*diversity*

MDPI

*Article*

# Towards a Comprehensive DNA Barcode Library of *Stenochironomus* Kieffer, 1919 (Diptera: Chironomidae) from China

**Hui Wang [1,2,†], Hai-Feng Xu [1,2,†], Chen-Hong Li [1,2], Hai-Xin Zhang [3], Yan-Ping Zhang [3], Bing-Jiao Sun [4,*] and Xiao-Long Lin [1,2,*]**

[1] Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, Shanghai 201306, China; d210100001@st.shou.edu.cn (H.W.)
[2] Shanghai Universities Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China
[3] Jiangxi Institute of Fisheries Science, Nanchang 330039, China
[4] China National Environmental Monitoring Centre, Beijing 100012, China
* Correspondence: sunbingjiao@cnemc.cn (B.-J.S.); xllin@shou.edu.cn (X.-L.L.)
† These authors contributed equally to this work.

**Abstract:** The adoption of DNA-based assessments for biodiversity monitoring has been on the rise. However, the effectiveness of DNA-based taxonomic assignments heavily relies on the availability and reliability of DNA barcode libraries. There is growing demand for a comprehensive understanding of aquatic biodiversity and the critical role of Chironomidae, specifically *Stenochironomus* in freshwater ecosystems. Therefore, our objective is to develop a reference barcode library for *Stenochironomus* in China. From 2016 to 2021, we collected *Stenochironomus* specimens in diverse Chinese landscapes using malaise traps, light traps, and sweep nets. These specimens were carefully preserved for DNA extraction and barcode sequencing. Our analysis unveiled 36 unique operational taxonomic units from 180 COI barcode sequences through a Neighbor-Joining tree and Automatic Barcode Gap Discovery program, highlighting a significant diversity within the *Stenochironomus* species. The findings emphasize the constraints of conventional morphological identification methods, especially for species with ambiguous morphologies. It also underscores the effectiveness of DNA barcoding in revealing hidden species diversity, known as cryptic species. Consequently, this study advocates for an integrated taxonomic approach, combining morphological and molecular data, to refine species identification and conservation strategies.

**Keywords:** DNA barcoding; Chironominae; cryptic species; COI



**Citation:** Wang, H.; Xu, H.-F.; Li, C.-H.; Zhang, H.-X.; Zhang, Y.-P.; Sun, B.-J.; Lin, X.-L. Towards a Comprehensive DNA Barcode Library of *Stenochironomus* Kieffer, 1919 (Diptera: Chironomidae) from China. *Diversity* **2024**, *16*, 257. https://doi.org/10.3390/d16050257

Academic Editor: Manuel Elias-Gutierrez

Received: 7 April 2024
Revised: 17 April 2024
Accepted: 20 April 2024
Published: 24 April 2024



## 1. Introduction

Habitat destruction, pollution, and climate change are major causes of biodiversity loss, especially in streams and rivers [1]. Despite the increasing risks, our understanding of its impacts on freshwater biodiversity remains limited. Unlike terrestrial habitats, inland waters are less protected from environmental impacts [2,3]. Improper land use practices, such as wastewater effluent, deforestation, and overgrazing, can lead to changes in river hydrology and water quality [4]. Freshwater invertebrates, especially aquatic insects, play a crucial role in the energy flow within aquatic ecosystems and are integral to the food chain as the dominant functional feeding group [5]. Moreover, due to their sensitivity to environmental changes, aquatic insects serve as important indicators for monitoring and protecting aquatic systems [6]. Thus, these are commonly used to assess the environmental quality of freshwater ecosystems [7].

The Chironomidae, also known as non-biting midges, are the most abundant aquatic insects in freshwater ecosystems. Over 6300 species occur in nearly every zoogeographical region on Earth, ranging from rainforests to the polar regions. They occupy a pivotal

niche and hold a lower position in the food chain, making them excellent indicators of environmental conditions [8,9]. Being the most widely distributed insect species in the world, they are vital models for studies in species evolution and global biogeography [9]. Meanwhile, their outstanding ability to act as environmental indicators has led to their widespread use in ecological monitoring and water quality assessment [10].

Accurate species identification is the most critical step in monitoring aquatic ecology [11]. Traditionally, species identification relies heavily on morphological features, which could be challenging, especially for macroinvertebrates that undergo metamorphosis, leading to differences between larvae and adult individuals [12]. To complement the traditional methods, DNA-based approaches have emerged as a powerful tool for species delimitation. DNA barcoding, first introduced in 2003 by Hebert and his collaborators, is commonly used in many research fields, such as species identification, diet analysis, and environmental DNA [13]. Compared to traditional methods, DNA barcoding offers several advantages. The usage of standardized molecular barcodes enables easy comparison between different individuals. For example, The Consortium for the Barcode of Life (CBOL) has proposed various working groups to identify universal barcoded genes, such as COI for Metazoa; rbcL, matK, and ITS for plants; ITS for fungi; and 16S rRNA genes for bacteria. DNA barcoding also can be used in a variety of mixed samples, such as soil and water samples. Nevertheless, the selection of a barcode gene is the most critical and it should satisfy certain criteria consisting of a clear "barcode gap" between maximum intraspecies and minimum interspecies distances among a group of species. Furthermore, the fragment should possess highly conserved flanking sequences to cover a wide range of species and a short amplification length to facilitate DNA amplification and sequencing [14]. The most widely accepted and standardized species-level DNA barcode is Cytochrome *c* oxidase subunit I (COI), a 658 bp DNA sequence located in the mitochondrial genome [9]. The cost of sequencing has decreased significantly in recent years, leading to the formation of more species-specific and public reference databases. One notable example is the Barcode of Life Data System (BOLD, http://www.barcodinglife.org, 20 March 2024) [15], a publicly available database that can store, analyze, and distribute DNA barcodes. There are also some regional databases for invertebrates [16,17]. BOLD, being the most famous online public barcode reference library, covers more than 3600 barcodes of chironomid species (up to 20 March 2024) [15].

