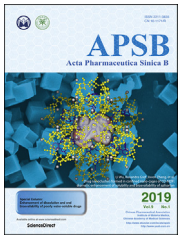
Acta Pharmaceutica Sinica B xxxx;xxx(xxx):xxx

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



ORIGINAL ARTICLE

A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong"

Zhimei Xing[a,b,†], Han Gao[a,c,†], Dan Wang[a,b,†], Ye Shang[a,b], Tenukeguli Tuliebieke[a,b], Jibao Jiang[d], Chunxiao Li[a,b], Hong Wang[a], Zhenguo Li[e], Lifu Jia[f], Yongsheng Wu[e], Dandan Wang[e], Wenzhi Yang[a,b], Yanxu Chang[a,b], Xiaoying Zhang[a,b], Liuwei Xu[a,b], Chao Jiang[g,*], Luqi Huang[g,*], Xiaoxuan Tian[a,b,*]

[a]*State Key Laboratory of Component-Based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China*
[b]*Haihe Laboratory of Modern Chinese Medicine, Tianjin 301617, China*
[c]*Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China*
[d]*School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China*
[e]*Mudanjiang YouBo Pharmaceutical Co. Ltd., Mudanjiang 157000, China*
[f]*Guizhou Ruihe Pharmaceutical Co. Ltd., Guizhou 550000, China*
[g]*State Key Laboratory of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100010, China*

Received 8 July 2022; received in revised form 28 September 2022; accepted 13 October 2022

KEY WORDS:
Traditional Chinese medicine;
"Guang Dilong";
Metabarcoding;

Abstract For wild natural medicine, unanticipated biodiversity as species or varieties with similar morphological characteristics and sympatric distribution may co-exist in a single batch of medical materials, which affects the efficacy and safety of clinical medication. DNA barcoding as an effective species identification tool is limited by its low sample throughput nature. In this study, combining DNA mini-barcode, DNA metabarcoding and species delimitation method, a novel biological sources consistency

*Corresponding authors.
E-mail addresses: jiangchao0411@126.com (Chao Jiang), huangluqi01@126.com (Luqi Huang), tian_xiaoxuan@tjutcm.edu.cn (Xiaoxuan Tian).
[†]These authors made equal contributions to this work.
Peer review under responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences

https://doi.org/10.1016/j.apsb.2022.10.024

Mini-barcode;
Subgroup

evaluation strategy was proposed, and high level of interspecific and intraspecific variations were observed and validated among 5376 *Amynthas* samples from 19 sampling points regarded as "Guang Dilong" and 25 batches of proprietary Chinese medicines. Besides *Amynthas aspergillum* as the authentic source, 8 other Molecular Operational Taxonomic Units (MOTUs) were elucidated. Significantly, even the subgroups within *A. aspergillum* revealed here differ significantly on chemical compositions and biological activity. Fortunately, this biodiversity could be controlled when the collection was limited to designated areas, as proved by 2796 "decoction pieces" samples. This batch biological identification method should be introduced as a novel concept regarding natural medicine quality control, and to offer guidelines for *in-situ* conservation and breeding bases construction of wild natural medicine.

1. Introduction

The evaluation of batch similarity/consistency is a fundamental issue in traditional Chinese medicine (TCM) quality control. Due to a lack of effective trait discovery and bioactivity assessment methods, biological sources evaluation has become an indispensable means in evaluating TCM within and between batches, particularly in animal-based traditional medicines. Specifically, approximately 80% of all Chinese herb sources are wild[1], and assessments of the inter-/intra-species level biodiversity of wild medicinal resources are constantly insufficient. Following the exploitation of wild resources, unanticipated biodiversity may exist in natural medical materials as species with similar morphological characteristics and sympatric distribution, resulting in high heterogeneity when processed to TCM and Chinese patent medicine. Establishing proper and efficient approach to evaluate the biological composition consistency of medicinal resources, at both of inter-/intra-species level, has become one of the key priorities in the internationalization and modernization of TCM.

Over the past decade, molecular methods such as DNA barcoding have provided an accurate and efficient strategy for identifying species and distinguishing the authentic ones from the adulterants in various types of traditional medicines[2−5]. Furthermore, DNA barcoding can also speed up the identification of distinct lineages, and accelerate the discovery of the cryptic/undescribed inter-/intra-species level biodiversity[6,7]. However, as traditional DNA barcoding identifies single specimens based on individual DNA sequencing reactions, it is often unsuitable for bulk commercial materials. The ongoing development of DNA metabarcoding which combines next-generation sequencing (NGS) with DNA barcoding has enabled the identification of a large number of samples with high throughput and low cost[8−15]. Aside from detecting species level diversity, recent studies have demonstrated its ability to distinguish intraspecific genetic variation from environmental samples[16−19]. Moreover, as cytochrome c oxidase I (COI) has a natural high intraspecies variability for metazoan[19], the mini-barcode (100−250 bp) designed within the COI region, combined with DNA metabarcoding, would simultaneously characterize the intra- and inter-species genetics of bulk animal-based medicinal materials[20], even in samples with degraded DNA[21−24].

The dried bodies of earthworms such as "Dilong" are widely used to treat inflammation, thrombus, asthma, and pyretic[25]. According to the *Chinese Pharmacopoeia*, species *Amynthas aspergillum* (Perrier, 1872) (synonym *Pheretima aspergillum*), *Metaphire vulgaris* (Chen, 1930) (synonym *P. vulgaris*), *M. guillelmi* (Michaelsen, 1895) (synonym *P. guillelmi*) and *A. pectieniferus* (Michaelsen, 1931) (synonym *P. pectinifera*) were official recorded as the original medicinal materials of "Dilong". The first one is known as "Guang Dilong", and mainly distributed in Guangdong province, Hainan province and Guangxi Zhuang Autonomous Region (Guangxi) in China. According to the authors' investigation, residents tend to collect this medical species in the wild at harvest time more than they breed them due to their relatively low price compared to human cost. In addition to *A. aspergillum*, there are a large number of sympatrically distributed Megascolecidae species that resemble each other in morphological characters[26], where new species are continually discovered[6,27,28]. Moreover, the rapid growth in price in recent years led to the emergence of intended adulterations and substitutions in medicinal markets[29−31], making the actual biological composition of "Guang Dilong" more confusing. The biological composition stability of medical materials should benefit from the fixation of producing areas. Nevertheless, while the alpha taxonomic classification of Megascolecidae in south China is still ongoing, information on the sympatric distribution of multiple cryptic species[32] and relative taxa abundance is still too limited to inform the selection of suitable collection or breeding sites. On the other hand, although multiple methods were developed for the authentication of *Amynthas* species[3,33−36], so far there is no reasonable biological composition assessment approach suitable for batch consistency evaluation of "Guang Dilong".

In this study, taking "Guang Dilong" as an example, a novel strategy based on DNA metabarcoding combined with mini-barcode was proposed for batch-to-batch biological sources consistency evaluation of TCM. High levels of interspecific and intraspecific genetic diversity were discovered in medical materials from various production regions and Chinese patent medicines. Notably, even the two subgroups of *A. aspergillum* that were revealed here differed significantly in chemical compositions and biological activities. Fortunately, the subsequent analysis of locally acquired "decoction pieces" samples confirmed the geographic areas with constant *Amynthas* species composition. The inter-/intra-biodiversity display in our approach could be taken as an innovative indicator of batch consistency evaluation, and our strategy should be conducive to the fixation of the producing area and the establishment of entire industrial chain origin quality consistency control and quality traceability systems.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

2. Materials and methods

2.1. Local cytochrome c oxidase I dataset preparation and evaluation of molecular markers

Previous reports have validated 658 bp Folmer region of COI gene amplified by LCO1490/HCO2198 as the "standard" marker for the DNA barcoding of earthworms[3,37]. To set up a local reference database, 1575 *Amynthas* and *Metaphire* COI sequences were first collected from BARCODE OF LIFE DATA (BOLD; http://boldsystems.org) and GenBank. After trimming with primers LCO1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′)[38], the sequences without species level taxonomic information, containing transcription terminators, or below 500 bp in length, were removed. Finally, 689 sequences from 99 species were preserved in our local dataset.

The LCO1490 and HCO1777 (5′-ACTTATATTGTTTATACGAGGGAA-3′) primer pair, which has been successfully used to identify degraded earthworm components in reptile feces samples, was selected as the marker, owing to the relative small size of the PCR product (281 bp)[39]. To further assess the universality of primer HCO1777 for *Amynthas* and *Metaphire* species, the nucleotide diversity (Pi) value of each locus in our database was calculated and screened using DnaSP to determine whether the primer binding sites among different species were conserved. To evaluate the taxonomic resolution of the LCO1490/HCO1777 marker, the amplicon regions were extracted in silico from our reference dataset, and after selecting the optimal fitting model determined by the ModelFinder[40] implemented in IQ-TREE v1.6.12[41], the ML phylogenetic tree was constructed using IQ-TREE with 1000 bootstrap replicates.

2.2. Sample collection and pretreatment

To investigate the biodiversity within "Guang Dilong", 5075 earthworm (Table 1) samples regarded as "Guang Dilong" were collected in the wild from 18 locations of Guangxi from September 2017 to June 2019, and made into dried decoction pieces by local farmers under the authors' supervision. Although the morphological characteristics of earthworms were lost in processed samples, this acquisition method ensured convenient sample collection, transportation, and storage, and maximized the numbers of samples. We assumed that this operable method could be adopted by other rapid biodiversity assessment practice on herbal medicine. In addition, 301 processed samples were bought from Baise Market of Traditional Chinese Medicine, Guangxi. To test the results of our species distribution pattern, three batches of 2796 dried earthworms (Table 1) collected from assigned sites were obtained from Youbo Pharmaceutical Co., Ltd. in April 2018. Finally, to uncover the cryptic biodiversity within the TCM product, three kinds of Chinese patent medicine containing "Dilong" (earthworm) were incorporated into the experiment in June 2019 as GX (batch number: 17083634), NX1-NX23 (batch numbers 130524, 141127, 1501116, 141016, 150534, 140817, 140968, 141042, 1411145, 150211, 150533, 1407137, 150535, 150543, 150544, 150545, 150546, 150547, 150848, 150556, 150557, 150558 and 1501122), and SH (batch number 17013914).

As metabarcoding allows for the pooling of hundreds of samples and their sequencing in a single sequencing run, for each collection site or product batch, 0.02 g of muscle was cut off individually from the abdomen and packed into a single tube. One point five percent (*w*/*v*) sodium dodecyl sulfate (SDS) lysate was used to pre-process samples after powdering. The mixture was incubated at 65 °C for 20 min and centrifuged at 3000 revolutions per minute (rpm) for 5 min. Approximately 200 μL of supernatant was stored for further DNA extraction.

