Apidologie (2022) 53:55
© INRAE, DIB and Springer-Verlag France SAS, part of Springer Nature, 2022
https://doi.org/10.1007/s13592-022-00958-x

Original article

Shorter, better, faster, stronger? Comparing the identification performance of full-length and mini-DNA barcodes for apid bees (Hymenoptera: Apidae)

Leonardo Tresoldi Gonçalves[1,2], Elaine Françoso[3], Maríndia Deprá[1,2]

[1] Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
[2] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
[3] Department of Biological Sciences, Royal Holloway University of London, Egham TW20 0EX, UK

Received 11 October 2021 – Revised 2 June 2022 – Accepted 27 July 2022

Abstract – Apid bees are essential ecosystem pollinators, occurring worldwide and comprising over 5900 species. Although they are identified mainly using morphology, DNA barcoding has been explored since its proposal as a supplementary tool in bee taxonomy. Smaller regions of barcode markers—mini-barcodes—were also successfully employed in corbiculate bee identification, but the performance of mini-barcodes was only tested in a narrow taxonomic scope. Here, we scrutinized all 18167 apid bee *cox1* sequences from the Barcode of Life Data System to provide an overview of the available data, search for barcoding gaps at genus level, test if full-length and mini-barcode regions perform similarly in specimen identification, and flag bee taxa that may benefit from studies implementing DNA barcodes. Our dataset encompassed five subfamilies, 25 tribes, 71 genera, and 1012 species, although it was biased towards corbiculate tribes. Most of the surveyed genera showed good performance in the barcoding gap analyses. Moreover, full-length and mini-barcodes displayed a similar probability of correct identification, demonstrating that both marker types are equivalent in bee identification. Finally, we discuss some examples to show how full-length and mini-barcodes can help solve taxonomic inconsistencies and foment future studies of apid bees.

Apoidea / Anthophila / Barcode of Life Data System / COI / *cox1* / integrative taxonomy

1. INTRODUCTION

The correct identification of species underpins most biological studies. In the last decades, molecular tools have been applied in organism identification, diversity surveys, and species delimitation (Roe et al. 2017). For animals, the 5′ region of the mitochondrial gene cytochrome *c* oxidase subunit I (*cox1*) was formalized as a DNA barcode, allowing a quick, efficient, and reliable tool for molecular identification (Hebert et al. 2003). DNA barcoding relies on comparing genetic distances of intraspecific and interspecific specimens. Generated sequences can be deposited in reference databases for future comparisons with novel data (Ratnasingham and Hebert 2007). Besides its applications in specimen identification, *cox1* barcodes may unveil cryptic diversity, shed light on species boundaries, and aid in phylogenetic and phylogeographic studies (DeSalle and Goldstein 2019).

Bees are known for their fundamental pollination role in ecosystems and their commercial and

Corresponding author: L. T. Gonçalves, tresoldigoncalves@gmail.com
Manuscript editor: Marina Meixner

Published online: 09 September 2022

scientific importance (Michener 2007). Among the bees, the Apidae form a clade comprising over 5900 valid species within 34 tribes (Table I; Bossert et al. 2019). Bee identification is based primarily on morphological characters and morphometric measurements (e.g., Michener 2007; Bustamante et al. 2021; Boustani et al. 2021; Schaller and Roig-Alsina 2021; Nogueira et al. 2022). However, morphology alone may be misleading since some bee species are challenging to distinguish, forming cryptic species complexes (e.g., some bumblebee species; Williams et al. 2012; Martinet et al. 2019). Specimens of certain life stages or castes may also lack informative characters, making it impossible to identify eggs, larvae, and most pupae to species (Michener 2007). Therefore, DNA barcoding has been employed as a supplementary tool to shed light on bee taxonomy (Schmidt et al. 2015; González-Vaquero et al. 2016; Packer and Ruz 2017).

The standard DNA barcode proposed by Hebert et al. (2003) corresponds to the so-called Folmer region, a 648 bp fragment at the 5′ end of the mitochondrial gene *cox1* amplified by the primers designed by Folmer et al. (1994). This region was initially chosen because it is informative and relatively easy to amplify, besides being sufficiently conserved within species yet variable between species (Hebert et al. 2003). Smaller regions of barcode markers—mini-barcodes—were developed for accurate identification in samples with degraded DNA (Hajibabaei et al. 2006). These markers can also be handy and cost-effective in high-throughput sequencing projects (Yeo et al. 2020). For bees, mini-barcodes may be applied to environmental samples, archived specimens (Françoso and Arias 2013), and commercial products such as honey (Schnell et al. 2010). A mini-barcode based on a 175 bp region of *cox1* was proposed for specimen identification of corbiculate bees (Figure 1) (Françoso and Arias 2013). However, the performance of this marker remains insufficiently tested: previous studies have only focused on a narrow taxonomic scope (e.g., Françoso and Arias 2013; Blasco-Lavilla et al. 2019), were limited to a regional bee fauna (e.g., Magnacca and Brown 2012; Sheffield et al. 2017), or evaluated mini-barcodes in a broader sense but did not include apid bees (e.g., Meusnier et al. 2008; Yeo et al. 2020).

In this context, we tested in silico if full-length barcodes and mini-barcodes perform similarly in specimen identification and species discovery of apid bees. We datamined all Apidae *cox1* sequences deposited in the Barcode of Life Data Systems (BOLD), aiming to (1) verify the existence of barcoding gaps in full-length and mini-barcodes at the generic level in all available apid genera, (2) test if full-length and mini-barcode regions have similar success rates of specimen identification, and (3) flag bee taxa that may benefit from integrative studies implementing DNA barcodes. Furthermore, we provide here an overview of the information available in BOLD concerning barcode sequences of apid bees.

2. MATERIALS AND METHODS

2.1. Data retrieval and filtering

We retrieved all BOLD sequences labeled as "Apidae" on November 27, 2020, generating separate FASTA files for each genus. To ensure robust analyses, we followed several filtering steps described by Bianchi and Gonçalves (2021a). Briefly, we only maintained sequences belonging to the 5′ region of *cox1* (labeled in BOLD as "COI-5P") and removed those entries without species-level identification (e.g., *Bombus* sp.). Then, we conducted preliminary alignments on MAFFT 7.0 (Katoh et al. 2019), seeking sequences with nonsense mutations, insertions, and deletions. These entries were removed from the datasets since we assumed they resulted from low-quality sequencing, erroneous amplification (e.g., nuclear mitochondrial DNA segments), or lab contamination. After the filtering steps, a final alignment round was conducted on MAFFT with default parameters.