The *Stenochironomus* Kieffer, 1919 (Figure 1) is a noteworthy genus in the Chironomidae family, known for its wide distribution and considerable species variety [18]. The larvae inhabit various habitats such as swamps, ponds, streams, rivers, and even trophic rainforests. According to the Catalogue of Life (https://www.catalogueoflife.org/, 20 March 2024), a total of 103 species was recorded up to Feb, 2023. However, due to convergent evolution, some morphological characteristics of larvae of this genus are too similar to distinguish. Furthermore, some adult individuals exhibit distinctive morphological features, such as patterned coloration [19]. In 2008, three new *Stenochironomus* species were recorded, and a total of ten Chinese *Stenochironomus* species were reviewed in 2011 [20,21,21]. Additionally, more and more new species were recorded and described, such as *Stenochironomus zhengi*, *Stenochironomus brevissimus*, and *Stenochironomus linanensis* [22,23]. Although a few DNA barcode records of *Stenochironomus* have been published, it is clear that further research is needed to develop a comprehensive COI DNA barcode reference library for this genus in China. In our study, we aim to address this gap by curating barcode sequences from 180 individuals representing 32 morphotypes of *Stenochironomus*. This will contribute to the development of a comprehensive DNA barcode reference library. Additionally, we will investigate barcode gaps and explore the possibility of cryptic species within *Stenochironomus*.


**Figure 1.** Pictures of adult male (**a**), female (**b**), and larvae (**c**) individuals of the *Stenochironomus* Kieffer. Photos (**a**,**b**) were photographed at the Summer Palace in Beijing by Ying-Chao Li, a Senior experimentalist at the Beijing Forestry University Museum. Photo (**c**) was the larvae of *Stenochironomus okialbus* and was photographed by Xiao-Long Lin, Shanghai Ocean University.

## 2. Materials and Methods

From 2016 to 2021, we conducted extensive fieldwork across China's mainland (longitude: 98.795°~128.144° E; latitude: 18.692°~42.4567° N) to collect *Stenochironomus* specimens. Various collection methods were employed, such as malaise traps, light traps, and sweep nets. The collection sites ranged in altitude from 16 m to 2558 m. Adults were stored in 85% ethanol, whereas larvae and other immature individuals were preserved in 95% ethanol. All samples were stored in dark conditions at a consistent temperature of 4 °C until they were ready for further analysis to preserve the integrity of DNA and samples. Selected specimens with intact morphology were chosen for detailed examination and prepared for DNA extraction. Materials were processed at Shanghai Ocean University. To minimize cross-contamination and ensure the integrity of the DNA samples for subsequent analysis, only legs or thoracic tissue from one side of the specimen were used for DNA extraction. Then, the remaining parts of the specimens were mounted on microscope slides with Euparal® and kept as vouchers. To ensure the robustness and reliability of the findings, each specimen was subjected to an analytical process that allowed for both targeted and broad-spectrum genetic insights. The DNA extraction was implemented with a CW2298M DNA extraction kit (CWBIO, Beijing, China). The DNA barcode sequence was amplified with the universal primers LCO1490 and HCO2198 and the PCR product was sent for Sanger sequencing [24]. The PCR was carried out with the reaction mix, including 12.5 μL of Taq Master Mix (Blue dye, Vazyme, Nanjing, China), 0.5 μL of both primers, 6.5 μL of PCR water, and 5 μL of the extracted DNA template. The PCR protocol was rigorously designed to optimize DNA amplification, starting with an initial denaturation at 94 °C for 2 min, followed by a series of temperature cycles tailored to the specific requirements of the COI barcode amplification [17].

Initial data processing was carried out using the Geneious (11.0.14.1+1, Biomatters, Auckland, New Zealand) software [25]. Two sequences were assembled after removing primers and low-quality bases. To verify the integrity of the coding sequences and ensure the presence of functional genes, the MUSCLE function was used for sequence alignment and the screening of stop codons [26]. Additionally, a comprehensive search was conducted in BOLD to identify more publicly available *Stenochironomus* barcode records within China (until 5 February 2024).

DNA barcode gaps refer to the genetic distances between and within species. For groups of species with limited information, DNA barcoding can be used directly to classify species into putative Operational Taxonomic Units (OTUs). Automatic Barcode Gap Discovery (ABGD) uses pairwise distance to identify the so-called "barcode gap" between species [27]. Additionally, it infers confidence for model-based intraspecific differentiation from the data by using the range of prior intraspecific differentiation. The method then detects barcode gaps as the first significant gap above this limit and uses it to partition

the data. The method recursively applies limit inference and gap detection to obtain finer partitions until no further partitions are available. Degrees of intra- and interspecific sequence divergence were quantified by the ABGD with the Kimura 2 Parameter (K2P) model (https://bioinfo.mnhn.fr/abi/public/abgd/, 10 February 2024). Finally, a Neighbor-Joining (NJ) tree was constructed within MEGA X (Pennsylvania, USA) using the K2P model with 1000 bootstrap replications.

## 3. Results

### *3.1. Sequence Information*

We successfully sequenced 114 COI barcode sequences of 24 *Stenochironomus* species from various provinces in China. Additionally, we collected 66 COI barcode sequences of 10 species from the BOLD system, resulting in a total of 180 barcode sequences. These sequences were obtained from larvae and adult individuals collected from the Jilin, Zhejiang, Jiangxi, Hunan, Hubei, Sichuan, Yunnan, Guangxi, and Taiwan provinces. All of the barcode sequences obtained were of high quality, with a length over 600 bp long without any deletion, insertion, or stop codons. The minimal and maximal lengths of those barcode sequences were 606 bp and 669 bp, respectively.

### *3.2. Inter- and Intraspecific K2P Distance*

The analysis of the genetic divergence of COI barcodes revealed interesting findings. At the species level, the average intraspecific divergence was 0.58% (SE:0.03), with the highest intraspecific divergence being 8.78%, found in *Stenochironomus okialbus* (Figure 2), indicating relatively low genetic variation within species. The average interspecific distance was 13.1% (SE: 0.05), and the smallest interspecific divergence was observed between *Stenochironomus* sp. 10XL and *Stenochironomus* sp. 11XL at 10.39%, and the highest interspecific divergence stood at 17.34% between *Stenochironomus* sp. 4XL and *Stenochironomus* sp. 6XL. These findings highlight the existing genetic demarcations between species within the genus. For the entire dataset, the distance of 32 morphotypes to their nearest neighbor was higher than 10%, with four species owning an intraspecific distance exceeding 2%, namely, *Stenochironomus hainanus* (7.22%), *Stenochironomus okialbus* (8.78%), *Stenochironomus* sp. 11XL (7.35%), and *Stenochironomus* sp. 8XL (7.78%).

