Bulletin of Entomological Research (2015) **105**, 545–554
doi:10.1017/S0007485315000413
© Cambridge University Press 2015

DNA barcoding of common soft scales (Hemiptera: Coccoidea: Coccidae) in China

X.-B. Wang[1,2]†, J. Deng[1]†, J.-T. Zhang[1], Q.-S. Zhou[2], Y.-Z. Zhang[2]* and S.-A. Wu[1]*

[1]The Key Laboratory for Silviculture and Conservation of Ministry of Education, Beijing Forestry University, Beijing 100083, China: [2]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

Abstract

The soft scales (Hemiptera: Coccoidea: Coccidae) are a group of sap-sucking plant parasites, many of which are notorious agricultural pests. The quarantine and economic importance of soft scales necessitates rapid and reliable identification of these taxa. Nucleotide sequences of the mitochondrial cytochrome c oxidase subunit I (COI) gene (barcoding region) and 28S rDNA were generated from 340 individuals of 36 common soft scales in China. Distance-based [(best match, Automated Barcode Gap Discovery (ABGD)], tree-based (neighbor-joining, Bayesian inference), Klee diagrams, and general mixed Yule coalescent (GMYC) models were used to evaluate barcoding success rates in the data set. Best match showed that COI and 28S sequences could provide 100 and 95.52% correct identification, respectively. The average interspecific divergences were 19.81% for COI data and 20.38% for 28S data, and mean intraspecific divergences were 0.56 and 0.07%, respectively. For COI data, multiple methods (ABGD, Klee, and tree-based methods) resulted in general congruence with morphological identifications. However, GMYC analysis tended to provide more molecular operational taxonomic units (MOTUs). Twelve MOTUs derived from five morphospecies (*Rhodococcus sariuoni*, *Pulvinaria vitis*, *Pulvinaria aurantii*, *Parasaissetia nigra*, and *Ceroplastes rubens*) were observed using the GMYC approach. In addition, tree-based methods showed that 28S sequences could be used for species-level identification (except for *Ceroplastes ceriferus* – *Ceroplastes pseudoceriferus*), even with low genetic variation (<1%). This report demonstrates the robustness of DNA barcoding for species discrimination of soft scales with two molecular markers (COI and 28S) and provides a reliable barcode library and rapid diagnostic tool for common soft scales in China.

Keywords: DNA barcoding, soft scales, pest, COI, 28S

(Accepted 09 April 2015; First published online 20 May 2015)

Introduction

The family Coccidae (Hemiptera: Coccoidea), or ‘soft scales’, is the third largest family of scale insects, with more than 1140 species described across approximately 169 genera (Hamon & Williams, 1984, Ben-Dov *et al.*, 2014). Soft scales are an economically important group including notorious agricultural pests such as *Ceroplastes rubens*, *Parasaissetia nigra*, *Saissetia coffeae*, *Saissetia oleae*, and *Coccus hesperidum* (Hamon & Williams, 1984; Gill, 1988). Soft scales suck plant sap and excrete copious honeydew covering the plant surface, which provides a medium for sooty mold (Hamon & Williams, 1984; Tang, 1991; Ben-Dov & Hodgson, 1997). In mainland China, 36 coccid species have been reported as serious pests of crop and ornamental plants (Wu, 2009). Soft scales also cause serious problems as invasive species. Of the 66 soft scales in the

*Authors for correspondence
Phone:
Fax: 86-10-62336596 (S.-A.W), 86-10-64807099 (Y.-Z.Z)
E-mail: sananwu@bjfu.edu.cn, zhangyz@ioz.ac.cn

†These authors contributed equally to this work.

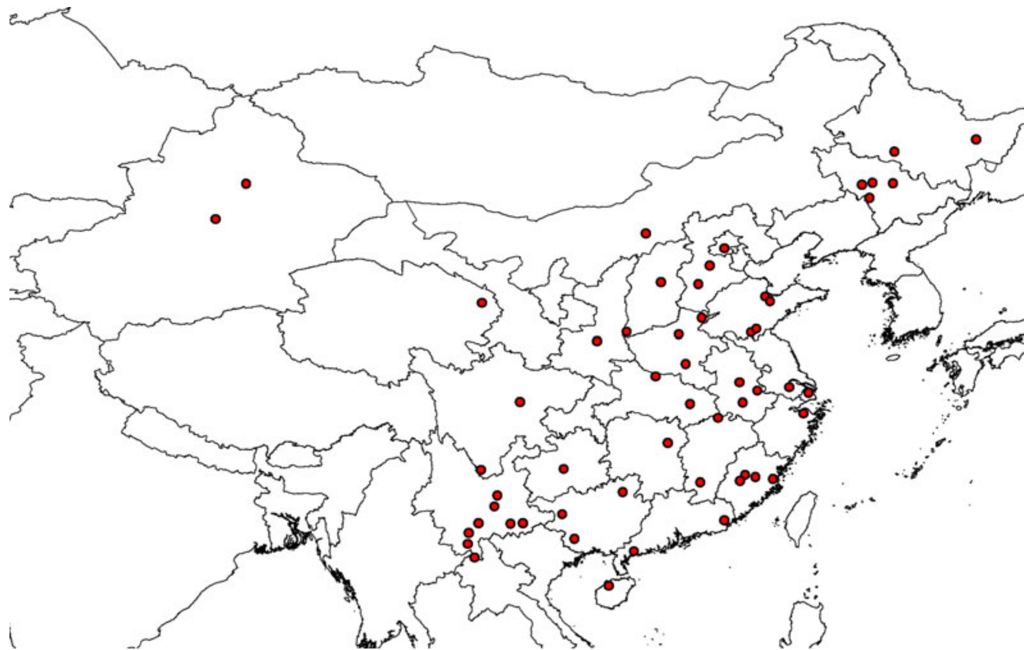

Fig. 1. Overview of geographic distribution of soft scales analyzed in this study. Collection sites are labeled with red circle. Materials from 26 provinces are concluded.

USA, 41 are invasive pests (for example, the fig wax scale *Ceroplastes rusci*, the green coffee scale *Coccus viridis*, and the European fruit lecanium *Parthenolecanium corni*, Miller & Miller, 2003).

