

GOPEN ACCESS

Citation: Niu M, Liu Y, Xue L, Cai B, Zhao Q, Wei J (2024) Improving DNA barcoding library of armored scale insects (Hemiptera: Diaspididae) in China. PLoS ONE 19(5): e0301499. https://doi.org/ 10.1371/journal.pone.0301499

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: December 27, 2023

Accepted: March 18, 2024

Published: May 30, 2024

Copyright: © 2024 Niu et al. This is an open access article distributed under the terms of the <u>Creative</u> Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the National Natural Science Foundation of China (No. 32100370), Science and Technology Innovation Funds of Shanxi Agricultural University (2020BQ79), the Excellent Doctoral Award of Shanxi Province for Scientific Research Project (SXBYKY2021024) and Science and Technology Innovation Projects of Universities in Shanxi Province (2021L097), the Natural Science

RESEARCH ARTICLE

Improving DNA barcoding library of armored scale insects (Hemiptera: Diaspididae) in China

Minmin Niu¹⁺, Yubo Liu¹⁺, Linjia Xue¹, Bo Cai², Qing Zhao¹, Jiufeng Wei¹*

1 College of Plant Protection, Shanxi Agricultural University, Taigu, China, 2 Hainan Province Engineering Research Center for Quarantine, Prevention and Control of Exotic Pests, Haikou, China

‡ MN and YL are co-first author to this work.
* wjfeng@nwsuaf.edu.cn

Abstract

DNA barcoding is used to identify cryptic species, survey environmental samples, and estimate phyletic and genetic diversity. Armored scale insects are phytophagous insects and are the most species-rich taxa in the Coccoidea superfamily. This study developed a DNA barcode library for armored scale insect species collected from southern China during 2021–2022. We sequenced a total of 239 specimens, recognized as 50 morphological species, representing two subfamilies and 21 genera. Sequencing analysis revealed that the average G + C content of the cytochrome oxidase subunit I (COI) gene sequence was very low (~18.06%) and that the average interspecific divergence was 10.07% while intraspecific divergence was 3.20%. The intraspecific divergence value was inflated by the high intraspecific divergence in ten taxa, which may indicate novel species overlooked by current taxonomic treatments. All the Automated Barcode Gap Discovery, Assemble Species by Automatic Partitioning, Taxon DNA analysis and Bayesian Poisson Tree Process methods yielded largely consistent results, indicating a robust and credible species delimitation. Based on these results, an intergeneric distance threshold of \leq 5% was deemed appropriate for the differentiation of armored scale insect species in China. This study establishes a comprehensive barcode library for the identification of armored scale insects, future research, and application.

Introduction

DNA barcoding is a method of species identification, using short standardized DNA fragments, first proposed by Herbert et al. in 2003 [1, 2]. It has since been used in several fields of biology, including taxonomy [3–5], ecology [6, 7], conservation biology [8], and evolution [9]. It can be used to elucidate cryptic species [10], survey environmental samples [11], and estimate phyletic and genetic diversities [12, 13]. Additionally, DNA barcoding can be used for species identification, when information on the morphology and taxonomy of the species is limited [14], such as for cryptic species and immature or mutilated specimens.

Scale insects are sap-sucking plant parasites that play an important role in the ecosystem [15]. Honeydew, the waste generated by scale insects feeding exclusively on the phloem sap of

Foundation of Shanxi, China (202103021224132), Hainan Province Science and Technology Special Fund (ZDKJ2021016).

Competing interests: The authors have declared that no competing interests exist.

host plants, is an important food source for birds, mammals and especially other insects [15]. However, they are common pests of perennial plants in managed systems and may cause chlorosis and leaf fall [16]. Moreover, some armored scale species can prey on more than 100 plant families, including fruit and nut crops, cotton, cereal crops, and forest and ornamental plants [17-20]. They belong to the superfamily Coccoidea and comprise more than 8000 described species from approximately 50 families [15]. Armored scale insects have a global distribution and are the most species-rich taxa in the Coccoidea, comprising more than 2600 species and approximately 400 genera in the family Diaspididae [15]. More specifically, there are 1108 species of scale insects in China, including 452 species of armored scale insects from 82 genera [15]. The Diaspididae possibly comprises the most invasive insect species, as their small size makes them cryptic and difficult to detect and identify [21]. At present, microscopic observations of adult female specimens are the most popular method for identifying armored scale species [22]; however, morphological identification requires an expert, and it cannot distinguish between two closely related species [23]. Additionally, it is difficult to identify species based on the morphological characteristics of specimens from different developmental stages, such as crawlers, second and third instar nymphs, or eggs [24].

DNA barcoding has been used with many taxa in the Coccoidea for primer designing [25], species identification [26–31], genetic diversity estimation [32], and quarantine inspection [33]. However, a search in the Barcode of Life Data System (BOLD) database in July 2023 using the term "Diaspididae" produced 2412 published records, whereas the term "Diaspididae China" produced only 8 records (~0.3%), suggesting a lack of data on the biodiversity of armored scale insects in China.

Fragments of the cytochrome c oxidase subunit I (*COI*) [25, 34] and 28S [35, 36] genes are primarily employed in the identification of armored scale insects. However, some studies suggest that 28S rDNA lacks sufficient variation to delimitate some species [26, 37]. Therefore, 28S is considered a complementary marker to *COI* in scale insects [31]. This study aimed to develop a comprehensive DNA barcode library for the armored scale insect species of China as well as assess the accuracy of *COI* barcodes in armored scale insects.

