

ORIGINAL ARTICLE

Borrowing the Pentatomomorpha tome from the DNA barcode library: Scanning the overall performance of *cox1* as a tool

Filipe Michels Bianchi^{1,2}  | Leonardo Tresoldi Gonçalves^{3,4} 

¹Laboratório de Entomologia Sistemática, Departamento de Zoologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

²Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

³Laboratório de Drosophila, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

⁴Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Correspondence

Filipe Michels Bianchi, Laboratório de Entomologia Sistemática, Departamento de Zoologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970, Porto Alegre, RS, Brazil. Email: bianchi.fm@hotmail.com

Funding information

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Conselho Nacional de Desenvolvimento Científico e Tecnológico

Abstract

Pentatomomorpha (Hemiptera: Heteroptera) occurs worldwide with almost 19,000 species within six superfamilies. Their wide morphological, behavioral, and ecological diversity is remarkable, making them subject of basic and applied studies. The mitochondrial *cytochrome c oxidase subunit I (cox1)* proposed as a DNA barcode for Metazoa is used in species discovery and identification, relying on threshold values to split intra- and interspecific sequences and a comprehensive library to accurate identification. Here, we scanned all the pentatomomorphan *cox1* sequences on Barcode of Life Data System library aiming to provide an overview of available data; verify barcoding gaps at generic level; infer individual empirical threshold values to distinguish congeneric species; and test efficiency of *cox1* based on the Probability of Correct Identification (PCI) analysis. Our final dataset comprised 12,189 sequences, covering five superfamilies, 32 families, 460 genera, and 1068 species. The dataset abundance and composition were biased to families with economic importance, that is, Pentatomidae, Lygaeidae, Scutelleridae, Coreidae, and Rhyparochromidae. Barcode gaps were detected for most of the analyzed genera, reaffirming the efficiency of *cox1* for Pentatomomorpha. We inferred threshold values for 131 genera and found a global PCI of 74.33%, suggesting that one out of four analyzed species suffer from operational biases or hide cryptic species. We brought examples to illustrate how *cox1* can be used to flag inconsistencies, refine, and shed light onto future studies on Pentatomomorpha. We emphasize the efficiency of *cox1* as DNA barcode for these true bugs, advocating for its combined use with, for example, morphology in integrative approaches.

KEYWORDS

Barcode of Life Data System, COI, Hemiptera, integrative taxonomy, true bugs

1 | INTRODUCTION

Pentatomomorpha (Heteroptera, the true bugs) are part of the most successful radiation of non-holometabolous insects, the Hemiptera (Weirauch & Schuh, 2011). They occur in all zoogeographical regions, except Antarctica, and thrive in tropical zones (Panizzi & Grazia, 2015). This infraorder encompasses six superfamilies (*sensu* Schuh & Weirauch, 2020): Aradoidea, Coreoidea, Idiostoloidea, Lygaeoidea,

Pentatomoidea, and Pyrrhocoroidea, including almost 19,000 species (Henry, 2017). The wide morphological, behavioral, and ecological diversity disclosed in Pentatomomorpha is remarkable (Figure 1).

These terrestrial true bugs comprise mycophagous, hematophagous, predators, and phytophagous species (Weirauch et al., 2019). Several predatory species are used as biocontrol agents, and phytophagous species—the majority of pentatomomorphans—feed on roots to flowers and seeds, with many species being serious crop

pests around the globe. Invasive pentatomomorphans intensify economic damages, mainly related to crops (McPherson, 2018). Since the Pentatomomorpha are subject of basic and applied studies, there is an urge to understand their lineage splits, diversification, and accurate identification.

Morphology is the foundation of pentatomomorph description and identification. But the morphological nuances that distinguish closely related species might be so complex that even expert taxonomists struggle in their diagnosis (Figure 1). Plastic phenotypes, sex-dependent characters, and the little knowledge for the identification of eggs and immature stages may further hinder this process. Alternative data sources have been expanding taxonomic studies, providing novelties to accurate identification, supporting synonyms, and consistently unveiling cryptic diversity (e.g., Hickmann et al., 2019; Zhang et al., 2017). In this scenario, molecular tools integrated to morphology may aid in organism identification and can be used to assess biodiversity quickly. DNA barcoding—the standardized use

of short nucleotide sequences for specimen identification (Hebert et al., 2003)—was already proved useful in pentatomomorph taxonomy (Park et al., 2011; Tembe et al., 2014).

Since the proposition of the 5' end of the mitochondrial cytochrome *c oxidase subunit I (cox1)* as a standard DNA barcode for Metazoa (Hebert et al., 2003), some studies built up sequence libraries for Pentatomomorpha, providing reference sequences to hundreds of species (e.g., Gwiazdowski et al., 2015; Raupach et al., 2014). Moreover, vast amounts of pentatomomorph barcoding *cox1* sequences are generated yearly (e.g., Gowande et al., 2018; Grebennikov & Heiss, 2014). These data supply Barcode of Life Data System (BOLD), a database dedicated to assembling barcode records that meet a series of criteria (Ratnasingham & Hebert, 2007). DNA barcoding as a tool impacts directly on pest management, forensics, conservation, environmental ecology, phylogeography, and many other biological disciplines. Since species are the fundamental unit for many biological subjects, DNA barcoding strongly improves

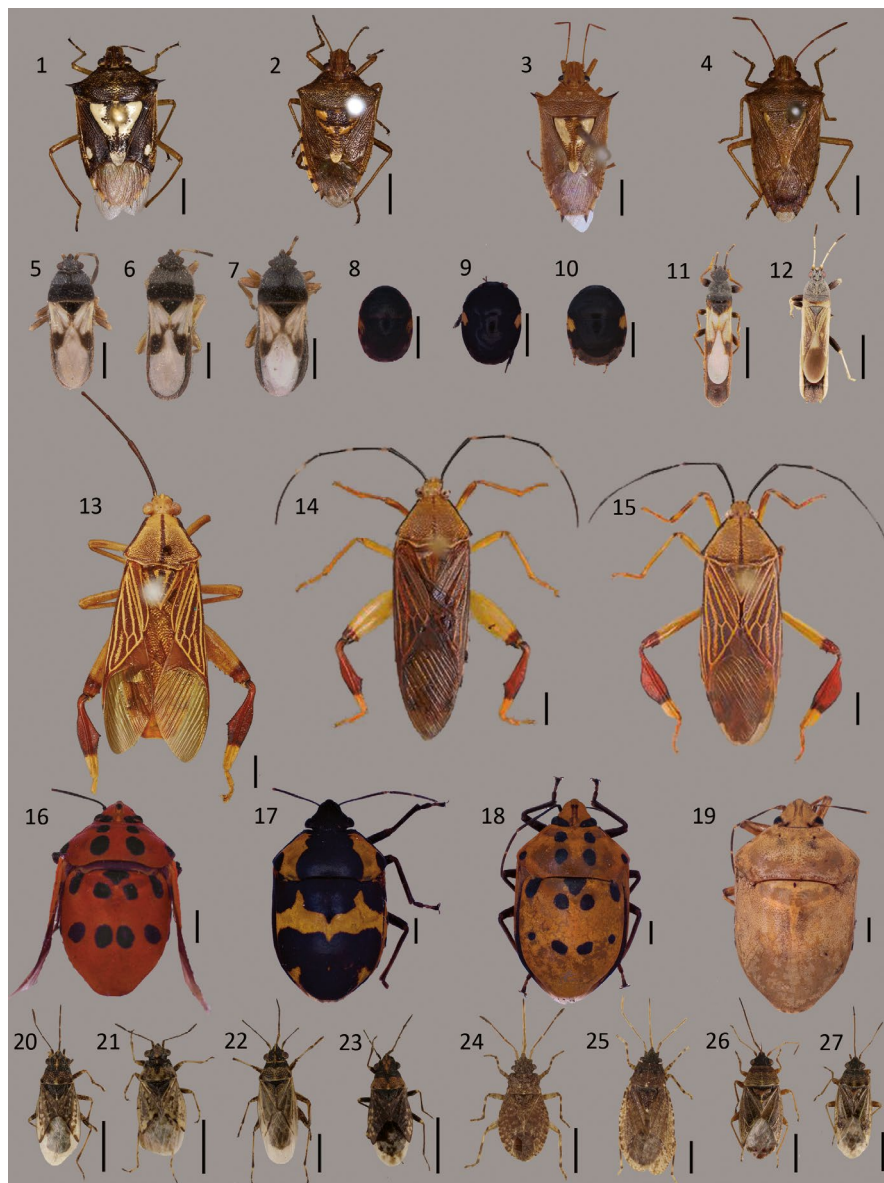


FIGURE 1 Morphological diversity of Pentatomomorpha, also showing intraspecific plasticity and similar taxa. 1–2 *Oebalus poecilus* (Dallas); 3–4 *Oebalus ypsilon* (De Geer); 5 *Blissus antillensis* Leonard; 6 *Blissus brasiliensis* Drake; 7 *Blissus leucopterus* (Say); 8 *Galgupha fossata* McAtee & Malloch; 9 *Galgupha difficilis* (Bredin); 10 *Galgupha differentialis* McAtee & Malloch; 11 *Ischnodemus lactipennis* Slater & Wilcox; 12 *Extrademus tylosis* Slater & Wilcox; 13 *Melucha lineatella* (Fabricius); 14 *Melucha dilatata* (Fabricius); 15 *Melucha chapadana* Brailovsky; 16–19 *Augocoris gomesii* Burmeister; 20 *Nysius nemorivagus* White; 21 *Nysius huttoni* White; 22 *Nysius delectus* White; 23 *Nysius blackburni* White; 24 *Nesocryptias villosa* (White); 25 *Metrarga nuda* White; 26 *Oceanides pteridicolus* (White); 27 *Oceanides vulcan* (White). Scale bars: 2 mm. Lygaeoidea species images available in Dellapé & Henry (2020), Lygaeoidea Species File. *Melucha* species images available in CoreoideaSF Team (2020), Coreoidea Species File Online

integrative taxonomy (Goulding & Dayrat, 2016). The increase of studies using *cox1* as a barcode mirrors the versatility of the marker and the method (DeSalle & Goldstein, 2019).

DNA barcoding manages two fundamental tasks: to diagnose species (i.e., specimen identification) and to discover new species (i.e., species delimitation) (DeSalle et al., 2005). Specimen identification is accomplished through comparisons with a reference database. In contrast, species discovery and delimitation use threshold values based on the interval between intra- and interspecific distances (i.e., the barcoding gap; Meyer & Paulay, 2005). Moreover, the barcoding gap may disclose outliers within a pool of congeneric samples, serving as a taxonomic red flag (e.g., Gonçalves et al., 2021). Since the availability of representative species and specimens (i.e., sequences) directly impact on DNA barcoding performance (Meyer & Paulay, 2005), threshold and gap values should be reviewed regularly whenever new samples are deposited in the databases.

Here, we mined all pentatomomorphan *cox1* sequences deposited on BOLD, aiming to (1) provide an overview of information available in BOLD; (2) verify the existence of a barcoding gap at generic level in Pentatomomorpha; (3) set up a threshold value that allows distinguishing intra- and interspecific taxa; and (4) test efficiency of *cox1* based on the Probability of Correct Identification (PCI) analysis. We also brought empirical data to illustrate how *cox1* can be used to flag inconsistencies and clue future studies on Pentatomomorpha.

