Facing the infinity: Tackling large samples of challenging Chironomidae (Diptera) with an integrative approach

Caroline Chimeno [Corresp., 1], **Björn Rulik** [2], **Alessandro Manfrin** [3], **Gregor Kalinkat** [4], **Franz Hölker** [4], **Viktor O Baranov** [5]

[1] SNSB-ZSM, Bavarian State Collection of Zoology, Munich, Germany

[2] Zoological Research Museum Alexander Koenig, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany

[3] Institute for Environmental Sciences, iES Landau, RPTU University of Kaiserslautern-Landau, Landau, Germany

[4] Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

[5] Estación Biológica de Doñana-CSIC / Doñana Biological Station-CSIC, Seville, Spain

Corresponding Author: Caroline Chimeno
Email address: chimeno@snsb.de

Background. Integrative taxonomy is becoming ever more significant in biodiversity research as scientists are tackling increasingly taxonomically challenging groups. Implementing a combined approach not only guarantees more accurate species identification, but also helps overcome limitations that each method presents when applied on its own. In this study, we present one application of integrative taxonomy for the highly abundant and particularly diverse fly taxon Chironomidae (Diptera). Although non-biting midges are key organisms in merolimnic systems, they are often cast aside in ecological surveys because they are very challenging to identify and extremely abundant.

Methods. Here, we demonstrate one way of applying integrative methods to tackle this highly diverse taxon. We present a three-level subsampling method to drastically reduce the workload of bulk sample processing, then apply morphological and molecular identification methods in parallel to evaluate species diversity and to examine inconsistencies across methods.

Results. Our results suggest that using our subsampling approach, identifying less than 10% of a sample's contents can reliably detect >90 % of its diversity. However, despite reducing the processing workload drastically, the performance of our taxonomist was affected by mistakes, caused by large amounts of material. We conducted misidentifications for 9% of vouchers, which may not have been recovered had we not applied a second identification method. On the other hand, we were able to provide species information in cases where molecular methods could not, which was the case for 14% of vouchers. Therefore, we conclude that when wanting to implement non-biting midges into ecological frameworks, it is imperative to use an integrative approach.

Facing the infinity: Tackling large samples of challenging Chironomidae (Diptera) with an integrative approach

Caroline Chimeno[1], Björn Rulik[2], Alessandro Manfrin[3], Gregor Kanlinkat[4], Franz Hölker[4], Viktor Baranov[5]

[1] SNSB-ZSM, Bavarian State Collection of Zoology, Munich, Germany
[2] Zoological Research Museum Alexander Koenig, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Bonn, Germany
[3] Institute for Environmental Sciences, iES Landau, RPTU University of Kaiserslautern-Landau, Landau, Germany
[4] Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
[5] Estación Biológica de Doñana-CSIC / Doñana Biological Station-CSIC, Seville, Spain

Corresponding Author:
Caroline Chimeno[1]
Münchhausenstraße 21, Munich, Bavaria, 81247, Germany
Email address: chimeno@snsb.de

Abstract

Background. Integrative taxonomy is becoming ever more significant in biodiversity research as scientists are tackling increasingly taxonomically challenging groups. Implementing a combined approach not only guarantees more accurate species identification, but also helps overcome limitations that each method presents when applied on its own. In this study, we present one application of integrative taxonomy for the highly abundant and particularly diverse fly taxon Chironomidae (Diptera). Although non-biting midges are key organisms in merolimnic systems, they are often cast aside in ecological surveys because they are very challenging to identify and extremely abundant.

Methods. Here, we demonstrate one way of applying integrative methods to tackle this highly diverse taxon. We present a three-level subsampling method to drastically reduce the workload of bulk sample processing, then apply morphological and molecular identification methods in parallel to evaluate species diversity and to examine inconsistencies across methods.

Results. Our results suggest that using our subsampling approach, identifying less than 10% of a sample's contents can reliably detect >90 % of its diversity. However, despite reducing the processing workload drastically, the performance of our taxonomist was affected by mistakes, caused by large amounts of material. We conducted misidentifications for 9% of vouchers, which may not have been recovered had we not applied a second identification method. On the other hand, we were able to provide species information in cases where molecular methods could not, which was the case for 14% of vouchers. Therefore, we conclude that when wanting to implement non-biting midges into ecological frameworks, it is imperative to use an integrative approach.

Introduction

Chironomidae (non-biting midges) is by far the most ecomorphologically diverse and widely distributed ingroup of aquatic insects (Hilsenhoff, Thorp & Covich, 2001; Armitage, Pinder & Cranston, 2012). Occurring in every zoo geographic region, including Antarctica, non-biting midges inhabit nearly all aquatic and semiaquatic, marine and terrestrial habitats (Armitage, Pinder & Cranston, 2012). Characteristic behavioral and physiological adaptations have enabled these flies to colonize extreme environments such as caves up to 1,000 meters deep, hot springs, high-altitude waters, glacial streams, and even highly polluted waters or sewage systems (Andersen et al., 2016; Gadawski et al., 2022). In aquatic systems, their abundance can be higher than that of all other macroinvertebrates combined, making them a flagship taxon in freshwater ecology (Gratton & Zanden, 2009; Marziali et al., 2010; Karima, 2021). The bottom-dwelling larvae not only represent almost every feeding group but, being ecosystem engineers, they also contribute enormously to sediment- and water-mixing, and to the global oxygen- and carbon-cycle (Hölker et al., 2015; Baranov, Lewandowski & Krause, 2016; Antczak-Orlewska et al., 2021). As ecosystem engineers, the Chironomidae are involved in modifying the availability of nutrients (chiefly phosphorous, but also nitrogen), as well as oxygen and carbon availability for other aquatic organisms (Hölker et al., 2015; Baranov, Lewandowski & Krause, 2016). All life stages (even the short-lived adults) play a vital role in aquatic and terrestrial food webs, serving as an important food source for fish, birds, bats and other arthropods (Gratton & Zanden, 2009; Raunio, Heino & Paasivirta, 2011; Armitage, Pinder & Cranston, 2012; Wirta et al., 2015; Herren et al., 2017). This combination of high ecosystem functionality, high abundance, and

habitat specificity of the Chironomidae to their environment makes them suitable biological indicators for ecological assessments (e.g. water quality control) (Sæther, 1977; Lencioni, Marziali & Rossaro, 2012; Dorić et al., 2021).

Despite this, only a limited subset of biodiversity studies or biomonitoring surveys of aquatic habitats incorporate species- or genus-level information of the Chironomidae and oftentimes, they are neglected altogether (Raunio, Heino & Paasivirta, 2011; Dorić et al., 2021). This is due to several factors: (i) non-biting midges are relatively difficult to identify (Cranston, 2008; Proulx et al., 2013), (ii) only few taxonomists with the required expertise are available for species-level identification (Cranston et al., 2013; Chan et al., 2014), (iii) traditional morphological-based species delimitations often require laborious dissection and mounting of specimens on microscope slides (Ekrem, Stur & Hebert, 2010; Gadawski et al., 2022), and (iv) they can be extremely species rich even in relative low-diversity temperate and boreal ecosystems (Lundström et al., 2010). The workload associated with the processing of non-biting midges from large bulk samples, common in ecological surveys, is immense when applying traditional identification methods (Rosenberg, 1992; Brodin et al., 2012). In humid climates, or during wetter years, the number of specimens to be processed can increase from hundreds of thousands to sometimes millions of specimens.

There are few methods that can help overcome the pitfall of processing an “infinite” number of specimens, with the most obvious one (and most resource-demanding) being the employment of more taxonomists or parataxonomists (Engel et al., 2021) to help accelerate specimen processing and identification. The availability of expert taxonomists, however, is in decline and even then, financing such manpower at a large scale is often not feasible and remains time-consuming (Hausmann et al., 2020; Chimeno et al., 2022). Therefore, researchers often subsample bulk samples to reduce the sorting effort, or limit sample processing to a few key families or species (Mandelik, Roll & Fleischer, 2010; Porter et al., 2014; Keck et al., 2017; Bohan et al., 2017; Chimeno et al.). One promising alternative that is currently in development is the use of automatic machine-based identification approaches for species identification (see Milošević et al., 2020). As demonstrated by Milošević and authors, after vigorously training their artificial neural network on 1,836 specimens belonging to ten similar-looking species of Chironomidae, they recovered 99% identification success when presenting their network new images. Despite

these promising results, this technology is not yet applicable at a large scale because it requires laborious sample preparation and a vigorous training-phase of the target taxa (Milošević et al., 2020).

Currently, one of the most common and promising methodologies for large-scale species identification is DNA barcoding, a molecular-based identification method (Brodin et al., 2012; Morinière et al., 2016). It uses a short DNA fragment to differentiate species from one another, and does so at a lower cost and faster pace than traditional morphological methods (Hebert et al., 2003; Ekrem, Willassen & Stur, 2007; Porter et al., 2014; Morinière et al., 2016). With the rise of DNA barcoding, high quality species-level information of Chironomidae is increasingly becoming more accessible to research (Ekrem, Stur & Hebert, 2010; Baloğlu, Clews & Meier, 2018), and studies examining the efficiency of this method in research of these insects reveal an overall congruence of 80-90%, making it a great complement to taxonomic methodologies (Carew, Pettigrove & Hoffmann, 2005; Pfenninger et al., 2007; Ekrem, Willassen & Stur, 2007; Carew et al., 2007; Carew, Marshall & Hoffmann, 2011; Lin, Stur & Ekrem, 2015). However, just as any identification method, DNA barcoding has its own limitations (Dayrat, 2005; Will, Mishler & Wheeler, 2005; Schlick-Steiner et al., 2010) and therefore, numerous studies resort to applying a combined methodological approach for species identifications (Pires & Marinoni, 2010; Sheth & Thaker, 2017).

With many studies highlighting the need for a smart and efficient integration of both morphological and molecular species identification methods (Hausmann et al., 2020; Hartop et al., 2022), our study aims to present and evaluate one way to do so for a particularly diverse and complicated group of insects: the Chironomidae. To tackle the large amounts of insect material, we apply a three-level subsampling technique that we present in the Methods. We also compare our DNA- and morphology based species identifications in terms of accuracy, to demonstrate how the use of each method on its own can provide discrepant results. We are processing bulk samples of Diptera that have been collected in the framework of the federal-funded field experiment “Verlust der Nacht” (https://www.igb-berlin.de/projekt/verlust-der-nacht) and the follow-up project “Artenschutz durch umweltfreundliche Beleuchtung” (https://www.igb-berlin.de/projekt/artenschutz-durch-umweltvertraegliche-beleuchtung-aube) located in the

Westhavelland Nature Park in northeast Germany. The project was launched in 2012 with the goal of studying the effects that artificial lighting at night has on species communities.