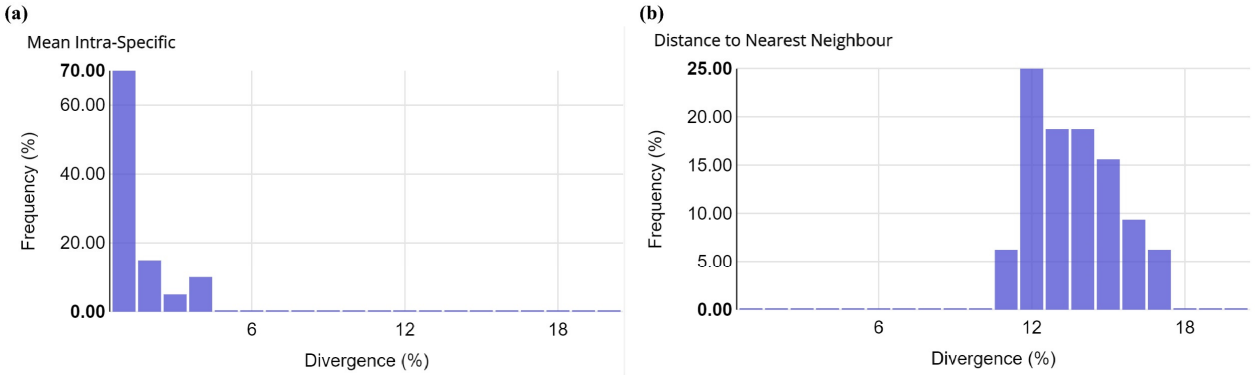


**Figure 2.** Distance distribution of mean intraspecific distance and distance to nearest neighbour. (**a**) The distance distribution histogram shows the mean intraspecific divergence. (**b**) The distance distribution histogram shows the distances to nearest neighbour.

According to the Barcode Gap analysis, all the plots were above the 1:1 oblique line, representing that the minimum interspecific divergence was found to be higher than the mean and max intraspecific distance (100%), which indicated the presence of a barcode gap (Figure 3a,b). Furthermore, based on species against max intraspecific distance, no strong sampling bias was observed (Figure 3c).

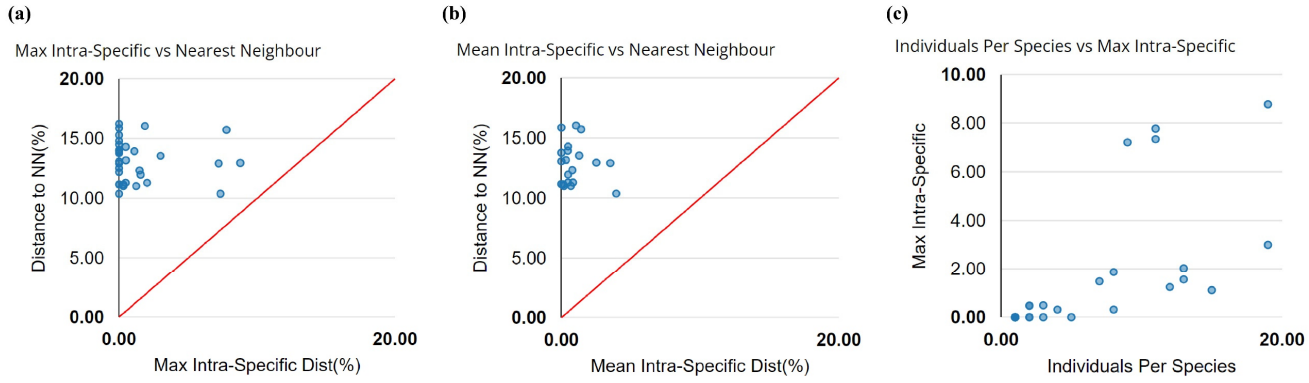


**Figure 3.** Barcode gap analysis of 180 COI barcode sequences of *Stenochironomus* species from China. (**a**) The maximum intraspecific distance to the nearest neighbor (NN) species (**b**) The mean intraspecific distance to the nearest neighbor (NN) species (**c**) The number of individuals in each species against their max intraspecific distances.

### *3.3. OTU Delineation Based on Distance Using ABGD*

The data showed a clear divergence in interspecific and intraspecific pairwise K2P distance (Figure 4). The number of pairwise variations decreased to 0 at 5% and increased from 6% to 9%. Then it increased from 10%, followed by serval peaks at 15% and 19%. All the pairs could be assigned to species levels under 5%. Within the 6% to 9% range, all the pairs were in agreement with the same genus. The lowest divergence to separate two species belonging to the same genus was at 9%, which was observed in *Stenochironomus okialbus* (max intraspecific distance at 8.78%). However, with similar morphological features, cryptic species could present when the intraspecific divergence is larger than 10%. While the prior intraspecific divergence increased gradually, the number of MOTUs converged continuously and stabilized to 36 species, indicating that cryptic species exist within our samples (Figure 4b, File S1).