2 | MATERIALS AND METHODS

2.1 | Data acquisition and filtering

Sequences were retrieved from BOLD on August 17, 2020, generating separate datasets for each pentatomomorphan family. We followed several filtering steps to ensure robust analyses (summarized

in Figure 2). First, we removed sequences belonging to other markers (or regions) than *COI-5P* (e.g., *COI-3P*, *COXII*) or without species-level identification (e.g., "*Euschistus*," "*Pentatomidae* sp."). Then, we made preliminary alignments of each dataset using MAFFT 7.0 (Katoh et al., 2019), keeping the parameters in default, and looking for non-sense mutations, insertions, and deletions. These sequences were removed from the dataset since we assumed they resulted from low-quality sequencing, erroneous amplification, or laboratory contamination. Additionally, misaligned sequences were used as a query for comparison with the NCBI database (<http://www.ncbi.nlm.nih.gov/>) using the Basic Local Alignment Search Tool—Nucleotide (BLASTn) to verify their identity. Non-pentatomomorphan sequences were purged from the dataset.

After these steps, we realigned the sequences (same parameters as above). The software AliView (Larsson, 2014) was used to visualize the alignments, verify the reading frame, and trim the sequences to restrict them to the canonical barcode region of *cox1* (Ratnasingham & Hebert, 2007). Sequences shorter than 400 base pairs were removed. As a final filtering step, we double-checked scientific names seeking for misspellings and non-valid names (i.e., synonyms). Then, we ran a final alignment for each dataset.

2.2 | Data analyses

For the barcoding gap analysis, we generated separate FASTA files for each genus represented in the datasets. We only analyzed genera featuring at least two species, and with at least one of the species being represented by more than one sequence, to ensure intra- and interspecific comparisons. For each genus, the R package SPIDER (Brown et al., 2012) was used to estimate uncorrected p-distances. Output values were visualized in a boxplot and we followed Badotti et al., (2017) to rank *cox1* efficiency based on three categories: *good*,

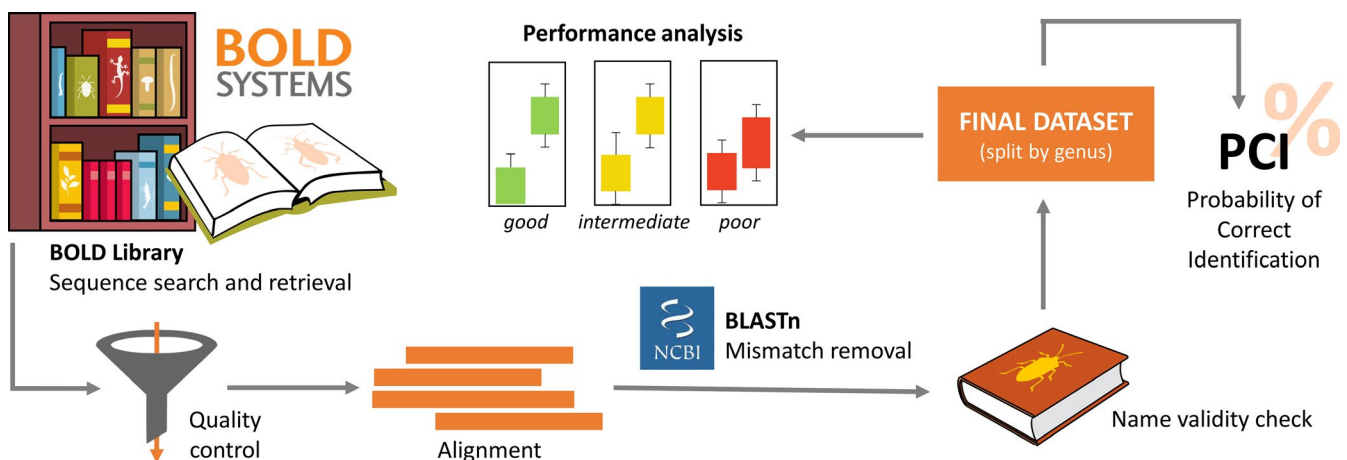


FIGURE 2 Workflow used in the present study to improve the final datasets before analyses. All sequences related to Pentatomomorpha were downloaded from Barcode of Life Data System; obtained dataset was filtered, maintaining only cytochrome *c* oxidase subunit I (*cox1*) sequences identified at species level; alignments conducted per family to check non-sense sequences; misaligned sequences submitted to Basic Local Alignment Search Tool—Nucleotide (in NCBI) to verify their identity affinity with Pentatomomorpha; validity of the names was checked

intermediate, and *poor*. Efficiency was considered *good* when boxplots disclosed a gap between intra- and interspecific distances, *intermediate* whenever the whiskers of intra- and interspecific distances overlapped, and *poor* when the boxes overlapped. We additionally used the function “localMinima” implemented in SPIDER to determine a threshold value for genus with a *good* *cox1* efficiency. This function optimizes putative threshold values from a density plot of the genetic distances (Brown et al., 2012).

Since the success of DNA barcoding for specimen identification is not necessarily related to a barcoding gap (see Discussion), we calculated the PCI (Hollingsworth et al., 2009) to measure the identification success when using *cox1* sequences. For each species, we considered the maximum intraspecific distance and the minimum interspecific distance (i.e., nearest-neighbor distance). If the maximum intraspecific distance of a species was less than its minimum interspecific distance, identification for that species was considered a success (Hollingsworth et al., 2009). PCI was calculated as the percentage of species correctly identified. Species represented by only one sequence were not analyzed, since it was not possible to obtain intraspecific distances in such cases. PCI rates were then graphically represented in a scatter plot, as suggested by Collins and Cruickshank (2012).

Since some family groups and their rankings are not consensual with current classification, we opted to present data for Thaumastellinae and Thyreocorinae, treated as subfamilies of Cydnoidea for some authors (e.g., Schuh & Weirauch, 2020) but as independent families in other classifications (e.g., Grazia et al., 2008; Lis et al., 2017; Rider et al., 2018). Here, we adopted the classification of Pentatomomorpha and numbers of species and genera per family presented by Schuh and Weirauch (2020).

3 | RESULTS

Our raw dataset consisted of 18,250 sequences. Removing sequences of non-barcode markers or without species-level identification reduced this number to 12,892. During the preliminary alignment step, we found a small number of sequences with insertions and deletions; we also detected nine sequences matching with non-pentatomomorphan taxa (e.g., bacteria, lepidopterans, and dipterans) in BLASTn. The removal of these entries led to 12,761 sequences. After purging sequences shorter than 400 bp, our final dataset comprised 12,189 pentatomomorphan sequences representing five superfamilies, 32 families, 460 genera, and 1068 species (Table S1). The global alignment of the analyzed sequences is available as Alignment S1.

The state of knowledge of sequences filtered from BOLD library cover 6% of the valid species within Pentatomomorpha. Sequence coverage by species range from 1 to 1116, with 82.30% of sampled species being represented by less than 10 sequences. The available sequences for families and the number of species sampled increase with their respective species richness (Table 1). The families with higher coverage of species were Stenocephalidae (20.00%), Rhopalidae (14.00%), Cymidae (13.00%), and Berytidae (11.00%). On the other hand, 10 of 40 families lack any sampled species. Regarding dataset composition,

the recovered sequences display a bias toward Pentatomoidea, representing almost half of analyzed sequences (47.05%), followed by Lygaeoidea (34.19%), Coreoidea (14.55%), Pyrrhocoroidea (2.55%), and Aradoidea (1.67%). The only sequence retrieved for Idiostoloidea (*Triseucus* sp.; GBA29199-16) was filtered out due to the lack of specific identification. The families with the greatest contribution of sequences for our final dataset were Pentatomidae ($n = 4160$), Lygaeidae ($n = 2450$), Rhyparochromidae ($n = 1192$), Scutelleridae ($n = 892$), and Coreidae ($n = 866$); for the total richness of 1070 species were Pentatomidae ($S = 361$), Rhyparochromidae ($S = 106$), Coreidae ($S = 95$), Lygaeidae ($S = 77$), and Aradidae ($S = 77$). See Table S1 for a list of all sampled species and Table S2 for a list of all sampled sequences.

For most of the evaluated genera, *cox1* efficiency was considered *good* (75.14%), although in some instances performance was *intermediate* (12.15%) or *poor* (12.71%). Figure 3 shows examples of such performances; boxplots for each genus are available as Data S1. Intra- and interspecific distances varied considerably among taxa, which directly affected the threshold values detected with “localMinima.” Threshold values ranged from 0.08% to 13.24%, with an average of 5.49%. For a detailed classification of each genus efficiency and the determined threshold values, see Appendix 1.

The PCI was estimated using a trimmed dataset, consisting of 11,853 sequences belonging to 339 genera and 737 species. The number of sequences decreased since we only considered species represented by more than one sequence, to guarantee intraspecific comparisons. The global PCI was 74.33%, but this value varied greatly among superfamilies and families (Figure 4, Appendix 1). Overall, genera that *cox1* efficiency was *good* disclosed higher PCI, whereas genera with *poor* performance showed lower values (Figure 5).

4 | DISCUSSION

In this paper, we scrutinized the BOLD library to detach the whole data related to the *cox1* barcode region of Pentatomomorpha. This broad approach detected barcode gaps for most of the analyzed genera, reaffirming the transversal efficiency of this marker as DNA barcode for Pentatomomorpha. These findings also allowed us to infer individual empirical thresholds values for 129 genera (Appendix 1), with a broad range of minimum interspecific distance and local minimum. For most genera, this is the first time that a threshold value is evaluated based on empirical data. The PCI analyses exposed a high number of sequences associated with potentially misidentified taxa—one out of four species on BOLD presents specimens wrongly labeled.

4.1 | How is the Pentatomomorpha tome sorted in the BOLD library?

