The European Zoological Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tizo21

The species identification problem in mirids (Hemiptera: Heteroptera) highlighted by DNA barcoding and species delimitation studies

L. Piemontese , I. Giovannini , R. Guidetti , G. Pellegri , P. Dioli , L. Maistrello , L. Rebecchi & M. Cesari

To cite this article: L. Piemontese , I. Giovannini , R. Guidetti , G. Pellegri , P. Dioli , L. Maistrello , L. Rebecchi & M. Cesari (2020) The species identification problem in mirids (Hemiptera: Heteroptera) highlighted by DNA barcoding and species delimitation studies, The European Zoological Journal, 87:1, 310-324, DOI: 10.1080/24750263.2020.1773948

To link to this article: https://doi.org/10.1080/24750263.2020.1773948

The European Zoological Journal, 2020, 310–324
Vol. 87, No. 1, https://doi.org/10.1080/24750263.2020.1773948

The species identification problem in mirids (Hemiptera: Heteroptera) highlighted by DNA barcoding and species delimitation studies

L. PIEMONTESE[1†], I. GIOVANNINI[1†], R. GUIDETTI[1*], G. PELLEGRI[1], P. DIOLI[2], L. MAISTRELLO[1], L. REBECCHI[1], & M. CESARI[1]

[1]*Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy, and* [2]*Department of Entomology, Museo Civico di Storia Naturale di Milano, Milan, Italy*

(Received 29 November 2019; accepted 18 May 2020)

Abstract
Due to the difficulties associated with detecting and correctly identifying mirids, developing an accurate species identification approach is crucial, especially for potential harmful species. Accurate identification is often hampered by inadequate morphological key characters, invalid and/or outdated systematics, and biases in the molecular data available in public databases. This study aimed to verify whether molecular characterization (i.e. DNA barcoding) is able to identify mirid species of economic relevance and if species delimitation approaches are reliable tools for species discrimination. Cytochrome c oxydase 1 (*cox1*) data from public genetic databases were compared with new data obtained from mirids sampled in different Italian localities, including an old specimen from private collection, showing contrasting results. Based on the DNA barcoding approach, for the genus *Orthops*, all sequences were unambiguously assigned to the same species, while in *Adelphocoris*, *Lygus* and *Trigonotylus* there were over-descriptions and/or misidentifications of species. On the other hand, in *Polymerus* and *Deraeocoris* there was an underestimation of the taxonomic diversity. The present study highlighted an important methodological problem: DNA barcoding can be a good tool for pest identification and discrimination, but the taxonomic unreliability of public DNA databases can make this method useless or even misleading.

Keywords: Miridae, DNA barcoding, species delimitation, species identification, genetic databases, integrative taxonomy

1. Introduction

The family Miridae (Hemiptera: Heteroptera) is comprised of 11,020 described species, subdivided into eight subfamilies; members of the family are distributed across every biogeographic region (Schuh 1995, 2013; Cassis & Schuh 2012). Despite the large number of species, some have been more closely studied than others because of their economic relevance as major agricultural pests as well as biological control agents of weeds and other pests. Mirids exhibit great trophic plasticity, ranging from phytophagy, zoophagy, and omnivory. The phytophagous members of the family include pests that encompass all of the feeding ranges from monophagy to polyphagy, covering a broad taxonomic range of host plants. Interestingly, under certain conditions, some pests are able to switch from phytophagy to zoophagy, turning themselves into useful natural enemies that are employed in biological control programs, along with zoophagous species (Schaefer & Panizzi 2000; Wheeler 2001; Ferreira et al. 2015).

The magnitude of the damage caused by these organisms is often enhanced by the difficulties associated with detecting and correctly identifying mirid pests, particularly in the early stages of their diffusion, leading to an underestimation of their presence and their potential to cause harm. Species identification of

*Correspondence: R. Guidetti, Department of Life Sciences, University of Modena and Reggio Emilia, Modena, 41125, Italy. Email: roberto.guidetti@unimore.it

†These authors contributed equally to this work

insects based on morphological characters is often problematic without the involvement of an expert taxonomist, but an accurate mirid species identification is crucial for early detection and for prompt and cost-effective pest management. The key morphological differences between species may be very elusive, difficult to describe, and applicable only to some life history stages or to only one gender. Moreover, the identification uncertainties could increase in presence of samples represented only by eggs and instars, or in the presence of a local phenotypic or an high intraspecific variability (Schaefer & Panizzi 2000; Floyd et al. 2010; Ashfaq et al. 2016). In addition, some collection methods (e.g. sticky trap; frappage) could damage the insects while some sample preservation methods (e.g. mass dry and alcohol storage) could alter the shapes and/or colours of specimens, leading to species misidentification because key/diagnostic characters may not be visible (Wagner & Weber 1964; Schuh 1995, 2013; Schwartz 2008; Ferreira et al. 2015). To overcome these problems, the molecular identification of species (e.g. DNA barcoding) may be utilised; this approach allows for species determination at each life stage and can even enable identification when the animals are absent. In the latter case, DNA traces that they have left in the environment can be used (e.g. DNA metabarcoding of eDNA). DNA barcode-based identification has been demonstrated to be effective at discriminating a limited set of species, such as those occurring in a small area, agricultural pests, and invasive species (Meier 2008; Kress et al. 2009); it has also been an important tool for arthropod plant pest bio-surveillance (Ashfaq et al. 2016). Public databases such as GenBank and Barcode of Life Data System (BOLD) provide a large number of genetic sequences thanks to the reduced cost of molecular techniques and the reliability of DNA sequences for taxonomic identification (Ratnasingham & Hebert 2007; Foottit et al. 2008). Despite the abundance of data, molecular information is not immune to errors; such errors can result from sequencing mistakes, sample mix-ups, contamination, incorrect species identification, or other possible reasons (Shen et al. 2013).

Resolving species boundaries between closely-related species is notoriously difficult especially in case of invalid and/or outdated systematics, as in the case of mirids. According to morphological and molecular revisions, the Miridae present a large number of taxonomically difficult lineages with polyphyletic or paraphyletic species groupings (Park et al. 2011; Cassis & Schuh 2012; Raupach et al. 2014; Gwiazdowski et al. 2015). Therefore, genetic sequences and the use of different species delimitation methods could provide additional data in outlining species relationships and allow a revisitation of the taxonomy of mirids as part of an integrated approach. This study tested the utility of DNA barcoding to identify mirid species from Italian samples to overcome the taxonomic hurdles associated with traditional morphology-based identification. In addition, the information that has been obtained has been used to address taxonomic issues as well as the presence of problems in the molecular identification of investigated mirid species due to the misidentification of specimens, the presence of inaccuracies in public databases, the lack of recognition of cryptic diversity and/or the over-splitting of species.

2. Material and methods

2.1 Insect collection and morphological identification

Forty-two mirid specimens were collected in 2014–2016 in alfalfa (*Medicago sativa* L.) crops in seven Italian localities: six were in Modena province (Emilia Romagna) and one was in Novara province (Piedmont) (I). Specimens from Emilia Romagna were stored at −20°C immediately after collection, while the specimens from Piedmont were dried and stored at private room temperature. A 45 year old specimen from the collection of P. Dioli (Italy) sampled in Pordenone province (Friuli Venezia Giulia, Italy) was also included in the analyses to test the approaches with old and dried preserved specimens (Table I).