AliView (Larsson 2014) was used to inspect the alignments and trim the sequences to the barcode region amplified by the primer pair BarbeeF (Françoso and Arias 2013) and MtD9 (Simon

Table I Taxonomic coverage of this study, sorted by subfamily and tribe following the revised generic classification of Bossert et al. (2019). In parentheses, the number of valid genera (sensu Bossert et al. 2019) and species (a rough estimative according to the Integrated Taxonomic Information System online database; http://www.itis.gov) for each tribe. Sampled sequence count is also provided, and estimated species coverage is given in percent (%)

Subfamily	Tribe	Genera	Species	Sequences	Species coverage (%)
Anthophorinae	Anthophorini	3 (7)	69 (794)	375	8.69
Apinae	Apini	1 (1)	8 (8)	2063	100.00
	Bombini	* 1 (1)	179 (280)	5566	63.93
	Centridini	2 (2)	18 (271)	49	6.64
	Euglossini	5 (5)	143 (248)	1615	57.66
	Meliponini	19 (51)	93 (518)	1712	17.95
Eucerinae	Ancylaini	0 (2)	0 (16)	0	0.00
	Emphorini	3 (10)	20 (120)	81	16.67
	Eucerini	6 (27)	131 (801)	549	16.35
	Exomalopsini	2 (5)	7 (156)	10	4.49
	Tapinotaspidini	5 (8)	23 (146)	36	15.75
Nomadinae	Ammobatini	1 (7)	2 (117)	3	1.71
	Ammobatoidini	2 (5)	6 (32)	10	18.75
	Biastini	1 (3)	2 (12)	3	16.67
	Brachynomadini	0 (5)	0 (26)	0	0.00
	Coelioxoidini	0 (1)	0 (4)	0	0.00
	Caenoprosopidini	0 (2)	0 (2)	0	0.00
	Epeolini	2 (8)	70 (309)	267	22.65
	Ericrocidini	3 (9)	9 (44)	14	20.45
	Hexepeolini	0 (1)	0 (1)	0	0.00
	Isepeolini	2 (2)	6 (21)	8	28.57
	Melectini	2 (9)	14 (206)	64	6.80
	Neolarrini	1 (1)	2 (16)	3	12.50
	Nomadini	1 (1)	95 (701)	738	13.55
	Osirini	1 (5)	2 (52)	7	3.85
	Protepeolini	1 (1)	2 (5)	3	40.00
	Rhathymini	0 (2)	0 (19)	0	0.00
	Townsendiellini	0 (1)	0 (3)	0	0.00
Xylocopinae	Allodapini	4 (16)	25 (248)	129	10.08
	Ceratinini	1 (1)	43 (339)	911	12.68
	Ctenoplectrini	1 (1)	5 (20)	11	25.00
	Manueliini	0 (1)	0 (3)	0	0.00
	Tetrapediini	0 (1)	0 (25)	0	0.00
	Xylocopini	1 (1)	38 (400)	351	9.50
		71 (203)	1012 (5963)	14,578	16.97

**Bombus* subgenera were also treated separately; our sample encompasses the 15 recognized Bombus subgenera sensu Williams et al. (2008)

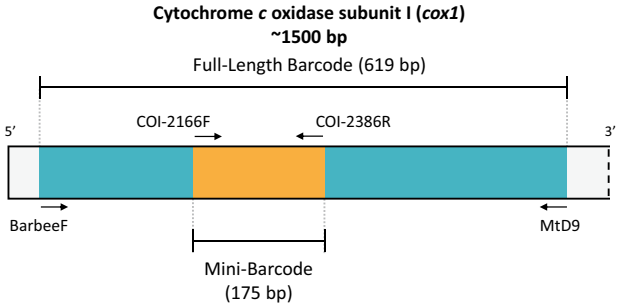

Figure 1. Schematic representation of the 5′ terminal of the mitochondrial gene cytochrome *c* oxidase subunit I (*cox1*). Primer pair BarbeeF and MtD9 amplify the full-length barcode in bees (619 bp); primer pair COI-2166F and COI-2386R amplify the mini-barcode (175 bp). Adapted from Françoso and Arias (2013)

et al. 1994), the first one specially developed to generate a 619 bp fragment of the Folmer region in bees (Figure 1). Sequences shorter than 400 bp were removed, and the remaining sequences composed our primary dataset (Full-Length Barcode Dataset). A secondary dataset (Mini-Barcode Dataset) was compiled, trimming the sequences from the Full-Length Barcode Dataset to a 175 bp mini-barcode region, which is amplified by the primer pair COI-2166F and COI-2386R (mini-barcode II; Françoso and Arias 2013) (Figure 1). As a final filtering step, we double-checked scientific names, correcting misspellings and non-valid names (i.e., synonyms) according to the Integrated Taxonomic Information System (www.itis.gov) and recent literature.

To guarantee intra- and interspecific comparisons, the analyses described below comprise only genera featuring at least two species, with at least one of the species represented by two or more sequences. Sequences identified as subspecies were treated at the species level. Although our analyses focused on the generic level, we also analyzed *Bombus* subgenera given the diversity of species and the high number of sequences recovered for this genus (see "Results"). We followed the simplified subgeneric classification of Williams et al. (2008). Our results are presented using the revised generic classification of Apidae from Bossert et al. (2019), and we include *Lanthanomelissa* as a valid genus of Tapinotaspidini (Ribeiro et al. 2021).

2.2. Data analysis

We generated separate FASTA files for each genus represented in the datasets for the barcoding gap analyses. The R package Spider (Brown et al. 2012) was used to estimate pairwise uncorrected p-distances for all sampled sequences within each genus. We opted to use uncorrected p-distances because they yield better or similar results in distance-based analyses when compared to other models of nucleotide substitution (e.g., K2P; see Collins et al. 2012; Srivathsan and Meier 2012). Intra- and interspecific distances of each genus were visualized in a boxplot. Boxplots are handy tools for data visualization: the line that divides the box into two parts represents the median of the data; box ends show the upper (Q3) and lower (Q1) quartiles; whiskers extend to Q3 + 1.5xIQR and Q1-1.5xIQR; dots show outlier values (McGill et al. 1978).