Materials & Methods

Study area and experimental design

The "Verlust der Nacht" experiment was conducted by the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) in a large-scale facility established in 2012 (see Holzhauer et al., 2015; Manfrin et al., 2017 for details). The facility is located in a 750-km² Dark-Sky Reserve within the Westhavelland Nature Park in the Berlin-Brandenburg Metropolitan Region (http://darksky.org/our-work/conservation/idsp/reserves/westhavelland/). The landscape is characterized by a system of drainage ditches (approximately 5 m wide, average annual water depth 50 ± 26 cm). In the grassland adjacent to the drainage ditch, we installed three parallel rows (3 m, 23 m and 43 m away from the drainage ditch) of four conventional 4.75 m high streetlights located 20 m apart. Each lamp post in the lit site was equipped with one 70-W high-pressure sodium lamp (VIALOX NAV-T Super 4Y, yellow 2000 K, Osram, Munich, Germany). In the control (dark) site only the lamp posts were installed (i.e., without bulbs) providing identical physical structure yet remaining dark. The lamps used in the lit site had a maximum illuminance of approximately 50 lx directly under the lamp, with the minimum illuminance between two adjacent streetlamps of the same row being approximately 10 lx, and a minimum illuminance between rows of streetlamps of ca. 1 lx (see Holzhauer et al. 2015 for further details about light distribution and spectral composition). From spring 2012 onward, the lit site was illuminated at night, i.e. between civil twilight at dusk and dawn. The lit and control sites are very similar in their environmental characteristics (e.g. water physico-chemistry, hydromorphology, riparian vegetation) and ~600 m (800 m along the drainage ditch) apart, separated by a row of trees.

Insect collection

We collected insects emerging from the drainage ditch from both lit and dark sites from May to October 2014. Emerging insects were sampled using four floating pyramidal emergence traps (0.85 × 0.85m, 300-µm mesh), placed in the drainage ditch *ca.* 1 m from the bank and directly in front of each streetlamp. Sampling duration ranged from 7 (one night samplings) to

approximately 185 hours (one week samplings) and occurred monthly except in July when the sampling was conducted twice. Flying adult non-biting midges were collected from the grassland adjacent to the drainage ditch using 24 flight interception traps, 12 at each site. Flight intercepting traps were placed 0.5 m below each lamp and consisted of two perpendicular acrylic panels (each 204 × 500 × 3 mm) mounted above a collecting funnel. The flight intercepting traps were collecting insects for one 24-h sampling period every month except in July when sampling was conducted twice. Based on astronomical sunset and sunrise, the 24-h sampling periods were always split into a night-sampling (8–14 hours, depending on the season) followed by a day-sampling (10–16 hours), replacing the collecting jars after each of them. Sampling always occurred on rainless days/nights within 24 h of either first- or third-quarter moon. Both emergence and flight intercepting traps were equipped with collecting jars containing 70% ethanol as a preservative medium (see Manfrin et al. 2017 for further details).

Morphotype sorting & subsampling for processing

We obtained bulk samples of pre-sorted adult "Nematoceran" flies (crane flies, midges, gnats, mosquitoes etc.) stored in 90% ethanol that were collected in the sampling year 2014 (see "Insect collection"). From these samples, our senior author, who is a trained expert of non-biting midges, sorted specimens using a stereo microscope and grouped them into different morphotypes. To do this, we used three different approaches based on the "difficulty" of specimen sorting (*Fig. 1*). Large and/or conspicuous species that are easy to recognize, such as *Prodiamesa olivacea* (Meigen, 1818) or *Ablabesmyia phatta* (Egger, 1863), were quickly sorted into their own distinct morphotypes and assigned a preliminary species name. Specimens that were more difficult to group (because they belong to genera that have similar-looking representatives when viewed under the stereo microscope) were sorted at the genus-level, hence, grouped into genera-morphotypes if possible. Hence, if several genera have similar-looking representatives under the stereo microscope, we sorted representatives of several genera into one morphotype. Lastly, for specimens that our expert taxonomist found difficult to address, subsets were mounted on temporary glycerol slides to be examined at x400 magnification in a first step, so that similar specimens can be assigned to the same morphotype in a second step. From every morphotype group, we selected a representative number of morphotype voucher specimens (about 10%). For very abundant morphotypes where 10% of specimens is still too much, we sampled fewer

186 individuals. Selected specimens were used for molecular and morphological species
187 identifications.

188 Sequencing of selected specimens

189 For specimens larger than 2 mm, we used a single leg or leg segment as a tissue sample that was
190 transferred to a 96-well plate. For smaller individuals, we extracted DNA non-destructively (i.e.,
191 subsequent voucher recovery) from the whole body. After lysis, we extracted genomic DNA
192 using the BioSprint96 magnetic bead extractor and the respective kits by Qiagen (Hilden,
193 Germany). We carried out a polymerase chain reaction (PCR) in a total reaction volume of 20 µl,
194 including 2 µl of undiluted DNA template, 0.8 µl of each primer (10 pmol/µl), 2 µl of 'Q-
195 Solution' and 10 µl of 'Multiplex PCR Master Mix', containing hot start Taq DNA polymerase
196 and buffers. The latter components are available in the Multiplex PCR kit by Qiagen (Hilden,
197 Germany).

198 Thermal cycling was performed on GeneAmp PCR System 2700 machines (Life Technologies,
199 Carlsbad, CA, USA) as follows: hot start *Taq* activation: 15 min at 95 °C; first cycle set (15
200 repeats): 35 s denaturation at 94 °C, 90 s annealing at 55 °C (−1 °C/cycle) and 90 s extension at
201 72 °C. Second cycle set (25 repeats): 35 s denaturation at 94 °C, 90 s annealing at 40 °C and 90 s
202 extension at 72 °C; final elongation 10 min at 72 °C. As established within GBOL at the ZFMK,
203 we used the standard degenerate barcoding primers LCO1490-JJ: 5′-
204 CHACWAAYCATAAAGATATYGG- 3′ and HCO2198-JJ: 5′-
205 AWACTTCVGGRTGVCCAAARAATCA- 3′ (Astrin & Stüben, 2008). Purification and
206 sequencing were conducted by the BGI Group (Hong Kong, China) using the amplification
207 primers.

208 Traces were semi-automatically edited, then assembled sequences using the MUSCLE alignment
209 approach (Edgar, 2004), and checked for the occurrence of stop-codons or hints of nuclear
210 mitochondrial DNA segments (NUMTs) in Geneious version 7.1.9 (http://www.geneious.com;
211 Kearse et al., 2012). Further details such as voucher information, primer pairs, sequence data and
212 trace files were deposited to BOLD and Genbank. These can be found under the following
213 information (http://doi.org/10.5883/DS-ALANCHIR; GenBank accession numbers OP927392 –
214 OP927685).

Morphological identifications

After DNA barcoding (or in parallel, depending on whether whole specimens were transferred to plates or just tissue samples), we mounted the specimens (or their empty shells) on permanent slides in Euparal and Hydromatrix following standard procedure (Kirk-Spriggs & Sinclair, 2017). Morphological identifications were conducted with aid of numerous identification keys and papers covering palaearctic Chironomidae (see (Lehmann, 1970; Saether, 1971; Hirvenoja, 1973; Wiederholm, 1989; Ekrem, 2002a; Langton & Pinder, 2007; Pillot, 2008; Giłka, 2011). These identifications were conducted by our senior author which has conducted various research on the taxonomy of Chironomidae (see Baranov, 2011a,b, 2013; Baranov & Perkovsky, 2013; Baranov & Przhiboro, 2014; Baranov, Andersen & Hagenlund, 2015; Baranov, Andersen & Perkovsky, 2015; Baranov, Góral & Ross, 2017; Baranov et al., 2019).

Data analysis

All sequence records including metadata were uploaded to the online database Barcode of Life Data System (BOLD; Ratnasingham & Hebert, 2007). Sequences ≥ 300 base pairs (bp) were automatically assigned a Barcode Index Number (BIN) on BOLD if sequence similarity based on the (RESL-) BIN algorithm was fulfilled. Sequences ≥ 500 bp which did not find a match served as founders of new BINs. The dataset was downloaded on April 11, 2022, for analysis and can be viewed on Figshare (https://doi.org/10.6084/m9.figshare.21803013). Therefore, the present results correspond to BINs assigned at that time (BIN assignments can change as new sequences are added to BOLD). In addition to using the RESL-algorithm that is implemented into BOLD, we also applied Assemble Species by Automatic Partitioning (ASAP; Puillandre et al., 2021) and SpeciesIdentifier version 1.9 (Meier et al., 2016) to cluster our sequences at 3%. ASAP uses pairwise genetic distances for hierarchical clustering without using information on intraspecific diversity, and SpeciesIdentifier is an algorithm that allows to cluster sequences based on their pairwise intra- and interspecific genetic distances. The outputs of all three algorithms were used to compare the number of Operational Taxonomic Units (OTUs) obtained with each and comparing diversity assessments. To compare all methodologies, we created a Neighbor-Joining in MEGA11 (version 11.0.13) of all sequence data and added morphological species-, ASAP-, RESL-, and SpeciesIdentifier labels *(Supplementary Data S1)*. Because all depict similar

performance (see results), subsequent taxonomic analyses were conducted only using the RESL outputs.

To assess our sampling effort, we calculated Chao1 and Chao2 estimates using the *ChaoSpecies* function of the *SpadeR* package (version 0.1.1; Chao, Ma & Chiu, 2016) in R (version 4.2.1) on abundance data and incidence data respectively (*Supplementary Data S2*). We did this to estimate the species diversity at the sampling site and to compare it to that which was empirically observed in our samples. Then, we used the *iNEXT* function from the *iNEXT* package (version 3.0.0; Hsieh, Ma & Chao, 2016; Hsieh & Chao, 2020) to extrapolate the species diversity obtained with each methodology (morphology, RESL, ASAP, and SpeciesIdentifier) to double the sampling effort. To depict the species diversity recovered per morphotype, we created accumulation curves using the *iNEXT* function on results derived from each identification method (morphological and molecular).

To double-check our identifications and to recover possible misidentifications, we created a dataset from BOLD containing 19,525 public COI-sequences of 1,035 species of non-biting midges collected throughout Europe (*Supplementary Data S3*). We applied the following selection criteria to build a neighbor-joining tree: Kimura 2 Parameter distance model, sequences $\geq$ 200 bp, and excluding contaminants, records flagged with stop codons, and records flagged as misidentifications. To facilitate review, we colored the tree based on barcode clusters (BINs). We added the names of identifiers along with the identification method to each entry to discriminate high-level taxonomists that used morphological methods to vouchers from parataxonomists relying on the BOLD engine for sequence identification. We considered expert identifications as those conducted by researchers with taxonomic experience of Chironomidae, such as Elisabeth Stur (Norwegian University of Science and Technology; Norway; see (Stur & Ekrem, 2000, 2006, 2011, 2015; Stur & Wiedenbrug, 2005; Stur & Spies, 2011), Torbjørn Ekrem (Norwegian University of Science and Technology, Norway; see Ekrem, 2002a,b, 2007; Ekrem & Stur, 2009; Ekrem, Stur & Hebert, 2010, Yngve Brodin (Swedish Museum of Natural History, Sweden; see Brodin, Lundström & Paasivirta, 2008; Siri & Brodin, 2014), Piotr Gadawski (University of Lodz; Poland; see Gadawski et al., 2022; Giłka & Gadawski, 2022), and Sophie Wiedenbrug (University of São Paulo, Brazil) (see Wiedenbrug, Lamas & Trivinho-Strixino, 2012, 2013; Silva & Wiedenbrug, 2015; Wiedenbrug & Silva, 2016).

Results

Identification of specimens

Overall, we sorted through 4,549 specimens of non-biting midges which made up the bulk (99.6%) of "Nematoceran" specimens in our samples. We recovered 48 morphotype groups, and in total selected 331 specimen-vouchers, of which more than half were females (*Supplementary Data S2*).

