



Yawen Mu^{1,2}, Chao Song³, Jianghua Yang^{1,*}, Yong Zhang² and Xiaowei Zhang¹



- Nanjing University Xianlin Campus, Nanjing University, 163 Xianlin Avenue, Nanjing 210023, China
 ² Jiangsu Provincial Environmental Monitoring Center Maniing 210019, China
- ² Jiangsu Provincial Environmental Monitoring Center, Nanjing 210019, China
- ³ Freshwater Fisheries Research Center, Chinese Academy of Fishery Science, Wuxi 214081, China
- * Correspondence: yangjianghua1012@163.com

Abstract: Tai Lake, an important biodiversity hotspot of the lower reaches of the Yangtze River in China, possesses its characteristic fish fauna. Barcoding on native species is important for species identification and biodiversity assessment with molecular-based methods, such as environmental DNA (eDNA) metabarcoding. Here, DNA-barcoding coupled with high-throughput sequencing (HTS) and traditional Sanger sequencing was introduced to barcoding 180 specimens belonging to 33 prior morphological species, including the most majority of fish fauna in Tai Lake. HTS technology, on the one hand, significantly enhances the capture of barcode sequences of fish. The successful rate of fish barcoding was 74% and 91% in Sanger and HTS, respectively. On the other hand, the HTS output has a large number (64%) of insertions and deletions, which require strict bioinformatics processing to ensure that the "true" barcode sequence is captured. Cross-contamination and parasites were the primary error sources that compromised attempts at the DNA barcoding of fish species. The barcode gap analysis was 100% successful at delimiting species in all specimens. The automatic barcode gap discovery (ABGD) method grouped barcode sequences into 34 OTUs, and some deep divergence and closed species failed to obtain corresponding OTUs. Overall, the local species barcode library established by HTS barcoding here is anticipated to shed new light on conserving fish diversity in Tai Lake.

Keywords: high-throughput sequencing; DNA barcoding; cytochrome c oxidase I; fish species identification; Tai Lake; metabarcoding

1. Introduction

It is well recognized that the loss of biodiversity caused by environmental deterioration has adverse ecological consequences [1], for example, nutrient over-enrichment in lakes leads to the simplification of flora and fauna [2]. Biodiversity conservation is a usual tool for sustaining the health of an ecosystem [3]. Species delimitation, especially identifying functional species, is the first step to understanding the relationship between biodiversity and ecological services [4–6].

A conventional taxonomist uses the morphological method to describe species, and misidentification often occurs because of phenotypic plasticity and differing life stages [7,8]. In addition, traditional fish surveys generally involve capturing organisms, are invasive for the biological community under study, and conflict with the original intention of biodiversity conservation. The DNA barcode, usually a short and standardized sequence, has emerged as a cost-effective proxy for species identification [9,10], which significantly improves the efficiency of biomonitoring [11,12]. Incomplete reference databases of the DNA barcode, especially for native species, are widely recognized as a major obstacle to the use of molecular-based methods.

The barcode database that was built needed a PCR procedure firstly, which amplified not only target region but also non-intended fragments. Subsequently, Sanger sequencing



Citation: Mu, Y.; Song, C.; Yang, J.; Zhang, Y.; Zhang, X. Next-Generation DNA Barcoding for Fish Identification Using High-Throughput Sequencing in Tai Lake, China. *Water* 2023, *15*, 774. https://doi.org/ 10.3390/w15040774

Academic Editor: Subba Rao Chaganti

Received: 30 September 2022 Revised: 2 November 2022 Accepted: 28 November 2022 Published: 16 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). was used to obtain the target barcode sequence. Although Sanger sequencing is viral and is low-cost, it can only provide a single sequence for each specimen. The single sequence from Sanger sequencing can be the PCR product of the co-amplification of contaminated DNA and may not represent the 'true' barcode sequence, which increases the risk of failure [13].

High-throughput sequencing (HTS) allows for sequencing millions of DNA fragments in parallel and provides an opportunity to reveal the sequence composition of the PCR product [14]. In addition, HTS allows for the generation of multiple sequences for a single specimen and provides an opportunity to identify the contamination. HTS barcoding not only enhances the accuracy of specimen identification but also accelerates the process of barcode capturing [14,15]. The Ion Torrent sequencing platform, in particular, can provide tens of millions of sequences within 10 h for the barcode testing of hundreds of specimens [16,17]. It dramatically improves the efficiency of database construction and reduces the cost. In this study, we established a DNA barcoding library of fish from Tai Lake using HTS and analyzed the cytochrome c oxidase I (COI) barcode gap among fish. The significance of the library was to provide a database for fishery resource surveillance and promote the taking of more careful measures in the conservation of fishery biodiversity.

