Annals of the Entomological Society of America, 109(3), 2016, 438–446
doi: 10.1093/aesa/saw009
Advance Access Publication Date: 27 March 2016
Research article

Genetics

DNA Barcoding of Mealybugs (Hemiptera: Coccoidea: Pseudococcidae) From Mainland China

Xu-Bo Wang,[1] Jiang-Tao Zhang,[1] Jun Deng,[1] Qing-Song Zhou,[3,4] Yan-Zhou Zhang,[3] and San-An Wu[1,2]

[1]The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, No.35 Tsinghua East Rd., Beijing, China 100083 (Wangxb_bjfu@163.com; 2469120608@qq.com; tiancai1282008@126.com; sananwu@bjfu.edu.cn), [2]Corresponding author, e-mail: sananwu@bjfu.edu.cn, [3]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Rd., Beijing, China 100101 (qinsn@qq.com; zhangyz@ioz.ac.cn), and [4]University of Chinese Academy of Sciences (UCAS), No.19A Yuquan Rd., Beijing 100049, China

Received 17 July 2015; Accepted 2 February 2016

Abstract

Mealybugs (Hemiptera: Pseudococcidae) are an insect group that feeds on plant sap; many are major pests of ornamental plants and crops worldwide. The difficulty of morphological identification of mealybugs points to a need for a rapid and effective identification method, like DNA barcoding, to assist morphological taxonomy. Here, we employed diverse methods (best close match [BCM], Neighbor-Joining [NJ] tree, Barcoding with LOGic formulas [BLOG], Poisson Tree Process [PTP] Species Delimitation Method) to test the efficiency of two molecular markers (mitochondrial cytochrome c oxidase I [COI] and large ribosomal subunit gene [28S]) that could be used for species identification of 54 mealybug species that commonly occur in China. Two hundred six COI barcodes (47 species) and 242 28S sequences (53 species) were recovered from 246 individuals. In both the COI and 28S data sets, species except for *Planococcus citri* and *P. minor* were unambiguously identified by all the methods. The PTP analysis based on COI sequences generated more putative species in *Antonina tesquorum*, *Atrococcus paludinus*, and *Formicococcus* sp. than morphological identification. Among these three cases, the sequences of *At. paludinus* showed 3.55% variation at the 28S locus, possibly reflecting cryptic diversity in this taxon. Our study corroborates the utility of the COI and 28S genes in the rapid identification of mealybugs, and the barcode library we provide will create an effective identification system for mealybug pest management in China.

Key words: 28S, COI barcode, PTP, mealybug identification, scale insect

Mealybug is the common name of a soft-bodied insect belonging to the family Pseudococcidae (Hemiptera: Coccoidea) that is usually coated in white powdery wax with lateral wax filaments (Downie and Gullan 2004, Williams 2004). This family currently includes 270 genera and about 2,000 described species worldwide (Downie and Gullan 2004, Ben-dov et al. 2015), with at least 151 species distributed in mainland China (Wu 2009). Mealybugs are mostly phloem-sucking plant parasites, rendering many economic plants unmarketable by causing physical and aesthetic damage to hosts, as well as transmitting destructive plant viruses to infested crops (Ben-Dov 1995, Hardy et al. 2008, Bethke and Cloyd 2009). In addition, excreted honeydew of mealybugs can allow the growth of sooty molds, further exacerbating the decline of host plants (Downie and Gullan 2004). For these reasons, many mealybugs are identified as agricultural pests; examples include the papaya mealybug (*Paracoccus marginatus* Williams and Granara de Willink), the long-tailed mealybug [*Pseudococcus longispinus* (Targioni-Tozzetti)], and the citrus mealybug [*Planococcus citri* (Risso)] (Miller et al. 2002, 2005). In mainland China, 40 mealybugs have been reported as pests of economic plants (Wu 2009). Moreover, mealybugs can create serious problems when introduced into new areas without natural predators, competitors, or parasites (Miller et al. 2002, 2005). Among the five most frequently intercepted scale insects on fruits from the Association of Southeast Asian Nations, four are mealybugs (Gu et al. 2013).

Traditional methods for distinguishing mealybug species are based on the examination of morphological traits of adult females mounted on slides. This is time-consuming and particularly difficult for nonspecialists, and is impractical for eggs, larvae, and adult males (Malausa et al. 2010, Beltrà et al. 2012). Even when keys are available, some related species are hard to identify to the species level, owing to a high degree of morphological similarity (Tang 1992, Williams 2004, Malausa et al. 2010). Additionally, the morphological characteristics of mealybugs are often influenced by environmental factors (Cox 1983, Charles et al. 2000, Hodgson et al. 2008), making traditional identification still more complicated and

difficult. Application of a rapid and effective identification method to complement morphological taxonomy of mealybugs would therefore be useful.

DNA barcoding is an appropriate alternative to morphological identification that has proven to be a reliable and rapid identification tool for many animals (Hebert et al. 2003, 2004a; Ward et al. 2005; Bucklin et al. 2011), including a diverse range of insect taxa (Hajibabaei et al. 2006, Park et al. 2011a, Hendrich et al. 2015). This approach employs partial cytochrome c oxidase subunit I (COI) sequences (DNA barcodes) to assign unidentified specimens to known species (Hebert et al. 2003). In many studies, DNA barcoding has served as a powerful tool for the identification of scale insect species (Ball and Armstrong 2007, Malausa et al. 2010, Abd-Rabou et al. 2012, Deng et al. 2012), including native mealybug pests and others intercepted from different countries (Park et al. 2011b, Sethusa et al. 2014). Furthermore, DNA barcodes have been successfully used to identify mealybug nymphs (Park et al. 2011b, Beltrà et al. 2012) and to uncover intraspecific variation, and as evidence for cryptic diversity in mealybugs (Abd-Rabou et al. 2012, Park et al. 2011b). One of the key factors for the successful application of DNA barcoding is the availability of reliable sequence reference libraries, like the Barcode of Life Data System (BOLD, Ratnasingham and Hebert 2007). An incomplete barcode library can lead to ambiguous identification. Although the COI barcodes of 52 mealybug species are available in the BOLD system (http://boldsystems.org/), the permanent addition of new barcode data is still essential to increasing its taxonomic resolution.

