



ORIGINAL RESEARCH

The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates

Yihao Ge^{1,2,3} | Chengxing Xia^{1,2} | Jun Wang^{1,3}  | Xiujie Zhang^{1,2} | Xufa Ma^{1,2}  | Qiong Zhou^{1,2} 

¹Key Laboratory of Freshwater Animal Breeding, Ministry of Agriculture and Rural Affairs/Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, College of Fisheries, Huazhong Agricultural University, Wuhan, China

²Engineering Research Center of Green Development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt, Ministry of Education, Wuhan, China

³The Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

Correspondence

Qiong Zhou, College of Fisheries, Huazhong Agricultural University, Wuhan 430070, China.

Email: hainan@mail.hzau.edu.cn

Funding information

Special Funds for the Foundation Work of Science and Technology, Grant/Award Number: 2012FY112700; Finance Special Fund of the Ministry of Agriculture and Rural Affairs (Fisheries Resources and Environment Survey in the Key Water Areas of Northwest China)

Abstract

Macroinvertebrates have been recognized as key ecological indicators of aquatic environment and are the most commonly used approaches for water quality assessment. However, species identification of macroinvertebrates (especially of aquatic insects) proves to be very difficult due to the lack of taxonomic expertise in some regions and can become time-consuming. In this study, we evaluated the feasibility of DNA barcoding for the classification of benthic macroinvertebrates and investigated the genetic differentiation in seven orders (Insecta: Ephemeroptera, Plecoptera, Trichoptera, Diptera, Hemiptera, Coleoptera, and Odonata) from four large transboundary rivers of northwest China and further explored its potential application to biodiversity assessment. A total of 1,144 COI sequences, belonging to 176 species, 112 genera, and 53 families were obtained and analyzed. The barcoding gap analysis showed that COI gene fragment yielded significant intra- and interspecific divergences and obvious barcoding gaps. NJ phylogenetic trees showed that all species group into monophyletic species clusters whether from the same population or not, except two species (*Polypedilum. laetum* and *Polypedilum. bullum*). The distance-based (ABGD) and tree-based (PTP and MPTP) methods were utilized for grouping specimens into Operational Taxonomic Units (OTUs) and delimiting species. The ABGD, PTP, and MPTP analysis were divided into 177 ($p = .0599$), 197, and 195 OTUs, respectively. The BIN analysis generated 186 different BINs. Overall, our study showed that DNA barcoding offers an effective framework for macroinvertebrate species identification and sheds new light on the biodiversity assessment of local macroinvertebrates. Also, the construction of DNA barcode reference library of benthic macroinvertebrates in Eurasian transboundary rivers provides a solid backup for bioassessment studies of freshwater habitats using modern high-throughput technologies in the near future.

Yihao Ge and Chengxing Xia contributed equally to this work.

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.

© 2021 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

KEYWORDS

benthic macroinvertebrates, biodiversity assessment, DNA barcoding, species identification

1 | INTRODUCTION

Macroinvertebrates (e.g., aquatic insects) are considered as one of the primary components in aquatic ecosystems and represent a key group as food resources for higher trophic levels (Castella et al., 1984). They provide unique ecosystem services in nutrient cycling as well as energy flow (Sivaramakrishnan et al., 2014). Macroinvertebrates have been frequently utilized to monitor the water quality worldwide (Chandler, 1970; Helson & Willians, 2013), especially for those water bodies (e.g., rivers and lakes) affected by anthropogenic activities (Wang et al., 2006; Zhang et al., 2011). However, these monitoring functions depend, to a large extent, on the accuracy and precision of species/specimen identification (Frézal & Leblois, 2008; Macher et al., 2016). Traditional taxonomy relies on the morphology-based traits to identify a species, but this approach proves to be difficult, time-consuming, and costly (Wong et al., 2014). Although the need for species-level identification in biomonitoring is controversial (Bailey et al., 2001; Lenat & Resh, 2001), DNA barcoding can provide the option of species-level identification when taxonomic discrimination at the species level is warranted. Moreover, the increased taxonomic resolution delivered by DNA barcoding will provide more sensitive measures of the magnitudes and types of environmental impacts (Pfenninger et al., 2007). The morphological delineation for aquatic insects, in particular of females and immatures, remains a challenging task, as some important traits for reliable identification are only available in a single sex or at a certain stage of development (Zhou et al., 2009a; Zhou et al., 2011). For example, the morphological identification of the mosquito *Anopheles funestus* and its sibling species can only be distinguished at specific stages of their development (Cohuet et al., 2003). Unlike those model species such as butterflies (Dinca et al., 2015) and beetles (Kang et al., 2012), the information on aquatic insects is quite limited due to the great variety of species and complex life history (Morse et al., 1984). Despite the increasing demand for well-trained taxonomists to support the assessment of aquatic ecosystems, the number of ecologists and researchers with taxonomic expertise is decreasing (Haase et al., 2006). Accordingly, such factors may hinder the accurate description of macroinvertebrates in terms of traditional classification.

To circumvent the morphological hurdles, DNA barcoding sensu Hebert et al. (2003) utilizes a short standard sequence of the mitochondrial genome for the species identification and classification, based on a part of the cytochrome c oxidase subunit I (COI) gene (Hebert, Cywinska, et al., 2003; Moriniere et al., 2017). DNA barcoding has the potential to identify cryptic species and highlight the diversity of macroinvertebrates in aquatic ecosystems (Bucklin et al., 2010), providing valuable information on the selection of taxa for further analyses (Hajibabaei et al., 2006) and allowing an identification of all life history stages and genders of a species (Casiraghi et al., 2010; Murria et al., 2010). Thus, DNA barcoding provides an effective way to overcome the difficulties in traditional morphological delineation, as this technology delivers fast, efficient, and reliable species identification

(Kress et al., 2012). However, it is noted that various problems may affect the use of mitochondrial DNA, for example, Wolbachia infections (Werren et al., 1995), heteroplasmy (Kavar et al., 2006), introgressive hybridization (Raupach et al., 2014), incomplete lineage sorting (Petit & Excoffier, 2009), pseudogenes (Ribeiro Leite, 2012), and the recent speciation events (Raupach et al., 2014), and thus affect the efficiency of DNA barcodes to discriminate the analyzed species (Havemann et al., 2018). Unknown specimens could be identified if their DNA barcode sequences match in comparison with a given barcode reference library (Moriniere et al., 2017). Previous studies illuminate the great reliability of DNA barcoding in species identification of aquatic insects, for example, Ephemeroptera (Ball et al., 2005; Curt et al., 2012), Plecoptera and Trichoptera (Gill et al., 2014), Hemiptera (Havemann et al., 2018) and Diptera (Brodin et al., 2012; Hernandez-Triana et al., 2014; Hunter et al., 2008; Lin et al., 2015). However, these studies focus primarily on the application of DNA barcoding in one to three orders, and few studies consider all the taxa of macroinvertebrates in a specific ecosystem and further apply to the environmental and biodiversity assessment (Ball et al., 2005; Moriniere et al., 2017).

Aquatic insects are predominant faunas in the transboundary rivers of northwest China (Wang et al., 2014). Despite extraordinarily abundant resources for aquatic insects, few studies have been carried out in these transboundary rivers (Wang et al., 2014). China exhibits highly different characteristics of climate and geography that harbors an enormous diversity in aquatic biota. Thus, great variability in species distribution and high diversity in aquatic biota occur in China. Until now, DNA barcoding has been applied to some specific taxa, for example, amphibians (Che et al., 2012), birds (Yoo et al., 2013), plants (Huang & Ke, 2015), Noctuoidea moths (Yang et al., 2014), mollusks (Barco et al., 2016), crickets (Hawiltschek et al., 2017), herpetofauna (Hawiltschek et al., 2017), fish (Smith et al., 2008), spiders (Ivanov et al., 2018), but rarely focuses on aquatic invertebrates (Zhou, 2009b; Zhou et al., 2009a). In this study, we attempted to test the feasibility of DNA barcoding in the classification of benthic macroinvertebrates in the transboundary rivers and their affiliated water bodies of northwest China. The main objectives of this study were to establish a reliable DNA barcode reference library for benthic macroinvertebrates in Eurasian transboundary rivers and sheds new light on the diversity status of local macroinvertebrates.