2. Materials and Methods

2.1. Sampling

Tai Lake, located in the lower reaches of the Yangtze River, is the third largest freshwater lake and harbors a large number of freshwater fish species (Figure 1). Algae bloom caused by eutrophication significantly threatens the lake's ecological health [2] and has led to a water crisis [18]. There were approximately 107 fish species belonging to 14 orders, 25 families, and 73 genera in Tai Lake historically, of which Cyprinidae accounted for more than 60% [19]. The diversity of fish fauna in Tai Lake decreased dramatically, and no more than 40 fish species were found in traditional fishery resource surveillance conducted by a local fishery institute in recent years [20]. In this study, 7 sampling sites were arranged, and a total of 180 specimens belonging to 34 species were collected with gill nets and ground cages. There were 130 specimens of Cyprinidae and 50 specimens representing another ten families [19] (Figure 2).

2.2. DNA Isolation and PCR Amplification

The fish specimens were identified by an empirical taxonomist and were preserved in 95% ethanol. About 0.5 mg tail fin tissue samples were collected, and the DNA was extracted using the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. PCR amplification was performed in a final volume of 50 μ L, made up of 1 μ L of 10 μ M of universal forward (GGWACWG-GWTGAACWGTWTAYCCYCC) and reverse (TAAACTTCAGGGTGACCAAARAAYCA) primers [9], 37.8 μ L of ultrapure water, 5 μ L of 10 \times PCR High Fidelity PCR buffer, 2 μ L of MgSO₄ (50 mM), 1 μ L of dNTP mix (10 mM), 0.2 μ L of Platinum Taq DNA polymerase, and 2 μ L of DNA template (Invitrogen, Waltham, MA, USA, USA). Amplification was performed using a "Touchdown" procedure. PCR conditions were 95 °C for 5 min, 16 cycles of 95 °C for 10 s, 62 °C for 30 s (-1 °C per cycle), and 72 °C for 60 s, followed by 20 cycles of 95 °C for 10 s, 46 °C for 30 s, and 72 °C for 60 s. Negative control was included in the experiment. PCR products were detected by 2% agarose gel.

2.3. Sequencing

Each amplicon was indexed with unique 10-mer multiple identifiers (MIDs). The PCR amplicons were pooled in equal volumes. The purified library was constructed using the Ion Torrent PGM template OT2 400 kit (Life Technologies, Carlsbad, CA, USA) and subsequently sequenced on the "318 V2" chip according to the protocols. At least 100 sequences were provided for each sample. The sequence of each amplicon was also obtained using Sanger sequencing (ABI 3730XL sequencer, Applied Systems Inc., USA) side by side.



Figure 1. Map of sampling sites in this study. Location details and specimens collected per site are provided in Supplementary Materials (Table S1).

2.4. Bioinformatics

All HTS reads were filtered for quality and length and reverse-complemented by the "Biostrings" package of R language [21]. Secondly, we programmed a pipeline using R language (involving the "DECIPHER", "seqinr", and "ShortRead" packages) to trim MIDs and forward and reverse primers and to assign the filtered reads to each specimen [22–24]. The barcode region of each sequence was retained for further processing.

De-replication was conducted by USEARCH [25], and the number of unique sequences each barcode contained, namely, its frequency, was counted simultaneously. Evidently, each specimen had more than one unique sequence owing to contamination and sequence error. Contamination was removed using a genetic distance method. The K2P (Kimura 2-parameter) distance of unique sequences in each specimen was calculated by the "seqinr" package and plotted by nMDS (non-metric multidimensional scaling) [23,26,27]. More than one cluster was observed if containment existed in this specimen. Among all the clusters, the biggest one was retained as a containment-free unique sequence, and other clusters were considered as containment and identified by BLAST against GenBank [28].



Figure 2. Historical list of fish species in Tai Lake, and specimens captured in this study. There were 107 species, historically belonging to 1 class, 14 orders, and 25 families; cyprinidae fish accounted for more than 60 percent of the total species. In total, 180 specimens were captured, belonging to 34 species and 6 orders in this study; among these specimens, cyprinidae also was the most dominant family.

