A New Primer Pair for Barcoding Bees (Hymenoptera, Anthophila) Without Amplifying the Homologous *Wolbachia* Gene

Christoph Bleidorn (cbleido@gwdg.de)

University of Göttingen Faculty of Biology and Psychology: Georg-August-Universitat Gottingen Fakultat fur Biologie und Psychologie https://orcid.org/0000-0001-7295-6616

Katharina Henze

University of Göttingen Faculty of Biology and Psychology: Georg-August-Universitat Gottingen Fakultat fur Biologie und Psychologie

Research note

Keywords: Apoidea, biomonitoring, DNA barcoding, misamplification

Posted Date: July 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-669019/v1

Abstract

Objectives

DNA barcoding became an important method for the identification and monitoring of bees. However, standard primer pairs used for barcoding often result in (co-) amplification of bacterial endosymbionts of the genus *Wolbachia*, which are widespread among bee species. Here we designed a new primer pair and compared it with the performance of the standard Folmer-primers for small sample set of bees representing the main taxonomic groups of bees.

Results

The newly designed primer pair (BeeCox1F1/BeeCox1R2) clearly outperformed the standard barcoding primer (LCO1490/HCO2198). By generating barcodes for a small test set of bees we found that the new primer pair produced in all cases clear sequences for unambiguous species identification using BOLD. In contrast, the standard barcoding primers often resulted in the amplification of the homologous *Wolbachia* gene, which either resulted in a mixed chromatogram signal or identification of the bacterial endosymbiont instead of the host.

Introduction

More than 20,000 species of bees (Hymenoptera, Anthophila) are estimated to occur worldwide [1] and they became a posterchild for conservation biology as they are regarded as the economically most important group of pollinators in both natural and managed ecosystems [2]. Unfortunately, bee diversity and abundance has been reported to decline at different levels across continents [3]. To better understand changes in occurrence of bee species across scales exhaustive faunistic inventories are necessary. However, bee taxonomy can be difficult and different ways to accelerate their biomonitoring have been suggested [4, 5]. The most prominent approach is DNA barcoding, where a specific segment of the mitochondrial cytochrome oxidase 1 gene (*cox1*) is used for species identification [6]. As such, several geographic region specific barcoding initiatives for bees have been launched or already successfully finished, e.g., Central Europe [7], Ireland [8], Canada [9], providing the necessary background for DNA-based identification.

While in general the barcoding approach has been found working well for the so far analysed regional bee faunae, several practical problems have been identified. Most notoriously, when using standard barcoding primers (eg., LCO1490 and HCO2198 [10], or LepF1 and LepR1 [11]) sequencing failures are reported and misamplification of endosymbiotic bacteria of the genus *Wolbachia* has been found to interfere with the amplification of the targeted bee DNA [8, 12]. Screening submitted data to the BOLD database revealed that Hymenoptera account for the highest number of cases where it came to un-intended amplification of *Wolbachia* DNA [13]. *Wolbachia* are maternally inherited intracellular Alphaproteobacteria that are found to be widespread in (terrestrial) arthropods, but also are also known from selected nematode taxa [14]. By screening the German bee fauna, a prevalence of > 60% of

Wolbachia infected species has been reported [15, 16]. There seem to be taxon-specific patterns regarding the frequency of infected species, and especially a high number of species from the species-rich genera *Andrena*, *Halictus*, *Lasioglossum*, *Nomada* or *Sphecodes* are infected [16]. In congruence with what was reported by [8] our own routine work in the lab showed that individuals of these genera are difficult to barcode using standard approaches, as sequencing revealed mixed signals or the *Wolbachia* sequence. This comes to no surprise, as the most common used standard barcoding primers (LCO1490/HCO2198, LepF1/LepR1) actually show a high similarity to the homologous region of the *Wolbachia coxA* gene (Fig. 1). Alternative primer pairs have been already suggested as a workaround, but these seem to need rather low (and therefore unspecific) annealing temperatures (the primer pair BarBeeF/MtD9, as proposed for corbiculate bees by Francoso & Arias [17]) or involve highly degenerate priming sites (Hymeno-1/HCO-Hymeno-2 as proposed by Vilalta et al. [12]), which often hampers ultimate sequencing success. Here we present a newly designed primer combination and test its suitability for a phylogenetically representative taxon sampling of bees, with a special focus on its suitability in the case of *Wolbachia* infected species.