In this study, 246 individual mealybugs were analyzed. To our knowledge, this study is the first large-scale report on DNA barcoding of the Pseudococcidae species distributed in China. We analyzed the COI barcode region and the nuclear 28S D2-D3 region, using tree-based, distance-based, and character-based methods, as well as the Poisson tree process (PTP) species delimitation method, to identify species of Pseudococcidae with the intention of 1) testing the efficiency of species discrimination of the two markers on a wide range of mealybug species and 2) establishing a reliable DNA barcode library of mealybugs in China.

Materials and Methods

Specimen Sampling

This study included 246 mealybug specimens representing 54 species in 23 genera. Sample sizes ranged from 1 to 13 specimens per species, with an average of 4.5 individuals. These specimens were collected from fruit trees, grasses, or ornamental plants in 49 locations within China. Two *Icerya* species, *I. purchasi* Maskell and *I. seychellarum* (Westwood), were chosen as outgroups (Downie and Gullan 2004). Detailed information, including sampling locations, host plants, and collecting dates, is given in Supp. Table 1 (online only; supporting information). Individuals were preserved in 95% ethanol and stored at −20°C until DNA extraction. Morphological identification was performed based on the taxonomic keys from Tang (1992) and Williams (2004). Slide-mounted voucher specimens were deposited in the Insect Collection of Beijing Forestry University (Supp. Table 1 [online only]).

DNA Extraction, Amplification, and Sequencing

A nondestructive protocol, which retains the cuticle of mealybugs for preparation of slides, was applied to extract DNA from individuals. This was essential to enable the vouchering of barcoded specimens (Rowley et al. 2007, Hunter et al. 2008) and to allow morphological re-examination where molecular results were inconsistent with morphology. Total genomic DNA was extracted from single specimens using a DNeasy Blood and Tissue Kit (Qiagen, Dalian, China) following the manufacturer's protocols. The cuticle was preserved in tubes for preparation of slides when pipetting the mixture into a DNeasy Mini spin column.

DNA amplification protocols for COI and 28S followed those of Deng et al. (2012). COI amplifications were performed with two primer pairs: PcoF1 (5′-CCTTCAACTAATCATAAAAATATYAG-3′)/LepR1 (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) (Park et al. 2011b) or C1-1554F (5′-CAGGAATAATAGGAACATC AATAAG-3′)/C1-2342R (5′-ATCAATGTCTAATCCGA TAGTAA ATA-3′) (Deng et al. 2012). DNA extracted from specimens was also processed for species identification using the 28S D2-D3 ribosomal DNA primer set: 28sF3633 (5′-TACCGTGAGGGAAAGTT GAAA-3′; Choudhury and Werren, 2006) and 28s-b (5′-TCGGAAGGAACCAGCTACTA-3′; Whiting et al. 1997). PCR products were visualized on 1% agarose gels, and the most intense products were sequenced bidirectionally using BigDye v3.1 on an ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, CA).

The COI and 28S sequences were aligned using the algorithm FFT-NS-i (Katoh et al. 2002) implemented in MAFFT 7 (Katoh and Standley 2013) and manually edited to adjust the aligned sequences by Bioedit (Hall 1999), resulting in 671 aligned sites for COI (range of sequence length: 617–671 nucleotides) and 940 aligned sites for 28S (range of sequence length: 624–865 nucleotides). All sequences were submitted to GenBank and accession numbers were obtained (Supp. Table 1 [online only]). Alignments of COI sequences were translated to amino acids using MEGA 6 (Tamura et al. 2013) to detect frame-shift mutations, which may indicate the presence of pseudogenes.

Analysis of Molecular Data

In order to allow comparisons between our results and a large part of the literature on DNA barcoding, we used the Kimura 2-parameter (K2P) model (Kimura 1980) to calculate genetic distances within and between species. We also checked for the presence of barcode gaps, which are defined by intraspecific vs. interspecific (nearest neighbor [NN]) genetic distances between species (Meyer and Paulay 2005). Singletons were excluded in the barcode gap analysis. Neighbor-joining trees (Saitou and Nei 1987) of K2P distances were constructed using 1,000 bootstrap replicates in MEGA 6 (Tamura et al. 2013) to provide a graphic representation of the patterning of divergence among species.

Distance-Based Method

The "best close match (BCM)" criteria from Meier et al. (2006) was employed to estimate the proportion of correct matches of COI and 28S data set. BCM relies on a threshold value of sequence similarity. The threshold was estimated by obtaining a frequency distribution of pairwise intraspecific distances and determining the distance below which 95% of all intraspecific distances are found. BCM first identified the best barcode match of a query and then assigned the species name of that barcode to the query only if the distance between query and barcode was below the threshold. All queries without a barcode match below the threshold value remained unidentified. Singletons were excluded from this analysis due to the need for at least two sequences (query and match). SpeciesIdentifier1.7 (Meier et al. 2006) was used to calculate pairwise K2P distances and to quantify the proportion of correct matches according to BCM.

Character-Based Method

A logic data mining method called BLOG (Barcoding with LOGic formulas) was used for character-based identification (Bertolazzi et al. 2009). This method classifies different species in two steps: 1) selecting the most relevant DNA base pairs that are best candidates to distinguish species and 2) extracting the logic formulas that are able to identify a species precisely (Bertolazzi et al. 2009). To evaluate the efficiency of this method, the program BLOG 2.0 (Weitschek et al. 2013) was employed to create classification logic formulas with a training file and apply them to a testing file. The training file included 80% of the sequences (the recommended proportion) and the testing file consisted of the remaining sequences. The BLOG analysis was implemented with input parameters as default values. The logic formula having the lowest false-positive rate on the reference data set was taken as the identification (Bertolazzi et al. 2009). Owing to the need for two sets of sequences (training file and testing file), singletons were removed from the character-based identification. The aligned COI sequences used in the BLOG analysis were trimmed to 589 bp for COI, homologous to the 5′ end of sequence HM474130 (Park et al. 2011b) from 43 to 631 bp. As for 28S, a fasta file of the final alignment was provided in the sup plementary materials for checking the specific classification formulas.

Poisson Tree Process (PTP) Species Delimitation Method

The PTP model (Zhang et al. 2013) is a recently developed method that models speciation or coalescent events relative to the number of substitutions, represented by branch lengths, and uses heuristic algorithms to estimate the most likely classification of branches into population- and species-level processes. The PTP model basically assumes that each substitution has a small probability of generating a speciation, and that the number of substitutions between species is significantly higher than the number of substitutions within species (Zhang et al. 2013). Corresponding analyses were implemented on the bPTP web server (http://species.h-its.org/) with the phylogenetic tree generated from Maximum Likelihood (ML) analyses as the input. We opted to remove distantly related outgroups to improve species delimitation. ML analysis was carried out with RAxML (Stamatakis et al. 2008) implemented in raxmlGUI (Silvestro and Michalak 2012) using rapid bootstrap analysis with 100 replicates and the GTRGAMMA substitution model.