2.3. DNA extraction and amplification

Total genomic DNA from each group sample was extracted using TIANamp Genomic DNA Kit (Tiangen Biotech Co., Ltd., China). The quality and concentration of the extracted DNA were determined using NanoDrop 2000. A 281 bp long segment of the 5′ terminus of the COI gene was amplified using primers LCO1490 and HCO1777. The amplification products were then pooled to construct the final libraries. To distinguish amplicons originating from different samples, tag oligonucleotides were ligated to each side of the primers, as shown in Supporting Information Table S1. As "tag jumps" between different indexed samples during library preparation and sequencing could lead to false assignment of sequences and artificially inflate diversity[42], each PCR amplification was carried out with matching tags to minimize tag jumps. PCR reactions were carried out at a final volume of 25 μL containing 12.5 μL 2 × Gflex buffer (containing 1 mmol/L of Mg^{2+} and 200 μmol/L of each dNTP), 0.5 μL Tks Gflex DNA Polymerase (1.25 units/μL) (Takara Dalian, China), 1 μL DNA template, 0.5 μL forward, and 0.5 μL reverse primers. The reaction condition that followed was predenaturation at 94 °C for 1 min; 40 cycles of denaturation at 98 °C for 10 s, annealing at 48 °C for 15 s, extension at 68 °C for 30 s; and final extension at 68 °C for 5 min. Extraction blanks and PCR blanks were included for each batch of DNA extraction to ensure no carryover contamination occurred. The PCR products were isolated on 1.5% agarose gel electrophoresis and purified using the Agarose Gel DNA Purification Kit Version 2.0 (Takara Dalian, China).

2.4. Amplicon sequencing data processing

The amplification products were sequenced using 2 × 250 bp paired-end protocol on the Illumina Hiseq 2500 platform. Pair-end reads were demultiplexed using fastq-multx and assigned to each sample according to the unique tags. Then the primer sequences were trimmed using bbduk[43]. The amplicon sequence variants (ASVs) were identified after quality control, joining of paired ends, removal of chimeras, removal of non-target-length sequences, and denoising using the DADA2 method[44]. In contrast with traditional Molecular Operational Taxonomic Units (MOTUs), which cluster multiple sequences to represent taxons at species level, ASVs generally stand for unique haplotypes. Rarefaction curves were inferred to define the rarefaction value for normalizing the ASVs table using the Vegan[45] R Package. Subsequently, the alpha-diversity indexes (Shannon) and beta-diversity were measured using R package phyloseq[46]. To analyze the phylogenetic relationship between ASV sequences, Bayesian inference (BI) and maximum-likelihood (ML) methods were performed using MrBayes and IQ-TREE and based on GTR+G and GTR+F+I+G4 model, respectively.

2.5. Taxonomy assignment of amplicon sequence variants and molecular operational taxonomic units

Two species delimitation methods were utilized to identify the MOTUs from the representative ASV sequences. The automatic

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

Table 1 Sample information.

Sample name	Sampling place/source	Coordinate	Sampling time	Number of individuals
BB	Bobai county	22°06'N, 109°87'E	2017.11	213
BL	Dayantang village, Beiliu city	22°43′27″N, 110°18′20″E	2018.11	428
BS	Bought from Baise drug market	23°54′18.21″N, 106°36′53.87″E	2018.11	301
CL	Changle town	21°49′21.67″N, 109°24′26.52″E	2017.11	340
CT	Caotang village, Changle town	21°88′N, 109°39′46″E	2017.11	398
DJ	Dujiao village, Changle town	21°49′21.67″N,1 09°24′26.52″E	2018.11	230
DQ	Daqiao town, Luchuan county	22°15′1.67″N, 110°12′58.36″E	2018.11	260
DX	Dongxing town, Dongxing county	21°32′20.25″N, 107°58′6.08″E	2018.11	260
DY	Dayantang village, Beiliu city	22°43′27″N, 110°18′20″E	2018.11	210
HX	Huoxing village, Changle town	21°87′16″N, 109°44′52″E	2017.11	307
LC	Luchuan county	22°19′26″N, 110°15′35″E	2017.11	410
MP	Mapo town, Luchuan county	22°29′36.74″N, 110°13′3.80″E	2018.11	240
PS	Poshan village, Changle town	21°49′21.67″N, 109°24′26.52″E	2018.11	220
SC	Songshuyuan village, Shikang town	21°80′04″N, 109°34′51″E	2017.11	307
ST	Santang town, Bobai county	22°11′2.83″N, 110°01′37.03″E	2018.11	170
TD	Tiandong county	23°35′58.99″N, 107°07′19.10″E	2017.11	260
WL	Pubei county	22°27"N, 109°55'E	2018.11	240
XWC	Xuwu village, Changle town	21°76′7″N, 109°42′58″E	2017.11	302
ZH	Zhanghuang town, Pubei county	22°0′37.79″N, 109°27′20.93″E	2018.11	280
dlA[a]	Hepu county, batch number: 170506	N/A	2018.04	1001
dlB[a]	Beiliu county, batch number: 170510	N/A	2018.04	902
dlC[a]	Bobai county, batch number: 170509	N/A	2018.04	893

[a]Samples purchased from assigned sites and provided by pharmaceutical company.

barcoding gap discovery (ABGD)[47] method was used to identify the genetic distance at which a "barcoding gap" occurs and then sort the ASVs into putative species, with the default settings being P_{min} = 0.001, P_{max} = 0.1, Steps = 10, X (relative gap width) = 1.5, and Nb bins = 20; and with K2P distances. The Bayesian version of Poisson Tree Processes (bPTP)[48] analysis was also run for delimiting species on a rooted MrBayes consensus tree. The Bayesian tree was uploaded to https://species.h-its.org/ptp/ and the number of MCMC generations was set as 500,000.

The taxonomic information of each ASV/MOTU was determined using two methods based on sequence similarity and phylogenetic placement, separately. The first strategy was BLAST-MEGAN. BLAST search against the local database was conducted as -outfmt 5, -evalue 10, -max_target_seqs 5, -gapopen 5, -gapextend 2, and -num_threads 16. Subsequently, the BLAST results were visualized using MEGAN community edition version 6.10.8[49] and parsed to assign the optimal hits to appropriate taxa in the NCBI taxonomy. MEGAN parameters were set as: minimum score = 50, maximum expected = 0.01, top percent = 10, minimum support percent = 0.01, minimum support = 1, and weighted LCA algorithm. The minimum identity to query would be set as 97% for species level identification, or 93% to obtain taxonomic information at a relatively high level for queries that could not be identified as exact species. The phylogenetic-based taxonomic assignment was performed as our second strategy. The ASVs queries were aligned with our local COI reference dataset and placed on the COI reference ML tree using pplacer followed by taxonomic classification with the pplacer guppy tool[50].

For that MOTUs failed to be identified at the species level, our cooperators in Shanghai Jiao Tong University were entrusted with the use of a professional database for identification. To verify the authenticity of MOTUs, whose Linnaean species name cannot be specified finally, 7 earthworms were randomly selected from the dried products purchased from the Baise Market of Traditional Chinese Medicine. After DNA extraction, PCR amplification and Sanger sequencing, the 658 bp Folmer regions of COI gene were obtained. Two individuals with 100% similarity to Megascolecidae sp2 were screened out. Whole-genome shotgun sequencing of 2 samples was carried out on the Illumina Hiseq X Ten platform (PE150; insert size, 350 bp).

Quality assessment of raw reads was performed using Trimmomatic version 0.36 by removing low-quality and adapter-contaminated reads. Subsequently, remaining high-quality reads were assembled into contigs using NOVOPlasty version 3.1[51]. The parameters were set as: insert size = 350, read length = 150, type = mito, gene range = 10,000−20,000, k-mer = 39. The preliminary mitochondrial genome annotations were conducted in MITOS web server[52], under default settings and the invertebrate genetic code for mitochondria, followed by manual corrections of the start and stop codons in Geneious based on the previously published mitochondrial genome of *Metaphire californica*[53].

To verify the phylogenetic status of the presumed species, protein-coding sequence (CDS) on mitochondrial genome of 2 Megascolecidae sp2 individuals and 18 other pheretimoid species were extracted and concatenated by PhyloSuite with default parameters, then the phylogenetic tree was constructed using ML algorithms with TIM2+F+R8 model in IQ-TREE v1.6.12[41] and visualized using iTOL (https://itol.embl.de/).

2.6. *Analysis of intraspecies subgroups within "Guang Dilong" samples that identified as A. aspergillum*

When inspecting the intraspecies biodiversity within *A. aspergillum*, two subgroups were discovered and the "barcoding gap" between them based on genetic distance was depicted by MEGA (10.1.8)[54] and R package ggplot2 (v3.3.5)[55]. Subsequently, a haplotype network was constructed to illustrate the relationships among haplotypes using Population Analysis of Reticulated Trees (PopART) version 1.7.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

To verify the authenticity of these two subgroups, one individual from each subgroup was selected for whole-genome sequencing. To compare the overall similarities of mitochondrial genomes, pairwise alignments were performed in the mVISTA program[56] (http://genome.lbl.gov/vista/mvista/submit.shtml), under LAGAN mode using the annotation of *A. aspergillum* (GenBank accession number: NC_025292, BINs ID: ACH7381) as reference.

To determine whether two subgroups of *A. aspergillum* have the same chemical composition and pharmacodynamic action, 19 samples (11 of subgroup 1 and 8 of subgroup 2) were selected randomly from our samples.

The UHPLC−Q-TOF/MS system used to describe the chemical profile was composed of Agilent 1290 UHPLC instrument (Agilent Technologies, Waldbronn, Germany) and Agilent 6546 Q-TOF mass spectrometer (Agilent Corporation, Santa Clara, CA, USA). The MS data were acquired in positive dual Agilent Jet Steam Electron Spray Ionization (AJS ESI) mode. Chromatographic separation was performed on a Waters ACQUITY BEH Amide column (2.1 mm × 100 mm, 1.7 μm) at 35 °C, and at a flow rate of 0.4 mL/min. The mobile phase comprised (A) 0.15% aqueous formic acid (contained 5 mmol/L ammonium formate and 5 mmol/L ammonium acetate) and (B) 0.15% acetonitrile formic acid (contained 1 mmol/L ammonium formate and 1 mmol/L ammonium acetate) using gradient elutions of 10% A at 0−3 min, 13%−25% A at 10−15 min, 50% A at 21−23 min, and 10% A at 24 min; the re-equilibration time of gradient elution was 4 min. The related Q-TOF/MS parameters were as follows: gas temp 350 °C, drying gas 10 L/min, nebulizer 50 psi, sheath gas temp 350 °C, sheath gas flow 10 L/min, vcap 4500 V, skimmer 65 V, and fragmentor 175 V. The acquisition rate was 1 spectra/s and the mass range was set as 100−1700 *m/z*. The collision energy was set as 10 V to obtain MS/MS information. Nineteen sample solutions were mixed at a certain volume to prepare for the QC sample for precision evaluation. To investigate the natural grouping situation and obtain the difference variable, the mass and retention time lists were used for principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA) using SIMCA-P software (Version 14.1).