2 | MATERIALS AND METHODS

2.1 | Study area

Our study was conducted in Xinjiang Uygur Autonomous Region (Figure 1), center of Eurasia (Figure 2). This region extends from the south slope of the Altai Mountains to the hinterland of the Tianshan Mountains. The Junggar basin between the two mountains

is covered by the Gurbantunggut desert, the second largest desert in China. This region belongs to semi-arid and arid climatic zone, with the elevation ranging from 189 to 7,435 m, and encompasses complex landforms and microclimate between mountains and basins. Due to the characteristics of geography, environment, and climate, Xinjiang region provides biological communities with a variety of habitats and has been listed as one of the priority areas for biodiversity conservation in China. The Irtysh River, Emin River, Ili River, and Bortala River originate from Xinjiang Uygur Autonomous Region, northwest China (Figure 2) and ultimately flow toward the Republic of Kazakhstan. In particular, the Ili River is the largest river in Xinjiang region. The Irtysh River is the only Chinese river that discharges into the Arctic Ocean. Ulungur Lake covers an area of 1,035 square kilometers (Wang, 2010).

2.2 | Sampling and experimental material

We collected macroinvertebrate samples in four transboundary rivers (the Irtysh River, Emin River, Ili River, and Bortala River) between China and Kazakhstan, Mongolia and Russia (Figure 2). Study sites were set in the main stems, tributaries and affiliated water bodies (lakes and reservoirs). Due to a long-frozen winter season (November to April), mayflies (Ephemeroptera), stoneflies (Plecoptera),

caddisflies (Trichoptera), true flies (Diptera), true bugs (Hemiptera), beetles (Coleoptera), and dragonflies (Odonata) were collected in May, July, August, and October of 2013–2017. Macroinvertebrates were collected by Surber net, Kick-net, Peterson grab, and D-framed dip net according to the habitat type of study areas. Standardized sampling protocols were executed at different habitats (i.e., lentic and lotic water bodies) from upstream to downstream (Stark et al., 2001). Approximately 30,000 specimens were preserved in 95% ethanol to allow for the morphological identification and molecular analyses. In some cases, ethanol was replaced for two or three times in order to guarantee a good preservation of the specimens that can be used for further molecular analysis. Macroinvertebrate samples were sorted and identified under a dissecting microscope in the laboratory, and stored at -20°C at the College of Fisheries, Huazhong Agricultural University (Brinkhurst, 1986; Dudgeon, 1999; Epler & Epler, 2001; Morse et al., 1994; Zhou et al., 2003). In the laboratory, specimens were identified based on morphological characteristics. Intact individuals were selected to conduct the DNA barcoding analysis.

2.3 | DNA extractions, amplification and sequencing

Total genomic DNA was extracted from legs, cerci, half or whole bodies according to a specimen's size, using the phenol–chloroform

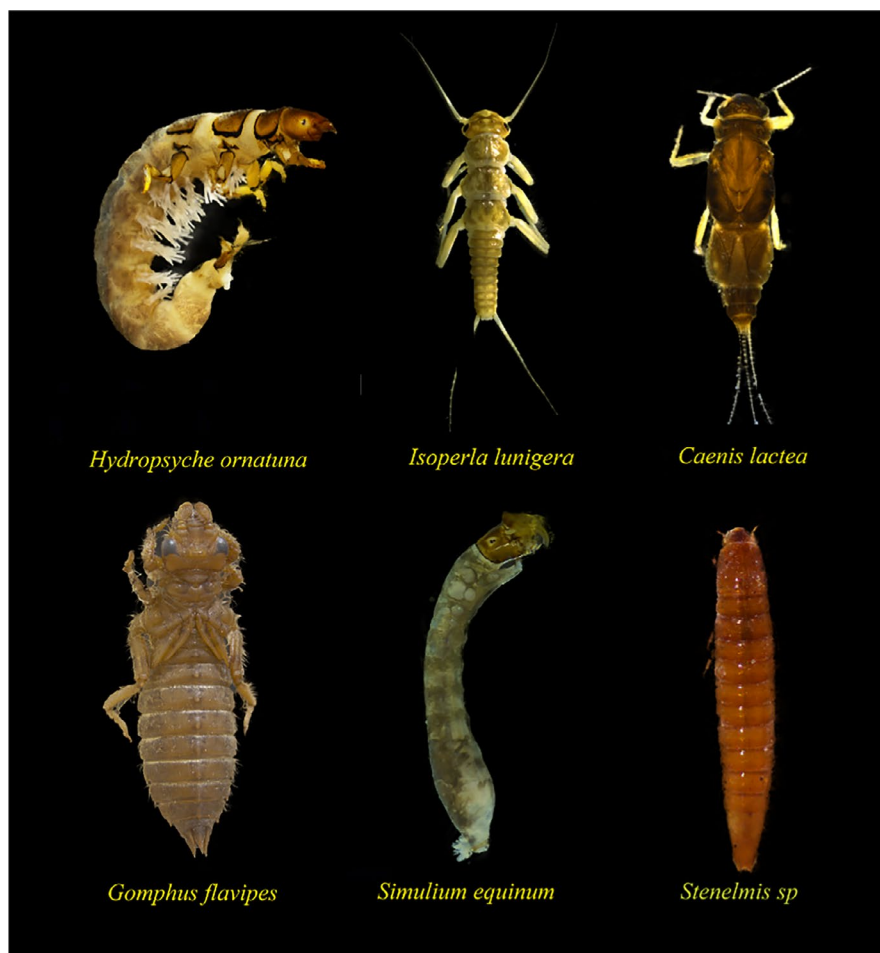


FIGURE 1 Selected common species of benthic macroinvertebrates in four transboundary rivers of northwest China