Results

Two hundred six COI barcodes (47 species) and 242 28S sequences (53 species, 28S amplification failed for *Trionymus bambusae* (Green), Supp. Table 1 [online only]) were obtained in this study. Using the COI primer set, we recovered 32 additional sequences from seven species (*Dysmicoccus alazon*, *D. boninsis*, *Ferresia virgata*, *Palmicultor lumpurensis*, *Pseudococcus viburni*, *T. bambusae*, and *T. perrisii*) that possessed two deletions of thymine residues at positions 149–150 in the standard COI barcode region (Supp. Fig. 1 [online only]). These cases represent possible nuclear copies of mitochondrial pseudogenes (known as numts, Lopez et al. 1994), as translation of the COI protein is disrupted by a frame-shift mutation induced by the deletions. They were removed from further analyses and submitted to Genbank without protein notation. The mean nucleotide frequencies of the COI and 28S data sets were A: 0.363, C: 0.103, G: 0.060, T: 0.474 and A: 0.189, C: 0.260, G: 0.332, T: 0.219, respectively.

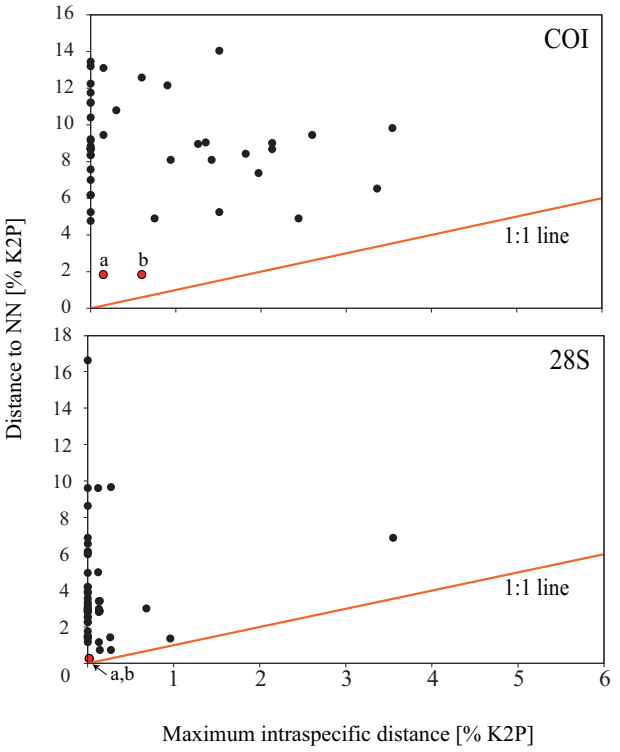

Fig. 1. Barcode gap plot showing the distance to the nearest neighbor (NN) vs. the maximum intraspecific distance for 54 species. Dots above the 1:1 line indicate the presence of a barcode gap. Dots of species exhibiting <2% congeneric divergence are in red: a. *Planococcus minor*, b. *Planococcus citri*. (Online figure in color.)

Genetic Variation and NJ trees

The K2P distances within species, genus, and family levels are summarized in Table 1. Genetic variation increased steadily with taxonomic level, supporting a marked change in genetic divergence at species boundaries. Genetic distances among the 54 species are shown in Table 2 (for details, see Supp. Table 2 [online only]). In the COI data set, values of divergence within species ranged from 0 to 3.54%, while divergence between species ranged from 1.96 to 21.90%. Of the 47 species analyzed, 45 were >2.0% divergent from their nearest neighbor. For two closely related species, *Planococcus citri* and *P. minor*, the distance to the NN was less than 2% (1.96%), but still exceeded the maximum intraspecific value (Fig. 1). A barcode gap was present for all species (Fig. 1) and this also produced high resolution between clusters in the NJ tree, grouping conspecific individuals into a distinct clade with >99% bootstrap support (Fig. 2A). More than 2% divergence was found in six species, *Antonina graminis*, *A. tesquorum*, *Atrococcus paludinus*, *Coccura suwakoensis*, *Crisicoccus matsumotoi*, and *Formicococcus* sp., for which the maximum divergence was 2.13–3.54% (Fig. 2A).

For the 28S data set, intraspecific distances ranged from 0 to 3.55% and interspecific distances ranged from 0.27 to 33.36% (Tables 2 and Supp. Table 2 [online only]). Overall, the 28S sequence divergences within species and between closely related species were lower than those of the COI barcodes (Fig. 1). All analyzed mealybug species had distinct 28S sequences that did not overlap with any other species (Fig. 1). NJ analysis of 28S sequences recovered 53 clades (Fig. 2B). Except for three species, *Coccura suwakoensis*, *P. citri*, and *P. minor*, they were supported in the NJ