Considering the medicinal properties of "Guang Dilong", the anti-coagulation activity of two subgroups was assessed using thrombin-fibrinogen assay. Furthermore, 0.2 g of the sample was weighed and extracted with 1 mL water in an ultrasonic water bath at 40 °C for 40 min. Afterwards, the solution was centrifuged at 14,000 rpm for 5 min. The supernatant was filtered using a 0.22 μm nylon membrane. Aliquot (100 μL) of the filtrate solution was transferred to a 2 mL centrifuge tube that contained 200 μL fibrinogen (5 mg/mL, PBS). Then, the mixture was incubated for 5 min in a 37 °C water bath, after which 2 μL thrombin (10 U/mL, normal saline) was added to the mixture every 4 min until a fibrous precipitate was observed. For the blanks (negative controls), the earthworm solution was replaced with distilled water and a procedure similar to the one above was carried out. The consumed thrombin solution was used to calculate the anti-coagulation activity of earthworms based on the following Eq. (1):

$$U = \frac{C_1 V_1}{C_2 V_2} \quad (1)$$

where U represented anti-coagulation potency per gram; C_1 represented the concentration of thrombin solution; V_1 represented the consumption volume of thrombin solution; C_2 represented the concentration of earthworm solution; and V_2 represented the volume of earthworm solution. If the consumption volume of thrombin was less than that of the negative control, the earthworm sample was regarded as lacking anti-coagulation activity. To verify whether the difference in efficacy between the two subgroups is stable, in addition to above samples, 12 subgroup 1 and 8 subgroup 2 *A. aspergillum* individuals were subsequently included.

2.7. Biodiversity analysis of earthworm samples

Prior to the biodiversity analyses at haplotype level, ASV total abundances in each sample were rarefied (resampled with replacement) to the minimum number of reads found (58,905 read counts). The relative read abundance (*i.e.*, the value of each taxon divided by the total reads per sample) of MOTUs were calculated and plotted using the R package ampvis2. Samples from the same collection site were grouped in the heat-map. The alpha- and beta-diversity of samples based on ASVs and MOTUs were analyzed. The Shannon index as a measure of entropy, indicated the richness and evenness of the taxa present. With respect to beta diversity, 18 batches of samples collected in the wild were included, and Principal co-ordinates analysis (PCoA) was performed using Bray-Curtis distance analysis.

To observe the influence of geographic distribution on the species level (MOTUs) biodiversity, the capture locations (geographical coordinates) were entered into the GenGIS[57] (v2.5.1) software to generate the distribution map.

For 3 batches of the samples of earthworm decoction pieces supplied by pharmaceutical companies and 25 batches of Chinese patent medicine, the relative abundance of taxa was analyzed as previously mentioned.

3. Results

3.1. Evaluation of LCO1490/HCO1777 marker for Amynthas and Metaphire species discrimination

Compared to the standard COI 5' region, the amplicon of LCO1490/HCO1777 with a length of 281 bp was more suitable for degraded samples, such as processed natural medicine products. While the universality of LCO1490 for metazoan had been validated by numerous studies[58−60], the conservation of the HCO1777 primer binding region was further assessed in our *Amynthas* and *Metaphire* local dataset (Supporting Information Appendix A). As shown in Fig. 1, HCO1777 was located in a region with low variability compared with its neighboring areas. Furthermore, the LCO1490/HCO1777 amplicon sequences of *A. aspergillum* clustered into one branch stably on the ML phylogenetic tree, and the species level discrimination ability of this short marker was acceptable for *Amynthas* and *Metaphire* species (Supporting Information Fig. S1). At the same time, after sanger sequencing of 15 individuals, it was found that the identification results of the two markers were consistent (Supporting Information Table S2). The 658 bp "classic" COI regions and the 232 bp regions of these samples were presented as Supporting Information Appendix B.

3.2. Sequencing data filtering and clustering for "Guang Dilong" samples collected in the wild

A total of 34,107,981 raw pair-end reads were generated, and 31,065,432 clean reads were obtained after trimming the adapter

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

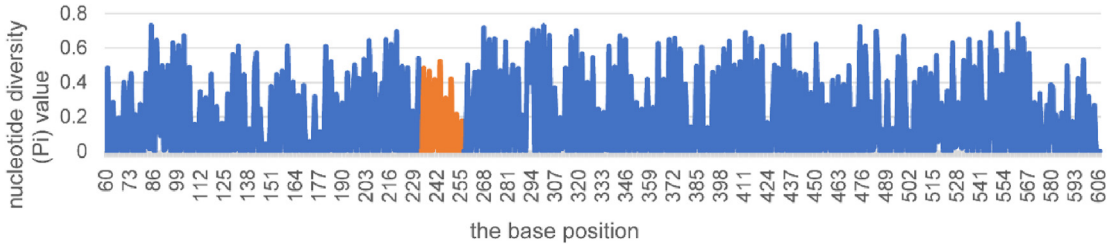


Figure 1 Nucleotide diversity (Pi) value of HCO1777 primer region across local COI barcode dataset comprising 689 sequences from 99 *Amynthas* and *Metaphire* species. The value on the horizontal axis represented the base position starting next to the 3′ end of LCO1490, and the orange section represented the location of primer HCO1777.

and removing low-quality sequences. The number of clean reads for each sample was presented in Supporting Information Table S3. First, 512 ASVs were detected using the DADA2 algorithm. After filtering out ASVs whose read count coverage was <100 and checking for stop codons, a total of 358 ASVs associated with particular COI haplotypes remained, while 302 of them were identified as earthworms (Supporting Information Table S4).

3.3. *Species level taxon identification for "Guang Dilong" samples*

Rarefaction analysis (ASVs counts based) showed that the sequencing effort covered the haplotypes present in the samples, as curves approached saturation (Supporting Information Fig. S2). To infer the taxonomic assignment of 302 "Guang Dilong" ASVs, both similarity-based and phylogeny-based identification procedures were performed. Not surprisingly, 233 ASVs were identified as *A. aspergillum* using both methods, *i.e.*, the certified "Guang Dilong" described in *China Pharmacopoeia*. Using the BLAST-MEGAN approach, only ASV_180 and ASV_189 were identified as *Amynthas* sp. and ASV_131 and ASV_170 as Megascolecidae sp. Among the rest, and the taxonomic status of other 65 ASVs could not be assigned. Furthermore, the phylogenetic placement strategy additionally identified 38 ASVs as *A. aspergillum* and provided additional taxonomic information for the 31 ASVs remaining (18 ASVs as *Amynthas* sp. and 13 as Megascolecidae sp.), as shown by Supporting Information Table S5.

For all haplotypes with exact or ambiguous taxon names revealed from "Guang Dilong" samples, species delimitation methods were applied to identify species level biodiversity. The ABGD method clustered 31 ambiguous ASVs into 7 molecular operational taxonomic units (MOTUs), and 271 *A. aspergillum* ASVs into the final one (Supporting Information Table S6). Meanwhile, the bPTP analysis of the COI gene data based on the ML tree identified 9 MOTUs as shown in Fig. 2. Conversely, given the ABGD result, the ASV_160 grouped with 6 other ASVs by ABGD was further classified as an independent MOTU (MOTU_7) by bPTP.

With the aid of collaborators in Shanghai Jiaotong University, typical ASVs of each MOTU were selected and used to search against a professional COI barcoding dataset constructed by taxonomists on pheretimoid earthworms. Finally, taxonomic identification results of 9 MOTUs containing 302 ASVs were presented in Table 2. In addition to MOTU_1 as *A. aspergillum*, MOTU_6, MOTU_7, MOTU_8 and MOTU_9 were identified as *Metaphire prominula* (Qiu & Jiang), *Metaphire* sp., *Amynthas dentiformis* (Sun & Jiang), and *Amynthas* sp., respectively. Notably, ASV112 and ASV119 within MOTU_5 matched with the COI region of a specimen called HN201702-01 and collected from Longtang Town, Haikou City, Hainan Province, China. Their respective identity percentages were 100% and 99.6%. According to the taxonomic description, this specimen should belong to the *Amynthas corticis*-group and be listed as a new species temporarily named as *Amynthas* A99, owing to its unique morphological

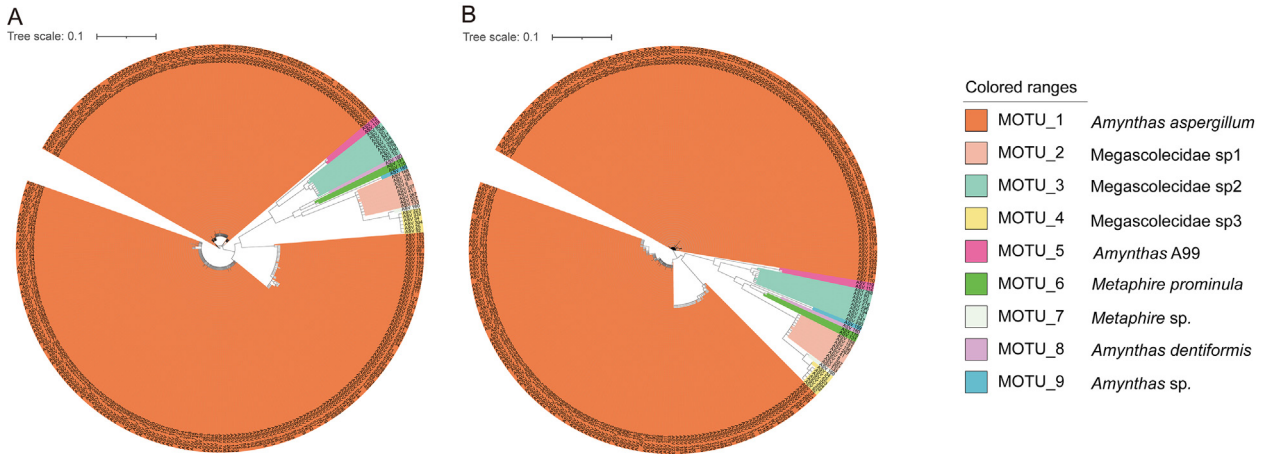


Figure 2 Species delimitation of "Guang Dilong" COI ASVs based on the bPTP method. (A) Bayesian tree; (B) ML tree. The light blue circular ball in the figure represents the node support rate of each branch, and the larger the sphere, the greater the node support rate. The ABGD result was similar to that of bPTP, except that MOTU_4 and MOTU_7 were clustered into a single MOTU by ABGD.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

Table 2 Taxonomic identification information of nine MOTUs.