Materials and methods

Specimen collection and identification

A total of 239 armored scale specimens, identified as 50 morphological species, representing two subfamilies and 21 genera were collected from 60 host plant species in seven provinces of southern China (Yunnan, Guangxi, Fujian, Zhejiang, Hainan, Sichuan, and Guizhou) during 2021–2022 (Fig 1). The specimens and their host plant tissues were stored at –20°C for further analysis. Detailed information on each specimen, including location, host plant, and GenBank accession numbers in NCBI (https://www.ncbi.nlm.nih.gov/), is provided in S1 Table. A combined molecular/morphological preparation protocol was performed on each specimen to obtain genomic DNA from the specimens and permanent slides of its cuticle [38]. Morphological identification was conducted independently by Jiufeng Wei and Minmin Niu, according to the morphological studies of the Diaspididae of Williams and Watson [39], Tang [20, 40, 41], Chou [42–44]. Specimen vouchers were deposited in Insect Specimen Museum, College of Plant Protection, Shanxi Agricultural University.

DNA extraction, PCR, and sequencing

Total genomic DNA of the specimens were extracted using the Ezup column-based animal genomic DNA extraction kit (Shanghai Bioengineering Co., Ltd, Shanghai, China). Thereafter, the DNA samples were used as templates for a PCR amplifying an approximately 650 base pair





(bp) fragment of the *COI* barcode region using the universal primers PcoF1 (5' –CCTTCAAC TAATCATAAAAATATYAG–3') [25] and LepR1 (5' –TAAACTTCTGGATGTCCAAAAAAT CA–3') [26]. The 30 μ L reaction contained 9.5 μ L ddH₂O, 13.5 μ L 2x Taq Master mix (1 mL), 1.5 μ L of each primer (10 μ M), and 4 μ L of template DNA. PCR reaction conditions were as follows: initial denaturation at 95°C for 3 min; followed by 5 cycles of 95°C for 1 min, 48°C for 2 min, and 72°C for 1 min and 35 cycles of 95°C for 1 min, 51°C for 2 min, and 72°C for 1 min are 32°C for 8 min. The PCR products were visualized using 1% agarose gel electrophoresis and sequenced using the forward primer by Qingke Biotechnology Co., Ltd (Shaanxi, China).

Sequence analysis

The sequencing results were viewed using Chromas v1.62 [45] and sorted using the TBtools software [46]. Subsequently, all sequences were aligned and trimmed using MEGA v11.0 [47] to obtain a matrix of 586-bp long sequences for further molecular analyses.

Genetic distance and phylogenetic analysis

The intraspecific and interspecific genetic distances were calculated using the Kimura 2 parameter (K2P) [48] and prior intraspecific divergence (P)-distance models and a neighborjoining (NJ) tree was constructed with the K2P model and 10000 bootstrap replicates in MEGA v11.0. The maximum likelihood (ML) phylogenetic tree was constructed using Phylo-Suite v1.2.3 [49] using the following settings: ML + standard bootstrap, 1000 bootstrap replicates, and the TIM + F + I + G4 model obtained by the ModelFinder software [50] under the BIC standard. A maximum parsimony (MP) tree was also constructed with 1000 bootstrap replicates using MEGA v11.0 and the Bayesian inference (BI) phylogenetic tree was constructed with PhyloSuite v1.2.3 using the following settings: 20000000 generations and the GTR+F+I+G4 model obtained by ModelFinder software under BIC standard. Two sequences of *Paracoccus marginatus* (Pseudococcidae, Hemiptera) were used as outgroups (GenBank accession numbers: OR544511 and OR544512). The layout of the NJ tree was edited using the Interactive Tree of Life v6 (https://itol.embl.de/) [51].

Species delimitation

Several methods have been proposed for species identification based on molecular data [52-56]. This study used the Automated Barcode Gap Discovery (ABGD) (https://bioinfo.mnhn.fr/ abi/public/abgd/abgdweb.html) [52], Bayesian Poisson Tree Process (bPTP) [53], Assemble Species by Automatic Partitioning (ASAP) (https://bioinfo.mnhn.fr/abi/public/asap/asapweb. html) [55] and Taxon DNA analysis [56] methods to assess species boundaries and delimit possible species. The ABGD method is a clustering approach based on genetic distances, and it can be performed using different models (JC69, K2P, and P-distance). In the ABGD approach, the relative gap width was set to 1.0 and the P-distance was set to 0.001–0.1. The bPTP method is based on interspecific and intraspecific substitutions, and it assumes that the number of substitutions within a species is lower than the number of substitutions between species. The distinction between species is then achieved by calculating these two values [57]. The ML tree was analyzed on the PTP website (https://species.h-its.org/ptp/) with the following settings: Rooted, Remove outgroups, and default settings for the other parameters. The ASAP analysis is a hierarchical clustering algorithm based on an evolutionary theory that avoids the computational burden of phylogenetic reconstruction by using only pairwise genetic distances [55]. In the ASAP approach, the default settings were selected and data analyzed by three models, Jukes-Cantor (JC69), Kimura (K80) ts/tv 2.0, and Simple Distance (p-distances). For the Taxon DNA Analysis method, parameters were set to Best match, Best close match, All species barcodes and Cluster methods to analyze the COI sequences.

Results

Genetic distance and phylogenetic analysis

Analysis of the sequencing data revealed 222 conserved sites, 364 variable sites, 334 parsimony information sites, and 30 singleton sites accounting for 37.88%, 62.12%, 56.99%, and 5.12% of the total *COI* gene sequence (586-bp), respectively. The average thymine (T), cytosine (C), adenine (A), and guanine (G) contents of the *COI* gene sequences were 41.32%, 11.44%, 40.62%, and 6.62%, respectively. The average A + T and G + C contents of the *COI* gene sequences were 81.94% and 18.06%, respectively, indicating a strong A + T bias.

