	**49**	**80**	**58**
Syrphidae	*Blera*	*fallax*	**1**	**1**	0	0	0	0	0	0
Syrphidae	*Chrysotoxum*	*cautum*	0	**1**	0	0	0	0	0	0
Syrphidae	*Chrysotoxum*	*festivum*	**1**	**1**	0	0	0	0	0	0
Syrphidae	*Criorhina*	*asilica*	**1**	0	0	0	0	0	0	0
Syrphidae	*Dasysyrphus*	*tricinctus*	0	**1**	0	0	0	0	0	0
Syrphidae	*Didea*	*intermedia*	0	**1**	0	0	0	0	0	0
Syrphidae	*Epistrophe*	*nitidicollis*	0	0	0	**1**	0	0	0	0
Syrphidae	*Epistrophella*	*euchroma*	0	0	0	**1**	0	0	0	0
Syrphidae	*Melanostoma*	*scalare*	0	**2**	0	0	0	0	0	0
Syrphidae	*Meligramma*	*triangulifera*	0	0	0	**1**	0	0	0	0
Syrphidae	*Pipiza*	*noctiluca*	0	**1**	0	0	0	0	0	0
Syrphidae	*Rhingia*	*campestris*	**1**	**1**	0	**1**	0	0	0	0
Syrphidae	*Syrphus*	*ribesii*	**1**	0	0	0	0	0	0	0
Syrphidae	*Temnostoma*	*bombylans*	0	**1**	0	**1**	0	**1**	0	0
Syrphidae	*Temnostoma*	*vespiforme*	0	0	0	0	0	**1**	0	0
Tipulidae	***Nephrotoma***	***sp.***	**1**	0	0	0	0	0	0	0
Tipulidae	*Tanyptera*	*atrata*	**1**	0	**2**	0	0	**1**	0	0
Tipulidae	*Tipula*	*irrorata*	0	0	**1**	0	0	0	0	0
Tipulidae	*Tipula*	*selene*	0	0	0	0	0	0	**1**	0
Tipulidae	**unknown**	**unknown**	0	**3**	**3**	**1**	0	0	0	**3**

Table 5.9: Diptera male (M) and female (F) species identifications based on morphological traits (identifier: Björn Rulik, Ximo Mengual). Marked in bold: Identifications that did not reach species level.

Family	Genus	Species	comment
Mycetophilidae	*Cordyla*	*insons*	Not in reference library
Syrphidae	*Meligramma*	*triangulifera*	Not in reference library
Tipulidae	*Tipula*	*irrorata*	Not in reference library
Chaoboridae	*Chaoborus*	*obscuripes*	One sequence in reference library
Syrphidae	*Epistrophella*	*euchroma*	One sequence in reference library
Syrphidae	*Melanostoma*	*scalare*	One sequence in reference library (also identical to *M. mellinum*)
Syrphidae	*Pipiza*	*noctiluca*	Two sequences in reference library
Mycetophilidae	*Ectrepesthoneura*	*hirta*	Two sequences in reference library
Mycetophilidae	*Acnemia*	*amoena*	Three sequences in reference library (one identical to *A. nitidicollis*)
Mycetophilidae	*Sciophila*	*sp.*	Six different sequences in refrence library

Table 5.10: Morphologically identified Diptera which could not be verified genetically were absent or rare in the reference database or problematic because of their identical barcodes with different species.

			SI								DI							
			3'- subset SEW				5'- subset SEW				3'- subset SEW				5'- subset SEW			
Family	**Genus**	**Species**	**2**	**3**	**7**	**9**	**2**	**3**	**7**	**9**	**2**	**3**	**7**	**9**	**2**	**3**	**7**	**9**
Agromyzidae	*Agromyza*	*pseudoreptans*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Agromyzidae	*Chromatomyia*	*milii*	0	X	0	0	0	X	0	0	X	X	0	0	X	X	0	0
Agromyzidae	*Chromatomyia*	*obscuriceps*	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Agromyzidae	*Liriomyza*	*taurica*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Anisopodidae	*Sylvicola*	*cinctus*	0	0	0	0	0	0	0	0	X	X	0	X	0	X	0	0
Anthomyiidae	*Botanophila*	*fugax*	X	X	0	X	X	0	0	0	X	0	0	0	X	0	0	0
Anthomyiidae	*Emmesomyia*	*grisea*	X	X	0	0	X	X	0	0	X	X	X	0	X	X	0	0
Anthomyiidae	*Hydrophoria*	*lancifer*	X	X	X	X	0	0	0	0	X	X	X	X	X	X	X	X
Anthomyiidae	*Hydrophoria*	*linogrisea*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Anthomyiidae	*Hylemya*	*nigrimana*	0	X	X	X	0	X	X	X	0	X	X	0	0	X	X	0
Anthomyiidae	*Mycophaga*	*testacea*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Anthomyiidae	*Pegomya*	*silacea*	0	0	0	X	0	0	0	X	0	0	0	X	X	0	0	X
Calliphoridae	*Bellardia*	*viarum*	X	0	X	X	X	X	X	X	X	0	0	X	X	X	0	X
Cecidomyiidae	*Aprionus*	*cardiophorus*	0	0	X	X	0	0	0	X	0	0	X	X	0	0	X	X
Cecidomyiidae	*Aprionus*	*similis*	0	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Cecidomyiidae	*Aprionus*	*spiniger*	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0
Cecidomyiidae	*Asynapta*	*pectoralis*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0
Cecidomyiidae	*Bryomyia*	*apsectra*	X	X	X	0	0	0	0	0	X	X	X	X	0	0	0	0
Cecidomyiidae	*Camptomyia*	*abnormis*	0	0	0	0	0	0	0	0	X	X	0	0	0	X	0	0
Cecidomyiidae	*Camptomyia*	*spinifera*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Cecidomyiidae	*Divellepidosis*	*hypoxantha*	0	0	0	0	0	0	0	0	X	X	0	0	X	X	0	0
Cecidomyiidae	*Lestremia*	*leucophaea*	0	0	0	0	0	0	0	0	0	0	X	X	0	0	X	X
Cecidomyiidae	*Peromyia*	*perpusilla*	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0
Cecidomyiidae	*Porricondyla*	*fulvescens*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Cecidomyiidae	*Winnertzia*	*curvata*	0	0	0	0	0	0	0	0	X	X	0	X	0	0	0	0
Cecidomyiidae	*Winnertzia*	*tridens*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	0
Cecidomyiidae	*Winnertzia*	*xylostei*	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0
Cecidomyiidae	*Xylopriona*	*atra*	0	0	X	0	0	0	X	0	0	0	X	X	0	0	X	X
Ceratopogonidae	*Atrichopogon*	*lucorum*	0	0	X	X	0	0	0	X	0	X	0	X	0	X	X	X
Ceratopogonidae	*Serromyia*	*femorata*	0	0	X	X	0	0	X	X	X	0	X	X	X	0	X	X
Chaoboridae	*Chaoborus*	*flavicans*	0	0	X	X	0	0	X	X	0	0	0	X	0	0	0	X
Chironomidae	*Limnophyes*	*angelicae*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Chironomidae	*Micropsectra*	*atrofasciata*	0	X	0	0	0	X	0	0	0	X	0	X	X	X	0	0
Chironomidae	*Xenopelopia*	*nigricans*	X	0	0	X	0	0	0	X	X	0	0	X	X	0	0	X
Chloropidae	*Chlorops*	*rossicus*	0	X	0	0	0	X	0	0	0	X	0	X	0	X	0	0
Chloropidae	*Dicraeus*	*styriacus*	0	0	X	0	0	0	0	0	0	0	0	0	0	0	0	0
Chloropidae	*Oscinella*	*frit*	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0
Conopidae	*Myopa*	*buccata*	0	X	0	0	X	X	0	0	0	X	0	0	0	X	0	0
Culicidae	*Aedes*	*cinereus*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Culicidae	*Culiseta*	*alaskaensis*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Diadocidiidae	*Diadocidia*	*ferruginosa*	X	X	0	X	X	X	0	X	X	X	0	X	X	X	0	X
Diadocidiidae	*Diadocidia*	*spinosula*	0	X	0	0	0	0	0	0	X	X	0	0	X	X	0	X
Ditomyiidae	*Ditomyia*	*fasciata*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Ditomyiidae	*Symmerus*	*annulatus*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Dolichopodidae	*Argyra*	*diaphana*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	0
Dolichopodidae	*Dolichopus*	*claviger*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Dolichopodidae	*Dolichopus*	*festivus*	0	0	X	0	0	0	X	X	0	0	X	0	0	0	X	X
Dolichopodidae	*Dolichopus*	*nigricornis*	0	0	X	X	0	0	X	X	0	0	0	X	0	0	X	X
Dolichopodidae	*Dolichopus*	*plumipes*	0	X	0	0	0	0	0	0	0	X	0	0	0	0	0	0
Dolichopodidae	*Hercostomus*	*aerosus*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	0
Dolichopodidae	*Neurigona*	*pallida*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Dolichopodidae	*Neurigona*	*quadrifasciata*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Dolichopodidae	*Sciapus*	*longulus*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Drosophilidae	*Drosophila*	*kuntzei*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	X	0
Dryomyzidae	*Dryomyza*	*decrepita*	0	0	0	X	0	0	X	X	0	0	0	X	0	0	0	X
Dryomyzidae	*Dryomyza*	*flaveola*	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	X
Empididae	*Empis*	*aestiva*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Empididae	*Empis*	*tessellata*	X	X	X	0	X	X	X	0	X	X	0	0	X	X	0	X
Empididae	*Hilara*	*interstincta*	0	0	X	X	X	0	X	X	0	0	X	X	0	0	X	X
Empididae	*Phyllodromia*	*melanocephala*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Empididae	*Rhamphomyia*	*longipes*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fanniidae	*Fannia*	*corvina*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Fanniidae	*Fannia*	*rondanii*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Fanniidae	*Fannia*	*sociella*	X	X	X	0	X	X	X	0	X	X	X	0	X	X	X	0
Heleomyzidae	*Morpholeria*	*ruficornis*	0	0	X	X	0	0	X	0	0	0	X	0	0	0	X	0
Heleomyzidae	*Suillia*	*bicolor*	X	X	X	X	X	X	X	X	X	0	X	X	X	0	X	X
Heleomyzidae	*Suillia*	*notata*	0	0	X	0	0	0	X	0	X	0	X	0	0	X	X	0
Heleomyzidae	*Suillia*	*ustulata*	0	0	0	0	0	0	0	0	0	0	X	X	0	0	0	0
Hybotidae	*Bicellaria*	*austriaca*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Hybotidae	*Euthyneura*	*myrtilli*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Hybotidae	*Oedalea*	*flavipes*	0	X	0	0	0	X	0	0	0	X	0	X	0	X	0	0
Hybotidae	*Platypalpus*	*exilis*	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0
Hybotidae	*Platypalpus*	*major*	X	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Hybotidae	*Tachypeza*	*nubila*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Hybotidae	*Trichina*	*bilobata*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Keroplatidae	*Keroplatus*	*testaceus*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Keroplatidae	*Orfelia*	*fasciata*	X	0	0	X	0	0	0	X	X	0	0	X	X	0	0	X

Keroplatidae	*Orfelia*	*nemoralis*	0	X	X	0	0	0	X	X	0	X	X	X	0	X	X	X
Keroplatidae	*Urytalpa*	*dorsalis*	0	X	0	0	0	X	0	0	0	X	0	X	0	X	0	0
Lauxaniidae	*Lyciella*	*platycephala*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Lauxaniidae	*Lyciella*	*rorida*	0	0	X	X	0	0	X	0	0	0	X	X	0	0	X	X
Lauxaniidae	*Tricholausearia*	*praeusta*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0
Lauxaniidae	*Tricholauxania*	*praeusta*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Limoniidae	*Achyrolimonia*	*decemmaculata*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Limoniidae	*Austrolimnophila*	*ochracea*	0	0	X	0	0	0	X	X	0	0	X	0	0	0	X	0
Limoniidae	*Epiphragma*	*ocellare*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Limoniidae	*Molophilus*	*appendiculatus*	0	0	0	X	0	0	0	X	X	0	0	X	0	0	0	X
Limoniidae	*Neolimonia*	*dumetorum*	X	X	0	X	X	X	0	X	X	X	0	X	X	X	0	X
Limoniidae	*Rhipidia*	*maculata*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Lonchaeidae	*Protearomyia*	*nigra*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Lonchaeidae	*Protearomyia*	*rameli*	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0
Micropezidae	*Neria*	*cibaria*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Milichiidae	*Phyllomyza*	*equitans*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Muscidae	*Coenosia*	*tigrina*	X	0	0	X	0	0	0	X	X	0	0	X	X	0	0	X
Muscidae	*Haematobosca*	*stimulans*	X	0	0	0	X	X	0	0	X	0	0	0	X	0	0	0
Muscidae	*Helina*	*impuncta*	X	X	X	0	X	X	X	0	X	X	X	X	X	X	X	X
Muscidae	*Helina*	*reversio*	X	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Muscidae	*Helina*	*trivittata*	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0
Muscidae	*Mydaea*	*corni*	X	0	X	X	X	0	X	0	X	0	X	0	X	0	X	0
Muscidae	*Mydaea*	*nebulosa*	X	X	0	0	X	X	0	0	X	X	0	0	X	X	0	0
Muscidae	*Phaonia*	*pallida*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Muscidae	*Polietes*	*lardarius*	X	0	0	0	X	X	0	0	X	0	0	0	X	0	0	0
Muscidae	*Thricops*	*semicinereus*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Acnemia*	*nitidicollis*	0	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Mycetophilidae	*Cordyla*	*semiflava*	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0
Mycetophilidae	*Docosia*	*fuscipes*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Docosia*	*gilvipes*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Mycetophilidae	*Leia*	*crucigera*	0	0	X	X	0	0	0	X	0	0	0	X	0	0	0	X
Mycetophilidae	*Monoclona*	*rufilatera*	0	0	0	X	0	0	0	X	X	0	0	X	X	0	0	X
Mycetophilidae	*Mycetophila*	*caudata*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Mycetophilidae	*Mycetophila*	*fungorum*	X	0	X	X	X	0	X	X	X	0	X	X	X	0	X	X
Mycetophilidae	*Mycetophila*	*ichneumonea*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X
Mycetophilidae	*Mycetophila*	*idonea*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Mycetophilidae	*Mycetophila*	*occultans*	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0	X
Mycetophilidae	*Mycetophila*	*perpallida*	X	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X
Mycetophilidae	*Mycomya*	*fimbriata*	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Phronia*	*basalis*	0	0	X	X	0	0	0	X	0	0	X	X	0	0	X	X
Mycetophilidae	*Phronia*	*nigricornis*	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Platurocypta*	*testata*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Synapha*	*vitripennis*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Mycetophilidae	*Tetragoneura*	*sylvatica*	X	X	X	X	0	X	X	X	X	X	X	X	X	X	X	X
Mycetophilidae	*Zygomyia*	*pseudohumeralis*	0	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Pallopteridae	*Toxoneura*	*quinquemaculata*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Phaeomyiidae	*Pelidnoptera*	*fuscipennis*	X	X	0	0	X	X	0	0	X	X	0	0	X	X	0	X
Phoridae	*Anevrina*	*thoracica*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Phoridae	*Conicera*	*tibialis*	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0
Phoridae	*Diplonevra*	*nitidula*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Phoridae	*Gymnophora*	*arcuata*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Phoridae	*Megaselia*	*campestris*	X	0	X	X	X	0	X	0	X	0	X	X	X	0	X	X
Phoridae	*Megaselia*	*ciliata*	0	0	X	X	0	0	X	0	0	0	X	0	0	0	X	0
Phoridae	*Megaselia*	*nigriceps*	X	0	0	X	0	0	0	X	X	X	0	X	X	X	0	X
Phoridae	*Megaselia*	*pleuralis*	0	0	X	X	0	0	X	X	0	X	X	X	0	X	X	0
Phoridae	*Megaselia*	*ruficornis*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Phoridae	*Megaselia*	*variana*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X
Phoridae	*Phora*	*atra*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	X
Pipunculidae	*Cephalosphaera*	*germanica*	0	0	0	0	0	X	0	0	0	0	0	0	0	X	0	0
Pipunculidae	*Eudorylas*	*subfascipes*	X	0	0	0	X	0	0	0	X	X	0	X	X	0	0	0
Pipunculidae	*Jassidophaga*	*beatricis*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Pipunculidae	*Nephrocerus*	*flavicornis*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Psychodidae	*Clytocerus*	*ocellaris*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	X
Psychodidae	*Parajungiella*	*longicornis*	0	0	0	0	0	0	0	0	X	X	0	X	X	X	0	X
Psychodidae	*Pneumia*	*nubila*	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X
Psychodidae	*Psychoda*	*albipennis*	0	0	0	X	0	0	0	X	0	X	0	X	0	X	0	X
Psychodidae	*Psychoda*	*gemina*	0	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Psychodidae	*Psychoda*	*lobata*	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0
Psychodidae	*Psychoda*	*phalaenoides*	X	X	X	X	0	X	X	X	X	X	X	X	X	X	X	X
Psychodidae	*Trichopsychoda*	*hirtella*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Rhagionidae	*Ptiolina*	*obscura*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Rhagionidae	*Rhagio*	*annulatus*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rhagionidae	*Rhagio*	*scolopaceus*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rhagionidae	*Rhagio*	*strigosus*	X	X	X	X	0	0	X	X	X	X	X	X	X	X	X	X
Sarcophagidae	*Sarcophaga*	*carnaria*	X	X	X	X	X	0	0	X	X	0	0	X	X	0	0	X
Sarcophagidae	*Sarcophaga*	*haemorrhoa*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Sarcophagidae	*Sarcophaga*	*subvicina*	X	X	0	X	X	0	0	X	X	0	0	X	X	0	0	X
Sarcophagidae	*Sarcophaga*	*variegata*	X	X	0	0	X	0	0	0	X	0	0	X	X	0	0	X
Scathophagidae	*Scathophaga*	*furcata*	0	X	0	0	X	X	0	0	0	X	0	0	X	X	0	0
Scathophagidae	*Scathophaga*	*stercoraria*	X	X	0	0	0	X	0	0	X	X	0	0	X	X	0	0
Sciaridae	*Bradysia*	*cinerascens*	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0

Sciaridae	*Bradysia*	*fungicola*	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Sciaridae	*Bradysia*	*insignis*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0
Sciaridae	*Bradysia*	*longicubitalis*	0	0	0	X	0	X	0	X	0	X	0	X	0	X	0	X
Sciaridae	*Bradysia*	*pectoralis*	0	0	X	X	0	0	X	X	0	0	0	X	0	0	X	X
Sciaridae	*Bradysia*	*placida*	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0	X
Sciaridae	*Bradysia*	*vagans*	X	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Sciaridae	*Camptochaeta*	*vivax*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Sciaridae	*Claustropyga*	*abblanda*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	X
Sciaridae	*Corynoptera*	*blanda*	0	0	0	0	0	0	0	0	0	0	X	X	0	0	0	X
Sciaridae	*Corynoptera*	*boletiphaga*	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0
Sciaridae	*Corynoptera*	*cuniculata*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	X
Sciaridae	*Corynoptera*	*dentata*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Sciaridae	*Corynoptera*	*forcipata*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sciaridae	*Corynoptera*	*furcifera*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Sciaridae	*Corynoptera*	*tetrachaeta*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Sciaridae	*Cratyna*	*harpata*	0	0	X	0	0	0	0	0	0	0	X	0	0	0	X	0
Sciaridae	*Cratyna*	*perplexa*	0	0	X	X	0	0	X	X	0	0	X	X	0	0	X	X
Sciaridae	*Cratyna*	*uliginosoides*	0	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0
Sciaridae	*Epidapus*	*alnicola*	0	0	X	X	0	0	0	X	0	0	X	X	0	0	X	X
Sciaridae	*Epidapus*	*atomarius*	0	0	0	0	0	0	0	0	0	0	X	0	0	0	X	0
Sciaridae	*Epidapus*	*ignotus*	0	0	0	0	0	0	0	X	0	0	0	0	0	0	X	X
Sciaridae	*Leptosciarella*	*scutellata*	0	0	X	0	0	0	X	0	0	0	X	X	0	0	X	X
Sciaridae	*Leptosciarella*	*subpilosa*	0	0	0	X	0	0	0	X	X	0	0	X	0	0	0	X
Sciaridae	*Phytosciara*	*flavipes*	0	0	0	0	0	0	0	0	0	0	0	0	0	X	0	0
Sciaridae	*Scatopsciara*	*atomaria*	0	0	0	X	0	0	0	X	0	X	0	X	0	X	0	X
Sciaridae	*Scatopsciara*	*edwardsi*	0	0	0	0	0	0	0	0	0	X	0	0	0	0	0	0
Sciaridae	*Scatopsciara*	*vitripennis*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Sciaridae	*Sciara*	*hebes*	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0
Sciaridae	*Sciara*	*lackschewitzi*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	X
Sciaridae	*Scythropochroa*	*radialis*	0	0	0	0	0	0	0	0	0	0	X	X	0	0	X	X
Sciaridae	*Trichosia*	*acrotricha*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Sciaridae	*Trichosia*	*edwardsi*	0	0	0	0	0	0	0	0	0	0	0	X	0	0	0	0
Sciaridae	*Trichosia*	*iota*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Sciaridae	*Trichosia*	*lengersdorfi*	0	0	0	X	0	0	0	0	X	0	0	X	X	0	0	X
Sciaridae	*Trichosia*	*splendens*	X	0	X	0	X	0	X	0	X	0	X	0	X	0	X	0
Sphaeroceridae	*Copromyza*	*stercoraria*	X	0	X	X	X	0	0	0	X	X	X	X	X	0	0	X
Sphaeroceridae	*Crumomyia*	*fimetaria*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Sphaeroceridae	*Crumomyia*	*nitida*	0	0	0	X	0	0	0	X	0	0	X	X	0	0	0	X
Sphaeroceridae	*Gigalimosina*	*flaviceps*	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0	X
Sphaeroceridae	*Spelobia*	*palmata*	0	0	X	X	0	0	X	X	X	X	X	X	X	X	X	X
Sphaeroceridae	*Spelobia*	*parapusio*	X	X	X	0	X	X	X	0	X	X	X	0	X	X	X	0
Stratiomyidae	*Clitellaria*	*ephippium*	X	X	0	0	X	X	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Blera*	*fallax*	X	X	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Chrysotoxum*	*cautum*	X	0	0	0	X	X	0	0	X	X	0	0	X	0	0	0
Syrphidae	*Chrysotoxum*	*festivum*	0	0	0	0	0	0	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Criorhina*	*asilica*	X	X	0	0	X	X	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Dasysyrphus*	*tricinctus*	X	0	0	0	X	X	0	0	X	0	0	X	X	X	0	0
Syrphidae	*Didea*	*fasciata*	X	0	X	0	X	X	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Didea*	*intermedia*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Epistrophe*	*nitidicollis*	X	X	0	0	X	X	0	0	X	X	X	0	X	X	0	0
Syrphidae	*Melanostoma*	*mellinum*	0	0	0	0	X	0	0	0	0	0	0	0	X	0	0	0
Syrphidae	*Rhingia*	*campestris*	X	X	0	0	X	X	0	X	X	X	0	0	X	X	0	X
Syrphidae	*Syrphus*	*ribesii*	X	X	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Syrphidae	*Temnostoma*	*bombylans*	X	X	X	X	X	X	X	X	X	X	X	0	X	X	X	0
Syrphidae	*Temnostoma*	*vespiforme*	0	0	X	X	0	X	X	X	X	X	X	0	0	X	X	0
Tabanidae	*Hybomitra*	*bimaculata*	X	X	X	0	X	X	0	X	X	0	X	0	X	0	0	X
Tabanidae	*Hybomitra*	*lundbecki*	X	0	0	0	0	0	0	0	X	0	0	0	0	0	0	0
Tachinidae	*Bessa*	*selecta*	0	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Tachinidae	*Gastrolepta*	*anthracina*	0	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0
Tachinidae	*Peribaea*	*fissicornis*	X	0	0	0	X	X	0	0	X	X	0	0	X	X	0	0
Tachinidae	*Phorocera*	*obscura*	0	X	0	0	0	X	0	0	0	X	X	0	0	X	0	0
Tachinidae	*Phryno*	*vetula*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Tachinidae	*Tachina*	*magnicornis*	X	X	X	X	X	X	X	X	0	X	X	0	0	X	X	0
Tachinidae	*Triarthria*	*setipennis*	0	X	0	0	0	X	0	0	0	X	X	0	0	X	0	0
Tipulidae	*Nephrotoma*	*analis*	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Tipulidae	*Nephrotoma*	*submaculosa*	0	0	0	0	X	0	0	0	X	0	0	0	X	0	0	0
Tipulidae	*Tanyptera*	*atrata*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tipulidae	*Tipula*	*nubeculosa*	0	0	X	X	0	0	X	X	0	0	0	X	0	0	0	X
Tipulidae	*Tipula*	*pseudovariipennis*	X	X	0	0	0	X	0	0	0	X	0	0	0	X	0	0

Table 5.11: 228 DNA based Diptera species identifications with Linnean name assignments. For single-indexed samples (SI) 130 species were identified. For double-indexed samples 227 species were identified. Species identifications were higher and also more consistent beween the 3'- and 5' sorted subsets for DI than for SI. Congruence between 3'- and 5'- sorted subset identifications ranged between 75% - 80% for DI and between 64% - 75% for SI. Several further species could also be identified aside the order of Diptera with a full list in the supplement (S3.24).

	MI	SI (tax)		DI (tax)		SI (full)		DI (full)	
Samples		**3'**	**5'**	**3'**	**5'**	**3'**	**5'**	**3'**	**5'**
SEW2	24	33	31	53	51	80	66	104	97
SEW3	23	34	28	50	46	77	70	94	91
SEW7	18	39	27	44	42	76	60	79	74
SEW9	25	52	47	65	63	99	92	117	112

Table 5.12: Sequencing greatly exceeded taxonomic species identifications. For this it did not matter if all genetically identified Diptera families were taken into account (full) or only those families that were previously identified by the taxonomists (tax). The table shows the number of identified species from the different samples based on the morphological identification (MI) and based on genetic sequencing using either a single-indexing or double-indexing approach (SI, DI).

5.4 Discussion

5.4.1 The automated interval sampler (AIS)

Great potential lies in the automation of traps for entomologists. Various traps have been used over time for general studies of insect diversity and for the detection and monitoring of invasions of pest insects (Nabli et al. 1999; Epsky et al. 2008; Selby et al. 2014; Becker et al. 2017). Especially for monitoring, standardized insect captures and regular trap maintenance are needed. An automation of different tasks can reduce the costs up to 80% (Selby et al. 2014). And the cost efficiency increases over time as the onetime investment into the material holds up against a constant payment for qualified human work.

AIS was designed to take samples for a whole season. The basic capture interval of one week allows autoated sampling for 3 months. After this, the sampling jars need to be changed. As the interval is free to choose, sampling periods can be stretched or shortened to either gain less intense manual maintenance or maybe a higher temporal resolution. The choosable settings also allow monitoring diurnal species shifts without continuous disturbances of the habitat.

5.4.1.1 Overall performance of the automated interval sampler (AIS)

In this study the duration of capturing extended over 6 months with a one-week sampling interval and included 6 Malaise traps with the most distant trap being about 700 km away from the research museum in Bonn (ZFMK). A 1,600 km drive for a weekly maintenance of all traps would have been unrealistic. The planned automation of the weekly change of the sampling jar was therefore very desirable. While the mechanical realization had been rather easy, the reliable powering of the remote electronics proved to be more challenging. Drawbacks seem to be a commonly shared experience in research (Reynolds and Riley 2002; Selby et al. 2014). While operating accurately under test-conditions, forest conditions proved to be different. Especially during summer less light than expected reached the solar panels that charged the batteries. A major reason for this was the increasing density of the foliage in the beech plots. Light incidence was also low in the evergreen pine plots. A while after vegetation period started the undergrowth also began to prevent the light from reaching the solar panels. Repeated power shortages were the consequence. Some of the different plot samples had therefore periods of unsynchronized sampling between the different traps. In some cases the AIS did periodically not capture at all, as it went to its stop position, the 13th jar. Repeated and previously unplanned battery changes were conducted to compensate for the energy shortages. Nevertheless, several samples were taken as planned, covered the same time period and were in good condition and suitable for

sequencing. For now their evaluation implicated their complete destruction to achieve a homogeneous distribution of DNA within the sampling liquid. Therefore only a small fraction was used to demonstrate the principal function of the workflow hoping that improvements in the DNA acquisition from large samples will improve sample treatment in the near future. Tests of using preservative ethanol as a non-destructive source for DNA have been promising (Hajibabaei et al. 2012).

And although the also collected unsynchronized samples do not allow direct comparisons in terms of identical weekly time periods, they still contain valuable plot data. These data could still be implemented for comparisons, when the time period to be analyzed is extended from a single week to several weeks. Even then should the temporal resolution be sufficient for valuable insights.

5.4.1.2 Improvements of the automated interval sampler (AIS)

Later applications on agricultural fields showed that the energy supply of the automated interval sampler can be more than sufficient when there is no shadow. Still most improvements concentrated on this issue to make the AIS widely applicable (Figure 5.13). The 12V, 9Ah lead-acid batteries were exchanged with higher capacity battery versions (12V, 12Ah) for extended stand by time. Additional batteries or different batteries with even higher capacities can also now be used. The usage of two 40W thin-film solar panels instead of the 20W crystalline solar panels now especially supports the employment in low light conditions. Thin-film solar panels are endorsed not only for general low light conditions but are also more forgiving when solar panels are mounted in suboptimal angles of incidence. The software controlling the sampler was also adapted to minimize its usage of energy.

Further improvements included the enhancement of the rotational mechanism for more robustness and the mounting of the construction onto a removable leveling tripod. The hardware now also allows a further software upgrade to measure and save different sorts of data. These data could include the own performance of the traps, the atmospheric pressure, temperature and humidity. While already capable to incorporate diurnal shifts, the interval sampler could also precisely record shifts in insect activity in response to local climate variations in the near future. Additional sensors could further broaden the station's application. The recoding of enviromental sound could document for example birds. External cameras with motion sensors could document the passing of larger mammals. Internal cameras with motion sensors could take pictures and document the captured arthropods when passing the funnel jar. The extended capturing data provided by this would certainly be beneficial for detailed analyzes and interlinked data evaluations.

Figure 5.13: Left: The automated interval sampler (AIS) as it was used in the Schorfheide-Chorin together with a Malaise trap on plot SEW15 (Photo: Struwe 2015). Right: The current AIS version with improvements (Photo: Kilian 2017).

5.4.2 How to efficiently sequence bulk samples

5.4.2.1 Choosing the adequate indexing approach

After a long career of Sanger based sequencing technologies, laboratory methods have improved further (Sanger et al. 1977; Schuster 2007; Metzker 2010; Kircher et al. 2012). Next-generation sequencing (NGS) produces large numbers of low-cost reads. This makes NGS valuable for many applications as gene discovery or for species classifications in metabarcoding as it was used in this study (Petrosino et al. 2009; Taberlet et al. 2012). To make economical use of the available sequencing capacities, multiplexing strategies are being used. Sample-specific indices attached to the sequences allow sequencing multiple samples in parallel (Meyer et al. 2007). However, these approaches come with a risk of misidentifications, also known as "cross-talk" (Wright and Vetsigian 2016). Cross contaminations or bulk amplification are proved to further impair the strategies accuracy (Kircher et al. 2012). Instead of only using a single index a double-indexing approach using two specific indices per sample has been proposed to eliminate these problems and increase both the scope and accuracy of multiplex sequencing.

Both indexing approaches were used for the samples from the Malaise traps to compare the results. While both approaches initially had similar read numbers it showed that after processing (e.g. quality filtering and read pairing) the amount of suitable sequences for further processing dropped more for single-indexed samples than for double-indexed samples (Figure 5.14). During processing, read numbers differed about tenfold between single- and double-indexing (Table 5.3 and Table 5.4). On the one hand this proves that sophisticated processing for high quality sequences indeed removes the expected larger error rates of single-indexed samples, but on the other hand also lowers the amount of evaluable data (Figure 5.14). Although the disparity lowers towards the end of processing, MOTU and species numbers remain significantly lower in single-indexing approaches. Species-specific recovery was also noticeably lower and also more inconsistent for single-indexed samples than for double-indexed samples (see also chapter 5.4.3).

The reason for the double-indexing success probably lies within the higher accuracy of the multiplex sequencing. Although highly distinguishable indices have been designed to avoid false assignments of reads to samples already in single-indexing, the

results remained behind expectations (Meyer and Kircher 2010; Kicher et al. 2012). In double-indexing results are from the beginning superior to the ones obtained by single-indexing. In double-indexing sample identification is performed twice for each template molecule, enabling an exponential decrease of the false-assignment rates and higher correct read to sample assignments (Kircher et al. 2012). Using accurate reads from the start also means that fewer errors need to be removed through various quality filters in later steps. And although the overall ecological evaluation seemed not to be drastically impaired by this (Figure 5.15), in detail several species were not recovered using single-indexing while they were recovered using double-indexing. Moreover, double-indexing greatly reduces the costs of highly multiplexed sequencing as it severely increases the number of unique sample identifiers through a high number of possible index combinations. Future samples should therefore always favour double-indexing before single indexing.

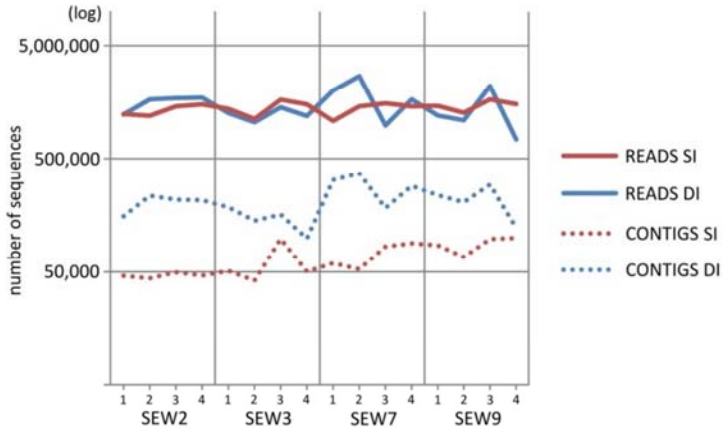


Figure 5.14: While the total read output after sequencing was about the same for single-indexing (SI) and double-indexing (DI) in all sequenced sub-samples (1, 2, 3, and 4) of the samples SEW2, SEW3, SEW7, and SEW9, the number of evaluable sequences experiences stronger reduction through sequence processing for samples using the SI approach than for the DI approach and stays noticeable until the end in the final MOTU and species numbers.