However, there was more than one containment-free unique sequence owing to substitution error. Subsequently, containment-free unique sequences were sorted at the frequency by USEARCH software [25], and one single highest frequency unique sequence was selected as the "true" barcode; if a medium-high frequency unique sequence was detected, a Numt sequence was suspected [29] and validated by comparison to GenBank; other verylow-frequency unique sequences were considered as substitution errors and validated by BLAST against the selected "true" barcode database. Through validation, substitution error unique sequences were generated by substituting one or two bases of a "true" barcode [30].

2.5. Phylogenetic Analysis

The effectiveness of the DNA barcode was validated by two exact opposite procedures, including (1) using the DNA barcode to discriminate between species or to check if the barcode gap exists; (2) successfully predicting unexplored specimens by grouping sequences into OTUs (operational taxonomic units).

The genetic K2P distance-based DNA barcoding method was used to discriminate between species [26]. To visually observe distance within species, two-dimensional nMDS was plotted using a "vegan" package [27]. A comparison between the maximum distance within species and the minimum distance between species in each specimen was performed using the "spider" package to examine the barcode gap [31]. K2P and p-distance-

based NJ (neighbor-joining) trees were implemented in MEGA 6.0 using 1000 bootstrap replicates [32].

Barcode sequences were further clustered into OTUs for the purpose of predicting unexplored specimens. The ABGD (automatic barcode gap discovery) approach on a web interface (www.abi.snvjussieu.fr/public/abgd, accessed on 26 November 2019) was applied using the default parameter (X = 1; K2P button was chosen). This approach is a model for the purpose of seeking the threshold to cluster sequences into hypothetical species [33].

3. Results

A total of 149,892 raw reads were generated by HTS (Figure 3). The raw reads ranged in length from 19 to 420 bp. After filtering quality and trimming MIDs and primers, 108,689 sequences remained, and an average of 603 sequences was assigned to each specimen. Following indel, contamination, and substitution removal, a total of 34,733 error-free reads remained. More than 64% of the error reads were generated by indel, and fewer than 500 reads were contamination reads, including fish cross-contamination and parasites. In each specimen, PGM HTS sequencing errors were proportional to sequencing reads (Figure 4). Owing to the degeneration of primers, some specimens had minor raw reads. Generally, at least 100 error-free reads could successfully be obtained by PGM sequencing per specimen (Figure 4). However, if fish tissues were contaminated by parasites, this would be judged a failure to get the "true" barcode. A total of 9% and 26% of specimens could not obtain "true" barcode sequences with HTS and Sanger sequencing, respectively (Figure 4). Finally, 163 "true" barcode sequences were recovered by PGM sequencing. All barcodes were of a full 313 bp length (Figure 3). Each specimen had only one unique sequence that remained, meaning that an average of 231 error-free reads was assigned to each "true" barcode.



Figure 3. Comparison of barcode sequences obtained by Sanger sequencing and HTS. In the horizontal barplot, the grey bar represents the number of specimens captured from Tai Lake, blue bars represent the number of specimens successfully barcoded by Sanger sequencing (up) and HTS (down), and orange bars represent the number of specimens failed to barcoded by Sanger sequencing (up) and HTS (down). In histogram, each step in bioinformatics is depicted.



Number of barcode region in each speciemen

Figure 4. The scatter plot demonstrates the relations between number of barcode regions in each specimen and the number of error-free reads finally retained. The red circle represents the specimen infected by parasites.

All 163 barcodes recovered by HTS belonged to 33 of 34 priori morphologically identified species, whereas the *Tridentiger bifasciatus* species could not be recovered. It was expected that K2P-based genetic variation hierarchically increased from within species (mean = 0.3%, SE = 0.03), to within genera (mean = 3.82%, SE = 0.23), and to within families (mean = 17.40%, SE = 0.09) (Table 1). Using nMDS to reduce the dimension of genetic distances within species, specimens within species were clustered respectively in the two dimensional plot. Cypriniformes was separated from other orders. Of specimens in the Cypriniformes order, the Acheilognathus family was separated from other families (Figure 5). Overall, a comparison between the maximum distance intra- species and the minimum distance inter-species demonstrated that a barcode gap existed in all analyzed specimens (Figure 6).



Table 1. Depiction of genetic divergences sorted by taxonomic level.