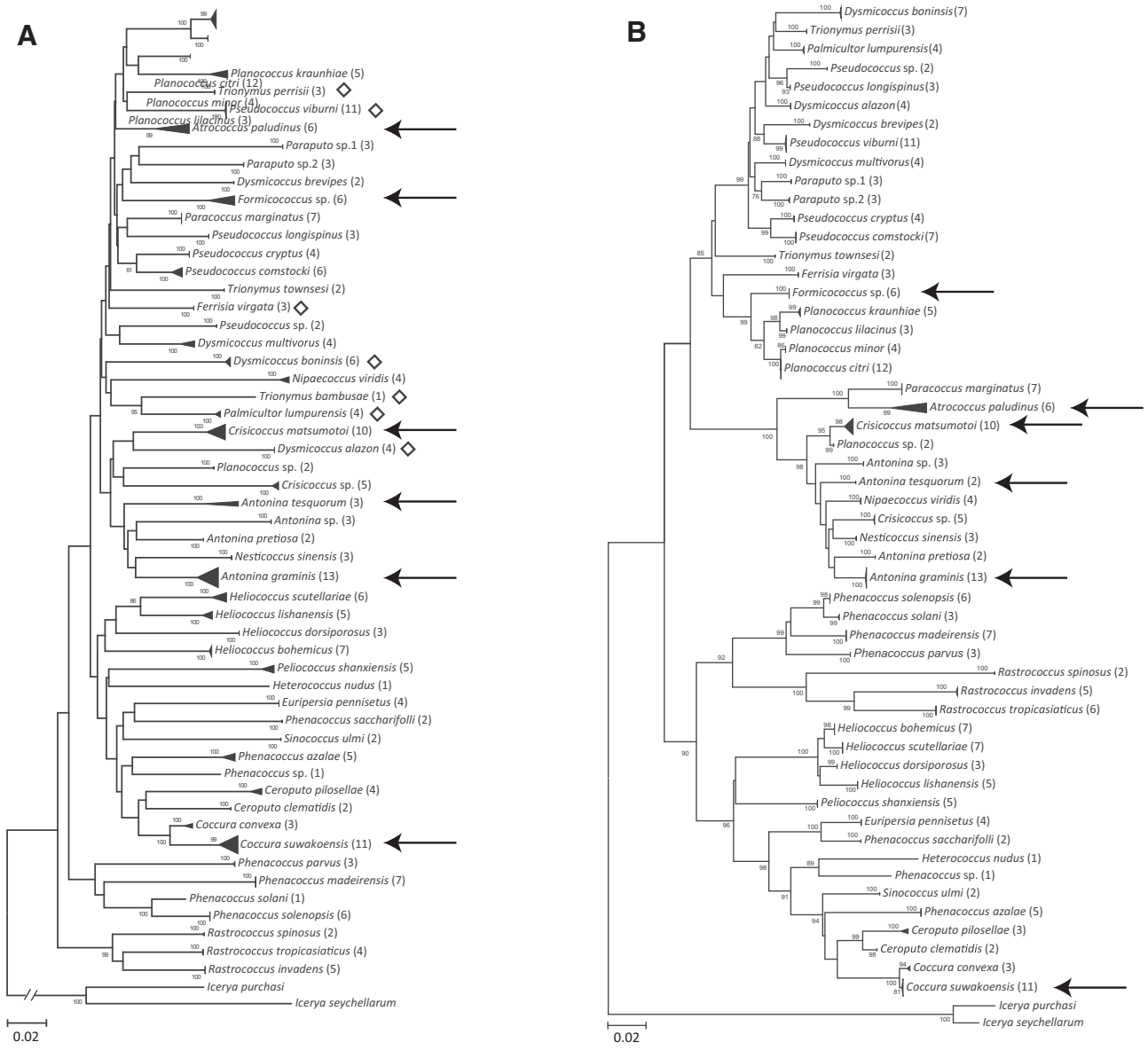

Fig. 2. Neighbor-joining tree of 206 COI barcodes from 47 mealybug species (A) and 242 nuclear 28S sequences from 53 mealybug species (B), using K2P distances. The number of individuals is indicated in parenthesis behind each species name. The node of each species with multiple specimens is collapsed to a vertical line or a triangle. The six instances with >2% COI intraspecific divergence are identified with arrows. Two species of Icerya (Hemiptera: Margarodidae) were chosen as outgroup. Numbers below or above branches refer to nodal support values inferred from NJ bootstrap. Values lower than 75% are hidden. The paralleled lines in the branch of outgroup, referring to 12% genetic distance, are used to cut down the branch length. Sequences with internal thymine residue deletions are labeled with diamonds.

Table 1. K2P pairwise distances (%) of the COI and 28S gene within different taxonomic levels of the investigated mealybugs

Comparisons within	COI			28S		
	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Species, between individuals	0.00	0.54	3.54	0.00	0.10	3.55
Genus, between species	1.96	11.53	19.48	0.27	7.76	21.67
Family, between species of different genus	6.18	14.37	21.90	1.37	16.75	33.36

tree with >90% bootstrap values (Fig. 2B). Among the five species (excluding *A. tesquorum*, see below) that contained relatively high COI variation, only *Formicococcus* sp. was invariant at the 28S gene, while the remaining four species possessed divergences of 0.14–3.55% at the nuclear loci (Fig. 2B, Table 2). The comparison was not conducted for *A. tesquorum*, because the 28S sequence of the specimen collected from Jilin province was not obtained (Supp. Table 1 [online only]).

Table 2 General barcode information and genetic variation (%) of COI barcodes and 28S sequences within (intra) and between (inter) mealybug species included in this study