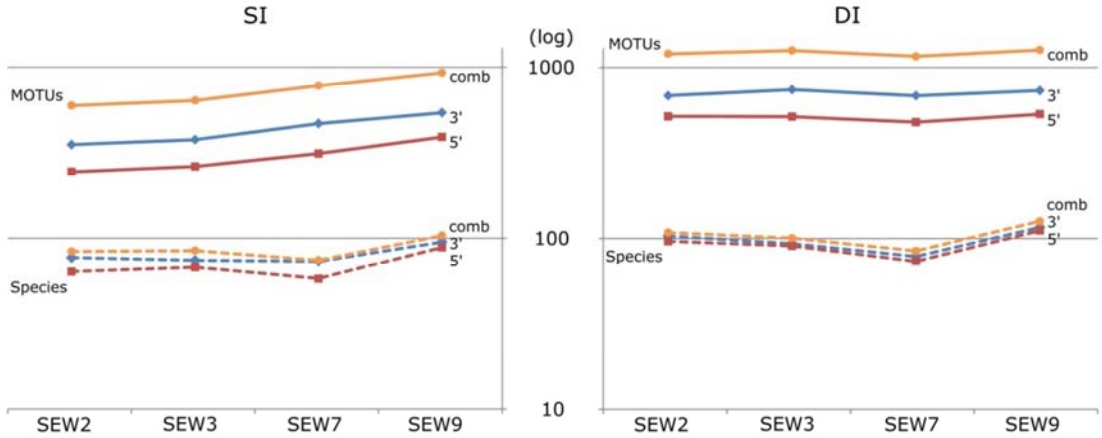


Figure 5.15: The number of MOTUs per sample (SEW2, SEW3, SEW7, and SEW9) was reduced by sorting the sequenced fragments into subsets according to their 3' and 5' positioning on a full length COI sequence. This slightly improved the MOTUs – species ratio. Species number was only marginally influenced by this for the 3'- and 5'- subsets. However, no information was lost as both subsets were preserved. Graphs labeled "comb" represent the combination of 3' and 5' data.

5.4.2.2 Suggestions for efficient high throughput sequencing

Bulk sequencing of mixed species samples for barcoding is challenging as current high throughput sequencing techniques concentrate on read output and not so much on read length (Metzker 2010; Xu 2014). However, since even very short fragments, often referred to as mini-barcodes, showed to lead to reliable results, full potential of these NGS methods can be exploited using shorther than full length (COI: 658 bp) reads (Meusnier 2008; Derocles et al. 2015; Lee et al. 2015; Villa et al. 2016). With a mean length of over 300 bp the evaluated fragment lengths exceeded all the minima found to be sufficient for adequate species determinations (Meusnier 2008; Derocles et al. 2015; Lee et al. 2015). The often used assembly of the shorther fragments is rather susceptible for errors. The wrong species fragments might be combined. Also parts can be omitted or become rearranged or might otherwise be deformed (Cheung et al. 2003; Schmutz et al. 2004; Salzberg and Yorke 2005; Phillippy et al. 2008; Alkan et al. 2011; Zhang and Backström 2014). The avoidance of assembling these fragments to what is hoped to resemble a full length COI sequence proved to still give detailed insight into diversity patterns and species distributions between the sampled forest plots. To avoid artefacts due to assembling errors and improve MOTU and species ratio, sequence fragments were aligned to full length COI sequences and then sorted into two subsets, according to their position in reference to the full length sequence. Forming subset indeed drastically reduced the total number of MOTUs for each subset by ~ 40% - 60% while species numbers identified from these MOTUs were only marginally reduced in each subset especially for double-indexing (Table 5.12 and Figure 5.15). The parallel evaluation of these 3' and 5' called subsets (named after the nucleic directions following the carbons of the sugar molecule) also showed that the different subsets still revealed highly similar results in diversity and richness patterns. This means informative evaluation is also possible using a single subset. This can save resources in hybrid capture where baits are designed to enrich the targets. As only one half of COI would now be used as a target the baits usally used for the second half can be designed to now also enrich a second gene-marker or further taxa aside from the order of Diptera.

5.4.2.3 Conclusion

Using a double-indexing approach on the targets that were enriched through hybrid capture yielded more valuable reads than using only single-indices on the targeted fragments. The higher in read numbers lead to more recovered Linnaean species for the samples with double-indexed sequences than for samples with single-indexed sequences. By this double-indexing gave a more detailed insight into the species composition of the samples.

The 3'- and 5'-sorted subsets yielded similar Linnaean species numbers in the different plot samples whereas double-indexing showed to be more consistent. On the one hand, these double results can be valuable as the subsets support each other; on the other hand this also indicates that using only a single subset can be sufficient to adequately evaluate a sample. This means that analyses can be focused on only a fragment of the original COI sequence. In the experiment informative results were gained using a fragment length of ~ 330bp. Results indicate that the targeted fragments should be situated in the sequence half closer to the 3'-end as this subset yielded more species identifications. Restricting the sequence length that needs to be analyzed also saves resources in the design of the baits used for the targeted hybrid enrichment.

These resources could be used to include further taxa next to the one of Diptera or to also include a second genetic marker next to COI.

This shows that the potential of DNA metabarcoding techniques are by for not yet exhausted. Improvements are for example still needed in its dependency on PCR, as this step is known for its biasing influence. Although hybrid enrichment already reduces the number of amplification cycles, especially the multiplexing step still requires a certain number of amplification cycles two introduce the sample specific indices. However, the impressive ongoing progress in DNA sequencing will hopefully soon eliminate the currently required DNA amplification steps. This would also benefit abundance estimations as read numbers would correlate with the biomass of species in a sample.

5.4.3 MOTU based assessment

5.4.3.1 Diversity of managed and unmanaged forests plots (pine vs beech)

Studies of microbial communities have declared that traditional diversity metrics calculated from MOTUS are hard to interpret (Edgar 2013). The reason for this is that MOTU frequencies generally had low correlation with species frequencies, which means that the most abundant MOTU usually did not contain the most abundant species. Furthermore multiplexed samples have also been prone for misidentifications, assigning reads to the wrong sample, also known as “cross-talk” (Wright and Vetsigian 2016). Besides that some diversity metrics use singletons to evaluate richness estimates, while at the same time singletons in sequencing are more likely based on sequencing errors than on rare species (Edgar 2013).

With this in mind the use of the ACE estimator is advisable as it makes use of variable abundance thresholds for its richness estimates aside the common focus on singletons (Chao and Lee 1992; Chao and Chen 2003; Chao and Chiu 2016). Furthermore double-indexing has significantly improved the accurarcy of read assignments, making misassignments rather uncommon (Kircher et al. 2012). Although it must be kept in mind that some models will generally be more reliable indicators of ecologically relevant changes than others and their validity is often in linked to the sample coverage (Hill TC et al. 2003). As for this it is the choice of tools that decides over the eglibility of MOTU based models in diversity assessments. MOTU based evaluations can still be considered valuable, as they aid sample comparison by revealing trends as well as specific changes (Hill TC et al. 2003).

Rarefaction curves showed that sequencing was especially efficient for the double-indexed samples; wheras the single-indexed samples in general had low sequence numbers (Figure 5.8). However, the Good’s coverage index yielded very high coverage rates for both approaches giving confidence to further evaluate the sequencing data (Table 5.5). When looking at the results of MOTU richness and diversity of the samples from the different plots it becomes apparent that the samples from the pine plots, SEW2 and SEW3, with their managed stands have relatively high richness and diversity values compared to the samples from the unmanaged beech stands, SEW7 and SEW9 (Figure 5.16). What might seem surprising at first has been already documented in literature. Previous studies have found higher numbers of individuals and species in managed forests than in unmanaged natural forests and that in general the fauna of

unmanaged forest is not richer than that of forests in traditional silvicultural use (Biström et al. 1988; Väisänen et al. 1993).

However, this is not the full picture. When the MOTU communities of the different plot samples were compared to each other, evidence was found that not only the communities of the two forest types differed from each other, but also the communities within the observed unmanaged beech plots, while the communities in the managed pine plots did not differ from each other. In contrast to the samples from the beech plots were the pine plots assigned to the same community type, indicating higher homogeneity within these two plots than within the two beech plots (Figure 5.16). This finding is also consistent with published research which indicated that forest managing and specifically logging tends to homogenize forest habitats (Niemelä 1997). An explanation for this finding can be that unmanaged forests might host more rare specialists that disappear when forest management starts, while generalists persist and numerous new species appear that can adapt to the new conditions in the managed plots.

As previously mentioned did the samples of the two unmanaged beech plots SEW7 and SEW9 not only differ from each other but also from the two managed pine plots SEW2 and SEW3 (Figure 5.16). This significant difference indicates (assuming that the unmanaged beech plots represent a more natural forest with its local original communities) that the observed managed pine forest were unable to maintain the original forest communities. This means, that although a high level of biodiversity can be achieved in managed forests, the original diversity of specialized species will be lost.

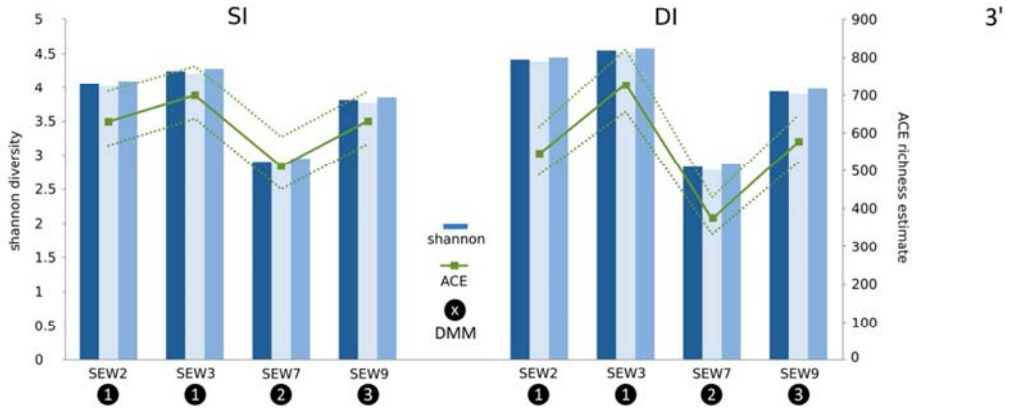


Figure 5.16: Richness and diversity estimates for the 3'-subset revealed similar patterns between the single (SI) and double (DI) indexing approaches. The highest MOTU richness can be found in the pine plot samples SEW2 and SEW3. In the beech plot samples the richness and diversity values were considerably lower for SEW7 than for SEW9. However, the DMM revealed that the MOTU communities are more different between the beech plots than between the pine plots. Dotted lines and light blue bars represent the 95% confidence intervals of ACE richness and Shannon diversity.

5.4.3.2 Ecological insights based on discriminative species characteristics

The MOTU-based recognition of characteristic species for pine and beech plots (using LEfSe: Segata et al. 2011) gave further interesting ecological insights. The discriminative species confirmed the characteristics of the sampled plots for the most parts (Table 5.7 and Table 5.8). In both plot types of managed pine (SEW2 and SEW3) and unmanaged beech (SEW7 ad SEW9) forests the general affiliation for "wetland", "forests" and "dead wood" were the dominating species characteristics encountered. Some (parasitic) species were more detailed determined to specific preferences as for certain host and

food plant species or host animals and/or their feces. These more general and species specific characteristics all reflect properties of the biosphere reserve. Its prevalent moist condition for example, is an inheritance of the last glacial period. The melting of the remaining ice created a plenitude of water bodies and marshlands allowing moist forests to grow in this region. Further species preferences also gave evidence of the agricultural use in the area around the sampled forest plots and in some cases possible habitat disturbances.

Most of the discriminative species from the different forest types had certain preferences for particular tree types, giving evidence of the composition and condition of the sampled forest types. *Blera fallax* (Syrphidae), a saproxylic Diptera species, was identified to be characteristic for the analyzed pine plot (Table 5.7). This species of hover fly is known to be associated with the decaying wood of *Pinus sylvestris*. Their larvae can be found in tree holes and trees with heart rot (Rotheray et al 2001). Opposing to this, *Temnostoma vespiforma* (Syrphidae) was identified as a discriminative species from the beech plots (Table 5.8). This hover fly species is known to occur in beech forests with aged trees and greater amounts of decaying wood (von der Dunk et al. 2006). *Cratyna perplexa* (Sciaridae), also a representative for the beech plots (Table5.8), is a species that is especially linked to deciduous forests containing *Fagus sylvaticus* (Menzal & Schulz 2007). However, the plots of the different forest types are naturally not exclusively made of pine and beech stands, as previously described in the outline of the study area. In addition to this in 2009 the TU Munich introduced several logs of further tree species to evaluate the long term effects of wood decay in different forest types (Wende 2014). For this freshly cut logs of 13 different tree species were introduced into each of the forest plots in the Biodiversity Exploratories (*Acer sp., Betula pendula, Carpinus betulus, Fagus sylvatica, Fraxinus excelsior, Larix decidua, Picea abies, Pinus sylvestris, Populus sp., Prunus avium, Pseudotsuga menziesii, Quercus sp., Tilia sp.*). Of course this experimental manipulation potentially changes the species composition in the surrounding area, but it might also be traceable by the performed analysis of discriminative species. However, the influence of the dead wood experiment can not be clearly discerned from the natural occurrence of different tree species within the forests. Rather unexpected species have also been recognized to be characteristic for the two plot types. Their preferences differed from what was expected from the forest type and the main tree species they were associated with in the analysis. *Ptiolina obscura* (Rhagionidae), a species identified to be characteristic for the sampled unmanaged beech plots (Table 5.8), is known to also need decaying wood. But instead of beech it favors the wood of trees belonging to the genera Acer and Quercus (Rotheray et al. 2001). And *Chrysotoxum cautum* (Syrphidae), found to be characteristic for the pine forest plots (Table 5.7) is actually known to rather occur in deciduous forests and not in pine stands (von der Dunk 2005).

Beyond the valuable information that certain sampled species were closely associated with either coniferous or deciduous forest types, insights into further habitat structures and also forest continuity were gained (Table 5.7 and Table 5.8). Information about adjacent biotopes gave for example the occurrence of *Empis aestiva* (Empididae) that was recognized as a discriminative species for the beech plots (Table 5.8). Empidoid flies are discussed as markers of landscape heterogeneity as the diversity of their life history traits plays an important role in their population dynamics (Delettre et al. 1997). Dry heathland appears to be the main site of larval growth of *E. aestiva* but

mating and feeding of the adults takes place near ponds and woodlots. The species of *Megaselia pleuralis* (Phoridae) has also been proposed as an indicator for habitat heterogeneity (Prescher et al. 1994; Brenner 2004; Durska et al. 2010). Their identification as a discriminative species for the sampled beech plots (Table 5.8) also indicates habitat heterogeneity. A source of heterogeneity can be habitat disturbances caused by cutting, wildfires or wind throw. One or more of these habitat structures should therefore be in a convenient distance or within the sampled plot. Areas with old wind throw were indeed observed in a distance of about 150-200 meter to the Malaise trap at the beech plot SEW9 (Figure 5.17). Further discriminative species also indicated the presence of cattle in adjacent areas to the forests. Especially the species *Haematobosca stimulans*, *Helina impuncta*, *Polietes lardarius* (Muscidae), *Scathophaga stercoraria* (Scathophagidae), and *Hybomitra bimaculata* (Tabanidae) that have identified as characteristic for the sampled pine plots (Table 5.7) can be taken as a strong evidence for cattle near these areas (Ball 1984; Skidmore 1985; Smith 1989; Krcmar et al. 2006; Stuke and Schacht 2009). Cattle were indeed observed in a noticeable close distance to the trap in the pine plot SEW2 (Figure 5.18). The presence of *Spelobia palmata* (Sphaeroceridae) might also hint to the presence of cattle close to the sampled beech plots (Table 5.8) although this indication is less distinct as *S. palmata* is also found to be associated with fox and deer feces (Hafez 1939; Coffey 1966; Buck 1994; Keiper et al. 2002; Papp 2002). Different Mycetophilidae were marked as discriminative species for the unmanaged beech plots (Table 5.8) but not so for the managed pine plots. Research showed that (semi-) natural forests are more sustaining for mycetophilids, compared to managed forests (Økland 1996). The historical continuity of forest patches seem to be especially important for their diversity. The restoration of the original diverse mycetophilid fauna after severe disturbance requires more than 70-120 years (Økland 1996). This emphasizes the high value of the old forests in the biosphere reserve which have never been clearcut, for preserving species diversity in the Schorfheide-Chorin.

Figure 5.17: Beech stand with areas of old wind throw close to the Malaise trap on the unmanaged beech plot SEW9 show plot heterogeneity through disturbance. (Photo: Struwe 2015)

Figure 5.18: Surrounding area of the managed pine plot SEW2. Location of the trap (left) with grazing cattle in near distance (right) and part of the Bugsin lake to the lower right. (Photos: trap and cattle, Struwe 2015; landscape, google maps)

5.4.4 Morphological versus sequence based species determinations

5.4.4.1 The impact of the unknown

As correlations between the number of MOTUs and the number of Linnean species are more than problematic, all MOTUs were identified to Linnean species when possible (Table 5.11). These species identifications were used to compare the success of species identifications based on morphological characters versus the identifications based on their sequences.

The differences in species numbers between the taxonomic and the sequencing approaches were striking. The number of species that were identified based on their sequences exceeded the taxonomic species identifications about 4.5 fold (Table 5.12 and Figure 5.19). And even when the families recovered by sequencing were restricted to the taxonomically identified families, species identifications were at least twice as high for sequencing as for the morphological approach. The differences in species numbers resulted not only from species belonging to families the taxonomists were not familiar with, but also from individuals where family and/or genus were known but further identification to species level was not possible (Table 5.9). Large proportions of these individuals were females, as these are especially hard to identify. Female Diptera often lack distinct or easily identifiable species characteristics (Ekrem et al. 2010; Cini 2012). In the history of taxonomy many species have been described from males only because of this impediment. And even from those species descriptions in which females have been used to describe the species, female individuals still are difficult to identify. As a consequence, females have suffered a general lack of taxonomic attention (Aagaard et al. 2005; Stur et al. 2005; Ekrem et al. 2010). But neglecting female individuals in ecological studies can bias species ratios, as rare species represented by females would stay unnoticed. This can also affect small sample sizes containing only a few individuals, as small samples can already be sexually biased in a way that not all species representatives are males. Especially parthenogenetic species, also existing in the order of Diptera, would be affected by this (Scholl 1960; Armitage et al 1995; Langton 1999). One of the major benefits of using a sequencing approach is that it allows assigning

sequence fragments to species independently from the originating gender (Casiraghi et. al. 2010). For some families at the observed forest plots the number of female individuals indeed exceeded those of males, being quite often twice as high, and sometimes exceeding them by even tenfold (Table 5.9). It is generally expected for species numbers to rise when females are included into the species inventories (Ekrem et al. 2010). Considering this, the doubling of species numbers in previous morphologically identified families seems less surprising. The detection of species in families the taxonomists had no expertise in only proves that the application of DNA based methods is highly valuable for general diversity estimates, as it gives a much more detailed insight into species communities in the investigated areas.

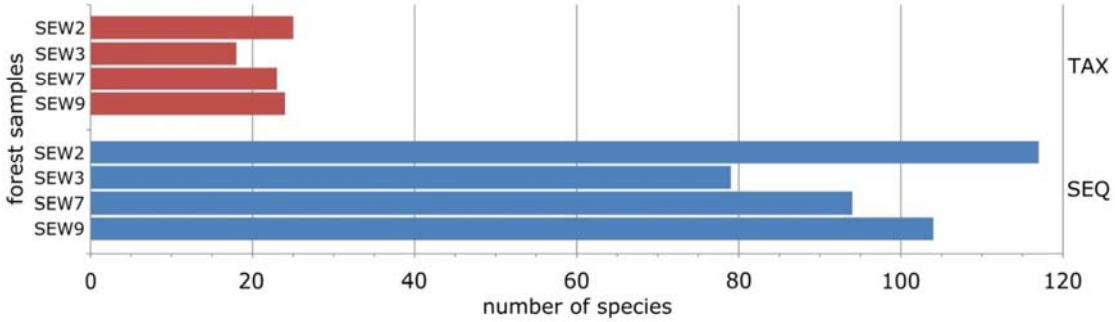


Figure 5.19: A comparison of Linnean species numbers from the different samples shows that more Diptera species were identified based on their sequences (SEQ) using the GBOL reference, than by the taxonomists (TAX), unsing morphological characters. Species numbers for sequencing were taken from the 3'-subset of the double-indexing approach (for detail see also: Table 5.12).

5.4.4.2 The downsides

A downside of the sequencing approach surely is that species missing in the sequence reference are not being named. However, the information about their presence is not lost, if species identifications are based on MOTUs. The use of MOTUs still allows statistical analyses and the comparison of different samples even when a Linnaean species name is not yet available. The constant growth of reference databases assures the inclusion of further species that permanently extend the databases with new references (Ratnasingham & Hebert 2007; O'Leary et al. 2016). Subsequent species identifications based on already acquired data from different samples are therefore very likely. The integration of computers for data management is also more efficient than the repeated manual sorting of individuals from a sampling jar whenever they need a morphological review.

When only a single genetic marker, often also reffered to as a genetic barcode, is used for species identification, species sharing identical genetic barcodes are of course problematic, while also different variants of this scenario are possible (Figure 5.20). The topic has already been addressed in Chapter 4.2.2 and is also frequently discussed in literature (Ferguson, 2002; Floyd et al., 2002; Quicke, 2004; Tautz et al., 2003; Meier et al. 2006). Although the effects of identical barcodes have to be seen critical in the context of DNA barcoding, their occurrence in species populations is estimated as rather moderate (Meier et al 2006). But shortening the identification barcode from the original 658 bp to ~330 bp leads to further identical specimen when the genetic distance is low and the characterizing substitution is outside of the queried sequence. However, time and cost efficiency with the still enormously detailed species list underline the usability of genetic approaches. Single adverse aspects should not be hastily taken as a reason to directly discard a method that has proved its applicability.

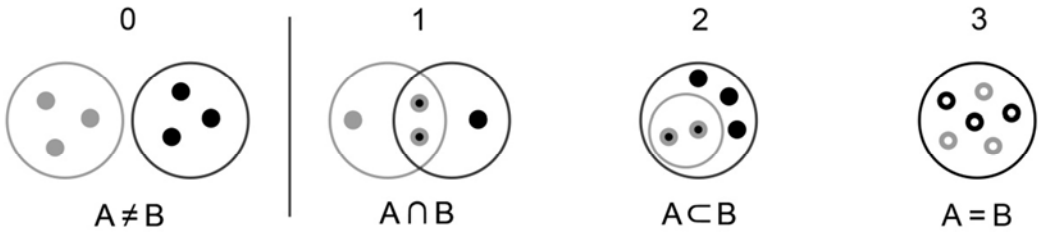


Figure 5.20: Exemplary cases of barcode distributions, where the circles represent the sum of all specimens of a single species. Grey circle: species A, black circle: species B. (0) Ideal scenario, where all specimens of both species A and B have their species specific barcode that easily distinguishes both. (1) Scenario of partially shared barcodes, where specimens of both species share specimens with the other species barcode, but also have specimens with their species specific barcode. (2) Scenario of partially shared barcodes, where the specimens of species B are still distinguishable from A based on their barcode but not vice versa, (3) Scenario where the specimens are indistinguishable based on their barcode.

5.4.5 Conclusion

5.4.5.1 Implementing DNA sequencing into ecological assessments

Results show that the implementation of DNA sequencing for diversity assessments is rewarding. It is being commonly used to assess microbial diversity, or species diversity in general, in enviroments where taxonomic knowledge is sparse (Moreira et al. 2002; Sogin et al. 2006; Holmes et al. 2012; Logares et al. 2013; Al-Rshaidat et al. 2016; Janzen et al. 2016; Bálint et al. 2017; Janzen et al. 2017). But in fact, not only in rather unexplored regions can diversity studies benfit from DNA approaches. Valuable assessments can start at the own doorstep (Geiger et al. 2016; Aagaard et al 2017). Especially when combined with automated workflows can these assessments be very useful in well-known ecosystems where continuous inventories and large scale species identification are needed. Such workflows can enhance biodiversity assessments by being faster and cheaper when automated capturing devices are integrated (Selby et al. 2014). Scientists stated already ten years ago that the costs for genetical processing of single specimens are low enough to compete with traditional morphological approaches; depending on laboratory facilities, equipment and the chosen approach (Cameron et al. 2006; Valentini et al. 2009). Since then costs have dropped further and new developments now allow the processing of multiple samples at the same time (Meier et al. 206). This progress allows the simultaneous identification of most species from a given biotope by a single researcher without the need to acquire taxonomic expertise in all the handled taxa. Using DNA barcodes enhances identification and differenciation by overcoming several taxonomic impediments which have been stated in literature (Gaston and O'Neill 2004). Different life stages or a specimen's gender do not exclude individuals anymore from being integrated into species assessments as it commonly happens in morphological studies (Aagaard et al. 2005; Stur et al. 2005; Casiraghi et al. 2010; Ekrem et al. 2010).

To compare the results of DNA sequencing in this study with a traditional morphological approach, samples were checked by two taxonomists before DNA extraction. These taxonomists were experts for the Diptera family Syrphidae and for the superfamily of Sciaroidea, containing Bolitophilidae, Diadocidiidae, Ditomyiidae, Keroplatidae, Mycetophilidae, Sciaridae and Cecidomyiidae. Although this enumeration might create the impression of covering a high number of Diptera families, it can only be considered a small fraction of the true diversity usually sampled in a Malaise trap

(Geiger et al 2016). Comparing a morphological approach with genetic sequencing might seem unfair as in the morphological approach species of families the taxonomist had no expertise in were knowingly left out. However, this approach is realistic as the content of extensive samples is rarely fully identified to date (Hallmann et al. 2017). Analyzing only fractions of the true diversity is also common when using indicator species to evaluate the ecological state of different sites. But this common approach has also already been criticized as inappropriate (Landres et al. 1988). A fixed set of indicator species can severely weaken the effectiveness and credibility of the results as the species biology and inferences from its properties might be inadequate to answer the addressed ecological question. Using only fractions of the true diversity also limits the amount of available data for statistical analysis, data interpretations, and possible conservational recommendations. If indicator species are still used it has been recommended that species should be identified to be characteristic for their environment from the samples taken to suit the addressed questions (Landres et al. 1988; Dufrêne and Legendre 1997). Analyses often concentrate on relative abundance and frequency of occurrence in various sites (Dufrêne and Legendre 1997, Segata et al. 2011). This procedure ensures that scientists are not overwhelmed by the collected data, but are able to reduce the available information to its informative characters.

5.4.5.2 Conservational recommendations

MOTU analyses showed that while in general species richness and diversity are still relatively high on managed forest plots it can be assumed that the original species composition of natural forests will be lost through forest management. Especially rare species and species with distinct habitat preferences and a low ability to adapt to changes can be expected to be endangered. Species that were identified from the MOTUs and classified to be representative for the sampled forest type almost all shared xylobiont or saproxylic life stages. But not only do most identified Diptera species depend on decaying wood, Diptera species are known to also constitute 80-90% of the insects reared from decaying wood (Hilt and Ammer 1994; Økland 1996). In contrast to this do xylobiont and saproxylic species usually suffer from low amounts of wooden debris in managed forests which can lead to a drastic reduction of original saproxylic species (Siitonen 2001). And although forest managment policies are being rethought, it shows that actions lying in the past still have consequences today. Decaying wood is crucial for dipteran diversity in both forest types but its insertion or reassemblage does not guarantee the preservation or the restoration of the original species communities in the managed forests. Important for diversity conservation is therefore the preservation of forests with long undisturbed histories.

Adjacent habitats also influence diversity. Forest edges are not necessarily also delimiting species distribution across and between habitats as results have proved. Some specimens might be passing through but others come in search for resources, mating spots or oviposition. Habitat heterogeneity resulting from windthrow increases diversity to some extent. However, larger disturbances like clear cutting changes habitats and leads to a homogenization of species diversity in the different forest patches.

6 General conclusion and future prospects

This study can be considered pioneering in context of future automated biodiversity assessments. It combines latest sequencing techniques with innovations on taking multiple standardized high quality samples with a minimum of trap maintenance and habitat disturbance. In its results it shows the applicability of even short (~ 330bp) barcodes and MOTUs for species identification and diversity estimations on the often ecologically neglected order of Diptera. In fact, Diptera species can indicate habitat disturbances and provide information about the historical continuity of forest patches. Although it was found that managed forest can yield a high diversit, maximum diversity is achieved by rather old unmanaged forests plots with a long continuity. While each of these unmanaged plots might have a lower diversity than the managed plots, their collective diversity can yield even higher species numbers, as unmanaged have more dissimilar species communites. Considering this, a special emphasis must be put on the preservation of these areas. A loss of continuity seems to cause excessive long-term effects on the species compositions and might even be irreversible when local species go extinct.

However, the evaluated sample size was low, as the study was used to demonstrate the feasibility of the concept of a DNA based and partially automated mass-sample evaluation. The great potential of automated biodiversity assessments lies in its large informative data output. Further more detailed results can be expected on the presence and the seasonal changes of species distributions when extensive numbers of samples are being analyzed over a longer period of time. DNA based automated biodiversity assessments prove to be highly valuable for exploring and understanding the diversity and ecology of species assemblages. The study showed that important knowledge about the species-rich order Diptera can be derived using relatively short DNA barcodes instead of time consuming morphological identifications used in conventional taxonomy. Beside information about species richness, knowledge about more complex ecological issues such as insights into habitat structure and forest continuity can be gained. DNA-sequencing allows the identification of both, Linnaean species, and species-like units, without previous taxonomic expertise in all identified taxa. Moreover, it is especially advantageous that the algorithm used for species-like MOTU definition can be explicit and largely deterministic allowing both hypothesis testing and the transferability of MOTU analyses between studies. A concept based on MOTUs allows combining data across different studies by robustly synonymizing taxa through for example co-clustering larval and adult specimens independently from their sex.

However, the morphological approach does not become obsolete. Especially as for now speciemens numbers can yet not be assessed using a DNA approach. Although the specimens' abundance and the resulting number of sequence reads within a sample are not totaly independent from each other, the variation found for specimens and species is still too large to draw any conclusions about the original specimens' numbers. Not only for this, the different approaches of DNA based and morphological based assessments need to be understood as being complementary. Accurate identifications based on a single genetic marker for taxa that do not resolve monophyletic remains difficult using DNA barcoding. Individuals may be genetically more similar to those of a

different species than to each other. In these rather complex species groups, joint analyses of nuclear and mitochondrial markers might help to distinguish between species. Fortunately, the techniques used in this study do have the potential to integrate a second genetic marker to compensate for inefficiencies of COI. However, each of these inconsistencies with common taxonomy should be checked by experienced taxonomists to differentiate between the valuable gain of knowledge and methodical artefacts. For the introduction of a large fraction of these artefacts PCR can be made responsible. Therefore a focus should also be put on the development of PCR free methods as it showed to severely bias all following analyses. And although target enrichment is highly efficient with a minimum use of PCR cycles its complete prevention should be striven for in the future.

A completion of the existing genetic reference databases with all existing species might be utopic; nevertheless, a focus should be put on the further extension of these databases with a special attention for their maintenance to guarantee only high quality sequences with accurate species determinations. At the same time the current standart of including several individuals of a species to validate a species sequence and to cover sequence variances of a species barcode should not be neglected.

At this point it must be mentioned that using a shorter barcode than the commonly used 658 bp of COI leads to increased numbers of identical barcodes between different species as their characteristic base differences might not be included in the fragment being analyzed. Current developments in NGS have put their focus rather on sequence output than on sequence length. Unfortunately Roche (Basel, Switzerland; https://www.roche.com) has shut down the 454 platform in 2013 putting an end to a technique that provided sequence reads with a length of ~500 bp. Very promising are the advances in the use of Nanopore DNA sequencing (Deamer and Akeson 2000; Deamer et al. 2016). The idea behind this is to sequence a single strand of DNA by measuring changes in the current while drawing it through a membrane with a nanoscopic pore by electrophoresis. With several thousand bases the reading length of Nanopore is unbeaten (Lazlo et al. 2014). However, the sequencing accuracy needed for a reasonable and standart application has not yet been reached. Nanopore still produced too many sequencing errors with deletion being the most common (Deamer et al. 2016). Still Nanopore seems to be the most promising techniques that will again reach read lengths suitable for COI barcoding or even full genome analyses. Until then even working with shorter than full length barcodes will still succeed anything that taxonomists are capable and willing to provide concerning routine mass-identifications of bulk samples.

Further great improvements can also be expected for the development of automated traps. The performance of the automated interval sampler (AIS) in the study yielded high quality samples that were all suited for DNA analyses. Especially important for the quality of the samples was also the use of high proof ethanol for the conservation of the specimens' right from the moment of collection. The construction showed to effectively prevent the evaporation of the ethanol during its deployment in the field. However, it showed to suffer from power shortage in the forests when its solar panels became shaded. In its latest version the new AIS was equipped with a new type of solar panel especially designed for diffuse light conditions. Also a battery with a higher capacity as well as a software adaption was provided to prevent power shortages in the future. Still, the automated interval sampler did not yet exploit its full potential as its

functional extensions and possible combinations with other capturing devices than Malaise traps make their possible applications numerous. As an autonomous system it could be easily used for Barber traps. Equipped with sensors it could also become an efficient portable weather station that provided samples with detailed environmental information. Sensors could be sensitive to sound and record a wide sonic spectrum that could be evaluated in spectrograms. Digital pictures could complement the genetic species information and provide reliable species abundances or be used to pre-evaluate a sample for its ecological and genetical value. Future developments in automation and digitalization techniques will help to meet the upcoming challenges of biodiversity assessments with its constant data growths and increased data management demands. Large-scale studies will benefit from this, because the barcoding approach allows the simultaneous identification of most species from a given biotope.

7 Bibliography

Aagaard K, Solem JO, Bongard T, Hanssen O (2004) Studies of aquatic insects in the Atna River 1987–2002. Hydrobiologia, 521, 87–105.

Aagaard K, BerggrenK, Hebert PDN, Sones J, McClenaghan B, Ekrem T (2017) Investigating suburban micromoth diversity using DNA barcoding of malaise trap samples. Urban Ecosystems. Volume 20, Issue 2, pp 353–361.

Abraham R. (1991). Fang und Präparation wirbelloser Tiere. Gustav Fischer Verlag, Stuttgart, New York. 132 S.

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71: 8966–8969. doi: 10.1128/AEM.71.12.8966-8969.2005

Adler PH and Foottit RG (2009) Introduction, in Insect Biodiversity: Science and Society (eds R. G. Foottit and P. H. Adler), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781444308211.ch1.

Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nature Methods, 8(1), 61-65. DOI: 10.1038/nmeth.1527

Al-Rshaidat MMD, Snider A, Rosebraugh S, Devine AM, Devine TD, Plaisance L, Knowlton N, Leray M (2016) Deep COI sequencing of standardized benthic samples unveils overlooked diversity of Jordanian coral reefs in the northern Red Sea. Genome 59:724–37. 10.1139/gen-2015-0208

Ammer U, Gossner M, Gruppe A, Simon U (2008) Integrating tree crown science with the development of 'near-to-nature' forest management practices: Examples from Bavaria. In: Floren A, Schmidl J, editors. Canopy Arthropod Research in Central Europe—basic and applied studies from the high frontier. Nürnberg: bioform entomology. pp. 531–549.