Despite their economic importance, Coccidae is considered a difficult group to identify at the species level because of their small size and high degree of similarity. General swelling of the body and sclerotization of the dorsum with increasing maturity of soft scales frequently make identification impossible (Hodgson, 1994). The group also lacks sufficient morphological characteristics to discriminate eggs or larvae at the species level. Only young adult females are available for species delimitation. Unfortunately, this stage is very short and challenging to collect. Meanwhile, intraspecific variation of morphological characteristics such as the stigmatic and dorsal setae is widespread in the soft scales (Gimpel *et al.*, 1974; Gullan & Kosztarab, 1997), making species delimitation more difficult. Traditional identification requires preservation of the adult female cuticle and preparation of slides, resulting in a time-consuming process of identification even for a trained taxonomist. These factors point to a need for a rapid method to effectively identify coccids, especially species common in quarantine work.

DNA barcoding has become a popular tool for species delimitation in vertebrates (Hebert *et al.*, 2004; Wong *et al.*, 2009) and invertebrates (Hajibabaei *et al.*, 2006; Costa *et al.*, 2007; Mikkelsen *et al.*, 2007), which makes it an ideal candidate for accurate and rapid identification of scale insects. However, universal primers fail to amplify the standard barcode region of the COI for any but a few taxa (Kondo *et al.*, 2008; Park *et al.*, 2011). Some recent studies on barcoding of scale insects have gradually expanded the range of application to several families, including Diaspididae, Pseudococcidae, Coccidae, and Margarodidae (Ball & Armstrong, 2007; Malausa *et al.*, 2011; Park *et al.*, 2011; Abd-Rabou *et al.*, 2012; Beltrà *et al.*, 2012; Deng *et al.*, 2012; Sethusa *et al.*, 2014). However, among the 1140 coccid species, only the COI barcode region of 41 (with scientific names) has been submitted to GenBank. This limited information calls for further investigation of the performance of DNA barcoding on a broader scale and development of a more efficient means of DNA barcoding identification in soft scales. Meanwhile, the 28S nuclear gene can identify species in various insect taxa (Campbell *et al.*, 1994; Smith *et al.*, 2008; Monaghan *et al.*, 2009). Although the 28S rDNA lacks sufficient variation to delimitate some species (Park *et al.*, 2011; Deng *et al.*, 2012), it is presently being proposed as a complementary marker to COI in scale insects (Sethusa *et al.*, 2014).

In this study, we sequenced the COI and 28S genes of 340 individuals belonging to 36 common soft scale species in China. The aim of this study was to: (1) explore the efficacy of DNA barcoding in Coccidae using multiple methods and (2) provide a comprehensive barcode library of common soft scales in China.

Materials and methods

Specimen sampling

A total of 340 individual soft scales representing 36 species in 17 genera were used for barcode analysis, with 292 newly collected and 48 from previous barcoding studies (six species of *Ceroplastes*, Deng *et al.*, 2012). The 292 collected specimens were obtained from 22 provinces in China and stored in 95% ethanol at −20°C. Morphological identification was based mainly on the taxonomic keys for Coccidae (Hamon & Williams, 1984; Gill, 1988; Tang, 1991). Slide-mounted voucher specimens were deposited in the Insect Collection of Beijing Forestry University. Details of collection including sampling locations, host plants, and date are available in Supplementary Table S1. The geographical distributions of sampling locations are provided in fig. 1.

DNA extraction, amplification, and sequencing

Total genomic DNA was extracted from each individual using DNeasy Blood & Tissue Kit (Qiagen, Dalian, China) following the manufacturer's protocols. Amplification of COI and 28S were performed in 50 μl reactions using the respective primer pairs: C1-1554F (5′-CAGGAATAATAGGAACATC AATAAG-3′)/C1-2342R (5′- ATCAATGTCTAATCCGAT AGTAAATA-3′, Deng *et al.*, 2012), and 28sF3633 (5-TACC GTGAGGGAAAGTTGAAA-3; Choudhury & Werren, 2006)/ 28b (5-TCGGAAGGAACCAGCTACTA-3; Whiting *et al.*, 1997). DNA amplification protocols of COI and 28S followed Deng *et al.* (2012). The amplification success rates for the COI and 28S genes were 96.3 and 90.2%, respectively. The 28S sequence of *Takahashia japonica* was not obtained either due to the low quality of DNA template extracted from dry specimen or failed amplification. Products were visualized on 1% agarose, and the most intense products were sequenced bidirectionally using BigDye v3.1 on an ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, California, USA). Sequences were aligned in Bioedit (Hall, 1999).

Analysis of molecular data

Similarity-based method

The BLAST programs are popular tools for searching DNA databases to determine the nearest neighbor to the query sequence using a raw similarity score (Altschul *et al.*, 1997). All haplotypes of COI and 28S sequence were queried in the National Center for Biotechnology Information (NCBI) nucleotide database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameters. Query COI sequences were assigned as the species associated with sequences with more than 90% coverage and 95% similarity. For 28S sequences, a higher similarity value (98%) was set because 28S lacks sufficient variation to resolve some species (Park *et al.*, 2011).

Distance-based method

Genetic interspecific and intraspecific distances were calculated using Mega 6 (Tamura *et al.*, 2013) with the Kimura two-parameter (K2P) model (Kimura, 1980). A frequency distribution histogram of inter- and intraspecific divergences of COI sequences was generated to identify the barcoding gap (Meyer & Paulay, 2005). To test the successful identification rate of COI and 28S, we employed the 'best match (BM)' criteria from Meier *et al.* (2006). This method assigns query sequences to species according to the best-matching barcode sequence. If query and match sequences are conspecific, the identification is considered a success, whereas mismatched names are considered failures. Several equally good best matches from different species are considered ambiguous (Meier *et al.*, 2006). TaxonDNA (Meier *et al.*, 2006) was used to estimate the proportion of correct matches according to BM.

Automatic barcoding gap discovery (ABGD) is a species discrimination tool based on clustering algorithms to distinguish partitions in the genetic distances (Puillandre *et al.*, 2012*a*), and was used in this study to assign sequences to candidate species. ABGD analysis was performed using the web interface (http://wwwabi.snv.jussieu.fr/public/abgd/, web version) using default parameters of relative gap width ($X = 1.5$) and K2P distance. The range of prior intraspecific divergence from 0.001 to 0.1 was recorded with 20 steps.