Species	NC	COI-Intra		COI-Inter		NS	28S-Intra		28S-Inter	
		Min	Max	Min	Max		Min	Max	Min	Max
Antonina graminis	13	0.00	2.13	8.69	17.27	13	0.00	0.14	2.88	30.22
Antonina pretiosa	2	0.00	0.00	8.69	17.18	2	0.00	0.00	3.87	29.63
Antonina sp.	3	0.00	0.00	10.41	21.05	3	0.00	0.00	3.93	29.66
Antonina tesquorum	3	1.36	3.54	9.83	18.92	2	0.00	0.00	3.21	29.11
Atrococcus paludinus	6	0.00	3.36	6.54	16.80	6	0.00	3.55	6.90	33.36
Ceroputo clematidis	2	0.00	0.00	7.58	18.22	2	0.00	0.00	3.02	24.96
Ceroputo pilosellae	4	0.16	1.26	8.97	19.86	3	0.00	0.68	3.02	26.76
Coccura convexa	3	0.00	0.75	4.91	15.84	3	0.00	0.27	0.74	26.73
Coccura suwakoensis	11	0.00	2.44	4.91	17.14	11	0.00	0.14	0.74	26.17
Crisicoccus matsumotoi	10	0.00	2.13	9.02	18.49	10	0.00	0.96	1.37	28.33
Crisicoccus sp.	5	0.00	0.60	12.59	21.52	5	0.00	0.14	3.43	31.28
Dysmicoccus alazon	4	—	—	—	—	4	0.00	0.00	2.82	24.60
Dysmicoccus boninsis	6	—	—	—	—	7	0.00	0.12	5.01	29.36
Dysmicoccus brevipes	2	0.00	0.00	8.70	17.55	2	0.00	0.00	3.58	23.76
Dysmicoccus multivorus	4	0.90	1.82	8.44	16.02	4	0.00	0.00	2.96	24.28
Euripersia pennisetus	4	0.00	0.00	12.26	19.09	4	0.00	0.00	4.20	26.04
Ferrisia virgata	3	—	—	—	—	3	0.00	0.00	6.58	23.05
Formicococcus sp.	6	0.00	2.60	9.46	19.31	6	0.00	0.00	3.36	25.17
Heliococcus bohemicus	7	0.00	0.30	10.80	18.57	7	0.00	0.00	1.50	24.62
Heliococcus dorsiporosus	3	0.00	0.00	11.23	18.85	3	0.00	0.00	1.77	23.90
Heliococcus lishanensis	5	0.00	0.94	8.10	17.01	5	0.00	0.00	2.87	25.77
Heliococcus scutellariae	6	0.00	1.42	8.10	18.82	7	0.00	0.00	1.50	25.13
Heterococcus nudus	1	—	—	12.70	19.02	1	—	—	8.99	27.16
Nesticoccus sinensis	3	0.00	0.00	8.70	18.32	3	0.00	0.00	2.88	31.15
Nipaecoccus viridis	4	0.00	0.90	12.16	21.90	4	0.00	0.00	2.94	29.47
Palmicultor lumpurensis	4	—	—	—	—	4	0.00	0.13	2.82	25.00
Paracoccus marginatus	7	0.00	0.00	6.18	16.04	7	0.00	0.00	6.90	31.64
Paraputo sp.1	3	0.00	0.00	11.21	19.86	3	0.00	0.00	3.00	25.36
Paraputo sp.2	3	0.00	0.00	9.23	18.69	3	0.00	0.13	3.00	24.35
Peliococcus shanxiensis	5	0.00	1.51	14.05	20.85	5	0.00	0.27	9.69	22.91
Phenacoccus azalae	5	0.00	1.35	9.05	19.23	5	0.00	0.00	8.66	29.34
Phenacoccus madeirensis	7	0.00	0.00	11.77	21.90	7	0.00	0.00	4.98	24.11
Phenacoccus parvus	3	0.00	0.15	13.11	19.66	3	0.00	0.00	6.09	24.27
Phenacoccus saccharifolli	2	0.00	0.00	13.28	21.52	2	0.00	0.00	4.20	25.64
Phenacoccus solani	1	—	—	4.77	18.72	3	0.00	0.13	1.17	23.56
Phenacoccus solenopsis	6	0.00	0.00	4.77	19.54	6	0.00	0.00	1.17	23.34
Phenacoccus sp.	1	—	—	9.43	17.70	1	—	—	8.99	26.23
Planococcus citri	12	0.00	0.60	1.96	17.78	12	0.00	0.00	0.27	22.16
Planococcus minor	4	0.00	0.15	1.96	16.70	4	0.00	0.00	0.27	22.54
Planococcus kraunhiae	5	0.15	1.97	7.38	17.67	5	0.00	0.26	1.43	23.77
Planococcus lilacinus	3	0.00	0.00	6.18	15.69	3	0.00	0.00	1.43	22.76
Planococcus sp.	2	0.00	0.00	8.69	17.37	2	0.00	0.00	1.37	27.62
Pseudococcus comstocki	6	0.00	1.51	5.25	16.62	7	0.00	0.00	2.55	24.54
Pseudococcus cryptus	4	0.00	0.00	5.25	16.90	4	0.00	0.00	2.55	24.35
Pseudococcus longispinus	3	0.00	0.00	7.00	16.02	3	0.00	0.00	2.28	24.41
Pseudococcus sp.	2	0.00	0.00	8.85	16.83	2	0.00	0.00	2.28	27.11
Pseudococcus viburni	11	—	—	—	—	11	0.00	0.13	3.42	23.85
Rastrococcus invadens	5	0.00	0.00	8.36	20.04	5	0.00	0.12	9.64	29.97
Rastrococcus spinosus	2	0.15	0.15	9.46	20.54	2	0.00	0.00	16.65	33.36
Rastrococcus tropicasiaticus	4	0.00	0.00	8.36	21.02	6	0.00	0.00	9.64	29.85
Sinococcus ulmi	2	0.00	0.00	13.46	20.40	2	0.00	0.00	6.00	25.55
Trionymus bambusae	1	—	—	—	—	0	—	—	—	—
Trionymus perrisii	3	—	—	—	—	3	0.00	0.00	3.09	24.50
Trionymus townsesi	2	0.00	0.00	9.17	18.70	2	0.00	0.00	6.16	22.49
	238					242				

NC/NS: the number of COI/28S sequences used in this analysis. Min/Max: the minimum/maximum genetic distance value. Species in bold are seven cases with putative COI pseudogene.

Performance of "Best Close Match" (BCM) and BLOG 2.0

For identification with BCM criteria and the BLOG program, the COI data set consisted of 202 sequences (43 species, excluding four singletons), whereas the 28S data set consisted of 240 sequences (51 species, excluding two singletons). Using the 95th percentile of intraspecific distances (COI-2.44%, 28S-0.4%) computed from a pairwise summary as the threshold for the BCM simulation, the correct identification using COI and 28S was 99.50 and 98.75%, respectively. Incorrect or ambiguous identifications were not found in either data set. As for the character-based method, the correct classification rate of the tested data was 94.12% for the COI barcode and 100% for the 28S gene. The logic formulas of all species are listed in Supp. Table 3 (online only).

Species Delimitation Through the PTP Model

Based on 89 COI haplotypes, the PTP method identified 49 putative species from the ML tree (Fig. 3), near the number (47 species) based on morphology. According to the PTP model, COI haplotypes of *P. citri* and *P. minor* clustered together, forming a monophyletic group (Fig. 3). *Antonina tesquorum*, *At. paludinus*, and *Formicococcus* sp. were each separately split into two putative species.

Discussion

According to the multiple methods used in this study, COI barcodes could accurately and effectively identify most mealybug species (45 of 47) collected in China. The results of this study corroborate the utility of COI gene analyses in previous studies of rapid mealybug identification (Park et al. 2011b, Abd-Rabou et al. 2012, Beltrà et al. 2012). Two closely related species (*Planococcus citri* and *P. minor*) were not unambiguously identified by all the methods. Although PTP analysis using COI barcodes well delimited the other species with high possibility values, these two were recognized as a single putative species (Fig. 3), showing the inability of the PTP model to delimit closely related species. These two species were well supported as separate clusters on the NJ tree of COI (Fig. 2A). For *P. citri* and *P. minor*, the distance to the NN was less than 2% (1.96%) for COI, but still exceeded the maximum intraspecific value (Fig. 1), meaning that there was a gap between the intra- and interspecific divergence. Additionally, the genetic distance of 1.96% is near the interspecific divergence between these two species (1.90%) reported in Rung et al. (2008, 2009). Considering that including a large proportion of closely related species would be likely to shrink the barcoding gap, perhaps resulting in less accurate identification through barcoding (Meyer and Paulay 2005, Meier et al. 2006), further study with additional specimens from sister species is needed to gain more specific insights into the intra- and interspecific COI variability of mealybugs. On the other hand, although the 28S gene lacks sufficient variation to resolve some closely related species (Park et al. 2011, Deng et al. 2012), this marker yielded sufficient identification accuracy (>98%) for distance-based and character-based methods for species from China.