Based on the boxplots obtained for each genus, we followed Badotti et al. (2017) to sort *cox1* efficacy into three categories: *good*, *intermediate*, and *poor*. Efficacy was considered *good* when whiskers displayed a gap between intra- and interspecific comparisons, *intermediate* whenever the whiskers of intra- and interspecific comparisons overlapped, and *poor* when the boxes overlapped. Moreover, we used the function localMinima() implemented in Spider to set a threshold value for the scrutinized genera that could serve as a reference in future DNA barcoding studies. This function optimizes a

putative threshold value based on a gap in the density plot of genetic distances, disregarding sequence labels (Brown et al. 2012). Additionally, we assessed the number of informative characters of full-length and mini-barcodes for each genus using the function pis() of the R package ips (Heibl et al. 2019). Last, a Pearson Correlation (function cor.test() in base R) was used to examine the relationship between the number of informative characters of each marker.

We compared identification success between full-length and mini-barcodes by calculating the Probability of Correct Identification (PCI). We specifically adopted here the PCI metrics classified by Erickson et al. (2008) as "discrete species assignment", which considers the maximum intraspecific distance and the minimum interspecific distance (nearest-neighbor distance) for each species. Identification of species was considered successful if the maximum intraspecific distance of a species was less than its minimum interspecific distance. Then, we calculated the PCI for each genus as the proportion of species successfully identified. If the PCI of full-length and mini-barcodes differed for a given genus, the observed proportions were converted to a 2×2 contingency table and compared with a Fisher's exact test using the function fisher.test() implemented in base R (R Core Team 2021).

3. RESULTS

Our raw dataset consisted of 18167 *cox1* sequences. After the filtering steps and maintaining only genera with viable intra- and interspecific comparisons, 14578 sequences remained (Table I). We used these sequences to compile the full-length barcode and the mini-barcode datasets. A total of 393 sequences (2.69% of the dataset) had their labels changed due to misspelled or invalid species names. These changes are appended as supplementary material with all sampled species and sequences (Online Resources 1 and 2).

Regarding species coverage, our sample encompasses five subfamilies, 25 tribes, 71 genera, and 1012 species (around 17% of valid apid species; Table I). Sequence coverage by species ranged from 1 to 1162, with 77.86% of sampled species represented by less than ten sequences. Concerning sequence abundance, our final datasets present a strong bias towards Apinae (75.49%), followed by Xylocopinae (9.62%), Nomadinae (7.68%), Eucerinae (4.64%), and Anthophorinae (2.57%) (Table I). Corbiculate tribes (Apini, Bombini, Meliponini, and Euglossini) contributed with most of the sequences for the final datasets, although absolute species richness was higher for Bombini, Euglossini, and Eucerini, respectively (Table I). Our datasets, however, lack sequence data for eight apid tribes.

For most genera, *cox1* efficacy was considered *good* for both full-length and mini-barcodes (77.02 and 71.62%, respectively). Some genera, however, displayed *intermediate* (17.57 and 25.67%) and *poor* (5.41 and 2.70%) performances. Figure 2 illustrates the barcoding gap classifications we adopted in this study. The PCI of the two markers sometimes differed among genera, with full-length barcodes performing better for some taxa—*Anthophora*, *B.* (*Bombus*), *B.* (*Megabombus*), *Euglossa*, and *Epeolus*—but worst for others, *Caenonomada*, *Diadasia*, and *Triepeolus*. Intra- and interspecific distances fluctuated considerably among genera and between markers, affecting threshold values inferred by function localMinima(). The average threshold for the full-length barcodes was 2.88%, whereas mini-barcodes displayed an average of 3.31%. Marker performance and threshold values of each genus are presented in Table II.

The number of informative sites varied among genera, ranging from 0 to 112 for mini-barcodes and 0 to 366 for full-length barcodes (Figure 3A). The relationship between the number of informative sites of the markers was positive and strong ($r = 0.968$, $p < 0.001$). However, the informativeness was heavily affected by the sample size. For instance, *Bombus* presented the highest number of informative sites for both markers and was also the genus with the highest number of sequences. In contrast, genera represented by less than five sequences (such as *Erichocis*, *Leiopodus*, *Melectoides*, and *Neolarra*) often lacked informative sites.

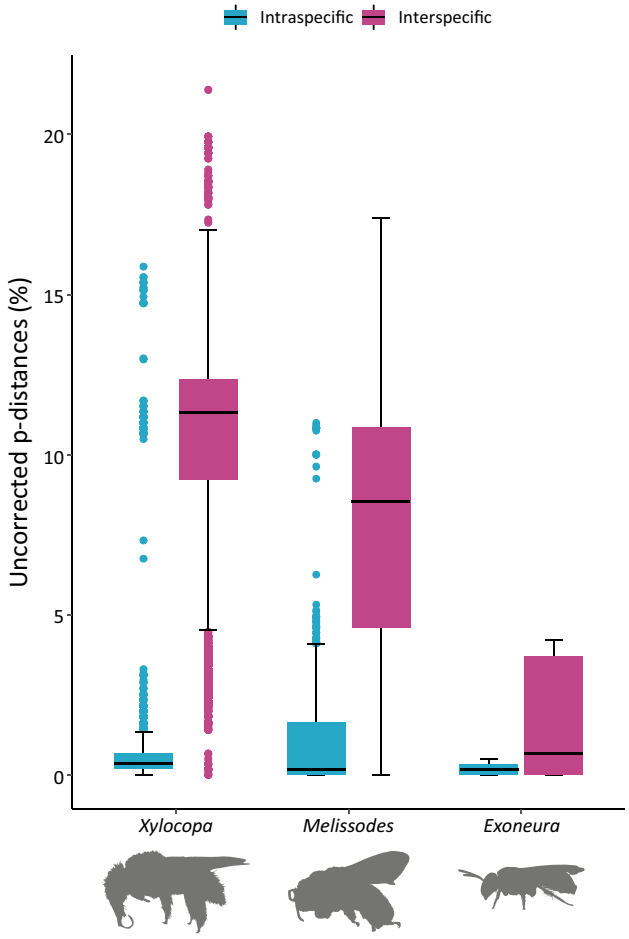

Figure 2. Examples of the barcoding gap performance classifications implemented in this study. The boxplots refer to comparisons done with the full-length dataset. *Xylocopa* displayed a *good* performance since intra- and interspecific boxes displayed a clear gap; the performance of *Melissodes* was classified as *intermediate* since the whiskers of the intra- and interspecific comparisons overlapped; intra- and interspecific boxes of *Exoneura* overlapped, implying a *poor* performance

Overall, full-length barcodes presented equal or higher PCI than mini-barcodes, whereas mini-barcodes of *Bombus (Psithyrus)* displayed a higher PCI than full-length barcodes (Table II). However, none of the differences were statistically significant (Table II). In general, PCI rates were higher for genera that displayed *good* and *intermediate* performances (Figure 3B). Full-length barcodes exhibited, on average, higher PCI rates for genera with *poor* performance. In short, our findings show that full-length barcodes and mini-barcodes have equivalent performance in bee identification.