Molecular Identifications

We applied DNA barcoding to all 331 specimens and obtained 315 COI-barcodes (95%) that we uploaded to BOLD. Five sequences contained cross contaminations, and another 16 were identified as not being non-biting midges, but species of the taxa Anisopodidae, Chaoboridae, Culicidae, Hybotidae, Psychodidae, Sciaridae, and Trichoceridae . The remaining COI-sequences were clustered into 77 BINs which provided coverage for 55 species and 4 interim species (essentially being morphotype analogs that are widely used in ecological studies)(*Ablabesmyia sp. 2ES*, *Smittia sp. 8ES*, *Smittia sp. 14ES*, and *Thienemanniella sp. 3TE*). Interim species names are assigned on BOLD when molecular analysis detects genetic differences, but no species name can be provided due to the lack of a taxonomic revision or of formal species description (Stur & Ekrem, 2011; Morinière et al., 2016). Seven BINs did not provide conclusive species-level identification and five BINs did not match to public data, providing no molecular identification. In five cases, two BINs were assigned to the same species (*Cladopelma viridulum* - BOLD:AAD7363 and BOLD:AAV3586; *Polypedilum cultellatum* - BOLD:AAH7761 and BOLD:ACX5929; *Polypedilum sordens* - BOLD:ACY3855 and BOLD:ADF3485; *Smittia stercoraria* - BOLD:AAN5358 and BOLD:AAN5355; *Smittia sp. 14ES* - BOLD:AAM7064 and BOLD:ACW5117). *Supplementary Data S2* provides an overview of the entire dataset.

We applied two other clustering algorithms (SpeciesIdentifier and ASAP) to our COI data. Although both SpeciesIdentifier (using 3% threshold) and ASAP (1st partition) did suggest slightly fewer clusters than the RESL-algorithm, the results were largely consistent across methods (*Table 1*) and all derived species diversities fall into the 95% confidence interval (*Fig. 2*).

Morphological Identifications

Using morphological methods, we identified a total of 76 species. 34 specimens were left unidentified at a higher taxonomic level: 22 at the genus-, and 12 at the family-level.

Assessing our sampling effort

Chao1 species richness estimates suggest that 79 ± 5 to 89 ± 7 species may have been present in the community that we sampled (*Table 1*). Sample-based Chao2 estimates were slightly higher, suggesting 92 ± 11 to 109 ± 15 species. Extrapolation to double the sampling effort would have increased the number of recovered entities by 11-17% (*Fig. 2*). Sample coverage was above 90% for all data (morphology, RESL, ASAP, SpeciesIdentifier).

Discrepancies between morphology- and DNA-based identifications

Overall, we recovered discrepant identifications among 103 specimens (*Table 2*), and categorized them as follows:

Type 1: Cases with complete incongruence in identifications across methods (27 specimens)

Type 2: Molecular methods provided higher taxonomic resolution than morphology (36 specimens)

Type 3: Morphology provided higher taxonomic resolution while molecular methods provided inconclusive or no identification at all (40 specimens)

Meticulous revision of our molecular and morphological data revealed that all type-1 discrepancies were caused by misidentifications that were performed by the senior author (Viktor Baranov), which involves 9% of all voucher specimens. For another 9% of vouchers, morphological identifications could not provide identifications at the species-level (type-2), meaning that for a total of 18% of vouchers, morphology did not provide accurate or comprehensive species-level identifications.

On the other hand, morphological identification methods did provide more comprehensive species information for a total of 40 specimens (14%). Here, we were able to provide species-level IDs for five BINs that did not provide public data on BOLD, and for 6 BINs that were linked to discrepant identifications by taxonomists.

Uncovering species diversity from morphotypes

Of the 48 morphotypes that we distinguished during sorting, we identified 77 species (including misidentifications) using morphology and 78 BINs using molecular methods (*Table 3*). The most abundant (and thus higher sampled) morphotypes within our samples were “MT Glyptotendipes”, “MT Parachironomus”, “MT Paratanytarsus / Rheotanytarsus”, “MT Cladopelma / Cryptochironomus / Harnischia", and "MT Cricotopus". These morphotypes encompass 42% (125) of all analyzed specimens. Species identification, revealed that each of these morphotypes comprise 4-7 different taxonomic entities. In 15 cases, more BINs than morphologically identified species were recovered per morphotype. Morphotypes that include morphological misidentifications are in red font.

We created accumulation curves based on our morphological (*Fig. 3a*) and molecular data (*Fig. 3b*), depicting the number of recovered taxonomic entities for the most diverse morphotypes (with at least 4 taxonomic entities), and extrapolating to double the sampling effort. Most morphotypes that depict an accumulation curve, reach an asymptote. Comparing graphs, we see that in some cases, too many species were identified morphologically per morphotype (see “MT Tanytarsus” and “MT Glyptotendipes”) and too few in others (see “MT Paratanytarsus / Rheotanytarsus”).

Discussion

In this study, we applied an integrative approach to facilitate sample processing of highly diverse non-biting midges. We applied a three-level subsampling technique and compared species recovered with each identification method (molecular and morphological) with the goal of assessing how an integrative approach can increase the incorporation of the Chironomidae into monitoring programs and biodiversity studies using a simplified approach (but without losing too much species information).

Morphotype sorting

Our results suggest that our morphotype sorting method was successful: We obtained a coverage of over 90% in species and cluster counts (*Table. 1*), and the plateauing accumulation curves in *Figure 2* indicate that we would not have captured substantially more species by increasing our sampling effort. This is interesting, because after sorting non-biting midges into morphotype

groups, we ultimately processed and identified only 7% of all specimens. Considering this, we believe that the task of grouping them into morphotypes, then selecting specimens for subsequent analysis can be easily delegated to parataxonomists. In-depth knowledge of Chironomidae morphology is not essential for this stage of sample processing, because sorting is based on observable traits such as size, coloration, venation, setation, and shapes of antennae which simply require having a good "eye" and patience (Krell, 2004; Ekrem et al., 2010). This approach was also applied by Ekrem (2010) and authors to subsample non-biting midges for analysis in their study. We are aware that in our case, sorting was not conducted by a parataxonomist, but by an experienced scientist (Ekrem, Stur & Hebert, 2010). However, our taxonomist sorted these directly from the ethanol fluid using a stereo microscope, which does not provide a high-enough resolution for distinguishing genus- or species-level morphological features, especially not in ethanol. When confronted with large numbers of especially challenging specimens, our taxonomist resorted to either mounting representatives on temporary slides for guidance, or grouping specimens in the very few genera that have distinct features even at low resolutions (e.g. *Cricotopus*, *Ablabesmyia* and *Tanypus*). Our identifications of voucher specimens recovered up to seven taxonomic entities per single morphotype, indicating that when in doubt, it is simply easier to merge more specimens into one larger morphotype and compensate by increasing the number of vouchers.

Using DNA barcoding: working with species proxies

In our study, we clustered our COI sequences using three delimitation algorithms, namely RESL, ASAP, and SpeciesIdentifier. Because the RESL algorithm and its BIN system is directly integrated into BOLD's interface, it is commonly used in DNA barcoding applications. However, there are varying opinions regarding the sole use of BINs for species delimitation (see Carstens et al., 2013; Meier et al., 2022), especially when assuming that BIN numbers are equal to species numbers in a 1:1 ratio. Therefore, as recommended by Carstens et al. (2013), we analyzed our sequence data with several delimitation methods that apply different clustering algorithms. It is important to note that regardless which method one chooses for analysis, clustering algorithms remain arbitrary. Our results indicate that all three algorithms performed well, with MOTU diversities derived from each depicting overlapping 95% confidence intervals. Overall, we obtained very comparable results for all three clustering methods. In fact, using the

NJ-tree to depict the assignment of specimens into clusters depicted almost identical results *(see Supplementary Data S1)*.

Using the RESL-algorithm led to the assignment to 77 BINs. Although BINs are a strong proxy for species boundaries (Zahiri et al., 2014; Hebert et al., 2016), it is important to keep in mind that they do not always reflect existing taxonomic systems (Raupach et al., 2010; Hausmann et al., 2013; Zahiri et al., 2014; Hawlitschek et al., 2017). Incongruences between BINs and traditional species names include multiple BIN assignments (more than one BIN is detected in a traditionally recognized species) and BIN sharing (the same BIN is detected across more than one recognized species) (Hawlitschek et al., 2017; Chimeno et al., 2022). Ideally, multiple BIN assignments would imply the presence of cryptic diversity whereas BIN sharing, which is commonly found among taxa with uncertain taxonomy or challenging species groups, is an indication for the need of species synonymization (Hausmann et al., 2013). However, ideal conditions are not the rule and there are various molecular factors (such as heteroplasmy, numts sequencing, introgression or homogenization of mtDNA haplotypes) that can challenge COI-based species identifications (Kmiec, Woloszynska & Janska, 2006; Dobson, 2007; Pamilo, Viljakainen & Vihavainen, 2007; Duron et al., 2008; Buhay, 2009; Hazkani-Covo, Zeller & Martin, 2010), making it important to incorporate morphological information whenever possible. Additionally, accurate species identification is only guaranteed provided that high quality reference libraries are being used as a backbone to analysis (Ekrem, Willassen & Stur, 2007; Chimeno et al., 2019). These, in turn, rely on the accuracy of morphological identifications conducted on voucher specimens (Ekrem, Willassen & Stur, 2007). Mistakes in reference databases are challenging to uncover, especially if one is working with molecular data only. Yet requesting taxonomists to meticulously revise identifications of vouchers is not feasible. Instead, we suggest that it is mandatory that all records uploaded to BOLD are provided with an identifier and identification method, so that others can rely on the data when no expert is available. As suggested by Brodin (2012) and authors, reference databases need to be expanded as best as possible in order to provide a better taxonomic coverage of species and their intraspecific variation (Brodin et al., 2012). Quantity, however, should not come at a cost of quality. In our case, we double-checked every molecular-based identification using a neighbor-joining tree of public sequence data of vouchers that were morphologically identified by a taxonomist and uploaded to BOLD. Sequence records that were either identified using the “BIN taxonomy

match" tool on BOLD, or that did not provide any information on the method of voucher identification whatsoever, were disregarded completely.

Discordances in our molecular dataset include multiple BINs assignments for a total of seven species, and the assignment of four interim species names. Although multiple BIN-assignments are an indication for cryptic diversity, extensive analysis is required to uncover the driving factors in the recovered genetic differences. On the other hand, interim species names are assigned to BINs when a genetic difference is detected, yet no species name can be provided. This can be an indication for the need of a taxonomic revision or a formal species description (Morinière et al., 2019; Ekrem et al., 2019). In other words: Interim species names provide species with an "intermediate name" until they obtain a formal species name. Because of this, such species can still be implemented into analyses, as in our study, because their BIN assignments act as "taxonomic handles" (see Morinière et al., 2016; Geiger et al., 2016).