MOTU ID	Contained ASVs ID (No. of ASVs)	BLAST against public dataset	Phylogenetic replacement using public dataset (support percent)	BLAST against professional dataset constructed by taxonomists on pheretimoid earthworms (ASVs as queries)	Final identification result
MOTU_1	ASV_1, ASV_2, ASV_3, ASV_4 …, ASV_360, ASV_361, ASV_363, ASV_365, ASV_366, ASV_368 (271)	233 ASVs identified as *Amynthas aspergillum,* while the other 38 ASVs could just be identified as *Amynthas* sp.	*A. aspergillum* (100)	*A. aspergillum* (ASV_5 ASV_6 ASV_13 ASV_21 ASV_79 ASV_90)	*A. aspergillum*
MOTU_2	ASV_8, ASV_41, ASV_133, ASV_156, ASV_168, ASV_181, ASV_207, ASV_267, ASV_327 (9)	Not assigned	*Amynthas* sp. (71.06−77.95)	Not assigned (ASV_8, ASV_41)	Megascolecidae sp1
MOTU_3	ASV_10, ASV_20, ASV_80, ASV_194, ASV_268, ASV_276, ASV_280, ASV_296, ASV_347 (9)	Not assigned	Megascolecidae sp. (91.22−99.99)	Not assigned (ASV_10, ASV_20, ASV_80)	Megascolecidae sp2
MOTU_4	ASV_25, ASV_34, ASV_92, ASV_304, ASV_323, ASV_332 (6)	Not assigned	*Amynthas* sp. (67.38−74.67)	Not assigned (ASV_25, ASV_34, ASV_92)	Megascolecidae sp3
MOTU_5	ASV_112, ASV_119 (2)	Not assigned	*Amynthas* sp. (63.22, 65.51)	*Amynthas* A99 (ASV112, ASV119)	*Amynthas* A99
MOTU_6	ASV_131, ASV_170 (2)	Megascolecidae sp.	Megascolecidae sp. (84.09, 75.61)	*Metaphire prominula* Qiu & Jiang (ASV_131, ASV_170)	*M. prominula* Qiu & Jiang
MOTU_7	ASV_160 (1)	Not assigned	Megascolecidae sp. (58.57)	*Metaphire* sp. (ASV_160)	*Metaphire* sp.
MOTU_8	ASV_180 (1)	*Amynthas* sp.	Megascolecidae sp. (72.79)	*Amynthas dentiformis* Sun & Jiang (ASV_180)	*A. dentiformis* Sun & Jiang
MOTU_9	ASV_189 (1)	*Amynthas* sp.	*Amynthas* sp. (99.99)	*Amynthas* sp. (ASV_189)	*Amynthas* sp.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

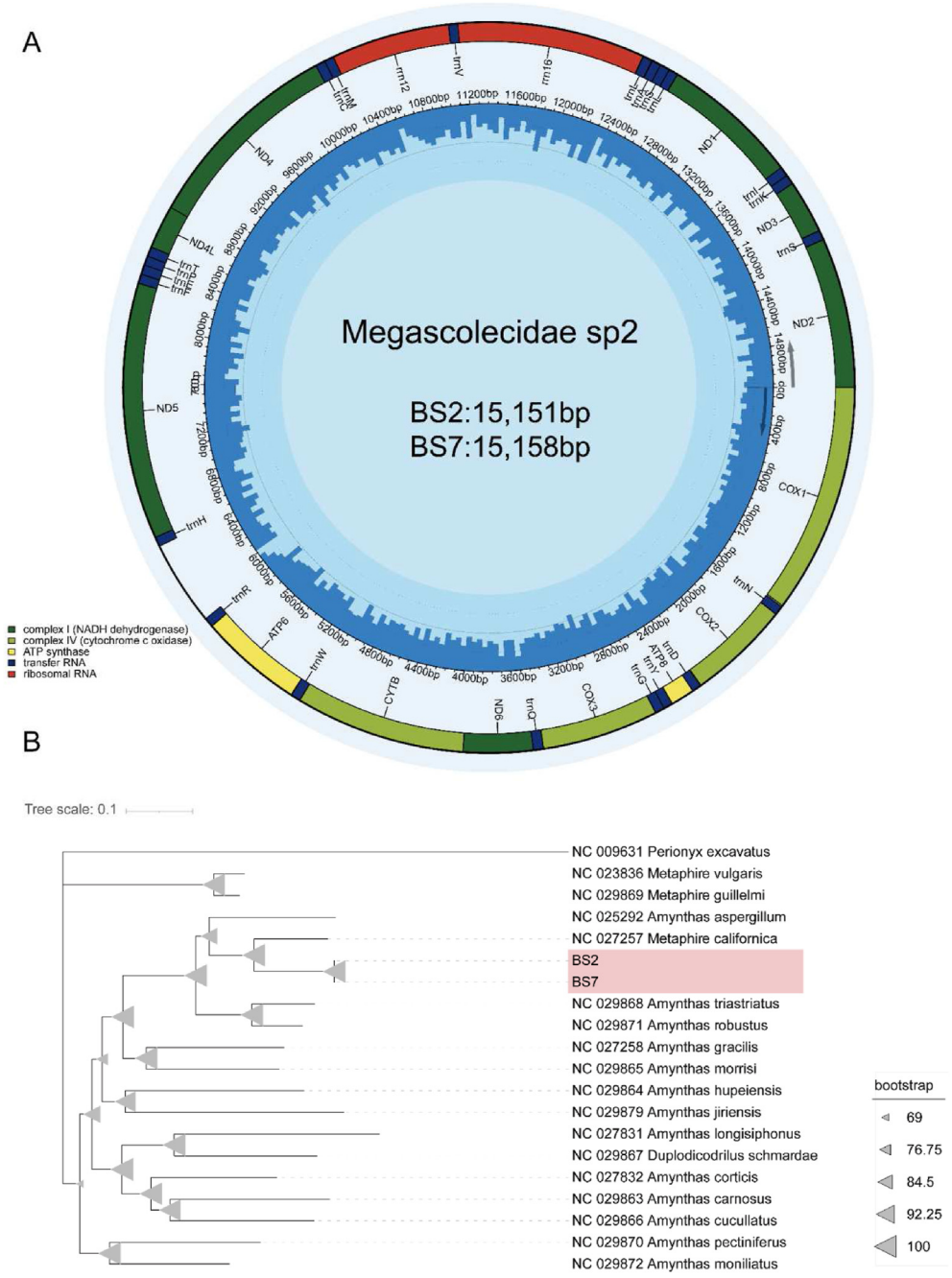


Figure 3 Mitochondrial genomes of two individuals within Megascolecidae sp2. (A) Physical map of the mitochondrial genome. The gray part in the inner ring indicates the percentage of GC content; (B) Phylogenetic analysis of Megascolecidae sp2 and 18 other pheretimoid species based on the mitochondrial genome.

characters of four pairs of spermathecal pores in 5/6−8/9, special male pore zone, cecum, and spermatheca caecus. The morphological description is presented in Supporting Information Fig. S3. Formal description of the new species will be published in another paper.

For the MOTUs that could not be assigned to the professional dataset, further validation on genome level was conducted. For MOTU_3 (Megascolecidae sp2), two corresponding individuals were determined, and the entire mitochondrial genome was constructed. The mitochondrial genomes of two individuals were 15,151 and 15,158 bp in length, respectively. These two genomes had a high degree of similarity (99.9%), and the same genetic composition and structure. Each of them encoded 37 unique genes, including 13 coding sequences, 22 tRNA, and 2 rRNA genes (Fig. 3A, Supporting Information Table S7). The "classic" 658bp-COI region was extracted from mitogenome and then searched against BOLD and GenBank nt databases. Both the ML phylogenetic analysis (Fig. 3B) and the BLAST identification results confirmed that, the nucleotide data of this putative species, at least, had never been recorded in the public database.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

3.4. Intraspecies subgroups hidden in "Guang Dilong" samples that identified as A. aspergillum

As shown in Fig. 4A, as the haplotype network, two major groups were uncovered within *A. aspergillum* ASVs, comprising 38 ASVs, which was not classified as *A. aspergillum* by BLAST-based method, and the 233 other ASVs. Moreover, the ambiguous "barcoding gap" (c, minimum interspecific distance exceeded the maximum intraspecific distance) was observed in Fig. 4B. Both the maximum intra-group and minimum inter-group genetic distance (Kimura 2-parameter model) were 5.40%, and the average intra- and inter-group genetic distances were 2.04% and 7.46%, respectively. Considering the sympatric distribution pattern and the species delimitation results described above, these two groups were regarded as subgroups within species *A. aspergillum*, rather than two distinct species, although the genetic distance between them was neither in the intraspecies nor interspecies range for earthworm species[61,62]. It is worth noting that, multiple genetic distance distribution peaks were observed in the "barcoding gap" plot, and three ASV clusters within subgroup 1 in the haplotype network directly corresponded with BOLD BINs ACH7381, ACB6697, and AAK1097 within *A. aspergillum*, whereas the subgroup 2 was consistent with BOLD BINs ACB6847.

Furthermore, 11 subgroup 1 samples (5 from BINs ACH 7381 and 6 from BINs ACB 6697) and 8 subgroup 2 samples were selected to assess the influence of these intraspecific diversities within *A. aspergillum* on the quality of "Guang Dilong". The 658 bp "classic" COI regions of these samples were presented as Supporting Information Appendix C. The maximum intra-group and minimum inter-group genetic distances (Kimura 2-parameter model) were 2.65% and 9.50%, respectively, and the average intra- and inter-group genetic distances were 1.32% and 1.02%, respectively. The comparison of mitochondrial genomes in two *A. aspergillum* subgroups was revealed using mVISTA, and we observed approximately identical gene order and organization among them (Fig. 4C). Notably, not only the COI gene, but also the other genes of mitochondrial genome of subgroup2 (BINs ID ACB6847) differed from the reference sequence (BINs ID ACH7381, subgroup1) much than subgroup1 (BINs ID ACB6697), especially for the noncoding regions.

However, we failed to identify sufficiently distinguishable morphological characters between the two subgroups with distinct genetic variation (Supporting Information Fig. S4).