Using the COI sequences of 47 mealybug species in this study, the intraspecific divergence of 0–3.54% was lower than that reported in another DNA barcoding study of mealybugs collected from 15 countries, which documented an intraspecific divergence of 0–5.95% (Park et al. 2011b). A possible explanation for our lower intraspecific divergence is the effect of sampling scale (Bergsten et al. 2012), as most mealybug species used in this study have a much wider distribution around the world than sampled here (Ben-Dov et al. 2015).

The presence of nuclear copies of mitochondrial DNA (numts) is considered a serious challenge to DNA barcoding (Song et al. 2008). Numts are nonfunctional fragments of mtDNA inserted into the nuclear genome; they are prevalent across eukaryotes and can be inadvertently coamplified with mitochondrial loci (COI in this study) using conserved primers via conventional polymerase chain reactions (Richly and Leister 2004, Bensasson et al. 2011, Song et al. 2014). Some have suggested that numts will make barcoding unreliable if they are not detected and taken into consideration (Song et al. 2008). We detected 2-bp deletions in every sequence of seven mealybug species (about 13% of the COI data set), of which three (*F. virgata*, *Pa. lumpurensis*, and *Ps. viburni*) were also reported in Park et al. (2011b). These deletions induced a frame-shift mutation, suggestive of numt pseudogenes (Triant and Dewoody 2009). Even though these pseudogenes were excluded from the final analyses, their inclusion did not affect the ability of the barcode to identify other species (Fig. 2A). In such cases, examining a nuclear gene is essential. In the present study, we used the 28S gene, a recommended nuclear marker for DNA barcoding of scale insects (Kondo et al. 2008). Except for *T. bambusae* (from which the 28S sequence was not recovered), the other six species were unambiguously identified according to the 28S NJ tree (Fig. 2B), demonstrating that this marker could be used as a complement to the COI barcode to distinguish mealybug species when putative numts are present.

In addition to providing specific identifications, detecting cryptic biodiversity is an appealing application of DNA barcoding (Hebert et al. 2004b; Janzen et al. 2005). In this study, deep divergence (>2%) of COI sequences was found in six species, of which three (*Antonina tequorum*, *Atrococcus paludinus*, and *Formicococcus* sp.) were each split into two putative species in the PTP analysis, possibly reflecting cryptic diversity in these taxa. However, such cases should be regarded with caution when nuclear loci show no variation. In the case of *Formicococcus* sp., the PTP model based COI haplotypes yielded two putative species, of which the sequences were invariant at the 28S gene (Supp. Table 4 [online only]). It may be that the 28S fragment is too conserved to distinguish the most recently diverged species. For example, two wax scale insects, *Ceroplastes ceriferus* and *C. pseudoceriferus*, have been documented as sharing the same 28S haplotype (Deng et al. 2012). Thus, it is necessary to apply more nuclear gene sequences for reliable species delimitation in cases with no 28S variation accompanied by deep divergence in COI. In contrast, the two groups (Gongzhuling vs. Jiaohe, Supp. Table 4 [online only]) of *Atrococcus paludinus* spilt by PTP showed a 3.55% 28S divergence. The congruent pattern between the two markers in showing distinct divergence may indicate cryptic speciation; an integrated taxonomic approach with more extensive sampling is needed for this to be conclusive (Tan et al. 2009).

While DNA barcoding has been widely used as an identification tool, a comprehensive reference database against which unknown sequences may be compared is critical for the accuracy of species identification (Yao et al. 2010). The present study provides 206 COI barcodes (47 species) and 242 sequences of 28S D2-D3 region (53 species) that will enrich the barcode reference library for mealybugs. The establishment of this reliable reference library will provide the foundation for further work focusing on mealybug management in China. Researchers and practitioners, especially in quarantine departments, could quickly and accurately identify mealybugs based on simple DNA sequence comparisons. Additionally, for species with high potential for becoming pests, like *Phenacoccus parvus*