The seven species involved in multiple-BIN cases are *Cladopelma viridulum*, *Polypedilum cultellatum*, *Polypedilum sordens*, *Psectrocladius oxyura*, *Psectrocladius limbatellus*, *Smittia stercoraria*, and *Smittia terrestris*. The genera that these species belong to (especially *Cladopelma*, *Polypedilum*, *Pscetrocladius* and *Smittia*) are known for their high intraspecific variations among species, challenging the discrimination of species boundaries (Pillot, 2008; Cranston, Hardy & Morse, 2012; Tang et al., 2022). These genera include species complexes whose taxonomic position is yet unsolved, and many traditional species are suspected to comprise more than one cryptic diverse species that are awaiting formal description (Lehmann, 1970; Saether, 1971; Carew, Pettigrove & Hoffmann, 2005; Song et al., 2018; Chimeno et al., 2022). Song (2018) and authors, for example, recovered a total of five BINs for *P. cultelatum* without finding any morphological discrepancies between adult specimens, and therefore concluded that they may be dealing with potential cryptic species within a species complex (Song et al., 2018). However, when Carew and authors (2005) did not find DNA marker-associated morphological variations among individuals of the genus *Cladopelma*, they realized that this was due to the fact that these variations are only present among immature stages (Carew, Pettigrove & Hoffmann, 2005).

With the increase in barcoding campaigns, more COI-data of the Chironomidae is being made publicly available. One valuable asset of DNA barcoding is the fact that different life stages of

the same species can be easily linked to one another without having to undergo larvae rearing which can be time-consuming, expensive, and for some species very challenging (Stoeckle, 2003; Blaxter, 2004; Ekrem, Willassen & Stur, 2007; Stur & Ekrem, 2011). Therefore, implementing more COI larvae data into studies would help enormously in resolving at least some taxonomic uncertainties (Carew, Pettigrove & Hoffmann, 2005; Sinclair & Gresens, 2008; Montagna et al., 2016).

Using morphology for species delimitation

In contrast to molecular identification methods, which use an algorithm for unbiased taxonomic clustering, accurate morphological identifications rely highly on (1) the availability and accuracy of species determination keys and (2) the identifier's ability to conduct identifications from an objective perspective (Ekrem et al., 2019). Chironomid identification requires extensive knowledge (which can generally only be provided by an expert) and ideally, as demonstrated by Carew (2005) and authors, more than one single life-stage (e.g. adults) of a single species should be assessed. Unfortunately, taxonomic expertise is overall in steady decline especially for those working on small-bodied and less conspicuous taxa (Engel et al., 2021; Chimeno et al., 2022). Still, the availability of a taxonomist does not automatically guarantee error-free species identifications, as demonstrated in this and other studies (Failla et al., 2016). Not only did we have a 9% error rate among morphological identifications, six of the "single species morphotypes" that were said to be distinguishable enough under the stereo microscope for direct species assignment were incorrectly identified. For another 9% of specimens, we could only provide identification to the family or to the genus-level.

False identifications were almost always within a given genus, hence, between closely related species whose morphological differences are often very subtle and therefore require specimen mounting and meticulous analysis (Ekrem, Stur & Hebert, 2010). For diverse morphotypes, the number of taxonomic entities recovered using morphology was often over- or underestimated. This reflects the fact that on one hand, these taxa can display high levels of intraspecific morphological variation (Carew et al., 2007; Carew, Marshall & Hoffmann, 2011), and on the other hand, closely related species exhibit strong similarities, leading to the erroneous synonymization of species (Anderson, Stur & Ekrem, 2013). Despite having drastically reduced our taxonomist's workload by analyzing only a small portion of collected individuals, our

taxonomist still spent about 500 active working hours processing, mounting, and identifying specimens, which was prone to errors over time (person. comment Baranov). This is a stark contrast to the 63 working hours for our molecular approach. Although females are known to be even more difficult to identify than males, misidentifications were much more frequent among male individuals (70% of all type-1 discrepancies).

Overall, despite applying a three-level subsampling approach, which reduced the processing workload drastically, the performance of our taxonomist was affected by mistakes, caused by large amounts of material. These large amounts of material, however, represent the everyday life conditions in ecological surveys. For almost 20% of selected vouchers, no species-level information was provided, and we therefore conclude that it is difficult to meet the requirements of ecological studies using morphology alone.

Conclusion

Our current contribution shows that while both morphological identification and DNA barcoding have their own limitations, they are highly complementary in tackling large insect samples. While DNA barcoding does not require difficult-to-acquire taxonomic knowledge and drastically fast-forwards the process of identification of non-biting midges, barcode registries are only as valuable as the quality of their vouchers. Hence, without morphological identifications, there is no DNA barcoding. We presented one way to apply an integrative approach on Chironomidae, and presented a three-level sorting method for large samples. We were able to demonstrate that DNA barcoding less than 10% of a sample's contents can reliably detect >90 % of its diversity, bringing us one step closer towards optimizing processing workflows for very large insect samples.

Acknowledgements

We are very grateful to the German Barcode of Life initiative (GBOL II) for providing the opportunity of specimen sequencing. We thank Martin Spies for the help with voucher verification, and Jörg Lewandowski (IGB Berlin) for supporting Viktor Baranov during his PhD, which allowed for side projects (such as this one) to take place.

References

Andersen T, Baranov V, Hagenlund LK, Ivković M, Kvifte GM, Pavlek M. 2016. Blind Flight? A New Troglobiotic Orthoclad (Diptera, Chironomidae) from the Lukina Jama – Trojama Cave in Croatia. *PLOS ONE* 11:e0152884. DOI: 10.1371/journal.pone.0152884.

Anderson AM, Stur E, Ekrem T. 2013. Molecular and morphological methods reveal cryptic diversity and three new species of Nearctic Micropsectra (Diptera:Chironomidae). *Freshwater Science* 32:892–921. DOI: 10.1899/12-026.1.

Antczak-Orlewska O, Płóciennik M, Sobczyk R, Okupny D, Stachowicz-Rybka R, Rzodkiewicz M, Siciński J, Mroczkowska A, Krąpiec M, Słowiński M, Kittel P. 2021. Chironomidae Morphological Types and Functional Feeding Groups as a Habitat Complexity Vestige. *Frontiers in Ecology and Evolution* 8.

Armitage PD, Pinder LC, Cranston PS. 2012. *The Chironomidae: Biology and ecology of non-biting midges*. Springer Science & Business Media.

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

Baloğlu B, Clews E, Meier R. 2018. NGS barcoding reveals high resistance of a hyperdiverse chironomid (Diptera) swamp fauna against invasion from adjacent freshwater reservoirs. *Frontiers in Zoology* 15:31. DOI: 10.1186/s12983-018-0276-7.

Baranov VA. 2011a. A preliminary annotated checklist of non-biting midges (Diptera, Chironomidae) of Ukraine. *Ukrainska Entomofaunistyka* 2:7–24.

Baranov VA. 2011b. New and rare species of Orthocladiinae (Diptera, Chironomidae) from the Crimea, Ukraine. *Вестник зоологии*.

Baranov VA. 2013. First records of several tanypodinae species (Diptera, Chironomidae) from Ukraine. *Вестник зоологии*.

Baranov V, Andersen T, Hagenlund LK. 2015. A new species of Bryophaenocladius Thienemann, 1934 (Diptera, Chironomidae, Orthocladiinae) from Baltic amber. *Norwegian Journal of Entomology* 62:53–56.

Baranov V, Andersen T, Perkovsky EE. 2015. Orthoclads from Eocene Amber from Sakhalin (Diptera: Chironomidae, Orthocladiinae). *Insect Systematics & Evolution* 46:359–378.

Baranov V, Gi\lka W, Zakrzewska M, Jarzembowski E. 2019. New non-biting midges (Diptera: Chironomidae) from Lower Cretaceous Wealden amber of the Isle of Wight (UK). *Cretaceous Research* 95:138–145.

Baranov V, Góral T, Ross A. 2017. A new genus of Buchonomyiinae (Diptera, Chironomidae) from Upper Cretaceous Burmese amber, with the phylogeny of the subfamily revisited. *Cretaceous Research* 79:146–152.

Baranov V, Lewandowski J, Krause S. 2016. Bioturbation enhances the aerobic respiration of lake sediments in warming lakes. *Biology Letters* 12:20160448. DOI: 10.1098/rsbl.2016.0448.

Baranov V, Perkovsky EE. 2013. New chironomids from Eocene Sakhalinian amber (Diptera; Chironomidae; Orthocladiinae). *Terrestrial Arthropod Reviews* 6:61–69.

Baranov VA, Przhiboro AA. 2014. New records of non-biting midges (Diptera: Chironomidae) from springs and streams of the Ukrainian Carpathians (Gorgany Massif). *Zoosystematica Rossica* 23:150–157.

Blaxter ML. 2004. The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. DOI: 10.1098/rstb.2003.1447.

Bohan DA, Vacher C, Tamaddoni-Nezhad A, Raybould A, Dumbrell AJ, Woodward G. 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution* 32:477–487. DOI: 10.1016/j.tree.2017.03.001.

Brodin Y, Ejdung G, Strandberg J, Lyrholm T. 2012. Improving environmental and biodiversity monitoring in the Baltic Sea using DNA barcoding of Chironomidae (Diptera). *Molecular Ecology Resources*:n/a-n/a. DOI: 10.1111/1755-0998.12053.

Brodin Y, Lundström JO, Paasivirta L. 2008. Tavastia yggdrasilia, a new orthoclad midge (Diptera: Chironomidae) from Europe. *Aquatic Insects* 30:261–267. DOI: 10.1080/01650420802448515.

Buhay JE. 2009. "COI-like" Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. *Journal of Crustacean Biology* 29:96–110. DOI: 10.1651/08-3020.1.

Carew ME, Marshall SE, Hoffmann AA. 2011. A combination of molecular and morphological approaches resolves species in the taxonomically difficult genus Procladius Skuse (Diptera: Chironomidae) despite high intra-specific morphological variation. *Bulletin of Entomological Research* 101:505–519. DOI: 10.1017/S000748531100006X.

Carew ME, Pettigrove V, Cox RL, Hoffmann AA. 2007. DNA identification of urban Tanytarsini chironomids (Diptera:Chironomidae). *Journal of the North American Benthological Society* 26:587–600. DOI: 10.1899/06-120.1.

Carew ME, Pettigrove V, Hoffmann AA. 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* 98:587–594. DOI: 10.1603/0013-8746(2005)098[0587:TUODMI]2.0.CO;2.

Chan A, Chiang L-P, Hapuarachchi HC, Tan C-H, Pang S-C, Lee R, Lee K-S, Ng L-C, Lam-Phua S-G. 2014. DNA barcoding: complementing morphological identification of mosquito species in Singapore. *Parasites & Vectors* 7:569. DOI: 10.1186/s13071-014-0569-4.

Chao A, Ma KH, Chiu TCH and C-H. 2016. SpadeR: Species-Richness Prediction and Diversity Estimation with R.

Chimeno C, Hausmann A, Schmidt S, Raupach MJ, Doczkal D, Baranov V, Hübner J, Höcherl A, Albrecht R, Jaschhof M, Haszprunar G, Hebert PDN. 2022. Peering into the darkness: DNA barcoding reveals surprisingly high diversity of unknown species of Diptera (Insecta) in Germany. *Insects* 13:82. DOI: 10.3390/insects13010082.

Chimeno C, Hübner J, Seifert L, Morinière J, Bozicevic V, Hausmann A, Schmidt S, Müller J. Depicting environmental gradients from Malaise trap samples: Is ethanol-based DNA metabarcoding enough? *Insect Conservation and Diversity* n/a. DOI: 10.1111/icad.12609.