Subsequently, the untargeted metabolomics strategy based on UHPLC−Q-TOF/MS was applied to compare chemical components, as shown in Fig. 4D. The QC samples as the data normalization tool were mainly distributed at the origin of the coordinates of the PCA plot, thereby indicating acceptable analytical precision. Significantly, samples of subgroup 1 were distributed in the second and third quadrants, and subgroup 2 samples mainly existed in the first and fourth quadrants. The OPLS-DA plot (Supporting Information Fig. S5A) displayed the obvious separation between two types of samples. To validate the accuracy of OPLS-DA, a 200-step permutation experiment was performed. As shown in Fig. S5B, the rightmost points of R^2 and Q^2 both exceeded those of the other points. Cumulative $R^2(Y)$ and $Q^2(Y)$ values ($R^2 = 0.98$, $Q^2 = 0.75$) close to 1 indicated a reliable model. Considering variable importance in projection (VIP) > 1 as the criteria, 161 variables (chemical compounds) (Supporting Information Table S8) were screened out, and within them, 18 compounds were unambiguously identified. Then, all variables were qualitatively analyzed using Agilent MassHunter Qualitative Analysis (version B.07.00). As illustrated in Fig. 4E, stachydrine, Lyso-PAF C-16, nicotinic acid, adenosine, and 3′-AMP/AMP had relatively high content in subgroup 2, while the 13 other ingredients such as L-pipecolic acid, Arg−Ala, Val−Arg, phenylalanine, Leu−Arg/Ile−Arg, alanine betaine, Ile−Val/Leu−Val, Pro−Arg, His−Gln, His−Val, His−Ala, Leu−Pro/Ile−Pro, and *O*-acetyl-L-serine exhibited relatively high content in subgroup 1. Although most of these compounds were regarded as primary metabolites of earthworms, certain ingredients such as Stachydrine[63], Lyso-PAF C-16[64], and nicotinic acid[65] have been reported to display bioactivities as anti-inflammatory, cardioprotective, or lipid metabolism regulation. Among 18 previously identified differential compounds, lyso-platelet-activating factor (lyso-PAF), the platelet-activating factor (PAF) precursor[66], could promote platelet aggregation by induction of free Ca^{2+} increase and coagulation factor-III secretion[67]. It is worth noting that, as the non-targeted metabolomics emphasized on the comprehensive analysis of detectable small molecules[68], the anticoagulant ingredients such as potential bioactive peptides may not be fully represented. The exact active substances of this kind of earthworm responsible for anticoagulant effects would require further research to be elucidated. With respect to anti-coagulation activity, the non-parmetric Mann-Whitney U-Test ($P = 0.0001$) showed significant bioactivity difference between two subgroups, and the average potency of subgroup 1 significantly exceeded that of subgroup 2 (Fig. 4F).

3.5. Overall biodiversity of the earthworm collected in the wild

As illustrated in Fig. 5A and B, except sample BS used as control and purchased at the local market, among the other 18 samples collected in the wild, the relative read abundance of *A. aspergillum* accounted for more than 80%. Fig. 5C showing the PCoA plot using Bray-Curtis distance and Supporting Information Fig. S6 showing the geographic distribution map present three samples: SC, CT, and XWC that were geographically adjacent in Hepu County and had similar earthworm species composition. While DQ in Luchuan had a relative severe unintentionally substitution by Megascolecidae sp1, the proportions of *A. aspergillum* in the other 14 sampling sites exceeded 98% and were grouped together in the beta-diversity analysis. Generally, our results indicated that the locations we selected, as part of the traditional producing areas for "Guang Dilong", had a relatively stable species composition.

Furthermore, considering the impact of genetic variation on pharmaceutical effect, the intraspecies level biodiversity was observed in these samples. Within the 18 samples, only 5 samples were occupied by single the *A. aspergillum* subgroup (subgroup 1), and 9 samples whose >10% of read abundance levels were occupied by the minor subgroup type. Samples that were subgroup 1-predominant, subgroup 2-predominant, or with middle type were clustered into different groups in Fig. 5D.

As illustrated in Fig. 5B, the distribution of the Shannon entropy index based on the MOTUs, subgroups, or ASVs were irrelevant to each other, which indicated the complication of intraspecific biodiversity, especially for the ASVs level. Six samples: BL, DY, BB, HX, PS, and LC, had an ASVs Shannon value that exceeded 1. Interestingly, sample PS comprising only *A. aspergillum* subgroup 1 had a relatively high alpha-diversity level based on ASVs, which implied the richness of germlines on this collection site. According to the composition of 302 earthworm ASVs, the 14 samples where *A. aspergillum* was absolutely

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

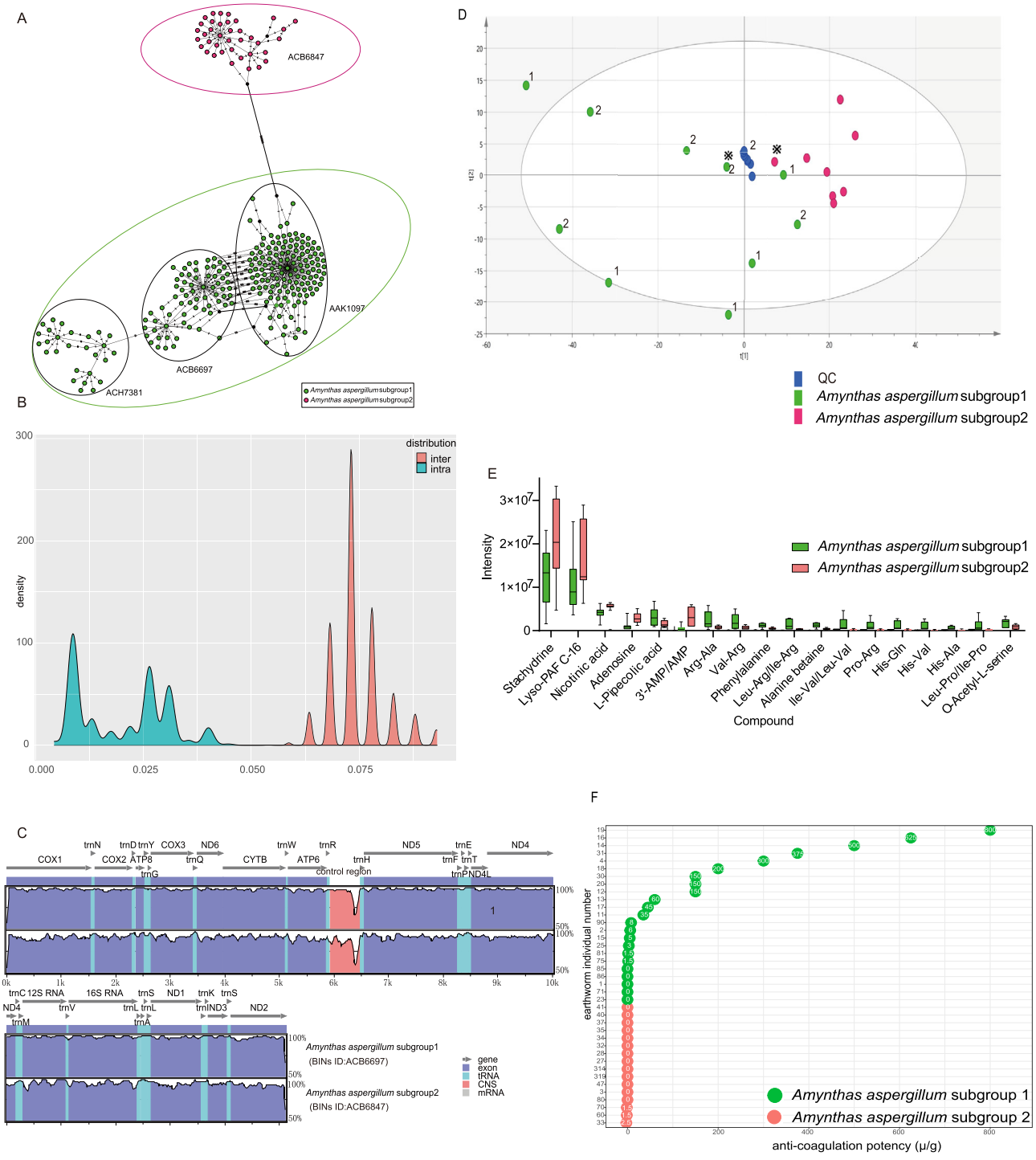


Figure 4 Intraspecies level biodiversity within *Amynthas aspergillum* samples. (A) Haplotype network of 302 ASVs within MOTU_1 (*Amynthas aspergillum*), and the corresponding BOLD BINs number was indicated; (B) Barcoding gap analysis between two subgroups; (C) Comparison of the mitochondrial genome sequences of two subgroups of *Amynthas aspergillum* using mVISTA alignment program, using the annotations of *A. aspergillum* (GenBank accession number: NC_025292, BINs ID: ACH7381) as reference. The vertical scale indicates the percentage of identity, ranging from 50% to 100%; (D) The PCA results between two subgroups: subgroup 1 samples with BINs ID: ACH 7381 (indicated as 1) and ACB 6697 (indicated as 2), and subgroup 2 samples with BINs ID: ACB6847. The QC samples were mainly distributed at the origin of the coordinates the PCA plot, which indicated acceptable analytical precision; (E) 18 differential chemical compositions were identified in two subgroups; (F) anti-coagulation potency of *Amynthas aspergillum* subgroup 1 and subgroup 2 samples. Two subgroups of individuals used for whole-genome skimming sequencing.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

predominant (>98%) were divided into 3 groups, as shown in Fig. 5E.

3.6. Testing traditional Chinese medicinal product samples containing earthworms

To further test our hypothesis that the source control of medical material "Guang Dilong" could benefit from the biodiversity pattern uncovered, three batches of 2796 dried earthworms collected from assigned sites were provided by a related pharmaceutical company for taxon identification. For these samples, a total of 5,753,904 paired-end reads of COI amplicon were generated. As a result, 105 ASVs were obtained, and rarefaction curve (Supporting Information Fig. S7) analysis indicated that the sequencing depths were sufficient for reliable results. The identification information of these ASVs was listed as Supporting Information Table S9. As shown in Fig. 6, while *A. aspergillum* subgroup 1 accounted for most of the proportion, *A. aspergillum* subgroup 2 and a few proportions of *A. dentiformis*, as unintentional substitution, were observed. This result was consistent with the species distribution we reported above.

For 25 Chinese patent medicine samples containing "Dilong" (earthworm), 10,593,960 paired-end reads and 304 ASVs were obtained, 37 ASVs of which were assigned to earthworms. As shown in the right section of Fig. 6, the sources of the earthworm components of GX and SH were subgroup 1 and subgroup 2 of *A. aspergillum*, individually. While the main component for NX was identified as *M. vulgaris*, one kind of source of "Dilong" included in *Chinese Pharmacopoeia*, *Metaphire tschiliensis* (Michaelsen, 1928) as common adulterant of "Dilong" recorded in the previous survey was also detected (percent ranging from 0 to 17.3%)[30,69,70]. Due to the limitation of resources yield, Chinese patent medicine companies prefer purchasing medical materials from different but relative stable geographic locations, and this may be one of the main causes of the species level diversity among different Chinese patent medicine samples.

4. Discussion

4.1. The advantage of DNA metabarcoding combined with mini-barcode in the discovery of biodiversity within batches of medical materials

As demonstrated in previous research studies[6,27,28,71,72], a large amount of cryptic biodiversity concealed amidst natural medical materials has not been discovered yet. DNA barcoding provided an effective tool for the discovery of cryptic species in the field of ecology and taxonomy[73,74]. However, because DNA degradation occurs during TCM manufacturing processing, the DNA fragments in extraction are frequently too short to be the template for "standard" DNA barcoding amplification, whose amplicon usually exceeds 500 bp[21]. DNA mini-barcode with short DNA segments ranging from 100 to 250 bp and sufficient variable sites could address the difficulties associated with "standard" DNA barcoding. As indicated by Lo et al.[75], DNA fragments of approximately 300 bp could be successfully amplified in herbs that have been boiled for 60 min, and PCR products of $\leq$120 bp were even observed for 120 min of the boiled sample. So far, several studies[75–77] have explicitly highlighted that the DNA mini-barcode expands the DNA barcoding method for assessing the quality of processed TCM materials.

