First DNA Barcodes of Arthropod Pests from Romania

Roxana Ciceoi¹, Adriana Radulovici²

¹Research Center for Studies of Food and Agricultural Products Quality HORTINVEST, University of Agronomic Sciences and Veterinary Medicine of Bucharest, 59 Marasti Blvd., District 1, Bucharest, Romania
²Centre for Biodiversity Genomics, 50 Stone Road E, Guelph, Canada

Corresponding author email: roxana.ciceoi@gmail.com

Abstract

DNA barcoding is a diagnostic method proposed by Paul Hebert and his team in 2003, using a short standardized genetic marker in an organism’s DNA to facilitate identification at a certain taxonomic level. Identification consists in finding the closest matching reference record in different databases. For arthropods, the mitochondrial cytochrome c oxidase I (COI) gene is used. The Barcode of Life Data System (BOLD) is the online facility created by Centre of Biodiversity Genomics as a freely available collaborative hub which supports the assembly and use of DNA barcode data. Currently ~6,650 k barcodes for specimens from 188 countries are available through the platform, of which 5,420 k represent arthropod specimens with barcodes. From Romania, 2,817 arthropod records are available, for 408 species, mainly butterflies (biodiversity data). Our present research made available the first DNA barcodes of arthropods plant pests from Romania, with emphasis on the invasive species. 85 insect specimens belonging to eight orders, 30 families, lead to 79 barcode compliant sequences. None of the barcoded species from Romania was previously recorded in BOLD, with the exception of one Autographa gamma specimen collected in 1980, deposited in the Smithsonian National Museum of Natural History.

Key words: DNA barcoding, arthropod pests, Romania, BOLD.

Introduction

In 2003, Paul Hebert and his team proposed the biological identification of species through DNA barcodes. Since then, millions of barcodes have been generated for hundreds of thousands of living species and are all freely available for comparisons and quicker identification for whoever might need, all around the world.

All these genetic data and additional metadata are stored and can be accessed through multiple databases such as BOLD (Barcode of Life Data Systems, boldsystems.org), the largest DNA barcodes database, GenBank, the largest molecular database (ncbi.nlm.nih.gov/genbank/), Q-bank (q-bank.eu/), the database on quarantine plant pests and diseases in the EPPO region, or even through taxon databases, such as Chironomid DNA Barcode database (nies.go.jp/yusurika/en/index.html), a database which provides information on dipteran species belonging to Chironomidae family and even country-dedicated databases, as GBOL for Germany (bolgermany.de/wp/en/).

The International Barcode of Life (iBOL; ibol.org) is a global research network aiming to serve and transform biodiversity science by building the DNA barcode reference libraries and everything else required cataloguing the entire Earth biodiversity. Many platforms dedicated to specific taxonomic groups were made available through iBOL initiatives, as mammaliabol.org, lepbarcoding.org, fishbol.org, formicidaebol.org, etc or as country inventorying databases, as finBOL.org for Finland, norbol.org for Norway etc., although today not all are still regularly updated.

DNA barcoding revolutionizes biodiversity research and taxonomical studies (Hebert and Gregory, 2005) by fast and accurate species identification, but new tools have been developed in the food control area, as in food traceability (Galimberti et al, 2013), food authenticity (Christiansen et al., 2018, Khaksar et al., 2015) and food security (Raclariu-Manolica et al., 2019). Onyia et al., 2014, mentions that DNA barcoding is used around the world to assist in species identification in all its life stages, but has some specific uses as
tool for: convicting illegal traffickers, proving illegal use of endangered species, identifying Branta canadensis, the Canada goose, as the bird species responsible for the US Airways Flight 1549 crash in 2009, rediscovering lost species as the predatory water beetle Graphoderus bilineatus believed extinct for 26 years, early warning and monitoring of pests from different crops, taxonomic tool, in order to discern species that have been misclassified into other groups, public health tool, bioassessment in environmental monitoring, protecting consumers from mislabelling.

Some important findings proved the utility of DNA barcoding method; while analysing seafood in Belgium, researchers discovered that Bluefin tuna fish was substituted by other tuna species in 95% of the cases (Christiansen et al., 2018) while others unmasked seafood mislabelling in U.S. markets and named DNA barcoding as a unique technology for food authentication and quality control (Khaksar et al., 2015).

DNA barcoding became in the last decade a trusted tool also for plant protection staff. Morphological identification might become a challenge for people without a strong background in taxonomy. Molecular methods as tools to identify unknown specimens, especially when plant health personnel must quickly identify different development stages of a species, become more and more convenient, cost-effective and reliable. For quarantine pests or in areas where some pests were not previously present, molecular techniques are imperative.