As expected, the mean genetic distances increased hierarchically with taxonomic categories based on the K2P and P-distance models. As seen in Fig 2, the genetic distances calculated by the K2P model were only slightly higher than those calculated by the P-distance model, with intraspecific, interspecific, and intergeneric distances of 3.20% and 3.02%; 10.94% and 10.07%; 20.91% and 18.03%, respectively. Since the results of the two models were very similar, the K2P distance model was used for further analysis. The intraspecific K2P distances ranged from 0 to 15.34%, and approximately 68.43% of the intraspecific distances were < 5%, while 99.88% of the interspecific distances were > 5%. This suggests an obvious barcoding gap in the data (Fig 3). Also, 10 species had K2P distances greater than 2% (S2 Table and Fig 4).

Based on the data in Table 1 and Fig 3, the intergeneric distance threshold of \leq 5% was appropriate for the differentiation of Diaspididae in China, where intergeneric distances > 5% suggest the presence of cryptic species.



Fig 2. Line chart of the mean genetic distances of the cytochrome oxidase subunit I (*COI*) gene sequences of armored scale insects from China at different taxonomic levels based on the Kimura 2 parameter and prior intraspecific divergence (P) distance models.

The NJ tree constructed from the 239 *COI* sequences can be seen in Fig 5. A total of 46 species formed monophyletic branches with a high bootstrap support at the species level, which was consistent with traditional taxonomic results. The ML tree constructed from the 239 *COI* sequences can be seen in Fig 6. A total of 46 species formed monophyletic branches at the species level with high bootstrap support, which was consistent with traditional taxonomic results.

The MP tree constructed from the 239 *COI* sequences is shown in Fig 7. A total of 46 species formed monophyletic branches at the species level with high bootstrap support, which was consistent with traditional taxonomic results. The BI tree constructed from the 239 *COI* sequences showed a total of 46 species on monophyletic branches at the species level with high bootstrap support, which was consistent with traditional taxonomic results (Fig 8).

Species delimitation

The results of the ABGD analysis performed using the JC69, K2P, and P-distance models were compared and data revealed that the number of groups based on the P-distance model ranged from 66 to 75, and the initial partition engendered 66 groups (P = 0.0028–0.0359) (Table 2) (S3 Table). The ABGD analysis categorized 13 sequences of *Chrysomphalus dictyospermi* into two groups, 12 sequences of *Lepidosaphes beckii* into three groups, 19 sequences of *Pseudaula-caspis cockerelli* into four groups, 11 sequences of *Chrysomphalus bifasciculatus* into four groups, eight sequences of *Hemiberlesia lataniae* into three groups, five sequences of *Pinnaspis theae* into five groups, five sequences of *Aonidiella inornata* into two groups, three sequences of *Parlatoria camelliae* into two groups, and six sequences of *Pseudaulacaspis celtis* along with



Fig 3. Frequency histogram of intraspecific and interspecific genetic distances of armored scale species from China, based on the cytochrome c oxidase subunit I (COI) sequences.

seven sequences of *P. cockerelli* into one group. These results indicate a species classification success rate (classification of the same species in the same group) of 65.69% for the ABGD analysis.

Table 3 shows the results of the ASAP analysis under the three models. The lower the ASAP score, the more reliable the partitioning results. The partitioning of the K2P model was closest to the morphological results (S4 Table). The ASAP analysis categorized 13 sequences of *C. dic-tyospermi* into two groups, 12 sequences of *L. beckii* into three groups, 19 sequences of *P. cock-erell* into four groups, 11 sequences of *C. bifasciculatus* into four groups, 11 sequences of *Aulacaspis rosae* into two groups, eight sequences of *H. lataniae* into three groups, seven sequences of *Parlatoria proteus* into three groups, five sequences of *P. theae* into five groups, and six sequences of *P. celtis* along with seven sequences of *P. cockerelli* into one group. These results indicate a species classification success rate of 61.51% for the ASAP analysis.

In the Taxon DNA analysis, the threshold of 239 *COI* sequences was calculated to be 6.82% by the Pairwise Summary method. With the Best match method, the number of accurately identified sequences was 222, with a success rate of 92.88%; the number of ambiguous sequences was 10, accounting for 4.18% of all the sequences; the number of incorrect identifications was seven, accounting for 2.92%. For the Best close match method, the number of accurately identified sequences was 215, with a success rate of 89.95%; the number of ambiguous sequences was eight, accounting for 3.34% of all sequences; the number of incorrect identifications was three, accounting for 1.25%; the number of sequences without any match closer than 6.82% was 13, accounting for 5.43%. With the All Species Barcodes method, the number of accurately identified sequences was six, the success rate was 2.51%; the number of





ambiguous sequences was 218, accounting for 91.21%; the number of incorrect identifications was two, accounting for 0.83%, and the number of sequences with no match closer than 6.82% was 13, accounting for 5.43%.

The Cluster method divided 239 sequences into 70 groups with a default threshold of 3% (S5 Table). The Cluster analysis categorized 13 sequences of *C. dictyospermi* into two groups, 12 sequences of *L. beckii* into three groups, 19 sequences of *P. cockerell* into four groups, 11 sequences of *C. bifasciculatus* into four groups, 11 sequences of *Aulacaspis rosae* into two groups, eight sequences of *H. lataniae* into three groups, seven sequences of *Parlatoria proteus* into three groups, five sequences of *P. theae* into five groups, five sequences of *A. inornata* into two groups, three sequences of *P. cockerelli* into two groups, and six sequences of *P. celtis* along with seven sequences of *P. cockerelli* into one group. These results indicate a species classification success rate of 58.16% for the Cluster analysis.