Tree-based method

Tree-based methods considered a species correctly identified if the query and all its conspecific sequences formed a monospecific clade (Virgilio *et al.*, 2010). Neighbor-joining (NJ) trees (Saitou & Nei, 1987) and Bayesian trees (BY) (Huelsenbeck & Ronquist, 2001) were constructed. The former represents the classical method of barcoding and the latter is recommended in further general mixed Yule coalescent (GMYC) analysis (Talavera *et al.*, 2013). NJ trees based on K2P distances were built in Mega6 (Tamura *et al.*, 2013) using 500 bootstrap replicates. Bayesian inference analysis was performed with MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003). The GTR + I + G model was selected for both COI and 28S data using jModelTest (Posada, 2008) based on the AICc criterion. Two independent runs (one hot and three cold chains) were performed for 5,000,000 generations by sampling one tree per 100 generations. The first 25% of trees were discarded as burn-ins. Bayesian posterior probabilities were used to evaluate tree robustness. *Nipponaclerda biwakoensis* (Hemiptera: Aclerdiae) was chosen as the outgroup.

Klee diagram

The Klee diagram approach is a recently described technique to assign sequences to known species and groups of organisms (Sirovich *et al.*, 2009, 2010). This method transforms nucleotide sequences into numerical vectors and compares the species- and group-distinguishing vectors with others. Klee diagrams distinguish differences between species with high information density, enabling accurate quantitative display of affinities amongst taxa at various scales and extending to large genomic data sets (Sirovich *et al.*, 2010). In this study, all 332 COI sequences were sorted based of the order of the BI tree and then transformed into vectors following Sirovich *et al.* (2009).

GMYC model

We also used the GMYC method (Pons *et al.*, 2006), a likelihood method that fits within- and between-species branching models to reconstructed gene trees, to delimit soft scale species. The BI tree was adjusted by non-parametric rate smoothing (Sanderson, 1997) to form an ultrametric tree using the r8s program (Sanderson, 2003). The evolutionary units on the BI tree were then inferred using the GMYC approach (Pons *et al.*, 2006). Single-threshold GMYC analysis was conducted in R (Team, 2012) using the APE (Paradis *et al.*, 2004) and SPLITS (Ezard *et al.*, 2009) packages. Haplotype sequences of COI and 28S were used in GMYC analyses.

Results

Sequence variation

The length of all 332 COI sequences was 543 bp after edge trimming, with 232 conserved sites, 311 variable sites, and 276 parsimony-informative sites. No insertions, deletions, or stop codons were found in any sequence. All COI sequences had a bias toward low GC content ($A = 41.4\%$, $T = 38.4\%$, $C = 14.2\%$, and $G = 6.0\%$), averaging about 20.2% (range 16.0–25.7%). The mean interspecific K2P distance of COI sequences was 19.81% (Table 1), ranging from 4.60% (*Ceroplastes ceriferus* vs. *Ceroplastes pseudoceriferus*) to 31.47% (*Pulvinaria vitis* vs. *Dicyphococcus ficicola*). Intraspecific divergences of COI sequences were 0–4.20%, with a mean divergence of 0.56%

Table 1. K2P distance information about 21 species with intraspecific divergence >0 and eight genera with multiple species.

	COI (mean ± SE, %)	Range (%)	28S (mean ± SE, %)	Range (%)
Intraspecific	**0.56 ± 0.02**	**0.00–4.20**	**0.07 ± 0.00**	**0.00–1.25**
C. ceriferus	0.93 ± 0.13	0.00–1.88	0.20 ± 0.05	0.00–1.25
C. floridensis	0.09 ± 0.01	0.00–0.19	0.00	0
C. japonicas	0.26 ± 0.04	0.00–0.56	0.07 ± 0.02	0.00–0.31
C. pseudoceriferus	0.33 ± 0.07	0.00–0.56	0.00	0
C. rubens	0.72 ± 0.13	0.00–2.06	0.34 ± 0.07	0.00–0.95
C. hesperidum	0.80 ± 0.17	0.00–1.50	0.00	0
Didesmococcus koreanus	0.10 ± 0.01	0.00–0.37	0.11 ± 0.01	0.00–0.26
Ericerus pela	0.93 ± 0.06	0.00–1.50	0.00	0
Eulecanium cerasorum	0.47 ± 0.06	0.00–0.93	0.06 ± 0.01	0.00–0.13
E. kuwanai	0.46 ± 0.02	0.00–0.93	0.04 ± 0.00	0.00–0.13
Megapulvinaria maxima	0.40 ± 0.10	0.00–0.74	0.00	0
P. nigra	1.49 ± 0.23	0.00–3.41	0.07 ± 0.01	0.00–0.15
P. corni	0.10 ± 0.02	0.00–1.12	0.00	0
Protopulvinaria pyriformis	0.12 ± 0.04	0.00–0.18	0.07 ± 0.03	0.00–0.15
Pulvinaria aurantii	1.34 ± 0.24	0.00–2.25	0.00	0
Pulvinaria psidii	0.05 ± 0.01	0.00–0.19	0.06 ± 0.01	0.00–0.15
P. vitis	2.90 ± 0.47	0.19–4.20	0.31 ± 0.06	0.00–0.56
R. sariuoni	1.34 ± 0.06	0.00–3.22	0.23 ± 0.02	0.00–0.81
Saissetia miranda	0.06 ± 0.01	0.00–0.19	0.20 ± 0.03	0.00–0.44
S. oleae	0.60 ± 0.15	0.00–1.12	0.28 ± 0.06	0.00–0.59
T. japonica	1.00 ± 0.25	0.00–1.87	null	null
Congeneric	**12.29 ± 0.08**	**4.60–23.17**	**8.82 ± 0.14**	**0.00–32.05**
Ceroplastes(8)	12.48 ± 0.06	4.60–16.40	7.96 ± 0.10	0.00–14.65
Coccus(3)	18.24 ± 0.05	17.43–18.81	16.62 ± 0.49	10.40–20.17
Eucalymnatus(2)	13.14 ± 0.00	13.14–13.14	1.54 ± 0.02	1.47–1.61
Eulecanium(3)	6.73 ± 0.03	5.77–8.31	2.24 ± 0.02	1.79–4.07
Parasaissetia(2)	11.22 ± 0.00	11.22–11.22	0.54 ± 0.01	0.44–0.58
Parthenolecanium(2)	9.70 ± 0.00	9.70–9.70	0.71 ± 0.00	0.71–0.71
Pulvinaria(4)	20.05 ± 0.12	17.23–23.17	21.40 ± 0.33	6.80–32.05
Saissetia(3)	12.52 ± 0.25	8.68–16.39	3.89 ± 0.23	0.44–6.50
Interspecific	**19.81 ± 0.02**	**4.60–31.47**	**20.38 ± 0.04**	**0.00–38.99**