For animals, 658 bp region in the gene encoding mitochondrial COI is one of the most important standard barcoding genes[78]. The key advantage of COI as a marker for DNA metabarcoding is that reference databases have been well established and are actively developed and extended. If the goal is to identify the species present in the sample, COI is the best choice due to the availability of extensive public reference databases[20,79]. Besides, the COI locus differs from many other metabarcoding loci (*e.g.*, 18S, 16S, 12S, ITS) in that it is a protein coding gene, imparting strict expectations of amplicon sequence read properties that can be exploited in metabarcoding bioinformatics[80]. In this study, to avoid the difficulty caused by DNA degradation, we selected a pair of primers that amplified the 281 bp fragment in COI 5′ region, which has been confirmed to be able to successfully identify earthworm components in feces.

A common concern for the mini-barcode is whether it can provide enough molecular polymorphism to differentiate species. However, Yeo et al.[81] recently observed that there was no significant difference between species-level identification performance of the full-length COI Folmer barcode and mini-barcode exceeding 200 bp across a large range of metazoan taxa, based on species delimitation algorithms. In our study, the 232 bp marker[39,82] we selected had a variation ratio that was consistent with that of the 658 bp "standard" COI sequence for *Amynthas* and *Metaphire* species, as shown in Fig. 1. Moreover, the mini-barcode ASVs identified as the same species based on the reference dataset, were clustered into single MOTU through species delimitation methods, indicating the reliable accuracy of the mini-barcode on species inference. In addition, 15 individuals were randomly selected to amplify a 658 bp Folmer region of the COI gene using LCO1490/HCO2198, and a 232 bp segment of the 5′ end of the COI gene with primers LCO1490/HCO1777, respectively, and the results of Sanger Sequencing showed that the identification results of the two markers were consistent (Table S2). In short, the 232 bp mini-barcode on COI genes was suitable for the biodiversity discovery of earthworms.

Another factor that limited the biodiversity survey on batched natural medicinal materials was the relatively high human resource and experimental costs of the regular individual identification method, such as DNA barcoding. DNA metabarcoding, simultaneous DNA barcoding of all specimens in a bulk sample using NGS platforms, allows for time- and cost-effective assessments of uncovered diversity[83,84]. To our knowledge, this study was the first application of metabarcoding on batched commercial TCM materials. In addition to species level (MOTUs) assignment, haplotype level (ASVs) information was obtained. Although unavoidably affected by PCR and sequencing error, metabarcoding bioinformatic pipeline as DADA2 could generate an error model based on the quality of sequencing run and use this model to distinguish between the predicted "true" biological variation (ASVs) and that was likely generated by systematic error. Ideally, these single-nucleotide level variations represent oligotypes within pool samples, which allow for description of intraspecific diversity[85–88]. Particularly for the COI barcode region, sequences that cannot be properly translated were excluded to further eliminate non-authentic ASVs, as nuclear mitochondrial pseudogenes (numts) and erroneous sequences.

Although the exact quantitative ability of metabarcoding has always been tested towing to primer bias, the relative abundances of MOTUs or ASVs were still comparable between samples

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

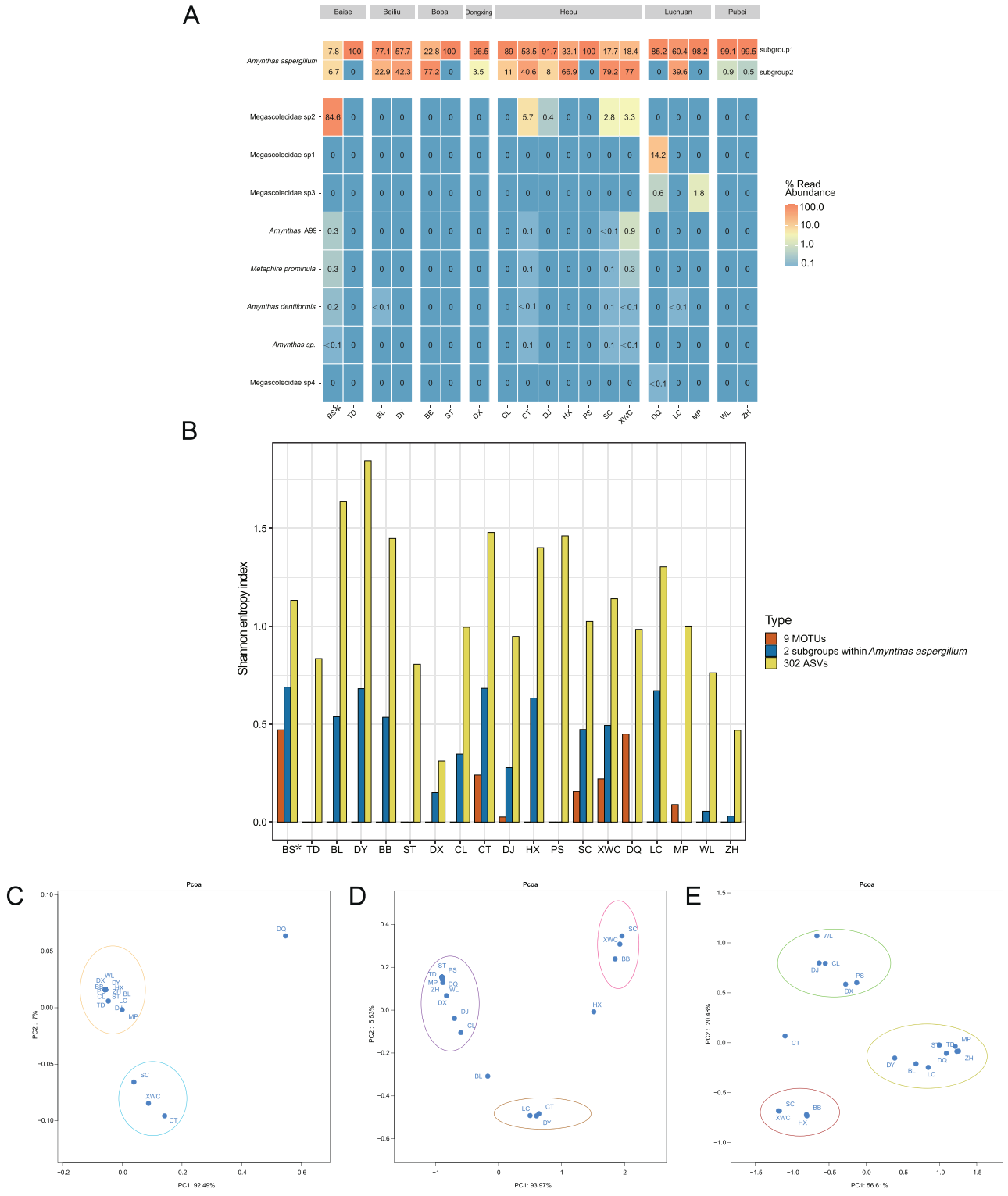


Figure 5 Biodiversity of the earthworm collected in the wild. (A) Heat map reflecting the species composition on 19 different collection sites; (B) Shannon index reflecting the alpha diversity within each sample; (C−E) Bray-Curtis PCoA (beta diversity) used to qualitatively examine differences in biological composition on MOTUs, *A. aspergillum* subgroups and ASVs level, separately. *Bought from Baise local market.

under the same experimental and bioinformatic processes. Diversity monitoring would benefit from this kind of read abundance comparison, even without a reference database and taxonomy assignment[89]. In this study, the species composition revealed from samples collected in the wild was largely consistent with that purchased from assigned areas.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as “Guang Dilong”, Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

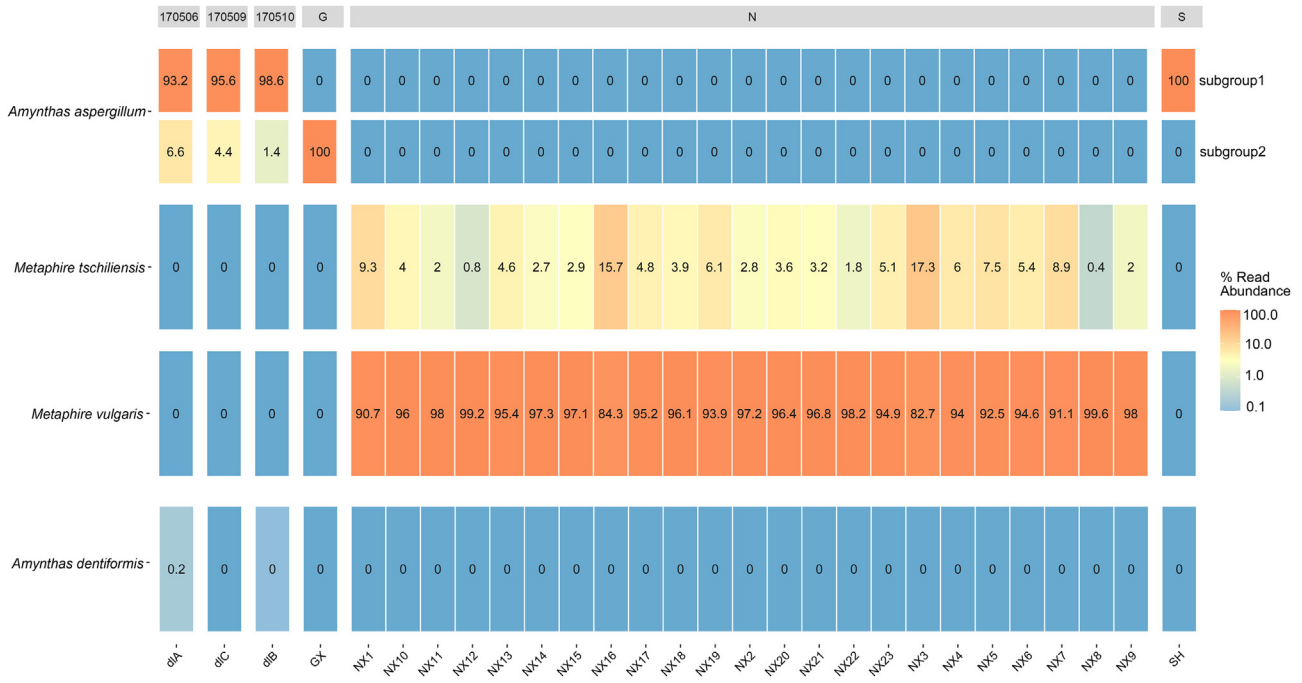


Figure 6 Heat map reflecting the species composition of 3 batches of decoction pieces and 25 Chinese patent medicines.