A total of 74 putative species were delimited using the bPTP analysis (S1 Fig). The bPTP analysis categorized five sequences of *P. theae* into five groups, 12 sequences of *L. beckii* into three groups, five sequences of *A. inornata* into two groups, eight sequences of *H. lataniae* into three groups, three sequences of *P. camelliae* into three groups, 19 sequences of *P. cockerell*

Table 1.	Kimura 2 parameter genetic distances (%) based on the cytochrome oxidase subunit I gene (COI) of the armored scale insects from China	a at different taxo-
nomic lev	els.	

Taxonomic level	Cytochrome oxidase subunit I				
	Minimum (%)	Mean (%)	Maximum (%)		
Intraspecific distance	0.00	3.20	15.34		
Interspecific distance	5.39	10.94	19.90		
Intergeneric distance	5.39	20.91	36.89		

https://doi.org/10.1371/journal.pone.0301499.t001





into five groups, 13 sequences of *C. dictyospermi* into two groups, 11 sequences of *C. bifasciculatus* into four groups, nine sequences of *Pseudaulacaspis pentagona* into three groups, 11 sequences of *Aulacaspis rosae* into two groups, seven sequences of *Parlatoria proteus* into three groups, three sequences of *Pseudaulacaspis prunicola* into two groups, and six sequences of *P. celtis* along with five sequences of *P. cockerelli* into one group. These results indicate a species classification success rate of 53.14% for the bPTP analysis.



Fig 6. Radial chronogram of the delimited armored scale insect species from China. The backbone tree represents the maximum likelihood (ML) tree based on the cytochrome c oxidase subunit I (*COI*) gene sequences, and the colored circles represent different morphospecies.

Discussion

Paul Hebert first introduced the concept of DNA barcoding in 2003 and suggested that the *COI* mitochondrial gene can be used as a universal barcode to identify all animals [1, 2]. Thereafter, *COI*-based DNA barcoding has been applied to delineate species in a wide range of animal taxa [58–63], by primarily utilizing the apparent gaps in genetic distances between the





COI sequences of different species. After analyzing 13320 organisms, Herbert et al. [1, 2] proposed an intraspecific genetic distance of $\leq 2\%$, which is still widely accepted. However, studies have found that the thresholds of genetic distance were not completely uniform across species [64]. The current identification threshold for the BOLD database is 3% [65] but specific species boundary thresholds have been identified for many insects. For instance, Ball & Armstrong [66] obtained a 12.8% interspecific divergence for the New Zealand sooty beech scale







insect, while Park et al. [26] found an average of 10.7% interspecific divergence and 0.97% intraspecific divergence for the Pseudococcidae and Diaspididae using *COI*-based DNA barcoding.

This study used five methods to determine whether *COI*-based DNA barcoding is effective in delimiting armored scale insect species from China. The K2P model provided a maximum intraspecific genetic distance of 15.34%, which significantly exceeds the 2% threshold proposed

Model	X	Partition	Prior intraspecific divergence (P)							
			0.001	0.0017	0.0028	0.0046	0.0077	0.0129	0.0215	0.0359
JC69	1	Initial	66	66	66	66	66	66	66	66
		Recursive	92	92	83	77	75	73	71	66
K2P	1	Initial	66	66	66	66	66	66	66	66
		Recursive	93	93	83	77	75	73	71	66
Р	1	Initial	66	66	66	66	66	66	66	66
		Recursive	75	75	75	72	72	69	69	66

Table 2.	ABGD analy	vsis of COI see	uences of armore	d scale insect spe	ecies from China.	based on three evolutional	v models.
Tuble 2.	mbob unu	y 313 01 0 01 3 0	quenees of armore	a scale moter spe	ceres from onnia,	bused on three evolutional	y moucis.

by Hebert and the 3% threshold of the BOLD database. Results revealed a significant overlap between the intraspecific and interspecific genetic distances, which may be because of insufficient data or inaccurate morphological identification of samples. However, these results are nonetheless informative for determining Diaspididae thresholds. The ABGD analysis showed a 65.69% species delineation success rate, does not require any input on genetic distance, and the results can be seen in its delineation of the threshold range of the species, to provide a reference for the selection of the correct threshold. The ASAP analysis showed a 61.51% species delineation success rate and is simple and easy to use, with clear results. For the Taxon DNA Analysis, the accurate identification rates of the Best match, Best close match and All species barcodes methods were 92.88%, 89.95% and 2.51%, respectively. The Cluster analysis showed a 58.16% species delineation success rate and is a commonly used method for analyzing the success rate of DNA barcoding [56]. The bPTP analysis showed a 53.14% success rate of species delineation and requires a prior input of a phylogenetic tree for species delineation, which is complicated and time-consuming.

In these analyses, most of the species could be clustered into groups. Therefore, a combination of genetic divergence analysis along with the NJ, ABGD, ASAP, Taxon DNA and bPTP analyses may help in the accurate identification of armored scale insect species. Based on the results of this study, an intergeneric distance of \leq 5% was considered an appropriate threshold for the identification of the Diaspididae in China using the *COI* gene region. The NJ tree, ML tree, MP tree and BI tree all showed a total of 46 species, forming monophyletic branches at the species level.