The available data on BOLD are biased to families with economic importance. Economic losses leveraged molecular studies through *cox1* to refine the knowledge on crop pest species (e.g., Barman et al.,

TABLE 1 Number of sequences of *cytochrome c oxidase subunit I* retrieved from BOLD per family before (raw data) and after filtering steps (final data), valid and sampled genera and species per family. Numbers for Thaumastellinae and Thyreocorinae showed apart from Cydnidae

Taxon	Data		Genera		Species	
	Raw	Final	Valid	Sampled	Valid	Sampled
Aradoidea						
Aradidae	544	203	233	29	1931	77
Termitaphididae	1	0	2	0	9	0
Coreoidea						
Alydidae	675	480	45	11	359	21
Coreidae	1150	866	267	52	3178	95
Hyocephalidae	0	0	2	0	3	0
Rhopalidae	556	422	21	12	295	36
Stenocephalidae	53	5	1	1	16	4
Idiostoloidea						
Henicocoridae	0	0	1	0	1	0
Idiostolidae	1	0	3	0	4	0
Lygaeoidea						
Artheneidae	33	33	8	1	20	1
Berytidae	197	140	36	10	174	21
Blissidae	78	32	51	3	436	8
Colobathristidae	12	2	23	1	84	1
Cryptorhamphidae	0	0	2	0	4	0
Cymidae	113	79	9	2	54	9
Geocoridae	164	78	25	2	280	15
Heterogastridae	10	8	24	2	100	2
Lygaeidae	3194	2450	102	28	968	77
Malcidae	25	4	3	2	29	2
Meschiidae	0	0	2	0	5	0
Ninidae	1	1	5	1	13	1
Oxycarenidae	79	26	26	4	150	11
Pachygronthidae	107	105	13	2	78	4
Piesmatidae	24	17	6	2	44	3
Rhyparochromidae	3170	1192	372	59	1850	106
Pentatomoidea						
Acanthosomatidae	407	363	57	13	317	52
Canopidae	1	0	1	0	16	0
Cydnidae	196	82	120	13	765	18
Thaumastellinae	2	2	1	1	3	2
Thyreocorinae	68	39	30	2	224	8
Dinidoridae	28	17	20	3	110	7
Lestoniidae	1	1	1	1	2	1
Megarididae	1	0	2	0	18	0
Pentatomidae	5368	4158	945	140	4855	361
Phloeidae	1	0	3	0	4	0
Plataspidae	475	135	66	4	604	7
Saileriolidae	0	0	3	0	4	0

(Continues)

TABLE 1 (Continued)

Taxon	Data		Genera		Species	
	Raw	Final	Valid	Sampled	Valid	Sampled
Scutelleridae	961	892	102	34	531	54
Tessaratomidae	43	20	62	5	250	6
Urostylididae	39	26	9	4	177	8
Pyrrhocoroidea						
Largidae	100	41	13	6	106	13
Pyrrhocoridae	372	270	33	10	340	37
Total	18,250	12,189	2750	460	18,411	1068

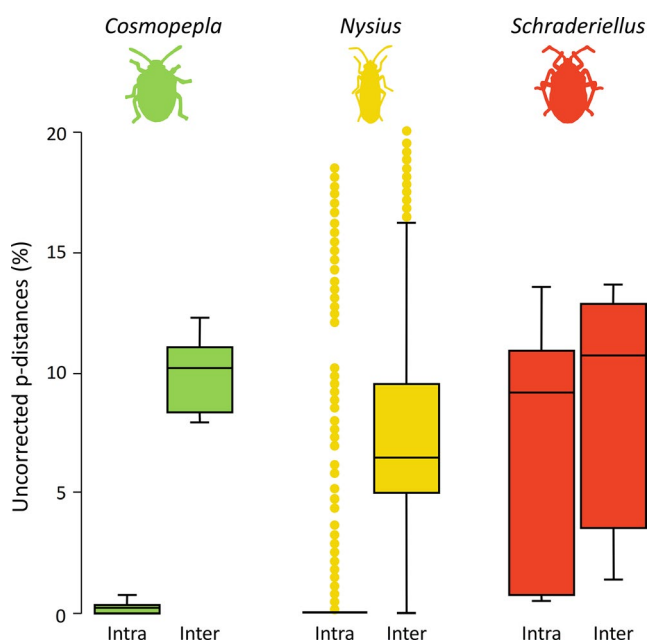


FIGURE 3 Examples of the barcode gap performance classifications used in this study. Performance was classified as good for *Cosmopepla* Stål, which boxplots disclosed a clear barcode gap. The whiskers of the intra- and interspecific distances of *Nysius* Dallas overlapped, implying an intermediate performance. *Schraderiellus* Rider displayed a poor performance, since the boxes of the intra- and interspecific distances overlapped

2017; Garipey et al., 2019; Soares et al., 2018). For instance, established biosafety and quarantine protocols use DNA barcoding to assess pest diversity on a global scale (Boykin et al., 2012; Lee et al., 2019). For pentatomomorphans, of which egg and immature identification is a challenging task (Lis et al., 2013; Matesco et al., 2014), the diagnosis using this molecular tool may promote a fast and accurate species identification, playing a pivotal role in crop pest management.

Besides being the most speciose family within Pentatomomorpha (Henry, 2017), Pentatomidae encompasses numerous crop pests (Schaefer & Panizzi, 2000). The pentatomids brown marmorated stink bug [*Halyomorpha halys* (Stål)], green stink bug [*Chinavia hilaris* (Say)], dusky stink bug [*Euschistus tristigmus* (Say)], and Neotropical brown stink bug [*Euschistus heros* (Fabricius)] sum alone almost

1400 sequences in our final datasets. They are aggressive crops in different regions. *Chinavia hilaris* and *E. tristigmus* occur in North America, while *E. heros* occurs mostly in South America (Schaefer & Panizzi, 2000; Smaniotto & Panizzi, 2015). The *H. halys* is native from Asia and has invaded other continents [i.e., America (Hoebeke & Carter, 2003) and Europe (Wermelinger et al., 2008)] The brown marmorated stink bug shows rapid spreading throughout invaded territories (Zhu et al., 2012) and seems to be the most prominent pentatomid pest in the past two decades.

The second most representative family in our sample was Lygaeidae, with *Nysius* Dallas comprising more than half of the sequences for the family, and in fact, being the genus with most analyzed sequences (1695 sequences) among our datasets. Although *Nysius* groups around a hundred valid species (Malipatil, 2010), our dataset presents only 17 species labels. *Nysius binotatus* (Germar) (as commonly found on BOLD) is currently a subspecies of *Nysius senecionis* (Schilling) (treated here only by the binomen), and this is the species with the greatest contribution of sequences for our dataset (1116 sequences). These data are from exploratory efforts on Saharo-Arabian region with year-long deployment of Malaise traps and show the high abundance of *N. senecionis* in the region. *Nysius* species occur worldwide and feed on the seeds of a broad range of economically important host plants (Ge & Li, 2019; Schaefer & Panizzi, 2000). The birch catkin bug [*Kleidocerys resedae* (Panzer)] is another lygaeid species prevalent in our datasets (396 sequences). This species has a Holarctic distribution and is a pest of birch and alder trees, feeding and breeding on the seed catkin (Dioli et al., 2019; Schaefer & Panizzi, 2000).

Only some small families lack *cox1* data in BOLD (Table 1). Most of them are endemic to one zoogeographical region, are rare, and have little knowledge of their natural histories (e.g., Canopidae, Cryptorhamphidae, Hyocephalidae). Although Termitaphididae is widespread, occurring in Australian, Afrotropical, Neotropical, and Oriental regions (Schuh & Weirauch, 2020), it is also a small family, with nine species without barcoding data. The lack of barcoding data for Idiostoloidea mirrors the scientific knowledge for its six species: Little is known concerning their natural history, systematics, and evolution (Schuh & Weirauch, 2020). Within the Pentatomomorpha tome, many taxa are blank pages waiting to be filled with barcode sequences.

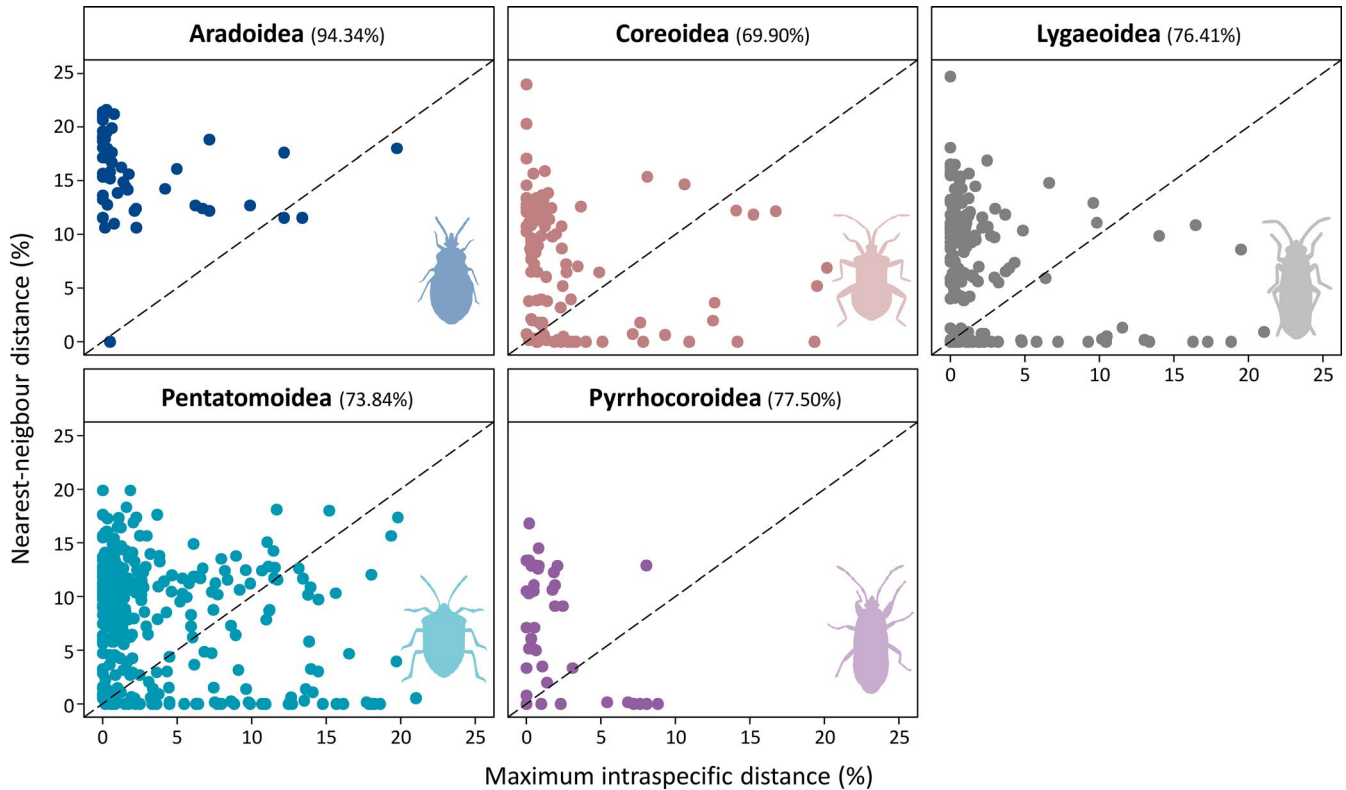
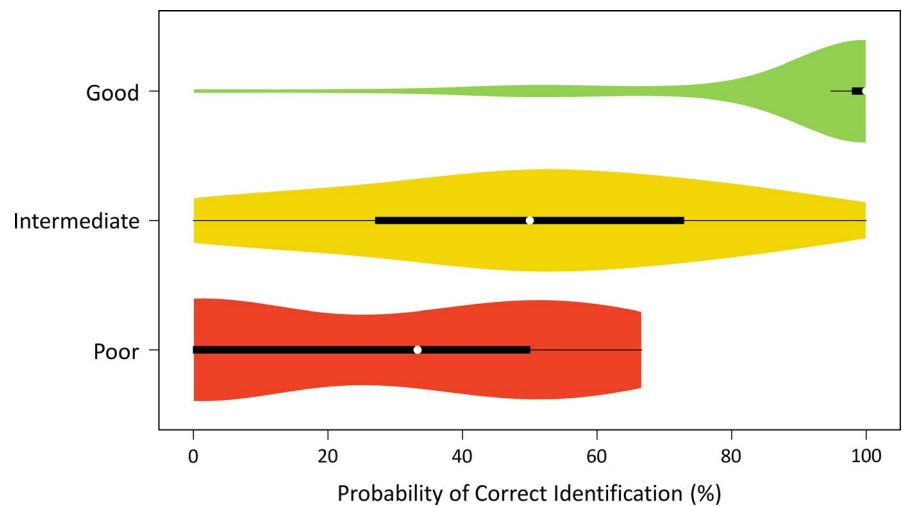


FIGURE 4 Graphic summary of the Probability of Correct Identification (PCI) analyses for Pentatomomorpha superfamilies. For each sampled species, we compared the maximum intraspecific distance with the nearest-neighbor distance. A 1:1 slope (dashed line) represents the point at which the difference between the two variables is zero. Above the slope, species are considered successfully identified; species below the slope represent identification failures. Numbers next to superfamilies represent PCI values

FIGURE 5 Violin plot showing density estimates of Probability of Correct Identification values according to barcoding gap categories. White circles show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles



4.2 | Using thresholds to discover species by the book

Some DNA barcoding analyses (e.g., gap discovery, threshold values) perform better if based on curated libraries (Meyer & Paulay, 2005) due to their sample dependency. The singular coalescent depths of each lineage (e.g., genus, species, population) produce a practical challenge to cluster/split sequences in operational taxonomic units (Fujita et al., 2012; Zinger & Philippe, 2016). Hence,

assuming a fixed threshold value for higher taxonomic levels is unrealistic and broadly accepted that a comprehensive haplotype sampling and species coverage leads to more reliable values (e.g., DeSalle & Goldstein, 2019; Meyer & Paulay, 2005). The use of empirical data is the best first step to look for barcoding gaps and threshold values (e.g., Gonçalves et al., 2021). The amassing of data for a lineage will progressively reduce inaccurate identification, false positives, false negatives (Meyer & Paulay, 2005) and shed light on cryptic species and specimen identification accuracy.

