



RESOURCE ARTICLE

A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring

Jérôme Morinière¹  | Michael Balke¹ | Dieter Doczkal¹ | Matthias F. Geiger² | Laura A. Hardulak¹ | Gerhard Haszprunar¹ | Axel Hausmann¹ | Lars Hendrich¹ | Ledis Regalado¹ | Björn Rulik² | Stefan Schmidt¹  | Johann-Wolfgang Wägele² | Paul D. N. Hebert³

¹SNSB-Zoologische Staatssammlung, München, Germany

²Zoological Research Museum Alexander Koenig - Leibniz Institute for Animal Biodiversity, Bonn, Germany

³Centre for Biodiversity Genomics, University of Guelph, Guelph, Ontario, Canada

Correspondence

Jérôme Morinière and Dieter Doczkal, SNSB-Zoologische Staatssammlung, München, Germany.
Emails: moriniere@snsb.de; doczkal@snsb.de

Funding information

Bavarian State Ministry of Science and the Arts; German Federal Ministry of Education and Research, Grant/Award Number: BMBF FKZ 01L1101 and BMBF FKZ 01L11501

Abstract

This study summarizes results of a DNA barcoding campaign on German Diptera, involving analysis of 45,040 specimens. The resultant DNA barcode library includes records for 2,453 named species comprising a total of 5,200 barcode index numbers (BINs), including 2,700 COI haplotype clusters without species-level assignment, so called “dark taxa.” Overall, 88 out of 117 families (75%) recorded from Germany were covered, representing more than 50% of the 9,544 known species of German Diptera. Until now, most of these families, especially the most diverse, have been taxonomically inaccessible. By contrast, within a few years this study provided an intermediate taxonomic system for half of the German Dipteran fauna, which will provide a useful foundation for subsequent detailed, integrative taxonomic studies. Using DNA extracts derived from bulk collections made by Malaise traps, we further demonstrate that species delineation using BINs and operational taxonomic units (OTUs) constitutes an effective method for biodiversity studies using DNA metabarcoding. As the reference libraries continue to grow, and gaps in the species catalogue are filled, BIN lists assembled by metabarcoding will provide greater taxonomic resolution. The present study has three main goals: (a) to provide a DNA barcode library for 5,200 BINs of Diptera; (b) to demonstrate, based on the example of bulk extractions from a Malaise trap experiment, that DNA barcode clusters, labelled with globally unique identifiers (such as OTUs and/or BINs), provide a pragmatic, accurate solution to the “taxonomic impediment”; and (c) to demonstrate that interim names based on BINs and OTUs obtained through metabarcoding provide an effective method for studies on species-rich groups that are usually neglected in biodiversity research projects because of their unresolved taxonomy.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Molecular Ecology Resources* Published by John Wiley & Sons Ltd

KEYWORDS

barcode library, biodiversity monitoring, CO1, cryptic diversity, Diptera, DNA barcoding, Germany, metabarcoding, mitochondrial DNA

1 | INTRODUCTION

Recent evidence for major declines in insect populations has provoked intense public concern. Detailed research on economically important groups, such as pollinators, have linked declines in wild bees to pesticide contamination, climate change, habitat fragmentation and degradation (Potts et al., 2010; Vanbergen & the Insect Pollinators Initiative, 2013). Other studies using mass collecting methods suggest the declines may be general, as evidenced by reductions in the biomass of flying insects by 75% over a few decades (Hallmann et al., 2017; Sorg, Schwan, Stenmans, & Müller, 2013) or even within a few years (Lister & Garcia, 2018). However, the evidence for general declines has failed to ascertain if impacts span all insect groups and all size ranges. The failure to track the status of individual lineages reflects the fact that despite advances in taxonomic practices (e.g., integrative taxonomy), our knowledge of most insect species is limited (Brix, Leese, Riehl, & Kihara, 2015; Cruaud, Rasplus, Rodriguez, & Cruaud, 2017; Pante, Schoelincx, & Puillandre, 2014; Riedel, Sagata, Suhardjono, Tänzler, & Balke, 2013; Wheeler, Raven, & Wilson, 2004). Even in Germany, a country with more than 250 years of taxonomic and faunistic research activity, many groups remain poorly known. This gap, which hampers ecological baseline research, is particularly serious for the two hyperdiverse insect orders, the Diptera and Hymenoptera (Geiger, Moriniere, et al., 2016; Klausnitzer, 2006). With at least 9,500 (Schumann, Bährmann, & Stark, 1999; Schumann, Doczkal, & Ziegler, 2011) and 9,600 (Dathe & Blank, 2004) recorded species in Germany, respectively, these two groups comprise over half of its insect alpha diversity (Völkl, Blick, Kornacker, & Martens, 2004). Moreover, it is likely that the true diversity of these two groups is seriously underestimated, a conclusion reinforced by the extraordinarily high numbers of DNA barcode clusters retrieved by simultaneous analysis of arthropods using high-throughput sequencing (HTS; metabarcoding) from insect collections at single monitoring sites (Moriniere et al., 2016). As only about 1,000 (Santos, Samprinha, & Santos, 2017) new species of Diptera are described each year from the million or more species awaiting description, the taxonomic impediment in this group will not be resolved without the adoption of new approaches, such as modern molecular genetic methods and integrative taxonomy (Fujita, Leache, Burbrink, McGuire, & Moritz, 2012; Padial, Miralles, Riva, & Vences, 2010; Schlick-Steiner, Arthofer, & Steiner, 2014; Schlick-Steiner et al., 2010).

The known dipteran fauna of Germany includes roughly half of the almost 20,000 species recorded for Europe (as defined in Fauna Europaea, <https://fauna-eu.org/>; Pape, 2009). Although this is the highest number of Diptera species recorded from any European country, the inventory is certainly very incomplete. A recent

checklist for the Empidoidea of Germany (Meyer & Stark, 2015) added 123 species new to Germany, an increase of 12.5%, Jaschhof (2009) added 34 species of Lestremiinae, an increase of 24.3%, and the collecting efforts for different barcoding campaigns resulted in more than 100 species from various families new to Germany among the identified material (Reimann & Rulik, 2015; Heller & Rulik, 2016; B. Rulik unpublished, D. Doczkal unpublished), with many more expected among the unidentified material. Rapid progress in inventorying is hampered by a lack of experts, also known as the taxonomic impediment (de Carvalho et al., 2007). For example, the German Dipterologist's working group (<http://www.ak-diptera.de/index.htm>, Accessed 18 December 2018) shows that experts were lacking for one-third of the dipteran families, and that most other families had just one or two experts, often voluntary (i.e., unpaid) taxonomists (in the sense defined by Fontaine et al., 2012). A few families such as the Culicidae (<https://mueckenatlas.com/>), the Asilidae (Wolff, Gebel, & Geller-Grimm, 2018) and the Syrphidae (Ssymank, Doczkal, Rennwald, & Dziock, 2011) are fairly well explored, but several of the species-richest families (e.g., Cecidomyiidae, Ceratopogonidae, Phoridae, Chloropidae, Sphaeroceridae, Anthomyiidae) have received little attention. Malaise traps are widely used as method of choice to collect arthropods and especially flying insects for biodiversity assessments in terrestrial ecosystems, with Diptera being among the most commonly caught taxa (Doczkal, 2017; Hallmann et al., 2017; Hebert et al., 2016; Karlsson, Pape, Johansson, Liljebblad, & Ronquist, 2005; Matthews & Matthews, 1971; Ssymank et al., 2018). The analysis of specimens from two Malaise traps deployed for a single summer in Germany within the Global Malaise Trap Program (GMTP; <http://biodiversitygenomics.net/projects/gmp/>) revealed similar trends; here Diptera was the most diverse order being represented by 2,500 species, slightly more than half of all the species that were collected and 70.3% of all individuals (26,189) that were analysed (Geiger, Moriniere, et al., 2016).

Taxonomists working on Diptera have long been well aware of the immense number of undescribed species (Bickel et al., 2009) with estimates of global Diptera species diversity ranging from 400,000 to 800,000 species compared with ~160,000 named species (Borkent et al., 2018; Pape, Blagoderov, & Mostovski, 2011). Hebert et al. (2016), applying DNA barcoding to Canadian insects, proposed that the actual number of species could be much higher, suggesting the possible presence of 1.8 million species in just one family, the Cecidomyiidae (gall midges) alone. Although this estimate may be too high, it is very likely that this single family includes more species than are currently described for the order.

At a time when hundreds or possibly thousands of species become extinct each year (Chivian & Bernstein, 2008), a comprehensive species inventory based on accurately identified specimens

represents the foundation for all conservation and biodiversity initiatives. However, the inventory of biodiversity cannot be completed through morphological approaches alone. Both the speed and costs associated with sequence characterization of a standardized DNA fragment can be improved using DNA barcoding. Usually DNA barcoding studies provide a basis for establishing the reference sequence libraries required to identify specimens of known species (Gwiżdowski, Footitt, Maw, & Hebert, 2015; Hebert, Cywinska, Ball, & Dewaard, 2003). Herein we additionally show that it is also an efficient method for registering unknown and taxonomically challenging species—so called “dark taxa” (Page, 2016). Sequenced taxa can subsequently be associated with established binomens by taxonomic specialists using a reverse taxonomy approach, based on accurately identified museum specimens (ideally type specimens) and expert knowledge. During this process, specimens that belong to unnamed molecular character-based units (operational taxonomic units [OTUs] or barcode index numbers [BINs]) will either be referenced to known species or they may represent overlooked species that are new to science (Geiger, Moriniere, et al., 2016). A curated and comprehensive DNA barcode reference library enables fast and reliable species identifications in those many cases where time, personnel and taxonomic expertise are limited. Furthermore, such a library also supports large-scale biodiversity monitoring that relies upon metabarcoding bulk samples (Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011; Hajibabaei, Spall, Shokralla, & Konynenburg, 2012; Shokralla, Spall, Gibson, & Hajibabaei, 2012), like those obtained from Malaise traps (Gibson et al., 2014; Leray & Knowlton, 2015; Morinière et al., 2016; Yu et al., 2012).

The results reported in this study derive from two major DNA barcoding campaigns: “Barcoding Fauna Bavarica” (BFB, www.fauna.bavarica.de; Haszprunar, 2009) and the “German Barcode of Life” project (GBOL, www.bolgermany.de; Geiger, Astrin, et al., 2016). Since 2009, DNA barcodes from over 23,000 German species of Metazoa have been assembled, reflecting the analysis of nearly 250,000 specimens that are curated at the SNSB-Zoologische Staatssammlung München (ZSM, see www.barcoding-zsm.de) and ~180,000 specimens curated at the Zoologisches Forschungsmuseum Alexander Koenig Bonn (ZFMK). These records represent a major contribution to the global DNA barcode library that is maintained in the Barcode of Life Data System (BOLD, www.boldsystems.org; Ratnasingham & Hebert, 2007). Currently, the DNA barcode library created by the ZSM researchers represents the second-most comprehensive library of any nation. Previous studies have reported on barcoding results for Coleoptera (Hendrich et al., 2015; Raupach, Hannig, Moriniere, & Hendrich, 2016; Raupach, Hannig, Moriniere, & Hendrich, 2018; Rulik et al., 2017), Ephemeroptera, Plecoptera and Trichoptera (Morinière et al., 2017), Heteroptera (Havemann et al., 2018; Raupach et al., 2014), Hymenoptera (Schmid-Egger et al., 2019; Schmidt, Schmid-Egger, Morinière, Haszprunar, & Hebert, 2015; Schmidt et al., 2017), Lepidoptera (Hausmann, Haszprunar, & Hebert, 2011; Hausmann, Haszprunar, Segerer, et al., 2011), Neuroptera (Morinière et al., 2014), Orthoptera (Hawlitcshek et al., 2017), Araneae and Opiliones (Astrin et al., 2016), and Myriapoda

(Spelda, Reip, Oliveira Biener, & Melzer, 2011; Wesener et al., 2015). Concerning DNA barcoding studies performed for Diptera, no comprehensive study encompassing this entire highly diverse order has been published, but data have been used to revise smaller units thereof: for example, for Calliphoridae (Jordaens et al., 2013; Nelson, Wallman, & Dowton, 2007; Reibe, Schmitz, & Madea, 2009), Ceratopogonidae (Stur & Borkent, 2014), Chironomidae (Carew, Pettigrove, Cox, & Hoffmann, 2007; Carew, Pettigrove, & Hoffmann, 2005; Cranston et al., 2013; Ekrem, Stur, & Hebert, 2010; Ekrem, Willassen, & Stur, 2007; Montagna, Mereghetti, Lencioni, & Rossaro, 2016; Pfenninger, Nowak, Kley, Steinke, & Streit, 2007; Sinclair & Gresens, 2008; Stur & Ekrem, 2011), Culicidae (Ashfaq et al., 2014; Cywinska, Hunter, & Hebert, 2006; Kumar, Rajavel, Natarajan, & Jambulingam, 2007; Versteirt et al., 2015; Wang et al., 2012), Hybotidae (Nagy, Sonet, Mortelmans, Vandewynkel, & Grootaert, 2013), Muscidae (Renaud, Savage, & Adamowicz, 2012), Psychodidae (Gutiérrez, Vivero, Vélez, Porter, & Uribe, 2014; Krüger, Strüven, Post, & Faulde, 2011; Kumar, Srinivasan, & Jambulingam, 2012; Nzulu et al., 2015), Sciaridae (Eiseman, Heller, & Rulik, 2016; Heller, Köhler, Menzel, Olsen, & Gammelo, 2016; Heller & Rulik, 2016; Latibari, Moravvej, Heller, Rulik, & Namaghi, 2015; Ševčík, Kaspřák, & Rulik, 2016), Simuliidae (Rivera & Currie, 2009), Syrphidae (Jordaens et al., 2015) and Tachinidae (Pohjoismäki, Kahanpää, & Mutanen, 2016).

This publication presents the first results of the Diptera campaign and it provides coverage for 5,200 BINs (Ratnasingham & Hebert, 2013). It covers ~55% of the known Diptera fauna from Germany. According to the checklist of German Diptera (Schumann et al., 1999) and the three additions published so far (Schumann, 2002, 2004, 2010) 9,544 species of Diptera have been recorded from Germany. The Diptera library now includes a total of 2,453 reliable species identifications, and 2,700 BINs, which possess either interim species names or just higher-level taxonomy (genus or family; “dark taxa”). Although it has been shown that BINs correspond closely to biological species of most insect orders (Hausmann et al., 2013), there are other studies reporting difficulties in determining species through DNA barcodes within Diptera. In particular, well-studied groups such as the syrphids represent a problem, because here additional genes for a clear type assignment must be consulted in many genera (Mengual, Ståhls, Vujić, & Marcos-García, 2006; Rojo, Ståhls, Pérez-Bañón, & Marcos-García, 2006). Further examples of problems in species delineation due to barcode gaps, at least for some genera, are the Tachinidae and the Calliphoridae (Nelson et al., 2012; Pohjoismäki et al., 2016; Whitworth, Dawson, Magalon, & Baudry, 2007). In one of the few studies dealing with DNA barcoding in Diptera it was shown, that less than 70% of a composition of about 450 species covering 12 families of Diptera could be reliably identified by DNA barcoding, as there was wide overlap between intra- and interspecific genetic variability on the *COI* gene (Meier, Shiyang, Vaidya, & Ng, 2006). However we find that more than 88% of the studied species, identified based on morphology or BIN matches to the BOLD database, can be unambiguously identified using their DNA barcode sequences. BINs enable the creation of an interim

taxonomic system in a structured, transparent and sustainable way and thus become a valuable foundation for subsequent detailed, integrative taxonomic studies. Furthermore, the BIN system enables analyses that are equivalent to studies based on named species, that is where the underlying specimens are identified by specialists using traditional methods (i.e., morphology). The latter will play a special role in the processing, classification and genetic inventorying of less-explored “dark taxa,” which have been treated and processed with less priority by previous DNA barcoding activities. Moreover, this automated approach of delineating species is less affected by operational taxonomic biases, so it can provide more objective identifications than conventional approaches (Mutanen et al., 2016; Packer, Gibbs, Sheffield, & Hanner, 2009; Schmidt et al., 2015). Using DNA extracts derived from bulk collections made by Malaise traps, we further demonstrate that species delineation using interim names based on BINs and OTUs constitutes an effective method for biodiversity studies using DNA metabarcoding. As the reference libraries continue to grow and gaps in the species catalogue are subsequently filled, BIN lists assembled by metabarcoding will provide improved taxonomic resolution.