Figure 5. The distribution of genetic divergences in all analyzed specimens was plotted by nMDS method (non-metric multidimensional scaling). The *x*-axis and *y*-axis together contribute 44.8% of the variation in the process of reducing dimensions. Species abbreviations are ordered according to Acheilognathus, other, and cypriniformes family. The species represented by capital letters are as follows: AC: *Acheilognathus chankaensis*; AM: *Acheilognathus macropterus*; AG: *Acheilognathus gracilis*; AI: *Acheilognathus imberbis*; RS: *Rhodeus sinensis*; RO: *Rhodeus ocellatus*; RG: *Rhinogbius giurinus*; HI: *Hyporhamphus intermedius*; OR: *Odontamblyopus rubicundus*; OP: *Odontobutis potamophila*; SA: *Silurus asotus*; CN: *Coilia nasus*; MS: *Mastacembelus sinensis*; MA: *Megalobrama amblycephala*; MS1: *Megalobrama skolkovii*; CD: *Culter dabryi*; CI: *Ctenopharyngodon idellus*; PS: *Pseudobrama simoni*; SN: *Sarcocheilichthys nigripinnis*; PD: *Paramisgurnus dabryanus*; MA2: *Monopterus albus*; MA1: *Misgurnus anguillicaudatus*; CA1: *Channa argus*; HM1: *Hemibarbus maculatus*; HM2: *Hypophthalmichthys molitrix*; MM: *Microphysogobio microstomus*; CC: *Cyprinus carpio*; CE: *Cultrichthys erythropterus*; CA2: Carassius auratus; HN: *Aristichthys nobilis*; HL: *Hemiculter leucisculus*; SN: *Sarcocheilichthys nigripinnis*; HD: *Pelteobagrus fulvidraco*.



The maximum distance within species

Figure 6. The maximum distance within species in each specimen compared with the minimum distance between species in each specimen. All specimens fall above the 1:1 line, indicating the existence of a barcode gap.

The phylogenetic tree based on the K2P distance contained 33 species clusters. The number of OTUs produced by the ABGD method was manifested by a red circle outside the NJ tree (Figure 7). The ABGD analysis produced nine initial partitions. The number of groups and the p distance were 32 to 49 and 0.059948 to 0.001000, respectively. The result of 34 OTUs (p distance = 0.012915) was chosen to set the threshold to delimit species boundaries since it was concordant with the outcome of NJ analysis. Of these 34 OTUs, two species, *Misgurnus anguillicaudatus* and *Monopterus albus*, were delimited to two theoretical species, respectively, whereas *Megalobrama amblycephala* and *Megalobrama skolkovii* were clustered into one candidate species. Other species boundaries represented by OTUs were concordant with morphologically identified species.



Figure 7. Neighbor joining diagram (NJ tree) of all analyzed specimens. The calculation of distance measurement is based on K2P method. Outside the tree, red curve fragments represent OTUs grouped by ABGD method; all 34 red curve fragments are in accordance with their prior morphological species.

4. Discussion

Some major questions in ecology, such as what constitutes the dietary range of a fish species and the assembly of ecological communities, are hampered by traditional morphological identification, owing to laborious work [34,35]. DNA barcoding, integrating ecological, morphological, and genetic data, is anticipated to bring the renaissance of taxonomy [36]. The barcode library of fish species established here not only provided an effective tool for identifying fish communities in Tai Lake [37] but also accurately measured the dietary range of some functional fish species.

Sanger sequencing is the dominant approach for obtaining barcode sequences and has been applied to establish a wide range of barcode libraries, from phytoplankton to vertebrates [11,37–40]. However, no-amplification and co-amplification of non-target sequences usually occur and decrease the efficiency and accuracy of the capture of the "true" barcode [41]. In this study, 26% of the analyzed specimens could not be sequenced by Sanger sequencing. When the HTS approach was used instead, the failure rate was reduced to 9%. Because "touch-down" PCR was performed, the HTS approach could obtain sequences from some specimens with low concentrations of amplifications where there was no measurable fluorescence detected by electropherogram [42]. The HTS approach increased the sensitivity of the capture of the DNA barcode. A previous study where an average of 143 sequences per specimen were generated by the HTS approach provided the proof for our research [15].

The Ion torrent PGM sequencer was chosen owing to more time-saving in comparison with other sequencers, such as Illumina HiSeq and 454 Junior [17]. However, every sequencer introduces errors in the read results. Because of the Ion Torrent based on pH flow call technology, the indel errors occur at very high frequencies when sequencing in massively parallel [43]. A SEAME tool was used to remove the error reads in bioinformatics [44,45]. We chose an irreproachable COI fragment sequenced by the Sanger approach in a bidirectional way as a gold template for BLAST, with all reads generated by the PGM [28]. Indel reads accounted for 64% of the total reads. In a previous study, the error rate was as high as 90%. The difference in the error rate may be due to the different uses of the sequencing kit or the chip density [43].