In bold are the intraspecific, interspecific, and congeneric divergences. The figures in parentheses refer to the number of species within genera.

(Table 1). There was no overlap between the maximum intraspecific and minimum interspecific divergence (fig. 2). The length of the 312 nuclear 28S sequences ranged from 665 bp in *Prococcus acutissimus* to 809 bp in *Eulecanium kuwanai*. The mean inter- and intraspecific divergences were 20.38 and 0.07%, respectively.

Blast query

Using the Blast program, our COI profile identified 17 out of 36 species and the 28S profile identified 16 out of 35, resulting in 47.2 and 45.7% success rates of identification, respectively. Two factors may explain the low success rate. One was a lack of conspecific sequences in GenBank, such as *E. kuwanai*, *Rhodococcus sariuoni*, and *Ceroplastes stellifer*. The other was that the best hits contained more than one species. The latter situation often occurred when querying 28S sequences.

BM and ABGD

The BM method yielded correct identification rates of 100 and 95.52% for COI and 28S data sets, respectively. For the 28S data set, equally good BMs of *C. ceriferus* and *C. pseudoceriferus* were from different species, resulting in 14 ambiguous identifications (4.48%).

The number of partitions varied from 34 to 69; both the lowest and highest results were produced by ABGD (fig. 3). A major barcode gap was evident at *a priori* genetic distance thresholds of 0.042 and 0.046, strongly supporting the presence of 35 genetically distinct partitions in the COI data set. The number of partitions produced by ABGD was generally in accord with morphological identifications, excluding *C. ceriferus* and *C. pseudoceriferus*. The remaining groups were partitioned unambiguously.

Tree-based method and Klee diagram

A total of 82 COI haplotypes and 59 nuclear 28S haplotypes were used to construct the BI (figs 4 and 5) and NJ trees (Supplementary figs S1 and S2). Both phylogenetic trees revealed similar topologies for most clades. The COI data set of common soft scales were split into 36 distinct clades according to the topologies and node supports, while the 28S data set (except *T. japonicas* because of polymerase chain reaction (PCR) failure) was split into 34 clades. *C. ceriferus* and *C. pseduceriferus* formed a monophyletic clade in 28S trees (fig. 5). A comparison between BI and NJ gene trees did not reveal obvious differences in the molecular operational taxonomic units (MOTUs). The affinities of 332 COI sequences are displayed in the Klee diagram (fig. 4b). Sequence clusters appeared as 36 blocks of high correlation along the diagonal and corresponded mutually to the 36 soft scale morphospecies. Two closely related species *C. ceriferus* and *C. pseudoceriferus*

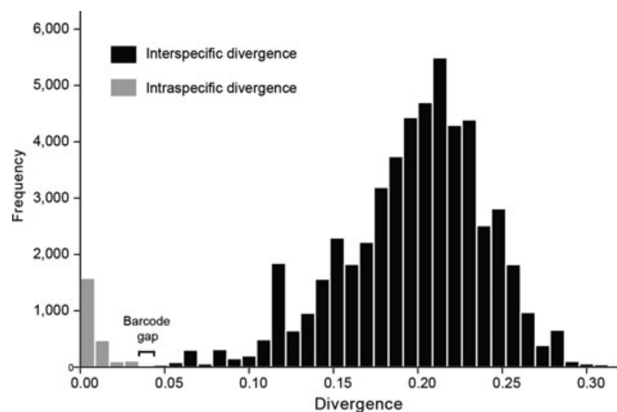

Fig. 2. Frequency distribution histogram of genetic distances based on 332 COI sequences for 36 soft scale species in China. The intraspecific and interspecific K2P distances are displayed using gray and black columns, respectively.

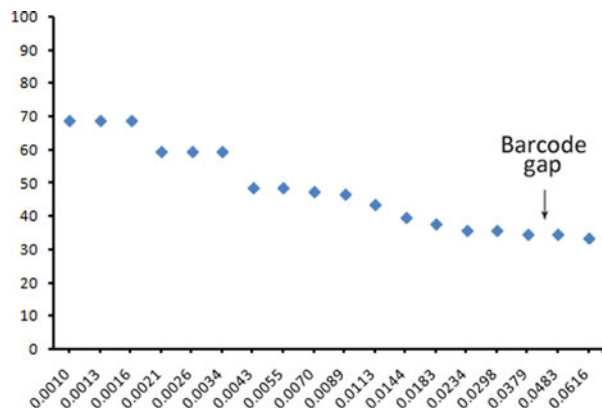

Fig. 3. Automatic partitions generated by ABGD using COI data set. Abscissa is the value of prior intraspecific divergence, while ordinate is the number of groups produced by ABGD.

possessed distinct blocks, and the similarity of the sequences between them was close to 0.8.

The GMYC model

The GMYC model based on COI data using a single-threshold method identified many morphological clusters as independent entities; its likelihood ($L_{GMYC} = 369.84$) was significantly superior to that of the null model ($L_{null} = 349.09$, P-value = 5.15×10^{-9}). The confidence interval for the number of entities ranged from 37 to 47, with the most conservative estimate being exactly 44, seven more than that based on morphology. For example, *R. sariuoni* and *P. vitis* were split into three MOTUs and *Pulvinaria aurantii*, *Pa. nigra*, and *C. rubens* were divided into two MOTUs (fig. 4a).