All plant bugs were observed with light microscopy (WILD M3Z, Heerbrugg) and individually identified at the morphological level by an expert taxonomist according to key characters described by Wagner and Weber (1964) and the following revisions: Schuh (1995, 2013), Schwartz (2008) and Ferreira et al. (2015). The coloration pattern of the head and the pronotum, the presence and morphology of the *arolia* and *pseudarolia* of the claws, the morphology of the male genitalia, and the setal abundance on elytrae were examined as morphological features. In addition, all specimens collected in field were photographed (Polaroid FCC) to obtain photo vouchers and retained at the Department of Life Sciences of the University of Modena and Reggio Emilia (Italy).

2.2 DNA extraction, PCR and sequencing

Total genomic DNA was isolated from one hind leg (from the third pair) of each specimens using the

Table I. List of mirid specimens used in this study: name of the morphospecies; retrieved haplotype (in bold haplotype definition); locality, date, geographic coordinates of sampling, and GenBank accession numbers. In brackets numbers of collected specimens.

Morphospecies	Haplotype	Locality	Collection date	Geographic coordinates	GenBank acc. no.
Adelphocoris lineolatus	(17) **A**	(1) Corlo, Modena	04.ix.2014	44°34′47”N 10°49′01”E	MH142526
		(1) Oleggio, Novara	03.ix.2015	45°35′27”N 08°37′55”E	MH142545
		(1) San Cesario sul Panaro, Modena	15.vii.2016	44°33′54”N 11°02′18”E	MH142542
		(7) Gaggio, Modena	15.vii.2016	44°37′44”N 11°00′42”E	MH142527
					MH142528
					MH142529
					MH142530
					MH142531
					MH142548
					MH142549
		(2) Nonantola, Modena	15.vii.2016	44°40′52”N 11°02′56”E	MH142536
					MH142537
		(4) Modena, Modena	15.vii.2016	44°38′07”N 10°59′00”E	MH142533
					MH142535
					MH142552
					MH142553
		(1) Budrione, Modena	22.vii.2016	44°49′01”N 10°51′24”E	MH142543
	(4) **B**	(1) Corlo, Modena	04.ix.2014	44°34′47”N 10°49′01”E	MH142550
		(1) San Cesario sul Panaro, Modena	15.vii.2016	44°33′54”N 11°02′18”E	MH142541
		(2) Modena, Modena	15.vii.2016	44°38′07”N 10°59′00”E	MH142534
					MH142551
	(3) **C**	(1) Oleggio, Novara	03.ix.2015	45°35′27”N 08°37′55”E	MH142547
		(2) Budrione, Modena	22.vii.2016	44°49′01”N 10°51′24”E	MH142539
					MH142540
	(1) **D**	(1) Oleggio, Novara	03.ix.2015	45°35′27”N 08°37′55”E	MH142546
	(2) **E**	(1) Oleggio, Novara	03.vi.2015	45°35′27”N 08°37′55”E	MH142544
		(1) Modena, Modena	15.vii.2016	44°38′07”N 10°59′00”E	MH142532
	(1) **F**	(1) Nonantola, Modena	15.vii.2016	44°40′52”N 11°02′56”E	MH142538
Lygus rugulipennis	(9) **G**	(3) San Cesario sul Panaro, Modena	15.vii.2016	44°33′54”N 11°02′18”E	MH142561

(*Continued*)

Table I. (Continued).

Morphospecies	Haplotype	Locality	Collection date	Geographic coordinates	GenBank acc. no.
					MH142562
					MH142563
		(3) Gaggio, Modena	15.vii.2016	44°37′44”N 11°00′42”E	MH142555
					MH142556
					MH142557
		(2) Nonantola, Modena	15.vii.2016	44°40′52”N 11°02′56”E	MH142558
					MH142559
		(1) Budrione, Modena	22.vii.2016	44°49′01”N 10°51′24”E	MH142560
Lygus wagneri	(1) **H**	(1) Passo Rest, Pordenone (PN), P. Dioli private collection	13.vi.1971		MH142554
Orthops kalmii	(1) **I**	(1) Modena, Modena	15.vii.2016	44°38′07”N 10°59′00”E	MH142567
Polymerus vulneratus	(1) **J**	(1) San Cesario sul Panaro, Modena	15.vii.2016	44°33′54”N 11°02′18”E	MH142564
Trigonotylus caelestialium	(1) **K**	(1) Modena, Modena	15.vii.2016	44°38′07”N 10°59′00”E	MH142568
Deraeocoris serenus	(1) **L**	(1) Gaggio, Modena	15.vii.2016	44°37′44”N 11°00′42”E	MH142566
	(1) **M**	(1) San Cesario sul Panaro, Modena	15.vii.2016	44°33′54”N 11°02′18”E	MH142565

EPICENTRE®MasterPureTM kit according to the manufacturer protocol. Three primers pairs found in the literature were tested for the amplification of the mtDNA gene *cox1*. The primer pair consisting of LepF1 (5′-ATT CAA CCA ATC ATA AAG ATA TTG G-3′, Hebert et al. 2004) and LepR1 (5′-TAA ACT TCT GGA TGT CCA AAA AAT CA-3′, Hebert et al. 2004) was initially used for some specimens of the morphospecies *Adelphocoris lineolatus* (Goeze, 1778), but then abandoned due to its limited ability to amplify specimens belonging to all other morphospecies. The pair LepF1 and LepR2 (5′-CTT ATA TTA TTT ATT CGT GGG AAA GC-3′, Hebert et al. 2004) was efficient but yielded short sequences, whereas the pair MHemF (5′-GCA TTY CCA CGA ATA AAT AAY ATA AG-3′, Park et al. 2011) and LepR1 led to the highest amplification success rates. Consequently, this last pair was used for the present analyses. Two PCR protocols were used based on the state of preservation of the specimens and their DNA. For well-preserved samples with good DNA quality, the following cycle was used: first denaturation step for 5 min at 94°C; 35 cycles with 30 s at 94°C; 30 s at 48°C, and 30 s at 72°C; and, a final elongation step at 72°C for 7 min. For degraded specimens with highly fragmented DNA, the PCR cycle was as follows: first denaturation step for 5 min at 94°C; 5 cycles with 30 s at 94°C, 40 s at 45°C, 1 min at 72°C; 35 cycles with 30 s at 94°C, 40 s at 51°C, and 1 min at 72°C; and a final elongation step at 72°C for 10 min. Amplicons were gel-purified using the Wizard Gel and PCR cleaning kit (Promega) and both strands were sequenced with an ABI Prism 3100 sequencer (Thermo Scientific). The chromatograms were checked for the presence of ambiguous bases as the sequences were translated to amino acids using the invertebrate mitochondrial code implemented in MEGA7 (Kumar et al. 2016) in order to check for the presence of stop codons and, therefore, of pseudogenes. Nucleotide sequences were aligned with the ClustalW algorithm as implemented in MEGA7 (pairwise and multiple alignment parameters; Gap opening penalty: 15; and Gap extend penalty: 6.66) and checked by visual inspection. All sequences have been deposited in GenBank (for accession numbers see Table I).