The biggest advantage of the HTS is the successful acquisition of hundreds of reads in parallel [14], allowing us to thoroughly understand the sequence composition in a single specimen, similarly to multiple clones in Sanger sequencing [46]. Previous studies have found that the "true" barcode was often confused with endosymbiotic bacteria (e.g., Wolbachia) [47], cross-contamination [15], and heteroplasmy [41]. In this study, one instance of parasite infection was detected in the species of cyprinidae. To our knowledge, parasites were first found as an error source that compromised attempts at the DNA barcoding of fish species. Cross-contamination was detected in all of the analyzed specimens. There are no intended fish infections might be due to contact within specimens in a single fish net, which traditional fishery inventory often used [48]. The Sanger-based barcoding method could not discriminate between the cross-contaminations unless enough clones were involved before sequencing. This was laborious work, which the HTS could have easily solved. In the process of bioinformatics, no medium-high frequency unique was detected that is, no heteroplasmy was detected in all specimens. In addition, the HTS process eliminates the need for post-clone sequencing, simplifies the procedure for obtaining barcodes, and reduces the cost of library construction by at least 30%.

This is the first comprehensive molecular assessment of the fish species in Tai Lake, including the most majority of known species up to now. The criteria for examining the discrimination power of DNA barcoding is based on comparing intraspecific and interspecific genetic distances [39], which were in the range of 0% to 9.35% and 0.32% to 23.9%, respectively, in this study. Here, the barcode gap analysis was 100% successful in all of the analyzed specimens. In previous studies, the species discrimination rate ranged from 88% to 95% [49–52]. The inability to differentiate between some instances may be due to cryptic species, which would lead to the incongruence between barcode and morphological identification [53]; another reason may be the haplotype sharing that caused by hybridization among species [37]. In any case, the 100% discrimination rate in this study demonstrated the perfect congruence between barcode sequences and adequate taxonomy.

The mean value of the intraspecific K2P distance of 0.3% (SE = 0.03) calculated here has also been shown for fish populations in the Nujiang River in Southwest China [37] and is in accordance with the value of below 1% calculated when COI was used as a barcoding maker [54,55]. Of the 33 species analyzed here, some extremely intraspecific distances with

the highest value of 9.35% were found in the two species *Misgurnus anguillicaudatus*, and *Monopterus albus* and displayed deep divergence in the NJ tree.

The barcode library is established for the purpose of predicting unexplored fish specimens when captured from the Tai Lake again. Some relative prediction models are designed for delimiting species based on the similarity of the barcode sequence. The "10-fold rule" that all barcode sequences differing at a species level by the 10 fold of the average value of intraspecific distance was introduced as a standard threshold [56]. In this study, the average K2P distance between congeneric species was 13-fold that of the overall intraspecific distance. The threshold calculated here was lower than those reported for other fish barcoding studies, where the value ranged from 15-fold to 70-fold [38,39,51,57].

ABGD is another important prediction model for the first set of species hypotheses performed on the web interface [33]. The ABGD method grouped barcode sequences into 34 OTUs. It was well known that the two species *Misgurnus anguillicaudatus* and *Monopterus albus* displayed deep divergence in the NJ tree. In contrast, the extremely closed distance between the other two species *Megalobrama amblycephala* and *Megalobrama skolkovii* led to the generation of only one OTU. This was due to the asymmetry of divergent COI sequences [56]. Moreover, the difference between the two *Megalobrama* species is only one base of the 313 bp COI sequences. A character- based identification method performed by the BLOG model could be used to delimit them [58]. Overall, when the unexplored specimen is identified, if ABGD is involved in grouping sequences into candidate species, another taxonomic approach should be complemented for the purpose of obtaining a 100% success rate in fish species identification [12,59].

5. Conclusions

With a strict bioinformatics process, high-throughput sequencing (HTS) can significantly improve the capture of barcode sequences of fish species. The successful rate of fish barcoding was 74% and 91% in Sanger and HTS, respectively. Cross-contamination and parasites were the primary error sources that compromised attempts at the DNA barcoding of fish species. Overall, the local species barcode library established by HTS barcoding here is anticipated to shed new light on the conservation of fish diversity in Tai Lake.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/w15040774/s1, Table S1: List of specimens collected from Tai Lake.