Discussion

PCR success rate is an important criterion for DNA barcodes (Kress & Erickson, 2007). The limited utility of DNA barcoding on scale insects is mainly attributed to the lack of universal primers in this group (Kondo *et al.*, 2008). Thus, many attempts have been made to overcome this challenge (mealybugs, Malausa *et al.*, 2011; mealybugs and armoured scales, Park *et al.*, 2011; wax scales, Deng *et al.*, 2012). In this study, primer pairs from Deng *et al.* (2012) were used to recover the COI barcodes of common soft scales in China. We successfully amplified and sequenced 96.3% samples, indicating that the primer set could be widely utilized in the barcoding work of soft scales. In addition, the 28S gene had a 90.2% PCR success rate, supporting its effectiveness as a complementary marker to the COI barcode (Sethusa *et al.*, 2014).

The present study assessed the use of DNA barcoding for common coccids in China. Overall, the success identification rates using BM were high, over 100% for COI sequences and above 95% for 28S sequences, supporting the utility of DNA barcoding for identification of soft scales in China. Only 14 (4.48%) of 313 nuclear 28S sequences were ambiguously identified, and were derived from two sibling species, *C. ceriferus* and *C. pseudoceriferus*. The two taxa formed a monophyletic cluster in 28S phylogenetic trees, as observed by Deng *et al.* (2012). They are morphologically similar and often difficult to identify (Deng *et al.*, 2012). However, COI barcodes unambiguously distinguished them based on divergence values (5.0% between the two taxa), well-supported trees, unique GMYC entities, and indicator vectors on the Klee diagram, all of which were highly correlated (fig. 4). One possible explanation for this phenomenon is that the nuclear 28S gene is more conserved than the mitochondrial COI gene (Park *et al.*, 2011), and closely related species often possess 28S sequences that are nearly identical. Beyond that, the 28S gene could specifically identify common soft scales in China, although the differentiations between some congeners was minor (<1%) (fig. 5).

Four other approaches (tree-based, ABGD, GMYC, and Klee diagram) were used to distinguish coccid species. They produced congruent results with morphological identifications, except for some taxa in the ABGD and GMYC analyses. ABGD is an effective identification method because it automatically detects the barcoding gap distance, greatly reducing the interference of artificial factors (Puillandre *et al.*, 2012*a*). In our COI data set, the 36 species were partitioned into 35 groups; *C. ceriferus* and *C. pseudoceriferus* were not successfully distinguished. This may be because the minimum distance between the two closely related species (4.6%) was near the maximum intraspecific divergence (4.2%), disturbing the analysis of ABGD. The GMYC model is generally considered an effective method to detect species boundaries (Leliaert *et al.*, 2009) with a tendency to deliver a higher MOTU count (Fontaneto *et al.*, 2009; Ceccarelli *et al.*, 2012; Puillandre *et al.*, 2012*b*; Tang *et al.*, 2012; Talavera *et al.*, 2013; Weigand *et al.*, 2013). The presence of cryptic taxa could explain splitting of species by GMYC. Bergsten *et al.* (2012) showed that expanding a study's geographic scale can increase intraspecific variation, meaning that the possibility of identifying cryptic species when sampled on a large geographical scale is high. In our study, among the five species with multiple GMYC entities, two species, namely, *C. rubens* and *R. sariuoni*, occupied vast geographic ranges. The other three species were also collected at two or three distant provinces (Supplementary Table S1). The presence of cryptic species of scale insects has been hypothesized because of their intimate relationship with host plants and the considerable intraspecific molecular divergence (Provencher *et al.*, 2005; Gwiazdowski *et al.*, 2011). The sedentary lifestyle of scale insects could allow local conditions to