586 Chimeno C, Morinière J, Podhorna J, Hardulak L, Hausmann A, Reckel F, Grunwald JE,
587 Penning R, Haszprunar G. 2019. DNA Barcoding in Forensic Entomology – Establishing a DNA
588 Reference Library of Potentially Forensic Relevant Arthropod Species,. *Journal of Forensic*
589 *Sciences* 64:593–601. DOI: 10.1111/1556-4029.13869.

590 Cranston P. 2008. Keys to The Adult Male Chironomidae of Britain and Ireland - Edited by P.H.
591 Langton and L.C.V. Pinder and A Guide to The Identification of Genera of Chironomid Pupal
592 Exuviae Occurring in Britain and Ireland (Including Common Genera From Northern Europe)
593 and Their Use In Monitoring Lotic and Lentic Fresh Waters - Edited by R.S. Wilson and L.P.
594 Ruse. *Systematic Entomology* 33:215–217. DOI: 10.1111/j.1365-3113.2007.00412_5.x.

595 Cranston PS, Ang YC, Heyzer A, Lim RBH, Wong WH, Woodford JM, Meier R. 2013. THE
596 NUISANCE MIDGES (DIPTERA: CHIRONOMIDAE) OF SINGAPORE'S PANDAN AND
597 BEDOK RESERVOIRS. :15.

598 Cranston PS, Hardy NB, Morse GE. 2012. A dated molecular phylogeny for the Chironomidae
599 (Diptera). *Systematic Entomology* 37:172–188. DOI: 10.1111/j.1365-3113.2011.00603.x.

600 Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*
601 85:407–415. DOI: 10.1111/j.1095-8312.2005.00503.x.

602 Dobson SL. 2007. Evolution of Wolbachia cytoplasmic incompatibility. *Evolution* 58:2156–
603 2166.

604 Dorić V, Pozojević I, Vučković N, Ivković M, Mihaljević Z. 2021. Lentic chironomid
605 performance in species-based bioassessment proving: High-level taxonomy is not a dead end in
606 monitoring. *Ecological Indicators* 121:107041. DOI: 10.1016/j.ecolind.2020.107041.

607 Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GD. 2008. The
608 diversity of reproductive parasites among arthropods: Wolbachiado not walk alone. *BMC*
609 *Biology* 6:27. DOI: 10.1186/1741-7007-6-27.

610 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
611 throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.

612 Ekrem T. 2002a. A Revision of the Tanytarsus Eminulus, Gregarius, Lugens and Mendax
613 Species Groups (Insecta: Diptera: Chironomidae). PhD Thesis Thesis. University of Bergen,
614 Museum of Zoology.

615 Ekrem T. 2002b. A review of selected South- and East Asian Tanytarsus v.d. Wulp (Diptera:
616 Chironomidae). *Hydrobiologia* 474:1–39. DOI: 10.1023/A:1016527603086.

617 Ekrem T. 2007. A taxonomic revision of the genus Stempellinella (Diptera: Chironomidae).
618 *Journal of Natural History* 41:1367–1465. DOI: 10.1080/00222930701437360.

619 Ekrem T, Stur E. 2009. A review of the genus Zavrelia (Diptera: Chironomidae). *European*
620 *Journal of Entomology* 106:119–144. DOI: 10.14411/eje.2009.016.

621 Ekrem T, Stur E, Hebert PDN. 2010. Females do count: Documenting Chironomidae (Diptera)
622 species diversity using DNA barcoding. *Organisms Diversity & Evolution* 10:397–408. DOI:
623 10.1007/s13127-010-0034-y.

624 Ekrem T, Stur E, Orton MG, Adamowicz SJ. 2019. DNA barcode data reveal biogeographic
625 trends in Arctic non-biting midges. *Trends in DNA Barcoding and Metabarcoding* 01:787–796.
626 DOI: 10.1139/gen-2018-0100@gen-dna.issue01.

627 Ekrem T, Willassen E, Stur E. 2007. A comprehensive DNA sequence library is essential for
628 identification with DNA barcodes. *Molecular Phylogenetics and Evolution* 43:530–542. DOI:
629 10.1016/j.ympev.2006.11.021.

630 Engel MS, Ceríaco LMP, Daniel GM, Dellapé PM, Löbl I, Marinov M, Reis RE, Young MT,
631 Dubois A, Agarwal I, Lehmann A. P, Alvarado M, Alvarez N, Andreone F, Araujo-Vieira K,
632 Ascher JS, Baêta D, Baldo D, Bandeira SA, Barden P, Barrasso DA, Bendifallah L, Bockmann
633 FA, Böhme W, Borkent A, Brandão CRF, Busack SD, Bybee SM, Channing A, Chatzimanolis S,
634 Christenhusz MJM, Crisci JV, D'elía G, Da Costa LM, Davis SR, De Lucena CAS, Deuve T,
635 Fernandes Elizalde S, Faivovich J, Farooq H, Ferguson AW, Gippoliti S, Gonçalves FMP,
636 Gonzalez VH, Greenbaum E, Hinojosa-Díaz IA, Ineich I, Jiang J, Kahono S, Kury AB, Lucinda
637 PHF, Lynch JD, Malécot V, Marques MP, Marris JWM, Mckellar RC, Mendes LF, Nihei SS,
638 Nishikawa K, Ohler A, Orrico VGD, Ota H, Paiva J, Parrinha D, Pauwels OSG, Pereyra MO,
639 Pestana LB, Pinheiro PDP, Prendini L, Prokop J, Rasmussen C, Rödel M-O, Rodrigues MT,

Rodríguez SM, Salatnaya H, Sampaio Í, Sánchez-García A, Shebl MA, Santos BS, Solórzano-Kraemer MM, Sousa ACA, Stoev P, Teta P, Trape J-F, Dos Santos CV-D, Vasudevan K, Vink CJ, Vogel G, Wagner P, Wappler T, Ware JL, Wedmann S, Zacharie CK. 2021. The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean Society* 193:381–387. DOI: 10.1093/zoolinnean/zlab072.

Failla AJ, Vasquez AA, Hudson P, Fujimoto M, Ram JL. 2016. Morphological identification and COI barcodes of adult flies help determine species identities of chironomid larvae (Diptera, Chironomidae). *Bulletin of Entomological Research* 106:34–46. DOI: 10.1017/S0007485315000486.

Gadawski P, Montagna M, Rossaro B, Giłka W, Pešić V, Grabowski M, Magoga G. 2022. DNA barcoding of Chironomidae from the Lake Skadar region: Reference library and a comparative analysis of the European fauna. *Diversity and Distributions*:1–20. DOI: 10.1111/ddi.13504.

Geiger MF, Moriniere J, Hausmann A, Haszprunar G, Wägele W, Hebert PDN, 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*. DOI: 10.3897/BDJ.4.e10671.

Giłka W. 2011. *Ochotkowate - Chironomidae : plemię Tanytarsini, postaci dorosłe, samce / oprac. Wojciech Giłka*. Biologica Silesiae.

Giłka W, Gadawski P. 2022. Tanytarsini (Diptera: Chironomidae) collected during Polish expeditions to North Korea. *Zootaxa* 5214:151–175. DOI: 10.11646/zootaxa.5214.2.1.

Gratton C, Zanden MJV. 2009. Flux of aquatic insect productivity to land: comparison of lentic and lotic ecosystems. *Ecology* 90:2689–2699. DOI: 10.1890/08-1546.1.

Hartop E, Srivathsan A, Ronquist F, Meier R. 2022. Towards Large-Scale Integrative Taxonomy (LIT): Resolving the Data Conundrum for Dark Taxa. *Systematic Biology* 71:1404–1422. DOI: 10.1093/sysbio/syac033.

Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, Nieukerken EJ van, Ratnasingham S, Hebert PDN. 2013. Genetic Patterns in European Geometrid Moths Revealed by the Barcode Index Number (BIN) System. *PLOS ONE* 8:e84518. DOI: 10.1371/journal.pone.0084518.

668 Hausmann A, Krogmann L, Peters R, Rduch V, Schmidt S. 2020. GBOL III: DARK TAXA.
669 *iBOL Barcode Bulletin* 10. DOI: 10.21083/ibol.v10i1.6242.

670 Hawlitschek O, Morinière J, Lehmann GUC, Lehmann AW, Kropf M, Dunz A, Glaw F,
671 Detcharoen M, Schmidt S, Hausmann A, Szucsich NU, Caetano-Wyler SA, Haszprunar G. 2017.
672 DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with
673 focus on Austria, Germany and Switzerland. *Molecular Ecology Resources* 17:1037–1053. DOI:
674 10.1111/1755-0998.12638.

675 Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular Poltergeists: Mitochondrial DNA
676 Copies (numts) in Sequenced Nuclear Genomes. *PLOS Genetics* 6:e1000834. DOI:
677 10.1371/journal.pgen.1000834.

678 Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA
679 barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:313–
680 321. DOI: 10.1098/rspb.2002.2218.

681 Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA,
682 Pedersen S, Jannetta P, deWaard JR. 2016. Counting animal species with DNA barcodes:
683 Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*
684 371:20150333. DOI: 10.1098/rstb.2015.0333.

685 Herren CM, Webert KC, Drake MD, Jake Vander Zanden M, Einarsson Á, Ives AR, Gratton C.
686 2017. Positive feedback between chironomids and algae creates net mutualism between benthic
687 primary consumers and producers. *Ecology* 98:447–455. DOI: 10.1002/ecy.1654.

688 Hilsenhoff WL, Thorp JH, Covich AP. 2001. Ecology and classification of North American
689 freshwater invertebrates. *Diversity and Classification of Insects and Collembola*:661–721.

690 Hirvenoja M. 1973. Revision der Gattung Cricotopus van der Wulp und ihrer Verdandten. *Ann.*
691 *Zool. Fen.* 10:1–363.

692 Hölker F, Vanni MJ, Kuiper JJ, Meile C, Grossart H-P, Stief P, Adrian R, Lorke A, Dellwig O,
693 Brand A, Hupfer M, Mooij WM, Nützmann G, Lewandowski J. 2015. Tube-dwelling
694 invertebrates: tiny ecosystem engineers have large effects in lake ecosystems. *Ecological*
695 *Monographs* 85:333–351. DOI: 10.1890/14-1160.1.

696 Holzhauer S, Franke S, Kyba C, Manfrin A, Klenke R, Voigt C, Lewanzik D, Oehlert M,
697 Monaghan M, Schneider S, Heller S, Kuechly H, Brüning A, Honnen A-C, Hölker F. 2015. Out
698 of the Dark: Establishing a Large-Scale Field Experiment to Assess the Effects of Artificial Light
699 at Night on Species and Food Webs. *Sustainability* 7:15593–15616. DOI: 10.3390/su71115593.

700 Hsieh TC, Chao KHM and A. 2020. iNEXT: Interpolation and Extrapolation for Species
701 Diversity.

702 Hsieh TC, Ma KH, Chao A. 2016. iNEXT: an R package for rarefaction and extrapolation of
703 species diversity (Hill numbers). *Methods in Ecology and Evolution* 7:1451–1456. DOI:
704 https://doi.org/10.1111/2041-210X.12613.

705 Karima Z. 2021. *Chironomidae: Biology, Ecology and Systematics*. IntechOpen. DOI:
706 10.5772/intechopen.95577.

707 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
708 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic:
709 an integrated and extendable desktop software platform for the organization and analysis of
710 sequence data. *Bioinformatics (Oxford, England)* 28:1647–1649. DOI:
711 10.1093/bioinformatics/bts199.