Author Contributions: Conceptualization, J.Y. and C.S.; methodology, C.S.; software, J.Y.; validation, J.Y., Y.M. and X.Z.; investigation, C.S.; writing—original draft preparation, C.S. and Y.M.; writing—review and editing, Y.Z., X.Z. and J.Y.; visualization, Y.M.; supervision, Y.Z. and X.Z.; funding acquisition, J.Y. and Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (41807482 and U1901220), and also supported by Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB811). X.Z. was supported by the Fundamental Research Funds for the Central Universities.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Loreau, M.; Naeem, S.; Inchausti, P.; Bengtsson, J.; Grime, J.P.; Hector, A.; Hooper, D.U.; Huston, M.A.; Raffaelli, D.; Schmid, B.; et al. Ecology—Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* 2001, 294, 804–808. [CrossRef] [PubMed]
- 2. Paerl, H.W.; Otten, T.G. Blooms Bite the Hand That Feeds Them. *Science* **2013**, *342*, 433–434. [CrossRef] [PubMed]
- 3. Cardinale, B.J. Biodiversity improves water quality through niche partitioning. *Nature* **2011**, 472, 86–89. [CrossRef]
- Hajibabaei, M.; Janzen, D.H.; Burns, J.M.; Hallwachs, W.; Hebert, P.D. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA 2006, 103, 968–971. [CrossRef] [PubMed]

- 5. Kress, W.J.; Garcia-Robledo, C.; Uriarte, M.; Erickson, D.L. DNA barcodes for ecology, evolution, and conservation. *Trends Ecol. Evol.* **2015**, *30*, 25–35. [CrossRef]
- 6. Valentini, A.; Pompanon, F.; Taberlet, P. DNA barcoding for ecologists. Trends Ecol. Evol. 2009, 24, 110–117. [CrossRef] [PubMed]
- 7. Meyer, C.P.; Paulay, G. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol. 2005, 3, 2229–2238. [CrossRef]
- 8. Moritz, C.; Cicero, C. DNA barcoding: Promise and pitfalls. PLoS Biol. 2004, 2, 1529–1531. [CrossRef]
- 9. Ratnasingham, S.; Hebert, P.D. bold: The Barcode of Life Data System. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef]
- 10. Ward, R.D.; Hanner, R.; Hebert, P.D. The campaign to DNA barcode all fishes, FISH-BOL. J. Fish Biol. 2009, 74, 329–356. [CrossRef]
- 11. Ward, R.D.; Zemlak, T.S.; Innes, B.H.; Last, P.R.; Hebert, P.D. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2005, 360, 1847–1857. [CrossRef] [PubMed]
- 12. Casiraghi, M.; Labra, M.; Ferri, E.; Galimberti, A.; De Mattia, F. DNA barcoding: A six-question tour to improve users' awareness about the method. *Brief. Bioinform.* 2010, *11*, 440–453. [CrossRef] [PubMed]
- 13. Yang, J.; Zhang, X.; Zhang, W.; Sun, J.; Xie, Y.; Zhang, Y.; Jr, G.A.B.; Yu, H. Indigenous species barcode database improves the identification of zooplankton. *PLoS ONE* **2017**, *12*, e0185697. [CrossRef] [PubMed]
- 14. Shokralla, S.; Porter, T.M.; Gibson, J.F.; Dobosz, R.; Janzen, D.H.; Hallwachs, W.; Golding, G.B.; Hajibabaei, M. Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* **2015**, *5*, 9687. [CrossRef]
- Shokralla, S.; Gibson, J.F.; Nikbakht, H.; Janzen, D.H.; Hallwachs, W.; Hajibabaei, M. Next-generation DNA barcoding: Using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol. Ecol. Resour.* 2014, 14, 892–901. [CrossRef]