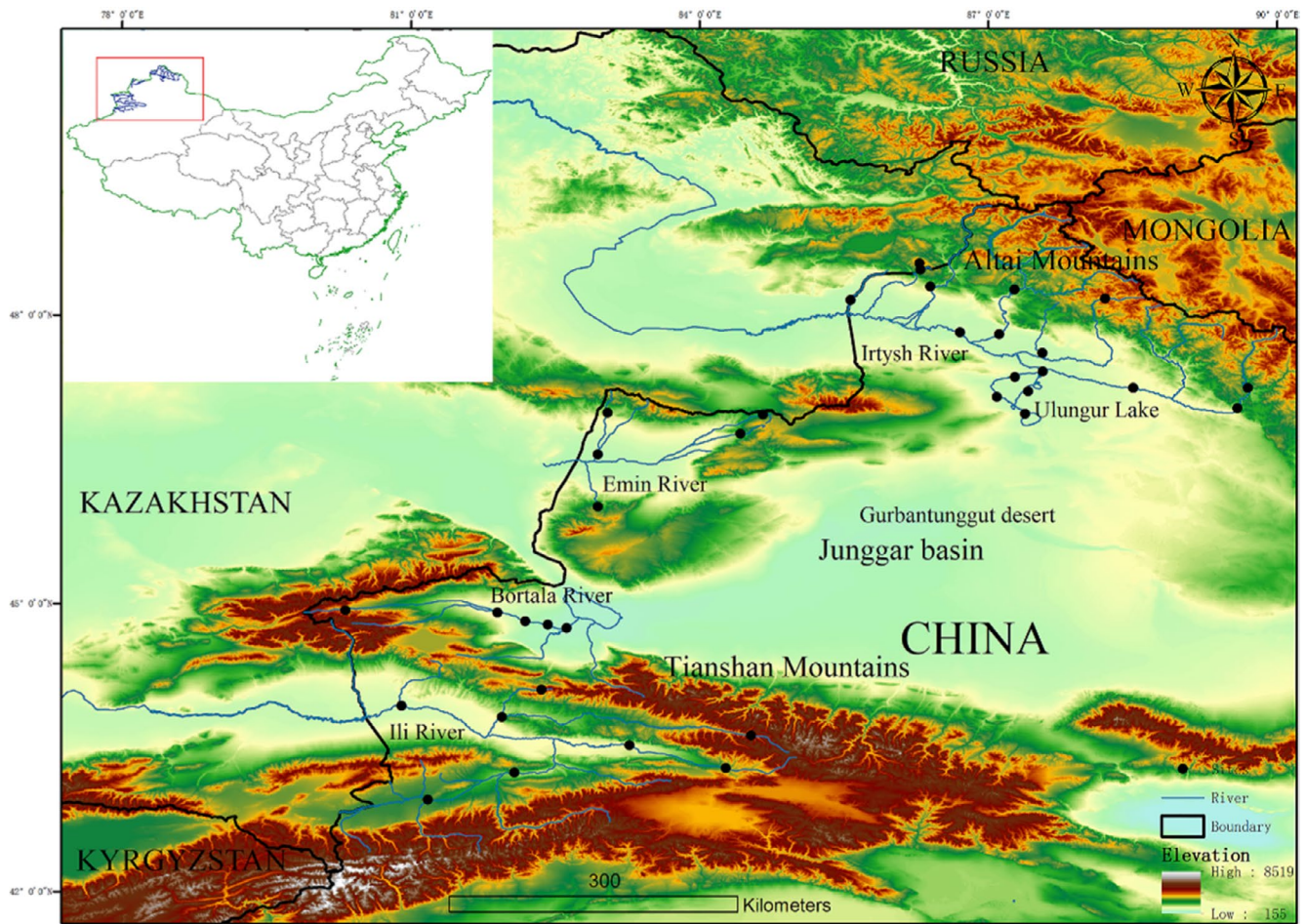


FIGURE 2 Sketch map showing the sampling locations of benthic macroinvertebrates in four transboundary rivers of northwest China. Solid black circles and solid black line represent the sampling sites of benthic macroinvertebrates and the border among different countries, respectively. The blue line in the middle of the map represents the rivers and main branches (Irtys River, Emin River, Bortala River, and Ili River)

protocol (Barnett & Larson, 2012) and quantified in a NanoDrop 2000c (Thermo Fisher Scientific, USA). The primer pair LCO-1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO-2198 (5'-TAA ACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al., 1994) was used to amplify the DNA barcode fragment with the length of about 658 base pairs (bp). Polymerase Chain Reactions were performed in 50 μ l volumes using the following reagents and concentrations: containing 10 \times PCR buffer 5 μ l, Mg^{2+} (2.5 mmol/L) 5 μ l, dNTP (2.5 mmol/L) 3.5 μ l, 1.5 μ l (10 nmol/L) each primer, 0.5 U Taq polymerase (Takara, Dalian, China), 2 μ l (50 ng/ μ l) DNA template and complete ddH₂O to 50 μ l. The PCRs were run as follows: 94°C for 3 min; then 94°C for 1 min, 45°C for 2 min, and 72°C for 3 min, for 40 cycles; and 72°C for another 5 min (Folmer et al., 1994). All PCR products were checked by electrophoresis at 1% agarose with an ethidium bromide stain, and if present, the PCR products were subsequently purified by Exol/FastAP (Fermentas) and directly sequenced by Invitrogen Corporation in China.

2.4 | Barcoding analysis based on COI

Bidirectional sequencing was employed to maximize the precision of sequencing. We aligned the COI gene using the ClustalW program from MEGA X (Kumar et al., 2018) package for each order with default parameters (Zhou et al., 2010). The online version of MAFFT v. 7.0 (Katoh & Standley, 2013) was utilized to align COI gene sequences under the algorithm Q-INS-I and the rest set as default. Amino acid translation was conducted to ensure that no gaps or stop codons existed in the alignment. We obtained 1,144 COI sequences from 176 species in total. Detailed information (locality data, habitat, altitude, collector, identifier, taxonomic classifications, habitus images, and DNA barcode sequences) for each voucher specimen was deposited to the Barcode of Life Data System at <http://www.boldsystems.org> (Ball et al., 2005), under the dataset project "XJDQD" (Process IDs: XJDQD001-18-XJDQD1275-18).

2.4.1 | Phylogenetic tree

Maximum likelihood (ML) analyses in MEGA 10.0 software was used for tree construction. When the multiple sequence alignment with ClustalW was complete, we used the “models” function to determine the best DNA/Protein Models. A total of 1,000 nonparametric bootstrap replicates were used (Felsenstein, 1985). To evaluate the accuracy of DNA barcoding for the classification and evolutionary relationship of species, we downloaded the sequencing data of relevant species from BOLD (Raupach et al., 2014; Stein et al., 2014) (Part I, Supplementary Material). A total of 1,000 replicates were utilized to detect the reliability of each branch of the tree in order to obtain bootstrap values.

2.4.2 | Barcode Gap Analysis and BIN analysis

Genetic distances within species and genera, determined using the Kimura 2-parameter (K2P) (Kimura, 1980) distance, were inputted into the MEGA X program (Kumar et al., 2018). Barcode Gap Analysis in the BOLD system was carried out to compare the distribution of distances within each species and the distance to the nearest neighbor of each species. Species were tested for the presence of barcode gaps between maximum intraspecific genetic distance and minimum interspecific distance according to Ratnasingham and Hebert (2007). Within the BOLD, the BIN system groups sequence data into clusters of closely COI barcode sequences that are assigned to a globally unique identifier, termed Barcode Index Number (BIN, Ratnasingham & Hebert, 2007). BIN analysis was restricted to sequences with a minimum length of 500 bp. Members of a BIN usually belong to a single species recognized by traditional taxonomy (Hendrich et al., 2015). Every case of disagreement/conflict is the starting point for reevaluation of both molecular and morphological data. We follow the concept of an integrative taxonomic approach to infer whether there are previously overlooked species in the samples or whether barcode divergences between species are large enough to enable delineation of species using the usual partial COI sequence (Moriniere et al., 2017). BINs for sequence clusters enable the delineation of geographical clades, which might reflect local environmental features (Moriniere et al., 2017).

2.4.3 | Automatic Barcode Gap Discovery analysis

Automatic Barcode Gap Discovery (ABGD) analysis (Puillandre et al., 2012) was used to group specimens into operational taxonomic units (OTUs) based on DNA barcodes and applied to compare the results of the BIN-based Barcode Gap Analysis, which was run by a web interface (www.abi.snv.jussieu.fr/public/abgd/). The default value for the relative gap-width was set as $x = 1.5$. Moreover, we analyzed each order in a separate analysis. All the assignments that the values (P) of intraspecific divergence ranged between 0.001 and

0.1 were recorded. Default settings were employed for all remaining parameters.

2.4.4 | Analyses of Poisson Tree Processes (PTP) and Multiple-threshold PTP (MPTP)

Phylogenetic inference analyses were conducted using maximum likelihood optimality criteria. ML analyses were conducted using RAXML 8.2.9 (Stamatakis, 2014) with 1,000 nonparametric bootstrap replicates. Appropriate substitution models were determined by JMODELTEST 2 (Darriba et al., 2012), and the best fitting substitution models were GTR + I + G. Poisson tree process (Zhang et al., 2013) was used in species delimitation analysis. Since PTPs belong to tree-based methods, they require species monophyly and are based on the analysis of branching rates (Fontaneto et al., 2015). Single-threshold Bayesian PTP analyses (Zhang et al., 2013) were conducted on the website <http://species.h-its.org/ptp>. The analyses were run for 1,000,000 MCMC generations with thinning value = 100 and burn-in = 0.25. The trace files were checked for the convergence of the MCMC. Multiple-threshold PTP (MPTP) analyses (Kapli et al., 2017) were run on the website <https://mptp.h-its.org>. Support information for the species delimitation hypotheses obtained with the ML implementation of PTP are shown in Appendix S5.

