

# The efficacy of DNA barcoding in the classification, biodiversity and environmental assessment of benthic macroinvertebrates

Yihao Ge<sup>1</sup>, Chengxin Xia<sup>2</sup>, Jun Wang<sup>3</sup>, Xiujie Zhang<sup>2</sup>, Xu-Fa Ma<sup>4</sup>, and Qiong Zhou<sup>2</sup>

<sup>1</sup>Institute of Hydrobiology, Chinese Academy of Sciences

<sup>2</sup>Huazhong Agricultural University

<sup>3</sup>Yunnan University

<sup>4</sup>College of Fisheries

May 21, 2020

## Abstract

Macroinvertebrates have been recognized as key ecological indicators of environmental and biodiversity assessment in aquatic ecosystems. However, species identification of macroinvertebrates (especially aquatic insects) proves to be very difficult due to lack of expertise. In this study, we evaluated the feasibility of DNA barcoding for the classification of benthic macroinvertebrates and investigated the genetic differentiation in nine taxonomic groups (Ephemeroptera, Plecoptera, Trichoptera, Diptera, Hemiptera, Coleoptera, Odonata, Mollusca and Annelida) from four large transboundary rivers of northwest China, and further explored its potential application to environment and biodiversity assessment. A total of 1227 COI sequences, belonging to 189 species, 122 genera and 59 families were obtained. The barcode gap analysis supported species status using the barcode gap approach. Meanwhile, NJ phylogenetic trees showed that all species group into single-species representing clusters whether from the same population or not, except two species (*Polypedilum. laetum* and *Polypedilum. bullum*). The ABGD analysis divided into 190 OTUs ( $P = 0.0599$ ) and BIN analysis generated 201 different BINs. Phylogenetic diversity (PD) metrics can reflect environmental stress and serve as a metrics of Index of Biotic Integrity (IBI) to reflect the degree of disturbance in river systems.

## Abstract

Macroinvertebrates have been recognized as key ecological indicators of environmental and biodiversity assessment in aquatic ecosystems. However, species identification of macroinvertebrates (especially aquatic insects) proves to be very difficult due to lack of expertise. In this study, we evaluated the feasibility of DNA barcoding for the classification of benthic macroinvertebrates and investigated the genetic differentiation in nine taxonomic groups (Ephemeroptera, Plecoptera, Trichoptera, Diptera, Hemiptera, Coleoptera, Odonata, Mollusca and Annelida) from four large transboundary rivers of northwest China, and further explored its potential application to environment and biodiversity assessment. A total of 1227 COI sequences, belonging to 189 species, 122 genera and 59 families were obtained. The barcode gap analysis supported species status using the barcode gap approach. Meanwhile, NJ phylogenetic trees showed that all species group into single-species representing clusters whether from the same population or not, except two species (*Polypedilum. laetum* and *Polypedilum. bullum* ). The ABGD analysis divided into 190 OTUs ( $P = 0.0599$ ) and BIN analysis generated 201 different BINs. Phylogenetic diversity (PD) metrics can reflect environmental stress and serve as a metrics of Index of Biotic Integrity (IBI) to reflect the degree of disturbance in river systems.

**Keywords** Benthic macroinvertebrates; Species identification; Environmental assessment; DNA barcoding;

## Introduction

Macroinvertebrates (e.g., aquatic insects) have been considered as one of primary components in aquatic ecosystems and a key food resources for higher trophic levels. 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 et al., 1970; Helson et al., 2013), especially for those water bodies (e.g., rivers and lakes) intervened by anthropogenic activities (Wang et al., 2006; Zhang et al., 2011). However, these ecological functions depend, to a larger extent, on the absolutely accuracy and precision of species identification (Frezal et al., 2008; Macher et al., 2016). Traditional taxonomy always relies on the morphology-based traits to identify a specific species, but this approach proves to be difficult, time-consuming and costly (Wong et al., 2014). Despite their important roles on water quality assessment, morphological delineation of aquatic insects (especially 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, Adamowicz, Jacobus, DeWalt, & Hebert, 2009; Zhou et al., 2011). For example, identification of Hydropsychidae caddisflies at the species level is only possible in adult organisms (Zhou et al., 2007). Unlike those model species such as butterflies (Dinca et al., 2015) and spiders (Blagoev et al., 2016), the information on aquatic insects is quite limited due to the complex life history. Moreover, despite the increasing demand for well-trained taxonomists to support the assessment of aquatic ecosystems, the number of ecologists and researchers with taxonomic expertise are decreasing (Wheeler, 2014). Accordingly, such subjective or objective factors may lead to an inaccurate description of macroinvertebrates in terms of traditional classification.