It is convenient to review these values whenever new samples are available in databases (see Qing et al., 2020).

Arbitrary fixed threshold values have been used and tested in Pentatomomorpha (e.g., Park et al., 2011; Zhao et al., 2018). Previous research with *Cletus* Stål species showed that the use of popular fixed threshold values [e.g., 1% BOLD (Ratnasingham & Hebert, 2007) and 2% (Hebert et al., 2003)] would overestimate the genus diversity (Zhang et al., 2017). The authors suggest a threshold value above 7% based on the minimum interspecific distance, which would help to solve the overestimation of diversity. Our result for the minimum interspecific distance was similar to that (6.86%), and the local minimum was 3.30%. Here we provide both minimum interspecific distance and local minima for three out of four tested genera (Appendix 1). Future studies may benefit from these empirical values, since they are more reliable than an arbitrary fixed threshold. As the taxonomic knowledge for the taxa under study is refined and new sequences are generated, re-assessing these values will increase their reliability and improve their practical applications.

The local minimum represents the barcoding gap, where a consistent dip in the density of genetic distances split intra- and interspecific distances is observed (Brown et al., 2012). Although the use of minimum interspecific distance is widespread in the literature as a threshold value (e.g., Meier et al., 2008; Song, Lin, Wang, & Wang, 2018), this may result in misidentifications when sequences from different species have distance values falling below this threshold, increasing the chance of false negatives. Since we understand DNA barcoding as a source of evidence and not the only way to make taxonomic decisions, the local minimum provides more rigor to identify molecular entities (see Casiraghi et al., 2010). It would flag more samples to be explored with other evidence, allowing more robust taxonomic decisions (e.g., identification, species description, synonyms).

The PCI and barcoding gap analysis are methods based on genetic distances that work in distinct scopes. While the PCI identifies discrepancies on the cohesion of the data and points out putative misidentification, the barcoding gap analysis shows the distribution of intra- and interspecific distances disclosing the existence (or not) of a gap. Higher PCI values tend to lead to good barcoding performances (Figure 5), but the relationship between both analyses is not obvious. For instance, *Leptocoris* Latreille (Alydidae), *Catorhintha* Stål (Coreidae), *Neortholomus* Hamilton (Lygaeidae), and *Padaeus* Stål (Pentatomidae) presented 0% of PCI although a good barcoding performance. It means the pool of samples for each taxa presented a clear gap, although these specimens seem to be wrongly labeled. The opposite scenario is also possible, where DNA barcoding can be used for specimen identification without a consistent barcoding gap (see Collins & Cruickshank, 2012). Thus, these methods are complementary to taxonomic approaches using barcoding.

Most superfamilies analyzed here presented PCI values around 70%, but Aradoidea showed 94.34%. Grebennikov and Heiss (2014) and Grebennikov and Heiss (2018) are responsible for most of

Aradidae sequences deposited in BOLD. The authors built a reference barcoding library for flat bugs (Aradidae) species and identified many cryptic new species exploring the Tanzanian fauna. High PCI values reflect rigorous taxonomic treatment. Although DNA barcoding may speed up the taxonomy, this is only possible (or at least should be) when molecular data meet taxonomic experts—but the training of such taxonomists is a long-term commitment that usually takes many years (Coleman & Radulovici, 2020). The slow increase in taxonomists is incompatible with the increase in articles dependent on taxonomic services (Wang et al., 2020). Specimen identification by non-specialists can be risky, and it is a source for many misidentification in databases.

Misidentification is an issue inherent to any public databases, where there is a claim for accurate identifications and rigorous curation (van den Burg et al., 2020). Ideally, all sequences in BOLD should be derived from vouchered specimens identified by taxonomic experts (Lis et al., 2016; Meiklejohn et al., 2019) and obtained through rigorous laboratory protocols. DNA barcoding success may also be affected by non-biological processes, referred by Mutanen et al., (2016) as sources of operational bias. These include inaccurate reference taxonomy, misidentifications, amplification of non-target sequences, alignment errors, erroneous sequencing due to contamination, and other methodological issues. Operational biases impact the reliability of the libraries and complicate the taxonomy and systematics of species (Mutanen et al., 2016).

We emphasize that the threshold values suggested here are not a “magic bullet” for species discrimination, but a source of evidence to this. Species delimitation based on DNA barcodes is hampered, for instance, in scenarios of hybridization or introgression, incomplete lineage sorting, and bacterial endosymbionts changing pathways of mtDNA inheritance (Magoga et al., 2018; Mutanen et al., 2016). Thus, this kind of approach as the single evidence for taxonomic decisions should be avoided. We advocate for the use of integrative approaches, which produce stronger species hypotheses (Padial et al., 2010).

4.3 | Empirical examples: How may these data clue future studies?

4.3.1 | Green flags: Fixing synonyms improves performance

We tracked the improvement of the results during filtering steps, and we found that misspellings and synonyms impacted the finding of barcoding gaps for some genera. On BOLD, *Cosmoepepla* Stål is represented by four species labels. Before our last filter (checking synonyms and looking for misspellings), the barcoding gap analysis for this genus disclosed a poor efficiency (data not shown). However, the two-spotted stink bug [*Cosmoepepla bimaculata* (Thomas)] labels 38 sequences. This name is a junior synonym of *Cosmoepepla lintneriana* (Kirkaldy), which in our final dataset presented 44 sequences due to the lumping of these species labels.

Another taxonomic issue in *Cosmopepla* was the misspelling of *Cosmopepla intergressus*, which probably refers to *Cosmopepla intergressa* (Uhler), a valid name. After solving these taxonomic issues, the final analysis for *Cosmopepla* resulted in *good* efficiency (Figure 3).

Misspellings and synonyms will inevitably worsen the performance of PCI and barcoding gap analyses. In both cases, disparate names for the same taxonomic unit will often lead to interspecific distance being masked as intraspecific distances. Thus, PCI would fail to distinguish the two putative names (i.e., species), suggesting a problematic identification or delimitation in the pool of sampled specimens; at the same time, it would distort the boxplots, decreasing the efficiency to detect a true barcoding gap.

4.3.2 | Yellow flags: Are speciose genera a bad omen?

We found that the efficiency of the barcoding gap was *intermediate* for some genera (e.g., *Euschistus* Dallas, *Geocoris* Fallen, *Leptoglossus* Leach). Some Pentatomomorpha need extensive efforts and expertise to accurate identification (e.g., Cioato et al., 2015; Packauskas & Schaefer, 2001). For instance, *Nysius* (Figure 3) was described to contain 11 species (Dallas, 1852). Since then, some species were transferred to other genera, and many others have been added to, resulting in *Nysius* as a speciose taxon (Barber, 1947; Malipatil, 2010). The interspecific morphological similarities and intraspecific coloration mislead taxonomists and have given taxa of this genus the reputation of being “hard to identify” (Nakatani, 2015). Reports of taxonomic misidentification are spread in the literature for decades (e.g., Barber, 1947; Beardsley, 1977; Nakatani, 2015; Schaefer & Panizzi, 2000; Swezey, 1942). The PCI results corroborate this chaotic scenario, suggesting that only three out of the 17 analyzed species have a pool of correctly identified specimens. The *intermediate* efficiency of barcoding may be related to a relatively high number of misidentification, noising a putative barcoding gap among species.

4.3.3 | Red flags: Poor performance—Don't panic!

While the *intermediate* efficiency of barcoding gap mirrors relatively scarce misidentification, the *poor* efficiency suggests gross taxonomic inconsistencies in the dataset (Figure 5). Most of the genera included in this rank of efficiency were, again, those with historically problematic taxonomy, such as *Blissus* Burmeister (Blissidae), *Mozena* Amyot & Serville (Coreidae), and *Physopelta* Amyot & Serville (Largidae) (Brailovsky & Barrera, 2001; Henry et al., 2015; Stehlik, 2013). Besides these emblematic scenarios, we also found the case of *Schraderiellus* Rider.

Schraderiellus is represented in our sample by specimens from Costa Rica of *Schraderiellus cinctus* (Ruckes) and *Schraderiellus*

hughesae (Ruckes). These sequences were deposited in BOLD in 2012. At the time, only these two species were valid in the genus. The barcoding gap analysis resulted in *poor* efficiency (Figure 3), and the PCI also failed to distinguish species (Appendix 1). These analyses raise a taxonomic red flag to this genus, indicating, for example, cryptic species or operational biases. Roell and Campos (2018) reviewed *Schraderiellus* and described five new species. Currently, this genus occurs from Nicaragua to Ecuador, being most species recorded to Costa Rica (Roell & Campos, 2018).

Checking the voucher pictures available in BOLD, we found misidentifications. The records ASIHE211-12, ASIHE214-12, ASIHE215-12, and ASIHE260-12 are identified as *S. hughesae*; however, only ASIHE211-12 seems to be correctly identified; the remain specimens indeed belong to other species (*S. cinctus* or *Schraderiellus rufilineatus* Roell & Campos). The records ASIHE196-12, ASIHE200-12, and ASIHE203-12 are identified as *S. cinctus*. Again, only the first seems to be correctly identified, while the last two likely belong to other species (i.e., *S. rufilineatus*) (T. Roell pers. comm.). This case exemplifies how barcoding analyses may clue taxonomic problems (e.g., misidentification, cryptic species) for in-depth studies, as already pointed out by many authors (e.g., Lis et al., 2016; Zhang et al., 2017).