Methods

Species collection and DNA extraction

During fieldwork for a faunistic inventory of the FFH-protected area “Ballertasche”, an active sand and gravel pit (DD 51.45827462120174, 9.63633266623266) 4 km North of Hann. Münden, Lower Saxony, Germany, some individuals of bees were collected for subsequent barcoding. First, specimens were identified based on morphology to the genus or species level and later processed for molecular lab work. Thorax muscle tissue was removed using forceps and transferred directly into an Eppendorf cap for DNA extraction using the Qiagen DNEasy Blood and Tissue Kit (Qiagen, Hilden).

Primer design and PCR

Initially, selected bees including difficult to identify groups of species (e.g., females of the *Andrena ovatula* group) were used for DNA barcoding using the standard Folmer primer pair (LCO1490/HCO2198). As the resulting sequences in many cases were either stemming from *Wolbachia* endosymbionts or have been impossible to analyse due to multiple signals (see below for details on results) a new pair of primers for barcoding bees was designed. For this purpose, we downloaded complete *cox1* sequences from published mitochondrial genomes for a phylogenetically representative taxon sampling of bees from NCBI GenBank. This dataset includes 66 species of bees (Anthophila). Moreover, we also included sequences from the homologues region of *Wolbachia* endosymbiotic Alphaprotebacteria for the common supergroups A (*wMel*) and B (*wAlb*), which show a high similarity with the primer regions of the standard barcoding fragments. We used the MAFFT online service [18] using default parameters and automatic strategy selection for alignment. We manually aligned the typically insect barcoding primer pairs (LCO1490/HCO2198, LepF1/LepR1) with this dataset to delimitate the barcoding fragment of the gene. The alignment including NCBI accession numbers is available at https://github.com/Animal-Evolution-and-Biodiversity/Design_Barcoding_Primer_Bees/tree/main/Alignment. Finally, we designed a primer pair

for bees which spans over the complete barcoding fragment, but uses a primer site which is more divergent from the *Wolbachia* sequences for the reverse fragment. The primer pair was named BeeCox1F1 (TAGTCAACAAATCATAAAGATATTGG) and BeeCox1R2 (CCAAATCCTGGTAGAATTAAAATATA). We used the NetPrimer online tool (http://www.premierbiosoft.com/netprimer/) to check the suitability of this primer pair (melting temperature, secondary structures and cross dimers) and finally checked usability in the lab by running PCR's using selected bees which partly had been already identified to show misamplifications with the standard primer pairs (Table 1).

Table 1

Species level taxonomic assignment success of barcodes from different primer pairs of selected bee species representing different families. Assignment success was either given as unambiguous (correct) identification in BOLD as CORRECT, misamplification of endosymbiont as WOLBACHIA, or a sequence that showed double or mixed signal and could not be identified as MIXED.

Taxon	LCO1490/HCO2198	BeeCox1F1/BeeCox1R2
Megachilidae		
Coelioxys alata (1)	CORRECT	CORRECT
Megachile circumcincta (2)	CORRECT	CORRECT
Andrenidae		
Andrena barbilabris (4)	CORRECT	CORRECT
Andrena flavipes (6)	WOLBACHIA	CORRECT
Andrena ovatula (5)	MIXED	CORRECT
Andrena ovatula (7)	WOLBACHIA	CORRECT
Andrena minutula (15)	CORRECT	CORRECT
Andrena nigroaenea (8)	WOLBACHIA	CORRECT
Andrena wilkella (35)	MIXED	CORRECT
Andrena wilkella (36)	MIXED	CORRECT
Andrena wilkella (37)	MIXED	CORRECT
Halictidae		
Lasioglossum minutulum (14)	MIXED	CORRECT
Lasioglossum nitidiusculum (41)	CORRECT	CORRECT
Lasioglossum pauxillum (49)	CORRECT	CORRECT
Lasioglossum semilucens (46)	WOLBACHIA	CORRECT
Lasioglossum sexstrigatum (16)	CORRECT	CORRECT
Lasioglossum villosulum (48)	CORRECT	CORRECT
Sphecodes miniatus (19)	CORRECT	CORRECT
Sphecodes monilicornis (18)	WOLBACHIA	CORRECT
Sphecodes puncticeps (17)	CORRECT	CORRECT
Colletidae		
Hylaeus brevicornis (13)	CORRECT	CORRECT