To circumvent the above-mentioned hurdles associated with morphology-based species identification, DNA barcoding emerged as a new tool that utilizes short standard sequences of the genome for species identification and classification, based on the mitochondrial gene cytochrome c oxidase subunit I (COI) (Hebert 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). This could provide valuable information on the selection of taxa for further analyses (Hajibabaei et al., 2007) and make it possible to associate with all life history stages and genders (Casiraghi et al., 2010). Thus, DNA barcoding offsets some shortcomings in traditionally morphological delineation (Kress et al., 2012). Unknown specimens could be identified by comparing their DNA barcode sequences with a barcode reference library (Moriniere et al., 2017). Previous studies illuminated the great reliability of DNA barcoding in species identification of aquatic insects, e.g., Ephemeroptera (Ball et al., 2005; Curt et al., 2012), Plecoptera and Trichoptera (Gillet et al., 2014) and Diptera (Brodin et al., 2013; Hernandez et al., 2014; Hunter et al., 2008). However, these studies focused primarily on the application of DNA barcoding in 1-3 orders, and few studies considered all the taxa of macroinvertebrates in a specific ecosystem and further applied to the environmental and biodiversity assessment.

China spans a great range from east to central Asia and stretches across Eurasia (e.g., neighboring to Russia), and exhibits very different characteristics of climate and geography. Thus, great differences and high diversity in aquatic biota occur in China. Until now, DNA barcoding has been applied to some organisms, e.g., amphibians (Che et al., 2012), birds (Yang et al., 2014), plants (Huang et al., 2015), Noctuoidea moths (Zheng et al., 2015), molluscs (Andrea et al., 2016), crickets (Hawlitshchek et al., 2016), herpetofauna (Hawlitshchek et al., 2016), fish (sonnet et al., 2018), spiders (Ivanov et al., 2018), but rarely focused on aquatic invertebrates (Zhou et al., 2009b; Zhou et al., 2009a), especially aquatic insects. 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 (1) establish a DNA barcoding reference library for benthic macroinvertebrates in northwest China; (2) apply phylogenetic diversity using community data to reflect the degree of disturbance in these rivers.

## Materials and Methods

### 2.1 Study area

Our study was conducted in Xinjiang Uygur Autonomous Region, center of Asia-Europe continent (Fig. 1). This region extends from the south slope of the Altai Mountains to the hinderland 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 7435 m, and encompasses complex landforms and microclimate between mountains and basins. Due to the characteristics of geography, environment and climate, Xinjiang region provides biological community 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 (Fig. 1) and ultimately flow towards 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 1035 square kilometers. These rivers and their affiliated water bodies are the most important water resources in northwest China and play an important role in maintaining the biodiversity of aquatic ecosystems (Wang et al., 2010).

## 2.2 Sampling and experimental material

Aquatic insects are absolutely 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). Furthermore, transboundary rivers such as the Ili River and Irtysh River are suffering from the intervening of anthropogenic activities, and hence accurate classification of macroinvertebrates is essential for the environmental and biodiversity assessment. As for the organisms used for water quality assessment, macroinvertebrates are one of the most optimal categories due to a large size, weak dispersal and the sensitiveness to the change of water quality (Resh et al., 1995).

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 (Fig. 1). Study sites were set in the main stems, tributaries and affiliated water bodies (lakes and reservoirs). Due to a long frozen season (November to next April), samples of Mayfly (order Ephemeroptera), stonefly (order Plecoptera), caddisfly (order Trichoptera), robber fly (order Diptera), true bugs (order Hemiptera), Coleoptera, Odonata, Mollusca and Annelida 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. Approximately 30,000 specimens were preserved in 95% ethanol to allow for the morphological identification and molecular analyses. In particular, when the color of ethanol solution for macroinvertebrate soaking became dark yellow during the field survey, ethanol needed to be replaced for two or three times in order to warrant a good preservation of the specimens 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 (China). In the lab, specimens were identified based on morphological characteristics, and then intact individuals were selected to conduct the DNA barcoding analysis for nine taxonomic groups (Ephemeroptera, Plecoptera, Trichoptera, Diptera, Hemiptera, Coleoptera, Odonata, Mollusca and Annelida).

## 2.3 DNA extractions, amplification and sequencing

Total genomic DNA was extracted from legs, cerci, half or whole body according to a specimen's size, using the phenol-chloroform protocol (Barnett and Larson, 2012). and quantified in a NanoDrop 2000c (Thermo Scientific, USA). The primer pair LCO-1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO-2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al., 1994) was used to amplify the fragment DNA of COI (mitochondrial cytochrome C oxidase I) with the length of about 658 bp. Polymerase Chain Reactions (PCR) were performed in 50  $\mu$ L volumes using the following reagents and concentrations: a volume of 50  $\mu$ L containing 10 $\times$ PCR buffer 5  $\mu$ L, Mg<sup>2+</sup> (2.5 mmol/L) 5  $\mu$ L, dNTP (2.5 mmol/L) 3.5  $\mu$ L, 1.5  $\mu$ L (10 nmol/L) each primer, 0.5 U Taq polymerase (Takara, Dalian, China), 2  $\mu$ L (50 ng/ $\mu$ L) DNA template and complete ddH<sub>2</sub>O to 50  $\mu$ 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). PCR products were checked by electrophoresis at 1% agarose with an ethidium bromide stain, and if successful, the PCR products were subsequently purified and directly sequenced by Invitrogen Corporation in China.