4. DISCUSSION

Since its formal proposal in 2003, DNA barcoding has achieved the status of a revolutionary and reliable tool to support taxonomic studies (DeSalle and Goldstein 2019). Shorter DNA barcode sequences—mini-barcodes—have been conveniently employed in the taxonomic

Table II Barcoding gap, probability of correct identification (PCI), and local minima results for full-length (full) and mini-barcode (mini) datasets. When PCI differed between markers, the *p*-value of the Fisher's exact test is also shown

	Barcoding gap		PCI (%)			Local minima (%)[a]	
	Full	Mini	Full	Mini	*p*	Full	Mini
Anthophorinae							
Anthophorini							
Amegilla	Intermediate	Intermediate	71.43	71.43		2.15	2.80
Anthophora	Good	Intermediate	71.43	66.67	0.649	1.68	1.64
Habropoda	Good	Good	100.00	100.00		4.89	4.94
Apinae							
Apini							
Apis	Good	Good	62.50	62.50		0.76	0.56
Bombini							
Bombus	Good	Good	63.64	62.09	0.603	1.15	0.27
B. (*Alpigenobombus*)	Good	Good	66.67	66.67		4.48	6.26
B. (*Alpinobombus*)	Intermediate	Intermediate	42.86	42.86		1.19	1.30
B. (*Bombias*)	Good	Good	66.67	66.67		2.43	3.36
B. (*Bombus*)	Good	Intermediate	73.33	62.50	0.846	1.67	0.42
B. (*Cullumanobombus*)	Good	Good	75.00	75.00		3.91	4.46
B. (*Megabombus*)	Good	Intermediate	93.75	93.75		2.38	2.35
B. (*Melanobombus*)	Good	Good	100.00	100.00		1.01	0.29
B. (*Mendacibombus*)	Good	Good	100.00	100.00		5.82	5.43
B. (*Orientalibombus*)	Good	Good	100.00	100.00		0.52	0.27
B. (*Psithyrus*)	Good	Good	66.67	75.00	0.782	1.96	2.32
B. (*Pyrobombus*)	Good	Good	58.82	50.00	0.947	1.69	1.00
B. (*Subterraneobombus*)	Intermediate	Intermediate	80.00	80.00		0.88	0.84
B. (*Thoracobombus*)	Intermediate	Intermediate	64.29	64.29		1.86	1.64
Centridini							
Centris	Good	Good	100.00	100.00		2.38	2.83
Epicharis	Good	Good	100.00	100.00		6.04	6.93
Euglossini							
Eufriesea	Intermediate	Intermediate	35.29	29.41	0.991	0.68	0.82
Euglossa	Good	Intermediate	19.35	14.52	1.000	0.82	0.86
Eulaema	Good	Good	33.33	25.00	0.988	0.08	0.30
Exaerete	Intermediate	Intermediate	25.00	25.00		2.82	2.56
Ctenoplectrini							
Ctenoplectra	Good	Good	50.00	50.00		5.73	5.96
Emphorini							
Diadasia	Intermediate	Good	77.78	77.78		1.93	5.71
Melitoma	Good	Good	100.00	100.00		5.52	6.37
Ptilothrix	Good	Good	100.00	100.00		4.32	6.77

Table II (continued)

	Barcoding gap		PCI (%)			Local minima (%)[a]	
	Full	Mini	Full	Mini	*p*	Full	Mini
Ericrocidini							
Ericrocis	Good	Good	100.00	100.00		4.94	5.76
Hopliphora	Poor	Poor	0.00	0.00		NA	NA
Mesoplia	Good	Good	100.00	100.00		4.42	5.96
Eucerini							
Alloscirtetica	Good	Good	100.00	100.00		4.93	6.76
Eucera	Good	Good	84.21	73.68	0.810	1.59	1.42
Florilegus	Good	Good	100.00	100.00		1.57	NA
Melissodes	Intermediate	Intermediate	50.00	44.44	0.990	1.12	1.08
Svastra	Good	Good	100.00	100.00		4.77	5.39
Thygater	Intermediate	Intermediate	77.78	66.67	0.822	0.69	1.71
Exomalopsini							
Anthophorula	Good	Good	100.00	100.00		8.76	NA
Exomalopsis	Good	Good	100.00	100.00		3.56	8.59
Isepeolini							
Isepeolus	Good	Good	100.00	100.00		2.79	3.12
Melectoides	Good	Good	100.00	100.00		4.98	6.69
Melectini							
Melecta	Good	Good	100.00	100.00		4.16	5.76
Thyreus	Good	Good	100.00	100.00		3.44	4.15
Meliponini							
Cephalotrigona	Good	Good	100.00	100.00		2.89	3.99
Liotrigona	Good	Good	100.00	50.00	0.879	1.13	1.35
Melipona	Intermediate	Intermediate	23.53	11.76	0.996	0.58	5.01
Partamona	Good	Good	40.00	0.00	1.000	1.40	2.08
Plebeia	Intermediate	Intermediate	0.00	0.00		1.68	0.38
Scaptotrigona	Intermediate	Intermediate	50.00	16.67	0.984	0.47	0.30
Scaura	Good	Good	50.00	50.00		2.87	5.17
Tetragona	Good	Good	100.00	100.00		3.16	4.20
Tetragonisca	Intermediate	Intermediate	0.00	0.00		0.11	0.28
Trigona	Good	Good	28.57	28.57		3.20	3.29
Osirini							
Epeoloides	Good	Good	100.00	100.00		4.35	5.40
Protepeolini							
Leiopodus	Good	Good	100.00	100.00		4.70	4.40
Tapinotaspidini							
Arhysoceble	Good	Good	100.00	100.00		3.56	4.70
Caenonomada	Poor	Intermediate	66.67	33.33	0.929	3.46	3.65
Chalepogenus	Good	Good	100.00	100.00		3.03	4.57
Lanthanomelissa	Good	Good	100.00	100.00		1.45	1.53
Paratetrapedia	Good	Good	100.00	100.00		2.28	3.26

Table II (continued)