In this study, fresh and full-bodied female armored scale insect specimens were collected and molecular and morphological analyses performed. The morphological integrity of the insects was maintained after genomic DNA extraction, which played an important role in post-experimental morphological review and species preservation. Although the obtained DNA concentration of the specimens was not high, it was sufficient for PCR amplification. This suggests that DNA barcoding can even be performed using small quantities of DNA from morphologically indistinguishable or mutilated specimens. This adds advantage to the

Table 3. ASAP analysis of (COI sequences of armored	scale insect species from	China, based on three e	volutionary models.
1				1

	Model	ASAP scores -10 best partitions									
	JC69	5.50	7.00	7.50	7.50	8.50	9.00	11.00	11.50	11.50	13.50
		-70	-70	-73	-74	-71	-79	-69	-68	-70	-70
	K2P	7.50	9.00	11.00	11.50	11.50	12.00	13.00	14.00	14.50	17.00
		-68	-70	-73	-71	-74	-78	-61	-83	-66	-72
	Р	6.00	7.00	7.00-	8.00	8.50	9.00	11.00-	11.00	11.50	13.50
		-70	-70	74	-73	-71	-79	67	-68	-69	-54

https://doi.org/10.1371/journal.pone.0301499.t003

delineation of cryptic species, where cryptic species diversity is being revealed by the development and application of DNA classification methods. This study found large genetic distances between the specimens that were identified as the same species, suggesting the presence of cryptic species.

Currently, it is widely believed that classification should reflect phylogeny [38]. The phylogenetic results obtained here are consistent with Normark et al., [38], with species mainly divided into the Diaspidinae and Aspidiotinae with the same level of strong statistical support observed for the species in both molecular phylogenetic trees. However, *P. centreesa* comprised a clade separate from the genus *Pseudaulacaspis*, and from a morphological perspective, the biggest difference between *P. centreesa* and other species of the genus *Pseudaulacaspis* is that its glandular spines are much longer. So, this species may not belong to the *Pseudaulacaspis* genus.

The main difference in morphology between *P. celtis* and *P. cockerelli* is that *P. celtis* has gland tubercles on the mesothorax and fewer dorsal macroducts (there are four pairs in *P. celtis* and five *pairs* in *P. cockerelli* in the 2nd stage female) [20]. However, *P. celtis* and some *P. cockerelli* group together in the phylogenetic analyses. Possible reasons for this include the fact that genes are extracted from female adults which cannot be compared with immature or male morphologies; and the presence or absence of dorsal ducts and the number of dorsal ducts are not the main distinguishing features in the submedian area of the second abdominal segment of *P. celtis*. Phylogenies also revealed that *Aspidiotinae*, *Chrysomphalus*, *Aonidiella* and *Hemiberlesia* are non-monophyletic and overlaps, forming a single clade. This suggests additional molecular and morphological work is needed for these genera.

The purpose of this study was to supplement the DNA barcoding library of armored scale insects in China, which can greatly improve the identification of scale insects, including immature and male scale insects. The results showed that *COI*-based DNA barcoding is a rapid and accurate technique for the identification of armored scale insect species. However, at present, *COI*-based DNA barcoding cannot be used independent of morphological analysis, and the two techniques should be combined for the accurate and efficient identification of armored scale insect species. Additionally, multi-molecular markers can also be used for species identification to improve accuracy.

Conclusion

This work increased the number of *COI* sequences available for common armored scale insects from China by adding 239 *COI* sequences from 50 morphological species representing 21 genera and two subfamilies. Analyses employed the Automated Barcode Gap Discovery, Assemble Species by Automatic Partitioning, Taxon DNA analysis and Bayesian Poisson Tree Process methods, which yielded largely consistent results, indicating a robust and credible species delimitation. Based on these results, an intergeneric threshold of \leq 5% is recommended for the identification of the Diaspididae in China. However, individual morphospecies may exist with cryptic species and more work is needed to elucidate these issues. Therefore, this study provides novel insights into the identification of armored scale insects in China and provides a DNA barcode library for future research and application.

Supporting information

S1 Table. Sample list and collection information of the 239 specimens, including accession numbers in NCBI (https://www.ncbi.nlm.nih.gov/), specimen ID, host plant, and collect information. OR544511 and OR544512 are outgroups downloaded from GenBank. (XLSX)

S2 Table. The intraspecific and interspecific genetic distances of congeneric species of armored scale insects. The genetic distances were calculated based on the K2P model. The intraspecific genetic distances were calculated when a species had at least two individuals. The interspecific genetic distances were calculated when a genus had at least two species. (XLSX)

S3 Table. Results of the automatic barcode gap discovery (ABGD) analyses. (XLSX)

S4 Table. Results of the assemble Species by Automatic Partitioning (ASAP) analyses. (XLSX)

S5 Table. Results of the cluster method. (XLSX)

S1 Fig. Results of the bayesian Poisson Tree Process (bPTP) analyses. (JPG)

Author Contributions

Conceptualization: Minmin Niu, Yubo Liu.

Data curation: Minmin Niu, Linjia Xue.

Funding acquisition: Minmin Niu, Bo Cai, Qing Zhao, Jiufeng Wei.

Investigation: Bo Cai.

Resources: Linjia Xue.

Writing - original draft: Minmin Niu, Yubo Liu.

Writing - review & editing: Minmin Niu, Jiufeng Wei.