Austerlitz F, Olivier D, Schaeffer B, Bleakley K, Olteanu M, Leblois R, et al. (2009) DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. BMC Bioinformatics. 10(Suppl 14):S10.

Andrews S, Lindenbaum P, Howard B, Ewels P (2011) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Armitage PD, Cranston PS, Pinder LCV (1995) The Chironomidae. Biology and ecology of non-biting midges. London: Chapman & Hall.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol., 71, 7724–7736.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New Screening Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain Chimeras . Applied and Environmental Microbiology. 72(9):5734-5741. doi:10.1128/AEM.00556-06.

Astrin JJ, Stüben PE (2008) Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). Invertebrate Systematics 22: 503–522. doi: 10.1071/IS07057

Astrin JJ, Stüben PE, Misof B, Wägele J, Gimnich F, Raupach MJ, Ahrens D (2012) Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation. Molecular Phylogenetics and Evolution, 63(1):1–14.

Baird DJ, Pascoe TJ, Zhou X, Hajibabaei M (2011) Building freshwater macroinvertebrate DNA-barcode libraries from reference collection material: formalin preservation vs specimen age. J North Am Benthol Soc 30: 125–130.

Bálint M, Nowak C, Márton O, Pauls S, Wittwer C, Aramayo JL, Schulze A, Chambert T, Cocchiararo B, Jansen M (2017) Twenty-five species of frogs in a liter of water: eDNA survey for exploring tropical frog diversity. bioRxiv 176065; doi: https://doi.org/10.1101/176065

Ball SG (1984) Seasonal abundance during the summer months of some cattle-visiting Muscidae (Diptera) in northeast England. Ecological Entomology, 9: 1–10. doi:10.1111/j.1365-2311.1984.tb00692.x

Bartak M (1998) Diptera of the Bavarian forest. Diptera Bavorskeho lesa. - Silva Gabreta 2: 239-258, Vimperk.

Bährmann R, Weber D (2013) Dungfliegen (Insecta, Diptera, Sphaeroceridae) aus Höhlen des Großherzogtums Luxemburg. Ferrantia 69, Musée national d'histoire naturelle, Luxembourg, 354–368.

Becker N, Schön S, Klein, A-M, Ferstl I, Kizgin A, Tannich E, Kuhn C, Pluskota B, Jöst A (2017) First mass development of Aedes albopictus (Diptera: Culicidae)—its surveillance and control in Germany. Parasitology Research, 116(3), 847–858. http://doi.org/10.1007/s00436-016-5356-z

Becker S, Hanner R, Steinke D (2011) Five years of FISH-BOL: brief status report. Mitochondrial DNA , 22 (Suppl. 1), 3 – 9.

Bellamy L, Chapman N, Fowler K, Pomiankowski A (2013) Sexual traits are sensitive to genetic stress and predict extinction risk in the stalk-eyed fly Diasemopsis meigenii. Evolution, 67, 2662–2673.

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology 10(1): 189.

Bensasson D, Zhang D, Hartl D, Hewitt G (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution, 16(6):314– 321.

Berg CO and Knutson L (1978) Biology and systematics of the Sciomyzidae. Annual Review of Entomology 23: 239–258.

Bickford D, Lohman D, Sodhi N, Ng P, Meier R, Winker K, Ingram K, Das I (2007) Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution, 22(3):148–155.

Biodiversity Exploratories. Exploratorien: Schorfheide-Chorin; [accessed 2017 Sept 21]. http://www.biodiversity-exploratories.de/1/exploratories/schorfheide-chorin/

Biosphärenreservat Schorfheide-Chorin. Landscape; [accessed 2017 Sept 21]. www.schorfheide-chorin.de/seite/109463/landscape.html

Birch JL, Walsh NG, Cantrill DJ, Holmes GD, Murphy DJ (2017) Testing efficacy of distance and tree-based methods for DNA barcoding of grasses (Poaceae tribe *Poeae*) in Australia. PLOS ONE 12(10): e0186259. https://doi.org/10.1371/journal.pone.0186259

Bisanti M, Ganassi S, Mandrioli M (2009) Comparative analysis of various fixative solutions on insect preservation for molecular studies. Entomologia Experimentalis et Applicata 130: 290–296.

Biström O, Väisänen R (1988) Ancient-forest invertebrates of the Pyhän-Häkki National Park in central Finland. Acta Zool. Fennica 185, 1-69.

Blaxter ML (2004) The promise of a DNA taxonomy. Phil. Trans. R. Soc. Lond. B 2004 359 669-679; DOI: 10.1098/rstb.2003.1447.

Blaxter ML (2016) Imagining Sisyphus happy: DNA barcoding and the unnamed majority. Philosophical Transactions of the Royal Society B: Biological Sciences. 371(1702):20150329. doi:10.1098/rstb.2015.0329.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. http://doi.org/10.1093/bioinformatics/btu170

Bonduriansky R (2006) Convergent evolution of sexual shape dimorphism in Diptera. J. Morphol., 267: 602–611. doi:10.1002/jmor.10426

Bonfield JK, Smith KF, Staden R (1995) A new DNA sequence assembly program, Nucleic Acids Research, Volume 23, Issue 24, Pages 4992–4999. doi:10.1093/nar/23.24.4992

Borkent A and Bissett B (1990) A revision of the Holarctic species of Serromyia Meigen (Diptera: Ceratopogonidae). Systematic Entomology, 15(2): 153-217.

Borkent C and Harder L (2007) Flies (Diptera) as pollinators of two dioecious plants: Behaviour and implications for plant mating. The Canadian Entomologist, 139(2), 235-246. doi:10.4039/n05-087.

Bouget C, Brustel H, Brin A, Valladares L (2009) Evaluation of window flight traps for effectiveness at monitoring dead wood-associated beetles: the effect of ethanol lure under contrasting environmental conditions. Agricultural and Forest Entomology 11: 143–152.

Bradshaw CJ, Sodhi NS, and Brook BW (2009) Tropical turmoil: a biodiversity tragedy in progress. Frontiers in Ecology and the Environment, 7(2):79–87.

Brenner S (2004) Buckelfliegen (Diptera, Phoridae) in Auwäldern entlang der Etsch (Südtirol, Italien) Gredleriana Vol. 4 / 2004 pp. 295 – 304.

Brodeur JC and Candioti JV (2017) Chapter 8: Impacts of Agriculture and Pesticides on Amphibian Terrestrial Life Stages: Potential Biomonitor/Bioindicator Species for the Pampa Region of Argentina. Ecotoxicology and Genotoxicology: Non-traditional Terrestrial Models, The Royal Society of Chemistry: 163-194.

Brown WM, George M, & Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America, 76(4), 1967–1971.

Brown WM (1985) The mitochondrial genome of animals. In: Molecular Evolutionary Genetics, R.J. Maclntyre (ed.). New York: Plenum Press, pp. 95-130.

Brown BV (2001) Flies, gnats, and mosquitoes. Pp. 815–826. In Levin SA (ed). Encylopedia of Biodiversity. Academic Press, London.
Buck M (1994) Sphaeroceridae (Diptera) reared from various types of carrion and other decaying substrates in Southern Germany, including new faunistic data on some rarely collected species. Eur. J. Entomol. 94:137–151.

Bucklin A, Steinke D and Blanco-Bercial L (2011) DNA barcoding of marine metazoa. Annu. Rev. Mar. Sci. 3:471–508.

Buhay J (2009) "COI-like" Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. Journal of Crustacean Biology, 29(1):96– 110.

Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (2015) Weltnaturerbe Buchenwälder Einzigartig und schützenswert. Druck- und Verlagshaus Zarbock GmbH & Co. KG, Frankfurt am Main.

Butchart SHM, Walpole M, Collen B, Van Strien A, Scharlemann JPW, Almond REA, Baillie JEM, Bomhard B, Brown C, Bruno J, Carpenter KE, Carr GM, Chanson J, Chenery AM, Csirke J, Davidson NC, Dentener F, Foster M, Galli A, Galloway JN, Genovesi P, Gregory RD, Hockings M, Kapos V, Lamarque JF, Leverington F, Loh J, McGeoch MA, McRae L, Minasyan A, Morcillo MH, Oldfield TEE, Pauly D, Quader S, Revenga C, Sauer JR, Skolnik B, Spear D, Stanwell-Smith D, Stuart SN, Symes A, Tierney M, Tyrrell TD, Vié JC & Watson R (2010) Global biodiversity: Indicators of recent declines. Science, vol 328, no. 5982, pp. 1164-1168. DOI: 10.1126/science.1187512

Byers JA (1992) Attraction of bark beetles, Tomicus piniperda, Hylurgops palliatus, and Trypodendron domesticum and other insects to short-chain alcohols and monoterpenes. Journal of Chemical Ecology 18: 2385–2402. doi: 10.1007/BF00984957 PMID: 24254878

Byers JA (2007) Chemical ecology of bark beetles in a complex olfactory landscape. In: Lieutier F, Day KR, Battisti A, Grégoire J-C, Evans HF, editors. Bark and Wood Boring Insects in Living Trees in Europe, A Synthesis. Dordrecht, The Netherlands: Springer. pp. 89–134.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High - resolution sample inference from Illumina amplicon data. Nat. Methods , 13 , 581.

Cameron S, Rubinoff D, Will K (2006) Who Will Actually Use DNA Barcoding and What Will It Cost? Systematic Biology, Volume 55, Issue 5, Pages 844–847.
https://doi.org/10.1080/10635150600960079

Caminade C, Medlock JM, Ducheyne E, McIntyre KM, Leach S, Baylis M & Morse AP (2012) Suitability of European climate for the Asian tiger mosquito Aedes albopictus: recent trends and future scenarios. Journal of the Royal Society Interface, 9(75), 2708–2717.
http://doi.org/10.1098/rsif.2012.0138

Canal NA, Hernández-Ortiz V, Salas JOT, Selivon D (2015) Morphometric study of third-instar larvae from five morphotypes of the Anastrepha fraterculus cryptic species complex (Diptera, Tephritidae). ZooKeys.(540):41-59. doi:10.3897/zookeys.540.6012.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail ,P Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S (2012) Biodiversity loss and its impact on humanity. Nature, 486, 59–67.

Cardinale BJ1, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S.

Cardoso P (2009) Standardization and optimization of arthropod inventories-the case of Iberian spiders. Biodiversity and Conservation 18: 3949–3962.
https://link.springer.com/article/10.1007%2Fs10531-009-9690-7

Casiraghi M, Labra M, Ferri E, Galimberti A, De Mattia F (2010) DNA barcoding: a six-question tour to improve users' awareness about the method. Brief Bioinform **11**(4): 440-453.

Cervantes-Cervantes M, Calderón-Salinas JV, Albores A, Muñoz-Sánchez J (2005) Copper increases the damage to DNA and proteins caused by reactive oxygen species. Biological Trace Element Research. 103: 229–248. PMID: 15784956

Chao A (1984) Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11, 265-270.

Chao A and Shen TJ (2003) User's Guide for Program SPADE (Species Prediction and Diversity Estimation). Available at: http://chao.stat.nthu.edu.tw (accessed 05 September 2017)

Chao A and Chiu CH (2016) Species richness: estimation and comparison. Wiley StatsRef: Statistics Reference Online. 1-26.

Chao A and Lee S-M (1992) Estimating the number of classes via sample coverage. Journal of the American Statistical Association 87, 210-217.

Chao A, Shen T-J, Hwang WH (2006) Application of Laplace's boundary-mode approximations to estimate species and shared species richness. Australian and New Zealand Journal of Statistics 48, 117-128.

Chapman NC, Siriwat P, Howie J, Towlson A, Bellamy L, Fowler K, Pomiankowski A (2017) The complexity of mating decisions in stalk-eyed flies. Ecol Evol. 2017;7:6659–6668. https://doi.org/10.1002/ece3.3225

Chase M, Salamin N, Wilkinson M, Dunwell J, Kesanakurthi R, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and longterm goals. Philosophical Transactions of the Royal Society B-Biological Sciences, 360(1462):1889–1895.

Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui L-C & Scherer SW (2003) Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. Genome Biology. 4(4):R25. doi:10.1186/gb-2003-4-4-r25.

Chippindale AK, Leroi AM, Kim SB, Rose MR (2004) Phenotypic plasticity and selection in Drosophila life-history evolution. I. Nutrition and the cost of reproduction. Methuselah Flies: pp. 122-144.

Cini A, Ioriatti C, Anfora G (2012) A review of the invasion of Drosophila suzukii in Europe and a draft research agenda for integrated pest management. Bull Insectol. 2012;65(1):149–60.

Clausnitzer V, Kalkman VJ, Ram M, Collen B, Bailliec JEM, Bedjanič M, Darwall WRT, Dijkstra K-DB, Dow R, Hawking J, Karube H, Malikova E, Paulson D, Schütte K, Suhling F, Villanueva RJ, von Ellenrieder N, Wilson K (2009). "Odonata enter the biodiversity crisis debate: The first global assessment of an insect group." Biological Conservation 142(8): 1864-1869.

Clutton-Brock, T. H. and G. A. Parker. 1992. Potential reproductive rates and the operation of sexual selection. Q. Rev. Biol 67:437–456.

Coffey MD (1966) Studies on the association of flies (Diptera) with dung in south- eastern Washington. Ann. Entomol. Soc. Am. 59:207–18, Hafez M. 1939. Some ecological observations of on the insect fauna of dung. Bull. Soc. Fouad. I Entomol. 23:241–87, Hammer O. 1942.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR and Tiedje JM (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research, 42(Database issue), D633–D642. http://doi.org/10.1093/nar/gkt1244

Colless DH and McAlpine DK (1991) Diptera (flies). Pp. 717–786. In Commonwealth Scientific and Industrial Research Organisation (Division of Entomology).The Insects of Australia. Second Edition. Volume 2. Cornell University Press, Ithaca, NY. vi + 543–1137 pp.

Collins RA, Cruickshank RH (2012) The seven deadly sins of DNA barcoding. Mol Ecol Resour. 13(6):969–75

Contreras Gutiérrez MA, Vivero RJ, Vélez ID, Porter CH, Uribe S (2014) DNA Barcoding for the Identification of Sand Fly Species (Diptera, Psychodidae, Phlebotominae) in Colombia. PLOS ONE 9(1): e85496. https://doi.org/10.1371/journal.pone.0085496.

Cornel AJ, Porter CH, Collins FH (1996) Polymerase Chain Reaction Species Diagnostic Assay for Anopheles quadrimaculatus Cryptic Species (Diptera: Culicidae) Based on Ribosomal DNA ITS2

Sequences, Journal of Medical Entomology, Volume 33, Issue 1, Pages 109–116, doi:10.1093/jmedent/33.1.109

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? Science 339, 413–416. (doi:10.1126/science.1230318)

Courtney GW and Merritt RW (2008) Aquatic Diptera: Part one: larvae of aquatic Diptera. Pp. 687–722. In Merritt RW, Cummins KW, and Berg MB (eds). An Introduction to the Aquatic Insects of North America. Fourth Edition. Kendall/Hunt Publishing, Dubuque, Iowa. 1158 pp.

Courtney GW, Pape T, Skevington JH and Sinclair BJ (2009) Biodiversity of Diptera, in Insect Biodiversity: Science and Society (eds R. G. Foottit and P. H. Adler), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781118945568.ch9

Cronn R, Cedroni M, Haselkorn T, Grover C. Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton." Theoretical and Applied Genetics 104(2): 482-489.

Crooks J (2002) Characterizing Ecosystem-Level Consequences of Biological Invasions: The Role of Ecosystem Engineers. *Oikos, 97*(2), 153-166. Retrieved from http://www.jstor.org/stable/3547405

Dahlem GA (1991) Sarcophagidae. In: Immature Insects, ed. FW Stehr, 2:871–73. Dubuque, IA: Kendall/Hunt.

Danks HV and Smith ABT (2009) Insect Biodiversity in the Nearctic Region, in Insect Biodiversity: Science and Society (eds R. G. Foottit and P. H. Adler), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781444308211.ch3

Darriet P, Bouchilloux P, Poupot C, Bugaret Y, Clerjeau M, Sauris P, et al. (2001) Effects of copper fungicide spraying on volatile thiols of the varietal aroma of Sauvignon blanc, Cabernet Sauvignon and Merlot wines. Vitis 40: 93–99.

Darwin C (1871) The descent of man and selection in relation to sex. John Murray, London.

Deamer DW and Akeson M (2000) Nanopores and nucleic acids: prospects for ultrarapid sequencing. Trends in Biotechnology. Volume 18, Issue 4, p 147-151. https://doi.org/10.1016/S0167-7799(00)01426-8

Deamer DW, Akeson M, Branton D (2016) Three decades of nanopore sequencing. Nature Biotechnology. Volume 34, pages 518–524.

DeBach P & Rosen D(1991) Biological Control by Natural Enemies. Second Edition. Cambridge University Press, Cambridge. 440 pp.

Delettre YR, Trehén P, Grootaert P (1992) Space heterogeneity, space use, space use and short range dispersal in Diptera: A case study. Landscape Ecol. 6, 175–181.

Delettre YR, Morvan N, Trehén P, Grootaert P (1997) Local biodiversity and multi-habitat use in empidoid flies (Insecta: Diptera, Empidoidea). Biodiversity and Conservation 7: 9. doi:10.1023/A:1008851511301

Derocles SAP, Evans DM, Nichols PC, Evans SA, Lunt DH (2015) Determining Plant – Leaf Miner – Parasitoid Interactions: A DNA Barcoding Approach. PLOS ONE 10(2): e0117872. doi:10.1371/journal.pone.0117872

Dillon N, Austin AD, Bartowsky E (1996) Comparison of preservation techniques for DNA extraction from hymenopterous insects. Insect Molecular Biology 5: 21–24. PMID: 8630531

Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJ and Collen B (2014) Defaunation in the Anthropocene. Science, 345(6195):401–406. https://www.ncbi.nlm.nih.gov/pubmed/25061202

Disney RHL (2006) Revision of the Palaearctic members of the species complex resembling Megaselia brevior (Schmitz) (Diptera: Phoridae). Fragmenta Faunistica. 49(1), 41–51.

Doledec S, Chessel D, terBraak CJF, Champely S (1996) Matching species traits to environmental variables: A new three-table ordination method. Environmental and Ecological Statistics 3: 143–166.

Dorow W, Blick T, Kopelke J (2010) Zoologische Forschung in hessischen Naturwaldreservaten-Exemplarische Ergebnisse und Perspektiven. Zoological research in Strict Forest Reserves in Hesse (Germany)- Exemplary results and prospects. Forstarchiv 8: 61–68.

Dowton M, Austin AD (2001) Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita — evolutionary transitions among parasitic wasps, Biological Journal of the Linnean Society, Volume 74, Issue 1, Pages 87–111. https://doi.org/10.1111/j.1095-8312.2001.tb01379.x

Dray S, Choler P, Dolédec S, Peres-Neto PR, Thuiller W, Pavoine S, ter Braak CJ (2014) Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation. Ecology 95: 14–21.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22: 1–20.

Dray S, Legendre P (2008) Testing the species traits-environment relationships: the fourth-corner problem revisited. Ecology 89: 3400–3412. PMID: 19137946

Dufrêne M and Legendre P (1997) Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs, 67: 345–366.

Durska E, Bonet J, Viklund B (2010). The scuttle fly (Diptera: Phoridae) assemblages of a wildfire- affected hemiboreal old-growth forest in Tyresta (Sweden). Entomologica Fennica. 21.Sweden.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST, Bioinformatics, Volume 26, Issue 19, Pages 2460–2461. doi:10.1093/bioinformatics/btq461

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads . Nat. Methods , 10, 996 –998

Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads, Bioinformatics, Volume 31, Issue 21, Pages 3476–3482. https://doi.org/10.1093/bioinformatics/btv401.

Edgar RC (2016) UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27 (16): 2194-2200. doi:10.1093/bioinformatics/btr381

Edwards FW (1938) On the British Lestremiinae, with notes on exotic species (Diptera, Cecidomyiidae). Proc. R. ent. Soc. Lond. (B), 7: 199-210.

Ekrem T, Stur E, Hebert PDN (2010) Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding. Organisms Diversity & Evolution 10(5): 397-408.

Elberling H and Olesen JM (1999) The structure of a high latitude plant-flower visitor system: the dominance of flies. Ecography, 22(3), 314-323. doi: 10.1111/j.1600-0587.1999.tb00507.x.

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(7): e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht V, Peinert B, Leese F (2017) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecol Evol. 7:6918–6926. https://doi.org/10.1002/ece3.3192

Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J, Coissac E, Boyer F, Leese F (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4:e1966 https://doi.org/10.7717/peerj.1966

Eldredge N, Cracraft J (1980) Phylogenetic Patterns and The Evolutionary Process: Method and Theory in Comparative Biology. New York: Columbia University Press.

Ence DD, Carstens BC (2011) SpedeSTEM: a rapid and accurate method for species delimitation. Mol. Ecol. Resour. 11, 473–480. doi:10.1111/j.1755-0998.2010.02947.x

Engel K, Ammer U (2003) Vergleich verschiedener Fanglösungen. Faunistisch-Ökologische Mitteilungen 8: 297–301.

Epsky ND, Morrill WL, Mankin RW (2008) Traps for Capturing Insects. In: Capinera J.L. (eds) Encyclopedia of Entomology. Springer, Dordrecht

Ernani FP and LeProust EM (2016) Agilent's SureSelect Target Enrichment System: Agilent's SureSelect Target Enrichment System: Bringing Cost and Process Efficiency to Next-Generation Sequencing. Product Note. Agilent Technologies, Inc., 2009, 2016. Publication Number 5990-3532EN PR7000-0328

Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J (2016) Identifying cryptic diversity with predictive phylogeography. Proceedings of the Royal Society B: Biological Sciences, 283(1841), 20161529. http://doi.org/10.1098/rspb.2016.1529

Evenhuis NL, Pape T, Pont AC and Thompson FC (eds) (2007) BioSystematics Database of World Diptera, Version 9.5. http://www.dipter a.org/biosys.htm [Accessed 20 January 2008].

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. Genetics. 1992;131(2):479-491.

Ferguson JWH (2002) On the use of genetic divergence for identifying species. Biol. J. Linn. Soc. 75:509–516.

Ferrar P (1987) A guide to the breeding habits and immature stages of Diptera Cyclorrhapha. Entomonograph 8: 1–907.

Fessl B, Sinclair BJ & Kleindorfer S (2006) The life-cycle of Philornis downsi (Diptera: Muscidae) parasitizing Darwin's finches and its impacts on nestling survival. Parasitology, 133(6), 739-747. doi:10.1017/s0031182006001089.

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsemair KE, Schulze E-D, Weisser WW (2010). Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology, 11(6), 473-485.

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. Mol Ecol; 11:839–50.

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. Molecular Marine Biology and Biotechnology, 3 294 -299.

Forsdyke, DR (2017) Base Composition, Speciation, and Why the Mitochondrial Barcode Precisely Classifies. Biol Theory. 12: 157. https://doi.org/10.1007/s13752-017-0267-5

Fountain-Jones NM, Baker SC, Jordan GJ (2014) Moving beyond the guild concept: developing a practical functional trait framework for terrestrial beetles. Ecological Entomology 40: 1–13.

Frouz J (1999) Use of soil dwelling Diptera (Insecta, Diptera) as bioindicators: a review of ecological requirements and response to disturbance, Agriculture, Ecosystems & Environment, Volume 74, Issue 1, Pages 167-186, ISSN 0167-8809, http://dx.doi.org/10.1016/S0167-8809(99)00036-5.

Funk D and Omland K (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology Evolution and Systematics, 34:397–423.

Gagné RJ, Blanco-Metzler H, Etienne J (2000) A new Neotropical species of Clinodiplosis (Diptera: Cecidomyiidae), an important new pest of cultivated peppers (Capsicum spp.: Solanaceae). Proceedings of the Entomological Society of Washington 102: 831–837.

Geiger M, Morinière J, Hausmann A, Haszprunar G, Wägele W, Hebert P, 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. http://dx.doi.org/10.3897/bdj.4.e10671.

Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M (2015) Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. PLoS ONE 10, e0138432.

Glavendekić MM, Mihajlović LS (2006) Exposure of noctuid and geometrid development stages in oak forests. IUFRO Working Party 7.03.10 Proceedings of the Workshop 2006, Gmunden/Austria.

Gminder A, Kaiser A, Krieglsteiner GJ, Winterhoff W (2000) Ständerpilze: Leisten-, Keulen-, Korallen- und Stoppelpilze, Bauchpilze, Röhrlings- und Täublingsartige. In: G. J. Krieglsteiner (Hrsg.): Die Großpilze Baden-Württembergs. Band 2. Eugen Ulmer, Stuttgart. ISBN 978-3-8001-3531-8.

Golding YC, Edmunds M (2000) Behavioural mimicry of honeybees (Apis mellifera) by droneflies (Eristalis spp., Diptera, Syrphidae). Proc. R. Soc. Lond. B 267:903–909.

Gonzalez A., Cardinale BJ, Allington GRH, Byrnes J, Endsley KA, Brown DG, Hooper DU, Isbell F, O'Connor MI, Loreau M (2016) Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. Ecology, 97: 1949–1960. doi:10.1890/15-1759.1

Gossner MM, Floren A, Weisser WW, Linsenmair KE (2013) Effect of dead wood enrichment in the canopy and on the forest floor on beetle guild composition. Forest Ecology and Management 302: 404– 413.

Gossner MM, Fonseca CR, Pašalić E, Türke M, Lange M, WeisserWW (2014) Limitations to the use of arthropods as temperate forests indicators. Biodiversity and Conservation 23: 945–962.

Gossner MM, Schall P, Ammer C, Ammer U, Engel K, Schubert H, Simon U, Utschick H, Weisser WW (2014) Forest management intensity measures as alternative to stand properties for quantifying effects on biodiversity. Ecosphere 5: art113.

Gossner MM, Struwe J-F, Sturm S, Max S, McCutcheon M, Weisser WW, Zytynska SE (2016) Searching for the Optimal Sampling Solution: Variation in Invertebrate Communities, Sample Condition and DNA Quality. PLoS ONE 11(2): e0148247. https://doi.org/10.1371/journal.pone.0148247

Gotelli NJ and Colwell RK (2011) Estimating species richness. – In: Magurran AE and McGill BJ (eds), Frontiers in measuring biodiversity. Oxford Univ. Press , pp. 39 – 54 .

Graham K (1968) Anaerobic induction of primary chemical attractancy for ambrosia beetles. Canadian Journal of Zoology 46: 905–908.

Green NM (1975). "Avidin". Advances in Protein Chemistry. 29: 85–133. doi:10.1016/s0065-3233(08)60411-8.

Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DZ, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, The Human Microbiome Consortium, Joseph F. Petrosino, Rob Knight, and Bruce W. Birren (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome research, 21(3), 494-504.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. PLoS ONE 6(4): e17497. https://doi.org/10.1371/journal.pone.0017497

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics, Trends in Genetics, Volume 23, Issue 4, Pages 167-172, ISSN 0168-9525, https://doi.org/10.1016/j.tig.2007.02.001.

Hajibabaei M, Spall J, 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, Vol. 12, No. 1. 28, doi:10.1186/1472-6785-12-28

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T, Goulson D, de Kroon H (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. Lamb EG, ed. PLoS ONE; 12(10):e0185809. doi:10.1371/journal.pone.0185809.

Hansen MC, Tolker-Nielsen T, Givskov M, Molin S (1998) Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region, FEMS Microbiology Ecology, Volume 26, Issue 2, Pages 141–149, https://doi.org/10.1111/j.1574-6941.1998.tb00500.x

Harbach RE (2007) The Culicidae (Diptera): a review of taxonomy, classification and phylogeny. Zootaxa 1668: 591–638.

Hardy D and McGuire J (1947) The Nearctic Ptiolina (Rhagionidae-Diptera) (Rhagionidae-Diptera). Journal of the Kansas Entomological Society, 20(1), 1-15.

Hardy S, Fallow K, Barkley P (2007) Using copper sprays to control diseases in citrus. Primefact 757: 1–5.

Harley E, Birge LM, Small J, Tazzyman SJ, Pomiankowski A, Fowler K (2013) Ejaculate investment and attractiveness in the stalk-eyed fly, Diasemopsis meigenii. Ecology and Evolution, 3, 1529–1538.

Harris MA, Gardner WA, Oetting RD (1996) A Review of the Scientific Literature on Fungus Gnats (Diptera: Sciaridae) in the Genus Bradysia. Journal of Entomological Science: July 1996, Vol. 31, No. 3, pp. 252-276.

Haszprunar, G (2009). Barcoding Fauna Bavarica–eine Chance für die Entomologie. *NachrBl bayer Ent*, *58*(1/2),45.

Haupt J, Haupt H (1995) Fliegen und Mücken. Beobachtung, Lebensweise. 1. Auflage. Naturbuch-Verlag, Jena und Stuttgart.

Hausdorf B, Hennig C (2010) Species delimitation using dominant and codominant multilocus markers. Syst. Biol. 59, 491–503. doi:10.1093/sysbio/syq039

Hauser M, Geller-Grimm F (1996) Schwebfliegen (Diptera: Syrphidae) des Messeler Forstes (Südhessen) mit besonderen Bemerkungen zu Mallota cimbiciformis (Fallén, 1817). Volucella 2:65–76.

Head SR, Komori HK, LaMere SA, Whisenant T, van Nieuwerburgh F, Salomon DR, & Ordoukhanian P (2014) Library construction for next-generation sequencing: Overviews and challenges. BioTechniques, 56(2), 61–passim. http://doi.org/10.2144/000114133

Heine K, Reuther AU, Thieke H, Schulz R, Schlaak N, Kubik PW (2009) Timing of Weichselian ice marginal positions in Brandenburg (northeastern Germany) using cosmogenic in situ ^{10}Be. Zeitschrift für Geomorphologie. 53. 433-454. 10.1127/0372-8854/2009/0053-0433.

Herbert PDN, Ratnasingham S, deWaard JR (2003a) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B: Biological Sciences 270 : S96 -S99 . https://doi.org/10.1098/rsbl.2003.0025

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003b) Biological identifications through DNA barcodes. Proceedings of the Royal Society B. 270 (1512): 313–321. https://doi:10.1098/rspb.2002.2218

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences. 101 (41) 14812-14817. DOI: 10.1073/pnas.0406166101

Hebert PDN and Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54(5): 852-859.

Hedges DJ, Guettouche T, Yang S, Bademci G, Diaz A, Andersen A, Hulme WF, Linker S, Mehta A, Edwards YJ, Beecham GW, Martin ER, Pericak-Vance MA, Zuchner S, Vance JM, Gilbert JR (2011) Comparison of three targeted enrichment strategies on the SOLiD sequencing platform. PLoS One.29;6(4).

Herbert PDN, deWaard JR, Landry J (2009). DNA barcodes for 1/1000 of the animal kingdom . Biology Letters 6 (3): 359 -362 . https://doi.org/10.1098/rsbl.2009.0848

Hernández-Ortiz V, Bartolucci A, Morales-Valles P, Frias D, Selivon D (2012) Cryptic species of the Anastrepha fraterculus complex (Diptera:Tephritidae): A multivariate approach for the recognition of South American morphotypes. Annals of the Entomological Society of America 105: 305–318. doi: 10.1603/AN11123.

Hessenmöller D, Nieschulze J, Lüpke Nv, Schulze E-D (2011) Identification of forest management types from ground-based and remotely sensed variables and the effects of forest management on forest structure and composition. Forstarchiv 82: 171–183.

Hevia V, Martín-López B, Palomo S, García-Llorente M, de Bello F, González JA. (2017). Trait-based approaches to analyze links between the drivers of change and ecosystem services: Synthesizing existing evidence and future challenges. Ecology and Evolution, 7(3):831-844. https://doi: 10.1002/ece3.2692.

Hill M, Holm K, Vel T, Shah NJ, Matyot P (2003) Impact of the introduced yellow crazy ant Anoplolepis gracilipes on Bird Island, Seychelles. Biodiversity and Conservation, 12: 1969. https://doi.org/10.1023/A:1024151630204

Hill TC, Walsh KA, Harris JA, Moffett BF (2003) Using ecological diversity measures with bacterial communities. FEMS Microbiol Ecol. ;43(1):1-11. doi: 10.1111/j.1574-6941.2003.tb01040.x.

Hillebrand H, Blasius B, Borer ET, Blasius B, Borer ET, Chase JM, Downing JA, Eriksson BK, Filstrup CT, Harpole WS, Hodapp D, Larsen S, Lewandowska AM, Seabloom EW, van de Waal DB, Ryabov AB (2017) Biodiversity change is uncoupled from species richness trends: consequences for conservation and monitoring. J Appl Ecol. 2017;00:1–16. https://doi.org/10.1111/1365-2664.12959

Holloway GJ (1993) Phenotypic variation in colour pattern and seasonal plasticity in Eristalis hoverflies (Diptera: Syrphidae). Ecological Entomology, 18: 209–217. doi:10.1111/j.1365-2311.1993.tb01092.x

Holmes I, Harris K, Quince C (2012) Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. PLOS ONE 7(2): e30126. https://doi.org/10.1371/journal.pone.0030126

Holopainen J (1990) Influence of ethylene glycol on the numbers of carabids and other soil arthropods caught in pitfall traps. In: Stork N, editor. The role of ground beetles in ecological and environmental studies. Andover, England: Intercept Publications. pp. 339–341.

Holopainen J (1992) Catch and sex ratio of Carabidae (Coleoptera) in pitfall traps filled with ethylene glycol or water. Pedobiologia 36: 257–261.

Holway DA, Lach L, Suarez A, Tsutsui N, Case TJ (2002) The causes and consequences of ant invasions. Annu Rev Ecol Syst 33:181–233.

Honek A (1993) Intraspecific variation in body size and fecundity in insects: a general relationship. Oikos 66:483–492.

Höfer H, Astrin J, Holstein J, Spelda J, Meyer F, Zarte N (2015) Propylene glycol–a useful capture preservative for spiders for DNA barcoding. Arachnologische Mitteilungen 50: 30–36.

Hövemeyer K (2000) 1.11. Ecology of Diptera. Pp. 437–489. In L. Papp and B. Darvas (eds). Contributions to a Manual of Palaearctic Diptera. Volume1. General and Applied Dipterology. Science Herald, Budapest. 978 pp.

Hövemeyer K and Schauermann J (2003) Succession of Diptera on dead beech wood: a 10-year study. – Pedobiologia 47: 61–75.

Huber K, Pluskota B, Jöst A, Hoffmann K & Becker N (2012) Status of the invasive species *Aedes japonicus japonicus* (Diptera: Culicidae) in southwest Germany in 2011. Journal of Vector Ecology, 37: 462–465. doi:10.1111/j.1948-7134.2012.00252.x

Huber T, Faulkner G and Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics, 20, 2317–2319.