	Barcoding gap		PCI (%)			Local minima (%)[a]	
	Full	Mini	Full	Mini	*p*	Full	Mini
Nomadinae							
Ammobatini							
Oreopasites	Good	Good	100.00	100.00		3.99	5.12
Ammobatoidini							
Ammobatoides	Good	Good	100.00	100.00		5.51	6.05
Holcopasites	Good	Good	100.00	100.00		4.31	6.63
Biastini							
Biastes	Good	Good	100.00	100.00		3.85	3.61
Epeolini							
Epeolus	Good	Intermediate	80.00	75.00	0.802	1.87	2.54
Triepeolus	Poor	Intermediate	50.00	50.00		1.31	1.53
Neolarrini							
Neolarra	Good	Good	100.00	100.00		4.75	6.85
Nomadini							
Nomada	Good	Good	62.20	52.44	0.983	0.70	0.29
Xylocopinae							
Allodapini							
Braunsapis	Good	Good	0.00	0.00		3.32	4.28
Exoneura	Poor	Poor	0.00	0.00		2.40	0.39
Exoneurella	Good	Good	100.00	100.00		2.87	2.90
Macrogalea	Good	Good	66.67	66.67		10.20	2.63
Ceratinini							
Ceratina	Good	Good	68.75	68.75		0.71	0.96
Xylocopini							
Xylocopa	Good	Good	78.26	69.57	0.852	1.64	2.12
						2.88	3.31

[a]NA: when local minima failed to return a threshold for the given dataset

identification of ancient or degraded DNA (Meusnier et al. 2008; Cardeñosa et al. 2017; Erickson et al. 2017). A specific mini-barcode was developed for corbiculate bees (Françoso and Arias 2013) and has been applied to identify century-old museum specimens (Françoso and Arias 2013) and to assess bee biodiversity (e.g., Brettell et al. 2020; Nakamura et al. 2020). Traditional Sanger sequencing of full-length barcodes can be expensive when dealing with thousands of samples or ancient material (Yeo et al. 2020, 2021). In the genomics era, mini-barcodes with good taxonomic resolution can be employed in large-scale biodiversity studies using next-generation sequencing pipelines, costing equal or less than a morphology-based diagnostic system (Stein et al. 2014; Roe et al. 2017; Xing et al. 2021).

Here we retrieved all Apidae *cox1* sequences from BOLD to compare barcoding gap performance and identification success of full-length barcodes and mini-barcodes. We detected barcoding gaps for most of the genera and both markers, suggesting that these markers perform equally well in species

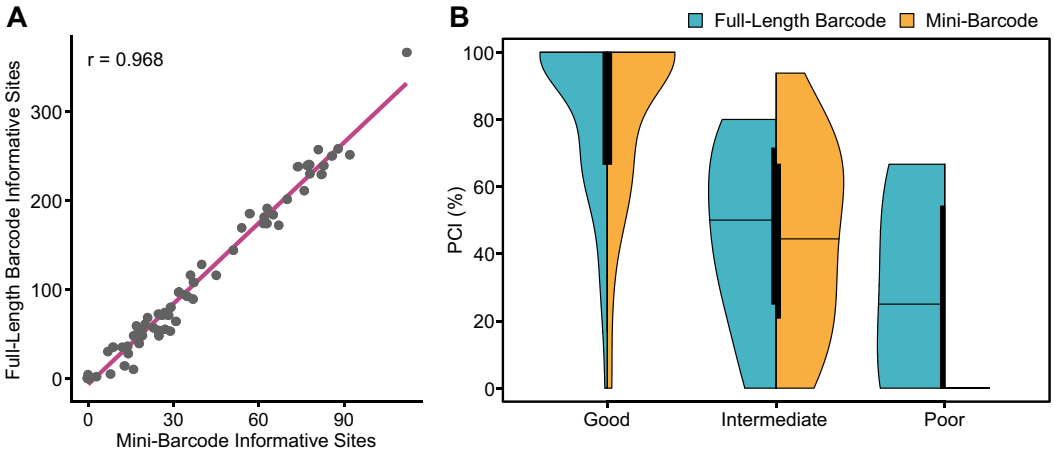

Fig. 3 **A** Relationship between the number of informative sites of full-length and mini-barcodes. Each point refers to a genus of the dataset. The solid line represents the regression line between variables. **B** Violin plot comparing identification success between full-length and mini-barcodes, sorted by barcoding gap categories. Lines show the medians; box limits indicate the 25th and 75th percentiles; colored areas extend 1.5 times the interquartile range from the 25th and 75th percentiles

discovery for apid bees—that is, when one of the barcode regions was considered *good*, the other region displayed the same performance. Furthermore, our results indicate that both markers also perform similarly concerning identification success, showing that mini-barcodes are a reliable supplementary tool to specimen identification and species delimitation considering a broader taxonomic scope. Here we also provide barcode threshold values for 78 bee genera that can aid future taxonomic projects (Table II), which can help determine clusters for further assessment with additional methods.

4.1. BOLD overview

Several DNA barcoding projects have been conducted for pollinators in the past years, especially in North America and Europe (e.g., Sheffield et al. 2009; Magnacca and Brown 2012; Packer and Ruz 2017). The available data on BOLD is overall biased towards corbiculate Apidae, often recognized as the commercially most important group of bees (Martins et al. 2014). Sequences belonging to *Apis* (Apini) and *Bombus* (Bombini) represented more than half of the dataset. These genera are widely studied due to their economic importance in crop pollination and honey production. Because of this commercial relevance, such taxa tend to be well characterized in reference databases (Virgilio et al. 2010).

After the filtration steps, eight apid tribes were absent in our dataset, reflecting the paucity of data for some taxa (Table I). These tribes comprise around 100 solitary and cleptoparasitic species. Even though most known bee species are solitary (Michener 2007), genetic and molecular studies usually focus on social species (Neumann and Seidelmann 2006), leading to an underrepresentation of solitary bee species in public barcode databases. Identifying solitary bees can be a nightmare even for experienced taxonomists because these species are usually small and exhibit cryptic morphology (Magnacca and Brown 2012). DNA barcoding can be a valuable tool to aid in specimen identification and species delimitation in this scenario. Since solitary bees also play a significant role in pollination services, DNA barcoding efforts focused on these species may also be fundamental to studying macroecological phenomena such as the worldwide decline in pollinators and animal-pollinated plant species (Pornon et al. 2017; Vamosi et al. 2017).

We detected 393 sequences (2.7% of the dataset) with invalid or misspelled species names during the filtration process. Although this number is somewhat concerning, previous surveys of BOLD sequences found a much higher proportion of invalid or misspelled names for other taxa (e.g., around 12% for true bugs; Bianchi and Gonçalves 2021a). A reliable reference database is fundamental to identifying specimens using DNA barcoding (DeSalle and Goldstein 2019), and incorrect taxonomy will inevitably hinder the effectiveness of this tool. Since identification errors are inherent to any public DNA repository (Meiklejohn et al. 2019; Bianchi and Gonçalves 2021b), data from these sources must be used with caution.