2.3 *DNA barcoding accuracy and specimen identification analysis*

Obtained *cox1* gene sequences were subjected to Nucleotide BLAST analyses (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to check for the presence of conspecific sequences in public genetic databases with over 95% similarities. All the sequences attributed to congeneric species found in GenBank and BOLD were included (Supplemental Table I; last download of sequences, January, 2017), in order to build up a more comprehensive taxonomic datasets for each identified genus. Interspecific and intraspecific Kimura's two-parameter (K2P) (Kimura 1980) distances for each dataset were calculated with the SPIDER (Species Identity and Evolution in R) v.1.5.0 package (Brown et al. 2012) implemented in R (R Core Team 2017). The reliability of datasets and the specimen identification accuracy were estimated through the *best close match analysis* and the *optimal threshold* optimization (Meier et al. 2006) in the SPIDER package. The barcoding identification efficacy of collected specimens was also tested using two specimen assignment methods implemented in the BarcodingR version 1.0.2 (Zhang et al. 2017) package implemented in R (R Core Team 2017): (a) the fuzzy set-based approach method (FZ), (Zhang et al. 2012), (b) the Bayesian-based (BI) method (Jin et al. 2013).

2.4 *Species delimitation analyses*

Species delimitation analyses were conducted on six separate datasets for each identified genus including the sequences obtained in the present study and the sequences retrieved in public databases. Pairwise nucleotide sequence divergences between scored haplotypes were computed in MEGA7. Relationships among haplotypes were estimated using a parsimony network by applying the method described by Templeton et al. (1992) and implemented in TCS 1.21 (Clement et al. 2000) using a 95% connection limit. The putative relationships were visualized using tcsBU (Múrias Dos Santos et al. 2015). The species delimitation analyses were performed using both the distance-based Automatic Barcode Gap Discovery (ABGD) method (Puillandre et al. 2012) and the tree-based Bayesian Poisson Tree Processes (bPTP) method (Zhang et al. 2013). Prior to analysis, datasets were adjusted to retain only one sequence for each haplotype, except those attributed to different species. ABGD settings were adjusted according to the characteristic of each dataset and were implemented on the ABGD website (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html). A Bayesian analysis was performed to obtain the consensus tree of the *cox1* sequences for each analysed plant bug genus by using MrBayes 3.2.3 (Ronquist et al. 2012) on Extreme Science and Engineering Discovery Environment (XSEDE 8.0.24) available on the

CIPRES Science Gateway v3.3 (Miller et al. 2010). MrModeltest 2.3 (Nylander 2004) was used to determine the best substitution model under Akaike's information criterion for each dataset. Two independent runs with four simultaneous Markov Chain Monte Carlo (MCMC) chains (1×10^6 generations, sampled every 100 generations, burn-in 25%) were executed; the GTR +I + G model was used for all analyses except in one case (*Orthops* dataset) for which the best substitution model resulted to be HKY. The bPTP analyses were conducted on a web server (http://species.h-its.org/ptp/) and were set up for 1×10^4 generations, with a thinning of 100 and burn-in of 0.1.

3. Results

The morphological and molecular data of this study unambiguously assigned the 43 specimens of Miridae to seven species belonging to six genera of the subfamilies Mirinae and Deraeocorinae: *Adelphocoris lineolatus, Lygus rugulipennis* (Poppius, 1912), *Lygus wagneri* (Remane, 1955), *Orthops kalmii* (Linnaeus, 1758), *Polymerus vulneratus* (Panzer, 1805), *Trigonotylus caelestialium* (Kirkaldy, 1902) and *Deraeocoris serenus* (Douglas & Scott, 1868). The *p*-distances between species range from a minimum of 7.2% (*O. kalmii* vs *L. rugulipennis*) to a maximum of 19.2% (*D. serenus* vs *P. vulneratus*) (Table II). In contrast, the *p*-distances within a single species were far lower (0–1.1%; Table II). The *cox1* data of these specimens were compared with sequences from public genetic databases (BOLD, GenBank) attributed to congeneric species showing contrasting results according to the genera.

3.1. *Genus* Adelphocoris

The complete *Adelphocoris* dataset was comprised of 28 sequences of *Adelphocoris lineolatus* obtained from Italian samples along with 215 sequences from GenBank and BOLD. The sequences belonged to 10 different species; only one was associated with an unidentified *Adelphocoris* species (Supplemental Table I).

Pairwise K2P nucleotide distances analysis showed mean values of 2.36% (range 0%–15%) of intraspecific and 2.38% (range 0%–5%) of interspecific distances (Supplemental Figure 1). The reliability of datasets and the specimen identification accuracy calculated for *Adelphocoris* dataset, using 0.89% for the optimal threshold, was 80%; the 20% error concerned sequences retrieved from public repositories (Supplemental Table II). The barcoding identification efficacy of newly analysed specimens showed success rates values of 63% (FMF scores 0) for the fuzzy set-based (FZ) method, and a posterior probability of 0.5 for the Bayesian-based (BI) method (Supplemental Table II). All the query sequences matched with reference sequences belonging to the same identified morphospecies.

Six haplotypes (A-F) were found among the 28 sequences of *Adelphocoris lineolatus* collected in Italy (Table I; Figure 1). These haplotypes were shared with specimens of *A. lineolatus* from Canada, U.S.A., Switzerland and Germany. The parsimony network analysis found four separate networks (Figure 1), three of which were formed by haplotypes belonging to single species: *Adelphocoris rapidus* (Say, 1832), *Adelphocoris seticornis* (Fabricius, 1775), and *Adelphocoris* sp. The other large network included haplotypes attributed to the other eight species. Each haplotype from these networks belonged to a single taxon with the only exception of one sequence

Table II. Uncorrected *p*-distance matrix of new analysed taxa. In brackets the number of specimens. d: average *p*-distance within single species; np: not possible.

	1 *Lygus wagneri*	2 *Adelphocoris lineolatus*	3 *Lygus rugulipennis*	4 *Polymerus vulneratus*	5 *Deraeocoris serenus*	6 *Trygonotylus caelestialium*	7 *Orthops kalmii*	d
1 *Lygus wagneri* (1)								np
2 *Adelphocoris lineolatus* (28)	0.154							0.011
3 *Lygus rugulipennis* (9)	0.072	0.145						0
4 *Polymerus vulneratus* (1)	0.156	0.171	0.156					np
5 *Deraeocoris serenus* (2)	0.155	0.166	0.152	0.192				0.002
6 *Trygonotylus caelestialium* (1)	0.169	0.185	0.164	0.186	0.177			np
7 *Orthops kalmii* (1)	0.117	0.148	0.104	0.169	0.158	0.164		np