The present study has three main goals: (a) to provide a DNA barcode library for 5,200 BINs of Diptera; (b) to demonstrate, based on the example of bulk extractions from a Malaise trap experiment, that DNA barcode clusters, labelled with globally unique identifiers (such as OTUs and/or BINs), provide a pragmatic, accurate solution to the “taxonomic impediment”; and (c) to demonstrate that interim names based on BINs and OTUs obtained through metabarcoding is an effective method for studies on species-rich groups that are usually neglected in biodiversity research projects because of their unresolved taxonomy.

2 | MATERIALS AND METHODS

2.1 | Fieldwork, specimens and taxonomy

A network of 130 (professional and voluntary) taxonomists and citizen scientists collected and contributed specimens to the DNA barcoding projects, primarily from various German states, but also from surrounding European countries (Austria, Belgium, Czech Republic, France, Italy). Most specimens (94.5%, 42,587 of 45,040 with *COI* sequences >500 bp) were collected by Malaise traps, which were deployed from 2009 to 2016. The study sites included more than 683 localities in state forests, public lands and protected areas such as the Nationalparks “Bayerischer Wald” and “Berchtesgadener Land,” the EU habitats directive site “Landskrone,” as well as alpine regions at altitudes up to 2,926 m (Zugspitze). Detailed information on collection sites and dates is available in Appendix S1. Since 2009, more than five million specimens of Diptera were collected by hand collecting, sweep netting, and by Malaise-, window- and pitfall-trapping. However, most voucher specimens have been extracted from Malaise trap samples. Twenty to 100 Malaise traps were deployed in each of seven years (2011–2017) mostly across habitats in Bavaria and Baden-Württemberg; one trap was placed

in Rhineland-Palatinate. Samples were screened morphologically to maximize the diversity of species submitted for sequence characterization. Most vouchers were derived from Germany (44,511), but others were collected in France (222), Czech Republic (147), Belgium (106), Austria (70) and other Central European countries (18). All samples and specimens are now stored in the SNSB-ZSM or ZFMK except for a few held in private collections. From the entire collection, ~3,000,000 specimens of potential interest, most of which derived from the huge Malaise trap experiments in the framework of the GMTP, were identified to family level mostly by D.D. and to a minor extent by B.R. and experienced specialists using appropriate literature (Oosterbroek, 2006 and references therein; Papp & Darvas, 1997, 1998, 2000a, 2000b, Schumann et al., 2011). From this material, 59,000 specimens were submitted for sequence analysis through the DNA barcoding pipeline (including sample preparation, high-quality imaging and metadata acquisition for each specimen) established at the ZSM to support its involvement in national and international DNA barcoding projects. Most samples (>99%) were stored in 96% EtOH before DNA extraction. Specimen ages generally ranged from 1 to 5 years (43,112 specimens, 96%); only 4% were more than 5 years old. The number of specimens analysed per species ranged from one to 1,356 (i.e., *Megaselia rufa*) (Wood, 1908; see Appendix S1). When taxonomic expertise was available, specimens were sent to specialists to obtain as many species-level identifications as possible.

2.2 | Laboratory protocols

A tissue sample was removed from each specimen and transferred into 96-well plates at the SNSB-ZSM for subsequent DNA extraction. For specimens with a body length >2 mm a single leg or a leg segment was removed for DNA extraction. The whole voucher was used for some very small specimens (e.g., ≤1 mm, such as small Cecidomyiidae, Chironomidae and Sciaridae), but replacement vouchers from the same locality were retained. In other cases (vouchers from Malaise traps), DNA was extracted from the whole voucher at the CCDB (Guelph, Canada) using “voucher-recovery” protocols (DeWaard et al., 2019) and the specimens were repatriated to the SNSB-ZSM and ZFMK for identification and curation. DNA extraction plates with the tissue samples were sent to the Canadian Center for DNA Barcoding (CCDB) where they were processed using standard protocols. All protocols for DNA extraction, PCR amplifications and Sanger sequencing procedures are available online (ccdb.ca/resources/). All samples were PCR-amplified with a cocktail of standard and modified Folmer primers CLepFoIF (5'-ATTCAACCAATCATAAAGATATTGG) and CLepFoIR (5'TAAACTTCTGGATGTCCAAAAAATCA) for the barcode fragment (5' *COI*; see Hernández-Triana et al., 2014), and the same primers were employed for subsequent bidirectional Sanger sequencing reactions (see also Ivanova, Dewaard, & Hebert, 2006; deWaard, Ivanova, Hajibabaei, & Hebert, 2008, DeWaard et al., 2019). Voucher information such as locality data, habitat, altitude, collector, identifier, taxonomic classifications, habitus images, DNA barcode sequences, primer pairs and trace files for 40,753

specimens are publicly accessible in the “DS-DIPBFGBL—A DNA Barcode reference library of German Diptera (BFB—Barcoding Fauna Bavarica & GBOL—German Barcode of Life)” data set on BOLD (<http://www.boldsystems.org> – data set DOI: [dx.doi.org/10.5883/DS-DIPBFGBL](https://doi.org/10.5883/DS-DIPBFGBL)), whereas 4,420 specimen records will be stored in the private data set “DS-DIPBFGBP—A DNA Barcode reference library of German Diptera (BFB—Barcoding Fauna Bavarica & GBOL—German Barcode of Life)—private records for future publication” for subsequent publication by the authors and associated taxonomists.

2.3 | Data analysis

Sequence divergences for the *COI*-5P barcode region (mean and maximum intraspecific variation and minimum genetic distance to the nearest-neighbouring species) were calculated using the “Barcode Gap Analysis” tool on BOLD, employing the Kimura 2-parameter (K2P) distance metric (Puillandre, Lambert, Brouillet, & Achatz, 2012). The program MUSCLE was applied for sequence alignment restricting analysis to sequences with a minimum length of 500 bp. Neighbour-joining (NJ) trees were calculated following alignment based on K2P distances. The “BIN Discordance” analysis on BOLD was used to reveal cases where species assigned to different species shared a BIN, and those cases where a particular species was assigned to two or more BINs. Sequences are grouped into clusters of closely similar *COI* barcode sequences, which are assigned a globally unique identifier, termed a “barcode index number” or BIN (Ratnasingham & Hebert, 2013). This system enables tentative species identifications when taxonomic information is lacking. The BIN system involves a three-step online pipeline, which clusters similar barcode sequences algorithmically into OTUs being “named” by a number. For the majority of studied insect orders, specimens sharing a BIN very often represent a close species-proxy as delineated by traditional taxonomy (e.g., for Lepidoptera, Hausmann et al., 2013). However, some genera or families throughout the insects exhibit problems with species delineation based on DNA barcodes, due to high intra- or low interspecific genetic distances (e.g., cryptic diversity, BIN sharing or the barcode gap; see Hubert & Hanner, 2015). Within the Diptera, this phenomenon has been well documented (Meier et al., 2006), at least in some families, such as caliphorid, syrphid and tachinid species (Mengual et al., 2006; Nelson et al., 2012; Pohjoismäki et al., 2016; Rojo et al., 2006; Whitworth et al., 2007), but may also occur in families of “dark taxa” as well.

Every other “disagreement/conflict” case is the starting point for re-evaluation of both molecular and morphological data. We follow the concept of Integrative Taxonomy (Fujita et al., 2012; Padial et al., 2010; Schlick-Steiner et al., 2014, 2010) to infer whether there are previously overlooked species (“cryptic taxa”) in the sample, or whether barcode divergence between species is too low or absent to allow valid species to be delineated using only *COI* characteristics.

2.4 | Reverse-taxonomy approach

When sequenced specimens could only be assigned to a category above the species level (family, subfamily or genus), we used interim

species names (such as *TachIntGen1* sp.BOLD:AAG2112) based on the corresponding BIN, so these specimens could be included in the “Barcode Gap Analysis” in order to provide more comprehensive estimates of the distribution of genetic divergences among both species assigned to Linnaean species and those with BIN assignments. This analysis was conducted on all specimens at the same time after updating the interim taxonomy where necessary. For specimen records, which lack lower taxonomy (e.g., those uploaded only as “Diptera”), we applied the highest “conflict-free” taxonomy—for example the genus name, when other specimens within that BIN had the same identification—using a BIN match with the public data on BOLD (e.g., *Melanagromyza* sp. BOLD:ACP6151). All specimens, which could not be identified to species or genus level, and where the vouchers were in acceptable condition (e.g., unbroken antennae and/or legs after retrieval from Malaise trap), were selected using the corresponding BINs for identification by taxonomic specialists. Interim names were subsequently moved into the “Voucher status” field in the BOLD metadata tables after all analyses were performed.

2.5 | Metabarcoding and bioinformatic data analysis

The potential utility of the DNA barcode library for biomonitoring Diptera was tested with field samples, focusing on an early warning system for pest and invasive species based on metabarcoding (L. A. Hardulak et al. in prep). In this study, nine Malaise traps were deployed in the Bayerischer Wald National Park and its surroundings during the vegetated period (May–September) in 2016. Trap bottles were changed twice monthly, producing a total of 90 bulk samples of macroinvertebrates. All specimens were dried and ground with a stainless steel pestle (no size-sorting step), and tissue lysis of insect powder per trap sample was performed overnight, using a solution of 90% insect lysis buffer and 10% proteinase K. DNA extraction was performed with the DNEasy Blood & Tissue kit (Qiagen). A minibarcode region was amplified by PCR, using forward and reverse NGS primers (Leray et al., 2013) targeting a 313-bp-long coding region of mitochondrial *COI*. High-throughput sequencing was performed on an Illumina MiSeq using version 2 (2 × 250 bp, 500 cycles, maximum of 20 million reads) chemistry at the Sequencing Service Unit of the Ludwig-Maximilians University (LMU, Munich, Germany; see Appendix S5 for a more detailed metabarcoding protocol).

Sequence processing was performed with the VSEARCH version 2.4.3 suite (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) and CUTADAPT version 1.14 (Martin, 2011). Forward and reverse reads in each sample were merged with the VSEARCH program “fastq_mergepairs” with a minimum overlap of 40 bp, yielding ~313-bp sequences. Forward and reverse primers were removed with CUTADAPT, using the “discard_untrimmed” option to discard sequences for which primers were not detected at ≥90% identity. Quality filtering was done with the “fastq_filter” in VSEARCH, keeping sequences with zero expected errors (“fastq_maxee” 1). Sequences were dereplicated with “derep_fulllength,” first at the sample level, and then concatenated into a fasta file, which was then dereplicated. Chimeric sequences were removed from the fasta file using “uchime_denovo.” The remaining

sequences were then clustered into OTUs at 97% identity employing “cluster_size,” a greedy, centroid-based clustering program. OTUs were BLASTed against the Diptera database downloaded from BOLD including taxonomy and BIN information in GENEIOUS (version 9.1.7; Biomatters) following the methods described in Morinière et al. (2016). The resulting csv file, which included BIN, Hit-%-ID value, family, genus and species information for each out, was exported from Geneious and combined with the OTU table generated by the bioinformatic pipeline. The combined results table was then filtered by Hit-%-ID value and total read numbers per OTU. All entries with identifications below 97% and total read numbers below 0.01% of the summed reads per sample were removed from the analysis. OTUs were then assigned to the respective BIN (Appendix S2). Presence-absence overviews of selected Diptera taxa (BINs) within the metabarcoding study were created; one-sided Pearson correlation coefficients were calculated to estimate the percentage of “dark taxa” with mid-range body size versus the number of species reported in Germany, both with the inclusion and with the exclusion of families with 0% “dark taxa.” (R version 3.4.4 [2018-03-15], R Core Team, 2018).

3 | RESULTS

3.1 | DNA barcoding/developing a reference library

From the 59,102 specimens submitted for Sanger sequencing, 50,963 COI-5P sequences (86.23%) were recovered. Length of the

recovered sequence varied with the sequencing protocol; 12.54% (7,410 specimens) were bidirectionally sequenced and yielded a full-length (658 bp) barcode while the rest (43,533) were unidirectionally sequenced yielding 69.95% (41,339) with sequences <658 to >500 bp and 3.75% (2,214 specimens) with sequences <500 bp. No sequence information was recovered from 13.77% (8,139) of the specimens. Barcode recovery was most successful for EtOH-preserved specimens less than 10 years old. For the subsequent analyses we selected 45,040 specimens with high-quality DNA barcode sequences (≥ 500 bp), which fulfilled the requirements for being assigned to a BIN. This data set included ~5,200 BINs (2,500 were assigned a total of 2,453 Linnean species while 2,700 lacked a species designation, 52.4% of the data set). These BINs included one or more representatives from 88 of the 117 (75%) dipteran families known from Germany (Figure 1, Table 1; Appendix S3, Krona graph in Figure S2). More than one-third (1,829) of the BINs were new to BOLD.

Inspection of the COI sequence clusters using NJ trees (created with analytical tools on BOLD) and using the TaxCl-approach for detecting taxonomic incongruences (Rulik et al., 2017) revealed high congruence with morphology-based identifications. Among the 2,453 taxa assigned a Linnean binomen based on morphological identifications and “conflict-free” BIN matches, 88.67% (2,138) were unambiguously discriminated by their COI sequences. Another 122 species (4.97%), representing 8.7% of all studied specimens (3,951 individuals), were assigned to more than one BIN, resulting in a total of 255 BINs (Table 1; Appendix S3). For purposes of re-identification, the species in this subset can also be unambiguously assigned

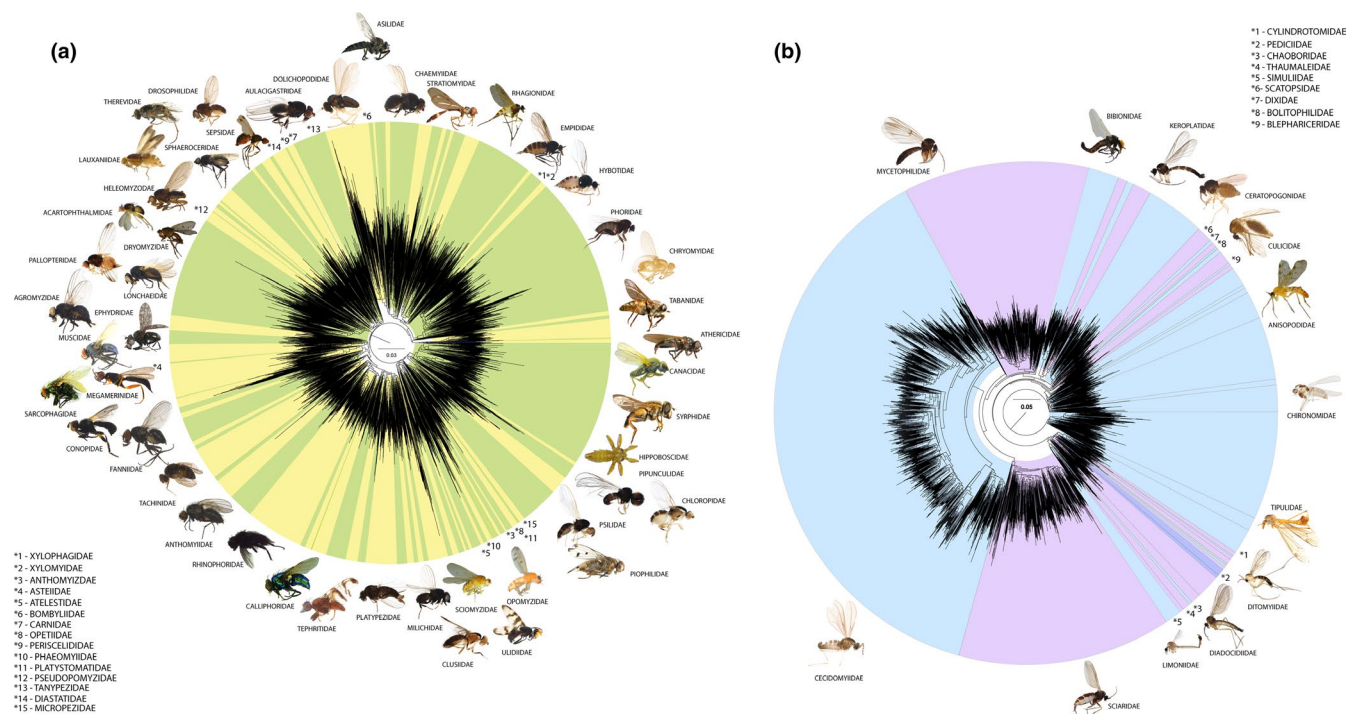


FIGURE 1 Illustrative circular neighbour-joining (NJ) trees for (a) all Brachycera and (b) all Nematocera within the Diptera barcode library; each line in the trees corresponds to one barcode index number (BIN). NJ tree calculations were performed on the BOLD database. A more detailed observation of the BIN diversity for each family can be studied within the Krona graph within the supporting information (Figure S2) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Families of Diptera reported in Germany. Information on BIN count, and on the numbers of named and unnamed species within the reference database