Taxon	LCO1490/HCO2198	BeeCox1F1/BeeCox1R2
Hylaeus communis (12)	WOLBACHIA	CORRECT
Apidae		
Bombus terrestris (72)	CORRECT	CORRECT
Nomada succincta (71)	MIXED	CORRECT
Mellitidae		
Melitta leporina (M10)	MIXED	CORRECT

We focused here comparing the suitability for using the new BeeCox1F1/BeeCox1R2 primer combination and the standard LCO1490/HCO2198. The PCR regime for the new BeeCox1F1/BeeCox1R2 was as follows: Initial denaturation at 94°C for 2 min; 40 cycles of 30 sec at 94°C, 45 sec at 50°C and 1 min at 72°C; Final extension for 10 min at 72°C. For the LCO1490/HCO219 the primer pair we used: Initial denaturation at 94°C for 2 min; 40 cycles of 30 sec at 94°C, 45 sec at 46°C and 1 min at 72°C; Final extension for 10 min at 72°C. All PCRs were carried out in a total volume of 25 µl, containing 1 µl of each primer (10 pM), 12.5 µl DreamTaq Green PCR Master Mix (Thermo Fisher), 1 µl of genomic DNA and 9.5 µl of ddH2O. Negative controls containing water instead of template DNA were included in all PCRs. Sanger sequencing for forward directions for all fragments was performed by Microsynth Seqlab GmbH (Germany). All chromatograms and fasta-files from Sanger sequencing are available at https://github.com/Animal-Evolution-and-Biodiversity/Design_Barcoding_Primer_Bees/tree/main/Sanger_sequence_files.

Taxonomic assignment

Sequence quality was verified by inspecting the chromatograms using Chromas 2.6.6 (http://technelysium.com.au/wp/chromas/) and sequences were subsequently submitted to the identification page of the BOLD database (https://www.boldsystems.org/). Results were summarized showing if the correct identification of the bee was possible with the barcode sequence, if the barcode produced a *Wolbachia*-hit, or if it was impossible to use the sequence due to a mixed signal.

Results And Discussion

PCR and sequencing results

We were able to produce PCR amplification products for all 25 bee individuals included in this study, representing six of the seven described families of the taxon. We focussed mainly on members of Andrenidae and Halictidae, as they had shown before to be more difficult to be barcoded (see above). After sequencing, we found seven sequences with a mixed signal, while 43 sequences could be used for identification using the BOLD database. Sequences with mixed signals were all amplicons from the

LCO1490/HCO2198 primer pair, and in four cases stem from individuals belonging to the *Andrena ovatula*-group.

Taxonomic assignment

Altogether 43 sequences were identified using the BOLD database. In the case of 30 barcodes from 15 individuals both primer pairs produced the same result and the corresponding taxonomic assignment of the bee species was unambiguous. In the case of 6 individuals, the sequence from the LCO1490/HCO2198 primer pair was taxonomically assigned to originate from a *Wolbachia* endosymbiont, while the corresponding barcodes from the BeeCox1F1/BeeCox1R2 primer pair allowed an unambiguous taxonomic assignment to a bee species. The same is true for the seven barcodes which could only be analysed for the BeeCox1F1/BeeCox1R2 primer pair, as the LCO1490/HCO2198 primer pair resulted in a mixed signal of the sequenced barcode.

Conclusion

By comparing the location of the standard barcoding primer pairs in the multiple sequence alignment of full *cox1*-sequences of bees and recognizing its similarity with the corresponding *Wolbachia* gene sequences, we designed a new primer pair circumventing these matches (Fig. 1). We were able to produce PCR amplification products for the new primer pair and the standard Folmer-primers for all 25 bee individuals included in our small test sample set, representing 6 of the seven higher taxa (“families”) recognized within Anthophila. Even among this small sample size we found 6 cases where the Folmer-primers amplified *Wolbachia* instead of the host cox-1 gene and 7 cases of mixed signal, which might also stem from simultaneous amplification of host and endosymbiont DNA. In contrast, with the newly designed BeeCox1F1/ BeeCox1R2-primer pair we were in all cases able to generate clear sequences which could be identified as the corresponding bee species in the BOLD database. As such the newly proposed primer pair should help to reduce problems when barcoding bees, which will be especially important for morphologically difficult to identify species, such as from the *Andrena ovatula* group.

Limitations

The new barcoding primer pair has been tested for a phylogenetically representative, but small taxon sampling of bees. It is also unclear how they will perform when barcoding related taxa of Hymenoptera. Whereas we are convinced that this barcoding primer pair will circumvent problems related to mis-amplification of endosymbionts of the genus Wolbachia, possible amplification of nuclear integrations of mitochondrial sequences (NUMTs) remains an issue for DNA barcoding [19].