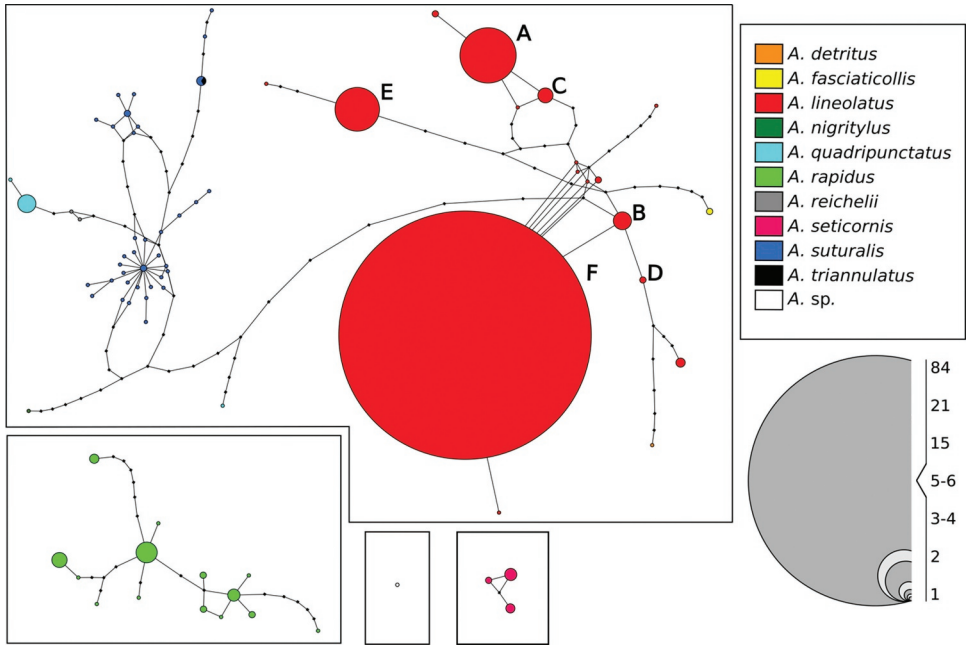

Figure 1. Parsimony network results for the genus *Adelphocoris*. Each circle represents a haplotype, while the size of the circle is proportional to the number of sequences. Different colours represent different species. Lines show single mutational step, while black squares denote missing/ideal haplotypes. Letters A-F indicate haplotypes retrieved in Italian specimens newly analysed in this study.

that was shared between *A. triannulatus* and *A. suturalis* (Figure 1). The largest network showed a well-structured clustering: all the haplotypes pertaining to a species grouped together and were linked to the group(s) of haplotype(s) of other species (Figure 1).

As mentioned above, each species delimitation analysis was conducted on a reduced dataset including only one sequence for each retrieved haplotype. For the genus *Adelphocoris*, out of 78 haplotypes, bPTP and ABGD analyses indicated the presence of 10 and 9 species, respectively (Figure 2). bPTP showed that the sequences attributed to four species [*Adelphocoris suturalis* (Jakovlev, 1882), *Adelphocoris triannulatus* (Stål, 1858), *Adelphocoris reichelii* (Fieber, 1836), and *Adelphocoris quadripunctatus* (Fabricius, 1794)] belonged to a single species, with the exception of a single sequence of *A. quadripunctatus* that was recognized as a separate species (Figure 2). The only difference in the ABGD analysis concerned the *A. detritus* haplotype, which was included in the species cluster with some *A. lineolatus* haplotypes.

3.2. *Genus* Lygus

The complete *Lygus* dataset comprised 1,572 sequences that were related to 32 species and several unidentified *Lygus* sp. (Supplemental Table I). A single sequence of *Lygus wagneri* and nine sequences of *L. rugulipennis* were obtained in this study. Mean K2P intra- and interspecific distances showed values of 1.93% (0–7.69%) and 0.72% (0–2.87%), respectively (supplemental Figure 1). The reliability of datasets and the specimen identification accuracy for *Lygus* dataset, using 0.21% for the optimal threshold, was 50% (Supplemental Table II). The barcoding identification efficacy of newly analysed specimens showed success rates values of 32% (FMF scores from 0 to 1) for the FZ method, and a posterior probability of 1 for the BI method. Not all the query sequences matched with reference sequences belonging to the same identified morphospecies (Supplemental Table II).

The haplotype of *L. wagneri* from Friuli Venezia Giulia (G, Figure 3) is shared with German specimens, whereas *L. rugulipennis* specimens from Emilia Romagna shared the same haplotype (H) with animals from Germany and Italy (Figure 3). One large and two small networks were found by the parsimony analysis (Figure 3). The first small one included exclusively sequences of the species *Lygus vanduzeei* (Knight, 1917). The second small network comprised haplotypes pertaining to four species: *Lygus pratensis* (Linnaeus, 1758), *Lygus gemellatus* (Herrich-Schaeffer, 1835), *L. wagneri* and a single sequence of *L. rugulipennis*. In this network, two haplotypes were shared among three species, one among *L. pratensis* - *L. gemellatus* - *L. wagneri*, and another one among *L. pratensis* - *L. gemellatus* - *L. rugulipennis*. The large