4.2. Barcoding gaps and taxonomic inconsistencies

As shown by the boxplots and the local minima analyses, barcoding gap values varied widely among the scrutinized genera, which could be explained by the different coalescence times of each lineage (Fujita et al. 2012). From a single-locus point of view, a recurrent debate is that the evolutionary story of a gene (like *cox1*) does not necessarily depict the evolution of the species (Knowles 2009). Furthermore, evolutionary events such as introgression, incomplete lineage sorting, heteroplasmy, and hybridization may further hinder single-locus approaches like traditional DNA barcoding (Moritz and Cicero 2004; Magnacca and Brown 2010). Inconsistencies that may arise from these biological factors may be mitigated, for instance, by multi-marker barcoding approaches (e.g., Cruaud et al. 2017). However, a myriad of operational biases—that is, non-biological factors—may be much more relevant to undermine DNA barcoding effectiveness, including (but not limited to) inaccurate reference taxonomy, misidentifications, spelling errors, contamination, and low-quality sequences (Mutanen et al. 2016). Since a gap between intra- and interspecific distances does not necessarily imply correct identification in DNA barcoding studies (see Collins and Cruickshank 2012), we separately evaluated identification success by calculating the PCI.

Most genera showed high PCI and a *good* barcoding gap performance for both full-length and mini-barcodes. This pattern is consistent with previous studies using similar metrics to evaluate barcode efficiencies for other taxonomic groups such as fungi (Badotti et al. 2017), nematodes (Gonçalves et al. 2021), and true bugs (Bianchi and Gonçalves 2021a). However, genera with *intermediate* or *poor* performance—which overall exhibited lower PCI—require special attention as these results may hint at operational biases and taxonomic inconsistencies. Although our objective here was not to discuss taxonomic details about the sampled taxa, we bring below some examples.

Melissodes (Eucerini) is a diverse genus of solitary bees (ca. 129 species) whose taxonomic determination at the species level is challenging (Wright et al. 2020). Moreover, identification keys for these species often refer to color and chaetotaxy (e.g., LaBerge 1956a, b), morphological characters that are not always available and are prone to deterioration in archived specimens. Thus, sequences generated for this genus are susceptible to misidentification in public databases due to problems in reference taxonomy, which may explain the PCI around 50%. Moreover, we found an *intermediate* performance in the barcoding gap analyses for *Melissodes* (Figure 2), and several of the sequences of this genus were labeled with invalid names (Table S2). These results may flag a relatively high number of misidentifications in BOLD and show that the taxonomic identity of sequences from this genus should be verified and analyzed for unknown cryptic diversity.

Exoneura (Allodapini), although common in temperate parts of Australia, is characterized by its intricate and volatile taxonomy (Michener 2007; Yong et al. 2020). Accordingly, *Exoneura* was one of the few genera that displayed a *poor* barcoding gap performance in our analysis, with a PCI of 0%. Although these results are partly explained by the low sample size (see Meyer and Paulay 2005), they emphasize the need for a taxonomic revision and deeper investigations concerning specimen identification

and species delimitation of this genus. To our knowledge, there is no recent taxonomic review for *Exoneura*, which could be one of the reasons behind the poor performance. Therefore, we believe that genera with *poor* barcoding gap and low PCI values, such as *Exoneura*, should be prioritized in future taxonomic works. As shown in multiple studies, DNA barcodes can be valuable resources in integrative approaches to unravel cryptic species complexes (e.g., Sheffield et al. 2017; Milam et al. 2020; Williams 2021).

4.3. Comparing barcodes: the longer, the better?

Although mini-barcodes represent roughly 25% of the canonical barcode region, we found that they still maintain significant information, explaining why the performance and identification success of full-length and mini-barcodes are very similar. Overall, both markers showed a high number of informative sites, and as new sequences become available for each genus, this number will inevitably increase. These results emphasize the remarks of Françoso and Arias (2013) about this mini-barcode region, reaffirming its effectiveness for specimen identification. Moreover, because of the greater relative genetic divergence of mini-barcodes, we found that their threshold values tend to be higher.

Since our work focused on the in silico evaluation of the barcodes, studies implementing other characters (such as morphology) must validate our results. However, our findings reinforce the utility of mini-barcodes in bee taxonomy, especially in scenarios where the amplification of full-length barcodes is compromised due to DNA degradation (e.g., archived specimens and environmental DNA). Moreover, genetic methods of specimen identification and species delimitation are prominent tools for discovering putative cryptic species; multiple evidence approaches are then necessary to corroborate these findings.

Studies related to bee identification are crucial to their conservation and to maintain ecosystem services (Zayed 2009; Lozier and Zayed 2017). DNA barcodes can be valuable allies to unveil cryptic diversity, flag taxonomic inconsistencies, and improve species discovery (DeSalle and Goldstein 2019; Bianchi and Gonçalves 2021a). By scrutinizing *cox1* sequences retrieved from BOLD, we showed that both full-length and mini-barcodes could be successfully employed in bee identification. We believe the reported results and analyses may help researchers identify species groups needing taxonomic revision as the first step in an integrative taxonomy workflow. Our findings may aid future research concerning apid bee diversity, taxonomy, and systematics.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1007/s13592-022-00958-x.

ACKNOWLEDGEMENTS

We thank Dr. Filipe Michels Bianchi for the feedback on an earlier draft of the manuscript and two anonymous reviewers for their valuable comments. LTG is supported by a doctoral fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

AUTHOR CONTRIBUTION

LTG and MD conceptualized the research; LTG conducted data analyses and wrote the manuscript draft; LTG, EF, and MD reviewed and edited the manuscript; MD supervised the research. All authors read and approved the final manuscript.

AVAILABILITY OF DATA AND MATERIAL

All sequences analyzed during this study are publicly available at the Barcode of Life Data System (https://www.boldsystems.org/). Accession numbers for these sequences are available as Online Resources.

CODE AVAILABILITY

Not applicable.

DECLARATIONS

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Competing interests The authors declare no competing interests.

