Systematic Entomology

Systematic Entomology (2013), **38**, 615–625 DOI: 10.1111/syen.12018

The *gnd* gene of *Buchnera* as a new, effective DNA barcode for aphid identification

RUI CHEN[1,2], LI-YUN JIANG[1], LIN LIU[1,2], QING-HUA LIU[1,2], JUAN WEN[1,2], RUI-LING ZHANG[1,2], XING-YI LI[1,2], YUAN WANG[1,2], FU-MIN LEI[1] and GE-XIA QIAO[1]

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, P.R. China and [2]University of Chinese Academy of Sciences, Beijing, P.R. China

Abstract. DNA barcoding uses a standard DNA sequence to facilitate species identification. Although the *COI* gene has been adopted as the standard, *COI* alone is imperfect due to several shortcomings. The primary endosymbiont of aphids, *Buchnera*, has higher evolutionary rates and interspecies divergence than its co-diverging aphid hosts, making it a potential tool for resolving the ambiguities in aphid taxonomy. We compared the effectiveness of employing two different DNA regions, *gnd* and *COI*, for the discrimination of over 100 species of aphids. The mean interspecific divergence of the *gnd* region was significantly higher than the mean intraspecific variation; there were nearly nonoverlapping distributions between the intra- and interspecific samples. In contrast, *COI* showed a lower interspecific divergence, which led to difficulties in identifying closely related species. Our results show that *gnd* can identify species in the Aphididae, which suggests that the *gnd* region of *Buchnera* is a potentially effective barcode for aphid species identification. We also recommend the 2-locus combination of *gnd* + *COI* as the aphid barcode. This will provide a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of aphids.

Introduction

Since Hebert *et al*. (2003a,b) proposed the use of a 'DNA barcode' to identify animals, DNA barcoding has attracted worldwide attention. Central to the efficacy of barcoding is selection of a suitable segment of DNA (Waugh, 2007). First, differences of barcoding segments should be sufficient for accurate species discrimination and specific for each species; secondly, universal robust primers for amplification and sequence acquisition must be available; thirdly, sequences should be aligned easily (rarity of indels and introns) (Hebert *et al*., 2003a; Waugh, 2007; Ferri *et al*., 2009; Floyd *et al*., 2009). The mitochondrial gene encoding the *cytochrome c oxidase subunit 1* (*COI*) possesses a high level of diversity, is easy to acquire and align, and has been favoured for most animals and certain fungi (Hebert *et al*., 2003b). The *COI* barcode has standardized characterization of life forms in numerous organismal groups (Hajibabaei *et al*., 2007). *COI*-based barcoding can contribute to the routine identification of species in applied settings, including detection of morphologically cryptic species and host-specific lineages, and discovery of associations between morphologically distinct forms in one life cycle of a species (Miller & Foottit, 2009). However, certain aspects are controversial: mitochondrial genes can overestimate the number of species due to amplification of nuclear mitochondrial pseudogenes (Song *et al*., 2008). Furthermore, the *COI* gene has limited utility in identifying certain organismal groups, such as plants (Chase *et al*., 2005; Kress *et al*., 2005; Fazekas *et al*., 2008), Diptera (Meier *et al*., 2006) and aphids (Foottit *et al*., 2008; Lee *et al*., 2011; Chen *et al*., 2012), due to its low level of variation between such species and thus it is necessary to seek other regions that are appropriate for DNA barcoding.

Correspondence: Ge-Xia Giao, Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing 100101, P.R. China. E-mail: qiaogx@ioz.ac.cn

The aphids (Insecta: Hemiptera: Aphididae) are a group of over 4700 known species of small, soft-bodied insects that feed on plant phloem with their slender mouthparts. They are divided into two groups, viviparous (Aphididae) and oviparous (Adelgidae and Phylloxeridae) (Heie, 1980). Many aphid species have complex life cycles involving several morphologically distinct forms, including a few parthenogenetic forms and sexual forms. Nearly 10% of aphid species are associated with host alternation (Heie, 1987). Local aphid populations can spread quickly, as parthenogenetic reproduction rapidly increases in number whereas wind assists dispersal of winged forms. As invasive pests, aphids have a major impact on global economies (Teulon & Stufkens, 2002; Foottit *et al*., 2006; Messing *et al*., 2007).

It is notably difficult to identify field samples because different morphological forms of a single species are present on different hosts at different times. Accurate methods of identification are needed because aphids not only cause direct damage in agriculture, forestry and horticulture but also transmit viral diseases to many important crops (Eastop, 1977; Blackman & Eastop, 1984; Minks & Harrewijn, 1987; van Emden & Harrington, 2007). Molecular taxonomic approaches have improved the accuracy of classification, as well as aided in the discovery of new species within the Aphididae (Foottit, 1997). However, morphological similarities between species limit accurate identification. Certain genera in particular – the largest aphid genus *Aphis* Linnaeus and the second largest *Cinara* Curtis – comprise many closely related species. Although some species in these groups can be recognized by diagnostic morphological traits, many cannot.

Mutualistic associations between insects and endosymbionts exist among diverse insect orders including Hemiptera (aphids, whiteflies, mealybugs, psyllids and cicadas), Blattaria (cockroaches) and Coleoptera (beetles) (Buchner, 1965; Baumann *et al*., 2000; Lefevre *et al*., 2004). In aphids, the primary endosymbiont *Buchnera aphidicola* is housed in bacteriocytes located in the abdominal haemocoel (McLean & Houk, 1973; Douglas & Dixon, 1987; Munson *et al*., 1991). *Buchnera* have been observed in almost all aphids, providing essential amino acids lacking in plant phloem (Shigenobu *et al*., 2000). It has higher evolutionary rates and interspecies divergence than its co-diverging aphid hosts (Moran *et al*., 1995; Clark *et al*., 1999; Jousselin *et al*., 2009). Aphids acquired *Buchnera c*. 150–200 ma (Moran *et al*., 1995; Martinez-Torres *et al*., 2001) and analyses of *Uroleucon* Mordvilko have suggested that *Buchnera* and aphids have undergone strict co-divergence (Clark *et al*., 2000; Funk *et al*., 2000; Wernegreen *et al*., 2001). Theoretically, the symbionts have undergone a cytoplasmic mode of inheritance with no horizontal transmission; the parallel cladogenesis of the symbionts and their hosts may be the result of long-term strict mother-to-daughter transmission (Funk *et al*., 2000; Jousselin *et al*., 2009). Due to the similarity of aphid and symbiont phylogenetic clusters and the higher interspecies divergence of *Buchnera*, *Buchnera* markers may replace or supplement aphid markers. The implementation of *Buchnera* DNA barcodes could resolve difficulties in aphid identification. The *gnd* gene encodes the third enzyme of the pentose phosphate pathway, 6-phosphogluconate dehydrogenase (6PGD). This pathway is one of two central and constitutive routes of intermediary carbohydrate metabolism in *Buchnera* (Nelson & Selander, 1994). Although 6PGD has an important metabolic function and the amino acid sequence of 6PGD and the nucleotide sequence of *gnd* would be expected to be highly conserved, yet several studies have found that *gnd* has an unusually high level of genetic diversity (Barcak & Wolf, 1988; Bisercić *et al*., 1991; Dykhuizen & Green, 1991). Despite this overall sequence variability, *gnd* has a conserved region that can be targeted for designing universal PCR primers. The *gnd* gene is one of the best candidate markers for *Buchnera* because of its high level of diversity and its ease of amplification with universal PCR primers.