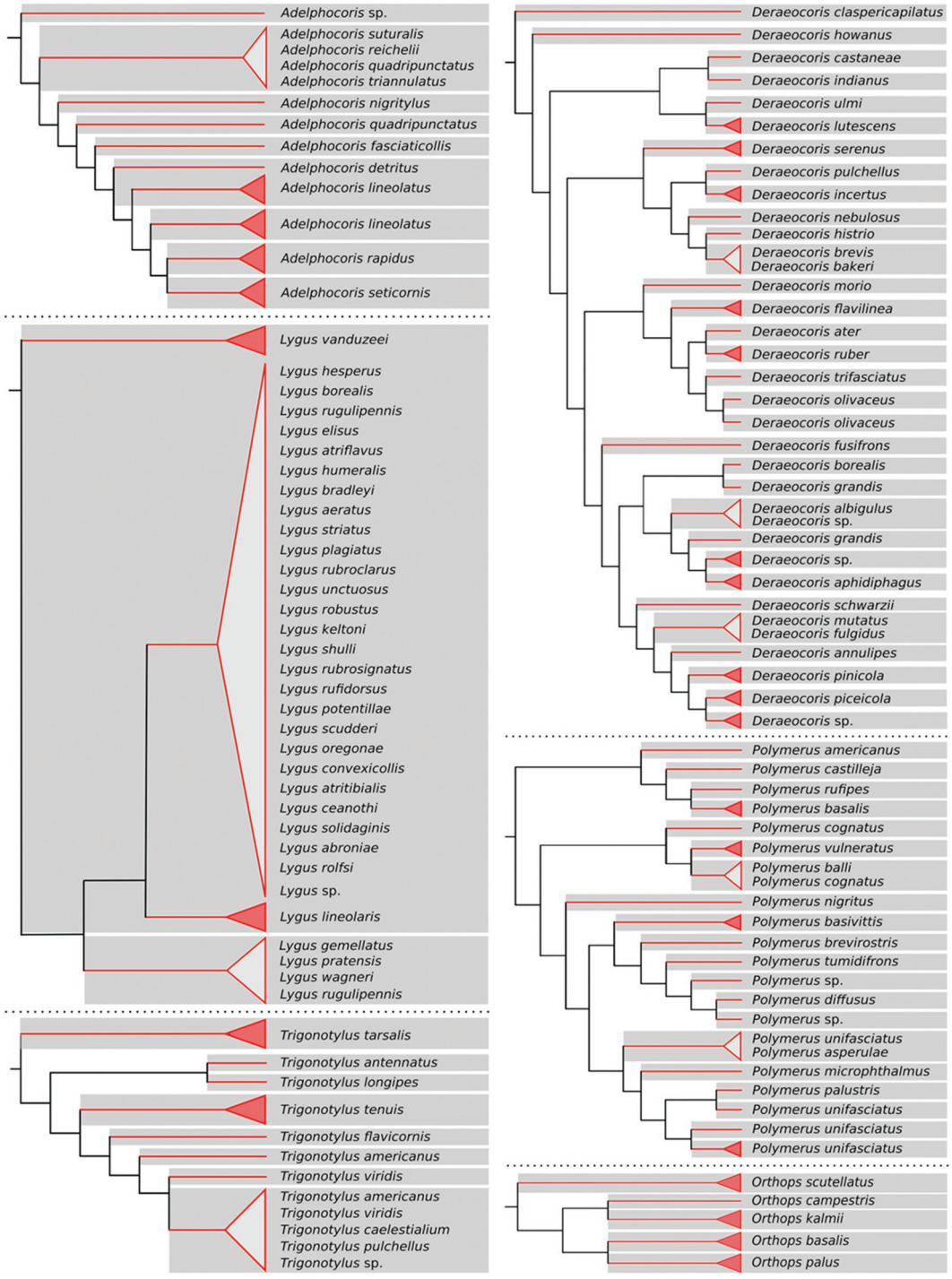

Figure 2. Comparison of species delimitation results for all analysed genera from bPTP and ABGD analyses. bPTP trees indicate putative species using transition from black-coloured branches to red-coloured branches. Triangles denote a group with more than one haplotype: red triangles group haplotypes attributed to the same morphospecies; grey triangles group haplotypes attributed to different morphospecies. Grey rectangles represent groups supported by ABGD.

network comprised sequences of all other species of *Lygus* and showed a very peculiar haplotype distribution pattern. Only two species, *Lygus lineolaris* (Palisot de Beauvois, 1818) and *L. rugulipennis* (although a haplotype belonged to a different network) formed well defined clusters in this network, whereas the

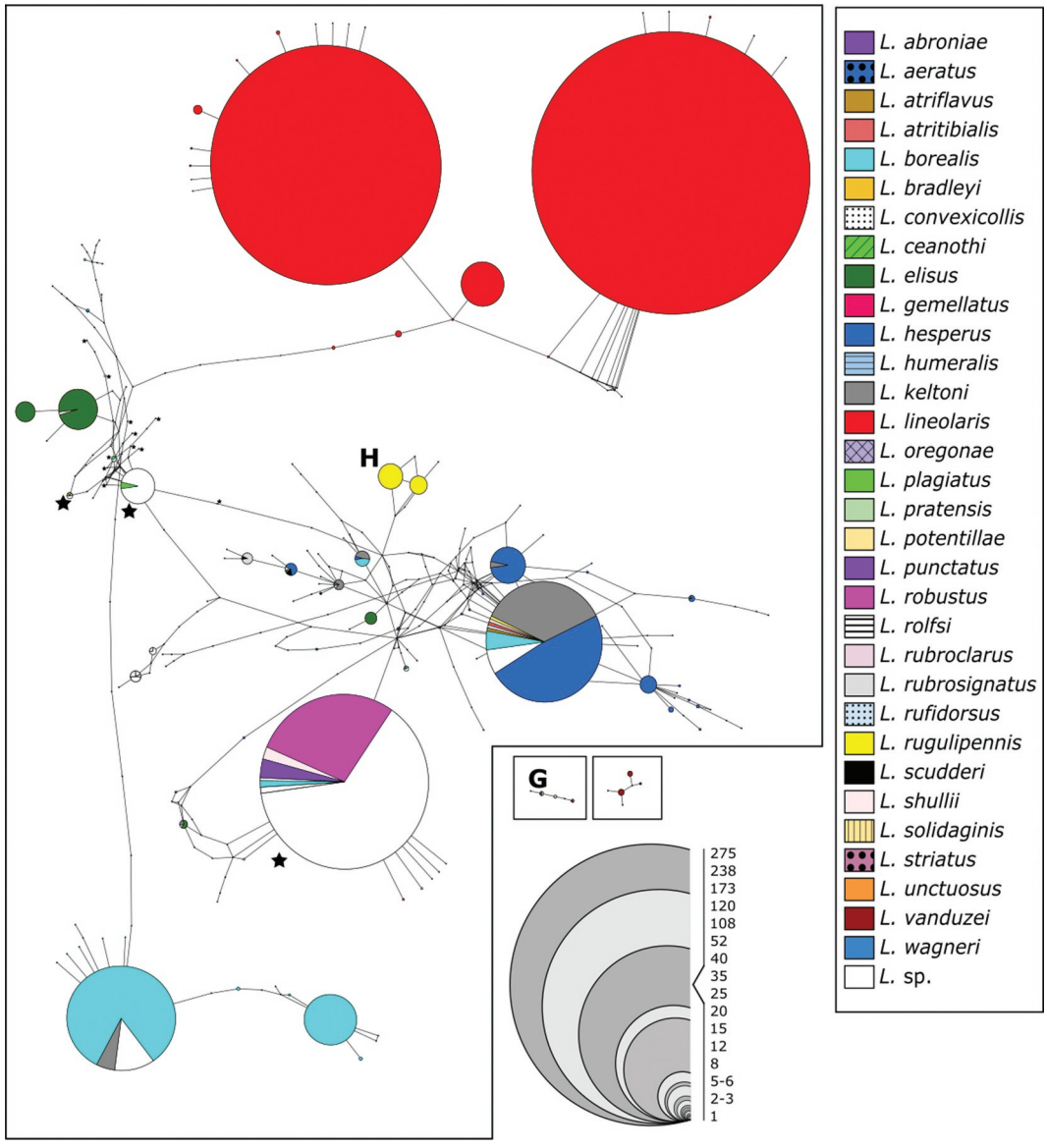

Figure 3. Parsimony network results of the genus *Lygus*. Each circle represents a haplotype. The size of the circle is proportional to the number of sequences, while different colours represent different species. Lines show a single mutational step, while black squares denote missing/ideal haplotypes. Letters G-H indicate haplotypes retrieved in Italian specimens newly analysed in this study. Stars indicate haplotypes also retrieved from Barcode of Life Data System (BOLD).

remaining species showed haplotypes shared with specimens attributed to different species, or haplotypes which did not cluster together with those of the same species.

A reduced dataset of 206 haplotypes was retained for species delimitation analyses (Figure 2). bPTP indicated the presence of four clusters: two of them corresponded to the two species *L. vanduzeei* and *L. lineolaris*, respectively, one to the four species *L. pratensis* - *L. gemellatus* - *L. wagneri* - *L. rugulipennis*, while the last one gathered all the remaining species. On the other hand, ABGD results indicated the presence of only two species, one comprising *L. pratensis* - *L. gemellatus* - *L. wagneri* - *L. rugulipennis*, and the other one composed by all the remaining sequences.

3.3. *Genus* Orthops

The complete dataset comprised 190 sequences related to five different species (Supplemental Table I). In this study, a single sequence of *Orthops kalmii* was retrieved. Mean K2P intra and interspecific distances showed values of 0.87% (0–2.15%) and 11.16%

(3.43–13.90%), respectively (Supplemental Figure 1). The reliability of datasets and the specimen identification accuracy was 100% using 3.47% for the optimal threshold (Supplemental Table II). The barcoding identification efficacy of the newly analysed specimen showed success rates values of 24% (FMF score 0.99) for the FZ method, and a posterior probability of 1 for the BI method (Supplemental Table II).

The retrieved haplotype grouped together with sequences of the same species from Germany (I, Figure 4). For this genus, the parsimony network analysis returned well separated networks for each species (Figure 4). The species delimitation analysis on 22 scored haplotypes supported the differentiation of the five species in both analyses (Figure 2).