## 2.4 Barcoding analysis based on COI

Bidirectional sequencing was employed to guarantee the precision of sequencing. We aligned the COI gene using the ClustalW program from MEGA 7.0 (Tamura et al., 2013) package for each order with default parameters (Zhou et al., 2010). The amino acid translation was conducted to ensure that no gaps or stop codons existed in the alignment. Totally, we obtained 1227 COI sequences from 189 species. Detailed information (locality data, habitat, altitude, collector, identifier, taxonomic classifications, habitus images and DNA barcode sequences etc.) for each voucher specimen was deposited to the Barcode of Life Data System (BOLD) at <http://www.boldsystems.org> (Ball et al., 2007), under the dataset project “XJDQD” (Process IDs: XJDQD001-18-XJDQD1275-18).

### 2.4.1 Phylogenetic tree

Neighbor-Joining (NJ) method in MEGA 7.0 software was used for tree construction (Satiou et al., 1987) with Kimura-2-Parameter (K-2-P) genetic distances (Kimura et al., 1980). Meanwhile, to evaluate the accuracy of DNA barcoding for classifying species as well as the evolutionary relationship of species, we downloaded the sequencing data of relevant species from BOLD, including those from Germany, United States, Mexico, Canada, Norway, Italy, Finland, Japan, Iran, etc. (Raupach et al., 2014; Stein et al., 2013) (Part I, Supporting information). A total of 1000 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

“Barcode Gap Analysis” tool 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 the Barcode Gap between maximum intraspecific genetic distance and minimum interspecific distance according to Ratnasingham & Hebert (2007). Within the BOLD, the BIN system group sequence data into clusters of closely COI barcode sequences that are assigned to a globally unique identifier, termed BIN (Barcode Index Number) (Ratnasingham & Hebert, 2013). BINs for species clusters also enable the delineation of geographical clades, which might reflect local environmental features (Jerome et al., 2017).

### 2.4.3 Automatic Barcode Gap Discovery *analysis*

Automatic Barcode Gap Discovery (ABGD) analysis (Puillandre et al., 2012) was 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/](http://www.abi.snv.jussieu.fr/public/abgd/)). The default value for the relative gap-width was set as  $X = 1.5$ . We also analyzed each order in a separate analysis and all the assignments that the values (P) of intraspecific divergence ranged between 0.001 and 0.100 were recorded. Default settings were employed for all remaining parameters.

## 2.5 Phylogenetic diversity analysis

$$PD = \sum_{\{c \in C\}} Lc \quad \setminus n$$

C is a phylogenetic tree, c is a branch on C and  $Lc$  is the branch length of C.

The box diagrams of the reference and disturbance sites were completed by origin 9.1 (Seifert et al., 2014). The “pwr” (Stephane et al., 2018) package in R was used for paired t test to clarify whether the reference sites were different from the disturbance sites.

## 2.6 Water quality assessment based on the biotic index of macroinvertebrates

$$BI = \sum_{i=1}^S n_i a_i / N \setminus n$$

where  $n_i$  represent the abundance of each species in each sampling sites,  $a_i$  represent the tolerance values of each species and the  $N$  represent the total abundance of each site. The key metrics of the B-IBI health assessment system include: phylogenetic diversity (PD), the number of EPT (Ephemeroptera, Plecoptera, Trichoptera) taxa, the percentage of Ephemeroptera, the percentage of tolerant groups, BI (Biotic Index) and the percentage of clingers. Based on BI values, sampling sites were classified as follows: excellent (0-2.97), good (2.98-4.72), fair (4.73-6.48) poor (6.49-8.24) and bad ( $>8.24$ ), accordingly to Satiou et al. (1987).

## Results

### 3.1 Sequencing analyses

Macroinvertebrates are absolutely predominant faunas in the international rivers of northwest China, but few studies on macroinvertebrates have been carried out in these rivers. In this study, 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 1227 sequences (516-Ephemeroptera, 146-Plecoptera, 81-Trichoptera, 66-Hemiptera, 273-Diptera, 55-Coleoptera, 23-Odonata, 66-Annelida, 17-Mollusca), belonging to 189 species, 123 genera and 59 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, suggesting that the COI PCR primers were efficient. All the species exhibited high adenosine and thymine (AT)-rich bias (averaged 63.1%). Specifically, the A+T % 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 15.07% with an average of 0.78%, whereas interspecific K2P distance within one genus ranged from 0.62% to 33.36% (average 16.37%). 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 neighbour overlapped partially (Fig. 2), 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 187 species was larger than the maximum intraspecific distance (98.9%; Fig. 3). For the species *P. laetum* and *P. 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 nine macroinvertebrate taxonomic groups (Table S1). In particular, Mollusca exhibited the largest mean intraspecific distance (1.93%), whereas Odonata yielded the lowest value of 0.28%.