3 | RESULTS

3.1 | Sequencing analyses

A total of 1,678 individuals were collected and utilized for the generation of DNA barcodes. However, 451 specimens showed poor quality and were excluded from further analyses. Accordingly, a total of 1,144 sequences (Ephemeroptera ($n = 516$), Plecoptera (146), Trichoptera (81), Hemiptera (66), Diptera (273), Coleoptera (55), and Odonata (23)), belonging to 176 species, 112 genera, and 53 families, were successfully generated. The lengths of all COI sequences were no less than 600 bp for sequencing analysis. Deletions, insertions, and stop codons were not detected in the sequences. All the species exhibited high adenosine and thymine (AT)-rich bias (averaged 63.1%) as it is typically known from arthropods. Specifically, the A + T content in Trichoptera was up to 69.9% and the lowest was shown in Diptera (55.6%).

3.2 | Genetic distance

The level of genetic divergence in the COI genes is summarized in Table S1. Intraspecific K2P distances ranged from 0% to 15.1% with an average of 0.78%, whereas interspecific K2P distance within a genus ranged from 0.6% to 33.4% (average 16.4%). Interspecific K2P distance within one genus was 21-fold higher than those of

intraspecific K2P distances. Although the maximum intraspecific distance and the distance to the nearest neighbor overlapped partially (Figure 3), the averages of the nearest neighbor distances were 28-fold higher than those of the maximum intraspecific distances. Based on the Barcode Gap analysis, the minimum interspecific distance for 174 species was larger than the maximum intraspecific distance (98.9%; Figure 4). For the species *Polypedilum laetum* and *Polypedilum bullum*, the maximum intraspecific variations overlapped with the NN distance, leading to the absence of a barcode gap. For the species *Chironomus heterodentatus* and *Chironomus* sp. XJ, the distances to the NN were less than 2% of sequence divergence, but more than the maximum intraspecific value. We compared the means of intraspecific K2P distance for seven macroinvertebrate taxonomic groups (Table S1). Mollusks exhibited the largest mean intraspecific distance (0.89%), whereas the analyzed Odonata yielded the lowest value of 0.28%.

3.3 | Phylogenetic tree-based identification and cluster analysis

Through the ML method, phylogenetic trees of seven orders show that all individuals grouped into monophyletic species clusters with high bootstrap values except *P. laetum* and *P. bullum* (Figure 5). For those species with two or more representatives, the conspecifics in monophyletic clades were associated with high confidence (100% bootstrap in the NJ tree). Although the very low interspecific variations (0.62%) were observed, species *Chironomus* sp. XJ and *Chironomus heterodentatus* could be well distinguished based on phylogenetic analyses. An unusual cluster was revealed for *P. bullum*, where it clustered with five individuals of *P. laetum* with a 0.63% mean interspecific distance (ranged 0.21%–1.06%). Species *Dicranota guerini*, *Ameletus montanus*, *Glyptotendipes* sp. XJ, *Euryhopsis* sp, *Cricotopus ornatus*, *Atherix* sp. XJ, and *Epeorus* sp5 exhibited large

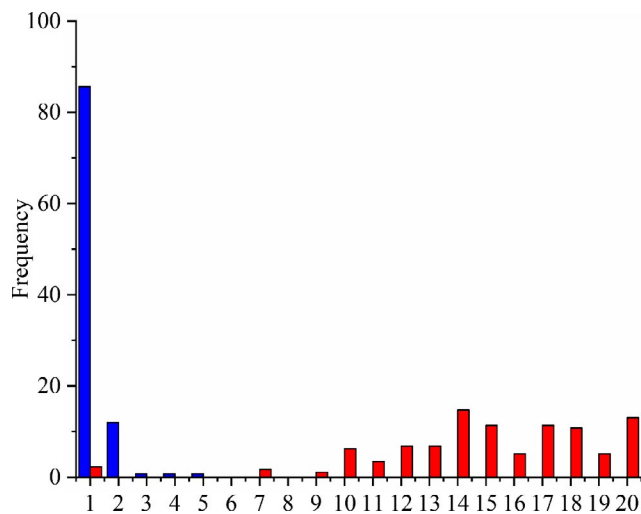


FIGURE 3 The distribution histograms of mean intraspecific distance (in blue) and the distance to nearest neighbor (in red) based on Kimura 2-parameter distance

internal splits in their monophyletic clusters (with > 90% support values), and large K2P distances were detected between the internal clusters (Figure S1). In addition, the clades of *Sigara striata* and *Chironomus pallidivittatus* were formed with some subclusters with relatively large mean K2P genetic distance (Figure S2).

The clustering analysis of 265 individual sequences from the BOLD data show that the species were clustered into species-specific groups with the homologous specimens (Figures 5a–5g). Firstly, the specimens of the same populations in our study were clustered together, and then were clustered with those from other areas (Germany, United States, Mexico, Canada, Norway, Italy, Finland, Japan, Iran etc). Based on the BLAST results, the query coverage and identity of COI sequences for the species in this study compared with those submitted in BOLD database did not reach 100%.

3.4 | OTU delineation based on distance using ABGD

ABGD software was utilized to delineate 176 morphological species and in total, 1,144 sequences with the assignments for intraspecific divergence (p) values between 0.001 and 0.1. The results included two cases: initial partition and recursive partition. As revealed in Part II (Supplementary Material), 1,144 samples were divided into 216 groups (p value: .0028–.0077) and 177 groups (p value: .0129–.0599) in two cases, respectively. The ABGD analysis derived a total of 177 OTUs with a prior intraspecific divergence of $P_{\max} = 0.0599$. Therefore, we selected the initial results to compare with those of morphological identification. We found that *A. montanus* and *D. guerini* were grouped into two different groups, whereas *Chironomus* sp. XJ and *C. heterodentatus*, *P. bullum* and *P. laetum*, *Diamesa* sp. XJC and *Diamesa* sp. XJB were clustered into the same group.

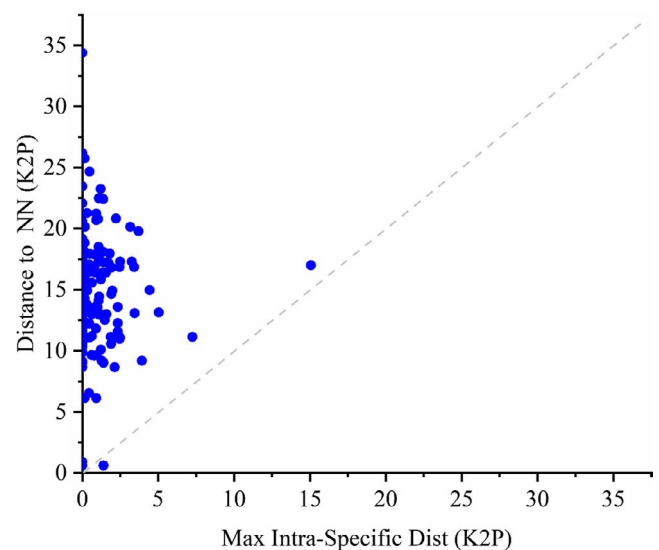
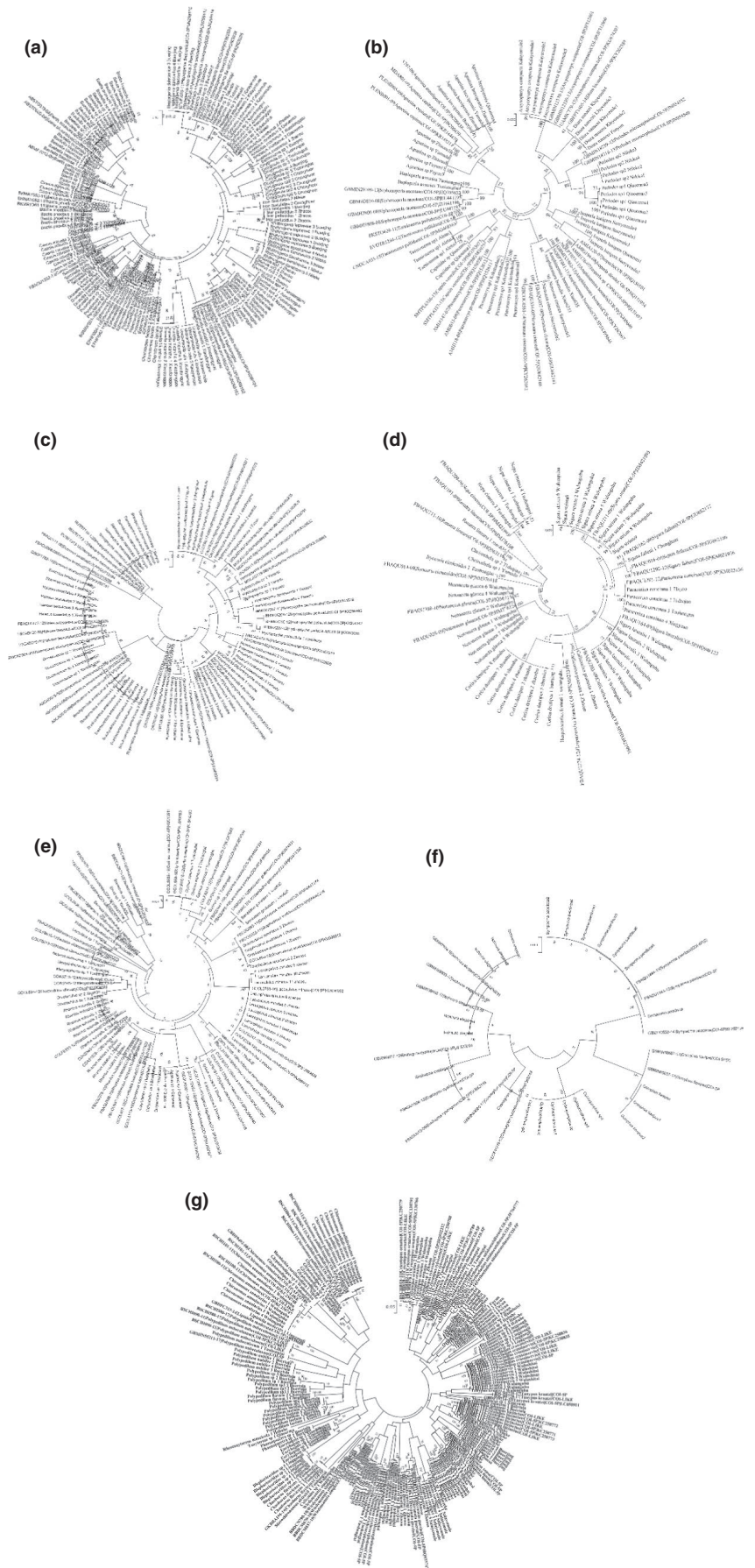