In previous studies of *COI* sequences, certain aphid species were identified successfully (Sabater-Munoz *et al*., 2005; Valenzuela *et al*., 2007; Coeur d'acier *et al*., 2008; Foottit *et al*., 2009; Wang & Qiao, 2009; Qiao *et al*., 2011), but many species were not clearly distinguished (Foottit *et al*., 2008; Lee *et al*., 2011; Chen *et al*., 2012). We have focused on the *gnd* region as a novel marker to complement *COI* DNA barcodes. We include a preliminary analysis of sequence variation in the *gnd* region and a comparison of the effectiveness of *gnd* and *COI* DNA barcoding in 120 species of Aphididae. Special emphasis was placed on taxa lacking diagnostic morphological characteristics, namely *Aphis* Linnaeus, *Cinara* Curtis, *Pseudoregma* Doncaster, *Ceratovacuna* Zehntner, *Greenidea* Schouteden and *Eutrichosiphum* Essig & Kuwana. The results suggest that the combination of *gnd* + *COI* as a standard 2-locus barcode is an ideal approach to identifying aphids.

Materials and methods

Taxa examined

All aphid samples were collected from China, Mongolia, USA, Korea and Japan in the past 10 years, and the detailed information can be found in File S1. Samples were selected to ensure coverage of most subfamilies of the Aphididae. For three genera – *Cinara* Curtis, *Aphis* Linnaeus and *Greenidea* Schouteden – we analysed as many species as were available. Meyer & Paulay (2005) concluded that the lack of broad geographical sampling for a single species was likely to result in a serious underestimation of within-species variation; additionally, the failure to survey closely related species would overestimate sequence divergence between congeneric taxa. To address intraspecific variation, we studied geographically distant samples of selected species, including *Cinara tujafilina* (del Guercio), *Cinara pinea* (Mordvilko), *Lachnus tropicalis* (van der Goot), *Greenidea kuwanai* (Pergande), *Aphis craccivora* Koch and *Eutrichosiphum pasaniae* (Okajima). The collection set includes 1008 sequences from 518 samples, covering 120 species, 45 genera and 15 subfamilies (File S2).

For *gnd*, the ingroup included 498 aphid samples from 120 species, 45 genera and 15 subfamilies. The outgroup consisted of *Escherichia coli* (Escherich) and *Klebsiella pneumoniae* (Schroeter). For *COI*, the ingroup included 510 aphid samples from 116 species, 41 genera and 14 subfamilies. The outgroup consisted of *Pineus armandicola* (Zhang) and *Phylloxerina salicis* (Lichtenstein).

All collection data, including locations, host plants and collection dates, are shown in File S1. With the exception of specimens for slide-mounting that were stored in 70% ethanol, all specimens were stored in 95% or 100% ethanol. All samples and voucher specimens (slide-mounted specimens) were deposited in the National Zoological Museum of China, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.

Taxon assignment follows the current world catalogue of aphids (Remaudière & Remaudière, 1997) with updates to the subfamily names according to Nieto Nafría *et al*. (1998). Species authorship and date of publication can be found in Remaudière & Remaudière (1997).

Most samples of each species included more than one individual. DNA was isolated for molecular studies from one to three individuals per sample, and three to five individuals per sample were prepared as slide-mounted specimens for morphological examination. Voucher specimens of all samples were identified by the main morphological diagnostic features and compared with the related denominate specimens. The species name of each sample is provided in File S1.

DNA extraction, PCR and sequencing

Total DNA was extracted from a single aphid preserved in 95% or 100% ethanol. Tissue homogenates were incubated at 55°C in lysis buffer [30 mM Tris HCl (pH 8.0), 200 mM EDTA, 50 mM NaCl, 1% SDS and 100 µg/ml Proteinase K] for 5–7 h, followed by a standard phenol-chloroform-isoamylalcohol (PCI) extraction with modifications (Sambrook *et al*., 1989). DNA was precipitated from the supernatant with two volumes of cold ethanol, centrifuged, washed, dried and dissolved in 15–20 µl TE buffer. The isolated DNA was stored at 4°C for later use.

The amplicon size of *gnd* is approximately 900 bp; the primers used (5'–3') were *Bam*HI: CGCGGATCCGGWCCWWSWATWATGCCWGGWGG and *Apa*I: CGCGGGCCCGTATGWGCWCCAAAATAATCWCKTTGWGCTTG (Clark *et al*., 1999). The amplicon size of *COI* is approximately 660 bp; the primers used (5'–3') were LepF: ATTCAACCAATCATAAAGATATTGG and LepR: TAAACTTCTGGATGTCCAAAAAATCA (Foottit *et al*., 2008). PCR for *gnd* was performed with an initial denaturation of 5 min at 95°C followed by 35 cycles of 95°C for 20 s, 53°C for 30 s and 72°C for 2 min and a final extension of 72°C for 7 min. PCR amplification of *COI* was performed with an initial denaturation of 5 min at 94°C followed by 40 cycles of 94°C for 30 s, 50°C for 1 min and 72°C for 1 min, and a final extension of 72°C for 10 min.

Sequencing reactions were performed bidirectionally with the appropriate amplification primers using a BigDye Terminator Cycle Sequencing Kit v2.0 (Applied Biosystems, USA) and an ABI 3730 automated sequencer (Applied Biosystems, USA). Each of the sequence reactions was repeated three times to confirm reproducibility.

Chromatograms of the sense and antisense sequences were assembled and analysed using the Lasergene Seqman software (DNASTAR, Inc., Madison, WI, USA), and a consensus sequence was obtained. Multiple alignments were generated using CLUSTALX (Thompson *et al*., 1997) and were subsequently pruned to lengths of 790 bp (*gnd*) and 658 bp (*COI*). We confirmed that these sequences were correct by translating them *in silico* using the Editseq software (DNASTAR, Inc.). Sequences were deposited in GenBank, and the accession numbers are provided in File S1.