3.4. *Genus* Polymerus

The complete *Polymerus* dataset comprised 54 sequences (Supplemental Table I), belonging to 16 species and five sequences of unidentified *Polymerus* sp. One sequence of a *Polymerus vulneratus* was found in this study. Mean K2P intra and interspecific distances showed values of 4.65% (0–19.34%) and 7.61% (0–15.37%), respectively (Supplemental Figure 1). The reliability of datasets and the specimen identification accuracy was 63% using 2.66% for the optimal threshold (Supplemental Table II). The barcoding identification efficacy of the newly analysed specimen showed success rates values of 51% (FMF score 0) for the FZ method, and a posterior probability of 0.5 for the BI method (Supplemental Table II).

The newly analysed sequence grouped together with sequences of the same species from Canada (haplotype J, Figure 4). Parsimony network analysis showed the presence of 20 networks (Figure 4), 12 of these networks each identified a single species. Sequences of *Polymerus cognatus* (Fieber, 1858) were present in two different networks, one composed by only one haplotype, and the other included also two haplotypes of *Polymerus balli* (Knight, 1925) and unidentified sequences of *Polymerus* sp. The sequences of *Polymerus unifasciatus* (Fabricius, 1794) were found in four different networks. In one of them, they were grouped with an unidentified sequence attributed to *Polymerus* sp., while in another one the *P. unifasciatus* haplotype was shared with *Polymerus asperulae* (Fieber, 1861), representing the only case of a haplotype shared between *Polymerus* species. Two networks were formed by one haplotype each of unidentified *Polymerus* sp. Species delimitation analyses performed on 28 haplotypes showed the presence of 20 species both in bPTP and ABGD, in agreement with the parsimony network results (Figure 2).

3.5. *Genus* Trigonotylus

The complete *Trigonotylus* dataset comprised 87 sequences related to 10 species and 12 unidentified species of this genus (Supplemental Table I). One sequence of *Trigonotylus caelestialium* was found in this study. Mean K2P intra and interspecific distances showed values of 2.13% (0–10.27%) and 2.19% (0–13.95%), respectively (Supplemental Figure 1). The reliability of datasets and the specimen identification accuracy was 24% using 3.56% for the optimal threshold (Supplemental Table II). The barcoding identification efficacy of the newly analysed specimen showed success rates values of 34% (FMF score 0) for the FZ method, and a posterior probability of 0.11 for the BI method (Supplemental Table II).

The retrieved sequence was identical to sequences of the same species found in Canada (haplotype K, Figure 4). The parsimony network analysis showed eight different networks (Figure 4), five of them identified well differentiated species. The sequences pertaining to the two species *Trigonotylus viridis* (Provancher, 1872) and *Trigonotylus americanus* (Carvalho, 1957) either formed private networks (with only one haplotype each), or were found in one large network that grouped together haplotypes of several unidentified sequences of *Trigonotylus* sp. and five different species. The same putative eight species were also found in the species delimitation analyses performed on 25 haplotypes (Figure 2).

3.6. *Genus* Deraeocoris

The complete dataset was made up of 124 sequences belonging to 30 different species and to unknown mirid members named *Deraeocoris* sp. (Supplemental Table I) and included the two sequences of *Deraeocoris serenus* found in this study. Mean K2P intra and interspecific distances showed values of 4.24% (0–20.72%) and 6.82% (0–16.48%), respectively (Supplemental Figure 1). The reliability of datasets and the specimens identification accuracy was 74% using 1.86% for the optimal threshold (Supplemental Table II). The barcoding identification efficacy of newly analysed specimens showed success rates values of 69% (FMF scores 0.89–0.95) for the FZ method, a posterior probabilities of 1 and for the BI method (Supplemental Table II).

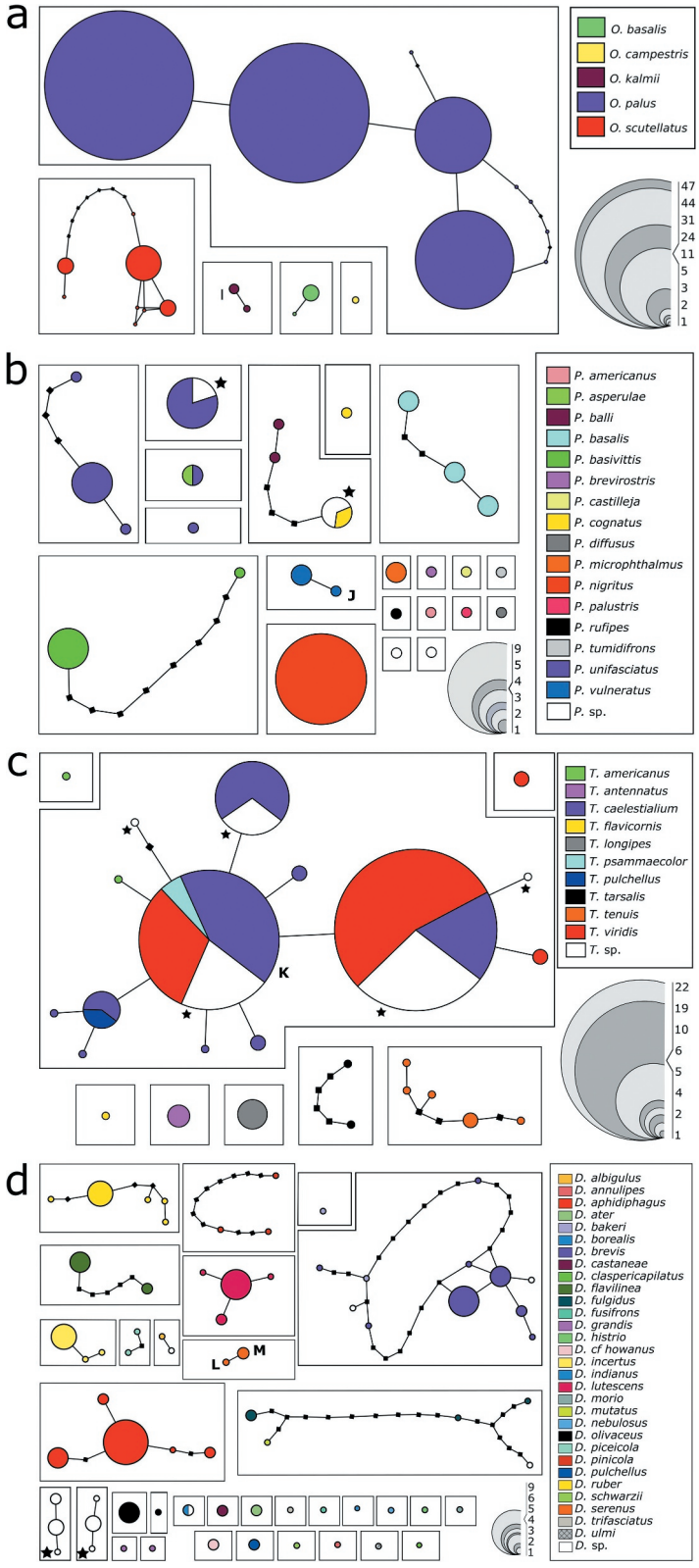

Figure 4. Parsimony network results. (a). Genus *Orthops*. (b). Genus *Polymerus*. (c). Genus *Trigonotylus*. (d). Genus *Deraeocoris*. Each circle represents a haplotype, while the size of the circle is proportional to the number of sequences. Different colours represent different species. Lines show a single mutational step, while black squares denote missing/ideal haplotypes. Letters I, J, K, L, M indicate haplotypes retrieved in Italian specimens newly analysed in this study. Stars indicate haplotypes also retrieved from Barcode of Life Data System (BOLD).