### 3.3 Phylogenetic tree-based identification and cluster analysis

Through the NJ method, phylogenetic trees of nine taxonomic groups show that all individuals grouped into species-specific clusters with high bootstrap values except *P. laetum* and *P. bullum* (Fig. 4). For those species with two or more representatives, the conspecifics in monophyletic clades was associated with high support (100% bootstrap in the NJ tree). Despite very low interspecific variations (0.62%), species *Chironomus* sp. XJ and *Chironomus heterodentatus* can be well distinguished by the phylogenetic analyses. One abnormal cluster occurred in an individual of *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, *Euryhapsis* sp, *Cricotopus ornatus*, *Atherix* sp. XJ, *Epeorus* sp5 and *Radix auricularia* exhibited large internal splits in their monophyletic clusters with  $>90\%$  support values and, simultaneously, large K2P distances between the internal clusters (Fig. S1). In addition, the clades of *Sigara striata* and *Chironomus pallidivittatus* formed with some subclusters with relatively large mean K2P genetic distance (Fig. S2).

The cluster analysis of 265 individual sequences from the BOLD data show that the species clustered well with the homologous specimens (Fig. 4B-4J). Firstly, specimens of the same species in our study clustered within the same cluster, and then clustered with those from other areas (Germany, United States, Mexico,

Canada, Norway, Italy, Finland, Japan, Iran etc). Based on the BLAST analysis, the query cover and identity of COI sequences for the same species from the present study and other reports in BOLD database did not reach 100%.

### 3.4 OTU delineation based on DNA barcodes using ABGD

ABGD software was utilized to delineate 189 morphological species and in total, 1227 sequences with the assignments for intraspecific divergence (p) values between 0.001 and 0.100. The results included two cases: initial partition and recursive partition. As revealed in Part II (Supporting information), 1227 samples were divided into 227 groups (p value: 0.0028-0.0077) and 190 groups (p value: 0.0129-0.0599). ABGD analysis of the whole sequencing data derived a total of 190 OTUs (Operational taxonomic units) with a prior intraspecific divergence of  $P_{max} = 0.0599$ . Therefore, we chose relatively stable initial results to compare with those of morphological identification. We found that *A. montanus* and *D. guerini* formed into two different groups, whereas *Chironomus* s p. XJ and *C. heterodontatus*, *P. bullum* and *P. laetum*, *Diamesa* sp. XJC and *Diamesa* sp. XJB clustered into the same group, respectively.

### 3.5 BIN analysis

In the present study, 1227 records generated 201 different BINs. In particular, 117 BINs were not recorded previously in the BOLD database and most sequences representing new endemic species received new BIN assignments. The Diptera species *Chironomus* sp. XJ and *C. heterodontatus* shared the same BIN [AAW4009], *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 was attributed to haplotype sharing and low interspecific divergence. A number of 141 BIN clusters were found to be taxonomically concordant with other barcode data on BOLD, that is, they were assigned to the same species. A number of 58 records are singletons, implying that these BINs were represented by only one sequence. Nine species were assigned to a total of 21 BINs (Table S2).

### 3.6 Water quality and biodiversity assessment

To test whether Phylogenetic diversity (PD) can be used as a metrics of Index of Biotic Integrity (IBI), the PD of each sampling site was calculated through community data and the phylogenetic tree constructed using sequences of mitochondrial COI gene (Table S3). The box plot and t-test were used to compare the PD between the reference and disturbance sites (Fig. 5). The PD values of reference sites were significantly higher than those in disturbance sites (One-way ANOVA,  $p < 0.05$ ), and there is no overlap between reference and disturbance box in box plot.

Macroinvertebrate assemblages were utilized to calculate the biological and diversity index. The grades of assessment in the Irtysh river were as follows:  $B-IBI > 2.04$  represents excellent,  $1.53 < B-IBI < 2.04$  represents good,  $1.02 < B-IBI < 1.53$  represents fair,  $0.51 < B-IBI < 1.02$  represents poor, and  $B-IBI < 0.51$  represents bad. For the Ili river, the grades were as follows:  $B-IBI > 1.31$  represents excellent,  $1.01 < B-IBI < 1.31$  represents good,  $0.68 < B-IBI < 1.01$  represents fair,  $0.34 < B-IBI < 0.68$  represents poor, and  $B-IBI < 0.34$  represents bad. The water quality and biodiversity of each site in the studied rivers were evaluated in Table S4. In the Irtysh River, four of eight sites were excellent and only one site was poor. However, in the Ili River, two of eight sites were excellent and two sites were Bad.