Data analysis

We used four parameters to determinate genetic divergence. Interspecific divergence was calculated from the average interspecific distance (K2P distance) between all species in each genus with more than one species. Intraspecific variation was evaluated by three additional parameters: average intraspecific difference, theta (θ), and average coalescent depth (Meyer & Paulay, 2005; Lahaye *et al*., 2008). The average intraspecific difference (K2P distance) was measured between all samples of each species with more than one individual. Theta (θ) was the mean pairwise distance within each species with at least two representatives; this measurement eliminates biases associated with unequal sampling within a species. The average coalescent depth was the maximum intraspecific distance within each species with at least two individuals. Intraspecific and interspecific sequence divergences were based on the K2P distances for aphid species; the divergence scores were calculated by MEGA v5.0. The K2P model provides the best metric when genetic distances are low (Nei & Kumar, 2000).

Neighbour-joining (NJ) analysis (Saitou & Nei, 1987) was used to examine the relationships among the taxa and population samples. The simple NJ algorithm was considered at this juncture to be an appropriate starting point for the analyses, given that specimen identification is based entirely on sequence similarity, rather than on strictly phylogenetic relationships, and the speed of analysis that is necessary for the large datasets.

Identifying species based on 'best match' results

We used Taxon DNA (Meier *et al*., 2006) to find each query's closest barcode match. If both sequences were from the same species, the identification was considered a success, whereas the identification was considered to be a failure if the species identities were mismatched. Several equally best matches from different species were considered ambiguous.

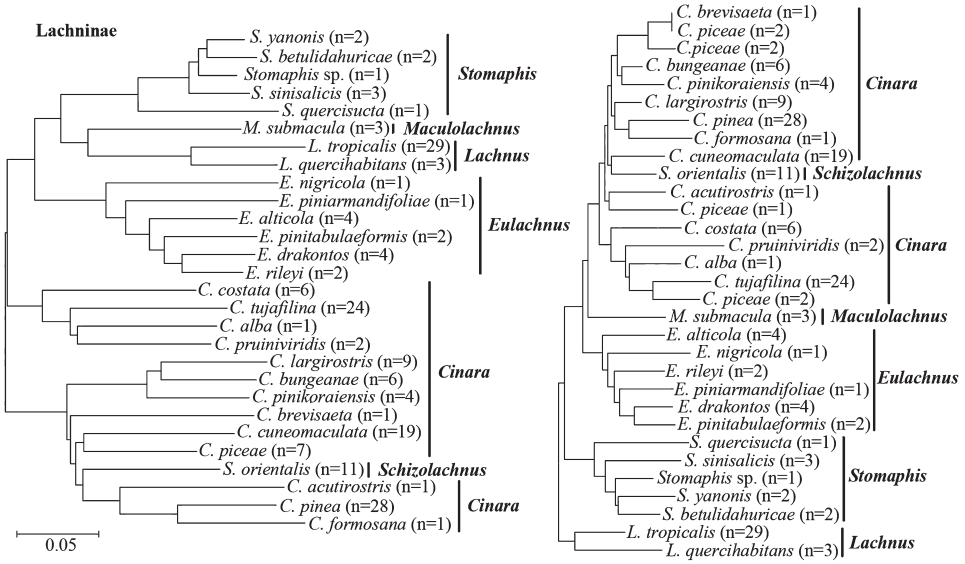

Fig. 1. Neighbour-joining trees based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model in Lachninae. *Buchnera* strains are represented by their host species names. The numbers of identical samples are given in brackets.

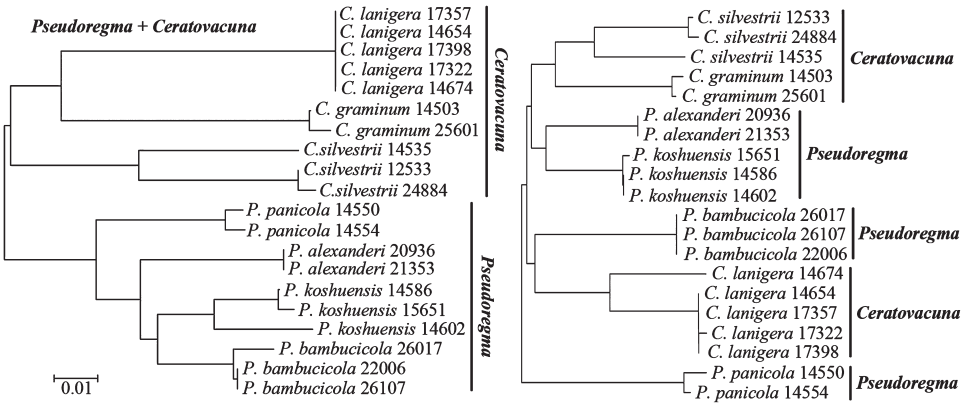

Fig. 2. Neighbour-joining trees based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model in *Pseudoregma* and *Ceratovacuna*. *Buchnera* strains are represented by their host species names.

Results

Efficiency of PCR amplification

We calculated the efficiency of PCR amplification of the *gnd* and *COI* sequences for all samples. For new specimens (characterized since 2008), the success rates for the *gnd* and *COI* sequences were 90.23 and 93.10%, respectively. For degraded DNA of old specimens (characterized before 2008), the success rate for *gnd* was higher than that of *COI* (88.17% vs 81.25%).

Taxonomic assignments and NJ tree structure

The results of the overall NJ analysis by the *gnd*/*COI* region of distances among the samples of 120 species are summarized in Figure S1. The trees represent the distance matrix only and should not be interpreted as a phylogenetic hypothesis. The node in Fig. 1 consisted of Lachninae and Hormaphidinae (*Pseudoregma* and *Ceratovacuna*) and is expanded in Fig. 2. Greenideinae (*Greenidea, Molitrichosiphum* and *Eutrichosiphum*) was expanded in Figure S2, and Eriosomatinae and Aphidinae (*Aphis* Linnaeus and *Toxoptera* Koch) are expanded in Figs 3 and 4, respectively.

In the subfamily Lachninae (Fig. 1), all species formed distinct clusters in the *gnd* analysis; a similar result was observed for *COI* with the exception of a few species (e.g. *Cinera piceae*) which may be polyphyletic. In Hormaphidinae (Fig. 2), *Pseudoregma* and *Ceratovacuna* were sister group. In the *COI* tree, the two genera were embedded in each other; however, for *gnd*, two genera formed a single cluster, respectively. To Eriosomatinae, there is the same as in Hormaphidinae (Fig. 3). In the subfamily Aphidinae (Fig. 4),

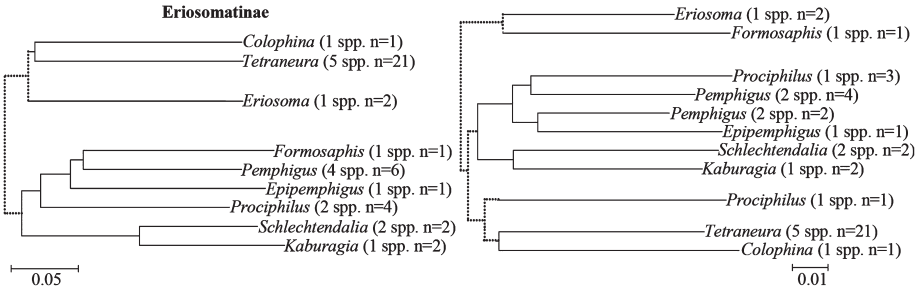

Fig. 3. Neighbour-joining trees based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model in Eriosomatinae. *Buchnera* strains are represented by their host species names. The numbers of identical samples are given in brackets.