Hugall AF, O'Hara TD, Hunjan S, Nilsen R, Moussalli A (2015) An exon-capture system for the entire class Ophiuroidea. Mol Bio Evol . 33:281–294.

Hulcr J, Miller S, Setliff G, Darrow K, Mueller N, Hebert P, Weiblen G (2007) DNA barcoding confirms polyphagy in a generalist moth, Homona mermerodes (Lepidoptera : Tortricidae). Molecular Ecology Notes, 7(4):549–557.

Hurst G and Jiggins F (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B-Biological Sciences, 272(1572):1525–1534.

Huse SM, Welch DM, Morrison HG and Sogin ML (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology, 12(7), 1889–1898. http://doi.org/10.1111/j.1462-2920.2010.02193.x

Huson DH and Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies, Molecular Biology and Evolution, 23(2):254-267.

IUPAC (1997) Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford.

Jakovlev J (2014) Checklist of the fungus gnats of Finland: Bolitophilidae, Diadocidiidae, Ditomyiidae, Keroplatidae and Mycetophilidae (Diptera). ZooKeys, (441), 119–149.

Janzen DH, Hallwachs W (2016) DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Area de Conservacion Guanacaste, northwestern Costa Rica. Genome, 2016, 59 (9), 641-660, 10.1139/gen-2016-0005@gen-iblf.issue01.

Janzen DH, Burns JM, Cong Q, Hallwachs W, Dapkey T, Manjunath R, Hajibabaei M, Hebert PDN, Grishin NV (2017) Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. Proceedings of the National Academy of Sciences. 114. 201621504. 10.1073/pnas.1621504114.

Jaschhof M (1998) Revision der "Lestremiinae" (Diptera, Cecidomyiidae) der Holarktis. Studia Dipterologica Supplement 4: 1–552.

Jaschhof M (2009) Eine aktualisierte Artenliste der Holzmücken Deutschlands, unter besonderer Berücksichtigung der Fauna Bayerns. - Spixiana 32 (1): 139-151, München.

Jaschhof M, Skuhravá M, Penttinen J (2014) Checklist of the Cecidomyiidae (Diptera) of Finland. ZooKeys, (441), 103–118. http://doi.org/10.3897/zookeys.441.7503.

Jhumur US, Dötterl S, Jürgens A (2006) Naïve and Conditioned Responses of Culex pipiens pipiens Biotype molestus (Diptera: Culicidae) to Flower Odors. Journal of Medical Entomology. 43(6): 1164-1170.

Jose G, Santra S, Mandal S, Sengupta T (2011) Singlet oxygen mediated DNA degradation by copper nanoparticles: potential towards cytotoxic effect on cancer cells. Journal of Nanobiotechnology 9: 9. doi: 10.1186/1477-3155-9-9 PMID: 21439072

Jud P, Schmidt-Entling MH (2008) Fluid type, dilution, and bitter agent influence spider preservation in pitfall traps. Entomologia Experimentalis et Applicata 129: 356–359.

Juliano SA, Lounibos LP (2005) Ecology of invasive mosquitoes: effects on resident species and on human health. Ecol Lett 8:558–574

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). J. Biosci. Bioeng. 96:317–323.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. doi: 10.1093/bioinformatics/bts199 PMID: 22543367

Kehlmaier C, Almeida JM (2014) New host records for European Acroceridae (Diptera), with discussion of species limits of *Acrocera orbiculus* (Fabricius) based on DNA-barcoding. Zootaxa 3780(1): 135-152.

Keiper JB, Walton WE, Foote BA (2002) Biology and ecology of higher diptera from freshwater wetlands. Annual Review of Entomology 2002 47:1, 207-232.

Kellogg SK, Fink LS, Brower LP (2003) Parasitism of Native Luna Moths, Actias luna (L.) (Lepidoptera: Saturniidae) by the Introduced Compsilura concinnata (Meigen) (Diptera: Tachinidae) in Central Virginia, and Their Hyperparasitism by Trigonalid Wasps (Hymenoptera: Trigonalidae), Environmental Entomology, Volume 32, Issue 5, Pages 1019–1027, https://doi.org/10.1603/0046-225X-32.5.1019

Kenis M, Auger-Rozenberg M-A, Roques A, Timms L, Péré C, Cock M, Settele J, Augustin S, Lopez Vaamonde C (2009) Ecological effects of invasive alien insects. Biological Invasions, 11 (1), 21-45. DOI : 10.1007/s10530-008-9318-y http://prodinra.inra.fr/record/32305

Kenis M, Adriaens T, Brown P, Katsanis A, San Martin G, Branquart E, Maes D, Eschen R, Zindel R, van Vlaenderen J, Babendreier D, Roy H, Hautier L, Poland LR (2016) Assessing the ecological risk posed by a recently established invasive alien predator: Harmonia axyridis as a case study. BioControl. . 10.1007/s10526-016-9764-x.

Khruleva OA (1987) Invertebrates. Pp. 6–35. In Sokolov VE (ed). Flora and Fauna of Reserves of the USSR. Fauna zapovednika ‘‘Ostrov Vrangelya’’. Nauka, Moscow.

King JR and Porter SD (2004) Recommendations on the use of alcohols for preservation of ant specimens (Hymenoptera, Formicidae) Insect. Soc. 51: 197. doi:10.1007/s00040-003-0709-x

Kircher M, Stenzel U, Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol., 10, R83.

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform, *Nucleic Acids Research*, Volume 40, Issue 1, Pages e3, https://doi.org/10.1093/nar/gkr771

Knapp S, Schweiger O, Kraberg A, Asmus H, Asmus R, Brey T, Frickenhaus S, Gutt J, Kühn I, Liess M, Musche M, Portner H-O, Seppelt R, Klotz S, Krause G (2017) Do drivers of biodiversity change differ in importance across marine and terrestrial systems or is it just different research communities’ perspectives? Sci Total Environ 574:191–203. doi:10.1016/j.scitotenv. 2016.09.002

Knutie SA, Herman JM, Owen JP, and Clayton DH (2017) Tri-trophic ecology of native parasitic nest flies of birds in Tobago. Ecosphere 8(1):e01670. 10.1002/ecs2.1670

Koivula M, Kotze DJ, Hiisivuori L, Rita H (2003) Pitfall trap efficiency: do trap size, collecting fluid and vegetation structure matter? Entomologica Fennica 14: 1–14.

Koop JAH, Kim PS, Knutie SA, Adler F and Clayton DH (2016) An introduced parasitic fly may lead to local extinction of Darwin's finch populations. J Appl Ecol, 53: 511–518. doi:10.1111/1365-2664.12575

Kozarewa I, Armisen J, Gardner AF, Slatko BE, Hendrickson CL (2015) Overview of target enrichment strategies. Curr. Protoc. Mol. Biol. 112:7.21.1-7.21.23.

Köhler F (2000) Totholzkäfer in Naturwaldzellen des nördlichen Rheinlands; Landesanstalt für Ökologie BuFLfAN-W, editor. Recklinghausen: Landesanstalt für Ökologie, Bodenordnung und Forsten / Landesamt für Agrarordnung Nordrhein-Westfalen. 352 p.

Kraushaar U and Blanckenhorn WU (2002) Population variation in sexual selection and its effect on body size allometry in two species of flies with contrasting sexual size dimorphism. Evolution 56:307–321.

Krcmar S, Mikuska A, Merdić E. (2006) Response of Tabanidae (Diptera) to different natural attractants. J Vector Ecol.31(2):262-5. PubMed PMID: 17249343.

Kremen C, Colwell R, Erwin T, Murphy D, Noss R, Sanjayan M (1993) Terrestrial arthropod assemblages: their use in conservation planning. Conservation Biology 7: 796-808. https://doi.org/10.1046/j.1523-1739.1993.740796.x

Krivosheina NP (2009) New data on the morphology of the larvae and pupae of xylophilous limoniids of the subgenus Austrolimnophila (s. str.) Alexander, 1920 (Diptera: Limoniidae). Russian Entomological Journal. Vol.18, No.3, pp. 221-228.

Krosch NM, Schutze MK, Armstrong KF, Boontop Y, Boykin LM, Chapman TA, Englezou A, Cameron SL, Clarke AR (2013) Piecing together an integrative taxonomic puzzle: microsatellite, wing shape and aedeagus length analyses of Bactrocera dorsalis s.l. (Diptera: Tephritidae) find no evidence of multiple lineages in a proposed contact zone along the Thai/Malay Peninsula. Systematic Entomology 38: 2–13. doi: 10.1111/j.1365-3113.2012.00643.x

Kruskal WH and Wallis WA (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* .47:583–621. doi:10.2307/2280779.

Krueger F, Andrews SR, Osborne CS (2011) Large Scale Loss of Data in Low-Diversity Illumina Sequencing Libraries Can Be Recovered by Deferred Cluster Calling. PLOS ONE 6(1): e16607. https://doi.org/10.1371/journal.pone.0016607.

Krüger A and Garms R (1999) Verification of the synonymy of *Simulium damnosum* cytoform 'Nyamagasani' with *Simulium kilibanum* Gouteux (Diptera: Simuliidae) together with descriptive data on related forms. Bulletin of Entomological Research, 89(6), 533-541. doi:10.1017/S0007485399000681

Kutz FW, Wade TG & Pagac BB (2003) A geospatial study of the potential of two exotic species of mosquitoes to impact the epidemiology of West Nile virus in Maryland. J. Am. Mosq. Contr. Assoc. 19: 190–198.

Kühner M (1992) Dipterengesellschaften (Brachycera - Ortorrhapha) in Landökosystemen Süddeutschlands. Teil II. Zool. Jb. Syst. 119, 169–223.

Lahr DJ, Katz LA (2009) Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a newgeneration high-fidelity DNA polymerase. Biotechniques 47: 857–866.

Landres PB, Verner J, Thomas JW (1988), Ecological Uses of Vertebrate Indicator Species: A Critique. Conservation Biology, 2: 316–328. doi:10.1111/j.1523-1739.1988.tb00195.x

Langton PH (1999) Micropsectra silvesterae n. sp. and Tanytarsus heliomesonyctios n. sp., (Diptera: Chironomidae), two parthenogenetic species from Ellesmere Island, Arctic Canada. Journal of the Kansas Entomological Society, 71, 208–215.

Larson B, Kevan P, Inouye D (2001) Flies and flowers: Taxonomic diversity of anthophiles and pollinators. The Canadian Entomologist, 133(4), 439-465. doi:10.4039/Ent133439-4.

Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, Craig JM, Langford KW, Samson JM, Daza R, Doering K, Shendure J, Gundlach JH (2014) Decoding long nanopore sequencing reads of natural DNA. Nature Biotechnology, 32(8), 829-833. DOI: 10.1038/nbt.2950

Laux HE (1985) Eßbare Pilze und ihre giftigen Doppelgänger. Franckh'sche Verlagshandlung, W. Keller & Co., Stuttgart.

Learn GH Jr, Korber BT, Foley B, Hahn BH, Wolinsky SM, Mullins JI (1996) Maintaining the integrity of human immunodeficiency virus sequence databases. J. Virol. vol. 70 no. 8 5720-5730

Lee PS, Sing KW, Wilson JJ (2015) Reading Mammal Diversity from Flies: The Persistence Period of Amplifiable Mammal mtDNA in Blowfly Guts (Chrysomya megacephala) and a New DNA Mini-Barcode Target. PLOS ONE 10(4): e0123871. doi:10.1371/journal.pone.0123871

Leclaire M & Brandl R (1994) Phenotypic plasticity and nutrition in a phytophagous insect: consequences of colonizing a new host. Oecologia. 100: 379. https://doi.org/10.1007/BF00317858

Lemmon IR, Emme SA, Lemmon EM (2012) Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics, Systematic Biology, Volume 61, Issue 5, Pages 727–744.

Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362: 709–715. PMID: 8469282

Lloyd DR, Phillips DH (1999) Oxidative DNA damage mediated by copper(II), iron(II) and nickel(II) Fenton reactions: evidence for site-specific mechanisms in the formation of double-strand breaks, 8-hydroxydeoxyguanosine and putative intrastrand cross-links. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 424: 23–36. PMID: 10064847

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, Hingamp P, Ogata H, de Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG (2013) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Metagenomics, and Biomes in Health and Disease. Metagenomics, Volume16, Issue9: 2659-2671.

Lopez J, Yuhki N, Masuda R, Modi W, Obrien S (1994) Numt, A Recent Transfer and Tandem Amplification of Mitochondrial-Dna to the Nuclear Genome of the Domestic Cat. Journal of Molecular Evolution, 39(2):174–190.

Lowe S, Browne M, Boudjelas S, De Poorter M (2000) 100 of the World's Worst Invasive Alien Species A selection from the Global Invasive Species Database . Published by: The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), 12pp. First published as special lift-out in Aliens 12, December 2000. Updated and reprinted version: November 2004.

Maddison WP (1997) Gene Trees in Species Trees. *Syst Biol*; 46 (3): 523-536. doi: 10.1093/sysbio/46.3.523

Maguire DY, Robert K, Brochu K, Larrivée M, Buddle CM, Wheeler TA (2014) Vertical Stratification of Beetles (Coleoptera) and Flies (Diptera) in Temperate Forest Canopies. Environmental Entomology 43: 9–17. doi: 10.1603/EN13056 PMID: 24472199

Maharaj R, Appleton CC, Miller RM (1992) Snail predation by larvae of Sepedon scapularis Adams (Diptera: Sciomyzidae), a potential biocontrol agent of snail intermediate hosts of schistosomiasis in South Africa. Medical and Veterinary Entomology 6: 183–187.

Maidak BL, Cole JR, Lilburn TG, Parker CT Jr, Saxman PR, Stredwick JM, Garrity GM, Li B, Olsen GJ, Pramanik S, Schmidt TM and Tiedje JM (2000) The RDP (Ribosomal Database Project) continues. Nucleic Acids Research, 28(1), 173–174.

Malaise R. (1937) A new insect-trap. Entomol. Tidskr. Stock.58: 148-160.

Mamaev BM, Økland B (1998) The genus Bryomyia Kieff. (Diptera, Cecidomyiidae): Palaearctic species and Fennoscandian records. Entomologica Fennica. Vol.9. p.147-152.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. Nature Methods. Volume 7, pages 111–118.

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 18:50–60. doi: 10.1214/aoms/1177730491.

Mayer C, Sann M, Donath A, Meixner M, Podsiadlowski L, Peters RS, Petersen M, Meusemann K, Liere K, Wägele J, Misof B, Bleidorn C, Ohl M, Niehuis O (2016) BaitFisher: A Software Package for Multispecies Target DNA Enrichment Probe Design. Molecular biology and evolution, 33 7, 1875-86.

McAlpine JF, Peterson BV, Shewell GE, Teskey HJ, Vockeroth JR, Wood DM (1981) Manual of Nearctic Diptera. Volume 1. Agriculture Canada Monograph No. 27. Research Branch, Agriculture Canada, Ottawa, ON.

McAlpine JF, Peterson BV, Shewell GE, Teskey HJ, Vockeroth JR, Wood DM (1987) Manual of Nearctic Diptera. Volume 2. Agriculture Canada Monograph No. 28. Research Branch, Agriculture Canada, Ottawa, ON.

McCravy KW, Willand JE (2007) Effects of pitfall trap preservative on collections of carabid beetles (Coleoptera: Carabidae). The Great Lakes Entomologist 40: 154–165.

McIntyre PB, Jones LE, Flecker AS, Vanni MJ (2006) Fish extinctions alter nutrient recycling in tropical freshwaters PNAS 2007 104 (11) 4461-4466. doi:10.1073/pnas.0608148104

McPheron BA & Steck GJ (1996) Fruit Fly Pests: A World Assessment of Their Biology and Management. St. Lucie Press, Delray Beach. 586 pp.

Meier R, Kwong S, Vaidya G, Ng PK (2006) DNA Barcoding and Taxonomy in Diptera: a Tale of High Intraspecific Variability and Low Identification Success. Systematic Biology, 55: 715-728.

Meier R, Wong W, Srivathsan A, Foo M (2016) $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. Cladistics, 32: 100–110. doi:10.1111/cla.12115

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, Volume 57, Issue 5, Pages 809–813, https://doi.org/10.1080/10635150802406343

Meyer C and Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. Plos Biology, 3(12):2229–2238.

Meyer M and Kircher M (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor protocols. pdb.prot5448. 10.1101/pdb.prot5448.

Menzel F and Schulz U (2007) Die Trauermücken in Deutschland – ökosystemare Bedeutung, zönologische Koinzidenzen und bioindikatorisches Potential (Diptera: Sciaridae). – Beiträge zur Entomologie 57 (1): 9-36, Keltern.

Merritt RW, Courtney GW and Keiper JB (2003) Diptera. Pp. 324–340. In Resh VH and Cardé RT (eds). Encyclopedia of Insects. Academic Press, London.

Merz B and Kofler A (2006) Zum Vorkommen von Hornfliegen in Osttirol und Kärnten (Österreich) (Diptera: Phaeomyiidae & Scimyz idae). - Ber. nat.-med. Ver.Innsbruck. 93:107-119 .

Merz B and Kofler A (2008) Zur Kenntnis der Lauxaniidae und Pallopteridae (Diptera, Acalyptratae) in Osttirol und Kärnten (Österreich). – Linzer biologische Beiträge 40 (2): 1225-1235, 8.

Messerli FH (2012) Chocolate consumption, cognitive function, and Nobel laureates New England Journal of Medicine In New England Journal of Medicine, Vol. 367, No. 16. pp. 1562-1564, doi:10.1056/nejmon1211064

Metzker ML (2010) Sequencing technologies—the next generation. Nature Reviews Genetics. doi: 10.1038/nrg2626

Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics. 9:214. doi:10.1186/1471-2164-9-214.

Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples, Nucleic Acids Res. Vol. 35 pg. e97

Miller RM (1977) Ecology of Lauxaniidae (Diptera: Acalyptratae). I. Old and new rearing records with biological notes and discussion. Ann. Natal Mus. 23:215– 38.

Miller S, Hausmann A, Hallwachs W, Janzen D (2016) Advancing taxonomy and bioinventories with DNA barcodes . Philosophical Transactions of the Royal Society B: Biological Sciences 371 (1702): 20150339. https://doi.org/10.1098/rstb.2015.0339

Munro VMW, Henderson IM (2002) Nontarget Effect of Entomophagous Biocontrol: Shared Parasitism Between Native Lepidopteran Parasitoids and the Biocontrol Agent Trigonospila brevifacies (Diptera: Tachinidae) in Forest Habitats, Environmental Entomology, Volume 31, Issue 2, Pages 388–396, https://doi.org/10.1603/0046-225X-31.2.388

Mursu J, Voutilainen S, Nurmi T, Rissanen TH, Virtanen JK, Kaikkonen J, Nyyssönen K, Salonen JT (2004) Dark Chocolate Consumption Increases HDL Cholesterol Concentration and Chocolate Fatty Acids May Inhibit Lipid Peroxidation in Healthy Humans, In Free Radical Biology and Medicine, Volume 37, Issue 9, 2004, Pages 1351-1359, ISSN 0891-5849. https://doi.org/10.1016/j.freeradbiomed.2004.06.002.

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 Biol 9: e1001127.

Mori AS, Lertzman KP and Gustafsson L (2017) Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. J Appl Ecol, 54: 12–27. doi:10.1111/1365-2664.12669

Morinière J, Cancian de Araujo B, Lam AW, Hausmann A, Balke M, Schmidt S, Hendrich L, Doczkal D, Fartmann B, Arvidsson S, Haszprunar G (2016) Species Identification in Malaise Trap Samples by DNA Barcoding Based on NGS Technologies and a Scoring Matrix. PLOS ONE 11(5): e0155497. https://doi.org/10.1371/journal.pone.0155497

Moreira D, Lopez-Garcia P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol; 10:31–8.

Morse H D (1971) The insectivorous bird as an adaptive strategy. Annual Review of Ecology and Systematics. p. 177–200.

Mullis KB, Faloona F (1987) Specific synthesis of DNA via a polymerase-catalyzed chain reaction, Methods in Enzymology, Volume 155, 1987, Pages 335-350, ISSN 0076-6879, http://dx.doi.org/10.1016/0076-6879(87)55023-6.

Nabli H, Bailey WC, Necibi S (1999) Beneficial Insect Attraction to Light Traps with Different Wavelengths, Biological Control, Volume 16, Issue 2, Pages 185-188, ISSN 1049-9644, https://doi.org/10.1006/bcon.1999.0748.

Newbold T, Hudson LN, Hill SL, Contu S, Lysenko I, Senior RA, Börger L, Bennett DJ, Choimes A, Collen B, Day J, De Palma A, Díaz S, Echeverria-Londoño S, Edgar MJ, Feldman A, Garon M, Harrison ML, Alhusseini T, Ingram DJ, Itescu Y, Kattge J, Kemp V, Kirkpatrick L, Kleyer M, Correia DL, Martin CD, Meiri S, Novosolov M, Pan Y, Phillips HR, Purves DW, Robinson A, Simpson J, Tuck SL, Weiher E, White HJ, Ewers RM, Mace GM, Scharlemann JP, Purvis A (2015) Global effects of land use on local terrestrial biodiversity. Nature 520:45–50.

Newton IL, Roeselers G (2012) The effect of training set on the classification of honey bee gut microbiota using the Naïve Bayesian Classifier , BMC Microbiol. Vol. 12 pg. 221.

Niemelä J (1997) Invertebrates and Boreal Forest Management. Conservation Biology, 11(3), 601-610.

Nilsson RH, Abarenkov K, Veldre V, Nylinder S, DE Wit P, Brosché S, Alfredsson JF, Ryberg M, Kristiansson E (2010) An open source chimera checker for the fungal ITS region. Mol Ecol Resour. 10(6):1076-81. doi: 10.1111/j.1755-0998.2010.02850.x.

Oberhauser K, Elmquist D, Perilla-López JM, Gebhard I, Lukens L, Stireman J (2017) Tachinid Fly (Diptera: Tachinidae) Parasitoids of Danaus plexippus (Lepidoptera: Nymphalidae), Annals of the Entomological Society of America, Volume 110, Issue 6, Pages 536–543, https://doi.org/10.1093/aesa/sax048

Obrist MK, Duelli P (2010) Rapid biodiversity assessment of arthropods for monitoring average local species richness and related ecosystem services. Biodiversity and Conservation 19: 2201–2220. https://link.springer.com/article/10.1007%2Fs10531-010-9832-y

O'Dowd DJ, Green PT and Lake PS (2003) Invasional 'meltdown' on an oceanic island. Ecology Letters, 6: 812–817. doi:10.1046/j.1461-0248.2003.00512.x

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) vegan: Community Ecology Package. R package version 2.0–10.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res; 44 (D1): D733-D745. doi: 10.1093/nar/gkv1189

Ollerton J, Winfree R, Tarrant S (2011) How many flowering plants are pollinated by animals? Oikos. 120(3):321–326.

Oliver L, Beattie AJ (1993) A possible method for the rapid assessment of biodiversity. Conservat Biol; 7:562–8.

O'Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59, 59–73. doi:10.1093/sysbio/syp077

Økland B (1996) Unlogged forests: Important sites for preserving the diversity of mycetophilids (Diptera: Sciaroidea). Biological Conservation, Volume 76, Issue 3, 1996, Pages 297-310, ISSN 0006-3207, https://doi.org/10.1016/0006-3207(95)00129-8.

Økland B, Götmark F and Nordén B (2008) Oak woodland restoration: testing the effects on biodiversity of mycetophilids in southern Sweden. Biodivers Conserv.17: 2599. https://doi.org/10.1007/s10531-008-9325-4.

Papp L (2002) Dipterous guilds of small-sized feeding sources in forest of Hungary. Acta Zoologica Academiae Scientiarum Hungarica. 48.

Peach DAH and Gries G (2016) Nectar thieves or invited pollinators? A case study of tansy flowers and common house mosquitoes. Arthropod-Plant Interactions 10(6): 497-506. https://doi.org/10.1007/s11829-016-9445-9

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. Clin. Chem. 55, 856–866.

Pfenninger M and Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evolutionary Biology, 7.

Phillippy AM, Schatz MC, Pop M. Genome assembly forensics: finding the elusive mis-assembly. Genome Biology. 2008;9(3):R55. doi:10.1186/gb-2008-9-3-r55.

Pietsch S, & Rulik B (2014) Eine genetische Nationalbibliothek der Artenvielfalt Deutschlands–Das German barcode of Life (GBOL) Projekt. Koenigiana, 8, 43-52.

Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2015) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–120.

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. Mol Ecol Resour, 15: 819–830. doi:10.1111/1755-0998.12355

Pluskota B, Artur J, Augsten X, Stelzner L, Ferstl I, Becker N (2016) Successful overwintering of Aedes albopictus in Germany. Parasitology Research. 115. Doi:10.1007/s00436-016-5078-2.

Poissant J, Wilson AJ, Coltman DW (2010) Sex-Specific Genetic Variance and the Evolution of Sexual Dimorphism: A Systematic Review of Cross-Sex Genetic Correlations. Evolution, 64(1):97–107.

Pokluda P, Čížek L, Stříbrná E, Drag L, Lukeš J, Novotný V (2014) A goodbye letter to alcohol: An alternative method for field preservation of arthropod specimens and DNA suitable for mass collecting methods. European Journal of Entomology 111: 175–179.

Pollard G (2000) FAO Assistance to Jamaica for improving hot pepper production, in part through development of an IPM programme for gall midges and broad mite. Caribbean IPM Knowledge Network, Current Awareness Bulletin 2: 1–6.

Pollet M, Meuffels H, Grootaert P (2003) Status of Dolichopodidae of the Flemish Red Data Book in the Netherlands (Insecta: Diptera), In: Reemer M, van Helsdingen PJ & Kleukers RMJC. (Eds.). Proceedings of the 13th International Colloquium of the European Invertebrate Survey, Leiden, 2-5 September 2001. EIS-Nederland, Leiden, pp. 61-68.

Polz MF and Cavanaugh CM (1998) Bias in Template-to-Product Ratios in Multitemplate PCR. Applied and Environmental Microbiology, 64(10), 3724–3730.

Porter TM, Gibson JF, Shokralla S, Baird D J, Golding GB and Hajibabaei M (2014) Rapid and accurate taxonomic classification of insect (class Insecta) cytochrome *c* oxidase subunit 1 (COI) DNA barcode sequences using a naïve Bayesian classifier. Mol Ecol Resour, 14: 929–942.

Post RJ, Flook PK, Millest AL (1993) Methods for the preservation of insects for DNA studies. Biochemical Systematics and Ecology. 21:1, 85-92. https://doi.org/10.1016/0305-1978(93)90012-G

Prescher S, Ellstedt BR (1994) Beitrag zur Kenntnis der Buckelfliegen Thüringens (Dipt., Phoridae). Entomologische Nachrichten und Berichte, 38: 45-51.

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular Ecology, 21: 1864-1877. doi:10.1111/j.1365-294X.2011.05239.x

Quan J, Lü X, Zhuang Z, Dai J, Deng J, Zhang Y (2001) Note: Low Genetic Variation of Penaeus chinensis as Revealed by Mitochondrial COI and 16S rRNA Gene Sequences. Biochem Genet. 39: 279. https://doi.org/10.1023/A:1010234816514

Quicke DLJ (2004) The world of DNA barcoding and morphology— Collision or synergism and what of the future? Systematist 23: 8–12.

Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods 6: 639.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics. 28;12:38. doi: 10.1186/1471-2105-12-38.

R Core Development Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rajaei Sh H, Struwe J-F, Raupach MJ, Ahrens D, Wägele JW (2013) Integration of cytochrome c oxidase I barcodes and geometric morphometrics to delimit species in the genus Gnopharmia (Lepidoptera: Geometridae, Ennominae), Zoological Journal of the Linnean Society, Volume 169, Issue 1, Pages 70–83. doi:10.1111/zoj.12053

Ratnasingham S, & Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes, 7(3), 355–364. http://doi.org/10.1111/j.1471-8286.2007.01678.x.

Ratnasingham S, Hebert PDN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8(7): e66213. https://doi.org/10.1371/journal.pone.0066213

Reiskind J (1965) The Taxonomic Problem of Sexual Dimorphism in Spiders and a Synonymy in Myrmecotypus (Araneae, Clubionidae). Psyche 72:279-281.

Renner K (1982) Coleopterenfänge mit Bodenfallen am Sandstrand der Ostseeküste, ein Beitrag zum Problem der Lockwirkung mit Konservierungsmitteln. Faunistisch-ökologische Mitteilungen 5: 137– 146.

Reusch H, Hohmann M (2009). Stelzmücken (Diptera: Limoniidae et Pediciidae) aus Emergenzfallen im "Nationalpark Harz" (Sachsen-Anhalt) . Lauterbornia 68 : 127 -134.

Reusch H and Weber D (2013) Stelzmücken (Diptera: Limoniidae und Pediciidae) aus Höhlen des Großherzogtums Luxemburg. Ferrantia 69: 268 - 275.

Reynolds DR, Riley JR (2002) Remote-sensing, telemetric and computer-based technologies for investigating insect movement: a survey of existing and potential techniques. Comput. Electron. Agric. 35:271–307.

Richly E and Leister D (2004) NUMTs in sequenced eukaryotic genomes. Molecular Biology and Evolution, 21(6):1081–1084.

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

Rondoni G, Ielo F, Ricci C, Conti E (2017) Behavioural and physiological responses to prey-related cues reflect higher competitiveness of invasive vs. native ladybirds. Scientific Reports. 7. 1-9. 10.1038/s41598-017-03471-9.

Ronquist F (2010) 250 Years of Swedish Taxonomy . In: Polaszek A (Ed.) Systema Naturae 250 - The Linnaean Ark . CRC Press , 239-248 pp. https://doi.org/10.1201/ebk1420095012-c18

Rosenberg NA and Tao R (2008) Discordance of species trees with their most likely gene trees: The case of five taxa. Systematic Biology, 57(1):131–140.

Ross HA, Murugan S, Li WL. (2008) Testing the reliability of genetic methods of species identifica tion via simulation. Syst Biol. 57(2):216–30. doi:10.1080/10635150802032990 PMID:18398767.

Rotheray GE, Hancock G, Hewitt S, Horsfield D, MacGowan I, Robertson D, Watt K (2001) The Biodiversity and Conservation of Saproxylic Diptera in Scotland. Journal of Insect Conservation 5: 77. https://doi.org/10.1023/A:1011329722100.

Rozkošný R (1986) A bioindicative significance of central European Anthomyiidae (Diptera). Dipterologica Bohemoslovaca 4, 75–78.

Rozkošný R (2006) Dryomyzidae Schiner, 1862. In: Jedlička L., Stloukalová V. & Kúdela M. (eds) 2006: Checklist of Diptera of the Czech Republic and Slovakia. http://zoology.fns.uniba. sk/Diptera and CD-RoM (ISBN 80-969629-0-6).

Rozkošný R, František G, Pont A (1997) The European Fanniidae (Diptera). Acta scientiarum naturalium Academiae scientiarum Bohemicae-Brno. 31.

Röder J, Bässler C, Brandl R, Dvorak L, Floren A, Gossner MM, Gruppe A, Jarzabek-Müller A, Vojtech O, Wagner C, Müller J (2010) Arthropod species richness in the Norway Spruce (Picea abies (L.) Karst.) canopy along an elevation gradient. Forest Ecology and Management 259: 1513–1521

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. Journal of Heredity, 97(6):581–594.

Salzberg SL, Yorke JA (2005) Beware of mis-assembled genomes. Bioinformatics. 21:4320–4321. doi: 10.1093/bioinformatics/bti769.

Samways M (1993) Insects in biodiversity conservation: some perspectives and directives. Biodiversity and Conservation 2 (3): 258-282. https://doi.org/10.1007/ bf00056672

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors Proceedings of the National Academy of Sciences.74 (12) 5463-5467.

Schaffner F, Kaufmann C, Hegglin D, and Mathis A (2009) The invasive mosquito Aedes japonicus in Central Europe. Med. Vet. Entomol. 23: 448–451.

Scheffer SJ (2000) Molecular Evidence of Cryptic Species within the Liriomyza huidobrensis (Diptera: Agromyzidae), Journal of Economic Entomology, Volume 93, Issue 4, Pages 1146–1151, doi:10.1603/0022-0493-93.4.1146

Scheffers BR, Joppa LN, Pimm SL, Laurance WF (2012) What we know and don't know about Earth's missing biodiversity. Trends in Ecology & Evolution; DOI: 10.1016/j.tree.2012.05.008

Scherber C, Vockenhuber E, Stark A, Meyer H, Tscharntke T (2014) Effects of tree and herb biodiversity on Diptera, a hyperdiverse insect order. Oecologia 174: 1387–1400. doi: 10.1007/s00442-013- 2865-7 PMID: 24394862

Schlichting CD and M Pigliucci (1998) Phenotypic Evolution a Reaction Norm Perspective. Sinauer Associates, Sunderland, MA.

Schlinger EI (1956) A revision of the Acrocerid flies of the genus Pialea Erichson with a discussion of their sexual dimorphism (Diptera). Proc. U. S. Nat. Mus. 16:359-375.

Schloss PD & Handelsman J (2005) Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. Appl. Environ. Microbiol. 71, 1501–1506.

Schloss PD (2008) Evaluating different approaches that test whether microbial communities have the same structure. The ISME journal, Vol. 2, No. 3. pp. 265-275, doi:10.1038/ismej.2008.5

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol, 75(23):7537-41.

Scholl H (1960) Die Oogenese einiger parthenogenetischer Orthocladiinen (Diptera). Chromosoma, 11, 380–401.

Schmidl J, Bußler H (2004) Ökologische Gilden xylobionter Käfer Deutschlands. Naturschutz und Landschaftsplanung 36: 202–218.

Schmidt MH, Clough Y, Schulz W, Westphalen A, Tscharntke T (2006) Capture efficiency and preservation attributes of different fluids in pitfall traps. Journal of Arachnology 34: 159–162.

Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, Escobar J, Flowers D, Fotopulos D, Garcia C, Gomez M, Gonzales E, Haydu L, Lopez F, Ramirez L, Retterer J, Rodriguez A, Rogers S, Salazar A, Tsai M, Myers RM (2004) Quality assessment of the human genome sequence. Nature, 429: 365-368. 10.1038/nature02390.

Schulp CJE, van Teeffelen AJA, Tucker G, Verburga PH (2016) A quantitative assessment of policy options for no net loss of biodiversity and ecosystem services in the European Union. Land Use Policy 57(Supplement C): 151-163.

Schumann H (1992) Systematische Gliederung der Ordnung Diptera mit besonderer Berücksichtigung der in Deutschland vorkommenden Familien. Deutsche Entomologische Zeitschrift 1–3: 103–116.

Schumann H, Bährmann R, Stark A (1999) Checkliste der Dipteren Deutschlands. Studia Dipterologica Suppl. 2, 7-354.