4.4 | Closing the book, for now

Molecular data changed drastically the way science produces and evaluates biological knowledge. Since the description and understanding of biodiversity are remarkably incomplete (Costello et al., 2013; Sheth & Thaker, 2017), the approaches to optimize robust species delimitation and specimen identification are desired. Here, we provide an overview of the available *cox1* information for Pentatomomorpha deposited in BOLD. After inspecting more than 12,000 sequences, we verified for the first time the existence of barcoding gaps and provided individual threshold values for 132 genera, also casting doubt on the label on many specimens. Our results may be used to mitigate the Linnean shortfall—the knowledge gap between formally described species and the number of extant species. The Linnean shortfall underpins the biological knowledge, since the underlying recognition of the biological units precedes any study (see Hortal et al., 2015).

The use of *cox1* as a taxonomic tool is copious, and our empirical examples reaffirm the efficiency of DNA barcoding to identify taxonomic inconsistencies. A careful handle of the data (i.e., samplings, specimens identification, laboratory protocols, or deposit of sequences in libraries) and interpretation of the results are mandatory. The simple act to double-check for invalid names and misspellings may denoise analyses and improve results. Here, we flag many genera for further investigations based on the state of knowledge of this infraorder. Future studies, including new sequences and species, may refine values and sharpen the accuracy of barcode tools.

ACKNOWLEDGEMENTS

We are thankful to Dr Paul Masonick and Dr Emília Wendt for the critical reading of this manuscript; and Dr Pablo Dellapé and MSc Wanessa Costa for helping us to compose Figure 1. We also would like to thank Dr Talita Roell for the comments on *Schraderiellus* and the two anonymous reviewers for their valuable comments. FMB is supported by a postdoctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (PNPD/CAPES)—Finance code 001; LTG is supported by a doctoral fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

ORCID

Filipe Michels Bianchi  <https://orcid.org/0000-0002-5842-8822>

Leonardo Tresoldi Gonçalves  <https://orcid.org/0000-0002-3593-8424>

REFERENCES

- Badotti, F., de Oliveira, F. S., Garcia, C. F., Vaz, A. B. M., Fonseca, P. L. C., Nahum, L. A., Oliveira, G., & Góes-Neto, A. (2017). Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiology*, 17(1), 42. <https://doi.org/10.1186/s12866-017-0958-x>
- Barber, H. G. (1947). Revision of the genus *Nysius* in the United States and Canada (Hemiptera Heteroptera: Lygaeidae). *Journal of the Washington Academy of Sciences*, 37(10), 354–366.
- Barman, A. K., Joyce, A. L., Torres, R., & Higbee, B. S. (2017). Assessing genetic diversity in four stink bug species, *Chinavia hilaris*, *Chlorochroa uhleri*, *Chlorochroa sayi*, and *Thyanta pallidovirens* (Hemiptera: Pentatomidae), using DNA Barcodes. *Journal of Economic Entomology*, 110(6), 2590–2598. <https://doi.org/10.1093/jee/tox227>
- Beardsley, J. W. J. (1977). The *Nysius* seed bugs of Haleakala National Park, Maui (Hemiptera: Lygaeidae: Orsillinae). *Proceedings of the Hawaiian Entomological Society*, 22(3), 443–450.
- Boykin, L. M., Armstrong, K. F., Kubatko, L., & De Barro, P. (2012). Species delimitation and global biosecurity. *Evolutionary Bioinformatics Online*, 8, 1–37. <https://doi.org/10.4137/EBO.S8532>
- Brailovsky, H., & Barrera, E. (2001). Six new species of *Mozena* from Mexico (Hemiptera: Coreidae: Coreinae: Nematopodini). *The Florida Entomologist*, 84(1), 99–111.
- Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M. C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12(3), 562–565. <https://doi.org/10.1111/j.1755-0998.2011.03108.x>
- Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., & De Mattia, F. (2010). DNA barcoding: A six-question tour to improve users' awareness about the method. *Briefings in Bioinformatics*, 11(4), 440–453. <https://doi.org/10.1093/bib/bbq003>
- Cioato, A., Bianchi, F. M., Eger, J., & Grazia, J. (2015). New species of *Euschistus* (*Euschistus*) from Jamaica, *Euschistus* (*Mitripus*) and *Ladeaschistus* from southern South America (Hemiptera: Heteroptera: Pentatomidae: Pentatominae). *Zootaxa*, 4048(4), 565–574. <https://doi.org/10.11646/zootaxa.4048.4.7>
- Coleman, C. O., & Radulovici, A. E. (2020). Challenges for the future of taxonomy: Talents, databases and knowledge growth. *Megataxa*, 1(1), 28–34. <https://doi.org/10.11646/megataxa.1.1.5>
- Collins, R. A., & Cruickshank, R. H. (2012). The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, 13(6), 969–975. <https://doi.org/10.1111/1755-0998.12046>
- CoreoideaSF Team (2020, October 26) Coreoidea species file online. Version 5.0/5.0. <http://Coreoidea.SpeciesFile.org>
- Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct? *Science*, 339(6118), 413–416. <https://doi.org/10.1126/science.1230318>
- Dallas, W. S. (1852). *List of the specimens of Hemipterous insects in the collection of the British Museum. Part II*. Taylor & Francis.
- Dellapé, P. M., & Henry, T. J. (2020). Lygaeoidea Species File. Version 5.0/5.0. <http://Lygaeoidea.SpeciesFile.org>
- DeSalle, R., Egan, M. G., & Siddall, M. (2005). The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1905–1916. <https://doi.org/10.1098/rstb.2005.1722>
- DeSalle, R., & Goldstein, P. (2019). Review and interpretation of trends in DNA barcoding. *Frontiers in Ecology and Evolution*, 7, 302. <https://doi.org/10.3389/fevo.2019.00302>
- Dioli, P., Boggio, C. M. T., & Limonta, L. (2019). Heteroptera collected in Valsesia, Northern Italy. *Journal of Entomological and Acarological Research*, 51(3), 82–86. <https://doi.org/10.4081/jear.2019.8752>
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27(9), 480–488. <https://doi.org/10.1016/j.tree.2012.04.012>
- Garipey, T. D., Bruin, A., Konopka, J., Scott-Dupree, C., Fraser, H., Bon, M.-C., & Talamas, E. (2019). A modified DNA barcode approach to define trophic interactions between native and exotic pentatomids and their parasitoids. *Molecular Ecology*, 28(2), 456–470. <https://doi.org/10.1111/mec.14868>
- Ge, X., & Li, J. (2019). Review of the genus *Nysius* Dallas from Mongolian Plateau (Hemiptera: Heteroptera: Orsillidae). *Zootaxa*, 4560(1), 171–183.
- Gonçalves, L. T., Bianchi, F. M., Deprá, M., & Calegario-Marques, C. (2021). Barcoding a can of worms: Testing *cox1* performance as a DNA barcode of Nematoda. *Genome* (in press). <https://doi.org/10.1139/gen-2020-0140>
- Goulding, T. C., & Dayrat, B. (2016). Integrative taxonomy: Ten years of practice and looking into the future. *Archives of Zoological Museum of Lomonosov Moscow State University*, 54, 116–133.
- Gowande, G. G., Tembe, S., & Ghate, H. V. (2018). Revisiting DNA barcoding of true bugs of the infraorder Pentatomomorpha (Hemiptera: Heteroptera) from India. *Mitochondrial DNA Part A*, 29(8), 1215–1223. <https://doi.org/10.1080/24701394.2018.1431229>
- Grazia, J., Schuh, R. T., & Wheeler, W. C. (2008). Phylogenetic relationships of family groups in Pentatomoidea based on morphology and DNA sequences (Insecta: Heteroptera). *Cladistics*, 24(6), 932–976.
- Grebennikov, V. V., & Heiss, E. (2014). DNA barcoding of flat bugs (Hemiptera: Aradidae) with phylogenetic implications. *Arthropod Systematics & Phylogeny*, 72(2), 213–219. [10.1111/1365-3113.12198](https://doi.org/10.1111/1365-3113.12198)
- Grebennikov, V. V., & Heiss, E. (2018). Survey and DNA barcoding of flat bugs (Hemiptera: Aradidae) in the Tanzanian Forest Archipelago reveal a phylogeographically structured fauna largely unknown at the species level. *European Journal of Entomology*, 115, 512–514. <http://doi.org/10.14411/eje.2018.051>
- Gwiazdowski, R. A., Footitt, R. G., Maw, H. E. L., & Hebert, P. D. N. (2015). The Hemiptera (Insecta) of Canada: Constructing a Reference Library of DNA Barcodes. *PLoS One*, 10(4), e0125635. <https://doi.org/10.1371/journal.pone.0125635>
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>
- Henry, T. J. (2017). Biodiversity of Heteroptera. In R. G. Footitt, & P. H. Adler (Eds.), *Insect biodiversity* (pp. 279–335). John Wiley & Sons.
- Henry, T. J., Dellapé, P. M., & de Paula, A. S. (2015). The big-eyed bugs, chinch bugs, and seed bugs (Lygaeoidea). In A. R. Panizzi, & J.