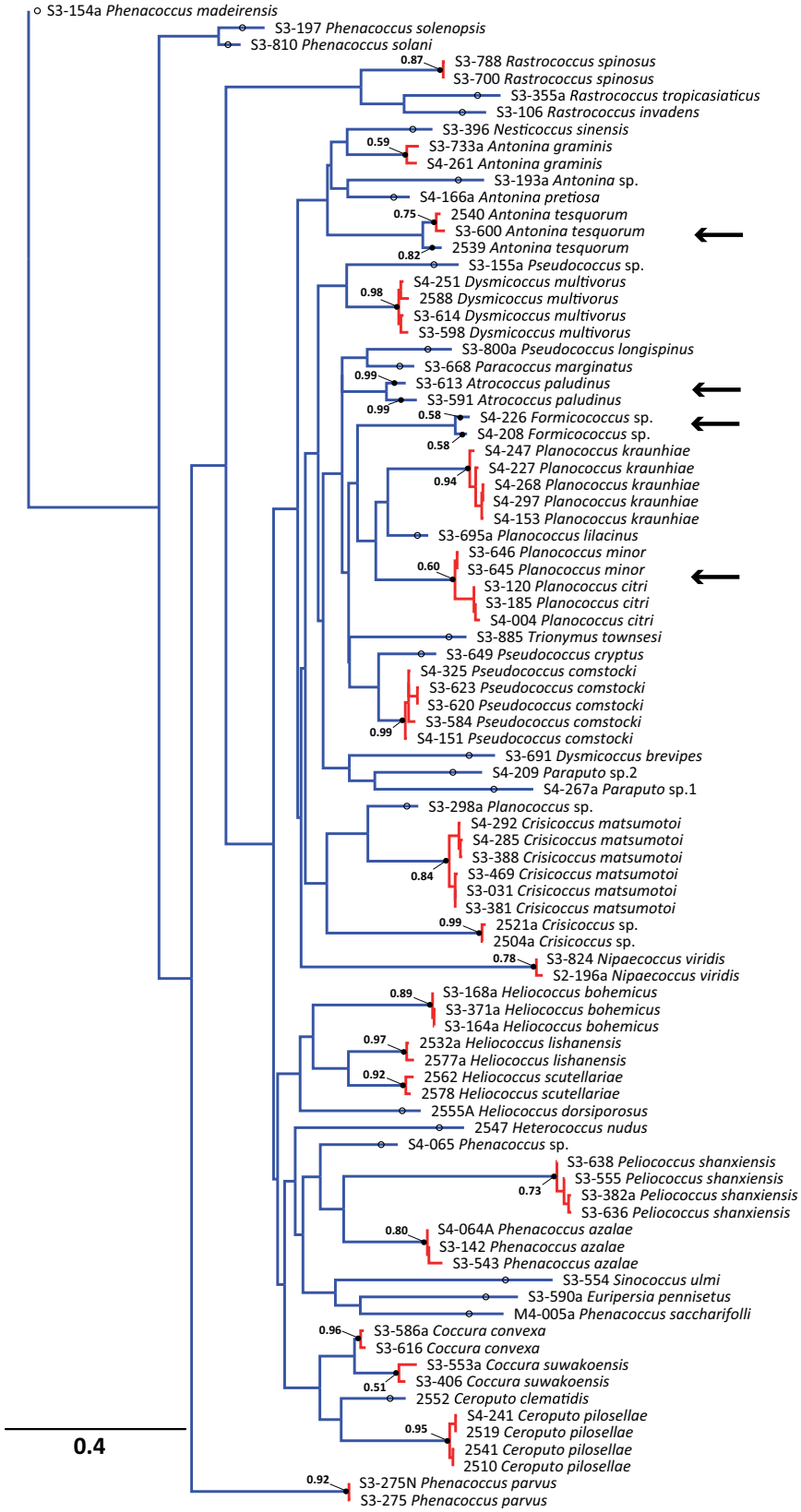

Fig. 3. Putative species delimitation of mealybug species based on the PTP model. Monophyletic groups in red indicate a single putative species as well as terminal branches in blue. Bayesian support values are shown near the branches and support value 1.00 is represented by a hollow circle. Arrows indicate the four cases different from the morphological delimitation. (Online figure in color.)

Morrison (Wang and Wu 2014), rapid identification kits (da Silva et al. 2014) could be designed based on the large set of raw sequences available.

Acknowledgments

We gratefully acknowledge the constructive comments of Editors, and two anonymous referees on an earlier version of the manuscript. We are grateful to the following people who helped us to collect mealybug samples: Fang Yu, Huan-Xi Cao, Qing-Tao Wu, Xiu-Wei Liu, Xu Zhang, Ying Wang (Institute of Zoology, Chinese Academy of Sciences, Beijing), Ju-Pu Chang (Puyang Academy of Forestry, Puyang), Hai-Bin Li, Min Chen, Nan Nan, Shu-Jing Xu, Wen-Chao Li, and Ying-Da Wu (Beijing Forestry University, Beijing), Chen-Yang Xie (University of Technology and Science Beijing), Fu-Zhong Wu (South China Agricultural University), Bo Cai (Hainan Entry-Exit Inspection and Quarantine Bureau), and Na Li (Shanxi Agricultural University). This project was supported by the Fundamental Research Funds for the Central Universities (BLYJ201609) and the National Natural Science Foundation of China (NSFC grant no. 31372151, 31272350).

References Cited

Abd-Rabou, S., H. Shalaby, J. F. Germain, N. Ris, P. Kreiter, and T. Malausa. 2012. Identification of mealybug pest species (Hemiptera: Pseudococcidae) in Egypt and France, using a DNA barcoding approach. Bull. Entomol. Res. 102: 515–523.

Ball, S. L., and K. F. Armstrong. 2007. Using DNA barcodes to investigate the taxonomy of the New Zealand sooty beech scale insect. Science and Technical Publishing, Department of Conservation, Wellington, New Zealand.

Beltrà, A., A. Soto, and T. Malausa. 2012. Molecular and morphological characterisation of Pseudococcidae surveyed on crops and ornamental plants in Spain. Bull. Entomol. Res. 102: 165–172.

Ben-Dov, Y. 1995. A systematic catalogue of the mealybugs of the World (Insecta: Homoptera: Coccoidea: Pseudococcidae, and Putoidae) with Data on Geographical Distribution, Host Plants, Biology, and Economic Importance. Intercept Ltd, Hampshire, United Kingdom.

Ben-Dov, Y., D. L. Miller, and G.A.P. Gibson. 2015. ScaleNet: A database of the scale insects of the world. (http://www.sel.barc.usda.gov/scalenet/scalenet.htm)

Bensasson, D., D. X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends. Ecol. Evol. 16: 314–321.

Bertolazzi, P., G. Felici, and E. Weitschek. 2009. Learning to classify species with barcodes. BMC Bioinformatics 10: S7.

Bethke, J. A., and R. A. Cloyd. 2009. Pesticide use in ornamental production: what are the benefits? Pest Manag. Sci. 65: 345–350.

Bucklin, A., D. Steinke, and L. Blanco-Bercial. 2011. DNA barcoding of marine metazoa. Ann. Rev. Mar. Sci. 3: 471–508.

Charles, J. G., K. J. Froud, and R. C. Henderson. 2000. Morphological variation and mating compatibility within the mealybugs *Pseudococcus calceolariae* and *P. similans* (Hemiptera: Pseudococcidae), and a new synonymy. Syst. Entomol. 25: 285–294.

Choudhury, R., and J. H. Werren. 2006. Unpublished primers. (http://research.amnh.org/FIBR/protocols.html)

Cox, J. M. 1983. An experimental study of morphological variation in mealybugs (Homoptera: Coccoidea: Pseudococcidae). Syst. Entomol. 8: 361–382.

da Silva, V.C.P., A. Bertin, A. Blin, J. F. Germain, D. Bernardi, G. Rignol, M. Botton, and T. Malausa. 2014. Molecular and morphological identification of mealybug species (Hemiptera: Pseudococcidae) in Brazilian vineyards. PLoS ONE 9: e103267.

Deng, J., F. Yu, T. X. Zhang, H. Y. Hu, C. D. Zhu, S. A. Wu, and Y. Z. Zhang. 2012. DNA barcoding of six *Ceroplastes* species (Hemiptera: Coccoidea: Coccidae) from China. Mol. Ecol. Resour. 12: 791–796.

Downie, D. A., and P. J. Gullan 2004. Phylogenetic analysis of mealybugs (Hemiptera: Coccoidea: Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher classification. Syst. Entomol. 29: 238–260.