Infraorder	Family	Species reported in Germany	BINs	Ratio barcoded/species (%)	Size (mm)	Total number of taxa/with barcode	Unnamed/with barcode	% of dark taxa
Brachycera	Acartophthalmidae	2	1	50	1.0–2.5	2	0	0
Brachycera	Acroceridae	11	0	0	2.5–20.0	N/A	N/A	N/A
Brachycera	Agromyzidae	552	218	39	1.0–6.0	214	149	70
Nematocera	Anisopodidae (& Mycetobiidae)	8	7	88	4.0–12.0	7	2	29
Brachycera	Anthomyiidae	227	188	83	4.0–12.0	178	64	36
Brachycera	Anthomyzidae	14	5	36	1.3–4.5	5	0	0
Brachycera	Asilidae	81	18	22	8.0–20.0	18	6	33
Brachycera	Asteiidae	7	3	43	1.0–3.0	3	0	0
Brachycera	Atelestidae	3	3	100	1.5–3.5	3	0	0
Brachycera	Athericidae	5	3	60	7.5–10.0	3	1	33
Brachycera	Aulacigastridae	1	0	0	2.0–5.0	0	0	N/A
Nematocera	Bibionidae (& Pleciidae)	21	12	57	2.0–15.0	10	2	20
Nematocera	Blephariceridae	7	2	29	3.0–15.0	2	1	50
Nematocera	Bolitophilidae	22	14	64	4.0–7.0	13	7	54
Brachycera	Bombyliidae	40	6	15	1.0–20.0	6	1	17
Brachycera	Braulidae	1	0	0	1.2–2.5	2	0	0
Brachycera	Calliphoridae	62	35	56	4.0–16.0	39	6	15
Brachycera	Camillidae	4	0	0	2.0–3.5	N/A	N/A	N/A
Brachycera	Campichoetidae	3	0	0	2.5–4.0	N/A	N/A	N/A
Brachycera	Canacidae	2	9	450	1.6–5.0	9	1	11
Nematocera	Canthylloscelidae	1	0	0	2.5–9.0	N/A	N/A	N/A
Brachycera	Carnidae	11	7	64	1.0–2.5	7	7	100
Nematocera	Cecidomyiidae	836	927	111	0.5–3.0	926	882	95
Nematocera	Ceratopogonidae	332	131	39	1.0–5.0	128	97	76
Brachycera	Chamaemyiidae	29	17	59	1.0–5.0	17	13	76
Nematocera	Chaoboridae	7	2	29	2.0–10.0	2	0	0
Nematocera	Chironomidae	696	455	65	1.0–10.0	438	286	65
Brachycera	Chloropidae	198	101	51	1.0–5.0	101	59	58
Brachycera	Chyromyidae	5	2	40	0.5–8.0	2	0	0
Brachycera	Clusiidae	9	6	67	1.5–8.0	7	3	43
Brachycera	Coelopidae	2	0	0	2.5–9.0	N/A	N/A	N/A
Brachycera	Coenomyiidae	1	0	0	14.0–20.0	N/A	N/A	N/A
Brachycera	Conopidae	52	9	17	5.0–15.0	9	0	0
Brachycera	Cremifaniidae	1	0	0	1.5–2.6	N/A	N/A	N/A
Brachycera	Cryptochetidae	1	0	0	2.0–4.0	N/A	N/A	N/A
Nematocera	Culicidae	46	8	17	3.0–9.0	7	0	0
Nematocera	Cylindrotomidae	4	1	25	11.0–16.0	1	0	0
Nematocera	Diadocidiidae	4	3	75	3–4.5.0	3	0	0
Brachycera	Diastatidae	6	8	133	2.5–4.0	8	2	25
Nematocera	Ditomyiidae	4	1	25	6.0–8.0	1	0	0
Nematocera	Dixidae	16	4	25	3.0–5.5	4	1	25
Brachycera	Dolichopodidae	356	112	31	1.0–9.0	112	58	52

(Continues)

TABLE 1 (Continued)

Infraorder	Family	Species reported in Germany	BINs	Ratio barcoded/species (%)	Size (mm)	Total number of taxa/with barcode	Unnamed/with barcode	% of dark taxa
Brachycera	Drosophilidae	59	28	47	1.5–7.0	28	5	18
Brachycera	Dryomyzidae	3	2	67	5.0–18.0	2	1	50
Brachycera	Eginiidae	1	0	0	2.0–18.0	N/A	N/A	N/A
Brachycera	Empididae (& Brachystomatidae)	383	161	42	1.0–12.0	161	107	66
Brachycera	Ephydriidae	177	130	73	1.0–11.0	132	16	12
Brachycera	Fanniidae	56	46	82	2.0–5.0	44	13	30
Brachycera	Gasterophilidae	4	0	0	9.0–16.0	N/A	N/A	N/A
Brachycera	Helcomyzidae	3	0	0	6.0–11.0	N/A	N/A	N/A
Brachycera	Heleomyzidae (& Heteromyzidae)	74	58	78	1.2–12.0	55	26	47
Nematocera	Hesperinidae	1	0	0	4.0–6.0	N/A	N/A	N/A
Brachycera	Hilarimorphidae	2	0	0	2.0–7.0	N/A	N/A	N/A
Brachycera	Hippoboscidae	12	7	58	2.5–10.0	7	1	14
Brachycera	Hybotidae	229	140	61	1.0–6.0	139	83	60
Brachycera	Hypodermatidae	5	0	0	10.0–22	N/A	N/A	N/A
Nematocera	Keroplastidae	60	30	50	4.0–15.0	30	12	40
Brachycera	Lauxaniidae	67	25	37	2.0–7.0	25	11	44
Nematocera	Limoniidae	280	96	34	2.0–11.0	91	50	55
Brachycera	Lonchaeidae	47	16	34	3.0–6.0	16	9	56
Brachycera	Lonchopteridae	9	5	56	2.0–5.0	6	0	0
Brachycera	Megamerinidae	1	1	100	6.0–9.0	1	0	0
Brachycera	Micropezidae	13	5	38	3.0–16.0	4	1	25
Brachycera	Microphoridae	6	0	0	1.5–3.0	N/A	N/A	N/A
Brachycera	Milichiidae	13	17	131	1.0–6.0	16	9	56
Brachycera	Muscidae	317	174	55	2.0–18.0	167	66	40
Nematocera	Mycetophilidae	573	306	53	2.0–15.0	301	89	30
Brachycera	Neottiophilidae	1	0	0	1.5–7.0	N/A	N/A	N/A
Brachycera	Nycteribiidae	8	0	0	1.5–5.0	N/A	N/A	N/A
Brachycera	Odiniidae	9	0	0	2.0–5.0	N/A	N/A	N/A
Brachycera	Oestridae	6	0	0	9.0–18.0	N/A	N/A	N/A
Brachycera	Opetiidae	1	1	100	2.0–5.0	1	0	0
Brachycera	Opomyzidae	15	4	27	2.0–5.0	4	1	25
Brachycera	Otitidae	26	0	0	2.5–11.0	N/A	N/A	N/A
Brachycera	Pallopteridae	16	8	50	2.5–7.0	7	0	0
Nematocera	Pediciidae	36	13	36	5.0–35.0	13	3	23
Brachycera	Periscelididae	6	1	17	1.0–5.0	1	0	0
Brachycera	Phaeomyiidae	3	2	67	3.0–11.0	2	0	0
Brachycera	Phoridae	364	289	79	0.5–6.0	276	166	60
Brachycera	Piophilidae	12	12	100	1.5–7.0	12	4	33
Brachycera	Pipunculidae	111	42	38	2.0–12.0	40	7	18
Brachycera	Platypezidae	23	4	17	1.5–6.0	4	0	0
Brachycera	Platystomatidae	3	2	67	3.0–11.0	2	0	0
Brachycera	Pseudopomyzidae	1	1	100	1.7–2.5	1	0	0

(Continues)

TABLE 1 (Continued)

Infraorder	Family	Species reported in Germany	BINs	Ratio barcoded/species (%)	Size (mm)	Total number of taxa/with barcode	Unnamed/with barcode	% of dark taxa
Brachycera	Psilidae	30	12	40	2.5–10.0	12	8	67
Nematocera	Psychodidae	143	51	36	2.0–6.0	50	25	50
Nematocera	Ptychopteridae	8	0	0	7.0–15.0	N/A	N/A	N/A
Brachycera	Pyrgotidae	1	0	0	8.0–9.0	N/A	N/A	N/A
Brachycera	Rhagionidae	35	20	57	2.0–20.0	20	10	50
Brachycera	Rhinophoridae	10	9	90	2.0–11.0	7	1	14
Brachycera	Sarcophagidae	130	49	38	3.0–22.0	49	17	35
Brachycera	Scatophagidae	57	0	0	3.0–12.0	0	0	N/A
Nematocera	Scatopsidae	47	30	64	0.5–4.0	30	24	80
Brachycera	Scenopinidae	3	0	0	2.0–7.0	N/A	N/A	N/A
Nematocera	Sciaridae	342	310	91	1.0–6.0	284	81	29
Brachycera	Sciomyzidae	78	19	24	2.0–14.0	18	4	22
Brachycera	Sepsidae	31	15	48	2.0–6.0	13	1	8
Nematocera	Simuliidae	50	19	38	1.2–6.0	18	9	50
Brachycera	Sphaeroceridae	137	79	58	0.7–5.5	77	31	40
Brachycera	Stratiomyidae	66	21	32	2.0–25.0	22	6	27
Brachycera	Strongylophthalmyiidae	1	0	0	3.0–5.5	N/A	N/A	N/A
Brachycera	Syrphidae	440	242	55	3.5–35.0	297	24	8
Brachycera	Tabanidae	58	46	79	6.0–30.0	45	3	7
Brachycera	Tachinidae	494	214	43	2.0–20.0	211	76	36
Brachycera	Tanypezidae	1	1	100	5.0–8.0	1	0	0
Brachycera	Tephritidae	110	28	25	2.5–10.0	27	5	19
Brachycera	Tethinidae	10	0	0	1.5–3.5	N/A	N/A	N/A
Nematocera	Thaumaleidae	15	13	87	3.0–5.0	13	1	8
Brachycera	Therevidae	32	4	13	2.5–15.0	4	1	25
Brachycera	Thyreophoridae	2	0	0	1.5–7.0	N/A	N/A	N/A
Nematocera	Tipulidae	123	46	37	7.0–35.0	46	15	33
Nematocera	Trichoceridae	18	24	133	3.0–9.0	24	17	71
Brachycera	Trixoscelididae	4	0	0	2.0–4.0	N/A	N/A	N/A
Brachycera	Ulidiidae	4	9	225	2.5–11.0	9	4	44
Brachycera	Xylomyidae	3	1	33	6.0–20.0	1	0	0
Brachycera	Xylophagidae	4	1	25	5.0–11.0	1	0	0

Note: Additional information on the average body size of the specimens in each family is included.

to a current species. For 34 of these taxa, the maximum intraspecific variation (maxISP) was <3% (range: 1.1%–3.0%), cases which may reflect either young sibling species or high intraspecific variation arising from secondary contact between phylogeographical lineages. Another 88 species showed considerably higher divergences with maxISP ranging from 3% to 6% in 48 species and from 6% to 12% in another 40 species, cases that are strong candidates for overlooked cryptic diversity. Most of these cases involved species whose members were assigned to two BINs (112 species), but specimens of nine species were assigned to three BINs and those of one other to four BINs. Another 156 species (6.56%), representing 2.9% of all specimens (1,316 specimens), involved two or more named species that

shared a BIN (Table 2). Ten of these species pairs possessed shallow but consistent divergences within the BIN, meaning that COI sequences enabled species identification (e.g., *Chrysotoxum bicinctum* Linnaeus, 1758 and *Chrysotoxum festivum* Linnaeus, 1758; *Sericomyia lappona* Linnaeus, 1758 and *Sericomyia silentis* Harris 1776; *Paragus majoranae* Rondani, 1857 and *Paragus pecchiolii*, Rondani, 1857). Interestingly, almost two-thirds (105/156) of the species exhibiting BIN sharing (168) were hoverflies (Syrphidae), a family that has seen intensive taxonomic study.

Appendix S1 provides species names, sample IDs, BIN assignment and collection information. All project data are available under the publicly accessible DOI: [dx.doi.org/10.5883/DS-DIPFBGL](https://doi.org/10.5883/DS-DIPFBGL).

3.2 | Performance of the reference library for metabarcoding of Malaise trap samples

Among the 90 Malaise trap samples from the Bavarian Forest National Park (L. A. Hardulak et al. in prep.), metabarcoding revealed 1,735 dipteran OTUs, comprising 536,376 reads: 5,960 average reads per sample, matching at 97% or higher to a taxon in the DNA barcode library downloaded from BOLD (average read count was 6,928 per sample with a total of 2,809 OTUs matched to the Diptera database with $\geq 90\%$). Multiple OTU matches to a single BIN were merged. Using the Diptera data, we identified a total of 1,403 BINs including representatives of 71 families (1,385 species) within the metabarcoding data set (Appendix S2). Almost one-third (498/1403) of these BINs belonged to “dark taxa.” Figure 2 illustrates examples of presence/absence overviews for the families Muscidae, Cecidomyiidae, Chironomidae and Syrphidae for selected Malaise trap sites.

Among families containing “dark taxa,” the percentage of unnamed taxa was inversely correlated with body size ($r = -0.41$, $p = 0.0004$) and positively with numbers of species reported from Germany ($r = 0.33$, $p = 0.0037$) (Figure 3; Appendix S4).