- Grazia (Eds.), *True bugs (Heteroptera) of the neotropics* (pp. 459–514). Springer.
- Hickmann, F., Moraes, T., Bianchi, F. M., Corrêa, A. S., & Schwertner, C. F. (2019). Integrating data to redescribe *Euschistus taurulus* Berg (Hemiptera: Pentatomidae). *Zootaxa*, 4688(1), 119–134. <https://doi.org/10.11646/zootaxa.4688.1.7>
- Hoebeke, E. R., & Carter, M. E. (2003). *Halyomorpha halys* (Stål) (Heteroptera: Pentatomidae): A polyphagous plant pest from Asia newly detected in North America. *Proceedings of the Entomological Society of Washington*, 105(1), 225–237.
- Hollingsworth, M. L., Andra clark, A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., Cowan, R., Chase, M. W., Gaudeul, M., & Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, 9(2), 439–457. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>
- Lee, T. R. C., Anderson, S. J., Tran-Nguyen, L. T. T., Sallam, N., Le Ru, B. P., Conlong, D., Powell, K., Ward, A., & Mitchell, A. (2019). Towards a global DNA barcode reference library for quarantine identifications of lepidopteran stemborers, with an emphasis on sugarcane pests. *Scientific Reports*, 9(1), 7039. <https://doi.org/10.1038/s41598-019-42995-0>
- Lis, B., Lis, J. A., & Ziaja, D. J. (2013). Identification of the nymphal stages of two European seed bugs, *L. equestris* and *L. simulans* (Hemiptera: Heteroptera: Lygaeidae), using DNA barcodes. *Zootaxa*, 3608, 147–150. <https://doi.org/10.11646/zootaxa.3608.2.5>
- Lis, J. A., Lis, B., & Ziaja, D. J. (2016). In BOLD we trust? A commentary on the reliability of specimen identification for DNA barcoding: A case study on burrower bugs (Hemiptera: Heteroptera: Cydnidae). *Zootaxa*, 4114(1), 83. <https://doi.org/10.11646/zootaxa.4114.1.6>
- Lis, J. A., Ziaja, D. J., Lis, B., & Gradowska, P. (2017). Non-monophyly of the "cydnoid" complex within Pentatomoidea (Hemiptera: Heteroptera) revealed by Bayesian phylogenetic analysis of nuclear rDNA sequences. *Arthropod Systematics & Phylogeny*, 75(3), 481–496.
- Magoga, G., Sahin, D. C., Fontaneto, D., & Montagna, M. (2018). Barcoding of Chrysomelidae of Euro-Mediterranean area: Efficiency and problematic species. *Scientific Reports*, 8(1), 13398. <https://doi.org/10.1038/s41598-018-31545-9>
- Malipatil, M. B. (2010). Review and revision of *Nysius* Dallas of Australia and South West Pacific (Hemiptera: Heteroptera: Orsillidae). *Zootaxa*, 2410(1), 29–44.
- Matesco, V. C., Bianchi, F. M., Fürstenau, B. B. R. J., Da Silva, P. P., Campos, L. A., & Grazia, J. (2014). External egg structure of the Pentatomidae (Hemiptera: Heteroptera) and the search for characters with phylogenetic importance. *Zootaxa*, 3768, 351–385. <https://doi.org/10.11646/zootaxa.3768.3.5>
- McPherson, J. E. (Ed.) (2018). *Invasive stink bugs and related species (Pentatomoidea): Biology, higher systematics, semiochemistry, and management*. CRC Press.
- Meier, R., Zhang, G., & Ali, F. (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the "Barcoding Gap" and leads to misidentification. *Systematic Biology*, 57(5), 809–813. <https://doi.org/10.1080/10635150802406343>
- Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLoS One*, 14(6), e0217084. <https://doi.org/10.1371/journal.pone.0217084>
- Meyer, C. P., & Paulay, G. (2005). DNA Barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, 3(12), e422. <https://doi.org/10.1371/journal.pbio.0030422>
- Mutanen, M., Kivelä, S. M., Vos, R. A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., Huemer, P., Dincă, V., van Nieuwerkerken, E. J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaëns, T., Efetov, K. A., Hebert, P. D. N., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., ... Godfray, H. C. J. (2016). Species-level para- and polyphyly in DNA Barcode gene trees: Strong operational bias in European Lepidoptera. *Systematic Biology*, 65(6), 1024–1040. <https://doi.org/10.1093/sysbio/syw044>
- Nakatani, Y. (2015). Revision of the lygaeid genus *Nysius* (Heteroptera: Lygaeidae: Orsillinae) of Japan, with description of a new species. *Entomological Science*, 18(4), 435–441. <https://doi.org/10.1111/ens.12141>
- Packauskas, R. J., & Schaefer, C. W. (2001). Clarification of some taxonomic problems in Anisoscnelini and Leptoscelini (Hemiptera: Coreidae: Coreinae). *Proceedings of the Entomological Society of Washington*, 103(1), 249–256.
- Padiál, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 7(1), 16. <https://doi.org/10.1186/1742-9994-7-16>
- Panizzi, A. R., & Grazia, J. (2015). Introduction to true bugs (Heteroptera) of the NEOTROPICS. In A. R. Panizzi, & J. Grazia (Eds.), *True bugs (Heteroptera) of the Neotropics* (pp. 3–20). Springer.
- Park, D.-S., Footitt, R., Maw, E., & Hebert, P. D. N. (2011). Barcoding Bugs: DNA-Based Identification of the True Bugs (Insecta: Hemiptera: Heteroptera). *PLoS One*, 6(4), e18749. <https://doi.org/10.1371/journal.pone.0018749>
- Qing, X., Wang, M., Karssen, G., Bucki, P., Bert, W., & Braun-Miyara, S. (2020). PPNID: A reference database and molecular identification pipeline for plant-parasitic nematodes. *Bioinformatics*, 36(4), 1052–1056. <https://doi.org/10.1093/bioinformatics/btz707>
- Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Raupach, M. J., Hendrich, L., Küchler, S. M., Deister, F., Morinière, J., & Gossner, M. M. (2014). Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany Reveals Taxonomic Uncertainties and Surprises. *PLoS One*, 9(9), e106940. <https://doi.org/10.1371/journal.pone.0106940>
- Rider, D. A., Schwertner, C. F., Vilimová, J., Rédei, D., Kment, P., & Thomas, D. B. (2018). Higher systematics of the pentatomoidea. In J. E. McPherson (Ed.), *Invasive stink bugs and related species (Pentatomoidea)* (pp. 25–201). CRC Press.
- Roell, T., & Campos, L. A. (2018). Phylogeny and revision of *Schraderiellus* Rider, 1998 with the description of five new species (Hemiptera, Pentatomidae, Discocephalinae, Ochlerini). *Zootaxa*, 4508(1), 47–67. <https://doi.org/10.11646/zootaxa.4508.1.3>
- Schaefer, C. W., & A. R. Panizzi (Eds.) (2000). *Heteroptera of economic importance*. CRC Press.
- Schuh, R. T., & Weirauch, C. (2020). *True Bugs of the World (Hemiptera: Heteroptera): Classification and Natural History* (2nd. ed.). Siri Scientific Press.
- Sheth, B. P., & Thaker, V. S. (2017). DNA barcoding and traditional taxonomy: An integrated approach for biodiversity conservation. *Genome*, 60(7), 618–628. <https://doi.org/10.1139/gen-2015-0167>
- Smaniotto, L. F., & Panizzi, A. R. (2015). Interactions of selected species of stink bugs (Hemiptera: Heteroptera: Pentatomidae) from leguminous crops with plants in the Neotropics. *The Florida Entomologist*, 98(1), 7–17. <https://doi.org/10.1653/024.098.0103>

- Soares, P. L., Cordeiro, E. M. G., Santos, F. N. S., Omoto, C., & Correa, A. S. (2018). The reunion of two lineages of the Neotropical brown stink bug on soybean lands in the heart of Brazil. *Scientific Reports*, 8(1), 2496. <https://doi.org/10.1038/s41598-018-20187-6>
- Song, C., Lin, X.-L., Wang, Q., & Wang, X.-H. (2018). DNA barcodes successfully delimit morphospecies in a superdiverse insect genus. *Zoologica Scripta*, 47(3), 311–324. <https://doi.org/10.1111/zsc.12284>
- Stehlík, J. L. (2013). Review and reclassification of the Old World genus *Physopelta* (Hemiptera: Heteroptera: Largidae). *Acta Entomologica Musei Natioalis Pragae*, 53(2), 505–584.
- Swezey, O. H. (1942). Some corrections in the use of names for the species of *Nysius* in Hawaii (Hem.). *Proceedings of the Hawaiian Entomological Society*, 11(2), 200–202.
- Tembe, S., Shouche, Y., & Ghate, H. V. (2014). DNA barcoding of Pentatomomorpha bugs (Hemiptera: Heteroptera) from Western Ghats of India. *Meta Gene*, 2, 737–745. <https://doi.org/10.1016/j.mgene.2014.09.006>
- van den Burg, M. P., Herrando-Pérez, S., & Vieites, D. R. (2020). ACDC, a global database of amphibian cytochrome-b sequences using reproducible curation for GenBank records. *Scientific Data*, 7(1), 268. <https://doi.org/10.1038/s41597-020-00598-9>
- Wang, K., Kirk, P. M., & Yao, Y. J. (2020). Development trends in taxonomy, with special reference to fungi. *Journal of Systematics and Evolution*, 58(4), 406–412. <https://doi.org/10.1111/jse.12538>
- Weirauch, C., & Schuh, R. T. (2011). Systematics and evolution of Heteroptera: 25 years of progress. *Annual Review of Entomology*, 56(1), 487–510. <https://doi.org/10.1146/annurev-ento-120709-144833>
- Weirauch, C., Schuh, R. T., Cassis, G., & Wheeler, W. C. (2019). Revisiting habitat and lifestyle transitions in Heteroptera (Insecta: Hemiptera): Insights from a combined morphological and molecular phylogeny. *Cladistics*, 35(1), 67–105. <https://doi.org/10.1111/cla.12233>
- Wermelinger, B., Wyniger, D., & Forster, B. (2008). First records of an invasive bug in Europe: *Halyomorpha halys* Stål (Heteroptera: Pentatomidae), a new pest on woody ornamentals and fruit trees? *Mitteilungen Der Schweizerischen Entomologischen Gesellschaft*, 81(1/2), 1–8.
- Zhang, H.-G., Lv, M.-H., Yi, W.-B., Zhu, W.-B., & Bu, W.-J. (2017). Species diversity can be overestimated by a fixed empirical threshold: Insights from DNA barcoding of the genus *Cletus* (Hemiptera: Coreidae) and the meta-analysis of COI data from previous phylogeographical studies. *Molecular Ecology Resources*, 17(2), 314–323. <https://doi.org/10.1111/1755-0998.12571>
- Zhao, Q., Wang, J., Wang, M.-Q., Cai, B., Zhang, H.-F., & Wei, J.-F. (2018). Complete Mitochondrial Genome of *Dinorhynchus dybowskyi* (Hemiptera: Pentatomidae: Asopinae) and Phylogenetic Analysis of Pentatomomorpha Species. *Journal of Insect Science (Online)*, 18(2), <https://doi.org/10.1093/jisesa/iey031>
- Zhu, G., Bu, W., Gao, Y., & Liu, G. (2012). Potential geographic distribution of brown marmorated stink bug invasion (*Halyomorpha halys*). *PLoS One*, 7(2), e31246. <https://doi.org/10.1371/journal.pone.0031246>
- Zinger, L., & Philippe, H. (2016). Coalescing molecular evolution and DNA barcoding. *Molecular Ecology*, 25(9), 1908–1910. <https://doi.org/10.1111/mec.13639>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Table S1. Species list and sequence number recovered to each genus and their taxonomic affiliations.

Table S2. List of analysed sequences with BOLD accession numbers.

Data S1. Boxplots.pdf. Barcode gap of all the 138 genera studied for *cox1* by plotting intra- and interspecific distances.

Alignment S1. Alignment.fasta. Global alignment of *cox1* sequences, consisting of 658 base pairs and 12,189 sequences that composed the final dataset

How to cite this article: Bianchi FM, Gonçalves LT. Borrowing the Pentatomomorpha tome from the DNA barcode library: Scanning the overall performance of *cox1* as a tool. *J Zool Syst Evol Res*. 2021;00:1–21. <https://doi.org/10.1111/jzs.12476>

APPENDIX 1

PERFORMANCE CATEGORY, MAXIMUM INTRASPECIFIC DISTANCE (LIMIT OF UPPER INTRASPECIFIC WHISKER), MINIMUM INTERSPECIFIC DISTANCE (LIMIT OF LOWER INTERSPECIFIC WHISKER), AND LOCAL MINIMA FOR ALL ANALYZED GENERA. IN PARENTHESES, THE PROBABILITY OF CORRECT IDENTIFICATION (PCI). VALUES ARE IN PERCENT