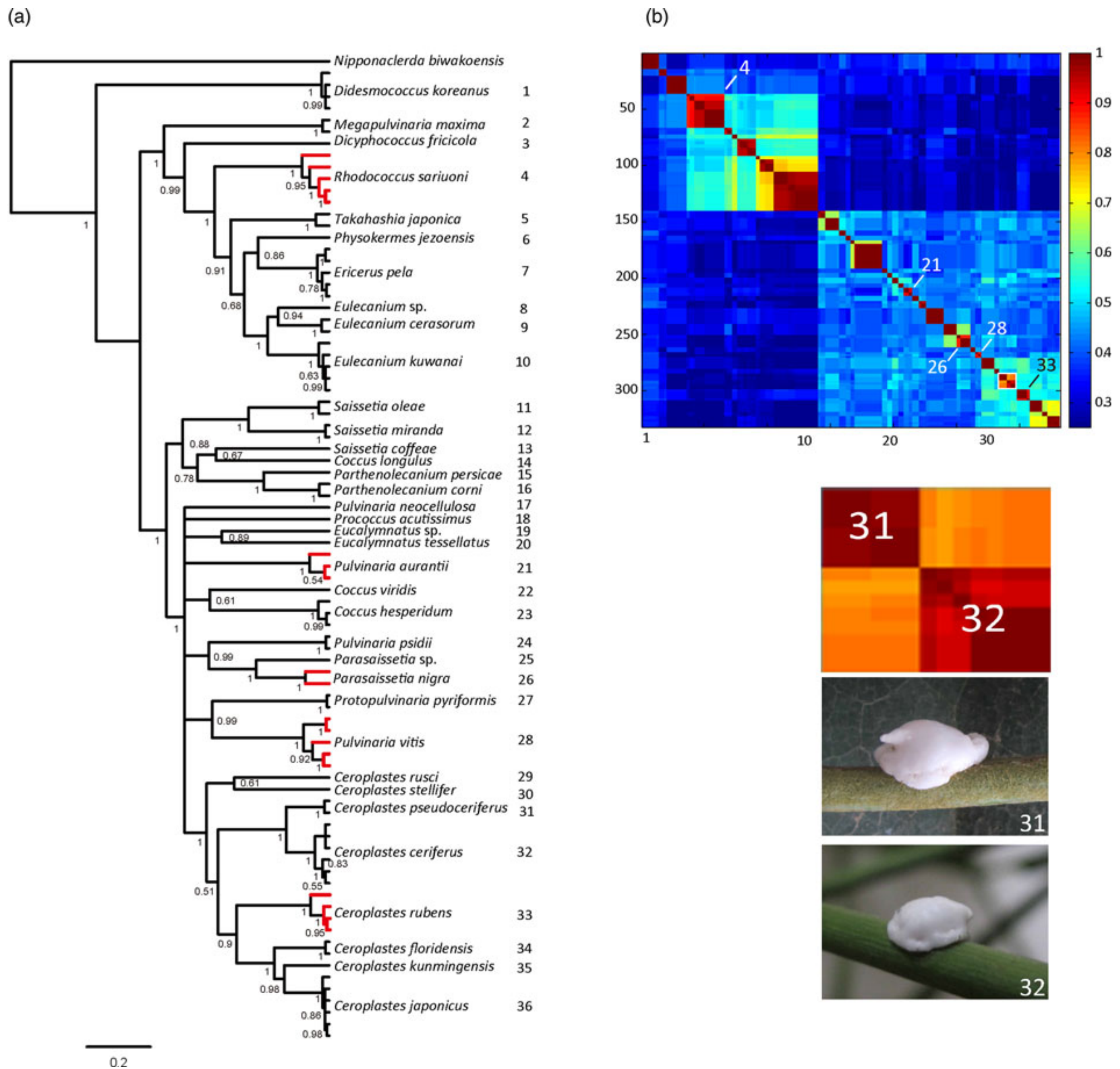

Fig. 4. Sequence clusters of 36 common soft scales in China according to the COI data set. (a) BT based on 83 haplotypes with posterior probabilities (>0.5) indicated next to each node. The 36 morphospecies are represented each by a monophyletic clade with scientific names at the tip of clade. Subclades in red refer to 12 GMYC entities in five morphospecies. (b) Klee diagrams of the 332 COI sequences (*y*-axis) showing the correlations among indicator vectors for the 36 soft scale species (*x*-axis). Sequences cluster as blocks with high correlation along the diagonal, corresponding to 36 morphospecies (the numbers beside blocks accord with those in fig. 4a and refer to the five species with multiple GMYC entities). In case of *C. ceriferus* and *C. pseudoceriferus*, magnifying Klee diagrams, along with their photographs, are showed below.

exert diversifying selection within and between populations (Gwiazdowski *et al.*, 2011).

DNA barcoding consists of constructing a barcode library from known species and then matching the barcode sequences of unknown samples (Kress & Erickson, 2012). However, query sequences from unknown samples can be difficult to identify using this approach because of the limited number of species in barcode libraries (Deng *et al.*, 2014; Jiang *et al.*, 2014). GenBank® is a comprehensive public database of nucleotide sequences that supports bibliographic and biological annotation (Benson *et al.*, 2012). COI sequences of 41 soft scales with scientific names were found in this database. Considering the 50% success identification rate of our Blast queries, its utility is limited for species identification of coccids in China. There are 36 soft scale pests in China (Wu, 2009), most of which were included in our study. The present study, which included 36 species of 17 genera, will enrich the barcode dataset for coccid pests in China and provide a reliable and rapid diagnostic method. Furthermore, as the adult females of some coccids have distinctive morphological features for generic-level identification, additional photographs of the tested 36 species are