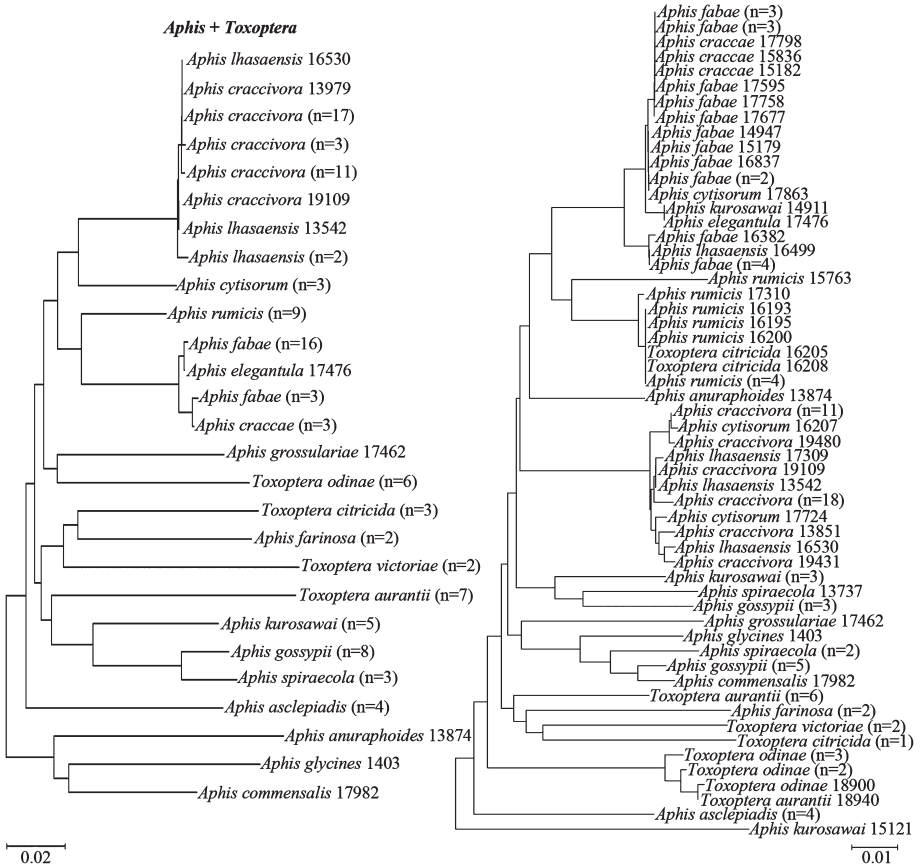

Fig. 4. Neighbour-joining trees based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model in *Aphis* and *Toxoptera*. *Buchnera* strains are represented by their host species names. The numbers of identical samples are given in brackets.

most species formed distinct clusters in the *gnd* tree. The exceptions were two species (*Aphis craccivora* Koch and *Aphis lhasaensis* Zhang) that may be paraphyly. However, when using *COI*, most species did not cluster.

Determination of genetic divergence

A favourable barcode should possess a high interspecific divergence to distinguish different species. The *gnd* region exhibited a higher interspecific divergence compared with the *COI* region (Table 1). Additionally, both the *gnd* and *COI* regions showed low levels of intraspecific variation by all three parameters (Table 1).

Success of similarity on the basis of DNA identification techniques

Success under 'best match' was 88.43% in *COI* and 99.00% in *gnd*. *COI* with 41 sequences and *gnd* with 5 were ambiguous (8.04% in *COI* and 1.00% in *gnd*); 18

Table 1. Analysis of intra- and interspecific divergences of congeneric species in certain genera of aphids for *gnd* and *COI*.

Genera	Analysis of intra- and interspecific divergences of congeneric species in Lachninae			
	Average interspecific distance (*gnd*/*COI*)	Average intraspecific distance (*gnd*/*COI*)	Theta (θ) (*gnd*/*COI*)	Coalescent depth (*gnd*/*COI*)
Cinara Curtis	0.3079 ± 0.0515/0.0895 ± 0.0215	0.0281 ± 0.0453/0.0243 ± 0.0216	0.0273 ± 0.0290/0.0197 ± 0.0273	0.0631 ± 0.0643/0.0362 ± 0.0432
Eulachnus del Guercio	0.1688 ± 0.0462/0.0743 ± 0.0095	0.0081 ± 0.0134/0.0086 ± 0.0108	0.0158 ± 0.0234/0.0130 ± 0.0177	0.0223 ± 0.0252/0.0148 ± 0.0177
Lachnus Burmeister	0.1589 ± 0.0056/0.0753 ± 0.0036	0.0062 ± 0.0078/0.0074 ± 0.0020	0.0078 ± 0.0023/0.0078 ± 0.0006	0.0174 ± 0.0046/0.0202 ± 0.0115
Stomaphis Walker	0.1070 ± 0.0503/0.0810 ± 0.0196	0.0018 ± 0.0019/0.0027 ± 0.0037	0.0013 ± 0.0013/0.0017 ± 0.0030	0.0017 ± 0.0019/0.0022 ± 0.0039
Aphis Linnaeus	0.1075 ± 0.251/0.0683 ± 0.0200	0.0026 ± 0.0031/0.0222 ± 0.0352	0.0024 ± 0.0018/0.0217 ± 0.0242	0.0072 ± 0.0097/0.0371 ± 0.0581
Toxoptera Koch	0.1567 ± 0.0068/0.0773 ± 0.0168	0.0073 ± 0.0145/0.0115 ± 0.0203	0.0077 ± 0.0042/0.0233 ± 0.0263	0.0121 ± 0.0209/0.0310 ± 0.0380
Pseudoregma Doncaster	0.0667 ± 0.0118/0.0596 ± 0.0121	0.0153 ± 0.0185/0.0013 ± 0.0024	0.0117 ± 0.0134/0.0020 ± 0.0033	0.0165 ± 0.0201/0.0022 ± 0.0033
Ceratovacuna Zehntner	0.1360 ± 0.0107/0.0712 ± 0.0070	0.0114 ± 0.0270/0.0208 ± 0.0226	0.0034 ± 0.0039/0.0176 ± 0.0134	0.0274 ± 0.0432/0.0320 ± 0.0250
Greenidea Schouteden	0.1205 ± 0.0260/0.0456 ± 0.0180	0.0198 ± 0.0183/0.0087 ± 0.0085	0.0213 ± 0.0242/0.0091 ± 0.0113	0.0256 ± 0.0314/0.0112 ± 0.0154
Eutrichosiphum Essig *et* Kuwana	0.1821 ± 0.0176/0.0739 ± 0.0100	0.0249 ± 0.0399/0.0143 ± 0.0187	0.0151 ± 0.0166/0.0076 ± 0.0080	0.0609 ± 0.0697/0.0295 ± 0.0340
Mollitrichosiphum Suenaga	0.1571 ± 0.0581/0.0767 ± 0.0280	0.0201 ± 0.0285/0.0087 ± 0.0154	0.0066 ± 0.0136/0.0054 ± 0.0106	0.0890 ± 0.4350/0.0106 ± 0.0225