Gu, Y. J., H. J. Liu, R. R. He, M. T. Wu, K. Chen, and X. N. Hu. 2013. Information analysis of scales on imported fruits from Association of Southeast Asian Nations (In Chinese). Plant Quarantine 27: 95–99.

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs, and P.D.N. Hebert. 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA 103: 968–971.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic. Acids. Symp. Ser. 41: 95–98.

Hardy, N. B., P. J. Gullan, and C. J. Hodgson. 2008. A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. Syst. Entomol. 33: 51–71.

Hebert, P.D.N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B. 270: 313–321.

Hebert, P.D.N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004a. Identification of birds through DNA barcodes. PLoS Biol. 2: 1657–1663.

Hebert, P.D.N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101: 14812–14817.

Hendrich, L., J., Morinière, G. Haszprunar, P. D. Hebert, A. Hausmann, F. Köhler, and M. Balke. 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. Mol. Ecol. Resour. 15: 795–818.

Hodgson, C., G. Abbas, M. J. Arif, S. Saeed, and H. Karar. 2008. *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Coccoidea: Pseudococcidae), an invasive mealybug damaging cotton in Pakistan and India, with a discussion on seasonal morphological variation. Zootaxa 1913: 1–35.

Hunter, S. J., T. I. Goodall, K. A. Walsh, R. Owen, and J. C. Day. 2008. Nondestructive DNA extraction from blackflies (Diptera: Simuliidae): retaining voucher specimens for DNA barcoding projects. Mol. Ecol. Resour. 8: 56–61.

Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P.D.N. Hebert. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Phil. Trans. R. Soc. B. 360: 1835–1845.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30: 772–780.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059–3066.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Lopez, J. V., N. Yuhki, W. Modi, R. Masuda, and S. J. O'Brien. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA in the nuclear genome of the domestic cat. J. Mol. Evol. 39: 171–190.

Malausa, T., A. Fenis, S. Warot, J. F. Germain, N. Ris, E. Prado, M. Botton, F. Vanlerberghe-Masutti, R. Sforza, C. Cruaud, et al. 2010. DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae). J. Appl. Entomol. 135: 142–155.

Meier, R., K. Shiyang, G. Vaidya, and P.K.L. Ng. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst. Biol. 55: 715–728.

Meyer, C. P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3: e422.

Miller, D. R., G. L. Miller, and G. W. Watson. 2002. Invasive species of mealybugs (Hemiptera: Pseudococcidae) and their threat to US agriculture. Proc. Entomol. Soc. Wash. 104: 825–836.

Miller, D. R., G. L. Miller, G. S. Hodges, and J. A. Davidson. 2005. Introduced scale insects (Hemiptera: Coccoidea) of the United States and their impact on US agriculture. Proc. Entomol. Soc. Wash. 107: 123–158.

Park, D. S., R. Foottit, E. Maw, and P. D. Hebert. 2011a. Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). PLoS ONE 6: e18749.

Park, D. S., S. J. Suh, P. D. N. Hebert, H. W. Oh, and K. J. Hong. 2011b. DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). Bull. Entomol. Res. 101: 429–434.

Ratnasingham, S., and P.D.N. Hebert. 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7: 355–364.

Richly, E., and D. Leister. 2004. NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. 21: 1081–1084.

Rowley, D. L., J. A. Coddington, M. W. Gates, A. L. Norrbom, R. A. Ochoa, N. J. Vandenberg, and M. H. Greenstone. 2007. Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. Mol. Ecol. Notes 7: 915–924.

Rung, A., S. J. Scheffer, G. Evans, and D. Miller. 2008. Molecular identification of two closely related species of mealybugs of the genus *Planococcus* (Homoptera: Pseudococcidae). Ann. Entomol. Soc. Am. 101: 525–532.

Rung, A., D. R. Miller, and S. J. Scheffer. 2009. Polymerase chain reaction-restriction fragment length polymorphism method to distinguish three mealybug groups within the *Planococcus citri–P. minor* species complex (Hemiptera: Coccoidea: Pseudococcidae). J. Econ. Entomol. 102: 8–12.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Sethusa, M. T., I. M. Millar, K. Yessoufou, A. Jacobs, M. Van der Bank, and H. Van der Bank. 2014. DNA barcode efficacy for the identification of economically important scale insects (Hemiptera: Coccoidea) in South Africa. Afr. Entomol. 22: 257–266.

Silvestro, D., and I. Michalak. 2012. RaxmlGUI: A graphical front-end for RaxML. Org. Divers. Evol. 12: 335–337.

Song, H., J. E. Buhay, M. F. Whiting, and K. A. Crandall. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. USA 105: 13486–13491.

Song, H., M. J. Moulton, and M. F. Whiting. 2014. Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). PLoS ONE 9: e110508.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57: 758–771.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Tan, D.S.H., Y. Ang, G. S. Lim, M.R.B. Ismail, and R. Meier. 2009. From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). Zool. Scr. 39: 51–61.

Tang, F. D. 1992. The Pseudococcidae of China (Homoptera: Coccinea of Insects). Chinese Agricultural Science Technology Press, Beijing, China.

Triant, D. A., and J. A. Dewoody. 2009. Demography and phylogenetic utility of numt pseudogenes in the southern red-backed vole (*Myodes gapperi*). J. Mammal. 90: 561–570.

Wang, X. B., and S. A. Wu. 2014. A new pest, *Phenacoccus parvus* Morrison (Hemiptera: Coccoidea: Pseudococcidae), in mainland China (In Chinese). Chin. J. Appl. Entomol. 51: 1098–1103.

Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. Hebert. 2005. DNA barcoding Australia's fish species. Phil. Trans. R. Soc. B. 360: 1847–1857.

Weitschek, E., R. Velzen, G. Felici, and P. Bertolazzi. 2013. BLOG 2.0: A software system for character-based species classification with DNA Barcode sequences. What it does, how to use it. Mol. Ecol. Resour. 13: 1043–1046.

Whiting, M. F., J. C. Carpenter, Q. D. Wheeler, and W. C. Wheeler. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46: 1–68.

Williams, D. J. 2004. Mealybugs of Southern Asia. Southdene, Kuala Lumpur, Malaysia.

Wu, S. A. 2009. Checklist and faunistic analysis of scale insect pests (Hemiptera: Coccoidea) in Chinese mainland. J. Beijing For. Univ. 31: 55–63.

Yao, H., J. Song, C. Liu, K. Luo, J. Han, Y. Li, X. Pang, H. Xu, Y. Zhu, P. Xiao, and S. Chen. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5: e13102.