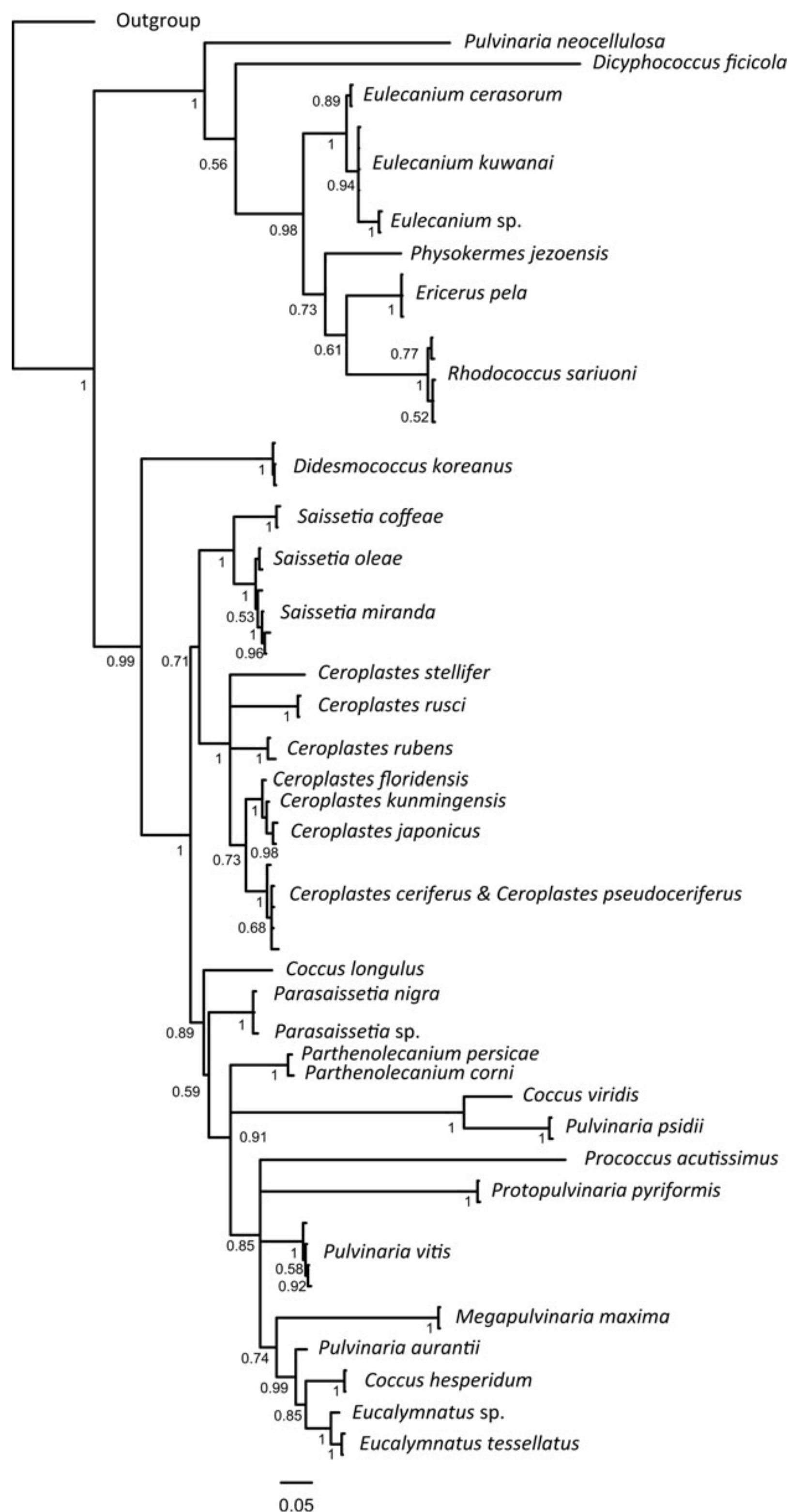

Fig. 5. Bayesian 28S gene tree of tested coccid species from 60 haplotypes. *Nipponaclerda biwakoensis* (Hemiptera: Aclerdiae) is chosen as the outgroup. Posterior probability for each haplogroup is shown near to the node. Values <50% are hidden.

provided in the Supplementary Material (figs S3–S5) to assist the primary distinction of soft scales in the wild.

In conclusion, our study suggests that DNA barcoding is a rapid and effective tool for identification of common soft scales in China. DNA barcoding with multiple methods not only accurately identifies species, but also quickly reveals species that require detailed inspection when conflicting results are available. Our results facilitate species identification and can be used to uncover new and cryptic species.

Supplementary material

The supplementary material for this article can be found at http://www.journals.cambridge.org/10.1017/S0007485315000413

Acknowledgements

We are grateful to two anonymous reviewers and editors for the helpful comments on an earlier version of the manuscript, and to the following people who helped us to collect coccids samples: Fang Yu, Lin-Lin Zheng, Qing-Tao Wu, Xin-Lei Huang, Xiu-Wei Liu, Xu Zhang, Xue-Mei Yang, Ying Wang (Institute of Zoology, Chinese Academy of Sciences, Beijing), Ju-Pu Chang (Puyang Academy of Forestry, Puyang), Xiao-Hua Dai (Gannan Normal University, Ganzhou), Yu-Qiang Xi (China Agricultural University, Beijing), Guo-Hua Huang (Hunan Agricultural University, Changsha), Shao-Bin Huang (Guangdong Forestry Vocational Technology College, Guangzhou), Jian-qin Wu (The Administrative Bureau of Tianbaoyan National Nature Reserve of Yong'an, Yong'an), Kai-Ju Wei (Youxi No.1 Middle school of Fujian Province, Youxi), Hong-Liang Li (Institute for Nutritional Sciences, SIBS, Chinese Academy of Sciences, Shanghai), Hu Li (Guizhou University, Guiyang), Xian Li (Forestry Protection Station of Chengdu, Sichuan), Qiang Shen (Forestry Protection Station of Yuyao, Ningbo), Xiu-Hao Yang (Forestry Protection Station of Guangxi, Nanning), Ying-Jie Zhang (Yunnan Agricultural University, Kunming), Fang-Ping Zhang (Chinese Academy of Tropical Agriculture Sciences, Haikou), Hai-Bin Li, Hui Wang, and Nan Nan (Beijing Forestry University, Beijing), Ping Zhang (Institute for Agricultural Sciences of the second division of Xinjiang production and construction corps, Korla), Xiu-Li Tang (Xinjiang University, Urumchi), Xu-Yang Zou (Henan Agricultural University, Zhengzhou). This project was supported by the National Natural Science Foundation of China (NSFC grant no. 31372151) and the Fundamental Research Funds for the Central Universities (BLYJ201305).

References

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

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Ball, S.L. & Armstrong, K.F. (2007) *Using DNA Barcodes to Investigate the Taxonomy of the New Zealand Sooty Beech Scale Insect*. Wellington, New Zealand: Science and Technical Publishing, Department of Conservation.

Beltrà, A., Soto, A. & Malausa, T. (2012) Molecular and morphological characterisation of Pseudococcidae surveyed on crops and ornamental plants in Spain. *Bulletin of Entomological Research* **102**, 165–172.

Ben-Dov, Y. & Hodgson, C.J. (1997) *Soft Scale Insects: Their Biology, Natural Enemies and Control*. World Crop Pests, Vol. **7A**. Amsterdam, Elsevier.

Ben-Dov, Y., Miller, D.R. & Gibson, G.A.P. (2014) Scalenet: *a database of the scale insects of the world*. United States Department of Agriculture (USDA) Available from: http://www.sel.barc.usda.gov/scalenet/scalenet.htm (accessed 16 October 2014).

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2012) Genbank. *Nucleic Acids Research* **40**, D48–D53.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G. & Vogler, A.P. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology* **61**, 851–869.

Campbell, B.C., Steffen-Campbell, J.D. & Werren, J.H. (1994) Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Molecular Biology* **2**, 225–237.

Ceccarelli, F.S., Sharkey, M.J. & Zaldívar-Riverón, A. (2012) Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach. *Molecular Phylogenetics and Evolution* **62**, 485–495.

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

Costa, F.O., DeWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M. & Hebert, P.D.N. (2007) Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 272–295.

Deng, J., Wang, X.B., Yu, F., Zhou, Q.S., Bernardo, U., Zhang, Y. Z. & Wu, S.A. (2014) Rapid diagnosis of the invasive wax scale, *Ceroplastes rusci* Linnaeus (Hemiptera: Coccoidea: Coccidae) using nested PCR. *Journal of Applied Entomology* **139**, 314–319.

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

Ezard, T., Fujisawa, T. & Barraclough, T.G. (2009) Splits: Species' limits by threshold statistics. R package version 1.0–11/r29.

Fontaneto, D., Kaya, M., Herniou, E.A. & Barraclough, T.G. (2009) Extreme levels of hidden diversity in microscopic animals (*Rotifera*) revealed by DNA taxonomy. *Molecular Phylogenetics and Evolution* **53**, 182–189.

Gill, R.J. (1988) *The Scale Insects of California. Part 1*. Florida, Analysis and Identification Branch, Division of Plant Industry, California Department of Food and Agriculture.