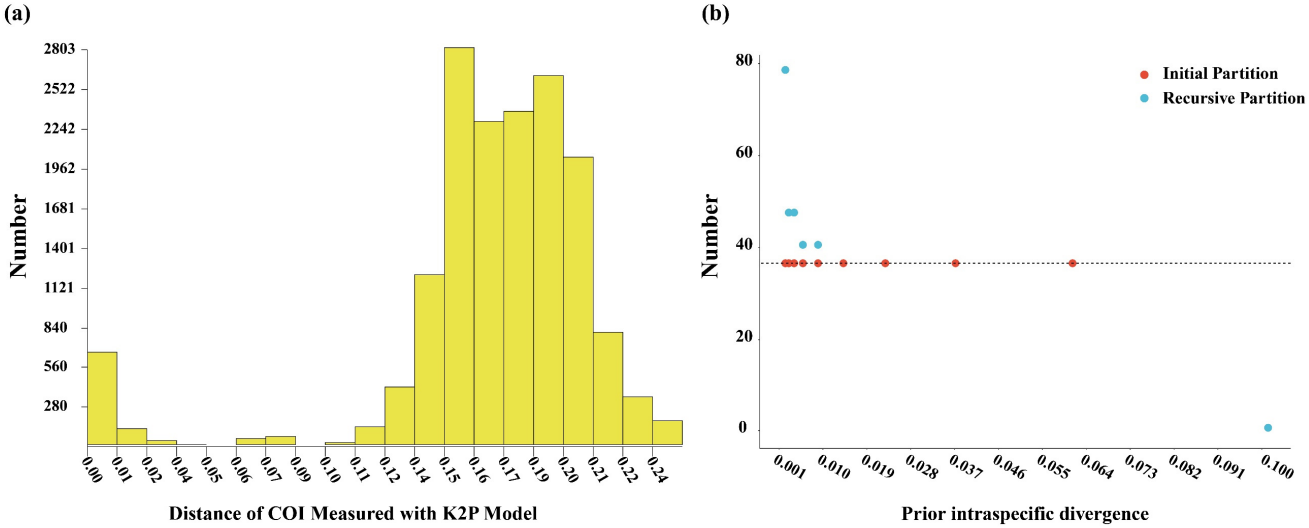


**Figure 4.** Histogram of pairwise K2P distance of COI sequences (**a**) and number of OTUs with the prior intraspecific divergence (**b**).

The BOLD implemented a BIN assignment by the *p*-distance seed threshold at 2.2% and clustered BINs according to genetic divergence among sequences. In total, 180 sequences of 32 morphotypes were assigned to 41 BINs. Besides 12 species that only have one individual assigned to a single BIN, there were 13 species with more than 2 individuals, also assigned to single BINs, such as *Stenochironomus annulus*: BOLD:ADK3619, *Stenochirono-*

*mus linanensis*: BOLD:AEG1837, *Stenochironomus macateei*: BOLD:ADL5641, *Stenochironomus* sp. 13XL: BOLD:AEC0447, *Stenochironomus* sp. 15XL: BOLD:AEC1431, *Stenochironomus* sp. 17XL: BOLD:AEB9437, *Stenochironomus* sp. 18XL: BOLD:AFN3855, *Stenochironomus* sp. 1XL: BOLD:ADK4980, *Stenochironomus* sp. 20XL: BOLD:AEN7975, *Stenochironomus* sp. 22XL: BOLD:ADD7020, *Stenochironomus* sp. 6XL: BOLD:AEC0006, *Stenochironomus* sp. 9XL: BOLD:AEJ1073, *Stenochironomus zhengi*: BOLD:AEG5343. In addition, we observed one-to-more BIN matches in our data; more than six species were assigned to more than one BIN (*Stenochironomus baishanzuensis*: BOLD:AEC0207, BOLD:AEL2939; *Stenochironomus hainanus*: BOLD:ADF2136, BOLD:AEJ4351, BOLD:AET5459; *Stenochironomus okialbus*: BOLD:ADC5272, BOLD:AEG6557; *Stenochironomus* sp. 11XL: BOLD:AEC0226, BOLD:AEO0097; *Stenochironomus* sp. 4XL: BOLD:ADP1847, BOLD:ADP3912, BOLD:AEN6782; *Stenochironomus* sp. 8XL: BOLD:AEG5843, BOLD:AET9193).

### *3.4. Phylogenetic Tree-Based Identification and Cluster Analysis*

An NJ-based phylogenetic tree of 32 morphotypes showed the conspecifics in monophyletic clades with high bootstrap values (Figure 5). Among these clusters, ten of them corresponded to previously identified species. The remaining 26 clusters pointed to a considerable number of unnamed species, emphasizing the rich, yet partially unexplored, biodiversity within the *Stenochironomus* sp. 16XL and *Stenochironomus annulus, Stenochironomus macateei,* and *Stenochironomus linanensis* stayed at close leaves in the NJ tree. Furthermore, sequences from different sampling areas of the same species tend to form separate branches according to the NJ tree (Figure S1). Species *Stenochironomus hainanus* were collected from Shaoguan (Guangdong), Yilan (Taiwan), and Honghe (Yunnan), and they formed three branches in the NJ tree. Interestingly, species within the Honghe population (Yunnan) were separated into two branches. *Stenochironomus okialbus* also showed the same situation and formed two separate branches consisting of populations from Zhejiang (Lishui, Taizhou, and Wenzhou), Guanggong (Shaoguan), Hunan (Chenzhou), and Yunnan provinces (Honghe). Additionally, those species were clustered into more than one BIN cluster.

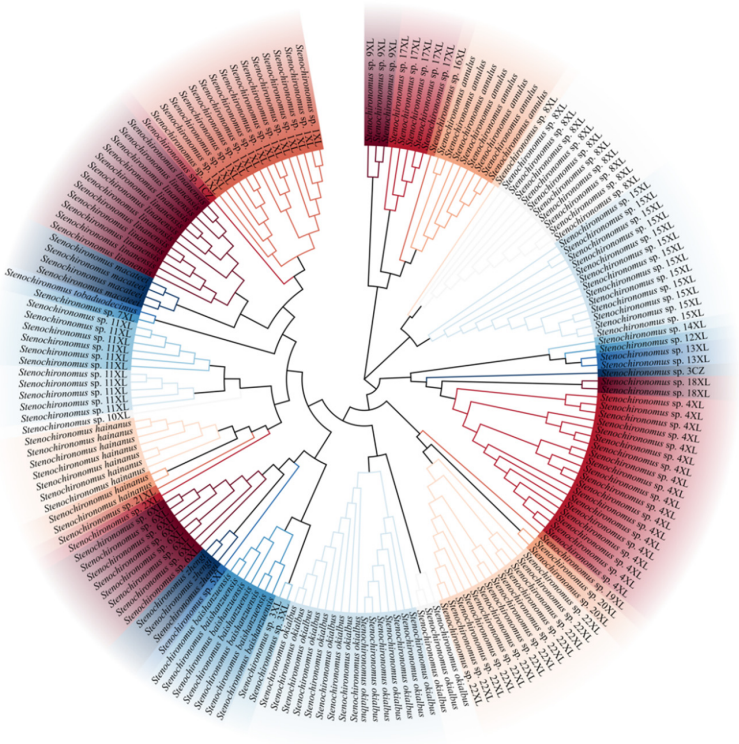

**Figure 5.** Neighbor-Joining tree based on K2P distance model of all barcode sequences generated in this study with 1000 bootstraps. Only a bootstrap value larger than 70% was shown.