Schumann H (2002) Erster Nachtrag zur „Checkliste der Dipteren Deutschlands". Studia Dipterologica 9, 437-444.

Schuster SC (2007) Next-generation sequencing transforms today's biology. Nature Methods 5, 16–18.

Segata N, Izard J, Waldron L, Izard, J, Waldron L, Gevers D, Miropolsky L, Garrett WS & Huttenhower C (2011) Metagenomic biomarker discovery and explanation. Genome Biology. 2011;12(6):R60. doi:10.1186/gb-2011-12-6-r60. http://huttenhower.sph.harvard.edu/lefse/

Selby RD, Gage SH, Whalon ME (2014) Precise and Low-Cost Monitoring of Plum Curculio (Coleoptera: Curculionidae) Pest Activity in Pyramid Traps With Cameras, Environmental Entomology, Volume 43, Issue 2, 1 April 2014, Pages 421–431, https://doi.org/10.1603/EN13136

Selivon D, Perondini ALP, Morgante JS (2005) A Genetic–Morphological Characterization of Two Cryptic Species of the Anastrepha fraterculus Complex (Diptera: Tephritidae), Annals of the Entomological Society of America, Volume 98, Issue 3, 1 May 2005, Pages 367–381. doi:10.1603/0013-8746(2005)098[0367:AGCOTC]2.0.CO;2

Sereda E, Blick T, Dorow WHO, Wolters V, Birkhofer K (2014) Assessing spider diversity on the forest floor: expert knowledge beats systematic design. Journal of Arachnology 42: 44–51.

Service M (1973) Spatial and Temporal Distributions of Aerial Populations of Woodland Tipulids (Diptera). Journal of Animal Ecology, 42(2), 295-303. doi:10.2307/3286.

Shen T-J, Chao A, Lin J-F (2002) Predicting the number of new species in further taxonomic sampling. Ecology 84, 798-804.

Siitonen J (2001) Forest Management, Coarse Woody Debris and Saproxylic Organisms: Fennoscandian Boreal Forests as an Example. Ecological Bulletins, (49), 11-41. http://www.jstor.org/stable/20113262

Simon C, Frati F, Beckenbach A, Crespi B, Liu H and Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87: 651Ð701.

Simpson JT and Pop M (2015) The Theory and Practice of Genome Sequence Assembly. Annual Review of Genomics and Human Genetics 16(1): 153-172.

Simon U, Pfütze J, Thömen D (2001) A time-sorting stem-eclector. Ecological Entomology 26(3): 325-329.

Singh JS (2002) The biodiversity crisis: A multifaceted review. Current Science, 82(6), 638-647. Retrieved from http://www.jstor.org/stable/24106689

Skidmore P (1985) The Biology of the Muscidae of the World. Junk, Dordrecht. Series Entomologica 29. xiv + 550 pp.

Small J, Cotton S, Fowler K, Pomiankowski A (2009) Male eyespan and resource ownership affect contest outcome in the stalk-eyed fly, Teleopsis dalmanni. Animal Behaviour, 78, 1213–1220.

Smith KGV (1989) An Introduction to the Immature Stages of British Flies: Diptera Larvae, with Notes on Eggs, Puparia and Pupae, Royal Entomological Society of London.

Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, Gibbs J, Hajibabaei M, Hallwachs W, Hind K, Hrcek J, Huang DW, Janda M, Janzen DH, Li Y, Miller SE, Packer L, Quicke D, Ratnasingham S, Rodriguez J, Rougerie R, Shaw MR, Sheffield C, Stahlhut JK, Steinke D, Whitfield J, Wood M, Zhou X (2012) Wolbachia and DNA Barcoding Insects: Patterns, Potential, and Problems. Plos One, 7(5).

Smith MA, Woodley NE, Janzen D, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies

(Diptera: Tachinidae). Proceedings of the National Academy of Sciences of the United States of America. 103. 3657-62. 10.1073/pnas.0511318103.

Smith MA, Wood DM, Janzen DH Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proceedings of the National Academy of Sciences of the United States of America, 104(12), 4967-4972. doi: 10.1073/pnas.0700050104

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences. 103 (32) 12115-12120. DOI: 10.1073/pnas.0605127103

Song H, Buhay J, Whiting M, Crandall K (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences of the United States of America, 105(36):13486–13491.

Spencer KA (1973) Agromyzidae (Diptera) of economic importance. Series Entomologica 9. xi+418 pp.

Spencer KA (1990) Host Specialization in the World Agromyzidae (Diptera). Kluwer Academic Publishers, Dordrecht. xii+444 pp.

Spencer KA, Steyskal GC (1986) Manual of the Agromyzidae (Diptera) of the United States. U.S. Dept. of Agriculture, Agricultural Research Service.

Stammer HJ (1933) Die Metamorphose der Syrphide Temnostoma vespiforme (L.) und die eigenartigen Anpassungen der Larve dieses Tieres an das Bohren im Holz (Dipt.). Z. Morph. u. Ökol. Tiere.26: 437. https://doi.org/10.1007/BF00446406.

Stein ED, White BP, Mazor RD, Miller PE, Pilgrim EM (2013) Evaluating Ethanol-based Sample Preservation to Facilitate Use of DNA Barcoding in Routine Freshwater Biomonitoring Programs Using Benthic Macroinvertebrates. Steinke D, ed. PLoS ONE;8(1):e51273. doi:10.1371/journal.pone.0051273.

Stein ED, Martinez MC, Stiles S, Miller PE, Zakharov EV (2014) Is DNA Barcoding Actually Cheaper and Faster than Traditional Morphological Methods: Results from a Survey of Freshwater Bioassessment Efforts in the United States? Casiraghi M, editor. PLoS ONE.9: e95525. pmid:24755838

Stewart CN and Excoffier L (1996) Assessing population genetic structure and variability with RAPD data: Application to *Vaccinium macrocarpon* (American Cranberry). Journal of Evolutionary Biology, 9: 153–171. doi:10.1046/j.1420-9101.1996.9020153.x

Stoeckle BC, Dworschak K, Gossner MM, Kuehn R (2010) Influence of arthropod sampling solutions on insect genotyping reliability. Entomologia Experimentalis et Applicata 135: 217–223.

Strenzke K (1950) Morphologie und Ökologie der terrestrischen Chironomiden. Archiv für Hydrobiologie. Suppl. 18, 209–412.

Stuke J-H (2003) Die Blasenkopffliegen (Diptera: Conopidae) Niedersachsens und Bremens. Drosera, 81–94.

Stuke J-H (2008) Die Ibisfliegen, Kugelfliegen, Hummelschweber, Schnepfenfliegen und Stilettfliegen Niedersachsens und Bremens (Diptera: Acroceridae, Athericidae, Bombyliidae, Rhagionidae, Therevidae). Braunschweiger naturkundliche Schriften 8: 235-259, Braunschweig.

Stuke J-H (2010) Bemerkenswerte Zweiflügler aus Niedersachsen und Bremen (Insecta: Diptera)- 3. Teil. – Drosera 2009: 143–150.

Stuke J-H and Schacht W (2009) Die Dungfliegen Niedersachsens und Bremens. Entomofauna, Bd. 30 (14): 221-256.

Stur E, Martin P, Ekrem T (2005) Non-biting midges as hosts for water mite larvae in spring habitats in Luxembourg. Annales de Limnologie, 41, 225–236.

Succow M, Jeschke L, Knapp HD (2012) Naturschutz in Deutschland. Ch. Links Verlag, Berlin.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). Molecular Biology and Evolution 13: 510– 524. PMID: 8742640

Suzuki MT and Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. February 1996 vol. 62 no. 2 625-630.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21: 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. TREE 18:70–74.

Teixeira RFM, de Souza DM, Curran PM, Assumpció A, Ottar M, Milà i Canals L (2016) Towards consensus on land use impacts on biodiversity in LCA: UNEP/SETAC Life Cycle Initiative preliminary recommendations based on expert contributions. Journal of Cleaner Production. 112. 4283-4287. 10.1016/j.jclepro.2015.07.118.

Teskey HJ (1976) Diptera larvae associated with trees in North America. Memoirs of the Entomological Society of Canada 100: 1–53.

Teviotdaleo BL, Sibbetto GS, Harper DH (1989) Several copper fungicides control olive leaf spot California Agriculture 43: 30–31.

Thapa S, Adler PH, Chhetri S, Varma R, Henry W (2017) Chromosomal evidence for a new cryptic species of black fly in the Simulium praelargum complex (Diptera: Simuliidae) from West Bengal, India. Zootaxa 4244(1). doi: http://dx.doi.org/10.11646/zootaxa.4244.1.8

Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR', Nucleic Acids Research, Volume 30, Issue 9, Pages 2083–2088. https://doi.org/10.1093/nar/30.9.2083

Thurber D, McClain W & Whitmore R (1994) Indirect Effects of Gypsy Moth Defoliation on Nest Predation. *The Journal of Wildlife Management, 58*(3), 493-500. doi:10.2307/3809321

Thien LB (1969) Mosquito pollination of Habenaria obtusata (Orchidaceae). Am. J. Bot. 56:232 - 37.

Thompson FC (1969) First record of Rhagio scolopaceus (Linné) in North America (Diptera: Rhagionidae). Proc. ent. Soc. Wash. 71: 141–143.

Thompson PA (1980) Germination Strategy of a Woodland Grass: Milium effusum L., Annals of Botany, Volume 46, Issue 5, 1 November 1980, Pages 593–602. doi.org/10.1093/oxfordjournals.aob.a085957.

Timms LL, Bowden JJ, Summerville KS and Buddle CM (2013) Does species-level resolution matter? Taxonomic sufficiency in terrestrial arthropod biodiversity studies. Insect Conserv Divers, 6: 453–462. doi:10.1111/icad.12004

Tkeshelashvili LK, McBride T, Spence K, Loeb LA (1991) Mutation spectrum of copper-induced DNA damage. Journal of Biological Chemistry 266: 6401–6406. PMID: 1826106

Trontelj P & Fišer C (2009) Perspectives: Cryptic species diversity should not be trivialised. Systematics and Biodiversity. 7. 1 - 3. doi:10.1017/S1477200008002909.

Turell MJ, O'Guinn ML, Dohm DJ, & Jones JW (2001) Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. J. Med. Entomol. 38: 130–134.

Vinje H, Liland KH, Almøy T, Snipen L (2015) Comparing K-mer based methods for improved classification of 16S sequences. BMC Bioinformatics. 16:205. doi:10.1186/s12859-015-0647-4.

United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, World Bank (2003) Handbook of National Accounting: Integrated Environmental and Economic Accounting. Studies in Methods, Series F, No.61, Rev.1, Glossary, United Nations, New York, para. 8.227

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. Trends in Ecology & Evolution, Volume 24, Issue 2, Pages 110-117. https://doi.org/10.1016/j.tree.2008.09.011.

Villa C, Costa J, Meira L, Oliveira M, Beatriz PP, Mafra I (2016) Exploiting DNA mini-barcodes as molecular markers to authenticate saffron (Crocus sativus L.). Food control, 65, 21-31. doi: 10.1016/j.foodcont.2016.01.008

Vink CJ, Thomas SM, Paquin P, Hayashi CY, Hedin M (2005) The effects of preservatives and temperatures on arachnid DNA. Invertebrate Systematics 19: 99–104.

Vitousek PM (1990). Biological Invasions and Ecosystem Processes: Towards an Integration of Population Biology and Ecosystem Studies. Oikos, *57*(1), 7-13. doi:10.2307/3565731

van der Weele R, Hrivniak L, Kappert J, Manko P, Shamshev I,Oboňa J (2017) Brachystomatidae, Empididae and Hybotidae (Diptera) of Uzh River Basin, with additions to checklists of Ukraine. Annales de la Société entomologique de France (N.S.). Vol. 53, Iss. 2.

Väisänen R, Biström O, Heliövaara K (1993) Sub-cortical Coleoptera in dead pines and spruces: is primeval species composition maintained in managed forests? Biodiversity and Conservation. 2. 95-113.

von der Dunk K (2005) Syrphidae Bayerns - eine kommentierte Checkliste (Insecta: Diptera). - Beiträge bayerischen Entomofaunistik 7: 97-114, Bamberg.

von der Dunk K, Hable J (2006) Verhalten und Nachweise xylobionter Schwebfliegen (Dipt. Syrphidae) - Galathea, Berichte des Kreises Nürnberger Entomologen e.V. 22: 69-78.

Walker ED, Olds EJ, Merritt RW (1988) Gut Content Analysis of Mosquito Larvae (Diptera: Culicidae) Using Dapi Stain and Epifluorescence Microscopy, Journal of Medical Entomology, Volume 25, Issue 6, 1, Pages 551–554, https://doi.org/10.1093/jmedent/25.6.551

Walker MJC, Bohncke SJP, Coope, O'Connell M, Usinger H, Verbruggen C (1994) The Devensian/Weichselian Late-glacial in northwest Europe (Ireland, Britain, North Belgium, The Netherlands, northwest Germany). Journal of Quaternary Science. 9. 109 - 118. 10.1002/jqs.3390090204.

Wang GC-Y, Wang Y (1996) The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology 142:1107-1114.

Wang GC-Y, Wang Y (1997) Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl. Environ. Microbiol. 63:4645-4650.

Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy . Applied and Environmental Microbiology, 73(16), 5261–5267. http://doi.org/10.1128/AEM.00062-07

Ward RD and Holmes BH (2007) An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. Molecular Ecology Notes, 7: 899–907. doi: 10.1111/j.1471-8286.2007.01886.x

Warren MS, Key RS (1991) Woodlands: past, present and potential for insects. In The Conservation of Insects and their Habitats, ed. Collins NM, Thomas JA, pp. 155–203. London: Academic.

Weeks RD, McIntyre NE (1997) A comparison of live versus kill pitfall trapping techniques using various killing agents. Entomologia Experimentalis et Applicata 82: 267–273.

Weiße R (1995) Die Potsdamer Glaziallandschaft - glazigene Sedimente und glaziäre Baustile. Brandenburgische Geowissenschaftliche Beiträge (1/1995), 13–26.

Wende B (2014). Diversity of saproxylic beetles and host tree specialisation in differently managed forests across Germany. URN:nbn:de:bvb:20-opus-107049

Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R, Ley RE (2012) Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. The ISME Journal, 6(1), 94–103. http://doi.org/10.1038/ismej.2011.82

Werren J and Windsor D (2000) Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proceedings of the Royal Society B-Biological Sciences, 267(1450):1277–1285.

Westcott SL & Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3:e1487. pmid:26664811. https://peerj.com/articles/1487

Westcott SL & Schloss PD (2017) OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*, *2*(2), e00073–17. http://doi.org/10.1128/mSphereDirect.00073-17

West-Eberhard, M. (1989). Phenotypic plasticity and the origins of diversity. Annual Review of Ecology, Evolution and Systematics, 20:249–278.

Wright ES and Vetsigian KH (2016) Quality filtering of Illumina index reads mitigates sample cross-talk. BMC Genomics, 2016, Volume 17, Number 1, Page 1.

Whitman DW & Agrawal A (2009) What is Phenotypic Plasticity and Why is it Important? Phenotypic plasticity of insects. doi:10.1201/b10201.

Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics.1:80–83. doi: 10.2307/3001968.

Will KW and Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55.

Wilson EO (1955) A monographic revision of the ant genus Lasius. Bull. Mus. Comp. Zool. 113: 1-201.

Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a Search-Based Approach to Chimera Identification for 16S rRNA Sequences. Applied and environmental microbiology, 78(3), 717-725. doi: 10.1128/AEM.06516-11

Xu J (2014) Next Generation Sequencing: Current Technologies and Applications. Norfolk, England: Caister Academic Press.

Yang ZH, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. Proc. Natl Acad. Sci. USA 107, 9264–9269. doi:10.1073/pnas.0913022107

Yao Y-G, Salas A, Logan I, & Bandelt H-J (2009) mtDNA Data Mining in GenBank Needs Surveying. American Journal of Human Genetics, 85(6), 929–933. http://doi.org/10.1016/j.ajhg.2009.10.023

Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J (2016) Introducing EzBioCloud: a taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst Evol Microbiol.67(5): 1613-1617. doi: 10.1099/ijsem.0.001755

Young AM (1986). Cocoa pollination. Cocoa Growers Bulletin 37: 5.

Young AM (1994) The Chocolate Tree, a Natural History of Cacao. Smithsonian Institution Press, Washington, London. 200 pp.

Yue JC, Clayton MK (2005) A similarity measure based on species proportions. Communications in Statistics-Theory and Methods 34: 2123–2131.

Zhang Q and Backström N (2014) Assembly errors cause false tandem duplicate regions in the chicken (Gallus gallus) genome sequence. Chromosoma,123: 165. https://doi.org/10.1007/s00412-013-0443-8

Zhang Z, Schwartz S, Wagner L, MillerW (2000) A greedy algorithm for aligning DNA sequences. Journal of Computational Biology 7: 203–214. PMID: 10890397

Ziegler, J. (2003). Diptera. In: Dathe H (ed) Lehrbuch der Speziellen Zoologie Band I: Wirbellose Tiere, 5. Teil: Insecta. Spektrum Akademischer Verlag Heidelberg – Berlin, pp. 756 – 860.

Zielke E and Banarp P (2017) First records of the genus Mydaea and of some other Muscidae species from Croatia (Diptera). Acta Musei Moraviae, Scientiae biologicae (Brno) 101(1): 43ñ48.

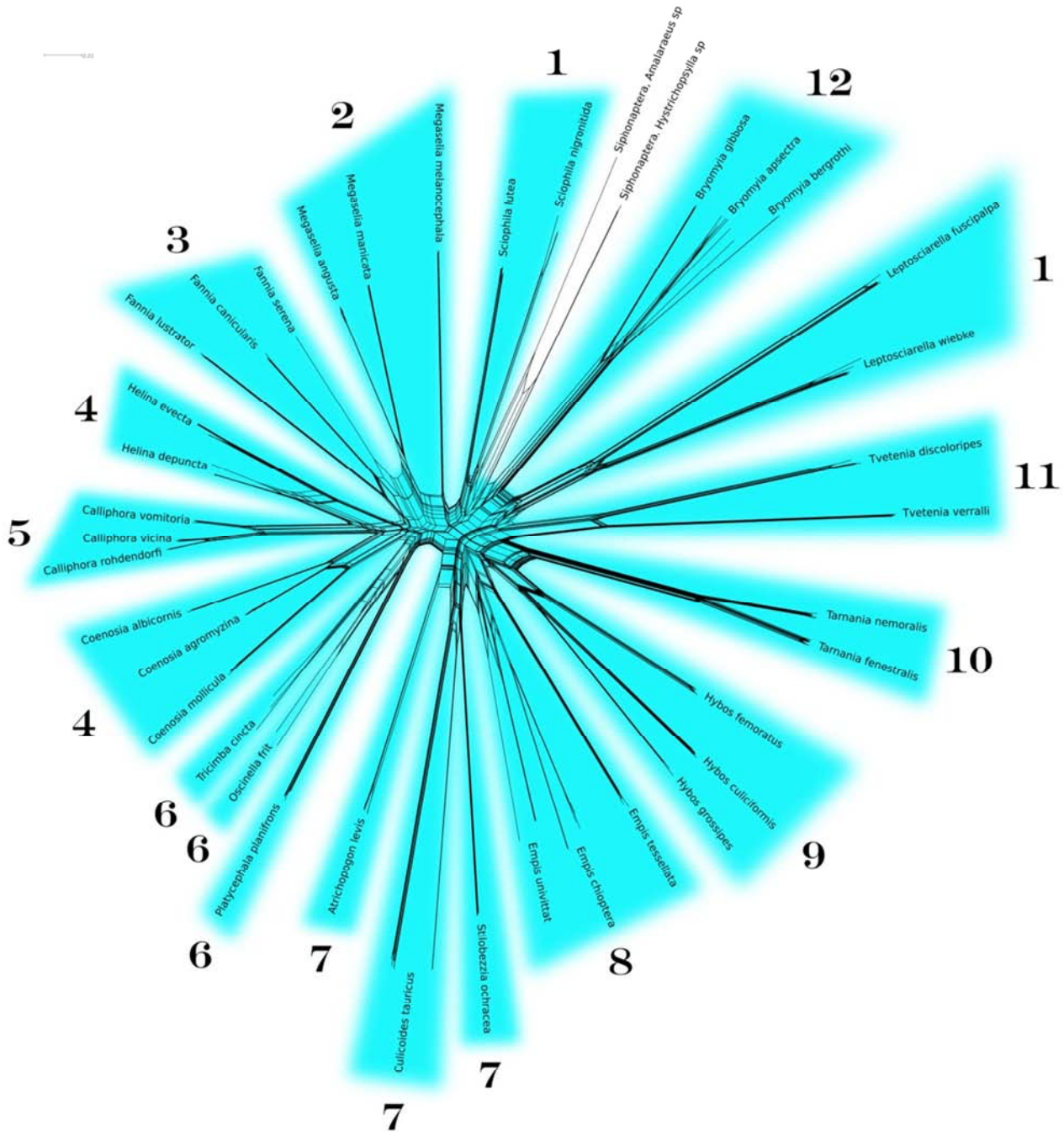
8 Supplement

S1

S1.1 Table: Overview of the number of individuals from different Diptera families. Sampled in 2008, in 30 forest "Very Intensive Plots" (VIPs) of the Exploratories. Data are from two flight-interception traps in the canopy and two in the understory in each of the plots that were repeatedly emptied between May and October.

Family	Number	Family	Number	Family	Number
Acartophthalmidae	2	Dixidae	2	Pseudopomyzidae	0
Agromyzidae	17	Dolichopodidae	602	Psilidae	5
Anisopodidae	880	Drosophilidae	540	Psychodidae	389
Anthomyiidae	570	Dryomyzidae	292	Ptychopteridae	2
Anthomyzidae	0	Empididae + Hybotidae	1639	Rhagionidae + Athericidae	81
Asilidae	31	Ephydridae	74	Sarcophagidae	776
Asteiidae	147	Fanniidae	1234	Scathophagidae	19
Aulacigastridae	13	Heleomyzidae	849	Scatopsidae	1097
Bibiondae	387	Hippoboscidae	4	Scenopinidae	1
Bolitophilidae	0	Lauxaniidae	893	Sciaridae	7286
Bombyliidae	0	Lonchaeidae	189	Sciomyzidae	60
Brachycera	158	Lonchopteridae	2	Sepsidae	22
Calliphoridae	1212	Megamerinidae	0	Simuliidae	4
Calyptratae	296	Micropezidae	1	Sphaeroceridae	31
Camillidae	0	Milichiidae	547	Stratiomyidae	4
Campiochoetidae	0	Muscidae	8011	Syrphidae	82
Carnidae	145	Mycetophilidae s.l.	15928	Tabanidae	13
Cecidomyiidae	4277	Nematocera	4	Tachinidae + Rhinophoridae	169
Ceratopogonidae	1924	Odiniidae	15	Tephritidae	1
Chamaemyiidae	6	Opetiidae	0	Thaumaleidae	0
Chaoboridae	2	Opomyzidae	2	Therevidae	2
Chironomidae	2771	Pallopteridae	12	Tipulidae + Limoniidae	1029
Chloropidae	2838	Periscelididae + Stenomicridae	45	Trichoceridae	17
Chyromyidae	1	Phaeomyiidae	0	Trixocelidae	0
Clusiidae	9	Phoridae	8612	Ulidiidae s.l.	24
Conopidae	2	Piophilidae	105	*unclear*	25
Culicidae	26	Pipunculidae	22	Xylomyidae	0
Cylindrotomidae	4	Platypezidae	1	Xylophagidae	3
Diastatidae	0	Platystomatidae	1	*Total*	66486

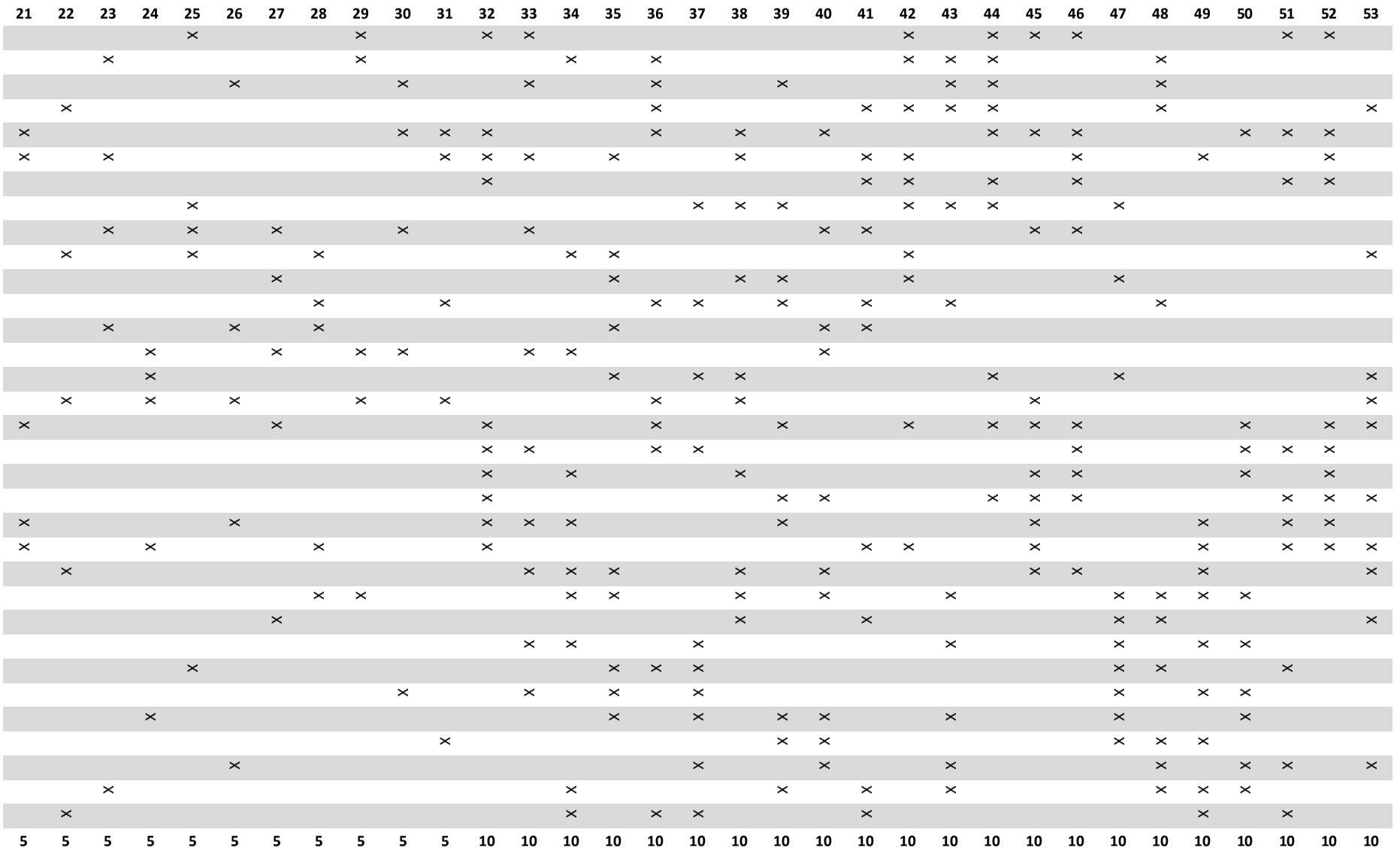
S1.2 Figure: Neighbor joining network shows the discriminative properties of COI for Diptera species. 166 specimen sequences where obtained from 12 Diptera families. Species from the order Siphonaptera were chosen as an outgroup. Each of the species names resembles up to five individuals. Matching genera are marked blue whereas families are marked by numbers. Diptera families: 1-Sciaridae, 2-Phoridae, 3-Fanniidae, 4-Muscidae, 5-Calliphoridae, 6-Chloropidae, 7-Ceratopogonidae, 8-Empididae, 9-Hybotidae, 10-Mycetophilidae, 11-Chironomidae, 12-Cecidomyiidae; Outgroup: Siphonaptera



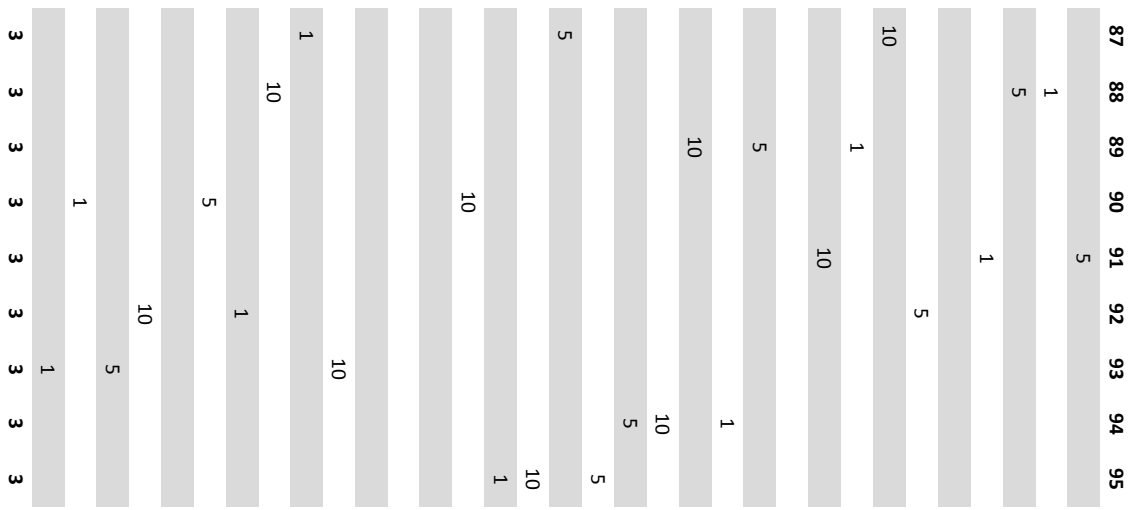
S2

S2.1 Table: Detailed list of species used for the sample mixtures in both treatments T1 and T2. All DNA quantities were leveled to guarantee the same amount of DNA for each species. X: leveled DNA of the species was used in the sample. 1: One unit of leveled DNA of the species was used in the sample. 5: Five units of leveled DNA of the species were used in the sample. 10: Ten units of leveled DNA of the species were used in the sample.

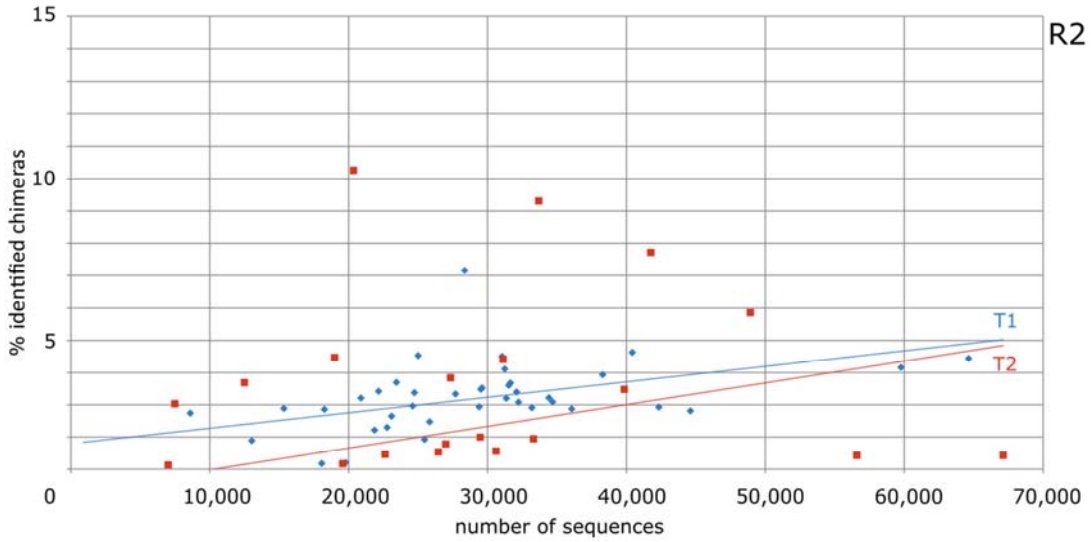
Genus	Species	Sample No.	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	Species N
Elachiptera	*cornuta*							X							X						X	X	5
Hydrotaea	*pilipes*										X				X		X			X			5
Eudasyphora	*zimini*					X			X														5
Muscina	*prolapsa*				X	X											X			X			5
Oscinella	*frit*			X																	X	X	5
Liriomyza	*intonsa*		X	X														X			X		5
Dicraeus	*ingratus*								X												X		5
Muscina	*levida*									X		X											5
Coprooica	*ferruginata*						X															X	5
Tachypeza	*nubila*												X										5
Helina	*impuncta*						X	X									X			X			5
Phaonia	*pallida*								X			X					X		X				5
Platypalpus	*exilis*											X											5
Dryodromia	*testacea*						X		X				X										5
Helina	*depuncta*			X						X	X						X		X				5
Suillia	*fuscicornis*						X	X														X	5
Oscinella	*vastator*				X																		5
Dicraeus	*styriacus*		X									X	X										5
Elachiptera	*tuberculifera*		X							X			X										5
Eutropha	*fulvifrons*		X				X											X				X	5
Phytomyza	*flavicornis*									X								X				X	5
Chromatomyia	*milii*							X										X			X		5
Agromyza	*cinerascens*										X							X				X	5
Episyrphus	*balteatus*												X										5
Eupeodes	*corollae*			X										X					X				5
Eupeodes	*latifasciatus*					X		X		X					X					X			5
Platycheirus	*clypeatus*					X								X									5
Syrphus	*torvus*										X			X									5
Melanostoma	*scalare*									X	X			X					X				5
Platycheirus	*peltatus*							X						X				X	X				5
Melanostoma	*mellinum*				X	X						X			X								5
Sphaerophoria	*scripta*							X							X					X			5
Pipizella	*viduata*										X				X								5


21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53
5 5 5 5 5 5 5 5 5 5 5 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10

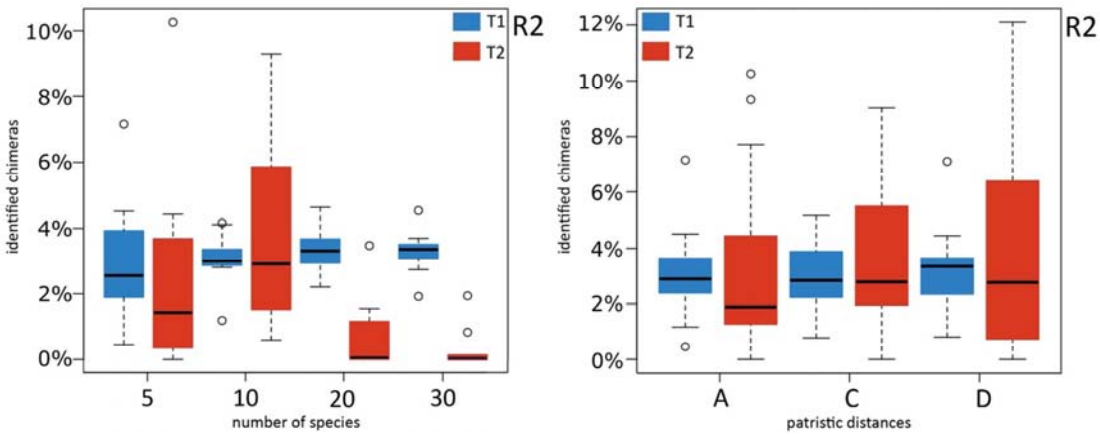
54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86
	×	×			×	×			×		×		×	×		×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	
			×				×		×				×	×	×	×	×	×	×	×	×	×	×	×		×		×	×	×		
	×			×	×			×	×	×		×			×	×	×			×	×	×	×		×		×	×	×	×	×	
		×							×		×	×		×	×		×	×		×	×	×	×	×		×		×	×	×	×	
							×		×	×	×	×	×			×	×		×	×	×	×	×		×	×	×	×	×	×		×	1
		×				×	×		×	×	×		×	×					×	×	×	×	×	×	×	×	×	×	×	×	×	
×		×							×		×		×	×		×	×	×		×		×	×	×	×	×	×	×	×	×	×	
×									×	×	×			×		×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	
×		×	×						×				×			×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	
		×	×		×		×			×	×	×	×			×	×	×				×		×	×	×	×		×	×		5
			×	×		×		×	×				×			×		×		×	×	×	×	×	×	×	×	×		×	×	
×		×							×	×		×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	
×	×			×		×		×			×	×		×	×						×	×	×	×	×	×	×	×	×	×	×	
	×	×		×				×			×	×		×	×	×		×	×			×	×	×	×	×	×	×	×	×	×	
	×			×	×				×	×				×		×	×		×	×	×	×	×	×		×	×	×	×	×	×	
	×				×				×	×	×			×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	
			×						×	×			×	×	×	×	×	×		×	×	×	×	×	×	×	×	×		×	×	
×			×		×		×		×	×		×			×			×	×		×	×	×	×	×	×	×	×	×	×	×	
							×	×	×	×	×	×	×	×						×	×	×	×	×	×	×	×	×	×		×	
									×	×		×	×	×			×	×	×	×								×	×	×	×	×	
									×	×		×	×	×		×			×	×		×	×	×		×	×	×	×	×	×	
										×	×		×		×		×		×	×	×	×	×	×	×	×	×	×	×	×	×	
×	×					×		×		×	×	×	×	×	×						×	×	×	×	×	×		×	×	×	×	
×	×		×	×			×			×	×		×	×	×			×	×		×	×	×		×	×	×	×	×	×	×	
			×		×	×	×	×			×	×	×	×	×	×		×			×	×	×	×	×	×		×	×	×	×	
				×		×	×	×				×		×			×		×		×	×	×	×	×	×	×	×	×	×	×	10
							×						×		×	×		×	×		×	×	×	×	×	×	×		×	×	×	
		×				×						×	×	×	×		×	×		×	×	×	×	×	×	×		×	×	×	×	
		×		×		×		×		×	×	×	×			×	×	×	×	×	×	×	×	×	×	×		×	×	×		
×	×		×	×	×			×					×			×	×			×	×	×	×	×	×	×	×	×	×	×	×	
10	10	10	10	10	10	10	10	10	20	20	20	20	20	20	20	20	20	20	20	30	30	30	30	30	30	30	30	30	30	30	33	3



S2.2 Figure: Trendline over the randomly chosen species samples from R2.