## 4. Discussion

### 4.1 Macroinvertebrate identification based on DNA barcoding

The species identification using DNA barcoding is based on the fact that the genetic distance between two species is much greater than that within a species. It has been proposed to take 2% as 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 et al., 2003; Ward et al., 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 of threshold value 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 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 187 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. 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.

Based on the NJ method, ABGD method and BIN analysis, high levels of genetic distance and genetic lineages were observed in nine taxa (*Radix auricularia*, *Epeorus* sp5, *Rhithrogena tianshanica*, *Ameletus montanus*, *Atherix* sp. XJ, *Glyptotendipes* sp. XJ, *Euryhapsis* sp, *Dicranota guerini* and *Cricotopus ornatus*), suggesting the presence of cryptic species diversity 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 et al., 2003; Ward et al., 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 (Tajima et al., 1983; Hickerson et al., 2006; Ward et al., 2009). In present study, the nine species exhibited high intraspecific genetic distance and multiple genetic lineages, and this was consistent with the conclusions of Ward et al. (2009). Meanwhile, our results support the conclusion that the genetic distance between different geographical populations of the conspecifics can exceed 2% (Hebert et al., 2003b; Ward et al., 2005). Coincidentally, the two or three respective molecular lineages/clusters observed in *R. auricularia*, *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 result in a great intraspecific divergence for these species. And geographical distance can play an important role in the formation of high intraspecific genetic distance or cryptic species.

Likewise, 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 sampling site of Bieliezeke in the Irtysh river is nearly 700 km away from that of Qiaoerma in the Ili river. However, for species *Ameletus montanus*, four individuals in site Bieliezeke showed high intraspecific divergence (up to 15.07%) with 23 individuals in site Qiaoerma, suggesting that the divergence reaches an interspecific level. Moreover, the BIN and ABGD analysis divided them into two different groups, and the NJ tree analysis also formed two main branches. Through rechecking the specimens, we did not find any morphological feature that represents different species. In contrast, the altitude (2294 m) of site Qiaoerma is higher than that of site Bieliezeke (640 m). And the habitats of these two sites are totally different as well as the nuptial flight, breeding time of the two populations. This can provide the two populations with ideal criteria for cryptic species due to long evolution time and the difference in the habitats. Likewise, the species *D. guerini* showed high intraspecific divergence (up to 7.26%) between the Irtysh River and Ili River populations. Even in the same river (Ili River), the species *R. auricularia* yielded an intraspecific divergence of 4.3% between site Nileke and site Zhaosu, that are separated by a distance of 150 km. Specifically, these two sites are situated at two different tributaries in the upstream of Ili River and isolated by the Wusun Mountain. In present study, DNA barcoding proves to be effective for the species identification of benthic macroinvertebrates in most cases. However, as a preliminary hypothesis of species classification, DNA barcodes can be supplemented by morphological, ecological nuclear DNA and other non-molecular data in what respects the existence of cryptic species.

DNA barcoding has been widely used for species identification (Andrea et al., 2016; Versteirt et al., 2015). However, whether DNA barcoding can distinguish the individuals from different geographical populations, subspecies or biotype, remains unknown. The NJ tree showed that the conspecifics of barcoding sequences in the present study first clustered together, and then clustered with those of other areas (Germany, United States, Mexico, Canada, Norway, Italy, Finland etc.). In the NJ tree, both *S. striata* and *C. pallidivittatus* covered two subclusters, which were in accordance with the sampling locations. The same geographical populations clustered together with high support values, and the phylogenetic tree 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 long enough to form a single pattern (Verheyen et al., 2003; Aliabadian and Kaboli., 2008). This phenomenon was also observed in our study. In terms of the genetic structure analysis among different geographic population, common haplotypes existed in three adjacent geographic population of three mayfly species, whereas different geographical populations generated a certain degree of gene flow, intra-population and inter-population genetic divergence. In the NJ tree, the geographical populations did not divide into different branches according to different geographical locations. The lacking of genetic differentiation among populations made COI gene unable to effectively distinguish intraspecific category. However, if the influence of geographical isolation and ecological environment lead to the accumulation of genetic differentiation among populations, DNA barcoding have the potential 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 rate of COI genes can be limited due to 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) should be considered, especially for those non-protein-coding genes with faster evolution rate.

#### 4.2 Environmental and biodiversity assessment

Phylogenetic diversity (PD) represents the adaptive potential of a community in that higher genetic diversity means greater adaptability to environmental changes (Vellend et al., 2011). In addition, PD indices measure the evolutionary relationship between species, and can reflect different aspects of biodiversity that are not captured by traditional indices based on taxonomic diversity (e.g., Petchey & Gaston, 2002; Cianciaruso et al., 2009; Vellend et al., 2011; Weiher, 2011). Therefore, PD metrics have the great potential to be used in biomonitoring of aquatic ecosystems (Vandewalle et al., 2010).