Table 2. Identification success based on 'best match' in analysis of interspecific divergences of aphid species, based on *COI* and *gnd*.

	COI (%)	*gnd* (%)
Success	88.43	99.00
Ambiguous	8.04	1.00
Misidentification	3.53	0.00

sequences were misidentified in *COI* (3.53%) (Table 2). The dataset contained 1008 sequences (*COI* 510 and *gnd* 498). The best match of each sequence was an identical one. In order to detect reliability of data, 170 *COI* and 166 *gnd* were random sampled to 'best match'. Success was 85.88% in *COI* (24 were misidentified or ambiguous) and 98.19% in *gnd* (3 were ambiguous). Similarly, we performed another three random samplings: successes in *COI* were 85.04, 86.27 and 85.88%, respectively, and in *gnd* 98.39, 100.00 and 96.39%, respectively. The results showed that successful identification efficiency of *gnd* is significantly higher than *COI* ($t = 14.730$, $P = 0.000$).

Discussion

Effective discrimination of closely related species

Aphis lacks diagnostic morphological characteristics that distinguish many closely related species. In our *COI* NJ tree most species in the genus were poorly separated (Fig. 4) because of their low interspecific divergence (0.0683, SE = 0.0200) and high intraspecific divergence (0.0222, SE = 0.0352), although some species were well defined with respect to the barcode sequence. In contrast, all but three samples of each species formed distinct clusters based on *gnd*, which showed higher interspecific divergence (0.1075, SE = 0.251) and lower intraspecific divergence (0.0026, SE = 0.0031). A similar result was observed for the other genera, such as *Toxoptera* and *Cinara*, in which the maximal intraspecific variation was lower than the minimum interspecific variation. This clearly demonstrates the better ability of *gnd* to identify closely related species in comparison with the *COI* region.

Effective discrimination of closely related genera

In Hormaphidinae (Fig. 2), *Pseudoregma* and *Ceratovacuna* are two closely related genera. In the *COI* tree, these genera were nested within each other. In contrast, the two genera both formed distinct clusters for *gnd*. In Eriosomatinae, the gall formers, a large number of closely related genera are difficult to identify because of the lack of morphological variation. The *COI* region, several genera did not form their respective monophyla (Fig. 3). Compared with *COI*, species within each genus formed a distinct cluster for *gnd*. This suggests that *gnd* had the potential to discriminate between closely related genera.

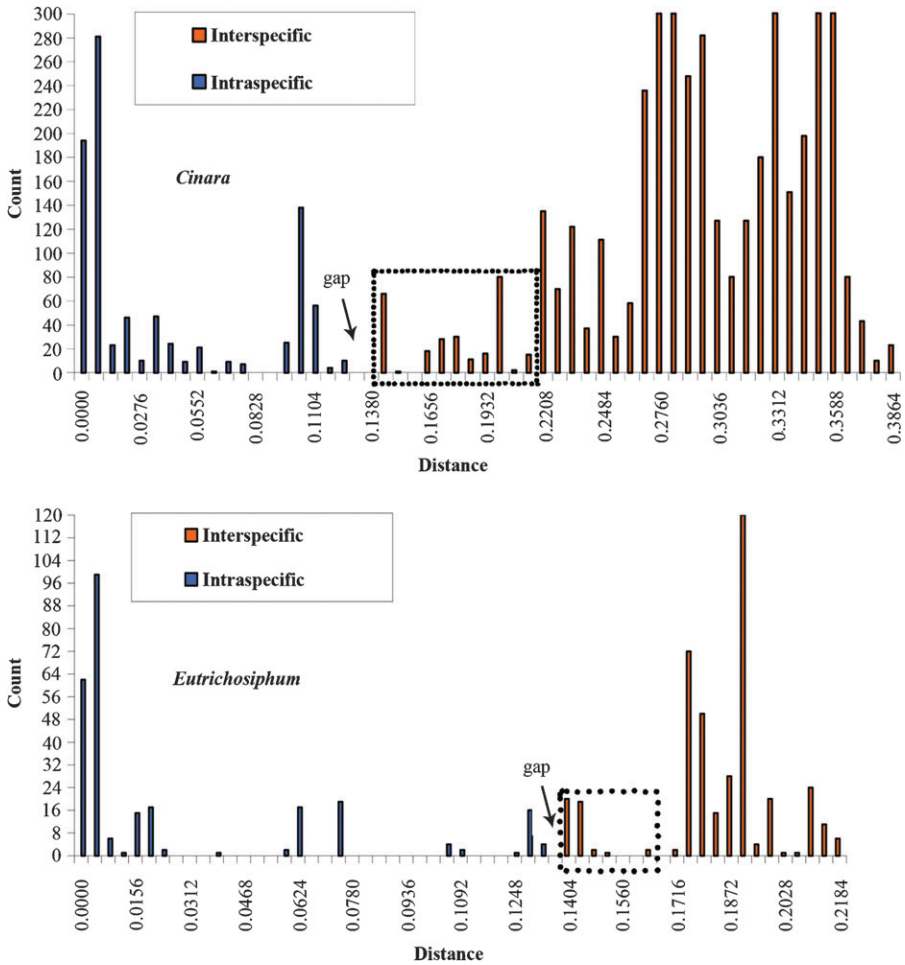

Fig. 5. Frequency histogram of intra- and interspecific genetic divergences in *Cinara* and *Eutrichosiphum* for the *gnd* region. The dashed box represents interspecific genetic divergences among closely related species. Divergences were calculated using the Kimura-2-parameter (K2P) model.