Abbreviations

BOLD: Barcode of Life Data System

PCR: Polymerase Chain Reaction

Declarations

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

Original chromatograms and sequence data for all barcoded bees in this study, as well as the multiple alignment for primer design can be found at https://github.com/Animal-Evolution-and-Biodiversity/Design_Barcoding_Primer_Bees.

Funding

Not applicable.

Acknowledgements

We would like to thank Bertram Preuschhof from the “Team Natur- und Bodenschutz” of the Landkreis Göttingen for issuing a permit to sample wild bees in the FFH-protected area “Ballertasche”.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

CB carried out study design, sampling of specimens and data analyses. KH performed all molecular laboratory work. Both authors read and approved the final manuscript.

References

1. Orr MC, Hughes AC, Chesters D, Pickering J, Zhu C-D, Ascher JS: **Global patterns and drivers of bee distribution.** *Current Biology* 2021, **31:**451-458.e454.
2. Kremen C, Williams NM, Aizen MA, Gemmill-Herren B, LeBuhn G, Minckley R, Packer L, Potts SG, Roulston Ta, Steffan-Dewenter I, et al: **Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change.** *Ecology Letters* 2007, **10:**299-314.

3. Zattara EE, Aizen MA: **Worldwide occurrence records suggest a global decline in bee species richness.** *One Earth* 2021, **4:**114-123.
4. Buschbacher K, Ahrens D, Espeland M, Steinhage V: **Image-based species identification of wild bees using convolutional neural networks.** *Ecological Informatics* 2020, **55:**101017.
5. Godfray HCJ, Knapp S, Gaston KJ, O'Neill MA: **Automated species identification: why not?** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359:**655-667.
6. Hebert PDN, Cywinska A, Ball SL, deWaard JR: **Biological identifications through DNA barcodes.** *Proceedings of the Royal Society of London Series B: Biological Sciences* 2003, **270:**313-321.
7. Schmidt S, Schmid-Egger C, Morinière J, Haszprunar G, Hebert PDN: **DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim).** *Molecular Ecology Resources* 2015, **15:**985-1000.
8. Magnacca KN, Brown MJF: **DNA barcoding a regional fauna: Irish solitary bees.** *Molecular Ecology Resources* 2012, **12:**990-998.
9. Sheffield CS, Heron J, Gibbs J, Onuferko TM, Oram R, Best L, deSilva N, Dumesh S, Pindar A, Rowe G: **Contribution of DNA barcoding to the study of the bees (Hymenoptera: Apoidea) of Canada: progress to date.** *The Canadian Entomologist* 2017, **149:**736-754.
10. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek RC: **DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.** *Molecular Marine Biology and Biotechnology* 1994, **3:**294-299.
11. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101:**14812-14817.
12. Villalta I, Ledet R, Baude M, Genoud D, Bouget C, Cornillon M, Moreau S, Courtial B, Lopez-Vaamonde C: **A DNA barcode-based survey of wild urban bees in the Loire Valley, France.** *Scientific Reports* 2021, **11:**4770.
13. Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, Gibbs J, Hajibabaei M, Hallwachs W, Hind K, et al: ***Wolbachia* and DNA barcoding insects: patterns, potential, and problems.** *PLOS ONE* 2012, **7:**e36514.
14. Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ: **The incidence of bacterial endosymbionts in terrestrial arthropods.** *Proceedings of the Royal Society B: Biological Sciences* 2015, **282:**20150249.
15. Gerth M, GeißLer A, Bleidorn C: ***Wolbachia* infections in bees (Anthophila) and possible implications for DNA barcoding.** *Systematics and Biodiversity* 2011, **9:**319-327.
16. Gerth M, Saeed A, White JA, Bleidorn C: **Extensive screen for bacterial endosymbionts reveals taxon-specific distribution patterns among bees (Hymenoptera, Anthophila).** *FEMS Microbiology Ecology* 2015, **91**.
17. Françoso E, Arias MC: **Cytochrome c oxidase I primers for corbiculate bees: DNA barcode and mini-barcode.** *Molecular Ecology Resources* 2013, **13:**844-850.

18. Katoh K, Rozewicki J, Yamada KD: **MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.** *Briefings in Bioinformatics* 2017, **20:**1160-1166.
19. Song H, Buhay JE, Whiting MF, Crandall KA: **Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified.** *Proceedings of the National Academy of Sciences* 2008, **105:**13486-13491.