FIGURE 4 Barcode gap plot showing the distance to the nearest neighbor (NN) versus the maximum intraspecific distance Kimura 2-parameter (K2P) for 189 species. Dots above the 1:1 line indicate the presence of a barcode gap

FIGURE 5 The maximum likelihood trees of Ephemeroptera (a), Plecoptera (b), Trichoptera (c), Hemiptera (d), Coleoptera (e), Odonata (f), and Diptera (g) based on K2P distance



3.5 | Tree-based (PTP and MPTP) delimitation analysis

The groups delimited by the PTP and MPTP analysis was much more than that of recognized morphospecies. The PTP and MPTP analyses, as the implementation of ML, generated 197 and 195 MOTUs, respectively (Supplementary Material Part III). Most incongruences between the delimited groups and morphology-based identifications could ascribe to multiple molecular lineages/clusters. These incongruence cases were generally associated with high intraspecific genetic distance and were mainly reported in ten taxa (*Epeorus sp5*, *Rhithrogena tianshanica*, *Ameletus montanus*, *Atherix sp. XJ*, *Glyptotendipes sp. XJ*, *Dicranota guerini*, *Caenis lactea*, *Ischnura elegans*, and *Tabanus cordiger*).

3.6 | BIN analysis

The 1,144 records generated a total of 186 different BINs. In particular, 117 BINs were not recorded previously in the BOLD database, and most sequences representing new endemic species had new BIN assignments. The Diptera species *Chironomus sp. XJ* and *C. heterodontatus* shared the same BIN [AAW4009], and *P. bullum* shared a BIN assignment [ACB4789] with *P. laetum*. BIN Discordance analysis was performed on BOLD (February 2018), and the results showed that two BINs were discordant with our prior taxonomic assignments, which indicated shared haplotypes and a low interspecific divergence. A number of 129 BIN clusters were found to be taxonomically concordant with other barcoding data on BOLD, and they were assigned to the same species. A number of 56 records are singletons, implying that these BINs were represented by only one sequence. Eight species were assigned to a total of 19 BINs (Table S2).

4 | DISCUSSION

The species identification using DNA barcoding is based on the principle that the genetic distance between two species is much greater than that within a species. It has been proposed that 2% is the threshold value of species delimitation and, in general, the average genetic distance between two species is over 10 times of that within a species (Hebert, Cywinska, et al., 2003; Ward, 2009). In this study, the average of interspecific K2P distance (16.37%) was 21-fold higher than that of intraspecific K2P distance (0.78%), which meets the criteria that the average of interspecific genetic distance is 10 more times than that of intraspecific genetic distance. The distribution histogram of intraspecific and interspecific distances shows that 85.19% and 97.04% of the intraspecific distances were less than 1% and 2%, respectively, and 97.88% of the interspecific distances were greater than 6%, implying a very little overlap between intraspecific and interspecific genetic distances. Based on the Barcode Gap analysis, the minimum interspecific distances to

the nearest neighbor were larger than the maximum intraspecific distance for 174 species (98.9% of all species). Only for two species (*P. laetum* and *P. bullum*), the maximum intraspecific distances overlapped with the NN distance, leading to the absence of a barcode gap. The multi-approach species delimitation showed a relatively high congruence between molecular groups and the Linnaean taxa with COI DNA barcodes. These results reveal that DNA barcoding based on COI gene is an effective method for the species identification of benthic macroinvertebrates in the transboundary rivers of northwest China.

DNA barcoding has been considered to be a successful molecular identification tool for insects such as mosquitoes (Cywinska, Hunter, & Hebert, 2006; Kumar et al., 2018; Weigand et al., 2019). Also, our study shows that DNA barcoding can offer a reliable framework for the identification of mosquito species even though the identifications of a few closely related species remain ambiguous (Versteirt et al., 2015). In this study, with the two closely related mosquito species (*Polypedilum. laetum* and *Polypedilum. bullum*), the maximum intraspecific distances overlapped with the NN distance, leading to the absence of a barcode gap. The absence of barcoding gap is not uncommon in mosquitoes (Cywinska, Hunter, & Hebert, 2006; Versteirt et al., 2015; Lin et al., 2015). These two mosquito species belong to the subgenus *Polypedilum* (Kieffer, 1912) and have overlapped distribution in western China. Based on the description of morphological characteristics, *P. laetum* (Meigen, 1804) should be the sister species of *P. bullum* (Zhang & Wang, 2004). The DNA barcode data show that recent speciation events as well as hybridization may represent important processes between *P. laetum* and *P. bullum*. Further in-depth studies including more specimens and other genetic markers should be investigated to resolve the eco-evolutionary events leading to the low interspecific variation.

Based on the NJ tree, ABGD, PTP, MPTP, and BIN analysis, high levels of genetic distance and multiple lineages were observed in eight taxa (*Epeorus sp5*, *Rhithrogena tianshanica*, *Ameletus montanus*, *Atherix sp. XJ*, *Glyptotendipes sp. XJ*, *Euryhopsis sp.*, *Dicranota guerini* and *Cricotopus ornatus*), suggesting the presence of cryptic species of benthic macroinvertebrates in these transboundary rivers of northwest China. Although Hebert and Ward proposed the threshold value (2%) of species differentiation based on DNA barcoding (Hebert, Cywinska, et al., 2003; Ward, 2009), the differences in genetic differentiation can occur in different geographical populations for the same species, and thus, the genetic distance can exceed the threshold value of 2% for species classification (Hickerson et al., 2006; Tajima, 1989; Ward, 2009). In the present study, the eight species exhibited high intraspecific genetic distance and multiple genetic lineages, which was consistent with the conclusions in Ward (2009). Meanwhile, our results supported the conclusion that the genetic distance between different geographical populations of the conspecifics can exceed 2% (Hebert Ratnasingham & deWaard, 2003; Hebert, Cywinska, et al., 2003; Ward et al., 2005). Coincidentally, the two or three respective molecular lineages/clusters observed in *Glyptotendipes sp. XJ*, *C. ornatus*, *D. guerini*, *Atherix*

sp. XJ, *R. tianshanica*, *Epeorus* sp5 and *A. montanus* corresponded to different geographical areas, implying that biogeographic events may result in a great intraspecific divergence for these species. Moreover, geographical isolation plays an important role in the formation of high intraspecific genetic distance or cryptic species.