4 | DISCUSSION

This study summarizes the results of a DNA barcoding campaign on German Diptera, work based on the characterization of 45,040 specimens. The resultant DNA barcode reference library included records for 5,200 BINs (2,453 named species comprising 2,500 BINs plus 2,700 unnamed BINs) belonging to 88 families, covering $\sim 50\%$ of the Diptera fauna reported for Germany (Schumann, 2002, 2004, 2010; Schumann et al., 1999). Until now, most of these families, especially some of the most diverse, have been taxonomically

TABLE 2 All cases of high intraspecific sequence variation at *COI*; cases of multiple BINs and/or cryptic diversity candidates (CDC)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Agromyzidae	<i>Napomyza cichorii</i>	CDC (2)	2.47	3.71	BOLD:AAP2990 BOLD:AAX3741
	<i>Phytomyza continua</i>	CDC (2)	2.84	5.44	BOLD:AAM6330 BOLD:AAY2701
	<i>Phytomyza ranunculi</i>	CDC (2)	3.26	6.43	BOLD:AAY3895 BOLD:ACL2003
Anthomyiidae	<i>Anthomyia liturata</i>	CDC (2)	0.87	1.98	BOLD:ACE4539 BOLD:ACE4540
	<i>Delia nuda</i>	CDC (2)	1.06	1.87	BOLD:ACJ0544 BOLD:ACJ0545
	<i>Hydrophoria lancifer</i>	CDC (2)	0.61	3.04	BOLD:AAG2460 BOLD:ADC1814
	<i>Pegomya flavifrons</i>	CDC (2)	2.5	8.83	BOLD:AAG2479 BOLD:AAG6754
	<i>Pegomya solennis</i>	CDC (2)	0.85	2.67	BOLD:ACD8686 BOLD:ACM6225
	<i>Pegomya winthemi</i>	CDC (2)	0.54	5.53	BOLD:AAG1783 BOLD:ABA6845
Bibionidae	<i>Biblio clavipes</i>	CDC (2)	1.2	2.46	BOLD:ACC6151 BOLD:ACR0881
	<i>Biblio nigriventris</i>	CDC (2)	1	3.13	BOLD:ABX1732 BOLD:ACU5368
Bolitophilidae	<i>Bolitophila austriaca</i>	CDC (2)	1.27	2.18	BOLD:AAG4863 BOLD:ACI5612
Ceratopogonidae	<i>Brachypogon sociabilis</i>	CDC (2)	1.24	2.31	BOLD:ABW3958 BOLD:ACE8195
	<i>Ceratopogon grandiforceps</i>	CDC (2)	2.63	3.94	BOLD:ABW3984 BOLD:ACP4327
	<i>Forcipomyia</i> sp. 4ES	CDC (2)	2.18	5.98	BOLD:AAM6200 BOLD:ACQ8860

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN	
Chironomidae	<i>Brillia bifida</i>	CDC (2)	2.31	6.93	BOLD:AAD7726 BOLD:ADI4999	
	<i>Cricotopus bicinctus</i>	CDC (2)	1.86	3.2	BOLD:AAI6018 BOLD:AAT9677	
	<i>Gymnometriocnemus brumalis</i>	CDC (2)	0.5	2.41	BOLD:ACD4501 BOLD:ACU9207	
	<i>Limnophyes natalensis</i>	CDC (2)	1.51	2.89	BOLD:AAB7361 BOLD:ACT1270	
	<i>Limnophyes</i> sp. 4SW	CDC (2)	1.49	4.03	BOLD:ACR9428 BOLD:ACU4225	
	<i>Mesosmittia flexuella</i>	CDC (2)	0.79	2.02	BOLD:ADE7569 BOLD:ACU4856	
	<i>Orthocladius fuscimanus</i>	CDC (2)	2	2.66	BOLD:AAV5075 BOLD:ACX3046	
	<i>Parametriocnemus stylatus</i>	CDC (2)	0.76	2.03	BOLD:AAI2687 BOLD:ACT9205	
	<i>Paraphaenocladus exagitans</i>	CDC (3)	2.54	5.88	BOLD:AAE3719 BOLD:ACQ4724 BOLD:ACT8523	
	<i>Paraphaenocladus impensus</i>	CDC (4)	6.85	11.99	BOLD:AAC4200 BOLD:ACT2714 BOLD:ACT5784 BOLD:ACU4175	
	<i>Paratanytarsus laccophilus</i>	CDC (2)	2.09	3.14	BOLD:AAC8842 BOLD:ACF2457	
	<i>Polypedilum convictum</i>	CDC (2)	2.45	4.61	BOLD:AAW4661 BOLD:ACT9278	
	<i>Smittia reissi</i>	CDC (2)	1.72	3.47	BOLD:ACS9748 BOLD:ACU4112	
	Conopidae	<i>Myopa testacea</i>	CDC (2)	3.72	3.72	BOLD:AAK8836 BOLD:AAK8838
	Dolichopodidae	<i>Microphor anomalus</i>	CDC (2)	5.47	11	BOLD:ACH9042 BOLD:ACH9043
<i>Microphor holosericeus</i>		CDC (2)	4.06	12.7	BOLD:ACB6469 BOLD:ACH6989	
Empididae	<i>Hemerodromia adulatoria</i>	CDC (2)	8.52	8.52	BOLD:ACJ6728 BOLD:ACJ6729	
	<i>Kowarzia barbatula</i>	CDC (2)	7.21	10.71	BOLD:ACJ6935 BOLD:ACJ7236	
	<i>Kowarzia tenella</i>	CDC (2)	5.39	10.8	BOLD:ACJ6935 BOLD:ACJ7236	

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Ephydriidae	<i>Allotrichoma laterale</i>	CDC (2)	6.44	6.44	BOLD:ABA8753 BOLD:ACF1575
	<i>Ditrichophora fuscella</i>	CDC (2)	3.81	7.62	BOLD:ABA8605 BOLD:ABA8606
	<i>Ditrichophora palliditarsis</i>	CDC (2)	3.87	6.57	BOLD:AA8675 BOLD:ABA8748
	<i>Halmopota salinarius</i>	CDC (2)	2.43	3.81	BOLD:ABA7826 BOLD:ABA7827
	<i>Hydrellia flaviceps</i>	CDC (2)	4.22	6.33	BOLD:ABA8652 BOLD:ABV8173
	<i>Philygria flavipes</i>	CDC (2)	1.19	2.03	BOLD:ABA8663 BOLD:ACK3229
	<i>Polytrichophora duplosetosa</i>	CDC (2)	2.05	4.11	BOLD:ABA8627 BOLD:ABA8628
	<i>Scatella obsoleta</i>	CDC (2)	1.25	2.5	BOLD:ABA7493 BOLD:ABA7494
	<i>Scatophila signata</i>	CDC (2)	3.3	3.3	BOLD:ABA7651 BOLD:ABA7652
	Fanniidae	<i>Fannia postica</i>	CDC (2)	2.35	7.03
Heleomyzidae	<i>Heleomyza serrata</i>	CDC (2)	0.37	3.54	BOLD:ABX8716 BOLD:ACV1127
Lauxaniidae	<i>Minettia longipennis</i>	CDC (2)	0.96	1.45	BOLD:ACR0546 BOLD:ACR0548
Limoniidae	<i>Chionea lutescens</i>	CDC (2)	1.1	1.1	BOLD:ABV5195 BOLD:ADD1050
	<i>Euphyllidorea meigenii</i>	CDC (2)	1.91	4.88	BOLD:ABV4905 BOLD:ACU9122
Milichiidae	<i>Phyllomyza equitans</i>	CDC (2)	1.39	4.05	BOLD:ACB3455 BOLD:ACD3072
Muscidae	<i>Helina evecta</i>	CDC (3)	1.83	4.27	BOLD:AAE3133 BOLD:ACB3279 BOLD:ADB5997
	<i>Mydaea humeralis</i>	CDC (2)	1.95	5.84	BOLD:AAE0058 BOLD:ACD1934

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Mycetophilidae	<i>Boletina dispecta</i>	CDC (3)	9.01	11.2	BOLD:AAY5579 BOLD:AAY5580 BOLD:AAY5581
	<i>Brevicornu griseicolle</i>	CDC(2)	9.06	13.6	BOLD:ACU9474 BOLD:ABA1563
	<i>Brevicornu sericoma</i>	CDC (2)	1.99	4.58	BOLD:AAY6368 BOLD:ABA1564
	<i>Phronia obtusa</i>	CDC (2)	0.83	1.18	BOLD:AAY8505 BOLD:ACJ2989
	<i>Stigmatomeria crassicornis</i>	CDC (2)	0.56	1.86	BOLD:AAY6370 BOLD:ACU7541
	<i>Zygomia angusta</i>	CDC (3)	3.29	14.88	BOLD:AAY5526 BOLD:AAY5527 BOLD:ABW0168
	<i>Zygomia valida</i>	CDC (2)	9.51	14.5	BOLD:AAY5526 BOLD:ABW0168
	Pallopteridae	<i>Toxoneura aff. modesta</i>	CDC (2)	3.41	5.13
Phoridae		<i>Megaselia consetigera</i>	CDC (2)	0.65	2.63
	<i>Megaselia glabrifrons</i>	CDC (2)	0.66	1.78	BOLD:ACG3433 BOLD:ACI6910
	<i>Megaselia longicostalis</i>	CDC (3)	1.32	5.72	BOLD:AAG3263 BOLD:ADA4916 BOLD:AAG7025
	<i>Megaselia lutea</i>	CDC (2)	2.14	6.46	BOLD:AAG3351 BOLD:ACG3608
	<i>Megaselia nigriceps</i>	CDC (3)	0.76	7.16	BOLD:AAG7022 BOLD:AAY6384 BOLD:ACF7950
	<i>Megaselia pulicaria complex</i>	CDC (3)	5.85	11.96	BOLD:AAL9073 BOLD:AAP4698 BOLD:AAU8534
	<i>Megaselia rufa</i>	CDC (2)	1.83	8.31	BOLD:ACD9573 BOLD:ACD9606
	<i>Megaselia ruficornis</i>	CDC (2)	5.46	17.53	BOLD:ACF7708 BOLD:ACG4585
	<i>Megaselia sepulchralis</i>	CDC (2)	2.27	4.27	BOLD:ACF7622 BOLD:ACZ9853
	<i>Megaselia subpalpalis</i>	CDC (2)	1.05	2.17	BOLD:AAL9083 BOLD:ACZ7449
	<i>Megaselia tarsella</i>	CDC (3)	0.45	5.61	BOLD:ACE0332 BOLD:ACF7226
	Psychodidae	<i>Psychoda nr. albipennis</i>	CDC (2)	1.55	3.45

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Rhinophoridae	<i>Rhinomorinia sarcophagina</i>	CDC (2)	0.75	1.78	BOLD:ACD9526 BOLD:ACG3259
Sciaridae	<i>Bradysia brevispina</i>	CDC (2)	2.86	8.4	BOLD:ACE4845 BOLD:ACI5443
	<i>Bradysia inusitata</i>	CDC (2)	6.61	6.61	BOLD:ACE7273 BOLD:ACH4332
	<i>Bradysia praecox</i>	CDC (2)	1.09	2.35	BOLD:ACF3561 BOLD:ACU9870
	<i>Bradysia regularis</i>	CDC (2)	0.1	1.67	BOLD:ACC1391 BOLD:ACQ7807
	<i>Bradysia tilicola</i>	CDC (2)	2.87	6.03	BOLD:AAN6444 BOLD:ACP0919
	<i>Bradysia trivittata</i>	CDC (2)	0.57	3.57	BOLD:AAH3947 BOLD:ACB1143
	<i>Bradysiopsis vittata</i>	CDC (2)	2.24	4.62	BOLD:ACC1999 BOLD:ACR0949
	<i>Corynoptera grothae</i>	CDC (2)	4.75	9.36	BOLD:ACK0158 BOLD:ACO7236
	<i>Corynoptera luteofusca</i>	CDC (2)	8.16	11.8	BOLD:ACJ1951 BOLD:ACQ8494
	<i>Corynoptera polana</i>	CDC (2)	1.95	3.81	BOLD:ACF6941 BOLD:ACF7764
	<i>Corynoptera subtilis</i>	CDC (2)	2.91	6.26	BOLD:ACD5314 BOLD:ACT9420
	<i>Corynoptera tetrachaeta</i>	CDC (2)	4.16	4.16	BOLD:ACG5327 BOLD:ACL4032
	<i>Corynoptera tridentata</i>	CDC (2)	9.95	9.95	BOLD:ACJ1561 BOLD:ACJ9791
	<i>Epidapus atomarius</i>	CDC (2)	0.07	3.98	BOLD:ACD4767 BOLD:ACX3063
	<i>Leptosciarella fuscipalpa</i>	CDC (2)	5.24	9.24	BOLD:ACE2641 BOLD:ACQ8733
	<i>Leptosciarella scutellata</i>	CDC (3)	4.84	7.98	BOLD:ACD6061 BOLD:ACG4078 BOLD:ACI9623
	<i>Phyxiopsis degener</i>	CDC (2)	1.83	5.17	BOLD:ACE2293 BOLD:ACF9729
	<i>Scatopsciara neglecta</i>	CDC (2)	0.53	1.78	BOLD:ACC7986 BOLD:ACQ2637
	<i>Scatopsciara subciliata</i>	CDC (2)	1.93	4.32	BOLD:AAH4004 BOLD:ACA8369
	<i>Sciara hemerobioides</i>	CDC (2)	4.1	4.1	BOLD:ACQ8933 BOLD:ACR4627
	<i>Trichosia morio</i>	CDC (2)	0.78	3.99	BOLD:ACD5342 BOLD:ACO9950

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Simuliidae	<i>Simulium cryophilum</i>	CDC (2)	1.49	3.14	BOLD:ACU9243 BOLD:AAU1818
Sphaeroceridae	<i>Opacifrons coxata</i>	CDC (2)	6.41	14	BOLD:ACP2618 BOLD:ACP5793
	<i>Spelobia clunipes</i>	CDC (2)	2.89	6.93	BOLD:AAG7312 BOLD:ACF9400
Syrphidae	<i>Cheilosia albipila</i>	CDC (2)	2.51	6.88	BOLD:AAW3610 BOLD:AAZ1026
	<i>Cheilosia chrysocoma</i>	CDC (2)	3.69	3.69	BOLD:ABY6892 BOLD:ACJ5068
	<i>Cheilosia derasa</i>	CDC (2)	0.58	3.47	BOLD:AAY9044 BOLD:AAW3649
	<i>Cheilosia flavipes</i>	CDC (2)	8.79	8.79	BOLD:AAW3610 BOLD:AAY9045
	<i>Cheilosia impressa</i>	CDC (2)	1.95	5.74	BOLD:AAW3651 BOLD:AAW3615
	<i>Cheilosia lenis</i>	CDC (2)	3.85	7.86	BOLD:AAY8876 BOLD:AAY8875
	<i>Cheilosia mutabilis</i>	CDC (2)	1.94	2.74	BOLD:AAY9746 BOLD:AAY9747
	<i>Cheilosia personata</i>	CDC (2)	1.35	1.88	BOLD:ACH1700 BOLD:ACX0819
	<i>Cheilosia proxima</i>	CDC (3)	3.28	6.91	BOLD:AAW3607 BOLD:AAW3651 BOLD:ABY8734
	<i>Cheilosia vernalis-agg.</i>	CDC (2)	2.07	3.84	BOLD:ACF0974 BOLD:ACJ5218
	<i>Eupeodes nitens</i>	CDC (2)	3.97	3.97	BOLD:AAB2384 BOLD:ACH1529
	<i>Melanogaster nuda</i>	CDC (2)	0.81	2.44	BOLD:AAY8880 BOLD:ACH5745
	<i>Merodon rufus</i>	CDC (2)	0.68	1.09	BOLD:ADI8358 BOLD:AAQ1380
	<i>Paragus pecchiolii</i>	CDC (2)	0.96	4.86	BOLD:ABA3664 BOLD:ACG8255
	<i>Parasyrphus punctulatus</i>	CDC (2)	1.11	2.65	BOLD:AAZ4514 BOLD:ACG4772
	<i>Pipiza noctiluca</i>	CDC (2)	1.54	3.92	BOLD:AAL4100 BOLD:ACG4983
	<i>Platycheirus albimanus</i>	CDC (2)	0.37	3.01	BOLD:AAL7898 BOLD:ACJ4919
	<i>Sericomyia lappona</i>	CDC (2)	2.06	3.9	BOLD:AAB1553 BOLD:ACH1641

(Continues)

TABLE 2 (Continued)

Family	Species	CDC rank	Mean intraspecific variation	Max. intraspecific variation	BIN
Tabanidae	<i>Tabanus bromius</i>	CDC (2)	2.04	2.93	BOLD:AAF3864 BOLD:ACJ5745
	<i>Tabanus glaucopis</i>		3.27	4.43	BOLD:AAF3858 BOLD:AAF3859
Tachinidae	<i>Actia dubitata</i>	CDC (2)	2.36	2.36	BOLD:ACP3766 BOLD:ACH1972
	<i>Bessa selecta</i>	CDC (2)	1.45	2.38	BOLD:ADK1760 BOLD:AAW3422
	<i>Cyzenis albicans</i>	CDC (2)	1.18	2.18	BOLD:ACB0896 BOLD:ACM9631
	<i>Kirbya moerens</i>	CDC (2)	1.22	1.86	BOLD:ACJ2730 BOLD:ACB0261
	<i>Peribaea fissicornis</i>	CDC (2)	2.22	8.17	BOLD:ACH1961 BOLD:ACJ2910
	<i>Phorinia aurifrons</i>	CDC (2)	3.76	11.2	BOLD:ADK4076 BOLD:ACB0795

inaccessible because of the lack of specialists. By contrast, within just a few years, this study provided an interim taxonomic identification system for half of the German Diptera fauna. Although half these species still lack a Linnean name, their BIN assignments are useful “taxonomic handles” for work in ecology, conservation biology and other biodiversity research (see Geiger, Moriniere, et al., 2016). The study demonstrates the efficiency of DNA barcoding in the identification of Central European Diptera, reinforcing the results of earlier studies. DNA barcode coverage was nearly complete for many species-poor families (e.g., Megamerinidae, Opetiidae, Phaeomyiidae) known from Germany and the incidence of “dark taxa” in these families was low. Overall, there was a strong inverse relationship between the number of “dark taxa” and average body size: the smaller the average body size of a family, the higher the ratio of “dark taxa” (Figure 3). Among families with the smallest body sizes, our results suggest a higher incidence of cryptic diversity and overlooked species, indicating the number of dipteran species in Germany is likely to be much higher than previously recognized. Among families, such as the “Iteaphila group” (Empidoidea; see Meyer & Stark, 2015), Milichiidae and Trichoceridae, DNA barcoding indicates unexpectedly high levels of diversity as their BIN count is substantially higher than the number of species known from Germany (Schumann et al., 1999). The Cecidomyiidae represent the most impressive example, as we encountered 930 BINs while only 886 species are known from Germany (Table 1; Jaschhof, 2009; Schumann et al., 1999). As such, they represent by far the largest family of Diptera in the studied area. When compared with the other families in Figure 2b, it is clear that the Cecidomyiidae show a lower average interspecific variation, indicating an increased evolutionary rate. As already proposed by Hebert et al. (2016), the extraordinary species—or BIN

number—might be linked to their unusual mode of reproduction, namely haplodiploidy. Here, paternally inherited genomes of diploid males are inactivated during embryogenesis (Normark, 2003). The phenomenon of haplodiploidy is known from Hymenoptera (Branstetter et al., 2018; Hansson & Schmidt, 2018) another group known to be rate accelerated, but it is largely unstudied throughout Diptera. Despite the need for more study, we conclude the true diversity of Diptera in Germany, Europe and the world has been seriously underestimated, a conclusion reached in several other studies (Erwin, 1982; Hebert et al., 2016; May, 1988; Ødegaard, 2000).