- Rothberg, J.M.; Hinz, W.; Rearick, T.M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J.H.; Johnson, K.; Milgrew, M.J.; Edwards, M.; et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011, 475, 348–352. [CrossRef]
- Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* 2012, 13, 341. [CrossRef]
- 18. Li, D.; Erickson, R.A.; Tang, S.; Zhang, Y.; Niu, Z.C.; Liu, H.L.; Yu, H.X. Structure and spatial patterns of macrobenthic community in Tai Lake, a large shallow lake, China. *Ecol. Indic.* 2016, *61*, 179–187. [CrossRef]
- 19. Ri, Y.Z.; Cheng, D. The Fisheries in Taihu; Shanghai Press of Scinece and Technology: Shanghai, China, 2005; p. 20.
- 20. Mao, Z.; Gu, X.; Zeng, Q.; Zhou, L.; Sun, M. Status and changes of fishery resources (2009–2010) in Lake Taihu and their responses to water eutrophication. *J. Lake Sci.* 2011, 23, 6.
- 21. Pages, H.A.P.; Gentleman, R.; DebRoy, S. Biostrings: String objects representing biological sequences, and matching algorithms. In *R Package Version* 2.30.1; R Foundation: Vienna, Austria, 2014.
- 22. Wright, E. Database Enabled Code for Ideal Probe Hybridization Employing R. In *R Version: 1.10.1;* R Foundation: Vienna, Austria, 2013.
- 23. Seqinr. Biological Sequences Retrieval and Analysis. In R Package Version 3.1.3; R Foundation: Vienna, Austria, 2014.
- 24. Morgan, M.; Lawrence, M.; Anders, S. FASTQ input and manipulation. In *R Package Version 1.22.0*; R Foundation: Vienna, Austria, 2014.
- 25. Phuong, T.M.; Do, C.B.; Edgar, R.C.; Batzoglou, S. Multiple alignment of protein sequences with repeats and rearrangements. *Nucleic Acids Res.* **2006**, *34*, 5932–5942. [CrossRef]
- 26. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef] [PubMed]
- 27. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Wagner, H. Community Ecology Package. In *R Package Version* 2.3.5; R Foundation: Vienna, Austria, 2016.
- Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 2000, 7, 203–214. [CrossRef] [PubMed]
- Song, H.; Buhay, J.E.; Whiting, M.F.; Crandall, K.A. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. USA* 2008, 105, 13486–13491. [CrossRef] [PubMed]
- 30. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
- 31. Brown, S.; Collins, R.; Boyer, S.; Lefort, M.-C.; Malumbres-Olarte, J.; Vink, C.; Cruickshank, R. Species Identity and Evolution in R. In *R Package Version* 1.3.0; R Foundation: Vienna, Austria, 2013.
- 32. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef]
- Puillandre, N.; Lambert, A.; Brouillet, S.; Achaz, G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* 2012, 21, 1864–1877. [CrossRef]
- 34. Port, J.A.; O'Donnell, J.L.; Romero-Maraccini, O.C.; Leary, P.R.; Litvin, S.Y.; Nickols, K.J.; Yamahara, K.M.; Kelly, R.P. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Mol. Ecol.* **2016**, *25*, 527–541. [CrossRef]
- Yu, D.W.; Ji, Y.Q.; Emerson, B.C.; Wang, X.Y.; Ye, C.X.; Yang, C.Y.; Ding, Z.L. Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* 2012, *3*, 613–623. [CrossRef]