We observed high intraspecific genetic distance (up to 15.07%) for species *Ameletus montanus*. Statistical methods (ABGD, BIN, PTP, and MPTP) divided them into two different OTUs. The presence of two clusters or OTUs with deep divergences that were observed in mayfly species was indicative of cryptic diversity. Actually, cryptic diversity is typical in mayfly species (Suh et al., 2019). High genetic divergences within nominal species can be interpreted as misidentification and, more importantly, as cryptic or unrecognized speciation events (Chen et al., 2015). Recognizing cryptic diversity contributes to improving our knowledge regarding the biodiversity of numerous taxa (e.g., benthic macroinvertebrates). The genetic differentiation within one species occurred at different sample sites or geographic scales for the Irtysh river, Emin river, and Ili river. For instance, the species *D. guerini* showed high intraspecific divergence (up to 7.26%) between the Irtysh River and Ili River populations. These findings probably suggest that geographical isolation and diversification events allow different populations to evolve in different directions, thus lead to a great increase in the diversity of benthic macroinvertebrates. This suggests that DNA barcodes could be supplemented in population genetics research, with morphological, ecological, nuclear DNA, and other nonmolecular data to explore the presence of cryptic species and assess intraspecific differentiation.

DNA barcoding has been widely used for species identification (Barco et al., 2016; Versteirt et al., 2015). However, whether DNA barcoding can distinguish the individuals from different geographical populations, subspecies or biotype, remains unknown. In this study, the NJ tree shows that the conspecifics of barcoding sequences firstly clustered together, and then clustered with those of other areas (Germany, United States, Mexico, Canada, Norway, Italy, Finland etc.). In the NJ tree, both *Sigara striata* and *Chironomus pallidivittatus* covered two subclusters, and this is in accordance with the sampling locations. The same geographical populations clustered together with high support values. It indicated that the evolution of geographical population was related to geographical distance. As a result, we inferred that the population differentiation of benthic macroinvertebrates in these four rivers was ascribed to geographical isolation. It has been reported that COI genes are not sensitive enough to identify intraspecific variation, especially when the geographical differentiation of populations is not high enough to form a single pattern (Aliabadian et al., 2009; Verheyen et al., 2003). This phenomenon was observed in our study. The genetic structure analysis among different geographic population show that the shared haplotypes existed in three adjacent geographic population of three mayfly species, whereas different geographical populations had a certain degree of gene flow, intrapopulation, and interpopulation genetic divergence. In the NJ tree, the geographical populations were not divided into different branches following geographical locations. The low level of genetic differentiation among populations

decreased the reliability of COI gene to effectively distinguish intraspecific category. In contrast, with the accumulation of genetic differentiation among populations, DNA barcoding could be used to distinguish the geographical populations, subspecies or biotype (Monaghan et al., 2006). Although COI gene has great potential of species identification at a species level, but for intraspecific identification, the evolution rates of COI genes is limited because it is a protein-coding gene. Therefore, COI gene is not sensitive enough to identify populations with tiny genetic differentiation, in which the geographical locations are adjacent and the formation of geographical isolation pattern is not long enough. In this case, more factors (e.g., increasing the length of DNA barcoding) could be considered, especially for those non-protein-coding genes with faster evolution rate.

This study is the first time to report the comprehensive DNA barcode reference library of benthic macroinvertebrates in Eurasian transboundary rivers. This library included nearly all of the dominant species appeared in environmental assessment studies (Figure 5). High-throughput sequencing (HTS) can provide taxonomic information at greater resolution, depth, and consistency, and at lower cost than morphologically identified samples (Gibson et al., 2015). However, the application of HTS for taxonomic identification of samples in a biomonitoring context is limited by the availability of cytochrome c oxidase subunit 1 (COI) sequence records in reference libraries databases. Numerous investigators have reported that, without adequate representation in a reference library, accurate taxonomic identification for a given sequence can be very difficult (Ekrem et al., 2007). Therefore, our study enables future applications such as environmental DNA barcoding (Baird & Hajibabaei, 2012) and metabarcoding (Gibson et al., 2014) based on HTS and provides a solid backup for effective bioassessment in river and stream ecosystems.

5 | CONCLUSION

In conclusion, our study revealed that DNA barcoding based on COI gene is an effective method to clarify species boundaries and quantitatively evaluate species diversity (e.g., taxa abundance and cryptic species). Population differentiation of benthic macroinvertebrates in four transboundary rivers was ascribed to geographical isolation. Geographical isolation and diversification events are two main factors for different populations to evolve in different directions and thus lead to a great increase in the diversity of benthic macroinvertebrates. Even so, DNA barcoding could be supplemented in population genetics studies, with morphological, ecological nuclear DNA, and other nonmolecular data regarding the existence of cryptic species and assessment of intraspecific divergence.

ACKNOWLEDGMENTS

This work was supported by the Special Funds for the Foundation Work of Science and Technology (2012FY112700) and the Finance Special Fund of the Ministry of Agriculture and Rural Affairs

(Fisheries Resources and Environment Survey in the Key Water Areas of Northwest China). We would greatly thank postgraduates Gengen Zhou, Lili Wei, and Guang Zhao for field sampling and Shiming Wan for help in laboratory analyses.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Yihao Ge: Investigation (equal); Writing-original draft (equal). **Chengxing Xia:** Investigation (equal); Writing-original draft (equal). **Jun Wang:** Investigation (supporting); Methodology (supporting). **Xiujie Zhang:** Data curation (supporting); Formal analysis (lead); Methodology (supporting). **Xufa Ma:** Investigation (supporting); Methodology (supporting); Resources (supporting). **Qiong Zhou:** Conceptualization (equal); Data curation (lead); Funding acquisition (lead); Project administration (lead); Supervision (lead); Writing-review & editing (lead).

DATA AVAILABILITY STATEMENT

Data used in this manuscript are catalogued in the Dryad Digital Repository (available at <https://orcid.org/0000-0001-6293-2928>).