The EPPO standard PM 7/129 (1) describes the use of DNA barcoding protocols to identify regulated pests and invasive plant species of importance to the EPPO area (Europe and the Mediterranean Region) and details all the steps required for molecular and analytical processing in order to arrive to a correct species identification.

In Romania, several steps have been performed in the DNA barcoding area, but almost all focus on biodiversity or food security. Romania is represented in iBOL by “Stejarul” Research Centre for Biological Sciences, a branch of the Romanian National Institute of Research and Development for Biological Sciences (NIRDBS), who is mainly interested in applying DNA barcoding to understand the “anthropogenic drivers influencing the biodiversity loss, focusing on protected areas and vulnerable environments” (iBOL, 2018).

The first molecular identification study done in Romania was performed by Nicolescu et al., 2004, in order to prove the existence of a new species of mosquitoes of the Anopheles maculipennis group (Diptera: Culicidae) in Romania.

In the biodiversity area, Dincă et al., (2011) made a complete DNA barcode reference library for the Romanian butterfly fauna, by analyzing 180 species of butterflies, which represent almost one third of the European butterfly fauna. The same author used DNA barcoding as a method to discriminate the European endemic butterfly Erebia oeme (Hübner, 1804) (Lepidoptera: Nymphalidae), considered extinct, from the woodland ringlet, Erebia medusa and demonstrated its presence in Retezat mountains (Dincă et al., 2013). Another study was published in 2012, by a Lithuanian researcher, mentioning a DNA barcode for Trifurcula (Glaucolepis) lituanaica sp. nov., (Lepidoptera: Nepticulidae), a new stem-miner collected also on Salvia pratensis from Romania (Ivinskis et al, 2012).

In the area of parasitology, DNA barcoding was used in a study about a parasite mite, Knemidocoptes jamaicensis, parasitising the Common Chaffinch, Fringilla coelebs (Dabert et al., 2013).

The DNA barcoding method started to be used recently by the Grigore Antipa National Museum of Natural History, to identify different specimen and for phylogeography studies (Popa, 2017). In 2018, a team used DNA barcoding to confirm the morphological identification of a native to the Far East Pacific oyster species, named Crassostrea gigas, which was found in the Black Sea at the Romanian littoral (Buhaciuc-Ionita et al., 2018). Lastly, in 2019, the museum organized a DNA day, an open day for public outreach in which the museum specialists explained to young people how molecular biology techniques are used to solve problems related to the taxonomic status of relict species, about the conservation of endangered species, the genetics of invasive species and DNA-
barcoding (DNA barcodes) of some alien species from the fauna of Romania (Antipa, 2019).
The present paper illustrates one of the first attempts in using DNA barcoding in the area of plant protection in Romania, by gathering DNA barcodes of arthropod pests of economic importance in our country.

**MATERIALS AND METHODS**

The 95 arthropod specimens were mainly collected in the period 2017-2018 from Bucharest area, from the experimental fields and the greenhouse of the Research Center for Study of Food and Agricultural Products Quality, University of Agronomic Sciences and Veterinary Medicine of Bucharest (44.38 N, 26.15 E). Xx specimens were collected during a biodiversity survey in Iligani de Jos (44.38 N, 26.15 E) area and Rosu (44.38 N, 26.15 E) village.

Two specimens of *Tuta absoluta* were collected in 2014 by a citizen-scientist, around Bucharest and kept dry in plastic tubes, which made them difficult to process. All insects were stored in 96% ethanol prior to DNA extraction.

All DNA barcoding steps were performed at the Centre for Biodiversity Genomics, University of Guelph, Canada (biodiversitygenomics.net) and followed the standard protocols (ccdb.ca/resources/).

Molecular processing used 96-well microplates. For tissue subsampling, one posterior leg of each specimen was used.

DNA was extracted using a manual silica-based protocol with glass fiber filtration plates, described by Ivanova et al. (2007). The 658-bp barcode region of the mitochondrial cytochrome c oxidase subunit 1 (COI) gene (the standard barcode marker in animals) was amplified using a primer cocktail (C_LepFoIF / C_LepFoIR, 1:1), described by Hernández-Triana et al. (2014).