Within the metabarcoded Malaise trap samples collected over just one season in one region of Germany, we identified 1,735 OTUs with a sequence identity higher than 97% to a dipteran record. This result indicates that metabarcoding analysis of bulk samples will be a valuable approach for assessing the diversity of Diptera in Germany (Appendix S2). Variation in overall biodiversity between sampling sites as well as annual phenologies of certain taxa can easily be visualized using presence–absence maps (Figure 2). This will be a useful feature for comparison of large data sets and for monitoring beneficial or pest insects (L. A. Hardulak et al. in preparation). Although a third of the OTUs within the metabarcoding data set could not be assigned to a Linnean species, interim names, such as BIN assignments, make it possible to compare sampling sites. OTUs with lower sequence similarities (<97%) to known taxa can be used to track “dark taxa,” those species missing from the reference sequence library. Although such taxa may only be assigned to a family or genus, their records are still valuable for evaluating differences between samples from various environments or sites. At present, dipteran species, although overall present in very high numbers, are extremely underrepresented within environmental assessments in Germany: ~2,000

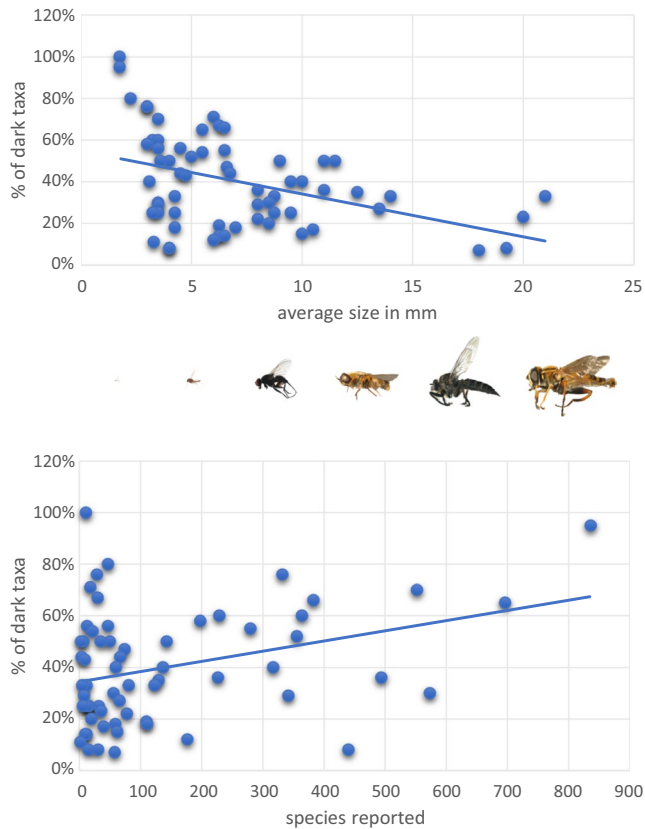


FIGURE 3 Illustration of the relationship between the percentage of “dark taxa” and average body size (mm), and in number of species reported for a family [Colour figure can be viewed at wileyonlinelibrary.com]

species from 11 families (Asilidae, Atelestidae, Ceratopogonidae, Chaoboridae, Dixidae, Dolichopodidae, Empididae, Hybotidae, Psychodidae, Syrphidae, Thaumaleidae) are included in the German red list (Gruttke et al., 2016), but not a single dipteran species is listed among the ~1,000 species being protected according to the European Flora-Fauna-Habitat directive (Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora, 1992), which ensures the conservation of a wide range of rare, threatened or endemic animal and plant species in Europe. The present study is a first step to permit the proper evaluation of the status of dipterans and the potential designation of some species as targets for conservation action.

Previous studies have shown the great potential of metabarcoding for biotic assessments in various contexts, including Malaise trap surveys (Morinière et al., 2016), biosurveillance of invasive and pest species (Ashfaq & Hebert, 2016; L. A. Hardulak et al. in prep), macrozoobenthos sampling for assessing water and stream health (Elbrecht & Leese, 2015; Serrana, Miyake, Gamboa, & Watanabe, 2018), faeces analyses for dietary inference (De Barba et al., 2014; Hawlitschek, Fernández-González, Balmori-de la Puente, & Castresana, 2018), species identification for forensic entomology (Chimeno et al., 2018) and for soil biology (Oliverio, Gan, Wickings, & Fierer, 2018). This approach combines the advantages of DNA barcoding, namely the capacity to identify any life stage, body fragment or even trace DNA in the

environment, with the ability of high-throughput sequencers to analyse millions of DNA fragments and thousands of specimens at a time. The application of this technology to biodiversity assessments will certainly enable species surveys at larger scales, shorter time and lower costs compared with classical morphological approaches (Douglas et al., 2012; Hajibabaei et al., 2011; Ji et al., 2013; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). The ability to upscale biomonitoring projects is crucial, as is the need to generate biodiversity data fast and with less dependence on often unavailable taxonomic experts. Additionally, data generated by ongoing metabarcoding studies, such as from annual national biomonitoring projects, can be combined and reanalysed, producing recursively more comprehensive species lists, when new reference sequences become available or when taxonomic annotations have been improved. While biomonitoring studies have traditionally employed small subsets of indicator species, metabarcoding will enable comprehensive assessments of biodiversity because even “dark taxa” can be tracked. Furthermore, metabarcoding can enhance the ability to rapidly assess biodiversity patterns to identify regions that are of most significance for conservation.

Although this project aimed to develop a comprehensive DNA barcode library, resource constraints meant that only half the specimens sorted to a family or better taxonomy could be analysed. It is certain that many species and genera currently absent from the reference library remain within this sorted material, making the remaining samples a valuable resource for future extension of the reference library. Our work has also highlighted the potential of DNA barcoding and metabarcoding to aid efforts to conserve the world's fauna. Because these technologies greatly enhance our ability to identify, and thus conserve, biodiversity, they should be pursued—vigorously. As our study has provided several thousands of voucher-based DNA barcode records, we invite the global community of dipteran taxonomists to improve identifications for the many “dark taxa” encountered in our study by identifying these vouchers using reverse taxonomic approaches.

The present study represents an important component of a decade of work directed toward creating a comprehensive DNA barcode library for German animal species. Because Diptera represents the largest and taxonomically most challenging insect order, they have received less attention than other orders (e.g., Lepidoptera, Coleoptera, freshwater orders) with lower species richness and more taxonomic expertise. Our work on Diptera has not only confirmed that this order is extremely species-rich, but also that several of its most diverse families include a large proportion of “dark taxa.” The present study represents a cornerstone for subsequent research on these unexplored groups of Diptera. This paper presents the results of one of the most comprehensive studies on DNA barcoding of Diptera, with a coverage of over 80% of German families. Due to the general lack of taxonomy in many groups of Diptera, only a fraction of the specimens could be identified to species level. Most specimens for the study were obtained from just three Malaise traps deployed as a component of the Global Malaise programme (see <http://biodiversitygenomics.net/projects/gmp/>). Voucher specimens are still being identified by external specialists, a process that

TABLE 3 All cases of low intraspecific sequence variation at COI; cases of BIN sharing (BS)

Family	Species	BS rank	Mean intraspecific variation	Max intraspecific	BIN
Anthomyiidae	<i>Hylemya nigrimana</i>	BS (2)	0.34	0.52	BOLD:ABA6492
	<i>Hylemya vagans</i>		0.37	1.58	
Calliphoridae	<i>Calliphora loewi</i>	BS (2)	1.07	1.07	BOLD:AAB6579
	<i>Calliphora vicina</i>		0.84	2.59	
	<i>Lucilia caesar</i>	BS (3)	0.95	3.07	BOLD:AAA7470
	<i>Lucilia caesarillustris</i>		0.7	2.43	
	<i>Lucilia illustris</i>		N/A	0	
Dolichopodidae	<i>Medetera</i> <i>petrophiloides</i>	BS (2)	0.35	1.22	BOLD:ACA1124
	<i>Medetera truncorum</i>		N/A	0	
	<i>Sphyrotarsus</i> <i>argyrostomus</i>	BS (2)	0.91	1.37	BOLD:ADB6106
	<i>Sphyrotarsus</i> <i>hygrophilus</i>		N/A	0	
Empididae	<i>Kowarzia madicola</i>	BS (2)	0	0	BOLD:ACJ7236
	<i>Kowarzia tenella</i>		5.39	10.8	
	<i>Kowarzia barbatula</i>	BS (2)	4.8	11.3	BOLD:ACJ6935
	<i>Kowarzia tenella</i>		5.39	10.8	
Ephydriidae	<i>Allotrichoma bezzii</i>	BS (4)	0.13	0.31	BOLD:ACF1575
	<i>Allotrichoma filiforme</i>		0.08	0.15	
	<i>Allotrichoma laterale</i>		6.44	6.44	
	<i>Allotrichoma schumanni</i>		0	0	
	<i>Ephydra macellaria</i>	BS (3)	N/A	0	BOLD:AAG2729
	<i>Ephydra murina</i>		N/A	0	
	<i>Ephydra riparia</i>		2.83	2.83	
	<i>Hydrellia nigricans</i>	BS (2)	0.23	0.31	BOLD:ABA8624
	<i>Hydrellia subalbiceps</i>		0.31	0.46	
	<i>Notiphila cinerea</i>	BS (2)	0.26	0.46	BOLD:ABA7513
	<i>Notiphila graecula</i>		0	0	
	<i>Notiphila riparia</i>	BS (2)	0.16	0.35	BOLD:AAX5585
	<i>Notiphila subnigra</i>		0.41	0.62	
	<i>Philygria flavipes</i>	BS (2)	1.19	2.03	BOLD:ACK3229
	<i>Philygria punctatonervosa</i>		0.15	0.15	
	<i>Psilopa compta</i>	BS (2)	0.08	0.16	BOLD:AAG6948
<i>Psilopa nitidula</i>		0.38	0.77		
Iteaphila-group	<i>Anthepiscopus indet.</i>	BS (2)	0.14	0.48	BOLD:ACD9492
	<i>Anthepiscopus</i> sp. 1		11.3	11.3	
	<i>Anthepiscopus</i> sp. 1	BS (2)	11.3	11.3	BOLD:ACJ7111
	<i>Anthepiscopus</i> sp. 4		0.91	1.58	
	<i>Iteaphila</i> sp. 1	BS (2)	0.07	0.15	BOLD:ACD3033
	<i>Iteaphila</i> sp. 2		4.49	9.77	
Lonchopteridae	<i>Lonchoptera lutea</i>	BS (2)	0.39	1.09	BOLD:ABX0277
	<i>Lonchoptera nitidifrons</i>		N/A	0	

(Continues)

TABLE 3 (Continued)

Family	Species	BS rank	Mean intraspecific variation	Max intraspecific	BIN
Muscidae	<i>Hydrotaea dentipes</i>	BS (2)	2.15	9.78	BOLD:AAZ9882
	<i>Hydrotaea similis</i>		0	0	
Mycetophilidae	<i>Boletina gripha</i>	BS (2)	0.52	0.9	BOLD:AAF6783
	<i>Boletina groenlandica</i>		N/A	0	
	<i>Mycetophila distigma</i>	BS (2)	N/A	0	BOLD:AAV8340
	<i>Mycetophila flava</i>		0.19	0.19	
	<i>Zygomyia angusta</i>	BS (2)	4.6	15.4	BOLD:AAV5526
	<i>Zygomyia valida</i>		14.5	14.5	
	<i>Zygomyia angusta</i>	BS (2)	4.6	15.4	BOLD:ABW0168
Phoridae	<i>Triphleba bicornuta</i>	BS (2)	N/A	0	BOLD:ACF0365
	<i>Triphleba</i> sp. BOLD:ACF0365		0.66	1.22	
Sarcophagidae	<i>Sarcophaga depressifrons</i>	BS (2)	0	0	BOLD:ABV4597
	<i>Sarcophaga haemorrhoea</i>		0.47	0.7	
Simuliidae	<i>Simulium balcanicum</i>	BS (2)	N/A	0	BOLD:AAM4036
	<i>Simulium equinum</i>		1.59	2.66	
Syrphidae	<i>Baccha elongata</i>	BS (6)	N/A	0	BOLD:ABA3006
	<i>Baccha elongata</i> s.s.		0	0	
	<i>Baccha obscuripennis</i>		1.23	2.02	
	<i>Baccha</i> sp. BOLD:ABA3006		N/A	0	
	<i>Brachypalpus laphriiformis</i>		0.56	1.54	BOLD:AAV9039
	<i>Brachypalpus valgus</i>		N/A	0	
	<i>Cheilosia albipila</i>	BS (2)	2.51	6.88	BOLD:AAW3610
	<i>Cheilosia flavipes</i>		8.79	8.79	
	<i>Cheilosia barbata</i>	BS (3)	0.1	0.3	BOLD:AAW3615
	<i>Cheilosia impressa</i>		1.95	5.74	
	<i>Cheilosia</i> sp. BOLD:AAW3615		0	0	
	<i>Cheilosia chloris</i>	BS (8)	0.57	1.42	BOLD:ACF0974
	<i>Cheilosia chlorus</i>		0.12	0.18	
	<i>Cheilosia chlorus</i> -group		N/A	0	
	<i>Cheilosia fraterna</i>		0.55	0.87	
	<i>Cheilosia melanura</i>		0.06	0.2	
	<i>Cheilosia ruficollis</i>		N/A	0	
	<i>Cheilosia</i> sp. BOLD:ACF0974		0.47	0.71	
	<i>Cheilosia vernalis</i> -agg.		2.07	3.84	

(Continues)

TABLE 3 (Continued)