S2.3 Figure: Boxplot diagram for R2 sequencing data. Left: Identified chimeric sequences in relation to the number of different species (5, 10, 20, and 30) within a sample of treatment T1 (mixed amplification) and T2 (single amplification). Right: Identified chimeric sequences in relation to the averaged patristic distances of, (A) arbitrary chosen species, (C) species that are taxonomically close, (D) and species that are taxonomically distant to each other.



S2.4 Table: Kruskal-Wallis rank sum test to test the influence of species numbers and patristic distances on the amount of chimeric sequences in the first sequencing run R1. Significant value is highlited.

	test of	chi-squared	df	p-value
T1	species number	3.6059	3	0.3073
T2	species number	16.446	3	**0.0009186***
T1	patristic distances	0.2875	2	0.5661
T2	patristic distances	4.3455	2	0.1139

S3

S3.1 Table: Measurements of the remaining material after evaporation.

set	subsample	SEW2	SEW3	SEW7	SEW9
I	1	25 mg	31 mg	30 mg	30 mg
I	2	22 mg	21 mg	17 mg	26 mg
I	3	21 mg	12 mg	08 mg	15 mg
I	4	12 mg	09 mg	08 mg	08 mg
II	1	26 mg	25 mg	22 mg	40 mg
II	2	18 mg	26 mg	16 mg	42 mg
II	3	12 mg	11 mg	13 mg	12 mg
II	4	10 mg	06 mg	04 mg	06 mg
	Total	146 mg	141 mg	118 mg	179 mg

S3.2 Protocol: Extraction protocol for the Qiagen DNeasy Blood & Tissue Kit (Qiagen N.V., Venlo, www.qiagen.com).

1. 180 µl ATL buffer in 1,7 ml tube
2. add material
3. 20 µl Proteinase K, vortex
4. Lyse overnight at 56°C.
5. 4 µl RNase (100 mg/mL), vortex, incubate 2 min at room temperature (rt)
6. Vortex 15 s
7. 200 µl AL buffer, incubate 10 min at 70°C, vortex
8. 200 µl ethanol (96 % - 100 %), vortex
9. load onto the DNeasy Mini spin column, centrifugate 1 min at 8.000 rpm
10. Put DNeasy Mini spin column onto new cap
11. 500 µl AW1 buffer, centrifugate 1 min at 8.000 rpm
12. Put DNeasy Mini spin column onto new cap
13. 500 µl AW2 buffer, centrifugate 3 min at 13.000 rpm
14. Put DNeasy Mini spin column onto 1,7 ml tube **- 2 x -**
15. 50 µl H20, incubate 1 min at rt, centrifugate 1 min at 8.000 rpm

S3.3 Protocol: Measuring DNA concentration using the Promega Quantus Fluorometer (Promega Corporation, Fitchburg, www.promega.com).

1. 98 µl TE (1:20 dilution of original 20xTE) in 0,5 ml tube
2. 2 µl of sample extract
3. 100 µl dsDNA Dye (1:200 dilution of original 200xDye)
4. Incubate 5 min in the dark

S3.4 Table: Measured DNA concentration of the combined subsamples.

subsample	SEW2	SEW3	SEW7	SEW9
1 (I+II)	62 ng/uL	56 ng/uL	71 ng/uL	62 ng/uL
2 (I+II)	33 ng/uL	49 ng/uL	29 ng/uL	47 ng/uL
3 (I+II)	34 ng/uL	22 ng/uL	25 ng/uL	23 ng/uL
4 (I+II)	28 ng/uL	19 ng/uL	21 ng/uL	20 ng/uL
Total	157 ng/uL	146 ng/uL	146 ng/uL	105 ng/uL

S3.5 Protocol: Fragmentation with Fragmentase for dsDNA (NEB, Ipswich, www.neb.com).

1. Samples were brought to a concentration of 1.000 ng in 16,0 µl and put on ice
2. Add 2 µl of 10 x Fragmentase buffer and 2 µl Fragmentase (NEB) per sample
3. Mix thoroughly by pipetting using a 20 µl pipette

4. Incubate at 37°C for as long as necessary (in our case 20min – 25 min) in PCR Thermocycler
5. Stop reaction with 5,0 µl of 0,5M EDTA

S3.6 Protocol: Purification with Ampure XP Beads(Agilent Technologies, Santa Clara, www.agilent.com).

1. Let Ampure Beads become rt (30 – 60 min)
2. Prepare fresh 80 % ethanol (2 x 200 µl) needed per sample
3. Mix bead suspension well and add Ampure Beads to the samples in a 1:1 proportion
4. Mix until homogenous
5. Incubate for 15 min at rt
6. Place samples into magnetic plate and wait 6 min
7. Remove supernatant
8. 2x -> Wash beads with 180 µl 80 % ethanol, wait 1 min and remove ethanol again
9. Remove all remaining ethanol and dry the beads for 2 - 4 min (not to long)
10. Add 40 µl dH2O outside the magnetic plate and mix well
11. Incubate for 4 min
12. Place again in magnetic plate and wait 6 min
13. Take 35 µl eluate and place into prepared tubes without any beads
14. Check quality of eluate with Fragment Analyzer

S3.7 Protocol: End repair.

1. Bring Ampure Beads to rt
2. Vaporize samples to 25µl at ~35°C using vaporizer
3. Hold sheared gDNA on ice
4. Prepare reaction mix on ice, containing 20 µl SureSelect End Repair Enzyme Mix and 5 µl SureSelect End Repair Oligo Mix per reaction.
5. Add 25 µl sheared gDNA and 25 µl reaction mix into prepared tube on ice and mix thoroughly
6. Purify samples with Ampure Beads in 1:1 concentration
7. Measure DNA concentration with Quantus Fluorometer

S3.8 Table: Measured DNA concentration of the single indexing samples after end repair. SI: samples for single-indexing; DI: samples for double-indexing.

subsample	SI SEW2	SI SEW3	SI SEW7	SI SEW9
1	2.49 ng/uL	1.83 ng/uL	3.56 ng/uL	3.23 ng/uL
2	2.19 ng/uL	2.10 ng/uL	1.27 ng/uL	2.84 ng/uL
3	1.78 ng/uL	1.30 ng/uL	3.40 ng/uL	1.74 ng/uL
4	1.11 ng/uL	1.21 ng/uL	2.37 ng/uL	1.17 ng/uL
Total	7.57 ng/uL	6.44 ng/uL	10.6 ng/uL	8.98 ng/uL

S3.9 Protocol: A-tailing

1. Centrifuge samples at rt
2. Vaporize samples to 10 µl – 15 µl (~24 min / 35° C) in vaporizer
3. Take 10 µl into prepared tube and put it on ice
4. Add 10 µl SureSelect dA-Tailing Master and mix thoroughly
5. Incubate samples at 37°C/30 min – 60°C/10 min – 4°C/hold, in a thermal cycler

S3.10 Protocol: Adapter ligation - Single indexing

1. Bring Ampure Beads to rt and hold samples and ligation reagents on ice
2. Prepare indexing adaptor: 1:3 dilution in nuclease-free water
3. Prepare reaction mix: Volume for 1 reaction

Reagent	Volume for 1 reaction (µl)

dA – tailed DNA samples	20
Indexed Adaptor	2.5
SureSelect Ligation Master Mix	2.5
Total	25.0

4. Mix well and incubate samples at 20°C for 15 min in thermal cycler
5.. Purify samples with Ampure Beads in 1:1 concentration
6. Store samples at 4°C over night or continue

S3.11 Protocol: Adapter ligation - Double indexing

1. Bring Ampure Beads to rt and hold samples and ligation reagents on ice
2. Prepare indexing adaptor: 1:10 dilution in nuclease-free water
3. Prepare Quick Ligation Mix

Reagent	Volume for 1 reaction (µl)
NEBNext Quick Ligation Reaction Buffer	10.0
Quick T4 DNA Ligase	5.0
Total	15.0

4. Prepare reaction mix: Volume for 1 reaction

Reagent	Volume for 1 reaction (µl)
dA – tailed DNA sample	25.0
NEBNext Adaptor	10.0
NEBNext Quick Ligation Mix	15.0
Total	50.0

5. Mix well and incubate samples at 20°C for 15 min in thermal cycler
6. Add 3 µl of USER Enzyme Mix by pipetting to each sample
7. Incubate samples at 37°C for 15 min in thermal cycler
8. Purify samples with Ampure Beads in 1:1 concentration
9. Store samples at 4°C over night or continue

S3.12 Protocol: Library amplification - Single indexing

1. Bring Ampure Beads to rt
2. Vaporize samples to 12 µl -15 µl (20-27 min/35°C) and hold on ice
3. Prepare pre-capture reaction mix

Reagent	Volume for 1 reaction (µl)
XT2 Primer Mix	0.5
Herculase II PCR Master Mix	12.5
Indexed DNA sample	12.0
Total	20.0

4. mix thoroughly by pipetting and gentle vortexing + short centrifugation
5. run following program in a thermal cycler
   98°C/2min – 98°C/30sec – 60°C/30sec – 72°C/1min (9 cycles) – 72°C/10min – 4°C (HOLD)
6. Purify samples with Ampure Beads in 1:1 concentration
7. If you do not continue, store samples at -20°C

S3.13 Protocol: Library amplification - Double indexing

1. Bring Ampure Beads to rt
2. Vaporize samples to 12 µl -15 µl (20-27 min/35°C) and hold on ice

3. Prepare pre-capture reaction mix

Reagent	Volume for 1 reaction (µl)
NEBNext i7XX Primer	5.0
NEBNext i5XX Primer	5.0
NEBNext Q5 Hot Start HiFi 5 Mix	25.0
Adaptor tagged DNA sample	15.0
Total	50.0

4. mix thoroughly by pipetting and gentle vortexing + short centrifugation
5. run following program in a thermal cycler
   98°C/30sec – 98°C/10sec – 65°C/75sec (9 cycles) – 65°C/5min – 4°C (HOLD)
6. Purify samples with Ampure Beads in 1:1 concentration
7. If you do not continue, store samples at -20°C

S3.14 Protocol: Baits hybridization

1. Prepare 0.2 mL tubes
2. Add 4.5 µl SureSelect XT2 Blocking Mix
3. Add 3.5 µl of each library sample
4. mix thoroughly + short centrifugation
5. run following program in a thermal cycler
   95°C/5min – 65°C/5min – 65°C/ (HOLD)
6. Prepare 0.2 mL tubes
7. Prepare 1:10 dilution of SureSelect RNase Block
8. Prepare mixture of Capture library (baits) and RNase Block dilution (1 µl and 2.5 µl)
9. Add 18.5 µl SureSelect XT2 hybridization buffer to the prepared 3.5 µl
10. Add 22 µl of Capture library mix to each sample pool with samples in the thermocycler
11. Incubate hybridization mixture for about 48 h at 65°C in the thermal cycler (evaporation should not exceed 8 µl)

S3.15 Protocol: Capturing the hybridized DNA

1. Prewarm SureSelect XT2 Wash 2 to 65°C in 5 mL tube in waterbath
2. Mix Dynabeads MyOne Streptavidin T1 thoroughly
3. Prepare 1,7 mL Lo-Bind Tubes - one for each sample
4. For each sample add 50 µL Dynabeads MyOne Streptavidin T1 to the prepared tubes
5. For each sample add 25 µL Dynabeads MyOne Streptavidin C1 to the prepared tubes
6. Add 200 µL SureSelect XT2 Binding Buffer to each tube and mix thoroughly
7. Place samples into the magnetic separator and wait 1 min
8. Remove the supernatant and discard it
9. Add 200 µL SureSelect XT2 Binding Buffer to each tube
10. Mix thoroughly by pipetting and gentle vortexing
11. Remove the supernatant and discard it
12. Remove the supernatant and discard it
13. Add 200 µL SureSelect XT2 Binding Buffer to each tube and mix thoroughly
14. Place samples into the magnetic separator and wait 1 min
15. Remove the supernatant and discard it
16. Resuspend the beads in 200 µL SureSelect XT2 Binding Buffer thoroughly by pipetting and gentle vortexing
17. Prepare the thermal mixer (1x 21°C with shaking and 1x 65°C without shaking)
18. keep the samples in the thermal cycler and estimate record the remaining volume

	SI1	SI2	DI1	DI2
Hybridization time (h)	~45 h	~45 h	~45 h	~45 h
Volume after hybridization	~10 µL	~10 µL	~10 µL	~10 µL

19. Transfer the entire volume to the 200 µL washed streptavidin beads and mix thoroughly
20. Incubate the samples for 30 minutes at 21°C in a thermal mixer 300 rpm
21. Prepare tubes for supernatant Ü1, Ü2, Ü3 (as a backup)
22. briefly spin the tubes, place the samples into the magnetic separator and wait 1 min

23.	Remove supernatant and add it into the previously prepared tubes for supernatant Ü1
24.	Resuspend the beads in 200 µL SureSelect XT2 Wash 1 thoroughly
25.	Briefly spin the tubes, place the samples into the magnetic separator and wait 1 min
26.	Remove supernatant and add it into the previously prepared tubes for supernatant Ü2
27.	Resuspend the beads in 200 µL of 65°C prewarmed SureSelect XT2 Wash 2 thoroughly
28.	Incubate the samples for 5 minutes at 65°C in a thermal mixer without shaking
29.	Briefly spin the tubes, place the samples into the magnetic separator and wait 1 min
30.	Remove supernatant and add it into the previously prepared tubes for supernatant Ü3 (only first wash)
31.	repeat step 25 through step 28 for 5 times
32.	Add 30 µL nuclease-free water to the beads and mix thoroughly
33.	If you do not continue, store samples at 4°C over night

S3.16 Protocol: Amplifying the captured libraries

1.	bring Ampure Beads to rt
2.	Prepare 0,2 mL tubes
3.	Prepare Post-Capture PCR Reaction Mix

Reagent	Volume for 1 reaction (µl)
Nuclease-free water	9.0
Herculase II PCR Master Mix	25.0
XT2 Primer Mix	1.0
Capture library	15.0
Total	50.0

4.	Mix reagents thoroughly by pipetting and gentle vortexing + short centrifugation
5.	run the following program in a thermal cycler 98°C/2min – 98°C/30sec – 60°C/30sec – 72°C/1min (9-12 cycles) – 72°C/10min – 4°C (HOLD)
6.	Purify samples with AmpBeads using a 1:0,7 concentration

S3.17 Table: Pooled single- and double indexing libraries for sequencing with concentration

Sample	Index 1	Index 2	Pool	Concentration	Volume
SI_SEW2_1	A01	/			
SI_SEW2_2	B01	/			
SI_SEW2_3	C01	/			
SI_SEW2_4	D01	/	SI1	2.50 ng/µL	20 µL
SI_SEW3_1	E01	/			
SI_SEW3_2	F01	/			
SI_SEW3_3	G01	/			
SI_SEW3_4	H01	/			
SI_SEW7_1	A02	/			
SI_SEW7_2	B02	/			
SI_SEW7_3	C02	/			
SI_SEW7_4	D02	/	SI2	2.97 ng/µL	20 µL
SI_SEW9_1	E02	/			
SI_SEW9_2	F02	/			
SI_SEW9_3	G02	/			
SI_SEW9_4	H02	/			
DI_SEW2_1	i701	i501			
DI_SEW2_2	i702	i502			
DI_SEW2_3	i703	i503			
DI_SEW2_4	i704	i504	DI1	1.57 ng/µL	20 µL
DI_SEW3_1	i705	i505			
DI_SEW3_2	i706	i506			
DI_SEW3_3	i707	i507			
DI_SEW3_4	i708	i508			
DI_SEW7_1	i709	i501			
DI_SEW7_2	i710	i502			

DI_SEW7_3	i711	i503			
DI_SEW7_4	i712	i504	DI2	1.57 ng/µL	20 µL
DI_SEW9_1	i701	i505			
DI_SEW9_2	i702	i506			
DI_SEW9_3	i703	i507			
DI_SEW9_4	i704	i508			

S3.18 Software script: Script used to run Trimmomatic on sequencing data for initial filtering

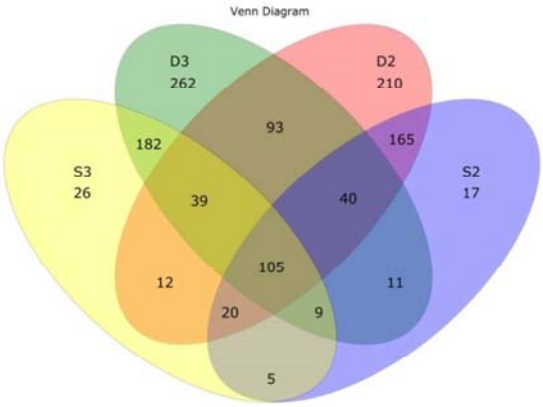
```
BATCH File:
@echo off

setlocal EnableDelayedExpansion

for %%X in (%1\*R1.fastq) do (
echo %%X
set _fastq_forward=%%X
set _fastq_reverse=!_fastq_forward:R1.fastq=R2.fastq!
set _fastq_forward_paired=!_fastq_forward:R1.fastq=R1_paired.fastq!
set _fastq_forward_unpaired=!_fastq_forward:R1.fastq=R1_unpaired.fastq!
set _fastq_reverse_paired=!_fastq_reverse:R2.fastq=R2_paired.fastq!
set _fastq_reverse_unpaired=!_fastq_reverse:R2.fastq=R2_unpaired.fastq!
start /wait java –jar C:\PATH-TO-FILE\trimmomatic-0.36.jar PE -phred33 !_fastq_forward! !_fastq_reverse!
!_fastq_forward_paired! !_fastq_forward_unpaired! !_fastq_reverse_paired! !_fastq_reverse_unpaired!
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:100
)
```

S3.19 Software script: Script used to run mothur on sequencing data

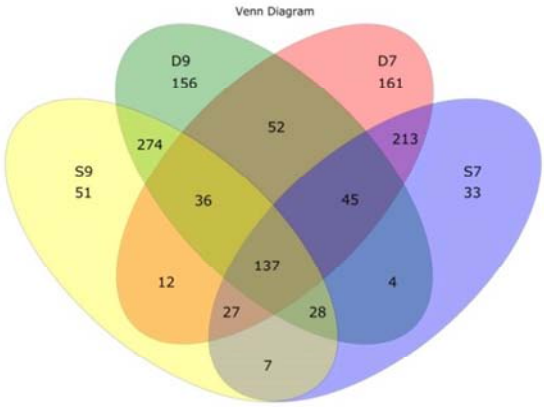
```
#!/bin/bash

#$ -cwd
#$ -j y
#$ -S /bin/bash
#$ -M janstruwe@msn.com
#$ -m be
#$ -N TE_DS_STOP

#PRESETS
#FILES
working_dir="/share/pool/jstruwe/TE"
mothur_path="/share/scientific_bin/mothur/1.38.1.1/mothur"
reference_sequences_alignment="/share/pool/jstruwe/GBOL_all_checked_filtered_mod_aln.fasta"
reference_sequences_database="/share/pool/jstruwe/GBOL_all_checked_filtered_mod.fasta"
taxfile="/share/pool/jstruwe/GBOL_all_checked_filtered_mod.tax"
files_DS="/share/pool/jstruwe/TE/TE_DS_STOPsub.files"
contigsreport="/share/pool/jstruwe/TE/TE_DS_STOPsub.contigs.report"

#VARIABLES
ambiguities=0
processors=10
minoverlap=50
maxhomop=10
startafter=468
endbefore=190
preclusterdiffs=3
taxpropcutoff=80
cluster=”agc”
#choose from
#abundance="agc"
#distance="dgc"
#opticlust="opti"
#average="average"
clustercutoff=0.03
removesingle=1
subsample_3=30000
subsample_5=9500
```

```
batchfile=${files_DS/.files/_mothur.batch}
echo "$batchfile"

#COMMANDS
echo "set.dir(input=$working_dir)" > $batchfile
echo "make.contigs(file=$files_DS, processors=$processors)" >> $batchfile
echo "screen.seqs(fasta=current, group=current, minoverlap=$minoverlap, contigsreport=$contigsreport,
maxambig=$ambiguities, maxhomop=$maxhomop, minlength=190)" >> $batchfile
echo "unique.seqs(fasta=current)" >> $batchfile
echo "count.seqs(name=current, group=current)" >> $batchfile
echo "align.seqs(fasta=current, reference=$reference_sequences_alignment, flip=t)" >> $batchfile
echo "screen.seqs(fasta=current, count=current, start=$startafter, end=$endbefore)" >> $batchfile
echo "filter.seqs(fasta=current, vertical=T)" >> $batchfile
echo "unique.seqs(fasta=current, count=current)" >> $batchfile
echo "pre.cluster(fasta=current, count=current, diffs=$preclusterdiffs)" >> $batchfile
echo "chimera.uchime(fasta=current, reference=$reference_sequences_alignment)" >> $batchfile
echo "remove.seqs(fasta=current, accnos=current, count=current)" >> $batchfile

***EINSCHUB***
*Seperation of the sequences into 3‘ and 5‘ subsets *
**********

echo "classify.seqs(fasta=current, count=current, reference=$reference_sequences_database, taxonomy=$taxfile,
cutoff=$taxpropcutoff)" >> $batchfile
echo "cluster(fasta=current, count=current, method=$cluster, cutoff=$clustercutoff)" >> $batchfile
echo "remove.rare(list=current, count=current, nseqs=$removesingle, label=$clustercutoff)" >> $batchfile
echo "classify.otu(list=current, count=current, taxonomy=current, label=$clustercutoff)" >> $batchfile
echo "make.shared(list=current, count=current)" >> $batchfile
echo "count.groups(shared=current)" >> $batchfile
echo " venn(shared=current, groups=S2-S3-S7-S9)" >> $batchfile
echo " venn(shared=current, groups=D2-D3-D7-D9)" >> $batchfile
echo " venn(shared=current, groups=S2-D2-S3-D3)" >> $batchfile
echo " venn(shared=current, groups=S7-D7-S9-D9)" >> $batchfile
echo "heatmap.bin(shared=current, scale=log2, numotu=25)" >> $batchfile
echo "summary.single(shared=current, calc=nseqs-coverage-sobs-invsimpson, subsample=T)" >> $batchfile
echo "rarefaction.single(shared=current, calc=sobs, freq=100, groupmode=T)" >> $batchfile
echo "dist.shared(shared=current, calc=thetayc-jclass, subsample=$subsample)" >> $batchfile
echo "heatmap.sim(phylip=*.jclass.0.03.lt.ave.dist)" >> $batchfile
echo "heatmap.sim(phylip=*. thetayc.0.03.lt.ave.dist)" >> $batchfile
echo "tree.shared(phylip=*.thetayc.0.03.lt.ave.dist)" >> $batchfile
echo "amova(phylip=*.thetayc.0.03.lt.ave.dist, design=type.design, iters=10000)" >> $batchfile
echo "homova(phylip=*.thetayc.0.03.lt.ave.dist, design=forest.type.design, iters=10000) "
echo "sub.sample(shared=current) " >> $batchfile
echo "rarefaction.single(shared=current, calc=sobs, freq=100) " >> $batchfile
echo "get.communitytype(shared=*.subsample.shared) " >> $batchfile
echo "lefse(shared=stability.opti_mcc.0.03.subsample.shared, design=forest.type.design) " >> $batchfile

$mothur_path $batchfile
```

S3.20 Figure: Venn diagrams of shared MOTUs comparing (S) single- and (D) double-indexing samples SEW2, SEW3, SEW7, SEW9. The diagram for 3' subsets shows that 96% of the MOTUs of the single-indexing samples are covered by the double-indexing samples from the pine plots and 93%'are covered from the beech plot.

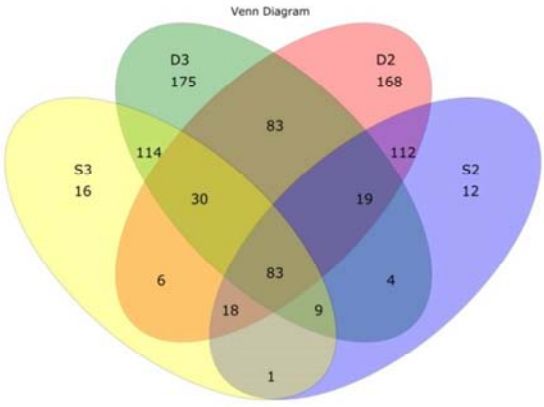


The number of species in group D2 is 684
The number of species in group D3 is 741
The number of species in group S2 is 372
The number of species in group S3 is 398
The number of species shared between groups D2 and D3 is 277
The number of species shared between groups D2 and S2 is 330
The number of species shared between groups D2 and S3 is 176
The number of species shared between groups D3 and S2 is 165
The number of species shared between groups D3 and S3 is 335
The number of species shared between groups S2 and S3 is 139
The number of species shared between groups D2, D3 and S2 is 145
The number of species shared between groups D2, D3 and S3 is 144
The number of species shared between groups D2, S2 and S3 is 125
The number of species shared between groups D3, S2 and S3 is 114
The total richness of all the groups is 1196

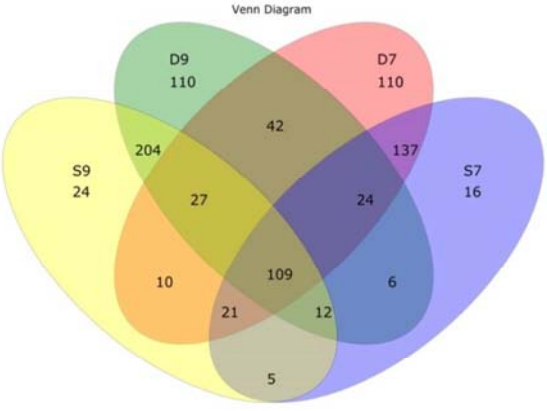


The number of species in group D7 is 683
The number of species in group D9 is 732
The number of species in group S7 is 494
The number of species in group S9 is 572
The number of species shared between groups D7 and D9 is 270
The number of species shared between groups D7 and S7 is 422
The number of species shared between groups D7 and S9 is 212
The number of species shared between groups D9 and S7 is 214
The number of species shared between groups D9 and S9 is 475
The number of species shared between groups S7 and S9 is 199
The number of species shared between groups D7, D9 and S7 is 182
The number of species shared between groups D7, D9 and S9 is 173
The number of species shared between groups D7, S7 and S9 is 164
The number of species shared between groups D9, S7 and S9 is 165
The total richness of all the groups is 1236

S3.21 Figure: Venn diagrams of shared MOTUs comparing (S) single- and (D) double-indexing samples SEW2, SEW3, SEW7, SEW9. The diagram for 5' subsets shows that 96% of the MOTUs of the single-indexing samples are covered by the double-indexing samples from the pine plots and 95%'are covered from the beech plot.

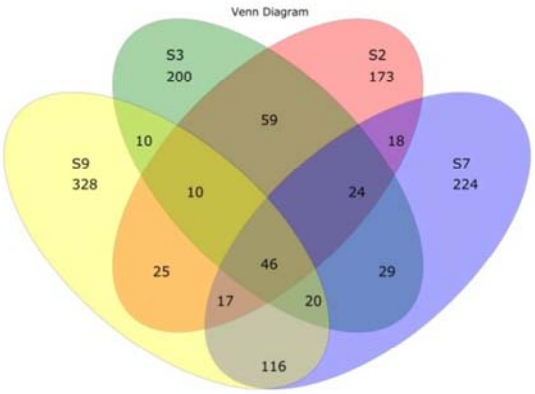


The number of species in group D2 is 519
The number of species in group D3 is 517
The number of species in group S2 is 258
The number of species in group S3 is 277
The number of species shared between groups D2 and D3 is 215
The number of species shared between groups D2 and S2 is 232
The number of species shared between groups D2 and S3 is 137
The number of species shared between groups D3 and S2 is 115
The number of species shared between groups D3 and S3 is 236
The number of species shared between groups S2 and S3 is 111
The number of species shared between groups D2, D3 and S2 is 102
The number of species shared between groups D2, D3 and S3 is 113
The number of species shared between groups D2, S2 and S3 is 101
The number of species shared between groups D3, S2 and S3 is 92
The total richness of all the groups is 850

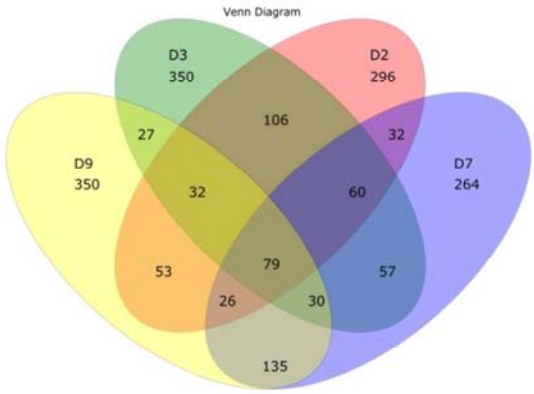


The number of species in group D7 is 480
The number of species in group D9 is 534
The number of species in group S7 is 330
The number of species in group S9 is 412
The number of species shared between groups D7 and D9 is 202
The number of species shared between groups D7 and S7 is 291
The number of species shared between groups D7 and S9 is 167
The number of species shared between groups D9 and S7 is 151
The number of species shared between groups D9 and S9 is 352
The number of species shared between groups S7 and S9 is 147
The number of species shared between groups D7, D9 and S7 is 133
The number of species shared between groups D7, D9 and S9 is 136
The number of species shared between groups D7, S7 and S9 is 130
The number of species shared between groups D9, S7 and S9 is 121
The total richness of all the groups is 857

S3.22 Figure: Venn diagrams of shared MOTUs comparing (S) single- and (D) double-indexing samples SEW2, SEW3, SEW7, SEW9. The diagram for 3' subsets shows that single- and double- indexing both affirm the highest convergence of MOTUs within the pine and beech plot samples than between them.

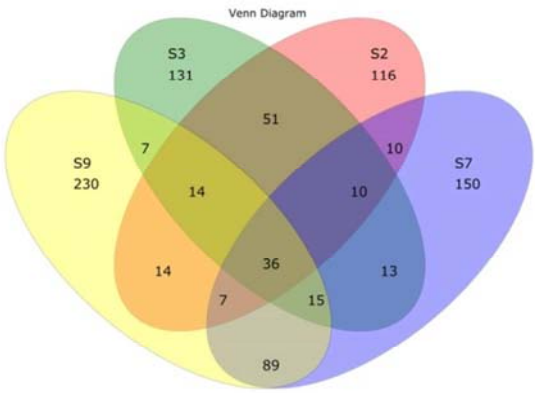


The number of species in group S2 is 372
The number of species in group S3 is 398
The number of species in group S7 is 494
The number of species in group S9 is 572
The number of species shared between groups S2 and S3 is **139**
The number of species shared between groups S2 and S7 is 105
The number of species shared between groups S2 and S9 is 98
The number of species shared between groups S3 and S7 is 119
The number of species shared between groups S3 and S9 is 86
The number of species shared between groups S7 and S9 is **199**
The number of species shared between groups S2, S3 and S7 is 70
The number of species shared between groups S2, S3 and S9 is 56
The number of species shared between groups S2, S7 and S9 is 63
The number of species shared between groups S3, S7 and S9 is 66
The total richness of all the groups is 1299

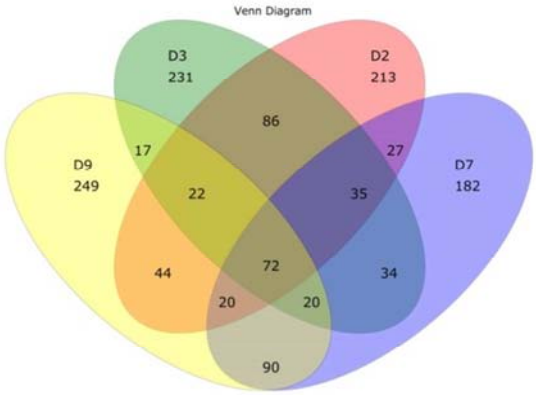


The number of species in group D2 is 684
The number of species in group D3 is 741
The number of species in group D7 is 683
The number of species in group D9 is 732
The number of species shared between groups D2 and D3 is **277**
The number of species shared between groups D2 and D7 is 197
The number of species shared between groups D2 and D9 is 190
The number of species shared between groups D3 and D7 is 226
The number of species shared between groups D3 and D9 is 168
The number of species shared between groups D7 and D9 is **270**
The number of species shared between groups D2, D3 and D7 is 139
The number of species shared between groups D2, D3 and D9 is 111
The number of species shared between groups D2, D7 and D9 is 105
The number of species shared between groups D3, D7 and D9 is 109
The total richness of all the groups is 1897

S3.23 Figure: Venn diagrams of shared MOTUs comparing (S) single- and (D) double-indexing samples SEW2, SEW3, SEW7, SEW9. The diagram for 5' subsets shows that single and double indexing both affirm the highest convergence of MOTUs within the pine and beech plot samples than between them.