- 36. Miller, S.E. DNA barcoding and the renaissance of taxonomy. Proc. Natl. Acad. Sci. USA 2007, 104, 4775–4776. [CrossRef]
- Chen, W.; Ma, X.; Shen, Y.; Mao, Y.; He, S. The fish diversity in the upper reaches of the Salween River, Nujiang River, revealed by DNA barcoding. *Sci. Rep.* 2015, 5, 17437. [CrossRef]
- Zhang, J.; Hanner, R. Molecular approach to the identification of fish in the South China Sea. *PLoS ONE* 2012, 7, e30621. [CrossRef] [PubMed]
- Knebelsberger, T.; Landi, M.; Neumann, H.; Kloppmann, M.; Sell, A.F.; Campbell, P.D.; Laakmann, S.; Raupach, M.J.; Carvalho, G.R.; Costa, F.O. A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. *Mol. Ecol. Resour.* 2014, 14, 1060–1071. [CrossRef] [PubMed]
- Liu, J.; Provan, J.; Gao, L.M.; Li, D.Z. Sampling strategy and potential utility of indels for DNA barcoding of closely related plant species: A case study in taxus. *Int. J. Mol. Sci.* 2012, *13*, 8740–8751. [CrossRef] [PubMed]
- Galan, M.; Pages, M.; Cosson, J.F. Next-generation sequencing for rodent barcoding: Species identification from fresh, degraded and environmental samples. *PLoS ONE* 2012, 7, e48374. [CrossRef] [PubMed]
- 42. Pratyush, D.D.; Tiwari, S.; Kumar, A.; Singh, S.K. A new approach to touch down method using betaine as co-solvent for increased specificity and intensity of GC rich gene amplification. *Gene* **2012**, *497*, 269–272. [CrossRef]
- Bragg, L.M.; Stone, G.; Butler, M.K.; Hugenholtz, P.; Tyson, G.W. Shining a light on dark sequencing: Characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 2013, 9, e1003031. [CrossRef]
- 44. Piry, S.; Guivier, E.; Realini, A.; Martin, J.F. |SE|S|AM|E| Barcode: NGS-oriented software for amplicon characterization– application to species and environmental barcoding. *Mol. Ecol. Resour.* **2012**, *12*, 1151–1157. [CrossRef]
- 45. Meglecz, E.; Piry, S.; Desmarais, E.; Galan, M.; Gilles, A.; Guivier, E.; Pech, N.; Martin, J.F. SESAME (SEquence Sorter & AMplicon Explorer): Genotyping based on high-throughput multiplex amplicon sequencing. *Bioinformatics* **2011**, *27*, 277–278.
- Minamoto, T.; Yamanaka, H.; Takahara, T.; Honjo, M.N.; Kawabata, Z.I. Surveillance of fish species composition using environmental DNA. *Limnology* 2012, 13, 193–197. [CrossRef]
- 47. Smith, M.A.; Bertrand, C.; Crosby, K.; Eveleigh, E.S.; Fernandez-Triana, J.; Fisher, B.L.; Gibbs, J.; Hajibabaei, M.; Hallwachs, W.; Hind, K.; et al. Wolbachia and DNA barcoding insects: Patterns, potential, and problems. *PLoS ONE* **2012**, *7*, e36514. [CrossRef]
- Viñas, L.; Besada, V.; Sericano, J.L. Sampling of fish, benthic species, and seabird eggs in pollution assessment A2—Pawliszyn, Janusz. In *Comprehensive Sampling and Sample Preparation*; Academic Press: Oxford, UK, 2012; pp. 349–372.
- April, J.; Mayden, R.L.; Hanner, R.H.; Bernatchez, L. Genetic calibration of species diversity among North America's freshwater fishes. Proc. Natl. Acad. Sci. USA 2011, 108, 10602–10607. [CrossRef] [PubMed]
- 50. Hubert, N.; Hanner, R.; Holm, E.; Mandrak, N.E.; Taylor, E.; Burridge, M.; Watkinson, D.; Dumont, P.; Curry, A.; Bentzen, P.; et al. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* **2008**, *3*, e2490. [CrossRef] [PubMed]
- 51. Mabragana, E.; Diaz de Astarloa, J.M.; Hanner, R.; Zhang, J.; Gonzalez Castro, M. DNA barcoding identifies Argentine fishes from marine and brackish waters. *PLoS ONE* **2011**, *6*, e28655. [CrossRef]
- 52. McCusker, M.R.; Denti, D.; Van Guelpen, L.; Kenchington, E.; Bentzen, P. Barcoding Atlantic Canada's commonly encountered marine fishes. *Mol. Ecol. Resour.* 2013, 13, 177–188. [CrossRef] [PubMed]
- Chen, J.; Li, Q.; Kong, L.; Yu, H. How DNA Barcodes Complement Taxonomy and Explore Species Diversity: The Case Study of a Poorly Understood Marine Fauna. *PLoS ONE* 2011, 6, e21326. [CrossRef] [PubMed]
- Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; deWaard, J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 2003, 270, 313–321. [CrossRef] [PubMed]
- 55. Hebert, P.D.; Ratnasingham, S.; deWaard, J.R. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci. R. Soc.* 2003, 270 (Suppl. S1), S96–S99. [CrossRef]
- 56. Hebert, P.D.N.; Stoeckle, M.Y.; Zemlak, T.S.; Francis, C.M. Identification of Birds through DNA Barcodes. *PLoS Biol.* 2004, 2, e312. [CrossRef]
- 57. Steinke, D.; Zemlak, T.S.; Hebert, P.D. Barcoding nemo: DNA-based identifications for the ornamental fish trade. *PLoS ONE* 2009, 4, e6300. [CrossRef]
- 58. Weitschek, E.; Van Velzen, R.; Felici, G.; Bertolazzi, P. BLOG 2.0: A software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. *Mol. Ecol. Resour.* **2013**, *13*, 1043–1046. [CrossRef]
- 59. Hebert, P.D.; Penton, E.H.; Burns, J.M.; Janzen, D.H.; Hallwachs, W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14812–14817. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.