ORCID

Jun Wang  <https://orcid.org/0000-0003-2481-1409>

Xufa Ma  <https://orcid.org/0000-0003-3262-9834>

Qiong Zhou  <https://orcid.org/0000-0001-5118-6683>

REFERENCES

- Aliabadian, M., Kaboli, M., Nijman, V., & Vences, M. (2009). Molecular identification of birds: Performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLoS ONE*, 4(1), e4119. <https://doi.org/10.1371/journal.pone.0004119>
- Bailey, R. C., Norris, R. H., & Reynoldson, T. B. (2001). Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society*, 20(2), 280–286. <https://doi.org/10.2307/1468322>
- Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21(8), 2039–2044. <https://doi.org/10.1111/j.1365-294X.2012.05519.x>
- Ball, S. L., Hebert, P. D. N., Burian, S. K., & Webb, J. M. (2005). Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society*, 24(3), 508–524. <https://doi.org/10.1899/04-142.1>
- Barco, A., Raupach, M. J., Laakmann, S., Neumann, H., & Kneibelsberger, T. (2016). Identification of North Sea molluscs with DNA barcoding. *Molecular Ecology Resources*, 16(1), 288–297. <https://doi.org/10.1111/1755-0998.12440>
- Barnett, R., & Larson, G. (2012). *A phenol-chloroform protocol for extracting DNA from ancient samples*.
- Brinkhurst, R. O. (1986). *Guide to the freshwater aquatic microdrile oligochaetes of North America* (p. 84). Canadian special publication of fisheries and aquatic sciences. Department of Fisheries and Oceans.
- 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*, 13(6), 996–1004. <https://doi.org/10.1111/1755-0998.12053>
- Bucklin, A., Hopcroft, R. R., Kosobokova, K. N., Nigro, L. M., Ortman, B. D., Jennings, R. M., & Sweetman, C. J. (2010). DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep Sea Research Part II Topical Studies in Oceanography*, 57(1–2), 40–48. <https://doi.org/10.1016/j.dsr2.2009.08.005>
- Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., & Mattia, F. D. (2010). DNA barcoding: Theoretical aspects and practical applications. In: EUT Edizioni Università di Trieste.
- Castella, E., Richardotcoulet, M., Roux, C., & Richoux, P. (1984). Macroinvertebrates as describers of morphological and hydrological types of aquatic ecosystems abandoned by the Rhone river. *Hydrobiologia*, 119(3), 219–225. <https://doi.org/10.1007/bf00015212>
- Chandler, J. R. (1970). A biological approach to water quality management. *Water Pollution Control*, 69, 415–422.
- Che, J., Chen, H.-M., Yang, J.-X., Jin, J.-Q., Jiang, K. E., Yuan, Z.-Y., Murphy, R. W., & Zhang, Y.-P. (2012). Universal COI primers for DNA barcoding amphibians. *Molecular Ecology Resources*, 12(2), 247–258. <https://doi.org/10.1111/j.1755-0998.2011.03090.x>
- Cohuet, A., Simard, F., Toto, J. C., Kengne, P., Coetzee, M., & Fontenille, D. (2003). Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. *American Journal of Tropical Medicine and Hygiene*, 69(2), 200–205. <https://doi.org/10.4269/ajtmh.2003.69.200>
- Cywinska, A., Hunter, F. F., & Hebert, P. D. N. (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>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772. <https://doi.org/10.1038/nmeth.2109>
- Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Hebert, P. D. N., & Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, 5, 12395. <https://doi.org/10.1038/srep12395>
- Dudgeon, D. (1999). *Tropical Asian streams: Zoobenthos, ecology and conservation* (Vol. 1). Hong Kong University Press.
- 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>
- Elderkin, C. L., Corkum, L. D., Bustos, C., Cunningham, E. L., & Berg, D. J. (2012). DNA barcoding to confirm morphological traits and determine relative abundance of burrowing mayfly species in western Lake Erie. *Journal of Great Lakes Research*, 38(1), 180–186. <https://doi.org/10.1016/j.jglr.2011.11.010>
- Epler, J. H., & Epler, J. H. (2001). *Identification manual for the larval Chironomidae (Diptera) of North and South Carolina. Version 1.0*.
- Felsenstein, J. (1985). Confidence-limits on phylogenies – An approach using the bootstrap. *Evolution*, 39(4), 783–791. <https://doi.org/10.2307/2408678>
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vejenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299. <https://doi.org/10.4028/www.scientific.net/DDF.7.460>
- Fontaneto, D., Flot, J.-F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity*, 45(3), 433–451. <https://doi.org/10.1007/s12526-015-0319-7>
- Frézal, L., & Leblois, R. (2008). Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, 8(5), 727–736. <https://doi.org/10.1016/j.meegid.2008.05.005>

- Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PLoS ONE*, *10*(10), e0138432. <https://doi.org/10.1371/journal.pone.0138432>
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8007–8012. <https://doi.org/10.1073/pnas.1406468111>
- Gill, B. A., Harrington, R. A., Kondratieff, B. C., Zamudio, K. R., Poff, L. R., & Funk, W. C. (2014). Morphological taxonomy, DNA barcoding, and species diversity in southern Rocky Mountain headwater streams. *Freshwater Science*, *33*(1), 288–301. <https://doi.org/10.1086/674526>
- Haase, P., Murray-Bligh, J., Lohse, S., Pauls, S., Sundermann, A., Gunn, R., & Clarke, R. (2006). Assessing the impact of errors in sorting and identifying macroinvertebrate samples. *Hydrobiologia*, *566*, 505–521. <https://doi.org/10.1007/s10750-006-0075-6>
- Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 968–971. <https://doi.org/10.1073/pnas.0510466103>
- Havemann, N., Gossner, M. M., Hendrich, L., Moriniere, J., Niedringhaus, R., Schaefer, 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>
- Hawlichschek, O., Moriniere, J., Lehmann, G. U. C., Lehmann, A. W., Kropf, M., Dunz, A., Glaw, F., Detcharoen, M., Schmidt, S., Hausmann, A., Szucsich, N. U., Caetano-Wyler, S. 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 B-Biological Sciences*, *270*(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit I divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(Suppl 1), S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>
- Helson, J. E., & Willians, D. D. (2013). Development of a macroinvertebrates multimetric index for the assessment of low-land streams in the neotropics. *Ecological Indicators*, *29*, 167–178. <https://doi.org/10.1016/j.ecolind.2012.12.030>
- Hendrich, L., Moriniere, J., Haszprunar, G., Hebert, P. D. N., Hausmann, A., Koehler, 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>
- Hernandez-Triana, L. M., Prosser, S. W., Rodriguez-Perez, M. A., Chaverri, L. G., Hebert, P. D. N., & Gregory, T. R. (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. <https://doi.org/10.1111/1755-0998.12208>
- Hickerson, M. J., Meyer, C. P., & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, *55*(5), 729–739. <https://doi.org/10.1080/10635150600969898>
- Huang, Z. H., & Ke, D. H. (2015). DNA barcoding and phylogenetic relationships in Timaliidae. *Genetics & Molecular Research*, *14*(2), 5943–5949. <https://doi.org/10.4238/2015.June.1.11>
- Hunter, S. J., Goodall, T. I., Walsh, K. A., Owen, R., & Day, J. C. (2008). Nondestructive DNA extraction from blackflies (Diptera: Simuliidae): Retaining voucher specimens for DNA barcoding projects. *Molecular Ecology Resources*, *8*(1), 56–61. <https://doi.org/10.1111/j.1471-8286.2007.01879.x>
- Ivanov, V., Lee, K. M., & Mutanen, M. (2018). Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. *Molecular Ecology*, *27*(7), 1681–1695. <https://doi.org/10.1111/mec.14564>
- Kang, B., Lee, J.-H., & Park, J.-K. (2012). Carabid beetle species as a biological indicator for different habitat types of agricultural landscapes in Korea. *Journal of Ecology and Environment*, *35*(1), 35–39. <https://doi.org/10.5141/JEFB.2012.006>
- Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, *33*(11), 1630–1638. <https://doi.org/10.1093/bioinformatics/btx025>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kavar, T., Pavlovcic, P., Susnik, S., Meglic, V., & Virant-Doberlet, M. (2006). Genetic differentiation of geographically separated populations of the southern green stink bug *Nezara viridula* (Hemiptera: Pentatomidae). *Bulletin of Entomological Research*, *96*(2), 117–128. <https://doi.org/10.1079/ber2005406>
- Kieffer, J. J. (1912). H. Sauter's Formosa-Ausbeute.) Tendipedidae (Chironomidae). *Supplementa Entomologica Berlin*, *1*, 27–43.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. <https://doi.org/10.1007/BF01731581>
- Kumar, S., Stecher, G., Li, M., Nnyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lenat, D. R., & Resh, V. H. (2001). Taxonomy and stream ecology—the benefits of genus-and species-level identifications. *Journal of the North American Benthological Society*, *20*(2), 287–298. <https://doi.org/10.2307/1468323>
- Lin, X., Stur, E., & Ekrem, T. (2015). Exploring genetic divergence in a species-rich insect genus using 2790 DNA barcodes. *Genome*, *58*(5), 246–247.
- Macher, J. N., Salis, R. K., Blakemore, K. S., Tollrian, R., Matthaei, C. D., & Leese, F. (2016). Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. *Ecological Indicators*, *61*, 159–169. <https://doi.org/10.1016/j.ecolind.2015.08.024>
- Monaghan, M. T., Balke, M., Pons, J., & Vogler, A. P. (2006). Beyond barcodes: Complex DNA taxonomy of a south pacific island radiation. *Proceedings of the Royal Society B-Biological Sciences*, *273*(1588), 887–893. <https://doi.org/10.1098/rspb.2005.3391>
- Moriniere, J., Hendrich, L., Balke, M., Beermann, A. J., König, T., Hess, M., Koch, S., Müller, R., Leese, F., Hebert, P. D. N., Hausmann, A., Schubart, C. D., & Haszprunar, G. (2017). A DNA barcode library for Germanys mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Molecular Ecology Resources*, *17*(6), 1293–1307. <https://doi.org/10.1111/1755-0998.12683>
- Morse, J. C., Yang, L., & Tian, L. (1984). *Aquatic insects of China useful for monitoring water quality*.