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
Aradoidea (94.34)				
Aradidae (94.34)				
<i>Afropictinus</i> (50)	Intermediate	13.37	11.55	
<i>Aneurus</i> (88.88)	Intermediate	19.73	18.02	
<i>Aradus</i> (100)	Good	2.24	10.64	4.01
<i>Brachyrhynchus</i> (100)	Good	9.88	12.69	8.57
<i>Ctenoneurus</i> (100)	–	–	–	
<i>Dysodius</i> (100)	Good	6.69	11.40	6.79
<i>Hesus</i> (100)	Good	0.25	17.48	7.69
<i>Isodermus</i> (100)	–	–	–	
<i>Jacobsaptera</i> (100)	–	–	–	
<i>Libiocoris</i> (100)	–	–	–	
<i>Malgasyaptera</i> (100)	–	–	–	
<i>Mauricicoris</i> (100)	–	–	–	
<i>Mezira</i> (100)	Good	0.00	17.17	6.67
<i>Neochelonoderus</i> (100)	Good	0.25	21.62	11.14
<i>Neuroctenus</i> (100)	Good	7.14	12.18	9.59
<i>Overlatiella</i> (100)	–	–	–	
<i>Pericartaptera</i> (100)	–	–	–	
<i>Pseudomezira</i> (100)	–	–	–	
<i>Stysaptera</i> (100)	Good	0.30	12.77	5.61
<i>Usumbaraia</i> (100)	–	–	–	
<i>Wuiessa</i> (100)	–	–	–	
Coreoidea (69.90)				
Alydidae (40)				
<i>Alydus</i> (40)	Intermediate	4.12	0.00	
<i>Heegeria</i> (0)	–	–	–	
<i>Leptocoris</i> (0)	Good	2.58	7.20	4.18
<i>Megalotomus</i> (100)	–	–	–	
<i>Noliphus</i> (100)	–	–	–	
<i>Protenor</i> (100)	–	–	–	
<i>Riptortus</i> (33.33)	Good	0.61	5.21	0.08
<i>Tollius</i> (0)	–	–	–	
Coreidae (74.55)				
<i>Acanthocephala</i> (100)	Good	1.16	12.52	
<i>Amblypelta</i> (100)	–	–	–	
<i>Anasa</i> (50)	Good	0.81	11.84	
<i>Anoplocnemis</i> (100)	Good	0	23.97	8.75
<i>Arenocoris</i> (100)	–	–	–	

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Bathysolen</i> (100)	—	—	—	
<i>Catorhintha</i> (0)	Good	10.92	11.09	5.81
<i>Ceraleptus</i> (100)	Good	0.76	8.97	6.42
<i>Chelinidea</i> (100)	Good	10.60	14.67	4.58
<i>Cletus</i> (44.44)	Good	2.78	6.86	3.10
<i>Coreus</i> (100)	—	—	—	
<i>Coriomeris</i> (66.66)	Poor	7.13	3.19	
<i>Enoplops</i> (100)	—	—	—	
<i>Euthochtha</i> (100)	—	—	—	
<i>Gonocerus</i> (100)	Good	0.30	8.36	4.32
<i>Homoeocerus</i> (100)	Good	1.22	16.37	4.78
<i>Hydaropsis</i> (100)	—	—	—	
<i>Hypselonotus</i> (66.66)	Poor	1.22	10.76	
<i>Leptoglossus</i> (83.33)	Intermediate	2.97	1.96	
<i>Madura</i> (100)	—	—	—	
<i>Mictis</i> (100)	—	—	—	
<i>Mozena</i> (33.33)	Poor	0.46	0.30	
<i>Notopteryx</i> (100)	—	—	—	
<i>Petillia</i> (100)	—	—	—	
<i>Piezogaster</i> (100)	Good	1.37	3.80	13.24
<i>Plinachtus</i> (100)	Good	0.65	13.40	7.48
<i>Prismatocerus</i> (100)	Good	0.00	12.77	6.29
<i>Spathocera</i> (100)	—	—	—	
<i>Syromastus</i> (100)	—	—	—	
<i>Thasus</i> (100)	—	—	—	
<i>Villasitocoris</i> (0)	—	—	—	
Rhopalidae (75)				
<i>Arhyssus</i> (20)	Good	2.23	9.86	5.01
<i>Aufeius</i> (100)	—	—	—	
<i>Boisea</i> (0)	Poor	2.78	0	
<i>Brachycarenum</i> (100)	—	—	—	
<i>Corizus</i> (100)	—	—	—	
<i>Harmostes</i> (100)	Good	2.21	11.25	3.76
<i>Leptocoris</i> (100)	Good	1.99	10.06	4.45
<i>Liorhyssus</i> (100)	Good	2.65	0	4.87
<i>Myrmus</i> (100)	—	—	—	
<i>Rhopalus</i> (100)	Good	0.33	10.97	5.69
<i>Stictopleurus</i> (75)	Intermediate	2.67	0.46	
Stenocephalidae (100)				
<i>Dicranocephalus</i> (100)	Good	0	10.26	2.28
Lygaeoidea (76.41)				
Artheneidae (100)				
<i>Chilacis</i> (100)	—	—	—	
Berytidae (78.95)				

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Berytinus</i> (100)	Good	0.50	9.27	4.98
<i>Gampsocoris</i> (100)	Good	0.46	14.24	6.66
<i>Hoplinus</i> (100)	–	–	–	
<i>Jalysus</i> (60)	Good	0.74	7.37	3.67
<i>Metatropis</i> (100)	Good	0.17	5.88	3.02
<i>Neides</i> (100)	–	–	–	
<i>Neoneides</i> (100)	–	–	–	
<i>Yemma</i> (0)	–	–	–	
<i>Yemmalysus</i> (0)	–	–	–	
Blissidae (50)				
<i>Blissus</i> (0)	Poor	1.38	0.15	
<i>Ischnodemus</i> (100)	Good	2.46	16.89	8.32
<i>Phaenacantha</i> (100)	–	–	–	
Cymidae (87.50)				
<i>Cymus</i> (87.50)	Good	0.37	7.13	3.53
Geocoridae (60)				
<i>Geocoris</i> (60)	Intermediate	1.99	0	
Heterogastridae (100)				
<i>Heterogaster</i> (100)	–	–	–	
<i>Platyplax</i> (100)	–	–	–	
Lygaeidae (53.85)				
<i>Arocatus</i> (25)	Good	1.22	8.55	5.87
<i>Aspilocoryphus</i> (100)	–	–	–	
<i>Belonochilus</i> (100)	–	–	–	
<i>Graptostethus</i> (100)	Good	1.85	5.62	Na
<i>Kleidocerys</i> (50)	Intermediate	6.09	5.10	
<i>Lygaeospilus</i> (100)	Good	1.91	6.98	4.33
<i>Lygaeus</i> (50)	Good	1.22	10.91	8.06
<i>Melacoryphus</i> (75)	Good	2.28	6.99	4.20
<i>Melanocoryphus</i> (100)	–	–	–	
<i>Melanopleurus</i> (100)	Good	3.00	13.32	5.92
<i>Neortholomus</i> (0)	Good	1.28	8.75	2.16
<i>Nithecus</i> (100)	–	–	–	
<i>Nysius</i> (20)	Intermediate	0	0	
<i>Ochrimnus</i> (100)	Good	0.46	9.97	5.47
<i>Oncopeltus</i> (100)	Good	0.30	11.83	6.75
<i>Orsillus</i> (100)	Good	0	5.50	11.00
<i>Ortholomus</i> (50)	Poor	13.02	0.16	
<i>Spilostethus</i> (33.33)	Poor	21.04	7.25	
<i>Tropidothorax</i> (100)	–	–	–	
<i>Xyonysius</i> (100)	–	–	–	
Malcidae (100)				
<i>Chauliops</i> (100)	–	–	–	
<i>Malcus</i> (100)	–	–	–	

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
Oxycarenidae (83.33)				
<i>Oxycarenus</i> (80)	Intermediate	0.77	0	
Pachygronthidae (100)				
<i>Oedanocala</i> (100)	Good	0.32	10.94	5.77
<i>Phlegyas</i> (100)	Good	1.23	10.57	6.27
Piesmatidae (100)				
<i>Parapiesma</i> (100)	Good	0.74	13.59	7.76
<i>Piesma</i> (100)	–	–	–	
Rhyparochromidae (89.16)				
<i>Acompus</i> (100)	–	–	–	
<i>Antillocoris</i> (100)	Good	0.15	8.04	4.45
<i>Atrazonotus</i> (100)	–	–	–	
<i>Beosus</i> (100)	–	–	–	
<i>Cnemodus</i> (100)	–	–	–	
<i>Cryphula</i> (100)	–	–	–	
<i>Drymus</i> (100)	Good	0.40	10.36	5.59
<i>Elasmolomus</i> (100)	–	–	–	
<i>Emblethis</i> (100)	Good	9.83	11.09	11.06
<i>Eremocoris</i> (50)	Intermediate	0	0	
<i>Gastrodes</i> (100)	Good	0.74	10.62	5.85
<i>Graptopeltus</i> (100)	–	–	–	
<i>Heraeus</i> (100)	–	–	–	
<i>Ischnocoris</i> (100)	Good	0.35	4.61	2.36
<i>Kolenetrus</i> (100)	–	–	–	
<i>Lanchnophorus</i> (100)	–	–	–	
<i>Lamprodema</i> (100)	–	–	–	
<i>Laryngodus</i> (100)	–	–	–	
<i>Ligyrocoris</i> (33.33)	Intermediate	0.48	0.17	
<i>Macrodema</i> (100)	–	–	–	
<i>Malezonotus</i> (100)	Good	1.40	9.27	4.72
<i>Megalonotus</i> (75)	Good	0.38	0.75	5.02
<i>Metochus</i> (100)	–	–	–	
<i>Myodocha</i> (100)	–	–	–	
<i>Neolethaeus</i> (100)	–	–	–	
<i>Neopamera</i> (100)	–	–	–	
<i>Neosuris</i> (100)	–	–	–	
<i>Ozophora</i> (100)	–	–	–	
<i>Pachybrachius</i> (100)	–	–	–	
<i>Paromius</i> (100)	Good	0	7.45	3.87
<i>Perigenes</i> (100)	–	–	–	
<i>Peritrechus</i> (100)	Good	3.69	9.27	6.84
<i>Plinthisus</i> (100)	Good	0	9.72	5.16
<i>Pseudocnemodus</i> (0)	–	–	–	
<i>Pseudopachybrachius</i> (100)	Good	0.49	5.52	3.08

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Pterotmetus</i> (100)	–	–	–	
<i>Raglius</i> (100)	Good	0.61	15.42	8.39
<i>Remaudiereana</i> (100)	–	–	–	
<i>Rhyparochromus</i> (100)	Good	0.61	9.22	5.20
<i>Scolopostethus</i> (100)	Good	3.71	4.10	5.61
<i>Sisamnes</i> (100)	–	–	–	
<i>Slaterobius</i> (0)	–	–	–	
<i>Sphragisticus</i> (0)	–	–	–	
<i>Stygnocoris</i> (100)	Good	1.32	12.09	6.66
<i>Trapezonotus</i> (0)	Poor	4.74	0.15	
<i>Uhleriola</i> (100)	–	–	–	
<i>Valtissius</i> (100)	–	–	–	
<i>Xanthochilus</i> (100)	–	–	–	
<i>Zeridoneus</i> (100)	–	–	–	
Pentatomoidea (73.84)				
Acanthosomatidae (85)				
<i>Acanthosoma</i> (100)	Good	1.32	8.25	2.97
<i>Cyphostethus</i> (100)	Good	0.26	9.70	4.34
<i>Elasmotethus</i> (100)	Good	1.33	1.65	1.48
<i>Elasmucha</i> (50)	Good	0.59	3.03	2.69
<i>Lindbergicoris</i> (100)	Good	0.82	4.53	8.57
<i>Sastragala</i> (100)	Good	5.91	7.22	Na
Cydnidae (92.86)				
<i>Adrisa</i> (100)	–	–	–	
<i>Aethus</i> (100)	–	–	–	
<i>Amnestus</i> (100)	Good	0.74	15.43	3.56
<i>Geotomus</i> (100)	–	–	–	
<i>Macroscythus</i> (66.66)	Poor	3.35	1.42	
<i>Melanaethus</i> (100)	–	–	–	
<i>Microporus</i> (100)	Good	3.65	17.63	7.84
<i>Pangaeus</i> (100)	–	–	–	
<i>Sehirus</i> (100)	–	–	–	
<i>Tritomegas</i> (100)	Good	3.19	13.98	7.19
Thyreocorinae (66.67)				
<i>Corimelaena</i> (80)	Poor	1.41	10.65	
<i>Galgupha</i> (0)	Poor	13.51	0.30	
Dinidoridae (60)				
<i>Coridius</i> (50)	Poor	13.95	3.27	
<i>Cyclopelta</i> (66.66)	Poor	3.08	1.57	
Pentatomidae (72.13)				
<i>Ablaptus</i> (100)	–	–	–	
<i>Acesines</i> (100)	–	–	–	
<i>Acrosternum</i> (0)	Poor	13.84	5.80	
<i>Adevoplitus</i> (100)	–	–	–	