Our results demonstrated that human disturbance led to a decrease of PD since values at reference sites were significantly higher than those in disturbed sites (Fig. 5). Lower PD values in disturbance sites suggest that human activities impact not only the taxonomic diversity of rivers, but also the evolutionary history shared by the component species. The PD metrics responded to environmental impact and complemented the information provided by classical metrics. This suggests that PD metrics can reflect a certain environmental stress and thus can be used as a metrics of Index of Biotic Integrity (IBI) to reflect the degree of disturbance in river systems.

We evaluated the water quality of the Irtysh River and Ili River based on the combination of macroinvertebrate assemblages and biotic index. Moreover, PD was included into the candidate parameters of Index of Biotic Integrity (IBI). The results of water quality assessment in the Irtysh River through chemical measurements were identical with those of Liu et al. (2002). Thus, the proposed IBI system is appropriate to the studied rivers, and can serve as an effective measure of environment monitoring in river systems. In summary, DNA barcoding can provide a quantitative method to differentiate between Good and Bad water quality. Moreover, our study showed that species identifications based on DNA barcoding have the potential to detect tiny changes in stream condition (e.g., taxa abundance, cryptic species and multiple species lineages).

In conclusion, for the first time, our study constructed a DNA barcoding reference library of benthic macroinvertebrates in transboundary rivers of northwest China, which provides a coverage for 1227 sequences, 189 species, nine taxa of macroinvertebrates. The integration of Barcode Gap analysis, Tree-based methods, ABGD analysis and BIN analysis was adopted to compare with those by morphological identification. Our results demonstrated that DNA barcoding based on COI is an effective method to clarify species boundaries and quantitatively evaluate species diversity, which can be used to evaluate lineages diversity and phylogenetic structure, as well as assess biodiversity and environmental condition for specific areas.

## Acknowledgements

This work was supported by Special Funds for the Foundation Work of Science and Technology (2012FY112700). We would greatly thank postgraduates Gengen Zhou, Lili Wei and Guang Zhao for field sampling and Shiming Wan for help in laboratory analyses.

## Contributions

Zh. Q. conceived and designed the work. G. Y. H. and X. C. X. performed the experiment. Zh. X. J., M. X. F. and W. J. provided help on field sampling or experimental analysis.

## Competing financial interests

The authors declare no competing research interests.

## Data Accessibility Statement

The sequences information used in this manuscript can be found in Barcode of Life Data System (BOLD) at <http://www.boldsystems.org>. BOLD Process IDs as follows: XJDQD001-18-XJDQD1275-18. We will upload the relevant data to Dryad after the manuscript is accepted.