The two retrieved sequences (haplotypes L, M) grouped together in the same network with the haplotype M also being found in specimens of the same species from France (Figure 4). The parsimony network analysis resulted in 33 different networks (Figure 4), 22 of which included haplotypes pertaining to single species. One network was represented by a single haplotype shared by *Deraeocoris albigulus* (Knight, 1921) and by unidentified *Deraeocoris* sp. One network included haplotypes of the three species *Deraeocoris mutatus* (Knight, 1921), *Deraeocoris fulgidus* (Van Duzee, 1914) and *Deraeocoris* sp., while another grouped *Deraeocoris brevis* (Uhler, 1904), *Deraeocoris bakeri* (Knight, 1921) and *Deraeocoris* sp. together. However, one haplotype of *D. bakeri* formed a separate network. The haplotypes of *Deraeocoris grandis* (Uhler, 1887) and *Deraeocoris olivaceus* (Fabricius, 1777) were found in two different networks for each of the two species, and not in a single private network. Species delimitation analyses were congruent with the network analysis (Figure 2). On a reduced dataset of 70 haplotypes, 32 different species were detected by bPTP and ABGD; this is one less compared to the previously described network analysis due to the merging of the single *D. bakeri* haplotype with the group of *D. bakeri* - *D. brevis* haplotypes.

4. Discussion

The results from the present study showed that the mitochondrial *cox1* sequence used for molecular characterization accurately identified the seven mirid species that were collected from the field. Moreover, molecular identification worked with a dried specimen that was older than 45 years stored at room temperature. This is a further demonstration of the importance of the use of material from old collections (Virgilio et al. 2010; Mitchell 2015; Timmermans et al. 2016). Despite showing a clear-cut agreement between the morphological and molecular identification of the analysed specimens, the overall results of the present study revealed contrasting results when comparing our data with those present in GenBank and BOLD. For some mirid species, our data pointed out the lack of reliability of public databases, thus jeopardising the effectiveness of DNA barcoding as a tool for species identification. Two main issues have surfaced from most of the analysed datasets. The first issue concerns the presence of many mismatches between the molecular sequence and the taxonomic name assigned to the specimen from which it had presumably been obtained (Figure 5). The second issue concerns the quality of the information linked to the retrieved sequences (e.g. missing or incomplete data regarding the taxonomy and the geographic origin). Except for the genus *Orthops*, for which all sequences were assigned to an unambiguous species, the incongruences retrieved in the remaining genera appeared to be the result of a combination of several problems.

In five out of six datasets, some erroneous assignments of the sequences to the corresponding species and/or several sequences unassigned to any species have been observed. The species delimitation approaches used in this study showed that the genera *Adelphocoris*, *Trigonotylus* and *Lygus* need in-depth

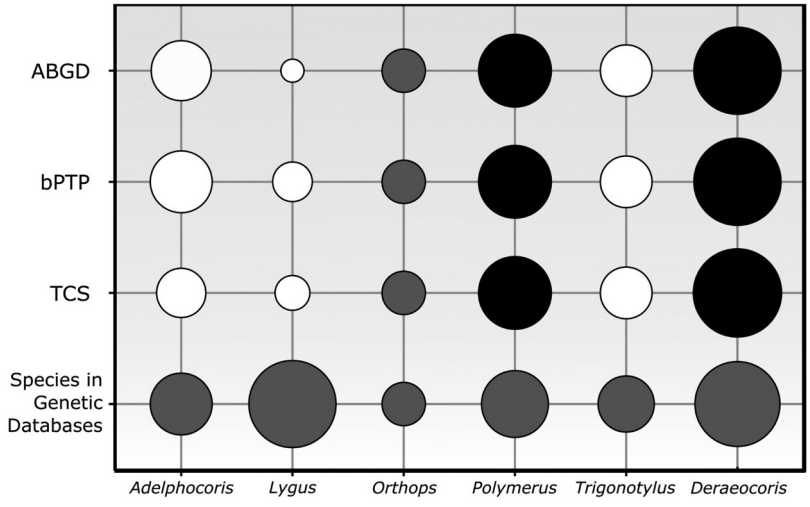

Figure 5. Graphic representation of species identification analysis. The circles are proportional to the number of species identified by each method (ABGD, bPTP, TCS) compared to the number of species in the datasets. Colours indicate lower (white) and higher (black) number of species with respect to those listed in datasets (grey).

taxonomic revisions applying an integrative approach. Furthermore, the two genera *Lygus* and *Adelphocoris* were described centuries ago by different authors and from different continents, thus enhancing the systematic over-splitting (Schwartz 2008). In the past, many pest species found on different plants have been considered as different taxa, even if they were likely to be pertaining to one taxon, as demonstrated by other studies that applied the DNA barcoding method (for a review see Ashfaq et al. 2016). On the other hand, molecular taxonomy suffers from the presence of species represented by a single barcode sequence and/or a high number of misidentified sequences, thus affecting molecular identification (Virgilio et al. 2010). The lack of characters and the difficulty of discrimination may have resulted in an underestimation of the taxa diversity in the genera *Polymerus* and *Deraeocoris*. In addition, the assessment of high or low intraspecific diversity could be explained by the presence of cryptic or sibling species. However, these cases are difficult to address because accurate references to the populations or exhaustive information about the characters are missing. Ultimately, the high number of unassigned sequences could be related to the difficulty of identifying Miridae species without revised identification keys by expert taxonomists, most notably when specimens have been collected with methodologies that do not allow proper preservation of the body parts of the animal, or when the samples are too degraded. In these cases, DNA barcoding is a useful tool to solve complex questions and to detect potential errors in GenBank, but only if the reference databases have been set up adequately (see also Ashfaq et al. 2016).

5. Conclusions

One of the main goal of DNA barcoding is to overcome the need for morphological identification of the specimens under study, so it works only when validated species identifications are present in databases. The DNA barcoding approach has been shown to be a good tool for identifying mirid species, particularly when all of the related sequences are correct and supported by solid information and references. Therefore, databases hampered by taxonomic/systematic problems can become reliable only when an integrated approach is applied and database quality control standards are more rigorously implemented.

This study supports the concluding statement of Meyer and Paulay (2005): "DNA barcoding is much less effective for identification in taxa where taxonomic scrutiny has not been thorough, […] many species will appear to be genetically non-monophyletic because of imperfect taxonomy, contributing to a high error rate for barcode-based identification." With the increase in non-curated molecular data available in the public databases and the decrease in the number of expert taxonomists (i.e. taxonomic impediment), this powerful method (DNA barcoding) will become unreliable or even misleading.

Acknowledgements

The authors wish to thank Prof. Aparna Palmer and Prof. Diane Nelson for the English revision of the text. We are grateful to the following peoples for providing specimens: Stefano Caruso, Giacomo Vaccari, Roberta Nannini, and Paolo Bortolotti of the "Consorzio Fitosanitario Provinciale" of Modena, Italy.

Funding

This work was supported by the "Fondazione Cassa di Risparmio di Modena", Italy, under Grant [2013.0665]: "Innovative tools and protocols for monitoring and sustainable control of the alien stink bug *Halyomorpha halys*, a new phytosanitary threat, and of other harmful heteropterans for the fruit crops of the territory of Modena".