The polymerase chain reaction was set in 12.5 μL total volume consisting of 10.5 μL PCR mix (6.25 μL 10% trehalose, 2 μL ddH2O, 1.25 μL 10× PCR buffer, 0.625 μL MgCl2 (50 mM), 0.125 μL of each primer (10 μm), 0.0625 μL dNTPs (10 mM), 0.06 μL Platinum Taq polymerase (5 U/μL)) and 2 μL DNA template. The amplification was performed in an Eppendorf Mastercycler Nexux Gradient machine, with the following thermocycling program: initial denaturation at 94°C for 1 min, 5 cycles of 94°C for 40 s, 45°C for 40 s and 72°C for 1 min, followed by 35 cycles of 94°C for 40 s, 51°C for 40 s and 72°C for 1 min, and 1 cycle at 72°C for 5 min.

Confirmation of PCR amplification was done by electrophoresis, the PCR products being visualized on pre-cast 96-well agarose E-Gels (Invitrogen). Bidirectional sequencing was performed on an ABI 3730XL DNA analyser. COI sequences were edited with CodonCode Aligner (www.codoncode.com) and manually inspected for ambiguities.

The aligned sequences were translated into aminoacids to check for stop codons (an indication of potential pseudogene amplification) in MEGA7.0 (ref?). Genetic divergences based on Kimura-2-parameter (K2P) were also calculated in MEGA 7.0 (MegaSoft, 2019).

Three molecular databases were used for sequence comparison, namely: BOLD), GenBank and Q-bank. All data was stored and additionally analysed in BOLD mainly by using the built-in neighbour-joining tree option.

**RESULTS AND DISCUSSIONS**

Out of the 95 sequenced specimens, 86 COI-5P sequences were obtained, 79 being barcode compliant, which mean not only they led to high quality electropherograms, but also fulfilled all criteria of BOLD database, as a minimum sequence length of 500bp, less than 1% ambiguous bases, the presence of two trace files, the presence of a country specification in the record as set out by the Consortium for DNA Barcoding (BoldSystem, 2013).

The DNA barcoded specimens belong to two classes, Insecta (94) and Arachnida (1), eight orders, Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Neuroptera, Orthoptera and Trombidiformes and 30 families (figure 1).

Forty-six arthropod species were barcoded (figure 2).

All details regarding taxonomy, vouchers, collection data together with images, DNA sequences and electropherograms can be found on BOLD, in the project – “Insects of..."
economic importance from Romania” (project code: MCCRG). The confirmation of PCR amplification by electrophoresis, on pre-cast 96-well agarose E-Gels was performed before preparation for sequencing. For nine specimens no sequences were obtained. Despite the fact that the PCR visualisation by electrophoresis confirmed the lack of DNA only in three of the wells (for the eggs of Halyomorpha halys, for one specimen of Tuta absoluta and one of Agriotes sp.), in the other 6 wells, DNA was present, but migrated in unclear bands. Another six specimens led to sequences having less than 658 bp (Harmonia axyridis, Tropinota hirta, Ceratitis capitata, 2 x Palpita vitrealis, Halyomorpha halys (larval stage 1).

Out of the other 80 DNA sequences with equal length of 658 bp, four were low quality and displaying 19-29 ambiguous nucleotide, belonging to Ceratitis capitata (3x) and Chrysoperla carnea (1x) species. No stop codons were observed upon translation into aminoacids.

Out of the 85 sequences, 60 sequences (70.58%) led to species identification based on 100% similarity and 18 sequences allowed species identification with more than 99.5% similarity. The 7 sequences that led to matches inferior to 99% were considered species identified only at genus level.

Figure 1. The distribution of arthropod pests specimens by family.
According to the percentage of similarity with the 1st match, when performing the species identification, 57 barcodes matched 100% the BOLD full database (68.23%) and 18 barcodes matched the full DB in percentages between 99.36 and 99.85% (21.17%), indicating a very likely level of certainty for taxon identification. The rest of 9 barcodes could not lead to a conclusion regarding the species identification using BOLD, not even at the genus level.

The largest insect family analysed was Pentatomidae, with 18 insect specimens and one trial from Halyomorpha halys eggs shells. No Pentatomidae records from Romania were previously existing in BOLD database. While blasting with all the other records in BOLD, out of the 10 specimens of H. halys nine samples had DNA barcodes that matched 100% BOLD full database and the number of overlap bases was higher than 599. The brown marmorated stink bug eggs shells led to no sequence. H. halys had at the time of study 2,246 specimens records, of which 2,181 (97.1%) had compliant barcodes.