## 4. Discussion

We collected 180 barcode sequences of the *Stenochironomus* from China and established a preliminary reference barcode library. Samples were collected from a vast range of altitudes, from 16 m above sea level to 2558 m, and a broad geographical range, including latitudes ranging from 18.692° N to 42.4567° N and longitudes ranging from 98.795° E to 128.144° E. These diverse sampling locations encompassed various terrains across China. The lengths of barcode sequences were all above 600 bp, including 35 sequences with a complete barcode length (658 bp). It is reported that the intraspecific distance of the COI barcode is less than 2%, and the interspecific distance is more than 4% among most species [28]. The BIN system combined the morphological characteristics and the barcode sequence with well-curated taxonomy, clustering barcoded sequences with a pairwise distance threshold of 2.2% [29]. The analysis of these sequences through the BIN system resulted in allocating the sequences into 41 distinct BINs. This categorization comprised 12 singleton BINs, which represent unique sequences without close matches in the dataset, and 13 consistent BINs, indicating groups of sequences with significant similarity. Notably, species like *S. hainanus* and *S. okialbus* were assigned to more than two BINs, suggesting the presence of potential cryptic diversity within these populations. Cryptic lineages are often defined as the occurrence of morphologically indistinguishable evolutionary branches within a species [30]. Despite morphological similarities, cryptic species may not only have unique evolutionary trajectories but also respond differently to environmental changes. The size of *Stenochironomus* may be higher than is currently recognized. Notably, each of the 12 species was associated with a single BIN, emphasizing the diversity within the *Stenochironomus* and the specificity of the DNA barcoding method in identifying unique species signatures. Owing to the absence of comprehensive descriptions of the larval or pupal stages of certain *Stenochironomus* species and the difficulty in distinguishing the females, it proved strenuous to accurately classify a larva of *Stenochironomus* sp. 9XL (BOLD:AEJ1073) and pupae of *Stenochironomus* sp. 22XL (BOLD:ADD7020) at the species level using morphological traits alone. This limitation in morphological differentiation emphasizes the indispensable role of molecular markers in complementing traditional taxonomy, especially in the context of distinguishing cryptic or closely related species [31]. While the identification results from the database only serve as a reference; it is important to know that a combination of traditional morphological data coupling with a DNA barcode should be used for accurate species identification [32].

In the past decade, various methods were developed for species delimitation, which could be broadly categorized into two types. The first category encompasses methods such as SpedesSTEM, BPP, and BFD, designed for multi-locus data. However, these methods are often computationally consuming [33–35] and are usually applied with data of limited size. The second category is based on the phylogenetic tree and genetic distance for a single barcode. Although single loci may not accurately represent the number of species well, this kind of method was widely applied to the DNA barcode since there is no need for predefined species hypotheses, such as GMYC (general mixed Yule-Coalescent model), PTP (Poisson tree process), and ABGD [36–38]. GMCY and PTP need a phylogenetic tree as the input, and ABGD inferred the so-called "barcode gap" only with genetic distance [39]. A threshold was inferred from the gap between the minimum intraspecific and the maximum interspecific distance, and was used for species delimitation. For generating the gap threshold, AGBD needs to accept a prior user-supplied value of intraspecific divergence and will provide multiple species partitions corresponding to different priori values for a single dataset. A barcode gap is typically used to delineate species; however, the overlap of intraspecific and interspecific distances could suggest either a high level of genetic variability within species (intraspecific divergence) or recent speciation events where species have not yet accumulated significant genetic differences. Although the above methods have their own limitations, they generally yield consistent species partitions [40]. It has been suggested that 2% of genetic distance divergence is the threshold for species delimitation, and interspecific distance is usually 8 to 10 times larger than intraspecific

distance [41]. The absence of a clear "barcode gap" also suggested possible cryptic species diversity, where morphologically similar species exhibit significant genetic differences. According to our study, most of the barcode divergence was lower than 2%. In fact, genetic markers exhibit varying genetic distances, even within the same type of gene marker. Nuclear ribosomal ITS regions on helminths exhibited the largest genetic distance, whereas the nuclear rRNA genes displayed the smallest genetic distance. Furthermore, the genetic distance observed in mitochondrial rRNA genes was comparable to that of mitochondrial protein-coding genes, but significantly higher than that of nuclear rRNA markers [42]. In addition, DNA-based species delimitation can be compromised by inadequate sampling and species rarity, including 'singletons', making it difficult to estimate evolutionary processes within and between species [16]. For example, *S. zhengi* only sampled three individuals at the same location. In addition, the frequent gene exchange may cause a low divergence within species, such as *S. annulus* (mean intraspecific divergence: 0.13) collected from Zhejiang Province and Fujian Province. While inadequate sampling, hybridization, or incomplete genealogical sorting can impact barcode identification results, DNA barcode analysis also needs to account for the effects of the geographic distribution of samples [43].

As the geographic distance between samples increases, intraspecific variation in the species increases while the distance to neighboring species decreases, leading to more ambiguous identification. Furthermore, the effect of geographic distance on genetic divergence is primarily based on the concept of genetic isolation by distance, like altitude and channel isolation, which may lead to population differentiation [44]. The genetic differentiation of the same species differs in different geographical populations and the genetic distance may exceed the species classification threshold of 2%. In this study, an intraspecific divergence greater than 2% was observed in three species, *S. hainanus* (3.53%), *S. okialbus* (2.52%), and *Stenochironomus* sp. 11XL (3.95%). Unlike other animals, the nonbiting midges do not have excellent motility [8,45]. The individuals of *S. hainanus* collected from Taiwan were assigned to a single BIN (BOLD:ADF2136), indicating the species isolation caused by the Taiwan Strait. Under some circumstances, genetic diversity is not uniformly distributed but concentrated in certain regions [46]. Therefore, for a given spatial distance between two DNA barcode sampling sites, the genetic distance may be greater if at least one of the sampling sites is from a genetic diversity hotspot [47]. For example, the individuals of *Stenochironomus* sp. 11XL collected from Honghe, Yunnan Province, a hotspot of total species, were assigned to two separate BINs. Furthermore, during the process of species identification with DNA barcodes, the size of the samples contained in the reference library also has an important impact on the identification accuracy [47]. Here, the sample size refers to the number of species of the taxon present in the database and the number of individuals within each species. The algorithm can process better results in species identification with more information from a larger sample size. Therefore, the database should be supplemented with species from multiple geographic regions in subsequent studies.