The number of species in group S2 is 258
The number of species in group S3 is 277
The number of species in group S7 is 330
The number of species in group S9 is 412
The number of species shared between groups S2 and S3 is **111**
The number of species shared between groups S2 and S7 is 63
The number of species shared between groups S2 and S9 is 71
The number of species shared between groups S3 and S7 is 74
The number of species shared between groups S3 and S9 is 72
The number of species shared between groups S7 and S9 is **147**
The number of species shared between groups S2, S3 and S7 is 46
The number of species shared between groups S2, S3 and S9 is 50
The number of species shared between groups S2, S7 and S9 is 43
The number of species shared between groups S3, S7 and S9 is 51
The total richness of all the groups is 893



The number of species in group D2 is 519
The number of species in group D3 is 517
The number of species in group D7 is 480
The number of species in group D9 is 534
The number of species shared between groups D2 and D3 is **215**
The number of species shared between groups D2 and D7 is 154
The number of species shared between groups D2 and D9 is 158
The number of species shared between groups D3 and D7 is 161
The number of species shared between groups D3 and D9 is 131
The number of species shared between groups D7 and D9 is **202**
The number of species shared between groups D2, D3 and D7 is 107
The number of species shared between groups D2, D3 and D9 is 94
The number of species shared between groups D2, D7 and D9 is 92
The number of species shared between groups D3, D7 and D9 is 92
The total richness of all the groups is 1342

S3.24 Table: Detailed list of all identifications of the Bayesian classifier for all (SI) single- and (DI) double-indexed samples and their 3' and 5' sorted subsets. Taxonomy: Different classes are written bold and in capitals. Different orders are written bold. Family levels are written regular. Genus and species are italic. The name affix "uncl." (for unclear) points out from which taxonomic level on a further classification was not possible for the MOTUs in the following collumns.

			SI 3'				SI 5'				DI 3'				DI 5'			
			SEW2	SEW3	SEW7	SEW9	SEW2	SEW3	SEW7	SEW9	SEW2	SEW3	SEW7	SEW9	SEW2	SEW3	SEW7	SEW9
ARACHNIDA																		
Arachnida uncl.																		
Arachnida uncl.	*Arachnida uncl.*	*Arachnida uncl.*	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Araneae																		
Anyphaenidae	*Anyphaena*	*accentuata*	1	1	0	2	0	0	0	1	2	2	0	2	2	3	0	2
Araneae uncl.	*Araneae uncl.*	*Araneae uncl.*	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0
Araneidae	*Araniella*	*cucurbitina*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Araneidae	*Cyclosa*	*conica*	0	2	0	0	0	0	0	0	0	3	0	0	0	1	0	0
Clubionidae	*Clubiona*	*pallidula*	0	0	2	2	0	0	1	1	0	0	2	2	0	0	2	2
Clubionidae	*Clubiona*	*terrestris*	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2	0
Linyphiidae	*Hypomma*	*cornutum*	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0
Linyphiidae	*Linyphia*	*triangularis*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Linyphiidae	*Tenuiphantes*	*flavipes*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Philodromidae	*Philodromus*	*albidus*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
Philodromidae	*Philodromus*	*aureolus*	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2	0
Philodromidae	*Philodromus*	*dispar*	2	2	0	0	0	0	0	0	3	3	0	0	3	2	0	0
Pisauridae	*Pisaura*	*mirabilis*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Tetragnathidae	*Metellina*	*mengei*	0	1	2	0	1	0	1	0	1	2	1	0	1	1	1	0
Tetragnathidae	*Tetragnatha*	*obtusa*	0	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0
Theridiidae	*Paidiscura*	*pallens*	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0
Theridiidae	*Platnickina*	*tincta*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Thomisidae	*Ozyptila*	*praticola*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Thomisidae	*Xysticus*	*audax*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Thomisidae	*Xysticus*	*cristatus*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Thomisidae	*Xysticus*	*lanio*	0	1	0	0	0	1	0	0	0	3	0	0	0	1	0	0
Thomisidae	*Xysticus*	*Xysticus uncl.*	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0	0
Opiliones																		
Phalangiidae	*Rilaena*	*triangularis*	0	0	1	5	0	0	0	3	0	0	0	7	0	0	0	5
Pseudoscorpionida																		
Chernetidae	*Pselaphochernes*	*scorpioides*	0	1	0	0	0	1	0	0	0	2	1	0	0	2	0	0
ENTOGNATHA																		
Entomobryomorpha																		
Entomobryidae	*Orchesella*	*flavescens*	0	2	0	2	0	1	0	1	0	2	0	2	0	1	0	1
INSECTA																		
Blattodea																		
Blattodea uncl.	*Blattodea uncl.*	*Blattodea uncl.*	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Ectobiidae	*Ectobius*	*sylvestris*	4	1	0	0	4	2	0	0	3	0	0	0	4	0	0	0
Coleoptera																			
Anobiidae	*Dorcatoma*	*dresdensis*	0	0	3	2	0	0	2	2	0	0	2	2	0	0	3	2
Anobiidae	*Ernobius*	*longicornis*	2	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0
Anobiidae	*Hedobia*	*imperialis*	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2
Anobiidae	*Xestobium*	*plumbeum*	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
Anthribidae	*Brachytarsus*	*nebulosus*	2	2	0	0	1	0	0	0	2	3	0	0	1	1	0	0
Aspidiphoridae	*Arpidiphorus*	*orbiculatus*	0	1	0	2	0	1	0	1	0	1	0	2	0	1	0	1
Buprestidae	*Agrilus*	*angustulus*	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2
Byturidae	*Byturus*	*tomentosus*	6	7	3	2	7	8	2	4	13	12	2	2	10	8	4	2
Cantharidae	*Cantharis*	*nigricans*	2	2	1	2	2	3	0	3	7	10	0	7	6	5	0	4
Cantharidae	*Cantharis*	*pellucida*	0	2	0	0	0	3	0	0	0	3	0	0	0	3	0	0
Cantharidae	*Rhagonycha*	*lignosa*	1	3	0	2	1	0	0	2	3	3	0	3	4	3	0	6
Cerambycidae	*Alosterna*	*tabacicolor*	11	1	7	12	7	0	6	8	14	0	8	12	6	0	7	6
Cerambycidae	*Anoplodera*	*sexguttata*	0	0	2	9	0	0	0	11	0	0	0	11	0	0	0	10
Cerambycidae	*Clytus*	*arietis*	0	0	2	7	0	0	2	5	0	0	0	5	0	0	0	6
Cerambycidae	*Clytus*	*Clytus uncl.*	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Cerambycidae	*Clytus*	*lama*	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	1
Cerambycidae	*Cortodera*	*femorata*	4	0	0	0	5	0	0	0	3	0	0	0	5	0	0	0
Cerambycidae	*Grammoptera*	*ruficornis*	1	0	0	1	1	1	0	1	3	3	0	2	2	2	0	2
Cerambycidae	*Rhagium*	*inquisitor*	5	2	0	0	4	2	1	0	7	3	0	0	3	2	0	0
Cerambycidae	*Rhagium*	*mordax*	0	4	0	0	1	2	0	0	0	6	0	0	0	3	0	0
Cerambycidae	*Tetrops*	*praeustus*	0	1	0	0	0	1	0	0	0	4	0	0	0	1	0	0
Cholevidae	*Catops*	*nigricantoides*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Cleridae	*Thanasimus*	*formicarius*	4	1	0	0	3	0	0	0	8	0	0	0	6	0	0	0
Cleridae	*Thanasimus*	*Thanasimus uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Cleridae	*Tillus*	*elongatus*	0	0	0	6	0	0	1	5	0	0	0	6	0	0	0	5
Coccinellidae	*Harmonia*	*axyridis*	0	0	3	0	0	0	2	0	0	0	3	0	0	0	2	0
Coccinellidae	*Propylea*	*quatuordecimpunctata*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Coccinellidae	*Psyllobora*	*vigintiduopunctata*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Coleoptera uncl.	*Coleoptera uncl.*	*Coleoptera uncl.*	12	16	9	23	9	9	5	16	19	18	11	21	15	18	7	20
Curculionidae	*Anthonomus*	*rubi*	0	0	0	0	0	0	0	0	1	2	0	0	1	1	0	0
Curculionidae	*Brachonyx*	*pineti*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Curculionidae	*Curculio*	*pyrrhoceras*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Curculionidae	*Hylobius*	*abietis*	1	6	0	0	0	2	0	0	0	8	0	0	0	4	0	0
Curculionidae	*Phyllobius*	*arborator*	4	1	0	0	3	0	0	0	5	0	0	0	3	0	0	0
Curculionidae	*Phyllobius*	*argentatus*	2	1	19	3	1	0	10	4	3	0	19	5	3	0	11	3
Curculionidae	*Strophosoma*	*capitatum*	0	0	0	2	0	0	0	2	0	0	0	4	0	0	0	2
Curculionidae	*Xyleborinus*	*saxeseni*	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Elateridae	*Ampedus*	*Ampedus uncl.*	1	1	0	5	2	1	0	6	3	3	0	3	4	0	0	4
Elateridae	*Ampedus*	*balteatus*	4	1	2	12	4	1	0	10	5	0	0	8	4	1	0	6
Elateridae	*Ampedus*	*elongatulus*	1	4	0	1	1	5	0	0	0	2	0	0	0	3	1	0
Elateridae	*Ampedus*	*pomorum*	0	7	2	7	1	7	2	9	0	7	0	6	1	6	0	8
Elateridae	*Athous*	*haemorrhoidalis*	12	15	0	0	3	5	1	0	7	11	0	0	4	5	0	0
Elateridae	*Athous*	*subfuscus*	3	12	6	1	2	11	5	1	0	10	7	0	0	8	6	0
Elateridae	*Cardiophorus*	*nigerrimus*	1	3	0	0	0	2	0	0	0	2	0	0	0	2	0	0

Elateridae	*Dalopius*	*marginatus*	6	7	7	8	6	7	9	8	5	6	6	5	5	8	9	6
Elateridae	*Denticollis*	*linearis*	5	2	2	10	8	1	1	8	7	0	0	8	7	0	0	6
Elateridae	*Dicronychus*	*cinereus*	1	2	0	0	0	2	0	0	0	2	0	0	0	3	0	0
Elateridae	*Ectinus*	*aterrimus*	14	11	10	23	12	12	10	22	8	7	7	20	8	9	9	16
Elateridae	*Elateridae uncl.*	*Elateridae uncl.*	11	4	5	17	4	4	2	7	15	5	5	14	10	4	2	13
Elateridae	*Elateridae-2569354*	*sp*	2	0	0	5	0	0	0	0	7	0	0	8	0	0	0	0
Elateridae	*Elateridae-2569355*	*sp*	0	0	0	0	3	0	1	9	0	0	0	0	7	0	0	10
Elateridae	*Elateridae-2569362*	*sp*	0	2	2	0	0	0	0	0	0	2	1	0	0	0	0	0
Elateridae	*Elateridae-2569363*	*sp*	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0
Elateridae	*Melanotus*	*castanipes*	0	0	2	22	0	0	2	20	0	0	0	16	0	0	0	15
Elateridae	*Melanotus*	*Melanotus uncl.*	0	0	0	3	0	0	0	3	0	0	0	3	0	0	1	4
Elateridae	*Melanotus*	*rufipes*	18	34	1	3	12	23	1	1	18	26	1	3	14	22	0	2
Erotylidae	*Tritoma*	*bipustulata*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Eucnemidae	*Melasis*	*buprestoides*	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Lagriidae	*Lagria*	*atripes*	0	1	0	0	0	4	0	0	0	2	0	0	0	5	0	0
Leiodidae	*Agathidium*	*seminulum*	0	1	0	0	1	1	0	0	0	3	0	0	0	1	0	0
Lymexylidae	*Hylecoetus*	*dermestoides*	1	4	1	0	0	3	1	1	0	4	2	0	0	6	2	0
Malachiidae	*Anthocomus*	*fasciatus*	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0
Melandryidae	*Orchesia*	*undulata*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	2
Melyridae	*Dasytes*	*aeratus*	0	2	0	0	0	1	0	0	0	2	0	0	0	3	0	0
Melyridae	*Dasytes*	*Dasytes uncl.*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Melyridae	*Dasytes*	*plumbeus*	0	0	2	0	1	0	1	0	2	0	2	1	1	0	1	0
Nitidulidae	*Meligethes*	*aeneus*	1	0	0	0	2	0	0	0	1	0	0	0	1	0	0	0
Ptinidae	*Hedobia*	*imperialis*	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Pyrochroidae	*Pyrochroa*	*coccinea*	4	1	0	0	2	0	0	0	5	0	0	0	3	0	0	0
Pyrochroidae	*Schizotus*	*pectinicornis*	1	2	0	2	0	2	1	2	0	2	0	3	0	2	0	2
Rhynchitidae	*Caenorhinus*	*germanicus*	0	2	0	0	0	1	0	0	1	2	0	0	1	1	0	0
Scirtidae	*Cyphon*	*coarctatus*	1	1	1	6	0	1	2	10	1	2	0	8	0	2	0	12
Scirtidae	*Cyphon*	*Cyphon uncl.*	0	0	0	6	0	0	0	1	0	0	0	5	0	0	0	2
Scirtidae	*Cyphon*	*variabilis*	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Scolytidae	*Xyleborus*	*germanus*	0	3	0	4	0	1	0	1	0	3	0	4	0	1	0	1
Scraptiidae	*Anaspis*	*frontalis*	2	3	0	0	1	2	0	0	3	2	0	1	3	2	0	0
Scraptiidae	*Anaspis*	*ruficollis*	1	3	0	0	0	2	0	0	0	3	0	0	0	2	0	0
Scraptiidae	*Anaspis*	*rufilabris*	2	0	5	6	0	0	7	5	2	0	5	5	1	0	4	4
Scraptiidae	*Anaspis*	*thoracica*	3	2	0	2	1	1	0	3	2	2	0	3	2	2	1	3
Silvanidae	*Uleiota*	*planata*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Staphylinidae	*Lordithon*	*lunulatus*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Staphylinidae	*Placusa*	*complanata*	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	1
Staphylinidae	*Scaphidium*	*quadrimaculatum*	2	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0
Staphylinidae	*Sepedophilus*	*littoreus*	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	2
Tenebrionidae	*Lagria*	*atripes*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Throscidae	*Aulonothroscus*	*brevicollis*	0	0	0	2	0	0	0	2	0	0	0	1	0	0	0	2
Throscidae	*Trixagus*	*carinifrons*	2	4	0	0	2	2	0	0	2	4	0	0	2	2	0	0
Throscidae	*Trixagus*	*dermestoides*	0	0	3	5	0	0	1	5	0	0	2	3	0	0	2	2
Throscidae	*Trixagus*	*leseigneuri*	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
Throscidae	*Trixagus*	*meybohmi*	2	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0

Diptera																			
Agromyzidae	*Agromyza*	*pseudoreptans*	2	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0
Agromyzidae	*Chromatomyia*	*milii*	0	1	0	0	0	1	0	0	2	2	0	0	1	1	0	0
Agromyzidae	*Chromatomyia*	*obscuriceps*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Agromyzidae	*Liriomyza*	*taurica*	2	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Anisopodidae	*Sylvicola*	*cinctus*	0	0	0	0	0	0	0	0	1	3	0	1	0	2	0	0
Anthomyiidae	*Anthomyiidae uncl.*	*Anthomyiidae uncl.*	1	2	6	5	2	0	4	4	2	5	8	7	2	1	4	4
Anthomyiidae	*Botanophila*	*fugax*	2	1	0	1	1	0	0	0	2	0	0	0	2	0	0	0
Anthomyiidae	*Emmesomyia*	*grisea*	6	4	0	0	4	3	0	0	7	9	1	0	4	3	0	0
Anthomyiidae	*Hydrophoria*	*Hydrophoria uncl.*	9	8	12	4	6	6	9	5	19	19	18	12	17	13	12	8
Anthomyiidae	*Hydrophoria*	*lancifer*	1	1	2	1	0	0	0	0	5	5	6	2	4	4	3	2
Anthomyiidae	*Hydrophoria*	*linogrisea*	1	3	1	1	1	1	1	1	4	4	3	2	1	2	3	2
Anthomyiidae	*Hylemya*	*nigrimana*	0	3	6	1	0	2	3	2	0	4	5	0	0	2	3	0
Anthomyiidae	*Mycophaga*	*testacea*	0	4	0	0	0	2	0	0	0	3	0	0	0	2	0	0
Anthomyiidae	*Pegomya*	*silacea*	0	0	0	3	0	0	0	1	0	0	0	4	1	0	0	2
Calliphoridae	*Bellardia*	*Bellardia uncl.*	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2
Calliphoridae	*Bellardia*	*viarum*	3	0	2	4	4	1	1	4	5	0	0	8	6	1	0	6
Calliphoridae	*Calliphoridae uncl.*	*Calliphoridae uncl.*	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0
Cecidomyiidae	*Aprionus*	*cardiophorus*	0	0	1	4	0	0	0	2	0	0	2	4	0	0	1	3
Cecidomyiidae	*Aprionus*	*similis*	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
Cecidomyiidae	*Aprionus*	*spiniger*	0	0	1	0	0	0	0	0	0	0	5	0	0	0	3	0
Cecidomyiidae	*Asynapta*	*pectoralis*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Cecidomyiidae	*Bryomyia*	*apsectra*	1	1	1	0	0	0	0	0	2	2	1	1	0	0	0	0
Cecidomyiidae	*Bryomyia*	*Bryomyia uncl.*	0	0	0	0	0	1	0	0	1	1	0	0	1	1	2	1
Cecidomyiidae	*Camptomyia*	*abnormis*	0	0	0	0	0	0	0	0	1	2	0	0	0	1	0	0
Cecidomyiidae	*Camptomyia*	*Camptomyia uncl.*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Cecidomyiidae	*Camptomyia*	*spinifera*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Cecidomyiidae	*Cecidomyiidae uncl.*	*Cecidomyiidae uncl.*	1	1	1	0	2	3	0	1	8	7	1	3	6	8	1	6
Cecidomyiidae	*Divellepidosis*	*hypoxantha*	0	0	0	0	0	0	0	0	2	4	0	0	2	3	0	0
Cecidomyiidae	*Lestremia*	*leucophaea*	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
Cecidomyiidae	*Lestremiinae*	*sp*	0	0	0	0	0	0	0	0	0	2	1	0	0	1	1	0
Cecidomyiidae	*Peromyia*	*perpusilla*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Cecidomyiidae	*Porricondyla*	*fulvescens*	0	0	2	1	0	0	1	1	0	0	3	1	0	0	4	1
Cecidomyiidae	*Winnertzia*	*curvata*	0	0	0	0	0	0	0	0	2	1	0	1	0	0	0	0
Cecidomyiidae	*Winnertzia*	*tridens*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Cecidomyiidae	*Winnertzia*	*xylostei*	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
Cecidomyiidae	*Xylopriona*	*atra*	0	0	1	0	0	0	1	0	0	0	3	1	0	0	1	1
Ceratopogonidae	*Atrichopogon*	*lucorum*	0	0	1	1	0	0	0	1	0	1	0	1	0	1	1	1
Ceratopogonidae	*Ceratopogonidae*	*Ceratopogonidae uncl.*	0	0	2	2	0	0	2	1	0	2	2	3	0	0	2	2
Ceratopogonidae	*Ceratopogonidae*	*sp6*	0	0	1	2	0	0	0	1	0	2	1	2	0	1	1	1
Ceratopogonidae	*Ceratopogonidae*	*sp7*	0	2	0	2	0	1	0	1	0	2	0	2	0	1	0	1
Ceratopogonidae	*Ceratopogonidae*	*sp9*	0	0	0	1	0	0	0	1	0	0	2	3	0	0	1	1
Ceratopogonidae	*Serromyia*	*femorata*	0	0	2	2	0	0	1	1	2	0	2	3	1	0	1	2
Chaoboridae	*Chaoborus*	*flavicans*	0	0	1	2	0	0	1	1	0	0	0	4	0	0	0	2
Chironomidae	*Ablabesmyia*	*Ablabesmyia uncl.*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Chironomidae	*Chironomidae*	*Chironomidae uncl.*	1	0	1	2	0	0	0	1	2	0	0	2	1	0	0	1

Chironomidae	*Chironomidae*	*sp19*	0	0	1	0	0	0	1	0	0	0	2	0	0	0	1	0
Chironomidae	*Chironomidae*	*sp2*	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0
Chironomidae	*Chironomidae*	*sp22*	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	3
Chironomidae	*Chironomidae*	*sp28*	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Chironomidae	*Chironomidae*	*sp32*	0	0	1	1	0	0	1	1	0	0	2	2	1	0	1	1
Chironomidae	*Chironomidae uncl.*	*Chironomidae uncl.*	0	0	2	0	0	0	4	1	0	0	2	0	0	0	4	2
Chironomidae	*Limnophyes*	*angelicae*	2	2	1	2	2	2	1	2	2	2	2	4	2	2	2	2
Chironomidae	*Micropsectra*	*atrofasciata*	0	1	0	0	0	1	0	0	0	2	0	0	0	1	0	0
Chironomidae	*Xanthochlorus*	*nigricans*	1	0	0	2	0	0	0	2	2	0	0	2	1	0	0	2
Chloropidae	*Chlorops*	*rossicus*	0	1	0	0	0	1	0	0	0	2	0	1	0	1	0	0
Chloropidae	*Dicraeus*	*styriacus*	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Chloropidae	*Oscinella*	*frit*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Conopidae	*Myopa*	*buccata*	0	4	0	0	1	3	0	0	0	8	0	0	0	4	0	0
Culicidae	*Aedes*	*cinereus*	0	0	0	3	0	0	0	2	0	0	0	3	0	0	0	2
Culicidae	*Culiseta*	*alaskaensis*	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2
Diadocidiidae	*Diadocidia*	*ferruginosa*	2	2	0	3	1	1	0	1	3	2	0	2	1	1	0	1
Diadocidiidae	*Diadocidia*	*spinosula*	0	1	0	0	0	0	0	0	2	2	0	0	1	1	0	1
Diptera uncl.	*Diptera uncl.*	*Diptera uncl.*	76	134	118	174	49	73	78	109	159	259	164	212	112	167	118	148
Ditomyiidae	*Ditomyia*	*fasciata*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Ditomyiidae	*Symmerus*	*annulatus*	0	0	0	4	0	0	0	3	0	0	0	4	0	0	0	3
Dolichopodidae	*Argyra*	*diaphana*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Dolichopodidae	*Dolichopodidae uncl.*	*Dolichopodidae uncl.*	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	2
Dolichopodidae	*Dolichopus*	*claviger*	0	0	3	3	0	0	3	2	0	0	3	3	0	0	3	2
Dolichopodidae	*Dolichopus*	*Dolichopus uncl.*	0	2	2	18	1	2	1	15	1	5	3	22	0	4	3	16
Dolichopodidae	*Dolichopus*	*festivus*	0	0	1	0	0	0	2	1	0	0	1	0	0	0	2	1
Dolichopodidae	*Dolichopus*	*nigricornis*	0	0	2	21	0	0	2	6	0	0	0	22	0	0	1	4
Dolichopodidae	*Dolichopus*	*plumipes*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Dolichopodidae	*Hercostomus*	*aerosus*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Dolichopodidae	*Hercostomus*	*Hercostomus uncl.*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Dolichopodidae	*Neurigona*	*pallida*	2	0	0	0	2	0	0	0	2	0	0	0	3	0	0	0
Dolichopodidae	*Neurigona*	*quadrifasciata*	13	14	20	19	10	5	12	9	16	23	22	19	13	14	13	16
Dolichopodidae	*Sciapus*	*longulus*	0	0	0	3	0	0	0	1	0	0	0	2	0	0	0	1
Dolichopodidae	*Xanthochlorus*	*sp*	0	0	0	1	0	0	0	1	0	0	0	3	0	0	0	2
Drosophilidae	*Drosophila*	*kuntzei*	0	2	0	0	0	2	0	0	0	2	0	0	0	2	1	0
Dryomyzidae	*Dryomyza*	*decrepita*	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	2
Dryomyzidae	*Dryomyza*	*decrepita*	0	0	0	2	0	0	0	0	0	0	0	3	0	0	0	0
Dryomyzidae	*Dryomyza*	*flaveola*	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
Dryomyzidae	*Dryomyza uncl.*	*Dryomyza uncl.*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Empididae	*Empididae uncl.*	*Empididae uncl.*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Empididae	*Empis*	*aestiva*	0	0	4	3	0	0	2	2	0	0	3	4	0	0	2	2
Empididae	*Empis*	*Empis uncl.*	1	3	0	0	1	1	0	0	2	2	0	0	2	2	0	0
Empididae	*Empis*	*tessellata*	17	9	1	0	14	11	1	0	23	21	0	0	21	16	0	1
Empididae	*Hilara*	*interstincta*	0	0	2	2	1	0	2	1	0	0	2	2	0	0	2	2
Empididae	*Phyllodromia*	*melanocephala*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Empididae	*Rhamphomyia*	*longipes*	2	2	3	2	2	2	2	2	2	2	3	2	2	2	2	2
Fanniidae	*Fannia*	*corvina*	1	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0

Fanniidae	*Fannia*	*Fannia uncl.*	2	1	0	0	1	1	0	0	2	1	0	0	1	0	0	0
Fanniidae	*Fannia*	*rondanii*	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0	0
Fanniidae	*Fannia*	*sociella*	3	6	3	0	4	6	3	0	3	10	5	0	5	9	4	0
Fanniidae	*Fanniidae uncl.*	*Fanniidae uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Heleomyzidae	*Heleomyzidae uncl.*	*Heleomyzidae uncl.*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Heleomyzidae	*Morpholeria*	*ruficornis*	0	0	4	1	0	0	7	0	0	0	11	0	0	0	7	0
Heleomyzidae	*Suillia*	*bicolor*	2	1	3	14	3	2	7	9	2	0	4	10	5	0	4	10
Heleomyzidae	*Suillia*	*notata*	0	0	3	0	0	0	6	0	1	0	3	0	0	1	8	0
Heleomyzidae	*Suillia*	*ustulata*	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Hybotidae	*Bicellaria*	*austriaca*	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	2
Hybotidae	*Euthyneura*	*myrtilli*	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Hybotidae	*Oedalea*	*flavipes*	0	2	0	0	0	1	0	0	0	3	0	1	0	3	0	0
Hybotidae	*Platypalpus*	*exilis*	0	0	1	0	0	0	1	0	0	0	2	0	0	0	1	0
Hybotidae	*Platypalpus*	*major*	1	2	0	0	0	2	0	0	0	2	0	0	0	2	0	0
Hybotidae	*Tachypeza*	*nubila*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Hybotidae	*Trichina*	*bilobata*	0	0	1	2	0	0	1	1	0	0	2	1	0	0	1	1
Keroplatidae	*Keroplatus*	*testaceus*	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Keroplatidae	*Orfelia*	*fasciata*	1	0	0	5	0	0	0	3	3	0	0	5	2	0	0	3
Keroplatidae	*Orfelia*	*nemoralis*	0	1	2	0	0	0	1	1	0	4	2	2	0	2	1	1
Keroplatidae	*Urytalpa*	*dorsalis*	0	1	0	0	0	2	0	0	0	2	0	1	0	2	0	0
Keroplatidae	*Urytalpa*	*Urytalpa uncl.*	0	2	0	0	0	1	0	0	0	4	1	0	0	2	0	0
Lauxaniidae	*Lyciella*	*platycephala*	0	0	16	14	0	0	8	10	0	0	23	18	0	0	11	8
Lauxaniidae	*Lyciella*	*rorida*	0	0	2	1	0	0	2	0	0	0	3	1	0	0	2	1
Lauxaniidae	*Tricholauxania*	*praeusta*	0	0	0	2	0	0	0	1	0	0	0	3	0	0	0	1
Limoniidae	*Achyrolimonia*	*decemmaculata*	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Limoniidae	*Austrolimnophila*	*ochracea*	0	0	7	0	0	0	3	1	0	0	5	0	0	0	3	0
Limoniidae	*Epiphragma*	*ocellare*	9	7	6	11	3	2	2	4	8	10	5	11	4	3	4	3
Limoniidae	*Molophilus*	*appendiculatus*	0	0	0	2	0	0	0	1	1	0	0	2	0	0	0	1
Limoniidae	*Neolimonia*	*dumetorum*	2	2	0	1	4	5	0	3	6	4	0	2	3	4	0	4
Limoniidae	*Rhipidia*	*maculata*	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Lonchaeidae	*Lonchaea*	*Lonchaea uncl.*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Lonchaeidae	*Lonchaea*	*sp*	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0
Lonchaeidae	*Lonchaeidae uncl.*	*Lonchaeidae uncl.*	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
Lonchaeidae	*Protearomyia*	*nigra*	0	2	0	0	0	2	0	0	0	2	0	0	0	6	0	0
Lonchaeidae	*Protearomyia*	*Protearomyia uncl.*	1	0	0	0	0	0	0	0	3	5	0	0	0	1	0	0
Lonchaeidae	*Protearomyia*	*rameli*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Lonchaeidae	*Protearomyia*	*sp*	1	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0
Micropezidae	*Neria*	*cibaria*	0	0	0	3	0	0	0	2	0	0	0	2	0	0	0	2
Milichiidae	*Phyllomyza*	*equitans*	0	1	0	0	0	1	0	0	0	2	0	0	0	1	0	0
Muscidae	*Coenosia*	*tigrina*	2	0	0	3	0	0	0	3	3	0	0	2	2	0	0	3
Muscidae	*Haematobosca*	*stimulans*	5	0	0	0	2	1	0	0	4	0	0	0	4	0	0	0
Muscidae	*Helina*	*Helina uncl.*	1	1	0	0	1	0	0	0	3	1	0	0	3	0	0	0
Muscidae	*Helina*	*impuncta*	2	2	4	0	6	3	3	0	7	5	2	1	12	9	6	1
Muscidae	*Helina*	*reversio*	1	1	0	0	0	1	0	0	0	3	0	0	0	1	0	0
Muscidae	*Helina*	*trivittata*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Muscidae	*Muscidae uncl.*	*Muscidae uncl.*	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0

Muscidae	*Mydaea*	*corni*	2	0	4	1	1	0	3	0	4	0	5	0	2	0	4	0
Muscidae	*Mydaea*	*Mydaea uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Muscidae	*Mydaea*	*nebulosa*	3	3	0	0	3	3	0	0	5	3	0	0	7	6	0	0
Muscidae	*Phaonia*	*pallida*	0	0	0	3	0	0	0	3	0	0	0	4	0	0	0	3
Muscidae	*Polietes*	*lardarius*	3	0	0	0	2	1	0	0	3	0	0	0	2	0	0	0
Muscidae	*Thricops*	*semicinereus*	0	2	0	0	0	1	0	0	0	3	0	0	0	2	0	0
Mycetophilidae	*Acnemia*	*nitidicollis*	0	1	2	2	1	1	1	1	2	2	2	1	1	1	1	1
Mycetophilidae	*Cordyla*	*semiflava*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Mycetophilidae	*Docosia*	*fuscipes*	0	1	0	0	0	1	0	0	0	3	0	0	0	2	0	0
Mycetophilidae	*Docosia*	*gilvipes*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Mycetophilidae	*Leia*	*crucigera*	0	0	1	3	0	0	0	1	0	0	0	3	0	0	0	2
Mycetophilidae	*Monoclona*	*rufilatera*	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0	1
Mycetophilidae	*Mycetophila*	*caudata*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Mycetophilidae	*Mycetophila*	*fungorum*	2	0	3	3	1	0	2	2	3	0	3	3	3	0	3	3
Mycetophilidae	*Mycetophila*	*ichneumonea*	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	1
Mycetophilidae	*Mycetophila*	*idonea*	0	0	0	1	0	0	0	1	0	0	0	2	0	0	0	1
Mycetophilidae	*Mycetophila*	*Mycetophila uncl.*	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	2
Mycetophilidae	*Mycetophila*	*occultans*	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	1
Mycetophilidae	*Mycetophila*	*perpallida*	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
Mycetophilidae	*Mycetophilidae uncl.*	*Mycetophilidae uncl.*	0	0	1	1	0	0	0	0	0	0	1	1	0	0	1	0
Mycetophilidae	*Mycomya*	*fimbriata*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Mycetophilidae	*Phronia*	*basalis*	0	0	2	3	0	0	0	2	0	0	3	2	0	0	2	2
Mycetophilidae	*Phronia*	*nigricornis*	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0
Mycetophilidae	*Phronia*	*Phronia uncl.*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Mycetophilidae	*Platurocypta*	*testata*	0	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0
Mycetophilidae	*Synapha*	*vitripennis*	0	1	0	0	0	1	0	0	0	2	0	0	0	1	0	0
Mycetophilidae	*Tetragoneura*	*sylvatica*	1	2	4	1	0	3	2	2	2	7	6	2	2	3	2	2
Mycetophilidae	*Zygomyia*	*pseudohumeralis*	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0
Pallopteridae	*Toxoneura*	*quinquemaculata*	0	2	0	0	0	1	0	0	0	3	0	0	0	2	0	0
Phaeomyiidae	*Pelidnoptera*	*fuscipennis*	3	5	0	0	2	2	0	0	2	3	0	0	2	2	0	1
Phoridae	*Anevrina*	*thoracica*	0	0	0	2	0	0	0	1	0	0	0	3	0	0	0	1
Phoridae	*Conicera*	*tibialis*	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Phoridae	*Diplonevra*	*nitidula*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Phoridae	*Gymnophora*	*arcuata*	1	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Phoridae	*Megaselia*	*campestris*	3	0	2	1	1	0	3	0	2	0	2	2	3	0	4	2
Phoridae	*Megaselia*	*ciliata*	0	0	3	1	0	0	1	0	0	0	3	0	0	0	2	0
Phoridae	*Megaselia*	*Megaselia uncl.*	1	0	0	0	0	2	0	0	1	1	0	0	1	2	0	0
Phoridae	*Megaselia*	*nigriceps*	1	0	0	1	0	0	0	1	1	2	0	3	1	1	0	1
Phoridae	*Megaselia*	*pleuralis*	0	0	1	2	0	0	1	1	0	3	2	2	0	2	2	1
Phoridae	*Megaselia*	*ruficornis*	1	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0
Phoridae	*Megaselia*	*variana*	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	2
Phoridae	*Phora*	*atra*	0	2	0	0	0	1	0	0	0	3	0	0	0	1	0	1
Phoridae	*Phoridae uncl.*	*Phoridae uncl.*	2	1	0	0	0	0	0	0	2	2	0	0	1	1	0	0
Pipunculidae	*Cephalosphaera*	*germanica*	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
Pipunculidae	*Eudorylas*	*subfascipes*	2	0	0	0	1	0	0	0	2	1	0	1	1	0	0	0
Pipunculidae	*Jassidophaga*	*beatricis*	2	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0