## References

1. Barco, A., Raupach, M. J., Laakmann, S., Neumann, H., & Knebelsberger, 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>
2. Barnett R, Larson G. A Phenol–Chloroform Protocol for Extracting DNA from Ancient Samples[J]. *Methods in molecular biology* (Clifton, N.J.), 2012, 840:13.
3. 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>
4. Blagoev, G.A., Dewaard, J.R., Ratnasingham, S., Dewaard, S.L., Lu, L., Robertson, J., Telfer, A.C., Hebert, P.D. (2016). Untangling taxonomy a DNA barcode reference library for Canadian spiders. *Molecular Ecology Resources*, 16 (1), 325-341. <https://doi.org/10.1111/1755-0998.12440>
5. Brodin, Y., Ejdung, G. & 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>
6. Bucklin, A., Hopcroft, R. R., Kosobokova, K. N., Nigro, L. M., Ortman, B. D., & Jennings, R. M., et al. (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>
7. Candek, K. & Klemen, M. (2015). DNA barcoding gap: reliable species identification over morphological and geographical scales. *Molecular Ecology Resources* . 15 : 268-277. <https://doi.org/https://doi.org/10.1111/1755-0998.12304>
8. Casiraghi, M., Labra, M., Ferri, E., Galimberti, A. & Mattia, F.D. (2010). DNA barcoding: theoretical aspects and practical applications. *Tools for Identifying Biodiversity: Progress and Problems* . 269-273. <https://doi.org/10.1002/jps.22820>
9. Chandler, J.R. (1970). A biological approach to water quality management. *Water Pollution Control* . 69: 415-422.
10. Che, J., Chen, H.-M., Yang, J.-X., Jin, J.-Q., Jiang, K., Yuan, Z.-Y., . . . Zhang, Y.-P. (2012). Universal COI primers for DNA barcoding amphibians. *Molecular ecology resources*, 12 (2), 247-258. <https://doi.org/doi:10.1111/j.1755-0998.2011.03090.x>
11. Chutter, F.M. (1972). An empirical biotic index of the quality of the water in south African streams and rivers. *Water Res* .6 (1), 19-30. [https://doi.org/10.1016/0043-1354\(72\)90170-4](https://doi.org/10.1016/0043-1354(72)90170-4)
12. Curt, L.E., Lynda, D.C., Claudia, B., Erica, L.C. & David, J.B. (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>
13. Danks, & H., V. (2007). How aquatic insects live in cold climates. *The Canadian Entomologist*, 139 (4), 443-471. <https://doi.org/10.4039/n06-100>
  14. Dinca, V., Montagud, S., Talavera, G., Hernandez-Roldan, J., Munguira, M. L., Garcia-Barros, E., . . . 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>
  15. Erik, M. *et al* . (2011). Bagley Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society* .30 (1), 217-231. <https://doi.org/10.1899/10-012.1>
  16. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Veijenhoek, 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>
  17. 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>
  18. 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>
  19. Gregory, J.P. *et al* . (2013). Calibration and validation of a regionally and seasonally stratified macroinvertebrates index for West Virginia adeable streams. *Environmental Monitoring and Assessment*. 185 (2), 1515-1540. <https://doi.org/10.1007/s10661-012-2648-3>
  20. Hajibabaei, M., Shhokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. (2011). Environmental barcoding: A next-Generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE* . 4 (6),1-7. <https://doi.org/10.1371/journal.pone.0017497>
  21. Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* ,23 (4), 167-172. <https://doi.org/10.1016/j.tig.2007.02.001>
  22. 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>
  23. 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>
  24. 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>
  25. Hao, J. , Zhang, X. , Wang, Y. , Liu, J. , Zhi, Y. , & Li, X.(2017). Diversity investigation and application of DNA barcoding of Acridoidea from Baiyangdian Wetland, *Biodiversity Science*, 25 (4), 409-417. <https://doi.org/10.17520/biods.2016331>
  26. Hawlitschek, O. , J. Moriniere, Lehmann, G. U. C. , Lehmann, A. W. , & 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>
  27. (17) (PDF) The importance of validated alpha taxonomy for phylogenetic and DNA barcoding studies: a comment on species identification of pygmy grasshoppers (Orthoptera, Tetrigidae). Available from: [https://www.researchgate.net/publication/317550152\\_The\\_importance\\_of\\_validated\\_alpha\\_taxonomy\\_for\\_phylogenetic\\_and\\_DNA\\_barcoding\\_studies\\_a\\_comment\\_on\\_species\\_identification\\_of\\_pygmy-grasshoppers\\_Orthoptera\\_Tetrigidae](https://www.researchgate.net/publication/317550152_The_importance_of_validated_alpha_taxonomy_for_phylogenetic_and_DNA_barcoding_studies_a_comment_on_species_identification_of_pygmy-grasshoppers_Orthoptera_Tetrigidae) [accessed Jan 03 2019].
  28. Huang, Z.H. & Ke, D.H. (2015). DNA barcoding and phylogenetic relationships in Timaliidae. *Genetics & Molecular Research Gmr* , 14 (2), 5943-5949. <https://doi.org/10.4238/2015.June.1.11>
  29. Huang, L., Que, H., Chen, C., Kuang, X., & Xue, F. (2012). Variation of mtDNA COI sequences