Family	Species	BS rank	Mean intraspecific variation	Max intraspecific	BIN
	<i>Cheilosia crassiseta</i>	BS (6)	N/A	0	BOLD:AAW3647
	<i>Cheilosia impudens</i>		N/A	0	
	<i>Cheilosia nigripes</i>		N/A	0	
	<i>Cheilosia</i> sp. BIOUG17085-G07		0.75	1.94	
	<i>Cheilosia</i> aff. <i>grisella</i>		N/A	0	
	<i>Cheilosia antiqua</i>		N/A	0	
	<i>Cheilosia faucis</i>	BS (2)	0.7	0.88	BOLD:AAW8874
	<i>Cheilosia nivalis</i>		0	0	
	<i>Cheilosia grisella</i>	BS (2)	0.18	0.18	BOLD:AAW3619
	<i>Cheilosia pubera</i>		0.49	0.87	
	<i>Cheilosia canicularis</i>	BS (2)	0.08	0.38	BOLD:ACI2500
	<i>Cheilosia montana</i>		N/A	0	
	<i>Cheilosia carbonaria</i>	BS (2)	0.37	0.37	BOLD:AAW8876
	<i>Cheilosia lenis</i>		3.85	7.86	
	<i>Chrysotoxum bicinctum</i>	BS (2)	0.86	2	BOLD:AAJ0967
	<i>Chrysotoxum festivum</i>		0	0	
	<i>Dasysyrphus hilaris</i>	BS (3)	0.35	0.52	BOLD:AAA7375
	<i>Dasysyrphus laskai</i>		0.3	0.3	
	<i>Dasysyrphus venustus</i>		N/A	0	
	<i>Dasysyrphus lenensis</i>	BS (3)	0.58	0.58	BOLD:AAB2865
	<i>Dasysyrphus pinastri</i>		1.25	2.1	
	<i>Dasysyrphus</i> sp. BOLDAAB2865		0.12	0.17	
	<i>Eupeodes bucculatus</i>	BS (5)	1.14	3.13	BOLD:AAB2384
	<i>Eupeodes nielsenii</i>		0.15	0.37	
	<i>Eupeodes nitens</i>		3.97	3.97	
	<i>Eupeodes</i> sp. BOLDAAB2384		0.39	1.03	
	<i>Eupeodes luniger</i>		0.53	1.05	
	<i>Melanogaster aerosa</i>	BS (2)	N/A	0	BOLD:AAQ4015
	<i>Melanogaster hirtella</i>		0.26	0.7	
	<i>Melanostoma dubium</i>	BS (7)	0	0	BOLD:AAB2866
	<i>Melanostoma mellinum</i>		0.58	1.21	
	<i>Melanostoma</i> <i>mellinum</i> -agg.		N/A	0	
	<i>Melanostoma scalare</i>		0.49	1.3	
	<i>Melanostoma</i> sp. A		0	0	
	<i>Melanostoma</i> sp. B		0.11	0.16	
	<i>Melanostoma</i> sp. BOLDAAB2866		0.63	2.69	
	<i>Merodon avidus</i>	BS (2)	N/A	0	BOLD:AAQ1379
	<i>Merodon avidus</i> B		0.55	1.03	

(Continues)

TABLE 3 (Continued)

Family	Species	BS rank	Mean intraspecific variation	Max intraspecific	BIN
	<i>Paragus aff. haemorrhous</i>	BS (5)	N/A	0	BOLD:ABZ4619
	<i>Paragus constrictus</i>		N/A	0	
	<i>Paragus haemorrhous</i>		0.26	0.87	
	<i>Paragus sp.</i> BOLDABZ4619		0.07	0.37	
	<i>Paragus tibialis</i>		N/A	0	
	<i>Paragus majoranae</i>	BS (2)	0.87	0.87	BOLD:ABA3664
	<i>Paragus pecchiolii</i>		0.96	4.86	
	<i>Parasyrphus lineola</i>	BS (2)	0.19	0.39	BOLD:ACE7140
	<i>Parasyrphus vittiger</i>		0.63	1.44	
	<i>Pipiza bimaculata</i>	BS (4)	N/A	0	BOLD:AAL4100
	<i>Pipiza nocticula</i>		N/A	0	
	<i>Pipiza noctiluca-agg.</i>		N/A	0	
	<i>Pipiza sp.</i> BOLDAAL4100		0.55	1.65	
	<i>Platycheirus angustatus</i>	BS (3)	0.84	2.02	BOLD:ACF4733
	<i>Platycheirus europaeus</i>		1.95	1.95	
	<i>Platycheirus sp.</i> BOLDACF4733		0.21	1.15	
	<i>Platycheirus clypeatus</i>	BS (5)	0.38	0.88	BOLD:AAA9506
	<i>Platycheirus fulviventris</i>		1.04	1.04	
	<i>Platycheirus occultus</i>		0.51	1.04	
	<i>Platycheirus perpallidus</i>		N/A	0	
	<i>Platycheirus sp.</i> BOLDAAA9506		0.9	2.03	
	<i>Platycheirus melanopsis</i>	BS (2)	0.25	0.62	BOLD:AAP0412
	<i>Platycheirus tatricus</i>		N/A	0	
	<i>Platycheirus nielseni</i>	BS (3)	0	0	BOLD:AAC6630
	<i>Platycheirus peltatus</i>		0.24	0.72	
	<i>Platycheirus peltatus-group</i>		N/A	0	
	<i>Platycheirus scutatus</i>	BS (3)	0.05	0.19	BOLD:AAG4665
	<i>Platycheirus scutatus-group</i>		0.44	0.71	
	<i>Platycheirus splendidus</i>		N/A	0	
	<i>Scaeva dignota</i>	BS (2)	N/A	0	BOLD:AAF2374
	<i>Scaeva pyrastris</i>		0.25	0.91	
	<i>Scaeva pyrastris</i>	BS (2)	0.25	0.91	BOLD:AAF2374
	<i>Scaeva dignota</i>		N/A	0	
	<i>Sericomyia lappona</i>	BS (2)	2.06	3.9	BOLD:AAB1553
	<i>Sericomyia silentis</i>		0.05	0.24	

(Continues)

TABLE 3 (Continued)

Family	Species	BS rank	Mean intraspecific variation	Max intraspecific	BIN
	<i>Sphaerophoria bankowskiae</i>	BS (9)	N/A	0	BOLD:AAA7374
	<i>Sphaerophoria infuscata</i>		0.24	0.38	
	<i>Sphaerophoria interrupta</i>		0	0	
	<i>Sphaerophoria interrupta-group</i>		0.49	0.75	
	<i>Sphaerophoria philanthus</i>		N/A	0	
	<i>Sphaerophoria rueppellii</i>		N/A	0	
	<i>Sphaerophoria</i> sp. BOLDAAA7374		0.31	6.54	
	<i>Sphaerophoria taeniata</i>		N/A	0	
	<i>Sphaerophoria virgata</i>		N/A	0	
	<i>Sphegina montana</i>	BS (2)	N/A	0	BOLD:ABX4867
	<i>Sphegina sibirica</i>		0.4	0.41	
	<i>Temnostoma apiforme</i>	BS (2)	0.52	0.52	BOLD:AAV6543
	<i>Temnostoma meridionale</i>		0.35	0.52	
Stratiomyidae	<i>Beris geniculata</i>	BS (2)	N/A	0	BOLD:AAW3384
	<i>Beris morrisii</i>		0.48	1.47	
Tachinidae	<i>Lydella stabulans</i>	BS (2)	0.12	0.44	BOLD:AAP8653
	<i>Lydella thompsoni</i>		0.68	1.31	
	<i>Medina luctuosa</i>	BS (3)	1.35	1.35	BOLD:AAG6902
	<i>Medina melania</i>				

is labour intensive and time consuming, especially for taxonomically challenging taxa.

Our study presents results from one of the most comprehensive DNA barcoding projects on Diptera, a megadiverse, and, almost certainly, most diverse insect order. Our results strongly support the conclusion that DNA barcoding will enable the discovery and identification of most dipteran taxa. Some cases of low interspecific variation were observed in the Syrphidae, Tachinidae and Calliphoridae where additional markers may be needed for species identification (Haarto & Ståhls, 2014; Nelson et al., 2012; Pohjoismäki et al., 2016; Whitworth et al., 2007). However, in most cases, there was congruence between BINs and species defined by traditional morphological methods, supporting the use of DNA barcoding as a species identification tool for Diptera. This conclusion and the finding that many of the species we encountered represent “dark taxa” indicates that DNA barcoding will speed the discovery of genetic entities that will eventually gain recognition as biological species. Our data release aims at making these results accessible to the scientific community through a public data portal so they will be available for taxonomic research, biodiversity studies and barcoding initiatives at national and international levels.

In summary, the application of DNA barcoding enabled a comprehensive assessment of German Diptera, including several highly diverse families, which would otherwise have been excluded due to a lack of taxonomic expertise. By selecting morphospecies from the pool of specimens collected by the year-long deployment of Malaise traps in ecosystems ranging from alpine to lowland settings, we constructed a reference library for most dipteran families known from Germany. Due to the diversity of sampling sites, we encountered a wide range of taxa from microendemics to wide-ranging generalists with varied seasonal phenologies. We emphasize that DNA barcoding and the resultant barcode reference libraries provide an easy, intuitive introduction to molecular genetics, an approach accessible to undergraduate students in a way that genome sequencing is not. Because DNA barcoding workflows have been implemented in many laboratories around the world and because current primer sets reliably generate amplicons, this method is ideal for educational purposes. Democratization of the method, the analytical tools and data through the BOLD database (Ratnasingham & Hebert, 2007) further facilitates its use in real world situations. The approach has the additional advantage of allowing students to not only work with “real organisms,” but also to solve long-standing taxonomic puzzles.

The latter work leads students to probe the historical literature, to regale in past expeditions in search of type locations or type material, and potentially to end the chase by describing a new species. However, it is critical that senior taxonomists and professors need to recognize these possibilities and encourage their students to embrace this approach as it offers such a clear solution to the taxonomic impediment.

Germany has a tradition of more than 250 years of entomological research, and the number of Diptera species recorded is the highest for any European country comprising almost half of the European fauna. Despite this long effort, knowledge of its Diptera fauna must be regarded as fragmentary. In accordance with the species accumulation curve presented by Pape (2009) for the British Isles, additional species were revealed from current collecting efforts for practically every species-rich family. Recording “new” species is slowed by the lack of experts for many of these families as well as by the lack of up-to-date identification keys. A particularly important result of our study is that the estimated number of dipteran species in Germany is certainly much higher than formerly thought. High proportions of unrecorded species were evident for the Agromyzidae, Anthomyiidae, Cecidomyiidae, Ceratopogonidae, Chironomidae, Chloropidae, Phoridae, Sciaridae and Sphaeroceridae, and to a lesser extent for the Empidoidea, Limoniidae, Mycetophilidae and others. Further studies point to an enormous under-estimation of the species diversity in the Cecidomyiidae (Borkent et al., 2018; Hebert et al., 2016). Although our data do not allow for an accurate projection for the size of the total species numbers, it seems quite likely that this single family contains thousands of unrecorded species in Germany.

ACKNOWLEDGEMENTS

We express our extreme gratitude to the taxonomists, citizen scientists and nature enthusiasts who supported this campaign by collecting thousands of dipteran species. The realization of this mammoth task would not have been possible without the help of Adaschkiewitz, W., Assum, Babiy, P. P., Bährmann, R., Baranov, V., Beermann, A., Behounek, G., Belleuvre, N., Blick, T., Bolz, R., Brandt-Floren, C., Brenzinger, S., Brown, A., Burmeister E. G., Charabidze, D., Chimeno, C., Claussen, C., Dettinger-Klemm, A., Diller, E., Drozd, P., Dunz, A., Duschl Miesbach, M., Dworschak, W., Eckert, I., Esser, J., Fahldieck, M., Fiedler, Fittkau, E. J. (+), Flügel, H. J., Forster, W., Forstner, P., Fünfstück, J., Fütterer, S., Fuhrmann, S., Gabriel, I., Gammelmo, O., Gerecke, R., Glaw, F., Guggemoos, T., Haberberger, S., Hable, J., Haeselbarth, E., Hansen, L. O., Hartop, E., Hawlitschek, O., Heller, K., Hessing, R., Hierlmeier, V., Hilbig, D., Höglund, J. R., Höhne, F., Honold, D., Jaschof, M., Jon, T., Kamin, J., Kappert, J., Kehlmaier, C., Kilian, D., Kirsch, H., Kjaerandsen, J., Kleiner, M., Koehler, F., Koehler, J., Kölbl, N., Koenig, T., Kolbeck, H., Kraus, G., Kraus, W., Kuehbandner, M., Kuehlhorn, F., Kuhlmann, M., Kusdas, K., Kvifte, G. M., Lindner, S., Loenneve, O. J., Lucas, W., Mair, K., Mandery, K., Mengual, X., Merkel-Wallner, G., Mortelmans, J., Müller, Mueller,

R., Mueller-Kroehling, S., Neumann, C., Olberg, S., Olsen, K. M., Pavlova, A., Pötter, L., Steven, M., Plassmann, E., Podhorna, J., Prescher, S., Prozeller, M., Pushkar, V., Reckel, F., Rehm, T., Reiff, N., Reimann, A., Reiso, S., Rennwald, K., Richter, B., Riedel, G., Rohmoser, S., Rozo, P., Rudzinski, H. G., Ruf, T., Salomon, C., Schacht, W. (+), Schäfer, A., Scheingraber, M., Scheler, Schmieder, F., Schödl, M., Schoenitzer, K., Schrott, S., Schubart, C., Schubert, C., Schubert, W., Schwarz, K., Schwemmer, R., Sedlak, G., Segerer, A., Sellmayer, G., Spelda, J., Spies, M., Ssymank, A., Steigemann, U., Stenhouse, G., Stoecklein, F., Stuke, J. H., Stur, E., Tänzler, K., Tänzler, R., Telfer, A., Toussaint, C., Toussaint, E., Treiber, R., Troester, M., v. Tschirnhaus, M., v. d. Dunk, K., Vallenduuk, H., van Ess, L., Velterop, J., Voith, J., Volf, M., Wachtel, F., Wagner, R., Warncke, K., Weber, D., Weiffenbach, H., Weigand, A. M., Weixler, K., Wiedenbrug, S., Windmaisser, T., Winqvist, K., Woodley, N., Zahn, A. The project was funded by grants from the Bavarian State Ministry of Science and the Arts (2009-2018: Barcoding Fauna Bavarica, BFB) and the German Federal Ministry of Education and Research (German Barcode of Life: 2012-2019, BMBF FKZ 01L11101 and 01L11501). We are grateful to the team at the Centre for Biodiversity Genomics in Guelph (Ontario, Canada) for their great support and help and particularly to Sujeewan Ratnasingham for developing the BOLD database (BOLD; www.boldsystems.org) infrastructure and the BIN management tools. The sequencing work was supported, in part, by funding from the Government of Canada to Genome Canada through the Ontario Genomics Institute, while the Ontario Ministry of Research and Innovation and NSERC supported development of the BOLD informatics platform. We also thank all the students who assisted in the ZSM-barcoding projects (barcoding-zsm.de) for picking countless legs and photographing countless specimens. We would like to express our thanks to Dr Vedran Bozicevic (AIM GmbH, Munich, Germany) for assisting with the KRONA file to enable inspection of BIN diversity. Fieldwork permits were issued by the responsible state environmental offices of Bavaria (Bayerisches Staatsministerium für Umwelt und Gesundheit, Munich, Germany, project: “Barcoding Fauna Bavarica”; confirmed by the regional governments “Bezirksregierungen”) and Rhineland-Palatinate (“Struktur- und Genehmigungsdirektion Nord”, Axel Schmidt [Koblenz, Germany]).

AUTHOR CONTRIBUTIONS

Obtained funding: G.H., W.W., A.H., P.D.N.H. Collected the samples: D.D., B.R. Conceived and designed the experiments: J.M., L.A.H., M.G., B.R. Analysed the data: J.M., L.A.H., M.F.G., L.R., B.R. Wrote the paper: J.M., L.A.H., S.S., M.B., D.D. Contributed (additions/corrections) to the manuscript: P.D.N.H., A.H., M.F.G., L.H., G.H.

DATA AVAILABILITY

All specimen data have been made publicly available within the BOLD workbench - a DOI for the dataset has been added.