- Morse, J. C., Yang, L. F., & Tian, L. X. (1994). Aquatic insects of China useful for monitoring water quality (pp. 92–102). Hohai University Press.
- Murria, C., Zamora-Munoz, C., Bonada, N., Ribera, C., & Prat, N. (2010). Genetic and morphological approaches to the problematic presence of three Hydropsyche species of the pellucidula group (Trichoptera: Hydropsychidae) in the westernmost Mediterranean Basin. *Aquatic Insects*, 32(2), 85–98. <https://doi.org/10.1080/01650424.2010.482939>
- Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, 24(7), 386–393. <https://doi.org/10.1016/j.tree.2009.02.011>
- 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, 1957–1968. <https://doi.org/10.1111/j.1365-294X.2006.03136.x>
- 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>
- Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7(3), 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678>
- 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>
- Ribeiro Leite, L. A. (2012). Mitochondrial pseudogenes in insect DNA barcoding: Differing points of view on the same issue. *Biota Neotropica*, 12(3), 301–308. <https://doi.org/10.1590/s1676-06032012000300029>
- Sivaramakrishnan, K. G., Janarthanan, S., Selvakumar, C., & Arumugam, M. (2014). Aquatic insect conservation: A molecular genetic approach. *Conservation Genetics Resources*, 6(4), 849–855. <https://doi.org/10.1007/s12686-014-0250-4>
- Smith, P. J., McVeagh, S. M., & Steinke, D. (2008). DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology*, 72(2), 464–471. <https://doi.org/10.1111/j.1095-8649.2007.01745.x>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stark, J. D., Boothroyd, I. K. G., Harding, J. S., Maxted, J. R., & Scarsbrook, M. R. (2001). Protocols for sampling macroinvertebrates in wadeable streams. *New Zealand Macroinvertebrate Working Group Report*, 1, i–v, 1–57.
- Stein, E. D., White, B. P., Mazor, R. D., Jackson, J. K., Battle, J. M., Miller, P. E., Pilgrim, E. M., & Sweeney, B. W. (2014). Does DNA barcoding improve performance of traditional stream bioassessment metrics? *Freshwater Science*, 33(1), 302–311. <https://doi.org/10.1086/674782>
- Suh, K. I., Hwang, J. M., Bae, Y. J., & Kang, J. H. (2019). Comprehensive DNA barcodes for species identification and discovery of cryptic diversity in mayfly larvae from South Korea: Implications for freshwater ecosystem biomonitoring. *Entomological Research*, 49(1), 46–54. <https://doi.org/10.1111/1748-5967.12334>
- Tajima, F. (1989). Statistical method for testing the natural mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.
- Verheyen, E., Salzburger, W., Snoeks, J., & Meyer, A. (2003). Origin of the superclade of cichlid fishes from Lake Victoria. *East Africa. Science*, 300(5617), 325–329. <https://doi.org/10.1126/science.1080699>
- Versteirt, V., Nagy, Z. T., Roelants, P., Denis, L., Breman, F. C., Damiens, D., Dekoninck, W., Backeljau, T., Coosemans, M., & 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>
- Wang, B. X., Yang, L. F., & Liu, Z. W. (2006). Index of biological integrity and its application in health assessment of aquatic ecosystem. *Chinese Journal of Ecology*, 25(6), 707–710.
- Wang, J., Zhou, Q., Xie, C. X., Li, J., & Wei, L. L. (2014). The community structure of macroinvertebrates and biological assessment of water quality in the Irtysh River of Xinjiang. *Chinese Journal of Ecology*, 33(9), 2420–2428.
- Wang, S. J. (2010). *A complete book of rivers and lakes in Xinjiang, China*. WaterPower Press.
- Ward, R. D. (2009). DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources*, 9(4), 1077–1085. <https://doi.org/10.1111/j.1755-0998.2009.02541.x>
- Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Začovičová, Z., Ferreira, S., ... Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, 678, 499–524. <https://doi.org/10.1016/j.scitotenv.2019.04.247>
- Werren, J. H., Zhang, W., & Guo, L. R. (1995). Evolution and phylogeny of wolbachia: Reproductive parasites of arthropods. *Proceedings of the Royal Society B: Biological Sciences*, 261(1360), 55–63.
- Wong, W. H., Tay, Y. C., Puniamoorthy, J., Balke, M., Cranston, P. S., & Meier, R. (2014). Direct PCR' optimization yields a rapid, cost-effective, nondestructive and efficient method for obtaining DNA barcodes without DNA extraction. *Molecular Ecology Resources*, 14(6), 1271–1280. <https://doi.org/10.1111/1755-0998.12275>
- Yang, F., Shi, Z. Y., Bai, S., Ward, R. D., & Zhang, A. B. (2014). Comparative studies on species identification of Noctuoidea moths in two nature reserve conservation zones (Beijing, China) using DNA barcodes and thin-film biosensor chips. *Molecular Ecology Resources*, 14(1), 50–59. <https://doi.org/10.1111/1755-0998.12165>
- Yoo, H. S., Eah, J. Y., Kim, J. S., Kim, Y. J., Min, M. S., Paek, W. K., & Kim, C. B. (2013). DNA barcoding Korean birds (Retraction of vol 22, pg 323, 2006). *Molecules and Cells*, 35(4), 357. <https://doi.org/10.1007/s10059-013-3151-6>
- Zhang, F., Zhang, M., Liu, Z., Chen, H., & Qi, S. (2011). A health assessment using a benthic-index of integrity in Ganjiang River basin. *Acta Hydrologica Sinica*, 35(6), 963–971.
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869–2876. <https://doi.org/10.1093/bioinformatics/btt499>
- Zhang, R. L., & Wang, X. H. (2004). Polypedium (Uresipedilum) Oyewo and Saether from China (Diptera : Chironomidae). *Zootaxa*, 565, 1–38.
- Zhou, C., Gui, H., & Zhou, K. (2003). Larval key to families of Ephemeroptera from China (Insecta). *Nanjing Shida Xuebao Ziran Kexue Ban*, 26(2), 65–68.
- Zhou, X. (2009b). The larvae of Chinese Hydropsychidae (Insecta: Trichoptera), Part I: *Arctopsyche shimianensis*, *Parapsyche* sp. A and *Dipletrona Obscura*. *Zootaxa*, 2174(2174), 1–17. <https://doi.org/10.1002/cne.22076>
- Zhou, X., Adamowicz, A. J., Jacobus, L. M., DeWalt, R. E., & Hebert, P. D. N. (2009a). Towards a comprehensive barcode library for arctic life – Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology*, 6(1), 30. <https://doi.org/10.1186/1742-9994-6-30>
- Zhou, X., Jacobus, L. M., DeWalt, R. E., Adamowicz, S. J., & Hebert, P. D. N. (2010). Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): Insights into biodiversity patterns from DNA barcoding. *Journal of the North American Benthological Society*, 29(3), 814–837. <https://doi.org/10.1899/09-121.1>
- Zhou, X., Robinson, J. L., Geraci, C. J., Parker, C. R., Flint, O. S., Etnier, D. A., Ruitter, D., DeWalt, R. E., Jacobus, L. M., & Hebert, P. D. N.

(2011). Accelerated construction of a regional DNA-barcode reference library: Caddisflies (Trichoptera) in the Great Smoky Mountains National Park. *Journal of the North American Benthological Society*, 30(1), 131–162. <https://doi.org/10.1899/10-010.1>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Ge Y, Xia C, Wang J, Zhang X, Ma X, Zhou Q. The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. *Ecol Evol.* 2021;11:5669–5681. <https://doi.org/10.1002/ece3.7470>