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Aelia</i> (100)	Good	1.83	11.21	7.41
<i>Aeliomorpha</i> (100)	–	–	–	–
<i>Agonoscelis</i> (100)	Good	0.71	13.27	7.34
<i>Agroecus</i> (100)	–	–	–	–
<i>Amaurochrous</i> (100)	–	–	–	–
<i>Amyotea</i> (100)	–	–	–	–
<i>Ancyrosoma</i> (100)	–	–	–	–
<i>Andrallus</i> (100)	–	–	–	–
<i>Antestiopsis</i> (100)	–	–	–	–
<i>Anthemina</i> (100)	–	–	–	–
<i>Apateticus</i> (100)	Good	3.71	10.96	6.32
<i>Arma</i> (50)	Poor	3.30	0.00	–
<i>Arocera</i> (100)	Good	2.48	14.14	6.59
<i>Arvelius</i> (100)	Good	9.18	14.14	5.16
<i>Bagrada</i> (100)	Good	1.39	12.72	1.81
<i>Banasa</i> (42.85)	Good	2.23	10.44	3.90
<i>Berecynthus</i> (100)	–	–	–	–
<i>Boea</i> (100)	–	–	–	–
<i>Brachystethus</i> (50)	Good	1.98	14.92	8.57
<i>Braunus</i> (100)	–	–	–	–
<i>Brepholoxa</i> (100)	–	–	–	–
<i>Brochymena</i> (80)	Good	1.24	3.67	2.51
<i>Carbula</i> (50)	Good	1.22	11.70	6.16
<i>Carpocoris</i> (80)	Good	1.53	7.34	2.31
<i>Cermatulus</i> (100)	–	–	–	–
<i>Chinavia</i> (50)	Good	1.06	6.77	1.35
<i>Chlorochroa</i> (28.57)	Good	0.92	3.50	2.78
<i>Chlorocoris</i> (50)	Intermediate	1.70	7.84	–
<i>Codophila</i> (100)	Good	0.33	15.98	8.53
<i>Coenus</i> (100)	–	–	–	–
<i>Cosmopepla</i> (100)	Good	0.75	7.95	1.36
<i>Cuspicona</i> (100)	–	–	–	–
<i>Cyptocephala</i> (100)	–	–	–	–
<i>Dalpada</i> (100)	Good	1.98	12.28	5.41
<i>Degonetus</i> (100)	–	–	–	–
<i>Dendrocoris</i> (100)	Intermediate	11.55	10.00	–
<i>Dictyotus</i> (100)	–	–	–	–
<i>Dinocoris</i> (0)	Intermediate	10.92	0	–
<i>Dinorhynchus</i> (100)	–	–	–	–
<i>Dolycoris</i> (50)	Poor	4.04	0.54	–
<i>Dryptocephala</i> (100)	–	–	–	–
<i>Edessa</i> (91.66)	Good	1.83	11.57	3.32
<i>Eludocoris</i> (100)	–	–	–	–
<i>Eocanthecona</i> (100)	–	–	–	–

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Eritrachys</i> (100)	–	–	–	
<i>Erthesina</i> (100)	Good	0.76	6.39	3.52
<i>Eurydema</i> (16.66)	Good	1.17	2.13	2.97
<i>Euschistus</i> (75)	Intermediate	17.69	4.84	
<i>Eysarcoris</i> (80)	Good	6.03	8.16	6.80
<i>Glaucias</i> (100)	Good	0.18	9.17	4.80
<i>Gonopsis</i> (100)	Good	0.00	7.46	2.72
<i>Graphosoma</i> (50)	Intermediate	3.19	0.00	
<i>Gynenica</i> (100)	–	–	–	
<i>Halyomorpha</i> (50)	Good	1.46	12.32	1.65
<i>Halys</i> (0)	Intermediate	8.81	5.93	
<i>Holcogaster</i> (100)	–	–	–	
<i>Holcostethus</i> (60)	Good	0.85	9.88	6.55
<i>Hymenarcys</i> (100)	Good	2.04	16.92	9.23
<i>Kermana</i> (0)	–	–	–	
<i>Loxa</i> (0)	Poor	23.57	2.98	
<i>Macropygium</i> (100)	–	–	–	
<i>Mecidea</i> (100)	–	–	–	
<i>Menecles</i> (100)	–	–	–	
<i>Menida</i> (100)	Good	0.31	13.39	7.54
<i>Mormidea</i> (60)	Good	2.73	9.44	4.90
<i>Murgantia</i> (50)	Good	3.20	15.67	9.32
<i>Neotibialis</i> (0)	–	–	–	
<i>Neottiglossa</i> (100)	Good	0.76	8.69	3.98
<i>Nezara</i> (0)	Intermediate	7.84	4.69	
<i>Odmalea</i> (100)	–	–	–	
<i>Oebalus</i> (100)	Good	2.58	10.42	6.76
<i>Oechalia</i> (0)	–	–	–	
<i>Padaeus</i> (0)	Good	0.33	4.87	2.89
<i>Pallantia</i> (100)	–	–	–	
<i>Palomena</i> (100)	Good	0.74	6.81	2.33
<i>Pantochlora</i> (100)	–	–	–	
<i>Parabrochymena</i> (0)	Poor	0.46	0.00	
<i>Parastalius</i> (3)	–	–	–	
<i>Pellaea</i> (100)	Good	11.66	18.11	6.33
<i>Pentatoma</i> (100)	Good	0.30	10.06	3.58
<i>Perillus</i> (100)	Good	1.83	2.69	1.84
<i>Peromatus</i> (50)	Poor	14.46	9.74	
<i>Pharypia</i> (100)	–	–	–	
<i>Picromerus</i> (100)	Good	0.00	10.86	4.70
<i>Piezodorus</i> (75)	Good	1.74	4.40	3.27
<i>Placosternum</i> (100)	–	–	–	
<i>Plautia</i> (66.66)	Good	1.03	8.78	3.92
<i>Podisus</i> (25)	Intermediate	2.34	1.80	

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Prionosoma</i> (100)	–	–	–	
<i>Proxys</i> (100)	Good	1.24	9.81	5.34
<i>Ramosiana</i> (100)	–	–	–	
<i>Rhaphigaster</i> (100)	–	–	–	
<i>Rhyncholepta</i> (100)	–	–	–	
<i>Rhysocephala</i> (100)	Good	0.20	12.90	6.14
<i>Rio</i> (100)	Good	0.99	10.42	5.12
<i>Roferta</i> (100)	–	–	–	
<i>Rubiconia</i> (100)	–	–	–	
<i>Sarju</i> (100)	–	–	–	
<i>Schraderiellus</i> (0)	Poor	13.61	1.38	
<i>Sciocoris</i> (100)	Good	1.85	15.26	8.94
<i>Scotinophara</i> (100)	Good	0.22	7.22	3.23
<i>Sibaria</i> (100)	–	–	–	
<i>Spermatodes</i> (100)	–	–	–	
<i>Stagonomus</i> (100)	Good	0	8.30	4.29
<i>Stiretrus</i> (100)	–	–	–	
<i>Taurocerus</i> (100)	Good	2.13	11.11	5.82
<i>Thyanta</i> (16.66)	Intermediate	1.38	0.00	
<i>Tolumnia</i> (100)	Good	0.66	7.91	4.25
<i>Trichocheila</i> (50)	Intermediate	7.44	5.63	
<i>Troilus</i> (100)	–	–	–	
<i>Vulsirea</i> (100)	–	–	–	
<i>Zicrona</i> (100)	–	–	–	
Plataspidae (60)				
<i>Brachyplatys</i> (100)	–	–	–	
<i>Coptosoma</i> (100)	Good	0.75	7.95	9.64
<i>Megacocta</i> (0)	Poor	2.13	0.15	
Scutelleridae (69.57)				
<i>Augocoris</i> (0)	–	–	–	
<i>Choerocoris</i> (100)	–	–	–	
<i>Chrysocoris</i> (0)	Poor	4.48	0.17	
<i>Coleotichus</i> (0)	–	–	–	
<i>Eucorysses</i> (100)	–	–	–	
<i>Eurygaster</i> (50)	Intermediate	0.87	2.11	
<i>Homaemus</i> (75)	Good	3.66	7.89	5.16
<i>Hotea</i> (100)	–	–	–	
<i>Odontotarsus</i> (100)	Good	6.23	10.94	8.28
<i>Pachycoris</i> (100)	Good	2.88	12.78	6.28
<i>Phimodera</i> (100)	–	–	–	
<i>Solenostethium</i> (100)	Good	1.58	18.32	9.68
<i>Stethaulax</i> (100)	–	–	–	
<i>Vanduzeeina</i> (100)	Good	0.46	13.01	6.18
Tessaratomidae (100)				

(Continues)

APPENDIX 1 (Continued)

Taxa (PCI)	Performance	Maximum intraspecific distance	Minimum interspecific distance	Local minima
<i>Eusthenes</i> (100)	—	—	—	
<i>Piezosternum</i> (100)	—	—	—	
<i>Tessarotoma</i> (100)	Good	0.00	11.29	5.91
Urostylididae (83.33)				
<i>Urochela</i> (50)	Poor	19.79	15.66	
<i>Urostylis</i> (100)	Good	1.72	14.29	7.08
Pyrrhocoroidea (77.50)				
Largidae (80)				
<i>Arhapse</i> (100)	Good	2.43	9.12	5.26
<i>Delacampius</i> (100)	—	—	—	
<i>Iphita</i> (100)	Good	0.48	10.53	5.14
<i>Largus</i> (100)	Good	1.37	1.98	4.96
<i>Macrocheraia</i> (100)	—	—	—	
<i>Physopelta</i> (100)	Poor	2.29	0.00	
Pyrrhocoridae (74.19)				
<i>Antilochus</i> (100)	Good	1.90	11.08	6.16
<i>Dindymus</i> (100)	Good	0.16	28.12	13.01
<i>Dysdercus</i> (53.33)	Good	1.86	3.16	2.41
<i>Euscopus</i> (100)	Good	1.82	12.27	6.81
<i>Melamphaus</i> (100)	Good	8.04	12.90	4.88
<i>Probergrothius</i> (100)	Good	0.32	11.92	6.64
<i>Pyrrhocoris</i> (100)	—	—	—	
<i>Pyrrhopleus</i> (100)	Good	3.09	3.34	0.82
<i>Scantius</i> (100)	—	—	—	