Assessment of the barcoding gap

Ideally the genetic variation of a DNA barcode should demonstrate separate, nonoverlapping distributions between intra- and interspecific samples. Moritz & Cicero (2004) and Meyer & Paulay (2005) reported that when the number of closely related species increased, the overlap of genetic variation without barcoding gaps increased accordingly. Our results showed that the distributions of intra- and interspecific variation of *gnd* exhibited gaps in most genera. In *Cinara*, the gap was small because the genus contained many closely related species, as in *Eutrichosiphum* (Fig. 5). The two genera still possessed gaps in the *gnd* analysis, but in the calculation of genetic distance using *COI*, there was significant overlap without gaps (Fig. 6). Furthermore, the mean interspecific divergence of *gnd* was obviously higher than that of the corresponding intraspecific variation (Fig. 5, Table 1). Consequently, *gnd* analysis possessed distinct intra- and interspecific variation gaps.

Comparison between gnd *and* COI

Both the *gnd* and *COI* genes demonstrated high efficiency of PCR amplification (90.23–93.10%). For older specimens, *gnd* had a higher amplification success rate. The *COI* primers used had a 6% probability of amplifying sequences from parasitic wasps that frequently lay their eggs in aphids. In contrast, the *gnd* primers amplified relatively conserved regions and showed specificity for *Buchnera*, although there was a small probability($< 1\%$) that sequences from other bacteria would be amplified (we only amplified five sequences from other bacteria). When we used 'best match' to identify species, *gnd* showed highly successful identification efficiency (99.00%) that was significantly higher than that of *COI* ($P = 0.000$) (Table 2). Thus, the *gnd* gene has the higher accuracy in aphid species identifications.

Hebert *et al*. (2003a,b) found that more than 98% of congeneric species have sufficient sequence divergence to ensure easy identification. However, the sequence divergence of *COI* for certain animal species, such as cnidarians (Hebert

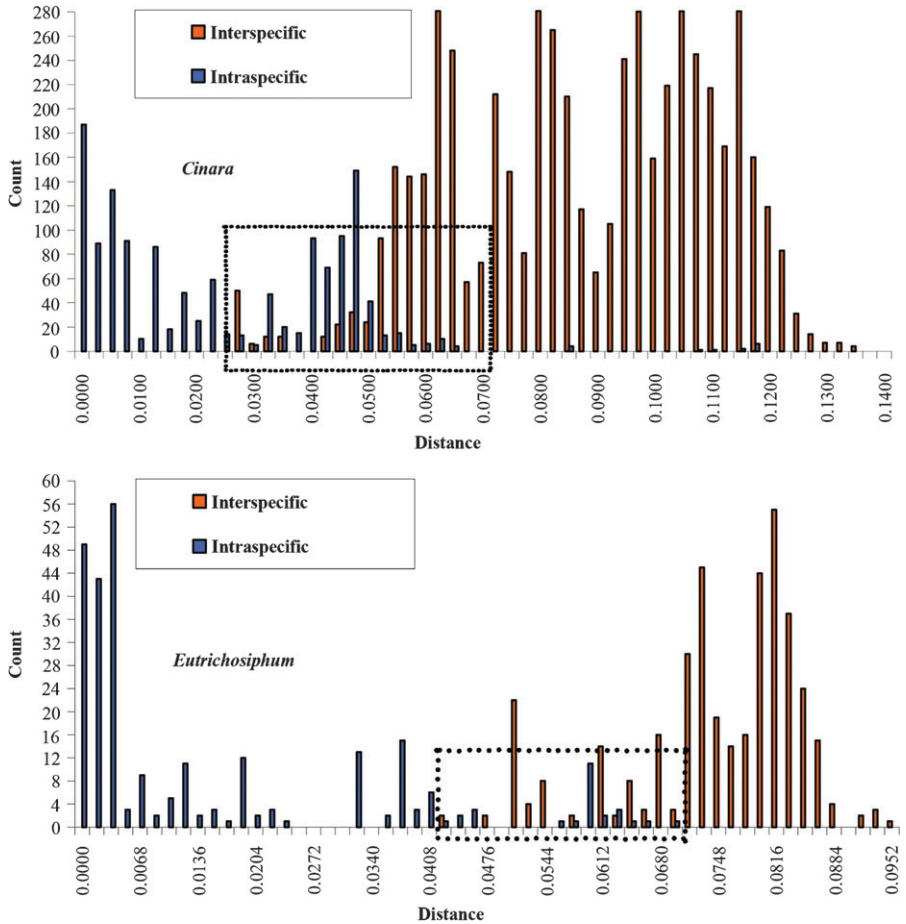

Fig. 6. Frequency histogram of intra- and interspecific genetic divergences in *Cinara* and *Eutrichosiphum* for the *COI* region. The dashed box represents interspecific genetic divergences among closely related species. Divergences were calculated using Kimura-2-parameter (K2P) model.

et al., 2003a,b) and the West Palaearctic *Pandasyopthalmus* taxa (Rojo *et al*., 2006), was much lower or even invariant. In Aphididae, *COI* showed lower interspecific divergence, and sometimes it was difficult to identify closely related species (Fig. 6, Table 1). In contrast to *COI*, for the *gnd* data, the intraspecific variations were much lower than the interspecific divergences, and the gap of genetic variation was also significant (Fig. 5, Table 1).

gnd *as an efficient aphid DNA barcode*

Our study strongly supports the *gnd* region as an efficient DNA barcode for aphids. *Buchnera* have undergone synchronous co-speciation with their animal hosts, as shown by phylogenetic relationships that track host matrilines (Wernegreen *et al*., 2001; Sabater-Munoz *et al*., 2005; Jousselin *et al*., 2009); therefore, the *gnd* region of *Buchnera* can represent its corresponding host. The *gnd* sequences were relatively easy to amplify using one pair of universal primers. The sequence variation of *gnd* was mostly derived from point mutations; therefore, it was also easy to generate the correct sequence alignment and distinguish closely related species. The *gnd* region possessed high interspecific divergence (Table 1) and was well separated. Analyses of the DNA barcoding gap supported the conclusion that the mean interspecific divergence of the *gnd* region was significantly higher than its mean intraspecific variation (Fig. 5, Table 1). Thus, the *gnd* region can be used to distinguish closely related species, as demonstrated in this study. Because synonymous sites in *Buchnera* genes evolve approximately twice as fast as those in the mitochondrial genes of their aphid host (Moran *et al*., 1995; Clark *et al*., 1999; Jousselin *et al*., 2009), the *gnd* region is a more effective and informative locus for species identification.

gnd + *COI as the standard barcode for aphids*

Recently, the CBOL (Consortium for the Barcode of Life) plant working group recommended using the 2-locus combination of *rbcL* + *matK* as a plant barcode. For animals, analysis of the *COI* region alone is insufficient for correct species identification. For example, Elias *et al*. (2007) recommended the addition of nuclear sequence data to identify problematic

Table 3. Accuracy of identification by using *gnd* +*COI*.