712 Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A. 2017. Freshwater biomonitoring in the
713 Information Age. *Frontiers in Ecology and the Environment* 15:266–274. DOI:
714 10.1002/fee.1490.

715 Kirk-Spriggs AH, Sinclair BJ. 2017. *Manual of Afrotropical Diptera Volume 1*.

716 Kmiec B, Woloszynska M, Janska H. 2006. Heteroplasmy as a common state of mitochondrial
717 genetic information in plants and animals. *Current Genetics* 50:149–159. DOI: 10.1007/s00294-
718 006-0082-1.

719 Krell F-T. 2004. Parataxonomy vs. taxonomy in biodiversity studies - Pitfalls and applicability of
720 "morphospecies" sorting. *Biodiversity and Conservation* 13:795–812. DOI:
721 10.1023/B:BIOC.0000011727.53780.63.

722 Langton PH, Pinder LCV. 2007. Keys to the adult male Chironomidae of Britain and Ireland,
723 Vol 1 and 2. *Keys to the adult male Chironomidae of Britain and Ireland, Vol 1 and 2.*

724 Lehmann J. 1970. Revision der Europäischen Arten (Imagines ♂♂) der Gattung Parachironomus
725 Lenz (Diptera, Chironomidae). *Hydrobiologia* 36:129–158. DOI: 10.1007/BF00751287.

726 Lencioni V, Marziali L, Rossaro B. 2012. Chironomids as bioindicators of environmental quality
727 in mountain springs. *Freshwater Science* 31:525–541. DOI: 10.1899/11-038.1.

728 Lin X, Stur E, Ekrem T. 2015. Exploring Genetic Divergence in a Species-Rich Insect Genus
729 Using 2790 DNA Barcodes. *PLOS ONE* 10:e0138993. DOI: 10.1371/journal.pone.0138993.

730 Lundström JO, Brodin Y, Schäfer ML, Vinnersten TZP, Östman Ö. 2010. High species richness
731 of Chironomidae (Diptera) in temporary flooded wetlands associated with high species turn-over
732 rates. *Bulletin of Entomological Research* 100:433–444. DOI: 10.1017/S0007485309990472.

733 Mandelik Y, Roll U, Fleischer A. 2010. Cost-efficiency of biodiversity indicators for
734 Mediterranean ecosystems and the effects of socio-economic factors. *Journal of Applied Ecology*
735 47:1179–1188. DOI: https://doi.org/10.1111/j.1365-2664.2010.01864.x.

736 Manfrin A, Singer G, Larsen S, Weiß N, van Grunsven RHA, Weiß N-S, Wohlfahrt S,
737 Monaghan MT, Hölker F. 2017. Artificial Light at Night Affects Organism Flux across
738 Ecosystem Boundaries and Drives Community Structure in the Recipient Ecosystem. *Frontiers*
739 *in Environmental Science* 5:61. DOI: 10.3389/fenvs.2017.00061.

740 Marziali L, Armanini DG, Cazzola M, Erba S, Toppi E, Buffagni A, Rossaro B. 2010. Responses
741 of Chironomid larvae (insecta, Diptera) to ecological quality in Mediterranean river mesohabitats
742 (South Italy). *River Research and Applications* 26:1036–1051. DOI: 10.1002/rra.1303.

743 Milošević D, Milosavljević A, Predić B, Medeiros AS, Savić-Zdravković D, Stojković Piperac
744 M, Kostić T, Spasić F, Leese F. 2020. Application of deep learning in aquatic bioassessment:
745 Towards automated identification of non-biting midges. *Science of The Total Environment*
746 711:135160. DOI: 10.1016/j.scitotenv.2019.135160.

747 Montagna M, Mereghetti V, Lencioni V, Rossaro B. 2016. Integrated Taxonomy and DNA
748 Barcoding of Alpine Midges (Diptera: Chironomidae). *PLOS ONE* 11:e0149673. DOI:
749 10.1371/journal.pone.0149673.

750 Morinière J, Araujo BC de, Lam AW, Hausmann A, Balke M, Schmidt S, Hendrich L, Doczkal
751 D, Fartmann B, Arvidsson S, Haszprunar G. 2016. Species identification in Malaise trap samples
752 by DNA barcoding based on NGS technologies and a scoring matrix. *PLOS ONE* 11:e0155497.
753 DOI: 10.1371/journal.pone.0155497.

754 Morinière J, Balke M, Doczkal D, Geiger MF, Hardulak LA, Haszprunar G, Hausmann A,
755 Hendrich L, Regalado L, Rulik B, Schmidt S, Wägele J-W, Hebert PDN. 2019. A DNA barcode
756 library for 5,200 German flies and midges (Insecta: Diptera) and its implications for
757 metabarcoding-based biomonitoring. *Molecular Ecology Resources* 19:900–928. DOI:
758 https://doi.org/10.1111/1755-0998.13022.

759 Pamilo P, Viljakainen L, Vihavainen A. 2007. Exceptionally High Density of NUMTs in the
760 Honeybee Genome. *Molecular Biology and Evolution* 24:1340–1346. DOI:
761 10.1093/molbev/msm055.

762 Pfenninger M, Nowak C, Kley C, Steinke D, Streit B. 2007. Utility of DNA taxonomy and
763 barcoding for the inference of larval community structure in morphologically cryptic
764 Chironomus (Diptera) species. *Molecular Ecology* 16:1957–1968. DOI: 10.1111/j.1365-
765 294X.2006.03136.x.

766 Pillot HKM. 2008. Identification and ecology of the genus Smittia Holmgren in the Netherlands
767 (Diptera: Chironomidae). *Tijdschrift voor Entomologie* 151:245–270.

768 Pires A, Marinoni L. 2010. DNA barcoding and traditional taxonomy unified through Integrative
769 Taxonomy: a view that challenges the debate questioning both methodologies. *Biota Neotropica*
770 *(Brasil) Num.2 Vol.10* 10. DOI: 10.1590/S1676-06032010000200035.

771 Porter TM, Gibson JF, Shokralla S, Baird DJ, Golding GB, Hajibabaei M. 2014. Rapid and
772 accurate taxonomic classification of insect (class Insecta) cytochrome c oxidase subunit 1 (COI)
773 DNA barcode sequences using a naïve Bayesian classifier. *Molecular Ecology Resources*
774 14:929–942. DOI: https://doi.org/10.1111/1755-0998.12240.

775 Proulx I, Martin J, Carew M, Hare L. 2013. Using various lines of evidence to identify
776 Chironomus species (Diptera: Chironomidae) in eastern Canadian lakes. *Zootaxa* 3741:401–458.
777 DOI: 10.11646/zootaxa.3741.4.1.

778 Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System
779 (http://www.barcodinglife.org). *Molecular Ecology Notes* 7:355–364. DOI: 10.1111/j.1471-
780 8286.2007.01678.x.

781 Raunio J, Heino J, Paasivirta L. 2011. Non-biting midges in biodiversity conservation and
782 environmental assessment: Findings from boreal freshwater ecosystems. *Ecological*
783 *Indicators*:1057–1064.

784 Raupach MJ, Astrin JJ, Hannig K, Peters MK, Stoeckle MY, Wägele J-W. 2010. Molecular
785 species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear
786 rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* 7:26. DOI: 10.1186/1742-
787 9994-7-26.

788 Rosenberg DM. 1992. Freshwater biomonitoring and Chironomidae. *Netherland Journal of*
789 *Aquatic Ecology* 26:101–122. DOI: 10.1007/BF02255231.

790 Saether OA. 1971. NOMENCLATURE AND PHYLOGENY OF THE GENUS HARNISCHIA
791 (DIPTERA: CHIRONOMIDAE). *The Canadian Entomologist* 103:347–362. DOI:
792 10.4039/Ent103347-3.

793 Sæther OA. 1977. *Female Genitalia in Chironomidae and Other Nematocera: Morphology,*
794 *Phylogenies, Keys*. Department of Fisheries and the Environment, Fisheries and Marine Service.

795 Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH. 2010. Integrative
796 Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology*
797 55:421–438. DOI: 10.1146/annurev-ento-112408-085432.

798 Sheth BP, Thaker VS. 2017. DNA barcoding and traditional taxonomy: an integrated approach
799 for biodiversity conservation. *Genome* 60:618–628. DOI: 10.1139/gen-2015-0167.

800 Silva F, Wiedenbrug S. 2015. Amazonimyia gigantea gen. n., sp. N., a new Tanypodinae
801 (Diptera: Chironomidae) from the Neotropical Region. *Zootaxa* 3947:275–281. DOI:
802 10.11646/zootaxa.3947.2.9.

803 Sinclair CS, Gresens SE. 2008. Discrimination of Cricotopus species (Diptera: Chironomidae) by
804 DNA barcoding. *Bulletin of Entomological Research* 98:555–563. DOI:
805 10.1017/S0007485308005865.

806 Siri A, Brodin Y. 2014. Cladistic analysis of Rheochlus and related genera, with description of a
807 new species (Diptera: Chironomidae: Podonominae).

808 Song C, Lin X-L, Wang Q, Wang X-H. 2018. DNA barcodes successfully delimit morphospecies
809 in a superdiverse insect genus. *Zoologica Scripta* 47:311–324. DOI: 10.1111/zsc.12284.

810 Stoeckle M. 2003. Taxonomy, DNA, and the Bar Code of Life. *BioScience* 53:796–797. DOI:
811 10.1641/0006-3568(2003)053[0796:TDATBC]2.0.CO;2.

812 Stur E, Ekrem T. 2000. Tanytarsus usambarae, spec. nov. from West Usambara Mts., Tanzania,
813 East Africa (Insecta, Diptera, Chironomidae). *Spixiana* 23:219–223.

814 Stur E, Ekrem T. 2006. A revision of West Palaearctic species of the Micropsectra atrofasciata
815 species group (Diptera: Chironomidae). *Zoological Journal of the Linnean Society* 146:165–225.
816 DOI: 10.1111/j.1096-3642.2006.00198.x.

817 Stur E, Ekrem T. 2011. Exploring unknown life stages of Arctic Tanytarsini (Diptera:
818 Chironomidae) with DNA barcoding. *Zootaxa* 2743:27–39. DOI: 10.11646/zootaxa.2743.1.2.

819 Stur E, Ekrem T. 2015. A review of Norwegian Gymnometriocnemus (Diptera, Chironomidae)
820 including the description of two new species and a new name for Gymnometriocnemus volitans
821 (Goetghebuer) sensu Brundin. *ZooKeys*:127–142. DOI: 10.3897/zookeys.508.9874.

822 Stur E, Spies M. 2011. Description of Chaetocladius longivirgatus sp. n., with a review of C.
823 suecicus (Kieffer) (Diptera: Chironomidae). *Zootaxa* 2762:37–48. DOI:
824 10.11646/zootaxa.2762.1.3.

825 Stur E, Wiedenbrug S. 2005. Two new Orthoclad species (Diptera: Chironomidae) from cold
826 water springs of the Nationalpark Berchtesgaden, Germany. *Aquatic Insects* 27:125–131. DOI:
827 10.1080/01650420400019817.