REFERENCES

Badotti F, de Oliveira FS, Garcia CF et al (2017) Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). BMC Microbiol 17:42. https://doi.org/10.1186/s12866-017-0958-x

Bianchi FM, Gonçalves LT (2021a) Borrowing the Pentatomomorpha tome from the DNA barcode library: scanning the overall performance of *cox1* as a tool. J Zool Syst Evol Res 59:992–1012. https://doi.org/10.1111/jzs.12476

Bianchi FM, Gonçalves LT (2021b) Getting science priorities straight: how to increase the reliability of specimen identification? Biol Lett 17:rsbl.2020.0874, 20200874. https://doi.org/10.1098/rsbl.2020.0874

Blasco-Lavilla N, Ornosa C, Michez D, De la Rúa P (2019) Contrasting patterns of genetic and morphological diversity in the bumblebee *Bombus lucorum* (Hymenoptera: Apidae: *Bombus*) along a European gradient. J Insect Conserv 23:933–943. https://doi.org/10.1007/s10841-019-00178-2

Bossert S, Murray EA, Almeida EAB et al (2019) Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol Phylogenet Evol 130:121–131. https://doi.org/10.1016/j.ympev.2018.10.012

Boustani M, Rasmont P, Dathe HH et al (2021) The bees of Lebanon (Hymenoptera: Apoidea: Anthophila). Zootaxa 4976:1–146. https://doi.org/10.11646/zootaxa.4976.1.1

Brettell LE, Riegler M, O'Brien C, Cook JM (2020) Occurrence of honey bee-associated pathogens in *Varroa*-free pollinator communities. J Invertebr Pathol 171:107344. https://doi.org/10.1016/j.jip.2020.107344

Brown SDJ, Collins RA, Boyer S et al (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12:562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x

Bustamante T, Fuchs S, Grünewald B, Ellis JD (2021) A geometric morphometric method and web application for identifying honey bee species (*Apis* spp.) using only forewings. Apidologie 52:697–706. https://doi.org/10.1007/s13592-021-00857-7

Cardeñosa D, Fields A, Abercrombie D et al (2017) A multiplex PCR mini-barcode assay to identify processed shark products in the global trade. PLoS ONE 12:e0185368. https://doi.org/10.1371/journal.pone.0185368

Collins RA, Boykin LM, Cruickshank RH, Armstrong KF (2012) Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. Methods Ecol Evol 3:457–465. https://doi.org/10.1111/j.2041-210X.2011.00176.x

Collins RA, Cruickshank RH (2012) The seven deadly sins of DNA barcoding. Mol Ecol Resour 13:969–975. https://doi.org/10.1111/1755-0998.12046

Cruaud P, Rasplus J-Y, Rodriguez LJ, Cruaud A (2017) High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. Sci Rep 7:41948. https://doi.org/10.1038/srep41948

DeSalle R, Goldstein P (2019) Review and interpretation of trends in DNA barcoding. Front Ecol Evol 7:302. https://doi.org/10.3389/fevo.2019.00302

Erickson DL, Reed E, Ramachandran P et al (2017) Reconstructing a herbivore's diet using a novel rbcL DNA mini-barcode for plants. AoB PLANTS 9:plx015. https://doi.org/10.1093/aobpla/plx015

Erickson DL, Spouge J, Resch A et al (2008) DNA barcoding in land plants: developing standards to quantify and maximize success. Taxon 57:1304–1316. https://doi.org/10.1002/tax.574020

Folmer O, Black M, Hoeh W et al (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299

Françoso E, Arias MC (2013) Cytochrome c oxidase I primers for corbiculate bees: DNA barcode and mini-barcode. Mol Ecol Resour 13:844–850. https://doi.org/10.1111/1755-0998.12135

Fujita MK, Leaché AD, Burbrink FT et al (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol 27:480–488. https://doi.org/10.1016/j.tree.2012.04.012

Gonçalves LT, Bianchi FM, Deprá M, Calegaro-Marques C (2021) Barcoding a can of worms: testing *cox1* performance as a DNA barcode of Nematoda. Genome 64:705–717. https://doi.org/10.1139/gen-2020-0140

González-Vaquero RA, Roig-Alsina A, Packer L (2016) DNA barcoding as a useful tool in the systematic study of wild bees of the tribe Augochlorini (Hymenoptera: Halictidae). Genome 59:889–898. https://doi.org/10.1139/gen-2016-0006

Hajibabaei M, Smith MA, Janzen DH et al (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964. https://doi.org/10.1111/j.1471-8286.2006.01470.x

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc B Biol Sci 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Heibl C, Cusimano N, Krah F-S (2019) Interfaces to Phylogenetic Software in R. https://CRAN.R-project.org/package=ips

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.

Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Knowles LL (2009) Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. Syst Biol 58:463–467. https://doi.org/10.1093/sysbio/syp061

LaBerge WE (1956a) A revision of the bees of the genus *Melissodes* in North and Central America. Part I (Hymenoptera, Apidae). Univ Kans Sci Bull 37:911–1194

LaBerge WE (1956b) A revision of the bees of the genus *Melissodes* in North and Central America. Part II (Hymenoptera, Apidae). Univ Kans Sci Bull 38:533–578

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Lozier JD, Zayed A (2017) Bee conservation in the age of genomics. Conserv Genet 18:713–729. https://doi.org/10.1007/s10592-016-0893-7

Magnacca KN, Brown MJ (2010) Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus (Nesoprosopis)* bees (Hymenoptera: Colletidae). BMC Evol Biol 10:174. https://doi.org/10.1186/1471-2148-10-174

Magnacca KN, Brown MJF (2012) DNA barcoding a regional fauna: Irish solitary bees. Mol Ecol Resour 12:990–998. https://doi.org/10.1111/1755-0998.12001

Martinet B, Lecocq T, Brasero N et al (2019) Integrative taxonomy of an arctic bumblebee species complex highlights a new cryptic species (Apidae: *Bombus*). Zool J Linn Soc 187:599–621. https://doi.org/10.1093/zoolinnean/zlz041

Martins AC, Melo GAR, Renner SS (2014) The corbiculate bees arose from New World oil-collecting bees: implications for the origin of pollen baskets. Mol Phylogenet Evol 80:88–94. https://doi.org/10.1016/j.ympev.2014.07.003

McGill R, Tukey JW, Larsen WA (1978) Variations of box plots. Am Stat 32:12. https://doi.org/10.2307/2683468

Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank – their accuracy and reliability for the identification of biological materials. PLoS ONE 14:e0217084. https://doi.org/10.1371/journal.pone.0217084

Meusnier I, Singer GA, Landry J-F et al (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9:214. https://doi.org/10.1186/1471-2164-9-214

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biol 3:e422. https://doi.org/10.1371/journal.pbio.0030422

Michener CD (2007) The bees of the world, 2nd edn. Johns Hopkins University Press, Baltimore