Out of the three specimens of Dolycoris baccarum, the larva (stored in freezer for 6 months) had seven matches of 100% similarity and 543 bp, the adult stored in freezer for six months had no 100% similarity (99.82% being the highest and 544 bp overlap) while the adult specimen kept dry and mixed together with the other insects led to 3 matches of 100% and 549 bp overlap. In BOLD, out of 91 specimens of D. baccarum, 82 (90%) had compliant barcodes.

For Aelia acuminata, 15 records matched 100% the BOLD records, with 600 bp overlap. 92 specimens (88.4%) had compliant barcodes. The second Aelia sp. specimen could not be confirmed by blast either in BOLD or NCBI databases, even at genus level, as the highest similarity was 88.53% and matches correspond to different genus. This specimen was collected from Tulcea, Iliganii de Sus, in September 2017. Further research is needed to identify this specimen (figure 3).
One red specimen of *Nezara viridula* was confirmed, with 19 matches of 100% similarity. Two *Graphosoma italicum* could not be confirmed as species using BOLD blast, as similarities of 100% were shown both for *G. italicum* and *G. lineatum*. The taxonomy of the two species is uncertain, as *G. italicum* had been regarded as either a subspecies, or a synonym to *G. lineatum* (Wiki, 2019) and its status has been, and still remains, controversial (Lupoli, 2017). Recent efforts have recently been made to establish the validity of *G. italicum* by Lupoli, using DNA barcoding and, according to the author, the species should be considered valid. According to the same author, *G. lineatum*’s distribution is limited to North Africa and Sicily while *G. italicum* is present all over Europe and the Middle East (Lupoli, 2017). This particular case demonstrates how DNA barcoding can contribute to insect taxonomy, helping to elucidate centuries of controversial debates. All 56 specimens recorded in BOLD have compliant barcodes, 600 bp overlap, leading to the conclusion that the actual issue is taxonomy.

One specimen previously identified as *Nezara viridula* showed no matches in BOLD or NCBI. Morphological identification showed the species to be *Acrosternum heegeri* (figure 4) (D. Rédei, 2018, pers. comm.), a species that is currently expanding in Europe (Károlyi and Rédei, 2017). This species has currently (August 2019) no specimen or barcode uploaded in BOLD and might be considered a new contribution, after additional analysis.

The Coccinellidae family was the second largest family to be analysed, with specimens belonging to three species: *Coccinella septempunctata*, *Harmonia axyridis* and *Tythaspis sedecimpunctata* (sin. *Coccinella sedecimpunctata* Linnaeus, 1758). All 5 specimens were confirmed as very likely to be the respective taxons, based on 100% similarity and 600 bp overlap. The 947 barcodes for *H. axyridis* and 557 barcodes for *C. septempunctata* show the importance of DNA barcoding in studies concerning the invasive species and their impact on biodiversity. As a remark, the identification tree also reveals how easily it is to identify errors using this technique - in a single branch tree of *H. axyridis*, the *Chilocorus renipustulatus* record indicates a misidentification (figure 5).
The Gelechiidae family was the third largest group analysed. Six tomato leafminer specimens were analysed, of which four captured by glue pheromone trap in March 2018 and transferred into 96% ethanol, with their body partially damaged by the contact with the glue and two specimens collected by a citizen-scientist in 2014 and kept dry in plastic tubes, no ethanol, for 4 years (Ciceoi and Radulovici, 2018). Although difficult to process, the DNA sequences of these specimens were identical, with equal length (658bp) and no ambiguous nucleotide, proving the versatility of DNA barcoding method in insect identification even when the morphological identification could not be reliable.

From Tephritidae family, four Ceratitis capitata specimens were barcoded. Although all electropherograms were medium and low quality and barcodes not compliant, as they had between 9 and 23 ambiguous bases, the taxon identification was possible, with very likely certainty, at 99.83% match and 575 bp overlap. This fact proves DNA barcoding may work for rapid identification even with non-specific primers, as C_LepFolF/ C_LepFolR. In previous studies amplification for C. capitata failed in some specimens therefore different primer sets were developed based on the full mitochondrial genomes of C. capitata (AJ242872) obtained from GenBank (Smit et al., 2013). The Romanian barcodes clustered with specimens from Mexico and Peru. Detailed research with specific primers is needed to confirm that such information may be reliable to establish the introduction pathway (figure 6).

In BOLD, 265 barcodes are available for C. capitata. DNA barcoding was previously mentioned as difficult for Tephritidae family, due to incomplete reference libraries, (Virgilio et al., 2012). As the pressure of the invasive species increases exponentially with the intensification of global trade, DNA barcoding may also serve as a tool for pest risk assessors. Considering the new Plant Health Regulation (EU) 2016/20311, on the protective measures against pests of plants, which will be enforced starting with December 2019, special attention should be given to commodity risk assessments (EFSA PLH Panel, 2019), introduction pathways understanding being of high importance in prevention.