828 Tang H, Cheng Q, Han W, Cranston PS. 2022. Integrative taxonomy: molecular phylogenetics of
829 Polypedilum (Cerobregma) and revisited morphology of Yaethauma and Collartomyia (Diptera:
830 Chironomidae) reveals synonymy and supports new classification. *Zoological Journal of the*
831 *Linnean Society* 194:102–119. DOI: 10.1093/zoolinnean/zlaa187.

832 Wiedenbrug S, Lamas C, Trivinho-Strixino S. 2012. A review of the genus Corynoneura
833 Winnertz (Diptera: Chironomidae) from the Neotropical region. *Zootaxa* 3574:1–61. DOI:
834 10.11646/zootaxa.3574.1.1.

835 Wiedenbrug S, Lamas C, Trivinho-Strixino S. 2013. A review of Neotropical species in
836 Thienemanniella Kieffer (Diptera, Chironomidae). *Zootaxa* 3670:215–237. DOI:
837 10.11646/zootaxa.3670.2.7.

838 Wiedenbrug S, Silva F. 2016. Diplosmittia caribensis, a new Orthocladiinae (Diptera:
839 Chironomidae) from the Dominican Republic. *Zootaxa* 4103:71. DOI:
840 10.11646/zootaxa.4103.1.8.

841 Wiederholm T. 1989. Chironomidae of the Holarctic region. Keys and diagnoses. Part 3-Adult
842 males. *Entomol. Scand. Suppl.* 34:1–532.

843 Will KW, Mishler BD, Wheeler QD. 2005. The Perils of DNA Barcoding and the Need for
844 Integrative Taxonomy. *Systematic Biology* 54:844–851. DOI: 10.1080/10635150500354878.

845 Wirta HK, Vesterinen EJ, Hambäck PA, Weingartner E, Rasmussen C, Reneerkens J, Schmidt
846 NM, Gilg O, Roslin T. 2015. Exposing the structure of an Arctic food web. *Ecology and*
847 *Evolution* 5:3842–3856. DOI: 10.1002/ece3.1647.

848 Zahiri R, Lafontaine JD, Schmidt BC, deWaard JR, Zakharov EV, Hebert PDN. 2014. A
849 Transcontinental Challenge — A Test of DNA Barcode Performance for 1,541 Species of
850 Canadian Noctuoidea (Lepidoptera). *PLOS ONE* 9:e92797. DOI: 10.1371/journal.pone.0092797.

851

Figure 1

Three-level sorting workflow that was used in this study for bulk sample processing.

For each morphotype distinguished in a bulk sample, we conducted morphological & molecular identifications of selected vouchers. The procedure was different based on the difficulty of the specimens involved in sorting.

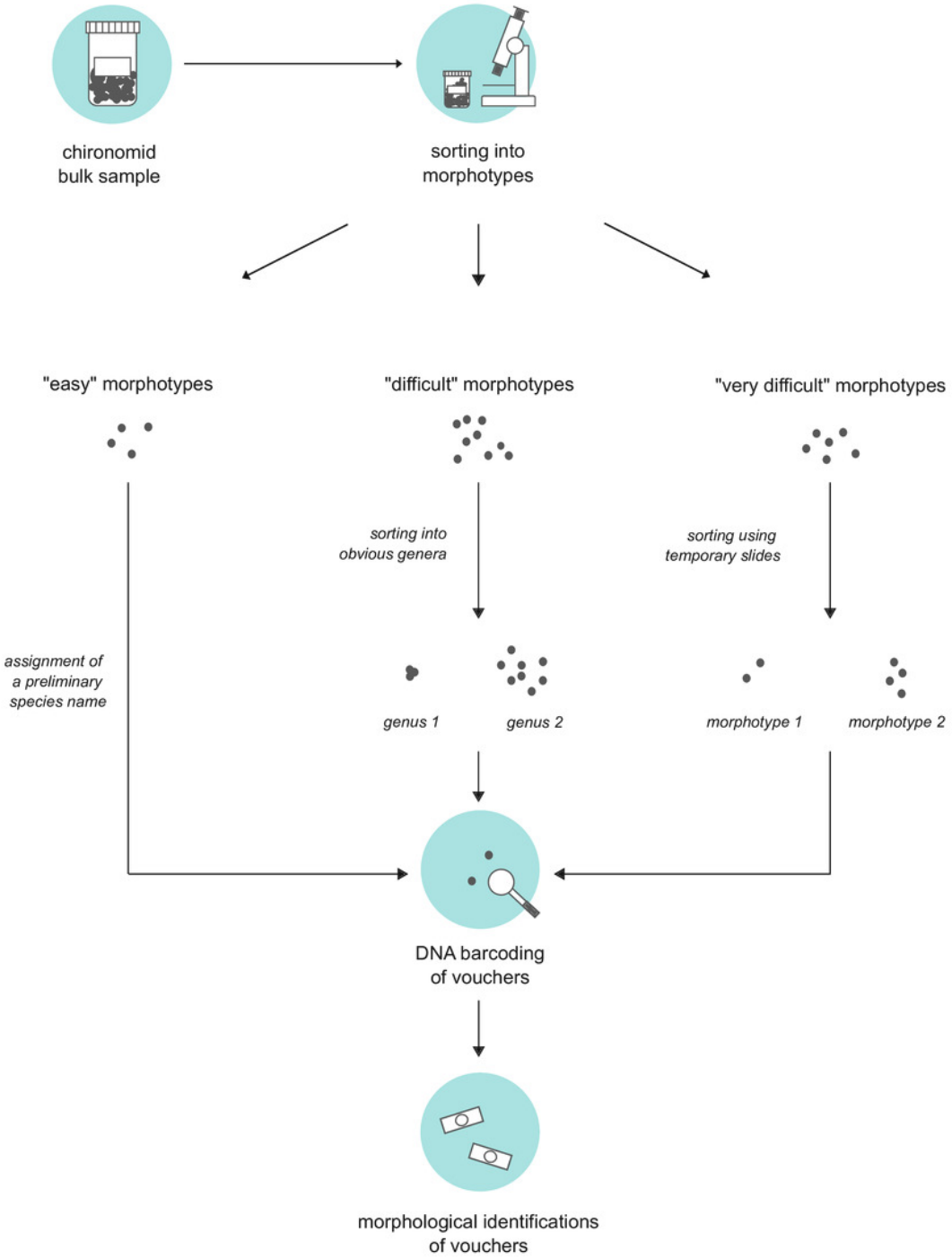
chironomid
bulk sample
sorting into
morphotypes
"easy" morphotypes
"difficult" morphotypes
"very difficult" morphotypes
sorting into
obvious genera
sorting using
temporary slides
assignment of
a preliminary
species name
genus 1
genus 2
morphotype 1
morphotype 2
DNA barcoding
of vouchers
morphological identifications
of vouchers

Figure 2

Accumulation curves of species and clusters recovered across methods.

Dotted lines represent extrapolated values (for up to double the sampling effort), bold lines represent interpolated values. Accumulation curves show the number of morphologically identified species and that of clusters recovered with RESL, ASAP, and SpeciesIdentifier.

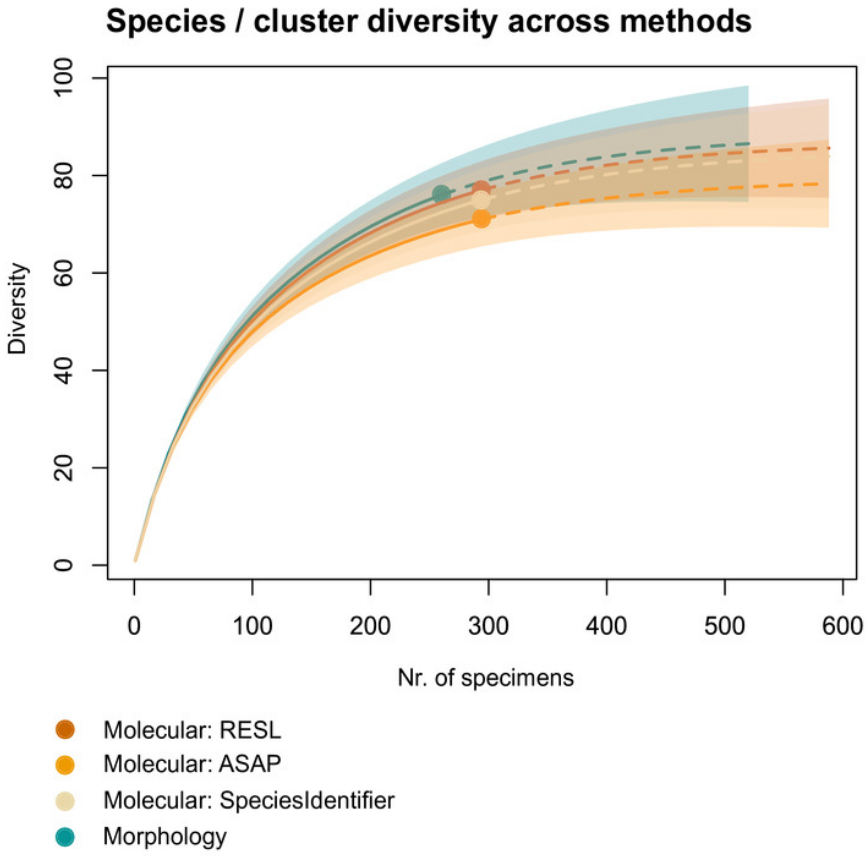
Species / cluster diversity across methods
Diversity
0
20
40
60
80
100
0
100
200
300
400
500
600
Nr. of specimens
Molecular: RESL
Molecular: ASAP
Molecular: SpeciesIdentifier
Morphology

Figure 3

Accumulation curves of (a.) species- and (b.) BIN-diversity recovered from each chironomid morphotype.

Dotted lines represent extrapolated values (up to double the sampling effort), bold lines represent interpolated values. Accumulation curve of number of morphologically identified species (a.) and BINs (b.) recovered per morphotype based on the number of sampled specimens. Multiple BIN cases have been accounted for and removed.

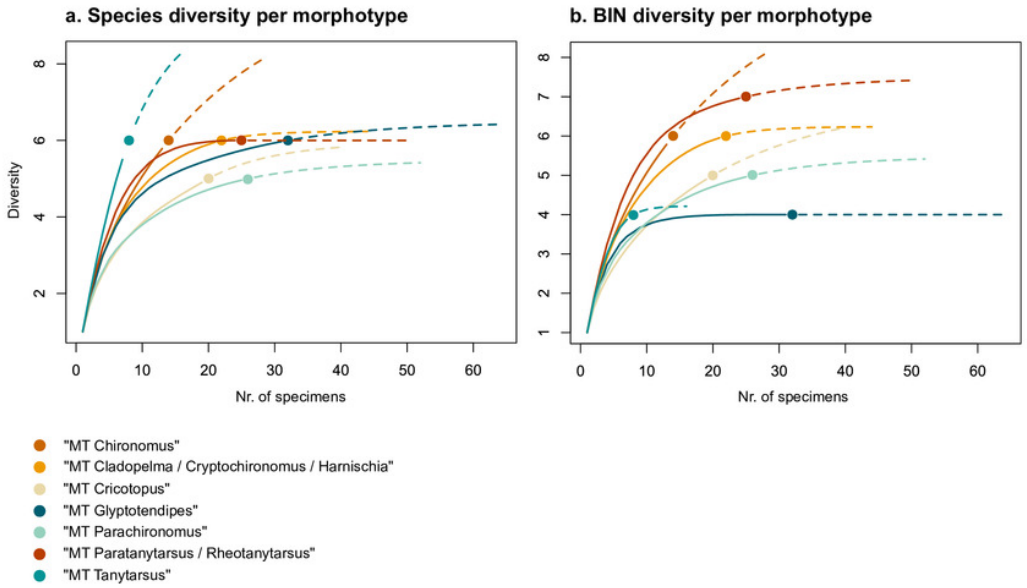
a. Species diversity per morphotype
b. BIN diversity per morphotype
Diversity
Nr. of specimens
Nr. of specimens
"MT Chironomus"
"MT Cladopelma / Cryptochironomus / Harnischia"
"MT Cricotopus"
"MT Glyptotendipes"
"MT Parachironomus"
"MT Paratanytarsus / Rheotanytarsus"
"MT Tanytarsus"

Table 1(on next page)

Chao1/2 estimates and iNEXT extrapolation values across methods.