Milam J, Johnson DE, Andersen JC et al (2020) Validating morphometrics with DNA barcoding to reliably separate three cryptic species of *Bombus* Cresson (Hymenoptera: Apidae). Insects 11:669. https://doi.org/10.3390/insects11100669

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. PLoS Biol 2:e354. https://doi.org/10.1371/journal.pbio.0020354

Mutanen M, Kivelä SM, Vos RA et al (2016) Species-level para- and polyphyly in DNA barcode gene trees: strong operational bias in European Lepidoptera. Syst Biol 65:1024–1040. https://doi.org/10.1093/sysbio/syw044

Nakamura S, Yamamoto S, Sawamura N et al (2020) Pollination effectiveness of European honeybee, *Apis mellifera* (Hymenoptera: Apidae), in an Oriental persimmon, *Diospyros kaki* (Ericales: Ebenaceae), orchard. Appl Entomol Zool 55:405–412. https://doi.org/10.1007/s13355-020-00696-5

Neumann K, Seidelmann K (2006) Microsatellites for the inference of population structures in the Red Mason bee *Osmia rufa* (Hymenoptera, Megachilidae). Apidologie 37:75–83. https://doi.org/10.1051/apido:2005060

Nogueira DS, De Oliveira FF, De Oliveira ML (2022) Revision of the *Tetragona clavipes* (Fabricius, 1804) species-group (Hymenoptera: Apidae: Meliponini). Zootaxa 5119:1–64. https://doi.org/10.11646/zootaxa.5119.1.1

Packer L, Ruz L (2017) DNA barcoding the bees (Hymenoptera: Apoidea) of Chile: species discovery in a reasonably well known bee fauna with the description of a new species of *Lonchopria* (Colletidae). Genome 60:414–430. https://doi.org/10.1139/gen-2016-0071

Pornon A, Andalo C, Burrus M, Escaravage N (2017) DNA metabarcoding data unveils invisible pollination networks. Sci Rep 7:16828. https://doi.org/10.1038/s41598-017-16785-5

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ribeiro TMA, Martins AC, Silva DP, Aguiar AJC (2021) Systematics of the oil bee genus *Lanthanomelissa* (Apidae: Tapinotaspidini) and its implications for the biogeography of South American grasslands. J Zool Syst Evol Res 59:1013–1027. https://doi.org/10.1111/jzs.12472

Roe A, Dupuis J, Sperling F (2017) Molecular dimensions of insect taxonomy in the Genomics Era. In: Foottit RG, Adler PH (eds) Insect Biodiversity. John Wiley & Sons, Ltd, Chichester, UK, pp 547–573

Schaller A, Roig-Alsina A (2021) A revision of the bee genus *Ancyloscelis* (Hymenoptera: Apidae) in

Argentina. Zootaxa 4980:521–540. https://doi.org/10.11646/zootaxa.4980.3.4

Schmidt S, Schmid-Egger C, Morinière J et al (2015) DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). Mol Ecol Resour 15:985–1000. https://doi.org/10.1111/1755-0998.12363

Schnell IB, Fraser M, Willerslev E, Gilbert MTP (2010) Characterisation of insect and plant origins using DNA extracted from small volumes of bee honey. Arthropod-Plant Interact 4:107–116. https://doi.org/10.1007/s11829-010-9089-0

Sheffield CS, Hebert PDN, Kevan PG, Packer L (2009) DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol Ecol Resour 9:196–207. https://doi.org/10.1111/j.1755-0998.2009.02645.x

Sheffield CS, Heron J, Gibbs J et al (2017) Contribution of DNA barcoding to the study of the bees (Hymenoptera: Apoidea) of Canada: progress to date. Can Entomol 149:736–754. https://doi.org/10.4039/tce.2017.49

Simon C, Frati F, Beckenbach A et al (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87:651–701. https://doi.org/10.1093/aesa/87.6.651

Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28:190–194. https://doi.org/10.1111/j.1096-0031.2011.00370.x

Stein ED, Martinez MC, Stiles S et al (2014) Is DNA barcoding actually cheaper and faster than traditional morphological methods? Results from a survey of freshwater bioassessment efforts in the United States. PLoS ONE 9:e95525. https://doi.org/10.1371/journal.pone.0095525

Vamosi JC, Gong Y-B, Adamowicz SJ, Packer L (2017) Forecasting pollination declines through DNA barcoding: the potential contributions of macroecological and macroevolutionary scales of inquiry. New Phytol 214:11–18. https://doi.org/10.1111/nph.14356

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

Williams PH (2021) Not just cryptic, but a barcode bush: PTP re-analysis of global data for the bumblebee subgenus *Bombus s. str.* supports additional species (Apidae, genus *Bombus*). J Nat Hist 55:271–282. https://doi.org/10.1080/00222933.2021.1900444

Williams PH, Brown MJF, Carolan JC et al (2012) Unveiling cryptic species of the bumblebee subgenus *Bombus s. str.* worldwide with COI barcodes (Hymenoptera: Apidae). Syst Biodivers 10:21–56. https://doi.org/10.1080/14772000.2012.664574

Williams PH, Cameron SA, Hines HM et al (2008) A simplified subgeneric classification of the bumblebees (genus *Bombus*). Apidologie 39:46–74. https://doi.org/10.1051/apido:2007052

Wright KW, Miller KB, Song H (2020) A molecular phylogeny of the long-horned bees in the genus *Melissodes* Latreille (Hymenoptera: Apidae: Eucerinae). Insect Syst Evol 1–16. https://doi.org/10.1163/1876312X-bja10015

Xing R-R, Hu R-R, Wang N et al (2021) Authentication of sea cucumber products using NGS-based DNA mini-barcoding. Food Control 129:108199. https://doi.org/10.1016/j.foodcont.2021.108199

Yeo D, Srivathsan A, Meier R (2020) Longer is not always better: optimizing barcode length for large-scale species discovery and identification. Syst Biol 69:999–1015. https://doi.org/10.1093/sysbio/syaa014

Yeo H, Yeoh TX, Ding H et al (2021) Morphology and mini-barcodes: the inclusion of larval sampling and NGS-based barcoding improves robustness of ecological analyses of mosquito communities. J Appl Ecol 1365–2664:13966. https://doi.org/10.1111/1365-2664.13966

Yong DL, Barton PS, Okada S et al (2020) Conserving focal insect groups in woodland remnants: the role of landscape context and habitat structure on cross-taxonomic congruence. Ecol Indic 115:106391. https://doi.org/10.1016/j.ecolind.2020.106391

Zayed A (2009) Bee genetics and conservation. Apidologie 40:237–262. https://doi.org/10.1051/apido/2009026

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.