References

- Hebert PDN, Cywinska A, Ball SL, Dewaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences. 2003; 270(1512): 313–321. https://doi.org/10. 1098/rspb.2002.2218 PMID: 12614582
- Hebert PDN, Ratnasingham S, Dewaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B: Biological Sciences. 2003; 270(suppl_1): S96–S99. https://doi.org/10.1098/rsbl.2003.0025 PMID: 12952648
- Song C, Lin XL, Wang Q, Wang XH. DNA barcodes successfully delimit morphospecies in a superdiverse insect genus. Zoologica Scripta. 2018; 47(3): 311–324. https://doi.org/10.1111/zsc.12284
- Evangelista D, Buss L, Ware JL. Using DNA barcodes to confirm the presence of a new invasive cockroach pest in New York City. Journal of Economic Entomology. 2013; 106(6): 2275–2279. <u>https://doi.org/10.1603/ec13402 PMID: 24498724</u>
- Madden MJL, Young RG, Brown JW, Miller SE, Frewin AJ, Hanner RH. Using DNA barcoding to improve invasive pest identification at US ports-of-entry. PLoS ONE. 2019; 14(9): e0222291. <u>https:// doi.org/10.1371/journal.pone.0222291 PMID: 31527883</u>
- Jurado-Rivera JA, Vogler AP, Reid CA, Petitpierre E, Gomez-Zurita J. DNA barcoding insect–host plant associations. Proceedings of the Royal Society B: Biological Sciences. 2009; 276(1657): 639–648. https://doi.org/10.1098/rspb.2008.1264 PMID: 19004756
- García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ. Tropical plantherbivore networks: reconstructing species interactions using DNA barcodes. PLoS ONE. 2013; 8(1): e52967. https://doi. org/10.1371/journal.pone.0052967 PMID: 23308128

- Krishnamurthy PK, Francis RA. A critical review on the utility of DNA barcoding in biodiversity conservation. Biodiversity and conservation. 2012; 21(8): 1901–1919. <u>https://doi.org/10.1007/s10531-012-0306-2</u>
- Kress WJ, García-Robledo C, Uriarte M, Erickson DL. DNA barcodes for ecology, evolution, and conservation. Trends in ecology and evolution. 2015; 30(1): 25–35. https://doi.org/10.1016/j.tree.2014.10. 008 PMID: 25468359
- Janzen DH, Burns JM, Cong Q, Hallwachs W, Dapkey T, Manjunath R, et al. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. Proceedings of the National Academy of Sciences. 2017; 114(31): 8313–8318. <u>https://doi.org/10.1073/pnas.1621504114</u> PMID: 28716927
- Schenk J, Kleinbölting N, Traunspurger W. Comparison of morphological, DNA barcoding, and metabarcoding characterizations of freshwater nematode communities. Ecology and evolution. 2020; 10(6): 2885–2899. https://doi.org/10.1002/ece3.6104 PMID: 32211163
- Liu J, Liu J, Shan YX, Ge XJ, Burgess KS. The use of DNA barcodes to estimate phylogenetic diversity in forest communities of southern China. Ecology and Evolution. 2019; 9(9): 5372–5379. <u>https://doi.org/10.1002/ece3.5128 PMID: 31110686</u>
- Dincă V, Dapporto L, Somervuo P, Vodă R, Cuvelier S, Gascoigne-Pees M, et al. High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. Communications Biology. 2021; 4(1): 315. https://doi.org/10.1038/s42003-021-01834-7 PMID: 33750912
- Yang CQ, Lv Q, Zhang AB. Sixteen years of DNA barcoding in China: What has been done? What can Be done? Frontiers in Ecology and Evolution. 2020; 8(57): 1–13. https://doi.org/10.3389/fevo.2020. 00057
- García Morales M, Denno BD, Miller DR, Miller GL, Ben-Dov Y, Hardy NB. ScaleNet: a literature-based model of scale insect biology and systematics. Database. 2016; 2016: bav118. <u>https://doi.org/10.1093/ database/bav118 PMID: 26861659</u>
- Smith-Pardo AH, Evans GA, Dooley JW. A review of the genus *Chrysomphalus Ashmead* (Hemiptera: Coccoidea: Diaspididae) with descriptions of a new species and a new, related genus. Zootaxa. 2012; 3570(1): 1–24. https://doi.org/10.11646/zootaxa.3570.1.1
- Wei JF, Schneider SA, Normark RD, Normark BB. Four new species of Aspidiotini (Hemiptera, Diaspididae, Aspidiotinae) from Panama, with a key to Panamanian species. ZooKeys. 2021; 1047: 1–25. https://doi.org/10.3897/zookeys.1047.68409 PMID: 34248364
- Henderson RC. Diaspididae (Insecta: Hemiptera: Coccoidea). Fauna of New Zealand. 2011; 66: 275pp. https://doi.org/10.7931/J2/FNZ.66
- Miller DR, Davidson JA. Armored scale insect pests of trees and shrubs (Hemiptera: Diaspididae). Cornell University Press. 2005: 422 pp.
- Tang FT. The scale insects of horticulture and forest of China. Vol III. Shanxi Agricultural University Press Taigu, Shanxi. 1986: 305 pp.
- Miller DR, Miller GL, Hodges GS, Davidson JA. Introduced scale insects (Hemiptera: Coccoidea) of the United States and their impact on US agriculture. Proceedings of the entomological Society of Washington. 2005; 107(1): 123–158. https://doi.org/10.1520/STP11985S
- Ahmed MZ, Moore MR, Rohrig EA, McKenzie CL, Liu Di, Feng JN, et al. Taxonomic and identification review of adventive *Fiorinia* Targioni Tozzetti (Hemiptera, Coccomorpha, Diaspididae) of the United States. ZooKeys. 2021; 1065: 141–203. <u>https://doi.org/10.3897/ZOOKEYS.1065.69171</u> PMID: 36452345
- 23. Gwiazdowski RA, Vea IM, Andersen JC, Normark BB. Discovery of cryptic species among North American pine-feeding *Chionaspis* scale insects (Hemiptera: Diaspididae). Biological Journal of the Linnean Society. 2011; 104(1): 47–62. https://doi.org/10.1111/j.1095-8312.2011.01716.x
- Shin S, Jung S, Heller K, Menzel F, Hong TK, Shin JS, et al. DNA barcoding of Bradysia (Diptera: Sciaridae) for detection of the immature stages on agricultural crops. Journal of Applied Entomology. 2015; 139(8): 638–645. https://doi.org/10.1111/jen.12198
- 25. Park DS, Suh SJ, Oh HW, Hebert PD. Recovery of the mitochondrial *COI* barcode region in diverse Hexapoda through tRNA-based primers. BMC genomics. 2010; 11: 423. <u>https://doi.org/10.1186/1471-2164-11-423</u> PMID: 20615258
- Park DS, Suh SJ, Hebert PD, Oh HW, Hong KJ. DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). Bulletin of entomological research. 2011; 101(4): 429–434. https://doi.org/10.1017/S0007485310000714 PMID: 21272395