Results after applying Chao1 and Chao2 biodiversity calculations to each datatype (morphological; molecular: RESL, ASAP, SpeciesIdentifier), including sample sizes (Nr. of specimens), taxonomical entities (Nr. of species for morphological data; clusters for molecular data), sample coverage, Chao1 and Chao2 estimates, jackknife validations, and extrapolations to double the sample size.

Method / Algorithm	Output	Values
Morphology	Sample size (n)	260
	Number of tax. entities	76
	Number of rare entities	44
	Sample coverage	0.91
	Chao1 estimate	89 ± 7 SE
	iNEXT Extrapolation (2n)	87 ± 12 SE
	Chao2 estimate	109 ± 15 SE
	Jackknife SE / bias	0.0036 / 0
Molecular: RESL	Sample size (n)	294
	Number of clusters	77
	Number of rare clusters	40
	Sample coverage	0.93
	Chao1 estimate	87 ± 6 SE
	iNEXT Extrapolation (2n)	86 ± 10 SE
	Chao2 estimate	100 ± 11 SE
	Jackknife SE / bias	0.0039 / $-2.3502e^{-14}$
Molecular: ASAP	Sample size (n)	294
	Number of clusters	71
	Number of rare clusters	34
	Sample coverage	0.94
	Chao1 estimate	79 ± 5 SE
	iNEXT Extrapolation (2n)	78 ± 9 SE
	Chao2 estimate	92 ± 11
	Jackknife SE / bias	0.0042 / 0
Molecular: SpeciesIdentifier	Sample size (n)	294
	Number of clusters	75
	Number of rare clusters	39
	Sample coverage	0.93
	Chao1 estimate	85 ± 6 SE
	iNEXT Extrapolation (2n)	84 ± 11 SE
	Chao2 estimate	98 ± 11
	Jackknife SE / bias	0.0040 / 0

1

Table 2(on next page)

Cases of discrepancies between morphological and molecular-based identifications.

Morphotypes, number of sequences, and identifications that were involved in discrepant results, namely complete incongruences in identification across methods (type 1), molecular methods provided more species-level information than morphology (type 2), and Morphology provided more species-level information while molecular methods provided inconclusive or no identification at all (type 3).

1

Discrepancy	Morphotype	Nr. of sequences	Morphological ID of specimen	BIN	Molecular ID linked to BIN
Type 1	*"Acricotopus lucens"*	2	*Acricotopus lucens*	BOLD:AAG5487	*Procladius crassinervis*
	"Chironomus"	1	*Chironomus plumosus*	BOLD:ACT6966	*Chironomus obtusidens*
	"Chironomus"	1	*Chironomus prasinatus*	BOLD:AAU4046	*Chironomus annularius*
	"Chironomus"	1	*Chironomus sp.*	BOLD:ADF1214	*Benthalia carbonaria*
	"Dicrotendipes"	1	*Dicrotendipes tritomus*	BOLD:AAU1021	*Dicrotendipes nervosus*
	"Endochironomus"	2	*Endochironomus albipennis*	BOLD:AAW5643	*Endochironomus tendens*
	"Endochironomus"	1	*Endochironomus stackelbergi*	BOLD:AAW5643	*Endochironomus tendens*
	"Glyptotendipes"	1	*Glyptotendipes cauliginellus*	BOLD:ACD4470	*Glyptotendipes pallens*
	"Glyptotendipes"	1	*Glyptotendipes glaucus*	BOLD:ACD4470	*Glyptotendipes pallens*
	"Glyptotendipes"	1	*Glyptotendipes glaucus*	BOLD:AAC0597	*Glyptotendipes paripes*

	"Parachironomus"	3	*Parachironomus gracilior*	BOLD:ACY5073	*Parachironomus monochromus*
	"Paratanytarsus / Rheotanytarsus"	1	*Paratanytarsus laetipes*	BOLD:AAI6018	*Cricotopus bicinctus*
	"Procladius ferrugineus"	2	*Procladius ferrugineus*	BOLD:AAG5487	*Procladius crassinervis*
	"Procladius pectinatus"	1	*Procladius pectinatus*	BOLD:ACW5385	*Procladius culiciformis*
	"Pseudosmittia obtusa"	1	*Pseudosmittia obtusa*	BOLD:ACP4407	*Pseudosmittia trilobata*
	"Smittia aterrima'"	2	*Smittia aterrima*	BOLD:AAN5358	*Smittia stercoraria*
	"Tanypus punctipennis"	1	*Tanypus punctipennis*	BOLD:ADJ7832	*Tanypus kraatzi*
	"Tanytarsus"	1	*Tanytarsus reei*	BOLD:ACF7553	*Tanytarsus heusdensis*
	"Tanytarsus"	2	*Tanytarsus dispar*	BOLD:ACG9929	*Tanytarsus medius*
	"Xenopelopia nigricans"	1	*Xenopelopia nigricans*	BOLD:ADJ7832	*Tanypus kraatzi*
Type 2	*"Ablabesmyia phatta"*	1	*Ablabesmyia phatta*	BOLD:ACK3818	*Ablabesmyia sp. 2ES*
	"Chironomidae"	12	*Chironomidae sp.*	BOLD:AAC0597	*Glyptotendipes paripes*

	"Cladopelma/Cryptochironomus/ Harnischia"	1	*Cladopelma sp.*	BOLD:AAV3586	*Cladopelma viridulum*
	"Cladopelma/Cryptochironomus/ Harnischia"	1	*Cladopelma sp.*	BOLD:AAV8096	*Cladopelma virescens*
	"Endochironomus"	9	*Endochironomus sp.*	BOLD:AAW5643	*Endochironomus tendens*
	"Glyptotendipes"	1	*Glyptotendipes sp.*	BOLD:ACD4470	*Glyptotendipes pallenses*
	"Psectrocladius"	1	*Psectrocladius sp.*	BOLD:AAU0273	*Psectrocladius limbatellus*
	"Smittia terrestris"	2	*Smittia terrestris*	BOLD:ACP4736	interim species *Smittia sp. 8ES*
	"Smittia terrestris"	7	*Smittia terrestris*	BOLD:ACW5117	interim species *Smittia sp. 14ES*
	"Thienemanniella"	1	*Thienemanniella vittata*	BOLD:AAV3048	interim species *Thienemanniella sp. 3TE*
Type 3	*"Acricotopus lucens"*	1	*Acricotopus lucens*	BOLD:AEO5089	no public data
	"Chironomus"	6	*Chironomus curabilis*	BOLD:ACD8415	*Chironomus curabilis / nuditarsis*
	"Cricotopus"	2	*Cricotopus sp.*	BOLD:AEO5089	no public data

	"Cricotopus"	3	*Cricotopus sylvestris*	BOLD:AAA5299	*Cricotopus sylvestris / glacialis*
	"Cricotopus"	1	*Cricotopus tricinctus*	BOLD:AEG4456	*Cricotopus tricinctus / sylvestris / trifasciatus*
	"Cricotopus"	9	*Cricotopus sylvestris*	BOLD:AAA5299	*Cricotopus sylvestris / glacialis*
	"Glyptotendipes"	4	*Glyptotendipes cauliginellus*	BOLD:AAF8348	*Glyptotendipes cauliginellus / lobiferus*
	"Metriocnemus"	1	*Metriocnemus sp.*	BOLD:ADV3586	no public data
	"Microtendipes chloris"	5	*Microtendipes chloris*	BOLD:ACY5270	*Microtendipes pedellus / chloris*
	"Parachironomus"	1	*Parachironomus sp.*	BOLD:ADV3586	no public data
	"Procladius crassinervis"	4	*Procladius crassinervis*	BOLD:ACB6320	*Procladius sp.*
	"Psectrocladius oxyura"	1	*Psectrocladius oxyura*	BOLD:AEO4348	no public data
	"Tanytarsus usmaensis"	2	*Tanytarsus usmaensis*	BOLD:AEO0788	no public data

2

3

Table 3(on next page)

Number of specimens, morphologically identified species, and BINs recovered per morphotype.

Morphotype-names are in quotation marks, and those that include morphological misidentifications are in red font.

1

Morphotype	Nr. specimens	Nr. species	Nr. BINs
"Ablabesmyia longistyla"	1	1	1
"Ablabesmyia monilis"	2	1	1
"Ablabesmyia phatta"	2	1	1
"Acricotopus lucens"	3	1	2
"Benthalia"	1	1	1
"Chironomidae"	12	0	1
"Chironomus"	14	5	6
"Cladopelma / Cryptochironomus / Harnischia"	22	5	6
"Coryneura"	6	2	2
"Cricotopus"	20	4	5
"Dicrotendipes"	5	2	1
"Endochironomus"	12	2	1
"Glyptotendipes"	32	5	4
"Guttipelopia guttipennis"	11	1	1
"Kiefferulus tendipediformis"	4	1	1
"Metriocnemus atriclava"	1	1	1
"Metriocnemus"	2	1	1
"Microchironomus"	5	1	1
"Microtendipes chloris"	5	1	1
"Microtendipes pedellus"	1	1	1
"Nanocladius dichromus"	1	1	1
"Orthocladius oblidens"	2	1	1
"Parachironomus"	26	4	5

"Paraphaenocladius impensus"	2	1	1
"Paratanytarsus / Rheotanytarsus"	25	6	7
"Polypedilum sordens"	6	1	2
"Polypedilum"	10	2	3
"Procladius crassinervis"	10	1	2
"Procladius culiciformis"	5	1	1
"Procladius ferrugineus"	2	1	1
"Procladius nigriventris"	2	1	1
"Procladius pectinatus"	1	1	1
"Procladius"	3	0	1
"Psectrocladius limbatellus"	5	1	2*
"Psectrocladius oxyura"	2	1	2*
"Psectrocladius"	1	0	1
"Pseudosmittia albipennis"	1	1	1
"Pseudosmittia obtusa"	1	1	1
"Smittia aterrima'"	2	1	1
"Smittia edwardsi"	2	1	1
"Smittia stercoraria"	1	1	1
"Smittia terrestris"	9	1	3*
"Tanypus punctipennis"	1	1	1
"Tanypus vilipennis"	1	1	1
"Tanytarsus usmaensis"	2	1	1
"Tanytarsus"	8	6	4
"Thienemanniella"	1	1	1
"Xenopelopia nigricans"	1	1	1
Total	**294**	**76**	**88**

*includes multiple BINs