	Genera (%)	Species (%)
Success	99.5	98.5

species; Raupach *et al*. (2010) combined three nuclear ribosomal expansions to discriminate ground beetles. Feng *et al*. (2011) analysed DNA barcoding in Pectinidae based on the mitochondrial *COI* and *16S* rRNA genes, and Yang *et al*. (2011) identified mites using the *ITS2* and *COI* regions. Thus, to distinguish species, it has proven necessary to use another locus to complement the analysis of *COI*.

We suggest the combination of *gnd* + *COI* as a standard 2-locus barcode in aphids for several reasons. *COI* should not be replaced by other markers because it has many proven advantages. *COI* is a barcoding standard for multiple levels of sequence variation. For example, congeneric interspecific variation in aphids averaged 7.4 vs 7.93% in North American birds (Hebert *et al*., 2004), 9.93% in marine fishes (Ward *et al*., 2005) and 4.48–6.02% in Lepidoptera (Hajibabaei *et al*., 2006). According to the level of sequence variation, the identified species can be exactly classified at the genus level. In contrast, *gnd* analysis can distinguish closely related species. The *COI* region can easily distinguish the systematic position of species in most cases. *Gnd* can make up for the shortcomings of *COI* in cases of low sequence divergence that result in inaccurate identification. Using *gnd* + *COI* in the sample set examined here, species discrimination was successful in 99.5% of genera and 98.5% of species (Table 3). This 2-locus barcode will provide a universal framework for the routine use of DNA sequence data to identify specimens, and it will contribute to the discovery of overlooked species of aphids.

In conclusion, the *gnd* gene has a number of characteristics mentioned above that are useful for DNA barcoding. The combination of *gnd* + *COI* can act as a standard 2-locus barcode for identifying aphids, as well as certain species that harbour endosymbionts, such as Hemiptera. However, there are few *gnd* sequences in public databases, such as GenBank, and the volume *gnd* sequence data should be expanded to enable its effective use in DNA barcoding.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/syen.12018

Figure S1. Basal nodes of the neighbour-joining tree based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model. The major subfamilies or genera are expanded in Figs 2–4 and Figure S2.

Figure S2. Neighbour-joining trees based on the *gnd* region (left) and the *COI* region (right) using the Kimura-2-parameter (K2P) model in *Greenidea*, *Mollitrichosiphum* and *Eutrichosiphum*. *Buchnera* strains are represented by their host species names. The numbers of identical samples are given in brackets.

File S1. The detailed collection information and GenBank accession numbers of aphid species included in this study.

File S2. List of the species included in this study sorted by genus and the number of sequences.

Acknowledgements

We thank all colleagues for their assistance in aphid collection and Fen-Di Yang for making slides of the voucher specimens. We thank Teng Sun for her editorial help. Many thanks to editor Thomas Simonsen and two anonymous reviewers for their valuable comments. This work was supported by the National Science Fund for Distinguished Young Scientists (No. 31025024), the National Natural Sciences Foundation of China (No. 31272348), the National Science Fund for Fostering Talents in Basic Research (No. J1210002), and a grant from the Ministry of Science and Technology of the People's Republic of China (MOST Grant No. 2011FY120200).

References

Barcak, G.J. & Wolf, R.E. Jr. (1988) Comparative nucleotide sequence analysis of growth-rate-regulated *gnd* alleles from natural isolates of Escherichia coli and from Salmonella typhimurium LT-2. *Journal of Bacteriology*, **170**, 372–379.

Baumann, P., Moran, N.A. & Baumann, L. (2000) Bacteriocyte-associated endosymbionts of insects. *The Prokaryotes, a Handbook on the Biology of Bacteria; Ecophysiology, Isolation, Identification, Applications*, Vol. **1**, pp. 403–438. Springer–Verlag, New York, New York.

Bisercić, M., Feutrier, J.Y. & Reeves, P.R. (1991) Nucleotide sequences of the *gnd* genes from nine natural isolates of Escherichia coli: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *Journal of Bacteriology*, **173**, 3894–3900.

Blackman, R.L. & Eastop, V.F. (1984) *Aphids on the World's Crops. An Identification and Information Guide*. Wiley, Chichester.

Buchner, P. (1965) *Endosymbiosis of Animals with Plant Microorganims*. Interscience Publishers, Inc., New York, New York.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N. & Savolainen, V. (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B*, **360**, 1889–1895.

Chen, R., Jiang, L.Y. & Qiao, G.X. (2012) The effectiveness of three regions in mitochondrial genome for aphid DNA barcoding: a case in Lachininae. *PLoS ONE*, **7**, e46190.

Clark, M.A., Moran, N.A. & Baumann, P. (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Molecular Biology and Evolution*, **16**, 1586–1598.

Clark, M.A., Moran, N.A., Baumann, P. & Wernegreen, J.J. (2000) Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. *Evolution*, **54**, 517–525.

Coeur d'acier, A., Cocuzza, G., Jousselin, E., Cavalieri, V. & Barbagallo, S. (2008) Molecular phylogeny and systematic in

the genus Brachycaudus (Homoptera: Aphididae): insights from a combined analysis of nuclear and mitochondrial genes. *Zoologica Scripta*, **37**, 175–193.

Douglas, A.E. & Dixon, A.F.G. (1987) The mycetocyte symbiosis of aphids: variation with age and morph in virginoparae of Megoura viciae and Acyrthosiphon pisum. *Journal of Insect Physiology*, **33**, 109–113.

Dykhuizen, D.E. & Green, L. (1991) Recombination in Escherichia coli and the definition of biological species. *Journal of Bacteriology*, **173**, 7257–7268.

Eastop, V.F. (1977) Worldwide importance of aphids as virus vectors. *Aphids as Virus Vectors* (ed. by K. F. Harris and K. Maramorosch), pp. 3–62. Academic Press, London.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J. & Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B*, **274**, 2881–2889.

Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R. *et al*. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE*, **3**, e2802.

Feng, Y., Li, Q., Kong, L. & Zheng, X. (2011) DNA barcoding and phylogenetic analysis of Pectinidae (Mollusca: Bivalvia) based on mitochondrial *COI* and 16S rRNA genes. *Molecular Biology Reports*, **38**, 291–299.

Ferri, G., Alù, M., Corradini, B., Licata, M. & Beduschi, G. (2009) Species identification through DNA "barcodes". *Genetic Testing and Molecular Biomarkers*, **13**, 421–426.

Floyd, R.M., Wilson, J.J. & Hebert, P.D.N. (2009) DNA barcodes and insect biodiversity. *Insect Biodiversity: Science and Society* (ed. by R. G. Foottit and P. H. Adler), pp. 417–431. Wiley–Blackwell, Oxford.

Foottit, R.G. (1997) *Recognition of Parthenogenetic Insect Species*. Chapman & Hall, London.