The construction of the NJ tree was to investigate the differences in the determination of the number of OTUs across these various methodologies, providing a comprehensive overview of species diversity within the collected *Stenochironomus* specimens. The NJ tree provides a clear and well-structured phylogenetic representation of the genetic variances within the *Stenochironomus,* making the relationships between different OTUs readily apparent. Such a visual representation not only facilitates the rapid identification of potential species but also provides a solid foundation for further phylogenetic and taxonomic research. Hence, the NJ tree serves as a means to explore species diversity and has appeared as a powerful tool to assess and compare the differences between various species delimitation methodologies [48]. Although our results have shown a high level of consistency, we acknowledge the limitations of using the DNA barcoding technique with only the COI gene for species discrimination. Factors such as significant intraspecific variation, potential underestimation of recent speciation events, and the effects of horizontal gene transfer should be considered when using a single molecular marker [49]. These challenges, coupled with issues such as incomplete lineage sorting and insufficient sampling, further

highlight the importance of integrating morphological data with molecular techniques. Such an integrated approach is crucial for accurately exploring species boundaries and identifying cryptic species.

## 5. Conclusions

In conclusion, our study has markedly refined the taxonomic understanding of the *Stenochironomus* within the diverse freshwater ecosystems of China. By integrating an extensive array of collection methods and the meticulous application of DNA barcoding, we have assembled a significant dataset that not only enhanced the existing genetic database but also revealed the cryptic species diversity. The assembly of 180 COI barcode sequences into 36 unique OTUs demonstrates the efficacy of DNA barcoding as a tool for species identification, particularly for those with ambiguous morphological features. This method has also been instrumental in differentiating 40 distinct BINs, revealing a substantial diversity within *Stenochironomus* that extends beyond the current taxonomic recognition. Our findings ease the complexities inherent in *Stenochironomus* taxonomy, showcasing the limitations of solely morphological features for species discrimination and emphasizing the importance of molecular techniques to clarify these ambiguities. The significant intraspecific variation, especially within species such as *Stenochironomus okialbus*, hints at the existence of yet undetected cryptic species, further highlighting the potential of DNA barcoding in uncovering hidden biodiversity. To obtain more accurate species delimitation and reveal more cryptic diversity, an approach combining morphology and multi-locus molecular data is advocated. Future research should focus on collecting species from multiple geographic regions. Our research is a crucial step in the ongoing effort to describe new taxonomic units and refine existing classifications, which is fundamental to preserving the rich biodiversity of the freshwater ecosystems of China.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d16050257/s1. Figure S1. Neighbor-Joining tree based on K2P distance model of all barcode sequences. File S1. File generated by Automatic Barcode Gap Discovery, including 36 detailed OTUs.

**Author Contributions:** Conceptualization, H.W., H.-F.X. and X.-L.L.; methodology, H.-F.X. and H.W.; software, H.W. and H.-F.X.; validation, H.-F.X., H.-X.Z., Y.-P.Z. and H.W.; formal analysis, B.-J.S.; investigation, H.-X.Z., Y.-P.Z., H.-F.X. and B.-J.S.; resources, X.-L.L. and C.-H.L.; data curation, X.-L.L. and H.W.; writing—original draft preparation, H.-X.Z., Y.-P.Z., H.W. and H.-F.X.; writing—review and editing, C.-H.L., X.-L.L. and B.-J.S.; visualization, H.-X.Z., Y.-P.Z., H.-F.X. and H.W.; supervision, X.-L.L. and C.-H.L.; project administration, X.-L.L.; funding acquisition, B.-J.S. and X.-L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Yangtze River Joint Research Phase II Program (2022-LHYJ-02-0102) and the National Natural Science Foundation of China (31900344).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** A list of all species, specimens, their individual images, georeferences, primers, sequences, and other relevant laboratory data of all *Stenochironomus* specimens are available through the public dataset "*Stenchironomus* COI DNA barcodes from China (DS-24STEN)" in the Barcode of Life Data System (http://www.boldsystems.org, 5 February 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Cardinale, B.J.; Duffy, J.E.; Gonzalez, A.; Hooper, D.U.; Perrings, C.; Venail, P.; Narwani, A.; MacE, G.M.; Tilman, D.; Wardle, D.A.; et al. Biodiversity Loss and Its Impact on Humanity. *Nature* **2012**, *486*, 59–67. [CrossRef] [PubMed]
2. Tickner, D.; Opperman, J.J.; Abell, R.; Acreman, M.; Arthington, A.H.; Bunn, S.E.; Cooke, S.J.; Dalton, J.; Darwall, W.; Edwards, G.; et al. Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan. *Bioscience* **2020**, *70*, 330–342. [CrossRef] [PubMed]