- in different geographic populations of the Asian corn borer, *Ostrinia furnacalis*. *Chinese Journal of Applied Entomology* , 49 (6), 1508-1512.
30. 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>
  31. Jun, Y. C., Won, D. H., Lee, S. H., Kong, D. S., & Hwang, S. J.. (2012). A Multi-metric Benthic Macroinvertebrates Index for the Assessment of Stream Biotic Integrity in Korea. *International Journal of Environmental Research and Public Health* , 9 (10), 3599-3628. <https://doi.org/10.3390/ijerph9103599>
  32. Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* . 1 6(2), 111-120. <https://doi.org/10.1007/BF01731581>
  33. Sekine, K., Bayartogtokh, B., & Bae, Y. J. (2017). Post-glacial distribution of a Mongolian mayfly inferred from population genetic analysis. *Freshwater Biology* . 62 (1), 102-110. <https://doi.org/10.1111/fwb.12853>
  34. Kjaerstad, *et al.* (2012). A review of the Ephemeroptera of Finnmark - DNA barcodes identify Holarctic relations. *Norwegian Journal of Entomology*. 59 (2), 182-195.
  35. Kress, W.J. & Erickson, D.L. DNA Barcodes. (Humana, 2012)
  36. Liu, K.H., Pan, X., Xie L.X. (2002). Present condition of water quality about Irtysh river. *Northwest Water Resources & Water Engineer* . 13 (1), 46-49.
  37. 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.scitotenv.2018.05.052>
  38. 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>
  39. Ratnasingham, S. & Hebert P.D.N. (2007). BOLD: the Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes* . 7 (3), 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678>
  40. Raupach, M. J. , Hendrich, L. , Kuchler, Stefan M., Deister, F. , Moriniere, Jerome, & 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). <https://doi.org/10.1371/journal.pone.0106940>
  41. Ren, M.L., *et al* . Fisheries resources and fishery of the ertixhe river in China. (Xinjiang Science and Health Press, 2002).
  42. Resh, V. (1995). Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Austral Ecology* . 20 , 108-112. <https://doi.org/10.1111/j.1442-9993.1995.tb00525.x>
  43. Morse, J.C., Yang, L.F., Tian, L.X. (1994). Aquatic Insects of China Useful For Monitoring Water Quality. 92-102.
  44. Satiou, N. & Nei, M.(1987). The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4 , 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
  45. Schmid-Egger, C., *et al.* (2017). vision of the West Palaearctic *Polistes* Latreille, with the descriptions of two species - an integrative approach using morphology and DNA barcodes (Hymenoptera, Vespidae). *Zookeys*. 713, 53-112. <https://doi.org/10.3897/zookeys.713.11335>
  46. 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>
  47. 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
48. Stahls, G. & Savolainen, E. (2008). MtDNA COI barcodes reveal cryptic diversity in the *Baetis vernus* group (Ephemeroptera, Baetidae). *Molecular Phylogenetics & Evolution*. 46 (1), 82-87. <https://doi.org/10.1016/j.ympev.2007.09.009>
  49. Stein, E.D., *et al* .(2014). Does DNA barcoding improve performance of traditional stream bioassessment metrics? *Freshwater Science*. 33 (1):302-311. <https://doi.org/10.1086/674782>
  50. Sweeney, B. W. , Battle, J. M. , Jackson, J. K. , & Dapkey, T. (2011). Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? *Journal of the North American Benthological Society* , 30 (1), 195-216. <https://doi.org/10.1899/10-016.1>
  51. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* . 30 (12): 2725-2729. <https://doi.org/10.1093/molbev/mst197>
  52. Tan, Y.L. (2011). Study on Benthic ecology in the e-er-qi-si river basin and the adjacent inland river basin. Shanghai ocean university.
  53. Wang, B.X. (2003). Water quality bioassessment using benthic macroinvertebrates. Nanjing Agriculture University.
  54. 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.
  55. 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.
  56. Wang, S. J . (2010). A complete book of rivers and lakes in Xinjiang, China . Beijing, WaterPower Press.
  57. Wong, W. H. , Tay, Y. C. , Puniamorthy, 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>
  58. 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>
  59. Zhou, C.F. (2002). A Taxonomic study on Mayflies from Mainland China (Insecta: Ephemeroptera). Naikai University. 44-45.
  60. 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 hydroiologica sinica* . 35(6), 963-971.
  61. 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>
  62. Zhou, X., Kjer, K.M. & Morse, J.C. (2007). Associating larvae and adults of Chinese Hydropsychidae caddisflies (Insecta:Trichoptera) using DNA sequences, *Journal of the North American Benthological Society*. 26 (4), 719-742. <https://doi.org/10.1899/06-089.1>
  63. 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), 30-30. <https://doi.org/10.1186/1742-9994-6-30>
  64. Zhou, X., Robinson, J. L., Geraci, C. J., Parker, C. R., Flint, O. S., Jr., Etnier, D. A., . . . 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>
  65. Zhou, X. (2009b). The larvae of Chinese Hydropsychidae (Insecta: Trichoptera), Part I: *Arcopsyche shimianensis* , *Parapsyche* sp. A and *Diplectrone obscura* . *Zootaxa*. 2174 (2174), 1-17.

<https://doi.org/10.1002/cne.22076>

66. Zou, S.M., Li, Q., Kong, L.F., Yu, H. & Zheng, X.D. (2011). Comparing the Usefulness of Distance, Monophyly and Character-Based DNA Barcoding Methods in Species Identification: A Case Study of Neogastropoda. *PLOS ONE* . 6 (10). <https://doi.org/10.1371/journal.pone.0026619>

### Figure captions

**Fig. 1** Sketch map showing the sampling locations of benthic macroinvertebrates in the 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 their main branches (a: Irtysh River, b: Emin River, c: Bortala River and d: Ili River). Grey areas represent the lakes that link these transboundary rivers.

**Fig. 2** The distribution histograms of mean intra-specific distance and the distance to nearest neighbor based on Kimura 2-parameter distance.

**Fig. 3** Barcode gap plot showing the distance to the nearest neighbor (NN) vs. the maximum intraspecific distance Kimura 2-parameter (K2P) for 189 species. Dots above the 1:1 line indicate the presence of a barcode gap.

**Fig. 4** The neighbor-joining trees of 189 macroinvertebrate species (A), Ephemeroptera (B), Plecoptera (C), Trichoptera (D), Hemiptera (E), Annelida (F), Coleoptera (G), Odonata (H), Mollusca (I) and Diptera (J) based on K2P distance.

**Fig. 5** Box diagram of phylogenetic diversity (PD) at the reference and disturbance sites.

### Hosted file

Figures.docx available at <https://authorea.com/users/324874/articles/452948-the-efficacy-of-dna-barcoding-in-the-classification-biodiversity-and-environmental-assessment-of-benthic-macroinvertebrates>