Foottit, R.G., Halbert, S.E., Miller, G.L., Maw, E. & Russell, L.M. (2006) Adventive aphids (Hemiptera: Aphididae) of America north of Mexico. *Proceedings of the Entomological Society of Washington*, **108**, 583–610.

Foottit, R.G., Maw, H.E.L., Von Dohlen, C.D. & Hebert, P.D.N. (2008) Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Molecular Ecology Resources*, **8**, 1189–1201.

Foottit, R.G., Maw, H.E.L., Havill, N.P., Ahern, R.G. & Montgomery, M.E. (2009) DNA barcodes to identify species and explore diversity in the Adelgidae (Insecta: Hemiptera: Aphidoidea). *Molecular Ecology Resources*, **9**, 188–195.

Funk, D.J., Helbling, L., Wernegreen, J.J. & Moran, N.A. (2000) Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proceedings of the Royal Society B*, **267**, 2517–2521.

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.

Hajibabaei, M., Singer, G., Clare, E. & Hebert, P.D.N. (2007) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*, **5**, 24.

Hebert, P.D.N., Cywinska, A. & Ball, S.L. (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B*, **270**, 313–321.

Hebert, P.D.N., Ratnasingham, S. & de Waard, J.R. (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B*, **270**, S96–S99.

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

Heie, O.E. (1980) The Aphidoidea (Hemiptera) of Fennoscandia and Denmark. 1. General part. The families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae and Pemphigidae. *Fauna Entomologica Scandinavica*, **9**, 1–236.

Heie, O.E. (1987) Palaeontology and phylogeny. *Aphids: their biology, natural enemies and control, World Crop Pests, Volume 2A* (ed. by A. K. Minks and P. Harrewijn), pp. 367–391. Elsevier, Amsterdam.

Jousselin, E., Desdevises, Y. & Coeur d'acier, A. (2009) Fine-scale cospeciation between Brachycaudus and *Buchnera* aphidicola: bacterial genome helps define species and evolutionary relationships in aphids. *Proceedings of the Royal Society B*, **276**, 187–196.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369–8374.

Lahaye, R., Van Der Bank, M., Bogarin, D. *et al*. (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Lee, W., Kim, H., Lim, J. *et al*. (2011) Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. *Molecular Ecology Resources*, **11**, 32–37.

Lefevre, C., Charles, H., Vallier, A., Delobel, B., Farrell, B. & Heddi, A. (2004) Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial replacement. *Molecular Biology and Evolution*, **21**, 965–973.

Martinez-Torres, D., Buades, C., Latorre, A. & Moya, A. (2001) Molecular systematics of aphids and their primary endosymbionts. *Molecular Phylogenetics and Evolution*, **20**, 437–449.

McLean, D.L. & Houk, E.J. (1973) Phase contrast and electron microscopy of the mycetocytes and symbiotes of the pea aphid, Acyrthosiphon pisum. *Journal of Insect Physiology*, **19**, 625–629 631–633.

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.

Messing, R.H., Tremblay, M.N., Mondor, E.B., Foottit, R.G. & Pike, K.S. (2007) Invasive aphids attack native Hawaiian plants. *Biological Invasions*, **9**, 601–607.

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

Miller, G.L. & Foottit, R.G. (2009) The taxonomy of crop pests: the aphids. *Insect Biodiversity: Science and Society* (ed. by R.G. Foottit and P.H. Adler), pp. 463–473. Wiley-Blackwell, Oxford.

Moran, N.A., Dohlen, C.D. & Baumann, P. (1995) Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *Journal of Molecular Evolution*, **41**, 727–731.

Moritz, C. & Cicero, C. (2004) DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**, e354.

Munson, M.A., Baumann, P. & Kinsey, M.G. (1991) *Buchnera* gen. nov. and *Buchnera* aphidicola sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *International Journal of Systematic Bacteriology*, **41**, 566–568.

Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Madison, New York.

Nelson, K. & Selander, R.K. (1994) Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 10227–10231.

Nieto Nafría, J.M., Mier Durante, M.P. & Remaudière, G. (1998) Les noms des taxa du groupe-famille chez les Aphididae (Hemiptera). *Revue Française d'Entomologie*, **19**, 77–92.

Qiao, G.X., Jianfeng, W. & Jiang, L.Y. (2011) Use of a mitochondrial *COI* sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). *ZooKeys*, **122**, 1–17.

Raupach, M.J., Astrin, J.J., Hannig, K., Peters, M.K., Stoeckle, M.Y. & Wägele, J.W. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology*, **7**, 26.

Remaudière, G. & Remaudière, M. (1997) *Catalogue des Aphididae du Monde*. Institut National de la Recherche Agronomique, Paris.

Rojo, S., Stahls, G., Perez-Banon, C. & Marcos-Garcia, M.A. (2006) Testing molecular barcodes: invariant mitochondrial DNA sequences vs the larval and adult morphology of West Palaearctic Pandasyopthalmus species (Diptera: Syrphidae: Paragini). *European Journal of Entomology*, **103**, 443–458.

Sabater-Munoz, B., Flores, B.N. & Jones, M.G.K. (2005) DNA barcodes: a useful character to assist in identification of aphid species in the new millennium. Proceeding of the 7th International Symposium on Aphids, Fremantle, Australia, pp. 20–21.

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

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Coldspring Harbour Laboratory Press, Cold Spring Harbour, New York.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**, 81–86.

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13 486–13 491.

Teulon, D.A.J. & Stufkens, M.A.W. (2002) Biosecurity and aphids in New Zealand. *New Zealand Plant Protection*, **55**, 12–17.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Valenzuela, I., Hoffmann, A.A., Malipatil, M.B., Ridland, P.M. & Weeks, A.R. (2007) Identification of aphid species (Hemiptera: Aphididae: Aphidinae) using a rapid polymerase chain reaction restriction fragment length polymorphism method based on the cytochrome oxidase subunit I gene. *Australian Journal of Entomology*, **46**, 305–312.

Van Emden, H.F. & Harrington, R. (2007) *Aphids as Crop Pests*. Cromwell Press, Trowbridge.

Wang, J.F. & Qiao, G.X. (2009) DNA barcoding of genus Toxoptera Koch (Hemiptera: Aphididae): identification and molecular phylogeny inferred from mitochondrial *COI* sequences. *Insect Science*, **16**, 475–484.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D.N. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, **360**, 1847–1857.

Waugh, J. (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, **29**, 188–197.

Wernegreen, J.J., Richardson, A.O. & Moran, N.A. (2001) Parallel acceleration of evolutionary rates in symbiont genes underlying host nutrition. *Molecular Phylogenetics and Evolution*, **19**, 479–485.

Yang, B., Cai, J. & Cheng, X. (2011) Identification of astigmatid mites using ITS2 and *COI* regions. *Parasitology Research*, **108**, 497–503.

Accepted 1 March 2013
First published online 12 May 2013