3. Reid, A.J.; Carlson, A.K.; Creed, I.F.; Eliason, E.J.; Gell, P.A.; Johnson, P.T.J.; Kidd, K.A.; MacCormack, T.J.; Olden, J.D.; Ormerod, S.J.; et al. Emerging Threats and Persistent Conservation Challenges for Freshwater Biodiversity. *Biol. Rev.* **2019**, *94*, 849–873. [CrossRef] [PubMed]
4. Albert, J.S.; Destouni, G.; Duke-Sylvester, S.M.; Magurran, A.E.; Oberdorff, T.; Reis, R.E.; Winemiller, K.O.; Ripple, W.J. Scientists' Warning to Humanity on the Freshwater Biodiversity Crisis. *Ambio* **2021**, *50*, 85–94. [CrossRef] [PubMed]
5. Abell, R.; Thieme, M.; Lehner, B. *Indicators for Assessing Threats to Freshwater Biodiversity from Humans and Human-Shaped Landscapes*; Cincotta, R.P., Gorenflo, L.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 103–124, ISBN 978-3-642-16707-2.
6. Brasil, L.S.; Luiza-Andrade, A.; Calvão, L.B.; Dias-Silva, K.; Faria, A.P.J.; Shimano, Y.; Oliveira-Junior, J.M.B.; Cardoso, M.N.; Juen, L. Aquatic Insects and Their Environmental Predictors: A Scientometric Study Focused on Environmental Monitoring in Lotic Environmental. *Environ. Monit. Assess.* **2020**, *192*, 194. [CrossRef]
7. Chowdhury, S.; Dubey, V.K.; Choudhury, S.; Das, A.; Jeengar, D.; Sujatha, B.; Kumar, A.; Kumar, N.; Semwal, A.; Kumar, V. Insects as Bioindicator: A Hidden Gem for Environmental Monitoring. *Front. Environ. Sci.* **2023**, *11*, 1146052. [CrossRef]
8. Cranston, P.S. *Biogeography BT—The Chironomidae: Biology and Ecology of Non-Biting Midges*; Armitage, P.D., Cranston, P.S., Pinder, L.C.V., Eds.; Springer: Dordrecht, The Netherlands, 1995; pp. 62–84, ISBN 978-94-011-0715-0.
9. Chac, L.D.; Thinh, B.B. Species Identification through DNA Barcoding and Its Applications: A Review. *Biol. Bull.* **2023**, *50*, 1143–1156. [CrossRef]
10. Martel-Cea, A.; Astorga, G.A.; Hernández, M.; Caputo, L.; Abarzúa, A.M. Modern Chironomids (Diptera: Chironomidae) and the Environmental Variables That Influence Their Distribution in the Araucanian Lakes, South-Central Chile. *Hydrobiologia* **2021**, *848*, 2551–2568. [CrossRef]
11. Baerwald, M.R.; Goodbla, A.M.; Nagarajan, R.P.; Gootenberg, J.S.; Abudayyeh, O.O.; Zhang, F.; Schreier, A.D. Rapid and Accurate Species Identification for Ecological Studies and Monitoring Using CRISPR-Based SHERLOCK. *Mol. Ecol. Resour.* **2020**, *20*, 961–970. [CrossRef]
12. Karima, Z. Chironomidae: Biology, Ecology and Systematics. In *The Wonders of Diptera—Characteristics, Diversity, and Significance for the World's Ecosystems*; IntechOpen: Rijeka, Croatia, 2021; ISBN 978-1-83968-882-9.
13. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. Biological Identifications through DNA Barcodes. *Proc. R. Soc. B Biol. Sci.* **2003**, *270*, 313–321. [CrossRef]
14. Antil, S.; Abraham, J.S.; Sripoorna, S.; Maurya, S.; Dagar, J.; Makhija, S.; Bhagat, P.; Gupta, R.; Sood, U.; Lal, R.; et al. DNA Barcoding, an Effective Tool for Species Identification: A Review. *Mol. Biol. Rep.* **2023**, *50*, 761–775. [CrossRef]
15. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System: Barcoding. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]
16. Lavrador, A.S.; Fontes, J.T.; Vieira, P.E.; Costa, F.O.; Duarte, S. Compilation, Revision, and Annotation of DNA Barcodes of Marine Invertebrate Non-Indigenous Species (NIS) Occurring in European Coastal Regions. *Diversity* **2023**, *15*, 174. [CrossRef]
17. Lin, X.L.; Mo, L.; Bu, W.J.; Wang, X.H. The First Comprehensive DNA Barcode Reference Library of Chinese *Tanytarsus* (Diptera: Chironomidae) for Environmental DNA Metabarcoding. *Divers. Distrib.* **2021**, *27*, 1932–1941. [CrossRef]
18. Song, C.; Zhu, B.Q.; Moubayed-Breil, J.; Lei, T.; Qi, X. Taxonomic Study on the Genus *Stenochironomus* Kieffer from the Baishanzu Nature Reserve, China (Diptera, Chironomidae). *Zookeys* **2022**, *2022*, 93–113. [CrossRef]
19. Oliver, D.R. The Adult Males of Diamesinae (Diptera: Chironomidae) of the Holarctic Region—Keys and Diagnoses. *Entomol. Scand. Suppl.* **1989**, *34*, 129–154.
20. Qi, X.; Shi, S.; Lin, X.L.; Wang, X.H. The Genus *Stenochironomus* Kieffer (Diptera : Chironomidae) with Three Newly Recorded Species from China. *Entomotaxonomia* **2011**, *33*, 220–230.
21. Qi, X.; Shi, S.D.; Wang, X.H. Two New Species and New Record of the Genus *Stenochironomus* Kieffer in China (Diptera, Chironomidae). *Acta Zootaxonomica Sin.* **2008**, *33*, 526–531.
22. Qi, X.; Lin, X.L.; Liu, Y.D.; Wang, X.H. Two New Species of *Stenochironomus* Kieffer (Diptera, Chironomidae) from Zhejiang, China. *Zookeys* **2015**, *479*, 109–119. [CrossRef]
23. Parise, A.G.; de Pinho, L.C. A New Species of *Stenochironomus* Kieffer, 1919 from the Atlantic Rainforest in Southern Brazil (Diptera: Chironomidae). *Aquat. Insects* **2016**, *37*, 1–7. [CrossRef]
24. Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA Primers for Amplification of Mitochondrial Cytochrome *c* Oxidase Subunit I from Diverse Metazoan Invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3*, 294–299. [PubMed]
25. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
26. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
27. Puillandre, N.; Lambert, A.; Brouillet, S.; Achaz, G. ABGD, Automatic Barcode Gap Discovery for Primary Species Delimitation. *Mol. Ecol.* **2012**, *21*, 1864–1877. [CrossRef] [PubMed]
28. Mutanen, M.; Hausmann, A.; Hebert, P.D.N.; Landry, J.F.; de Waard, J.R.; Huemer, P. Allopatry as a Gordian Knot for Taxonomists: Patterns of DNA Barcode Divergence in Arctic-Alpine Lepidoptera. *PLoS ONE* **2012**, *7*, e47214. [CrossRef] [PubMed]
29. Ratnasingham, S.; Hebert, P.D.N. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* **2013**, *8*, e66213. [CrossRef]