Pipunculidae	*Jassidophaga*	*Jassidophaga uncl.*	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Pipunculidae	*Jassidophaga*	*sp*	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Pipunculidae	*Nephrocerus*	*flavicornis*	1	0	0	0	2	0	0	0	2	0	0	0	3	0	0	0
Psychodidae	*Clytocerus*	*ocellaris*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Psychodidae	*Parajungiella*	*longicornis*	0	0	0	0	0	0	0	0	2	2	0	2	1	1	0	1
Psychodidae	*Pneumia*	*nubila*	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	2
Psychodidae	*Psychoda*	*albipennis*	0	0	0	1	0	0	0	1	0	2	0	2	0	2	0	2
Psychodidae	*Psychoda*	*gemina*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Psychodidae	*Psychoda*	*lobata*	0	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0
Psychodidae	*Psychoda*	*phalaenoides*	2	1	1	3	0	2	1	2	4	3	2	4	3	2	2	2
Psychodidae	*Psychoda*	*sp*	0	0	0	0	0	0	0	0	2	2	0	0	1	1	0	0
Psychodidae	*Psychodidae*	*sp1*	0	1	2	1	0	0	0	0	0	2	2	2	0	2	2	1
Psychodidae	*Psychodidae uncl.*	*Psychodidae uncl.*	0	0	1	0	0	0	1	0	0	0	1	2	0	0	1	2
Psychodidae	*Trichopsychoda*	*hirtella*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Rhagionidae	*Ptiolina*	*obscura*	0	0	3	3	0	0	2	2	0	0	4	2	0	0	2	2
Rhagionidae	*Rhagio*	*annulatus*	10	4	5	5	8	6	8	5	14	13	11	7	16	9	9	7
Rhagionidae	*Rhagio*	*Rhagio uncl.*	4	7	24	11	3	3	14	9	11	14	34	22	7	8	19	14
Rhagionidae	*Rhagio*	*scolopaceus*	8	9	43	24	5	10	26	23	15	17	59	38	14	12	36	27
Rhagionidae	*Rhagio*	*strigosus*	1	1	7	5	0	0	6	4	3	3	8	6	1	2	8	4
Rhagionidae	*Rhagionidae uncl.*	*Rhagionidae uncl.*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Sarcophagidae	*Sarcophaga*	*carnaria*	8	3	3	12	8	0	0	12	12	0	0	17	12	0	0	12
Sarcophagidae	*Sarcophaga*	*haemorrhoa*	0	2	0	0	0	2	0	0	0	2	0	0	0	2	0	0
Sarcophagidae	*Sarcophaga*	*Sarcophaga uncl.*	1	1	0	1	0	0	0	1	4	1	0	3	4	0	0	1
Sarcophagidae	*Sarcophaga*	*subvicina*	7	1	0	1	2	0	0	1	8	0	0	2	6	0	0	1
Sarcophagidae	*Sarcophaga*	*variegata*	3	1	0	0	2	0	0	0	6	0	0	1	5	0	0	1
Sarcophagidae	*Sarcophagidae uncl.*	*Sarcophagidae uncl.*	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0
Scathophagidae	*Scathophaga*	*furcata*	0	2	0	0	1	2	0	0	0	2	0	0	1	2	0	0
Scathophagidae	*Scathophaga*	*stercoraria*	1	2	0	0	0	3	0	0	1	4	0	0	1	3	0	0
Sciaridae	*Bradysia*	*Bradysia uncl.*	0	0	1	2	0	0	0	2	0	0	0	3	0	0	0	3
Sciaridae	*Bradysia*	*cinerascens*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sciaridae	*Bradysia*	*fungicola*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Sciaridae	*Bradysia*	*insignis*	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Sciaridae	*Bradysia*	*longicubitalis*	0	0	0	3	0	1	0	2	0	2	0	2	0	2	0	2
Sciaridae	*Bradysia*	*pectoralis*	0	0	2	2	0	0	1	2	0	0	0	2	0	0	1	1
Sciaridae	*Bradysia*	*placida*	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1
Sciaridae	*Bradysia*	*vagans*	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Sciaridae	*Camptochaeta*	*vivax*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Sciaridae	*Claustropyga*	*abblanda*	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Sciaridae	*Corynoptera*	*blanda*	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Sciaridae	*Corynoptera*	*boletiphaga*	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0
Sciaridae	*Corynoptera*	*cuniculata*	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Sciaridae	*Corynoptera*	*dentata*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Sciaridae	*Corynoptera*	*forcipata*	2	1	4	2	1	1	1	1	5	3	7	5	2	1	3	2
Sciaridae	*Corynoptera*	*furcifera*	0	0	1	3	0	0	1	1	0	0	3	2	0	0	3	1
Sciaridae	*Corynoptera*	*tetrachaeta*	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Sciaridae	*Cratyna*	*harpata*	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0

Sciaridae	*Cratyna*	*perplexa*	0	0	1	3	0	0	1	4	0	0	5	9	0	0	4	5
Sciaridae	*Cratyna*	*uliginosoides*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Sciaridae	*Epidapus*	*alnicola*	0	0	1	2	0	0	0	1	0	0	3	3	0	0	1	1
Sciaridae	*Epidapus*	*atomarius*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Sciaridae	*Epidapus*	*Epidapus uncl.*	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0
Sciaridae	*Epidapus*	*ignotus*	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1
Sciaridae	*Epidapus*	*sp*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Sciaridae	*Leptosciarella*	*Leptosciarella uncl.*	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1
Sciaridae	*Leptosciarella*	*scutellata*	0	0	1	0	0	0	1	0	0	0	3	1	0	0	2	1
Sciaridae	*Leptosciarella*	*subpilosa*	0	0	0	2	0	0	0	2	2	0	0	3	0	0	0	3
Sciaridae	*Phytosciara*	*flavipes*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Sciaridae	*Pnyxia*	*sp*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sciaridae	*Scatopsciara*	*atomaria*	0	0	0	1	0	0	0	1	0	1	0	2	0	1	0	1
Sciaridae	*Scatopsciara*	*edwardsi*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Sciaridae	*Scatopsciara*	*vitripennis*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Sciaridae	*Sciara*	*hebes*	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
Sciaridae	*Sciara*	*lackschewitzi*	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2
Sciaridae	*Sciaridae uncl.*	*Sciaridae uncl.*	0	0	0	2	0	0	0	3	0	2	1	2	0	1	0	4
Sciaridae	*Scythropochroa*	*radialis*	0	0	0	0	0	0	0	0	0	0	3	3	0	0	3	2
Sciaridae	*Trichosia*	*acrotricha*	0	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0
Sciaridae	*Trichosia*	*edwardsi*	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Sciaridae	*Trichosia*	*iota*	0	1	0	0	0	1	0	0	0	2	0	0	0	3	0	0
Sciaridae	*Trichosia*	*lengersdorfi*	0	0	0	1	0	0	0	0	2	0	0	2	1	0	0	1
Sciaridae	*Trichosia*	*splendens*	1	0	2	0	1	0	1	0	2	0	3	0	1	0	1	0
Sphaeroceridae	*Copromyza*	*stercoraria*	1	0	1	2	1	0	0	0	2	1	1	1	2	0	0	2
Sphaeroceridae	*Crumomyia*	*fimetaria*	0	1	0	0	0	2	0	0	0	1	0	0	0	2	0	0
Sphaeroceridae	*Crumomyia*	*nitida*	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	1
Sphaeroceridae	*Gigalimosina*	*flaviceps*	0	0	0	1	0	0	0	1	0	0	0	2	0	0	0	2
Sphaeroceridae	*Spelobia*	*palmata*	0	0	2	5	0	0	1	1	3	1	5	6	2	1	3	2
Sphaeroceridae	*Spelobia*	*parapusio*	1	1	1	0	1	2	1	0	2	2	1	0	2	2	1	0
Stratiomyidae	*Clitellaria*	*ephippium*	11	1	0	0	8	2	0	0	13	0	0	0	11	0	0	0
Syrphidae	*Blera*	*fallax*	4	1	0	0	3	0	0	0	9	0	0	0	4	0	0	0
Syrphidae	*Chrysotoxum*	*cautum*	2	0	0	0	2	2	0	0	5	1	0	0	3	0	0	0
Syrphidae	*Chrysotoxum*	*Chrysotoxum uncl.*	9	3	0	0	6	0	0	0	17	0	0	0	11	0	0	1
Syrphidae	*Chrysotoxum*	*festivum*	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Syrphidae	*Criorhina*	*asilica*	4	2	0	0	4	1	0	0	7	0	0	0	6	0	0	0
Syrphidae	*Dasysyrphus*	*tricinctus*	4	0	0	0	2	1	0	0	4	0	0	1	2	1	0	0
Syrphidae	*Didea*	*fasciata*	3	0	1	0	2	1	0	0	4	0	0	0	2	0	0	0
Syrphidae	*Didea*	*intermedia*	3	0	0	0	2	0	0	0	3	0	0	0	2	0	0	0
Syrphidae	*Epistrophe*	*nitidicollis*	9	8	0	0	2	3	0	0	11	11	2	0	6	4	0	0
Syrphidae	*Melanostoma*	*Melanostoma uncl.*	2	1	0	0	1	0	0	0	2	0	1	0	1	0	0	0
Syrphidae	*Melanostoma*	*mellinum*	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
Syrphidae	*Pipiza*	*Pipiza uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Syrphidae	*Pipiza*	*sp*	2	0	0	0	1	0	0	0	3	0	0	0	2	0	0	0
Syrphidae	*Rhingia*	*campestris*	4	4	0	0	2	3	0	1	9	7	0	0	2	5	0	1
Syrphidae	*Rhingia*	*Rhingia uncl.*	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Syrphidae	*Syrphidae uncl.*	*Syrphidae uncl.*	3	4	14	0	2	4	6	0	7	6	24	1	4	5	12	0
Syrphidae	*Syrphus*	*ribesii*	5	1	0	0	2	0	0	0	5	0	0	0	3	0	0	0
Syrphidae	*Temnostoma*	*bombylans*	8	16	30	3	6	12	14	2	17	27	32	0	15	20	21	0
Syrphidae	*Temnostoma*	*Temnostoma uncl.*	1	2	7	0	1	2	8	0	5	7	10	0	3	4	11	0
Syrphidae	*Temnostoma*	*vespiforme*	0	0	20	4	0	1	12	3	1	1	30	0	0	2	17	0
Tabanidae	*Hybomitra*	*bimaculata*	7	1	1	0	3	2	0	1	12	0	2	0	5	0	0	1
Tabanidae	*Hybomitra*	*Hybomitra uncl.*	0	0	1	6	1	0	1	6	2	0	0	10	1	0	0	6
Tabanidae	*Hybomitra*	*lundbecki*	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Tabanidae	*Tabanidae uncl.*	*Tabanidae uncl.*	0	0	1	15	0	0	2	21	2	0	0	14	0	0	0	22
Tachinidae	*Bessa*	*selecta*	0	0	0	0	1	0	0	0	2	0	0	0	2	0	0	0
Tachinidae	*Gastrolepta*	*anthracina*	0	1	0	0	0	1	0	0	0	3	0	0	0	2	0	0
Tachinidae	*Peribaea*	*fissicornis*	1	0	0	0	1	2	0	0	3	3	0	0	2	3	0	0
Tachinidae	*Phorocera*	*obscura*	0	4	0	0	0	3	0	0	0	3	1	0	0	3	0	0
Tachinidae	*Phryno*	*vetula*	2	0	0	0	3	0	0	0	4	0	0	0	3	0	0	0
Tachinidae	*Tachina*	*magnicornis*	1	10	9	2	2	7	11	2	0	15	15	0	0	12	14	0
Tachinidae	*Tachina*	*Tachina uncl.*	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Tachinidae	*Tachinidae uncl.*	*Tachinidae uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	2	1	1	0
Tachinidae	*Triarthria*	*setipennis*	0	3	0	0	0	2	0	0	0	4	1	0	0	2	0	0
Tipulidae	*Nephrotoma*	*analis*	2	0	0	0	6	0	0	0	4	0	0	0	7	0	0	0
Tipulidae	*Nephrotoma*	*Nephrotoma uncl.*	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Tipulidae	*Nephrotoma*	*submaculosa*	0	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0
Tipulidae	*Tanyptera*	*atrata*	12	8	12	16	6	4	9	10	16	15	17	16	19	16	17	18
Tipulidae	*Tanyptera*	*Tanyptera uncl.*	0	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1
Tipulidae	*Tipula*	*nubeculosa*	0	0	1	6	0	0	1	5	0	0	0	6	0	0	0	4
Tipulidae	*Tipula*	*pseudovariipennis*	2	4	0	0	0	2	0	0	0	7	0	0	0	4	0	0
Tipulidae	*Tipula*	*Tipula uncl.*	0	1	0	1	0	1	0	4	0	3	0	1	0	5	0	4
Tipulidae	*Tipulidae uncl.*	*Tipulidae uncl.*	2	10	1	2	1	6	1	3	1	11	1	3	2	11	1	3
Hemiptera																		
Acanthosomatidae	*Acanthosoma*	*haemorrhoidale*	0	0	3	1	0	0	3	0	0	0	5	0	0	0	4	0
Aphididae	*Aphididae uncl.*	*Aphididae uncl.*	0	1	0	0	0	0	0	0	2	5	2	3	0	3	1	1
Aphididae	*Aphis*	*sambuci*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Aphididae	*Eriosoma*	*anncharlotteae*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Aphididae	*Euceraphis*	*betulae*	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
Aphididae	*Eulachnus*	*agilis*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Aphididae	*Sitobion*	*Sitobion uncl.*	0	0	0	0	1	0	0	0	1	1	0	0	1	2	0	0
Cicadellidae	*Balclutha*	*punctata*	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0
Cicadellidae	*Cicadellidae uncl.*	*Cicadellidae uncl.*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Cicadellidae	*Edwardsiana*	*flavescens*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Cicadellidae	*Empoasca*	*vitis*	1	1	0	0	0	0	0	0	1	1	0	0	1	1	0	0
Cicadellidae	*Eupteryx*	*atropunctata*	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	1
Cicadellidae	*Eupteryx*	*aurata*	1	1	0	1	1	1	1	1	1	1	0	2	1	1	0	1
Cicadellidae	*Eupteryx*	*Eupteryx uncl.*	0	0	2	2	0	0	2	1	0	0	2	2	0	0	1	1
Cicadellidae	*Eupteryx*	*urticae*	0	0	0	3	0	0	0	0	0	0	0	3	0	0	0	0
Cicadellidae	*Eupteryx*	*urticae*	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	3
Cicadellidae	*Fagocyba*	*cruenta*	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0
Cicadellidae	*Fagocyba*	*Fagocyba uncl.*	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0

Cicadellidae	*Fagocyba*	*sp*	0	0	5	5	0	0	3	2	0	0	5	4	0	0	3	3
Cicadellidae	*Macrosteles*	*laevis*	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	2
Cicadellidae	*Speudotettix*	*subfusculus*	2	2	0	0	1	3	0	0	2	2	0	0	3	3	0	0
Cicadellidae	*Zygina*	*Zygina uncl.*	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Cixiidae	*Cixiidae uncl.*	*Cixiidae uncl.*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Cixiidae	*Tachycixius*	*pilosus*	1	2	0	2	1	2	0	2	2	3	0	3	3	4	0	2
Delphacidae	*Javesella*	*Javesella uncl.*	2	2	0	0	1	2	0	0	2	3	0	0	1	1	0	0
Hemiptera uncl.	*Hemiptera uncl.*	*Hemiptera uncl.*	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	1
Lygaeidae	*Kleidocerys*	*resedae*	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
Miridae	*Psallus*	*varians*	0	0	2	2	0	0	3	1	0	0	2	2	0	0	2	2
Pentatomidae	*Palomena*	*prasina*	0	5	0	0	1	3	0	0	0	6	0	0	0	3	0	0
Psyllidae	*Cacopsylla*	*pulchra*	0	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Hymenoptera																		
Andrenidae	*Andrena*	*Andrena uncl.*	3	4	1	1	2	3	0	3	3	3	0	1	5	5	0	3
Andrenidae	*Andrena*	*minutula*	4	4	0	2	4	4	0	0	11	10	0	0	9	7	0	0
Apidae	*Apidae uncl.*	*Apidae uncl.*	3	1	0	0	1	0	0	0	4	3	0	0	1	0	0	0
Apidae	*Apis*	*mellifera*	0	1	0	0	0	1	0	0	0	7	0	0	0	6	0	0
Apidae	*Bombus*	*Bombus uncl.*	3	2	0	0	3	0	0	0	10	6	0	0	14	12	1	0
Apidae	*Bombus*	*hypnolum*	2	0	0	0	1	0	0	0	4	0	0	0	3	0	0	0
Apidae	*Bombus*	*terrestris*	1	0	0	0	0	0	0	0	3	3	0	0	3	3	0	0
Braconidae	*Braconidae uncl.*	*Braconidae uncl.*	0	0	0	0	0	0	0	0	3	8	3	1	3	7	1	1
Braconidae	*Braconidae-1803037*	*sp*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Braconidae	*Braconidae-1803038*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Braconidae	*Braconidae-223959458*	*sp*	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Braconidae	*Braconidae-223961430*	*sp*	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Braconidae	*Braconidae-223970726*	*sp*	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Braconidae	*Braconidae-223970727*	*sp*	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Diapriidae	*Diapriidae uncl.*	*Diapriidae uncl.*	0	0	1	0	0	0	0	0	1	0	2	3	0	1	2	1
Eurytomidae	*Eurytomidae uncl.*	*Eurytomidae uncl.*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Formicidae	*Formica*	*Formica uncl.*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Formicidae	*Formica*	*fusca*	0	1	0	0	0	0	0	0	0	2	1	0	0	1	0	0
Formicidae	*Formica*	*lemani*	0	0	3	0	0	0	1	0	0	0	4	0	0	0	2	0
Formicidae	*Formicidae uncl.*	*Formicidae uncl.*	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
Formicidae	*Lasius*	*fuliginosus*	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Formicidae	*Myrmica*	*ruginodis*	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
Hymenoptera uncl.	*Hymenoptera uncl.*	*Hymenoptera uncl.*	7	11	4	7	5	6	3	3	27	43	32	32	24	31	17	22
Ichneumonidae	*Aperileptus*	*albipalpus*	0	0	0	1	0	0	0	1	0	2	0	2	0	1	0	1
Ichneumonidae	*Cryptus*	*armator*	1	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0
Ichneumonidae	*Ctenochira*	*marginata*	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0
Ichneumonidae	*Diadegma*	*consumptor*	0	1	1	1	0	0	1	1	0	2	2	2	0	1	1	1
Ichneumonidae	*Hemichneumon*	*subdolus*	0	1	0	2	1	0	0	1	1	0	0	2	1	0	0	1
Ichneumonidae	*Ichneumon*	*emancipatus*	0	1	2	0	0	1	1	0	3	2	3	0	3	1	1	0
Ichneumonidae	*Ichneumonidae uncl.*	*Ichneumonidae uncl.*	46	72	49	83	37	64	31	56	191	264	128	179	150	189	87	126
Ichneumonidae	*Ichneumonidae-1800006*	*sp*	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1800008*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ichneumonidae	*Ichneumonidae-1800068*	*sp*	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0

Ichneumonidae	*Ichneumonidae-1800223*	*sp*	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-1800225*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ichneumonidae	*Ichneumonidae-1800455*	*sp*	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1800457*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0
Ichneumonidae	*Ichneumonidae-1801011*	*sp*	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1801013*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ichneumonidae	*Ichneumonidae-1801132*	*sp*	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1801134*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ichneumonidae	*Ichneumonidae-1801270*	*sp*	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-1801272*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ichneumonidae	*Ichneumonidae-1801275*	*sp*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1801277*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Ichneumonidae	*Ichneumonidae-1803022*	*sp*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Ichneumonidae	*Ichneumonidae-1803024*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Ichneumonidae	*Ichneumonidae-1803422*	*sp*	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-1803424*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ichneumonidae	*Ichneumonidae-223954931*	*sp*	0	0	2	0	0	0	1	0	0	0	3	0	0	1	3	0
Ichneumonidae	*Ichneumonidae-223955556*	*sp*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-223957367*	*sp*	0	0	0	2	0	0	0	0	0	0	0	3	0	0	0	0
Ichneumonidae	*Ichneumonidae-223957369*	*sp*	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	4
Ichneumonidae	*Ichneumonidae-223961333*	*sp*	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-223961335*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Ichneumonidae	*Ichneumonidae-223962085*	*sp*	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-223962087*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Ichneumonidae	*Ichneumonidae-223966007*	*sp*	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-223966009*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Ichneumonidae	*Ichneumonidae-223967673*	*sp*	0	0	1	1	0	0	0	0	0	0	2	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-223967675*	*sp*	0	0	0	0	0	0	2	1	0	0	0	0	0	0	2	1
Ichneumonidae	*Ichneumonidae-223968660*	*sp*	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-223968662*	*sp*	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2
Ichneumonidae	*Ichneumonidae-223968672*	*sp*	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
Ichneumonidae	*Ichneumonidae-223968673*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Ichneumonidae	*Ichneumonidae-223968687*	*sp*	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-223968689*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Ichneumonidae	*Ichneumonidae-223969983*	*sp*	0	0	1	1	0	0	0	0	0	2	2	3	0	0	0	0
Ichneumonidae	*Ichneumonidae-223970024*	*sp*	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0
Ichneumonidae	*Ichneumonidae-223970645*	*sp*	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-223970647*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
Ichneumonidae	*Ichneumonidae-223970926*	*sp*	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0
Ichneumonidae	*Ichneumonidae-223970928*	*sp*	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Ichneumonidae	*Ichneumonidae-223971190*	*sp*	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
Ichneumonidae	*Ichneumonidae-223972532*	*sp*	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-224010873*	*sp*	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Ichneumonidae	*Ichneumonidae-224010875*	*sp*	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Ichneumonidae	*Olesicampe*	*tarsator*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Ichneumonidae	*Ophion*	*longigena*	0	0	0	1	0	0	0	1	0	0	0	3	0	0	0	2

Ichneumonidae	*Orthizema*	*gravipes*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Ichneumonidae	*Polytribax*	*arrogans*	1	1	0	0	1	1	0	0	4	4	0	0	1	1	0	0
Ichneumonidae	*Xylophrurus*	*augustus*	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Tenthredinidae	*Aneugmenus*	*coronatus*	0	2	0	0	0	2	0	0	0	3	0	0	0	3	0	0
Tenthredinidae	*Claremontia*	*waldheimii*	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Insecta uncl.																		
Insecta uncl.	*Insecta uncl.*	*Insecta uncl.*	244	345	270	464	143	205	177	281	526	725	448	608	351	477	302	396
Lepidoptera																		
Adelidae	*Nematopogon*	*swammerdamella*	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Adelidae	*Nemophora*	*degeerella*	1	0	0	2	1	0	0	1	5	0	0	2	4	0	0	3
Hepialidae	*Phymatopus*	*hecta*	0	0	13	1	0	0	12	0	0	1	19	0	0	0	13	0
Lepidoptera uncl.	*Lepidoptera uncl.*	*Lepidoptera uncl.*	3	3	0	0	1	2	0	2	4	3	1	2	1	2	0	3
Noctuidae	*Noctuidae uncl.*	*Noctuidae uncl.*	2	1	0	0	3	2	0	1	5	1	0	1	5	7	0	1
Mecoptera																		
Panorpidae	*Panorpa*	*communis*	3	3	0	0	2	2	0	0	4	4	0	0	2	2	1	0
Panorpidae	*Panorpa*	*Panorpa uncl.*	0	0	0	0	0	0	0	0	0	3	0	0	1	1	0	0
Chrysopidae	*Chrysopa*	*perla*	3	0	0	0	2	0	0	0	3	0	0	0	2	0	0	0
Neuroptera																		
Hemerobiidae	*Hemerobius*	*humulinus*	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
Odonata																		
Coenagrionidae	*Coenagrion*	*pulchellum*	0	0	1	6	0	0	1	7	0	0	0	4	0	0	0	5
Coenagrionidae	*Valenzuela*	*flavidus*	0	0	0	0	0	0	2	2	0	0	0	0	0	0	2	2
Psocoptera																		
Caeciliusidae	*Valenzuela*	*flavidus*	0	0	3	4	0	0	0	0	0	0	6	5	0	0	0	0
Peripsocidae	*Peripsocus*	*parvulus*	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1	0
Peripsocidae	*Peripsocus*	*phaeopterus*	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	1
Psocidae	*Loensia*	*fasciata*	2	2	0	0	1	2	0	0	2	3	0	0	2	2	0	0
Trichoptera																		
Leptoceridae	*Limnephilus*	*Limnephilus uncl.*	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Limnephilidae	*Glyphotaelius*	*pellucidus*	0	0	0	5	0	0	1	3	0	0	0	7	0	0	0	3
Limnephilidae	*Grammotaulius*	*Grammotaulius uncl.*	0	0	0	2	0	0	0	2	0	0	0	3	0	0	0	2
Limnephilidae	*Grammotaulius*	*nigropunctatus*	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
Limnephilidae	*Limnephilidae uncl.*	*Limnephilidae uncl.*	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
Polycentropodidae	*Holocentropus*	*Holocentropus uncl.*	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Trichoptera uncl.	*Trichoptera uncl.*	*Trichoptera uncl.*	0	0	0	3	0	0	1	2	0	0	0	3	0	0	0	2
UNKNOWN																		
unknown uncl																		
unknown uncl.	*unknown uncl.*	*unknown uncl.*	20	20	16	35	13	13	20	32	33	39	25	50	14	18	15	31

9 APPENDIX

1. Zusammenfassung
2. Erklärung

Zusammenfassung

Der Verlust an Biodiversität, dem die Welt gegenwärtig gegenübersteht ist beträchtlich. Diese desaströse Entwicklung zeichnet sich nicht nur durch einen Rückgang lokaler Diversität aus, sondern auch durch einen massiven Rückgang an Biomasse. Habitatverlust, Klimawandel, invasive Arten, extensive Landnutzung und -ausbeutung treiben diese Veränderungen voran. Die Inventarisierung und Differenzierung von Arten bildet die Grundlage vieler notwendiger ökologischer Studien zum Verständnis dieser Vorgänge. Dennoch scheinen Datensätze in ihrer räumlichen und zeitlichen Abdeckung nicht ausreichend zu sein und wiederholt werden neue Anstrengungen unternommen die Ansätze des Monitoring zu verbessern. Die erforderliche räumliche und zeitliche Abdeckung und die Forderung nach regelmäßigen Aktualisierungen der Daten weisen dabei stark auf die Notwendigkeit der Automatisierung der Prozesse hin. Dies schließt auch die Identifizierung der in Proben gefundenen Arten mit ein. Aber bereits das derzeitige Aufkommen an Proben überfordert die traditionellen Methoden zur Identifizierung sowohl zeitlich, als auch finanziell. Grundsätzlich benötigt ein morphologischer Ansatz oft zahlreiche Experten für jede taxonomische Ordnung in einer Probe. Im Gegensatz dazu bietet ein genetisch basierter Ansatz, wie das DNA-Barcoding, die Möglichkeit das bisher auf Taxonomen beschränkte Expertenwissen universell anwendbar zu machen. DNA-Barcoding bietet dabei die Möglichkeit auch noch unbekannte Arten in die Auswertung mit aufzunehmen. Publikationen haben bereits gezeigt, wie Sequenzdaten dafür gehandhabt werden können; indem nämlich diese als operative taxonomische Einheiten (OTUs) oder molekulare taxonomische Einheiten (MOTUs) definiert werden, bis Informationen zur Art verfügbar sind. Die Entwicklung eines Barcode Indexing Number (BIN) -Systems in der Barcode of Life Database (BOLD) unterstützt diesen Prozess und passt sich an das an, was man als „postlinnaeisches“ taxonomisches System bezeichnen könnte. Diese MOTUs, OTUs oder BINs eignen sich auch dazu bei ökologischen Fragestellungen eingesetzt zu werden. Und da ein großer Teil der globalen Biodiversität noch immer unbekannt ist, ist die Bewertung und Dokumentation auch dieses Teils ein unschätzbarer Beitrag.

Vor diesem Hintergrund war eine grundlegende Frage die Eignung von DNA-Barcoding, mit Cytochrom-*c*-Oxidase (COI) als Marker, für ein umfangreiches Biodiversitäts-Monitoring und die Realisierbarkeit eines vorerst Diptera-fokussierten Workflows für eine schnelle Bewertung von Biodiversität. Dies beinhaltete die Sammlung von DNA-Daten bereits identifizierter Arten zur Bereitstellung von Art-Referenzen, die Prüfung der Effektivität des Markers, die Bereitstellung und Auswertung geeigneter gemischter Proben, die Arbeit mit MOTUs für eine Auswertung unabhängig von beschriebenen Arten und die Entwicklung einer auf diese Aufgabe ausgerichteten automatisierten Fangvorrichtung, für einen umfassenden Arbeitsablauf, beginnend bei der Datensammlung und Aufbereitung und endend bei der Auswertung der Sequenzen. Unter Berücksichtigung von Kosteneffizienz bei hohem Durchsatz und gleichzeitiger Gewinnung qualitativ hochwertiger Daten, wurde unter anderem die Methode des Hybrid Enrichment, so wie Methoden der parallelisierenden Sequenzierung verwendet und bewertet. Ein Einblick in die identifizierten Arten, so wie die Ökologie und den Unterschied der Diversitäten zweier auf diesem Wege experimentell beprobter

Waldtypen, fungiert als exemplarische Demonstration des Workflows und bildet den Abschluss der Arbeit.

Die Auswertung der Ergebnisse zeigte, dass COI als „Barcode“ durchaus geeignet ist Dipterenarten zu differenzieren und korrekt zu identifizieren; und, dass die Implementierung von DNA-Sequenzierung für Diversitätserfassungen durchaus lohnend ist, insbesondere in Verbindung mit automatisierten Arbeitsabläufen. Durch die Kombination neuer Sequenziertechnologien mit Automatisierungen bei der Beprobung wird eine hohe Proben- und Datenqualität bei einem Minimum an Beprobungsaufwand und Habitatstörung erreicht. Viele der Artenfunde, basierend auf Sequenzdaten, wurden durch vorherige morphologische Bestimmungen eines Teils der Probe validiert. Die Sequenzdaten waren jedoch, wie erwartet, weitaus differenzierter. Der große Anteil von weiblichen, morphologisch nur schwer bestimmbaren, Individuen in den Proben ist dabei nur ein Grund. Bei gemischten Proben wurden die negativen Auswirkungen bei der Verwendung von PCR untersucht und bestätigt. Primer und Amplifizierung wirken selektiv und erzeugen Artefakte, je mehr Arten sich in einer Probe befinden. Eine gute Alternative stellte das Hybrid Enrichment dar. Der Einsatz des entwickelten „Automatisierten Intervall-Samplers“ (AIS) in der Studie ergab qualitativ hochwertige Proben, die für DNA-Analysen sehr gut geeignet sind. Wichtig für die Qualität einer Probe zeigte sich die Verwendung von hochprozentigem Ethanol zur Konservierung der DNA, ab dem Zeitpunkt des Fangs. Die Konstruktion konnte im Einsatz das Verdampfen des Ethanols in den Proben wirksam reduzieren. Energieengpässe für die Steuerung ergaben sich jedoch aus der Abschattung der Kollektoren bei fortgeschrittener Vegetationsperiode durch Blattwerk und Unterwuchs. In seiner neuesten Version wurde das AIS daher mit Solarpanelen ausgestattet, die speziell für diffuse Lichtverhältnisse geeignet sind. Ein Akku mit höherer Kapazität, sowie eine Softwareanpassung stellen sicher, dass Stromausfälle in Zukunft vermieden werden. Zahlreiche funktionale Erweiterungen und die mögliche Kombination mit anderen Fallentypen können seine Anwendung in Zukunft erweitern. Ausgestattet mit Sensoren kann das AIS Proben zusammen mit detaillierten Umweltinformationen liefern, welche helfen könnten eine Probe bereits im Vorfeld auf ihren ökologischen und genetischen Wert hin einzuschätzen. Bei der Auswertung der Proben zeigte sich die Möglichkeit zur Verwendung von relativ kurzen (~ 330bp) Sequenzstücken für eine Artidentifizierung und Einschätzungen der Diversität. Einzelne Dipterenarten bewiesen ihre wertvolle Funktion als ökologische Indikatoren, da sie durch ihre Kombination von Sessilität und Mobilität nicht nur Informationen über die untersuchten Waldflächen selbst lieferten, sondern auch über die daran angrenzenden Gebiete. Auch gab es einen Hinweis auf die historische Kontinuität der Waldflächen. Obwohl festgestellt wurde, dass ein einzelnes bewirtschaftetes Waldstück hohe Diversität aufweisen kann, kann trotzdem angenommen werden, dass die maximale Diversität durch unbewirtschaftete Waldflächen erreicht wird, die im Idealfall noch nie vollständig abgeholzt wurden. Der Verlust der Kontinuität scheint langfristige Auswirkungen auf die Arten-zusammensetzung in Waldflächen zu haben und könnte irreversibel sein, wenn lokale Arten dabei aussterben. Es kann abschließend davon ausgegangen werden, dass großflächig angelegte Langzeitstudien von den Entwicklungen im Bereich Automatisierung und Digitalisierung profitieren werden, da diese die differenzierte und dabei zeit- und kosteneffiziente Auswertung großer Datenmengen ermöglichen.

Erklärung

Ich versichere, dass ich diese Arbeit selbständig verfasst, keine anderen Quellen und Hilfsmittel als die angegebenen benutzt und die Stellen der Arbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen sind, kenntlich gemacht habe. Diese Arbeit hat in dieser oder ähnlichen Form keiner anderen Prüfungsbehörde vorgelegen.

Bonn, ____ . ____ . _____

_____