4.2. *The batch consistency control of "Guang Dilong" would benefit from proper production area selection and effective biological composition assessment approach*

The taxonomic identification of earthworm mainly relies on the morphological characteristics such as spermathecal pores, female pores and male pores, genital papillae, etc.[6,27,61,90], and requires the expertise of an experienced professional taxonomist[91]. More importantly, most of these characteristics would be destroyed during the preparing process of medicinal materials. Significantly, DNA barcoding uses DNA fragments exhibiting sufficient variation between different species to identify species[4,5]. Except for the identification of unknown specimens, DNA barcoding can speed up the designation of distinct lineages, and accelerate the discovery of the cryptic/undescribed earthworm species[6,7]. However, as traditional DNA barcoding identifies single specimens based on individual DNA sequencing reactions, it is often unsuitable for bulk commercial materials.

In this study, through mini-barcode and DNA metabarcoding, thousands of *Amynthas* earthworm individuals used as medicinal materials were treated in batches, and high interspecies and intraspecies diversity was demonstrated. Besides the intentional and unintentional adulteration, even the subgroups within *A. aspergillum* revealed here differed in chemical compositions and biological activity, indicating the need for origin control for this kind of wild source medicine.

Fortunately, this study identified the producing areas where *A. aspergillum* as the genuine source in the Pharmacopoeia taking overwhelming majority of the local earthworm community. As the discrimination result of subsequent commercial samples implied, it was feasible to stabilize the species origin of wild TCM by controlling their geographical origin. This study also provided an identification tool for earthworm medical materials, such as mini-barcode primers pair and reasonable species genetic distance threshold[92]. Combined with the power of DNA metabarcoding, it could support the large-scale sampling needed for time- and cost-efficient assessment of commercial materials.

Recently, the rising demand for TCM products in treating cardiovascular and cerebrovascular diseases has led to additional consumption of earthworm medicinal materials. Driven by growing market demand, over-harvesting may lead to a decline or loss of the biodiversity among earthworms. In the long term, more attention should be paid to the domestication and breeding of earthworms. The producing areas selected by this study could be regarded as artificial breeding bases. Furthermore, the intraspecies level diversity obtained elucidates existing germplasm resources, offers guidelines for *in-situ* conservation, and the core germplasm collection construction.

From a biology point of view, the Megascolecidae MOTUs that could not be identified at the species level also suggested the presence of numerous cryptic species or species without standard barcodes yet. For MOTU_5, identified as new species *Amynthas* A99, which was first collected in Hainan Province, China, a new geographic distribution record was reported. In addition, two subgroups whose genetic distance was neither in the intraspecies nor interspecies range were discovered within species *A. aspergillum*. Based on high-throughput method as DNA metabarcoding, the biodiversity assessment on abundant commercial TCM materials collected in the wild may elucidate studies on taxonomy and ecology.

4.3. *Data availability statement*

The datasets presented in this study are accessible from online repositories. The names of the repository/repositories and accession number(s) are accessible here: https://www.ncbi.nlm.nih.gov/genbank/, (NCBI accession number: ON553959), https://www.ncbi.nlm.nih.gov/genbank/, (NCBI accession number:

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

ON553960). Illumina data sets have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers: SRR19586215, SRR19446212 and SRR19855328.

5. Conclusions

To our knowledge, this was the first biological sources consistency evaluation of batched natural medicinal materials based on the well-evaluated mini-barcode and DNA metabarcoding. The earthworms as "Guang Dilong" were distinguished at haplotype level, from both commercially available medical materials and Chinese patent medicine. Although there were significant differences between two subgroups within *A. aspergillum* in terms of chemical compositions and biological activity, the hypothesis that the stability of biological composition would benefit from producing areas fixation was proved when the biodiversity pattern was uncovered at relatively fine geographical levels (e.g., village or town). We believe that, due to the advancement of handheld high-throughput sequencing technology and the declining sequencing cost, the origin control strategy described in this study should be introduced to improve the batch consistency of TCM, and to obtain the species community diversity information of medical materials depending on wild resources, particularly for taxa for which basic taxonomic and biodiversity works are woefully insufficient. In addition to stabilizing the biological source, the intraspecific biodiversity revealed by this method also offers guidelines for *in-situ* conservation and breeding bases construction.

Acknowledgments

This study was supported by Foundation Science and Technology Program of Tianjin (No. 22ZYJDSS00040, 20ZYJDJC00120, China), and Foundation CACMS Innovation Fund (No. CI2021A04104, China), Foundation Key project at central government level: The ability establishment of sustainable use for valuable Chinese medicine resources (No. 2060302, China). The authors would like to thank Yinglong Lv, Junhua Lv, Liqiang Zhang and Zilong He for their help in the collection of "Guang Dilong" samples.

Author contributions

Xiaoxuan Tian and Luqi Huang designed the study; Zhimei Xing and Han Gao performed the experiment and bioinformatic analysis; Zhimei Xing and Dan Wang drafted the manuscript; Ye Shang and Jibao Jiang provided chemical and morphological data to the manuscript; Chao Jiang, Luqi Huang and Xiaoxuan Tian revised the manuscript; Chunxiao Li, Hong Wang, Zhenguo Li, Lifu Jia, Yongsheng Wu, Dandan Wang, Wenzhi Yang, Yanxu Chang, Xiaoying Zhang, Tenukeguli Tuliebieke and Liuwei Xu coordinated the related research works. All authors approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supporting data to this article can be found online at https://doi.org/10.1016/j.apsb.2022.10.024.

References

1. Chen S, Wang Y, Zhao Z, Leon CJ, Henry RJ. Sustainable utilization of TCM resources. *Evid Based Complement Alternat Med* 2015;**2015**:613836.
2. Yang F, Ding F, Chen H, He M, Zhu S, Ma X, et al. DNA Barcoding for the identification and authentication of animal species in traditional medicine. *Evid Based Complement Alternat Med* 2018;**2018**:5160254.
3. Liu X, Liu H, Zhang C, Wei A, Ao H, Liu F, et al. Combination of c oxidase subunit I based deoxyribonucleic acid barcoding and HPLC techniques for the identification and quality evaluation of *Pheretima aspergillum*. *J Sep Sci* 2020;**43**:2989−95.
4. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc Biol Sci* 2003;**270**:313−21.
5. Chen S, Pang X, Song J, Shi L, Yao H, Han J, et al. A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol Adv* 2014;**32**:1237−44.
6. Dong Y, Law MMS, Jiang J, Qiu J. Three new species and one subspecies of the *Amynthas corticis*-group from Guangxi Zhuang autonomous region, China (Oligochaeta, Megascolecidae). *Zookeys* 2019;**884**:23−42.
7. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos Trans R Soc Lond B Biol Sci* 2005;**360**:1805−11.
8. Gueuning M, Ganser D, Blaser S, Albrecht M, Knop E, Praz C, et al. Evaluating next-generation sequencing (NGS) methods for routine monitoring of wild bees: metabarcoding, mitogenomics or NGS barcoding. *Mol Ecol Resour* 2019;**19**:847−62.
9. Xie H, Zhao Q, Shi M, Kong W, Mu W, Li B, et al. Biological ingredient analysis of traditional herbal patent medicine Fuke Desheng Wan using the shotgun metabarcoding approach. *Front Pharmacol* 2021;**12**:607197.
10. Zhang G, Liu J, Gao M, Kong W, Zhao Q, Shi L, et al. Tracing the edible and medicinal plant *Pueraria montana* and its products in the marketplace yields subspecies level distinction using DNA barcoding and DNA metabarcoding. *Front Pharmacol* 2020;**11**:336.
11. Liu J, Mu W, Shi M, Zhao Q, Kong W, Xie H, et al. The species identification in traditional herbal patent medicine, Wuhu San, based on shotgun metabarcoding. *Front Pharmacol* 2021;**12**:607200.
12. Liu J, Shi M, Zhao Q, Kong W, Mu W, Xie H, et al. Precise species detection in traditional herbal patent medicine, Qingguo Wan, using shotgun metabarcoding. *Front Pharmacol* 2021;**12**:607210.
13. Sa Q, Bao S, Wu R, Ao W, Bao G, Xu L, et al. Identification of mixed powder of 11 species of Aconitum Linnaeus medicinal materials based on high-throughput sequencing technology. *J Pharm Anal* 2021;**41**:1598−604.
14. Xing Y, Chen S, Xu L, Liang Y, Wang J, Wang B, et al. Study on high throughput sequencing identification of Fructus Arctii and five counterfeit species mix power. *Chin J Chin Mater Med* 2018;**43**:3862−6.
15. Zhang T, Liang Y, Xu L, Yang Y, Xing Y, Liu T, et al. Study on DNA molecular identification of mix samples of five species of Baitouweng medicinal materials based on high-throughput sequencing technology. *Acta Pharm Sin* 2018;**53**:1918−23.
16. Miya M. Environmental DNA metabarcoding: a novel method for biodiversity monitoring of marine fish communities. *Ann Rev Mar Sci* 2022;**14**:161−85.
17. Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, et al. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol Ecol* 2017;**26**:5872−95.
18. Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat Ecol Evol* 2016;**1**:4.
19. Turon X, Antich A, Palacín C, Præbel K, Wangensteen OS. From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecol Appl* 2020;**30**:e02036.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

20. Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Mol Ecol* 2018;**27**:3968−75.
21. Yang X, Yu X, Zhang X, Guo H, Xing Z, Xu L, et al. Development of mini-barcode based on chloroplast genome and its application in metabarcoding molecular identification of Chinese medicinal material Radix *Paeoniae Rubra* (Chishao). *Front Plant Sci* 2022;**13**:819822.
22. Liu Z, Xu J, Sun W, Shi Y, Chen S. Application of DNA metabarcoding in species identification of Chinese herbal medicine. *Chin J Chin Mater Med* 2019;**44**:1−8.
23. Pandit R, Travadi T, Sharma S, Joshi C, Joshi M. DNA metabarcoding using rbcL based mini-barcode revealed presence of unspecified plant species in Ayurvedic polyherbal formulations. *Phytochem Anal* 2021;**32**:804−10.
24. Yu X, Tan W, Gao H, Miao L, Tian X. Development of a specific mini-barcode from plastome and its application for qualitative and quantitative identification of processed herbal products using DNA metabarcoding technique: a case study on Senna. *Front Pharmacol* 2020;**11**:585687.
25. State Pharmacopoeia Commission of the People's Republic of China. *Pharmacopoeia of the People's Republic of China*. Beijing: Chinese medicines and Technology Press; 2020. p. 127.
26. Wu W. *Studies on the germplasm resources and quality estimation of Dilong (Earthworm)*. Guangzhou, China: Guangzhou University of Chinese Medicine; 2008.
27. Sun J, Jiang J, Bartlam S, Qiu J, Hu F. Four new *Amynthas* and *Metaphire* earthworm species from nine provinces in southern China. *Zootaxa* 2018;**4496**:287−301.
28. Dong Y, Yuan Z, Jiang J, Zhao Q, Qiu J. Two new species of earthworms belonging to the genus Amynthas (Oligochaeta: Megascolecidae) from Guangxi province, China. *Zootaxa* 2018;**4496**:259−68.
29. Huang Q, Li Z, Ma Z, Li H, Mao R. Specific and rapid identification of the Pheretima aspergillum by loop-mediated isothermal amplification. *Biosci Rep* 2019;**39**. BSR20181943.
30. Ge X, Jiang C, Tian N, Wei Y, Huang L, Yuan Y, et al. DNA sequencing to identify zoological origin of commercial *Pheretima* from Chinese. *Mod Chin Med* 2019;**21**:1206−14.
31. Sun J, Tian F, Mao R, Wu M, Zhang Y, Cao H, et al. Ldentification of commercial DiLong pieces by DNA barcode technology. *Strait Pharm J* 2021;**53**:1729−38.
32. Ma M. *The study on identification of Pheretima by highly specific PCR*. Guangzhou, China: Guangzhou University of Chinese Medicine; 2014.
33. Liu Q, Bi Q, Tan N. Authentication of three main commercial *Pheretima* based on amino acids analysis. *Amino Acids* 2021;**53**:1729−38.
34. Sun J, Tian F, Zhang Y, Wu M, Mao R, Le Z, et al. Chromatographic fingerprint and quantitative analysis of commercial *Pheretima aspergillum* (Guang Dilong) and its adulterants by UPLC−DAD. *Int J Anal Chem* 2019;**2019**:4531092.
35. Liu Q, Bi Q, Zhang J, Qin W, Yi S, Hu Q, et al. A rapid and simple signature peptides-based method for species authentication of three main commercial *Pheretima*. *J Proteomics* 2022;**255**:104456.
36. Gu Y, Zhang J, Sun J, Yu H, Feng R, Mao X, et al. Marker peptide screening and species-specific authentication of *Pheretima* using proteomics. *Anal Bioanal Chem* 2021;**413**:3167−76.
37. Latif R, Malek M, Aminjan AR, Pasantes JJ, Briones MJI, Csuzdi C. Integrative taxonomy of some Iranian peregrine earthworm species using morphology and barcoding (Annelida: Megadrili). *Zootaxa* 2020;**4877**. zootaxa.4877.1.7.
38. 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**:294−9.
39. Brown DS, Jarman SN, Symondson WOC. Pyrosequencing of prey DNA in reptile faeces: analysis of earthworm consumption by slow worms. *Mol Ecol Resour* 2012;**12**:259−66.
40. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587−9.
41. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268−74.
42. Schnell IB, Bohmann K, Gilbert MTP. Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol Ecol Resour* 2015;**15**:1289−303.
43. Bushnell B, Rood J, Singer E. BBMerge—Accurate paired shotgun read merging *via* overlap. *PLoS One* 2017;**12**:e0185056.
44. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from *Illumina amplicon* data. *Nat Methods* 2016;**13**:581−3.
45. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003;**14**:927−30.
46. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;**8**:e61217.
47. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol Ecol* 2012;**21**:1864−77.
48. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 2013;**29**:2869−76.
49. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* 2007;**17**:377−86.
50. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformat* 2010;**11**:538.
51. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 2017;**45**:e18.
52. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 2013;**69**:313−9.
53. Zhang L, Jiang J, Dong Y, Qiu J. Complete mitochondrial genome of four pheretimoid earthworms (Clitellata: Oligochaeta) and their phylogenetic reconstruction. *Gene* 2015;**574**:308−16.
54. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: molecular evolutionary genetics analysis across computing rlatforms. *Mol Biol Evol* 2018;**35**:1547−9.
55. Wickham H. *ggplot2: Elegant graphics for data analysis, 1st ed. 2009. Corr. 3rd printing 2010 edition*. New York: Springer; 2010.
56. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, et al. Vista : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 2000;**16**:1046−7.
57. Parks DH, Porter M, Churcher S, Wang S, Blouin C, Whalley J, et al. GenGIS: a geospatial information system for genomic data. *Genome Res* 2009;**19**:1896−904.
58. Morinière J, Hendrich L, Balke M, Beermann AJ, König T, Hess M, et al. A DNA barcode library for Germany's mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera and Trichoptera). *Mol Ecol Resour* 2017;**17**:1293−307.
59. Albo JE, Marelli JP, Puig AS. Rapid molecular identification of Scolytinae (Coleoptera: Curculionidae). *Int J Mol Sci* 2019;**20**:E5944.
60. Geller J, Meyer C, Parker M, Hawk H. Redesign of PCR primers for mitochondrial cytochrome *c* oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol Ecol Resour* 2013;**13**:851−61.
61. Dong Y, Jiang J, Yuan Z, Zhao Q, Qiu J. Population genetic structure reveals two lineages of *Amynthas triastriatus* (Oligochaeta: Megascolecidae) in China, with notes on a new subspecies of *Amynthas triastriatus*. *Int J Environ Res Public Health* 2020;**17**:E1538.
62. Yuan Z. *Taxon, differentiation and dispersal of earthworms in the hengduan mountains*. Shanghai, China: Shanghai Jiao Tong University; 2020.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024

63. Meng J, Zhou C, Zhang W, Wang W, He B, Hu B, et al. Stachydrine prevents LPS-induced bone loss by inhibiting osteoclastogenesis *via* NF-κB and Akt signalling. *J Cell Mol Med* 2019;**23**:6730−43.
64. Riaz MS, Kaur A, Shwayat SN, Behboudi S, Kishore U, Pathan AA. Direct growth inhibitory effect of platelet activating factor C-16 and its structural analogs on mycobacteria. *Front Microbiol* 2018;**9**:1903.
65. Blond E, Rieusset J, Alligier M, Lambert-Porcheron S, Bendridi N, Gabert L, et al. Nicotinic acid effects on insulin sensitivity and hepatic lipid metabolism: an *in vivo* to *in vitro* study. *Horm Metab Res* 2014;**46**:390−6.
66. Denizot Y, Rougier F, Dupuis F, Trimoreau F, Dulery C, Laskar M, et al. Presence and metabolism of lyso platelet-activating factor in human bone marrow. *J Lipid Mediat Cell Signal* 1997;**16**:53−62.
67. Sun AH, Liu XX, Yan J. Leptospirosis is an invasive infectious and systemic inflammatory disease. *Biomed J* 2020;**43**:24−31.
68. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. *J Chromatogr A* 2014;**1353**:99−105.
69. Guo L, Wang Q, Zhang D, Zheng K, Zheng Y, Hou F. Identification of micro-traits and quality of Di Long herbs from different origins. *Chin Med Mat* 2018;**41**:66−9.
70. Gao X, Zhao Y, Guo Y, Nian J, Zhang H, Wu Y, et al. Morphological and DNA double barcode identification of *Pheretima* and its adulterants. *Chin Tradit Herbal Drugs* 2020;**51**:2530−7.
71. Wang C, Zhang Y, Ding H, Song M, Yin J, Yu H, et al. Authentication of zingiber species based on analysis of metabolite profiles. *Front Plant Sci* 2021;**12**:705446.
72. Zhu S, Li Q, Chen S, Wang Y, Zhou L, Zeng C, et al. Phylogenetic analysis of Uncaria species based on internal transcribed spacer (ITS) region and ITS2 secondary structure. *Pharm Biol* 2018;**56**:548−58.
73. Saccò M, Guzik MT, van der Heyde M, Nevill P, Cooper SJB, Austin AD, et al. eDNA in subterranean ecosystems: applications, technical aspects, and future prospects. *Sci Total Environ* 2022;**820**:153223.
74. Kulik T, Bilska K, Żelechowski M. Promising perspectives for detection, identification, and quantification of plant pathogenic fungi and oomycetes through targeting mitochondrial DNA. *Int J Mol Sci* 2020;**21**:E2645.
75. Lo YT, Li M, Shaw PC. Identification of constituent herbs in ginseng decoctions by DNA markers. *Chin Med* 2015;**10**:1.
76. Su Y, Ding D, Yao M, Wu L, Dong G, Zhang D, et al. Specific DNA mini-barcoding for identification of *Gekko gecko* and its products. *Chin Med* 2020;**15**:103.
77. Ragupathy S, Faller AC, Shanmughanandhan D, Kesanakurti P, Shaanker RU, Ravikanth G, et al. Exploring DNA quantity and quality from raw materials to botanical extracts. *Heliyon* 2019;**5**:e01935.
78. Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, et al. Advances in DNA metabarcoding for food and wildlife forensic species identification. *Anal Bioanal Chem* 2016;**408**:4615−30.
79. Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel JN, et al. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ* 2016;**4**:e1966.
80. Creedy TJ, Andújar C, Meramveliotakis E, Noguerales V, Overcast I, Papadopoulou A, et al. Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Mol Ecol Resour* 2022;**22**:847−61.
81. Yeo D, Srivathsan A, Meier R. Longer is not always better: optimizing barcode length for large-scale species discovery and identification. *Syst Biol* 2020;**69**:999−1015.
82. Pearson CE, Symondson WOC, Clare EL, Ormerod SJ, Iparraguirre Bolaños E, Vaughan IP. The effects of pastoral intensification on the feeding interactions of generalist predators in streams. *Mol Ecol* 2018;**27**:590−602.
83. Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, et al. Reliable, verifiable and efficient monitoring of biodiversity *via* metabarcoding. *Ecol Lett* 2013;**16**:1245−57.
84. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, et al. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol* 2014;**29**:358−67.
85. Buttigieg PL, Ramette A. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol Ecol* 2014;**90**:543−50.
86. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* 2015;**9**:968−79.
87. Hanson CA, Müller AL, Loy A, Dona C, Appel R, Jørgensen BB, et al. Historical factors associated with past environments influence the biogeography of thermophilic endospores in arctic marine sediments. *Front Microbiol* 2019;**10**:245.
88. Morinière J, Balke M, Doczkal D, Geiger MF, Hardulak LA, Haszprunar G, et al. A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Mol Ecol Resour* 2019;**19**:900−28.
89. Apothéloz-Perret-Gentil L, Cordonier A, Straub F, Iseli J, Esling P, Pawlowski J. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Mol Ecol Resour* 2017;**17**:1231−42.
90. Novo M, Almodóvar A, Fernández R, Trigo D, Díaz Cosín DJ. Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Mol Phylogenet Evol* 2010;**56**:507−12.
91. Sigsgaard EE, Torquato F, Frøslev TG, Moore ABM, Sørensen JM, Range P, et al. Using vertebrate environmental DNA from seawater in biomonitoring of marine habitats. *Conserv Biol* 2020;**34**:697−710.
92. Ma Z, Ren J, Zhang R. Identifying the genetic distance threshold for Entiminae (Coleoptera: Curculionidae) species delimitation *via* COI barcodes. *Insects* 2022;**13**:261.

Please cite this article as: Xing Zhimei et al., A novel biological sources consistency evaluation method reveals high level of biodiversity within wild natural medicine: A case study of *Amynthas* earthworms as "Guang Dilong", Acta Pharmaceutica Sinica B, https://doi.org/10.1016/j.apsb.2022.10.024