30. Theodoridis, S.; Nogués-Bravo, D.; Conti, E. The Role of Cryptic Diversity and Its Environmental Correlates in Global Conservation Status Assessments: Insights from the Threatened Bird's-Eye Primrose (*Primula farinosa* L.). *Divers. Distrib.* **2019**, *25*, 1457–1471. [CrossRef]
31. Ahmed, S.; Ibrahim, M.; Nantasenamat, C.; Nisar, M.F.; Malik, A.A.; Waheed, R.; Ahmed, M.Z.; Ojha, S.C.; Alam, M.K. Pragmatic Applications and Universality of DNA Barcoding for Substantial Organisms at Species Level: A Review to Explore a Way Forward. *BioMed Res. Int.* **2022**, *2022*, 1846485. [CrossRef]
32. Yang, B.; Zhang, Z.; Yang, C.Q.; Wang, Y.; Orr, M.C.; Wang, H.; Zhang, A.B. Identification of Species by Combining Molecular and Morphological Data Using Convolutional Neural Networks. *Syst. Biol.* **2022**, *71*, 690–705. [CrossRef]
33. Ence, D.D.; Carstens, B.C. SpedeSTEM: A Rapid and Accurate Method for Species Delimitation. *Mol. Ecol. Resour.* **2011**, *11*, 473–480. [CrossRef]
34. Leaché, A.D.; Fujita, M.K.; Minin, V.N.; Bouckaert, R.R. Species Delimitation Using Genome-Wide SNP Data. *Syst. Biol.* **2014**, *63*, 534–542. [CrossRef] [PubMed]
35. Yang, Z.; Rannala, B. Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Mol. Biol. Evol.* **2014**, *31*, 3125–3135. [CrossRef] [PubMed]
36. Zhang, J.; Kapli, P.; Pavlidis, P.; Stamatakis, A. A General Species Delimitation Method with Applications to Phylogenetic Placements. *Bioinformatics* **2013**, *29*, 2869–2876. [CrossRef] [PubMed]
37. Pons, J.; Barraclough, T.G.; Gomez-Zurita, J.; Cardoso, A.; Duran, D.P.; Hazell, S.; Kamoun, S.; Sumlin, W.D.; Vogler, A.P. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Syst. Biol.* **2006**, *55*, 595–609. [CrossRef]
38. Puillandre, N.; Modica, M.V.; Zhang, Y.; Sirovich, L.; Boisselier, M.C.; Cruaud, C.; Holford, M.; Samadi, S. Large-Scale Species Delimitation Method for Hyperdiverse Groups. *Mol. Ecol.* **2012**, *21*, 2671–2691. [CrossRef] [PubMed]
39. Magnusdottir, H.; Palsson, S.; Westfall, K.M.; Jonsson, Z.O.; Goodall, J.; Ornolfsdottir, E.B. Revised Phylogeography of the Common Whelk Buccinum Undatum (Gastropoda: Buccinidae) across the North Atlantic. *Biol. J. Linn. Soc.* **2019**, *127*, 890–899. [CrossRef]
40. Puillandre, N.; Brouillet, S.; Achaz, G. ASAP: Assemble Species by Automatic Partitioning. *Mol. Ecol. Resour.* **2021**, *21*, 609–620. [CrossRef]
41. Avise, J.C.; Walker, D.E. Species Realities and Numbers in Sexual Vertebrates: Perspectives from an Asexually Transmitted Genome. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 992–995. [CrossRef]
42. Chan, A.H.E.; Chaisiri, K.; Saralamba, S.; Morand, S.; Thaenkham, U. Assessing the Suitability of Mitochondrial and Nuclear DNA Genetic Markers for Molecular Systematics and Species Identification of Helminths. *Parasites Vectors* **2021**, *14*, 233. [CrossRef]
43. Vella, A.; Mifsud, C.M.; Magro, D.; Vella, N. DNA Barcoding of Lepidoptera Species from the Maltese Islands: New and Additional Records, with an Insight into Endemic Diversity. *Diversity* **2022**, *14*, 1090. [CrossRef]
44. Goudarzi, F.; Hemami, M.R.; Rancilhac, L.; Malekian, M.; Fakheran, S.; Elmer, K.R.; Steinfartz, S. Geographic Separation and Genetic Differentiation of Populations Are Not Coupled with Niche Differentiation in Threatened Kaiser's Spotted Newt (Neurergus Kaiseri). *Sci. Rep.* **2019**, *9*, 6239. [CrossRef] [PubMed]
45. Floss, E.C.S.; Secretti, E.; Kotzian, C.B.; Spies, M.R.; Pires, M.M. Spatial and Temporal Distribution of Non-Biting Midge Larvae Assemblages in Streams in a Mountainous Region in Southern Brazil. *J. Insect Sci.* **2013**, *13*, 156. [CrossRef] [PubMed]
46. Schmitt, T. Molecular Biogeography of Europe: Pleistocene Cycles and Postglacial Trends. *Front. Zool.* **2007**, *4*, 11. [CrossRef] [PubMed]
47. Gaytán, Á.; Bergsten, J.; Canelo, T.; Pérez-Izquierdo, C.; Santoro, M.; Bonal, R. DNA Barcoding and Geographical Scale Effect: The Problems of Undersampling Genetic Diversity Hotspots. *Ecol. Evol.* **2020**, *10*, 10754–10772. [CrossRef] [PubMed]
48. Pešić, V.; Asadi, M.; Cimpean, M.; Dabert, M.; Esen, Y.; Gerecke, R.; Martin, P.; Savić, A.; Smit, H.; Stur, E. Six Species in One: Evidence of Cryptic Speciation in the Hygrobates Fluviatilis Complex (Acariformes, Hydrachnidia, Hygrobatidae). *Syst. Appl. Acarol.* **2017**, *22*, 1327–1337. [CrossRef]
49. Chaisiri, C.; Liu, X.; Lin, Y.; Fu, Y.; Zhu, F.; Luo, C. Phylogenetic and Haplotype Network Analyses of Diaporthe Eres Species in China Based on Sequences of Multiple Loci. *Biology* **2021**, *10*, 179. [CrossRef]