Authors' contributions

R.G and M.C. designed and conceived the experiments. L.P., I.G. and G.P. conducted the experiments. L.P and I.G. analysed the data. P.D. and L. M. provided specimens and discussed the results together with the other Authors. L.P., I.G. and R. G. wrote the first draft of the manuscript. L.R., R. G. and L.M. provided specimens reagents, instruments, and funds. All authors participated in revising the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplementary material

Supplemental data for this article can be accessed here.

ORCID

R. Guidetti http://orcid.org/0000-0001-6079-2538
P. Dioli http://orcid.org/0000-0002-4274-0926
L. Maistrello http://orcid.org/0000-0002-2996-8993
L. Rebecchi http://orcid.org/0000-0001-5610-806X
M. Cesari http://orcid.org/0000-0001-8857-3791

References

Ashfaq M, Hebert PDN, Naaum A. 2016. DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. Genome 59:933–945. doi:10.1139/gen-2016-0024.

Brown ADJ, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, Cruickshank RH. 2012. Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources 12:562–565. doi:10.1111/j.1755-0998.2011.03108.x.

Cassis G, Schuh RT. 2012. Systematics, biodiversity, biogeography, and host associations of the miridae (Insecta: Hemiptera: Heteroptera: Cimicomorpha). Annual review of entomology 57:377–404. doi:10.1146/annurev-ento-121510-133533.

Clement M, Posada D, Crandall KA. 2000. TCS: A computer program to estimate gene genealogies. Molecular ecology 9:1657–1659. doi:10.1046/j.1365-294x.2000.01020.x.

Ferreira PSF, Henry TJ, Coelho LA. 2015. Plant Bugs (Miridae). In: Panizzi A, Grazia J, editors. True bugs (Heteroptera) of the Neotropics. Dordrecht: Springer Netherlands. pp. 237–286.

Floyd R, Lima J, de Waard J, Humble L, Hanner R. 2010. Common goals: Policy implications of DNA barcoding as a protocol for identification of arthropod pests. Biological invasions 12:2947–2954. doi:10.1007/s10530-010-9709-8.

Foottit RG, Maw HEL, Von Dohlen CD, Hebert PDN. 2008. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. Molecular Ecology Resources 8:1189–1201. doi:10.1111/j.1755-0998.2008.02297.x.

Gwiazdowski RA, Foottit RG, Maw HEL, Hebert PDN. 2015. The Hemiptera (Insecta) of Canada: Constructing a reference library of DNA barcodes. PLoS one 10:1–20. doi:10.1371/journal.pone.0125635.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences 101:14812–14817. doi:10.1073/pnas.0406166101.

Jin Q, Han HL, Hu XM, Li XH, Zhu CD, Ho SYW, Ward RD, Zhang AB. 2013. Quantifying Species Diversity with a DNA Barcoding-Based Method: Tibetan Moth Species (Noctuidae) on the Qinghai-Tibetan Plateau. PloS one 8:e644.

Kimura A. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120. doi:10.1007/BF01731581.

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences 106:18621–18626. doi:10.1073/pnas.0909820106.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution 33:1870–1874. doi:10.1093/molbev/msw054.

Meier R. 2008. DNA sequences in taxonomy: Opportunities and challenges. Systematics association special 76:95.

Meier R, Shiyang K, Vaidya G, Ng PKL. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Systematic Biology 55:715–728. doi:10.1080/10635150600969864.

Meyer CP, Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS biology 3(12):e422. doi:10.1371/journal.pbio.0030422.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Gateway computing environments workshop, 14-20 November 2010, New Orleans, Louisiana. pp. 1–8.

Mitchell A. 2015. Collecting in Collections: A PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. Molecular ecology resources 15:1102–1111. doi:10.1111/1755-0998.12380.

Múrias Dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M. 2015. TcsBU: A tool to extend TCS network layout and visualization. Bioinformatics 32:627–628. doi:10.1093/bioinformatics/btv636.

Nylander JA. 2004. MrModeltest v2.3. Program distributed by the author. Sweden: Evolutionary biology centre, Uppsala University. Issue January pp. 2–4.

Park DS, Foottit R, Maw E, Hebert PDN. 2011. Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). PLoS one 6:e18749. doi:10.1371/journal.pone.0018749.

Puillandre N, Lambert A, Brouillet S, Achaz G. 2012. ABGD, automatic barcode gap discovery for primary species delimitation. Molecular ecology 21:1864–1877. doi:10.1111/j.1365-294X.2011.05239.x.

R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org/.

Ratnasingham S, Hebert PDN. 2007. Barcoding, BOLD: The barcode of life data system. Molecular ecology notes 7:355–364. Available: www.barcodinglife.org.

Raupach MJ, Hendrich L, Chler SMK, Deister F, Rome J, Re M, Gossner MM. 2014. Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. PLoS one 9: e106940. doi:10.1371/journal.pone.0106940.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61:539–542. doi:10.1093/sysbio/sys029.

Schaefer CW, Panizzi AR. (Eds.). 2000. Heteroptera of economic importance. In: Schaefer CW, Panizzi AR, editors. Boca Raton, Florida: CRC Press. pp. 828

Schuh RT. 1995. Plant bugs of the world (Insecta, Heteroptera, Miridae): Systematic catalog, distributions, host list, and bibliography. New York: Entomological Society. pp. 1329.

Schuh RT. 2013. On-line systematic catalog of plant bugs (Insecta: Heteroptera: Miridae). Available: http://research.amnh.org/pbi/catalog/.

Schwartz MD. 2008. Revision of the Stenodemini with a review of the included genera (Hemiptera: Heteroptera: Miridae: Mirinae). Proceedings of the entomological society of Washington 110:1111–1201. doi:10.4289/0013-8797-110.4.1111.

Shen YY, Chen X, Murphy RW. 2013. Assessing DNA barcoding as a tool for species identification and data quality control. PLoS one 8:e57125. doi:10.1371/journal.pone.0057125.

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633.

Timmermans MJTN, Viberg C, Martin G, Hopkins K, Vogler AP. 2016. Rapid assembly of taxonomically validated mitochondrial genomes from historical insect collections. Biological journal of the Linnean Society 117:83–95. doi:10.1111/bij.12552.

Virgilio M, Backeljau T, Nevado B, De Meyer M. 2010. Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 11:206. doi:10.1186/1471-2105-11-206.

Wagner E, Weber HH. 1964. Hétéroptères miridae. In: Faune de France 67. Paris: Fédération Française des Sociétés de Sciences Naturelles. pp. 589

Wheeler AG. 2001. Biology of the plant bugs (Hemiptera: Miridae): Pests, predators, opportunists. Ithaca, New York: Comstock publishing associates. pp. 507.

Zhang AB, Hao MD, Yang CQ, Shi ZY. 2017. BarcodingR: An integrated r package for species identification using DNA barcodes. Methods in Ecology and Evolution 8:627–634. doi:10.1111/2041-210X.12682.

Zhang AB, Muster C, Liang HB, Zhu CD, Crozier R, Wan P, Feng J, Ward RD. 2012. A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. Molecular Ecology 21 (8):1848–1863. doi:10.1111/j.1365-294X.2011.05235.x.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876. doi:10.1093/bioinformatics/btt499.