ORCID

Jérôme Morinière  <https://orcid.org/0000-0001-9167-6409>

Stefan Schmidt  <https://orcid.org/0000-0001-5751-8706>

REFERENCES

- Ashfaq, M., & Hebert, P. D. N. (2016). DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. *Genome*, 59(11), 933–945. <https://doi.org/10.1139/gen-2016-0024>
- Ashfaq, M., Hebert, P. D. N., Mirza, J. H., Khan, A. M., Zafar, Y., & Mirza, M. S. (2014). Analyzing mosquito (Diptera: Culicidae) diversity in Pakistan by DNA barcoding. *PLoS ONE*, 9(5), e97268. <https://doi.org/10.1371/journal.pone.0097268>
- Astrin, J. J., Höfer, H., Spelda, J., Holstein, J., Bayer, S., Hendrich, L., ... Muster, C. (2016). Towards a DNA barcode reference database for spiders and harvestmen of Germany. *PLoS ONE*, 11(9), e0162624 (24 pp, supplements). <https://doi.org/10.1371/journal.pone.0162624>
- Bickel, D., T. Pape, & R. Meier (Eds.) (2009). *Diptera diversity: status, challenges and tools* (pp. 459). Leiden, Netherlands: Brill.
- Borkent, A., Brown, B., Adler, P. H., Amorim, D. D. S., Barber, K., Bickel, D., ... Capellari, R. S. (2018). Remarkable fly (Diptera) diversity in a patch of Costa Rican cloud forest. *Zootaxa*, 4402(1), 53–90.
- Branstetter, M. G., Childers, A. K., Cox-Foster, D., Hopper, K. R., Kapheim, K. M., Toth, A. L., & Worley, K. C. (2018). Genomes of the Hymenoptera. *Current Opinion in Insect Science*, 25, 65–75. <https://doi.org/10.1016/j.cois.2017.11.008>
- Brix, S., Leese, F., Riehl, T., & Kihara, T. C. (2015). A new genus and new species of Desmosomatidae Sars, 1897 (Isopoda) from the eastern South Atlantic abyss described by means of integrative taxonomy. *Marine Biodiversity*, 45(1), 7–61. <https://doi.org/10.1007/s12526-014-0218-3>
- Carew, M. E., Pettigrove, V., Cox, R. L., & Hoffmann, A. A. (2007). DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae). *Journal of the North American Benthological Society*, 26(4), 587–600. <https://doi.org/10.1899/06-120.1>
- Carew, M. E., Pettigrove, V., & Hoffmann, A. A. (2005). The utility of DNA markers in classical taxonomy: Using cytochrome oxidase I markers to differentiate Australian *Cladopelma* (Diptera: Chironomidae) midges. *Annals of the Entomological Society of America*, 98(4), 587–594.
- Chimeno, C., Morinière, J., Podhorna, J., Hardulak, L., Hausmann, A., Reckel, F., ... Haszprunar, G. (2018). DNA barcoding in forensic entomology – Establishing a DNA reference library of potentially forensic relevant arthropod species. *Journal of Forensic Sciences*, 64(2), 593–601. <https://doi.org/10.1111/1556-4029.13869>
- Chivian, E., & Bernstein, A. (Eds.) (2008). *Sustaining life: How human health depends on biodiversity*. Oxford, UK: Oxford University Press.
- Cranston, P. S., Ang, Y. C., Heyzer, A., Lim, R. B. H., Wong, W. H., Woodford, J. M., & Meier, R. (2013). The nuisance midges (Diptera: Chironomidae) of Singapore's Pandan and Bedok reservoirs. *Raffles Bulletin of Zoology*, 61(2), 779–793.
- Cruaud, P., Rasplus, J. Y., Rodriguez, L. J., & Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific Reports*, 7, 41948. <https://doi.org/10.1038/srep41948>
- Cywinska, A., Hunter, F. F., & Hebert, P. D. (2006). Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology*, 20(4), 413–424. <https://doi.org/10.1111/j.1365-2915.2006.00653.x>
- Dathe, H. H., & Blank, S. M. (2004). Nachträge zum Verzeichnis der Hautflügler Deutschlands, Entomofauna Germanica Band 4 (Hymenoptera). (1). *Entomologische Nachrichten Und Berichte*, 48(3–4), 179–182.
- De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14(2), 306–323. <https://doi.org/10.1111/1755-0998.12188>
- de Carvalho, M. R., Bockmann, F. A., Amorim, D. S., Brandão, C. R. F., de Vivo, M., de Figueiredo, J. L., ... Nelson, G. J. (2007). Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. *Evolutionary Biology*, 34(3), 140–143. <https://doi.org/10.1007/s11692-007-9011-6>
- DeWaard, J. R., Ivanova, N. V., Hajibabaei, M., & Hebert, P. D. N. (2008). Assembling DNA barcodes. *Environmental Genomics*, 410, 275–294.
- DeWaard, J. R., Levesque-Beaudin, V., deWaard, S. L., Ivanova, N. V., McKeown, J. T. A., Miskie, R., ... Hebert, P. D. N. (2019). Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. *Genome*, 62(3), 85–95. <https://doi.org/10.1139/gen-2018-0093>
- Doczkal, D. (2017). Vorsortierung der Proben und Vollständigkeit der Erfassung. In A. Ssymank, & D. Doczkal (Eds.), *Biodiversität des südwestlichen Dinkelbergrandes und des Rheintals bei Grenzach-Whylen, eine Bestandsaufnahme im südwestlichen Einfallstor Deutschlands für neue Arten in der Folge des Klimawandels*. Mauritiana (Altenburg) 34, 1–910.
- Douglas, W. Y., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>
- Eiseman, C. S., Heller, K., & Rulik, B. (2016). A new leaf-mining dark-winged fungus gnat (Diptera: Sciaridae), with notes on other insect associates of marsh marigold (Ranunculaceae: *Caltha palustris* L.). *Proceedings of the Entomological Society of Washington*, 118(4), 519–533.
- Ekrem, T., Stur, E., & Hebert, P. D. N. (2010). Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding. *Organisms Diversity & Evolution*, 10(5), 397–408. <https://doi.org/10.1007/s13127-010-0034-y>
- Ekrem, T., Willassen, E., & Stur, E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution*, 43(2), 530–542. <https://doi.org/10.1016/j.ympev.2006.11.021>
- 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>
- Erwin, T. L. (1982). Tropical forests: Their richness in Coleoptera and other arthropod species. *The Coleopterists Bulletin*, 36(1), 74–75.
- Fontaine, B., van Achterberg, K., Alonso-Zarazaga, M. A., Araujo, R., Asche, M., Aspöck, H., ... Bouchet, P. (2012). New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st Century taxonomy. *PLoS ONE*, 7(5), e36881. <https://doi.org/10.1371/journal.pone.0036881>
- Fujita, M. K., Leache, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27(9), 480–488. <https://doi.org/10.1016/j.tree.2012.04.012>
- Geiger, M. F., Astrin, J. J., Borsch, T., Burkhardt, U., Grobe, P., Hand, R., ... Monje, C. (2016). How to tackle the molecular species inventory for an industrialized nation—lessons from the first phase of the German Barcode of Life initiative GBOL (2012–2015). *Genome*, 59(9), 661–670.
- Geiger, M. F., Morinière, J., Hausmann, A., Haszprunar, G., Wägele, W., Hebert, P. D. N., & 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. <https://doi.org/10.3897/BDJ.4.e10671>
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., ... Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasyntematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 8007–8012.
- Gruttke, H., Binot-Hafke, M., Balzer, S., Haupt, H., Hofbauer, N., Ludwig, G., & Ries, M. (2016). Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands. Band 4: Wirbellose Tiere (Teil 2). *Naturschutz Und Biologische Vielfalt*, 70(4), 598.
- Gutiérrez, M. A. C., Vivero, R. J., Vélez, I. D., Porter, C. H., & Uribe, S. (2014). DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE*, 9(1), e85496.
- Gwiazdowski, R. A., Footitt, R. G., Maw, H. E. L., & Hebert, P. D. N. (2015). The Hemiptera (Insecta) of Canada: Constructing a reference library of DNA barcodes. *PLoS ONE*, 10(4), e0125635.
- Haarto, A., & Ståhls, G. (2014). When mtDNA COI is misleading: Congruent signal of ITS2 molecular marker and morphology for North European *Melanostoma Schiener*, 1860 (Diptera, Syrphidae). *ZooKeys*, 431, 93–134.
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6(4), e17497 (7 pp).
- Hajibabaei, M., Spall, J. L., Shokralla, S., & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12(1), 28.
- Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., ... Goulson, D. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, 12(10), e0185809.
- Hansson, C., & Schmidt, S. (2018). Revision of the European species of *Euplectrus* Westwood (Hymenoptera, Eulophidae), with a key to European species of *Euplectrini*. *Journal of Hymenoptera Research*, 67, 1.
- Hardaluk, L. (in prep.). Metabarcoding in the Nationalpark Bayerischer Wald - screening for invasive and pest invertebrates in bulk samples.
- Haszprunar, G. (2009). Barcoding Fauna Bavarica—eine Chance für die Entomologie. *Nachrichtenblatt Der Bayerischen Entomologen Bayer Ent*, 58(1/2), 45.
- Hausmann, A., Godfray, H. C. J., Huemer, P., Mutanen, M., Rougerie, R., van Nieuerken, E. J., ... Hebert, P. D. N. (2013). Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) system. *PLoS ONE*, 8(12), e84518. <https://doi.org/10.1371/journal.pone.0084518>
- Hausmann, A., Haszprunar, G., & Hebert, P. D. N. (2011). DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): Successes, surprises, and questions. *PLoS ONE*, 6(2), e17134. <https://doi.org/10.1371/journal.pone.0017134>
- Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G., & Hebert, P. D. N. (2011). Now DNA-barcoded: The butterflies and larger moths of Germany. *Spixiana*, 34(1), 47–58.
- Havemann, N., Gossner, M. M., Hendrich, L., Morinière, J., Niedringhaus, R., Schäfer, P., & Raupach, M. J. (2018). From water striders to water bugs: The molecular diversity of aquatic Heteroptera (Gerromorpha, Nepomorpha) of Germany based on DNA barcodes. *PeerJ*, 6, e4577. <https://doi.org/10.7717/peerj.4577>
- Hawllitschek, O., Fernández-González, A., Balmori-de la Puente, A., & Castresana, J. (2018). A pipeline for metabarcoding and diet analysis from fecal samples developed for a small semi-aquatic mammal. *PLoS ONE*, 13(8), e0201763. <https://doi.org/10.1371/journal.pone.0201763>
- Hawllitschek, O., Morinière, J., Lehmann, G. U. C., Lehmann, A. W., Kropf, M., Dünz, A., ... Haszprunar, G. (2017). DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria. *Germany and Switzerland. Molecular Ecology Resources*, 17(5), 1037–1053. <https://doi.org/10.1111/1755-0998.12638>
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Hebert, P. D. N., Ratnasingham, S., Zakharov, E. V., Telfer, A. C., Levesque-Beaudin, V., Milton, M. A., ... Jannetta, P. (2016). Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702), 20150333.
- Heller, K., Köhler, A., Menzel, F., Olsen, K. M., & Gammelø, Ø. (2016). Two formerly unrecognized species of Sciaridae (Diptera) revealed by DNA barcoding. *Norwegian Journal of Entomology*, 63(1), 96–115.
- Heller, K., & Rulik, B. (2016). *Ctenosciara alexanderkoenigi* sp. n. (Diptera: Sciaridae), an exotic invader in Germany? *Biodiversity Data Journal*, 4, e6460.
- Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Köhler, F., & Balke, M. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: Adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, 15(4), 795–818. <https://doi.org/10.1111/1755-0998.12354>
- Hernández-Triana, L. M., Prosser, S. W., Rodríguez-Perez, M. A., Chaverri, L. G., Hebert, P. D. N., & Ryan Gregory, T. (2014). Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths. *Molecular Ecology Resources*, 14(3), 508–518.
- Hubert, N., & Hanner, R. (2015). DNA barcoding, species delineation and taxonomy: A historical perspective. *DNA Barcodes*, 3(1), 44–58. <https://doi.org/10.1015/dna-2015-0006>
- Ivanova, N. V., Dewaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, 6(4), 998–1002. <https://doi.org/10.1111/j.1471-8286.2006.01428.x>
- Jaschhof, M. (2009). Eine aktualisierte Artenliste der Holzmücken Deutschlands, unter besonderer Berücksichtigung der Fauna Bayerns (Diptera, Cecidomyiidae, Lestremiinae). *Spixiana*, 32(1), 139–151.
- Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16(10), 1245–1257. <https://doi.org/10.1111/ele.12162>
- Jordaens, K., Goergen, G., Virgilio, M., Bäckeljau, T., Vokaer, A., & De Meyer, M. (2015). DNA barcoding to improve the taxonomy of the Afrotropical hoverflies (Insecta: Diptera: Syrphidae). *PLoS ONE*, 10(10), e0140264. <https://doi.org/10.1371/journal.pone.0140264>
- Jordaens, K., Sonet, G., Braet, Y., De Meyer, M., Bäckeljau, T., Goovaerts, F., ... Desmyter, S. (2013). DNA barcoding and the differentiation between North American and West European *Phormia regina* (Diptera, Calliphoridae, Chrysomyinae). *ZooKeys*, 365, 149–174. <https://doi.org/10.3897/zookeys.365.6202>
- Karlsson, D., Pape, T., Johansson, K. A., Liljebäck, J., & Ronquist, F. (2005). The Swedish Malaise Trap Project, or how many species of Hymenoptera and Diptera are there in Sweden? *Entomologisk Tidskrift*, 126, 43–53.
- Klausnitzer, B. (2006). Stiefkinder der Entomologie in Mitteleuropa. *Beiträge Zur Entomologie*, 56, 360–368.
- Krüger, A., Strüven, L., Post, R. J., & Faulde, M. (2011). The sandflies (Diptera: Psychodidae, Phlebotominae) in military camps in northern Afghanistan (2007–2009), as identified by morphology and DNA 'barcoding'. *Annals of Tropical Medicine & Parasitology*, 105(2), 163–176. <https://doi.org/10.1179/136485911X12899838683241>
- Kumar, N. P., Rajavel, A. R., Natarajan, R., & Jambulingam, P. (2007). DNA barcodes can distinguish species of Indian mosquitoes (Diptera:

- Culicidae). *Journal of Medical Entomology*, 44(1), 01–07. <https://doi.org/10.1093/jmedent/41.5.01>
- Kumar, N. P., Srinivasan, R., & Jambulingam, P. (2012). DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. *Molecular Ecology Resources*, 12(3), 414–420. <https://doi.org/10.1111/j.1755-0998.2012.03117.x>
- Latibari, M. H., Moravvej, G., Heller, K., Rulik, B., & Namaghi, H. S. (2015). New records of Black Fungus Gnats (Diptera: Sciaridae) from Iran, including the reinstatement of *Bradysia cellarum* Frey. *Studia Dipterologica*, 22(1), 39–44.
- Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences USA*, 112(7), 2076–2081. <https://doi.org/10.1073/pnas.1424997112>
- Leray, M., Yang, Y. J., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10(1), 34. <https://doi.org/10.1186/1742-9994-10-34>
- Lister, B. C., & Garcia, A. (2018). Climate-driven declines in arthropod abundance restructure a rainforest food web. *Proceedings of the National Academy of Sciences*, 115(44), E10397–E10406. <https://doi.org/10.1073/pnas.1722477115>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- Matthews, R. W., & Matthews, J. R. (1971). The Malaise trap: Its utility and potential for sampling insect populations. *The Great Lakes Entomologist*, 4(4), 4.
- May, R. M. (1988). How many species are there on earth? *Science*, 241(4872), 1441–1449.
- Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology*, 55(5), 715–728. <https://doi.org/10.1080/10635150600969864>
- Mengual, X., Ståhls, G., Vujić, A., & Marcos-Garcia, M. A. (2006). Integrative taxonomy of Iberian *Merodon* species (Diptera, Syrphidae). *Zootaxa*, 1377, 1–26.
- Meyer, H., & Stark, A. (2015). Verzeichnis und Bibliografie der Tanzfliegenverwandten Deutschlands (Diptera: Empidoidea: Atelestidae, Brachystomatidae, Dolichopodidae s. l., Empididae, Hybotidae, "Iteaphila-Gruppe", Oreogetonidae). *Studia Dipterologica Supplement* 19.
- Montagna, M., Mereghetti, V., Lencioni, V., & Rossaro, B. (2016). Integrated taxonomy and DNA barcoding of alpine midges (Diptera: Chironomidae). *PLoS ONE*, 11(3), e0149673. <https://doi.org/10.1371/journal.pone.0149673>
- Morinière, J., Cancian de Araujo, B., Lam, A. W., Hausmann, A., Balke, M., Schmidt, 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>
- Morinière, J., Hendrich, L., Balke, M., Beermann, A. J., König, T., Hess, M., ... Haszprunar, G. (2017). A DNA barcode library for Germany's mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Molecular Ecology Resources*, 17(6), 1293–1307. <https://doi.org/10.1111/1755-0998.12683>
- Morinière, J., Hendrich, L., Hausmann, A., Hebert, P., Haszprunar, G., & Gruppe, A. (2014). Barcoding Fauna Bavarica: 78% of the Neuropterida fauna barcoded!. *PLoS ONE*, 9(10), e109719. <https://doi.org/10.1371/journal.pone.0109719>
- Mutanen, M., Kivelä, S. M., Vos, R. A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., ... Godfray, H. C. J. (2016). Species-level para- and polyphyly in DNA barcode gene trees: Strong operational bias in European Lepidoptera. *Systematic Biology*, 65(6), 1024–1040. <https://doi.org/10.1093/sysbio/syw044>
- Nagy, Z. T., Sonet, G., Mortelmans, J., Vandewynkel, C., & Grootaert, P. (2013). Using DNA barcodes for assessing diversity in the family Hybotidae (Diptera, Empidoidea). *ZooKeys*, 365, 263–278. <https://doi.org/10.3897/zookeys.365.6070>
- Nelson, L. A., Lambkin, C. L., Batterham, P., Wallman, J. F., Dowton, M., Whiting, M. F., ... Cameron, S. L. (2012). Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene*, 511(2), 131–142. <https://doi.org/10.1016/j.gene.2012.09.103>
- Nelson, L. A., Wallman, J. F., & Dowton, M. (2007). Using COI barcodes to identify forensically and medically important blowflies. *Medical and Veterinary Entomology*, 21(1), 44–52. <https://doi.org/10.1111/j.1365-2915.2007.00664.x>
- Normark, B. B. (2003). The evolution of alternative genetic systems in insects. *Annual Review of Entomology*, 48(1), 397–423.
- Nzulu, C. O., Cáceres, A. G., Arrunátegui-Jiménez, M. J., Lañas-Rosas, M. F., Yañez-Trujillano, H. H., Luna-Caipo, D. V., ... Kato, H. (2015). DNA barcoding for identification of sand fly species (Diptera: Psychodidae) from leishmaniasis-endemic areas of Peru. *Acta Tropica*, 145, 45–51. <https://doi.org/10.1016/j.actatropica.2015.02.003>
- Ødegaard, F. (2000). How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society*, 71(4), 583–597. <https://doi.org/10.1111/j.1095-8312.2000.tb01279.x>
- Oliverio, A. M., Gan, H., Wickings, K., & Fierer, N. (2018). A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biology and Biochemistry*, 125, 37–43. <https://doi.org/10.1016/j.soilbio.2018.06.026>
- Oosterbroek, P. (2006). *The European Families of the Diptera*. Uitgeverij: KNNV-Vereniging voor Veldbiologie.
- Packer, L., Gibbs, J., Sheffield, C., & Hanner, R. (2009). DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, 9(Supplement 1), 42–50. <https://doi.org/10.1111/j.1755-0998.2009.02631.x>
- Padijal, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 7(1), 16.
- Page, R. D. M. (2016). DNA barcoding and taxonomy: Dark taxa and dark texts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1702), 20150334.
- Pante, E., Schoelinc, C., & Puillandre, N. (2014). From integrative taxonomy to species description: One step beyond. *Systematic Biology*, 64(1), 152–160.
- Pape, T. (2009). Palaeartic Diptera - from tundra to desert. In T. Pape, D. Bickel, & R. Meier (Eds.), *Diptera diversity: Status, challenges and tools* (pp. 121–154). Leiden, The Netherlands: Brill.
- Pape, T., Blagoderov, V., & Mostovski, M. B. (2011). Order Diptera Linnaeus, 1758. In Z.-Q. Zhang (Ed.), *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness* (pp. 222–229). Woodcroft, South Australia: Magnolia Press.
- Papp, L., & B. Darvas (Eds.) (1997). *Contribution to a Manual of Palaeartic Diptera. Vol. 2, Nematocera and Lower Brachycera*. Budapest, Hungary: Science Herald.
- Papp, L., & B. Darvas (Eds.) (1998). *Contribution to a Manual of Palaeartic Diptera. Vol. 3, Higher Brachycera*. Budapest, Hungary: Science Herald.
- Papp, L., & B. Darvas (Eds.) (2000a). *Contribution to a Manual of Palaeartic Diptera. Vol. 1, General and Applied Dipterology*. Budapest, Hungary: Science Herald.
- Papp, L., & B. Darvas (Eds.) (2000b). *Contribution to a Manual of Palaeartic Diptera. Appendix*. Budapest, Hungary: Science Herald.
- Pfenninger, M., Nowak, C., Kley, C., Steinke, D., & Streit, B. (2007). Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. *Molecular Ecology*, 16(9), 1957–1968.

- Pohjoismäki, J. L., Kahanpää, J., & Mutanen, M. (2016). DNA barcodes for the northern European tachinid flies (Diptera: Tachinidae). *PLoS ONE*, 11(11), e0164933.
- Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., & Kunin, W. E. (2010). Global pollinator declines: Trends, impacts and drivers. *Trends in Ecology & Evolution*, 25(6), 345–353.
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864–1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364.
- Ratnasingham, S., & Hebert, P. D. N. (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>
- Raupach, M. J., Hannig, K., Morinière, J., & Hendrich, L. (2016). A DNA barcode library for ground beetles (Insecta, Coleoptera, Carabidae) of Germany: The genus *Bembidion* Latreille, 1802 and allied taxa. *ZooKeys*, 592, 121–141. <https://doi.org/10.3897/zookeys.592.8316>
- Raupach, M. J., Hannig, K., Morinière, J., & Hendrich, L. (2018). A DNA barcode library for ground beetles of Germany: The genus *Amara* Bonelli, 1810 (Insecta, Coleoptera, Carabidae). *ZooKeys*, 759, 57–80. <https://doi.org/10.3897/zookeys.759.24129>
- Raupach, M. J., Hendrich, L., Küchler, S. M., Deister, F., Morinière, J., & Gossner, M. M. (2014). Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. *PLoS ONE*, 9(9), e106940. <https://doi.org/10.1371/journal.pone.0106940>
- Reibe, S., Schmitz, J., & Madea, B. (2009). Molecular identification of forensically important blowfly species (Diptera: Calliphoridae) from Germany. *Parasitology Research*, 106(1), 257–261. <https://doi.org/10.1007/s00436-009-1657-9>
- Reimann, B., & Rulik, B. (2015). *Dasiops calvus* Morge (Diptera: Lonchaeidae), a lance fly new to the German fauna, revealed by the GBOL-project. *Studia Dipterologica*, 21(2), 283–287.
- Renaud, A. K., Savage, J., & Adamowicz, S. J. (2012). DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. *BMC Ecology*, 12(1), 24. <https://doi.org/10.1186/1472-6785-12-24>
- Riedel, A., Sagata, K., Suhardjono, Y. R., Tänzler, R., & Balke, M. (2013). Integrative taxonomy on the fast track-towards more sustainability in biodiversity research. *Frontiers in Zoology*, 10(1), 15. <https://doi.org/10.1186/1742-9994-10-15>
- Rivera, J., & Currie, D. C. (2009). Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). *Molecular Ecology Resources*, 9, 224–236. <https://doi.org/10.1111/j.1755-0998.2009.02648.x>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
- Rojo, S., Ståhls, G., Pérez-Bañón, C., & Marcos-García, M. Á. (2006). Testing molecular barcodes: Invariant mitochondrial DNA sequences vs the larval and adult morphology of West Palearctic *Pandasyopthalmus* species (Diptera: Syrphidae: Paragini). *European Journal of Entomology*, 103(2), 443. <https://doi.org/10.14411/eje.2006.058>
- Rulik, B., Eberle, J., von der Mark, L., Thormann, J., Jung, M., Köhler, F., ... Ahrens, D. (2017). Using taxonomic consistency with semi-automated data preprocessing for high quality DNA barcodes. *Methods in Ecology and Evolution*, 8(12), 1878–1887. <https://doi.org/10.1111/2041-210X.12824>
- Santos, D., Sampranha, S., & Santos, C. M. D. (2017). Advances on dipterology in the 21st century and extinction rates. *Papéis Avulsos De Zoologia*, 57(33), 433–444. <https://doi.org/10.11606/0031-1049.2017.57.33>
- Schlick-Steiner, B. C., Arthofer, W., & Steiner, F. M. (2014). Take up the challenge! Opportunities for evolution research from resolving conflict in integrative taxonomy. *Molecular Ecology*, 23(17), 4192–4194. <https://doi.org/10.1111/mec.12868>
- Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55, 421–438. <https://doi.org/10.1146/annurev-ento-112408-085432>
- Schmid-Egger, C., Straka, J., Ljubomirov, T., Blagoev, G. A., Morinière, J., & Schmidt, S. (2019). DNA barcodes identify 99 per cent of apoïd wasp species (Hymenoptera: Ampulicidae, Crabronidae, Sphecidae) from the Western Palearctic. *Molecular Ecology Resources*, 19(2), 476–484. <https://doi.org/10.1111/1755-0998.12963>
- Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., & Hebert, P. D. N. (2015). DNA barcoding largely supports 250 years of classical taxonomy: Identifications for Central European bees (Hymenoptera, Apoidea, partim). *Molecular Ecology Resources*, 15(4), 985–1000.
- Schmidt, S., Taeger, A., Morinière, J., Liston, A., Blank, S. M., Kramp, K., ... Stahlhut, J. (2017). Identification of sawflies and hornails (Hymenoptera, 'Symphyta') through DNA barcodes: Successes and caveats. *Molecular Ecology Resources*, 17(4), 670–685. <https://doi.org/10.1111/1755-0998.12614>
- Schumann, H. (2002). Erster Nachtrag zur „Checkliste der Dipteren Deutschlands“. *Studia Dipterologica*, 9(2), 437–445.
- Schumann, H. (2004). Zweiter Nachtrag zur „Checkliste der Dipteren Deutschlands“. *Studia Dipterologica*, 11(2), 619–630.
- Schumann, H. (2010). Dritter Nachtrag zur „Checkliste der Dipteren Deutschlands“. *Studia Dipterologica*, 16(1/2), 17–27.
- Schumann, H., Bährmann, R., & Stark, A. (1999). Checkliste der Dipteren Deutschlands. *Entomofauna Germanica 2. Studia Dipterologica Supplement*, 2, 1–354.
- Schumann, H., Doczkal, D., & Ziegler, J. (2011). Diptera - Zweiflügler. In: B. Klausnitzer (Ed.), *Stresemann, Exkursionsfauna von Deutschland. Vol. 2, Wirbellose: Insekten*. 11 (pp. 832–932). Auflage: Spektrum Akademischer Verlag.
- Serrana, J. M., Miyake, Y., Gamboa, M., & Watanabe, K. (2018). Comparison of DNA metabarcoding and morphological identification for stream macroinvertebrate biodiversity assessment and monitoring. *bioRxiv*, 436162. <https://doi.org/10.1101/436162>
- Ševčík, J., Kasprák, D., & Rulik, B. (2016). A new species of *Docosia* Winnertz from Central Europe, with DNA barcoding based on four gene markers (Diptera, Mycetophilidae). *ZooKeys*, 549, 127–143. <https://doi.org/10.3897/zookeys.549.6925>
- Shokralla, S., Spall, J., Gibson, J., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8), 1794–1805. <https://doi.org/10.1111/j.1365-294X.2012.05538.x>
- Sinclair, C. S., & Gresens, S. E. (2008). Discrimination of *Cricotopus* species (Diptera: Chironomidae) by DNA barcoding. *Bulletin of Entomological Research*, 98(6), 555–563. <https://doi.org/10.1017/S0007485308005865>
- Sorg, M., Schwan, H., Stenmans, W., & Müller, A. (2013). Ermittlung der Biomassen flugaktiver Insekten im Naturschutzgebiet Orbroicher Bruch mit Malaise Fallen in den Jahren 1989 und 2013. *Mitteilungen Entomologischer Verein Krefeld*, 1, 1–5.
- Spelda, J., Reip, H. S., Oliveira Biener, U., & Melzer, R. R. (2011). Barcoding Fauna Bavarica: Myriapoda – a contribution to DNA sequence-based identifications of centipedes and millipedes (Chilopoda, Diplopoda). *ZooKeys*, 115, 123–139. <https://doi.org/10.3897/zookeys.115.2176>

- Ssymank, A., Doczkal, D., Rennwald, K., & Dziöck, F. (2011). Rote Liste und Gesamtartenliste der Schwebfliegen (Diptera: Syrphidae) Deutschlands. *Naturschutz Und Biologische Vielfalt*, 70(3), 13–83.
- Ssymank, A., Sorg, M., Doczkal, D., Rulik, B., Merkel-Wallner, G., & Vischer-Leopold, M. (2018). Praktische Hinweise und Empfehlungen zur Anwendung von Malaisefallen für Insekten in der Biodiversitätserfassung und im Monitoring. *Series Naturalis*, 1, 1–12.
- Stur, E., & Borkent, A. (2014). When DNA barcoding and morphology mesh: Ceratopogonidae diversity in Finnmark, Norway. *ZooKeys*, 463, 95–131. <https://doi.org/10.3897/zookeys.463.7964>
- Stur, E., & Ekrem, T. (2011). Exploring unknown life stages of Arctic Tanytarsini (Diptera: Chironomidae) with DNA barcoding. *Zootaxa*, 2743(1), 27–39. <https://doi.org/10.11646/zootaxa.2743.1.2>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Vanbergen, A. J., & Insect Pollinators Initiative. (2013). Threats to an ecosystem service: Pressures on pollinators. *Frontiers in Ecology and the Environment*, 11(5), 251–259. <https://doi.org/10.1890/120126>
- Versteirt, V., Nagy, Z. T., Roelants, P., Denis, L., Breman, F. C., Damiens, D., ... Van Bortel, W. (2015). Identification of Belgian mosquito species (Diptera: Culicidae) by DNA barcoding. *Molecular Ecology Resources*, 15(2), 449–457. <https://doi.org/10.1111/1755-0998.12318>
- Völkl, W., Blick, T., Kornacker, P. M., & Martens, H. (2004). Quantitativer Überblick über die rezente Fauna von Deutschland. *Natur Und Landschaft*, 79(7), 293–295.
- Wang, G., Li, C., Guo, X., Xing, D., Dong, Y., Wang, Z., ... Zhao, T. (2012). Identifying the main mosquito species in China based on DNA barcoding. *PLoS ONE*, 7(10), e47051. <https://doi.org/10.1371/journal.pone.0047051>
- Wesener, T., Voigtländer, K., Decker, P., Oeyen, J. P., Spelda, J., & Lindner, N. (2015). First results of the German Barcode of Life (GBOL)-Myriapoda project: Cryptic lineages in German *Stenotaenia linearis* (Koch, 1835) (Chilopoda, Geophilomorpha). *ZooKeys*, 510, 15–29. <https://doi.org/10.3897/zookeys.510.8852>
- Wheeler, Q. D., Raven, P. H., & Wilson, E. O. (2004). Taxonomy: Impediment or expedient? *Science*, 303, 285–285. <https://doi.org/10.1126/science.303.5656.285>
- Whitworth, T. L., Dawson, R. D., Magalon, H., & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B: Biological Sciences*, 274(1619), 1731–1739.
- Wolff, D., Gebel, M., & Geller-Grimm, F. (2018). *Die Raubfliegen Deutschlands*. Quelle & Meyer Bestimmungsbücher.
- Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Morinière J, Balke M, Doczkal D, et al. A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Mol Ecol Resour*. 2019;19:900–928. <https://doi.org/10.1111/1755-0998.13022>