Figures

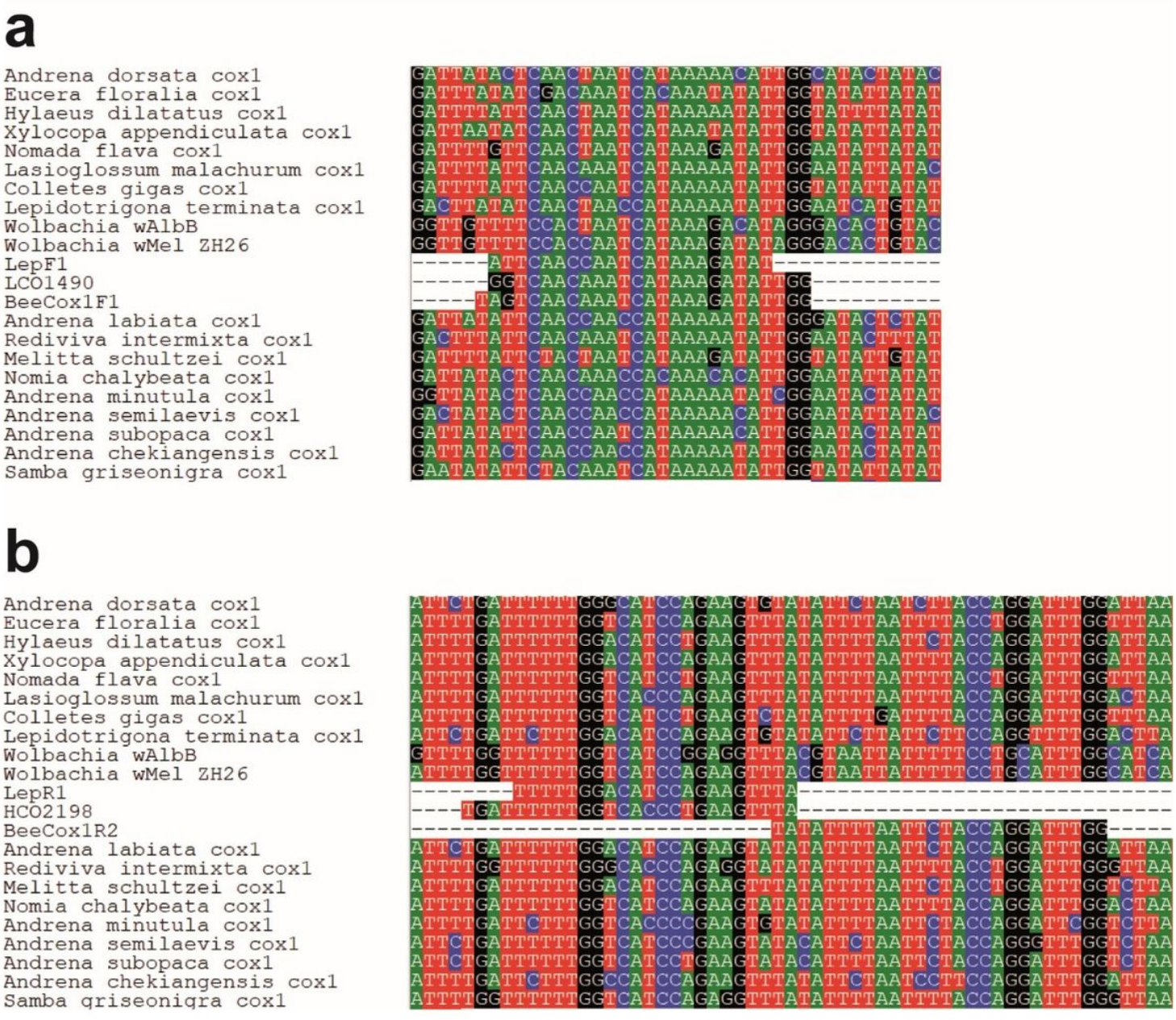

Figure 1

Multiple sequence alignment. Snap shot of the multiple alignment of bee cox1-sequences (and two sequences of the homologous Wolbachia gene) indicating the position of the forward (a) and reverse (b)

primer sequences of the standard (LCO1498/HCO2198, LepF1/LepR1) and new (BeeCox1F1/BeeCox1R2) primer pairs.