Gimpel, W.F., Miller, D.R. & Davidson, J.A. (1974) *A Systematic Revision of the Wax Scales, Genus Ceroplastes, in the United States (Homoptera: Coccoidea: Coccidae)*. Maryland, Agricultural Experiment Station, University of Maryland.

Gullan, P.J. & Kosztarab, M. (1997) Adaptations in scale insects. *Annual Review of Entomology* **42**, 23–50.

Gwiazdowski, R.A., Vea, I.M., Andersen, J.C. & Normark, B.B. (2011) Discovery of cryptic species among North American pine-feeding *Chionaspis* scale insects (Hemiptera: Diaspididae). *Biological Journal of the Linnean Society* **104**, 47–62.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. & Hebert, P.D.N. (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 968–971.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editorand analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Hamon, A.B. & Williams, M.L. (1984) *The Soft Scale Insects of Florida (Homoptera: Coccoidea: Coccidae) In: Arthropods of Florida and Neighboring Land Areas*. Gainesville, Florida, Florida Department of Agriculture & Consumer Services, Division of Plant Industry.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. (2004) Identification of birds through DNA barcodes. *PLoS Biology* **2**, e312.

Hodgson, C.J. (1994) *The Scale Insect Family Coccidae: An Identification Manual to Genera*. Wallingford, Oxon, UK, CAB International.

Huelsenbeck, J.P. & Ronquist, F. (2001) Mrbayes: bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Jiang, F., Jin, Q., Liang, L., Zhang, A.B. & Li, Z.H. (2014) Existence of species complex largely reduced barcoding success for invasive species of Tephritidae: a case study in *Bactrocera* spp. *Molecular Ecology Resources* **14**, 1114–1128.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.

Kondo, T., Gullan, P.J. & Williams, D.J. (2008) Coccidology. The study of scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Revista Corpoica–Ciencia y Tecnología Agropecuaria* **9**, 55–61.

Kress, W.J. & Erickson, D.L. (2007) A two-locus global DNA barcode for land plants, the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* **2**, 508.

Kress, W.J. & Erickson, D.L. (2012) *DNA Barcodes: Methods and Protocols*. Berlin, Germany, Springer.

Leliaert, F., Verbruggen, H., Wysor, B. & Clerck, O.D. (2009) DNA taxonomy in morphologically plastic taxa: algorithmic species delimitation in the *Boodlea* complex (Chlorophyta: Cladophorales). *Molecular Phylogenetics and Evolution* **53**, 122–133.

Malausa, T., Fenis, A., Warot, S., Germain, J.F., Ris, N., Prado, E., Botton, M., Vanlerberghe-Masutti, F., Sforza, R. & Cruaud, C. (2011) DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae). *Journal of Applied Entomology* **135**, 142–155.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55**, 715–728.

Meyer, C.P. & Paulay, G. (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3**, e422.

Mikkelsen, N.T., Schander, C. & Willassen, E. (2007) Local scale DNA barcoding of bivalves (Mollusca): a case study. *Zoologica Scripta* **36**, 455–463.

Miller, G.L. & Miller, D.R. (2003) Invasive soft scales (Hemiptera: Coccidae) and their threat to US agriculture. *Proceedings of Entomological Society of Washington* **105**, 832–846.

Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J.G., Lees, D.C., Ranaivosolo, R., Eggleton, P. & Barraclough, T.G. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* **58**, 298–311.

Paradis, E., Claude, J. & Strimmer, K. (2004) Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290.

Park, D.S., Suh, S.J., Hebert, P.D.N., Oh, H.W. & Hong, K.J. (2011) DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). *Bulletin of Entomological Research* **101**, 429–434.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* **55**, 595–609.

Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253–1256.

Provencher, L.M., Morse, G.E., Weeks, A.R. & Normark, B.B. (2005) Parthenogenesis in the *Aspidiotus nerii* complex (Hemiptera: Diaspididae): a single origin of a worldwide, polyphagous lineage associated with Cardinium bacteria. *Annals of the Entomological Society of America* **98**, 629–635.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012a) ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology* **21**, 1864–1877.

Puillandre, N., Modica, M.V., Zhang, Y., Sirovich, L., Boisselier, M.C., Cruaud, C., Holford, M. & Samadi, S. (2012b) Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology* **21**, 2671–2691.

Ronquist, F. & Huelsenbeck, J.P. (2003) Mrbayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Sanderson, M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**, 1218–1231.

Sanderson, M.J. (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302.

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

Sirovich, L., Stoeckle, M.Y. & Zhang, Y. (2009) A scalable method for analysis and display of DNA sequences. *PLoS ONE* **4**, e7051.

Sirovich, L., Stoeckle, M.Y. & Zhang, Y. (2010) Structural analysis of biodiversity. *PLoS ONE* **5**, e9266.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W. & Hebert, P.D.N. (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12359–12364.

Talavera, G., Dincă, V. & Vila, R. (2013) Factors affecting species delimitations with the gmyc model: insights from a butterfly survey. *Methods in Ecology and Evolution* **4**, 1101–1110.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G. & Fontaneto, D. (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16208–16212.

Tang, F.T. (1991) *The Coccidae of China*. Taiyuan, China, Shanxi United Universities Press.

Team, R.C. (2012) *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing.

Virgilio, M., Backeljau, T., Nevado, B. & De Meyer, M. (2010) Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics* **11**, 206.

Weigand, A.M., Jochum, A., Slapnik, R., Schnitzler, J., Zarza, E. & Klussmann-Kolb, A. (2013) Evolution of microgastropods (Ellobioidea, Carychiidae): integrating taxonomic, phylogenetic and evolutionary hypotheses. *BMC Evolutionary Biology* **13**, 1471–2148.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D. & Wheeler, W.C. (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Systematic Biology* **46**, 1–68.

Wong, E.H.K., Shivji, M.S. & Hanner, R.H. (2009) Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. *Molecular Ecology Resources* **9**, 243–256.

Wu, S.A. (2009) Checklist and faunistic analysis of scale insect pests (Hemiptera: Coccoidea) in Chinese mainland. *Journal of Beijing Forestry University* **31**, 55–63.