- Deng J, Wang XB, Yu F, Zhou QS, Bernardo U, Zhang YZ, et al. Rapid diagnosis of the invasive wax scale, *Ceroplastes rusci* Linnaeus (Hemiptera: Coccoidea: Coccidae) using nested PCR. Journal of Applied Entomology. 2015; 139(4): 314–319. https://doi.org/10.1111/jen.12155
- Oliveira PV, Santos AR, Olive EL, Britto KB, Almeida FAN, Silva VC, et al. Molecular Species Delimitation Using CO/Barcodes of Mealybugs (Hemiptera: Pseudococcidae) from Coffee Plants in Espírito Santo, Brazil. Diversity. 2023; 15(2): 305. https://doi.org/10.3390/d15020305
- Wang XB, Deng J, Zhang JT, Zhou QS, Zhang YZ, Wu SA. DNA barcoding of common soft scales (Hemiptera: Coccoidea: Coccidae) in China. Bulletin of entomological research. 2015; 105(5): 545– 554. https://doi.org/10.1017/S0007485315000413 PMID: 25989705
- Wang XB, Zhang JT, Deng J, Zhou QS, Zhang YZ, Wu SA. DNA barcoding of mealybugs (Hemiptera: Coccoidea: Pseudococcidae) from mainland China. Annals of the Entomological Society of America. 2016; 109(3): 438–446. https://doi.org/10.1093/aesa/saw009
- Sethusa MT, Millar IM, Yessoufou K, Jacobs A, Bank MVD, Bank HVD. DNA barcode efficacy for the identification of economically important scale insects (Hemiptera: Coccoidea) in South Africa. African Entomology. 2014; 22(2): 257–266. https://hdl.handle.net/10520/EJC155711
- Amouroux P, Crochard D, Germain JF, Correa M, Ampuero J, Groussier G, et al. Genetic diversity of armored scales (Hemiptera: Diaspididae) and soft scales (Hemiptera: Coccidae) in Chile. Scientific reports. 2017; 7(1): 2014. https://doi.org/10.1038/s41598-017-01997-6 PMID: 28515435
- Ren JM, Ashfaq M, Hu XN, Ma J, Liang F, Hebert PDN, et al. Barcode index numbers expedite quarantine inspections and aid the interception of nonindigenous mealybugs (Pseudococcidae). Biological Invasions. 2018; 20: 449–460. https://doi.org/10.1007/s10530-017-1546-6
- Ramasubramanian T, Ramaraju K, Nirmala R. COI gene-based species diagnostic kit for sugarcane scale insect, Melanaspis glomerata (Green) (Homoptera: Diaspididae). Sugar Tech. 2016; 18: 441– 446. https://doi.org/10.1007/s12355-015-0394-x
- Latina RA, Lantican DV, Guerrero MS, Rubico EC, Laquinta JF, Caoili BL. Species-specific PCR-based marker for rapid detection of *Aspidiotus rigidus* Reyne (Hemiptera: Diaspididae). Journal of Asia-Pacific Entomology. 2022; 25(1): 101848. https://doi.org/10.1016/j.aspen.2021.101848
- Campbell AM, Lawrence AJ, Hudspath CB, Gruwell ME. Molecular identification of Diaspididae and elucidation of non-native species using the genes 28s and 16s. Insects. 2014; 5(3): 528–538. <u>https://doi.org/10.3390/insects5030528 PMID: 26462823</u>
- Deng J, Yu F, Zhang TX, Hu HY, Zhu CD, Wu SA, et al. DNA barcoding of six *Ceroplastes* species (Hemiptera: Coccoidea: Coccidae) from China. Molecular Ecology Resources. 2012; 12(5): 791–796. https://doi.org/10.1111/j.1755-0998.2012.03152.x PMID: 22548741
- Normark BB, Okusu A, Morse GE, Peterson DA, Itioka T, Schneider SA. Phylogeny and classification of armored scale insects (Hemiptera: Coccomorpha: Diaspididae). Zootaxa. 2019; 4616(1): 1–98. https://doi.org/10.11646/zootaxa.4616.1.1 PMID: 31716328
- Williams DJ, Watson GW. The Scale Insects of the Tropical South Pacific Region. Pt. 1. The Armoured Scales (Diaspididae). CAB International Wallingford, U.K. 1988; 290 pp.
- 40. Tang FT. The scale insects of horticulture and forest of China. Vol. I. The Institute of Gardening, Forestry Science of Shenyang. Liaoning, China. 1977; 259 pp.
- Tang FT. The scale insects of horticulture and forests of China. Volume II. Shanxi Agricultural University Press, Taigu, Shanxi. 1984; 115pp.
- **42.** Chou I. Monograph of the Diaspididae of China. Vol. 1. Shaanxi Science and Technology Press. 1982; 1–195 +6.
- **43.** Chou I. Monograph of the Diaspididae of China. Vol. 2. Shaanxi Science and Technology Press. 1985; 196–432 +9.
- **44.** Chou I. Monograph of the Diaspididae of China. Vol. 3. Shaanxi Science and Technology Press 1986; 433–771 + 9.
- Ewert S, Meinard M. Chroma Toolbox: MATLAB implementations for extracting variants of chromabased audio features. International Society for Music Information Retrieval Conference. 2011.
- 46. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular plant. 2020; 13(8): 1194–1202. https://doi.org/ 10.1016/j.molp.2020.06.009 PMID: 32585190
- Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Molecular biology and evolution. 2021; 38(7): 3022–3027. https://doi.org/10.1093/molbev/msab120 PMID: 33892491
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution. 1980; 16: 111–120. https://doi.org/10. 1007/BF01731581 PMID: 7463489

- Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular ecology resources. 2020; 20(1): 348–355. <u>https://doi.org/10.1111/1755-0998.13096</u> PMID: 31599058
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14(6): 587–589. <u>https://doi.org/10.1038/nmeth.4285 PMID: 28481363</u>
- Letunic I, Bork P. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research. 2021; 49(W1): W293–W296. <u>https://doi.org/10.1093/nar/gkab301</u> PMID: 33885785
- Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Molecular ecology. 2012; 21(8): 1864–1877. https://doi.org/10.1111/j.1365-294X. 2011.05239.x PMID: 21883587
- Zhang JJ, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 2013; 29(22): 2869–2876. https://doi.org/10.1093/ bioinformatics/btt499 PMID: 23990417
- 54. Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Systematic biology. 2013; 62(5): 707–724. https://doi.org/10.1093/sysbio/syt033 PMID: 23681854
- 55. Puillandre N, Brouillet S, Achaz G. ASAP: assemble species by automatpartitioning. Molecular Ecology Resources. 2021; 21(2): 609–620. https://doi.org/10.1111/1755-0998.13281 PMID: 33058550
- 56. Meier R, Shiyang K, Vaidya G, Ng PK. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology. 2006; 55(5): 715–728. <u>https://doi.org/10.1080/10635150600969864</u> PMID: 17060194
- Schwarzfeld MD, Sperling FAH. Comparison of five methods for delimitating species in *Ophion* Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae). Molecular phylogenetics and evolution. 2015; 93: 234–248. https://doi.org/10.1016/j.ympev.2015.08.003 PMID: 26265257
- Tembe S, Shouche Y, Ghate HV. DNA barcoding of Pentatomomorpha bugs (Hemiptera: Heteroptera) from western ghats of India. Meta Gene. 2014; 2: 737–745. <u>https://doi.org/10.1016/j.mgene.2014.09</u>. 006 PMID: 25606457
- Yin Y, Yao LF, Hu Y, Shao ZK, Hong XY, Hebert PDN, et al. DNA barcoding uncovers cryptic diversity in minute herbivorous mites (Acari, Eriophyoidea). Molecular Ecology Resources. 2022; 22(5): 1986– 1998. https://doi.org/10.1111/1755-0998.13599 PMID: 35178894
- 60. Pešić V, Jovanović M, Oliveira AE, Pedro A, Freira M, Morais MM. New records of water mites (Acari, Hydrachnidia) from Portugal revealed by DNA barcoding, with the description of *Atractides marizae* sp. nov.. ZooKeys. 2023; 1151: 205–222. https://doi.org/10.3897/zookeys.1151.100766 PMID: 37214175
- Pinto IdS, Rodrigues BL, de Araujo-Pereira T, Shimabukuro PHF, de Pita-Pereira D, Britto C, et al. DNA barcoding of sand flies (Diptera, Psychodidae, Phlebotominae) from the western Brazilian Amazon. PLoS ONE. 2023; 18(2): e0281289. https://doi.org/10.1371/journal.pone.0281289 PMID: 36730314
- Tsoupas A, Papavasileiou S, Minoudi S, Gkagkavouzis K, Petriki O, Bobori D, et al. DNA barcoding identification of Greek freshwater fishes. PLoS ONE. 2022; 17(1): e0263118. https://doi.org/10.1371/ journal.pone.0263118 PMID: 35081163
- **63.** Pramual P, Jomkumsing P, Wongpakam K, Wongwian P. DNA barcoding of tropical black flies (Diptera: Simuliidae) in Thailand: One decade of progress. Acta Tropica. 2021; 224: 106116. https://doi.org/10. 1016/j.actatropica.2021.106116 PMID: 34464589
- Collins R, Cruickshank R. The seven deadly sins of DNA barcoding. Molecular ecology resources. 2013; 13(6): 969–975. https://doi.org/10.1111/1755-0998.12046 PMID: 23280099
- 65. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (<u>http://www.barcodinglife.org</u>). Molecular Ecology Notes. 2007; 7(3): 355–364. <u>https://doi.org/10.1111/j.1471-8286.2007.01678.x</u> PMID: 18784790
- Ball S, Armstrong KF. Using DNA barcodes to investigate the taxonomy of the New Zealand sooty beech scale insect. New Zealand Department of Conservation. 2007: 9–14. <u>https://hdl.handle.net/</u> 10182/1636