From Crambidae family, four specimens belonging to Cydalima perspectalis, Palpita vitrealis and Ostrinia nubilalis species were barcoded. Cydalima perspectalis and Ostrinia nubilalis specimens were both confirmed with 100% similarity, for a 600 bp sequence. Palpita vitrealis’s sequences led to a false ambiguous result, as both 99.84% similarities were obtained for P. vitrealis and P. unionalis names. Further taxonomical research revealed a disagreement concerning the name of this species, Fauna Europaea mentioning P.
vitrealis as the single species inside the *Palpita* genus.

For *Corythucha arcuata* (Hemiptera: Tingidae), four specimens were analysed, 2 trapped in 2018, in the greenhouse, with yellow sticky traps, (bad preserved insect bodies) and 2 kept dry for 5 months, collected from outdoors. Different tissue sampling was used, in order to determine the method versatility: entire body, only two legs, only abdomen (the one presumed to have the chance to bite humans) and whole body (Ciceoi and Radulovici, 2018b). High quality traces were obtained for three out of four specimens (legs and entire body), while the sequencing failed for the abdomen tissue. Analysing the identification tree, the Romanian barcodes clustered with *C. pallipes*, a species known to occur only in North America. Further studies are needed (figure 7).

![Identification Tree](image)

Figure 7. The identification tree for the *Corythucha* sp. barcodes present BOLD.

The specimens belonging to the families Chrysomelidae (*Chrysolina polita*, *Chrysolina graminis* and *Cassida viridis*), Scarabaeidae (*Tropinota hirta*, *Anisoplia agricola*, *Oxythyrea funesta*), Tabanidae (3x *Chrysops* sp.), Flatidae (3x *Metcalfa pruinosa*), Cicadellidae (*Cicadella viridis*, 2x *Scaphoideus cyrius*), Tortricidae (3x *Cydia pomonella*), and Pyralidae (3x *Plodia interpunctella*) were each represent by three specimens and the majority of them was confirmed with 100% similarity. The exception is represented by the 3 specimens of *Chrysops* sp. (figure 8), where the similarities of 94% indicate that actually this specimen cannot be identified using the obtained sequences and, most probably, this species is not yet present in BOLD, NCBI or Q-bank.

The specimens belonging to the families Curculionidae (*Otiorhynchus armadillo*, *Larinus planus*), Cerambycidae (*Semanotus* sp., *Aegosoma* sp.) Phoridae (2x *Megaselia scalaris*), Aphididae (2x *Cinara tujafilina*), Aleyrodidae (2x *Trialeurodes vaporariorum*), Noctuidae (2x *Autographa gamma*), Chrysopidae (*Chrysoperla carnea*, *Chrysoperla sp.*) and Gryllotalpidae (2x *Gryllotalpa stepposa*) were identified with 100% similarity, except the two Cerambycidae specimens, belonging to *Semanotus* (figure 9) and *Aegosoma* genus, with 89.91% and respectively 98.15 % similarity.

The specimens belonging to the families Cixiidae (*Reptalus quinquecostatus*), Coreidae (*Coreus marginatus*), Membracidae (*Ceresa bubalus*), Pyrrhocoridae (*Pyrrhocoris apterus*), Rhopalidae (*Myrmus miriformis*), Sphecidae (*Sceliphron caementarium*), Vespidae (*Vespula germanica*), Gracillariidae (*Phyllonorycter platani*) were all identified with similarities higher than 99.54%.

Figure 8. One specimen of *Chrysops* sp. which could not be identified.
The only non-insect arthropod that was analysed was a specimen of Panonychus citri, identified with 100% similarity.

CONCLUSIONS

DNA barcoding proved to be a very useful tool in confirming or identifying some of the main invasive pests from Romania. Besides the 86 barcodes of pest of economic importance from Romania that are now available for future studies in BOLD, new uses of DNA barcoding and future research studies have been foreseen.

ACKNOWLEDGEMENTS

This research was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, project number PN-III-P1-1.1-MC-2017-2338, within PNCDI III.

REFERENCES


Ivinskis, P., Van Nieukerken, E.J., Rimsaite, J. (2012). Trifurcula (Glaucolepis) lituanica sp. nov., an unexpected new stem-miner on Salvia pratensis
occurring in eastern Europe (Lepidoptera: Nepticulidae). *Zootaxa* 3570, 41–55.


